Classification of Voting Patterns to Improve the Generalized Hough Transform for Epiphyses Localization

Ferdinand Hahmann[a], Gordon Böer[a], Eric Gabriel[a], Thomas M. Deserno[b], Carsten Meyer[a,c], and Hauke Schramm[a,c]

[a]Kiel University of Applied Sciences, Germany
[b]RWTH Aachen University, Germany
[c]Christian-Albrechts University Kiel, Germany

ABSTRACT

This paper presents a general framework for object localization in medical (and non-medical) images. In particular, we focus on objects of well-defined shape, like epiphyseal regions in hand-radiographs, which are localized based on a voting framework using the Generalized Hough Transform (GHT). We suggest to combine the GHT voting with a classifier which rates the voting characteristics of the GHT model at individual Hough cells. Specifically, a Random Forest Classifier rates whether the model points, voting for an object position, constitute a regular shape or not, and this measure is combined with the GHT votes. With this technique, we achieve a success rate of 99.4% for localizing 12 epiphyseal regions of interest in 412 hand- radiographs. The mean error is 6.6 pixels on images with a mean resolution of 1185×2006 pixels. Furthermore, we analyze the influence of the radius of the local neighborhood which is considered in analyzing the voting characteristics of a Hough cell.

Keywords: Epiphyses Localization, Localization, Detection, Discriminative Generalized Hough Transform, DGHT, GHT, Shape Consistency Measure, Bone Age Assessment

1. INTRODUCTION

Bone Age Assessment (BAA) is an important method in diagnostic radiology which is used for evaluating the skeletal maturity in order to diagnose growth disorders in children and adolescents. Since manual BAA techniques (e.g. Tanner & Whitehouse (TW)[1]) are time consuming, subjective, and require expert knowledge from a physician a number of automatic methods have been developed in recent years. Many of these approaches follow the basic concept of TW, classifying only certain extracts from the radiograph, as this substantially reduces the complexity of the classification problem. An important prerequisite for BAA is the availability of a reliable and robust object detection method to enable the extraction of the required region-of-interest (ROI). This task can be solved by individually adjusted methods with heavy usage of expert knowledge about the searched objects. For instance Hsieh *et al.*[2] and Pietka *et al.*[3] analyze the image columns and search for bright lines representing the bones of the fingers (phalanges). More general solutions are presented in Thodberg *et al.*,[4] which uses active appearance models for bone reconstruction, and Fischer *et al.*,[5] where a graph-based structural prototype, representing the phalanges and metacarpal bones, is registered to the image.

In this paper, we employe a general object detection framework, using the idea of the Generalized Hough Transform[6] (GHT). The concept is to model an object by feature points, which describe different object parts, e.g. based on edge or salient point detection or image patch classification, in relation to the target landmark. To analyze an image with a given model, the utilized features are extracted and vote with corresponding model points for hypothetical target point locations in a transformation parameter space called Hough space.

A voting-based procedure is also the basis of some current, robust object detection methods, like Hough Forests.[7] The basic idea is to use a Random Forest to determine the displacement vector from an image patch in order to obtain the target landmark. The Random Forest is generated based on image patch / displacement

Further author information: (Send correspondence to Ferdinand Hahmann)
Ferdinand Hahmann: Ferdinand.Hahmann@fh-kiel.de, Telephone: +49 (0)431 210 4143
Hauke Schramm: Hauke.Schramm@fh-kiel.de, Telephone: +49 (0)431 210 4140

vector pairs learned from training data, such that similar image patches, describing comparable object parts, yield similar displacement vectors. This idea is often used for various tasks of medical image processing,[8,9] including the epiphyses detection.[10,11]

The Discriminative Generalized Hough Transform (DGHT), used in this paper, is also based on the GHT. Whereas the Hough Forests use a quite complex feature extraction algorithm, namely the image patch extraction, the DGHT has up to now only been explored using simple (e.g. edge) features but with a sophisticated discriminative training procedure for the model generation.

Voting-based approaches have, in general, the drawback of an independent treatment of individual model points. This may pose a problem if the object to localize appears in different variants, e.g. the epiphyseal distance on left hand-radiographs resulting from different patient ages. If these variants are used to generate a single voting model, image patches or edges, representing different object variants, could incidentally vote for the same Hough cell, which may result in false-positive object localization. To handle this drawback, the original GHT implementation applies linear model transformations and Hough space splitting to address target object variability rather than using a model containing different variants.

Another possible solution, which can also be applied for non-linear model transformations, is to train separate models for the observed object variants. This can be achieved by grouping the training images into variation classes, e.g. by manually defined criteria, such as head pose,[12] size[13] or depth information,[14] but also automatic grouping procedures[15,16] have been presented in the past.

An alternative is to train possible object variants into a single model, and to use the pattern of model points voting for a particular Hough cell for further assessments such as segmentation,[17] comparison with training images,[18] or using a classifier for rating whether the pattern represents a coherent variability class or not.[19] Here, the GHT voting pattern for a localization hypothesis is analyzed by a Random Forest Classifier[20] to generate a Shape Consistency Measure (SCM) which is used to assign a score, representing the coherence of the model points voting for a Hough cell. The aim of this score is to assign low weights to Hough cells whose votes do not represent a coherent object variant. This prevents false positives originating from mutually exclusive object variants. In this paper, we apply our object localization technique consisting of DGHT and SCM to a medical image processing task, namely epiphysis localization in hand-radiographs. Furthermore, we analyze the influence of the size of the local neighborhood which is considered in the assessment of the voting pattern of a Hough cell on the localization performance.

2. METHOD

2.1 Generalized Hough Transform

The Generalized Hough Transform (GHT), introduced by Ballard in 1981,[6] is a general and well-known model-based approach for object localization, which belongs to the category of template-matching techniques. Each model point $\mathbf{m}_j$ is represented by a displacement vector to the reference point.

The GHT transforms a feature image, in our case an edge 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 directly increases the computational complexity of the algorithm, we restrict the model transformation to a simple translation in this work. Moderate object variability with respect to shape, size, and rotation is not explicitly parameterized, but implicitly learned into the model by appropriately placing model points as indicated by the training data.

The voting procedure, which transforms a feature image $\mathcal{X}_n$ into the Hough space $\mathcal{H}$ (with discrete elements $\mathbf{c}_i$) by using the shape model $\mathcal{M}$, can be described by

$$\mathcal{H}(\mathbf{c}_i, \mathcal{M}, \mathcal{X}_n) = \sum_{j=1}^{|\mathcal{M}|} f_j(\mathbf{c}_i, \mathcal{X}_n) \tag{1}$$

with*

$$f_j(\mathbf{c}_i, \mathcal{X}_n) = \sum_{\forall \mathbf{e}_l \in \mathcal{X}_n} \begin{cases} 1, & \text{if } \mathbf{c}_i = \lfloor (\mathbf{e}_l - \mathbf{m_j})/\varrho \rfloor \text{ and } |\phi_{e_l} - \phi_{m_j}| < \Delta\varphi. \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

The quantized Hough space $\mathcal{H}$ (with quantization parameter ϱ) consists of Hough cells $\mathbf{c}_i$ that accumulate the number of matching pairs of all model points $\mathbf{m}_j$ and feature points $\mathbf{e}_l$. Each Hough cell $\mathbf{c}_i$ represents a target hypothesis whose coordinates in image space are given by $\lfloor (\mathbf{c}_i + 0.5) \cdot \varrho \rfloor$.

$f_j(\mathbf{c}_i, \mathcal{X}_n)$ determines how often model point $\mathbf{m}_j$ votes for Hough cell $\mathbf{c}_i$ for the given feature image $\mathcal{X}_n$. However, a voting is only possible, if the orientation of the model and feature point, ϕ_{m_j} and ϕ_{e_l}, respectively, has a small difference of below $\Delta\varphi$. Then, the displacement vector of model point $\mathbf{m}_j$ is subtracted from the edge point coordinates $\mathbf{e}_l$, and the resulting vector points to a potential location of the target point which generates a vote for the corresponding Hough cell $\mathbf{c}_i$ (in units of the quantization parameter ϱ).

The most likely target point location results from the Hough cell $\tilde{\mathbf{c}}_n$ with the highest number of votes, corresponding to the best match between the model $\mathcal{M}$ and the feature image $\mathcal{X}_n$:

$$\tilde{\mathbf{c}}_n = \arg\max_{\mathbf{c}_i} \mathcal{H}(\mathbf{c}_i, \mathcal{M}, \mathcal{X}_n) \tag{3}$$

2.2 Discriminative Generalized Hough Transform

The Discriminative Generalized Hough Transform (DGHT) extends the Generalized Hough Transform by discriminatively trained model point specific weights λ_j that reflect their importance for a correct localization and discrimination from confusable objects. These weights, which may also be negative, are incorporated as follows into the voting procedure (1):

$$\mathcal{H}(\mathbf{c}_i, \mathcal{M}, \mathcal{X}_n) = \sum_{j=1}^{|\mathcal{M}|} \lambda_j f_j(\mathbf{c}_i, \mathcal{X}_n) \tag{4}$$

The optimization of λ_j is based on a Minimum Classification Error training,[21,22] which aims at minimizing the sum of localization errors over all training images.

In GHT-based approaches, the quality of the localization highly depends on the quality of the model. A good model has to fulfill two important conditions: A high correlation with the feature image on the target point location and a small correlation with confusable objects. In the DGHT, this is achieved by an iterative training procedure. It starts with an initial model that is generated by superimposing annotated feature images at the reference point. The model point weights λ_j are optimized using a Minimum Classification Error approach, and model points with a low absolute weight are eliminated. At last, the model is extended by target structures from training images which still have a high localization error. This procedure is repeated until all training images are used or have a low localization error. A more detailed description of the technique can be found in Ruppertshofen.[22]

Although this optimization reduces the model fuzziness by focusing on comparably few key model points, the technique can only be applied for target objects with limited variability since the final DGHT model contains model points from different, and possibly mutually exclusive, variants and a common voting of these model points can still occur.

2.3 Shape Consistency Measure (SCM)

In order to handle the problem of false-positives that are induced by unlikely model point combinations, the Hough votes of a localization hypothesis are analyzed to determine whether they stem from a coherent object variant, observed in the training data. To this end, a Shape Consistency Measure is introduced that assesses the GHT voting pattern, described as $F(\mathbf{c}_i, \mathcal{X}_n) = \{f_1(\mathbf{c}_i, \mathcal{X}_n), f_2(\mathbf{c}_i, \mathcal{X}_n), ..., f_{|\mathcal{M}|}(\mathbf{c}_i, \mathcal{X}_n)\}$, as an expected (regular) or irregular (non-coherent) pattern.

* $\lfloor \mathbf{a} \rfloor$ denotes the floor of each component of $\mathbf{a}$.

Since each individual localization hypothesis contains votes from a comparably small set of model points, the voting pattern in each cell may be partly coincidental.[19] To overcome this issue, it is reasonable to consider the common voting behavior of model points for a Hough cell and its neighborhood within a certain distance. However, in case of a too large neighborhood, the set of model points loses its explanatory power. Thus, treating all model points voting in a fixed neighborhood, its size would be an important parameter. To avoid that, we define a feature function, which captures the closest distance of a vote of model point $\mathbf{m}_j$ in a given neighborhood area of cell $\mathbf{c}_i$ as

$$r_j(\mathbf{c}_i, \mathcal{X}_n) = \min_{\mathbf{c}_k} \begin{cases} d(\mathbf{c}_i, \mathbf{c}_k), & \text{if } f_j(\mathbf{c}_k, \mathcal{X}_n) \geq 1 \text{ and } d(\mathbf{c}_i, \mathbf{c}_k) \leq \vartheta \\ \vartheta + 1, & \text{otherwise.} \end{cases} \tag{5}$$

with $d(\mathbf{a}, \mathbf{b}) = max_t|a_t - b_t|$. Thus, a value $\alpha = r_j(\mathbf{c}_i, \mathcal{X}_n) \leq \vartheta$ specifies the minimum neighborhood of $(2\alpha + 1) \times (2\alpha + 1)$ around $\mathbf{c}_i$ in which the model point $\mathbf{m}_j$ has voted. In r_j we also introduce a new parameter ϑ. But as long as $\vartheta > 4$ the exact choice is only relevant for runtime performance, as we will show in Section 5. With the feature function $r_j(\mathbf{c}_i, \mathcal{X}_n)$ the GHT voting pattern is extended by

$$R(\mathbf{c}_i, \mathcal{X}_n) = \{r_1(\mathbf{c}_i, \mathcal{X}_n), r_2(\mathbf{c}_i, \mathcal{X}_n), ..., r_J(\mathbf{c}_i, \mathcal{X}_n)\} \tag{6}$$

containing information about the voting behavior in cell $\mathbf{c}_i$ and its neighborhood.

The vector $R(\mathbf{c}_i, \mathcal{X}_n)$ is used as feature vector to discriminate two classes: Class Ω_r comprises feature vectors describing "regular" voting patterns belonging to true object positions, class Ω_i contains feature vectors originating from "irregular (non-coherent)" voting patterns, expected at false positive locations. However, instead of a pure two-class classification, the Random Forest Classifier determines the posterior probability $p(\Omega_r|R(\mathbf{c}_i, \mathcal{X}_n))$ to belong to class Ω_r, given the attribute vector $R(\mathbf{c}_i, \mathcal{X}_n)$. We use this probability as an additional factor to weight the votes in Hough space, such that Hough votes from irregular voting patterns are downweighted by a small posterior probability $p(\Omega_r|R(\mathbf{c}_i, \mathcal{X}_n))$:

$$\tilde{\mathbf{c}}_n = \arg\max_{\mathbf{c}_i} p(\Omega_r|R(\mathbf{c}_i, \mathcal{X}_n)) \cdot \mathcal{H}(\mathbf{c}_i, \mathcal{M}, \mathcal{X}_n). \tag{7}$$

In summary, during the training of the SCM, a previously generated DGHT model is applied to all training images. For each image, we select the 50 hypotheses with the highest votes, which will be used as training samples for the Random Forest Classifier. For these hypotheses, the feature vector $R(\mathbf{c}_i, \mathcal{X}_n)$ is generated according to Equation (6). To determine the class label, we use the Euclidean distance $\varepsilon(\mathbf{c}_i, \hat{\mathbf{c}}_n) = \|\mathbf{c}_i, \hat{\mathbf{c}}_n\|_2$ between the hypothesis in question $\mathbf{c}_i$ and the ground truth localization $\hat{\mathbf{c}}_n$. Hypotheses with an error less or equal 3 Hough cells are regular structures and determined as class Ω_r wheres as hypotheses with an error larger than 10 Hough cells are from class Ω_i (irregular shape). Hypotheses with an error between 3 and 10 Hough cells are not considered during training to ensure a better discrimination between both classes. Then, the Random Forest Classifier is trained as described in Breiman,[20] to separate the two classes Ω_r and Ω_i (Figure 1).

For localizing the target object in an unknown image, at first the DGHT model is applied. Then for the 50 hypotheses with the highest DGHT-Votes, the feature vector $R(\mathbf{c}_i, \mathcal{X}_n)$, generated according to Equation (6), is used as input into the previously trained Random Forest Classifier. This determines the posterior probability $p(\Omega_r|R(\mathbf{c}_i, \mathcal{X}_n))$ that the hypothesis in question belongs to class Ω_r, i.e. that it is the target object. Finally, according to Equation (7) the best localization hypothesis is selected as the estimated target landmark (Figure 2).

3. EXPERIMENTS

We evaluated the SCM on an inhouse corpus from the University Hospital RWTH Aachen consisting of 812 unnormalized hand-radiographs with an average size of 1185×2066 pixel. The age of the subjects ranged from 3 to 19 years and the objective was the localization of the 12 epiphyses, illustrated in Fig. 3. 400 images were randomely selected to train the DGHT-Model as well as the Random Forest for the SCM. The remaining 412 images constituted the evaluation corpus.

In order to speed up the process and to increase the localization performance we used a multi-level localization approach with two zoom levels.[23,24] In the first level, the image resolution was reduced to one-eighth ensuring a

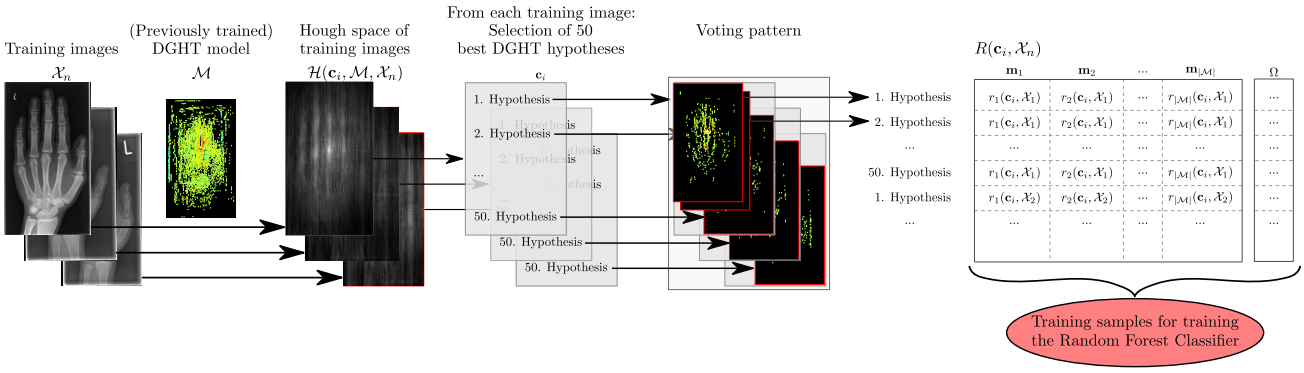

Figure 1. Scheme of the SCM training procedure: For each image $\mathcal{X}_n$, the feature vectors and corresponding class labels from the 50 best DGHT hypotheses constitute the training samples for training the Random Forest Classifier. Note, that hypotheses with an error between 3 and 10 Hough cells are not considered for the Random Forest training.

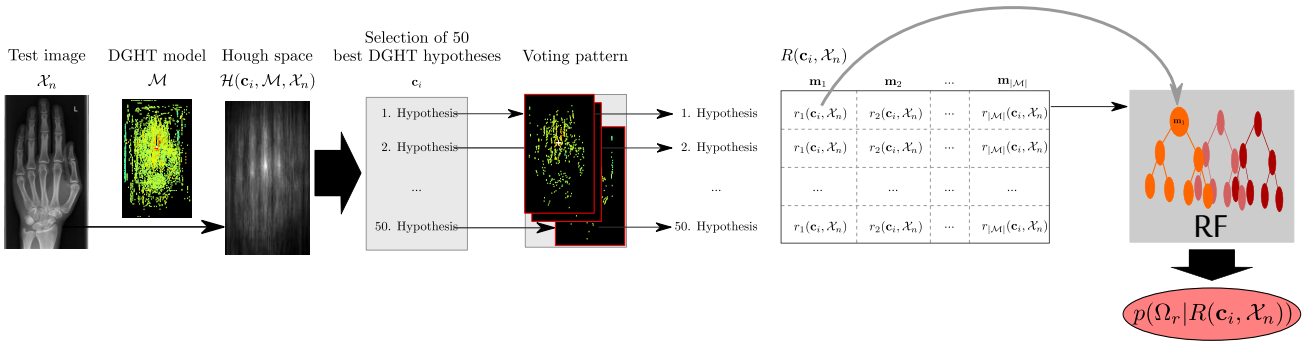

Figure 2. Scheme of the SCM test procedure: For an unknown test image $\mathcal{X}_n$, the feature vectors $R(\mathbf{c}_i, \mathcal{X}_n))$ from the 50 best DGHT hypotheses are fed into the Random Forest Classifier which determines $p(\Omega_r|R(\mathbf{c}_i, \mathcal{X}_n))$. Note, the used DGHT model $\mathcal{M}$ is the same as during the training procedure (see Figure 1).

robust albeit coarse localization. Subsequently an image extract of 192×256 pixels with the original resolution was selected around the detected point and used for a more accurate localization in the second level. Thus, the size of the image extract was large enough to compensate small errors from the first level but also small enough to exclude confusable objects like other epiphyses.

For both zoom levels as well as for each epiphysis we train a specific DGHT model by using the iterative training procedure, described in Ruppertshofen.[22] Thus, in total 24 different DGHT models were generated. These models were restricted to a maximum of 4000 model points, which is a good trade-off between capturing the main structures of the target while avoiding model points with negligible influence increasing only the processing time.

Since the Random Forest, implementing the SCM, is related to the voting pattern of a DGHT model, a separate Random Forest has to be trained for each of the 24 DGHT models. Each trained Random Forest consists of 500 unpruned trees. Since the Random Forest training involves a random component, also the results are partly coincidental. Therefore, all experiments, involving the SCM, were performed four times, and we report the mean and standard deviation over the four runs.[†]

After training the DGHT model and the SCM, each epiphysis was evaluated independently of any other epiphysis as follows: (I) The DGHT model for the first zoom level was applied, followed by (II) the application of the corresponding Random Forest Classifier to calculate the SCM for each of the 50 best localization hypotheses provided by the DGHT. Based on Equation (7) the best hypothesis was selected. Around this localization result

[†]Experiments only involving the DGHT (without SCM) do not use a random component and are only performed once.

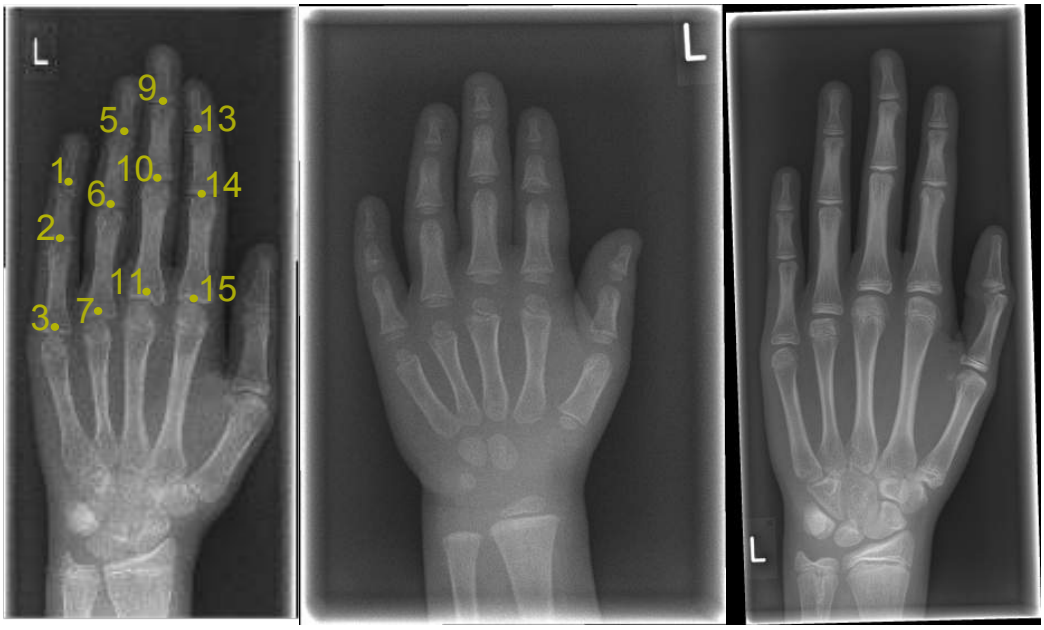

Figure 3. Epiphyses considered in this paper together with their identification number (left image) and some examples of hand-radiographs. Note that only the 12 epiphyses, which are used in this paper have been labeled.

(III) an image patch was extracted on which (IV) the DGHT model for the second zoom level was applied. (V) The corresponding SCM was calculated again for each of the 50 best localization hypotheses of this level and the best hypothesis according to Equation (7) was the final result. Note that the exact number of (50) DGHT localization hypotheses, considered for SCM assessment, had only a minor influence on the classification performance (see Section 5).

According to Fisher *et al.*,[5] a human observer perceives an epiphyseal localization as correct if the Euclidean distance to the center is less than 6 pixels for hand-radiographs normalized to an image height of 256 pixels. Since our images were not normalized and substantially larger, the human observer tolerance threshold was set to $\frac{6}{256} \cdot h_i$ pixel in the experiments, with h_i being the image height. A localization result with a Euclidean distance to the annotated point smaller than this threshold was considered as correct. Thus, we define the success rate for a given epiphysis to be the percentage of images where this epiphysis was correctly localized with our approach. The mean success rate is just the average of the 12 individual success rates.

4. RESULTS

The experiments showed a mean success rate of 99.4% for the localization of the 12 epiphyses (see Table 1), using an error tolerance perceived as correct by a human observer. This is a significant improvement of 2.9% compared to the DGHT baseline result without using the SCM methodology.

The mean localization error over all 412 test images decreased from 14.6 to 6.6 pixel (see Table 1) using the new SCM technique which enabled a successful localization of all 12 epiphyses in 96.2% of the images (Figure 5). The worst localization result for a single image on all four test runs still contained 7 successfully localized epiphyses which is expected to be sufficient for a subsequent automatic bone age assessment step (Figure 4).

5. DISCUSSION

Due to the dependency of the Random Forest tree generation on a random number generator it was necessary to study its robustness for different random initializations. It turned out that this had little effect on the variance of the localization result. The mean localization rate over all epiphyses varied only between 99.3% and 99.4%. Furthermore, 99.2% of epiphyses in all test images were correctly localized in all four test runs. This means that

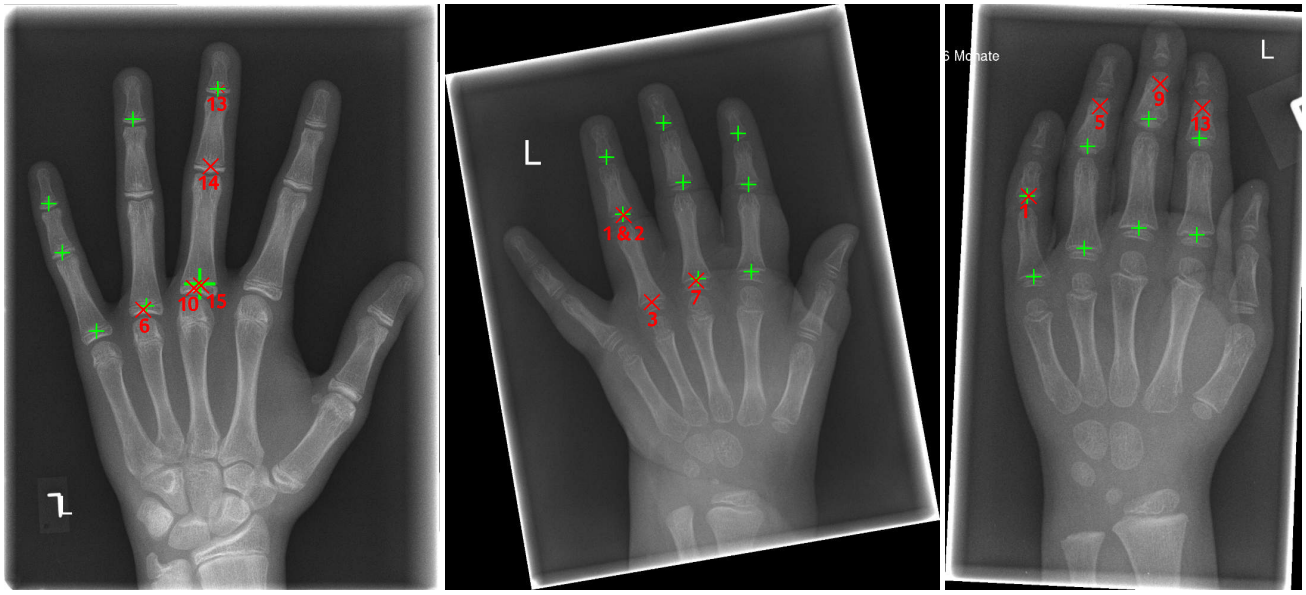

Figure 4. The 3 images with the most incorrect localizated epiphyses. The green "+" is a correct localization. Incorrect results are marked with a red "x" and the ID of the epiphyses. Note, on the second image, the epiphysis "1" and "2" were localized at the same position.

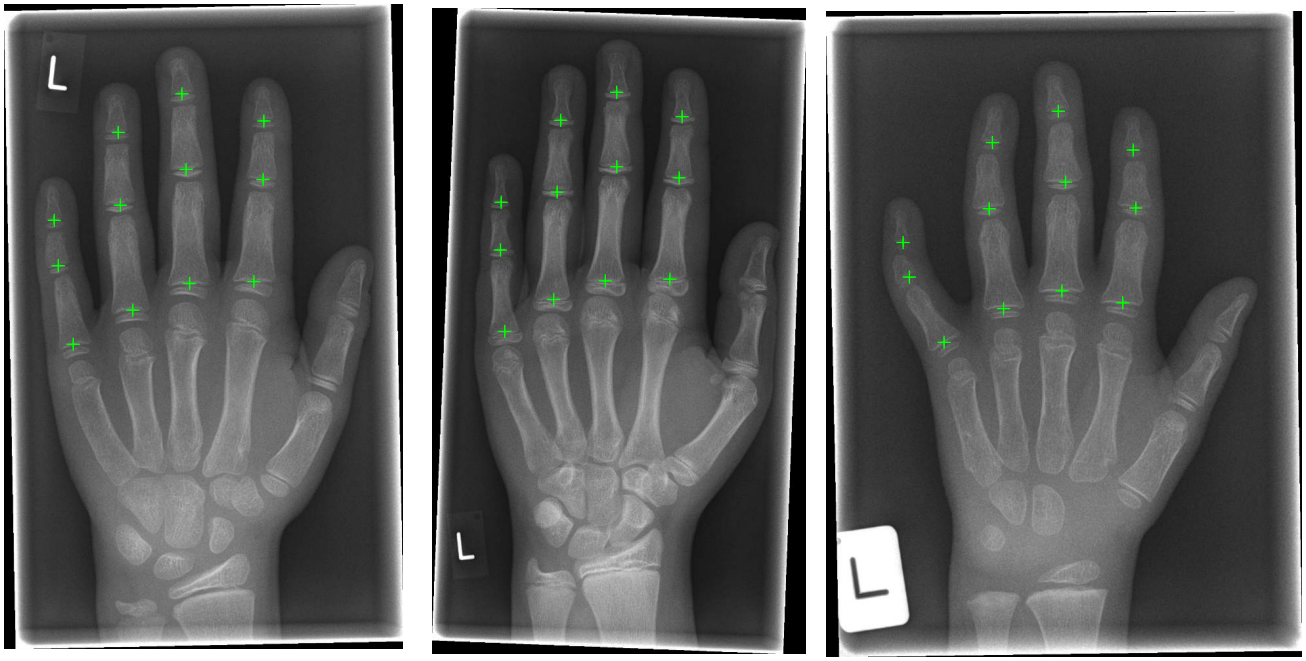

Figure 5. Some examples of correct localization results

	epiphysis ID													ø Error
	1	**2**	**3**	**5**	**6**	**7**	**9**	**10**	**11**	**13**	**14**	**15**	**Mean**	(Pixel)
DGHT	93.4	94.4	96.8	97.8	96.4	97.1	95.6	98.3	98.8	93.9	96.8	98.1	**96.5**	14.6
DGHT	99.5	99.1	99.3	99.7	99.3	99.3	99.0	99.5	99.8	99.0	99.3	99.8	**99.4**	6.6
+ SCM	±.1	±.1	±.0	±.1	±.0	±.0	±.3	±.2	±.2	±.4	±.0	±.0		

Table 1. Comparison of success rates (%) for epiphysis localization using (1) the standard DGHT, (2) and the DGHT in combination with the SCM. Results are provided for the considered 12 epiphyses, illustrated in Figure 3. The column "Mean" provides the average over the individual epiphysis localization results.

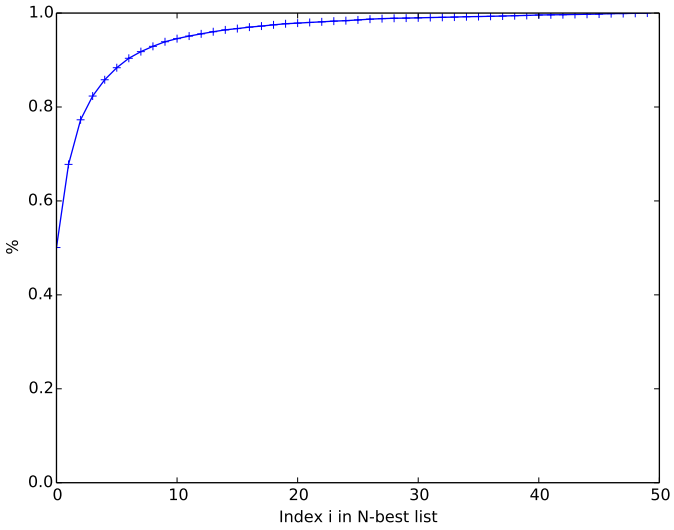

Figure 6. Percentage of cases in which the i-th entry in the N-best list was chosen by SCM (Equation (7))

most of the wrong localization results occurred in all test runs and thus the random number generator had a negligible influence not only on the mean success rate but also directly on the localization results.

As mentioned before, only the 50 best DGHT-localization hypotheses were used during evaluation. This is reasonable since the DGHT already achieved a good localization performance of 96.5% so that we assumed that the correct localization was among the best 50 alternatives. The analysis revealed that in 50% of all cases the SCM indeed confirms the best DGHT hypothesis. Furthermore, only very few hypotheses (1.5%), that were not among the 25 best DGHT results, were chosen by the SCM (see also Figure 6). Therefore, it can be assumed that rating only 50 DGHT hypotheses in the evaluation did not significantly degrade the achieved results.

On a standard PC with x64 Intel Xeon CPU E5-1650 @ 3.2 GHZ and 32 GB RAM the processing time for the GHT voting procedure, including the selection of the 50 best results from the Hough space, was found to be 1.5s on average per epiphysis and zoom level. The subsequent generation of the feature vector for the SCM took 8.9ms and the Random Forest classification additional 9.7ms, while further 0.7ms were required for other administrative tasks, such as initialization and result sorting. Note that no time measurements have been conducted for image pre-processing, like edge detection and downsampling, and that no runtime optimization has been performed, yet. Thus, the overall processing time is only slightly increased by the SCM but results in a large improvement of the localization performance.

Hahmann *et al.*[19] suggested that considering the GHT voting characteristics based on a single, individual Hough cell is not robust and may result in overfitting. To analyze this aspect, we set $\vartheta = 0$ in Equation (5). As expected, the localization rate decreased to 98.0%. At the same time, the processing time for classifying the feature vectors increased to 31.6ms. This is due to a larger mean tree depth of 72.5 nodes for $\vartheta = 0$ compared to 22.2 nodes for $\vartheta = 7$ (Figure 7). Thus, feature vectors referring to a single Hough cell contain less

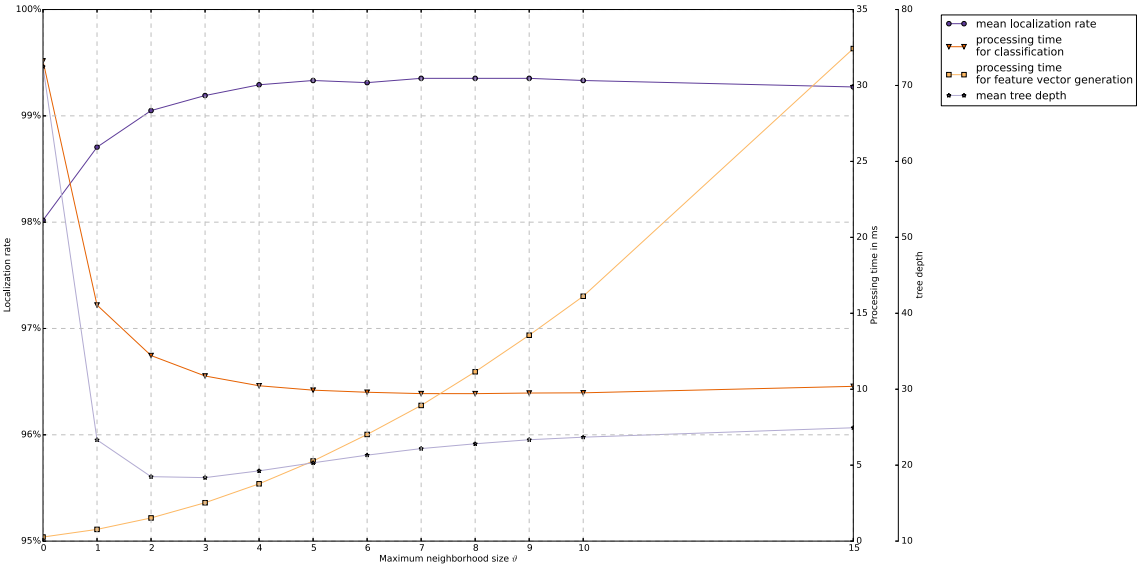

Figure 7. Comparison of mean success rate in %, processing time in ms (separately for feature vector generation and classification), and mean tree depth depending on the maximum neighborhood size ϑ.

information, and in order to achieve a good representation of the training samples, the Random Forest needs a larger tree depth. At the same time, the test samples are insufficiently represented by the trees resulting in a worse localization rate. Hence, without including a local neighborhood in the feature vectors of the SCM, the SCM tends to overfitting .

By contrast, even with a small neighborhood of $\vartheta = 1$, the tree depth reduced to 23 nodes while the localization rate increased to 98.7%. By using $\vartheta \geq 4$, the localization rates saturated at approximately 99.3% but the processing time for generating the feature vectors increased squarely with the maximum neighborhood size. Thus, a value of 4 or 5 seems to be a good trade-off between localization rate and processing time for the given task (Figure 7).

The presented localization results can be used for extracting epiphyseal regions of interest, which is required for subsequent Bone Age Assessment. For instance, this task can be solved by the Classifying Generalized Hough Transform[25,26] or a Support Vector Machine.[27–29]

6. CONCLUSIONS

A general problem of GHT-based object localization frameworks, dealing with large shape variabilities, is the independent treatment of model points during the voting procedure. When using a GHT model, incorporating different target shape variants, points from mutually exclusive variations may vote for the same cell in the Hough space which may lead to false positive localizations. This problem can be addressed by (1) rating the pattern of model points, voting for a cell in the Hough space and its neighborhood, with a Random Forest Classifier and (2) using the derived shape consistency measure to weight the votes in the Hough space. In this work, we successfully applied the described method to medical image processing. A substantial improvement of the localization rate from 96.5% to 99.4% could be achieved for the task of extracting the 12 epiphyseal regions of interest in hand-radiographs of patients with an age range between 3 and 19 years. Since the subsequent Bone Age Assessment step of the overall system relies on a combined classification of the 12 epiphyseal regions, the small amount of remaining localization errors is expected to have only a minimal effect on the final system performance.

Furthermore, we analyzed the dependence of the SCM on the radius of the local neighborhood which is considered around each Hough cell to analyze the voting pattern. It was found that a non-zero neighborhood is crucial and that the localization performance is improved with increasing radius. However, also the processing time of the algorithm increases due to a different depth of the trained Random Forest, so that there is a trade-off. Optimal values for the radius have been found to be in the range from 4 to 5.

ACKNOWLEDGMENTS

This work is funded by the Central Innovation Program SME from the Federal Ministry for Economic Affairs and Energy.

REFERENCES

[1] Tanner, J. M., Whitehouse, R. H., Marshall, W. A., Healty, M. J. R., Goldstein, H., Tanner, J. M., Whitehouse, R. H., Marshall, W. A., Healty, M. J. R., and Goldstein, H., [*Assessment of Skeleton Maturity and Maturity and Prediction of Adult Height (TW2 Method)*], Academic Press, London (1975).

[2] Hsieh, C.-W., Jong, T.-L., and Tiu, C.-M., "Bone age estimation based on phalanx information with fuzzy constrain of carpals," *MED BIOL ENG COMPUT* **45**(3), 283–295 (2007).

[3] Pietka, E., Gertych, A., Pospiech, S., Cao, F., Huang, H., and Gilsanz, V., "Computer-assisted bone age assessment: image preprocessing and epiphyseal/metaphyseal ROI extraction," *IEEE T MED IMAGING* **20**(8), 715–729 (2001).

[4] Thodberg, H., Kreiborg, S., Juul, A., and Pedersen, K., "The BoneXpert Method for Automated Determination of Skeletal Maturity," *IEEE T MED IMAGING* **28**(1), 52–66 (2009).

[5] Fischer, B., Brosig, A., Deserno, T. M., Ott, B., and Günther, R. W., "Structural scene analysis and content-based image retrieval applied to bone age assessment," *Proc. SPIE Medical Imaging* , 726004–726004–11 (2009).

[6] Ballard, D. H., "Generalizing the Hough transform to detect arbitrary shapes," *PATTERN RECOGN* **13**(2), 111–122 (1981).

[7] Gall, J., Yao, A., Razavi, N., Van Gool, L., and Lempitsky, V., "Hough Forests for Object Detection, Tracking, and Action Recognition," *IEEE T PATTERN ANAL* **33**(11), 2188–2202 (2011).

[8] Criminisi, A., Shotton, J., Robertson, D., and Konukoglu, E., "Regression Forests for Efficient Anatomy Detection and Localization in CT Studies," *Proc. MICCAI* , 106–117 (2010).

[9] Criminisi, A., Robertson, D., Konukoglu, E., Shotton, J., Pathak, S., White, S., and Siddiqui, K., "Regression forests for efficient anatomy detection and localization in computed tomography scans," *Med Image Anal* **17**(8), 1293–1303 (2013).

[10] Donner, R., Menze, B. H., Bischof, H., and Langs, G., "Global localization of 3D anatomical structures by pre-filtered Hough Forests and discrete optimization," *Med Image Anal* **17**(8), 1304–1314 (2013).

[11] Stern, D., Ebner, T., Bischof, H., Grassegger, S., Ehammer, T., and Urschler, M., "Fully Automatic Bone Age Estimation from Left Hand MR Images," *Proc. MICCAI* , 220–227 (2014).

[12] Dantone, M., Gall, J., Fanelli, G., and Van Gool, L., "Real-time facial feature detection using conditional regression forests," *Proc. CVPR* , 2578–2585 (2012).

[13] Felzenszwalb, P., Girshick, R., McAllester, D., and Ramanan, D., "Object Detection with Discriminatively Trained Part-Based Models," *IEEE T PATTERN ANAL* **32**(9), 1627–1645 (2010).

[14] Sun, M., Bradski, G., Xu, B.-X., and Savarese, S., "Depth-Encoded Hough Voting for Joint Object Detection and Shape Recovery," *Proc. ECCV* , 658–671 (2010).

[15] Razavi, N., Gall, J., Kohli, P., and van Gool, L., "Latent Hough Transform for Object Detection," *Proc. ECCV* , 312–325 (2012).

[16] Ruppertshofen, H., Lorenz, C., Beyerlein, P., Salah, Z., Rose, G., and Schramm, H., "A multidimensional model for localization of highly variable objects," *Proc. SPIE Medical Imaging* (2012).

[17] Leibe, B., Leonardis, A., and Schiele, B., "Robust Object Detection with Interleaved Categorization and Segmentation," *INT J COMPUT VISION* **77**(1-3), 259–289 (2008).

[18] Razavi, N., Gall, J., and Gool, L. V., "Backprojection Revisited: Scalable Multi-view Object Detection and Similarity Metrics for Detections," *Proc. ECCV* , 620–633 (2010).

[19] Hahmann, F., Böer, G., Gabriel, E., Meyer, C., and Schramm, H., "A Shape Consistency Measure for Improving the Generalized Hough Transform," *Proc. VISAPP* (2015).

[20] Breiman, L., "Random Forests," *Machine Learning* **45**(1), 5–32 (2001).

[21] Juang, B.-H. and Katagiri, S., "Discriminative learning for minimum error classification [pattern recognition]," *IEEE T SIGNAL PROCES* **40**(12), 3043–3054 (1992).

[22] Ruppertshofen, H., *Automatic Modeling of Anatomical Variability for Object Localization in Medical Images*, PhD thesis, Universität Magdeburg (2012).

[23] Hahmann, F., Boer, G., and Schramm, H., "Combination of facial landmarks for robust eye localization using the Discriminative Generalized Hough Transform," *Proc. BIOSIG* , 1–12 (2013).

[24] Hahmann, F., Böer, G., Deserno, T. M., and Schramm, H., "Epiphyses localization for bone age assessment using the discriminative generalized hough transform," *Proc. BVM* , 66–71 (2014).

[25] Brunk, M., Ruppertshofen, H., Schmidt, S., Beyerlein, P., and Schramm, H., "Bone Age Classification Using the Discriminative Generalized Hough Transform," *Proc. BVM* , 284–288 (2011).

[26] Hahmann, F., Berger, I., Ruppertshofen, H., Deserno, T., and Schramm, H., "Bone Age Assessment Using the Classifying Generalized Hough Transform," *Proc. GCPR* , 313–322 (2013).

[27] Harmsen, M., Fischer, B., Schramm, H., Seidl, T., and Deserno, T., "Support Vector Machine Classification Based on Correlation Prototypes Applied to Bone Age Assessment," *IEEE J Biomed Health Inform* **17**(1), 190–197 (2013).

[28] Kashif, M., Jonas, S., Haak, D., and Deserno, T. M., "Bone age assessment meets SIFT," *Proc. SPIE Medical Imaging* , 941439–941439–7 (2015).

[29] Kashif, M., Deserno, T. M., Haak, D., and Jonas, S., "Feature description with SIFT, SURF, BRIEF, BRISK, or FREAK? A general question answered for bone age assessment," *COMPUT BIOL MED* **68**, 67–75 (2016).