Epiphyses Localization for Bone Age Assessment Using the Discriminative Generalized Hough Transform

Ferdinand Hahmann[1], Gordon Böer[1], Thomas M. Deserno[2], Hauke Schramm[1]

[1]Institute of Applied Computer Science, University of Applied Sciences Kiel
[2]Department of Medical Informatics, RWTH Aachen University
ferdinand.hahmann@fh-kiel.de

Abstract. This paper presents the Discriminative Generalized Hough Transform (DGHT) as a robust and accurate method for the localization of epiphyseal regions in radiographs of the left hand. The technique utilizes a discriminative training approach to generate shape models with individual positive and negative model point weights for the Generalized Hough Transform. The framework incorporates a multi-level approach which reduces the searched region in two zooming steps, using specifically trained DGHT shape models. In addition to the standard method, a novel landmark combination approach is presented. Here, the N-best lists of individual landmark localizations are combined with anatomical constraints to achieve a globally optimal localization result for all 12 considered epiphyseal regions of interest. The technique has been applied to extract 12 epiphyseal regions of interest for a subsequent automatic bone age assessment. It achieved a localization success rate of 98.1% on a corpus with 412 left hand radiographs covering the age range from 3 to 19 years.

1 Introduction

Bone Age Assessment (BAA) is an important method in diagnostic radiology which is used for evaluating the skeletal maturity 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 opinion from a physician a number of automatic methods have been developed in recent years. Many of these approaches follow the basic concept of TW to classify only certain extracts from the radiograph since this substantially reduces the complexity of the classification problem. However, an important prerequisite for all these techniques is the availability of a reliable and robust object detection method to enable the extraction of the required region-of-interest.

Object detection problems in medical image analysis are often solved by individually adjusted methods, which make heavy use of expert knowledge about the shape and neighborhood of the searched object. [2], for example, analyzes image lines to assign the maxima in the vertical stripe pattern of a hand radiograph to the individual phalanges. The localization approach in [3] is based on

the same idea using an arc scanning procedure to generate a gray value matrix whose maximal column sums refer to the phalanges. A more general approach has been presented in [4] where an active appearance model is used for bone reconstruction. This method can also be used for epiphysis localization but requires a high number of manually defined shaped points for the training process. [5] presents another segmentation approach used for epiphysis localization. Here, a graph-based structural prototype, representing the phalanges and metacarpal bones, is registered to the image and successfully provides the epiphyseal regions on 77% of 137 images.

There are only a few general object detection techniques which have shown a good performance in the field of medical image analysis. Marginal space learning [6] trains probabilistic boosting trees and gradually estimates the translation, rotation and scaling parameters of the searched object. The technique has successfully been applied to a number of anatomical objects but requires a large number of training images which are difficult to obtain in the field of medical imaging. Another general object detection approach is described in [7] where random forests are used to map image patches to Hough space votes for possible target point location. The work of [8] is based on the same idea but utilizes so called regression forests.

In this paper, the Discriminative Generalized Hough Transform (DGHT) is introduced as a robust method for the detection of epiphyseal regions of interest (eROI), located around the finger joints (cf. Fig. 1a), in hand radiographs. The DGHT is a general object localization approach, which has been successfully applied to the localization of anatomical structures in medical images and several non-medical tasks [9]. To assure a reasonable combination of the localization results for all visible eROIs the technique is applied in conjunction with some simple geometrical constraints taken from the hand anatomy.

2 Materials and Methods

In this contribution, the DGHT is used in the first step to generate a ranked list of localization candidates for each searched landmark (Sect. 2.1). The obtained set of lists is searched for an optimal combined solution for all eROI positions. This solution must meet a number of predefined constraints derived from the hand anatomy (Sect. 2.2).

2.1 Discriminative Generalized Hough Transform

The technique extends the well-known Generalized Hough Transform (GHT) by an iterative and discriminative training approach for the generation of optimal shape models with individually weighted model points [9]. The availability of models compensating moderate target object variability is crucial due to the expected anatomical variability. Apart from the object's translation no additional GHT transformation parameters are considered in this work, since we follow the idea of learning medium variability into the model.

The DGHT training procedure starts with the generation of an initial shape model which is extracted from a set of training images by overlaying their edge features inside a predefined region-of-interest with respect to a target point. The individual contributions of each initial model point to the Hough space are combined in a Maximum Entropy Distribution. This step introduces model point specific weights which are optimized in the next step using a Minimum Classification Error (MCE) training approach [10]. The optimized weights reflect the importance of each individual model point for supporting the correct localization and suppressing votes at similar but false structures. The latter aspect is possible since model point weights may also have negative values. After the model point weights in the initial shape model have been optimized with the described procedure, points with low absolute weights may be eliminated since their influence on the localization result is negligible.

The optimized and thinned initial model is further enhanced by an iterative procedure. This technique evaluates the model on the training images and extends it with image features from the target region and confusable areas in training images with high localization error. The procedure is repeated until the localization error on all training images is below some given threshold.

In addition to the model optimization procedure the applied framework uses a coarse-to-fine localization strategy to split different levels of anatomical detail into different localization models. This technique may also speed-up the processing in case of high resolution images. In this work, only two zooming steps are applied since the confusability of individual fingers in intermediate levels is quite high. In zoom level 0, showing the whole hand, the resolution is reduced to one-eight and a specifically trained DGHT model is used to obtain a coarse localization result. An image extract of size 192×256 with the original resolution is cut around the localized point and used for a precise eROI detection in zoom level 1 with a specific and more detailed DGHT model.

2.2 Constrained Localization

To avoid the confusion of eROIs, which is a frequent source of error, the individual eROI localization results are combined with 133 simple anatomical constraints obtained from the hand anatomy. These describe (1) the minimum distance of eROIs (50 pixels), (2) the positioning of the fingers with respect to other fingers (e.g. index finger is right to middle finger), and (3) the positioning of eROIs inside a single finger with respect to the other eROIs in this finger (e.g. metacarpophalangeal is below proximal interphalangeal).

To derive a confidence measure from these requirements, each of the 12 considered eROI localization hypotheses (Fig. 1a) is assigned an individual error score which is determined by the number of unmet conditions. With this score the localization hypotheses are corrected in the following iterative manner: First, the eROI localization hypothesis with the highest error score is identified. Second, the hypothesis is rejected and replaced by a concurrent one from the 10-best list. The replacement is selected to have at least 95% of the GHT votes of the best hypothesis and the smallest error score of all remaining 10-best entries.

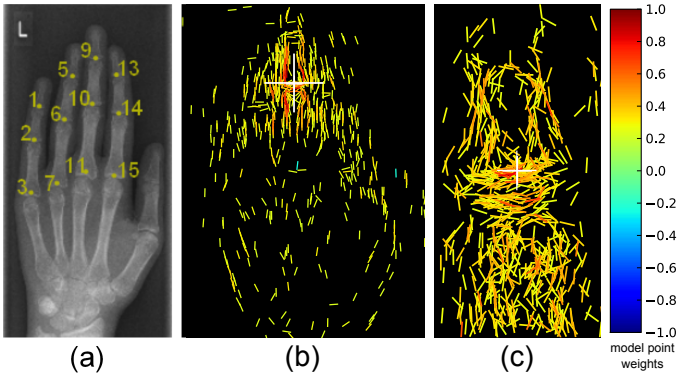

Fig. 1. (a) Location and identification number of eROIs considered in this paper. Examples of (b) the global DGHT localization model of zoom-level 0 and (c) the DGHT localization model of zoom-level 1 with color-coded model point weights.

Third, the error score is recalculated using the replacement hypothesis and the next iteration is started. The iteration stops if the error scores of all eROIs are zero or if all landmarks have been changed. If it is not possible to fulfill all constraints, the system may ask the user for manual correction.

2.3 Experimental Setup

The aim of this contribution is to set up a framework for localizing the 12 eROIs labelled in Fig. 1a. Since these regions are extracted for later usage in an automatic BAA system [11] it is necessary that the complete epiphysis is visible inside the identified image fragment. Due to the anatomy of the epiphyseal region the image extract chosen here is a narrow upright rectangle. Consequently, the applied error distance for measuring the success of the localization is asymmetric and tolerates a larger deviation from the target point in vertical direction than in horizontal direction. Thus, the localization is defined as being successful if the localized point differs less than 50 / 100 pixels in horizontal / vertical direction from the given target point. This assures the containment of the eROI inside the resulting image extract and therefore allows for a subsequent BAA classification step. An alternative error measure for eROI detection has been given in [5]. Here, it was stated that a human observer may accept a localization error of 6 pixel for hand radiographs with a height of 256 pixel. Below we will also refer to this error mesasure as 'Fischers measure' for better comparability with the literature.

For training and test of the described framework an inhouse corpus from the University Hospital Aachen, consisting of 812 unnormalized hand radiographs with an average size of 1185 × 2066 pixel ranging from 3 to 19 years, has been used. Note that radiographs of children younger than 3 years were not considered since only little data with low image quality, especially with respect to contrast and hand positioning, was available. The corpus, which contained target point annotations for all eROIs, was split into 400 randomly selected training and 412 evaluation images. For each eROI a single DGHT localization model was trained for both genders and the complete age range from 3 to 19 years.

Table 1. Success rates (%) for eROI localization using (1) the standard DGHT, (2) the method of constrained localization, and (3) two zooming-levels. Results are provided for the considered 12 eROIs which label IDs can be seen in Fig. 1a.

	eROI (ID)													**ø Error**
	1	**2**	**3**	**5**	**6**	**7**	**9**	**10**	**11**	**13**	**14**	**15**	**All**	**(Pixel)**
DGHT	94.9	94.9	96.6	98.1	96.1	96.8	95.4	98.3	98.8	93.9	96.6	97.8	**96.3**	23.2
Constrained localization	96.4	98.1	97.6	98.8	98.5	98.5	96.4	99.0	99.8	95.4	98.6	98.1	**97.8**	20.1
+ 2nd zoom-level	96.6	98.3	98.5	98.8	98.5	98.8	96.8	99.0	99.8	95.4	98.1	98.5	**98.1**	11.4

3 Results

With the standard procedure using a global DGHT model (Fig. 1b) an overall localization success rate of 96.3% was obtained for all eROIs (Tab. 1). The average localization error for this experiment was 23 pixel corresponding to about 1% of the image height. This result could be significantly improved to an overall success rate of 97.8% by applying the constrained localization method (Sect. 2.2). The technique also improved the average localization error to about 20 pixel. A much more exact localization with an average error of only 11 pixel could be achieved with a second zooming step utilizing a refined model, specifically trained on small image extracts around the eROI. This model has been optimized to represent fine details instead of global characteristics and is therefore much more exact than the global model (Fig. 1c). An additional gain of the success rate of 0.3% was achieved and led to an overall rate of 98.1%. This value is only slightly decreased to 97.6% when using the stricter 'Fischers measure'.

4 Discussion and Conclusion

The experimental results show that the DGHT can be used as a robust and accurate method for the localization of epiphyseal regions of interest in radiographs of the left hand. The high localization rates of the basic technology can be substantially improved by the introduced novel landmark combination approach. It utilizes a set of simple anatomical constraints and successfully identifies globally optimal eROI localization results from the N-best lists of the individual DGHT-based localizations. In order to further improve the localization accuracy a zoom-in strategy with a specialized high-detail DGHT model was applied which nearly halved the average localization error to about 11 pixel.

Considering the few remaining errors of the system nearly two-thirds can be identified by unfulfilled anatomical conditions (Sect. 2.2) which allows for eliminating those results in a subsequent BAA classification step. Since the applied BAA framework requires only a subset of eROIs, the elimination of single regions from the combined decision is probably not critical for the overall classification

success. It is additionally expected that the influence of isolated undetected localization errors on the BAA result is low since several eROIs are utilized for the final decision.

The experiments were performed on radiographs of the complete age range from 3 to 19 years and both genders. The achieved high localization rates demonstrate that most of the resulting large object variability could be successfully trained into a single weighted DGHT shape model. An important source of variability for this task is, however, the spreading of fingers which lead to a clear error rate increase in the upper eROIs (especially No. 1, 9, and 13 in Fig. 1a). Although this specific kind of shape variability might in general be learned by the training process it is not sufficiently represented in the used training corpus and therefore lead to a substantial amount of the remaining errors.

For the addressed BAA task, the achieved localization accuracy is sufficient. A combination of the presented DGHT based localization framework with the Classifying GHT for bone age assessment [11] will be addressed in the next step of the investigations.

References

1. Tanner J, Healy M, Goldstein H, et al. Assessment of skeletal maturity and prediction of adult height (TW3). WB Saunders. 2001;.
2. Pietka E, Gertych A, Pospiech S, et al. Computer-assisted bone age assessment: Image preprocessing and epiphyseal/metaphyseal ROI extraction. IEEE Trans Med Imaging. 2001;20:715–729.
3. Hsieh C, Jong T, Tiu C. Bone age estimation based on phalanx information with fuzzy constrain of carpals. Med Biol Eng Comput. 2007;45:283–295.
4. Thodberg H, Kreiborg S, Juul A, et al. The BoneXpert method for automated determination of skeletal maturity. IEEE Trans Med Imaging. 2009;28:52–66.
5. Fischer B, Brosig A, Deserno T, et al. Structural scene analysis and content-based image retrieval applied to bone age assessment. SPIE Medical Imaging. 2009;7260:726004–1.
6. Zheng Y, Georgescu B, Comaniciu D. Marginal space learning for efficient detection of 2D/3D anatomical structures in medical images. Inf Process Med Imaging. 2009;21:411–422.
7. Gall J, Yao A, Razavi N, et al. Hough forests for object detection, tracking, and action recognition. IEEE Trans Pattern Anal Mach Intell. 2011;33:2188–2202.
8. Criminisi A, Shotton J, Robertson D, et al. Regression forests for efficient anatomy detection and localization in CT studies. Medical Computer Vision. 2011;.
9. Ruppertshofen H. Automatic Modeling of Anatomical Variability for Object Localization in Medical Images. Ph.D. thesis, University Magdeburg; 2013.
10. Juang BH, Katagiri S. Discriminative learning for minimum error classification. IEEE Transactions on Signal Processing. 1992;40:3043–3054.
11. Hahmann F, Berger I, Ruppertshofen H, et al. Bone Age Assessment Using the Classifying Generalized Hough Transform. Pattern Recognit. 2013; p. 313–322.