Bone Age Assessment Using the Classifying Generalized Hough Transform

Ferdinand Hahmann[1], Inga Berger[1], Heike Ruppertshofen[2],
Thomas Deserno[3], and Hauke Schramm[1]

[1] University of Applied Sciences Kiel
[2] Philips Technologie GmbH
[3] Department of Medical Informatics, RWTH Aachen University
Ferdinand.Hahmann@fh-kiel.de

Abstract. A theoretical description and experimental validation of the Classifying Generalized Hough Transform (CGHT) is presented. This general image classification technique is based on a discriminative training procedure that jointly estimates concurrent class-dependent shape models for usage in a GHT voting procedure. The basic approach is extended by a coarse-to-fine classification strategy and a simple classifier combination technique for a combined decision on several regions of interest in a given image. The framework is successfully applied to the task of automatic bone age assessment and produces comparable results to other state-of-the-art techniques on a public database. For the most difficult age range of 9 to 16 years the automatic system achieves a mean error of 0.8 years compared to the average rating of two physicians. Unlike most other image classification techniques, the trained CGHT models can be visually interpreted, unveiling the most relevant anatomical structures for class discrimination.

1 Introduction

Bone Age Assessment (BAA) based on left hand radiographs is a well-established procedure for determining the skeletal maturity which is mainly applied for diagnosing growth disorders or forensic age estimation. Manual BAA is usually performed with one of two common methods: Greulich & Pyle (GP) [7] developed an approach, in which the radiologist determines the bone age by comparing the radiograph with a standard atlas. In contrast, Tanner & Whitehouse (TW) [17] have proposed to consider only regions of interest (ROI) around the epiphyses and the carpal bones. For each of these so-called eROIs a score based on the gap and shape of the epiphysis is assigned. The sum of all scores determines the age.

Since the manual assessment is time consuming, subjective, and requires expert knowledge, an automatic method is desirable. In recent years, various automatic techniques have been proposed, which are usually based on some kind of image feature extraction in combination with a standard classification technique. While some of these approaches employ heuristic features, like the length and size of phalanges [6,9] or the distance between metaphysis and diaphysis

J. Weickert, M. Hein, and B. Schiele (Eds.): GCPR 2013, LNCS 8142, pp. 313–322, 2013.

[13], other methods directly utilize the TW rules, for example, by using a decision tree [1] or an artificial neural network [4]. The leading commercial product, BoneXpert, employs the rules from TW after applying an active shape model for the segmentation of 15 bones [18].

More general BAA approaches without any kind of heuristic feature selection are Kim & Kim [10] and Harmsen et al. [8]. Kim & Kim classify discrete cosine transform coefficients, computed from pixel intensity values in epiphyseal regions, with a linear discriminant analysis. Harmsen et al. analyze 14 epiphyseal regions of interest (eROIs) using the cross-correlation with 30 class-specific prototypes as features and employing a k-Nearest Neighbor algorithm (kNN) or a Support Vector Machine (SVM) for classification.

In this work, the Discriminative Generalized Hough Transform (DGHT) [14] is extended. The DGHT utilizes a discriminative training technique to estimate optimal shape models for usage in a standard Generalized Hough Transform (GHT) approach and achieves high localization rates for well-defined objects with medium shape variability. An unsupervised training method can be used to learn parallel variation-specific GHT models to deal with stronger variabilities [16] and returns the variation class together with the localization result. This approach can be modified towards a general image classification technique, called Classifying Generalized Hough Transform (CGHT) [15]. A first proof-of-concept has already been presented in [5], where CGHT-based models were successfully applied to the BAA task of separating the age classes 11-12 and 14-15 years. In this paper the method is theoretically described and comprehensively evaluated.

2 Method

2.1 Discriminative Generalized Hough Transform

The classification technique, presented in this paper, is based on the Generalized Hough Transform (GHT) [2] which is a general model-based localization method. For 2D images, a point model $M := \{\mathbf{m}\} \subset \mathbb{R}^2$ is used to represent the shape of the searched-for object in relation to a reference point, which is the target point for localization. Using this model, a voting procedure transforms a feature image X_n, usually a binary edge image, into a parameter space H, called Hough space. The Hough space is usually quantized and consists of so called Hough cells $\mathbf{c}$, which accumulate the votes in the respective region. The cells represent possible target point locations and, potentially, shape model transformations. The latter are not considered in this work, since moderate shape variations are learned into the shape model. Thus, the voting procedure may be simplified as follows:

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

Here, $\mathbf{e}_i$ represents the i-th feature point while $\mathbf{m}_j$ is the vector from the reference point to the j-th model point. A pair $(\mathbf{e}_i, \mathbf{m}_j)$ is allowed to vote if the difference of the gradient direction of $\mathbf{e}_i$ and the orientation of $\mathbf{m}_j$ is below the threshold ϑ_φ. The number of votes per accumulator cell $\mathbf{c}$ after the quantization reflects the degree of matching between the feature image X_n and the model M at this point. The best positioning of the model in the image is given by the Hough cell $\hat{\mathbf{c}} = \arg\max_{\mathbf{c}} H(\mathbf{c})$ with the highest degree of matching.

Since the accuracy of GHT localization highly depends on the quality of the shape model, the Discriminative Generalized Hough Transform (DGHT) additionally includes a machine learning approach for generating discriminative models. This procedure, which is described in detail in [14], assigns individual positive and negative weights to model points based on their importance for correct localizations on training images. The GHT-based classification technique, explained in the next section, is an extension of the DGHT which employs a number of competitively trained submodels.

2.2 Classifying Generalized Hough Transform

Given a classification task with K classes, the CGHT [15] combines a set of K competitive submodels M_k into a 3D GHT Model $M = \{M_k\}$, $k \epsilon \{1, ..., K\}$. Each submodel in M represents one class and the whole set $\{M_k\}$ is jointly optimized with respect to a minimum classification error (Section 2.3). This competitive training procedure assigns large absolute weights to model points supporting the class discrimination while eliminating irrelevant model parts.

Applying the optimized 3D GHT model M on a 2D image results in a 3D Hough space $H(\mathbf{x}, k) = \{H_k(\mathbf{x})\}$, $k \epsilon \{1, ..., K\}$ (Fig. 1), whereas the individual $H_k(\mathbf{x})$ have been obtained by applying the voting procedure in Equation 1 to the submodels M_k:

$$H_k(\mathbf{x}) = \sum_{\forall \mathbf{e}_i \in X_n} \sum_{\forall \mathbf{m}_j \in M_k} \begin{cases} 1, & \text{if } \mathbf{x} = \mathbf{e}_i - \mathbf{m}_j \text{ and } |\varphi_i - \varphi_j| \leq \vartheta_\varphi \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

After the voting procedure has finished, the classification result $\hat{k}$ is given by the submodel with the highest degree of matching $\hat{k} = \arg\max_k [\max_{\mathbf{x}} H_k(\mathbf{x})]$.

Note that this procedure is flexible enough to compensate for a moderate variability of the object's position in the image. As long as the object to be classified is completely visible, localizing the peak in the Hough space does not effect the classification result.

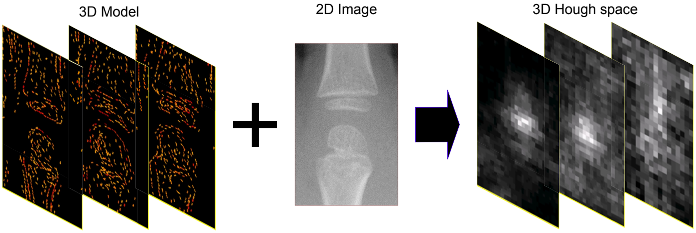

Fig. 1. Classification of an epiphyseal region of interest (eROI) using the Classifying Generalized Hough Transform

2.3 Training

The training procedure starts with an initial model $M = \{M_k\}$ composed of identical submodels $M_k \equiv M_{k'}$, $\forall(k', k)$, which is obtained by overlaying the features of several training images in a predefined region-of-interest around a manually annotated landmark. Other methods for obtaining an initial shape model are described in [14].

With this model the modified voting procedure in Equation (2) is applied producing a 3D Hough space $H(\mathbf{x}, k)$. Note that initially the $H_k(\mathbf{x})$ are identical for all K classes. To determine individual model point weights it is necessary to capture the influence of each single model point on the Hough space, which is achieved by the following feature function:

$$f_j^k(\mathbf{c}_i^k, X_n) = v_{i,j}^k. \tag{3}$$

For a given class k, this function denotes the number of votes $v_{i,j}^k$ from model point $\mathbf{m}_j^k$ in Hough cell $\mathbf{c}_i^k$. Considering the constraints of the GHT voting procedure for the entire model, the individual contributions of all model points have to be recombined into an overall distribution. To assure maximum objectivity, the maximum entropy distribution

$$p_{\Lambda_k}(\mathbf{c}_i^k|X_n) = \frac{\exp\left(\sum_j \lambda_j^k \cdot f_j^k(\mathbf{c}_i^k, X_n)\right)}{\sum_l \exp\left(\sum_j \lambda_j^k \cdot f_j^k(\mathbf{c}_l^k, X_n)\right)} \tag{4}$$

is used, which introduces class and model point specific weights $\Lambda_k = \{\lambda_1^k, ..., \lambda_{J_k}^k\}$. Note that this probabilistic representation of the Hough space is in line with the standard GHT theory, as the Hough space can be easily transferred into a probability distribution by using relative frequencies.

Since this work aims at a minimum classification error instead of a Hough space with maximized entropy, the λ_j^k are optimized using a Minimum Classification Error (MCE) training approach [3], which minimizes the smoothed error measure

$$E(\Lambda) = \sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{i=1}^{I} \varepsilon(\mathbf{c}_i^k, \widetilde{\mathbf{c}}_n^{k_n}) \cdot \frac{p_{\Lambda_k}(\mathbf{c}_i^k|X_n)^\eta}{\sum_l p_{\Lambda_k}(\mathbf{c}_l^k|X_n)^\eta}. \tag{5}$$

Here, the error is summed over the N images in the training corpus, the K classes and I Hough cells providing the votes $v_{i,j}^k$ in each class specific layer $H_k(\mathbf{x})$. For a given training image n, the error function $\varepsilon(\cdot)$ measures the distance of each Hough cell $\mathbf{c}_i^k$ to a given target cell $\widetilde{\mathbf{c}}_n^{k_n}$, which might be the center of the object to be classified in the Hough space layer of the correct class k_n. While this function is realized as a Euclidean distance in the standard DGHT method, the CGHT may additionally employ a fixed inter-class penalty to enforce discrimination between the different class layers. However, since a focused peak in the layers $H_k(\mathbf{x})$ is not the target criterion, a simplified error measure has been applied finally which equally penalizes Hough cells of wrong classes:

$$\varepsilon(\mathbf{c}_i^k, \widetilde{\mathbf{c}}_n^{k_n}) = \begin{cases} 0, & \text{if } k = k_n \\ 1, & \text{otherwise} \end{cases} \tag{6}$$

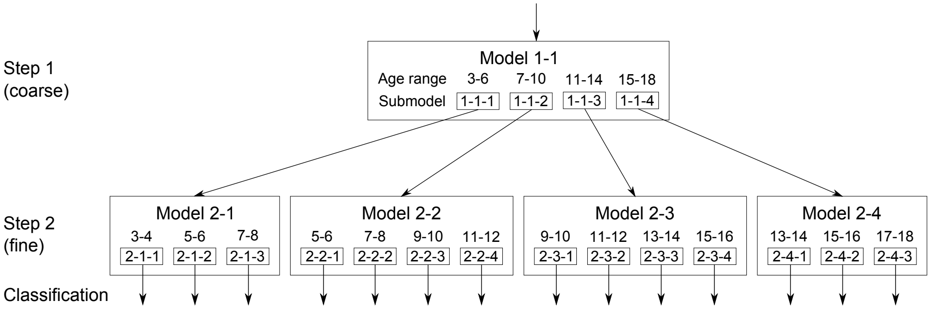

Fig. 2. Illustration of the used coarse-to-fine classification strategy

The second term in (5) is a sigmoidal smoothing function that controls the influence of the best hypotheses on the overall error measure with the parameter η. Consequently, the optimization procedure is adjusting the model weights particularly suppressing votes in the most likely cells of wrong classes.

For optimization of $E(\Lambda)$ over the model point weights $\Lambda = \{\Lambda_1, ..., \Lambda_K\}$, the method of steepest descent is used, which is not assuring a global minimum but has shown significant improvements compared to other weighting strategies in recent experiments [14].

The described optimization procedure assigns individual weights to each model point in M and allows for eliminating model parts with small absolute weights and therefore low influence on the overall classification result. The procedure can be repeated in an iterative manner [14] to gradually enhance the model with structures from training images not yet correctly classified. This has, however, not yet been studied.

3 Bone Age Classification

The proposed BAA procedure is solely based on analyzing eROIs, according to Tanner & Whitehouse [17]. The eROI extraction is done based on the given annotations from the corpus although it could be shown in [5] that a robust automatic localization of those regions can be done automatically by using the DGHT. Therefore, it is planned to combine the two techniques into a fully automated BAA system at a later time.

3.1 Coarse-to-Fine Classification Strategy

The BAA task is characterized by a large object variability in combination with a rather large number of classes. Therefore, it is necessary to restrict the interclass confusion, which can be achieved by utilizing a coarse-to-fine classification strategy comprising of two levels (Fig. 2). The first level classifies a given image into one of four coarse age classes. The second level decides more precisely between age ranges of 2 years only. Due to the significant differences of the epiphyseal shapes between the coarse age ranges of the first level, the respective CGHT models may focus here on global characteristics, such as the size, while the models of the second level represent details to discriminate from neighboring classes of similar age.

It is apparent that in this scenario, misclassifications are more likely to occur at ages close to class boundaries. Therefore, the second level operates with overlapping classes (Fig. 2). For the sake of clarification, let us consider a patient with a skeletal maturity of 12 years who should be assigned to class 11-14 years in the first level. Due to a misclassification during the first level it may occur that instead of the correct class, the class 7-10 years is selected which induces the utilization of Model 2-2 for the second level. This mistake may be corrected due to the incorporation of submodel 11-12 years in Model 2-2.

3.2 Combination of Classifiers

A combined decision based on several eROIs clearly improves the bone age classification performance [8]. This idea has been tested in a first experiment to improve the Model 2-3, which corresponds to the age range 9 to 16 years. To this end, the three epiphyseal plates of the long finger have been analyzed with individually trained CGHT models producing four class-specific Hough spaces $H_k^a(\mathbf{x})$ per joint a for the submodels 2-3-1 to 2-3-4. Afterwards, a normalization step is applied, eliminating any bias from different model point numbers. Finally, the peaks in the normalized Hough spaces are linearly combined for the three joints and a decision is made for the class with the highest combined vote. As an alternative to this heuristic approach, a log linear combination of the classifiers [11] could be used in future attempts which, however, requires additional training data.

4 Data

Training as well as evaluation is performed on images of male patients in an overall age range between 3 and 18 years. The models are trained on non-public data from the University Hospital Aachen and the University Medical Center Schleswig-Holstein. To assure comparability with other studies, evaluation is performed on the public database from the University of Southern California (USC), where each image is assigned an individual age assessment from two radiologists. In order to eliminate debatable cases from our experiments, 156 images with an inter-observer variability of more than 1 year have been removed from the evaluation database as well as images with strong rotation (18 images), atypical positioning (2 images) or unsuitable spacing (5 images). In order to clarify the degree of deviation for these cases, some examples are provided in Figure 3. The remaining 481 images were annotated using the average of both expert readings.

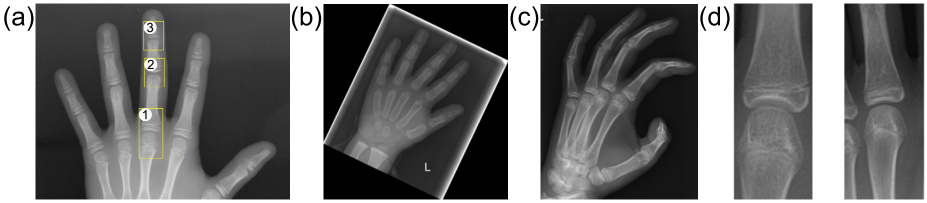

Fig. 3. Illustration of eROIs used in this paper. ① is used for single eROI classification, ② and ③ are additionally used for combined classification. (b-d) Examples of strong rotation (b), atypical positioning (c) and unsuitable spacing (d).

5 Experiments

In order to evaluate the GHT-based classification technique described above, two different experiments have been done. In the first investigation, described in Section 5.1, a single epiphysis has been analyzed to perform the age classification while the second experiment (Section 5.2) analyzes the combination of the classification results of three eROIs.

5.1 Single eROI Classification

In this experiment, the coarse-to-fine strategy (Section 3.1), has been applied to classify the metacarpophalangeal of the middle finger (① in Fig. 3). The training of each submodel in classification level 1, covering an age range of 4 years, could be performed using 84 images. Reducing the age range to 2 years, the amount of training data decreased to only 42 images per submodel. The trained models have been evaluated on 481 images and achieved a mean classification error of 1.11 years.

Figure 4 shows the resulting CGHT submodels of the first classification level. It can be seen that the training procedure has automatically learned reasonable representations of the 4 different age classes. The models have captured size and anatomical differences between the classes while at the same time preserving some level of intra-class shape variability. Studying the learned anatomical structures, it is interesting to note that the fusion of the epiphyseal cartilage can be observed in the submodels 11-14 and 15-18 years while the first two submodels 3-6 and 7-10 years are characterized by a clearly visible gap in this area, emphasized by highly weighted model points shown in red color.

Since a larger number of training images was available for the age range 9-16 years the training was repeated with an amount of 56, instead of 42, images per submodel in classification level 2. In this experiment a slight gain of the mean error to 1.15 years could be observed.

5.2 Combined Classification

The classifier combination described in Section 3.2 has been trained on 56 training images per class from the restricted age range 9 to 16 years and evaluated on

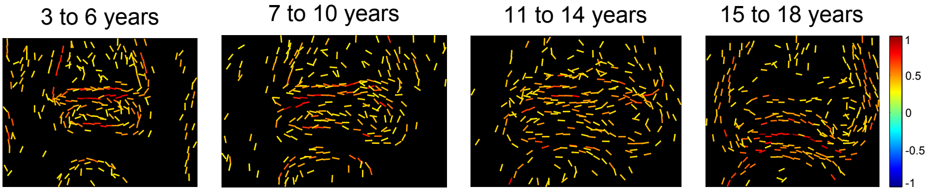

Fig. 4. CGHT submodels of classification level 1. The colors indicate positive and negative model point weighting.

253 test images. A combination of the metacarpophalangeal, proximal and distal interphalangeal of the middle finger (Fig. 3) led to a significant improvement of the mean error from 1.15 years, when using a single classifier, to 0.84 years.

Table 1 gives a comparison of the presented technique with other recently published methods which have been evaluated on the USC database. Note that these publications eliminate a similar amount of problematic images from the evaluation dataset and apply a scaling and orientation normalization prior to the actual classification. Note further, that the age range restriction in our experiments was only necessary due to the shortage of training data. The addressed age range for our combined classification experiments (9-16 years) is considered the most difficult [8] due to the fact that the growth differentials are significantly lower than in younger children.

Table 1. Comparison of BAA methods

Method	Database	Age range	#Images	Mean Error
BoneXpert[12]	Subset of USC[1]	2-17	1083	0.72
Harmsen[8] - six eROIs combined	Subset of USC[1]	0-18	1097	0.83
CGHT - single eROI	Subset of USC[1]	3-18 (male)	481	1.11
CGHT - single eROI	Subset of USC[1]	9-16 (male)	253	1.15
CGHT - three eROIs combined	Subset of USC[1]	9-16 (male)	253	0.84

6 Discussion

In the presented validation experiments, the proposed image classification method has shown good performance. A crucial aspect for the success of the proposed discriminative training technique is, however, the availability of sufficient training images since the data must reflect the large shape variability contained in clinical data. Apart from the general problem of finding comprehensive corpora with annotated hand radiographs three restricting factors have to be addressed when using the CGHT. First, all submodels M_k of a model M should be trained on the same number of images in order to prevent a preference for submodels

[1] Corpus from the University of South California: `http://www.ipilab.org/BAAweb/`

with a larger amount of training data. This is a direct consequence of using the smoothed error measure (Equation 5), which does not compensate for biased data. Therefore, the class with the smallest amount of available images determines the training amount for all classes of the same model. Second, working with the described coarse-to-fine strategy makes the training data situation worse since narrowing the age ranges in the second classification level reduces the available data per class. Due to the training data shortage the installation of a third classification level for an age range of 1 year is currently not feasible. The usage of age ranges of 2 years, however, already induces a best-case expected error of 0.5 years if the data is equally distributed with respect to the age. Third, the method is currently limited to images with restricted orientation and scaling variability, which required the omission of some strong outlier images as shown in Figure 3. A normalization step, based on the results of the prior epiphysis localization procedure, may allow for a better treatment of those cases and will probably additionally improve the classification rate by reducing the contained shape variability to mostly anatomical factors. As a consequence of the three mentioned aspects, it is expected that the observed results can be further improved by (1) increasing the amount of training data, (2) reducing the scaling and orientation variability by a prior normalization step, and (3) introducing further classification levels into the coarse-to-fine framework.

7 Conclusion

This contribution has, for the first time, presented a mathematical description and comprehensive experimental validation of the novel Classifying Generalized Hough Transform (CGHT). It could be shown that this general image classification method can be successfully applied to the task of automatic bone age assessment and achieves comparable results to other state-of-the-art techniques on a public database. In contrast to most other image classification methods, the learned models can be visually interpreted and unveil the most relevant anatomical structures for class discrimination. The basic approach has been extended by a coarse-to-fine classification strategy and a simple classifier combination framework for a combined decision based on several epiphyseal regions of interest. The latter method was shown to significantly improve the mean classification error which confirms the findings of other authors [8].

Since the success of the applied discriminative model training heavily depends upon the amount and quality of available training data, it is expected that further improvements can be achieved by using a larger amount of training images, providing high quality annotations and employing a normalization step prior to the actual classification. Besides these aspects, our future work will consider the combination of a larger number of eROIs, a more sophisticated, e.g. log-linear [11], classifier combination, and the integration of this classification approach with the automatic landmark detection technique based on the Discriminative Generalized Hough Transform (DGHT) [14].

Acknowledgement. The authors would like to thank the University Medical Center Schleswig-Holstein and the University of South California for providing the X-ray images used in this study. Additionally, the authors are grateful to the anonymous reviewers for their valuable comments.

References

1. Aja-Fernández, S., de Luis-Garcıa, R., Martın-Fernandez, M.A., Alberola-López, C.: A computational tw3 classifier for skeletal maturity assessment. A computing with words approach. Journal of Biomedical Informatics 37(2), 99–107 (2004)
2. Ballard, D.: Generalizing the Hough transform to detect arbitrary shapes. Pattern Recognition 13(2), 111–122 (1981)
3. Beyerlein, P.: Discriminative model combination. In: ICASSP, pp. 481–484 (1998)
4. Bocchi, L., Ferrara, F., Nicoletti, I., Valli, G.: An artificial neural network architecture for skeletal age assessment. In: ICIP, pp. 1077–1080 (2003)
5. Brunk, M., Ruppertshofen, H., Schmidt, S., Beyerlein, P., Schramm, H.: Bone age classification using the discriminative generalized hough transform. In: BVM, pp. 284–288 (2011)
6. Chang, C.H., Hsieh, C.W., Jong, T.L., Tiu, C.M.: A fully automatic computerized bone age assessment procedure based on phalange ossification analysis. In: IPPR, pp. 463–468 (2003)
7. Greulich, W.W., Pyle, S.I.: Radiographic atlas of skeletal development of the hand and wrist. The American Journal of the Medical Sciences 238(3), 393 (1959)
8. Harmsen, M., Fischer, B., Schramm, H., Seidl, T., Deserno, T.M.: Support vector machine classification based on correlation prototypes applied to bone age assessment. IEEE Transaction on Information Technology in Biomedicine (2012)
9. Hsieh, C., Jong, T., Chou, Y., Tiu, C.: Computerized geometric features of carpal bone for bone age estimation. Chinese Medical Journal 120(9), 767–770 (2007)
10. Kim, H.-J., Kim, W.-Y.: Computerized bone age assessment using DCT and LDA. In: Gagalowicz, A., Philips, W. (eds.) MIRAGE 2007. LNCS, vol. 4418, pp. 440–448. Springer, Heidelberg (2007)
11. Klakow, D.: Log-linear interpolation of language models. In: ICSLP (1998)
12. Martin, D.D., Deusch, D., Schweizer, R., Binder, G., Thodberg, H.H., Ranke, M.B.: Clinical application of automated greulich-pyle bone age determination in children with short stature. Pediatric Radiology 39(6), 598–607 (2009)
13. Pietka, E., Gertych, A., Pospiech, S., Cao, F., Huang, H., Gilsanz, V.: Computer-assisted bone age assessment: Image preprocessing and epiphyseal/metaphyseal roi extraction. IEEE Transactions on Medical Imaging 20(8), 715–729 (2001)
14. Ruppertshofen, H.: Automatic Modeling of Anatomical Variability for Object Localization in Medical Images. Ph.D. thesis, Otto-von-Guericke University Magdeburg (2013)
15. Ruppertshofen, H., Schramm, H.: The classifying generalized hough transform. German Patent Submission (2011)
16. Ruppertshofen, H., Bülow, T., von Berg, J., Schmidt, S., Beyerlein, P., Salah, Z., Rose, G., Schramm, H.: A multi-dimensional model for localization of highly variable objects. In: SPIE, vol. 8314, p. 88 (2012)
17. Tanner, J., Healy, M., Goldstein, H., Cameron, N.: Assessment of skeletal maturity and prediction of adult height (tw3). WB Saunders, London (2001)
18. Thodberg, H.H., Kreiborg, S., Juul, A., Pedersen, K.D.: The bonexpert method for automated determination of skeletal maturity. IEEE Transactions on Medical Imaging 28(1), 52–66 (2009)