Bone age assessment meets SIFT

Muhammad Kashif[1], Stephan Jonas, Daniel Haak, Thomas M. Deserno

Department of Medical Informatics, RWTH Aachen University, Germany

ABSTRACT

Bone age assessment (BAA) is a method of determining the skeletal maturity and finding the growth disorder in the skeleton of a person. BAA is frequently used in pediatric medicine but also a time-consuming and cumbersome task for a radiologist. Conventionally, the Greulich & Pyle and the Tanner & Whitehouse methods are used for bone age assessment, which are based on visual comparison of left hand radiographs with a standard atlas. We present a novel approach for automated bone age assessment, combining scale invariant feature transform (SIFT) features and support vector machine (SVM) classification. In this approach, (i) data is grouped into 30 classes to represent the age range of 0-18 years, (ii) 14 epiphyseal ROIs are extracted from left hand radiographs, (iii) multi-level image thresholding, using Otsu method, is applied to specify keypoints on bone and osseous tissues of eROIs, (iv) SIFT features are extracted for specified keypoints for each eROI of hand radiograph, and (v) classification is performed using a multi-class extension of SVM. A total of 1101 radiographs of University of Southern California are used in training and testing phases using 5-fold cross-validation. Evaluation is performed for two age ranges (0-18 years and 2-17 years) for comparison with previous work and the commercial product BoneXpert, respectively. Results were improved significantly, where the mean errors of 0.67 years and 0.68 years for the age ranges 0-18 years and 2-17 years, respectively, were obtained. Accuracy of 98.09 %, within the range of two years was achieved.

Keywords: Bone age assessment, Epiphyseal region of interest (eROIs), Scale invariant feature transform (SIFT), Feature extraction, Support vector machine, Classification.

1. INTRODUCTION

Bone age assessment (BAA) is a process of determining the skeletal maturity of a person. For example, the bone age of a child (developmental age of the bones) is assessed based on a radiological examination of skeletal development of the left-hand & wrist and compared to the chronological age. This procedure allows anticipating the adult height, as well as diagnosis and management of endocrine disorder and pediatric syndromes [1]. Moreover, BAA methods can also be used in forensic medicine to determine the age of an unidentified corpse.

Besides medicine, a relevant application of BAA is in social field. According to UNICEF, only half of the children under five years old in the developing world have their births registered. In sub-Saharan Africa and South Asia, about 65% of all births go unregistered. Without documented proof of children's age, they can be recruited into fighting forces, exposed to hazardous forms of work, forced into early marriages, and treated as an adult rather than a child in criminal proceedings. Similarly, asylum seekers without documented proof of their age cannot use the advantage of childhood to obtain a residential permit [2] . Skeletal maturity can help in all these cases to determine the true age of a person.

In radiology, BAA is a frequently used but time-consuming and cumbersome task. Two conventional methods for bone age assessment are the Greulich & Pyle method [3] and the Tanner & Whitehouse method [4]. In Greulich & Pyle method, a radiologist compares all bones of the left hand to the radiographs in a standard atlas and assesses the bone age according to his visual perception. In Tanner & Whitehouse method, only a certain subset of left hand bones is examined. The Greulich & Pyle method is more subjective while the Tanner & Whitehouse method is more complex and both are time consuming methods. Therefore, an automated BAA method is desired in order to assist the radiologists.

Many approaches have already been adopted to automate the process of bone age assessment. Al-Taani et al. presented an automatic BAA approach in 1996 that was based on point distribution model (PDM) of 130 feature points [5]. The

[1] Corresponding Author: Muhammad Kashif, Department of Medical Informatics, RWTH Aachen University, Pauwelsstr. 30, 52057 Aachen, Germany, email: muhammad.kashif@rwth-aachen.de, phone: +49 241 8088790, fax: +49 241 803388790

Medical Imaging 2015: Computer-Aided Diagnosis, edited by Lubomir M. Hadjiiski, Georgia D. Tourassi, Proc. of SPIE Vol. 9414, 941439 · © 2015 SPIE · CCC code: 1605-7422/15/$18 · doi: 10.1117/12.2074572

distal and middle phalanxes of third finger were classified. A set of 120 images was used for evaluation and classification rates for two experiments were 70.5% and 73.7% respectively. Then in 2001, Pietka et al. comprehensively reviewed early approaches for BAA and presented a method for features extraction from left hand radiograph by measuring the gap between metaphyses and diaphyses [6]. In a method by Chang et al. the back propagation of neural networks was applied to train the features of phalanges. In total, 501 female and 416 male hand radiographs were used for evaluation. An error of 1.5 years was reported [7]. Then, Kim and Kim gave the idea of using epiphyseal regions of interest (eROIs). Discrete cosine transform and linear discriminant analysis were applied on nine relevant eROIs segmented from the left hand radiographs. Experiments were performed on a private data set of 93 males and 303 female (total 396) radiographs and an average error of 0.6 years was reported [8].

These approaches were either based on the use of Greulich & Pyle or Tanner & Whitehouse methods or they used private data set that restricts the comparability of BAA approaches. To improve this drawback, A new digital hand atlas was established at the University of Southern California (USC) that is termed as USC hand atlas. It reflects a standard reference data base for evaluation and improves the comparability of automated BAA approaches. First experiment on the image data set was performed by Gertych et al. where fuzzy classifier was applied on carpal bone and phalangeal ROIs [9].

In our previous work, a method based on content-based image retrieval (CBIR) was presented [10, 11], where eROI patches were automatically extracted from the hand radiographs of USC data and similar and labeled patches were retrieved from Image Retrieval and Medical Application (IRMA) framework. Classification was done using k-nearest neighbor (kNN) method. A mean error of 0.97 years, on the age range of 0-18 years, was reported. This method was extended by Harmsen et al. by introducing class prototypes and applying support vector machine (SVM) [12]. A mean error of 0.83 years with variance 0.50 years was achieved. Haak et al. improved this work by replacing SVM with support vector regression (SVR) as a classifier [13]. A mean error of 0.768 years with standard deviation 0.657 years was achieved.

In contrast, the leading commercial product BoneXpert uses active shape models for BAA [14]. BoneXpert obtains a root mean square error of 0.61 years within the range of 2.5-17 years and 2-15 years for boys and girls respectively [15]. Since our previous work applies cross correlation function for obtaining features, and the features were dependent on the similarity measure between test and reference images [11, 12, 13], we here have specified features points using multi-thresholding and the scale invariant feature transform (SIFT) method [16], which has been recently introduced and proven outstanding performance in many registration and retrieval applications [17, 18].

2. MATERIAL AND METHODS

BAA comprises four steps: (i) eROIs extraction, (ii) feature points identification, (iii) SIFT features extraction, and (iv) SVM classification (Fig. 1).

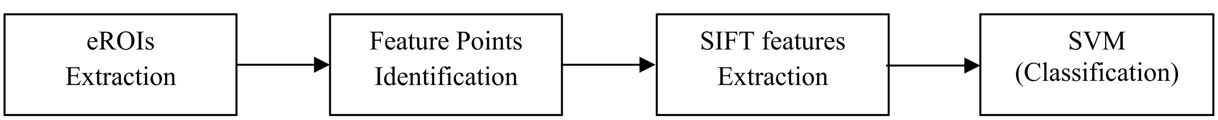

Figure 1: Different processing steps in the proposed method.

2.1 EROIs Extraction

Using the previously presented eROI extraction approach [11], where the user hits the centers of relevant epiphyses for proper epyiphseal centers localization, 14 eROIs from each hand radiograph are extracted and rotated into an upright position. Hence all eROIs are in reference position, disregarding their position in original radiograph (Fig. 2).

2.2 Features Extraction

Features extraction can be divided into two parts, i.e., feature detection and feature description. In feature detection, key points are detected. In feature description, a local image descriptor is calculated for every key point. The SIFT detector is not used in this work. Instead, key points are identified by applying multilevel thresholding. However, the SIFT descriptor is then used to describe the features of the specified keypoints.

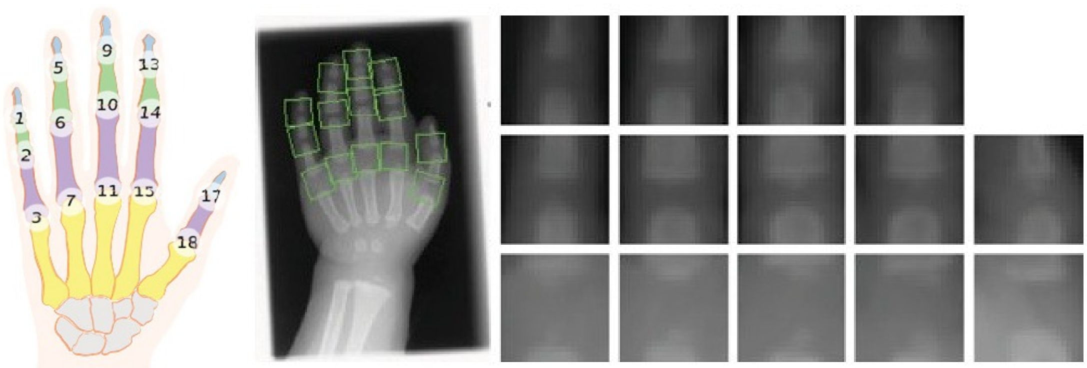

Figure 2: (i) Corresponding eROI numbers, (ii) eROIs are extracted and rotated into upright position [12].

Here, each eROI is quantized into four discrete levels using Multi Otsu method [19, 20]. The two higher levels represent the bone and osseous tissues. Six key point locations are specified on upper, middle and lower parts of the bone. Next, a SIFT descriptor is computed for each of the six key points. Each SIFT descriptor contains 128 numerical features. These 6 descriptors are combined to form a 768 dimensional SIFT feature vector (Fig. 3). One SIFT feature vector represents one eROI.

Figure 3: From left: (i) Original image of an eROI, (ii) RGB segmented Image obtained using multi-thresholding, and (iii) SIFT features descriptors: Green circles and green lines in the circles represent key points and their orientations respectively. The yellow box and arrows show feature descriptors.

2.3 Classification

A support vector machine with polynomial kernel is used for classification purpose [21]. Due to reasonably large number of features and the assumption of non-linear relation between features and classes, we used a three-degree polynomial kernel for non-linear classification. SVM is extended for multiclass classification using one-against-one approach.

2.4 Age Classes

Since the epiphyses of a child gradually ossify in predictable order from birth to the age of 18 years and it allows the data to be grouped according to growth spurts [1]. Hence the data is grouped into 30 classes that represent the whole age range [12] (Tab. 1).

2.5 Validation Experiments

The publically available USC hand atlas is used for validation experiments. It is composed of 1101 left hand radiographs from different ethnics and age. Fourteen eROIs are extracted from each hand radiograph (Figure 2). After extracting eROIs, SIFT feature vector is computed for each eROI. SIFT features are extracted from a rescaled eROI of 48x48 pixel image. This feature vector is then applied for SVM. Analysis is performed using 5-fold cross validation. The parameter values (C = 1 and $\gamma = 0.25$) are used for SVM. The outcome of these experiments is the number of correct classes, age class accuracy, accuracy in 2 years age range, mean error and class error variance.

Table 1: Age Class and Corresponding Age Range (years)

Class	Age range	Class	Age range	Class	Age range
01	0.00 - 0.65	11	2.50 - 2.99	21	09.00 - 09.99
02	0.66 - 0.82	12	3.00 - 3.49	22	10.00 - 10.99
03	0.83 - 0.99	13	3.50 - 3.99	23	11.00 - 11.99
04	1.00 - 1.15	14	4.00 - 4.49	24	12.00 - 12.99
05	1.16 - 1.32	15	4.50 - 4.99	25	13.00 - 13.99
06	1.33 - 1.49	16	5.00 - 5.49	26	14.00 - 14.99
07	1.50 - 1.65	17	5.50 - 5.99	27	15.00 - 15.99
08	1.66 - 1.99	18	6.00 - 6.99	28	16.00 - 16.99
09	2.00 - 2.32	19	7.00 - 7.99	29	17.00 - 17.99
10	2.33 - 2.49	20	8.00 - 8.99	30	18.00 - 99.00

2.6 Implementation

The implementation has been done in MATLAB. SIFT features are computed using VL-SIFT library version 0.9.17, which is developed by VL-Feat and is available open source [22]. Multiclass SVM is implemented using libSVM library version 3.17 that has built-in support for multiclass classification using one-against-one approach [23].

3. RESULTS

Experiments are performed on single region, subset of regions and on a complete set of regions with and without gender specification. Here, variances to errors are given in brackets. Results for individual regions showed age class accuracy from 29 % to 38% and a mean age error from 0.845 (±0.052) to 1.524 (±0.413) years (Tab. 2). For the complete set of regions, the age class accuracy was 41.69% and the mean age error was 0.675 (±0.047) years. It gave the mean errors 0.753 (±0.084) and 0.630 (±0.088) years for male and female radiographs, respectively (Table 3). The subset of 8 regions (3,6,7,9,10,11,15,18) gave the best result (age class accuracy 41%, mean age error 0.669 (±0.051) years) in terms of mean error. Considering best subsets of regions, it gave the mean errors 0.725 (±0.060) and 0.619 (±0.108) years for male and female radiographs, respectively (Tab. 3).

4. DISCUSSION

Region 11 that belongs to the middle finger gave the best results among all individual regions. This is in line to findings of others [4, 11, 12]. Lower error rates are achieved when individual regions are combined to form a subset of regions. For instance, the subset of the best four regions (3, 7, 11, 15) gives an age class accuracy 41.24% and a mean age error of 0.707 (±0.057) years. Consistent and almost best results are obtained when experiments are performed on the complete set of regions. However, few subsets of regions have also achieved the similar or marginally better results than the complete set of regions. In addition, better results are obtained in case of female radiographs as compared to the male radiographs (Tab. 3).

The best result is obtained when male and female images are classified separately on the age range of 0-18 years. Thereby, our approach outperforms all prior published methods on USC data excluding the commercial product BoneXpert (Tab. 4).

Table 2: Experiment outcome for individual regions

Region	Accuracy (%)	Mean Error (y)	Error Variance
1	29.34	1.524	0.413
2	30.70	1.270	0.173
3	33.33	0.952	0.076
5	34.70	1.111	0.113
6	34.97	1.063	0.086
7	36.24	0.888	0.066
9	31.52	1.082	0.086
10	35.06	0.959	0.113
11	38.78	**0.845**	0.052
13	31.06	1.282	0.184
14	33.61	1.159	0.168
15	35.79	0.894	0.157
17	32.70	1.173	0.160
18	33.15	1.022	0.143

Table 3: Experiment outcome for combined and separate male and female radiographs

Age Range	Gender	Regions/[Best Regions]	Accuracy (%)	Accuracy in 2 Yrs Range (%)	Mean Error (y)	Error Variance
0-18	M + F	All	41.69	97.91	0.675	0.047
	Male	All	42.75	97.64	0.753	0.084
	Female	All	40.62	98.18	0.630	0.088
	M + F	[3,6,7,9,10,11,15,18]	40.96	98.09	0.669	0.051
	Male	[3,7,10,11,15,18]	42.57	98.55	0.725	0.060
	Female	[3,5,6,7,9,10,11,13,14,15,17,18]	40.98	97.45	0.619	0.108

The commercial product BoneXpert was evaluated over a range of 2.5-17 years and 2-15 years for boys and girls, respectively. BoneXpert reached a RMS error 0.61 years [15]. We considered age range 2-17 years for both boys and girls, and reached the mean and the RMS errors 0.68 years and 0.7166 years respectively. This is much better compared to prior publications but still slightly inferior to BoneXpert. However, since the evaluation set between BoneXpert and our approach differs, a final conclusion cannot be deducted.

In this work, a semi-automatic approach for eROI extraction was used [11], where the user hits the centre of relevant epiphyses for the localization. However, Hahmann et al. recently presented a new approach for the localization of epiphyseal regions of interest (eROIs) [24]. In this approach, Discriminative Generalized Hough Transform (DGHT) was used, in conjunction with some simple geometrical constraints, to locate 12 epiphyseal ROIs in left hand radiographs. This approach can be adopted in future work for fully automatic bone age assessment.

Table 4: Comparison to Published Results - Mean Error

Age Range	Used Gender	All (14) Regions	Best Regions	Region Numbers	Harmsen et al.	Fischer et al.	BoneXpert
0 - 18	Yes	0.6746	0.6693	8	0.8320	-	-
0 - 18	No	0.8456	0.8348	13	0.9637	0.97	-
2 - 17	Yes	0.7029	0.6834 0.7166(RMS)	9	0.8265 0.9887(RMS)	-	0.61(RMS)
2 - 17	No	0.9486	0.9486	All	0.9850	-	-

5. CONCLUSION

We have presented an effective and robust method for automatic bone age assessment. Age class accuracy 41% is achieved and the accuracy within the range of two years is 98.09%. Most of the misclassification lies within the range of one or two classes. In comparison, the differences between two expert readings in the USC data reaches up to 2.5 years [12]. Therefore, 98.09 % accuracy within the range of 2 years is a good performance.

The presented approach is novel and easy to implement, where SIFT features are extracted directly from the eROIs of left hand radiographs and classification is performed using SVM. In contrast to previous work, it does not require comparing given eROI with all 1101 eROIs stored in database or with 30 prototype eROIs as suggested by Fischer et al. [11] and Harmsen et al. [12] respectively. It does not need semantic features or atlases unlike method by Pietka et al., or the convential methods by Greulich & Pyle or Tanner & Whitehouse.

6. REFERENCES

[1] Gilsanz V, Ratib O. Hand Bone Age: A digital atlas of skeletal maturity. Berlin, Germany: Springer-Verlag, 2005.

[2] Smith T, Brownlees L. Age assessment practices: a literature review & annotated bibliography. United Nations Children’s Fund (UNICEF), New York 2011.

[3] Greulich WW, Pyle SI. Radiographic atlas of skeletal development of hand wrist. Stanford, CA: Stanford Univ. Press, 1971.

[4] Tanner JM, Whitehouse RH. Assessment of skeletal maturity and prediction of adult height (TW2 method). London, U.K: Academic, 1975.

[5] Al-Taani AT, Ricketts IW, Cairns AY. Classification of hand bones for bone age assessment. Proc IEEE ICECS. 1996;2:1088–91.

[6] Pietka E, Gertych A, Pospiech S, Cao F, Huang HK. Computer assisted bone age assessment: image preprocessing and epiphyseal/ metaphyseal ROI extraction. IEEE Trans Med Imaging 2001; 20(8):715–29.

[7] Chang CH, Hsieh CW, Jong TL, Tiu CM. A fully automatic computerized bone age assessment procedure based on phalange ossification analysis. Proc IPPR. 2003;16:463–68.

[8] Kim HJ, Kim YW. Computerized bone age assessment using DCT and LDA. Proc ICC Vis/CGCT. 2007; 4418:440–48.

[9] Gertych A, Zhang A, Sayre J, Pospiech-Kurkowska S, Huang H. Bone age assessment of children using a digital hand atlas. Comput Med Imaging Graph 2007;31(4-5):322–31.

[10] Lehmann TM, Güld MO, Thies C, Fischer B, Spitzer K, Keysers D, et al. Content-based image retrieval in medical applications. Methods Inf Med 2004;43(4):354-61.

[11] Fischer B, Welter P, Günther RW, Deserno TM. Web-based bone age assessment by content-based image retrieval for case-based reasoning. Int J Comput Assist Radiol Surg 2012;7:389–99.
[12] Harmsen M, Fischer B, Schramm H, Seidl T, Deserno TM. Support vector machine classification based on correlation prototypes applied to bone age assessment. IEEE J Biomed Health Inform 2013;17(1).
[13] Haak D, Simon H, Yu J, Harmsen M, Deserno TM. Bone age assessment using support vector machine regression. Meinzer HP, Deserno TM, Handels H, Tolxdorff T (Ed) Bildverarbeitung für die Medizin 2013. Springer-Verlag, Berlin 2013;164-9.
[14] Thodberg HH, Kreiborg S, Juul A, Pedersen KD. The bonexpert method for automated determination of skeletal maturity. IEEE Trans Med Imaging 2009; 28(1): 52–66.
[15] Thodberg HH, Sävendahl L. Validation and reference values of automated bone age determination for four ethnicities. Acad Radiol 2010;17(11):1425–32.
[16] Lowe DG. Distinctive image features from scale invariant keypoints. Int J Comput Vis 2004;60(2):91-110.
[17] Seok J, Hyun B, Kasa-Vubu J, Girard A. Automated classification system for bone age X-ray images. IEEE ICSMC, Seoul, Korea. 2012.
[18] Kaur B, Jindal S. Content based image retrieval with graphical processing unit. Proc Int Conf Recent Trends Inform Telecom Comput 2014;364-73.
[19] Otsu N. A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 1979;9(1):62–66.
[20] Liao P, Chen T, Chung P. A Fast Algorithm for Multilevel Thresholding, J Inf Sci Eng 2001;17(5):713–27.
[21] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(3):273-97
[22] Vedaldi A, Fulkerson B. Vlfeat: an open and portable library of computer vision algorithms. MM '10 Proc ICM. 2010;1469-72.
[23] Chang CC, Lin CJ. Libsvm: A library for support vector machines. ACM Trans Intell Syst Technol. 2011; 2(3):1-27.
[24] Hahmann F, Böer G, Deserno TM, Schramm H. Epiphyses localization for bone age assessment using the discriminative generalized Hough transform. Bildverarbeitung für die Medizin 2014. Informatik Aktuell. Springer-Verlag, Berlin 2014;66-71