Evaluating different structures for predicting skeletal maturity using statistical appearance models

Steve A. Adeshina,* Timothy F. Cootes† and Judith E. Adams‡

Imaging Science and Biomedical Engineering, The University of Manchester, Manchester UK.

Abstract. We compare the utility of models of different structures in the hand for predicting skeletal maturity in young people. Bone age assessment is important for diagnosing and monitoring growth disorders. Statistical models of bone shape and appearance have been shown to be useful for estimating skeletal maturity. In this work we investigate the effect of the choice of region to model on the prediction performance. By analysing the performance on a dataset of 170 digitized radiographs of normal children we show that improvements in accuracy can be achieved by using models of the joint complexes and bones constructed by automatic registration, compared to ones built from manual annotation. We also investigate how sets of joint complexes and bones can best be combined to estimate overall skeletal age. Our results show that the best predictions are obtained from 13 RUS complexes and their combination into eight bone complexes. The mean absolute prediction errors of 0.80 (13 joint complexes) , 0.81(eight bone complexes) years for females and 0.93, 0.94 years for males respectively, demonstrates that similar results are obtained whether we use many local models of individual joints, or a smaller number of models of groups of bones.

1 Introduction

A widely used method of estimating skeletal maturity is to examine the bones and joints of the non-dominant hand in a radiograph. A significant difference between the bone age and the actual age of a child is an indication of growth abnormalities. The main methods used in clinical practice are those of Greulich and Pyle(GP) [5] and Tanner and Whitehouse(TW2/3) [12]. The GP method involves comparing the whole image with an atlas, while the TW2 method involves scoring each of a number of different bone complexes - it is less subjective but significantly slower. Both methods are subject to inter- and intra- observer variability. While there is no controversy as to points where growth occurs, there is a difference of opinion as to coverage areas and specific bones to observe while estimating skeletal maturity. The GP method observes 28 growth points while the TW3 observes 13 growth points.

There have been many attempts to automate the bone age assessment procedure. These range from classical image analysis methods [8, 10], machine learning techniques [1, 6, 15], and model based methods [7, 9, 16]. The majority of these studies are restricted to isolated areas in the hand. This process of reducing skeletal maturity to a single isolated bone complex is not entirely wrong, but it loses vital complimentary information available from the estimation of other growth complexes [12]. It is indeed desirable to have as many bones as are biologically important for analysis of maturity information to get a good estimation of skeletal maturity. There are a range of possible choices of models, from using single models of the whole hand, to averaging the predictions of local models of individual joints and bones. This paper investigates the effect of the choice of region to model on the prediction performance. The most closely related work to that presented here is that of Thodberg *et al.* [14], who showed how Active Appearance Models [3, 13] can be used to locate the bones of the hand and how the parameters of the associated appearance models can be combined with other texture measures to predict skeletal age. Whereas Thodberg *et al.* used models of the individual bones, we build local appearance models of the regions around the bones and joints and compare with models of individual bones containing the complexes. We perform a set of systematic experiments to investigate which structures are most significant for prediction of skeletal age.

In our approach we manually annotated a set of points on the bones in a set of digitized radiographs of children's hands. We constructed shape and appearance models of each of 20 joints and bone complexes. We also construct combined bone models of bones using equivalent combination bones and complexes. The parameters of the resulting models were used in a linear regressor to predict the chronological age of the child. The best results were obtained by averaging over 13 bone complexes and 8 equivalent bones. The result of the average of seven Carpal bones was also compared with that of a single model of the seven bones. In the following we describe the key components of the work and present quantitative results.

---

*steve.adeshina@postgrad.manchester.ac.uk

†t.cootes@manchester.ac.uk

‡judith.adams@manchester.ac.uk

2 Methods

2.1 Data Set

We have access to a database of radiographs of the non-dominant hand of normally developing children. The children were enrolled on a bone ageing study at the University of Manchester. Their ages ranged between 5 years and 20 years. In the following work we used a subset of 170 (87 male and 83 female) digitized radiographs of normal children.

2.2 Construction of Statistical Appearance Models

Statistical appearance models [3, 4] were generated by combining a model of shape variation with a model of texture variation. Each radiograph was manually annotated with 330 points around important structures (Figure 1a). Statistical models of shape and texture (intensities in the reference frame) were constructed by applying Principal Component Analysis (PCA) to the resulting annotations, leading to linear models of the form

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}_s \mathbf{b}_s \quad \mathbf{g} = \bar{\mathbf{g}} + \mathbf{P}_g \mathbf{b}_g \tag{1}$$

where $\bar{\mathbf{x}}$ is the mean shape, $\bar{\mathbf{g}}$ is the mean texture, $\mathbf{P}_s, \mathbf{P}_g$ are the main modes of shape and texture variation and $\mathbf{b}_s, \mathbf{b}_g$ are the shape and texture model parameter vectors. Combining the shape and texture models gives a combined appearance model of the form

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Q}_s \mathbf{c} \quad \mathbf{g} = \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{c} \tag{2}$$

where $\mathbf{Q}_s$, $\mathbf{Q}_g$ are matrices describing the modes of variation derived from the training set and $\mathbf{c}$ is a combined vector of appearance parameters controlling both shape and texture.

2.3 Groupwise registration

The manual annotation only uses a few points for each local bone complex model, so does not represent details of the bone shape. To improve the density of the correspondences we applied a 'groupwise' non-rigid registration algorithm, similar to that in [2, 11], initialised with the manual points. For each structure we defined a dense triangulated mesh on one image, then used the manual annotation to propagate this to the other images using thin-plate spline interpolation. We then estimated the mean shape and texture and applied a non-rigid registration approach to improve the correspondence between each image and the mean. The process is repeated until convergence, leading to an accurate, dense correspondence across the set. Models of shape, texture and appearance were then constructed from the resulting points. We built a set of models of 7 local carpal models and a combined one, 13 RUS local model and 8 combined bone models of bones and growth complexes known to be important in estimating bone age [12] – see Figure 1b. Figure 2 shows examples of modes of variation of the appearance models for four of the local models and the mode of variation of the bone model that combines them. The figure also shows separate and combined models for Carpal bones.

2.4 Estimation of skeletal maturity

Given the appearance models we can compute shape, texture and appearance parameter vectors for each structure on each image. We performed experiments that showed that appearance parameters correlate with age better than either shape or texture parameters alone. We use classical linear regression of the form:

$$A = \mathbf{w}^T \mathbf{p} + A_0 \tag{3}$$

where A is the predicted age, $\mathbf{w}$ is a vector of weights, $\mathbf{p}$ is the parameter vector and A_0 is the intercept constant. In the following we describe experiments comparing the performance of different models and combinations of models.

3 Experiments

A total of 170 images (87 male children and 83 female children) were annotated with 330 points (Figure 1a). Shape, texture and appearance models were built based on the results of the (manually initialised) automatic registration. Images of males and females were pooled to create the models.

For each model we computed the shape, texture and appearance $\mathbf{c}$ parameters for every image. We then evaluated the utility of linear age prediction models using a Leave-One-Out (LOO) paradigm. We trained linear regressors to predict age on all but one image, then tested the prediction on the left-out image. Since male and female children are known
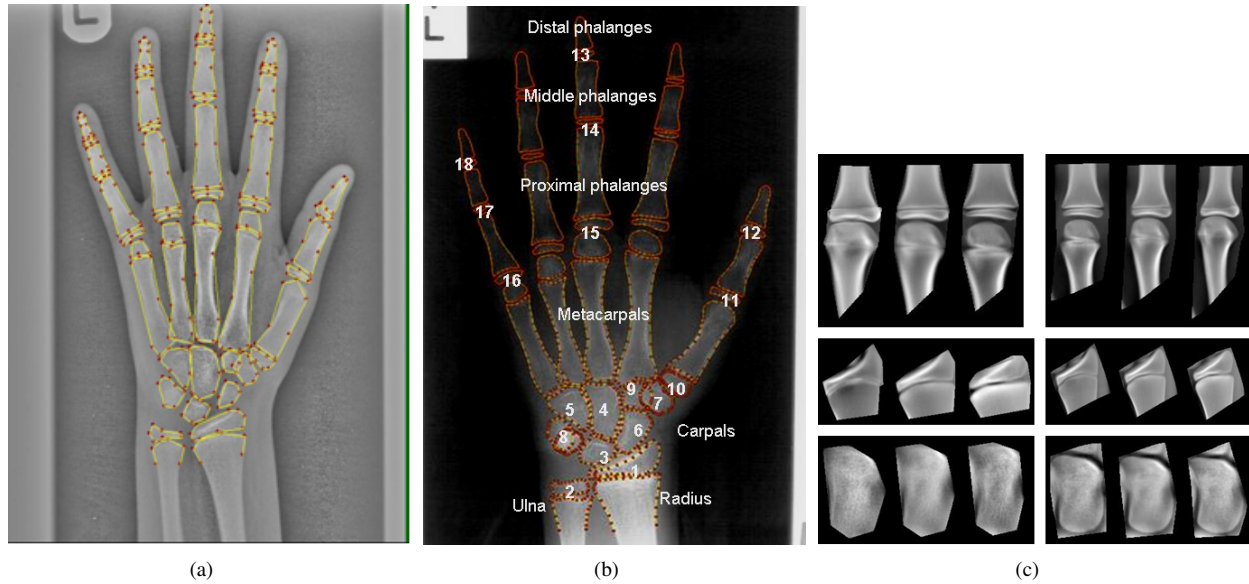
Figure 1. (a) Radiograph with manually annotated points. (b) Skeletal maturity growth points based on TW method. **RUS bones:** Radius(1), Ulna(2), Metacarpal I, III, V, Proximal phalanges I, III, V (10,15,16) , Middle phalanges III, V (14,17), Distal phalanges I, III, V (12,13,18); **Carpal bones:** Capitate(4), Hamate(5), Triquetral(8), Lunate(3), Scaphoid (6), Trapezium(7) and Trapezoid(9). (c) The first mode appearance variation of models from three joint complexes (Metacarpal III, Radius and Capitate) from manual markup(left) and after automatic registration (right).
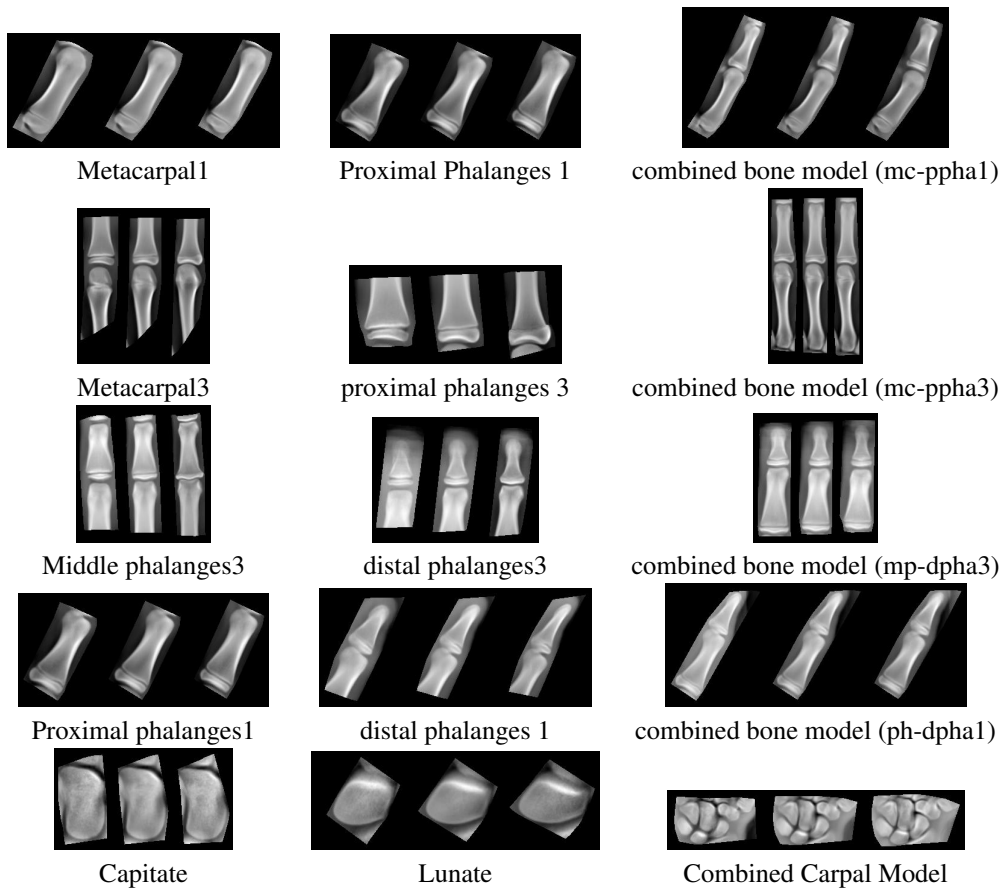

Figure 2. The first mode appearance variation of models from five joint complexes with their corresponding combined bone model.

to develop at different rates, different regressor models were used for the male and the female sets. We evaluated performance using the mean absolute error between prediction and chronological age. This is a more robust measure than using the RMS error, which is prone to corruption by outliers.

We performed initial experiments based on established growth complex considering the 13 RUS bone complexes and 20 Bone complexes. Table 1 compares the average performance of single local appearance models build with and without automatic registration for 20 growth complexes. It demonstrates that overall the registration improves the quality of the predictions.

	Female	Male
Models from manual markup	1.47±0.08	1.26±0.07
Automatic registration	1.35±0.08	1.20±0.04

Table 1. Average performance of single bone complex models, with and without automatic registration (Mean absolute error in years).

Table 2 shows the prediction errors for the average performance of single local models, Overall this shows that the predictions based on appearance parameters give the best performance, though the improvement is not clearly significant with this size of dataset. The texture parameters in female show a marginally better performance, this may be due to the large number of older female children in our dataset. We intend to further explore the possibilities from texture parameters in future.

	Female			Male		
	Shape	Tex.	App.	Shape	Tex.	App.
Average - RUS13	1.35±0.07	1.18±0.07	1.18±0.05	1.23±0.03	1.23±0.05	1.16±0.05
Average - 20 Complexes	1.48±0.07	1.31±0.07	1.35±0.08	1.32±0.05	1.24±0.04	1.20±0.04

Table 2. Average performance of local models - Mean absolute predictions error (years) of RUS13 complexes and 20 bone complexes.

We performed experiments comparing a single model of all carpals bones with the prediction based on the average of that from each individual carpal bone model. See row 5 of Figure 2 for the image of the models. Table 4 (rows 2 and 3) shows the resulting prediction errors. The difference between the individual bones average mean absolute errors is not significant. However the errors are comparatively high. This justifies the exclusion of carpal bones from skeletal maturity estimation in line with recent medical studies [12].

We performed further experiments to compare the performance of individual bone complexes (joints) with combined model of bones which combine two or more complexes. We use the 13 RUS bone complexes shown in Figure 1b. Proximal joints were combined within the 13 RUS Complexes to give a total of eight models. Eight examples of RUS 13 bone complexes and how they are constituted into four bone models are shown in Figure 2 (first 4 rows). Additional combinations of complexes not shown are metacarpal5 + proximal phalanges5 (mc-ppha5) and middle phalanges 5 + distal phalanges 5 (mp-dpha5). Six combined bone models plus Radius and Ulna are compared with the 13 RUS complexes. The mean absolute errors of the combined bone models versus the average error of the two equivalent complexes are shown in Table 3 (rows 6 and 7). The results show that the combined bone models are slightly more effective than the individual based models, but the difference is not statistically significant.

	Female		Male	
	Combined model	Equivalent average .	Combined model	Equivalent average .
mc-ppha1	1.01±0.09	1.18±0.12	1.00±0.09	1.12±0.09
mc-ppha3	1.18 ±0.10	1.00±0.10	1.16±0.09	0.95±0.08
mc-ppha5	0.98±0.11	1.23±0.10	1.15±0.11	1.15±0.10
mp-dpha3	1.26±0.13	1.52±0.14	1.28±0.09	1.34±0.11
mp-dpha5	1.27±0.12	1.11±0.10	1.07±0.10	1.31±0.11
ph-dpha1	1.10±0.10	1.14±0.11	1.21±0.10	1.14±0.10
Average error	1.13±0.11	1.19±0.11	1.14±0.10	1.17±0.10

Table 3. Mean absolute predictions error (years) for six combined bone models and the equivalent average of 2 single RUS Complex.

Overall age prediction can be improved by averaging the ages estimated from each local bone model over the set ($A_\mu = \frac{1}{n} \sum_{i=1}^{N} A_i$, where A_i is the prediction from the ith local model). We computed the average from predictions

of eight combine bone models versus 13 RUS complexes and results are shown in Table 4 (rows 4 and 5).

	Female	Male
Average predictions from 20 local models (RUS and Carpals)	0.92 ±0.09	0.92 ±0.07
Average predictions from individual Carpal bones' model	1.30±0.10	1.00±0.08
Predictions from a combined carpal model	1.32±0.14	1.16±0.09
Average predictions from 8 Combined bone models	0.81 ±0.08	0.94 ±0.08
Average predictions from 13 RUS local models	0.80 ±0.09	0.93±0.08

Table 4. Mean absolute predictions error (years) using average predictions of constituent bone models.

4 Discussion and Conclusion

This work confirms earlier work [14] that good predictions of chronological age can be made using simple linear predictors based on the parameters of appearance models of bones and joint complexes of the hand. The mean absolute prediction errors of 0.80, 0.81 years for males and 0.93, 0.94 years for females are encouraging. The average of predictions from Table 4 shows a small difference between between the 8 combined models and the 13 RUS joint complexes. The results of a combined Carpals' model versus average of single Carpal models shown in Table 4 also shows a small difference. This suggests that it may not matter whether a combined bone models or single complex models are used. The mean absolute prediction errors correspond to root mean square errors of about 1.1 and 1.2 years. Thodberg *et al.* [14] describe a system based on appearance models of the bones, which achieves an RMS error of 0.87 years between predicted age and chronological age on a large dataset, using a set of linear predictors for each bone (one for each of several age ranges). We are currently extending our dataset, and will explore whether non-linear models and multiple predictors can further improve our results.

References

1. L. Bocchi, F. Ferrara, I. Nicoletti, and G. Valli. An artificial neural network architecture for skeletal age assessment. In *Proc. 2003 International Conference on Image Processing*, volume 1, pages 1077–1080, Sept. 2003.
2. T. Cootes, C. Twining, V.Petrović, R.Schestowitz, and C. Taylor. Groupwise construction of appearance models using piece-wise affine deformations. In *16th British Machine Vision Conference*, volume 2, pages 879–888, 2005.
3. T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.
4. T. F. Cootes, C. J. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, Jan. 1995.
5. W. W. Greulich and S. I. Pyle. *Radiographic Atlas of Skeletal Development of Hand Wrist*. Palo Alto, CA: Stanford Univ. Press, 1971.
6. T. S. Levitt, M. W. Hedgcock, D. N. Vosky, and V. M. Shadle. Model-based prediction of phalanx radiograph boundaries. In *Proc. SPIE Medical Imaging*, volume 1898, pages 670–678, 1993.
7. S. Mahmoodi, B. S. Sharif, E. G. Chester, J. P. Owen, and R. Lee. Skeletal growth estimation using radiographic image processing and analysis. *IEEE Transactions on Information Technology in Biomedicine*, 4(4):292 – 297, Dec. 2000.
8. D. J. Michael and A. C. Nelson. Handx: a model-based system for automatic segmentation of bones from digital hand radiographs. *IEEE Transactions on Medical Imaging*, 8(1):64–69, 1989.
9. M. Niemeijer, B. van Ginneken, C. Maas, F. Beek, and M. Viergever. Assessing the skeletal age from a hand radiograph: automating the tanner-whitehouse method. In *Proceedings of SPIE – Volume 5032*, pages 1197–1205, May 2003.
10. E. Pietka, A. Gertych, S. Pospiech, F. Cao, H. K. Huang, and V. Gilsanz. Computer-assisted bone age assessment: image preprocessing and epiphyseal/metaphyseal roi extraction. *IEEE Transactions on Medical Imaging*, 20(8):715–729, 2001.
11. D. Rueckert, A. Frangi, and J. Schnabel. Automatic construction of 3D statistical deformation models using non-rigid registration. In *MICCAI*, pages 77–84, 2001.
12. J. M. Tanner, R. H. Whitehouse, W. A. Marshall, M. R. Healy, and H. Goldstein. *Skeletal Maturity and Prediction of Adult Height (TW2 Method)*. New York, NY: Academic, 1975.
13. H. Thodberg. Hands-on experience with active appearance models. In *SPIE Medical Imaging*, Feb. 2002.
14. H. Thodberg, S. Kreiborg, A. Juul, and K. Pedersen. The bonexpert method for automated determination of skeletal maturity. *Medical Imaging, IEEE Transactions on*, 1(1309):52–66, 2009.
15. A. Tristan-Vega and J. Arribas. A radius and ulna tw3 bone age assessment system. *Biomedical Engineering, IEEE Transactions on*, 55(5):1463–1476, 2008.
16. F. Vogelsang, M. Kohnen, H. Schneider, F. Weiler, M. W. Kilbinger, B. B. Wein, and G. R. W. Skeletal maturity determination from hand radiograph by model-based analysis. In *SPIE Medical Imaging*, pages 294–305, Feb 2000.