Automatic Annotation of Radiographs using Random Forest Regression Voting for Building Statistical Models for Skeletal Maturity

Steve A. Adeshina, Claudia Lindner, and Timothy F. Cootes.

Abstract—Statistical Models of Shape and Appearance require annotation of the bones of the hand of children and young adults. Due to very large variation in the shape and appearance of these bones, automatic annotation is particularly challenging. Statistical Models of Shape and Appearance have been found useful in several medical image analysis and other applications. In this work we locate sparse points on the bones of the hand with an automatic system which uses a Constrained Local Model with Random Forest Regression Voting. These sparse points were then used as input to a groupwise registration algorithm. The control point of the groupwise algorithm can then be used to propagate manually annotated points to other images. The resulting propagation may be used to build Statistical models. By analysing performance on dataset of 537 digitized images of normal children we achieved an automatic annotation accuracy of a mean point to curve error of $0.94mm \pm 0.01$ and a median error $0.92mm$.

Index Terms—Skeletal maturity assessment, Random Forest regression voting, Constrained Local Models, Random Forest in a Constrained Local Model framework (RV-CLM).

1 Introduction

The annotation of images is very important in building Statistical models and in medical diagnosis. In many applications though, these annotations are carried out manually [9]. This is indeed a very tedious exercise as it entails putting hundreds of landmark points across the images. With these manual points, it is indeed almost impossible to ensure that each manual point actually correspond to the equivalent manual points on other images. In addition typical annotations may take several hundreds of man hours to achieve the desired purpose. The annotation of images for building statistical models and other models is usually done off-line in a supervised manner. This definitely will delay the time to production and it is indeed limiting the number of images being used for building models thereby reducing the generalization ability of the models. Statistical Models have been found useful in several medical image analysis, facial image analysis, verification and recognition and several other applications.

In this work, we extended [1] [3] to deal with very large variation. We did this by dividing the data-set into three groups and running the algorithm for each group. In addition we provided methods for placing as many as 2,797 landmarks on several images. We then obtained annotations that is usable for building models that requires corresponding points and in others applications in classical medicine. This work also extend our earlier work [2] where we used a similar technique to segment the carpal area of the bones of the hand. Whereas in that work we segmented a single region of interest i.e the Carpal area bones, this work applies a similar method to segment the 28 bones of the hand.

- *Steve Adeshina is with the Department of Electrical and Electronics Engineering, Nile University of Nigeria, Abuja, Nigeria. E-mail: steve.adeshina@nileuniversity.edu.ng*
- *Claudia Lindner and Timothy Cootes are with Centre for Imaging Sciences, The University of Manchester, U.K*

2 Related Work

For many years there had been several efforts to automate this process. Recently there

has been considerable research into automated methods of achieving correspondence, such as from boundaries (eg [10]) in 2D or surfaces in 3D (eg [11]), or more generally by directly registering images using non-rigid registration methods [15] or 'groupwise' techniques [4], [6], [20], [21]. Other methods employ Random Forest regression voting [5], [8].

In our earlier work [1], we dealt with the problem of registering images of objects with considerable shape variation and multiple similar sub-parts, for instance radiographs of the human hand, such as those in Figure 1. This was done by finding initialization points for a groupwise registration using a semi-automatic method. The frontiers of this work was extended further by Zhang and Cootes with a fully automatic method [21] to locate a number of sparse points in images of large variation. Further work had also been done in this respect by Cootes *et al.*[8], [18] using Random Forest regression for finding optimal points with Statistical Shape models.

Random Forests (RF) [8] describe an ensemble of decision trees trained independently on a randomized selection of features. They have been shown to be effective in a range of classification and regression problems [8]. Gall and Lempitsky show in Hough Forests [16] that objects can be accurately located using RF regressors to predict the position of a point relative to the sampled region, then running the regressors over the region and accumulating votes for the likely position. Cootes *et al.*[8] show how this method (RF) can be combined with Statistical Shape model to accurately segment a variety of complex dataset.

Lindner *et al.*[19] using the methods of [8], applied RF regression in a Constrained Local Model(CLM) to accurately segment the femur in a pelvic radiograph. Following the approach of Lindner *et al.*[19] we apply the method of Cootes *et al.*to locate salient points in hand radiograph by applying RF regression in a Constrained Local Model (CLM) framework to vote for optimal position of each model point. This is done by running feature detectors independently to generate a response image for each point. A shape model is used to find the best combination of points.

Donner *et al.*presented an impressive work in [13] where they use a top-down image patch regression to perform a fast anatomical structure localization. They obtained very impressive result though they tested their model on 20 images. This work extends the work of Lindner *et al.*[18] to full annotation of Radiographic image of Children and young adults for the purposes of determining skeletal maturity. Figure 1 shows a typical radiograph showing the growth points and local models build from full annotation.

Whereas most of these methods got sub-millimeter results in locating sparse points on images of the Radiograph of the hand and other data-sets, they only located sparse points for initialising groupwise registration and whereas the algorithms were able to deal with variation in the radiographs of the hand, they are all limited in dealing with the very large variation that is often required when skeletal maturity is the goal. Most of the work cited above did not cover the age range required for skeletal maturity, so their effectiveness for this purpose is limited. In addition most of the methods also stopped at finding sparse points for dense registration.

3 METHODS

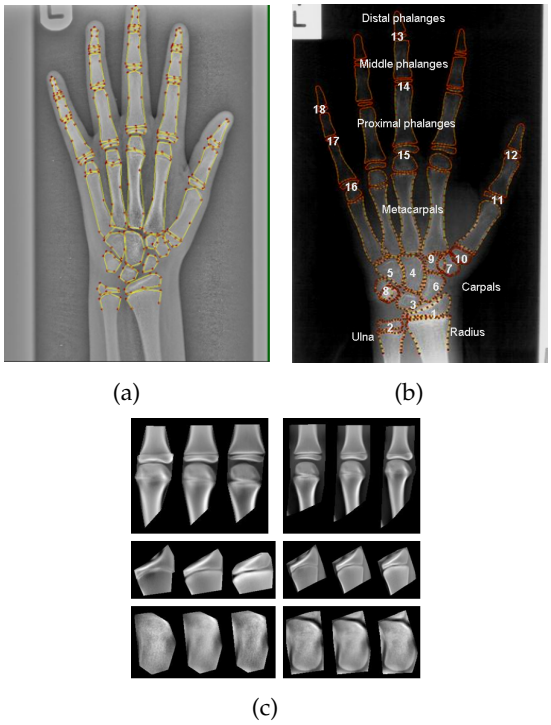
3.1 Data Set

We have access to a database of radiographs of the non-dominant hand of normally developing children. The data is being collect for a different Bone Ageing project at a University. [1]. Their ages ranged between 5 years and 19 years. In the following work we used a subset of 536 digitized radiographs of normal children. The images were divided into three groups of ages 5-7 (63 images), ages 7-13 (284 images) and ages 14-18 (189 images) years.

3.2 Constrained Local Models (CLM)

We segment the region around the carpal bones using Constrained Local Models (CLMs) of [12]. We follow the method of Cootes *et al.*[8]. CLM combines global constraints with local

1. Bone ageing program at the University of Manchester



(a) (b)

(c)

Fig. 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).

models to segment an object form an image. This it does by considering the pattern of intensities. Based on a number of landmark points outlining the contour of the object in a set of images, we train a statistical shape model by applying PCA to the aligned shapes [9]. This yields a linear model model of shape variation which represent the position of each landmark point using $\mathbf{x}_i = T_o(\bar{\mathbf{x}}_i + \mathbf{P}_i\mathbf{b} + \mathbf{r})$ where $\bar{\mathbf{x}}_i$ gives the mean in the reference frame, $\mathbf{P}_i$ is a set of modes of variation, $\mathbf{b}$ are the shape model parameters, $\mathbf{r}$ allows small deviation from the model, and T_o apply a global transformation with parameters θ. To match a CLM to a new image we seek the shape and pose parameters $\mathbf{p} = \{\mathbf{b}, \theta\}$, which optimize the fit to the model. These parameters seek optimize $\sum_{i=1}^{n} \mathbf{R}_i(T_\theta(\bar{\mathbf{x}}_i + \mathbf{P}_i\mathbf{b} + \mathbf{r}))$ where at every position i, $\mathbf{R}_i$ is the stored value of the quality of fit at every position representing the similarity between template texture at this landmark leaned from the model and the texture at the same position.

3.3 Voting with Random Forest (RF) Regressors

We applied RF similar to the Hough Forest approach of [16],but we did not require voting to be dependent on a class labels. We adopted the method used in [8] and [19] where votes are gathered from regions around every point. During training a set of decision trees (a Random Forest) is trained so that each predicts the displacement from a given image patch to the target point. When searching, each tree is scanned over nearby patches in a grid, and produces a vote for where the target point is. All votes for point i are combined in an array, $\mathbf{R}_i$. For further details see [8] and [19] [18]

3.4 Construction of Statistical Appearance Models

Statistical appearance models (SAM) [7] were generated by combining a model of shape variation with a model of texture variation. Each radiograph was automatically annotated with points around important structures. 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

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

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

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

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.

3.5 Dense Correspondence

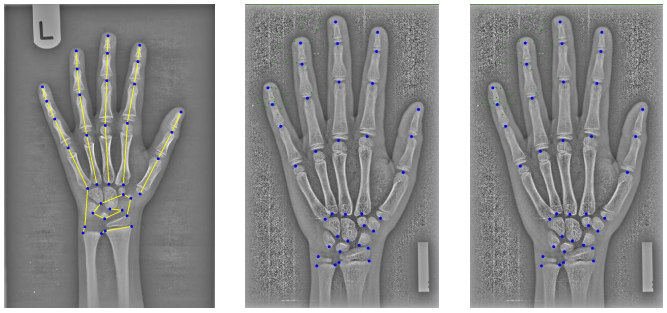
At convergence we obtain a model of parts and geometry, together with a sparse annotation of every image in the training set. The centres of each part region define correspondences.We use these to initialise a groupwise registration. We Place a dense mesh of control points on the first image, use a thin-plate spline based on the sparse annotation to propogate these points to all other images. We then compute the mean shape and warp each example into the mean. Furthermore we perform non-rigid registration [6] to modify the control points on each image to best match to the mean. Finally we re-compute the mean and iterate

4 Experiments and Results

4.1 Finding the initial 37 points using RV-CLM

In order to locate 37 sparse points automatically the model was initialized by automatically detecting nine points (four around the palm and one at the base of each finger). This was achieved by first detecting the object in the image and initialising two reference points within the detected bounding box as in [18]. We used these two points to initialise the mean shape of a 9- point RFRV-CLM and ran a single iteration to locate the nine points. The 37-point RFRV-CLM was then initialised using these nine points. Note that the positions of the points used for initialisation were refined during model matching. We simply applied the method used in [18].

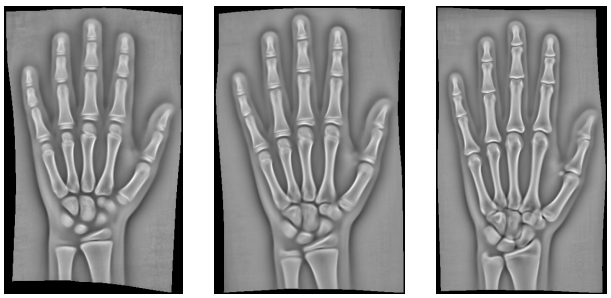
The system combined a Hough Forest-like global search with local refinement (as in [19]). To evaluate the system we trained two models, one on the males, one on the females, and applied each model to the images from the other sex. The images were also manually annotated and the annotation was compared with the automatic points resulting into mean point to point errors of $\mathbf{0.87mm}$. This is slightly higher than what was reported in [18] whose method we adopted.



(a) (b) (c)

Fig. 2. Annotation example of a radiograph and 37 found points from $RFV - CLM$ shape models.

4.2 Groupwise Registration Experiments



a) b) c)

Fig. 3. Final mean images after groupwise registration. a) agegroup 1, b) agegroup 2 and c) agegroup 3.

4.3 Dense annotation experiments

We divided the dataset into three age-groups. Agegroup1 -63 images (5 - 7 yrs), Agegroup2 -284 images (8-13 yrs) and Agegroup 3-189 images (14 -18 years)

Groupwise registration algorithm was initialized with the 37 found points from $RFRV - CLM$

The found points in each of the groups were used to initialise a groupwise algorithm as described above. Qualitative results of the registration is shown in Figures 3. The crispness of the images indicate a good alignment.

We evaluated the accuracy of the points location by comparing with manual annotations.

This is the same approach adopted in [2]. In order to evaluate the accuracy of the point correspondences we manually annotated every image with 37 landmarks at the major joints. The sets of points found by the model were mapped using a thin-plate spline (TPS) into a reference frame defined by the aligned mean of the manual landmarks. The mean position for each part was calculated, then mapped back into the original image using the TPS. The absolute difference between each found point and the estimate of the mean position was calculated. The mean distance errors for sparse point errors was found to be $1.08 \pm 0.18mm$, $0.91 \pm 0.15mm$, $0.75 \pm 0.09mm$ for agegroup 1, agegroup2, agegroup 3 respectively. The result of agegroup 3 14 -19, a very difficult group, is comparable to the original result obtained in [1]. Table 1 show the Statistics of mean distance errors of models for the three age-groups from the $RFRV - CLM$ Models.

Age-group	Mean distance $\pm se$ (mm)	% $d > 2mm$
Age-group 1(63)	1.08 $\pm$0.18	6%
Age-group 2(284)	0.91 $\pm$0.15	7%
Age-group 3(189)	0.75 $\pm$0.09	4%

TABLE 1
Statistics of mean distance errors of models for the three age-groups from the Part Based Models $RFRV - CLM$

However, we wish to use these points to initialise a dense annotation. This can be done by using the found points as control points in a TPS warp. To evaluate the accuracy of the resulting deformation field, we measure the distance between the manual points and the estimate of their mean warped to each image (reversing the roles of the found points and manual points compared to the previous experiment). Table 2 show the mean distance errors of models for the three age-groups after dense registration Please note that in both cases errors are highest for agegroup1. The few number of images and very large variation may be responsible. Sometimes there is no correspondence amongst the bones.

In order to accurately locate the boundaries of more than 28 bones of the hand, we manually annotated the borders of three images with 2,797 points. One image was selected from each of the three groups. Figure 4 shows representative annotation for each of the three groups. These dense points are then propagated to other images using the parameters of the groupwise registration for the 37 automatically found points. Essentially 2,797 points are propagated to all images in the three age groups.

Age-group	Mean distance $\pm se$ (mm)	% $d > 2mm$
Age-group 1(63)	0.83 $\pm$0.02	6%
Age-group 2(284)	0.98 $\pm$0.01	6%
Age-group 3(189)	1.05 $\pm$0.01	10%

TABLE 2
Statistics of mean distance errors of models for the three age-groups after dense registration

Age Groups	Mean $\pm se$ (mm)	Median (mm)	90%-ile
Age-group 1	0.88 $\pm$0.01	0.87	1.04
Age-group 2	1.19 $\pm$0.01	1.16	1.49
Age-group 3	0.75 $\pm$0.01	0.72	0.96

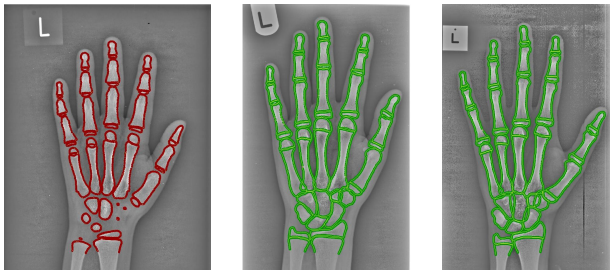
TABLE 3
Statistics of point to curve error after dense propagation for Age-group 1 ,2 3 (mm)

We measure the errors by comparing how well a correspondence defined by control points and mesh in the densely propagated set agrees with a set of dense manual annotations with 330 points in equivalent positions. We compute the mean distance to curve error between warped version of manual points and the manual annotations for each image. The result of this evaluation is shown in Table 3, for the three age groups.

Quantitative result of annotating the 28 bones of the hand is shown in Figure 5 for two images in Age-group 3, while further results of enlarged version of an area, typical success and failure annotated images are shown in Figure 6

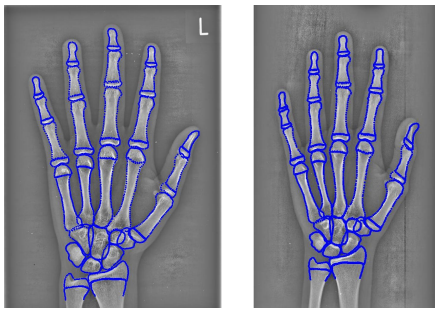
5 Discussion and Conclusions

We have proposed an approach for automatically locating sparse correspondences across a set of images, by constructing a parts and



a) b) c)

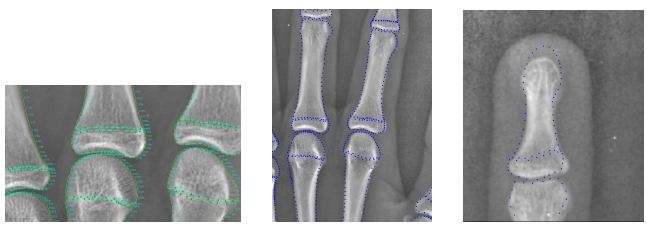
Fig. 4. (a) A dense manual markup with 2,797 points on a group 1 image (b) A dense manual markup with 2,797 points on a group 2 image (c)A dense manual markup with 2,797 points on a group 3 image



a) Image1 b) Image2

Fig. 5. Quantitative results of propagation of 2,797 points based on the Part based models' automatic initialization for two examples in group 3

geometry model with an extended dataset. We achieve an accuracy of 0.80mm on the positioning of the chosen parts. This work compares favourably with what obtains in the literature [14] [8] [1], [17]. The closest work to this is that of Zhang *et al.*[21] though a fully automatic method, the algorithm was run for a limited age range(10-13years). Most of the cited work are often dealing with model matching for sparse points placement but rarely proceeds to the point of annotating data-set in a way to accurately separate the different bones. This is indeed a requirement for skeletal maturity where global and local model of bone complexes are often required. This work therefore extends our earlier work [1] in a way to provide a tool for annotating hand radiograph for use in skeletal maturity. We have achieved a point to curve error of approximately $1mm$ for a dense annotation of close to 3000 points. This is one of the best results for this application. It is envisaged that this work will be useful to professional who have need for accurate bone annotations.



a) b) c)

Fig. 6. Quantitative results of point annotations (a)Points annotations zoomed in (b) Successful example (c) Enlarged typical failure

ACKNOWLEDGMENTS

The authors would like to thank Prof Judith Adams for providing the Radiograph images..

Steve A. Adeshina Dr Steve A Adeshina is with the Nile University of Nigeria, Abuja Nigera.

Claudia Lindner Dr Claudia Lindner is with Centre for Imaging Sciences, The University of Manchester, Manchester, UK.

Timothy F. Cootes Prof Tim Cootes is with Centre for Imaging Sciences, The University of Manchester, Manchester, UK.

REFERENCES

[1] S. A. Adeshina and T. F. Cootes. Constructing part-based models for groupwise registration. In *Proc. 2010 IEEE International Symposium on Biomedical Imaging*, pages 733–740, 2010.

[2] S. A. Adeshina and T. F. Cootes. Automatic annotation of radiographs using parts and geometry models for building statistical models for skeletal maturity. In *Proc. 2014 IEEE International Conference on Electronics Computers and Computation*, pages 733–740, 2014.

[3] S. A. Adeshina and T. F. Cootes. Automatic segmentation of carpal area bones with random forest regression voting for estimating skeletal maturity in infants. In *Proc. 2014 IEEE International Conference on Electronics Computers and Computation*, pages 733–740, 2014.

[4] S. Baker, I. Matthews, and J.Schneider. Automatic construction of active appearance models as an image coding problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1380–84, 2004.

[5] L. C., T. S., J. Wilkinson, T. Consortium, G. Wallis, and . Cootes. Fully automatic segmentation of the proximal femur using random forest regression voting. *Medical Imaging, IEEE Transactions on*, 32(8):1462–1472, 2013.

[6] 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.

[7] 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.

[8] T. F. Cootes, M. C. Ionita, C. Lindner, and P. Sauer. Robust and accurate shape model fitting using random forest regression voting. In *ECCV 2012*, pages 278–291. Springer-Verlag, 2012.

[9] 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.

[10] R. Davies, C.Twining, T. Cootes, and C. Taylor. A minimum description length approach to statistical shape modelling. *IEEE Transactions on Medical Imaging*, 21:525–537, 2002.

[11] R. H. Davies, C. Twining, T. F. Cootes, J. Waterton, and C. Taylor. 3D ststistical shape models using direct optimisation of description length. In 7^{th} *European Conference on Computer Vision*, volume 3, pages 3–20. Springer, 2002.

[12] D.Cristinacce and T.F.Cootes. Automatic feature localisation with constrained local models. *Pattern Recognition*, 41(10):3054–3067, 2008.

[13] R. Donner, B. Menz, H. Bischof, and G. Langs. Fast anatomical structure localization using top-down image patch regression. In *Proc. MICCAI*, volume Workshop, pages 133–141, 2013.

[14] R. Donner, H. Wildenauer, H. Bischof, and G. Langs. Weakly supervised group-wise model learning based on discrete optimization. In *Proc. MICCAI*, volume 2, pages 860–868, 2009.

[15] A. Frangi, D. Rueckert, J. Schnabel, and W. Niessen. Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modeling. *IEEE-TMI*, 21:1151–66, 2002.

[16] V. Gall. J., Lempitsky. Class specific hough forests for object detection. In *IEEE Proc Computer Vision and Pattern Recognition*, pages 415–422, 2009.

[17] M. Harmsen, B. Fischer, H. Schramm, T. Seidl, and T. M. Deserno. Support vector machine classification based on correlation prototypes applied to bone age assessment. *IEEE J. Biomedical and Health Informatics*, 17(1):190–197, 2013.

[18] C. Lindner, P. Bromiley, M. Ionita, and T. Cootes. Robust and accurate shape model matching using random forest regression-voting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 01(99):1862–1874, 2015.

[19] C. Lindner, S. Thiagarajah, J. M. Wilkinson, G. A. Wallis, and T. F. Cootes. Accurate fully automatic femur segmentation in pelvic radiographs using regression voting. In *MICCAI*, pages 353–360, 2012.

[20] M. Miller. Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms. *NeuroImage*, 23:S19–S33, 2004.

[21] P. Zhang and T. Cootes. Automatic construction of parts+geometry models for initializing groupwise registration. *Medical Imaging, IEEE Transactions on*, 31(2):341–358, 2012.