Accurate Bone Segmentation in 2D Radiographs Using Fully Automatic Shape Model Matching Based On Regression-Voting

C. Lindner[1], S. Thiagarajah[2], J.M. Wilkinson[2], arcOGEN Consortium,
G.A. Wallis[3], and T.F. Cootes[1]

[1] Centre for Imaging Sciences, University of Manchester, UK
[2] Department of Human Metabolism, University of Sheffield, UK
[3] Wellcome Trust Centre for Cell Matrix Research, University of Manchester, UK

Abstract. Recent work has shown that using Random Forests (RFs) to vote for the optimal position of model feature points leads to robust and accurate shape model matching. This paper applies RF regression-voting as part of a fully automatic shape model matching (FASMM) system to three different radiograph segmentation problems: the proximal femur, the bones of the knee joint and the joints of the hand. We investigate why this approach works so well and demonstrate that the performance comes from a combination of three properties: *(i)* The integration of votes from multiple regions around the model point. *(ii)* The combination of multiple independent votes from each tree. *(iii)* The use of a coarse to fine strategy. We show that each property can improve performance, and that the best performance comes from using all three. We demonstrate that FASMM based on RF regression-voting generalises well across application areas, achieving state of the art performance in each of the three segmentation problems. This FASMM system provides an accurate and time-efficient way for the segmentation of bony structures in radiographs.

Keywords: Computational anatomy, Random Forests, Constrained Local Models, statistical shape models, bone segmentation

1 Introduction

Shape model matching plays an important role in a variety of application areas when segmenting structures in medical images. Random Forest (RF) regression-voting has emerged as a powerful and promising technique to predict objects and image properties [4, 7–9]. RF regression-voting combines predictions from a number of independent decision trees (each trained on a randomised selection of features [1]) and accumulates predictions of object or feature points made from multiple regions of the image. Recent work has shown that using RF regression to vote for the optimal position of model feature points leads to robust and accurate shape model matching [2, 10].

In this paper, we apply RF regression-voting as part of a fully automatic shape model matching (FASMM) system to three different radiograph segmentation problems: the proximal femur, the bones of the knee joint and the joints of the hand. We investigate why RF regression-voting works so well and which properties of the regressor are important for achieving the best performance. In particular, we analyse the importance of:

1) The integration of votes from multiple regions around the model point.
2) The forests combining multiple independent estimates, one from each tree.
3) The use of a coarse to fine strategy.

We show that all three properties contribute individually as well as collectively to the performance of RF regression-voting and that the latter generalises well across application areas. By incorporating all three properties, *(i)* we significantly improve upon previously published state of the art results on proximal femur segmentation [10]; *(ii)* we obtain what we believe to be the best yet published results on knee joint segmentation; and *(iii)* we achieve excellent fully automatic results for hand joint annotation.

2 Methods

To match a shape model to a new image, we explore the performance of RF regression-voting in the Constrained Local Model (CLM) [2] framework.

2.1 Shape model matching

CLMs combine global shape constraints with local models of the pattern of intensities. Based on a number of model points in a set of images, a statistical shape model is trained by applying principal component analysis to the aligned shapes [3]. This yields a linear model of shape variation which represents the position of each model point l using $\mathbf{x}_l = T_\theta(\bar{\mathbf{x}}_l + \mathbf{P}_l\mathbf{b} + \mathbf{r}_l)$ where $\bar{\mathbf{x}}_l$ is the mean position of the point in a suitable reference frame, $\mathbf{P}_l$ is a set of modes of variation, $\mathbf{b}$ are the shape model parameters, $\mathbf{r}_l$ allows small deviations from the model, and T_θ applies a global transformation (e. g. similarity) with parameters θ. Based on the 2D histograms of votes V_l from a RF regressor (see Section 2.2 below), one for every model point, we aim to combine the votes in all histograms given these shape constraints via optimising

$$Q(\{\mathbf{b}, \theta\}) = \Sigma_{l=1}^{n} V_l(T_\theta(\bar{\mathbf{x}}_l + \mathbf{P}_l\mathbf{b} + \mathbf{r}_l)). \tag{1}$$

We apply the technique described in [2] to solve this optimisation problem.

2.2 Random Forest regression-voting

RFs are composed of multiple independent trees [1]. All trees are trained independently on a random subset of features, and every tree will cast independent votes to make predictions. The forest prediction is computed by combining all independent tree predictions.

In the RF regression-voting approach as presented in [2], a set of points in a grid over a region of interest is evaluated. We train the RF regressors from sets of images, each of which is annotated with the feature points of interest on the object, $\mathbf{x}$. The region of interest of the image that captures all feature points of the object is re-sampled into a standardised reference frame. For every point in $\mathbf{x}$, a set of features $\mathbf{f}_i(\mathbf{x})$ is sampled at a set of random displacements $\mathbf{d}_i$ from the true position in the reference frame, and a regressor $\delta = R(\mathbf{f}(\mathbf{x}))$ is trained to predict the most likely position of the model point relative to $\mathbf{x}$. Displacements are drawn from a flat distribution in the range $[d_{max}, +d_{max}]$ in x and y. Each tree leaf stores the mean offset and the standard deviation of the displacements of all training samples that arrived at that leaf. In this work, we use Haar features [12] as they have been found to be effective for a range of applications and can be calculated efficiently from integral images.

During model matching, given an initial estimate of the pose of the object, the region of interest of the image is re-sampled into the reference frame, an area around each feature point is searched and the relevant feature values at every position are extracted. These will be used for the RF regressor to vote for the best position in an accumulator array yielding a 2D histogram of votes. To blur out impulse responses we slightly smooth the histogram with a Gaussian.

3 Experiments and evaluation

We perform a series of experiments to analyse the performance of the RF regression-voting approach across application areas and to evaluate the contribution of each of the properties. In each case, we use a *fully automatic* shape model matching system: We use our own implementation of Hough Forests [6] to estimate the position, orientation and scale of the object in the image, and use this to initialise the shape model matching (starting at the mean shape).

The most extensive experiments were performed on the segmentation of the proximal femur from pelvic radiographs. We also give fully automatic segmentation results for a subset of experiments performed on knee bone and hand joint segmentation. All evaluations are based on manually annotated ground truth.

3.1 Data sets

We used anteroposterior (AP) pelvic radiographs of 839 subjects suffering from unilateral hip osteoarthritis, each annotated with 65 points. We also used AP knee radiographs of 500 subjects suffering from knee osteoarthritis, each annotated with 87 points. Both the pelvic and knee radiographs were provided by the arcOGEN Consortium and were collected under relevant ethical approvals. In addition, we used 564 AP hand radiographs of children of ages between five and eighteen years, each annotated with 37 points. Each data set varies widely in terms of resolution levels (pelves: 555-4723 pixels wide; knees: 765-4280 pixels wide; hands: 485-1168 pixels wide) and intensity levels. Fig. 1 gives annotation examples for each of the three data sets. For the scope of the experiments described below, the femur and knee data sets are split randomly into a training and a testing subset of equal size. The hand data set is split randomly into a training set containing 200 images and a testing set containing 364 images.

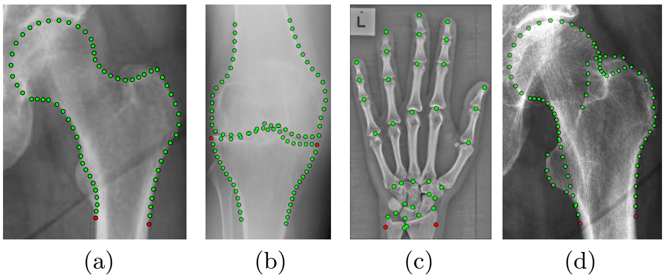


Fig. 1. Annotation examples showing the 95%ile of the fully automatic shape model matching results for: (a) the proximal femur excluding the trochanters with 65 points; (b) the knee with 87 points; (c) the hand with 37 points; (d) the proximal femur including the trochanters with 81 points. Red points define the reference length.

3.2 Segmenting the proximal femur in pelvic radiographs

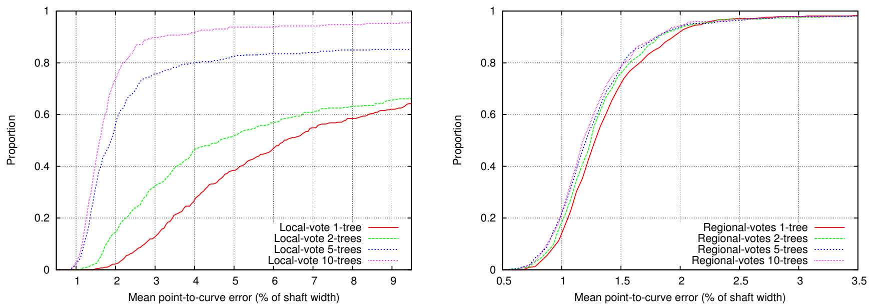
We run a series of systematic experiments to place 65 points along the contour of the left proximal femur as in Fig. 1(a). The reference frame image is 200 pixels wide, and Haar features are sampled from patches of size 20×20. The regression functions for all feature points are trained using 15 random displacements of up to 20 pixels in x and y in the reference image, as well as random displacements in scale (in the range of $\pm 5\%$) and rotation (in the range of $\pm 6^\circ$).

Based on the estimated pose and scale of the proximal femur resulting from the global search scanner, we run 5 search iterations (i. e. generating vote histograms and matching the shape model) updating shape and pose parameters $\{\mathbf{b}, \theta\}$ (see Equation 1). If not stated otherwise each RF has 10 trees.

Results are presented as cumulative density functions (CDFs) of the mean point-to-curve error as a percentage of the shaft width (see red points in Fig. 1(a)) – to provide invariance to image scaling. This can be used as a reference length to evaluate segmentation performance in mm as it tends to be relatively constant across individuals. Our proximal femur data set contains 15 calibrated images which suggest an average length of $37mm$.

Integrating votes from multiple regions around the model point In the case of only one tree, Fig. 2(a) shows the results when using only a single local vote from the patch at the current estimate of the feature point. Comparing these to the single tree results in Fig. 2(b), it shows that significant improvements can be achieved when combining votes from multiple regions around the model point.

Combining multiple independent votes for each model point To investigate the effect of multiple independent votes on the performance of the RF regressor, we vary the number of trees. Fig. 2(a) shows the results when every tree uses a single local vote only and Fig. 2(b) when we combine multiple votes from different regions. This demonstrates that increasing the number of trees has



(a) (b)

Fig. 2. Effect of combining multiple independent estimates of position by varying the number of trees in the Random Forest (RF) when every tree: (a) votes only from the current point; (b) combines votes from multiple regions around the model point.

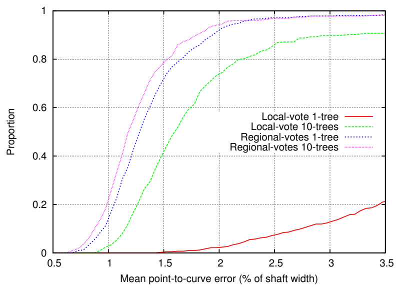


Fig. 3. Effect of combining votes from multiple different regions around the model point along with multiple independent estimates of position by varying the number of trees in the Random Forest.

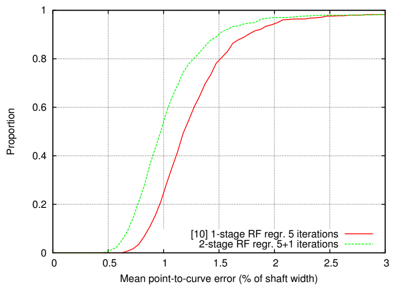


Fig. 4. Effect of introducing a second searching stage to locally refine point placements. Every Random Forest (RF) uses 10 trees.

a significant impact when voting from a single position. Using multiple regional votes *and* increasing the number of trees, the main improvement happens when using more than one tree – additional trees only slightly improve the performance. Linking votes from multiple different regions and independent estimates from multiple trees shows that using multiple regional votes around the model point has a greater effect than increasing the number of trees; see Fig. 3. The best performance can be achieved by combining multiple regional votes with votes from multiple independent trees in the RF.

Using a coarse to fine strategy Further improvements can be achieved when using the regression-voting scheme in a multi-stage coarse to fine approach. We run a rather coarse model with a lower resolution reference frame first (as is described above), and use its results to initialise a higher resolution (fine) model. The latter uses a reference image that is 500 pixels wide and the displacement range during training is within 15 pixels in x and y in the reference image. Patch size and number as well as scale and angle perturbation ranges are the same as for the coarse model. Since the coarse model provides a very good initialisation, we only apply a single search iteration of the fine model (i. e. the coarse to fine two stage model runs 5 coarse + 1 fine iterations). Fig. 4 compares the model-matching performance of a one stage and a two stage regression-voting strategy. Here, we perform two-fold cross-validation experiments training on one half of the data and testing on the other and then doing the same with the sets switched (using all 839 images). Results reported are the average of the two runs.

In [10], a fully automatic proximal femur segmentation system was presented achieving the best results yet published. These results were based on only a single stage RF regression-voting approach. By introducing a second stage, we can significantly improve upon those results, decreasing the mean 95%ile point-to-curve error from 2.1 to 1.8 and hence achieving an increased accuracy with errors of less than 0.7mm for 95% of all 839 images. We achieve the same performance when including the trochanters in the model as illustrated in Fig. 1(d).

3.3 Segmenting the bones of the knee joint

All experiments on segmenting the bones of the knee use the same parameter values as were used for segmenting the femur. We run two-fold cross-validation experiments to place 87 points along the contour of the right knee. Fig. 5 shows the CDFs of the mean point-to-curve errors as a percentage of the tibial plateau width (see red points in Fig. 1(b)) – comparing one and two stage RF regression models. We assume an average tibial plateau width of 75mm. Using the very same settings as for the femur shows that the parameter settings generalise well across application areas. However, the 99%ile when running 5+1 iterations is significantly higher for the knee ($\approx 10mm$) than for the femur ($\approx 2mm$). This indicates that the initialisation resulting from the global search scanner does not perform as well as for the femur. Increasing the number of search iterations from 5+1 to 20+1 overcomes this issue: The best results relate to a mean point-to-curve error of less than 1mm for 99% of all 500 images.

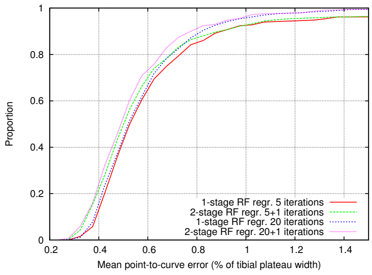


Fig. 5. Performance of Random Forest (RF) regression-voting for fully automatically segmenting the knee bones.

3.4 Segmenting the joints of the hand

We initialise the model by automatically locating 9 points (four around the palm and one at the base of each finger). This is achieved by finding a box around the palm with a Hough Forest, and then running a two stage RF-CLM to locate the 9 points. The 37 point model (see Fig. 1(c)) is initialised from these 9 points.

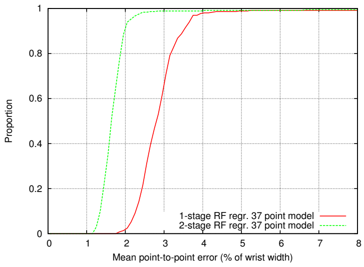


Fig. 6. Performance of Random Forest (RF) regression-voting for fully automatically segmenting the hand joints.

For the first stage, the reference frame image is 70 pixels wide, and Haar features are sampled from patches of size 13 × 13. The regression functions for all feature points are trained using 10 random displacements of up to 11 pixels in x and y in the reference image, as well as random displacements in scale (in the range of $\pm 5\%$) and rotation (in the range of $\pm 3°$). The second stage model uses a reference frame image that is 210 pixels wide, patches of size 17 × 17 and the displacement range during training is within 7 pixels. Each RF has 10 trees. The values above were found to provide the best performance in an initial pilot experiment sweeping through the parameters. We run a single search iteration with the first stage model followed by a single search iteration with the second stage model.

Fig. 6 shows the CDFs of the mean point-to-point error as a percentage of the wrist width (see red points in Fig. 1(c)) for one and two stage models. This shows that a coarse to fine strategy significantly improves the results. The best results are very similar to those presented in [2], however in our case they are produced using a fully automatic system rather than assuming a reasonable local initialisation. Assuming a mean wrist width of $50mm$, the best results relate to a mean point-to-point error of within $1.1mm$ for 95% of all images.

4 Discussion and conclusions

We have investigated which properties are important when aiming at robust and accurate shape model matching based on RF regression-voting. In our experiments, the latter was incorporated in a FASMM system and tested on comprehensive mixed-quality data sets. We have shown that combining votes from multiple regions around the model point with multiple independent estimates of position while following a coarse to fine strategy achieves state of the art segmentation performance across application areas. Fig. 1 shows the 95%ile of our fully automatic matching results for each of the applications.

We have demonstrated that applying a coarse to fine strategy to the proximal femur segmentation problem gives an improvement over the best published results [10] achieving a mean point-to-curve error of less than $0.7mm$ for 95% of all images. We have also shown state of the art performance when fully automatically segmenting the bones of the knee joint achieving errors of within $0.7mm$ for 95% of all images. A direct comparison to other reported results seems difficult as most 2D knee joint segmentation findings are either only given in pixels or relate to a different ROI (see e. g. [11]). However, we believe this to be the best results yet published. In addition, when segmenting the joints of the hand our FASMM system achieved results comparable to previously published results [2, 5] with a mean point-to-point error of within 1.1mm for 95% of all images. The latest version of our FASMM system can process one image on average in 28s for the femur, 30s for the knee and less than 1s for the hand (on a 3.3GHz Intel Core Duo PC using 1.5GB RAM); note that the hand models are of lower resolution and only need to search a small area. The presented approach shows great promise for many medical applications such as early diagnosis or quantitative assessment of treatment response/progression of disease.

Acknowledgements. The arcOGEN Consortium is funded by Arthritis Research UK and C. Lindner by the Medical Research Council. We thank K. Ward, R. Ashby, Z. Mughal and Prof. J. Adams for providing the hand radiographs and S. Adeshina for the hand annotations. We thank X. Milidonis for the annotation of the knees.

References

1. Breiman, L.: Random Forests. Machine Learning 45, 5–32 (2001)
2. Cootes, T., Ionita, M., Lindner, C., Sauer, P.: Robust and Accurate Shape Model Fitting using Random Forest Regression Voting. In: ECCV 2012 - Part VII. LNCS, vol. 7578, pp. 278–291. Springer, Heidelberg (2012)
3. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models - their training and application. Computer Vision and Image Understanding 61(1), 38–59 (1995)
4. Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E.: Regression forests for efficient anatomy detection and localization in CT studies. In: Menze, B., Langs, G., Tu, Z., Criminisi, A. (eds.) MICCAI 2010 Workshop MCV. LNCS, vol. 6533, pp. 106–117. Springer, Heidelberg (2010)
5. Donner, R., Menze, B., Bischof, H., Langs, G.: Fast anatomical structure localization using top-down image patch regression. In: Menze, B., Langs, G., Lu, L., Montillo, A., Tu, Z., Criminisi, A. (eds.) MICCAI 2012 Workshop MCV. LNCS, vol. 7766, pp. 133–141. Springer, Heidelberg (2013)
6. Gall, J., Lempitsky, V.: Class-specific Hough forests for object detection. In: CVPR. pp. 1022–1029. IEEE Press (2009)
7. Girshick, R., Shotton, J., Kohli, P., Criminisi, A., Fitzgibbon, A.: Efficient Regression of General-Activity Human Poses from Depth Images. In: ICCV. pp. 415–422. IEEE Press (2011)
8. Glocker, B., Feulner, J., Criminisi, A., Haynor, D., Konukoglu, E.: Automatic Localization and Identification of Vertebrae in Arbitrary Field-of-View CT Scans. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part III. LNCS, vol. 7512, pp. 590–598. Springer, Heidelberg (2012)
9. Konukoglu, E., Glocker, B., Zikic, D., Criminisi, A.: Neighbourhood Approximation Forests. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part III. LNCS, vol. 7512, pp. 75–82. Springer, Heidelberg (2012)
10. Lindner, C., Thiagarajah, S., Wilkinson, J., arcOGEN Consortium, T., Wallis, G., Cootes, T.: Accurate fully automatic femur segmentation in pelvic radiographs using regression voting. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part III. LNCS, vol. 7512, pp. 353–360. Springer, Heidelberg (2012)
11. Seise, M., McKenna, S., Ricketts, I., Wigderowitz, C.: Learning active shape models for bifurcating contours. IEEE Trans. on Medical Imaging 26(5), 666–677 (2007)
12. Viola, P., Jones, M.: Rapid Object Detection Using a Boosted Cascade of Simple Features. In: CVPR. pp. 511–518. IEEE Press (2001)