# Accurate Fully Automatic Femur Segmentation in Pelvic Radiographs Using Regression Voting

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

[1] 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.** Extraction of bone contours from radiographs plays an important role in disease diagnosis, pre-operative planning, and treatment analysis. We present a fully automatic method to accurately segment the proximal femur in anteroposterior pelvic radiographs. A number of candidate positions are produced by a global search with a detector. Each is then refined using a statistical shape model together with local detectors for each model point. Both global and local models use Random Forest regression to vote for the optimal positions, leading to robust and accurate results. The performance of the system is evaluated using a set of 519 images. We show that the fully automated system is able to achieve a mean point-to-curve error of less than $1mm$ for 98% of all 519 images. To the best of our knowledge, this is the most accurate automatic method for segmenting the proximal femur in radiographs yet reported.

**Keywords:** automatic femur segmentation, femur detection, Random Forests, Hough Transform, Constrained Local Models, radiographs.

## 1 Introduction

In clinical practice, plain film radiographs are widely used to assist in disease diagnosis, pre-operative planning and treatment analysis. Extraction of the contours of the proximal femur from anteroposterior (AP) pelvic radiographs plays an important role in diseases such as osteoarthritis (e. g. diagnostics and joint-replacement planning) or osteoporosis (e. g. fracture detection and bone density measurements). In addition, accurately segmenting the contours of the proximal femur in radiographs allows monitoring of disease progression.

Manual segmentation of the femur is time-consuming and hard to do consistently. Our aim is to automate the segmentation procedure. Fully automatic proximal femur segmentation is challenging for several reasons: *(i)* The quality of radiographs may vary a lot in terms of contrast, resolution and the region of the pelvis shown. *(ii)* AP pelvic radiographs only give a 2D projection, and hence are susceptible to rotational issues; the same 3D shape may yield a different 2D projection depending on the view point. *(iii)* Plain film radiographs do not provide homogeneous values for the same structure due to overlapping body parts.

N. Ayache et al. (Eds.): MICCAI 2012, Part III, LNCS 7512, pp. 353–360, 2012.

*(iv)* Deformities of the proximal femur may cause the loss of distinguishable radiographic key features.

Automatically extracting the contours of the proximal femur comprises two key steps: Firstly, the femur is detected in the image and secondly, the contours are segmented. Behiels et al. [1] have shown the suitability of statistical shape models for proximal femur segmentation. Recent work on automatically segmenting the femur in radiographs using statistical shape models includes [11,12]. Object detection in the latter as well as in the atlas-based approach of Ding et al. [8] is based on edge detection. We use Random Forest regression in a sliding window approach to automatically segment the proximal femur.

Random Forests (RF) [2] describe an ensemble of decision trees trained independently on a randomised selection of features. They have been shown to be effective in a range of classification and regression problems [6,10]. Recent work on Hough Forests [9] has shown that objects can be effectively located by training RF regressors to predict the position of a point relative to the sampled region, then running the regressors over a region and accumulating votes for the likely position. To detect the femur, our global search uses a RF regressor that votes for the centre of a reference frame, resulting in a response image of accumulated votes. The approximated position is then used to initialise a local search to segment the femur, combining local detectors with a statistical shape model. Following [3], we apply RF regression in the Constrained Local Model (CLM) framework to vote for the optimal position of each model point. Here, feature detectors are run independently to generate response images for each point and then a shape model is used to find the best combination of points [7].

Using RF regression voting for both object detection and CLM-based contour extraction yields a robust and fully automatic segmentation system. We use the latter to segment the femur in pelvic radiographs, and demonstrate that results are very accurate. The local search and the fully automatic search outperform alternative matching techniques such as Active Shape Models [5], CLMs using normalised correlation and RF classification-based search. We believe this to be the most accurate fully automatic femur segmentation system yet published.

## 2 Methods

The fully automated segmentation system comprises a global search detecting the object and a local search segmenting the contours. Both global and local search use RF regression voting to predict object and point positions.

### 2.1 Voting with Random Forest Regression

We use RF regression in a similar manner to the Hough Forests approach [9]. However, we do not require voting to be dependent on a class label, allowing all image structures to vote. In the voting-regression approach, we evaluate a set of points in a grid over a region of interest. At each point $\mathbf{z}$, a set of features $\mathbf{f}(\mathbf{z})$ is sampled. A regressor, $R(\mathbf{f}(\mathbf{z}))$, is trained to predict the most likely position(s) of the target point relative to $\mathbf{z}$. During training, given the samples at a particular node, we seek to select a feature and threshold to best split the data. Let $f_i$ be the value of one feature associated with sample $i$. The best threshold, $t$, for this feature at this node is the one which minimises $G_T(t) = G(\{\mathbf{d}_i : f_i < t\}) + G(\{\mathbf{d}_i : f_i >= t\})$ where $G(S)$ is a function evaluating the set of vectors $S$, and $\mathbf{d}_i$ the predicted displacement of sample $i$. We aim at minimising the entropy in the branches when splitting the nodes using $G(\{\mathbf{d}_i\}) = Nlog|\Sigma|$, where $N$ is the number of displacements in $\{\mathbf{d}_i\}$ and $\Sigma$ the respective covariance matrix. Criminisi et al. [6] showed that a related measure of information gain was effective for regression.

Hough Forests use RFs whose leaves store multiple training samples. Thus each sample produces multiple votes, allowing for arbitrary distributions to be encoded. Each leaf of our decision trees only stores the mean offset and the standard deviation of the displacements of all training samples that arrived at that leaf. During search, these predictions are used to vote for the best position in an accumulator array. Predictions are made using a single vote per tree yielding a Hough-like response image. To blur out impulse responses we slightly smooth the response image with a Gaussian.

Below, we use Haar features [13] as they have been found to be effective for a range of applications and can be calculated efficiently from integral images.

### 2.2 Object Detection

**Training.** A reference frame, or bounding box, is set to capture the object of interest. For each training image, a number of random displacements (scale, angle and position) of the bounding box are sampled. To train the detector, for every sample we extract features $\mathbf{f}_i$ at a set of random positions within the sampled patch and store displacement $\mathbf{d}_i$ from the original centre of the reference frame. We then train a RF on the pairs $\{\mathbf{f}_i, \mathbf{d}_i\}$. To train a single tree, we take a bootstrap sample of the training set, and construct the tree by recursively splitting the data at each node as described in Section 2.1. The extracted features are a random subset of all possible Haar features and at each node, we choose the feature and associated threshold which minimise $G_T$ to split the data.

**Search.** To detect the object in an image, we scan the image at a set of coarse angles and scales in a sliding window approach. The search is speeded up by evaluating only positions on a sparse grid rather than at every pixel. For every angle-scale combination, we scan the bounding box across the image. We obtain the relevant feature values from each box and get the RF to make predictions on the reference frame centre. Predictions are made using a single Gaussian weighted vote per tree, where the weights relate to the spread of the displacements of the training samples that arrived at the particular leaf. The resulting response image is then searched for local maxima. Once a response image has been obtained for every angle-scale combination, all maxima are ranked according to their total votes. Every maxima is associated with an angle, a scale and a prediction of the reference frame centre. This results in candidate positions for the object.

### 2.3 Segmentation Using Constrained Local Models

CLMs combine global shape constraints with local models of 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 [5]. This yields a linear model of shape variation which represents the position of each landmark point using $\mathbf{x}_i = T_\theta(\bar{\mathbf{x}}_i + \mathbf{P}_i\mathbf{b} + \mathbf{r})$ 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 deviations from the model, and $T_\theta$ applies a global transformation (e. g. similarity) with parameters $\theta$. Similar to Active Appearance Models [4], CLMs combine this shape model with a texture model but only sample a local patch around each landmark rather than the whole object.

To match the CLM to a new image, we seek the shape and pose parameters, $\mathbf{p} = \{\mathbf{b}, \theta\}$, which optimise the fit of the model to the image. Given an initial estimate of every landmark's position, an area around each landmark point is searched. At every position $i$, a quality-of-fit value, describing the similarity between the template texture for this landmark learned from the model and the texture at that position, is obtained and stored in a response image $\mathbf{R}_i$. We then find the shape and pose parameters which optimise $\Sigma_{i=1}^{n}\mathbf{R}_i(T_\theta(\bar{\mathbf{x}}_i + \mathbf{P}_i\mathbf{b} + \mathbf{r}))$.

In [3] it is shown how RF regression voting produces useful response images for the CLM framework. Here we summarise the key steps.

**Training.** CLMs in their original form use normalised correlation as quality-of-fit measurement for each response image. In the RF regression approach, we train a regressor to predict the position of a landmark point based on a random set of Haar features. The quality-of-fit values here relate to the votes of the RF.

For every landmark $i$, we sample local patches at a number of random displacements $\mathbf{d}_i$ from the true position. For every sample we extract features $\mathbf{f}_i$ and train a RF on the pairs $\{\mathbf{f}_i, \mathbf{d}_i\}$. As with the global search, we train every tree taking a bootstrap sample and constructing it recursively by splitting the data at each node as described in Section 2.1.

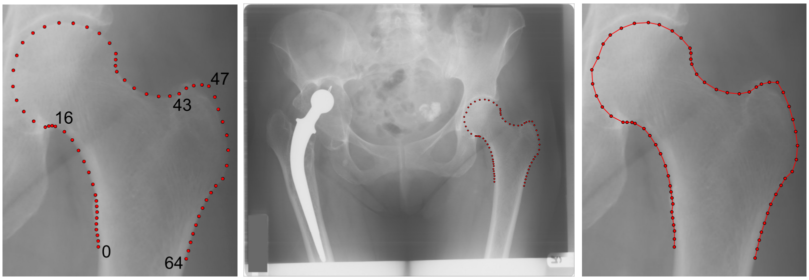
**Search.** To match the RF regression-based CLM to a new image, for every landmark $i$, we sparsely sample local patches in the area around an initial estimate of the landmark's position. We extract the relevant features for each sample and get the RF to make predictions on the true position of the landmark. Predictions are made using a single vote per tree. This yields a response image $\mathbf{R}_i$ for every landmark $i$. We then aim to combine voting peaks in the response images with the global constraints learned by the shape model.

### 2.4 Automated System

The fully automated system performs a global search at multiple scales and orientations to produce a number of candidate poses which are ranked by total votes. The local search is then applied at each of the best $l$ search candidates, and the final results are ranked by the total CLM fit (sums of votes).

## 3 Experiments and Evaluation

The aim is to fully automatically segment the femur by putting a dense annotation of 65 landmarks along its contours as demonstrated in Figure 1; (a) gives the manual annotation and (b)-(c) the result of the fully automated system. We use a *front-view femur* model that excludes both trochanters and approximates the superior lateral edge (points 43 to 47) from an anterior perspective. All points were defined using anatomical features mixed with an evenly spaced subset.



(a) (b) (c)

**Fig. 1.** Segmentation of the proximal femur: (a) 65 landmarks outlining the 'front-view' femur (ground truth); (b)-(c) automatically segmented femur in AP pelvic radiograph

Our data set comprises AP pelvic radiographs of 519 females suffering from unilateral hip osteoarthritis. All images were provided by the arcOGEN Consortium and were collected under relevant ethical approvals. The images have been collected from different radiographic centres resulting in varying resolution levels (555-4723 pixels wide) and large intensity differences. In addition, the displayed pelvic region and the pose of the femur in the images vary a lot. For each image, a manual annotation of 65 landmarks as in Figure 1(a) is available. In the following we performed two-fold cross-validation experiments, averaging results from training on each half of the data and testing on the other half.

### 3.1 Global Search: Automatic Femur Detection

We set up a detector that samples the whole proximal femur and three regions of interest (shaft, femoral head, greater trochanter). For each of the latter, we train a RF of 10 trees using samples at 20 random pose and scale displacements.

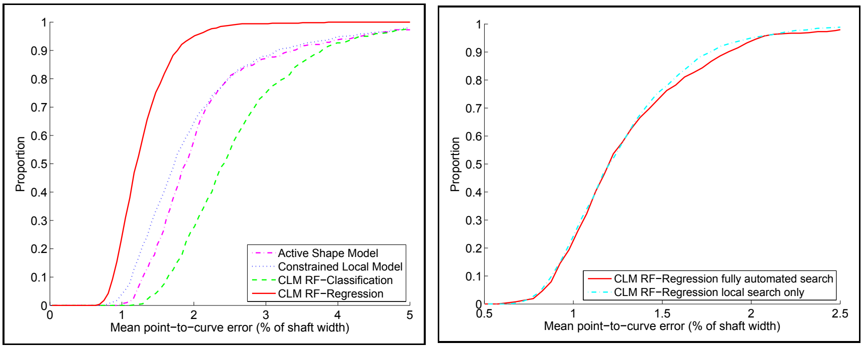
During search, the object detector scans the image at a range of coarse orientations and scales, and provides the 40 best fits. Each match determines candidate positions for points 16 and 43 (see Figure 1), defining a reference length. All candidates are clustered using a cluster radius of 10% of the reference length. We evaluate the mean point-to-point error as a percentage of the reference length, and give results for the *best* (minimal mean error) cluster only. When averaging over both reference points, the detector yields an error of less than 11.4%

for 95% of all 519 images. Our data set contains 15 calibrated images suggesting an average reference length of 57$mm$. Using the latter, the error of the global search relates to less than 6.5$mm$ for 95% of all images.

### 3.2 Local Search: Accurate Femur Segmentation

We train a RF regression-based CLM using a reference frame that is 200 pixels wide and a patch size of 15x15 pixels within the reference frame. For each training image and every landmark, we sample 20 patches using random displacements of up to 20 pixels in $x$ and $y$ in the reference image, as well as random displacements in scale (±5%) and rotation (±6°). We train a RF of 10 trees for every landmark.

To compare the performance of the RF regression-based CLM with alternative techniques, we train a correlation-based CLM and a RF classification-based CLM using the same settings, as well as an ASM. All models are trained to explain 95% of the shape variation given by the training set, and start searching from the mean shape at the correct pose. Figure 2(a) shows the mean point-to-curve error as a percentage of the shaft width. We define the latter as the distance between landmarks 0 and 64 (see Figure 1). We use this as a reference length as it tends to be relatively constant across individuals; our calibrated subset suggests an average length of 37$mm$. Results show that the RF regression-based CLM performs best with a mean point-to-curve error of within 2.0% for 95% of all images, which relates to a local search accuracy of within 0.7$mm$.



(a) (b)

**Fig. 2.** Quantitative evaluation: (a) local search results starting from mean shape at true pose; (b) fully automated search showing results for the best clustered candidate

### 3.3 Full Search: Accurate Automatic Femur Segmentation

For the fully automated system, we use the clustered candidates obtained via the global search to initialise the local search. Every candidate predicts the positions of points 16 and 43. This initialises the scale and pose of the RF regression-based CLM. We test all candidates for every image, and run 20 search iterations from the initialised mean model. We choose the candidate that gives the best final quality-of-fit value to give the fully automatic segmentation result.

Figure 2(b) gives the mean point-to-curve error of the fully automated system as a percentage of the shaft width. This shows that the global search works sufficiently well for the fully automated system to be very accurate with errors of less than 2.1% for 95% of all 519 images, relating to $0.8mm$. The overlapping plots indicate that the fully automated system yields almost equally high accuracy as a local search starting from the mean shape at the correct pose.

Figure 3 shows various segmentation results of the fully automated system, ranked according to mean point-to-curve percentiles: (a) gives the median result (50% of the images have a mean error of less than $0.5mm$); (b) is based on the second highest global search error yielding a mean segmentation error of $0.7mm$; (c)-(d) show the two highest mean segmentation errors where (c) achieved an accuracy of $1.6mm$ and (d) is the only case out of 519 images where the global search failed to initialise the local search sufficiently well.

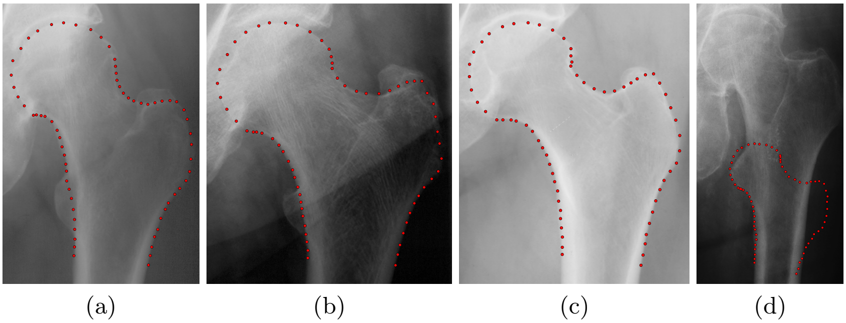


**Fig. 3.** Examples of segmentation results of the fully automated system (sorted by the mean point-to-curve percentiles): (a) median; (b) 92.1%, based on second highest global search error; (c) 99.8%, second highest overall error; (d) maximal overall error, only example where global search failed to sufficiently initialise the local search. (Due to space we only show the proximal femur; all searches were run on full pelvic images.)

A direct comparison to other reported results seems difficult as most findings are either given qualitatively, or are not easy to interpret in more general terms. The best reported results appear to be the ones by Pilgram et al. [11] with a point-to-curve error of within $1.6mm$ for 80% of the 117 test cases (estimated on the basis of likely shaft width relative to image width).

## 4 Discussion and Conclusions

We have presented a system to segment the proximal femur in AP pelvic radiographs which is fully automatic, does not make any assumptions about the femur pose, and is very accurate. We have shown that the system achieves excellent performance when tested on a set of 519 images of mixed quality. The femur detector, achieving an accuracy of a mean point-to-point error of less than $8.4mm$

for 99% of all images, works generally sufficiently well to initialise the local model used for segmentation. In our experiments, the fully automatic segmentation system achieved an overall mean point-to-curve error of less than $1mm$ for 98% of all images. We believe that this is the most accurate fully automatic system for segmenting the proximal femur in AP pelvic radiographs so far reported.

All experiments were run on a 3.3 GHz Intel Core Duo PC using 2GB RAM. The global search took on average 15s per image, and the local search 10s per image and cluster; we searched on average 10 clusters. Note that running times vary depending on image size and search settings. The fully automated system is sufficiently general to be applied to other medical segmentation problems.

**Acknowledgements.** The arcOGEN Consortium is funded by Arthritis Research UK and C. Lindner by the Medical Research Council.

## References

1. Behiels, G., Maes, F., Vandermeulen, D., Suetens, P.: Evaluation of image features and search strategies for segmentation of bone structures in radiographs using Active Shape Models. Medical Image Analysis 6(1), 47–62 (2002)
2. Breiman, L.: Random forests. Machine Learning 45, 5–32 (2001)
3. Cootes, T., Ionita, M., Lindner, C., Sauer, P.: Robust and accurate shape model fitting using random forest regression. Tech. Rep. 2012-01, Uni. Manchester (2012)
4. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active Appearance Models. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1407, pp. 484–498. Springer, Heidelberg (1998)
5. 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)
6. 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 (2011)
7. Cristinacce, D., Cootes, T.: Automatic feature localisation with Constrained Local Models. Journal of Pattern Recognition 41(10), 3054–3067 (2008)
8. Ding, F., Leow, W.-K., Howe, T.S.: Automatic Segmentation of Femur Bones in Anterior-Posterior Pelvis X-Ray Images. In: Kropatsch, W.G., Kampel, M., Hanbury, A. (eds.) CAIP 2007. LNCS, vol. 4673, pp. 205–212. Springer, Heidelberg (2007)
9. Gall, J., Lempitsky, V.: Class-specific Hough forests for object detection. In: CVPR, pp. 1022–1029. IEEE Press (2009)
10. 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)
11. Pilgram, R., et al.: Knowledge-based femur detection in conventional radiographs of the pelvis. Computers in Biology and Medicine 38, 535–544 (2008)
12. Smith, R., Najarian, K., Ward, K.: A hierarchical method based on active shape models and directed Hough transform for segmentation of noisy biomedical images. BMC Medical Informatics and Decision Making 9(suppl. 1), 2–12 (2009)
13. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: CVPR, pp. 511–518. IEEE Press (2001)