

Automatic segmentation of lumbar vertebrae on digitised radiographs using linked active appearance models

M.G. Roberts^a, T.F. Cootes^a, J.E. Adams^a

^aDepartment of Imaging Science and Biomedical Engineering
University of Manchester
Manchester, M13 9PL, UK

*

Abstract. Manual point placement for vertebral morphometry is time-consuming and imprecise, and morphometric methods for vertebral fracture diagnosis are unreliable. Automatic computer determination of the detailed vertebral shape could enable more powerful quantitative classifiers of osteoporotic vertebral fracture. The shape and appearance of vertebrae on 250 digitised lumbar radiographs were statistically modelled, using a sequence of active appearance models (AAMs) of overlapping triplets of vertebrae. To automatically locate the vertebrae, the sequence of models was matched to previously unseen scans. Accuracy results (0.64mm mean point-to-line error) were found to be similar to previously published results for dual-energy X-ray absorptiometry (DXA), but a low fracture prevalence meant that the shape models were undertrained for the few moderate and severe fractures. However mild fractures were fitted with good accuracy (mean 0.84mm). The results confirm the feasibility of substantially automating vertebral morphometry measurements on radiographs, despite the projective effects of the divergent X-ray beam. Use of the shape and appearance parameters of the models could in future provide a quantified form of some of the more subtle aspects of visual or semi-quantitative expert reading of vertebral fractures.

1 Introduction

Osteoporosis is a progressive skeletal disease characterised by a reduction in bone mass, resulting in an increased risk of fractures. Vertebral fractures are the most common, and occur in younger patients. The presence of vertebral fractures significantly increases the risk of further vertebral and non-vertebral fractures [1, 2]. The accurate identification of prevalent vertebral fractures is therefore clinically important, and the detection of incident vertebral fractures is important in evaluating new osteoporosis therapies. However there is no precise definition of exactly what constitutes a vertebral fracture, though a variety of methods of describing them have been developed [3]. These include semi-quantitative methods [4] involving some subjective judgement by an expert radiologist, and fully quantitative morphometric methods [5, 6]. The latter require the manual annotation of six (or more) points on each vertebra. This annotation is time consuming, and subtle shape information is lost in the reduction of shape to 6 points. Our ultimate aim is to define more reliable quantitative fracture classification methods based on a complete definition of the vertebra's shape and the surrounding texture. The first step must therefore be to achieve a reliable automatic segmentation. Some success in automatically locating vertebrae has been reported by several authors [7–10]. Most previous work on model-based vision applied to vertebrae has used DXA images, though [9, 10] used lumbar radiographs. DXA has many advantages, such as low radiation dose, and a lack of projective effects, but despite the growing use of DXA for vertebral fracture assessment, radiographs (or computed radiography) remain the definitive means of diagnosis.

We have already developed Active Appearance Model (AAM) [11, 12] based algorithms for segmenting vertebrae, and successfully applied these to DXA images [8, 13]. The purpose of this study was to assess the accuracy of this approach to segmenting vertebrae on radiographs. Although radiographs typically have better resolution and signal to noise ratio, the shape and appearance of the vertebrae is more complex due to projectional parallax effects. The divergent beam used in conventional radiography causes a variable scaling across the image, and can cause severe apparent tilting of the vertebral bodies. Also as the more extreme vertebral bodies tend to be obliquely irradiated, their superior and inferior endplates typically appear as elliptical rims, rather than the more linear edge typical of DXA. Figure 1 shows a typical lumbar radiograph, with some contrast enhancement to ensure all vertebrae are simultaneously visible.

2 Materials and Methods

2.1 Data

The images used were obtained from radiographs collected in a previous epidemiological study. We have thoracic and lumbar radiographs, but have initially just used lumbar radiographs as these are the more straightforward case

*martin.roberts@manchester.ac.uk

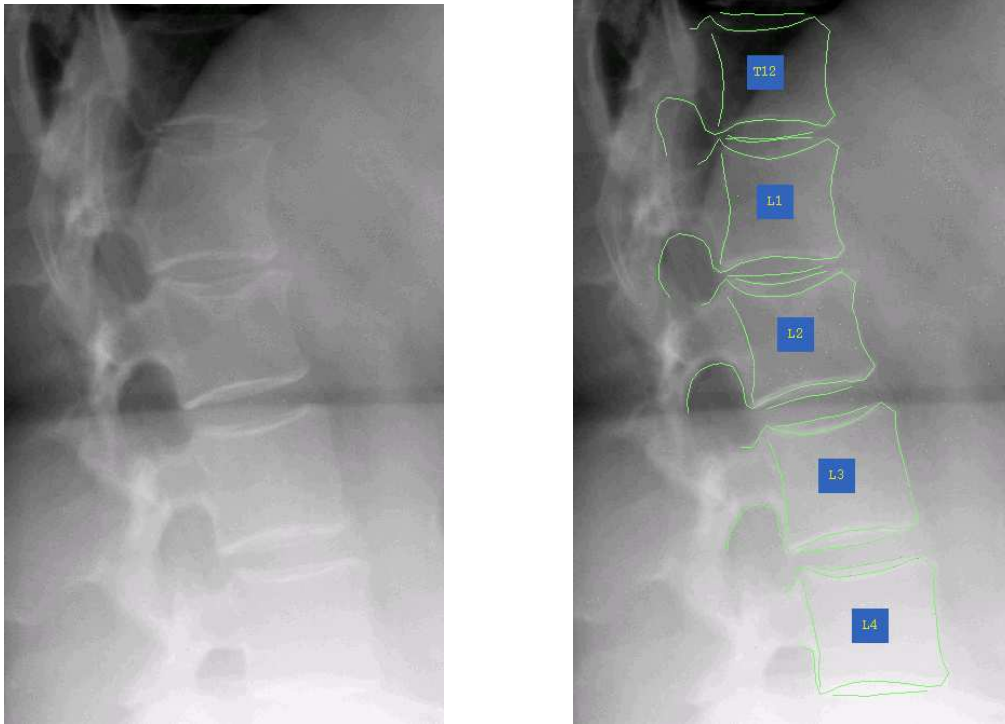


Figure 1. Lumbar radiograph. a) shows the raw image (contrast enhanced); b) shows the automatically located vertebral contours superimposed.

due to less clutter (e.g. from lungs and ribs), and a lower fracture prevalence. The dataset consisted of 250 lumbar radiographs, digitised using a Vidar ¹ Diagnostic Pro Advantage digitiser at 300dpi and 12 bit intensity resolution. This Vidar digitiser allows a variety of analogue to digital conversion mappings. As there is typically a large range of brightness/contrast at different vertebral levels in the radiographs it was important to select a transform that preserved information across a large dynamic range. The default logarithmic transform did not work well on these images, as it typically “washed out” the often brighter vertebrae in the lower lumbar, whereas using a more nearly linear transform had the opposite effect of losing information in the typically darker upper portion (T12/L1). After some initial experimentation it appeared that the “power 3” ² option gave the best compromise performance.

The digitised images were manually annotated using an in-house tool, by an experienced radiographer, supervised by the first author. Each vertebral contour uses 60 points around the vertebral body with 8 further points around the pedicles. The endplate rims were modelled using a quasi-elliptical shape, rather than the single edge previously used for DXA images. No images were included where the projectively induced tilting was so severe that lumbar vertebrae appeared to interpenetrate each other (with the occasional exception of the extreme T12/L1 or L4/L5 pairs). Such images essentially represent a setup error, and are extremely difficult to read, even by an expert radiologist, and lead to unreliable diagnosis. Figure 2 shows a zoomed in view of L3 with its shape model points displayed.

2.2 AAM approach

The method of [13] was used to fit a sequence of three AAMs composed of overlapping vertebral triplets covering the spine from L4 up to T12. Note that L5 is not normally used in vertebral fracture assessment as it is very rare for L5 to suffer osteoporotic fracture, and it may be obscured by the iliac crest. The three triplet models used were T12/L1/L2, L1/L2/L3 and L2/L3/L4. Results on DXA [8, 13] indicate that triplets of vertebrae are the optimal structure to model. If the modelled structure is too large then it is more prone to undertraining problems and latent non-linearities, and does not cope with local displacement pathologies such as scoliosis; whereas a smaller structure is too unconstrained. The algorithm of [8] combines the results of multiple sub-models with overlapping regions by fitting the models in a sequence, and using the constrained AAM [14] with constraints associated with the overlapping points. The extension to this algorithm in [13] used a “best-fit first” heuristic to obtain a close-to-optimal ordering of the model fitting sequence. In effect each (non-extreme) vertebrae is fitted using the triplet sub-model in which it is central, and the

¹Vidar Systems Corp, Herndon VA, USA

²manufacturer’s designation, in fact it appears to be a cube root

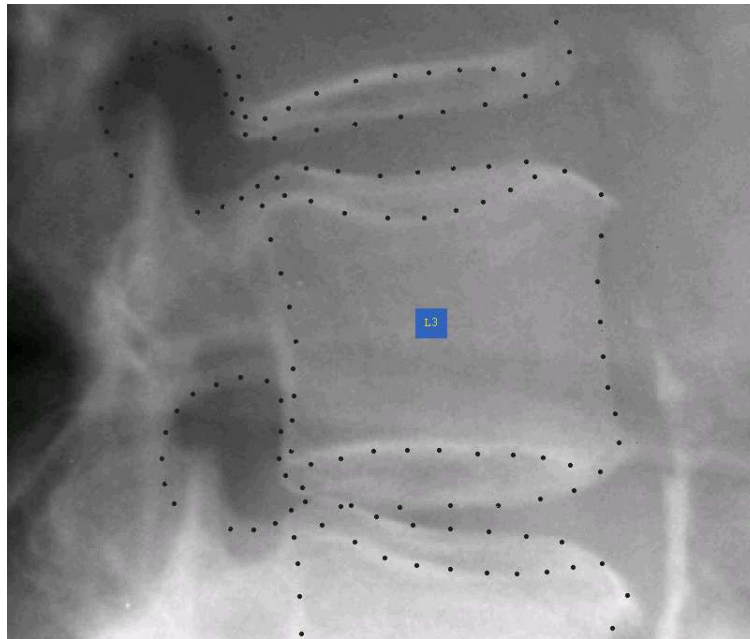


Figure 2. Zoomed in view of L3 showing its shape model points

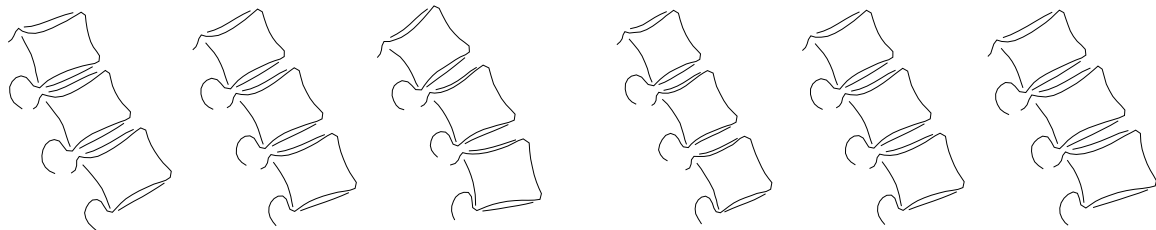


Figure 3. L2 triplet 3SD variation in first (left) and second (right) shape modes

neighbours provide helpful constraints and linkage to the other sub-models. See [8] for details. Each triplet sub-model has its own affine pose parameters. Figure 3 shows the variation in the first two shape modes of the L2-centred triplet.

T12 was included to form the uppermost L1-centred triplet, although there were some lumbar radiographs in which T12 was not fully visible. Nevertheless in general T12 should be visible on a lumbar radiograph, and results from DXA lead us to believe that it is helpful in fitting L1 to also include the neighbouring T12 in the model. In fact sometimes T12 is better visualised on the lumbar radiograph than on the thoracic. There was often a high variation in brightness and contrast across the different vertebral levels. For example T12 or even L1 were often very dark, and could often not be seen without some local contrast optimisation, whereas L4 typically had an over-bright “washed-out” appearance. Figure 1 is typical in this respect. Another advantage of decomposing the overall shape into sub-structures is that the texture normalisation can be better tuned to the local brightness and contrast, where there is substantial variation in these across the image.

As there is little useful information inside the vertebral body we used profile samplers for the AAM texture model, rather than the triangulated region samplers classically used with an AAM. The profile samplers extracted the gradient perpendicular to the local shape, and this was non-linearly renormalised using a sigmoidal function tuned to the mean absolute gradient [15] over the entire profile set. We used a 4-level multi-resolution pyramid search, to extend the convergence zone, with 8 samples either side of the shape. The finest level step size was 0.375mm, and the images were pre-smoothed up to a resolution of 0.1694 mm per pixel (i.e. one level of Gaussian pyramid up-smoothing). Thus the profile step size represents about 2 pixels (at each level of the pyramid). The extracted gradient is Gaussian smoothed across the local tangent, with a smoothing window equal to the step length. We also experimented with a profile sampler which concatenated this profile with a similar profile sampler extracting a measure of image corner strength

Profile Sampler	Normal				Fractured			
	Mean Acc	Median Acc	75%ile Acc	%ge errors over 2mm	Mean Acc	Median Acc	75%ile Acc	%ge errors over 2mm
Gradient Only	0.71	0.46	0.89	6.2%	1.11	0.62	1.34	14.1%
Gradient & Corner	0.64	0.43	0.82	4.6%	1.06	0.61	1.32	13.3%

Table 1. Search Accuracy Percentiles by Fracture Status for the two profile samplers used

(“corneriness”), as in [15]. As the corners of the vertebrae are of physical interest in standard morphometry, it was thought that including a corneriness measure in the AAM might improve the accuracy at points of important diagnostic interest. Furthermore the projective parallax and oblique beam orientation tend to introduce curved features in the region of the profile. The corneriness measure has the further advantage that it implicitly includes feature information from a somewhat larger region, as the measure is based on the structure tensor $(\nabla I \nabla I^T)^3$, which is Gaussian smoothed over a square region with semi-width twice the profile step length. See [15] for details.

2.3 Experiments

Leave-25-out tests were performed over the 250 images. As AAMs perform local search an approximate initialisation somewhere in the vicinity of the vertebrae is needed. When the algorithm is run interactively in an associated prototype clinical tool, the clinician initialises the solution by clicking on the approximate centres of each vertebra. A global shape model is then used to give an approximate starting solution with best least squares fit of its vertebral centres to the input points. The same method was used in this study, and on each experiment the user imprecision was simulated by using the known equivalent marked points and adding random offsets to them. These were zero-mean Gaussian errors with SD of 2mm in the y-direction (along the spine) and 3mm in the x-direction. Twenty replications (i.e. random initialisations) of each image were performed.

3 Results

The accuracy of the search was characterised by calculating the absolute point-to-line distance error for each point on the vertebral body. Table 1 compares results for the two profile samplers used with the data separated into points within normal or fractured vertebrae. Each row gives the mean, median and 75th percentiles, and the percentage of point errors in excess of 2mm. The threshold of 2mm would be around 2.5 SDs of manual precision, and can be viewed as a point failure indicator.

Table 1 shows that the results are worse for fractured than normal vertebrae. A more detailed examination by fracture grade gives mean accuracies of 0.84mm, 1.79mm and 3.35mm for fracture grades 1, 2 and 3⁴ respectively (for the gradient and corner sampler). However there was a low fracture prevalence in the lumbar region in the sample, and these figures are based on 17 grade 1 fractured vertebrae, 2 grade 2 and only a single grade 3 fracture.

The more sophisticated profile sampler including a corner measure appears to produce a small improvement in accuracy of around 0.07mm. This represents a 10% reduction in mean error. We confirmed that this difference is statistically significant at the 1% level by bootstrap resampling of the differences in errors between the two profiles. This enables the derivation of a symmetric (in probability) 99% bootstrapped confidence interval on the mean difference of [0.048,0.082]. As this interval does not span zero, the difference is significant at the 1% level.

4 Discussion

4.1 Overall Accuracy Performance

The mean segmentation accuracy of 0.64mm on normal vertebrae is comparable to manual precision in point placement, and to previous results on DXA images [13]. Over 95% of points in normal vertebrae are located to within 2mm of the manually annotated outline. However the dataset contained a very low prevalence of fractured vertebrae, and so the shape models are evidently undertrained, for fractures above grade 1. Therefore within the limitations of the small fractured sample it appears that the results deteriorate with increasing fracture grade. However given previous reasonable accuracy achieved on fractured vertebrae with DXA images [13], we believe that this problem could be

³with the Cartesian image gradient ∇I as a column vector

⁴i.e. mild, moderate and severe fractures, see [4]

solved by adding more fractured training examples. The mean accuracy is better than other comparable cited figures in the literature [9, 10]. For example de Bruijne *et al* [10] obtained a mean point-to-contour accuracy of 1.4mm on lumbar radiographs using shape particle filtering, which is more than double the size of error achieved by our AAM approach. On the other hand this was for a fully automatic search with no approximate manual initialisation such as we use. Howe *et al* [9] state that 68% of points on lumbar radiographs were located to within 25 pixels. We understand the dataset used had a resolution of 0.174mm per pixel, so this is equivalent to a 68th percentile of 4.35mm, clearly substantially worse than our 75th percentile of 0.85mm. However again Howe *et al* were using a completely automatic method, with the AAM being initialised to the best template match found by an initial Generalised Hough Transform.

4.2 Conclusion

In conclusion the results confirm the feasibility of substantially automating vertebral morphometry measurements on radiographs, although the shape models need better training on fractured vertebrae. Within the limitations of the dataset, the projective effects of spinal radiography do not appear to present any substantial problem to an AAM-based approach.

4.3 Future Work

We intend to extend the work to the thoracic spine, which tends to contain more osteoporotic fractures. Use of the shape and appearance parameters of the fitted models could in future provide a means of classifying vertebrae as normal, fractured, or otherwise deformed. Current simplistic quantitative morphometric methods are unreliable, especially for mild fractures, but the appearance parameters may provide a quantified form of some of the more subtle aspects of visual or semi-quantitative expert reading of vertebral fractures. We therefore view obtaining a reliable automatic segmentation as the first step in achieving a Computer Aided Diagnosis (CAD) system for vertebral fracture.

5 Acknowledgements

The authors would like to thank Stephen Capener who performed the manual annotation of the vertebrae.

References

1. D. Black, N. Arden, L. Palermo et al. "Prevalent vertebral deformities predict hip fractures and new vertebral deformities but not wrist fractures." *Study of Osteoporotic Fractures Research Group, J Bone Miner Res* **14**, pp. 821–828, 1999.
2. L. Melton, E. Atkinson, C. Cooper et al. "Vertebral fractures predict subsequent fractures." *Osteoporosis Int* **10**, pp. 214–221, 1999.
3. A. Guermazi, A. Mohr, M. Grigorian et al. "Identification of vertebral fractures in osteoporosis." *Seminars in Musculoskeletal Radiology* **6(3)**, pp. 241–252, 2002.
4. H. Genant, C. Wu, C. van Kuijk et al. "Vertebral fracture assessment using a semi-quantitative technique." *J Bone Miner Res* **8**, pp. 1137–1148, 1993.
5. R. Eastell, S. Cedel, H. Wahner et al. "Classification of vertebral fractures." *J Bone Miner Res* **6(3)**, pp. 207–215, 1991.
6. E. McCloskey, T. Spector, K. Eyres et al. "The assessment of vertebral deformity: a method for use in population studies and clinical trials." *Osteoporosis Int* **3**, pp. 138–147, 1993.
7. P. Smyth, C. Taylor & J. Adams. "Vertebral shape: automatic measurement with active shape models." *Radiology* **211**, pp. 571–578, 1999.
8. M. Roberts, T. Cootes & J. Adams. "Linking sequences of active appearance sub-models via constraints: an application in automated vertebral morphometry." In *14th British Machine Vision Conference*, pp. 349–358. 2003.
9. B. Howe, A. Gururajan, H. Sari-Sarraf et al. "Hierarchical segmentation of cervical and lumbar vertebrae using a customized generalized hough transform and extensions to active appearance models." In *Proc IEEE 6th SSIAP*, pp. 182–186. 2004.
10. M. de Bruijne & M. Nielsen. "Image segmentation by shape particle filtering." In *International Conference on Pattern Recognition*, pp. 722–725. IEEE Computer Society Press, 2004.
11. T. Cootes & C. Taylor. "Statistical models of appearance for medical image analysis and computer vision." *Proc SPIE Medical Imaging* **3**, pp. 138–147, 2001.
12. T. Cootes, G. Edwards & C. Taylor. "Active appearance models." In *5th European Conference on Computer Vision*, volume 2, pp. 484–498. Springer (Berlin), 1998.
13. M. Roberts, T. Cootes & J. Adams. "Vertebral shape: Automatic measurement with dynamically sequenced active appearance models." In *8th MICCAI Conference*, volume 2, pp. 733–740. 2005.
14. T. Cootes & C. Taylor. "Constrained active appearance models." In *8th International Conference on Computer Vision*, volume 1, pp. 748–754. IEEE Computer Society Press, July 2001.
15. I. Scott, T. Cootes & C. Taylor. "Improving active appearance model matching using local image structure." In *18th Conference on Information Processing in Medical Imaging*, pp. 258–269. 2003.