

Segmentation of Lumbar Vertebrae Using Part-Based Graphs and Active Appearance Models*

Martin G. Roberts¹, Tim F. Cootes¹, Elisa Pacheco¹, Teik Oh¹,
and Judith E. Adams¹

School of Imaging Science, University of Manchester, U.K.
`martin.roberts@manchester.ac.uk`

Abstract. The aim of the work is to provide a fully automatic method of segmenting vertebrae in spinal radiographs. This is of clinical relevance to the diagnosis of osteoporosis by vertebral fracture assessment, and to grading incident fractures in clinical trials. We use a parts based model of small vertebral patches (e.g. corners). Many potential candidates are found in a global search using multi-resolution normalised correlation. The ambiguity in the possible solution is resolved by applying a graphical model of the connections between parts, and applying geometric constraints. The resulting graph optimisation problem is solved using loopy belief propagation.

The minimum cost solution is used to initialise a second phase of active appearance model search. The method is applied to a clinical data set of computed radiography images of lumbar spines. The accuracy of this fully automatic method is assessed by comparing the results to a gold standard of manual annotation by expert radiologists.

1 Introduction

The accurate identification of prevalent vertebral fractures is clinically important in the diagnosis of osteoporosis. However there is no precise definition of exactly what constitutes a vertebral fracture, though a variety of methods of describing such fractures have been developed [1]. These include semi-quantitative methods, requiring some subjective judgement by an expert radiologist; and fully quantitative morphometric methods. The latter require the manual annotation of six (or more) points on each vertebra. The manual marking is time consuming, and such methods are lacking in specificity. More sophisticated classification methods based on statistical models have been reported in [2,3], but these require an accurate segmentation method. Active appearance models [4] (AAM) have been used to segment dual energy X-ray absorptiometry (DXA) images in [5], but the method required a manual initialisation on the centre of each vertebra. Also in clinical practise the use of spinal radiographs is still the gold standard, despite the increasing use of DXA. Radiographs are more challenging as the fan beam

* We are thankful to the UK Arthritis Research Council for funding.

used in conventional radiography can lead to parallax errors and apparent scale changes. Some success in automatically locating vertebrae in lumbar radiographs has been reported in [6].

In this work we apply an AAM-based approach similar to [5] to computed radiography images of the lumbar spine. In applications such as large clinical drug trials, it is desirable to eliminate the manual initialisation of the AAM used in [5]. This paper describes a two-phase approach. First a global search is conducted using a set of feature detectors combined with a graph model of their pair-wise geometry. This is then used to initialise a set of AAMs which provide a more detailed segmentation.

A potential problem in the case of the spine is that many vertebrae have a similar appearance. A natural approach is thus to use a parts+geometry model [7,8] as is widely used in the object recognition literature [9]. This work is partly inspired by that of Donner et al.[10], who demonstrate that sparse local models, together with a network of inter-part relationships, can be very useful for locating structures in medical images.

The novelty of the work lies in: a) the combined use of the parts+geometry model with an AAM search, thus providing a fully automatic segmentation method for vertebrae; b) substantial evaluation on a clinical and challenging dataset of spinal radiographs.

2 Methods

2.1 Multi-resolution Patch Models

We first describe the components of the models of parts and geometry used for initial vertebrae location. Given a set of training images in which a particular region has been annotated, we can construct a statistical model of the region. Firstly gross brightness and contrast variation are removed by normalising each pixel using the locally smoothed mean and variance, derived by exponential smoothing with a 50% response radius of 4mm. We then use a simple oriented rectangle, centred on a point, with the axis direction and scales defined by other annotated points (see figure 1).

We search new images using this model by running normalised correlation at a range of positions, orientations and scales in an exhaustive search. To reduce the combinatorial complexity, we perform a multi-resolution search. The coarser resolution models sub-sample smoothed versions of the image, thus covering the same image region with a reduced number of pixels. A fully exhaustive search for local optima is performed only at the coarsest scale to get a set of plausible candidates. Then smaller regions around these candidates are searched, to refine each candidate at the finer resolutions. Only the best candidate is retained when multiple candidate patches overlap. This coarse-to-fine approach usually results in between 10-40 candidates.

We used 6 patches (parts) for each lumbar vertebra: the four corners, a patch on the curved inferior junction of the vertebra with the pedicle; and a wide patch across the spine aligned with the lower pedicle. Note the latter two can

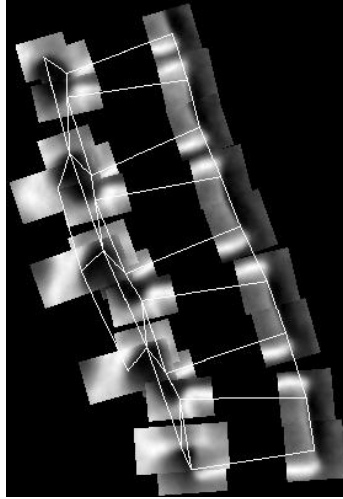


Fig. 1. Mean normalised patches with arc links

overlap somewhat. We chose corner features as they tend to be more distinctive in appearance than other edges. Also the middle of the vertebral endplates tend to have a much more variable appearance due to both parallax tilting (a “bean-can” appearance), and collapse due to fracture. It is doubtful whether normalised correlation of whole vertebral edges would be reliable due to the intrinsic variability of appearance, whereas the corners tend to be more stable in appearance. The curvature of the pedicle join also has a distinctive appearance, and sections of the spine are typically the brightest thing in the image, and so give coarse lateral positional clues.

The node cost of each candidate patch is assigned using a normalised mean absolute deviation from the training set mean patch. For each patch pixel vector \mathbf{g} with mean μ_g and standard deviation σ_g (over the patch), we compute the normalised vector of pixels \mathbf{v} , where $\mathbf{v} = \frac{\mathbf{g} - \mu_g}{\max(\sigma_g, \sigma_{min})} \mathbf{1}$.

The mean of each normalised patch is calculated over the training set, to provide a set of patch means $\bar{\mathbf{v}}$. For ease of subscript notation we consider one such patch form. A candidate $\mathbf{v}^{(k)}$ in image j is assigned a cost C_{jk} thus by summing over the absolute deviations from the mean of each of n pixels, with the normalisation factors $\{\hat{\sigma}\}$ evaluated over the training set of N images:

$$C_{jk} = \sum_{i=1}^n \frac{|v_{ji}^{(k)} - \bar{v}_i|}{\hat{\sigma}_i}, \quad \hat{\sigma}_i = \frac{1}{N} \sum_{j=1}^N |v_{ji} - \bar{v}_i| \quad (1)$$

Finally the actual cost assigned is further normalised as

$c_{jk} = \frac{C_{jk}}{\alpha}$ where alpha is the standard deviation of $\{C_j\}$, $j \in \{1 \dots N\}$ over the training set. Note that this normalisation step ensures that the distribution of

¹ Where σ_{min} is a lower-bound, to avoid ill-conditioned behaviour in flat regions.

c (over correctly aligned patches) has approximately unit variance. This ensures a compatible scale for the node and arc costs in the subsequent graph model.

2.2 Geometric Relationships

There are many responses for each part, some of them being duplicated over the similar-looking vertebrae. To resolve the ambiguity, we use a geometric model containing multiple nodes (one per part), together with a model of the pairwise geometrical relationships between them. This is a widely used technique [7,8,10]. We can then use graph algorithms to locate the optimal solution for the combination of feature response candidates and pairwise geometry. The set of vertebral corner and spine features we used form a natural grid with arc set E (see figure 1).

Geometric Constraints. Each candidate response for part r has a central reference point \mathbf{p}_r , a scale s_r and an orientation θ_r . The geometric constraints between parts r and q can be represented in the cost function $f_{rq}(\mathbf{p}_r, s_r, \theta_r, \mathbf{p}_q, s_q, \theta_q)$. In general this could be based on a joint probability density function of the parameters.

To allow for the varying scale and orientation between patch candidates, we use similarity transforms. Let the similarity transformation $T(\mathbf{x} : \mathbf{d}, s, \theta)$ apply to \mathbf{x} a scaling of s , a rotation of θ , followed by a translation of \mathbf{d} . The corrected relative position of a candidate for patch q is obtained by mapping the position of \mathbf{p}_q into the co-ordinate frame defined by a candidate for r using $\mathbf{t}_{rq} = T^{-1}(\mathbf{p}_q : \mathbf{p}_r, s_r, \theta_r)$.

We compute the 2D mean and variances of the relative position defined by the positional components of $\mathbf{t}_{rq}(\mathbf{p}_q)$ across a training set. We then use the Mahalanabis distance for the cost function f_{rq} . The node cost $c_r(k_r)$ of candidate k_r is obtained as for the $\{c_{jk}\}$ above, but we now drop the image subscript j and use an explicit parts subscript r . The final selection of optimal solution $\tilde{\mathbf{k}}$ from the set of candidate patches $\{P_{rk}\}$ is given by finding the minimum sum of node and arc costs:

$$\tilde{\mathbf{k}} = \operatorname{argmin}_{\mathbf{k}} \left\{ \sum_{r=1}^R c_r(k_r) + \sum_{(r,q) \in E} f_{rq}(k_r, k_q) \right\} \quad (2)$$

We trivially convert this to a maximisation problem and use a loopy belief propagation algorithm [11]² to solve for the optimal candidate vector $\tilde{\mathbf{k}}$. Note the ‘‘loopy’’ nature of the grid topology means that this is not guaranteed to converge to the true global optimum.

2.3 Data

We obtained anonymised computed radiography (CR) lumbar images from a local hospital, from patients having had spinal radiography over the previous

² The max-product variant, equivalent to max-sum with log probabilities.

12 months, and with the approval of the local ethics committee. We selected a training subset of 135 images, particularly including those images with evidence of vertebral fracture, to try and ensure a sufficient training of the AAM shape models to cope with fractured cases.

The images were manually annotated using an in-house tool by an experienced radiologist. 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 to cope with the dual-edge appearance induced by projective tilting.

2.4 Active Appearance Models

The linked AAM approach of [5] 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. The AAM combination algorithm also uses a global shape model of the entire spine, but this is used to guide the initialisation of later sub-models in the fitting sequence (given the earlier sub-model solutions). T12 was included to form the uppermost L1-centred triplet.

As there is little useful information inside the vertebral body we used profile gradient samplers for the AAM texture model, as was used on DXA data [5], with a non-linear renormalisation using a sigmoidal function tuned to the mean absolute gradient. We used a 4-level multi-resolution pyramid search to extend the convergence zone.

3 Experiments

3.1 Feature Detection

A train/test split was performed by picking out 10 images in turn from the CR data. The patch correlators were trained on the remaining CR data, and similarly the means and variances of the relative patch poses were calculated in order to construct relative position functions for the arc costs. The patch correlation followed by graph optimisation was then run on the respective test subsets.

There is a fundamental ambiguity in determining the vertebral levels. This was worsened by the fact that our modelling does not yet include L5, as L5 is not normally assessed for osteoporosis. We therefore cropped the images at 20 mm below L4, which tends to include about half of L5. Because the lumbar images typically include several thoracic vertebrae (e.g. up to T10), we found that the optimal solution is frequently displaced by one vertebral level (i.e. L3 appears to be L4 etc). Even such cases represent a kind of success, as the feature detectors are still locating vertebrae. So we classified each image detection result into three categories: successful, level-displaced, and unsuccessful. Successful cases require at least 3 lumbar centres to be located within 10mm of the true centre (i.e. about half a vertebral height), so the putative centre is normally inside the

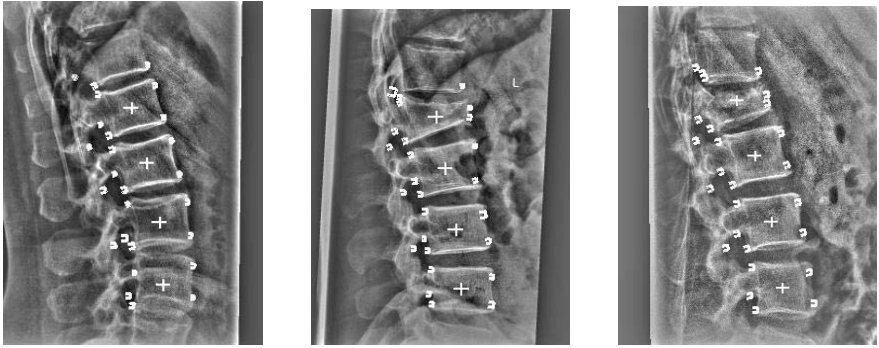


Fig. 2. Successful automatic initialisations. Crosses mark the estimated vertebral centres, other smaller 3-sided markers are drawn around located feature points. Note the L1 wedge fractures in the images ii) and iii).

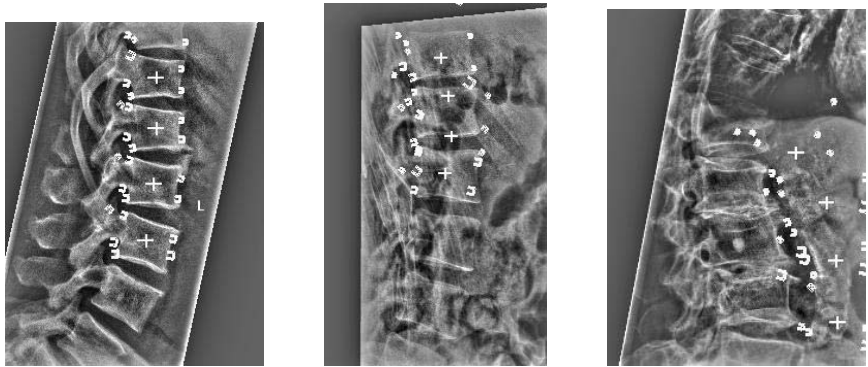


Fig. 3. Unsuccessful automatic initialisations, a)-c) (left first): a) vertebral level displaced upwards by one; b) the middle two vertebrae are also compressed into one (edge confusion); c) Completely failure due to scoliosis and anterior calcification

correct vertebra. A level-displaced case is similar but with the levels shifted by one vertebrae; and a failure case is all others. Only successful cases were subject to the secondary stage of AAM search.

3.2 AAM Search

We used a similar train/test cycle of miss-10-out to run the AAM searches. We then took the centre of the 4 lumbar vertebral corner patches and used these vertebral centres to initialise the next phase of AAM search, using a global shape model of T12-L4 (plus vertebral centres). Note that shape constraints are applied in the usual AAM manner, so this initialisation will tend to smooth out outliers. Point-to-line errors were calculated of the converged AAM positions with respect to the gold-standard radiologist manually placed points.

Table 1. Number of Images / (Percentage) for which vertebrae location algorithm succeeds

Successful	Level-Shifted	Failures
98 (73%)	27 (20%)	10 (7%)

Table 2. Search error statistics (point-to-line) for AAM

Vertebra Status	Search Error Statistic				
	Mean (mm)	Median (mm)	75%-ile mm	95%-ile mm	%ge errors >2mm
Normal	0.93	0.53	0.98	1.68	7.3%
Fractured	2.27	1.16	2.44	5.50	30.6%

4 Results

Figure 2 shows three examples of the located features (and vertebral centres), while Figure 3 shows a level-shifted result and two failures.

Table 1 gives the numbers of successful, level-displaced (by one vertebral level), and unsuccessful phase 1 localisations. It can be seen that the ambiguity of vertebral levels is a problem, but only 7% of images result in outright failure. The mean central point error of the successful cases was 3.6mm.

Table 2 gives point-to-line accuracy results from the phase 2 AAM segmentation given the automatic (“successful” only cases) initialisations. These results are decomposed into points within normal or fractured vertebrae. Data are given for the mean, median, 75th and 95th percentiles; and the percentage of point errors in excess of 2mm, which can be viewed as a point failure threshold. The overall mean error was 1.06mm.

5 Discussion and Conclusions

Although there is some degree of failure of the automatic initialisation process, the algorithm successfully locates a plausible set of vertebrae in over 90% of cases. The fundamental ambiguity of vertebral levels is a problem (20% level-shifted), but we hope that by more fully modelling L5 and the upper sacrum we will be able to better resolve this ambiguity. We may have to extend the algorithm to include optional thoracic vertebra (e.g. up to T10), and thus deal with potentially missing structure. In some applications it may be possible for the level issue to be resolved by manual pre-cropping of the images. The use of more sophisticated feature representations of the images to reduce spurious feature detections is also under investigation.

Successful cases give adequate accuracy for a secondary phase of detailed AAM segmentation. The overall mean AAM accuracy of 1.06mm is comparable to results on DXA [5], and other methods for lumbar vertebrae (e.g. [6], mean error 1.4mm). Good accuracy is obtained for normal (unfractured) vertebrae, but there

is a higher point failure rate and more skewed errors for fractured vertebrae. If severe fractures³ are excluded, then the mean error over the remaining fractured vertebrae reduces to 1.7mm (median 0.94mm); whereas for severe (grade 3) fractures it is 3.9mm. This deterioration may be partly due to undertraining of the shape models for the more severe fractures, a problem which should be resolved by more training examples.

A parts+geometry model trained on vertebral patches with a grid topology successfully located a sequence of vertebrae in 93% of images. However issues remain over the ambiguity of vertebral level. In successful cases this vertebral detection method provides adequate initialisation for an accurate AAM segmentation, but further AAM training on severe fractures is needed.

References

1. Guerhazi, A., Mohr, A., Grigorian, M., Taouli, B., Genant, H.K.: Identification of vertebral fractures in osteoporosis. *Seminars in Musculoskeletal Radiology* 6(3), 241–252 (2002)
2. de Bruijne, M., Lund, M.T., Tankó, L.B., Pettersen, P.P., Nielsen, M.: Quantitative vertebral morphometry using neighbor-conditional shape models. In: Larsen, R., Nielsen, M., Sporning, J. (eds.) *MICCAI 2006*. LNCS, vol. 4190, pp. 1–8. Springer, Heidelberg (2006)
3. Roberts, M.G., Cootes, T.F., Pacheco, E.M., Adams, J.E.: Quantitative vertebral fracture detection on DXA images using shape and appearance models. *Academic Radiology* 14, 1166–1178 (2007)
4. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 681–685 (2001)
5. Roberts, M.G., Cootes, T.F., Adams, J.E.: Vertebral morphometry: semi-automatic determination of detailed shape from DXA images using active appearance models. *Investigative Radiology* 41(12), 849–859 (2006)
6. de Bruijne, M., Nielsen, M.: Image segmentation by shape particle filtering. In: *International Conference on Pattern Recognition*, pp. 722–725. IEEE Computer Society Press, Los Alamitos (2004)
7. Fergus, R., Perona, P., Zisserman, A.: A visual category filter for google images. In: Pajdla, T., Matas, J.(G.) (eds.) *ECCV 2004*. LNCS, vol. 3021, pp. 242–256. Springer, Heidelberg (2004)
8. Felzenszwalb, P., Huttenlocher, D.: Pictorial structures for object recognition. *Int. Journal of Computer Vision* 61(1), 55–79 (2005)
9. Ponce, J., Hebert, M., Schmid, C., Zisserman, A. (eds.): *Towards Category-Level Object Recognition*. Springer, Heidelberg (2006)
10. Donner, R., Micusik, B., Langs, G., Bischof, H.: Sparse MRF appearance models for fast anatomical structure localisation. In: *Proc. British Machine Vision Conference*, vol. 2, pp. 1080–1089 (2007)
11. Weiss, Y., Freeman, W.: On the optimality of solutions of the max-product belief propagation algorithm in arbitrary graphs. *IEEE Trans. Inf. Theory* 47, 736–744 (2001)

³ i.e. grade 3 fractures in the Genant system [1], suffering over 40% height loss.