Automatic Labeling and Segmentation of Vertebrae in CT Images

Anonymous Author(s)
Affiliation
Address
email

Abstract

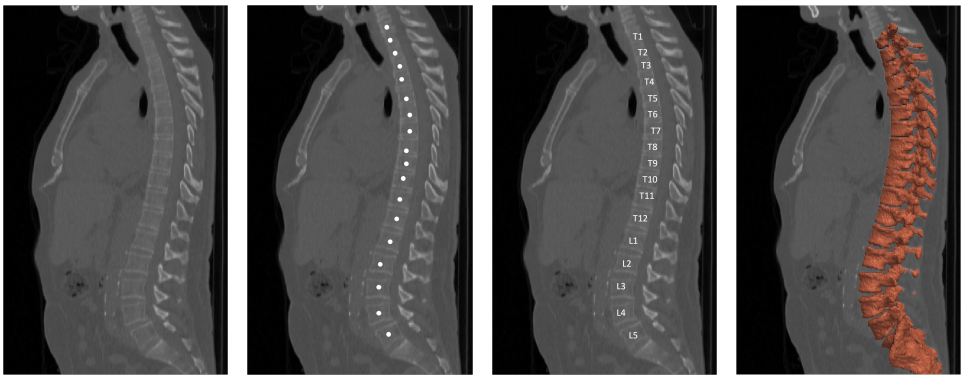
Detection, Labeling, and segmentation of the spinal column from CT images is a pre-processing step for a range of image-guided interventions. State-of-the art techniques have focused either on image feature extraction or template matching for labeling of the vertebrae followed by segmentation of each vertebra. Recently, statistical multi-object models have been introduced to extract common statistical characteristics among several anatomies. These models have also been used for joint labeling and segmentation of the lumbar spine and were shown to be robust, accurate, and computationally tractable. In this paper, I reconstruct a statistical multi-vertebrae pose+shape model and utilize it in a novel framework for labeling of an arbitrary vertebra in a CT image. I also use the model for segmentation of the entire vertebral column. I validate my technique in terms of accuracy of the labeling and segmentation of CT images acquired from 61 subjects. The vertebral column is correctly labeled in 97% of the subjects and mean distance error achieved for the segmentation is 2.1 ± 0.7 mm.

1 Introduction

Accurate segmentation and labeling (sometimes referred to as identification) of individual vertebrae from Computed Tomography (CT) images is a necessary pre-processing step in a range of image-guided therapy applications such as insertion of pedicle screws or spinal implants. It is also beneficial for many other applications that use vertebrae as anatomical landmarks.

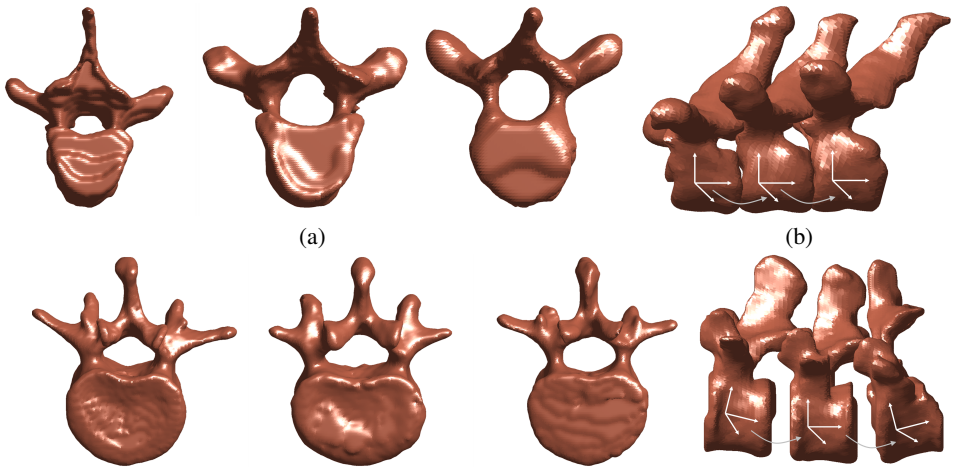
Figure 1 shows an illustration of these procedure that are performed on 3D CT images. Detection of the vertebrae is usually performed by finding the characteristic shape of the vertebrae in the CT images using Generalized Hough Transform [4], which is computationally intensive and generate many false positives. The automatic identification of the vertebrae is also challenging due to the repetitive nature of these structures and the variability of images in resolution and field-of-view. The latter becomes more important when the reference structures that are typically used for labeling (e.g., the first cervical or the first thoracic vertebra) are missing from the image. Previous research has addressed the identification problem [5, 7]. However, to the best of my knowledge, only two works have handled labeling of the vertebral column in arbitrary scans [3, 4]. Klinder *et al.* build a template (mean shape) of each vertebrae and performed affine alignment between each template and vertebrae that are detected in the CT image. They reported the best match as the label [4]. Given the large number of alignment, the identification phase was reported to take up to 45 minutes for 12 thoracic vertebrae. Glocker *et al.* addressed the identification problem by performing a random forest classification on the image features [3]. Although the results were promising, this technique might not be applicable to CT images with a small field-of-view or those that are reduced in-plane showing only a small region around the vertebrae with little structural context. Moreover, this technique only provides labeling, not segmentation.

The segmentation task remains a challenge despite the high contrast of bony structures in CT im-



(a) (b) (c) (d)

Figure 1: a) A slice of a 3D CT image which includes thoracic (T1-T12) and lumbar (L1-L5) vertebrae. b) Detection of the vertebrae on the CT images. c) Identification (labeling) of the detected vertebra. d) segmentation of the identified vertebra.



(a) (b) (c) (d)

Figure 2: a) three examples of T1 (first thoracic) vertebrae from three different patient. b) An example of T1-T2-T3 vertebrae of a patient. c) Thee examples of L1 (first lumbar) vertebrae from three different patients. d) An example of L1-L2-L3 vertebrae of a patient.

ages. This is due to the presence of unclear boundaries, the complex structure of vertebra, and substantial inter-subject variability. A rich body of literature exists on segmentation of the vertebral column from CT images [1, 4, 5, 6, 10, 11, 12]. Conventinoally, segmentation of the vertebral column is performed for the individual vertebrae separately, which has several disadvantages. A clear boundary may not exist in regions between two vertebrae such as in the intervertebral disk and facet joints, which may lead to mis-segmentation or an overlap between the segmentation of consecutive vertebrae. Moreover, some anatomically useful information is discarded, such as the common shape variations among vertebrae.

1.1 Proposed framework

Although previous works tried to solve the problem of identification by finding image feature, I believe that such complex task can be performed by shape comparison. Looking at Figures 2a and 2c, there are difference within the population of each vertebrae. However, some characteristic differences can also be detected between the shape of different vertebrae. For example, thoracic vertebrae are more stretched in certain direction or have larger transverse processes. Differences can also be detected in other aspects of the data. Looking at Figures 2b and 2d, the relative position

of vertebrae can be used for identification. The spine curvature is usually larger around Lumbar vertebrae. Given that, I propose to generate a statistical model that can capture the variations of shapes and poses among different vertebrae. I will illustrate my technique to align these model to the CT images to best fit the target vertebra. I will also demonstrate my labeling technique that uses random forest where hyper-parameters are optimized using Bayesian optimization.

2 Methods

2.1 Construction of the multi-vertebrae model

For construction of the model, pose statistics are separated from the shape statistics since they are not necessarily correlated and do not belong to the same space. Pose, which is represented by a similarity (rigid+scale) transformation, form a Lie group, G, which is a group and a differentiable manifold and thus linear analysis is not applicable. However, the exponential mapping, $\mathfrak{g} : \exp(x) \rightarrow G$, and its inverse, logarithm mapping $G : \log(x) \rightarrow \mathfrak{g}$, can be used to transfer the elements back and forth from a tangent space, $\mathfrak{g}$ (which is a linear space), defined at identity element of the group. Analogous to principal components in the Euclidean space, Principal Geodesics (PG) are defined for Lie groups. The approximation is as follows [2]: for a set of elements, $x_1, \ldots, x_n$, the mean, μ, is found using an iterative approach suggested by Pennec [8]. Principal Component Analysis (PCA) is then applied to the residuals in the tangent space at the mean, $\log(\mu^{-1}x_i)$. The results are orthonormal principal components, v_l, which give the PGs (modes of variations) by exponential mapping, $\mu \exp(v_l)$.
Assume that the training set contains N instances of an ensemble of L anatomies. A group-wise GMM-based alignment technique [9] is used to establish dense correspondences across the training set. Generalized Procrustes analysis is then used to generate the mean shape for all the anatomies, and their transformation, $\mathbf{T}_{n,l}$, to each instance. The transformation for all anatomies are concatenated and PGs are then extracted. The results are principal geodesics, which can be written for each anatomy. Common statistics between shapes are also extracted with the same technique.
Assume that θ_k^s is the weight applied to the kth shape PG and θ_k^p is applied to the kth pose PG. A new instance of the model is generated as follows:

$$S = \Phi(\theta^s, \theta^p) = \Phi^p\big(\Phi^s(\theta^s); \theta^p\big). \tag{1}$$

Note that $\Phi^p(.; \theta^p)$ and $\Phi^s(.; \theta^p)$ denote a pose and shape respectively, built by combination of the PGs with corresponding weights.
The constructed model is capable of generating any n consecutive vertebrae. In other words, the model can be used to describe the shape of an unseen target in an arbitrary CT image. The only unknowns are then the weights that are applied to shape and pose modes of variations.
For the rest of this paper, I set the number of vertebrae in the model to three (referred to as generic 3-vertebrae model hereafter), since larger number will result in larger computational time and also it is sufficient to encode the variabilities of a certain vertebrae and its relation to its neighbouring vertebrae.

2.2 Alignment of the model to CT images

The model is aligned to a target CT image using the following technique. Initially a preprocessing step is performed to extract the boundary of bony anatomies from the CT images. To this end, CT images are smoothed and thresholded with a value of 100 Hounsfield units and a Canny edge detector is applied to the 2D transverse planes. The result is a point set representing a rough segmentation of the bony anatomy. Figure 4 shows an example of such enhancement. Next, an iterative Expectation Maximization (EM) [9] alignment technique is utilized where the alignment is considered as a probability density estimation problem. The points in the surface of the model are assumed to be centroids of a Gaussian Mixture Model (GMM) and the edge points extracted from the CT image are the observations. The model is then deformed using the weghts that are applied to the modes of variation to maximize the likelihood of the GMM generating the observations.

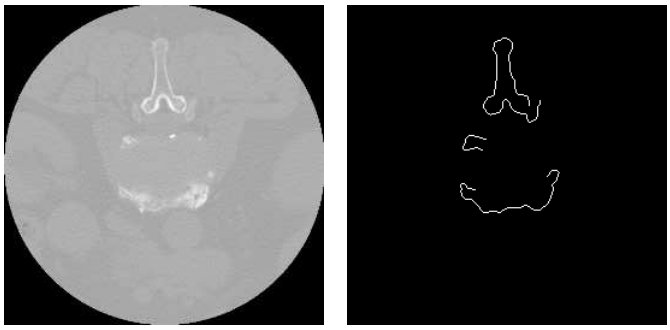


Figure 3

Figure 4: Result of the canny edge detection in CT images

2.3 Labeling of the target vertebra

The shape and pose coefficients derived from the alignment, $[\theta^{s\mathbf{T}}, \theta^{p\mathbf{T}}]^{\mathbf{T}}$ are used as features for the labeling. I use random forest (Matlab standard implementation) to classify these features. I also use the Gaussian process to adjust the hyper-parameters of the random forest.
In detail, random forests is an ensemble of threes that each is a weak classifier. Each tree is constructed by a random sub-sample of the data based on a random sub-sample of features. Each tree contains several branches where in each branch a decision is made based on the value of a certain feature. Usually, leaves of each tree have a minimum number of members which is set initially by the user.
In this work, I use Gaussian process with GP-UCB as criterion to optimize the hyper-parameters of the random forest, i.e. the number of trees, the portion of the samples used for training of each tree, and the minimum number of elements for each leaf. In each iteration of the GP, a ten-fold cross-validation is performed to identify the classifier accuracy with the selected parameters.

2.4 Labeling and segmentation of the entire vertebral column

Segmentation of the entire vertebral column is performed as follows. Initially a single point is selected on the vertebral column regardless of its level. To do this automatically, the standard deviation of the distribution of edges, which are found using the Canny edge detector, is computed for each transverse plane. The plane with the smallest standard deviation is selected and the center of the masses of the edges is chosen as the detected point. This process ensures the point falls inside a single vertebra. Next, the generic 3-vertebrae model is initialized on the selected point and is aligned to the edges. The labeling is then performed on the pose and shape coefficients to detect the level of the vertebra. Labeling of a single vertebra might still be incorrect as I will show in the next section. To address this problem and also to segment the rest of the vertebral column I perform an iterative technique. In each iteration, the previous aligned model is used to initialize another model, either one level superior or one level inferior. The new models are aligned and also classified. The procedure is continued till the models are out of the field-of-view. The algorithm steps are shown in Figure 5. The labeling is further improved by constructing a matrix, $\mathbf{P}$, where its elements, $p_{n,l}$, are the probabilities computed using random forest and represent the similarity of nth found vertebrae in the CT image to the level, l. For each diagonal of this matrix, the average posterior probability is calculated and the diagonal with the highest mean posterior probability is selected as the true configuration of labels.

3 Experiments and Results

Experiments were carried out on CT images of 61 patients (41 lumbar and 20 thoracic scans), which include 496 thoracic and lumbar vertebrae in total. Written informed consent was obtained from all patients. Manual CT segmentations were performed using ITK-SNAP. The CT imaging resolution ranged from 0.6 mm$\times$0.6 mm$\times$0.6 mm to 0.9 mm$\times$0.9 mm$\times$3.2 mm spacing.

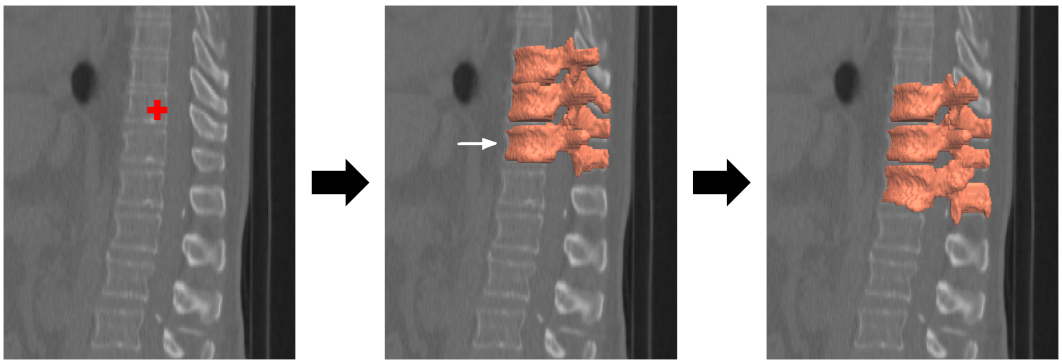


Figure 5: Initially a point is automatically found on the vertebral column (red cross). Next, a 3-vertebrae model is aligned. The last vertebra of the aligned model (white arrow) is then used to initialize the next model. This process continues until it reaches the extents of the field-of-view.

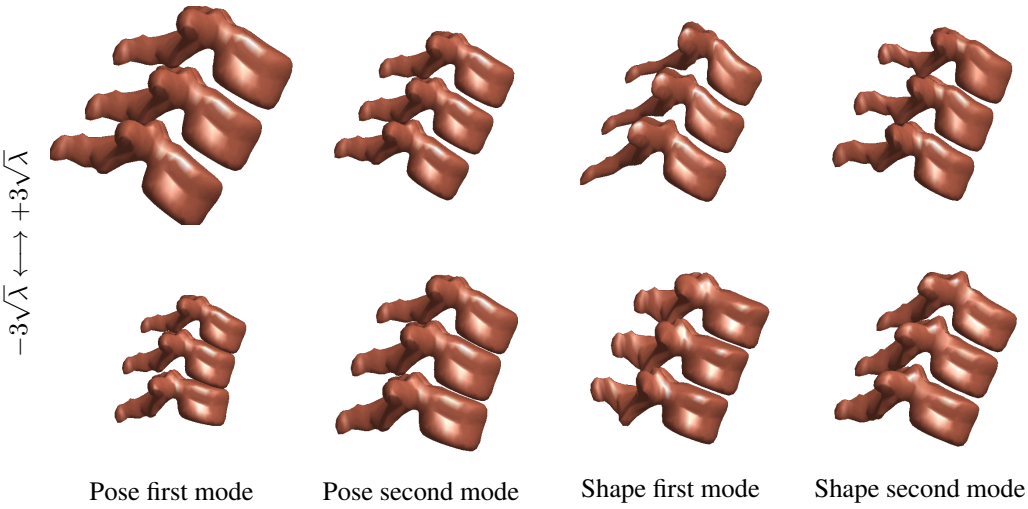


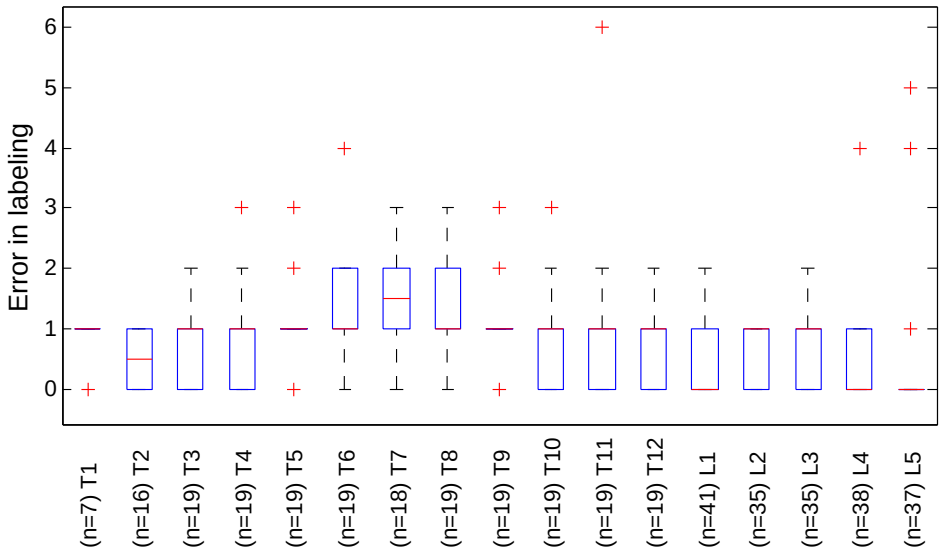
Figure 6: Graphical representation of the generic 3-vertebrae model described by changing the weights corresponding to the first two principal modes of pose and shape variation.

3.1 Generic 3-vertebrae model

Using the manual segmentation of the vertebrae, the generic 3-vertebrae model is reconstructed. Figure 6 illustrates the changes in the shape of the model that result from changing the weights corresponding to the first two pose and shape modes of the variation. 90% of the variations are achieved by the first 3 pose modes and the first 55 shape modes. As expected, the model is not compact in the shape space since it represents the shape of an arbitrary vertebra. I found that levels are more identifiable from the shape coefficients than the pose coefficients.

3.2 Labeling of an arbitrary vertebra

To assess the accuracy of the labeling, the center of the mass of the generic 3-vertebrae model and the manual segmentation are aligned. Next, the model is aligned to the edges extracted by Canny edge detector. The shape and pose coefficients are then given as input to the classifier. I used the first 8 shape and pose modes for the labeling (total 16 features). Bayesian optimization usually takes between 10-20 iteration to converge. The error is reported as the difference between the correct level and the one predicted by the classifier; e.g., for an actual level of T1 and predicted level of T3, the error is 2 levels.
five-fold cross-validation is used for the assessment of the labeling. The level detection error has



(a)

Figure 7: a) Error statistics for labeling of an arbitrary vertebra, which is defined as the number of vertebrae away from corrected label. Box and whisker shows mean and standard deviation. Outliers are indicated with '+'. The number of each vertebra in the study is given in parenthesis below the horizontal axis.

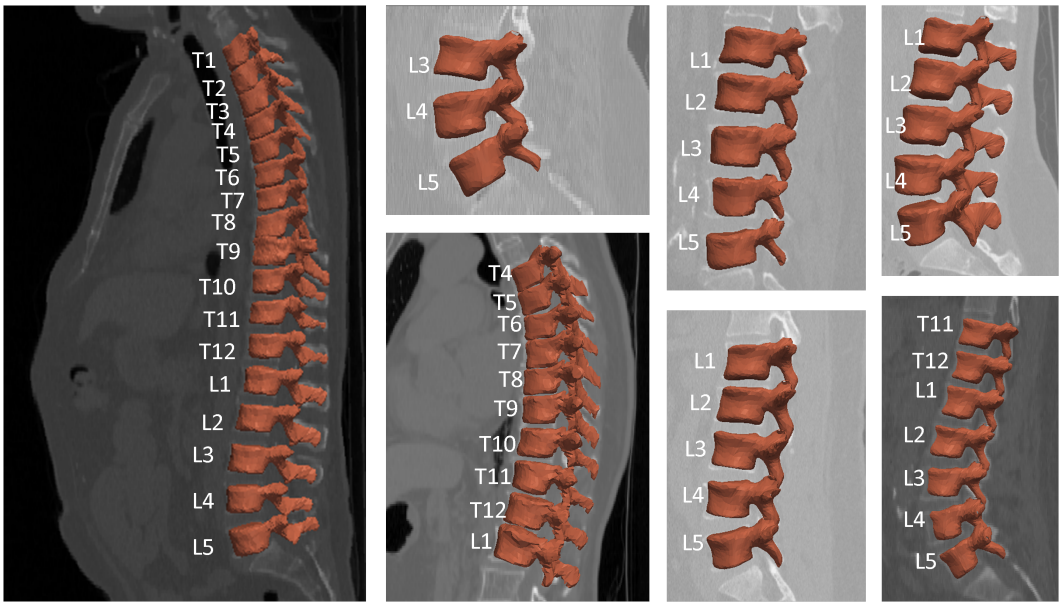


Figure 8: Examples of segmentation result

the mean of 0.76 with a standard deviation of 0.85. Separate results for each vertebra is reported in Figure 7.

3.3 Labeling and segmentation of the entire vertebral column

Leave-one-out experiments were performed on CT images to assess the accuracy of labeling and segmentation. Segmentation was performed on the entire vertebral column using the algorithm explained in the previous section. Examples of the segmentation are shown in Figure 8. Construction

Table 1: Accuracy of the segmentation (in mm) for each vertebra.

T1	T2	T3	T4	T5	T6	T7	T8	T9
3.3±0.7	2.2±0.2	2.5±1.0	2.0±0.5	1.9±0.6	2.2±0.4	2.1±0.7	2.5±1.0	2.6±0.8
T10	T11	T12	L1	L2	L3	L4	L5	Average
1.7±0.4	2.0±0.5	2.7±0.8	1.9±0.5	1.5±0.4	1.8±0.6	2.2±0.9	2.8±0.8	2.1±0.7

of the probability matrix, $\mathbf{P}$, between the detected vertebrae and finding the best labeling configuration, results in 97% accuracy. In other words, the vertebral column in all the patients were correctly labeled except one patient where the labeling was one level off. To assess the segmentation, the distances from vertices of the manual and fitted model were computed. The mean of the distances for each vertebra separately is detailed in Table 1.

4 Discussion and Conclusion

I proposed a novel technique for construction of a generic n-vertebrae model. This model enjoys joint representation of n arbitrary consecutive vertebrae by embedding statistics derived from a training set which contain thoracic and lumbar vertebrae. I used such a representation for characterizing an arbitrary vertebra in a CT image and for labeling. I showed that shape characteristics of three consecutive vertebrae can be used robustly for vertebra identification. I also used the generic n-vertebrae model to segment the entire vertebral column and used the result of the labeling of the entire vertebral column to correct the computed labels.
The current unoptimized MATLAB code running on an Intel Xeon X5650 2.67 GHz, requires one minute for segmentation of each vertebra. Given that, the segmentation of a CT with thoracic vertebrae (12 vertebrae) takes around 12 minutes. This is approximately four times faster than previously reported work [4]. The segmentation error is 2.1 mm which makes it suitable for a wide range of image-guided spinal interventions.

References

[1] S. S. C. Burnett, G. Starkschall, C. W. Stevens, and Z. Liao. A deformable-model approach to semi-automatic segmentation of CT images demonstrated by application to the spinal canal. *Medical Physics*, 31(2):251–263, 2004.

[2] P.T. Fletcher, Conglin Lu, and S. Joshi. Statistics of shape via principal geodesic analysis on lie groups. In *IEEE CVPR*, volume 1, pages 95–101, 2003.

[3] Ben Glocker, Johannes Feulner, Antonio Criminisi, D Haynor, and Ender Konukoglu. Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans. *Medical Image Computing and Computer-Assisted Intervention (MICCA)*, pages 590–598, 2012.

[4] T. Klinder, J. Ostermann, M. Ehm, A. Franz, R. Kneser, and C. Lorenz. Automated model-based vertebra detection, identification, and segmentation in CT images. *Medical Image Analysis*, 13(3):471 – 482, 2009.

[5] Jun Ma, Le Lu, Yiqiang Zhan, Xiang Zhou, Marcos Salganicoff, and Arun Krishnan. Hierarchical segmentation and identification of thoracic vertebra using learning-based edge detection and coarse-to-fine deformable model. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 19–27, 2010.

[6] A. Mastmeyer, K. Engelke, C. Fuchs, and W. A. Kalender. A hierarchical 3D segmentation method and the definition of vertebral body coordinate systems for QCT of the lumbar spine. *Medical Image Analysis*, 10(4):560 – 577, 2006.

[7] Zhigang Peng, Jia Zhong, William Wee, and Jing-huei Lee. Automated vertebra detection and segmentation from the whole spine MR images. In *IEEE-EMBS*, pages 2527–2530, 2006.

[8] X. Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25:127–154, 2006.

[9] A. Rasoulian, R. Rohling, and P. Abolmaesumi. Group-wise registration of point sets for statistical shape models. *IEEE TMI*, 31(11):2025–2034, 2012.

[10] S. Schmidt, J. Kappes, M. Bergtholdt, V. Pekar, S. Dries, D. Bystrov, and C. Schnoerr. Spine detection and labeling using a parts-based graphical model. In *IPMI*, volume 4584, pages 122–133. 2007.

[11] H. Shen, A. Litvin, and C. Alvino. Localized priors for the precise segmentation of individual vertebras from CT volume data. In *MICCAI*, volume 5241, pages 367–375. 2008.

[12] T. Vrtovec, B. Likar, and F. Pernus. Automated curved planar reformation of 3D spine images. *Physics in Medicine and Biology*, 50:4527–4540, 2005.