Hierarchical Segmentation and Identification of Thoracic Vertebra Using Learning-Based Edge Detection and Coarse-to-Fine Deformable Model

Jun Ma, Le Lu, Yiqiang Zhan, Xiang Zhou,
Marcos Salganicoff, and Arun Krishnan

CAD & Knowledge Solutions, Siemens Healthcare, Malvern, PA 19355
Center for Imaging Science, Johns Hopkins University, Baltimore, MD 21218

Abstract. Precise segmentation and identification of thoracic vertebrae is important for many medical imaging applications whereas it remains challenging due to vertebra's complex shape and varied neighboring structures. In this paper, a new method based on learned bone-structure edge detectors and a coarse-to-fine deformable surface model is proposed to segment and identify vertebrae in 3D CT thoracic images. In the training stage, a discriminative classifier for object-specific edge detection is trained using steerable features and statistical shape models for 12 thoracic vertebrae are also learned. In the run-time, we design a new coarse-to-fine, two-stage segmentation strategy: subregions of a vertebra first deforms together as a group; then vertebra mesh vertices in a smaller neighborhood move group-wise, to progressively drive the deformable model towards edge response maps by optimizing a probability cost function. In this manner, the smoothness and topology of vertebra's shapes are guaranteed. This algorithm performs successfully with reliable mean point-to-surface errors 0.95 ± 0.91 mm on 40 volumes. Consequently a vertebra identification scheme is also proposed via mean surface meshes matching. We achieve a success rate of 73.1% using a single vertebra, and over 95% for 8 or more vertebra which is comparable or slightly better than state-of-the-art [1].

1 Introduction

A precise vertebra segmentation and identification method is in high demand due to its important impact in many orthopaedic, neurological and oncological applications. In this paper, we focus on thoracic vertebra where accurate segmentation and identification of them can directly eliminate false findings on lung nodules in computer aided diagnosis system [2]. However, this task remains challenging due to vertebra's complexity, i.e., within-class shape variation and different neighboring structures.

Several methods have been reported addressing segmentation and/or identification of vertebra under different modalities, e.g., magnetic resonance imaging (MRI) or computed tomography (CT). Yao et al. [3] present a method to automatically extract and partition the spinal cord in CT images, and a surface-based

T. Jiang et al. (Eds.): MICCAI 2010, Part I, LNCS 6361, pp. 19–27, 2010.

registration approach for automatic lumbar vertebra identification is described in [4], where no identification was carried out in both work. Recently, Klinder et al. [1] propose a model-based solution for vertebra detection, segmentation and identification in CT images. They achieved very competitive identification rates of $> 70\%$ for a single vertebra and 100% for 16 or more vertebrae. However, their identification algorithm is based on vertebra appearance model (i.e., an averaged volume block) spatial registration and matching which is very computationally consuming ($20 \sim 30$ minutes).

In this paper, we present a new automatic vertebra segmentation and identification method. Although this work mainly focuses on thoracic vertebra (for potential lung applications), our approach can be easily extended to cervical and lumbar vertebrae. The main contributions of this paper are summarized as follows. First, we introduce a learning based bone structure edge detection algorithm, including efficient and effective gradient steerable features and robust training data sampling. Second, a hierarchical, coarse-to-fine deformable surface based segmentation method is proposed based on the response maps from the learned edge detector, followed with an efficient vertebra identification method using mean shapes. Finally, the promising results of segmentation and identification are presented, compared with the state-of-the-art [1].

2 Method

Due to complex neighboring structures around vertebra and imaging noise, common edge detectors, e.g., Canny operator, often produce leaking and spurious edge. To achieve robust edge detection, we develop a learning-based object specific edge detection algorithm, similar to semantic object-level boundary lineation in natural images [5].

2.1 Supervised Bone Edge Detection

We manually segmented 12 thoracic vertebrae from 20 CT volumes for training, and generated corresponding triangulated surfaces using Marching Cube algorithm, with about 10,000 triangular faces per vertebra model. It is observed that along the normal direction of the bone boundary, the intensity values roughly form a ridge pattern. Our new set of steerable features is designed to describe the characteristics of boundary appearance, which make it feasible for statistical training.

Gradient steerable features: For each triangle face of the surface mesh, we take 5 sampling points (called **a sampling parcel**) along the face normal direction with one voxel interval. Specially, given x a point on the normal line and n the unit normal vector, the sampling parcel associated with x is

$$\mathcal{P}(x) = \{x - 2n, x - n, x, x + n, x + 2n\}$$

For each of the 5 sampling points we compute three features: intensity I, projections of gradient onto the normal direction $\nabla_1 I \cdot n, \nabla_2 I \cdot n$, where $\nabla_1 I$ and $\nabla_2 I$

are gradient vectors computed using derivative of Gaussian with two different kernel scales. Totally, the **feature vector** of a point x, denoted by $\mathcal{F}(x)$, has 15 elements:

$$\mathcal{F}(x) = \{I(y), \nabla_1 I(y) \cdot n, \nabla_2 I(y) \cdot n | y \in \mathcal{P}(x)\}$$

Fig. 1 illustrates the sampling parcel and its associated features. Our steerable features are indeed oriented-gradient pattern descriptor with easy computation.

Vertebra edge detector training: The training samples of positive and negative boundary voxels are obtained from manually segmented vertebra mesh as below. For a triangle face center c, we define the boundary parcel as

$$\mathcal{P}(c) = \{c - 2n, c - n, c, c + n, c + 2n\}$$

interior parcel as

$$\mathcal{P}(c - 3n) = \{c - 5n, c - 4n, c - 3n, c - 2n, c - n\}$$

and exterior parcel as

$$\mathcal{P}(c + 3n) = \{c + n, c + 2n, c + 3n, c + 4n, c + 5n\}$$

That is, the interior parcel is 3 voxels away backward from boundary parcel while exterior parcel is the 3 voxels forward, where 3 is adjustable. The corresponding feature vectors $\mathcal{F}(c), \mathcal{F}(c - 3n), \mathcal{F}(c + 3n)$ can be also computed. Then we label $\mathcal{F}(c)$ as positive class (i.e., boundary), and assign both $\mathcal{F}(c - 3n)$ and $\mathcal{F}(c + 3n)$ as negative class (i.e., non-boundary), as Fig. 2 (*left*). Thus, each triangle face provides one positive data and two negative data. Given one vertebra surface mesh with about 10,000 faces, sufficient and adequate training feature vectors are obtained. Note that a single and unified bony edge detector will be learned for all 12 thoracic vertebrae. Compared with implicit, object"inside-outside" learning[1] [6], our boundary/non-boundary delineation strategy directly focuses on modeling the runtime boundary localization process (i.e., explicitly moving towards classified boundary positives), and is expected to have higher precision.

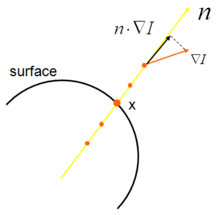


Fig. 1. Steerable features of x. Five red dots indicate sampling parcel associated with x. Yellow arrow indicates the normal direction. Red and black arrows indicate gradient ∇I and projection $\nabla I \cdot n$.

The feature vectors depend on the normal direction of triangle faces so that the edge detector is sensitive to the initialization of the surface template. In our experimental setup, the surface model is first roughly registered with images by automatic detection [7,8] or manual alignment, thus the normal direction of the surface model can not perfectly coincide with the true bony normal. To make the detector more robust to mis-alignment errors

[1] The boundary has to be further inferred from the transition of (object) internal positives and external negatives [6] which may not be trivial.

and the later deformable model convergent, it is important that we synthesize some "noisy" training samples by stress testing. Particularly, we add some random disturbances to the orientations and scales of the template model so that the template surface model does not accurately overlap with the manual segmentation. Considering a similarity transform, a random number between 0.9 and 1.1 for each of the three scales, and a random angle between $-\frac{\pi}{10}$ and $\frac{\pi}{10}$ for each of the three orientation angles are used. The true boundary parcels, as well as interior and exterior parcels are defined using ground truth positions but with disturbed template surface normals. Refer to Fig. 2 (*middle*) for an illustrative example. Their corresponding feature vectors are consequently calculated (with the disturbed face normals) and added into our training sets. The random disturbance process is repeated 10 times for each training mesh to guarantee we get enough noisy samples. We then train an Linear or Quadartic Discriminant (LDA, QDA) classifier based on the combined non-disturbed and disturbed feature vectors. Both LDA and QDA are evaluated and we find that LDA yields more robust results. The experiment results are computed with LDA. Finally, given a voxel x and its feature vector $\mathcal{F}(x)$, our classifier will assign a value $\mathcal{L}(x) \in [0, 1.0]$ which indicates the likelihood of x being boundary point.

2.2 Segmentation: Coarse-to-Fine Deformation

The main idea of segmentation is to deform the surface template mesh towards boundary points detected by the learned edge detector. After the surface template is initially positioned into a new volume, (The template can be initialized using similar strategies as marginal space learning [7,8]) edge detector calculates the edge likelihoods $\mathcal{L}(x)$ for voxels along the normal directions of all mesh faces, where a response map can be generated. As shown in Fig. 2 (**Right**), this response map is informative but unavoidably noisy. To guarantee the surface shape topology and smoothness during deformation/segmentation, we propose a hierarchical deformation scheme of first performing deformation of subregions;

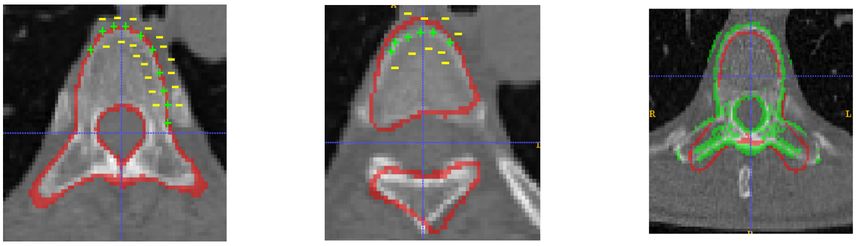


Fig. 2. Left: Surface template perfectly aligned with the true boundary. **Middle:** Disturbed Surface template overlapped within the volume. Green plus and Yellow minus signs are positive or negative sample samples, respectively. **Right:** Response map of vertebra edge detection in the section view of 3D CT volume. The red curve indicates the template surface while the green dots are the voxels classified as boundary points with likelihood values > 0.8.

then performing patch-wise deformation, i.e., points in the same neighborhood move together.

Deformation of subregions: We manually divide the surface mesh into 12 subregions, as indicated by Fig. 3. In order to maintain the shape of these subregions, a similarity transformation to each subregion is applied such that the total response of edge detection is maximum in the transformed configuration. For a subregion S and some face center f on it, we intend to find a similarity transformation $\hat{T}$ satisfying

$$\hat{T} = \arg\max_{T \in \mathbf{T}} \sum_{f \in S} \mathcal{L}(T(f)) \tag{1}$$

where $\mathbf{T}$ is the set of similarity transformations T. Searching the optimal T involves the 9-dimensional parameters space of $(T_x, T_y, T_z, S_x, S_y, S_z, \theta_x, \theta_y, \theta_z)$. If we perform a exhaustive search with 5 grid steps for each parameters, then possible transformation is 5^9 which is computationally infeasible. To reduce the search space, we perform a three-stage search. First, search for (T_x, T_y, T_z) with displacement $\{-4, -2, 0, 2, 4\}$ voxels for each translation; second, with fixed $(\hat{T}_x, \hat{T}_y, \hat{T}_z)$, search for (S_x, S_y, S_z) with discretization grids of $\{0.8, 0.9, 1.0, 1.1, 1.2\}$ for each scaling; third, with fixed optimal translation and scaling, search for $(\theta_x, \theta_y, \theta_z)$ with intervals of $\{-\pi/10, -\pi/20, 0, \pi/20, \pi/10\}$ for each orientation. In this way, we need to only consider $5^3 \times 3 = 375$ possible poses and select the one with the strongest response as $\hat{T}$. This heuristic searching strategy turns out to be effective in capturing the true pose of subregions though it might be suboptimal. Fig. 4(a) illustrates the searching process.

After the optimal similarity transformation is found for each subregions, a smooth deformation of the whole surface can be obtained using simple Gaussian smoothing. Let $S_1, S_2, ..., S_{12}$ denote the twelve subregions, and $T_1, T_2, ..., T_{12}$ be the corresponding optimal transform. Denote v an arbitrary vertex in the template surface and u a vertex in a certain subregion. Then the new position of v is

$$v' = v + \lambda \sum_{i=1}^{12} \sum_{w \in S_i} (T_i(w) - w) K(w - v)$$

where $K(x) = e^{-\frac{x^2}{2\sigma^2}}$ is the Gaussian kernel and λ is a regulation parameter. Fig. 4 (b) shows the result of "deformation of subregion" stage. One can see the surface mesh is more closely aligned with the true boundary through "articulated" similarity moves, although in several area, the surface mesh still has a certain distance from the true boundary. This will be solved by the finer-scale deformation strategy described below.

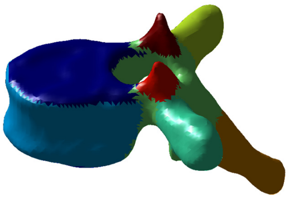


Fig. 3. Subregions of the surface. Subregions are illustrated in different colors.

Deformation of patches: After deforming the subregions, the surface mesh is approximately overlap with the vertebra's boundary in CT volume. Next, we

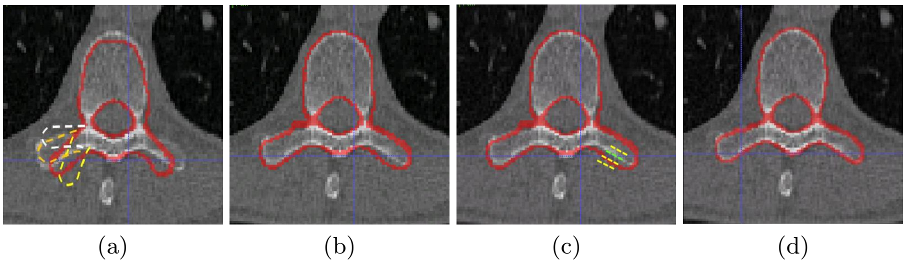


Fig. 4. (a,b) Deformation of left transverse process. (a) Dot curves indicate searching of transformations of this subregion. In this case, the orange curve indicates the optimal position. (b) Subregion deformation result. (c,d) Deformation of patches. (c) Dot curve indicate displacing a patch in the normal direction for search of strongest response. The green dots indicates the optimal displacement. (d) Patch deformation result.

perform deformation on local neighborhoods of 200 patches divided from each vertebra mesh surface (each patch may contain 50 faces approximately). For each patch (denoted as PT), we compute its mean normal by this formula:

$$\bar{n} = \frac{1}{N} \sum_{f \in PT} n(f) \tag{2}$$

where f is a face in the patch and $n(f)$ is the unit normal of the face. Then the patch is moved along its mean normal direction in search of the strongest response, that is, we optimize this term:

$$\hat{i} = \arg\max_{i} \sum_{f \in S} \mathcal{L}(f + i\bar{n}) \tag{3}$$

where the search range is limited in $i = -6, -5, ...5, 6$. Fig. 4(c) shows the a patch is displaced along its mean normal direction in search for the boundary. After all patches find their optimal displacement, a smooth deformation of surface is again obtained by Gaussian smoothing. Fig. 4 (d) shows the segmentation result of “deformation of patches” stage. Clearly, the surface mesh now can accurately capture the true boundary of the vertebra. The two-stage, coarse-to-fine deformation of surface model guarantees the accuracy of segmentation as well as the smoothness of the shapes, using articulated similarity transforms and nonrigid transform respectively.

2.3 Identification Using Mean Shapes

We applied the segmentation algorithm to 40 volumes at 1mm by 1mm by 1mm resolution, and $15 \sim 20$ surface meshes are obtained per thoracic vertebra, due to missing vertebra in some volume. Vertex correspondence across meshes for each vertebra is also directly available since surface meshes are deformed by the same template. Therefore we can compute the mean vertebra shapes by simply taking the arithmetical mean of corresponding vertices' positions. There are 12 thoracic vertebrae, namely T1, T2, ..., T12. Vertebra identification is to label a segmented vertebra to be one of the twelve. In this context, given a single vertebra subvolume, we carry out the identification process by testing which mean shape has the maximum response. Specially, we feed the 12 mean shapes to the vertebra volume one after another, and calculate the supervised edge response scores without deformation. The mean shape with the strongest response is determined as the label of this vertebra.

Let $M_1, M_2, ..., M_{12}$ denote the twelve mean shapes and f is an arbitrary face center in the mean shapes. One way to calculate the responses is computing the overall likelihood of boundary,

$$\hat{i} = \arg\max_i \sum_{f \in M_i} \mathcal{L}(f) \tag{4}$$

Another way is to count the number of faces with high probability to be boundary point,

$$\hat{i} = \arg\max_i \sum_{f \in M_i} \mathbf{1}_{\mathcal{L}(f) > \alpha} \tag{5}$$

where α is a threshold. We find the second method is more robust against outliers and noise, by tolerating up to $(1 - \alpha)$ portion of data being polluted or not at the correct spatial configuration, and take $\alpha = 0.8$ which is used for following experiments. We also extend the identification method to multiple vertebrae, i.e., a vertebra string. By using more context, multiple vertebrae identification is expected to have higher success rate.

3 Result

We apply our automatic segmentation algorithm to 40 volumes of thoracic scans and the evaluation is performed using four-fold cross validation. In implementation, we run the subregion deformation step multiple (m) times followed by patch-based deformation n times, where m and n are empirically optimized to be 3 and 4, respectively. The supervised edge detection is performed at each iteration to reflect the runtime vertebra mesh surface configuration. In Fig. 5, we show some segmentation examples in axial, sagittal or coronal view, for visual inspection. To quantitatively evaluate our segmentation algorithm, we use the distance of a vertex on the fitted mesh to the closest mesh point (not necessarily a vertex) of the ground truth mesh which is generated from manual segmentation. The mean point-to-surface error and the standard deviation for individual

Table 1. Mean point-to-surface error and standard deviation for individual vertebra

vertebra	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
mean error (mm)	1.05	1.11	1.03	0.93	0.99	0.92	0.83	0.75	0.89	0.79	0.94	1.21
std deviation(mm)	0.96	0.97	1.04	1.03	1.31	0.92	0.56	0.59	0.68	0.50	0.63	1.16

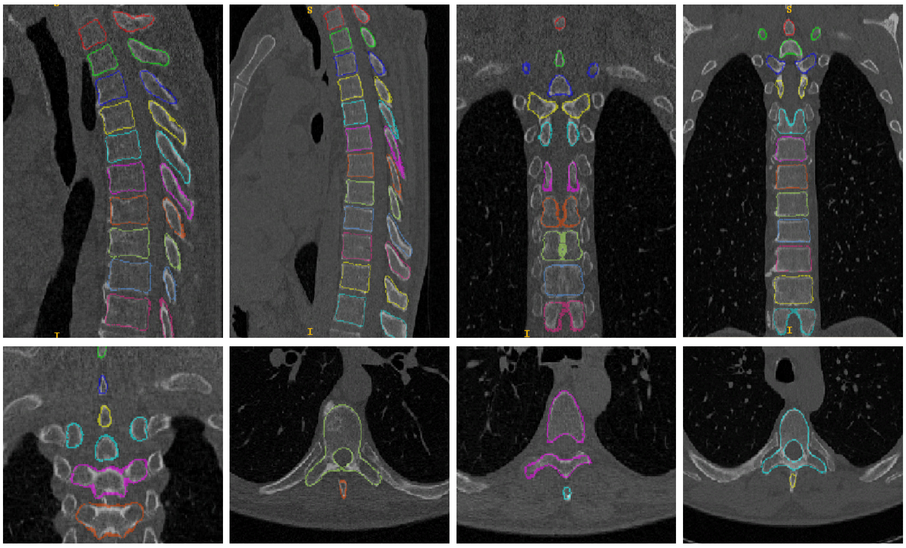
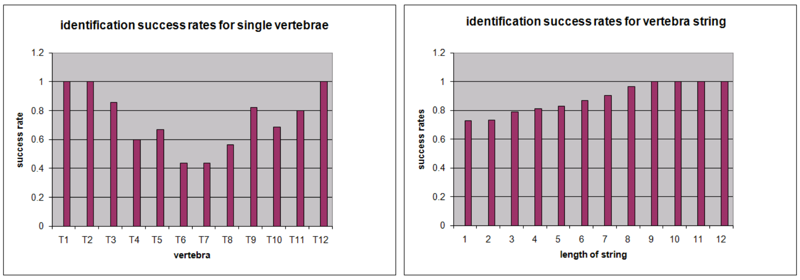


Fig. 5. Segmentation results of chosen volume in axial or sagittal or coronal view. Different vertebrae are featured in different colors.



individual success rates

string success rates

Fig. 6. Identification success rate of individual vertebra and stringed vertebrae

vertebra is shown in Table 1. Highly reliable and accurate segmentation results have been achieved, with the overall final mean error of 0.95 ± 0.91 mm. [1] reports a comparable accuracy level at 1.12 ± 1.04 mm.

For identification, we have an average success rate of 73.1% using single vertebra. This success rate also varies regarding to a specific vertebra where the rates for $T5, T6, T7, T8$ as $\leq 60\%$ are especially lower than others because these four vertebrae look alike. Furthermore, when exploiting vertebra string for identification, the success rate is improved and increases with longer string. With a string of 7 or 8 and more vertebrae, we achieve over 91% or $> 95\%$ success rates, whereas rates are $\approx 71\%$ for one vertebra, $\approx 87\%, 89\%$ for 7 or 8 vertebra strings in [1]. The success rates of individual and stringed vertebra identification (via mean mesh shapes) are comparable or better than [1] using intensity based matching, as shown in Fig. 6.

A volumetric mean appearance model is used for vertebra identification in [1], which seems more comprehensive than our shape information alone. However we observe that in real cases, the variability of neighboring structures is quite large due to patients' pose variation. The adjacent vertebrae can be so close to each other where the boundary even can not be clearly distinguished; or, successive vertebrae are apart from each other with a large distance. Thus, the neighboring structures are not necessarily positive factors in the identification procedure. A clean shape model without surrounding structures may be of advantage and our identification results are indeed slightly better.

4 Conclusion

In this paper, a hierarchical thoracic vertebra segmentation and identification method is presented. We propose learning-based edge detectors using steerable gradient features. The segmentation applies a surface deformable model by adopting a new two-stage "coarse-to-fine" deformation scheme: first subregion based articulated similarity deformation and then nonrigid local patch deformation. The segmentation result is satisfying with point-to-surface error 0.95 ± 0.91 mm. We also use the generated mean shape model of each thoracic vertebra for identification process. Both our segmentation and identification performance is compared favorably against the state-of-the-art [1].

References

1. Klinder, T., Ostermann, J., Ehm, M., Franz, A., Kneser, R., Lorenz, C.: Automated model-based vertebra detection, identification, and segmentation in ct images. Medical Image Analysis 13, 471–481 (2009)
2. Murphy, K., et al.: A large-scale evaluation of automatic pulmonary nodule detection in chest ct using local image features and k-nearest-neighbour classification. Medical Image Analysis 13, 757–770 (2009)
3. Yao, J., O'Connor, S., Summers, R.: Automated spinal column extraction and partitioning. In: Proc. of IEEE ISBI, pp. 390–393 (2006)
4. Herring, J., Dawant, B.: Automatic lumbar vertebral identification using surface-based registration. Computers and Biomedical Research 34(2), 629–642 (2001)
5. Dollar, P., Tu, Z., Belongie, S.: Supervised learning of edges and object boundaries. In: CVPR (2006)
6. Zhan, Y., Shen, D.: Deformable segmentation of 3-d ultrasound prostate images using statistical texture matching method. IEEE Trans. on Medical Imaging (2006)
7. Lu, L., Barbu, A., Wolf, M., Salganicoff, M., Comaniciu, D.: Simultaneous detection and registration for ileo-cecal valve detection in 3d ct colonography. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 465–478. Springer, Heidelberg (2008)
8. Zheng, Y., Barbu, A., et al.: Four-chamber heart modeling and automatic segmentation for 3-d cardiac ct volumes using marginal space learning and steerable features. IEEE Trans. Medical Imaging 27(11), 1668–1681 (2008)