

Hierarchical segmentation and identification of thoracic vertebra using learning-based edge detection and coarse-to-fine deformable model



Jun Ma, Le Lu *

Siemens Molecular Imaging, 2501 Barrington Road, Hoffman Estates, IL 60192, United States
Siemens Corporate Research, 755 College Road East, Princeton, NJ 08540, United States

ARTICLE INFO

Article history:

Received 13 February 2012
Accepted 9 November 2012
Available online 28 April 2013

Keywords:

Supervised edge detection
Coarse-to-fine shape deformation model
Shape based medical image segmentation
Vertebra segmentation
Vertebra identification
Computed tomography

ABSTRACT

Precise segmentation and identification of thoracic vertebrae is important for many medical imaging applications though it remains challenging due to the 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. For the run-time testing, we design a new coarse-to-fine, two-stage segmentation strategy: subregions of a vertebra first deform 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 vertebrae 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 mesh 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 [5].

Published by Elsevier Inc.

1. Introduction

A precise vertebra segmentation and identification method is in high demand due to its importance in many orthopaedic, neurological and oncological applications. In this paper, we focus on thoracic vertebra where accurate segmentation and identification can directly eliminate false findings on lung nodules in computer aided diagnosis (Lung CAD) system [14,25,6]. However, this task remains challenging due to the significant complexity of vertebrae, i.e., within-class shape variation and different neighboring structures. An illustration of our motivation is shown in Fig. 1.

Several methods have been reported addressing segmentation and/or identification of vertebra under different modalities [24,17,4,13,15,5], e.g., magnetic resonance imaging (MRI) or computed tomography (CT). Yao et al. [24] present a method to automatically extract and partition the spinal cord in CT images, and a surface-based registration approach for automatic lumbar vertebra identification is described in [4], where no identification was carried out in either work. A hierarchical 3D segmentation method of the lumbar spine in CT images with the definition of vertebral

body coordinate systems is described in [13]. In [15], Peng et al. propose a vertebra detection and segmentation algorithm processing 2D slices of MRI images, instead of the 3D volume as a whole. Recently, Klinder et al. [5] 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 vertebrae active appearance model (i.e., an averaged intensity volume block) for spatial registration and matching which is very computationally expensive (20–30 min).

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, we propose a hierarchical, coarse-to-fine deformable surface based segmentation method based on the response maps from the learned edge detector, followed with an efficient vertebra identification method using mean shapes. Finally, promising results of both segmentation and identification are presented, compared with the state-of-the-art [5]. An earlier version of this paper appears in [11].

* Corresponding author at: National Institutes of Health, Clinical Center, 10 Center Drive, Bethesda, MD 20892-1182, United States.

E-mail addresses: jun.jm.ma@siemens.com (J. Ma), le.lu@nih.gov (L. Lu).

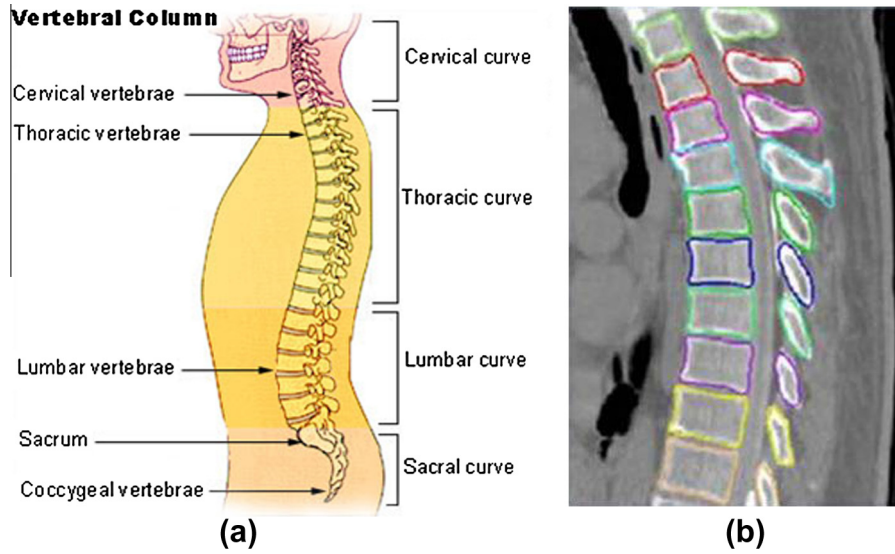


Fig. 1. (a) Illustration of the human vertebra system: cervical, thoracic, lumbar and sacrum sections (http://en.wikipedia.org/wiki/Vertebral_column). (b) Expected thoracic vertebra segmentation result to delineate 3D vertebra boundary from surrounding soft tissues, including lung.

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 and medical images [2,12,28,10].

2.1. Supervised bone edge detection

Bone-structures normally show high intensity/attenuation in CT images. However, due to complex neighboring structures around vertebra and imaging noise, common edge detectors (e.g., Canny operator) do not perform well, often with leaking and spurious edge.

We manually segmented 12 thoracic vertebrae from 20 CT volumes for training, and generated corresponding triangulated surfaces using the Marching Cube algorithm [8], 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 are designed to describe the characteristics of boundary appearance, which make it feasible for statistical training.

2.1.1. 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. Specifically, given x as 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: (1) intensity I , (2) and (3) projections of gradient onto the normal direction $\nabla_1 I \cdot n$ and $\nabla_2 I \cdot n$, where $\nabla_1 I$ and $\nabla_2 I$ are gradient vectors computed using derivative of Gaussian with two different kernel scales. In total, 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. 2 illustrates the sampling parcel and its associated features. Our steerable features are indeed oriented-gradient pattern descriptor with easy computation.

2.1.2. Vertebra edge detector training

The training samples of positive and negative boundary voxels are obtained from manually segmented vertebra meshes. 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 three voxels away backward from boundary parcel while exterior parcel is the three 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 (i.e., boundary), and label both $\mathcal{F}(c - 3n)$ and $\mathcal{F}(c + 3n)$ as negative (i.e., non-boundary), as Fig. 3 (top-left). Thus,

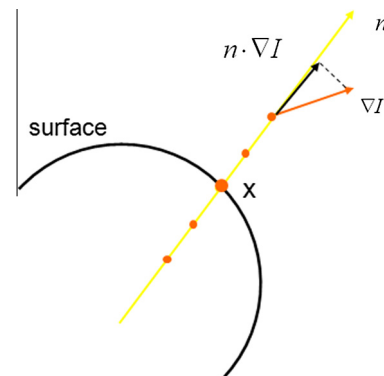


Fig. 2. Steerable features of x . Five red dots indicate sampling parcel associated with x . Yellow arrow indicates the normal direction. Red and black arrows indicate gradient ∇I and projection $\nabla I \cdot n$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

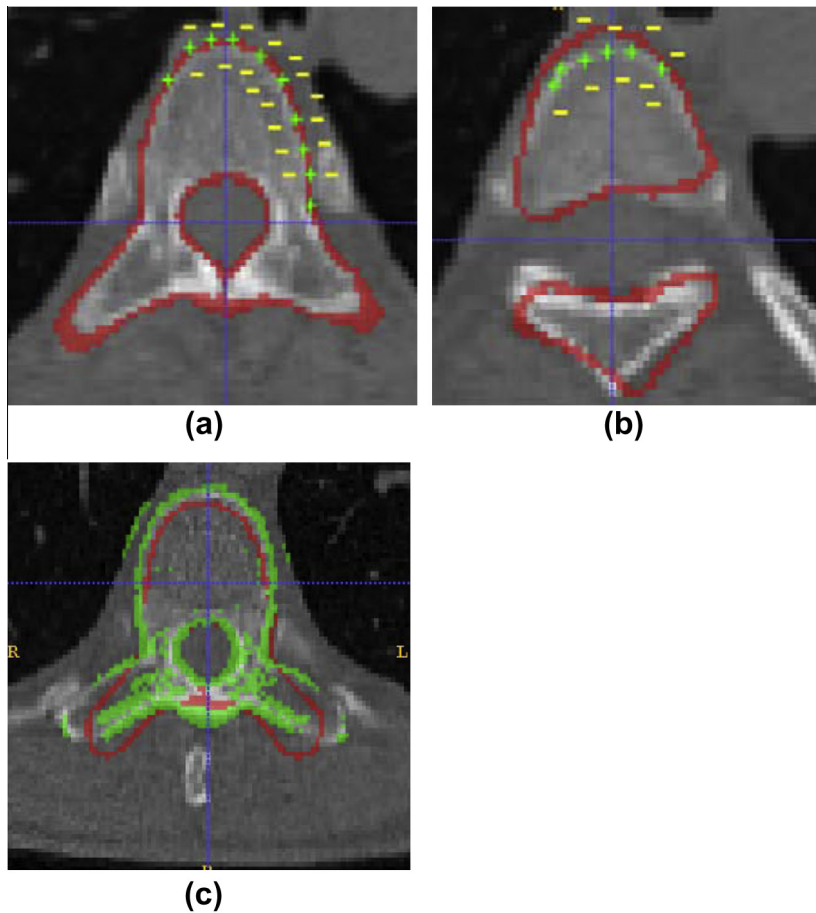


Fig. 3. (a) Surface template perfectly aligned with the true boundary. (b) Disturbed surface template overlapped within the volume. Green plus and yellow minus signs are positive or negative sample samples, respectively. (c) Response map of vertebra edge detection in the section view of 3D CT volume. The red curve indicates the template surface while the green dots are the voxels classified as boundary points with likelihood values >0.8. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

each triangle face provides one positive data example and two negative examples. 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” learning¹ [27,10,16], 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.

The feature vectors depend on the normal direction of triangle faces and consequently 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 [9,28,23] or manual alignment, thus the normal direction of the surface model cannot perfectly coincide with the true bony normal. To make the detector more robust to mis-alignment errors and the later deformable model convergent, it is important that we synthesize some “noisy” training samples by stress testing. Particularly, we add some random perturbations 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 perturbed template surface normals. Refer to Fig. 3 (top-right) for an illustrative example. Their corresponding feature vectors are consequently calculated (with the disturbed face normals) and added into our training sets. The random perturbation process is repeated 10 times for each training mesh to guarantee sufficient noisy training samples. We then train a Linear or Quadratic Discriminant (LDA, QDA) classifier [3] based on the combined non-perturbed and perturbed 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 assigns a value $\mathcal{L}(x) \in [0, 1.0]$ which indicates the likelihood of x being boundary point.

Our method is conceptually simple and easy to compute compared with boundary detectors learned using tens of thousands of 3D Haar features [20] or steerable features [28,10,7] with hierarchical boosting framework [22,19]. Due to our effective coarse-to-fine deformable model, it also alleviates the requirement or burden of generating perfect boundary response maps where even more sophisticated methods [28,10,7] still produce noisy responses. In general, per-voxel class conditional labeling is non-trivial and needs spatial regularization for smooth results. Note that here we only need to train one classifier for detecting all 3D (thoracic) vertebra bone boundaries coupled with our deformation model (Non-Gaussian), while active shape model [1] based deformation

¹ The boundary has to be further inferred from the transition of (object) internal positives and external negatives in prostate or polyp segmentation [27,10] which may not be trivial.

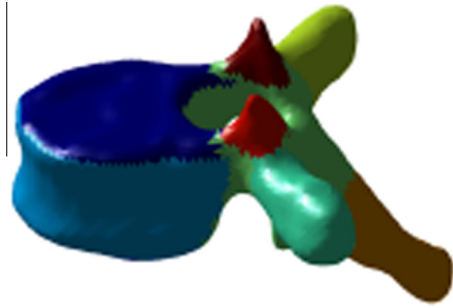


Fig. 4. Subregions of the surface. Subregions are illustrated in different colors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Gaussian) needs to build several individual boundary detectors for successive organ parts of 3D liver segmentation in CT images [7].

2.2. Segmentation: Coarse-to-fine deformation

In this work, our research is focused on providing an automatic vertebra segmentation and identification module instead of detecting the pose of vertebra. We manually initialize the position and orientation of the template surface using landmarks so that the surface model roughly coincides with the vertebrae to be segmented in the 3D volume. The automatic algorithm for detection and localizing vertebra [30,18] is also used for evaluation later, using similar strategies as sparse redundant anatomical landmark

detections [29] or marginal space learning [9,28] in the full 3D volumetric domain.

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 in a new volume (the template can be initialized using similar strategies as marginal space learning [9,28] for 3D organ localization), the edge detector calculates the edge likelihoods $\mathcal{L}(x)$ for voxels along the normal directions of all mesh faces so that a response map is generated. As shown in Fig. 3 (bottom-right), this response map is informative but unavoidably noisy. If we just deform each face toward its most-likely boundary position independently, the shape of vertebrae may not be retained. Hence some constraints to guarantee the surface shape topology and smoothness during deformation/segmentation must be enforced. To guarantee the surface shape topology and smoothness during deformation/segmentation, we propose a hierarchical deformation scheme of first performing deformation of subregions; then performing patch-wise deformation (i.e., points in the same neighborhood move together).

2.2.1. Deformation of subregions

We manually divide the surface mesh into 12 subregions, as indicated by Fig. 4. 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 S , we intend to find a similarity transformation \hat{T} satisfying

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

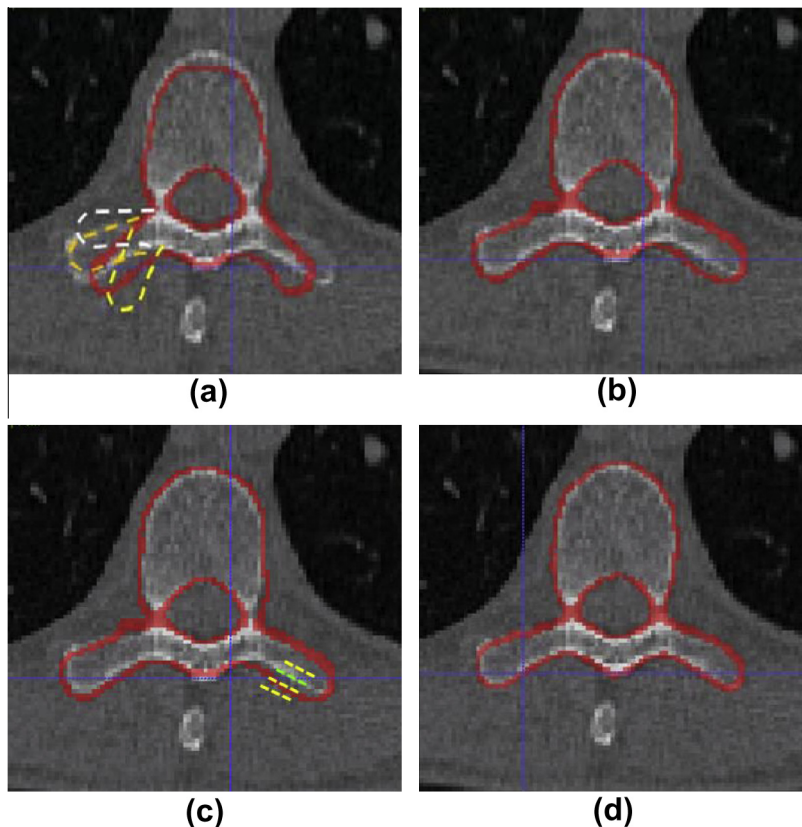


Fig. 5. (a and b) Deformation of left transverse process. (a) Dot curves indicate searching of transformations of this subregion. In this case, the orange curve indicates the optimal position. (b) Subregion deformation result. (c and d) Deformation of patches. (c) Dot curve indicate displacing a patch in the normal direction for search of strongest response. The green dots indicates the optimal displacement. (d) Patch deformation result. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where \mathbf{T} is the set of similarity transformations T :

$$T = R \begin{pmatrix} S_x & & \\ & S_y & \\ & & S_z \end{pmatrix} + \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix} \quad (2)$$

Here R is a rotation matrix parameterized by three angles $\theta_x, \theta_y, \theta_z$. 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 an exhaustive search with 5 grid steps for each parameter, then possible transformation is 5^9 which is computationally prohibitive. 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. 5a 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, \dots, S_{12} denote the twelve subregions, and T_1, T_2, \dots, T_{12} be the corresponding optimal transform. Denote v to be an arbitrary vertex in the template surface and w 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. 5b 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 areas, the surface mesh still has a certain distance from the true boundary. This will be solved by the finer-scale deformation strategy described below.

2.2.2. Deformation of patches

After deforming the subregions, the surface mesh is approximately overlapped with the vertebra’s boundary in CT volume. Next, we perform deformation on local neighborhoods of 200 patches divided from each vertebra mesh surface (each patch may contain approximately 50 faces). For each patch (denoted as PT), we compute its mean normal by this formula:

$$\bar{n} = \frac{1}{N} \sum_{f \in PT} n(f) \quad (3)$$

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:

$$\hat{i} = \arg \max_i \sum_{f \in S} \mathcal{L}(f + i\bar{n}) \quad (4)$$

where the search range is limited in $i = -6, -5, \dots, 5, 6$. Fig. 5c shows the a patch is displaced along its mean normal direction in search of the boundary. After all patches find their optimal displacement, a smooth deformation of the surface is again obtained by Gaussian smoothing. Fig. 5d shows the segmentation result of the “deformation of patches” stage. Clearly, the surface mesh now can accurately capture the true boundary of the vertebra. The two-

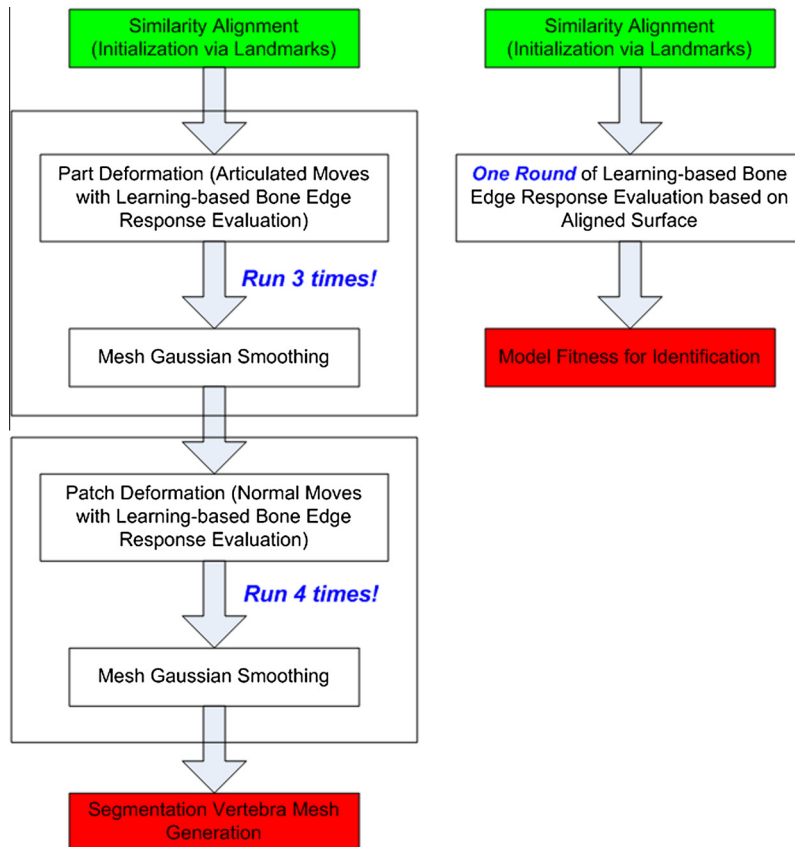


Fig. 6. The flowcharts of our vertebra segmentation (left) and identification (right) algorithms.

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. In practice, we execute multiple rounds of Part- and Patch-based deformation schemes. The iteration numbers are empirically calibrated using cross-validation. In this paper, we choose the iteration numbers as $Pt = 3$ (Part) and $Ph = 4$ (Patch) by default.

Remark. Though the 12 subregions and patches are selected manually during the model building, this does not require much time (e.g., takes about 30 min for each thoracic vertebra 3D mesh model). Once these models are selected, all other surface meshes can be generated by deforming the first template against other vertebra instances (i.e., 12 thoracic vertebra models are learned by fitting one template mesh on one exemplar case and then deforming on the other 9 instances²; each of the 12 vertebra models has 12 subgroups). Hence, the subregions and patches are inherited.

2.3. Identification using mean shapes

We applied the segmentation algorithm to 40 volumes at 1 mm by 1 mm by 1 mm resolution, and 31–40 surface meshes are obtained per thoracic vertebra, due to missing vertebrae in some volumes. 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 arithmetic 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. Specifically, 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, \dots, M_{12} denote the twelve mean shapes and f , an arbitrary face center in the mean shapes. One way to calculate the responses is computing the overall likelihood of boundary:

$$\hat{i} = \arg \max_i \sum_{f \in M_i} \mathcal{L}(f) \quad (5)$$

Another way is to count the number of faces with high probability values as valid boundary points:

$$\hat{i} = \arg \max_i \sum_{f \in M_i} \mathbf{1}_{\mathcal{L}(f) > \alpha} \quad (6)$$

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 at the correct spatial configuration where $\alpha = 0.8$ 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. Refer to Fig. 7 for the mean shapes. This process is illustrated in Fig. 8.

The algorithm flowcharts of our proposed deformable vertebra segmentation and identification algorithms are summarized in Fig. 6.

² In our fourfold cross validation setting, ten volumes are used to learn the mean mesh shape model for each thoracic vertebra. Then these fitted models are applied on other 30 volumes for deformable segmentation. This is that we use onefold for training and three folds for testing.

3. Results

3.1. Data and experiment setup

We apply our automatic segmentation algorithm to 40 volumes of thoracic scans and the evaluation is performed using fourfold cross validation. These 40 volumes are randomly sampled from our Lung CAD dataset [6] that was collected from multiple hospitals in Asia, Europe and United States. Many types of Siemens, GE, Philips and Toshiba scanners were used for the original data acquisition. All 40 volumes have slice thickness ranging from 1.00 mm to 1.25 mm, with 512 by 512 (voxel) axial-slice resolution. The voxel size in axial-planes ranges from 0.782 mm to 1.12 mm. In Fig. 9, we illustrate the process for single vertebra manual annotation and building surface mesh models of the ground truth generation. From an original 3D Lung CT scan, we manually localize and crop any thoracic vertebra. This step can be replaced by the recent development of highly robust and fully



(a) Mean shapes of 12 thoracic vertebrae in Axial view



(b) Mean shapes of 12 thoracic vertebrae in Sagittal view

Fig. 7. Mean shapes of 12 thoracic vertebrae: (a) axial view and (b) sagittal view.

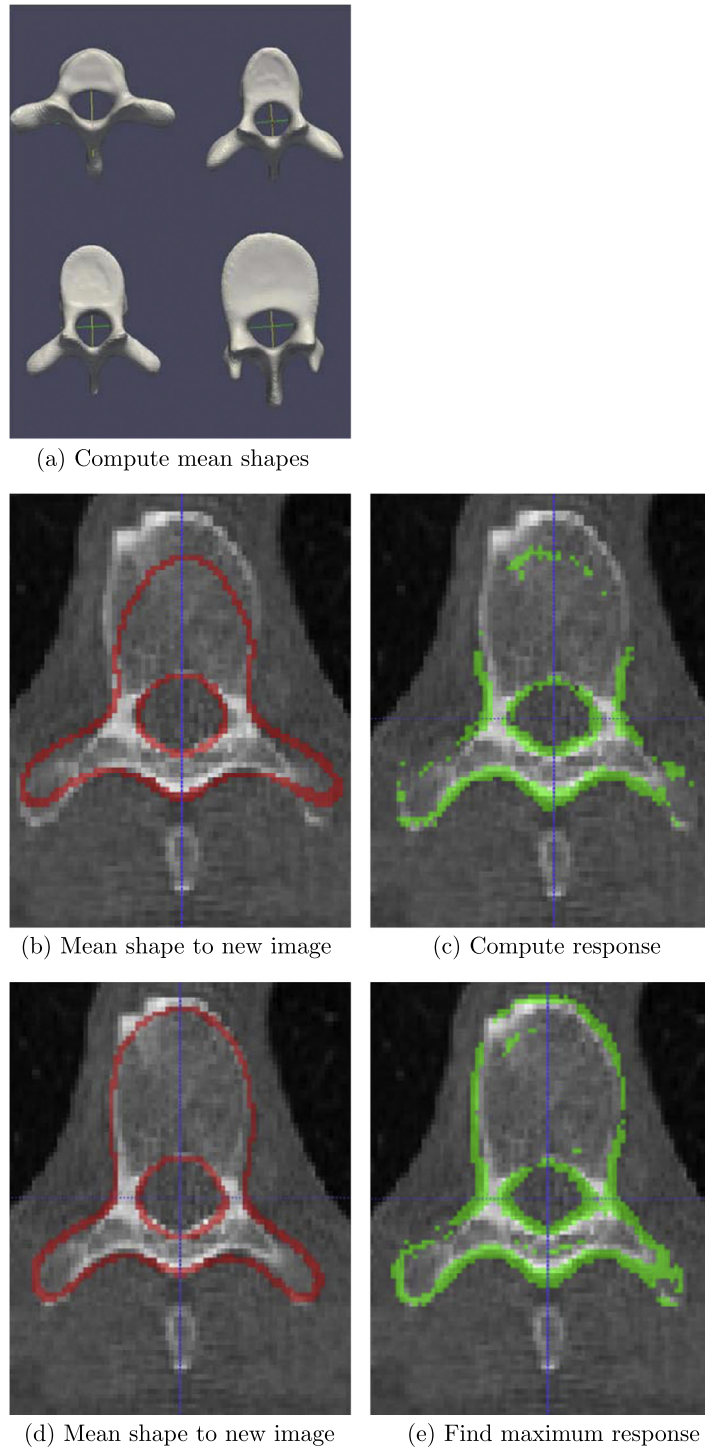


Fig. 8. The vertebra identification process pipeline.

automatic spine labeling system (FAST-Spine) [18,30,23]. Then the ITK Snap Tool is used to annotate or “paint” all vertebra voxels, followed by surface generation M and smoothing from the annotated vertebra mask. Optional mesh editing is used to fit the image boundary as accurately as possible. The manual annotation approximately takes 1 h per vertebra by a medical student. The mesh models M (generated using the Marching Cubes algorithm [8] from the annotated binary volume masks) of the same vertebra across CT volumes (or different patients) do not have mesh correspondence, but can be used for training our supervised bone boundary detector (as in Section 2.1). We further select one M_0 per vertebra from

an anchor volume and deform M_0 to align or register with M_k of the same vertebra in another volume k to obtain M'_k , via our coarse-to-fine deformation method. Therefore the mesh correspondence relationships are set up from M_0 to any M'_k where we also treat them as ground truth. The mean mesh model per vertebra can be further computed using the collection of $\{M_0, \dots, M'_k\}$.

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

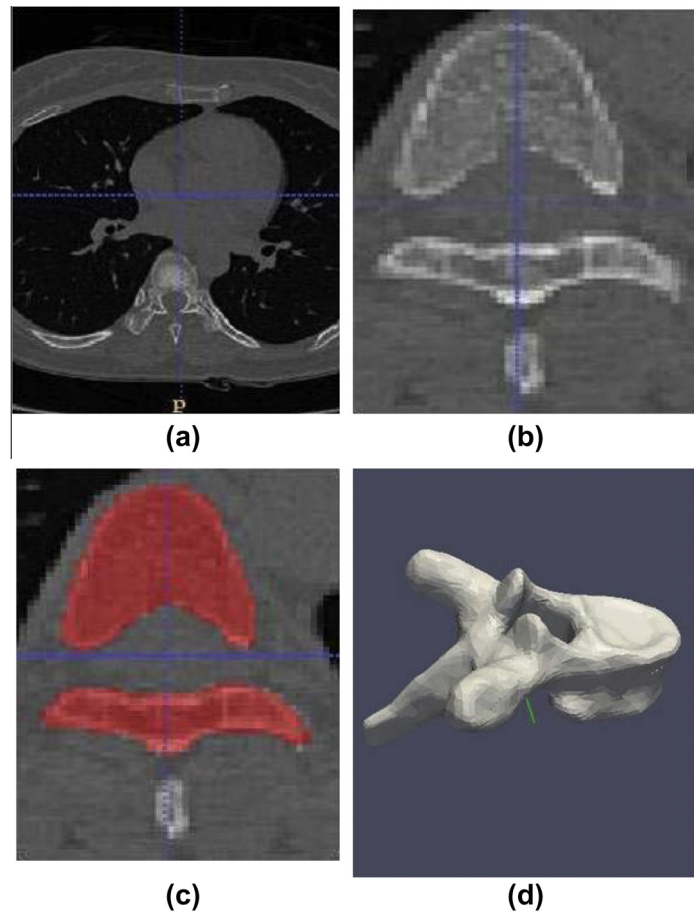


Fig. 9. The pipeline for single vertebra manual annotation and building surface mesh models of ground truth generation. (a) Original 3D CT image, (b) pre-processing of vertebra localization, (c) manual segmentation via ITK Snap Tool (<http://www.itksnap.org/>), and (d) mesh generation and optional mesh editing.

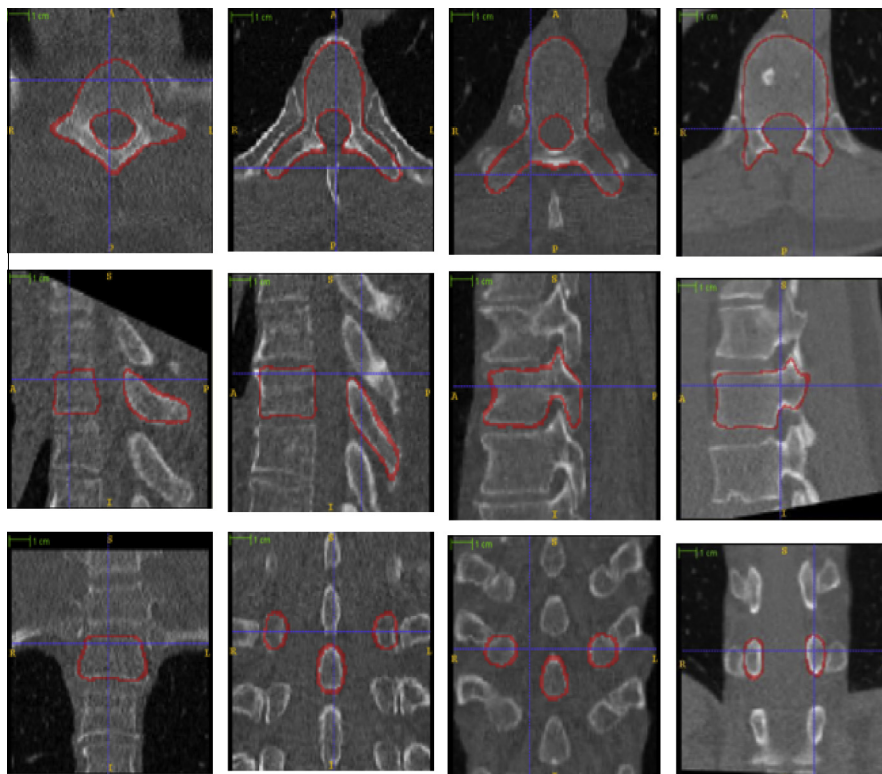


Fig. 10. Segmentation results of four individual vertebrae (columnwise) in axial or sagittal and coronal views. Different vertebrae are featured in different colors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

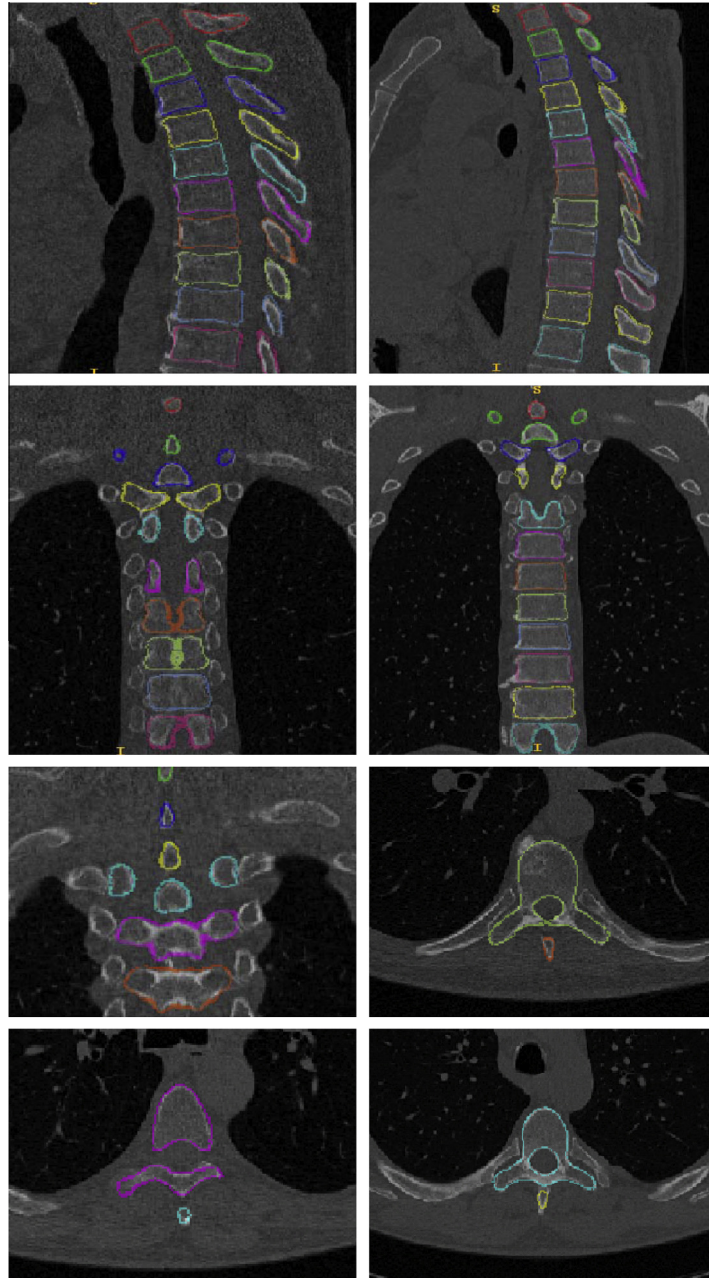


Fig. 11. Segmentation results of chosen volume in axial or sagittal or coronal view. Different vertebrae are featured in different colors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Figs. 10 and 11, we show some segmentation examples in axial, sagittal or coronal views, 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 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 [5] reports a comparable accuracy level at 1.12 ± 1.04 mm.

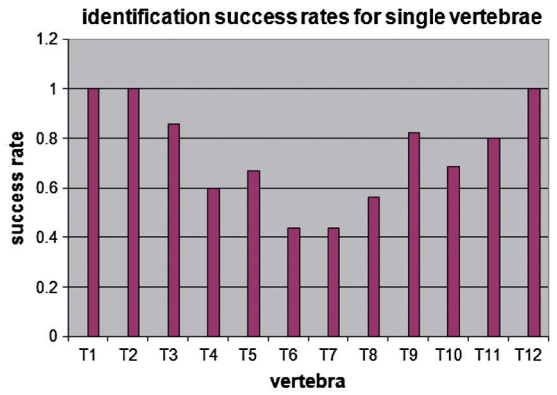
For identification, we have an average success rate of 73.1% using single vertebra. This success rate also varies regarding to a specific vertebra where the rates for T5, T6, T7, T8 as $\leq 60\%$ are especially lower than others because these four vertebrae look alike. Furthermore, when exploiting the vertebra strings or chains

for identification, the success rates are improved to increase with longer string lengths. With a string of 7 or 8 and more vertebrae, we achieve over 91% or $>95\%$ success rates, whereas rates are $\approx 73.1\%$ for one vertebra, $\approx 87\%$, 89% for 7 or 8 vertebra strings in

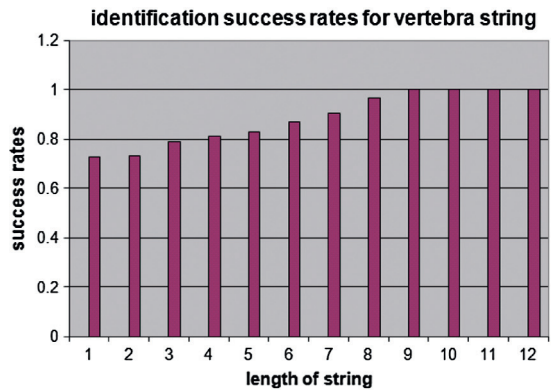
Table 1

Mean point-to-surface error and standard deviation for individual vertebra by manual initialization.

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



(a) individual vertebra identification rates



(b) vertebra string identification rates

Fig. 12. Identification success rates of individual vertebra and stringed vertebrae.

[5]. The success rates of individual and stringed vertebra identification (via mean mesh shapes) are comparable or better than [5] using intensity based matching, as shown in Fig. 12.

A volumetric mean appearance model is used for vertebra identification in [5], 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 is difficult to 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 advantageous and our identification results are indeed slightly better. Lastly, our vertebra segmentation method is applicable to other orthopedic bone structures as well, e.g., manubrium, as shown in Fig. 13.

3.2. Evaluation on automatic initialization via fast spine

We evaluate our vertebra segmentation and identification based on the initialization of the Siemens Fast Spine [18,30] software application. Fast Spine is developed using Learning Ensembles of Anatomical Patterns (LEAP) framework [29] through redundant anatomy feature detection, parsing and fusion [30,23]. An illustrative example is shown in Fig. 14. Its extension from CT spine labeling to MR is shown in [26]. Fast Spine can achieve high robustness and accuracy on 3D localization of the center origins $C_T(x,y,z)$, $C_B(x,y,z)$ of each vertebra body top plane and bottom plane, and the three orthogonal axes $\Upsilon_T(x,y,z)$, $\Upsilon_B(x,y,z)$ representing its local vertebra anatomy coordinates [5]. This anatomical information can uniquely determine the position and pose or orientation of each individual vertebra. In practice, we use translation + uniform scaling transformation to align the vertebra mesh model onto a new detection by Fast Spine. The geometry of vertebra detection is defined as the centroid point $C_C(x,y,z) = (C_T(-x,y,z) + C_B(x,y,z))/2$ and mean local anatomical coordinates $\Upsilon_C(x,y,z) = (\Upsilon_T(x,y,z) + \Upsilon_B(x,y,z))/2$. Note that Fast Spine also identifies and labels vertebra with very high accuracy $\sim 100\%$, as it always uses the full length of vertebra string information. Our vertebra identification is based on very different cues of robust shape matching, not landmark detection and counting in Fast Spine. Our method achieve about 100% identification accuracy when having strings of 9 vertebra or longer (Fig. 12). Therefore the performances of two methods are comparable and complementary for further enhanced robustness.

To increase the robustness, as discussed in Section 2.1, we add random perturbations on manually initialized vertebra pose variations in the range of $[-\pi/10; \pi/10]$ during training the edge response classifier and corresponding vertebra edge response map. This largely assures that our coarse-to-fine deformation scheme will converge correctly even under moderate to large perturbations on initialization. Using Fast Spine software application for automatic initialization (i.e., replacing our manual initialization), we achieve consistently more accurate segmentation results (Table 2) and comparable identification performance. In summary, we take the stratified, divide-and-conquer approach by solving vertebra detection via Fast Spine and segmentation problems in this paper.

3.3. Evaluation on different learning based bone boundary edge detectors

We study and compare different learning based boundary detectors through ROC analysis, specifically comparing our method with [28,10,7]. Note that 3D Haar based supervised edge detection does not work well in general since it cannot be steered efficiently in 3D space to fit edges with different orientations [21]. Thus, as in

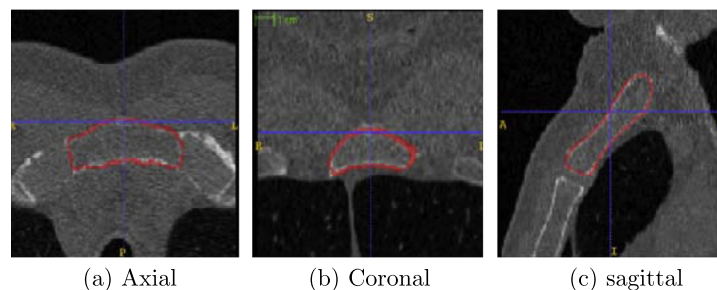


Fig. 13. Segmentation results of manubrium in axial or sagittal or coronal view.

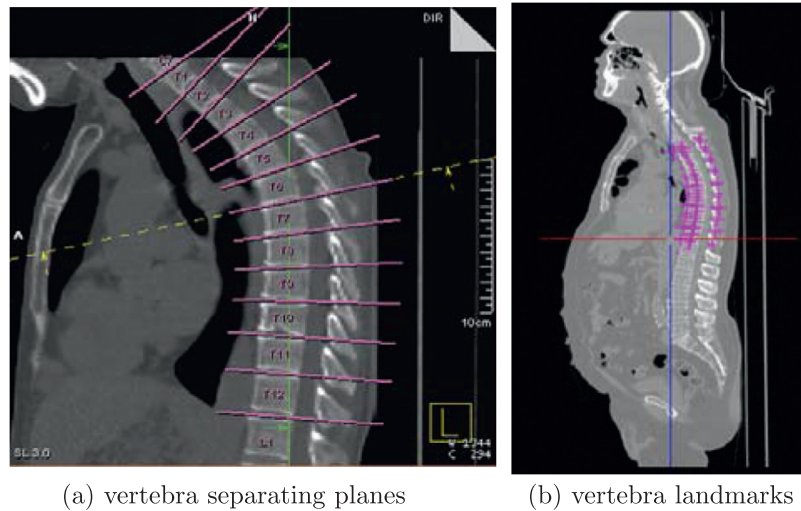


Fig. 14. Fast Spine application: (a) vertebra separating planes and (b) vertebra landmarks, derived from detected $C_T(x,y,z)$, $C_B(x,y,z)$, $C_C(x,y,z)$ and local 3D coordinates $Y_T(x,y,z)$, $Y_B(x,y,z)$, as output geometric results.

Table 2

Mean point-to-surface error and standard deviation for individual vertebra by automatic initialization via Siemens Fast Spine [30].

Vertebra	T1	T2	T3	T4	T5	T6
Mean error (mm)	1.02	1.02	0.98	0.93	0.96	0.91
Std deviation (mm)	0.94	0.99	1.01	1.11	1.23	0.88
Vertebra	T7	T8	T9	T10	T11	T12
Mean error (mm)	0.82	0.77	0.83	0.76	0.87	1.17
Std deviation (mm)	0.56	0.64	0.63	0.52	0.72	1.08

Table 3

Mean point-to-surface error (mm) of all vertebrae (averaged) by different number of iterations for part- ($Pt = 1, 2, 3, 4$) and patch- ($Ph = 1, 2, 3, 4, 5$) based deformable models.

Iteration #	$Ph = 1$	$Ph = 2$	$Ph = 3$	$Ph = 4$	$Ph = 5$
$Pt = 1$	4.21	3.88	3.65	3.54	3.46
$Pt = 2$	2.49	2.31	2.15	2.02	1.95
$Pt = 3$	1.54	1.40	1.25	1.12	1.11
$Pt = 4$	1.55	1.41	1.22	1.13	1.13

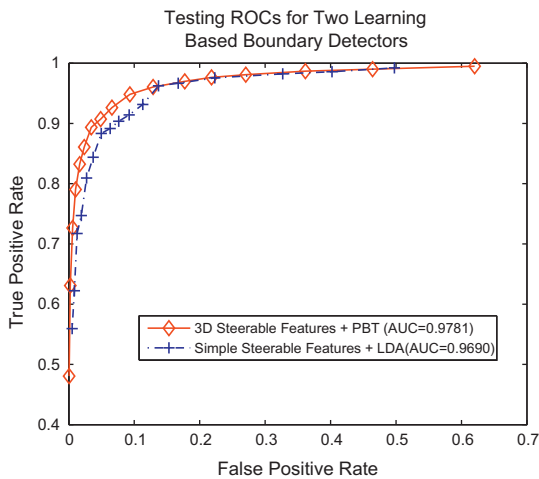


Fig. 15. Comparison of testing ROC curves for supervised vertebra edge detection using our simple steerable gradient vector + LDA classifier versus 3D steerable features + PBT scheme in [28,10,7].

[28,10,7], the organ boundary detectors are learned by boosted 3D steerable features using probabilistic boosting tree (PBT) classifier [19]. In brief, 3D Haar features [21] are generally believed to be unsuitable for voxel-level boundary learning with varying orientations, but are more common for global scale organ localization [28,7].

In our experiments, the very simple steerable gradient feature vector + LDA classifier performs slightly worse than 3D steerable features + PBT scheme in [28,10,7], for vertebra bone edge detec-

tion task via ROC analysis,³ as shown in Fig. 15. The testing and training ROC curves appear similar and only testing curves are shown for better clarity. Even the area under ROC curve (AUC) measurements are both reasonably high at 0.9781 and 0.9690 respectively, which still represents noisy edge responses in run time. On the other hand, our version of steerable feature + LDA classifier has only 15 computed features with a linear classifier, thus it has much lower model and computational complexity which may also indicate good generality to unseen data. As discussed in this paper, our deployed deformable model is sufficiently comprehensive and robustly designed to handle imperfect, noisy bone edge response maps. For segmentation purpose, the accuracy results are statistically different using the current detector or detectors in [28,10,7]. However it slightly helps improve the identification performance, especially for lower order vertebra strings, e.g., comparing 74.6% from 73.1% for one vertebra, 76.7% for strings of two vertebrae from Fig. 12.

3.4. Evaluation on different iteration numbers of part- and patch-based deformable models

As mentioned in Section 2.2, we perform multiple rounds of Part- and Patch-based deformable processes. The numerical relationship between segmentation accuracy (in terms of Mean point-to-surface error (mm) of all vertebrae averaged) and these iteration numbers ($Pt = 1, 2, 3, 4$; $Ph = 1, 2, 3, 4, 5$) are given in Table 3. We find that, with more iterations in both Pt and Ph , the averaged segmentation accuracy improves gradually until the respective iteration numbers reach $Pt = 3$ (Part) and $Ph = 4$ (Patch). There are no iterations in the vertebra identification task.

³ It could be significantly worse for low-contrast organ boundaries (e.g., liver, heart), which is not the focus of this paper.

4. Conclusion and discussion

In this paper, a hierarchical thoracic vertebra segmentation and identification method is presented. We propose steerable gradient features for learning based edge detection with intended misalignment perturbations for training dataset construction, to ensure robustness and convergence. 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 highly competitive with the point-to-surface error 0.95 ± 0.91 mm. We also use the generated mean shape model of each thoracic vertebra for identification process, where success rates of 73.1% for a single vertebra, and over 95% for 7 or more vertebra are achieved. Both our segmentation and identification performance compare favorably with the state-of-the-art [5], as surface based versus intensity based approach [5].

Our proposed hierarchical thoracic vertebra segmentation and identification method is directly extensible to other cervical or lumbar vertebrae or bone structures. For example [23], utilizes 2D Curved Multi-planar projection of CT images and Orthogonal Matching Pursuit (OMP) classification to distinguish thoracic versus lumbar Vertebrae. Therefore thoracolumbar vertebrae modeling can be solved in a divide-and-conquer manner.

References

- [1] T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, Active shape models-their training and application, *Comput. Vis. Image Underst.* 61 (1) (1995) 38–59.
- [2] P. Dollár, Z. Tu, S. Belongie, Supervised learning of edges and object boundaries, in: *IEEE Conf. on Computer Vision and Pattern Recognition*, 2006, pp. 1964–1971.
- [3] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, Wiley-Interscience, second ed., 2000.
- [4] J. Herring, B. Dawant, Automatic lumbar vertebral identification using surface-based registration, *Comput. Biomed. Res.* 34 (2) (2001) 629–642.
- [5] T. Klinder, J. Ostermann, M. Ehm, A. Franz, R. Kneser, C. Lorenz, Automated model-based vertebra detection, identification, and segmentation in CT images, *Med. Image Anal.* 13 (2009) 471–481.
- [6] T. Kubota, A.K. Jerebko, M. Dewan, M. Salganicoff, A. Krishnan, Segmentation of pulmonary nodules of various densities with morphological approaches and convexity models, *Med. Image Anal.* 15 (1) (2011) 133–154.
- [7] H. Ling, S.K. Zhou, Y. Zheng, B. Georgescu, M. Snhling, D. Comaniciu, Hierarchical, learning-based automatic liver segmentation, in: *IEEE Conf. on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.
- [8] W.E. Lorensen, H.E. Cline, Marching cubes: a high resolution 3d surface construction algorithm, *Comput. Grap.* 21 (4) (1987).
- [9] L. Lu, A. Barbu, M. Wolf, J. Liang, L. Bogoni, M. Salganicoff, D. Comaniciu, Simultaneous detection and registration for ileo-cecal valve detection in 3d CT colonography, in: *Proc. of European Conf. on Computer Vision*, vol. 4, 2008a, pp. 465–478.
- [10] L. Lu, A. Barbu, M. Wolf, J. Liang, M. Salganicoff, D. Comaniciu, Accurate polyp segmentation for 3d CT colonography using multi-staged probabilistic binary learning and compositional model, in: *IEEE Conf. on Computer Vision and Pattern Recognition*, 2008b, pp. 1–8.
- [11] J. Ma, L. Lu, Y. Zhan, X.-S. Zhou, M. Salganicoff, A. Krishnan, Hierarchical segmentation and identification of thoracic vertebra using learning-based edge detection and coarse-to-fine deformable model, *MICCAI* (2010) 19–27.
- [12] D.R. Martin, C. Fowlkes, J. Malik, Learning to detect natural image boundaries using local brightness, color, and texture cues, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (5) (2004) 530–549.
- [13] A. Mastmeyer, K. Engelke, C. Fuchs, W. Kalender, A hierarchical 3d segmentation method and the definition of vertebral body coordinate systems for QCT of the lumbar spine, *Med. Image Anal.* 10 (4) (2006) 560–577.
- [14] K. Murphy et al., A large-scale evaluation of automatic pulmonary nodule detection in chest ct using local image features and k-nearest-neighbour classification, *Med. Image Anal.* 13 (2009) 757–770.
- [15] Z. Peng, J. Zhong, W. Wee, J. Lee, Automated vertebra detection and segmentation from the whole spine mr images, in: *Proc. of IEEE EMB*, 2005, pp. 2527–2530.
- [16] D. Shen, C. Davatzikos, An adaptive-focus deformable model using statistical and geometric information, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (8) (2000) 906–913.
- [17] H. Shen, A. Litvin, C. Alvino, Localized priors for the precise segmentation of individual vertebrae from ct volume data, *MICCAI* (2008) 367–375.
- [18] Siemens, 2011. Siemens Fast Spine Application, Siemens Medical Solutions.
- [19] Z. Tu, Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering, in: *IEEE Int. Conf. on Computer Vision*, 2005, pp. 1589–1596.
- [20] Z. Tu, S. Zheng, A. Yuille, A. Reiss, R. Dutton, A. Lee, A. Galaburda, I. Dinov, P. Thompson, A. Toga, Automated extraction of the cortical sulci based on a supervised learning approach, *IEEE Trans. Med. Imag.* 26 (4) (2007) 514–552.
- [21] Z. Tu, X. Zhou, L. Bogoni, A. Barbu, D. Comaniciu, Probabilistic 3d polyp detection in ct images: the role of sample alignment, in: *IEEE Conf. on Computer Vision and Pattern Recognition*, 2006, pp. 1544–1551.
- [22] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: *IEEE Conf. on Computer Vision and Pattern Recognition*, 2001, pp. 511–518.
- [23] T. Wu, B. Jian, X.S. Zhou, Automated identification of thoracolumbar vertebrae using orthogonal matching pursuit, *Mach. Learn. Med. Imag.* (2011) 126–133.
- [24] J. Yao, S. O’Connor, R. Summers, Automated spinal column extraction and partitioning, in: *Proc. of IEEE ISBI*, 2006, pp. 390–393.
- [25] X. Ye, X. Lin, J. Dehmeshki, G. Slabaugh, G. Beddoe, Shape based computer-aided detection of lung nodules in thoracic ct images, *IEEE Trans. Biomed. Eng.* 56 (7) (2009) 1810–1820.
- [26] Y. Zhan, M. Dewan, M. Harder, X. Zhou, Robust MR spine detection using hierarchical learning local articulated model, *MICCAI* (1) (2012) 141–148.
- [27] Y. Zhan, D. Shen, Deformable segmentation of 3-d ultrasound prostate images using statistical texture matching method, *IEEE Trans. Med. Imag.* 25 (3) (2006) 256–272.
- [28] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, D. Comaniciu, Four-chamber heart modeling and automatic segmentation for 3-d cardiac CT volumes using marginal space learning and steerable features, *IEEE Trans. Med. Imag.* 27 (11) (2008) 1668–1681.
- [29] X. Zhou, Z. Peng, Y. Zhan, M. Dewan, B. Jian, A. Krishnan, Y. Tao, M. Harder, S. Grosskopf, U. Feuerlein, Redundancy, redundancy, redundancy: the three keys to highly robust anatomical parsing in medical images, *Multimedia Inform. Retrieval*. (2010) 175–184.
- [30] X. Zhou, Y. Zhan, Z. Peng, M. Dewan, B. Jian, A. Krishnan, M. Harder, R. Schwarz, L. Lauer, H. Meyer, S. Grosskopf, U. Feuerlein, H. Ditt, Redundancy, Redundancy, Redundancy: The Three Keys to Highly Robust Anatomical Parsing in Medical Images, *Cost-Sensitive Machine Learning*, CRC Press, 2011, pp. 219–240.