Localized Priors for the Precise Segmentation of Individual Vertebras from CT Volume Data

Hong Shen[1], Andrew Litvin[2], and Christopher Alvino[1]

[1] Siemens Corporate Research, Inc., 755 College Road East, Princeton, NJ 08540
{shen.hong,christopher.alvinvo}@siemens.com
[2] Analogic Corporation, 8 Centennial Drive, Peabody, MA 01960
alitvin@analogic.com

Abstract. We present algorithms for the automatic and precise segmentation of individual vertebras in CT Volume data. When a local surface evolution method such as the level set is applied to such a complex structure, global shape priors will not be sufficient to avoid the leakage and local minima problems, particularly if precise object boundary is desired. We propose a prior knowledge base that contains localized priors—a group of high-level features whose detection will augment the surface model and be the key to success. Base on this a set of context blockers are applied to prevent the leakages. Carefully designed initial surface when registered with the data helps avoid the local minimum problem. The results of segmentation well approximate the human delineated object boundaries. We also present the validation result of the segmentation of 150 vertebras.

1 Introduction

The vertebra in high-contrast CT images as shown in Fig.1(a)-(i) represents a type of 3D structure that is complex yet prominent. The images differ from many other medical images, such as organs in MR or Ultrasound data, in which the shape is relatively simple, but with obscure boundaries even for a human observer. The vertebra in CT data has well-defined boundaries from a human observer's point of view, in that an educated observer can perform consistent and relatively precise delineation of the object boundaries on 2D slice images. On the other hand, the boundaries when viewed in detail are full of challenges for any segmentation algorithm. The vertebra of interest is usually adjacent or even connected to neighboring structures that have similar intensities. The internal regions are highly inhomogeneous, and the weak and diffused boundaries contain many gaps. Segmentation of the vertebra is also a typical 3D problem not to be solved on a slice-by-slice basis, due to the large variation of 2D cross-sectional shape from the 3D structural complexity, as shown in Fig.1(e)-(i). To appreciate the complexity, our timing for a human observer to delineate the 2D boundaries of a single vertebra slice-by-slice using a digital tablet is typically around 45 minutes.

The segmentation of the vertebra is of high importance to orthopedic applications. Interactive [1] and automatic methods were reported [2-5]. For an overview please see

D. Metaxas et al. (Eds.): MICCAI 2008, Part I, LNCS 5241, pp. 367–375, 2008.

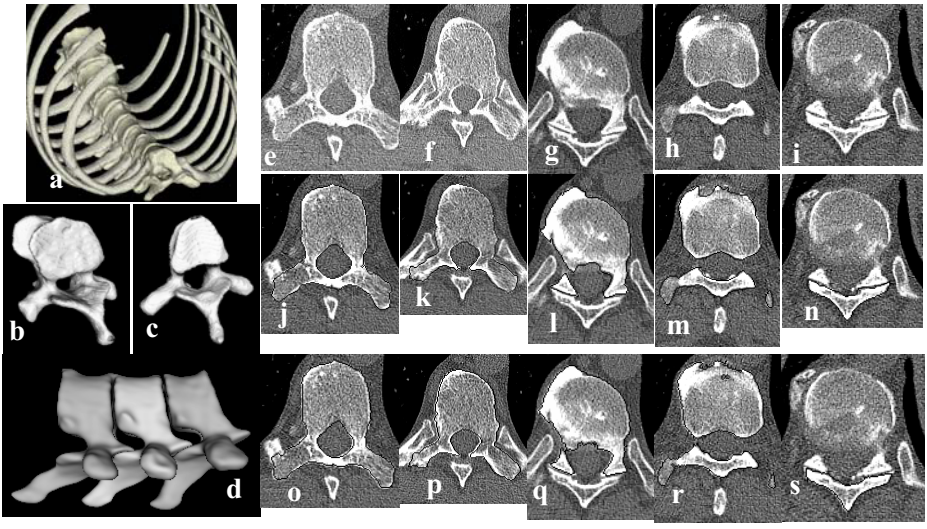

Fig. 1. 3D views of (a) the vertebras and rib structures form the thoracic cage. (b)(c) two segmented vertebras (d) three individually segmented and adjacent vertebras. (e)-(i): Cross-sections of one vertebra on sample axial slices, together with its neighboring vertebra and rib structures. (j)-(n): Corresponding 2D boundaries of the vertebra of interest delineated by a human observer on axial slices. The regions not enclosed by the boundaries belong to a neighboring vertebra or rib structure.(o)-(s) 2D boundaries of the vertebra segmented by the automatic algorithm.

Naegel [3]. None of the works however were able to automatically and precisely segment the complete individual vertebras from the complex neighboring structures.

In our previous work [6], we presented a preliminary solution for such a problem. We also discussed the problems of such typical method as described in [7], in which the level set is constrained by a global PCA model. The summary of discussion is as follows. First, the high consistency of human manual segmentation (Fig. 1(j)-(n)) requires precise matching of the auto-segmentation (Fig. 1(o)-(s)) to the actual object boundary, including any pathological deformation (e.g. Fig. 1(g); We recently measured the typical consistency between two independent human observers on the manual segmentation of a vertebra as over 94%). Second, a representative PCA model space covering the high variation and complex structure as well as unpredictable pathologies is theoretically and practically unachievable. Third, therefore we cannot solve the leakage and local minimum problem of level set using tight constraints from a global PCA model. Fourth, adding the model force and local image force together--and using a global multiplier to adjust relative strength [7] --creates competition between global model and local data forces. The dilemma lies in whether the model should be made stronger to prevent leakage or weaker to allow the surface convergence to the local data that is not covered in the incomplete model space. From the above, we believe the global model should serve as the optimal initialization for the level set rather than as a soft constraint applied during surface evolution. In this paper, we further propose the concept of *localized priors* that will guide the level set to avoid leakage and local minima at the places where most necessary.

2 Methods

2.1 The Mean Shape Model and the Vertebra Coordinate System

As shown in Fig. 2(a), the centerline of the spinal canal formed by the central cavity of all the vertebras was automatically extracted. Our system also fitted a set of planes to the inter-vertebral disc spaces at every interface of the adjacent vertebra bodies. The automated extraction of the spinal cord and inter-vertebra planes from a CT volume are yet to be detailed in a future publication. On any vertebra, a vertebra coordinate system ($\boldsymbol{O}$, $\boldsymbol{u}_x$, $\boldsymbol{u}_y$,$\boldsymbol{u}_z$) can be defined from the extracted local centerline section and the top and bottom inter-vertebra planes, as shown in Fig. 2(b). The spinal canal centerline makes two intersections $\boldsymbol{T}$ and $\boldsymbol{B}$ at the top and bottom inter-vertebra planes, and we define the origin $\boldsymbol{O}$ as the middle point of $\boldsymbol{T}$ and $\boldsymbol{B}$. We define the z-axis unit vector $\boldsymbol{u}_z$ as the average of the two plane normals. On the plane orthogonal to $\boldsymbol{u}_z$ and passing through $\boldsymbol{O}$, $\boldsymbol{u}_y$ is defined as the symmetrical axis of the 2D image on the plane. This comes from the 2D axial symmetric property of the vertebra and the rib structures. $\boldsymbol{u}_x$ is obtained by applying the right-hand rule. The scales on the three directions can be computed from the original axial, coronal, and sagittal resolutions.
We manually segmented a small number of vertebras and registered the surfaces of the vertebras such that $\boldsymbol{u}_x$, $\boldsymbol{u}_y$, $\boldsymbol{u}_z$ and $\boldsymbol{O}$ of all the vertebras coincided. In this common coordinate system, we constructed from the transformed surfaces a set of signed distance functions $S = \{\phi_1, \phi_2, ..., \phi_n\}$ such that the vertebra surfaces were represented as $\Gamma = \{\vec{x} \mid \vec{x} \in R^3, \phi(\vec{x}) = 0\}$. The mean shape shown in Fig. 2(b) was constructed as

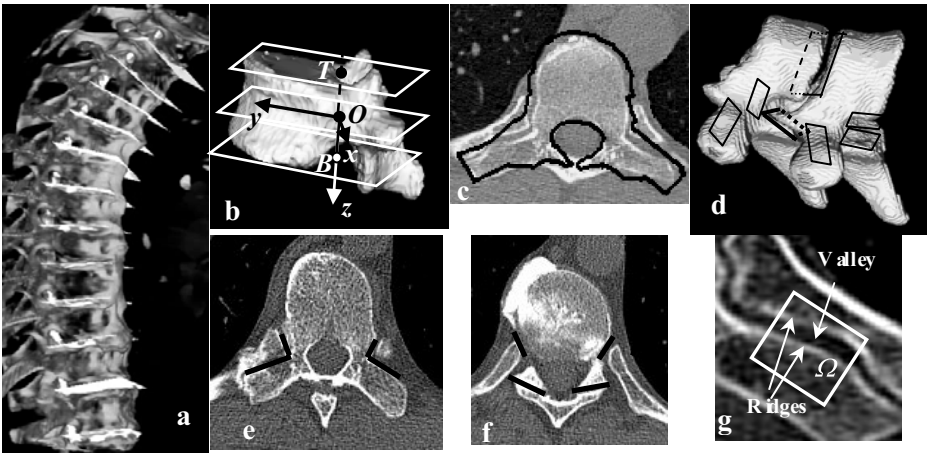

Fig. 2. (a) Spinal canal centerline and inter-vertebra planes overlaid on the spine. (b) The vertebra coordinate system based on the mean shape model. (c) The mean shape registered to data being used as the initial front. (d) Plane models at object interfaces representing high-level context features registered with the shape model. (e)(f) The axial projections of the plane models (g) Sandwich structure of the interface and the search ROI for the plane fitting.

$$\bar{\phi} = \frac{1}{n}\sum_{i=1}^{n}\phi_i \tag{1}$$

The mean shape model was not intended to cover the shape variation among the vertebras, but it can be used to construct the initial front for level set evolution. Prior to the level set evolution, the whole data volume is automatically divided into volume of interests (VOI) centered at the vertebras, using the inter-vertebra planes (Fig. 2(a)) as references. The mean shape model was transformed into each of the VOI's. The registered surface was augmented such that its surface is close to and outside of the vertebra boundaries. [6] The augmented shape as shown in Fig. 2c was used as the initial front, thus to avoid the energy minimums inside the vertebra from the start.

2.2 Localized Priors

At the interfaces between two bright bone regions, where level set leakage usually occurred due to close contact of the surfaces, we defined plane models as illustrated in Fig. 2(d)-(f). These plane models were approximations of the curved surfaces at these interfaces, and we record them as high-level context features in the vertebra coordinate system associated with the mean shape model. In contrast to a global shape model, they represent prior knowledge at key locations and therefore are *localized priors*. The plane model was a simple and hence more robust representation, even though the interface may actually be a curved surface.

Given a VOI, the localized priors were transformed together with the mean shape model. In the local neighborhood of each mapped prior, we applied 3D steer-able filter [8] to detect the local primitives. We tuned the steer-able filter such that it only detected the primitives with orientations in the neighborhood of the prior orientation, thus reducing the noise and irrelevant primitives. We also switched the parameters of the steer-able filter to detect edges, ridges, and valleys, respectively. [8] As can be seen on Fig.2(g), the typical interface was a sandwich structure, with two layers of intensity ridges and one intensity valley between the two ridges. Instead of only fitting the plane model to the valley, three parallel planes with a fixed distance s apart were jointly fitted to the sandwich structure to improve robustness.

The search for the optimal location of the planes was conducted in a limited local neighborhood Ω, using gradient descent method. We denoted the set of valley primitives as V, ridge primitives as R, and the three sandwich planes to be detected as L, L^s and L^{-s}, respectively. The cost function was given by

$$E = \sum_{p\in V\cap\Omega} S_V(p)D(p,L) + 0.5\sum_{p\in R\cap\Omega} S_R(p)D(p,L^s) + 0.5\sum_{p\in R\cap\Omega} S_R(p)D(p,L^{-s}) \tag{2}$$

where $S_v(p)$ and $S_R(p)$ were the filter responses of the valley and ridge primitives, respectively. $D(p, L)$ was the robust distance computed from the Euclidean point-to-plane distance $d(p, L)$ through differentiable p-norm with cut-off re-descending:

$$D(p,L) = \begin{cases} \left(d(p,L)^2 + \alpha\right)^{m/2} & d(p,L) < T \\ 0 & else \end{cases} \tag{3}$$

where m was the p-norm parameter, and T was the distance threshold.

Most of the optimal planes fitted using gradient decent method closely approximated the actually interfaces. However, the interface between the rib and the transverse process of the vertebra was a rather curved surface. We therefore performed refinement in the neighborhood of the fitted plane. As shown in Fig. 3(a), we represented the initial surface with a mesh computed by uniform sampling of the plane. The element (i, j) was allowed only to move in the direction orthogonal to the plane, i.e., within the parallelepiped C_{ij}. The final mesh was estimated by solving the optimization problem

$$\{\hat{d}_{ij}\} = \underset{d_{ij}\in[-R,R]}{\arg\min} \left[\sum_{i,j} \sum_{k\in V\cap C_{ij}} S_V(k) \mid d(k,L) - d_{ij} \mid + \sum_{ij} d_{ij}^2\right] \tag{4}$$

where d_{ij} was the displacement of the mesh element with respect to the starting plane L, $S_v(k)$ was the strength of the valley primitive set V, and R was the maximum allowable displacement. The first term measures the weighted distance sum of the valley points in the parallelepiped to the mesh element, and becomes minimal when the mesh element is close to or on the valley primitives. The second term restricts the mesh element from being too far away from the original plane. The energy is the sum over all mesh elements. The estimated optimal displacements of all mesh elements define the final surface mesh. Shown in Fig. 3(b)(c) are the two final surfaces at the interface of the transverse process and the ribs.

Fig. 3. Surface refinement based on fitted planes. (a) Mesh refinement scheme. (b) 3D rendering of the final surface (c) Axial projection of the surface. (d) The blocker regions formed based on the high-level features.

As shown in Fig. 3(d), the fitted planes and surfaces as detected high-level features are used to form blocker regions. To decide if a point is inside or outside of the blocker region, the origin ***O*** of the vertebra coordinate system is used as the reference. [6] Given the block region Ω_B, we define the "blocker speed" as

$$v_B(\vec{x}) = \begin{cases} 1, & \vec{x} \in \Omega_B \\ 0, & else \end{cases} \tag{5}$$

2.3 Level Set Formulation

The image speeds included both region-based terms and the primitive-bases terms. [6]

$$F_{image}(\vec{x}) = \alpha[(I(\vec{x}) - c_1)^2 - (I(\vec{x}) - c_2)^2] + \beta \nabla g(\vec{x}) \cdot \nabla \phi(\vec{x}, t), \tag{6}$$

where c_1 and c_2 were the intensity constants for soft tissues and cortical bones, respectively, and

$$g(\vec{x}) = \frac{1}{1 + D_E(\vec{x}) + D_R(\vec{x})}, \tag{7}$$

D_E and D_R represented the edge and ridge responses from the 3D steerable filter [8],

$$F_{final}(\vec{x}) = [1 - \nu_B(\vec{x})]F_{image}(\vec{x}) - \nu_B(\vec{x}) + \varepsilon\kappa \tag{8}$$

where κ was the mean curvature of the level set. The evolution of the surface was embedded in the evolution of a level set function $\phi(t)$:

$$\frac{\partial \phi}{\partial t} = F_{final}(\vec{x}) \; |\nabla \phi| \tag{9}$$

3 Results

We tested our algorithm on 21 thoracic CT volumes with axial resolution of 0.6~1 mm, each containing 6-8 thoracic vertebras. The total number of vertebras came up to 150. To provide quantitative validation for the automatic segmentations, we compared to human delineated results. The complete manual delineation of the vertebra by a human observer was very tedious and time consuming (e.g., 45 minutes for a single vertebra traversing 78 axial slices), which was the very reason that we needed an automatic algorithm. It was even more prohibitive to have a physician to perform such a task. As an alternative, we used the automatic segmentation result as the starting point of human delineation and the edited result were used as the ground truth for evaluation. To justify this, we tested the consistency between the edited results and the completely manual results on a couple of vertebras. Two human observers independently delineated these vertebras, and the results were examined by an orthopedic surgeon as acceptable. A third observer edited the automatic results of these same vertebras. The consistency between two result point sets S_1 and S_2 using S_2 as the reference was measured by the under-segment rate

$$u(S_1|S_2) = |U|/|S_2| \times 100\%, \quad where \quad U = \{p | p \in S_2 \; and \; p \notin S_1\}, \tag{10}$$

and the over-segment rate

$$o(S_1|S_2) = |O|/|S_2| \times 100\%, \quad where \quad O = \{p | p \in S_1 \; and \; p \notin S_2\}, \tag{11}$$

and $|.|$ was the size of the point set. Shown in Table 1 are the under-segment and over-segment rates using the completely manual results as the references. From these we concluded that the inter-personal consistency was high enough that any educated human observer would be able to perform a consistent manual segmentation that was

acceptable to an orthopedic surgeon. Further, the edited result based on the auto segmentation result was as good as a completely manual segmentation. Therefore to use the edited result as the reference for validation was reasonable and sound.

The 150 vertebras were segmented using our algorithm. The result of each vertebra was manually edited by a human observer and recorded as the reference. Shown in Fig. 4 are the histograms of the under-segment and over-segment rates of the automatic results using the edited results as the references. From the histogram we observed that the general segmentation results were satisfactory, but under-segmentation happened more frequently than over-segmentation.

Shown in Fig. 5 is a typical segmentation result. The 2D boundaries of the results were overlaid on the original data, which matched reasonably to the vertebra boundaries despite the adjacent and connected neighboring structures. Strictly speaking, the level set surface does appear to be a bit over-smoothed to preserve every details of the vertebra surface. Minor under-segmentation also existed at a few places. The blocker regions and initial level set front shown in Fig. 5(k)-(o) helps us understand the reason for success: The initial front was close to and outside of the vertebra boundaries, which avoided the internal local minimums; the blocker region forced the level-set to evolve outside of the neighboring structures so that leakage was prevented.

The un-optimized automatic algorithm took about 7 minutes in average for each vertebra, and the manual editing based on the automatic result took about 5 minutes in average.

Table 1. Consistency measures of manual and edited results

	Manual result 1		Manual result 2	
S_1	$u(S_1 \mid S_2)$	$o(S_1 \mid S_2)$	$u(S_1 \mid S_2)$	$o(S_1 \mid S_2)$
Manual result 1	n/a	n/a	5.3%	5.1%
Manual result 2	5.7%	5.5%	n/a	n/a
Edited result	6.2%	5.9%	5.8%	6.4%

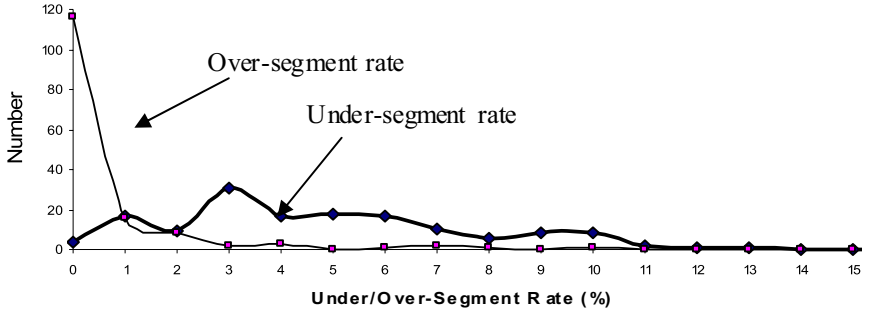

Fig. 4. Histogram of the under and over-segment rates

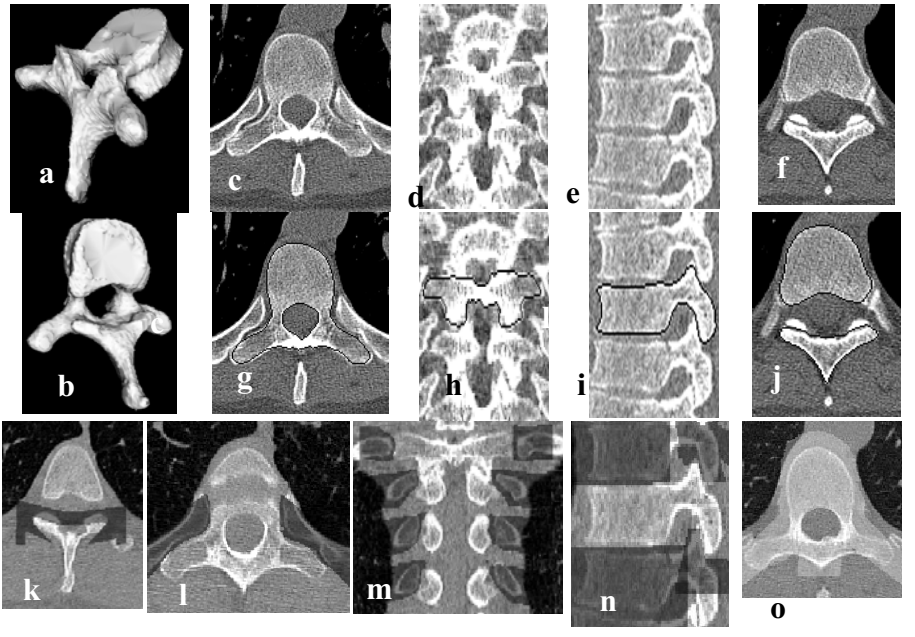

Fig. 5. A sample automatic segmentation result before human editing. (a)(b) 3D surface rendering of the segmented vertebra. (c)-(f) Sample axial, coronal and sagittal views of the original data (g)-(j) 2D boundaries of the segmentations. (k)-(n) Blocker regions (dark) as an intermediate result of the algorithm. (o) The initial front of the level set (white).

4 Conclusions

We reported in this paper an automatic method and the validated results for the precise segmentation of individual vertebras from a CT volume. The algorithm can be further improved; we should also investigate the effect of the key components such as the blocker extraction and the initial front in more detail. The current result, however, is useful with minor user edits. The concept of localized priors should be further explored to provide a more general framework, e.g. a statistical description to cover the variations. It is imaginable that such a set of localized statistical models should form a much smaller model space compared to a global model.

References

1. Kaminsky, J., Klinge, P., Rodt, T., Bokemeyer, M., Luedemann, W., Samii, M.: Specially adapted interactive tools for an improved 3D-segmentation of the spine. Computerized Medical Imaging and Graphics 28, 119–127 (2004)
2. Tan, S., Yao, J., Ward, M.M., Yao, L., Summers, R.M.: Computer aided evaluation of ankylosing spondylitis. In: 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, pp. 339–342 (2006)

3. Naegel, B.: Using mathematical morphology for the anatomical labeling of vertebrae from 3D CT-scan images. Computerized Medical Imaging and Graphics 31(3), 141–156
4. Smyth, P.P., Taylor, C.J., Adams, J.E.: Automatic measurement of vertebral shape using active shape models. Image Vis. Computing 15, 575–581 (1997)
5. Lorenz, C., Krahnstover, N.: 3D Statistical Shape Models for Medical Image Segmentation. In: Second International Conference on 3-D Imaging and Modeling, p. 0414 (1999)
6. Shen, H., Shi, Y., Peng, Z.: Applying Prior Knowledge in the Segmentation of 3D Complex Anatomic Structures. In: The 1st International Workshop on Computer Vision for Biomedical Image Applications, Beijing, China (October 2005)
7. Leventon, M.E., Grimson, W.E., Faugeras, O.: Statistical shape influence in geodesic active contours. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition, vol. 1(1), pp. 316–323 (2000)
8. Freeman, W.T., Adelson, E.H.: The design and use of steerable filters. IEEE Transactions on Pattern Analysis and Machine Intelligence 13(9), 891–906 (1991)