Region Detection by Minimizing Intraclass Variance With Geometric Constraints, Global Optimality, and Efficient Approximation

Xiaodong Wu[*] **[Senior Member, IEEE]**,
Department of Electrical and Computer Engineering and the Department of Radiation Oncology, The University of Iowa, Iowa City, IA 52242 USA

Xin Dou,
Department of Electrical and Computer Engineering, The University of Iowa, Iowa City, IA, 52242 USA. He is now with EDDA Technology Inc., Princeton, NJ 08540 USA

Andreas Wahle [Member, IEEE], and
Department of Electrical and Computer Engineering, The University of Iowa, Iowa City, IA, 52242 USA

Milan Sonka [Fellow, IEEE]
Department of Electrical and Computer Engineering, the Department of Ophthalmology and Visual Sciences, and the Department of Radiation Oncology, The University of Iowa, Iowa City, IA, 52242 USA

Xin Dou: douxin98@gmail.com; Andreas Wahle: andreas-wahle@uiowa.edu; Milan Sonka: milan-sonka@uiowa.edu

Abstract

Efficient segmentation of globally optimal surfaces in volumetric images is a central problem in many medical image analysis applications. Intraclass variance has been successfully utilized for object segmentation, for instance, in the Chan–Vese model, especially for images without prominent edges. In this paper, we study the optimization problem of detecting a region (volume) between two coupled smooth surfaces by minimizing the intraclass variance using an efficient polynomial-time algorithm. Our algorithm is based on the shape probing technique in computational geometry and computes a sequence of minimum-cost closed sets in a derived parametric graph. The method has been validated on computer-synthetic volumetric images and in X-ray CT-scanned datasets of plexiglas tubes of known sizes. Its applicability to clinical data sets was also demonstrated. In all cases, the approach yielded highly accurate results. We believe that the developed technique is of interest on its own. We expect that it can shed some light on solving other important optimization problems arising in medical imaging. Furthermore, we report an approximation algorithm which runs much faster than the exact algorithm while yielding highly comparable segmentation accuracy.

Index Terms

Global optimization; image segmentation; intra-class variance; optimal region detection; parametric search; shape probing

---

[*]xiaodong-wu@uiowa.edu).

I. Introduction

Efficient segmentation of globally optimal surfaces in volumetric images is a central problem in many medical image analysis applications. While edges defined by image gradients are commonly used for segmentation, many object boundaries in medical image data may lack strong edges. For example, multiple adjacent objects with similar intensity profiles may be locally noisy and may not exhibit distinct edge properties. Image segmentation having the capability of handling weak edges is crucially important in medical image analysis. Intraclass variance has been successfully used in the Chan–Vese active contour model *without* using image gradient [1], which is based on a piecewise constant minimal variance criterion of the Mumford-Shah functional [2]. The following formula captures the intraclass variance, which is the data-driven term of the energy function used by Chan and Vese:

$$\mathcal{E}(\mathcal{S}) = \int_{\text{inside}(\mathcal{S})} |u_0(x, y, z) - c_1|^2 dxdydz + \int_{\text{outside}(\mathcal{S})} |u_0(x, y, z) - c_2|^2 dxdydz \quad (1)$$

where u_0 is the image, $\mathcal{S}$ is a variable closed boundary surface, and the constants c_1 and c_2, depending on $\mathcal{S}$, are the average intensity of u_0 inside and outside $\mathcal{S}$, respectively. This energy function (intraclass variance) was proven capable of producing promising results [1]. However, Chan and Vese's method lacks the ability of finding the global optimality. Chan and Vese also considered two regularization terms in their energy function, which regularize the length of the boundary and the area of the region. The regularization terms tend to smooth the boundary of the target object.

We report a novel algorithm for finding a globally optimal solution to segmentation by minimizing the *intraclass variance*. Our approach detects an optimal region between two coupled terrain-like surfaces in a volumetric image in a low-order polynomial time. Instead of adding the smoothness regularization term to the objective function as in the Chan–Vese model, we explicitly enforce the smoothness of the target surfaces with geometric constraints between neighboring voxels on the surfaces (see details in Section III). The proposed method is limited to handling those target objects that can be "unfolded" into two coupled terrain-like surfaces, which may seem to highly limit the application scope of the method. However, as we will demonstrate, the guarantee of global optimality and the freedom to design problem-specific cost functions allow the method to be applied to various medical image segmentation problems. Examples include the delineation of inner and outer airway wall surfaces in pulmonary CT images and the detection of endocardial and epicardial boundaries of the left ventricle from cardiac MR, both of which are difficult to solve by previous techniques. We show that the optimal solution can be obtained via the construction of the convex hull for a set of $O(n)$ unknown 2-D points using the shape probing technique [3], [4] in computational geometry, where n is the size of the input image. The probing oracles are implemented by computing a minimum s-t cut in a weighted directed graph. The intraclass variance can then be minimized by a sequence of calls to the minimum s-t cut algorithm. The shape probing technique has been used for image segmentation in the past [5], [6]. We also believe that the developed technique is of interest on its own. We expect that it can help solving other important optimization problems in medical imaging. We further develop an approximation method which runs much faster than the exact algorithm while yielding highly comparable segmentation accuracy. Part of the preliminary results related to this research has been presented in a conference paper [7].

II. Previous Work

A. Chan–Vese Related Optimization Methods

Several methods have been published for globally optimizing the Chan–Vese functional. Chan *et al.* [8] developed global minimizers to the Rudin–Osher–Fatemi model for denoising and extended the idea to the Mumford–Shah and Chan–Vese models. Based on that research, Bresson *et al.* [9] defined new variational models (including the Chan–Vese functional) unifying the classical snake and the geodesic active contour models. They developed global minimizers to their segmentation variational models based on a dual formulation of the total variational (TV) norm. In their approach, the mean intensities of the target object and the background are fixed in each iteration and updated occasionally, aiming at converging to the actual values of the average intensities. Li *et al.* [10] developed a graph searching based approach for globally optimal detection of multiple surfaces, which only utilizes the image gradient information. As a by-product, they derived an approximation method for optimizing the Chan–Vese cost functional by estimating the mean intensity. This approach, however, does not guarantee a globally optimal solution. Furthermore, the previously presented method for the Chan–Vese functional works only for the detection of a target object bounded by a single surface.

Global optimization for other variational segmentation methods has recently attracted intensive research. For example, Cohen and Kimmel [11] and Appleton and Talbot [12] developed methods to compute globally optimal solutions to the active contour and the geodesic active contour models, respectively. Appleton and Talbot further extended their approach in [13] to compute minimal surfaces using continuous maximal flows. Ardon *et al.* [14] generalized globally minimal paths for curve segmentation in 2-D to surface segmentation in 3-D. Their approach exploits the solution to the Eikonal equation and generates a function whose zero level set contains all the globally minimal paths between two user-supplied curves through a linear partial differential equation. However, all those variational methods inherently rely on approximating numerical schemes which must be very carefully designed to ensure robustness and convergence.

B. Graph Search and Graph Cut Methods

Many graph-based methods have been developed for optimal segmentation of medical image data. Methods using the graph-searching principle [15]–[19] have become important medical image segmentation tools in 2-D. But it was not until recently that the graph-search method was extended to d-D (d 3). Wu and Chen [20] first developed a graph-search based method for single surface detection and then Li *et al.* [10] extended the work for detecting multiple interacting surfaces. Generalized version of this method is dubbed LOGISMOS, which stands for layered optimal graph-based image segmentation of multiple objects and surfaces. Note that the LOGISMOS approach has been developed independently of and in parallel to the well-known approach of Boykov and Jolly [21]. The basic idea of the LOGISMOS framework is to formulate the tasks as a minimum-cost closed set problem in graph theory, which can be solved by a minimum *s-t* cut algorithm. The two approaches, despite sharing the same underlying *s-t* cut graph optimization, are substantially different in the way how image surface segmentation problem is represented and especially in the ability to simultaneously deal with multiple objects and multiple surfaces.

Recently, the graph-cut based segmentation method of Boykov *et al.* (e.g., [22]), which represents an option for optimally performing segmentation tasks in 3D, has attracted a lot of attention. The cost function employed in their work follows the "Gibbs model," which is general enough to include both the region and boundary properties of the target objects. Their approach, which is topologically flexible and shares some elegance with the level set

methods, has been quite successful. Though desirable for segmenting objects of unknown structures in many applications, topology flexibility in medical image segmentation may sometimes be considered a liability rather than an advantage since many medical structures have known topologies. Furthermore, Boykov *et al.*'s method is non-trivial to be extended to simultaneous detection of coupled surfaces. Delong and Boykov's work [23] is one step towards that goal. More recently, Grady [24] presented a method for minimal surface computation based on the minimum-cost circulation network flow algorithm, which is an extension of the shortest path segmentation method to 3D.

C. Coupled Surface Segmentation

Due to the imperfections of medical imaging techniques, insufficient image-derived information may be available for defining an object boundary or surface. This insufficiency can be remedied by using clues from other mutually related boundaries or surfaces. Co-optimization of coupled surfaces thus frequently yields superior results compared to the traditional single-surface detection approaches. Several methods for handling coupled surfaces have been proposed in recent years (e.g., [25]–[28]). They all demonstrate good performance in a wide variety of medical image segmentation problems. None of them, however, guarantees a globally optimal solution. The method in [28] is essentially 2-D and the performance depends on the initial model. It is based on coupled parametric deformable models with self-intersection avoidance, which requires a complex objective function and is computationally expensive. The methods in [26], [27] utilize level-set formulations that can take advantage of efficient time-implicit numerical schemes [29]. They are not topology-preserving [25]. Further, the local boundary-based formulation in [26] can be trapped in a local minimum that is arbitrarily far away from the global optimum. While the introduction of a weighted balloon-force term may alleviate this difficulty [27], it exposes the method to a "leaking" problem.

III. Problem Modeling

Let I be a given 3-D volumetric image of $n = X \times Y \times Z$ voxels, where $X, Y,$ and Z denote the image sizes in $\mathbf{x}, \mathbf{y}$, and $\mathbf{z}$ directions, respectively. The intensity level of every voxel (x, y, z) ($1 \le x \le X, 1 \le y \le Y,$ and $1 \le z \le Z$) is denoted by $I(x, y, z)$. We consider the desired region (target object) R that is bounded by two coupled terrain-like surfaces, S_l and S_u, and oriented as shown in Fig. 1. Each of the bounding surfaces intersects with exactly one voxel of every *column* parallel to the $\mathbf{z}$-axis. We look for an *optimal region* by minimizing the intraclass variance among all feasible regions that can be defined in the 3-D volumetric image I. Let μ_0 (respectively, μ_1) be the average intensity of the desired region R (respectively, the background $\bar{R} = I - R$), that is $\mu_0 = 1/|R|\Sigma_{(x,y,z)\in R} I(x, y, z)$, and $\mu_1 = 1/|\bar{R}| \Sigma_{(x,y,z)\in\bar{R}} I(x, y, z)$. The intraclass variance is

$$\mathscr{E}_{\mathrm{Intra}}(R) = \sum_{(x,y,z)\in R} (I(x, y, z) - \mu_0)^2 + \sum_{(x,y,z)\in\bar{R}} (I(x, y, z) - \mu_1)^2. \quad (2)$$

The feasibility of a region in I is constrained by two sets of application-specific parameters: 1) *surface smoothness parameters,* $\Delta_{\mathbf{x}}$ and $\Delta_{\mathbf{y}}$, and 2) *surface separation parameters,* δ^l and δ^u. The surface smoothness parameters guarantee the continuity of the bounding surfaces of R. More precisely, if (x, y, z) and $(x + 1, y, z')$ (respectively, $(x, y + 1, z')$) are two voxels on a feasible bounding surface, then $|z - z'| \le \Delta_{\mathbf{x}}$ (respectively, $|z - z'| \le \Delta_{\mathbf{y}}$). The surface separation parameters ensure that the two bounding surfaces, S_l and S_u, of the desired region R are at a certain distance range apart, that is, for every pair (x, y), $0 < \delta^l \le S_u(x, y) - S_l(x, y) \le \delta^u$, where $S(x, y)$ denotes the z-coordinate of the intersection voxel of the surface S with the column (x, y) of I. Comparing to the regularizing terms used in Chan and Vese's

method [1], our geometric constraints not only regulate the smoothness of the bounding surfaces, they also incorporate essential shape information: the guarantee of mono-tonicity and topological constraints.

IV. The Algorithm

Although minimizing the intraclass variance for general object shapes is computationally intractable, we are able to *optimally* detect the region bounded by two coupled terrain-like surfaces (or those regions that can be "unfolded" into two coupled terrain-like surfaces) in low-order polynomial time using the techniques of parametric search [30], hand probing [3], [4] in computational geometry, and 3-D graph-searching [10], [20], [31].

Let $\mu = 1/n \sum_{(x,y,z) \in I} I(x, y, z)$ be the average intensity of the entire image I. It is known that minimizing the intraclass variance $\mathcal{E}_{\text{Intra}}(R)$ is equivalent to the maximization of the following objective function [32]

$$\mathcal{E}_{\text{Inter}}(R)= |R|(\mu - \mu_0)^2 + |\overline{R}|(\mu - \mu_1)^2$$

which is actually the *interclass variance* of R and $\overline{R}$.

The equivalency of the two objective functions can be shown, as follows:

$$\mathcal{E}_{\text{Intra}}(R)= \sum_{(x,y,z)\in R} (I(x, y, z)^2 - 2\mu_0 I(x, y, z) + \mu_0^2) + \sum_{(x,y,z)\in \overline{R}} (I(x, y, z)^2 - 2\mu_1 I(x, y, z) + \mu_1^2)$$
$$= \sum_{(x,y,z)\in I} I(x, y, z)^2 - 2|R|\mu_0^2 + |R|\mu_0^2 - 2|\overline{R}|\mu_1^2 + |\overline{R}|\mu_1^2 \tag{3}$$
$$= \sum_{(x,y,z)\in I} I(x, y, z)^2 - |R|\mu_0^2 - |\overline{R}|\mu_1^2$$

and

$$-\mathcal{E}_{\text{Inter}}(R)= - n\mu^2 + 2\mu(|R|\mu_0 + |\overline{R}|\mu_1) - |R|\mu_0^2 - |\overline{R}|\mu_1^2$$
$$= \frac{\left(\sum_{(x,y,z)\in I} I(x,y,z)\right)^2}{n} - |R|\mu_0^2 - |\overline{R}|\mu_1^2. \tag{4}$$

Noticing that both $\sum_{(x,y,z)\in I} I(x, y, z)^2$ and $(\sum_{(x,y,z)\in I} I(x, y, z))^2/n$ are constants for a given image, the two objective functions differ by a constant, and thus minimizing $\mathcal{E}_{\text{Intra}}(R)$ is equivalent to maximizing $\mathcal{E}_{\text{Intra}}(R)$.

Note that the objective function $\mathcal{E}_{\text{Intra}}(R)$ is invariant if we replace $I(x, y, z)$ by $\tilde{I}(x, y, z) = I(x, y, z) - \mu$ for every voxel (x, y, z) in I. We thus, without loss of generality (WLOG), assume that $\mu = 0$ and, accordingly

$$\mathcal{E}_{\text{Inter}}(R)= |R|\left(\frac{U(R)}{|R|}\right)^2 + |\overline{R}|\left(\frac{-U(R)}{|\overline{R}|}\right)^2$$
$$= \frac{n}{|R|\cdot|\overline{R}|}(U(R))^2 \tag{5}$$

where $U(R) = \sum_{(x,y,z) \in R} I(x, y, z)$. Note that $U(R)$ could be negative. If $U(R) < 0$ for an optimal region R, then we can define a new image such that the intensity of each of its voxel (x, y, z) is $-I(x, y, z)$. It is not difficult to see that an optimal region in this new image is also

an optimal one in the original image. Hence, WLOG, we can assume $U(R)$ 0, and thus minimizing $\varepsilon_{\text{Intra}}(R)$ is equivalent to maximizing

$$\Psi(R) \equiv \frac{U(R)}{\sqrt{|R|(n - |R|)}} = \frac{\sum\limits_{(x,y,z) \in R} I(x, y, z)}{\sqrt{|R|(n - |R|)}}. \quad (6)$$

Let us further demonstrate how to find an optimal region R while maximizing $\Psi(R)$, where R is bounded by two coupled terrain-like surfaces.

A. Overview of the Algorithm

To maximize $\Psi(R)$, the following straightforward observation holds: for each $k = 0, 1, \ldots, n$, if an optimal region R_k^* of size k can be computed so that it maximizes the total sum of intensity of all voxels in the region [denoted by $U(R_k^*)$], the problem is solved. Unfortunately, finding a region maximizing the sum of intensity for a given size is not an easy task at all. However, viewing the problem in this way provides a basis for further exploitation of the intrinsic geometric structure of the problem.

For each $k = 0, 1, \ldots, n$, the pair $(k, U(R_k^*))$ defines a point on the 2-D plane, on which the x-axis represents the number of voxels of a desired region R and the y-axis represents $U(R)$, thus forming a set $\mathcal{P}$ of $n + 1$ points. A key observation here is that it may not be necessary to compute all points in $\mathcal{P}$. A classical concept in computational geometry [33], called *convex hulls,* plays an important role. The convex hull $CH(\mathcal{P})$ of a point set $\mathcal{P}$ is the unique convex polygon which contains $\mathcal{P}$ and all vertices of which are points from $\mathcal{P}$. Using similar techniques in [5], [6], we can prove the following lemma.

Lemma 1—The point $(|R^*|, U(R^*))$ defined by an optimal region R^* in I (i.e., $\Psi(R^*) = \max_R \Psi(R)$), must be a vertex of the convex hull $CH(\mathcal{P})$.

<u>Proof:</u> Let $\alpha^* = \Psi(R^*) = U(R^*) / \sqrt{|R^*|(n - |R^*|)}$. Consider the curve ξ: $y = \alpha^* \sqrt{x(n - x)}$ in the 2-D plane (see Fig. 2). Since $U(R^*) = \alpha^* \sqrt{|R^*|(n - |R^*|)}$, the point $(|R^*|, U(R^*))$ is on the curve ξ. Notice that $\alpha^* = \max_R \{U(R) / \sqrt{|R|(n - |R|)}\}$. Thus, for any region R bounded by two coupled terrain-like surfaces, we have $U(R) \leq \alpha^* \sqrt{|R|(n - |R|)}$, i.e., every point $(k, U(R_k^*)) \in \mathcal{P}$ $(k = 0, 1, \ldots, n)$ lies below or on the curve ξ. Furthermore, due to the concavity of the curve ξ: $y = \alpha^* \sqrt{x(n - x)}$, all points in $\mathcal{P}$ lie below or on the tangent line l to ξ at the point $(|R^*|, U(R^*))$. Hence, $(|R^*|, U(R^*))$ is a vertex of the upper chain of the convex hull $CH(\mathcal{P})$ of $\mathcal{P}$.

Thus, finding the optimum can be simplified to examining all convex hull vertices. Inspired by the *shape probing* method [3], [4] which can be viewed as recognizing a convex polygon by "touching with lines," we use the following *probing oracle* to construct $CH(\mathcal{P})$ when the coordinates of the points in $\mathcal{P}$ are unknown. The probing oracle is:

Given a slope θ, report the tangent line with slope θ to $CH(\mathcal{P})$ and the tangent point.

Using this probing oracle, the convex hull $CH(\mathcal{P})$ can be constructed as follows. Start with slopes $+\infty$ and $-\infty$ to find the two endpoints of $\mathcal{P}$ (left-most and right-most points, which are always $(0, 0)$ and $(n, 0)$ in this problem). Note that the convex hull $CH(\mathcal{P})$ is always an upper portion (or upper convex chain). Now suppose that two vertices u and v were computed on the hull and that so far there is no vertex of $CH(\mathcal{P})$ between u and v. Let θ be

the slope of the line through u and v. Then, employ a probing oracle with respect to θ (see Fig. 3). Consequently, we either find a new vertex on $CH(\mathcal{P})$ between u and v or conclude that uv is an edge of $CH(\mathcal{P})$. Thus, employing a probing oracle results in either a new vertex or a new edge of $CH(\mathcal{P})$. Hence, the convex hull $CH(\mathcal{P})$ with k vertices can be computed in $O(k)$ probing oracle application steps.

A major challenge is to implement this oracle for a given slope θ. The parametric search [30] in computational geometry is utilized. For a given real-valued parameter θ, we define a *parametric image* I_θ with $I_\theta(x, y, z) = I(x, y, z) - \theta$ for every (x, y, z) tuple. The *parametric intensity sum* of a region R, denoted by $U_\theta(R)$, equals to the total sum of the intensities of all voxels in R minus $\theta|R|$ (i.e., $U_\theta(R) = U(R) - \theta|R|$). We show in Lemma 2 that, the tangent point of the tangent line with slope θ to $CH(\mathcal{P})$ is defined by the optimal feasible region with a maximized intensity sum in the parametric image I_θ. This last step of finding the optimal region in I_θ can be solved using the graph-based segmentation method [10].

The key shape probing procedure can therefore be summarized as follows.

ShapeProbe (I, k_{left}, U_{left}, k_{right}, U_{right})

1. $\theta \leftarrow (U_{\text{right}} - U_{\text{left}})/(k_{\text{right}} - k_{\text{left}})$

2. Find the region $R^*(\theta)$ such that $U_\theta(R^*(\theta)) = max_R U_\theta(R)$.

3. **if** $|R^*(\theta)|$ k_{left} **then do**

4. ShapeProbe (I, k_{left}, U_{left}, $|R^*(\theta)|$, $U(R^*(\theta))$)

5. ShapeProbe (I, $|R^*(\theta)$, $U(R^*(\theta))$, k_{right}, U_{right})

The input to the subroutine ShapeProbe includes two known hull vertices $u(k_{\text{left}}, U_{\text{left}})$ and $v(k_{\text{right}}, U_{\text{right}})$, with no known hull vertices in between so far. The subroutine finds all hull vertices of $CH(\mathcal{P})$ between u and v. Line 1 calculates the slope θ of the probing line. Line 2 computes an optimal region $R^*(\theta)$ whose parametric intensity sum is maximized to implement the probing oracle. Then, if a new hull vertex ($|R^*(\theta)|$, $U(R^*(\theta))$) between u and v is found, as shown in Line 3, the procedure recursively computes all hull vertices on both left and right convex hull chains divided by ($|R^*(\theta)|$, $U(R^*(\theta))$). Thus, the convex hull $CH(\mathcal{P})$ is computed. Based on Lemma 1, we can examine every vertex of $CH(\mathcal{P})$ to find the optimum. We next show the efficient implementation of the probing oracle.

B. Implementation of the Probing Oracle

Given a real-valued parameter θ, we show in this section that the probing oracle can be implemented by computing in I an optimal region $R^*(\theta)$ whose parametric intensity sum $U_\theta(R^*(\theta))$ is maximized. We call $R^*(\theta)$ an *optimal parametric region* associated with the parameter θ. Recall that R_k^* denotes an optimal region in I whose total intensity sum $U(R_k^*)$ is maximized and the size of the region is exactly k.

Lemma 2—There exists a tangent line to $CH(\mathcal{P})$ at the point $(\bar{k}, U(R_{\bar{k}}^*))$ with a slope θ if and only if $|R^*(\theta)| = \bar{k}$ and $U(R^*(\theta))=U(R_{\bar{k}}^*)$.

Proof: The proof of this lemma is similar to that in [5], [6].

"$\Rightarrow$" Suppose that l: $y = \theta x + b$ is a tangent line to $CH(\mathcal{P})$ at the point $(\bar{k}, U(R_{\bar{k}}^*))$. This implies that $b=U(R_{\bar{k}}^*) - \bar{k}\theta=U(R_{\bar{k}}^*) - |R_{\bar{k}}^*|\theta$. Note that $CH(\mathcal{P})$ is actually the upper chain of the convex hull. Thus, for any k $\bar{k}$, the point $(k, k\theta + b)$ on l is on or above $CH(\mathcal{P})$.

Alternatively, $k\theta + b \geq U(R_k^*)$, that is $U(R_{\bar{k}}^*) - \bar{k}\theta \geq U(R_k^*) - k\theta$ for any k k (see Fig. 4). Hence, the region $R_{\bar{k}}^*$ achieves $\max_k\{U(R_k^*) - k\theta\}$. Note that $U_\theta(R^*(\theta)) = \max_{R(\theta)}\{U(R(\theta)) - |R(\theta)| \cdot \theta\} = \max_k\{U(R_k^*) - k\theta\}$. Thus, $U(R^*(\theta)) = U(R_{\bar{k}}^*)$.

"⇐" The fact that $|R^*(\theta)| = \bar{k}$ indicates that, for any feasible region $R(\theta)$ bounded by two coupled terrain-like surfaces, if $|R(\theta)|$ $\bar{k}$, then $U(R(\theta)) - |R(\theta)| \cdot \theta$ $U_\theta(R^*(\theta))$. Thus, for any k $\bar{k}$, $U(R_k^*) - k\theta \leq U_\theta(R^*(\theta))$. Based on the assumption that $U(R^*(\theta)) = U(R_{\bar{k}}^*)$ and $|R^*(\theta)| = \bar{k}$, we have $U(R_k^*) - k\theta \leq U(R_{\bar{k}}^*) - \bar{k}\theta$ for any k $\bar{k}$.Consider the line $l: y - \theta x = b$ with $b = U(R_{\bar{k}}^*) - \bar{k}\theta$. Obviously, the point $(\bar{k}, U(R_{\bar{k}}^*))$ is on Line l and any other point $(k, U(R_k^*))$ with k $\bar{k}$ is on or below Line l (see Fig. 4). Hence, Line l is a tangent line to $CH(\mathcal{P})$ at the point $(k, U(R_{\bar{k}}^*))$ with a slope θ.

This proves Lemma 2.

Consequently, for a given slope θ, we need to compute an optimal parametric region $R^*(\theta)$ bounded by two coupled terrain-like surfaces in I. If the size of $R^*(\theta)$ is $\bar{k}$, based on Lemma 2, the line $l: y = \theta x + (U(R^*(\theta)) - \bar{k} \cdot \theta)$ is a tangent line to $CH(\mathcal{P})$ at the point $(\bar{k}, U(R^*(\theta)))$ with slope θ. Let $R_{\bar{k}}^* = R^*(\theta)$. We thus recognize a hull vertex on $CH(\mathcal{P})$. Next, we develop an efficient algorithm for computing such an optimal parametric region $R^*(\theta)$ in I.

C. Computing an Optimal Parametric Region

Given a parameter θ, we reduce the problem of computing an optimal parametric region $R^*(\theta)$ in I to the problem of finding two coupled terrain-like 3-D surfaces on the transformed images while minimizing the total sum of the cost on both surfaces. The detection of two coupled terrain-like surfaces can be formulated as a surface segmentation problem [10].

First, we perform the following transformations on the image I:

$$I_\theta'(x, y, z) = \begin{cases} 0, & \text{if } z = 0 \\ \sum_{0 \leq z' < z}\{I(x, y, z') - \theta\}, & \text{otherwise} \end{cases} \quad (7)$$

and

$$I_\theta''(x, y, z) = \sum_{0 \leq z' \leq z} -\{I(x, y, z') - \theta\}. \quad (8)$$

Hence, for any feasible region $R(\theta)$ bounded by two coupled terrain-like surfaces, S_l and S_u with S_u above S_l, we have

$$\sum_{(x,y,z) \in S_l} I_\theta'(x, y, z) + \sum_{(x,y,z) \in S_u} I_\theta''(x, y, z) = -U_\theta(R(\theta)). \quad (9)$$

Note that both bounding surfaces S_l and S_u satisfy the smoothness constraint and the surface separation constraint.

In this way, we convert the optimal parametric region problem to a surface segmentation problem, which can be solved optimally using Li *et al.*'s method [10].

The basic idea of Li *et al.*'s surface segmentation method is to transform the problem into computing a minimum *s-t* cut in a derived arc-weighted directed graph with $O(n)$ nodes and $O(n)$ arcs. For a directed graph with $O(n')$ nodes and $O(m')$ arcs, a minimum *s-t* cut can be computed in $O(m'n' \log n'^2/m')$ time using Goldberg and Tarjan's algorithm [34].

Lemma 3—For a given θ, an optimal parametric region $R^*(\theta)$ in *I* can be computed in $O(n^2 \log n)$ time.

In summary, it suffices to compute the convex hull *CH*($\mathcal{P}$) to detect in *I* an optimal region while minimizing the intraclass variance by Lemma 1. Based on Lemma 2, we can perform $O(n)$ probing oracles to obtain all vertices on *CH*($\mathcal{P}$). Each probing oracle can be implemented in $O(n^2 \log n)$ time by Lemma 3. Thus, the total running time of our algorithm for minimizing the intraclass variance is $O(n^3 \log n)$. However, in our experimentation, the number of the employed probing oracles was much less than *n*.

V. Approximating Minimum Intraclass Variance

Although our exact algorithm for minimizing the intraclass variance in Section IV is efficient for the moderate size of input images, the large image size may prevent the method from being computationally practical. In this section, we develop an approximation method that improves the running time while still producing close-to-optimal solutions.

Our method is based on the property of interclass variance explored by Asano *et al.* [5]. They utilized the property to obtain a $(1 - \varepsilon)$-approximation algorithm for computing an optimal connected rectilinear region with maximum interclass variance in 2-D. We characterize a similar property for a region bounded by two coupled terrain-like surface in 3-D, as stated in Lemma 4. The proof of the lemma is similar to that in [5].

Lemma 4

Let θ_O denote the optimal parameter value with which the optimal parametric region $R^*(\theta_O)$ maximizes $\Psi(R)$. If $\theta_O \ge 0$, then an optimal parametric region $R^*((1 + \varepsilon)\theta_O)$ gives an $(1 - \varepsilon)$-approximate solution to the problem of maximizing $\Psi(R)$, that is, for any $0 < \varepsilon < 1$, $\Psi(R^*((1 + \varepsilon)\theta_O)) \ge (1 - \varepsilon)\Psi(R^*(\theta_O))$.

We assume that the cost $I(x, y, z)$ of each voxel in the input image *I* is an integer, and $\mathcal{L}$ is the total sum of the absolute values of the voxel costs in *I*. Due to the integrality of $I(x, y, z)$, it is not difficult to see that $1/n \le |\theta_O| \le \mathcal{L}$. Our idea is to partition the θ-space $[-\mathcal{L} \cdots - 1/n] \cup [1/n \cdots \mathcal{L}]$ into intervals $[\theta_i \cdots \theta_{i+1}]$ such that either $\theta_i/\theta_{i+1} = 1 + \varepsilon$ or $\theta_{i+1}/\theta_i = 1 + \varepsilon$. Note that Lemma 4 actually indicates that for any θ in between $(1 + \varepsilon) \theta_O$ and θ_O, an optimal parametric region $R^*(\theta_O)$ gives an $(1 - \varepsilon)$-approximate solution to the problem of maximizing $\Psi(R)$. Thus, if $\theta_O \in [\theta_i \cdots \theta_{i+1}]$, then either $R^*(\theta_O{}_i)$ or $R^*(\theta_O{}_{i+1})$ is an $(1 - \varepsilon)$-approximate solution. Hence, we partition the θ-space into the following intervals:

$$\left(\cup_{i=0}^{\lfloor \log_{1+\varepsilon} n\mathcal{L} \rfloor} [-1/n(1+\varepsilon)^{i+1} \cdots -1/n(1+\varepsilon)^i] \right) \cup \left(\cup_{i=0}^{\lfloor \log_{1+\varepsilon} n\mathcal{L} \rfloor} [1/n(1+\varepsilon)^i \cdots 1/n(1+\varepsilon)^{i+1}] \right).$$

Considering the sequence $\{-\theta_i, \theta_i\}$ of parameters for θ, with $\theta_i = 1/n(1 + \varepsilon)^i$ ($i = 0, 1, \ldots, \lfloor \log_{1+\varepsilon} n \mathcal{L} \rfloor$), we compute an optimal parametric region $R^*(\theta_i)$ for each θ_i. Among those optimal parametric regions, the one that maximizes $\Psi(R)$ is chosen as the approximation solution. It is clear that such a solution is an $(1 - \varepsilon)$-approximation solution. The number of parameters that we search is $O(\log_{1+\varepsilon} n \mathcal{L}) = O(\varepsilon^{-1} \log n \mathcal{L})$. Note that an optimal parametric region $R^*(\theta)$ for a given parameter θ can be computed in $n^2 \log n$ time using

Goldberg and Tarjan's minimum *s-t* cut algorithm. Thus, the running time of this approximation algorithm is $O(\log n \, \mathcal{L}/\varepsilon n^2 \log n)$.

Lemma 5

A $(1 - \varepsilon)$ -approximation solution to the problem of maximizing $\Psi(R)$ can be computed in $O(\log n \, \mathcal{L}/\varepsilon n^2 \log n)$ time.

We then output the $(1 - \varepsilon)$ -approximation solution to the problem of maximizing $\Psi(R)$ as our approximate solution to the problem of minimizing the intraclass variance $\varepsilon_{\text{Intra}}(R)$. Although we have not yet proven a tight bound of the approximation ratio of this approximation algorithm for minimizing the intraclass variance, our implementation demonstrated its high segmentation accuracy with much less execution times for all data we tested for our exact algorithm.

VI. Experimental Methods

The experiments were performed on computer phantoms, physical phantoms and 3-D CT and MR medical images. Both regions bounded by terrain-like surfaces and by tubular surfaces are used.

A. Data

1) Computer Phantoms—To validate the correctness of the modeling techniques, we tested our method on a set of computer-generated phantoms, with sizes of $256 \times 256 \times 3$ voxels, containing differently textured regions or shapes (Fig. 5). For the evaluation of the execution times, a second set of computer phantoms was produced that contained a region bounded by two coupled terrain-like surfaces with various shapes and mutual positions. Two sets of different patterns were used to differentiate the target region from the background. The sizes of the phantom images range from $30 \times 30 \times 40$ to $100 \times 100 \times 50$. To test the performance of our algorithm in noisy images, we added Gaussian noise of $\sigma = 0.5, 2.0, 3.0$, respectively, to this set of computer phantoms. In addition, we generated a third set of 160 computer phantoms, each sized $100 \times 100 \times 10$, to evaluate the performance of the approximation algorithm. Each phantom image includes a region bounded by two coupled terrain-like surfaces with specific positions, distances, and smoothness randomly generated.

2) Physical Phantoms—To quantitatively assess the performance of our method, a physical phantom was imaged by multidetector CT and analyzed. The phantom contained six plexi-glas tubes, numbered 1–6, with nominal inner diameters of 1.98, 3.25, 6.40, 6.50, 9.50, and 19.25 mm, respectively. The corresponding outer diameters were 4.45, 6.30, 9.70, 12.60, 15.60, and 25.50 mm, respectively. The phantom was scanned using Philips Mx8000 4-slice CT scanner with three different scan settings (low dose, regular dose, and high dose). Under each setting, the scans were taken at four distinct angles of $0°, 5°, 30°$, and $90°$, rotated in the coronal plane, resulting in a total of 12 datasets for use in the validation. The regular dose scanning was intentionally repeated, yielding additional four datasets used for initial calibration of the cost functions. In all cases, a resolution of $0.39 \times 0.39 \times 0.6$ mm^3 was used, images consisted of 200–250 slices, 512×512 pixels each.

3) CT Images of Pulmonary Airways—To demonstrate the utility of our method in quantitative analysis of medical images, the method was applied to human pulmonary CT images to concurrently segment the inner and outer wall surfaces of intrathoracic airways imaged by the multidetector CT. Twenty airway wall segments extracted from 12 *in vivo* CT scans of six human objects were used for the experiments. The airway wall segments had a resolution of $0.7 \times 0.7 \times 0.6$ mm^3 and consisted of 30–50 slices, 50×50 pixels each.

The inner and outer surfaces needed to be unfolded for our method to be employed. The centerlines of the airways were manually identified for this unfolding process, which utilized *B*-spline interpolation. Segmentation was performed only on the trunks of the airway tree between bifurcations since our method can not handle topological changes at this point. The inner and outer walls were segmented simultaneously.

4) MR Images of Cardiac Ventricular Walls—To show the applicability of the method to a broader range of medical image segmentation applications, we have utilized our method for segmenting MRI images of left human cardiac ventricles. 20 MRI scans of normal human hearts and 10 MRI scans of patient hearts were segmented. The images had an in-plane resolution of approximately 2.0×2.0 mm^2 but a large slice thickness of about 6–8 mm. To get better image quality we performed image fusion on the raw datasets as proposed in [35]. After the image fusion, the images had a resolution of approximately $2.0 \times 2.0 \times 2.0$ mm. Each dataset consisted of 30–40 mid-papillary slices and each slice had an average size of 40×40 pixels. To facilitate the unfolding process of the tubular shape of ventricular walls, the centerlines (long axes) of the cardiac ventricle were manually determined.

5) MR Images of Cardiac Ventricular Walls From Miccai 2009 Cardiac MR Left Ventricle Segmentation Challenge—To compare the performance of our method with other contemporary segmentation methods, our method was applied to MRI images of the left human cardiac ventricle from the Cardiac MR Left Ventricle Segmentation Challenge images (MICCAI 2009 Workshop). 15 MRI scans of preworkshop cardiac left ventricle datasets were segmented. Each dataset consisted of around 6–12 slices per phase and 20 phases over the heart cycle and each slice had an average size of 256×256 pixels. The centerlines (long axes) of the cardiac ventricle were manually determined and only a 101×101-pixel region surrounding the centerline were used in the segmentation. Refer to [36] for details.

B. Cost Functions

Cost function design is very important in graph-based segmentation. Since our method minimizes the intraclass variance, a cost function reflecting the homogeneity was used. For the experiments, image intensity or its linear transformation was utilized as voxel-related cost. For the texture-related phantom images, the cost function also included orientation and/ or curvature information [1]. For the clinical data, our algorithm was run on the original images to get the estimated positions of the bounding surfaces and then this estimated position was combined with the voxel intensities to form the cost function.

C. Execution Time

All experiments were conducted on a Pentium-D 2.4 GHz PC with 3.5 GB of memory. The execution time included the graph initialization time and the actual computation time. The execution times varied across the test cases even with the same image size. We thus averaged the running times over 20 test cases for each image size. The execution time for each test case was measured three times and the results were averaged. In our implementation, Boykov and Kolmogorov's algorithm [37] was used to compute the minimum *s-t* cut.

D. Performance Indices

Besides visual inspection, which was carried out on computer phantom images, we used the following quantitative indices to measure the performance of our method.

1) Diameter Calculation for Physical Phantom—For the physical phantoms, the average diameter of the segmentation result of each slice was calculated using the circle fitting method to minimize the mean-squared error, and then the diameter was compared with the known tube diameter. The diameter errors are reported as mean ± standard deviation in absolute measurements and as percentages of the diameter.

2) Positioning Error Measure for Clinical Test Cases—Surface detection accuracy was determined in clinical test cases in comparison with the independent standard. The mean surface positioning errors were computed. Corresponding points were defined as pairs of points, the first point being from a computer detected border and the second point from the reference standard border that was closest to each other using the Euclidean distance metric. The positioning errors were calculated both signed and unsigned, where signed errors were calculated by adding a sign to each pair-wise unsigned errors, and the sign was positive while the point from a computer detected border was outside the reference standard contour, and negative otherwise. These errors were reported as mean ± standard deviation.

3) Dice Similarity Coefficient for Left Ventricle Segmentation of the Datasets From Miccai 2009 Challenge—The Dice similarity coefficient is used for the verification of left ventricular segmentation and measures the similarity between the automated segmentation volume V_A and the expert-traced volume V_{GT} with $s(V_A, V_{GT}) = 2\|V_A \cap V_{GT}\|/(\|V_A\| + \|V_{GT}\|)$, where $\|X\|$ denotes the cardinality of X.

4) Optimizing Intraclass Variance With Iterations—Since our reported approach and the Chan–Vese original method [1] both optimize the intraclass variance, it is of interest to compare their performance. Considering that the regularization terms and constraints are different, a fair comparison is very difficult. We thus compared our approach with a modified Chan–Vese method, which keeps the iteration-based principle of the Chan–Vese method to optimize the intraclass variance while incorporating the geometric constraints of our method. We refer to this approach as to an *iterative intraclass variance* (briefly, an *iterative ICV*) method.

The basic idea of the Chan–Vese method [1] is to first find an initial region and calculate the average intensities of the initial region and the background as μ_0^e and μ_1^e. These two parameters are used as constants in the next iteration to find an optimal region optimizing the objective function

$$\mathcal{E}_{\text{ChanVese}}(R) = \sum_{(x,y,z) \in R} (I(x,y,z) - \mu_0^e)^2 + \sum_{(x,y,z) \in \overline{R}} (I(x,y,z) - \mu_1^e)^2 + T_{\text{regularization}} \quad (10)$$

where $T_{\text{regularization}}$ is a regularization term. The new average intensities μ_0^e and μ_1^e of the newly computed region and the background are calculated. This process iterates until both μ_0^e and μ_1^e converge.

The iterative ICV method "simulates" the Chan–Vese method to find an optimal region bounded by two coupled terrain-like surfaces while minimizing the intraclass variance of the region. With given average intensities μ_0^e and μ_1^e of the initial region and the background, the graph searching method of Li *et al.* [10] was employed to find the desired region with the same geometric constraints as used in the presented method. The ICV method shares the same geometric constraints and objective function with our method. Comparison between these two methods can demonstrate the improvement from global optimality.

We performed the following transformations on image I

$$I^{'}(x, y, z) = \sum_{0 \leq z^{'} < z} (I(x, y, z^{'}) - \mu_1^e)^2 + \sum_{z \leq z^{'} < Z} (I(x, y, z^{'}) - \mu_0^e)^2. \quad (11)$$

and

$$I^{''}(x, y, z) = - \sum_{z < z^{'} < Z} (I(x, y, z^{'}) - \mu_0^e)^2 + \sum_{z < z^{'} < Z} (I(x, y, z^{'}) - \mu_1^e)^2. \quad (12)$$

Hence, for any feasible region R bounded by two coupled terrain-like surfaces, S_l and S_u with S_u above S_l, we have

$$\sum_{(x,y,z) \in S_l} I^{'}(x, y, z) + \sum_{(x,y,z) \in S_u} I^{''}(x, y, z)$$
$$= \sum_{(x,y,z) \in R} (I(x, y, z) - \mu_0^e)^2 + \sum_{(x,y,z) \in \overline{R}} (I(x, y, z) - \mu_1^e)^2. \quad (13)$$

Note that both bounding surfaces S_l and S_u satisfy the surface smoothness and separation constraints. Using Li *et al.*'s method [10], the optimal surfaces S_l^* and S_u^* are simultaneously detected in $I^{'}$ and $I^{''}$, respectively. From (13), the region R^* bounded by S_l^* and S_u^* is the desired region in I. This process iterates until both μ_0^e and μ_1^e converge, as in the Chan–Vese method.

5) Paired T-Test—Our method is compared with Chan–Vese method and iterative ICV method. In these cases, paired t-test is performed to show that our method performs better. When doing experiments on left ventricle segmentation of the datasets from MICCAI 2009 challenge, our method is compared with two other methods [38], [39]. Again paired t-test is performed but in this case to show our method can achieve comparable results.

VII. Results

A. Computer Phantoms

The general capability of our approach is demonstrated in Fig. 5, in which the object and background differ only in texture with no obvious edge-based inter-object boundary information. The regional properties (texture pattern description) were represented in the cost function using curvature and edge orientation. As can be seen from comparison of the middle and right panels of Fig. 5, our method and the true Chan–Vese method with the same cost images yielded very similar results.

B. Accuracy Assessment in CT-Imaged Physical Phantoms

By comparing achieved segmentation results with the known truth about phantom tube diameters, the signed errors of the inner and outer wall diameter were 0.03 ± 0.07 mm and 0.04 ± 0.09 mm, respectively, with the voxel sizes of $0.39 \times 0.39 \times 0.60$ mm^3. The corresponding unsigned errors were 0.06 ± 0.06 mm and 0.08 ± 0.07 mm. The signed percent errors of the computer-determined diameters are presented in Fig. 6, where mean errors $\pm$ standard deviations are shown as a function of the true phantom tube diameter.

Using the Chan–Vese method (using the same cost image as used in our method), we obtained the segmentation results with signed errors of 0.31 ± 0.09 mm and 0.10 ± 0.06 mm,

and unsigned errors of 0.41 ± 0.09 mm and 0.36 ± 0.06 mm for the inner- and outer-diameters, respectively. Comparing with the Chan–Vese method, our approach was significantly more accurate ($p < 0.0048$ in all cases). Fig. 7 shows example segmentation results using the Chan–Vese method and our approach obtained in the 3-D CT-imaged physical phantom.

C. Airway Wall Segmentation—Human in vivo 3-D CT

While the inner airway wall surfaces are well visible in *in vivo* CT images, the outer airway wall surfaces are very difficult to segment due to their blurred and discontinuous appearance. The results showed subvoxel accuracy and 3-D consistency of our method (see Figs. 8 and 9). Compared with the available manual tracings, our method yielded signed border positioning errors of 0.42 ± 0.38 and −0.13 ± 0.46 voxel for the inner and outer wall surfaces, respectively. The corresponding unsigned errors were 0.66 ± 0.15 and 0.57 ± 0.03 voxel, respectively. We attempted to perform the Chan–Vese method on these data, but without enforcing the separation constraints, the Chan–Vese method failed to produce meaningful segmentations on these data sets.

D. Cardiac Ventricular Wall Segmentation—Human in vivo 3-D MR

Our 3-D method for cardiac ventricular wall segmentation demonstrated low surface positioning errors as well as robust performance when compared with the expert-traced results, although the simultaneous inner- and outer-wall segmentation is challenging due to image blurriness and discontinuity of the boundaries in the cardiac ventricular images. Detailed results are given in Table I(a).

E. Cardiac Ventricular Wall Segmentation Using DatasetsFrom Miccai 2009 Challenge

Fig. 10 shows examples of our segmentation compared with expert-traced results. Tables II and III show comparisons of average positioning errors and Dice similarity coefficients between our method and the top two methods reported in the segmentation contest of the MICCAI 2009 Challenge [38], [39]. The p-values between our method and the method in [38] are 0.373 and 0.922 for endo- and epi-cardial walls. The p-values between our method and the method in [39] are 0.750 and 0.133 for endo- and epi-cardial walls. All show that our method can produce comparable results with the two methods.

F. Comparison With Iterative ICV Method

Results presented in this section show that our method outperformed the iterative ICV method (Section VI-D-IV) when detecting a region bounded by two coupled terrain-like surfaces. The experiments were conducted on both phantoms and 3-D MR cardiac ventricular datasets.

Fig. 11 shows an example of the segmentation results produced by the iterative ICV method and by our method.

For the 3-D MR left ventricular datasets, our method outperformed the iterative ICV method by a factor of about 9% on average in terms of the surface positioning errors. For the fairness of the comparison, the cost functions and geometric constraints used for the iterative ICV method were the same as those used in our method.

Table I(b) shows the signed and unsigned positioning errors of results obtained by the iterative ICV method.

G. Execution Time

The execution time of our method mainly depends on the number of shape probings performed (i.e., the number of hull vertices of the constructed convex hull) and the time needed for employing each probing oracle. The average execution times for employing each probing oracle are shown in Table IV. The minimum *s-t* cut was computed using Boykov and Kolmogorov's maximum flow algorithm [37] with a "forward-star" graph representation.

For the airway wall segmentation, the numbers of the hull vertices of *CH*($\mathcal{P}$) ranged from 1.50% to 2.86% of the image size; while for the 3-D ventricular datasets, the percentages ranged between 1.64% and 2.89%. The average total execution times of our algorithm are shown in Fig. 12.

For the computer phantoms with no Gaussian noise added ($\sigma = 0.0$), the number of hull vertices of *CH*($\mathcal{P}$) was ranging from 1.32% to 1.61% of the image size. Our experiments showed that the added Gaussian noise affected the performance of our algorithm substantially. The average execution times on the computer phantoms with different levels of Gaussian noise are shown in Fig. 13.

H. Performance of the Approximation Method

The performance of our developed approximation method was tested on computer phantoms, physical phantoms, 3-D CT images of pulmonary airway walls, and 3-D MR images of cardiac ventricular walls, and was compared to that of our exact algorithm. Our experiments demonstrated that the approximation method ran much faster than the exact algorithm while the segmentation accuracy was highly comparable to that of the exact algorithm.

We chose $\varepsilon = 0.01, 0.02, 0.05, 0.1, 0.2, 0.5$ to test the approximation method with respect to segmentation accuracy. In addition to the surface positioning errors, the so-called *approx-opt ratio*, which is the ratio of the objective function value produced by the approximation algorithm over that yielded by the exact algorithm, was also used to measure the segmentation performance. Interestingly, while $\varepsilon \leq 0.05$, the approximation method always computed the optimal solutions to the problem of minimizing intraclass variance $\varepsilon_{\text{Intra}}(R)$ for all the datasets we tested. For the six physical phantoms, even while we took $\varepsilon = 0.1, 0.2, 0.5$, we were still able to obtain the optimal solutions by using the approximation method. Table V presents, for the different ε values, the average approx-opt ratios, the maximum approx-opt ratios, and the percentage of the test cases that the approximation algorithm did not produce the optimal solution. Tables VI and VII show the average approx-opt ratio and the average surface positioning errors with different ε values for the airway wall images and the ventricular wall images, respectively. Note that for $\varepsilon = 0.05$ in both tables, the objective function values were optimal and the average surface positioning errors were the same as those obtained by the exact algorithm. Thus, our experiments showed that the approximation method was able to produce highly comparable segmentation results as those produced by the exact algorithm even for relatively large ε values.

The running times of the approximation method were evaluated on computer phantoms and clinical data. Our experiments demonstrated great improvement of running time over the exact algorithm. Fig. 14 shows the comparison of the execution times of the approximation algorithm with different ε's on the computer phantom images against the exact algorithm. We did not explicitly add Gaussian noise to those phantoms. Our further experiments on those phantom images with added Gaussian noise ($\sigma = 0.5, 2.0, 3.0$) revealed that the efficiency of the approximation method did not decrease with the presence of noise. In contrast, the performance of the exact algorithm on running time deteriorated significantly

for noisy images. The dramatic improvement on the execution times of the approximation algorithm over the exact algorithm are shown in Figs. 15 and 16, respectively, for 3-D airway CT images and 3-D ventricular MR images, which show less intensity homogeneity.

In addition, we conducted experiments on the 20 MR images of cardiac ventricular walls of normal human hearts to compare the iterative ICV method against the approximation algorithm. Table VIII shows the surface positioning errors of the iterative ICV method and the approximation algorithm as well as those of the exact algorithm. The results demonstrated that the approximation algorithm outperformed the iterative ICV method, even with a large approximation error parameter $e = 0.5$.

VIII. Discussion

In the following discussion, we focus on several important issues potentially influencing the utility of the presented method as well as outline their possible solutions.

A. Global Optimality

One of the accomplishment we achieved is the global optimality of the method. In fact, our method produces a more "optimal" solution than the iterative ICV method (Section VI-D-IV does. Although the iterative ICV method has always converged, it frequently could not find the globally optimal solution with respect to the intraclass variance objective function.

For the physical phantom images tested, our method outperformed the iterative ICV method by a factor of 0.23% in terms of the intraclass variance (i.e., the value of the objective function) obtained with p-value of 0.0094. For the test case in Fig. 11, our method gave a segmentation with an intraclass variance of 9.6071×10^6, compared with 9.5909×10^6 achieved by the iterative ICV method. Although these are not big improvement, it demonstrates that iterative ICV method does not guarantee global optimality. The global optimality of our method is theoretically proved.

B. Limitations

One obvious limitation of the presented method is that it only allows optimal detection of a single object (region). This may prevent the method from being used for a wider range of medical applications, in which multiple mutually interacting objects are present. The extension of our method for dealing with simultaneous multiple object segmentation is an interesting and challenging future work. The second limitation is that the method can only detect those surfaces that can be unfolded to be terrain-like, including cylindrical or tubular surfaces. Unfolding techniques have been developed for objects of relatively complex shapes, such as liver [40], knee bone and cartilage [41], [42], heart [43], and pulmonary airway and vascular trees [44]. However, one may experience difficulty to unfold more complex objects into terrain-like surfaces. Thus, another challenge is to make the method topologically adaptive.

The approximation algorithm was demonstrated as having much faster running times while still producing highly comparable segmentation results as the exact algorithm. Unfortunately, we have not yet proven a tight bound of the approximation ratio, although the experiments indicated that the bound was pretty small in all tested cases. We plan to further study the approximation ratio of our approximation algorithm.

C. Improving Efficiency

We view the first contribution of this work as a theoretical advance for optimizing the intraclass variance of a target object (which is the data-driven term of the Chan–Vese cost

funtional) bounded by two terrain-like surfaces. To the best of our knowledge, this is the first *polynominal time combinatorial algorithm* solving this problem that guarantees globally optimal solutions.

The current implementation (on a Pentium-D 2.4 GHz PC with 3.5 GB of memory) of the exact algorithm though requires a relatively long computing time for large-size datasets. However, achieving polynominal-time solvability is practically important. It indicates that with the advance of the computing technology, the *execution time* required for solving the problem will be substantially shortened. For instance, with the application of the multicore computing technology, we believe that the execution time of our algorithm for minimizing the intraclass variance can be substantially reduced—our algorithm is highly parallelizable. The algorithm recursively calls the probing oracle to recognize a vertex on the convex hull of a set of $O(n)$ unknown points on the 2-D plane. Each probing oracle can be implemented by computing a minimum *s-t* cut in a derived graph. In the worst case, it may take $O(n)$ runs of the minimum *s-t* cut algorithm. Those $O(n)$ runs form the nodes of the recursion tree with each node requiring a minimum *s-t* cut solution. The height of such a tree is $O(log\ n)$. It is not difficult to observe that all the nodes on the same level of the recursion tree can attempt to solve the minimum *s-t* cut problem in parallel. Theoretically, if we have $O(n)$ processing units, the running time of our algorithm is $O(log\ n)$ runs of the minimum *s-t* cut algorithm, substantially reducing the running time of our algorithm that is now running in sequence. In practice, we can take advantage of the emerging GPU-based technology, to make sure that all the nodes on the same level of the recursion tree run in parallel as much as possible, thus significantly improving performance of our algorithm.

D. Incorporating Edge Information

Using only region information for segmentation may not fully utilize the available image information. Image edges defined by image gradients, even rather weak gradients, may be very helpful for detecting the object boundary. In fact, combining both image edges and image regional information is becoming increasingly important in image segmentation. We thus specify the following problem description to incorporate additional terms associated with the image edges. In addition to the intraclass variance, edge terms may be added to the objective function as follows:

$$\mathcal{E}(R) = \underbrace{\sum_{(x,y,z) \in S_l} c_l(x,y,z) + \sum_{(x,y,z) \in S_u} c_u(x,y,z)}_{\text{edge term}} + \underbrace{\sum_{(x,y,z) \in R} (I(x,y,z) - \mu_0)^2 + \sum_{(x,y,z) \in \overline{R}} (I(x,y,z) - \mu_1)^2}_{\text{intraclass variance term}} \quad (14)$$

where S_l and S_u are two bounding surfaces of the region R, and $c_l(x, y, z)$ [respectively, $c_u(x, y, z)$] is the edge-based cost of voxel $I(x, y, z)$ measuring the inverse likelihood that the desired surface S_l (respectively, S_u) contains the voxel. Unfortunately, it seems at least nontrivial to extend our algorithm for solving this generalized problem.

IX. Conclusion

An algorithm to find globally optimal solution to segmentation by minimizing the intraclass variance was reported. Our approach optimally detects a region bounded by two coupled terrain-like surfaces in a volumetric image in a low-order polynomial time. We employ the techniques of parametric search, shape probing in computational geometry, and 3-D graph-searching.

The developed approximation algorithm exhibited a significantly improved running times while still produced highly close-to-optimal solutions when compared with our optimality guaranteeing algorithm.

Acknowledgments

References

1. Chan T, Vese L. An active contour model without edges. IEEE Trans Image Process. Feb; 2001 10(2):266–277. [PubMed: 18249617]

2. Mumford D, Shah J. Optimal approximation by piecewise smooth functions and associated variational problems. Commun Pure Appl Math. 1989; 42:577–685.

3. Cole R, Yap C. Shape from probing. J Algorithms. 1987; 8(1):19–38.

4. Dobkin D, Edelsbrunner H, Yap C. Probing convex polytopes. Proc 18th Annu ACM Symp Theory Comput (STOC). 1986:387–392.

5. Asano T, Chen D, Katoh N, Tokuyama T. Efficient algorithms for optimization-based image segmentation. Int J Computat Geometry Appl. 2001; 11(2):145–166.

6. Wu X. Efficient algorithms for the optimal-ratio region detection problems in discrete geometry with applications. Int J Computat Geometry Appl. 2009; 19(2):141–159.

7. Dou, X.; Wu, X.; Wahle, A.; Sonka, M. Globally optimal surface segmentation using regional properties of segmented objects. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR); 2008. p. 1-8.

8. Chan T, Esedoglu S, Nikolova M. Algorithms for finding global minimizers of image segmentation and denoising models. SIAM J Appl Math. 2006; 66(5):1632–1648.

9. Bresson X, Esedoglu S, Vandergheynst P, Thiran J, Osher S. Fast global minimization of the active contour/snake model. J Math Imag Vis. 2007; 28(2):151–167.

10. Li, K.; Wu, X.; Chen, D.; Sonka, M. Globally optimal segmentation of interacting surfaces with geometric constraints. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR); 2004. p. 394-399.

11. Cohen L, Kimmel R. Global minimum for active contour models: A minimal path approach. Int J Comput Vis. 1997; 24:57–78.

12. Appleton B, Talbot H. Globally optimal geodesic active contours. J Math Imag Vis. 2005; 23:67–86.

13. Appleton B, Talbot H. Globally minimal surfaces by continuous maximal flows. IEEE Trans Pattern Anal Mach Intell. Jan; 2006 28(1):106–118. [PubMed: 16402623]

14. Ardon R, Cohen L, Yezzi A. A new implicit method for surface segmentation by minimal paths: Applications in 3D medical images. Proc EMMCVPR05. 2006:520–535.

15. Montanari U. On the optimal detection of curves in noisy pictures. Commun ACM. 1971; 14:335–345.

16. Martelli A. Edge detection using heuristic search methods. Comput Graphics Image Process. 1972; 1:169–182.

17. Pope D, Parker D, Clayton P, Gustafson D. Left ventricular border detection using a dynamic search. Radiology. 1985; 155:513–518. [PubMed: 3885315]

18. Sonka M, Zhang X, DeJong S, Collins S, McKay C. Segmentation of intravascular ultrasound images: A knowledge-based approach. IEEE Trans Imaage Process. Dec; 1995 14(4):719–732.

19. Falcao A, Udupa J, Miyazawa F. An ultra-fast user-steered image segmentation paradigm: Live wire on the fly. IEEE Trans Image Process. Jan; 2000 19(1):55–62.

20. Wu, X.; Chen, D. Optimal net surface problems with applications. Proc. 29th Int. Colloquium Automata, Languages Programm. (ICALP); 2002. p. 1029-1042.

21. Boykov Y, Jolly M-P. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. Int Conf Comput Vis (ICCV). 2001; I:105–112.

22. Boykov Y, Funka-Lea G. Graph cuts and efficient N-D image segmentation. Int J Comput Vis. 2006; 70(2):109–131.

23. Delong, A.; Boykov, Y. Globally optimal segmentation of multi-region objects. Int. Conf. Comput. Vis. (ICCV); 2009.

24. Grady G. Minimal surfaces extend shortest path segmentation methods to 3D. IEEE Trans Pattern Anal Mach Intell. Feb; 2010 32(2):321–334. [PubMed: 20075461]

25. MacDonald D, Kabani N, Avis D, Evans A. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. NeuroImage. 2000; 12(3):340–356. [PubMed: 10944416]

26. Zeng X, Staib L, Schultz R, Duncan J. Segmentation and measurement of the cortex from 3D MR images using coupled surfaces propagation. IEEE Trans Med Imag. 1999; 18:100–111.

27. Goldenberg R, Kimmel R, Rivlin E, Rudzsky M. Cortex segmentation: A fast variational geometric approach. IEEE Trans Med Imag. Dec; 2002 21(12):1544–1551.

28. Spreeuwers L, Breeuwer M. Detection of left ventricular epi-and endocardial borders using coupled active contours. Comput Assist Radiol Surg. Jun.2003 :1147–1152.

29. Goldenberg R, Kimmel R, Rivlin E, Rudzsky M. Fast geodesic active contours. IEEE Trans Med Imag. Dec; 2001 10(10):1467–1475.

30. Katoh N, Ibaraki T. A parametric characterization and an ε-approximation scheme for the minimization of a quasiconcave program. Discrete Appl Math. 1987; 17:39–66.

31. Wu X, Chen D, Li K, Sonka M. The layered net surface problems in discrete geometry an medical image segmentation. Int J Computat Geometry Appl. 2007; 17(3):261–296.

32. Hand, D. Discrimination and Clasification. New York: Wiley; 1981.

33. de Berg, M.; van Kreveld, M.; Overmars, M.; Schwarzkopf, O. Computational Geometry: Algorithms and Applications. 2. New York: Springer-Verlag; 2000.

34. Goldberg A, Tarjan R. A new approach to the maximum-flow problem. J Assoc Comput Mach. 1988; 35:921–940.

35. Zhang H, Wahle A, Jhonson R, Scholz T, Sonka M. Cardiac MR image analysis: Left and right ventricular morphology and function. IEEE Trans Med Imag. Feb; 2010 29(2):350–364.

36. Radau, P.; Lu, Y.; Connelly, K.; Paul, G.; Dick, A.; Wright, G. Evaluation framework for algorithms segmenting short axis cardiac MRI, MIDAS J.-Cardiac MR Left Ventricle Segment. Challenge. 2009. [Online]. Available: http://hdl.handle.net/10380/3070

37. Boykov Y, Kolmogorov V. An experimental comparison of mincut/max-flow algorithms for energy minimization in vision. IEEE Trans Pattern Anal Mach Intell. Sep; 2004 24(9):1124–1137. [PubMed: 15742889]

38. Wijnhout J, Hendriksen D, van Assen H, van der Geest R. LV challenge LKEB contribution: Fully automated myocardial contour detection. MICCAI 2009's 3D Segmentat Challenge Clin Appl. 2009

39. Lu Y, Radau P, Connelly K, Dick A, Wright G. Automatic image-driven segmentation of left ventricle in cardiac cine MRI. MICCAI 2009's 3D Segmentat Challenge Clin Appl. 2009

40. Heimann T, Munzing S, Meinzer H, Wolf I. A shape-guided deformable model with evolutionary algorithm initialization for 3D soft tissue segmentation. Proc 20th Int Conf Inf Process Med Imag (IPMI). 2007:1–12.

41. Li K, Millington S, Wu X, Chen D, Sonka M. Simultaneous segmentation of multiple closed surfaces using optimal graph searching. Proc 19th Int Conf Inf Process Med Imag (IPMI). 2005:406–417.

42. Kainmueller D, Lamecker H, Zachow S, Heller M, Hege H. Multi-object segmentation with coupled deformable models. Proc 12th Annu Conf Med Image Understand Anal (MIUA). 2008:34–38.

43. Schwarz T, Heimann T, Wolf I, Meinzer H. 3D heart segmentation and volumetry using deformable shape. Comput Cardiol. 2007; 34:741–744.

44. Liu X, Chen D, Wu X, Sonka M. Optimal graph-based segmentation of 3D pulmonary airway and vascular trees across bifurcations. Proc 1st Int Workshop Pulmonary Image Anal. 2008:103–111.
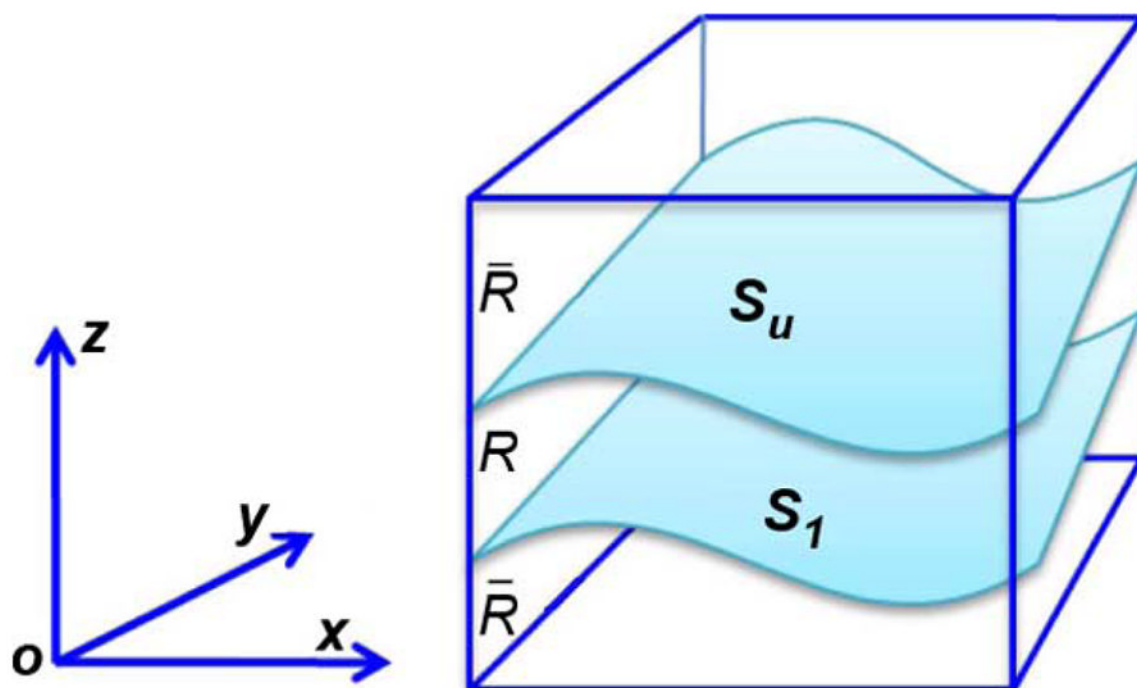
Fig. 1.
Region R enclosed by two coupled terrain-like surfaces S_i and S_u.
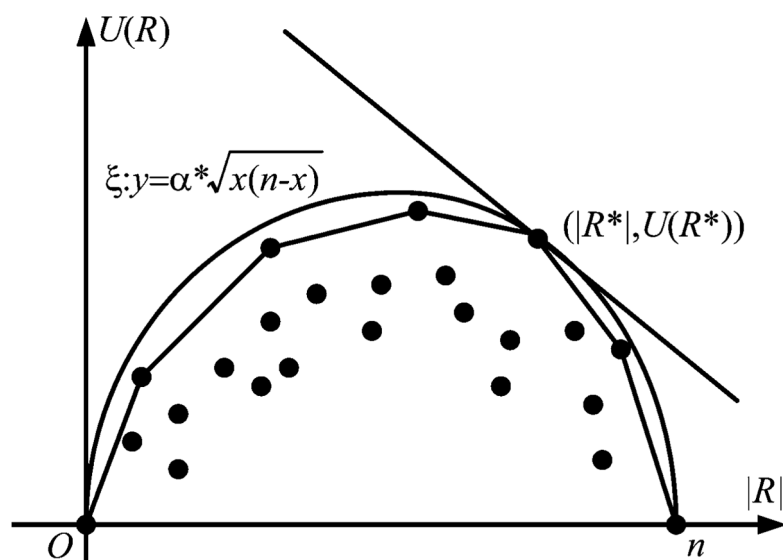
Fig. 2.
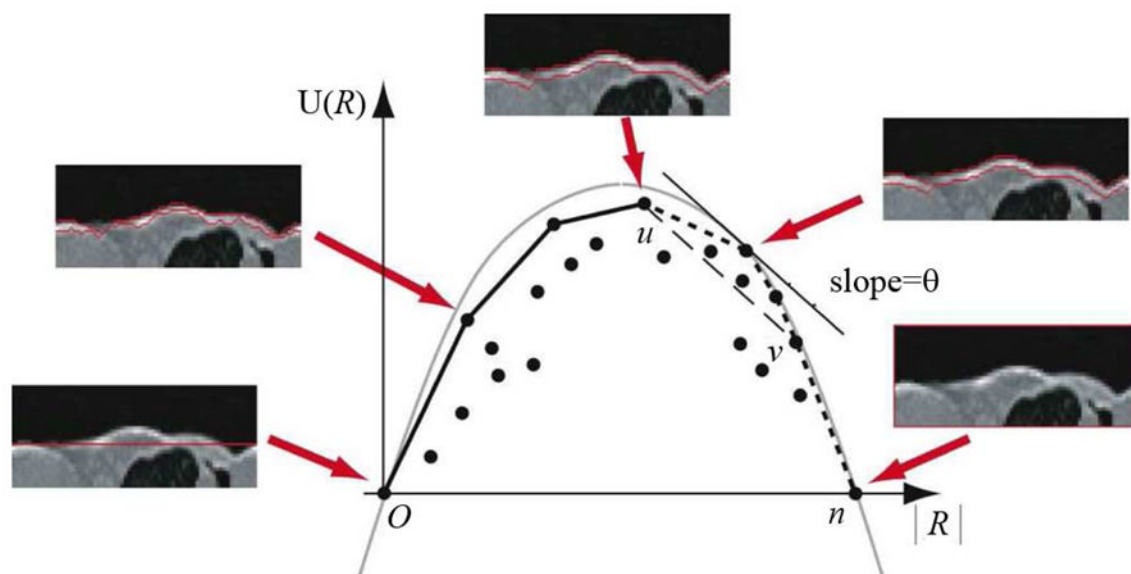Illustrating the proof of Lemma 1.

Fig. 3.
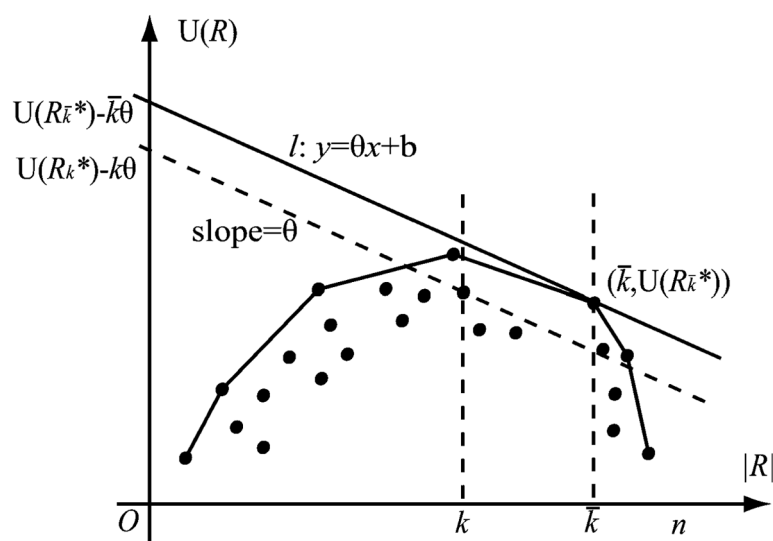Illustrating the construction of a convex hull using the shape probing technique.

Fig. 4.
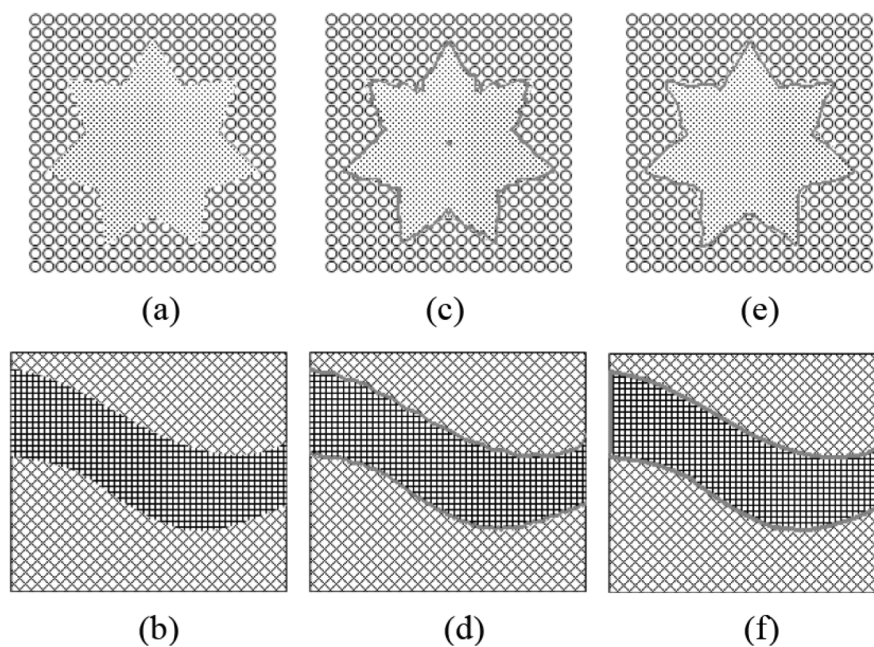Illustrating the proof of Lemma 2.

Fig. 5.
Computer phantoms with textures and segmentation results. (a), (b) Original images. (c), (d) The segmentation results from our method. (e), (f) The segmentation results from the Chan–Vese method.
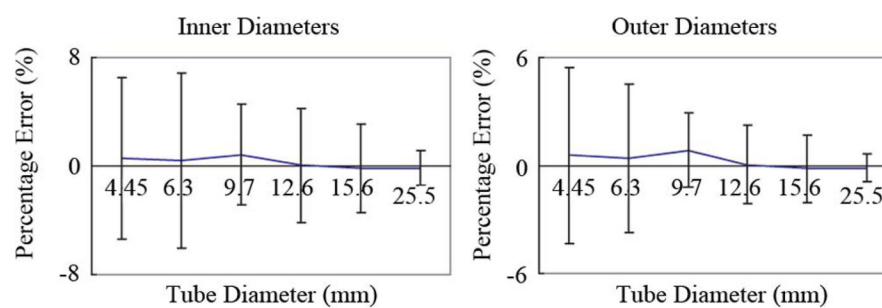
Fig. 6.
Signed percent errors of the inner- and outer-diameter measurements in the CT-imaged physical phantom tubes.
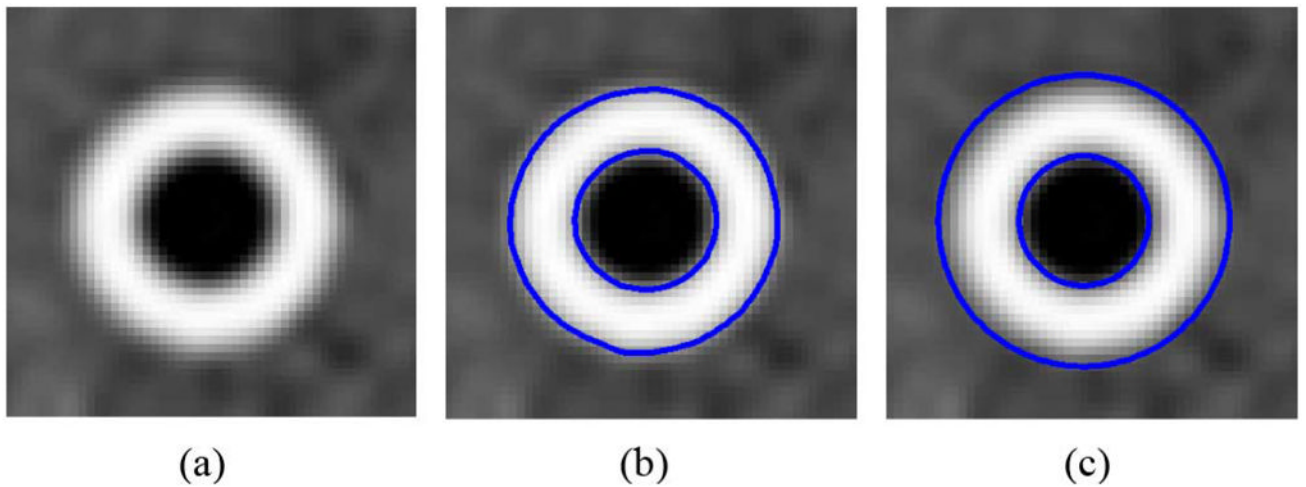
(a)   (b)   (c)

Fig. 7.
Segmentation result on CT-imaged physical phantom tubes. (a) An example slice from a 3-D CT-imaged physical phantom. (b) Segmentation result achieved using Chan–Vese method. The errors of the measured diameters of the inner and outer walls were −0.14 mm and −0.31 mm, respectively. (c) Segmentation result obtained using our method. The errors of the measured diameters of the inner and outer walls were 0.04 mm and 0.07 mm, respectively. Five-point moving average is applied to both methods to smooth the results.
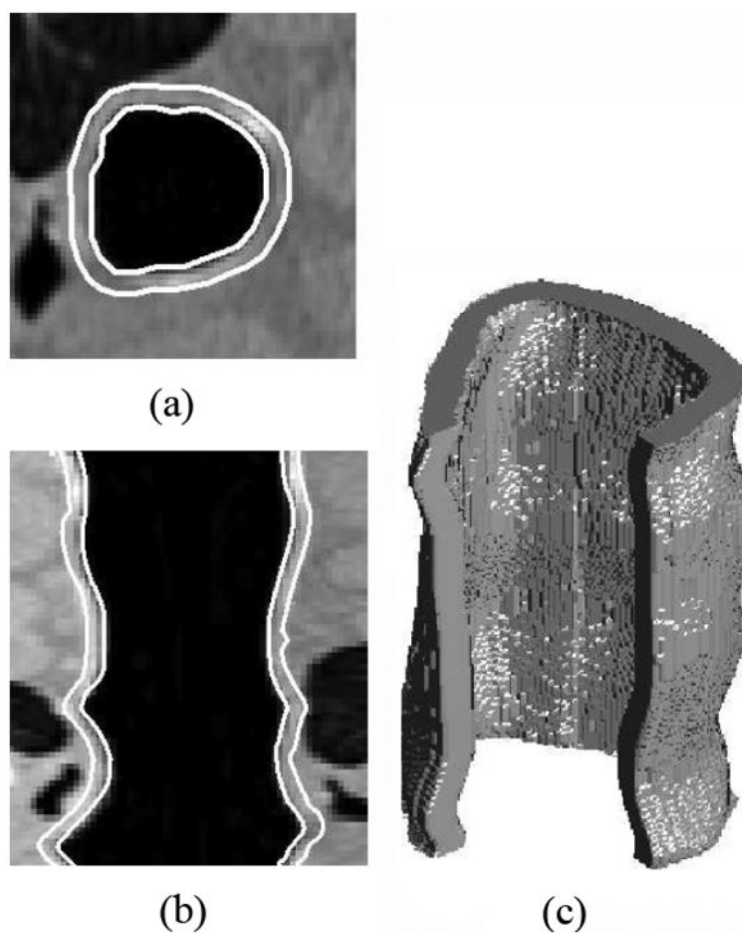
Fig. 8.
Segmented inner and outer walls of human pulmonary airways imaged with multidetector CT. (a) Transverse slice. (b) Sagittal cross-section. (c) 3-D view of segmented airway, cut-open for visibility.
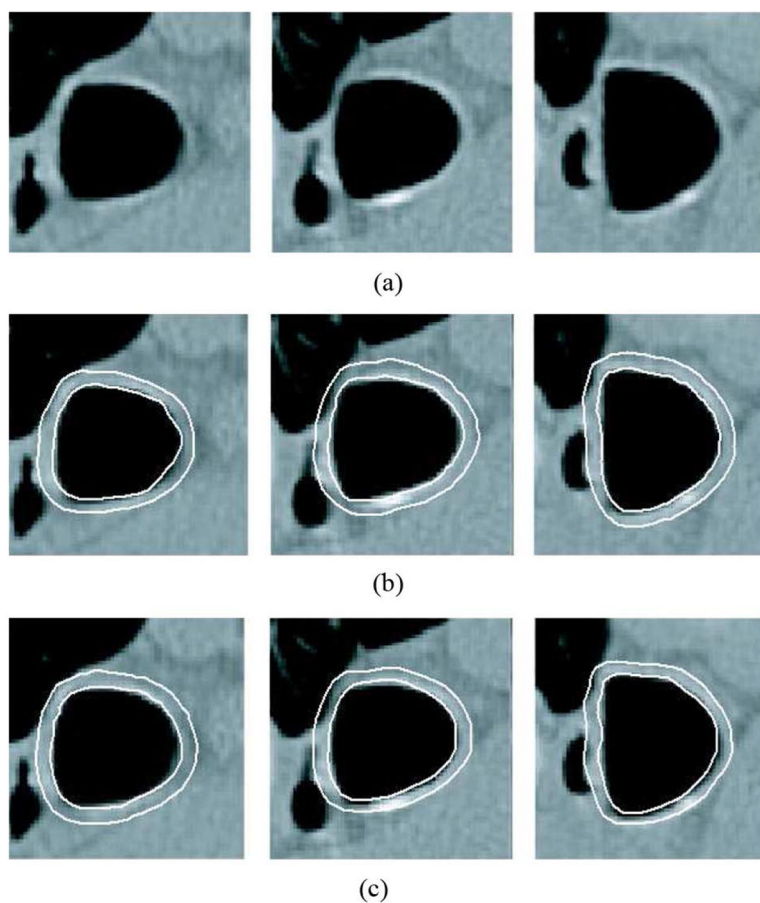
Fig. 9.
Comparison of computer-segmented and expert-traced inner and outer airway wall borders in *in vivo* human 3D CT images. (a) Original image. (b) Expert-traced borders. (c) Segmentation obtained using our 3-D method.
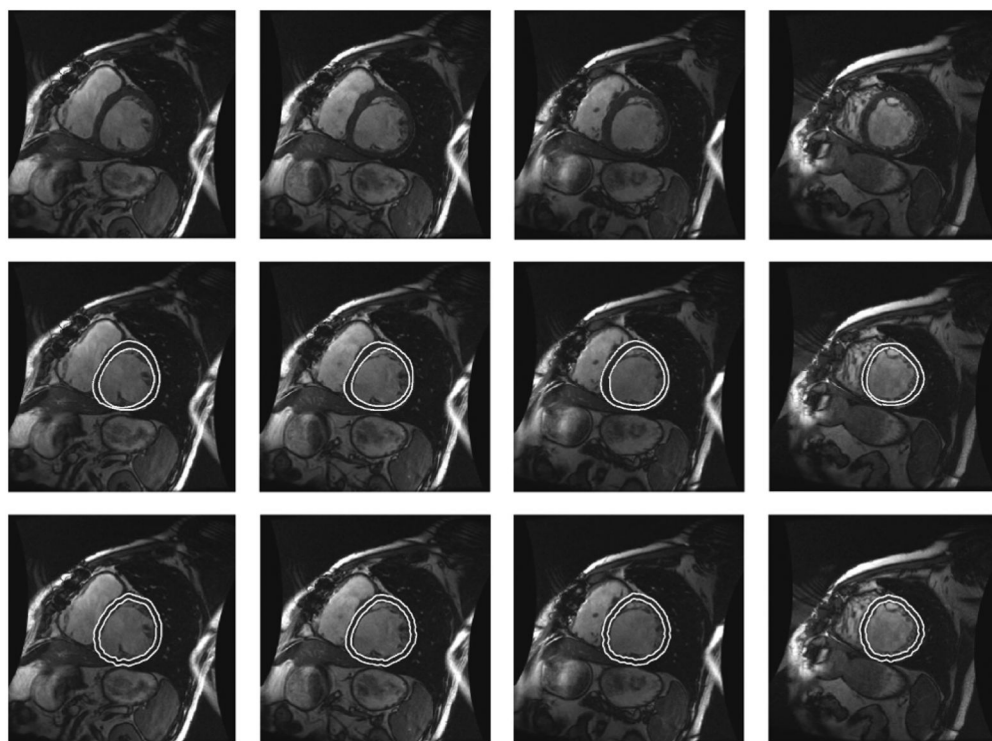
Fig. 10.
Segmentation of MR ventricular walls. Top row shows the original images. Middle row shows expert-traced endo- and epi-cardial walls. Bottom row shows results of the reported approach.
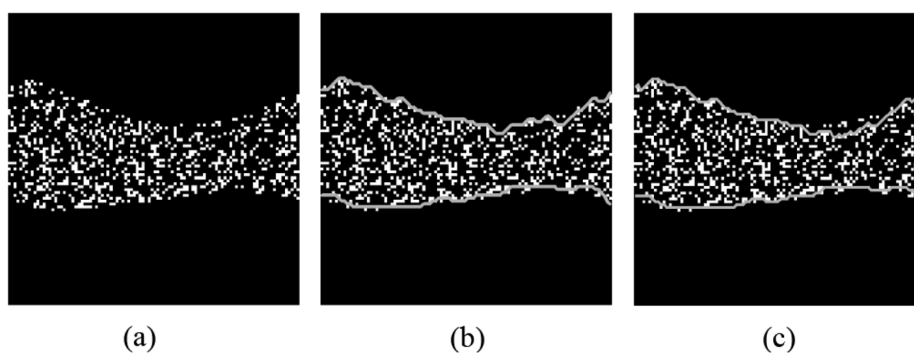
Fig. 11.
Comparison of segmentation results for an example computer phantom image. (a) Original image. (b) Segmentation produced by our method. (c) Segmentation result obtained by the iterative ICV method.
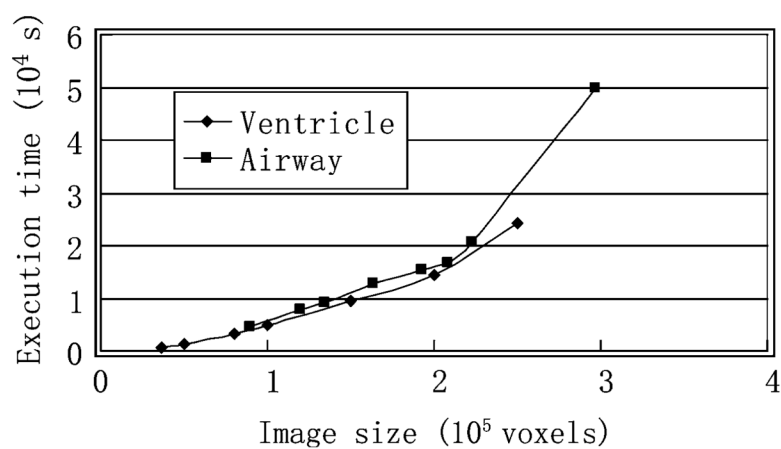
Fig. 12.
Execution times for 3-D airway CT images and 3-D ventricular MR images as a function of image size.
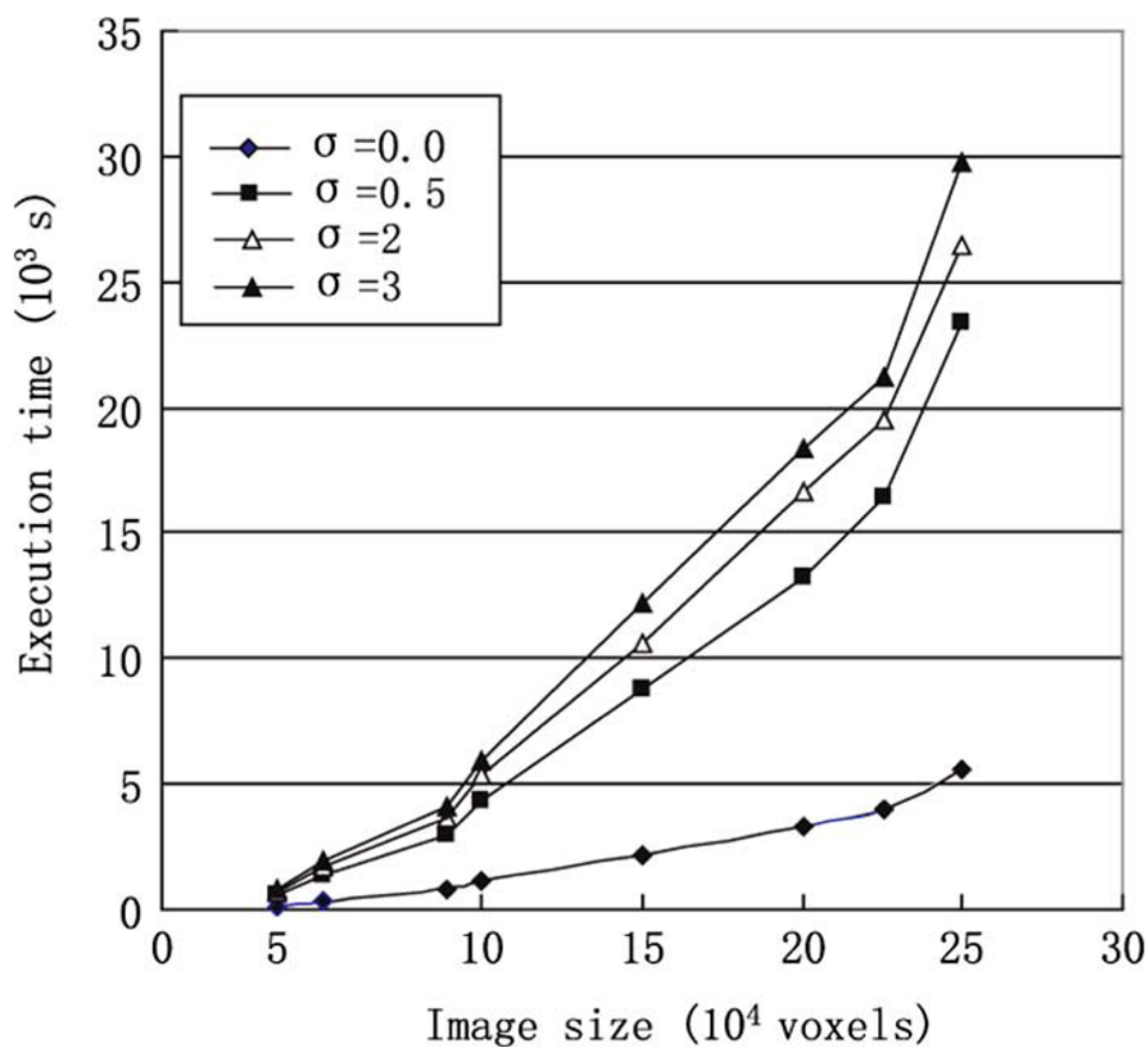
Fig. 13.
Execution times for computer phantoms with different levels of Gaussian noise ($\sigma = 0.0$, 0.5, 2.0, 3.0) as a function of image size.
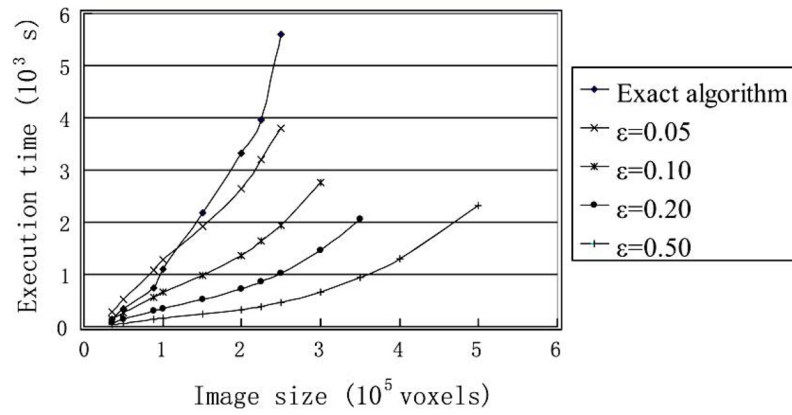
Fig. 14.
Comparison of execution times on computer phantoms by the approximation algorithm against the exact algorithm.
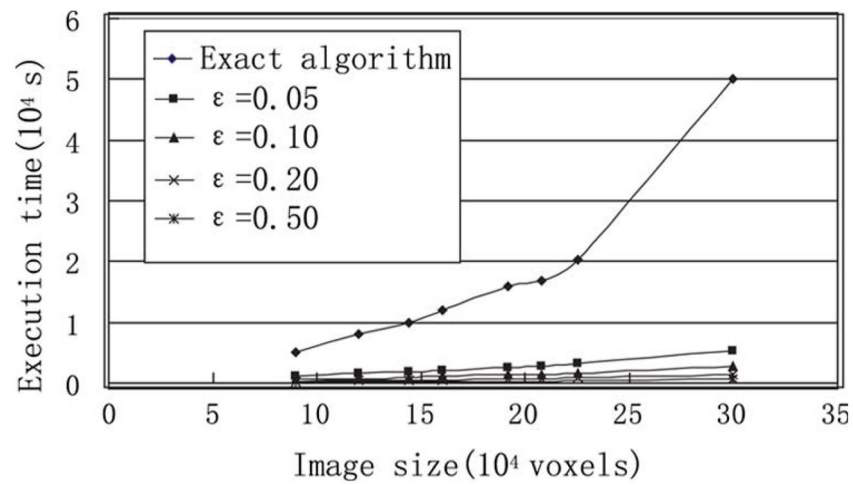
Fig. 15.
Comparison of execution times on 3-D airway CT image data by the approximation algorithm against the exact algorithm.
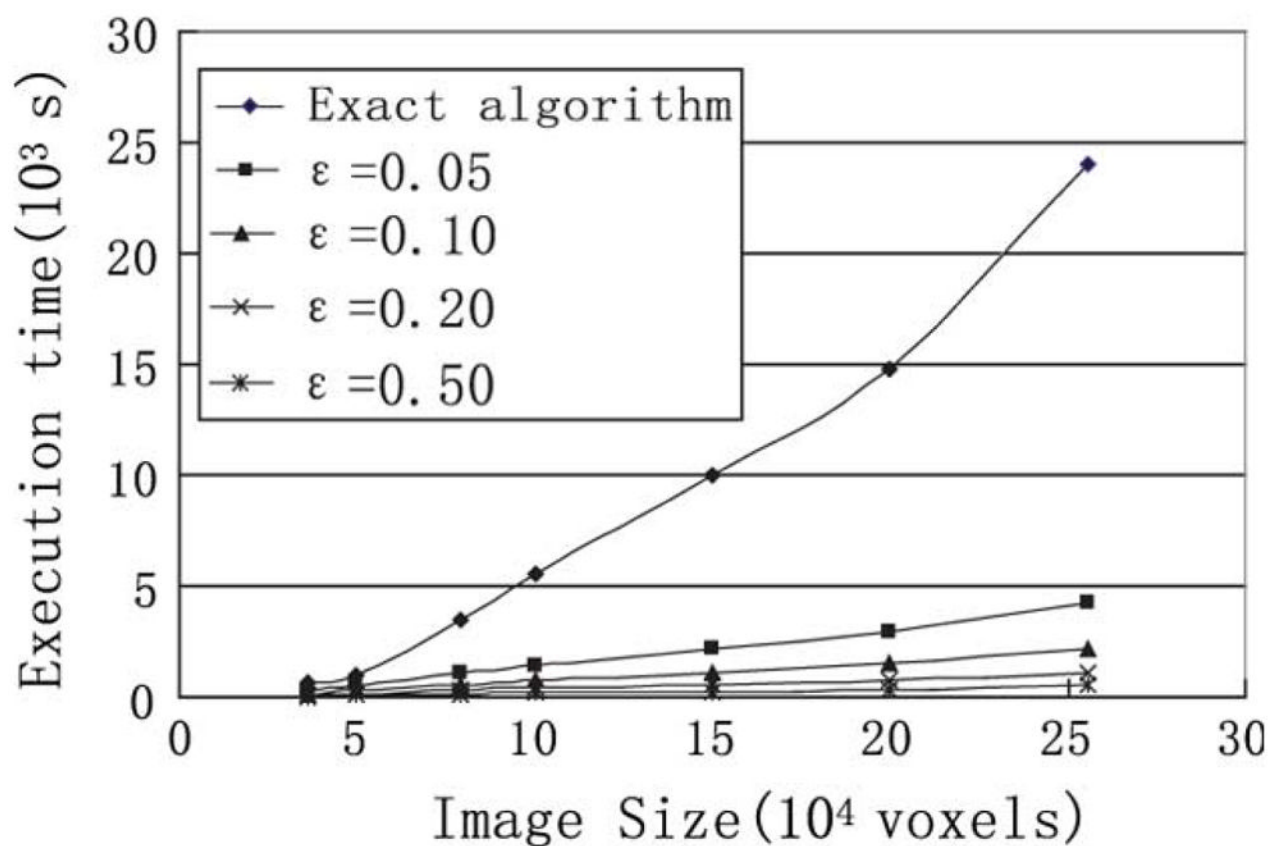
Fig. 16.
Comparison of execution times on 3-D ventricular MR images by the approximation algorithm against the exact algorithm.

TABLE I

Surface Positioning Errors of 3-D Left Ventricular Wall Segmentation: (a) Our Exact Algorithm, (b) Iterative ICV Method

|	Signed Error(Pixels)		Unsigned Error(Pixels)	
	Inner	**Outer**	**Inner**	**Outer**
Normal	−0.710 ± 0.538	−0.584 ± 0.437	0.778 ± 0.436	0.793 ± 0.437
Patient	−0.494 ± 0.773	0.528 ± 0.589	0.784 ± 0.476	0.946 ± 0.589
		(a)		
Normal	−0.727 ±0.591	−0.622 ± 0.447	0.806 ± 0.568	0.862 ± 0.564
Patient	−0.544 ±0.769	−0.936 ±0.725	0.794 ±0.507	1.110 ±0.725

(b)

TABLE II

Comparison of Average Positioning Error

ID	Our Method		Method 1 [38]		Method 2 [39]	
	Inner	Outer	Inner	Outer	Inner	Outer
SC-HF-I-05	1.83	2.59	1.7	1.6	1.37	1.06
SC-HF-I-06	2.47	2.38	1.8	2.4	1.32	2.15
SC-HF-I-07	1.92	2.25	2.8	2.2	2.44	1.57
SC-HF-I-08	2.69	2.04	2.1	2.3	1.61	2.59
SC-HF-NI-07	2.10	1.94	2.9	2.3	2.61	1.09
SC-HF-NI-11	2.02	2.69	2.3	2.0	2.19	1.57
SC-HF-NI-31	2.24	1.94	2.1	2.6	2.41	1.51
SC-HF-NI-33	2.21	2.06	2.3	2.9	2.04	2.01
SC-HYP-06	2.20	1.81	2.0	2.1	2.30	3.26
SC-HYP-07	2.23	2.47	2.2	2.4	1.87	2.81
SC-HYP-08	2.18	3.88	3.9	2.2	3.77	1.81
SC-HYP-37	1.71	2.31	2.3	3.0	1.93	2.45
SC-N-05	1.69	2.06	2.0	2.1	1.66	1.38
SC-N-06	2.27	1.55	2.1	1.7	1.77	1.77
SC-N-07	2.17	2.07	1.7	2.5	1.79	1.63
Average	2.12	2.27	2.29	2.28	2.07	1.91
Stdev	0.27	0.54	0.57	0.39	0.61	0.63

TABLE III

Comparison of Dice Metric

ID	Our Method		Method 1 [38]		Method 2 [39]	
	Inner	Outer	Inner	Outer	Inner	Outer
SC-HF-I-05	0.94	0.92	0.94	0.96	0.94	0.97
SC-HF-I-06	0.91	0.95	0.92	0.94	0.94	0.93
SC-HF-I-07	0.95	0.94	0.88	0.93	0.87	0.95
SC-HF-I-08	0.93	0.95	0.92	0.93	0.94	0.94
SC-HF-NI-07	0.92	0.95	0.89	0.93	0.89	0.97
SC-HF-NI-11	0.94	0.89	0.91	0.95	0.90	0.96
SC-HF-NI-31	0.91	0.94	0.91	0.93	0.90	0.96
SC-HF-NI-33	0.88	0.94	0.87	0.91	0.89	0.94
SC-HYP-06	0.88	0.95	0.89	0.93	0.85	0.90
SC-HYP-07	0.90	0.94	0.90	0.93	0.90	0.92
SC-HYP-08	0.94	0.84	0.82	0.94	0.85	0.95
SC-HYP-37	0.85	0.87	0.83	0.90	0.84	0.92
SC-N-05	0.87	0.92	0.84	0.92	0.87	0.95
SC-N-06	0.89	0.93	0.89	0.94	0.88	0.93
SC-N-07	0.89	0.92	0.91	0.92	0.89	0.93
Average	0.91	0.92	0.89	0.93	0.89	0.94
Stdev	0.03	0.03	0.03	0.01	0.03	0.02

TABLE IV

Average Execution Times of Each Shape Probing Process

Image Size	Execution Time (s)	Image Size	Execution Time (s)
$30^2 \times 40$	0.5	$80^2 \times 40$	5.9
$50^2 \times 40$	2.1	$100^2 \times 40$	15.7

TABLE V

Segmentation Accuracy of the Approximation Method on Computer Phantoms

e	Average Approx-opt Ratio	Max Approx-opt Ratio	% of Non-Optimal Results
0.05	100.00%	100.00%	1.25%
0.1	100.03%	100.15%	7.50%
0.2	100.22%	100.80%	25%
0.5	101.02%	104.35%	51.25%

TABLE VI

Segmentation Accuracy of the Approximation Method on 3-D CT Airway Wall Images

e	Average Approx-opt Ratio	Unsigned Error(Pixels)		Signed Error(Pixels)	
		Inner Wall	Outer Wall	Inner Wall	Outer Wall
0.05 *	100.00%	0.422 ± 0.381	-0.127 ± 0.460	0.657 ± 0.149	0.572 ± 0.032
0.1	100.10%	0.422 ± 0.381	-0.127 ± 0.460	0.656 ± 0.149	0.573 ± 0.032
0.2	100.52%	0.422 ± 0.381	-0.127 ± 0.460	0.657 ± 0.149	0.572 ± 0.032
0.5	100.97%	0.422 ± 0.381	-0.127 ± 0.460	0.657 ± 0.149	0.572 ± 0.032

*
For $e = 0.05$ in the table, the objective function value was optimal and the average surface positioning errors were the same as those obtained by the exact algorithm.

TABLE VII

Segmentation Accuracy of the Approximation Method on 3-D MR Ventricular Images

ε	Average Approx-opt Ratio	Signed Error(Pixels)		Unsigned Error(Pixels)	
		Inner Wall	Outer Wall	Inner Wall	Outer Wall
0.05 *	100.00%	−0.710 ± 0.538	−0.584 ± 0.437	0.778 ± 0.436	0.793 ± 0.437
0.1	100.14%	−0.709 ± 0.538	−0.586 ± 0.443	0.774 ± 0.437	0.807 ± 0.443
0.2	100.86%	−0.708 ± 0.535	−0.586 ± 0.444	0.767 ±0.440	0.821 ± 0.444
0.5	101.43%	−0.720 ± 0.547	−0.587 ±0.428	0.803 ± 0.429	0.752 ± 0.428

*
For $\varepsilon = 0.05$ in the table, the objective function value was optimal and the average surface positioning errors were the same as those obtained by the exact algorithm.

TABLE VIII

Comparison of Surface Positioning Errors on 20 MR Images of Left Ventricular Walls of Normal Human Hearts: (a) Our Exact Algorithm, (b) Iterative ICV Method, (c)The Proposed Approximation Algorithm

	Signed Error(Pixels)		Unsigned Error(Pixels)	
	Inner	Outer	Inner	Outer
	−0.710 ± 0.538	−0.584 ± 0.437	0.778 ± 0.436	0.793 ± 0.437

(a)

	−0.727 ± 0.591	−0.622 ± 0.447	0.806 ± 0.568	0.862 ± 0.564

(b)

	Inner	Outer	Inner	Outer
$\varepsilon = 0.05$	−0.710 ± 0.538	−0.584 ± 0.437	0.778 ± 0.436	0.793 ± 0.437
$\varepsilon = 0.1$	−0.709 ± 0.538	−0.586 ± 0.443	0.774 ± 0.437	0.807 ± 0.443
$\varepsilon = 0.2$	−0.708 ± 0.535	−0.586 ± 0.444	0.767 ± 0.440	0.821 ± 0.444
$\varepsilon = 0.5$	−0.720 ± 0.547	−0.587 ± 0.428	0.803 ± 0.429	0.752 ± 0.428

(c)