A Geometric Approach to Segmentation and Analysis of 3D Medical Images

R. Malladi,* R. Kimmel,* D. Adalsteinsson,* G. Sapiro,† V. Caselles,‡ J. A. Sethian§

Abstract

A geometric scheme for detecting, representing, and measuring 3D medical data is presented. The technique is based on deforming 3D surfaces, represented via level-sets, towards the medical objects, according to intrinsic geometric measures of the data. The 3D medical object is represented as a (weighted) minimal surface in a Riemannian space whose metric is induced from the image. This minimal surface is computed using the level-set methodology for propagating interfaces, combined with a narrow band technique which allows fast implementation. This computation technique automatically handles topological changes. Measurements like volume and area are performed on the surface, exploiting the representation and the high accuracy intrinsic to the algorithm.

Key words: Deformable models, minimal surfaces, segmentation, measurements, level sets, narrow-band methods, fast implementation.

1. Introduction

One of the basic problems in medical imaging is to detect the boundaries of the objects of interest, efficiently represent them, and perform measurements significant for diagnosis, surgery, or other applications. Active contours, initially proposed by Kass et al., and deformable surfaces introduced by Terzopoulos et al [44], can be used to solve the first part of this goal, i.e. the segmentation task. These models are based on deforming an initial contour or surface towards the boundary of the object to be detected. The deformation is obtained by minimizing a functional designed such that its (local) minima is obtained at the boundary of the object [4, 44].

Implicit surface-evolution models for medical image segmentation were proposed by Caselles et al. [5] and by Malladi et al. [28, 29]. In these models, the surface propagates by an implicit velocity containing two terms, one related to the regularity of the deforming shape and the other attracting it to the boundary. The model is given by a geometric flow based on mean curvature motion, and is not the result of minimizing an energy functional. At their core, these models rely on the Osher-Sethian level-set model for evolving surfaces [34, 39]. The level-set approach offers a robust, stable, and efficient numerical algorithm that allows for complex changes in topology, accurate evaluation of curvatures, and straightforward implementation. In addition, fast implementations of this approach were introduced by Adalsteinsson and Sethian [1] and Malladi, Sethian, and Vemuri [30]. In the energy based approach, the flexibility in topology can be obtained only if special tracking procedures are implemented [31, 42]. For details about a wide collection of level-set applications, see [40].

In [6], formal mathematical relation between these two approaches for 2D object detection is shown, proposing the "geodesic active contours." As shown in [6], the geodesic active contours model has the following main properties:

1. It connects energy and curve evolution approaches of active contours.
2. Presents the snake problem as a geodesic computation.
3. Improves existing models as a result of the geodesic formulation.
4. Allows simultaneous detection of interior and exterior boundaries of several objects without special contour tracking procedures.
5. Holds formal existence, uniqueness, and stability results.

6. Stops automatically.

The 2D model in [6] was then extended to 3D surfaces in [7, 8]. The proposed 3D approach has the same important properties as its 2D analogous one. The original model in Malladi et al. [29] (and in [5]) has also been extended to 3D and coupled with fast tube methods for approximating interface position in [24, 27]. In this paper, we combine the model in [7], which is revisited here, with the fast numerical approach first described in [1, 30] and extended to 3D in [2, 24, 27] and present a complete approach for 3D medical data analysis. Then, following the 2D analysis started in [36], we exploit the representation and the high accuracy intrinsic to the algorithm to compute important characteristics such as volume and surface area of the segmented 3D medical organs.

We should note that the deformable surfaces model used here is related to a number of previously or simultaneously developed results. It is of course closely related to the problems of Terzopoulos and colleagues on energy based deformable surfaces, and the works by Malladi et al. and Caselles et al. [5, 24, 29]. It is an extension of the 2D model derived in [6]; see [7, 8]. The basic equations in this paper, as well as the corresponding 2D ones in [6], were simultaneously developed in [20, 21, 41]. In [20, 21], the authors base their approach on gradient flows, while in [41], the development partially follows from the Mumford-Shah segmentation technique [32]. Similar 3D models are studied in [45, 46] as well. In [43], the authors work with multiple initializations and multiple parameter-space, partially motivated by the shape theory in [22], using the equations in [5, 28, 29]. In this paper, we develop a complete and fast approach for 3D segmentation and area and volumetric measurements on the segmented objects; see Whitaker [46] for fast multi-scale implementation. For more details on the similitude and differences between these approaches, see [6, 7, 8].

The details of tube method for moving 3D shapes and a comparison to the traditional full matrix method may be found in [24, 27].

2. Basic approaches of deformable surfaces

The 3D deformable surface model was introduced by Terzopoulos et al. [44] and further extended by others (e.g. [11, 12, 13]). A parameterized surface \( v(r, s) = (x(r, s), y(r, s), z(r, s)), (r, s) \in [0,1] \times [0,1], \)

is considered, and the energy functional is given by

\[
E(v) = \int_{\Omega} \left[ \omega_{10} \left| \frac{\partial u}{\partial r} \right|^2 + \omega_{01} \left| \frac{\partial v}{\partial s} \right|^2 + \omega_{11} \left| \frac{\partial^2 v}{\partial r \partial s} \right|^2 + 
\omega_{20} \left| \frac{\partial^2 v}{\partial r^2} \right|^2 + \omega_{02} \left| \frac{\partial^2 v}{\partial s^2} \right|^2 + P \right] dr ds,
\]

where \( P = -\left| \nabla v \right|^2, \) or any related decreasing function of the gradient. The first terms are related to the smoothness of the surface, while the last one is responsible for attracting it to the object. The evolution of the surface is expressed via Euler-Largange equations. From an initial surface \( \Sigma_0, \) generally near the desired 3D boundary \( \Omega, \) the algorithm tries to deform \( \Sigma_0 \) towards a local minimum of \( E. \)

The geometric models proposed in [5, 29] can easily be extended to 3D object detection. Let \( Q = [0, a] \times [0, b] \times [0, c] \) and \( I : Q \rightarrow \mathbb{R} \) be a given 3D data image. Let \( g(I) = \frac{1}{1 + \left| \nabla I \right|^p}, \) where \( I \) a regularized version of \( I, \) and \( p = 1 \) or 2. \( g(I) \) acts as an edge detector so that the object boundary we are looking for is ideally given by the equation \( g = 0. \) Our initial active surface \( \Sigma_0 \) will be embedded as a level-set [34, 39] of a function \( u_0 : Q \rightarrow \mathbb{R}, \) say \( \Sigma_0 = \{ x : u_0(x) = 0 \} \) with \( u_0 \) being positive in the exterior and negative in the interior of \( \Sigma_0. \) The evolving active surface is defined by \( \Sigma_t = \{ x : u(t, x) = 0 \} \) where \( u(t, x) \) is the solution of

\[
\frac{\partial u}{\partial t} = g(I)\left| \nabla u \right| \text{div} \left( \frac{\nabla u}{\left| \nabla u \right|} \right) + \nu g(I)\left| \nabla u \right|
\]

\[
= g(I)(\nu + H)\left| \nabla u \right|, \quad (1)
\]

with initial condition \( u(0, x) = u_0(x) \) and Neumann boundary conditions. Here \( H = \text{div} \left( \frac{\nabla u}{\left| \nabla u \right|} \right) \) is the sum of the two principal curvatures of the level sets of \( u, \) (twice its mean curvature) and \( \nu \) is a positive real constant. The 2D version of this model was heuristically justified in [5, 29]. It contains:

- A smoothing term: Twice the mean curvature in the case of (1). More efficient smoothing velocities as those proposed in [3, 9, 33, 25, 27] can be used instead of \( H^1 \).

- A constant balloon-type force similar to [12]; \( \nu \left| \nabla u \right|. \)

- A stopping factor \( g(I). \) (Related edge detectors can be used, e.g. [47].) Note that in this model, \( g \) must be zero for the surface to stop (see next section).

---

1Although curvature flows smooth 2D curves [17, 18, 37, 38], a 3D geometric flow that smooths all possible surfaces was not found [35]. Frequently used are mean curvature or the positive part of the Gaussian curvature flows [5, 9].
The sign conventions here are adapted to active contours propagating inwards. For active contours evolving from the inside outwards, we take $\nu < 0$. This is a drawback of this model: the active contours cannot go in both directions (see also [43]). Moreover, we always need to select $\nu \neq 0$ even if the surface is close to the object's boundary.

The goal in [7, 8] was to define a 3D geometric model (with level-set formulation) corresponding to the minimization of a meaningful and intrinsic energy functional. It is motivated by the extension of 2D geometric model to the geodesic active contours as done in [6] and briefly described in the following section.

3. Weighted minimal surfaces

In [6], a model for 2D object detection based on the computation of geodesics in a given Riemannian space was presented. This means that we are computing paths or curves of minimal (weighted) length. That model also shows the exact mathematical relation between energy-based active contours and those based on curvature motion, improving on those approaches. This idea may be extended to 3D surfaces [7, 8], computing surfaces of minimal area, where the area is defined in a given Riemannian space. In the case of surfaces, arc-length is replaced by surface area $A = \int \int da$, and weighted arc-length by “weighted” area,

$$A_R = \int \int g(\rho) da,$$

where $da$ is the (Euclidean) element of area. Surfaces minimizing $A$ are denoted as minimal surfaces [35]. In the same manner, we will denote by minimal surfaces those surfaces that minimize (2). The area element $da$ is given by the classical area element in Euclidean space, while the area element $da_R$ is given by $g(\rho) da$. Observe that $da_R$ corresponds to the area element induced on a surface of $\mathbb{R}^3$ by the metric of $\mathbb{R}^3$ given by $g_{ij} dx_i dx_j$ with $g_{ij} = g(\rho)^2 \delta_{ij}$. This is the 3D analogue of the metric used in [6] to construct the geodesic active contour model. The energy $A_R$ can be formally derived from the original energy formulation using basic principles of dynamical systems [6, 7, 8], further justifying this model. The basic element of the deformable model will be given by minimizing (2) by means of an evolution equation obtained from its Euler-Lagrange. Let us point out the basic characteristics of this flow.

The Euler-Lagrange of $A$ is given by the mean curvature $H$, resulting a curvature (steepest descent) flow

$$\frac{\partial S}{\partial t} = H N,$$

where $S$ is the 3D surface and $N$ is its inner unit normal. With the sign conventions explained above, the corresponding level-set [34] formulation is

$$u_t = \nabla u |\nabla u| \frac{\nabla u}{|\nabla u|} = \nabla u H. \text{ Therefore, the mean curvature motion provides a flow that computes (local) minimal surfaces [10]. Computing the Euler-Lagrange of } A_R, \text{ we get}

$$S_t = (g H - \nabla g \cdot \vec{N}) \vec{N}. \quad (3)$$

This is the basic weighted minimal surface flow. Taking a level-set representation, the steepest descent method to minimize (2), yields

$$\frac{\partial u}{\partial t} = |\nabla u| \nabla \left( g(\rho) \frac{\nabla u}{|\nabla u|} \right)$$

$$= g(\rho) |\nabla u| \nabla \left( \frac{\nabla u}{|\nabla u|} \right) + \nabla g(\rho) \cdot \nabla u. \quad (4)$$

We note that comparing with previous geometric surface evolution approaches for 3D object detection, the minimal surfaces model includes a new term, $\nabla g \cdot \nabla u$. This term is fundamental for detecting boundaries with fluctuations in their gradient, since it will cause the surfaces to stop even if $g$ is not zero. See [6, 7, 8] for details.

As in the 2D case, we can add a constant force to the minimization problem (minimizing volume), obtaining the general minimal surfaces model for object detection:

$$\frac{\partial u}{\partial t} = |\nabla u| \nabla \left( g(\rho) \frac{\nabla u}{|\nabla u|} \right) + \nu g(\rho) |\nabla u|. \quad (5)$$

This is the flow we will use for 3D object detection. It has the same properties and geometric characteristics as the geodesic active contours, leading to accurate numerical implementations and topology free object segmentation. The flow also holds important existence and completeness properties. Basically, there is a unique solution to the flow in the viscosity sense, and the flow converges to ideal objects [7, 8] (see also [20, 21]). It is also possible to bound the size of the gaps that can be successfully avoided by the flow [6, 8, 15].

As pointed out in the introduction, see [14, 16, 20, 21, 41, 43, 45, 46] for related approaches and [6, 7, 8, 24, 26, 27] for full comparisons.

4. The Tube method

4.1. Motivation

For a two-dimensional interface evolving in three space dimensions, the level set algorithm is at least an $O(N^3)$ method per time step, where $N$ is the number of points in the spatial direction. One drawback of the technique stems from the expense; by embedding the
interface as the zero-level set of a higher dimensional function, a two-dimensional interface problem has been transformed into a three-dimensional problem.

The main idea of the tube method is to modify the level set method so it only affects points close to the region of interest, namely the cells where the front is located. This will save considerable computer time, since it reduces by an order of magnitude the number of points where terms must be evaluated. This is particularly beneficial in cases where there is significant computation to be made at each point, as well as simpler cases involving fronts propagating with a constant normal speed.

Our tube is constructed by choosing points that lie less than some given distance away from the curve. The resulting tube-like domain contains the zero-level set. This method was used before in [29] and described in more detail in [1]. A three dimensional version of the tube method was used in [2].

As an illustration, Figure 1 shows the initial surface for a circle in two dimensions. On the left the surface is defined on a square that contains the circle. On the right the surface is only defined in a neighborhood of the circle.

4.2. Implementation

The definition of a tube used here is just a set of points inside a rectangular array. There are no restrictions about the shape of the tube, however most of the time it looks like a border around the current front. Initially the surface is defined as the signed distance from the front for all points closer then some predefined maximum value. Outside that tube the surface is constant, chosen to make the surface continuous on all of space.

The front is then moved the same way as in the standard level set method, however only those points belonging to the tube are updated. The tube is rebuilt whenever the front gets too close to an edge. Boundary conditions assume fixed values of u at the tube edge; a more sophisticated implementation is given in [1].

4.3. Speed comparison

To check the speed of the algorithm, we studied the collapse of a sphere with unit speed and according to its curvature. Collapse under mean curvature involves more computations than just the gradient, and better reflects the abilities of the narrow band method. Furthermore, inclusion of such parabolic terms requires smaller time steps than those required by the advection term. The initial sphere is chosen so that it almost fills the computational box. We perform timings with several different time steps and compare the execution time. The full-matrix method and the tube method require approximately the same time when there are 60 cells per side; however, as the number of grid points increases, the narrow band method becomes significantly faster than the full matrix method. Reinitializing the function values in the tube takes up sizable fraction of time but that happens only in the starting stages. The number of reinitializations is independent of the time step, so when the time step is decreased the reinitialization cost becomes a smaller fraction of the total time. We tabulate the execution times of the full matrix and the tube methods until the sphere collapses in Table 1.

5. Experimental results

The experiments described here are based on implementing the level-set flow (4) using the narrow band method [1, 30, 24, 27] and the 3D segmentation results are reproduced from Malladi and Sethian [27].

The volume is computed by counting the interior voxels as one volume unit (this is straightforward from the level-sets representation), and proportionally adding the contribution of boundary voxels. Linear
<table>
<thead>
<tr>
<th>Grid size</th>
<th>$\Delta t=0.8h$</th>
<th>$\Delta t=0.1h$</th>
<th>$\Delta t=0.05h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$60 \times 60 \times 60$</td>
<td>61s</td>
<td>465s</td>
<td>925s</td>
</tr>
<tr>
<td>$100 \times 100 \times 100$</td>
<td>480s</td>
<td>3856s</td>
<td>7710s</td>
</tr>
<tr>
<td>$150 \times 150 \times 150$</td>
<td>2570s</td>
<td>20500s</td>
<td>41100s</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The tube method</th>
</tr>
</thead>
<tbody>
<tr>
<td>$60 \times 60 \times 60$</td>
</tr>
<tr>
<td>$100 \times 100 \times 100$</td>
</tr>
<tr>
<td>$150 \times 150 \times 150$</td>
</tr>
</tbody>
</table>

Table 1. Speed comparison between the regular full matrix method and the tube method on a Sun Sparc 10 machine.

interpolation was used in this case, but more precise interpolators can be implemented if better accuracy is needed. The surface area is computed by adding the areas of all the triangles representing the zero level-set, that are extracted from the implicit representation.

The first example presents the segmentation and volume measurement of the soft tissue in a 3D CT data of two thighs (see Fig. 2). The computation was done by starting from a collection of spheres inside the thighs and growing outwards thereby reconstructing the thighs and the femur. Fig. 3 presents the final segmentation result in which we can observe the outer surface of the bones and the skin. The 3D data is given on a grid of $128 \times 128 \times 61$ which subsequently was mapped onto a unit cube; i.e. $1 \times 1 \times 1$. The measured volume of the soft tissue is 0.268492.

In Fig. 4, the left bone of the first example is extracted by using simple region growing on the segmented thighs. The volume of the bone is measured to be 0.00844465.

In the next example, we present the segmentation results of heart chambers as the heart goes through a diastolic and systolic cycle. The time varying 3D MRI heart image data is given on a $256 \times 256 \times 8$ grid. In Fig. 5, we show two cross-sections from the third and sixth data set from our sequence. Again, computation is made to proceed outward from a spherical initialization in the domain. Fig. 6 presents 10 samples of the cycle of a heart beat. The measurements of the area and the volume of each of the segmented heart states are presented in the two graphs in Fig. 7.

6. Concluding remarks

In this paper we presented a complete approach for detecting, representing, and measuring 3D medical data, extending work reported in [36, 27]. The medical object is represented as a weighted minimal surface, following [7, 8]. This is computed from geometric measures on the image, via a fast numerical implementation of the level-set approach as in [1, 24, 27]. The high accuracy of the results of the algorithm, enable us to perform measurements like volume and surface area of the detected objects, as was shown in our experiments. Interested readers are invited to visit the web site http://www.lbl.gov/~malladi for movies depicting the 3D simulations presented in this paper.

Figure 2. The 5th, 35th, and 59th cross-sections from a CT image of human thighs.

Acknowledgements

The research of RM, RK, DA, and JAS are supported in part by the Applied Mathematical Sciences Subprogram of the Office of Energy Research, U.S. Dept. of Energy under Contract DE-AC03-76SD00098. All calculations shown here were performed at University of California at Berkeley and the Lawrence Berkeley National Laboratory.

References

Figure 3. Detection of thighs in $128 \times 128 \times 61$ CT data images. The volume of the soft tissue was measured to be 0.266492.

Figure 4. The left thigh bone of Fig. 3, is extracted and its volume is 0.00844465.

Figure 5. Two cross-sections each from third and sixth heart MRI data set.


Figure 6. 10 stages along one heart beat, 3D segmentation results, presented left to right top to bottom.

Figure 7. Volume and area (upper and lower graphs) along the heart beat cycle in Fig. 6.


