

# A Statistically Based Surface Evolution Method for Medical Image Segmentation: Presentation and Validation

Eric Pichon<sup>1</sup> Allen Tannenbaum<sup>1</sup> and Ron Kikinis<sup>2</sup>

<sup>1</sup> Georgia Institute of Technology, Atlanta GA 30332, USA  
{eric, tannenba}@ece.gatech.edu,  
<http://users.ece.gatech.edu/~eric>

<sup>2</sup> Harvard Medical School, Boston, MA 02115, USA  
kikinis@bwh.harvard.edu

**Abstract.** In this paper we present a new algorithm for 3D medical image segmentation. The algorithm is fast, relatively simple to implement, and semi-automatic. It is based on minimizing a global energy defined from a learned non-parametric estimation of the statistics of the region to be segmented. Implementation details are discussed and source code is freely available as part of the 3D Slicer project. In addition, a new unified set of validation metrics is proposed. Results on artificial and real MRI images show that the algorithm performs well on large brain structures both in terms of accuracy and robustness to noise.

## 1 Introduction

The problem of segmentation, that is finding regions in an image that are homogeneous in a certain sense, is central to the field of computer vision. Medical applications, visualization and quantification methods for computer-aided diagnosis or therapy planning from various modalities typically involve the segmentation of anatomical structures as a preliminary step.

In this paper we will consider the problem of finding the boundaries of only one anatomical region with limited user interaction. Interactivity is very desirable since the user will be given the opportunity to make use of often implicit but absolutely necessary external knowledge to guide the algorithm towards a result that would make sense for her task. The segmentation process can be repeated in order to identify as many different regions as necessary.

Many different approaches have been proposed to address the segmentation problem which can be dually considered as finding regions or finding boundaries. Focusing only on the boundaries is less complex computationally but also less robust since information inside the region is discarded. Typically this is the approach of the snake and active contours variational methods [7, 16, 17].

While the original region-growing algorithm [11] formalism is extremely crude, interesting extensions have been proposed in [9] where some statistical information is derived from the region as it expands. These techniques have been applied to medical image analysis [12, 14]. The relation between region-growing and active contours has been studied in [15] and more recently active contours have been extended to an elegant active regions formalism [8] where the boundaries of regions are deformed according to an evolution equation derived to minimize an energy based on some statistics of the regions.

## 2 Basic Flow

In this section, we state the fundamental flow underpinning the segmentation method. Let  $\Omega$  be an open connected bounded subset of  $\mathbb{R}^n$  with smooth boundary  $\partial\Omega$ . Let  $\psi^t : \Omega \rightarrow \mathbb{R}^n$  be a family of embeddings, such that  $\psi^0$  is the identity. Let  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$  be a positive  $C^1$  function. We set  $\Omega(t) := \psi^t(\Omega)$  and  $S(t) := \psi^t(\partial\Omega)$ . We consider the family of  $\phi$ -weighted volumes

$$\mathcal{H}(t) := \int_{\Omega} \phi(\psi^t(x)) d\psi^t(x) = \int_{\Omega(t)} \phi(y) dy.$$

Set  $X = \frac{\partial\psi^t}{\partial t}|_{t=0}$ . then using the area formula [6] and then the divergence theorem, the first variation is  $\frac{d\mathcal{H}}{dt}|_{t=0} = \int_{\Omega} \text{div}(\phi X) dx. = - \int_{\partial\Omega} (\phi X) \cdot \mathcal{N} dy$ , where  $\mathcal{N}$  is the inward unit normal to  $\partial\Omega$ . Consequently the corresponding  $\phi$ -weighted volume minimizing flow is

$$\frac{\partial S}{\partial t} = \phi \mathcal{N}.$$

A different derivation of the same result has previously been proposed in [4].

## 3 Method

In what follows we will only consider the 3D case. A region  $R$  will be a subset of  $\mathbb{R}^3$  with smooth boundary  $S = \partial R$ . As above,  $\mathcal{N}$  denotes the corresponding inward unit normal vector to  $S$ .

Given an image  $I$ , a non-negative weighting function  $w(\cdot, \cdot)$  and a region  $R$  we define the energy:

$$E(I, w, R) := \int_R w( I(\underline{x}), \|\nabla I(\underline{x})\| ) d\underline{x}. \quad (1)$$

$E$  is the weighted volume of the region  $R$ . The weight of a voxel  $\underline{x}$  is determined by the function  $w(\cdot, \cdot)$  of the local properties  $I(\underline{x})$  and  $\|\nabla I(\underline{x})\|$  of the image. Ideally,  $w$  should reflect the local properties of the region we want to segment. As this is not known *a priori* we will heuristically estimate  $w$  as we evolve  $R$  to maximize  $E$ .

**Proposition 1.** *Notation as above. Then for a given weighting function  $w$ , the evolution in which the energy  $E(I, w, R)$  is decreasing as fast as possible (using only local information) is  $\frac{\partial S}{\partial t} = w \mathcal{N}$ .*

**Proof.** Follows immediately from the discussion in Section 2.  $\square$   
Since  $w$  is a non-negative function, the flow is reversible. In particular, the flow in the reverse direction,

$$\frac{\partial S}{\partial t} = -w \mathcal{N}, \quad (2)$$

gives the direction in which the energy is increasing as fast as possible (using local information). In the context of segmentation, one may think of (2) as a *bubble* and of the original flow as a *snake*.

Given an approximation  $R_0$  of the region to be segmented we can use a maximum likelihood-like approach to determine the weighting function  $w_0$  which would *a posteriori* justify the segmentation of  $R_0$ .

**Proposition 2.** *For a given fixed region  $R_0$ , the energy  $E(I, w, R_0)$  is maximized by setting  $w$  to  $p_0$  the conditional probability on that region:*

$$w_0 = \arg \max_p E(I, w, R_0) = Pr( I(\underline{x}), \|\nabla I(\underline{x})\| \mid \underline{x} \in R_0 ). \quad (3)$$

**Proof.** We can rewrite the energy as:

$$E(I, p, R_0) = \int_I \int_{\|\nabla I\|} N_{R_0}(u, v) \cdot w(u, v) \, dudv,$$

where  $N_{R_0}(u, v)$  is the volume of the set of points  $\underline{x} \in R_0$  such that  $I(\underline{x}) = u$  and  $\|\nabla I(\underline{x})\| = v$ . But this is just a constant multiple of  $Pr(I(\underline{x}), \|\nabla I(\underline{x})\| \mid \underline{x} \in R_0)$  which by the Schwartz's inequality is the maximizer of  $E$ .  $\square$

As the region evolves,  $p$  will be periodically updated according to (3). This will change the definition of the energy (1) and therefore (2) can only be considered a gradient flow for every time interval when  $w$  is fixed.

## 4 Implementation

We implemented our method as a module of the open-source software 3D Slicer. It is freely available for download at <http://www.slicer.org>.

### 4.1 Surface evolution

As the flow (2) is unidirectional (the surface can only expand since  $w \geq 0$ ) any voxel  $\underline{x}$  will eventually be reached at a time  $T(\underline{x})$ . Knowing  $T$  is equivalent to knowing  $R$  or  $S$  since by construction:

$$R(t) = \{ \underline{x}, T(\underline{x}) \leq t \} \quad \text{and} \quad S(t) = \partial R(t). \quad (4)$$

Solving the flow (2) for  $S(t)$  is equivalent to solving the Eikonal equation (5) for  $T(\underline{x})$ :

$$\|\nabla T(\underline{x})\| \cdot w(\underline{x}) = 1. \quad (5)$$

This can be done very efficiently using the Fast Marching method [3]. Starting from known seed points which define the initial surface, the algorithm marches outwards by considering neighboring voxels and iteratively computing arrival times  $T$  in increasing order. The seed points are set by the user inside the structure to be segmented. By construction, when computing  $T(\underline{x})$ , the surface contains the voxel  $\underline{x}$  as well as all voxels for which  $T$  has already been computed. The algorithm will terminate when  $T$  is known for all points and using (4) we will know  $S(t)$  for any  $t$ . We will then let the user determine what time  $t_0$  of the evolution corresponds best to the region she wants.

Note that with a very different formalism, our method is, in its implementation, very reminiscent of region growing. For example, the min-heap data structure which makes Fast Marching efficient is the direct equivalent of the sequentially sorted list in the seeded region growing algorithm [9]. In fact our algorithm could be made a direct non-parametric extension of seeded region growing simply by artificially forcing arrival times to zero for all points inside the surface  $S$ . Relations between region growing and variational schemes have been previously exposed in [15].

### 4.2 Estimation of probability density function

The probability has been modified to  $p = p_M(m) \cdot p_H(h)$  where  $M$  and  $H$  are the median and interquartile range (the difference of between the first and last quartile) operators on a  $3 \times 3 \times 3$  neighborhood.  $M$  and  $H$  convey more or less the same information as  $I$  (gray level) and  $\|\nabla I\|$  (local homogeneity) but their non-linear nature makes them more robust to noise and allow them to respect better the edges of the image.

We use Parzen windows [10] to estimate the probability density functions. It is a non-parametric technique and therefore no assumption is required on the shape

of the distributions. Given a window function  $\phi$  and  $N$  samples  $m_1, \dots, m_N$  and  $h_1, \dots, h_N$  the densities are estimated by:

$$p_M(m) = \frac{1}{N} \sum_{i=1}^N \phi(m - m_i) \quad \text{and} \quad p_H(h) = \frac{1}{N} \sum_{i=1}^N \phi(h - h_i)$$

## 5 Validation

Objective and quantitative analysis of performance is absolutely crucial (but often overlooked) when proposing a segmentation algorithm. Since designing a segmentation method is challenging (lack of unifying formalism, high diversity in the applications, subjectivity, implicitness, etc.) it does not come as a surprise that the validation of such an algorithm will also be challenging. Different methods have been studied (see [20] and references therein). We will propose a unifying framework for discrepancy measures based on the number and the position of mis-segmented voxels and show how it relates to classical measures. We will then apply it to the validation of segmentation of realistic synthetic images (for which the ‘‘ground truth’’ i.e. perfect segmentation is known) at different levels of noise for accuracy and robustness assessment as well as to manual expert segmentation of real datasets.

### 5.1 Classical discrepancy measures

Different measures have been proposed to assess the resemblance between a proposed segmentation  $S$  and the corresponding ground truth  $G$ . The Dice Similarity Coefficient has been widely used and it can be derived as an approximation of the kappa statistic (see [1]). It is defined as:

$$DSC(S, G) := \frac{V(S \cap G)}{\frac{1}{2}(V(S) + V(G))}$$

Where  $V(\cdot)$  is the volume (number of voxels) of a set.

One disadvantage of this coefficient is that it only takes into account the number of mis-segmented pixels and disregards their position and therefore the severities of errors. This was corrected in Yasnoff’s normalized discrepancy measure (ND, see [18]) and the Factor of Merit (FOM, see [5]):

$$ND := \frac{1}{N} \sum_{i=1}^N d(i)^2 \quad \text{and} \quad FOM_e := \frac{1}{N} \sum_{i=1}^N \frac{1}{1 + d(i)^2}$$

Where  $N$  is the number of mis-segmented voxels and  $d(i)$  is the error on the  $i^{\text{th}}$  voxel. Another popular measure is the Hausdorff distance:

$$H(S, G) := \max \left\{ \max_{s \in S} \min_{g \in G} \|s - g\|, \max_{g \in G} \min_{s \in S} \|s - g\| \right\}$$

$H(S, G)$  is the maximum distance we would have to move the boundaries of one set so that it would encompass completely the other set. As this is extremely sensitive to extreme errors, the partial Hausdorff distance  $H_f(S, G)$  can be introduced (see [2]) as the maximum distance we would have to move the boundaries of one set so that it would cover  $f\%$  of the other set.

## 5.2 Proposed framework

Consider now the error-distance:

$$d(\underline{x}) := \begin{cases} 0 & \text{for } \underline{x} \text{ correctly segmented } (\underline{x} \in S \cap G) \\ \min_{s \in S} \|\underline{x} - s\| & \text{for } \underline{x} \text{ under-segmented } (\underline{x} \in G \setminus S) \\ \min_{g \in G} \|\underline{x} - g\| & \text{for } \underline{x} \text{ over-segmented } (\underline{x} \in S \setminus G) \end{cases}$$

Assuming that all points  $\underline{x} \in S \cup G$  are equally likely  $d$  can be seen as a random variable  $D$  which describes completely the discrepancy between  $S$  and  $G$ . We can study  $D$  using the standard statistical tools:

$$\text{probability of error: PE} := Pr(D > 0)$$

$$\text{mean error: } \mu_{D>0} := \text{mean}(D \mid D > 0)$$

$$\text{standard deviation of error: } \sigma_{D>0} := \text{stdev}(D \mid D > 0)$$

$$\text{partial distance-error: } D_f := f - \text{quantile}(D)$$

These measures receive a natural intuitive interpretation.

- PE is the probability for a voxel  $\underline{x} \in S \cap G$  to be misclassified (either over- or under-segmented).
- An erroneous voxel is on average  $\mu_{D>0}$  pixels off. This value is or is not typical depending on the standard deviation  $\sigma_{D>0}$ .
- $D_{1-f}$  is the error distance of the worst  $f\%$  voxels or equivalently the maximum distance we would need to move erroneous voxels for the error to be improved to PE =  $f$ .

As an example, PE = 10%,  $\mu_{D>0} = 3.1$ ,  $\sigma_{D>0} = 0.3$  and  $D_{0.99} = 14$  would mean that the overlap between the ground truth and the proposed segmentation is 90%. The 10% remaining pixels are either under-segmented or over-segmented pixels (“false positive” i.e. pixels that are in  $S$  and not in  $G$ ). On average these pixels are 3.1 pixels off. This value is very typical since the standard variation is low (0.3). However there is no reason for the error to be Gaussian and, here, the tail probability is not negligible since the worst 1% pixels are at least 14 pixels off. This could be due to a thin, long finger of mis-segmented pixels.

The following proposition justifies the definition of these new unified measures.

**Proposition 3.** *These measures are related to the measures presented in Section 5.1 according to:*

$$1 - DSC \leq PE = (1 - DSC) / (1 - \frac{DSC}{2}) \quad (6)$$

$$\frac{1}{FOM_e} - 1 \leq (\mu_{D>0}^2 + \sigma_{D>0}^2) = ND \quad (7)$$

$$H_{1-f/(1-PE)} \leq D_{1-f} \leq H_{1-\frac{f}{2}} \quad (8)$$

**Proof.** in future publication (in particular,  $D_1 = H$ ) □

## 5.3 Results on simulated datasets

The publicly available *Brain Web* [19] datasets have been generated from a known ground truth using a sophisticated physical modeling [13] of the MRI process. We can assess in a perfectly objective way the performance of our method by comparing the result of our segmentation with the underlying ground truth. Note that even though these datasets are computer-generated they are very realistic (see figure 1(b)) Another interesting aspect of this project is that from

the same ground truth, datasets with different levels of noise can be simulated which allows us to study the robustness of our method with respect to noise. We segmented the lateral ventricle, white matter (WM) and white matter and gray matter (WM+GM) on 2 datasets:

- Normal brain, T1,  $1 \times 1 \times 1$  mm ( $181 \times 181 \times 217$  voxels), 3% noise, 20% intensity non-uniformity ("RF") (standard parameters of the *Brain Web* model).
- Normal brain, T1,  $1 \times 1 \times 1$  mm ( $181 \times 181 \times 217$  voxels), 9%, 40% (highest levels of noise available).

Our results (Table 1) show that the proposed algorithm gives very good results on these structures ( $DSC > 0.7$  has been described as a good agreement in the literature, see for example [1]). The complex structure of the white matter makes it more challenging and explains the somewhat mediocre performance (in the case of the maximum noise dataset, the cerebellum was not perfectly segmented). In the highest level of noise, connectivity between the lateral and the third ventricles was lost (the intraventricular foramen of Monro disappeared in the noise). This increased the strength of the ventricle edges in the noisy dataset and, paradoxically, simplified the segmentation. Overall the algorithm appears extremely robust to noise.

	DSC		PE		$\mu_{D>0}$		$\sigma_{D>0}$		$D_{0.95}$		$D_{0.99}$	
Ventricle	<b>92.0%</b>	<u>95.1%</u>	<b>14.9%</b>	<u>9.4%</u>	<b>1.07</b>	<u>1.13</u>	<b>0.48</b>	<u>0.61</u>	<b>1.00</b>	<u>1.00</u>	<b>1.00</b>	<u>1.41</u>
WM	<b>91.9%</b>	80.3%	<b>15.0%</b>	32.0%	<b>1.59</b>	2.03	<b>1.58</b>	1.94	<b>1.00</b>	2.83	<b>3.61</b>	8.25
WM+GM	<b>96.2%</b>	95.2%	<b>7.4%</b>	9.2%	<b>1.42</b>	1.40	<b>1.25</b>	1.15	<b>1.00</b>	1.00	<b>1.41</b>	2.00

**Table 1.** Performance measure on artificial dataset. Left bold, with standard noise, right, with maximum noise. Underlined results are illustrated by figures 1(b),1(d),1(f).

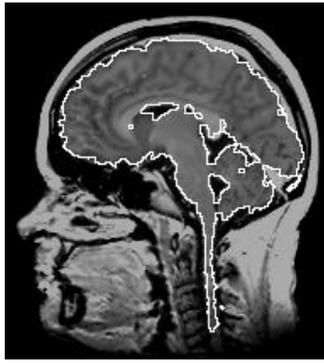
#### 5.4 Results on real datasets

In this real case, the pathological diagnoses are meningiomas (brain tumor). Patients' heads were imaged in the sagittal and axial plane with a 1.5 T MRI system<sup>3</sup> with a postcontrast 3D sagittal spoiled gradient recalled (SPGR) acquisition with contiguous slices. The resolution is  $0.975 \times 0.975 \times 1.5$  mm ( $256 \times 256 \times 124$  voxels). These datasets were manually segmented by one expert. Because of inter- and intra-expert variability we should expect these results *not* to be as good as in the synthetic case. It should also be noted that the arbitrary conventions of the manual segmentations are responsible for a lot of the observed error since for example the ventricle was labeled as gray matter, the medulla oblongata and the spinal cord have been left out etc. (compare Fig. 1(a) and 1(c)). Overall, nonetheless, results are consistent with the artificial case.

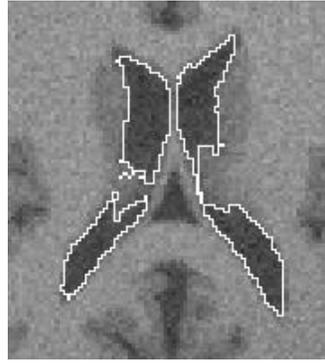
	DSC		PE		$\mu_{D>0}$		$\sigma_{D>0}$		$D_{0.95}$		$D_{0.99}$	
Tumor	78.0%	88.0%	36.0%	21.4%	1.97	1.34	1.63	0.94	3.32	1.41	7.00	2.83
WM+GM	<u>96.1%</u>	<u>92.4%</u>	<u>7.5%</u>	14.2%	<u>1.69</u>	1.28	<u>1.99</u>	0.75	<u>1.00</u>	1.00	<u>2.00</u>	2.24

**Table 2.** Performance measure on 2 real datasets. Underlined results are illustrated by figures 1(a),1(c),1(e).

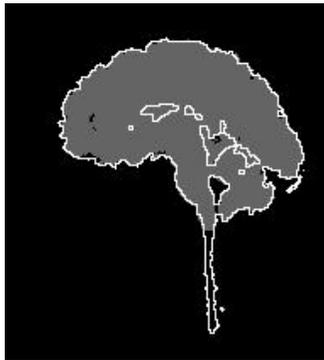
<sup>3</sup> Signa, GE Medical Systems, Milwaukee, WI.



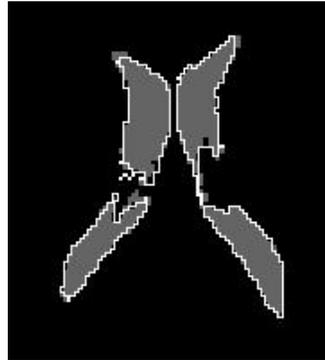
(a) Sagittal slice of real dataset and proposed white and gray matter segmentation (white)



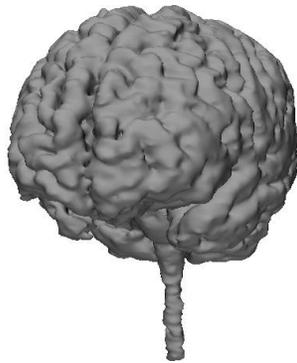
(b) Axial slice of noisy artificial dataset and proposed ventricle segmentation (white)



(c) Expert segmentation (gray) and proposed white and gray matter segmentation (white)



(d) Underlying ground truth (gray) and proposed ventricle segmentation (white)



(e) Rendered surface of proposed white and gray matter segmentation



(f) Rendered surface of proposed ventricle segmentation

## 6 Conclusion

We presented a new curve evolution flow based on learned non-parametric statistics of the image. Implementation is straightforward and efficient using the Fast Marching algorithm and is freely available as part of the 3D Slicer project. An extensive validation study as well as a new unified set of validation metrics have also been proposed.

Future work will focus on extending our formalism into a purely variational framework, adding some regularizing constraints and extending the validation study.

## Acknowledgements

Eric Pichon and Allen Tannenbaum are supported by NSF, NIH, AFOSR, MRI-HEL and ARO.

Ron Kikinis is supported by grants PO1 CA67165, R01EB000304 and P41RR13218.

## References

1. Zijdenbos A., Dawant B., and Margolin R. Morphometric analysis of white matter lesions in MR images: Method and validation. *IEEE TMI*, 13(4):716–724, 1994.
2. Huttenlocher D., Klanderma G., and Rucklidge W. Comparing images using the Hausdorff distance. *PAMI*, 15(9):850–863, 1993.
3. Sethian J. *Level Set Methods and Fast Marching Methods*. Cambridge University Press, 1999.
4. Siddiqi K., Lauziere Y., Tannenbaum A., and Zucker S. Area and length minimizing flows for shape segmentation. *IEEE TMI*, 7:433–443, 1998.
5. Strasters K. and Gerbrands J. Three-dimensional segmentation using a split, merge and group approach. *Pattern Recognition Letters*, 12:307–325, 1991.
6. Simon L. Lectures on geometric measure theory. In *Proceedings of the Centre for Mathematical Analysis, Australian National University, Canberra*, 1983.
7. Kass M., Witkin A., and Terzopoulos D. Snakes: Active contour models. *Int. J. Computer Vision*, 1:321–332, 1988.
8. Paragios N. and Deriche R. Geodesic active regions: A new paradigm to deal with frame partition problems in computer vision. *Journal of Visual Communication and Image Representation*, 13:249–268, 2002.
9. Adams R. and Bischof L. Seeded region growing. *PAMI*, 16(6):641–647, 1994.
10. Duda R., Hart P., and Stork D. *Pattern Classification*. Wiley-Interscience, 2001.
11. Gonzalez R. and Woods R. *Digital Image Processing*. Prentice Hall, 2001.
12. Justice R., Stokely E., Strobel J., Ideker R., and Smith W. Medical image segmentation using 3-D seeded region growing. *Proc. SPIE Symposium on Medical Imaging Volume*, 3034:900–910, 1997.
13. Kwan R., Evans A., and Pike G. MRI simulation-based evaluation of image-processing and classification methods. *IEEE TMI*, 18(11):1085–1097, 1999.
14. Pohle R. and Toennies K. Segmentation of medical images using adaptive region growing. In *Proc. SPIE Medical Imaging*.
15. Zhu S. and Yuille A. Region competition: Unifying snakes, region growing, and bayes/MDL for multiband image segmentation. *PAMI*, 18(9):884–900, 1996.
16. McInerney T. and Terzopoulos D. Deformable models in medical image analysis: A survey. *Medical Image Analysis*, 1(2):91–108, 1996.
17. Caselles V., Kimmel R., and Sapiro G. Geodesic active contours. In *Proc. ICCV*, pages 694–699, 1995.
18. Yasnoff W., Miu J., and Bacus J. Error measures for scene segmentation. *Pattern Recognition*, 9:217–231, 1977.
19. World Wide Web. <http://www.bic.mni.mcgill.ca/brainweb/>.
20. Zhang Y. A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29(8):1335–1346, 1996.