Nonlinear Diffusion Scale-Space and Fast Marching Level Sets for Segmentation of MR Imagery and Volume Estimation of Stroke Lesions

Jerod Weinman[1], George Bissias[1], Joseph Horowitz[2], Edward Riseman[1], and Allen Hanson[1]

[1] Dept. of Computer Science, Univ. of Massachusetts-Amherst
http://vis-www.cs.umass.edu
[2] Dept. of Mathematics and Statistics, Univ. of Massachusetts-Amherst
http://www.math.umass.edu

Abstract. We combine nonlinear diffusion scale-space and geometric deformable models for segmenting lesions in MR images of ischemic stroke patients. Region and boundary information are integrated in a speed function for robust segmentation with the fast marching level set method. A confidence-based model of segmentation captures the significant variability in human segmentation and the ambiguity inherent in many lesions, and it provides a testbed for a new measure of variance with sets as random variables. This method offers users a family of segmentations, requires less user input than previous methods, and its volume estimates effectively match those of doctors' hand segmentations.

1 Introduction

Treatment for stroke requires almost immediate intervention, and quick, consistent computer segmentations would be useful for evaluating alternative therapies [1]. Given an MRI of a stroke victim, the primary task is to identify voxels (volume elements) clinically deemed to be lesion. According to physicians, volume estimates within 20% of physician manual segmentations are clinically useful. The primary difficulty is that the segmentation occurs over a noisy, indirect observation, rather than directly from the generating process (i.e. tissue). Other difficulties include tissue mixtures, ambiguous boundaries, multi-modal lesion intensity, and complex shape.

Creating and evaluating automatic methods for segmenting lesions is challenging for two major reasons. First, the concept of a ground-truth segmentation is elusive. Different doctors may give substantially different segmentations of the same lesion; the same doctor on different occasions might produce disparate segmentations; a doctor might not even exhibit a consistent strategy throughout all slices of an MRI. Ambiguity in a lesion boundary is one cause of this segmentation variation, while some other lesions are so convoluted that accurate manual segmentation is prohibitively tedious.

The second challenge is that lesions themselves vary quite widely. Much of the finesse in diagnosing tissue as stroke lesion relies upon anatomical knowl-

edge, which can be difficult to incorporate in automatic methods. Some imaging methods show distinct tissue types differently, requiring an allowance for multiple intensities within a lesion. In other cases, doctors know that stroke does not stop at cortical boundaries, so even the faintest signal in regions neighboring the primary affected area are likely to be labeled stroke, whereas in other lesions a similar drop in intensity might signal a boundary with healthy tissue or an imaging artifact. The result is that simple rules relying on intensity patterns are inconsistent across lesions, even on the small dataset explored in this work.

The simplest solution is to find a consistent strategy that gives the least error. This works reasonably well on average but is unsatisfying on lesions where there is ambiguity or subtle anatomical knowledge is required. Alternatively, one might opt for the overhead of image registration and attempt to tie in more medical domain knowledge in hopes of eliminating the error. We argue, however, that with minimal additional user interaction, segmentations of these subtleties can be proposed automatically and confirmed manually.

The intensity nuances found in stroke lesions are well-modeled by the scale-space theory of computer vision, which facilitates the inspection of an image at varying levels of detail. For segmenting both complex and regular shapes quite easily, we employ geometric deformable models, or level sets. These require a single click to initialize and inflate like balloons that fill to lesion boundaries. We give a measure of confidence in the classification of each voxel by segmenting the image at several scales and with several threshold parameters that stop the inflation of the deformable model at different places.

2 Segmentation Model

2.1 Nonlinear Diffusion Scale-Space

We desire to model the segmentation of lesions in a framework that encapsulates the variation in lesion boundary criteria. Scale-space is a formal method of blurring an image so that details disappear as the scale increases. Our goal is to localize lesion boundaries, so we use the semantically meaningful (i.e. edge-preserving) scale-space provided by nonlinear diffusion [2]. Edges are preserved and enhanced by encouraging blurring within regions while restricting blurring between them. The regularized [3] nonlinear isotropic diffusion equation governing this scale-space evolution is

$$\partial_t u = \text{div}\left(g\left(|\nabla u_\sigma|^2\right)\nabla u\right),\tag{1}$$

where $u\left(\mathbf{x}; t\right)$ is the image at scale t, $u\left(\mathbf{x}; 0\right)$ is the original image, and u_σ is u convolved with a Gaussian of standard deviation σ. We use the diffusivity [4]

$$g\left(s\right) = 1 - \exp\left(-3.315\left(s/\lambda\right)^{-4}\right),\tag{2}$$

where λ is a contrast parameter so that structures with $|\nabla u_\sigma| > \lambda$ are preserved, and those with $|\nabla u_\sigma| \leq \lambda$ are blurred. Since $g\left(\cdot\right) > 0$, regions of the image slowly blend together as the scale t increases, thus gradually eradicating edges.

2.2 Fast Marching Level Set Method Speed Function

The implicit nature of the fast marching level set method allows it to effectively segment lesions of complex shape. A front propagates outward from an initial point within the lesion at some speed $F(\mathbf{x})$. This front's arrival time at any image location is given by $T(\mathbf{x})$, which is defined [5] as the solution to

$$F\,|\nabla T| = 1\,.\tag{3}$$

The process is complete when the speed function prevents the front from reaching any more points within a reasonable time, and the segmentation is given by classifying as lesion those voxels that have been reached.

Because F determines the speed of the propagating front at every image location, it makes sense to assign F a large value at voxels believed to be part of the lesion and a small value elsewhere. Two of the best clues we have regarding the classification of a particular voxel are its intensity and gradient magnitude. Intensities consistent with healthy tissue or boundaries indicated by large gradient magnitude are good clues for where to stop the front. We combine this information into a speed function that is a hybrid of region (intensity) and boundary (gradient) based segmentation strategies.

Diffusion will not always blend a lesion to a uniform region because multi-modal lesions may have relatively dark areas that blend with the background first. However, if we have a strict lower bound on lesion intensity, i.e. that of healthy tissue, we can introduce a threshold β, such that the front slows when it encounters intensities below the threshold. Such an intensity-based speed function is given by

$$H(\mathbf{x}) = \exp\left(-\left(\frac{\beta}{u(\mathbf{x};t)}\right)^{a}\right)\,.\tag{4}$$

where $u(\mathbf{x};t)$ is the image intensity at diffusion scale t, and a is a parameter controlling slope of the speed about the intensity threshold β, which we assume some confidence in. To reduce the front's speed at a strong gradient, which potentially indicates a lesion boundary, we use

$$G(\mathbf{x}) = \exp\left(-b\,|\nabla u(\mathbf{x};t)|\right)\,,\tag{5}$$

where b is only a linear parameter controlling the sensitivity to gradient magnitude, which we have little confidence in. Our hybrid speed function is then

$$F(\mathbf{x}) = H(\mathbf{x})\,G(\mathbf{x}) = \exp\left(-\left(\frac{\beta}{u(\mathbf{x};t)}\right)^{a} - b\,|\nabla u(\mathbf{x};t)|\right)\,.\tag{6}$$

This function allows large speeds where the intensity on the front is above the threshold and there is little boundary evidence but slows when either the intensity drops or a large gradient is encountered. The function approaches zero rapidly when both events occur (Figure 1). Our hybrid speed function allows for multi-modal lesions to be segmented correctly by not relying solely on gradient strength and prevents a front from propagating outside the lesion when there is no strong boundary.
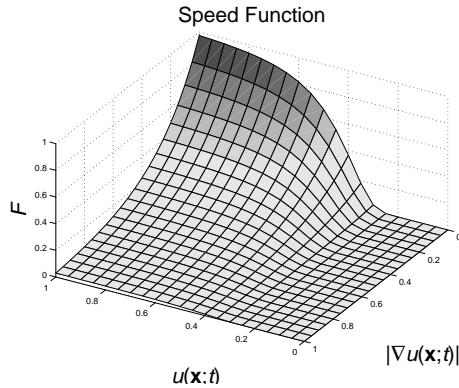
Fig. 1. Hybrid gradient and intensity speed function ($\beta = 0.4$, $a = 4$, $b = 3.5$).
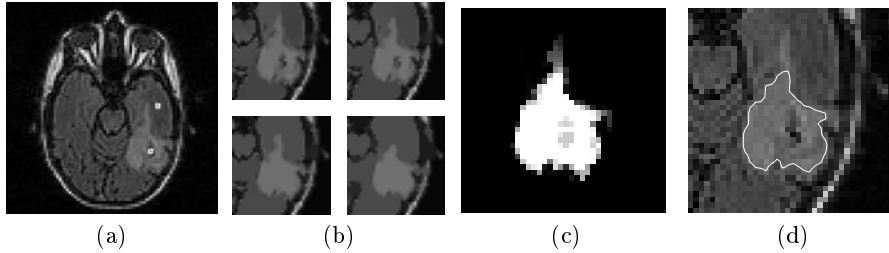
2.3 Confidence

Different lesions exhibit different intensity patterns and thus often require different parameters for correct segmentation; we don't know a priori what scale t and intensity threshold β are best for segmenting a particular lesion. We therefore segment an MR image at several scales and intensity thresholds. In this two-dimensional space of control parameters, the frequency of a voxel being labeled lesion by the algorithm will be regarded as a one-sided measure of confidence in the classification of that voxel. As the scale or the threshold is increased, only the most persistent or brightest (darkest) areas of the lesion (background) are included (excluded) across segmentations. Thus, the more often a voxel is included, the more certain the algorithm, without prior anatomical knowledge, can be about the results.

The segmentation process begins when a user selects a seed point inside the lesion and an exterior point somewhere in healthy tissue, giving a lower bound on lesion intensity (Figure 2). While the locations of these points remain fixed throughout the segmentation process, the intensities at the points vary with scale. The intensity threshold is implicitly varied between the intensities of the given interior and exterior points. Let $\beta\left(0;t\right)$ be the intensity of the exterior point at scale t and $\beta\left(1;t\right)$ be the intensity of the interior seed point at scale t. The explicit segmentation parameter we vary is $p \in [0,1]$ and we use

$$\beta\left(p;t\right) := \beta\left(0;t\right) + p\left(\beta\left(1;t\right) - \beta\left(0;t\right)\right) \tag{7}$$

for $0 < p < 1$ as the speed function intensity threshold parameter. As p approaches 1, the threshold increases, making segmentations more conservative. For the most conservative possible segmentation, the interior seed point should lie in the lesion's brightest area.

If $\mathcal{T}$ and $\mathcal{P}$ are the (finite) sets of scales and threshold parameters, respectively, then let a particular segmentation be $S\left(t,p\right)$ for $t \in \mathcal{T}$ and $p \in \mathcal{P}$. The

Fig. 2. Segmentation Process. (a) Interior and exterior points selected. (b) Segmentation over several scales ($t = 37$, 221, 493, 4447 shown) and thresholds. (c) Segmentation inclusion frequency. (d) Contour for confidence $c = 0.71$, which most closely matches doctors' volumes.

relative frequency of a voxel's inclusion in the lesion is

$$\phi\left(\mathbf{x}\right) = \frac{1}{|\mathcal{T}||\mathcal{P}|} \sum_{t \in \mathcal{T}} \sum_{p \in \mathcal{P}} \chi_{S(t,p)}\left(\mathbf{x}\right) , \tag{8}$$

which is the total number of times a voxel $\mathbf{x}$ appears in the segmentations divided by the number of segmentations. We define a confidence segmentation to be the set of voxels above some minimum relative frequency $0 \leq c \leq 1$,

$$C\left(c\right) = \left\{\mathbf{x} \mid \phi\left(\mathbf{x}\right) \geq c\right\} . \tag{9}$$

3   Experiments

Experiments were performed on 7 MR images of various modes and voxel dimensions from 5 patients. The volumes from two hand segmentations by two physicians are reported for 5 of these images in [1]. For 3 additional images, the segmentations themselves were obtained, thus allowing a more careful evaluation of automatic segmentations by directly comparing voxel classifications.

Three metrics are used to evaluate the results. As in [1], we compare the volumes of automatic and physician segmentations using the relative error

$$\text{Error} = 100 * \frac{\text{Volume}_{\text{Estimate}} - \text{Volume}_{\text{Physician}}}{\text{Volume}_{\text{Physician}}} . \tag{10}$$

However, similar volumes might correspond to different segmentations (sets of voxels). Where possible, we measure the precision P (percentage of included voxels that are correct) and recall R (percentage of correct voxels that are included) of the estimate with respect to a physician segmentation and consider their harmonic mean, $F = 2PR/\left(P + R\right)$, known as the F-measure.

To test parameter sensitivity, some measure of segmentation variance is needed, but the usual variance of volumes is susceptible to the same problem

Table 1. For doctors 1 and 2, the Individual column gives the minimum volume error on each image and the parameters where it occurs; the Total column gives each image's error at the parameters where minimum total absolute volume error occurred ($t = 6634$, $p = 0.95$). Maximum F-measures and corresponding volume error for each image are given for doctor 3.

	Doctor 1				Doctor 2				Doctor 3			
	Individual			Total	Individual			Total	Individual			
Img	%Err	Scale	Thresh	%Err	%Err	Scale	Thresh	%Err	F-msr	%Err	Scale	Thresh
2	-0.1	1636	0.95	1.8	0.0	40135	0.95	-3.9	84.3	-3.9	1998	0.45
3P	-	-	-	-	-	-	-	-	88.1	-10.9	330	0.35
3Q	-	-	-	-	-	-	-	-	88.1	-10.8	3641	0.15
3R	0.0	330	0.95	-15.3	0.0	1998	0.85	-6.6	-	-	-	-
4	0.7	8103	0.95	1.6	0.4	8103	0.85	-6.2	-	-	-	-
5	0.0	270	0.75	-0.5	-0.4	45	0.75	7.5	-	-	-	-
6	0.0	55	0.95	-4.6	0.5	4447	0.75	-6.0	-	-	-	-

mentioned above. The variance of a numerical sample $y_1, \ldots, y_n$ may be written $s^2 = \frac{1}{2n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} (y_i - y_j)^2$. Since our segmentation sample $S_1, \ldots, S_n$ consists of sets rather than numbers, we propose to use the analogous measure of "set variance"

$$s_{\ominus}^2 := \frac{1}{2n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} |S_i \ominus S_j|^2 \, , \tag{11}$$

where $S_i \ominus S_j$ is the symmetric difference $(S_i \cup S_j) \setminus (S_i \cap S_j)$. When sets in a sample all have nearly the same volume but contain distinct elements, the usual standard deviation of volume will be misleadingly small, while (11) will be large. As an example, for n disjoint segmentations each of volume v, $s^2 = 0$ while $s_{\ominus}^2 = \frac{2n}{n-1} v^2$. It can be shown that $s_{\ominus}^2 \geq s^2$ and $s_{\ominus}^2 = s^2$ iff either $S_i \subseteq S_j$ or $S_j \subseteq S_i \, \forall i, j$ [6].

The results of Table 1 show that it is possible to achieve almost zero volume error for each image and each physician at some parameter setting. If we fix the scale and threshold parameters to the pair that minimizes the total absolute error over both of the first two doctors' segmentations, the results are still almost entirely within a 10% error. Table 1 also gives the best F-measure on images with physician segmentations. For these images, the best automatic segmentations are similar in appearance to those given by physicians. All volume differences are well within our clinically useful bounds of 20%.

We tested the sensitivity of the segmentation process to the location and intensity of the interior point (Table 2) by using several seed points chosen at random from a conservative segmentation. Each image was segmented with the individual parameters from Table 1 for doctor 1 or doctor 3. While the error is within clinically useful bounds on average, when the seed point comes from darker regions of multi-modal lesions but p remains fixed, the threshold parameter β changes, which alters the segmentation and increases standard deviation. By holding β constant, we see that nearly all the variation is due to inten-

Table 2. LEFT: Segmentation volumes (in mm^3) and the percentage error of the mean using several initial points. RIGHT: Individual errors at the confidence ($c = 0.53$) with zero mean error (standard deviation 22.0%), and confidence values where the percentage error is minimum for each individual image with corresponding error.

	Varying β				Controlled β				Zero Mean	Individual	
Img	Mean	Sd	Set Sd	Mean %Err	Mean	Sd	Set Sd	Mean %Err	%Err	Conf	%Err
2	3710	922	922	8.6	3420	4	4	0.11	8.6	0.67	-0.1
3P	28534	993	993	-12.2	28948	37	37	-10.96	-30.1	0.02	-2.2
3Q	27578	111	111	-10.5	27502	40	40	-10.69	-23.6	0.05	0.6
3R	21701	5882	5882	-7.9	23437	82	84	-0.50	8.7	0.71	0.2
4	11164	1833	1833	12.1	10030	8	8	0.68	35.2	0.91	-0.2
5	10652	495	495	5.1	10132	24	24	0.01	- 4.9	0.45	-0.1
6	2880	1031	1031	-16.6	3458	4	4	0.16	5.5	0.76	-0.1
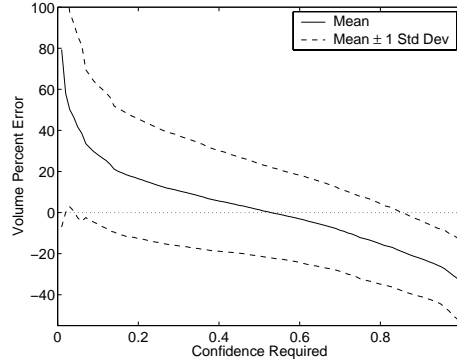
Fig. 3. Mean volume error for all images versus confidence.

sity, rather than location. There is little difference between the usual standard deviation over volumes and our proposed definition over sets because most segmentations are nested, indicating the sets vary only as much as the volumes. Accuracy is sensitive to the intensity threshold, which is difficult to control since it is derived from user input. We therefore use the confidence measure to consider several thresholds.

Figure 3 confirms the intuition that over-segmentation occurs when low confidence is required and under-segmentation occurs when high confidence is required. Therefore, we might set a default confidence requirement at the zero crossing $c = 0.53$, where the average error over images is zero. That the confidence giving zero mean error is almost exactly one half is intuitively appealing. Additionally, the standard deviation of volume error is 22.0%, barely above the error deemed clinically useful by physicians.

Although they average to zero, the individual errors in Table 2 indicate that a single confidence level for all lesions might not be the best option, especially for the variety of lesions expected to occur in practice. For example, the large negative errors on images 3P and 3Q at $c = 0.53$ are because both contain ambiguous regions that doctors classified as lesion. These areas of uncertainty are only included at lower confidence levels, and similar uncertain structures will always be eliminated at the "average" confidence level. Conversely, image 4 contains a mix of healthy and lesion tissue that doctors excluded from their lesion segmentations, while the confidence segmentation only excludes the mix and isolates the correct region at higher confidence levels.

4 Conclusions

We have introduced a model for segmenting MR lesion images that requires little user input. Clinically useful volumes are achieved on most lesions even when all parameters are fixed, but some types of lesions do not lend themselves to accurate segmentation with a single parameter setting. Rather than ignoring these anatomical nuances, we allow a user to choose a segmentation corresponding to the confidence level he or she feels is the best representative of actual lesion boundaries, giving more precise volumes. This requires still less user interaction than [1], which implicitly required the user to choose the scale of the segmentation by drawing an initial contour.

The statistical analysis of random sets is a relatively unexplored area. We presented a new direction for thinking about the variability in segmentations by proposing a measure of set variance that is near the value of the usual variance of volume measurements when the segmentations are similar yet is much larger when they are dissimilar.

References

1. D. Lisin, E. Riseman, G. Whitten, D. Geman, J. Horowitz, and B. Stein. Development of statistical and computer vision techniques to support the clinical study of ischemic stroke treatment. Technical Report UM-CS-2001-019, University of Massachusetts-Amherst, 2001.
2. P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.
3. F. Catté, P.L. Lions, J.M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. Numer. Anal.*, 29(1):182–193, 1992.
4. J. Weickert, B.M. ter Haar Romeny, and M.A. Viergever. Efficient and reliable schemes for nonlinear diffusion filtering. *IEEE Trans. Image Processing*, 7(3):398–410, 1998.
5. J.A. Sethian. A fast marching level set method for monotonically advancing fronts. *Proc. National Academy of Sciences*, 93(4):1591–1595, 1996.
6. J. Weinman, G. Bissias, E. Riseman, A. Hanson, and J. Horowitz. Confidence-based segmentation of MR imagery using region and boundary information with nonlinear scale-space and fast marching level sets. Technical Report UM-CS-2003-017, University of Massachusetts-Amherst, 2003.