Statistical and Deformable Model Approaches to the Segmentation of MR Imagery and Volume Estimation of Stroke Lesions

Benjamin Stein[1], Dimitri Lisin[2], Joseph Horowitz[1], Edward Riseman[2], and Gary Whitten[2]

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

Abstract. We propose two 3D methods to segment magnetic resonance imagery (MRI) of ischemic stroke patients into lesion and background, and hence to estimate lesion volumes. The first is a hierarchical, regularized method based on classical statistics that produces a rigorous confidence interval for lesion volume. This approach requires a limited amount of user interaction to initialize, but this step can be time-consuming. The second method integrates the first into the deformable models framework. This hybrid approach combines intensity-based information provided by the statistical method and shape-based information given by the deformable model. It also requires less initialization than the statistical method. Both procedures have been tested on real MR data, with volume estimates within 20% of those derived from doctors' hand segmentations. According to the physicians with whom we are working, these results are clinically useful to evaluate stroke therapies.

1 Introduction

In evaluating therapies for ischemic stroke patients, many physicians are interested in finding consistent, reliable estimates of lesion volume from MR images. We introduce two new methods to segment three-dimensional (3D) images into lesion and background, and thus to estimate lesion volumes. In section 2 we present the first procedure, called "packing," which is a hierarchical, regularized method based on classical statistics. Several other research groups have used statistical approaches to segment tissue types in MRI (see [1, 8, 12, 13], e.g.), with varying degrees of user interaction, but with no single method emerging as superior. As in those methods, we are concerned with producing consistent estimates with limited user interaction, but our procedure goes beyond them in producing an assessment of the *error* of our estimate, in the form of a rigorous confidence interval for lesion volume. While this idea has been explored in [11] for near-infrared imaging, we do not know of any such results for MRI.

The packing method requires sometimes time-consuming user interaction to initialize, and uses only the statistical information in the image. By integrating

packing with the well-known deformable models ("snakes") framework [6], we reduce user interaction and take advantage of of geometric as well as statistical information in the image. Our second procedure (see section 3) is therefore a 3D "hybrid" method that uses a version of packing as an additional external energy term. Also, the addition of statistical information into the deformable models framework helps to suppress some of the limitations of snakes, including sensitivity to strong gradients produced by other nearby objects and the inability to adjust to large changes in lesion size from one slice to another.

We have tested the new procedures on actual MRI, and found that both methods are consistently estimating lesion volumes to within 20% of those derived from doctors' hand segmentations. According to the doctors, these results (see section 4) are clinically useful.

2   The Packing Method

We assume the lesion is imaged as a "bright spot" lesion, which means that the mean lesion intensity μ_L is higher than the mean of the intensities of any other tissue type in the region of interest (ROI). The diffusion-weighted pulse sequence shows an ischemic stroke lesion as the only bright object, so that the ROI can be the entire set of imagery. However, for pulse sequences such as T_2 and FLAIR, the user must extract the ROI manually.

Aside from choosing a ROI, the only other manual step necessary to initialize the process is to sample the data. The user chooses a "base slice" from the stack of two-dimensional images, and two regions bounded by closed contours in that slice: one region completely inside the lesion and the other in the background. The pixels inside these contours constitute the lesion and background samples that our statistical analysis will be based upon. This step tupically requires less than a minute of user interaction.

2.1   Coarse-Grid Segmentation

The method continues with no more outside assistance. We will describe this multi-step procedure by first explaining the mechanics of it at step $i, i = 1, ..., D$ (usually $D = 2$ or 3), with L_{i-1} the set of voxels classified as lesion in previous steps, and L_0 the lesion sample. We cover the ROI by a grid of cubes of edge length d_i, so that each cube contains d_i^3 voxels. In the initial steps of the method, we use large cubes, typically $d_1 = 8$ or 4, and decrease the size for later steps, making this a hierarchial procedure. We will discuss this further in section 2.2.

At step i, for each such cube C that borders but does not intersect L_{i-1}, we consider the following hypothesis test:

$$H_0 : C \text{ is entirely inside the lesion}$$
$$H_A : C \text{ is not entirely inside the lesion.}$$

More precisely, H_A says that at least one voxel in C is not completely inside the lesion. We write these qualitative hypotheses more precisely as follows:

$$H_0 : \mu_C \geq \mu_L - k_{i,L} \tag{1}$$

$$H_A : \mu_C < \mu_L - k_{i,L}, \tag{2}$$

where μ_C is the mean intensity of the cube being tested, and $k_{i,L}$ is a pre-set parameter that depends on i and the pulse sequence used to acquire the imagery.

Assuming that intensity follows a normal distribution [9], we use the standard two-sample t-test of size $\frac{\alpha}{D N_i}$ [5] to test H_0 vs. H_A, where α is a fixed constant between 0 and 1, and N_i is the number of cubes being tested at step i. A type I error occurs if C is actually inside the lesion, but is not accepted into it.

Using Bonferroni's inequality [5], it can be shown that the *overall* probability of a type I error can be controlled when all the cubes in the grid are tested. That is, let $C_{im}, m = 1, ..., N_i$, denote the cubes in the covering of the ROI at step i that border but do not intersect L_{i-1}, and let $\mathcal{L}_i$ denote the set of cubes C_{im} that are actually in the lesion. Then

$$P(\text{all } C_{im} \in \mathcal{L}_i \text{ are accepted into the lesion}) \geq 1 - \frac{\alpha}{D}. \tag{3}$$

2.2 Coarse-to-Fine Aspect

After step i is completed, we have L_i, the set of voxels classified as lesion in steps $1, ..., i$, and we now use it as the lesion sample for step $i + 1$, yielding the sample statistics $\overline{x}_{L_i}$ and $s_{L_i}^2$. In step $i + 1$, we cover the ROI with (smaller) cubes of edge $d_{i+1} < d_i$. This coarse-to-fine aspect of the method allows us to update our coarse segmentation from the previous steps into a more accurate one by "packing" it with smaller cubes.

At step $i + 1$, we test each cube $C_{(i+1)m}, m = 1, ..., N_{i+1}$, that borders but does not intersect L_i, for acceptance into the lesion via the two-sample t-test. We repeat for all steps $i = 1, ..., D$, so that L_D is the final segmentation of the lesion. Equation 3 says that we can control the overall probability of a type I error at each step; similarly we can also show that we can control the type I error for the entire multi-step procedure, namely,

$$P(\bigcap_{i=1}^{D} [\text{all } C_{im} \in \mathcal{L}_i \text{ are accepted into the lesion}]) \geq 1 - \alpha. \tag{4}$$

Proofs of the results (3) and (4) are in [10]. We could continue the procedure to the finest resolution, so that $d_D = 1$. However, in practice we stop before this finest level to diminish the effect of the high variability present in individual voxels; this helps us to regularize the segmentation.

2.3 One-sided Confidence Bound for Volume

A lesion volume can be estimated by counting the number of voxels that are accepted into the lesion. From (4), we know (with $(1 - \alpha)100\%$ confidence) that

Fig. 1. Coarse-to-fine aspect of packing. Left: Sample slice (restricted to a ROI). Center: Lesion segmentation after $4 \times 4 \times 4$ boxes are tested. Right: Final segmentation after $2 \times 2 \times 2$ boxes are tested, which overestimates the lesion (see section 2.3).

we will admit all cubes that actually belong in the lesion. In terms of the volume estimate V_L and the actual volume V, we have that

$$P(V_L \geq V) \geq 1 - \alpha. \tag{5}$$

However, we do not control the possibility of admitting cubes that do not belong in the lesion. Cubes with a mixture of lesion and background voxels are especially susceptible of being incorrectly admitted. Thus V_L is not an unbiased estimator of the lesion volume, but rather an upper confidence bound of V.

2.4 Two-Sided Confidence Interval and Point Estimate for Volume

Using the same method as above to pack the *background*, we obtain another segmentation of the image. The resulting lesion volume V_B provides a lower confidence bound for V: $P(V_B \leq V) \geq 1 - \alpha$. Combining this with (5), we have

$$P(V_B \leq V \leq V_L) \geq 1 - 2\alpha. \tag{6}$$

Therefore, the interval (V_B, V_L) is a $(1 - 2\alpha)100\%$ confidence interval for the lesion volume. Any combination of V_B and V_L (the average, e.g.) is a valid point estimate for V. We will apply this method to real MRI in section 4.

Fig. 2. Results for the packing method. Left: Sample slice. Center: Lesion segmentation via forward packing (an overestimate). Right: Lesion segmentation using background packing (an underestimate).

3 Hybrid Method

3.1 Deformable Models

A modeling techinique known as "active contours" or "snakes," is a semi-automatic approach to segmentation, originally proposed in [6]. It belongs to a class of methods known as deformable models. The idea is that a contour may be placed near some image feature and then deformed to optimally fit the feature.

The contour, or snake, is deformed automatically by minimizing an "energy functional," given by

$$E_{total} = E_{int} + E_{ext} = \int F_{int} + \int F_{ext}. \qquad (7)$$

The internal energy, E_{int}, is typically defined as a smoothness constraint on the contour, E_{ext}, the external energy, is commonly defined as the negative of the gradient of the image, which pushes each portion of the snake to the strongest and nearest edge. F_{int} and F_{ext} are the corresponding forces, which we integrate over the contour.

External forces other than the image gradient have been proposed (see [4], e.g.). In section 3.2 we propose a new force, based on statistical packing, to help regularize the model. When the snake has converged to a minimal energy state, we classify the area inside it as lesion; see [6] for the computational details.

To go from a two-dimensional model to a three-dimensional one we use an approach similar to that of [2] and [14]. We propagate a final contour from one slice to serve as the initial contour in the next slice. Contours in adjacent slices are then connected with an additional force, imposing an overall smoothness constraint on the model. This 3D force assumes that only small changes in lesion shape and size occur in neighboring slices.

3.2 The Hybrid Algorithm

The approaches in sections 2 and 3.1 each have some limitations. Packing sometimes requires a large amount of user interaction, since the user must manually select the 3D ROI and the lesion and background samples. It is also sensitive to the quality of the initial samples. Meanwhile, the snake approach can be affected by other nearby objects creating strong edges, and it can perform poorly if there are significant changes in the size of the lesion in adjacent slices. It also can get stuck in a local energy minimum.

To minimize the problems of the two approaches, as well as to reinforce their strengths, we have combined them into a "hybrid" method: use a 2D version of packing described in [7] to produce a single lesion estimate, and use the gradient of the resulting binary image to define an additional external energy. The full algorithm is outlined as follows:

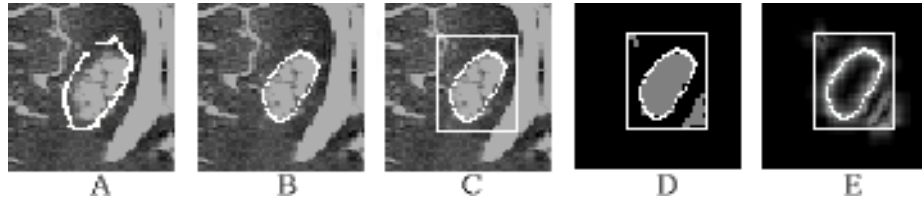1. Manually initialize a snake in the base slice (Fig. 3A), and let it deform until it converges (Fig. 3B).

Fig. 3. Steps of the hybrid method (see text). A: Manual initialization of the contour. B: Converged contour. C: ROI. D: Statistical segmentation. E: Statistical energy.

2. Define a bounding box around the resulting contour, which will serve as the ROI in the current slice (Fig. 3C). The interior of the contour is used as the lesion sample, and the remainder of the box is the background sample.
3. Run the packing method and make the result into a binary image (Fig. 3D).
4. Define the negative of the gradient of the image as an additional external energy term (Fig. 3E).
5. Run the snake with this added energy to produce a segmentation of the slice.
6. Use the finished contour as the initialization for the adjacent slice. Repeat steps 2–5 until the number of pixels classified as lesion in a slice is less than some threshold. This will result in a 3D model of the lesion (Fig. 4).
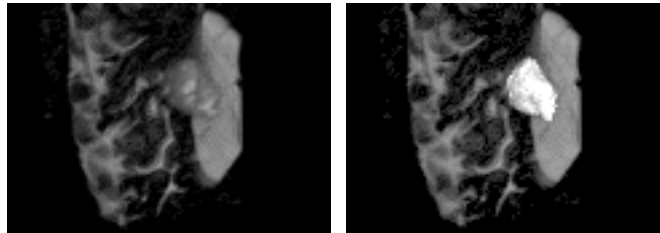

Fig. 4. Result of the hybrid method. Left: A portion of a 3D MR scan. Right: The lesion estimate superimposed (in white) onto the original image. 3D visualization software courtesy of Biomedical Imaging Group, UMass Medical Center, Worcester, MA.

With this approach we greatly reduce the amount of user interaction that the statistical procedure normally requires. All the user must do is draw a contour in the base slice; the ROI and the samples will then be obtained automatically. It also helps deal with the limitations of snakes, by adding the intensity-based statistical information to the snake model. This helps to keep it from being distracted by nearby normal, bright structures, and also helps to keep it on track in case of a non-incremental change from one slice to another. For implementation details see [7].

4 Results

We have tested the above methods on lesions in axial MRI. We obtained all of our data sets from Baystate Medical Center (West Springfield, MA) using a 1.5 T Picker Edge machine. One patient was imaged using a T_2 sequence with 1 mm slice thickness and no gaps. A FLAIR sequence with 2.5 mm slice thickness and no gaps was used for the other five.

To evaluate our results, we obtained two hand segmentations from each of two physicians at Baystate for each data set. We compare the volumes derived from these segmentations to the results for five different initializations of the packing and hybrid methods. All results are in Table 1.

| | Hand Segmentation | | | | Packing | | | | Hybrid | |
| | Doctor 1 | | Doctor 2 | | Lower CB | | Upper CB | | | |
Patient	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
1	2069	120	1947.5	24.5	1914.6	56.3	2337	108.6	1927.6	49.3
2	3416	139.7	3617.5	183.8	2608	310.1	4709	142.7	3841.5	277.8
3	23555	762.3	21359	618.7	20019	1020	25891	760	21821.9	617.7
4	9962.5	67.2	10787.5	449.0	9434.5	320.4	11524.5	633.7	10031	906.8
5	10131	740.7	9377.5	1064.2	8603.5	320.2	11443.5	294.4	9536.9	251.8
6	3452.5	35.3	3504	160.9	2699.5	94.5	4205	476.2	3525	117.9

Table 1. Results for six MRI scans (in mm^3), including the means and standard deviations of lesion volumes from two physicians, those for the packing method (CB stands for a 90% confidence bound), using five different initializations, and for the hybrid method, also using five initializations.

The results show that the methods are consistently estimating lesion volume to within 20% of the physicians. The results from packing give only the lower and upper 90% confidence bounds for lesion volume; we use the average of the bounds to compare with the other estimates. The worst error, compared to the overall average of the doctors' estimates, was 6% for packing and 9% for the hybrid method, well within the 20% threshold that the physicians required. Comparing standard deviations, we see that the two methods generally are as consistent as the doctors, and more so in some cases. All pre-set parameters in the methods were the same for data imaged with the same pulse sequence; changes were necessary for packing between T_2 and FLAIR imagery, whereas no adjustments were necessary for the hybrid method. We have also tested the procedures on synthetic imagery, to evaluate them in the situation when the volume is known. Similar to real MRI, the error of our volume estimates did not, on average, exceed 5% in all cases.

5 Conclusion

We have introduced two new methods to segment MR imagery and estimate lesion volumes, one statistical and another that incorporates the statistical approach into a deformable models framework. Both methods are working well for actual patient data. The first procedure gives a rigorous confidence interval for lesion volume, but requires manual ROI extraction. The hybrid method reduces the user interaction, and generates a model of the lesion, insuring connectivity and smoothness, but a confidence interval has not yet been derived.

In the next step of our research we plan to integrate a full 3D version of packing into the snake framework, as well as to utilize the confidence interval to bound the error of the hybrid method. These steps will further increase the reliability and robustness of the hybrid method.

References

1. R. Adams and L. Bischof. Seeded region growing. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 16:641–647, 1994.
2. I. Carlbom, D. Terzopoulos, K. Harris. Computer-assisted registration, segmentation, and 3D reconstruction from images of neuronal tissue sections, *IEEE Transactions on Medical Imaging*, 13(2), 351-362, 1994
3. H. Cline, W.E. Lorenson, R. Kikinis, F. Jolesz. Three-dimensional segmentation of MR images of the head using probability and connectivity. *Journal of Computer Assisted Tomography*, 14(6):1037–1045, 1990.
4. L.D. Cohen. On active contour models and balloons. *CVGIP: Image Understanding*, 53(2):211–218, 1991.
5. R. Johnson and D. Wichern. *Applied Multivariate Statistical Analysis*. Prentice-Hall, New Jersey, 1992.
6. M. Kass, A. Witkin and D. Terzopoulos. Snakes: Active contour models. *Intl. Journal of Computer Vision*, 1:321–331, 1987.
7. D. Lisin, B. Stein, J. Horowitz, G. Whitten, E. Riseman, D. Geman, R. Hicks, and B. Pleet. Statistical and computer vision techniques to support the clinical study of ischemic stroke treatment. Technical Report. UM-CS-2001-019. UMass, 2001.
8. A. Martel, S. Allder, G. Delay, P. Morgan, and A.R. Moody. Measurement of infarct volume in stroke patients using adaptive segmentation of diffusion weighted MR images. *Proc MICCAI Conference*, 1999.
9. J. Sijbers, A.J. Den Dekker, P. Scheunders, and D. Van Dyck. ML estimation of Rician distribution parameters. *IEEE Trans on Medical Imaging*, 17:357–361, 1998.
10. B. Stein. *Signal formation, segmentation, and lesion volume estimation in magnetic resonance imagery*. Ph.D. dissertation, University of Massachusetts, 2001.
11. T. Tosteson, B. Pogue, E. Demidenko, T. McBride, and K. Paulsen. Confidence maps and confidence intervals for near infrared images in breast cancer. *IEEE Trans on Medical Imaging*, 18:1188–1193, 1999.
12. C. Watson, C. Jack Jr., and F. Cendes. Volumetric magnetic resonance imaging. *Archives of Neurology*, 54:1521–1531, 1997.
13. W. Wells, R. Kikinis, W. Grimson, and F. Jolesz. Adaptive segmentation of MRI data. *IEEE Trans on Medical Imaging*, 15:429–442, 1996.
14. J. M. Whitaker and M. Braun. 3D image segmentation using active contours with interslice energy. *Proc APRS/CBT Image Segmentation Workshop*, 47–51, 1996.