Measurement of Infarct Volume in Stroke Patients Using Adaptive Segmentation of Diffusion Weighted MR Images.

Anne L. Martel[1], Steven J. Allder[2], Gota S. Delay[3], Paul S. Morgan[3] , and Alan R. Moody[3]

[1] Dept. Medical Physics, Queens Medical Centre, Nottingham, NG7 2UH, UK
[2] Dept. Clinical Neurology, University of Nottingham, NG7 2UH, UK
[3] Dept. of Academic Radiology, University of Nottingham, NG7 2UH, UK

Abstract. This paper describes a semi-automatic method of determining the infarct volume, an important parameter in the assessment of stroke patients, from MRI Diffusion Weighted Images (DWI). An adaptive thresholding algorithm incorporating a spatial constraint was used to segment the images. The relationship between adjacent pixels was modeled using a Markov Random Field (MRF) and the Iterative Conditional Modes (ICM) method was used to find a locally optimum solution. In order to improve the robustness of the ICM method, initial threshold levels were determined automatically using a non-spatial method. Preliminary results showed that the completely automatic technique failed if the infarct was too small or if the contrast was too low. The operator was therefore given a choice of modifying the initial threshold levels manually. It was also necessary to edit the final segmentation results in some cases as nerve tracts may also appear as bright regions on the images. Simulation studies were used to determine the accuracy of the technique. Reproducibility studies were carried out to determine the effect of inter and intra observer variability and patient positioning. The semi-automatic technique was quicker and more reproducible than manual segmentation and allowed the infarct volumes to be measured with a repeatability coefficient of < 6 cc.

1. Introduction

There has been considerable interest in the use of diffusion weighted imaging (DWI) to detect the site and size of ischaemic lesions in stroke patients [1, 2]. This information assists in classifying the stroke sub-type and may also be useful in predicting the clinical condition and eventual outcome [3]. Changes in infarct volumes over time have also been measured in a number of studies in an attempt to understand the natural history of stroke [4]. In future, measurements of infarct volume may have an important role to play in the assessment of stroke therapies in clinical trials.

In previous studies the DWI images have been segmented using manual region tracing methods in order to calculate infarct volumes [4, 5]. This is a laborious task and results are operator dependent. The aim of this work was to devise a method of

C. Taylor, A. Colchester (Eds.): MICCAI'99, LNCS 1679, pp. 22-31, 1999.

segmenting the DWI images with a minimum of operator intervention that is quicker and more reproducible than existing methods.

We have used an adaptive segmentation technique [6] to automatically determine the brain and infarct volumes. This is based on the Iterative Conditional Modes (ICM) method [7] which allows the relationship between adjacent voxels to be taken into account during the segmentation. Since this method only converges to a local solution we have estimated the initial threshold values using an automatic technique [8]. In order to make the program completely robust it is designed to run under operator supervision. The operator is able to modify the initial threshold values if necessary and some editing of the final segmented image is also possible. The segmentation program was tested using simulation studies and inter and intra observer variability and inter-scan variability were determined using patient studies.

2. Theory

The contrast in a diffusion-weighted image is due to differences in mobility of water molecules in the tissue. If water is freely diffusible then the MRI signal is attenuated, if the movement of the water molecules is impeded in some way, for example along nerve bundles, then the MRI signal is larger. Acute cerebral ischaemia causes cell damage which results in edema due to the accumulation of intracellular water. This causes a restriction in the movement of water molecules therefore the infarct shows up as a region of increased signal intensity. The result is an image with three distinct regions: air, which has a signal intensity of close to zero; normal brain which has an intermediate grey scale value and infarcted tissue which has the highest signal intensity and is typically more heterogeneous than the other two regions. The problem is therefore to estimate the "true" image $X = \{x_s, s = 1...N\}$ (where N is the number of voxels in the image) with discrete values $x_s \in \{1,2,...K.\}$ (where K is the number of regions) from the noisy observed image $Y = \{y_s, s = 1...N\}$. $\mu(k)$ is the mean intensity of all pixels belonging to the k'th region and $\mu(1) < \mu(2) ... < \mu(K)$. From Bayes' Theorem $P(X|Y)$, the conditional probability of X given the observed image Y is given by

$$P(X|Y) \propto P(Y|X)\, P(X) \qquad (1)$$

Where $P(Y|X)$ is the likelihood of observing the image Y given the true image X and $P(X)$ is the prior probability of X. The maximum *a posteriori* (MAP) estimate of the true image is the value of X which maximizes $P(X|Y)$. If we assume that the observed intensity values y_s are independently distributed with a mean x_s and variance σ^2 then $P(Y|X)$ is given by

$$P(Y \mid X) = \prod_{s=1}^{N} f(y_s \mid x_s) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left\{\frac{-1}{2\sigma^2}\sum_{s=1}^{N}\{y_s - \mu(x_s)\}^2\right\} \qquad (2)$$

If the probability of a pixel x_s having a particular value depends only on the pixel values in the neighborhood of s (η_s) then X corresponds to a Markov Random Field (MRF). We have used

$$P(x_s) \propto \exp\{\beta Z(x_s)\} \tag{3}$$

where $Z(x_s)$ is the number of pixels where $x_n = x_s$ for $n \in \eta_i$ and η_i is the 3-D, 2nd order neighborhood of i. The probability $P(X|Y)$ can be maximised using the method of iterated conditional modes (ICM) [7]. An initial estimate of the image X is generated and then each pixel s is updated in turn so that the new value of x_s is optimised for the current estimate of the true image, X_c. The function to be maximised is given by

$$p(x_s \mid Y, X_c) = p(x_s \mid y_s, x_n, n \in \eta_s) \propto \exp\left(\frac{-1}{2\sigma^2} \sum_{s=1}^{N} \{y_s - \mu(x_s)\}^2 + \beta Z(x_s) \right) \tag{4}$$

Several sweeps of the image are made until convergence is reached. This technique converges to a local rather than a global minimum but it is computationally simpler than alternative minimization schemes such as simulated annealing [9].

3. Methods

3.1 Adaptive Segmentation

Mardia and Hainsworth [6] describe an algorithm for adaptive thresholding based on ICM. From equation (4) it can be shown that the intensity level y for which $p(x_s=i| y_s, X_c)$ is equal to $p(x_s=j| y_s, X_c)$ is given by

$$y = \frac{1}{2}\{\mu(i) + \mu(j)\} + \beta\sigma^2 \{Z(j) - Z(i)\} / \{\mu(i) - \mu(j)\} \tag{5}$$

where i,j = 1..K, i<j and $Z(i)$ is the number of pixels in the neighborhood of s with $x_n = i$. The decision variable is B defined as

$$B = y_s + \beta\sigma^2 \{Z(j) - Z(i)\} / \{\mu(i) - \mu(j)\} \tag{6}$$

and the threshold level separating region i from region j is given by

$$t_{i,j} = \frac{1}{2}\{\mu(i) + \mu(j)\} \tag{7}$$

Using these definitions the algorithm is as follows:

1. Estimate the initial threshold values $\{t_{i,j}, i,j=1..K, i<j\}$ and segment the image into K regions to obtain an initial estimate of X
2. Estimate the mean grey level $\mu(k)$ for each region.
3. For each region subtract the mean grey level $\mu(x_s)$ from the observed grey level y_s. and estimate the variance σ^2 from this modified data set.
4. Re-evaluate the threshold values, $t_{i,j} = \frac{1}{2}\{\mu(i) + \mu(j)\}$
5. For each pixel in the image do the following:

Set i=1
For j=2 to K do
If $B > t_{i,j}$, set $i=j$: next j
Set $x_s = i$

6. Repeat steps 2-5 until the solution is stable. We have defined convergence as occurring when the percentage change in the number of pixels in the infarct region is < 0.1 %.

The ICM method finds a local rather than a global solution and the final segmentation may therefore be affected by the starting values [6]. The number of iterations required will also depend on how close the initial threshold values are to the final solution. For this reason we have used the iterative method described by Ridler and Calvard [8] to generate the initial threshold values. This method is equivalent to the adaptive segmentation algorithm described above for the special case when $\beta=0$. Since no spatial information is used the technique can be applied directly to the image histogram which makes it considerably faster to run than ICM which has to update each pixel in the image in turn. The algorithm used is as follows:

1. Generate the image histogram $h(v)$ from the observed image Y. $v = 0,...\ m$ and m is the maximum grey level in the image.
2. Estimate the initial threshold values by dividing the range of grey level values into K equal partitions so that $t_{i,j} = \{i \times m\}/K$ where $i < j = 1...K$
3. Calculate the mean intensity values for the regions defined by the threshold values.

$$\mu(k) = \frac{\sum_{v=t_{k-1,k}}^{v<t_{k,k+1}} h(v) \times v}{\sum_{v=t_{k-1,k}}^{v<t_{k,k+1}} v} \qquad (8)$$

4. Calculate the new threshold values $t_{i,j} = \frac{1}{2}\{\mu(i) + \mu(j)\}$
5. Repeat steps 3-5 until the threshold values converge to a stable solution.
This method always converges but the solution is dependent on starting points.

Figure 1 illustrates the whole segmentation process.

3.2 Semi-automatic Segmentation.

The fully automatic technique described above worked well in most cases. However preliminary trials with patient data showed that in some images where the infarct was small and/or the contrast between the infarct and normal brain tissue was low the initial thresholding technique failed to separate normal brain from infarct. An example of this is shown in figure 2. The operator was therefore presented with a display showing the threshold contours superimposed on the DWI images, together with the intensity histogram. The thresholds could be modified manually by defining new threshold levels on the histogram.

Manual editing of the segmented images was also required in some cases. This was due to the presence of high intensity regions in the corpus callosum (due to the

presence of nerve tissue) and in the base of the brain (due to susceptibility artifacts) which were incorrectly classified as infarct. A simple graphical user interface written in IDL (Boulder, CO) enabled the operator to reclassify these regions as normal brain.

3.3 Simulation Studies.

The adaptive segmentation technique described in section 3.1 assumes that true image *X* is contaminated with white gaussian noise. This is a resonable assumption for the brain and infarct regions where the signal to noise ratio is high but in the low intensity background region where there is no signal the noise has a Rician distribution [10]. It is also assumed that the pixels in the true image *X* can only take one of three discrete values. This is only an approximation and in practice both the normal brain and the infarct regions will be heterogeneous.

The accuracy of the segmentation technique was assessed using simulation studies which attempt to model the true patient data. A patient scan with a moderate sized infarct (28 cc) was segmented manually into background, brain and infarct regions. The mean and standard deviation of pixel values within each region were calculated. Simulation studies were created by filling in the 3 regions with intensity values drawn from the appropriate distribution. 3 high contrast simulations were generated using μ=[0,130,430], σ = [20,30,80]. A second patient image with a much smaller, less distinct infarct was analyzed and the mean values from this were used to create 3 low contrast images (μ=[0,130,230], σ = [20,35,40]).

The effect of varying β between 0.5 and 2 and the effect of different initial threshold values on the segmentation results were investigated.

3.4 Image Acquisition

Patients diagnosed with a recent anterior circulation stroke were imaged on a 1.5 T Siemens Magnetom Vision using an echo planar DWI sequence. Imaging parameters were as follows: b value, 1100 s/mm^2; field of view, 250 mm; matrix size, 128x128 (sinc interpolated to 256x256); pixel size, 0.977 mm; slice thickness, 5 mm; number of slices, 25 (acquired interleaved with no slice gap); effective TE, 123 ms; total scan time, 7.044 s (for 25 slices).

A total of 63 patient studies were carried out with a median time between onset of stroke and acquisition of images of 26 hours. 10 of these studies were randomly selected for the inter- and intra- observer reproducibility studies (median time to scan: 11 hours). In a further 5 patients a second DWI image was acquired with different tilt angles and table positions to simulate the effect of changing the patients position in the scanner (median time to scan: 24 hours).

4. Results

4.1 Simulation Studies

The results in Table 1. give the mean values from 3 noisy simulations. The % miss-classification is the number of pixels incorrectly classified*100 / total number of pixels. The % error in the infarct volume is also given.

For the high contrast simulation the classification accuracy over the whole image was very good although the infarct volume was consistently underestimated. There was little difference between using the automatic and manually selected thresholds as starting values.

For the lower contrast simulations, the segmentation failed when $\beta<1$. The starting threshold levels selected by the automatic technique were clearly incorrect but despite this the ICM method converged to a reasonable solution. However this took an average of 11 iterations compared to 4 iterations needed for the manually selected thresholds (for $\beta = 1$)

There is little difference in the results for $\beta = 1$ and $\beta = 1.5$ but convergence occurred in fewer iterations when $\beta = 1$ so this value was selected for the patient studies.

Table 1. Summary of results for simulation studies.

		High Contrast		Low Contrast	
Initial Threshold	β	% Miss-classification (± 1SD)	% error in infarct volume (± 1SD)	% Miss-classification (± 1SD)	% error in infarct volume (± 1SD)
automatic	0	**0.60** (0.01)	**-3.2** (0.2)	**6.84** (0.02)	**1399.7** (0.1)
	0.5	**0.087** (0.002)	**-2.73** (0.08)	**0.301** (0.003)	**23.5** (0.7)
	1	**0.049** (0.001)	**-2.45** (0.03)	**0.104** (0.001)	**-1.0** (0.4)
	1.5	**0.049** (0.002)	**-2.13** (0.04)	**0.156** (0.008)	**1.4** (2.3)
manual	0.5	**0.081** (0.005)	**-2.6** (0.1)	**0.31** (0.01)	**18.0** (2.5)
	1	**0.045** (0.002)	**-2.24** (0.07)	**0.115** (0.003)	**-4.7** (0. 8)
	1.5	**0.048** (0.001)	**-1.89** (0.09)	**0.113** (0.006)	**-8.7** (1.9)

4.2 Patient Studies

The inter-observer, intra-observer and intra-scan reproducibility of the semi-automatic segmentation was assessed using the approach described in [11]. The difference between the repeated measures of infarct volume was calculated for each data set and the mean and standard deviation of these differences was calculated. The repeatability coefficient is defined as the 95% confidence interval, i.e. ±2SD.

The intra and inter observer variability for manual segmentation was also assessed. Manual segmentation was carried out using a graphical tool that provided access to a combination of thresholding, region growing and manual tracing techniques. The results show that the reproducibility is poor compared with the semi-automatic technique. The infarct volumes estimated by observer 1 were significantly higher than those estimated by observer 2 for the manual segmentation method. There was no significant bias with the semi-automatic technique.

The average time taken to segment both the normal brain and the infarct was 6 minutes for the semi-automatic technique and 15 minutes for the manual technique. Of the 6 minutes it took to carry out the semi-automatic segmentation, approximately 3 minutes were spent waiting for the ICM method to converge.

Table 2. Results of the reproducibility studies.

		Semi-Automatic		Manual	
	Mean volume	Mean difference	2SD	Mean difference	2SD
Intra-observer	94.0 cc (n=10)	0.9 cc	5.5 cc	1.8 cc	17.9 cc
Inter-observer	"	-0.5 cc	5.8 cc	11.0 cc	27.7 cc
Inter-scan	71.9 cc (n=6)	-0.2 cc	1.4 cc		

The median infarct volume for all 63 studies was 73cc. Automatic segmentation failed completely in 6 studies due to poor image quality. The initial thresholding step failed in a further 23 cases making it necessary to manually define the threshold level; the infarct size in this group was significantly lower than the remaining studies (11cc compared with 115cc).

The importance of the final manual editing step depended on the size of the infarct: in 27 cases the manual editing modified the infarct volume by < 1% (median infarct volume=125cc), in 12 cases the change was between 1-10% (median infarct volume=92cc) and in 18 cases the change was > 10% (median infarct volume=11cc). For all studies the median change in volume after manual editing was 1.2cc or 1.2% of infarct volume. There was no significant relationship between the time at which the scan was carried out and the quality of the segmentation.

Discussion

Previous studies have used manual segmentation techniques to determine the infarct volume from DWI images of the brain. The semi-automatic segmentation method

described here is faster than the manual technique and has significantly improved reproducibility.

With the present technique the initial thresholding step tends to fail for small infarcts. The manual selection of a threshold level only takes seconds but in order to completely automate the process an alternative method is needed. One possibility would be to fit gaussian functions to the brain and background peaks in the intensity histogram and to derive the threshold values from the fitted parameters. Alternatively, if the b value and other acquisition parameters are kept the same for all images then the initial threshold levels could be based on intensity values obtained from training data. Improving the initial selection of the threshold would allow the segmentation to be carried out automatically. The initial results suggest that this would remove the need for any operator intervention at all for larger infarcts. For smaller infarcts the operator would only have to check the segmentation results and edit any miss-classified regions if necessary. We anticipate that this final editing step would only take 1-2 minutes depending on the quality of the initial segmentation. It may also be possible to predict the spatial distribution of the nerve tracts in order to exclude them automatically.

Previous studies have defined changes in infarct volumes of > 20% as significant°[4]. If we assume a mean infarct volume of 35 cc [12] then this corresponds to an increase of 7 cc. Our results show that with the semi-automatic segmentation technique infarct volumes can be measured with a reproducibility of < 6°cc. Other groups [4, 5] have reported an inter-observer reproducibility of 5% using a manual region tracing technique. This is considerably better than our results and this may be due to better training of the observers or to differences in the manual region tracing method. It is difficult to compare results directly as details of how the reproducibility was measured are not given and the percentage error will depend on the infarct size.

Apparent diffusion coefficient (ADC) maps can be generated by acquiring DWI images with several different b values. ADC images provide quantitative information and may be of value in distinguishing between old and new lesions, however the contrast between infarcted and non-infarcted tissue is poor compared to the diffusion weighed images. We have therefore used the DWI images to estimate infarct volume. Some attempts have also been made to use multispectral segmentation techniques to segment the DWI images acquired with different b values [13, 14]. The disadvantage of this approach is that the multiple DWI images must be registered exactly so corrections must be made due to the presence of distortions in the image. It is also important that no patient movement occurs between images and this may be difficult to achieve when dealing with acute stroke patients.

References

[1] R. G. Gonzalez, P. W. Schaefer, F. S. Buonanno, L. H. Schwamm, R. F. Budzik, G. Rordorf, B. Wang, A. G. Sorensen, and W. J. Koroshetz,: Diffusion-weighted MR imaging: diagnostic accuracy in patients imaged within 6 hours of stroke symptom onset. Radiology **210** (1999) 155-62.

[2] K. O. Lovblad, H. J. Laubach, A. E. Baird, F. Curtin, G. Schlaug, R. R. Edelman, and S. Warach,: Clinical experience with diffusion-weighted MR

in patients with acute stroke [see comments]. AJNR Am J Neuroradiol **19** (1998) 1061-6.

[3] K. O. Lovblad, A. E. Baird, G. Schlaug, A. Benfield, B. Siewert, B. Voetsch, A. Connor, C. Burzynski, R. R. Edelman, and S. Warach,: Ischemic lesion volumes in acute stroke by diffusion-weighted magnetic resonance imaging correlate with clinical outcome. Ann Neurol **42** (1997) 164-70.

[4] A. E. Baird, A. Benfield, G. Schlaug, B. Siewert, K. O. Lovblad, R. R. Edelman, and S. Warach,: Enlargement of human cerebral ischemic lesion volumes measured by diffusion-weighted magnetic resonance imaging [see comments]. Ann Neurol **41** (1997) 581-9.

[5] P. A. Barber, D. G. Darby, P. M. Desmond, Q. Yang, R. P. Gerraty, D. Jolley, G. A. Donnan, B. M. Tress, and S. M. Davis,: Prediction of stroke outcome with echoplanar perfusion- and diffusion- weighted MRI. Neurology **51** (1998) 418-26.

[6] K. V. Mardia and T. J. Hainsworth,: Spatial thresholding method for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **10** (1988) 919-927.

[7] J. Besag,: On the statistical analysis of dirty pictures. J. R. Statist. Soc. B **48** (1986) 259-302.

[8] T. W. Ridler and S. Calvard,: Picture Thresholding Using an Iterative Selection Method. IEEE Transactions on Systems, Man and Cybernetics **SMC-8** (1978) 630-632.

[9] S. Geman and D. Geman,: Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. IEEE Trans. Pattern Anal. Machine Intell. **PAMI-6** (1984) 721-741.

[10] H. Gudbjartsson and S. Patz,: The Rician distribution of noisy MRI data [published erratum appears in Magn Reson Med 1996 Aug;36(2):332]. Magn Reson Med **34** (1995) 910-4.

[11] J. M. Bland and D. G. Altman,: Statistical methods for assessing agreement between two methods of clinical measurement. Lancet **1** (1986) 307-10.

[12] K. J. van Everdingen, J. van der Grond, L. J. Kappelle, L. M. Ramos, and W. P. Mali,: Diffusion-weighted magnetic resonance imaging in acute stroke. Stroke **29** (1998) 1783-90.

[13] R. A. Carano, K. Takano, K. G. Helmer, T. Tatlisumak, K. Irie, J. D. Petruccelli, M. Fisher, and C. H. Sotak,: Determination of focal ischemic lesion volume in the rat brain using multispectral analysis. J Magn Reson Imaging **8** (1998) 1266-78.

[14] V. Nagesh, K. M. Welch, J. P. Windham, S. Patel, S. R. Levine, D. Hearshen, D. Peck, K. Robbins, L. D'Olhaberriague, H. Soltanian-Zadeh, and M. D. Boska,: Time course of ADCw changes in ischemic stroke: beyond the human eye! Stroke **29** (1998) 1778-82.

Fig. 1. Adaptive Segmentation. Top left: a slice taken from a DWI 3D image. Top right: the image histogram has a background, normal brain and infarct peak (the y axis has been clipped to show the infarct peak). The two automatically determined threshold levels (85 and 237) are displayed. Bottom left: the segmentation using the threshold levels determined using β=0. Bottom right: the segmentation with β=1.

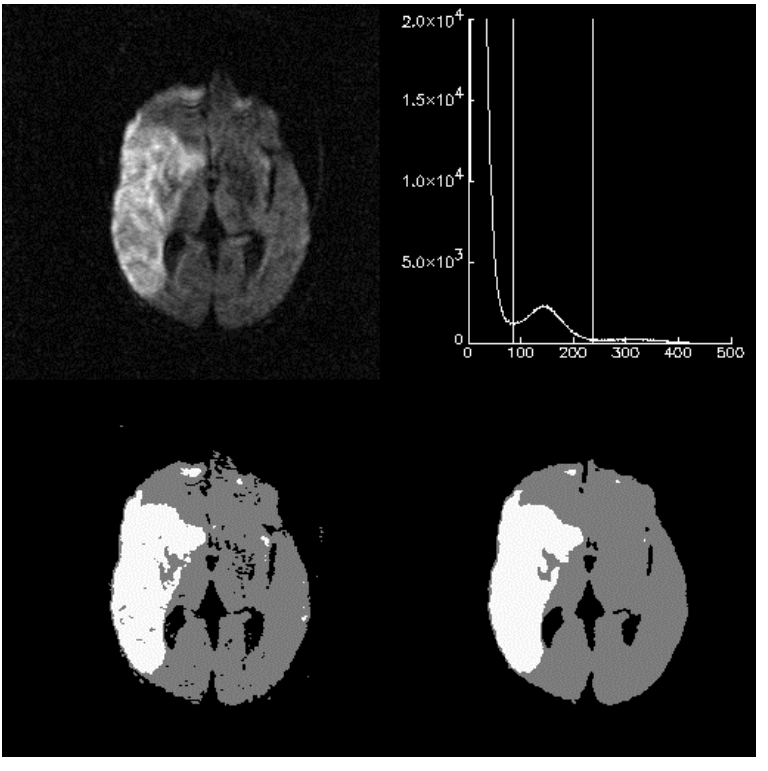

Fig. 2. In cases where the infarct is small (left) the image histogram only has two peaks (centre) and the automatically determined threshold values include normal brain (right)