Tina Memo No. 2008-003
Internal Memo

Performance Evaluation of the TINA Medical Image Segmentation Algorithm on Brainweb Simulated Images

P. A. Bromiley

Last updated
20 / 12 / 2007

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

Performance Evaluation of the TINA Medical Image Segmentation Algorithm on Brainweb Simulated Images

P. A. Bromiley
Dept. of Medical Biophysics
Imaging Science and Biomedical Engineering Division
Medical School, University of Manchester
Manchester, M13 9PT, UK
paul.bromiley@man.ac.uk

Abstract

This memo describes the performance evaluation of the TINA medical image segmentation algorithm described in Memo 2004-009 when applied to simulated images produced by the Brainweb MRI simulator. In order to allow Monte-Carlo experiments to be performed using independent image noise fields, and to avoid problems introduced by the presence of histogram artefacts in the Brainweb simulated images as identified in TINA Memo no 2008-002, images were re-simulated from the Brainweb tissue phantoms. Monte-Carlo experiments were performed by applying the segmentation algorithm both with and without the use of gradient terms. The results were evaluated by measuring the rates of voxel misclassification, treating the algorithm as a simple classifier, and by measuring the χ^2 of the tissue volume estimates, in order to evaluate the accuracy of partial volume estimation. We conclude that the TINA medical image segmentation algorithm achieves misclassification rates close to the Bayes error for intensity-based segmentation, and that the use of gradient terms improves the accuracy of tissue volume estimates in partial volume voxels by an average of 31%.

1 Introduction

One of the most common tasks in medical image analysis is that of segmentation: separating an image into regions representing the various tissues it contains. The TINA software contains a medical image segmentation algorithm, described in [4], which fits a model consisting of Gaussian distributions, describing pure tissues, and Gaussian distributions convolved with triangular distributions, describing partial volume contributions i.e. voxels containing more than one tissue, to medical image volumes. The fit to the image data is optimised using the Expectation-Maximisation algorithm. The algorithm accepts input data of arbitrary dimensionality i.e. multiple images of the same scene, acquired using different modalities. In addition, the algorithm can use image gradients, as well as grey levels, to increase the separability of pure tissue and partial volume contributions. The development of the algorithm through the addition of each of these capabilities is described in [3], [1], [6], and [2].

2 Method

The 1.0mm slice thickness Brainweb phantoms were downloaded in gzipped raw short format, uncompressed using gunzip 1.3.12, and read into TINA using the raw image input tool. Nine structure-rich slices (slice 142 to slice 150 inclusive) from an area of the brain expected to show significant partial voluming were used for analysis. Only the tissues within the brain (i.e. cerebro-spinal fluid (CSF), grey matter (GM) and white matter (WM)) were considered. The Brainweb simulator includes glial matter as a separate class, and the glial matter phantom consists of a thin surface lying along the inside of the ventricles. Ignoring this class would therefore remove much of the partial voluming from this area of the brain. Therefore, the glial matter was treated as white matter: the glial matter phantom was added to the white matter phantom to produce the final white matter phantom used in these experiments All phantoms were scaled to lie between 0 and 1, such that the intensities represented tissue volume estimates within each voxel. The noise-free, inhomogeneity-free, 1.0mm slice thickness Brainweb simulations of T1-, T2- and PD-weighted images were then downloaded and loaded into TINA using the same procedure adopted for the phantoms. Pure CSF, WM and GM voxels were identified manually and their intensities measured (see Table 1), scaled such that the highest intensity tissue had a value of 250 grey levels.

In order to re-simulate the Brainweb images, the scaled CSF, WM and GM+glial phantoms were multiplied by the mean tissue intensities listed in Table 1, and added together. The summed images were then blurred by convolution with a Gaussian kernel of $\sigma = 0.8$ pixels. This replicated the point spread function simulation included in Brainweb. The kernel size was determined by matching the edge gradients in the re-simulated images to those in the Brainweb simulated images (see Fig.2). Finally, Gaussian image noise was added with standard deviations of 7.2, 12 and 16.8 grey levels. These values were chosen to match the absolute values of the noise present in the Brainweb 3%, 5% and 7% noise simulations, taking account of the scaling applied, such that results derived using the re-simulated images can be directly compared to those derived on Brainweb simulated images. Example images and intensity histograms at 5% noise are shown in Fig.1A Gaussian noise distribution was used here, in contrast to the Rician noise present in Brainweb simulations. However, differences between the two distributions are insignificant at signal-to-noise ratios greater than 3, and since only tissue within the brain were considered in the experiments described here the signal-to-noise ratio never fell below 4.2, so the use of Gaussian rather than Rician noise is justified. Figure 3 shows the intensity histograms of slice 142 of the noise-free, inhomogeneity-free, T1-weighted Brainweb simulated image and the equivalent re-simulated image. The histograms are essentially identical with the exception of the artefactual secondary peaks in the histogram identified in TINA Memo no. 2008-002.

Monte-Carlo experiments were performed on each set of re-simulated images. In order to limit the processor time required, segmentations were performed on 9 slices from each volume (slices 142-150 inclusive) rather than the entire volumes. This amounted to 353,493 image voxels, of which 186,742 contained tissues and the remainder were background voxels. These particular slices were chosen as they covered the top of the cerebellum, so were rich in structure and contained significant amounts of partial voluming, and thus presented the most difficult segmentation problem available from any set of slices in the whole image volumes. Initial tissue models were prepared containing the parameters used in the re-simulation, and tissue volumes measured from the phantoms, such that they represented the correct solution. This prevented fit failures, so the Monte-Carlo experiments measured the effects of image noise on segmentation accuracy. 50 segmentations were performed for each set of images, with independent Gaussian noise fields for each segmentation, both with and without the use of gradient terms in the model.

Two performance metrics were applied to the results. First, partial volume maps of the CSF, GM and WM+glial tissues were produced, and binarized at the 0.5 level to produce tissue classifications (i.e. voxels were classified as the tissue comprising the majority of their volume). These were compared to the original phantoms, and the number of misclassified pixels was counted. Misclassification rate considers all tissues present in the images, and so provides a more easily interpretable measure of performance than alternatives such as the Dice coefficient, which only consider single tissues [5]. However, since the TINA medical image segmentation algorithm was designed to measure partial volumes accurately, and the additional information utilised by including gradient terms in the tissue model is limited to the partial volume voxels (since these are the voxels occurring at tissue boundaries i.e. points of non-zero gradient), simple classification-based measures do not reveal the increased accuracy of segmentations performed using gradient terms in the tissue model. Therefore, a second performance metric was also applied. The segmentation result was used to produce noise-free estimates of the original images, by multiplying the partial volume estimates by the pure tissue mean intensities from the fitted tissue models. This produced representations of the fitted tissue models that could be directly compared to the original images. The squared difference between the two was divided by the squared image noise to produce a chi-squared type measure. Note that this is not equivalent to a true chi-squared metric as the segmentation errors on pure tissue voxels are expected to be zero: instead, it is dominated by the accuracy of the tissue volume estimates in partial volume voxels.

In order to provide a benchmark for the misclassification rate, the Bayes error was measured. This was done by measuring the number of voxels in each re-simulated image that were moved across the midpoints between pure tissue mean intensities by the Gaussian blurring and Gaussian noise addition, therefore representing the minimum number of misclassifications for a perfect model fit compared to the tissue phantoms.

Sequence	CSF Mean	GM Mean	WM mean
T1	78	187	250
T2	250	100	71
PD	250	218	185

Table 1: Pure tissue mean intensities used in the re-simulation of the Brainweb images.

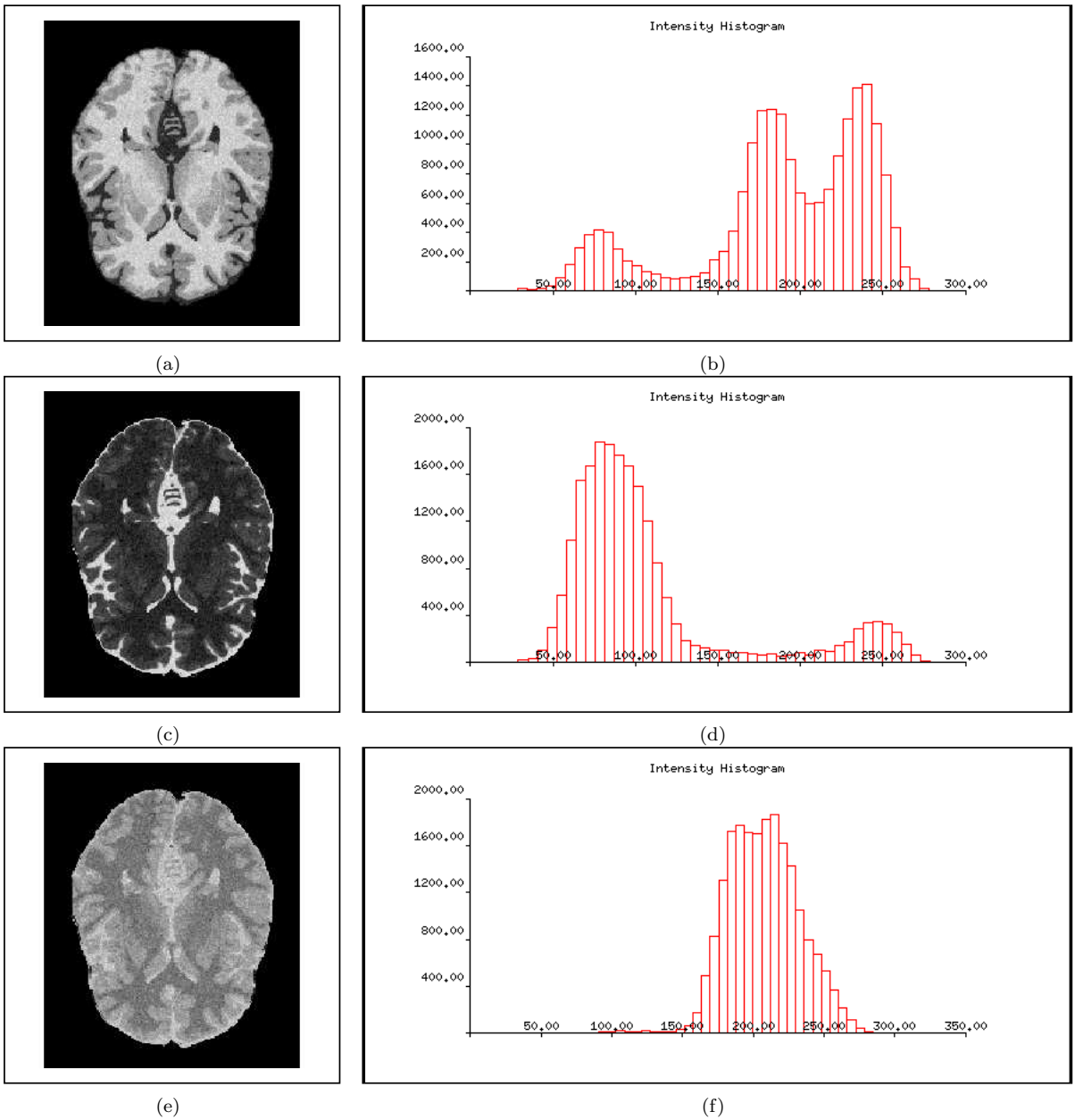
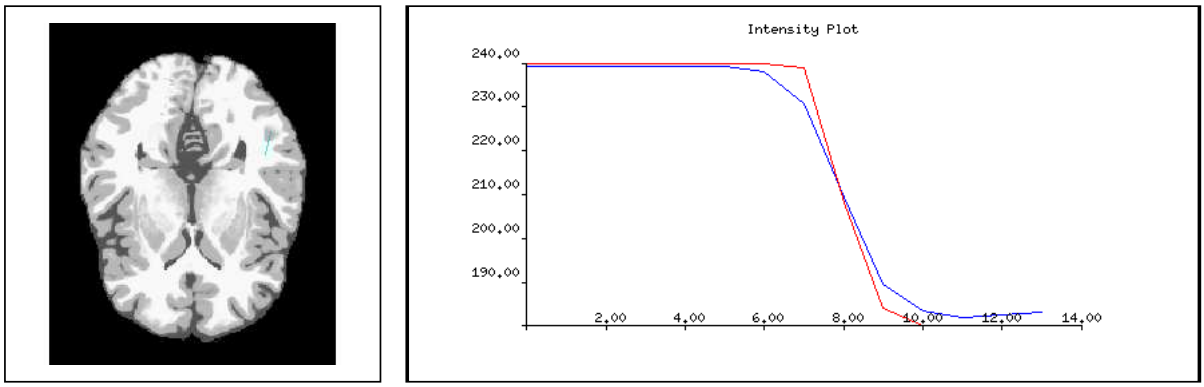


Figure 1: Example images and intensity histograms of the re-simulated Brainweb T1 (a,b) T2 (c,d) and PD (e,f) images. The noise level is 5%.

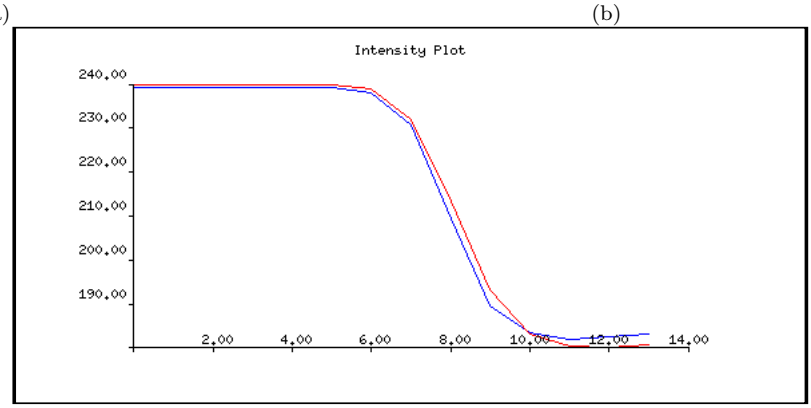
3 Results

In order to set the stopping criterion for the EM optimisation, an initial experiment was performed using the T1 weighted re-simulated images with 5% noise, in which the optimisation was run for 50 iterations and the intensity model parameters recorded after each iteration. The results are shown in Fig.4. They demonstrate that the optimisation has broadly converged after a few tens of iterations. Therefore, in all subsequent experiments the optimisation was halted after 40 iterations. These results also show that the errors on the fitting of the tissue model mean intensity parameters are roughly halved when gradient terms are used in the model, compared to when they are not used.

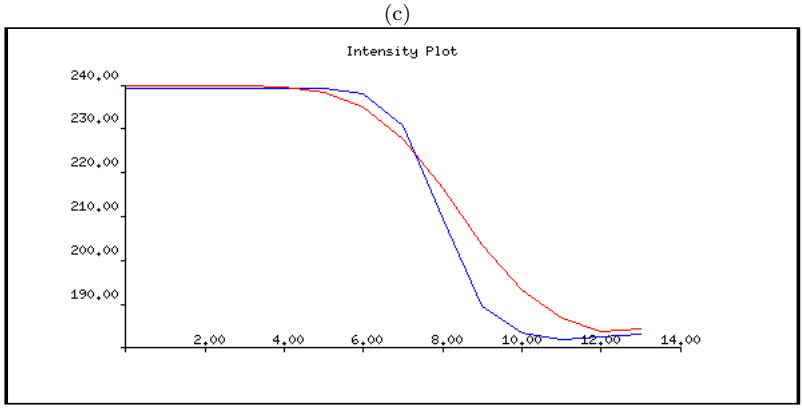


(a)

(b)



(c)

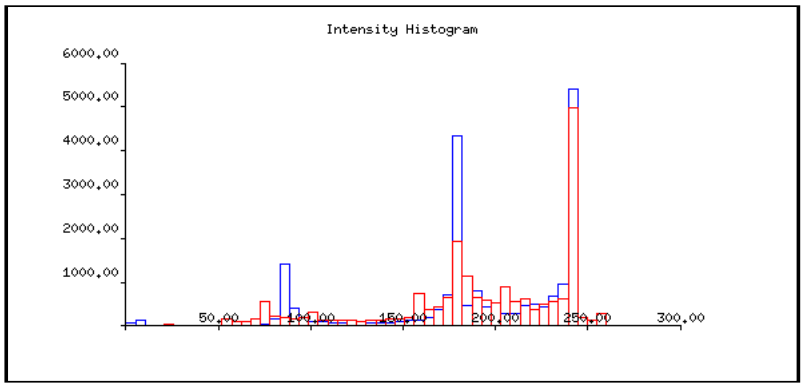


(d)

Figure 2: Intensity profiles along the line shown in (a) in the original Brainweb simulations (blue) and re-simulated images (red) after no blurring (b), Gaussian blurring at $\sigma = 0.8$ voxels (b), and at $\sigma = 1.6$ voxels (c).

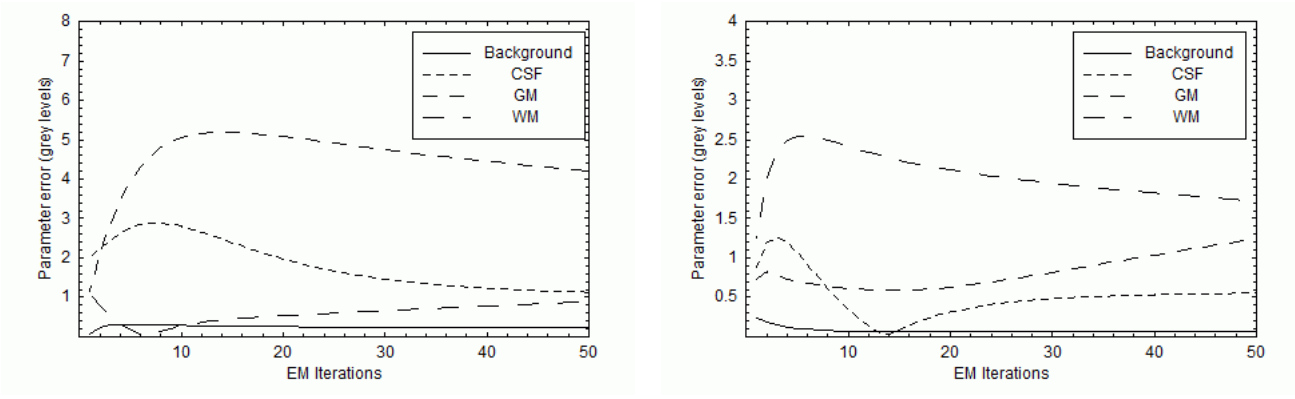
Noise Level (%)	T1	T2	PD
No gradients			
3	7.282 ± 0.038	13.292 ± 0.054	14.738 ± 0.054
5	10.213 ± 0.054	19.817 ± 0.074	23.840 ± 0.095
7	14.301 ± 0.072	25.689 ± 0.082	31.491 ± 0.101
With gradients			
3	6.972 ± 0.041	12.307 ± 0.060	12.916 ± 0.054
5	10.143 ± 0.058	19.603 ± 0.162	21.816 ± 0.085
7	14.226 ± 0.073	25.460 ± 0.104	29.497 ± 0.096
Bayes error			
3	7.211 ± 0.040	13.159 ± 0.053	15.141 ± 0.068
5	10.146 ± 0.056	19.373 ± 0.089	23.967 ± 0.087
7	14.267 ± 0.061	25.272 ± 0.081	31.370 ± 0.089

Table 2: Voxel misclassification rates (in percentages of the WM, GM and CSF voxels combined i.e. ignoring the background).



(a)

Figure 3: Intensity histogram of the noise-free, Brainweb T1 simulated image (red) and the equivalent re-simulated image (blue).



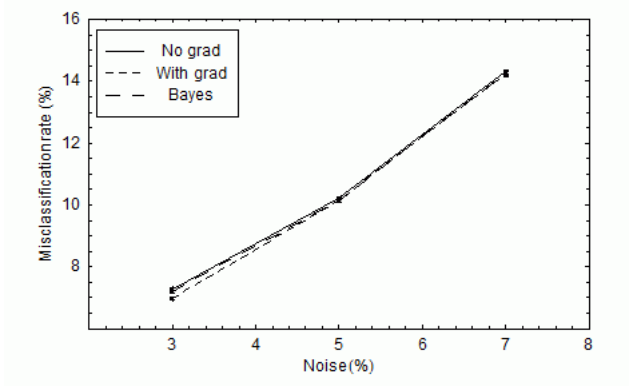
(a) (b)

Figure 4: Values of the tissue mean intensity parameters during 50 iterations of EM optimisation using the T1 weighted images with 5% noise without gradient terms in the tissue model (a) and with gradient terms (b).

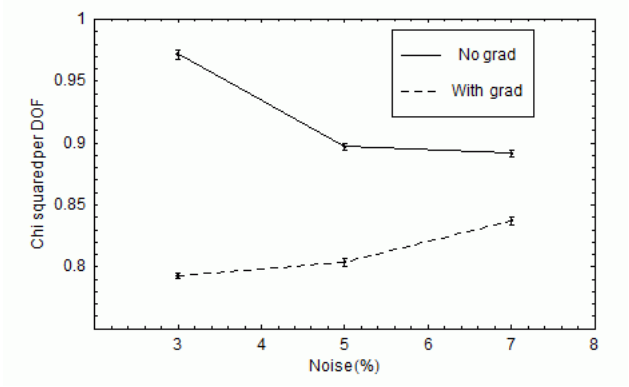
Noise Level (%)	T1	T2	PD
No gradients			
3	0.9717 ± 0.0039	1.9727 ± 0.0618	0.8559 ± 0.0033
5	0.8973 ± 0.0027	1.1717 ± 0.0136	0.7866 ± 0.0026
7	0.8918 ± 0.0027	1.0592 ± 0.0073	0.6688 ± 0.0022
With gradients			
3	0.7927 ± 0.0023	0.8306 ± 0.0047	0.7834 ± 0.0034
5	0.8037 ± 0.0032	0.7665 ± 0.0045	0.7109 ± 0.0027
7	0.8373 ± 0.0033	0.8773 ± 0.0048	0.5913 ± 0.0020

Table 3: Chi-squared per DOF between the reconstructed noise-free tissue model and the noise-free simulated data.

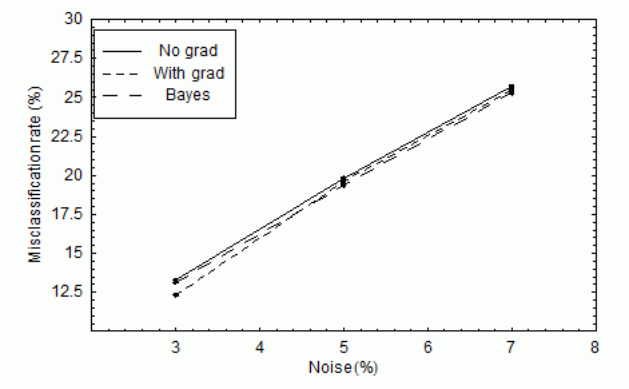
Table 2 and Fig. 5 show the misclassification rates produced by the segmentation algorithm, as percentages of the number of CSF, GM and WM voxels (i.e. ignoring voxels in the background). The Bayes errors for intensity-based segmentations of these images are also shown. When gradient terms are not used, the TINA medical image segmentation algorithm achieves almost optimal misclassification rates. The difference is due to errors on the tissue model parameter estimation introduced by the presence of image noise. In the cases of the T1- and T2-weighted images, inclusion of gradient terms does not reduce the misclassification rates. However, in the PD-weighted images the pure tissue intensity distributions overlap to a considerable extent (see Fig.1), obscuring the partial volume distributions. The inclusion of gradient information in the tissue model disambiguates the pure tissue and partial volume distributions in this case, since partial volume voxels may have the same intensity as pure tissue voxels but will occur at tissue boundaries and so have higher gradients. This leads to an improvement in misclassification



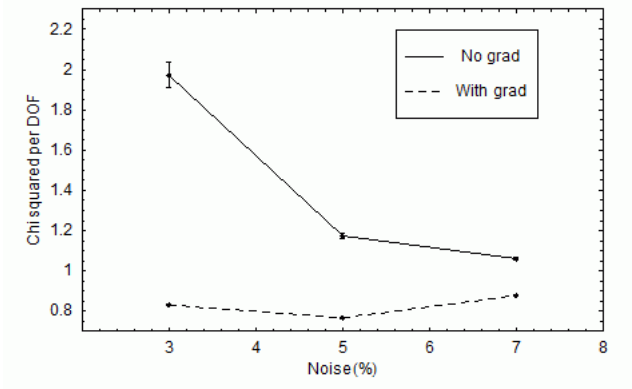
(a) T1 % Misclassification



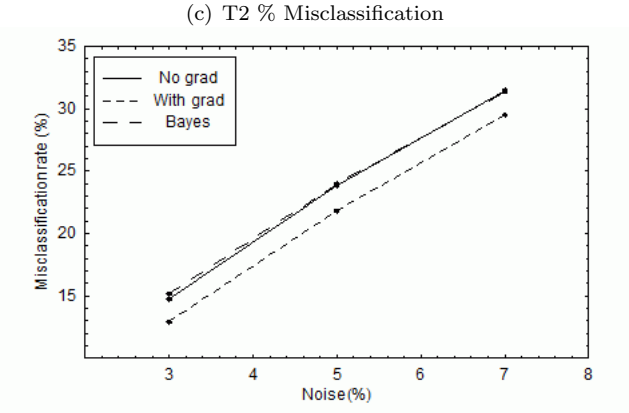
(b) T1 χ^2 per DOF



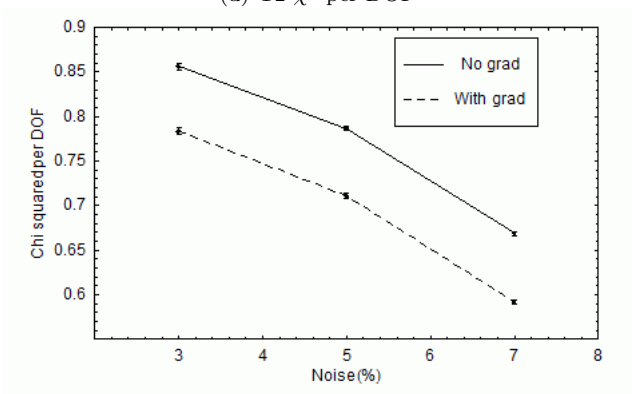
(c) T2 % Misclassification



(d) T2 χ^2 per DOF



(e) PD % Misclassification



(f) PD χ^2 per DOF

Figure 5: Voxel misclassification rates (a,c,e) and χ^2 per DOF (b,d,f) for segmentations of the re-simulated Brainweb T1 (a,b) T2 (c,d) and PD (e,f) images. The abscissa shows the image noise in all cases (see main text for details).

rates of around 2% when segmentation of the PD-weighted images is performed using gradient terms in the tissue model compared to when they are not used, exceeding the Bayes error rates for intensity-only segmentations.

Simple misclassification rates fail to reveal most of the difference between segmentations with and without gradient terms, since the use of gradient terms only increases the accuracy of partial volume estimation in voxels at tissue boundaries, and simple measurements of misclassification rates largely ignore this. Table 3 and Fig 5 show the chi-squared between images reconstructed from the fitted tissue model and partial volume maps and the original image data. This measure fully considers the accuracy of the tissue volume estimates, and so is more relevant when considering the differences introduced by the use of gradient terms. The inclusion of the gradient terms improves the quality of the segmentation result in all cases: the average improvement in the chi-squared when the gradient terms are used is 31.7%.

4 Conclusions

The results presented here demonstrate that the TINA medical image segmentation algorithm is capable of achieving close to optimal performance on simulated images when gradient terms are not included, and that the use of gradient terms in the tissue model can increase the accuracy of tissue classification results by several percent in PD-weighted images, where the intensity distributions for the pure tissues overlap to a considerable extent, obscuring the partial volume distributions. Inclusion of gradient terms improves the accuracy of tissue volume estimation by an average of around 31.7% in partial volume voxels. Whilst this is a relatively small difference when tissue volume measurements over the whole brain are required, it may be significant when measurements of small volume changes (e.g. in small tumours, or when applying algorithms such as the boundary shift integral) are required.

References

[1] M Pokric, N A Thacker, M L J Scott, and A Jackson. Tina memo 2001-009: Multi-dimensional medical image segmentation with partial voluming. Technical report, Imaging Science and Biomedical Engineering Division, University of Manchester, 2001.

[2] N A Thacker. Tina memo 2004-006: Parameter estimation for em mixture modelling and its relationship to likelihood and eml. Technical report, Imaging Science and Biomedical Engineering Division, University of Manchester, 2004.

[3] N A Thacker, M Pokric, and A J Lacey. Tina memo 2004-009: Model selection and convergence of the em algorithm. Technical report, Imaging Science and Biomedical Engineering Division, University of Manchester, 2001.

[4] N A Thacker, M Pokric, and D C Williamson. Tina memo 2004-009: Multi-dimensional medical image segmentation with partial voluming and gradient estimation. Technical report, Imaging Science and Biomedical Engineering Division, University of Manchester, 2004.

[5] J Tohka, E Krestyannikov, I D Dinov, A MacKenzie Graham, D W Shattuck, U Ruotsaainen, and A W Toga. Genetic algorithms for finite mixture model based voxel classification in neuroimaging. *IEEE Trans Med Imag*, 26:696–711, 2007.

[6] D C Williamson, N A Thacker, S R Williams, and M Pokrić. Partial volume tissue segmentation using grey-level gradient. In *Proc. MIUA*, pages 17–20, 2002.