Tina Memo No. 2005-013
Internal Memo

Stability Testing of the TINA Medical Image Segmentation Algorithm

P. A. Bromiley and N.A. Thacker

Last updated
19 / 9 / 2005

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

Stability Testing of the TINA Medical Image Segmentation Algorithm

P. A. Bromiley and N.A. Thacker.
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 stability testing of the TINA medical image segmentation algorithm described in Memo 2004-009. The performance of the algorithm when applied to a common medical image segmentation problem, the segmentation of structural MR image volumes of the brain into CSF, grey matter and white matter regions, was evaluated. We examine the stability of the optimised tissue model after the optimisation is initiated from multiple random starting points. This method allows comparative evaluation of the results without the need to define a gold-standard segmentation. A straw-man algorithm, based on simplex optimisation of a χ^2 histogram fit of the model to the images, is used to provide a baseline for this comparison. The superiority of the EM-based algorithm over this simpler alternative, and the improvements gained through the inclusion of gradient information in the tissue model, are demonstrated.

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 [8], 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 [7], [3], [9], and [5].

In order to evaluate the stability of the TINA medical image segmentation algorithm, a common medical segmentation task was used: the identification of CSF, grey matter and white matter in structural MR image volumes of the brain. The evaluation was performed by collecting MR image volumes from a large number of normal subjects, segmenting them, and calculating the standard deviation of the final model parameters for the tissue means. The drawback of this approach is that it supplies no quantitative measure of the accuracy of the segmentation, such as would be obtained through comparison with a gold-standard segmentation. However, the results can be evaluated by comparing the stability of the model parameters across several different segmentation algorithms, and thus provide a qualitative measure of the accuracy of the algorithms in comparison with one another, without having to address the problem of defining a gold-standard result.

Three segmentation algorithms were compared in the work described here. The main goal of the work was to compare the EM-based algorithm with and without the inclusion of image gradient information. The inclusion of image gradient information increases the separability of pure tissue and partial volume components. Consider images containing three tissue types: partial volume voxels containing mixtures of the tissues having the highest and lowest mean grey-level values can themselves have grey-levels at any value between the pure tissue mean grey-levels. The third tissue must also have a mean grey-level value somewhere within this range, leading to confusion between these pure tissue and partial volume components. However, pure tissue voxels will occur within tissue regions, at locations of low image gradient, whereas partial volume voxels will occur at the boundaries of tissue regions, at locations of high image gradient. Therefore, the inclusion of gradient information will increase the separability of pure tissue and partial volume components.

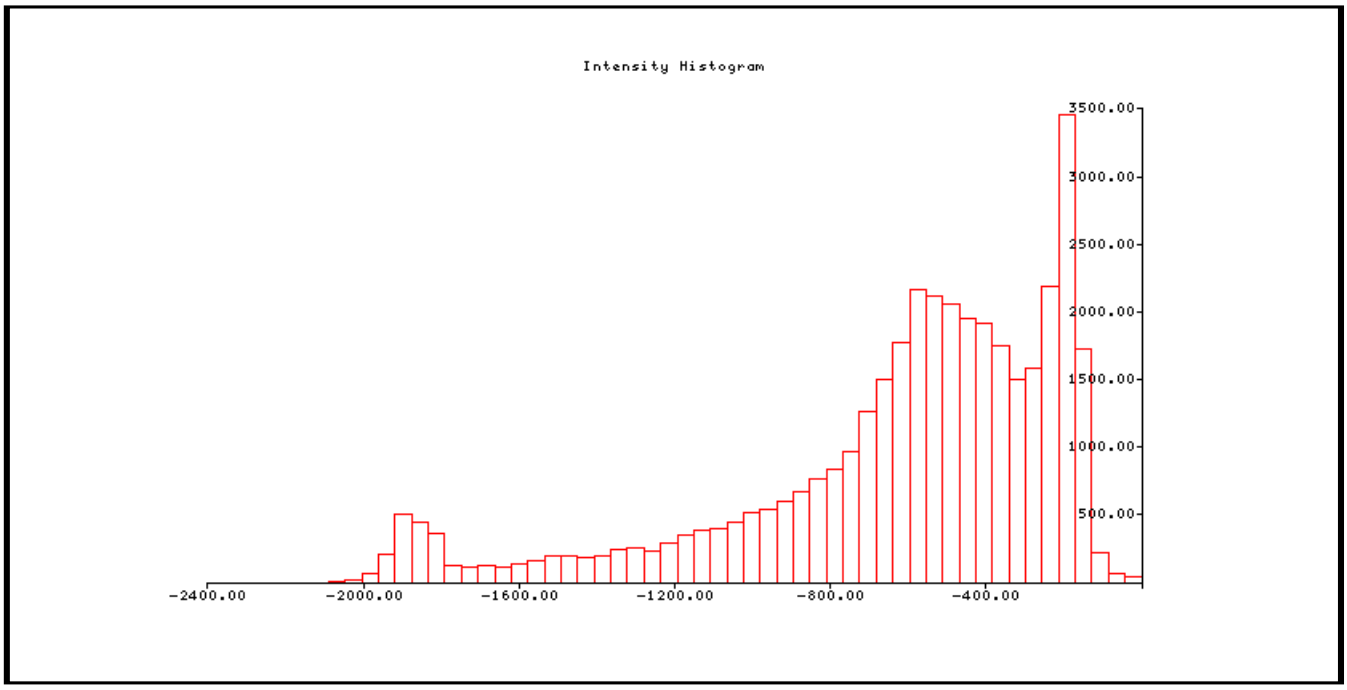

Figure 1: Typical image histogram of the data used in the stability analysis, showing peaks for CSF (at the lower end), grey matter (in the middle) and white matter (at the upper end).

A straw-man algorithm was included in the evaluation, in order to provide a base-line for the comparison. This consisted of fitting the same model used in the EM-based algorithms to the histogram of an image volume. A χ^2 goodness-of-fit metric was calculated, and optimised using the simplex algorithm [4].

2 Method

The stability analysis was performed by comparing the results from three segmentation algorithms: the EM-based algorithm applied to grey-levels alone i.e. without gradient information, the EM-based algorithm applied to both grey-levels and image gradients, and the straw-man algorithm. The model was the same in all three cases, and consisted of three pure tissue components (CSF, grey matter and white matter) represented by Gaussian distributions, and three partial volume components (for mixtures of each pair of tissues) represented by Gaussian distributions convolved with triangular distributions. Each component contained parameters for the mean, the standard deviation, and a prior term representing frequency of occurrence, giving a total of twelve parameters. All three segmentation algorithms required a starting point or initial model. Therefore, a standard initial model was defined for all three algorithms, with the CSF, grey matter and white matter peaks set to -1800, -650 and -120 grey-levels respectively.

The data used in the evaluation consisted of structural MR image volumes of the brain obtained from 67 normal volunteers (33 male and 34 female) ranging in age from 19 to 85 years, with a mean age of 66.5 years. The youngest nine of these, ranging in age from 19 to 53 years with a mean age of 35.4 years, were each scanned four times: two scans an hour apart, followed by another two scans an hour apart a year later. All other patients were scanned once. All subjects underwent MR imaging with a 1.5-T system (ACS-NT, with PowerTrack 6000 gradient subsystem; Phillips Medical Systems, Hamburg, Germany) with a birdcage head coil receiver. Fast spin-echo inversion-recovery images (repetition time, 6850 msec; echo time, 18 msec; inversion time, 300msec; echo train length, 9) were acquired. The image volumes consisted of 3-mm thick sections throughout the brain, with an in-plane resolution of $0.89mm^2$ (matrix, 256×204, field of view, $230 \times 184mm$). All MR volumes were coregistered to a single standard volume using rigid Mutual Information based coregistration [1, 2], allowing the definition of consistent regions across all image volumes.

The image volumes consisted of 50 images of 256 by 256 voxels, covering the whole head. A sub-region was defined in the standard coordinate system, containing only CSF, grey matter and white matter, spanning the front six slices containing the ventricles. The sub-region consisted of a total of 46,200 voxels. Its purpose was to limit the

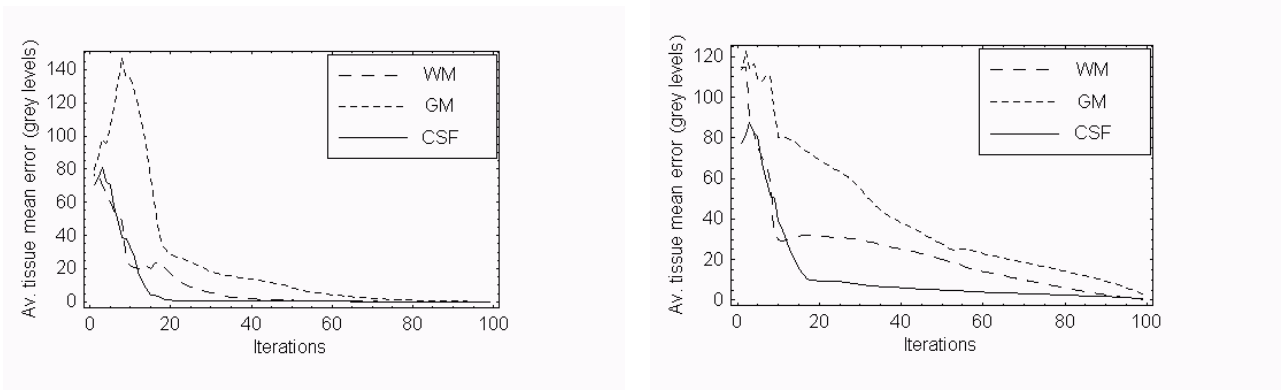

Figure 2: The average difference between the CSF, grey matter and white matter tissue means after 100 iterations, and the means at each iteration of the optimisation, averaged over ten data sets, for the EM-based segmentation with (a) and without (b) gradient.

number of model components required to fit the data, reducing the required processor time and simplifying the analysis of the results. A typical histogram of the data contained in the sub-region is shown in Fig.1: the peaks representing CSF, grey matter and white matter are clearly visible.

The stability analysis consisted of several experiments, involving segmentation of sub-sets of the MR image data, followed by calculation of the standard deviations of the final model parameters for the tissue means. In each case, outlier filtering was performed in an iterative manner, through calculation of an initial estimate of the standard deviation, then removal of results which deviated from the mean result by more than two standard deviations i.e. at the 95% confidence limit, followed by a calculation of the final standard deviation from this reduced data set. This procedure removed any results representing fit failures.

Several preparatory experiments were performed in order to test various aspects of the implementation of the algorithms. The first evaluated the stopping criteria of the EM optimisation. The simplex-based segmentation incorporated a stopping criteria based on the step-to-step change in the model parameters during the optimisation, terminating when this quantity fell below a user-defined accuracy goal (set to 1×10^{-6} in terms of the proportional difference between the highest cost and lowest cost points of the simplex). However, due to the considerable computational load of the EM-algorithm it was impractical to implement such an accuracy goal. Therefore, a stopping criteria defined on the basis of processor time was defined empirically. Ten image volumes were selected at random from those collected, and the EM-based segmentation was applied both with and without the use of gradient information. The algorithm was run for 100 iterations, and the model recorded after each iteration. Figure 2 shows the absolute difference between the final tissue mean parameters and those recorded after each iteration, averaged over all ten data sets. It is immediately clear that the inclusion of gradient information leads to faster convergence. In the experiments where gradient information was included, the model parameters for CSF and white matter converged to within one grey-level of their final value within 40 iterations of the algorithm. Convergence of the peak for grey matter was slower, due to the greater width of the grey matter peak as seen in Fig.1. However, convergence to within 20 grey-levels of the final result occurred within 40 iterations for both algorithms. Therefore, in subsequent experiments the EM-based segmentation algorithms were run for 40 iterations.

The second preparatory experiment evaluated the effects of applying smoothing to the data in order to stabilise the histogram. Iterated tangential smoothing [6] was applied in order to avoid significant broadening of the peaks in the histogram. Various numbers of iterations of tangential smoothing were applied to the data prior to segmentation, and all image volumes were segmented with all three algorithms. The standard deviations, after outlier rejection, for the CSF, grey matter and white matter tissue means, and the mean of the standard deviations for all three tissues, are shown in Fig.3. It is clear from the results that the optimal blurring level occurs at between one and two iterations of smoothing. Therefore, in all subsequent experiments, one iteration of tangential smoothing was applied to the data prior to segmentation.

Once the implementation of the segmentation algorithms had been optimised in this way, two experiments were performed in order to evaluate the stability of the algorithms. The first involved repeated segmentations of the same data set. The initial tissue mean parameters were perturbed through the addition of a random offset, distributed uniformly between $\mu_t \pm n$ where μ_t was the initial tissue mean, and n was a constant. 40 segmentations were performed with each algorithm, for each of several values of n between 50 and 200 grey-levels. The standard deviations of the optimised tissue means were then calculated for each level of initial model randomisation. The

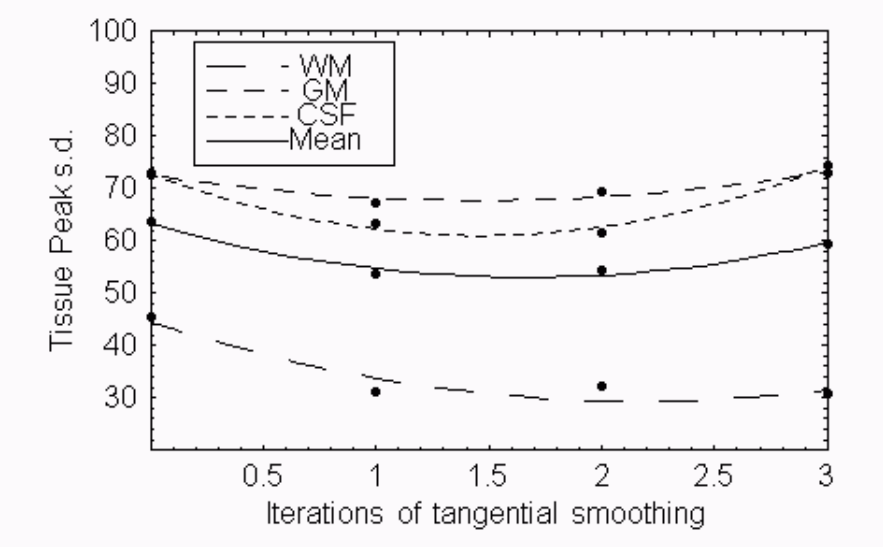

Figure 3: Standard deviations of the CSF, grey matter, and white matter tissue means of fits to all 97 image volumes, and the mean of the three sets of results, plotted against the number of iterations of tangential smoothing applied to the data prior to model fitting. The curves show quadratic fits to the results.

experiment was repeated across ten data sets, chosen at random from the 97 data sets collected, and the average standard deviations of the final tissue means across the ten data sets were calculated. The aim of this experiment was to evaluate the inherent stability of the algorithm, by eliminating the biological variability that would be present in segmentations of multiple data sets. The maximum level of randomisation applied was chosen such that it was less than the differences between the tissue means, in order to avoid interchanging the components.

The second experiment was designed to demonstrate the improvements gained by using gradient information in the EM-based algorithm in a scenario which more accurately reflected practical medical image analysis tasks. All 97 image volumes were segmented both with and without the inclusion of gradient information. The standard deviations of the CSF, grey matter and white matter peaks, after outlier rejection, were calculated. The experiment was repeated at levels of initial model randomisation between 0 and 225 grey-levels.

3 Results

Figure 4 shows the standard deviations, after outlier rejection, of the final CSF, grey mater and white matter tissue means for 40 segmentations of single data sets at each of various levels of initial model randomisation, averaged over ten data sets. Since the underlying data remained the same across all segmentations, any variation in the results reflects the ability of the segmentation algorithm to converge to a stable result. A clear difference between the three algorithms is demonstrated: the EM-based algorithm without gradient information is roughly an order of magnitude more stable than the straw-man algorithm, and the EM-based algorithm with gradient information is roughly an order of magnitude more stable than that without gradient information.

The two EM-based algorithms were identical in every respect except for the inclusion of gradient information, and the differences between them were greater than could be accounted for due to the slower rate of convergence in the algorithm without gradient information (as shown in Fig.2). Therefore, the difference between them must reflect the additional stability gained through the inclusion of that information, and so these results demonstrate the importance of the inclusion of partial volume components in models used for medical image segmentation.

The EM-based algorithm without gradient information and the straw-man algorithm were based on identical models: therefore the difference between these two algorithms must reflect the superiority of the EM-based approach over naive histogram fitting segmentation algorithms. The most probable cause of the difference in the results is premature convergence in the straw-man algorithm i.e. the simplex optimisation terminates on locating a local minimum, representing a peak in the histogram due to a noise fluctuation, rather than the global minimum representing the true tissue mean. This behaviour would be expected, given that the EM-based approach has a greater ability to move around the cost function space.

The superiority of the EM-based approaches over naive histogram fitting were demonstrated in the first experiment.

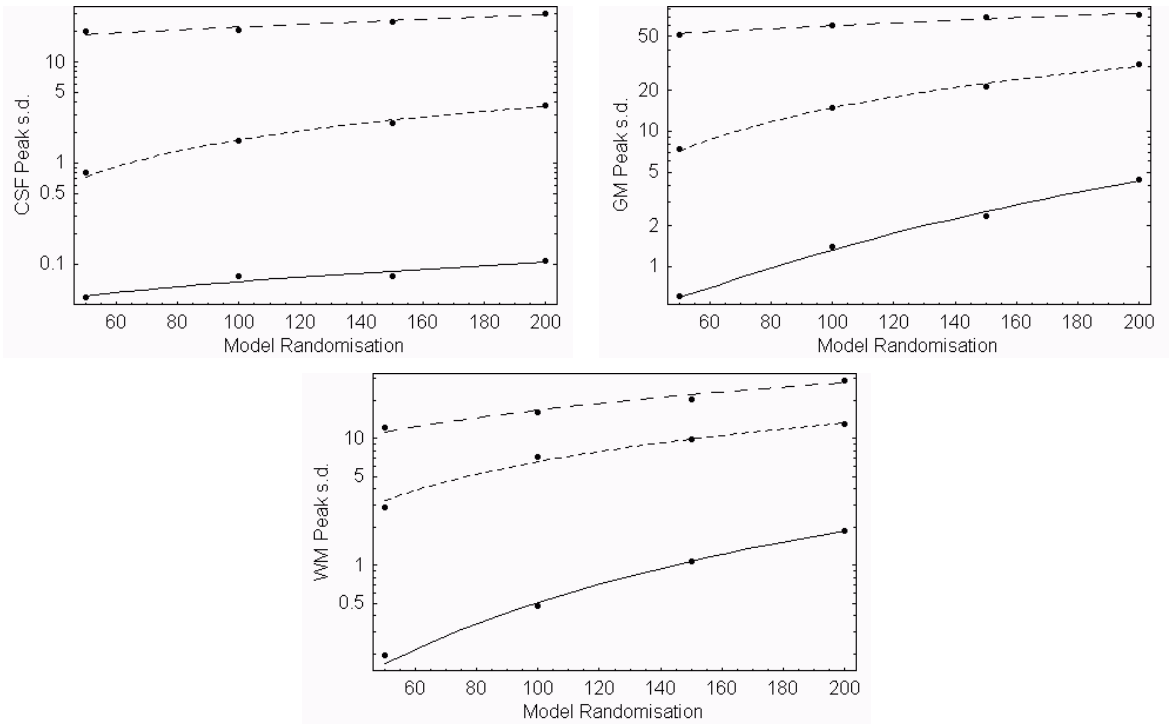

Figure 4: Standard deviations of the CSF (a), grey matter (b), and white matter (c) tissue means for 40 fits to a single MR image volume, plotted against the level of randomisation applies to the initial model. In each case, the upper curve shows the results from the χ^2 fit optimised using simplex, the middle curve the results from the EM-based segmentation without gradient, and the lower curve the results from the EM-based algorithm with gradient. The curves show linear fits to the data (note that the abscissa is plotted on a logarithmic scale).

Therefore, the second experiment was limited to comparisons of the EM-based segmentation with and without gradient information. The experiment involved the segmentation of all 97 MR data sets, with various levels of initial model randomisation. The results therefore reflect the expected improvements, gained through the inclusion of gradient information in the segmentation, for large medical image analysis projects which involve a segmentation stage.

The standard deviation of the optimised CSF, grey matter and white matter tissue means, after outlier rejection, plotted against the level of initial model randomisation, are shown in Fig5. The results show that the inclusion of gradient information results in roughly a 50% reduction in the standard deviations: again the difference is larger than could be accounted for by differences in rates of convergence.

4 Conclusions

Segmentation of structural MR images of the brain is a common task in medical image analysis, and a large variety of segmentation algorithms intended to address this task have been described in the literature. Evaluation of the accuracy of such algorithms usually requires data accompanied by gold-standard segmentation, against which the algorithmic results can be compared. However, the definition of such gold-standard results is problematic. In the work presented here, this issue has been avoided by comparing segmentation results from several different algorithms when applied to a large data set.

The work described here evaluated three segmentation algorithms: an EM-based algorithm applied both with and without the inclusion of gradient information, and a straw-man algorithm consisting of simplex optimisation of a χ^2 goodness-of-fit measure of the partial volume model to the histogram of the image data, providing a baseline for the comparative analysis. The first stability analysis experiment involved repeated segmentations of single data sets with various levels of randomisation applied to the initial model. The results indicated that the TINA medial image segmentation algorithm, based on the Expectation-Maximisation algorithm, was roughly an order of magnitude more stable than naive histogram fitting approaches, and furthermore that the inclusion of gradient information resulted in a further order of magnitude improvement in stability. The inclusion of gradient

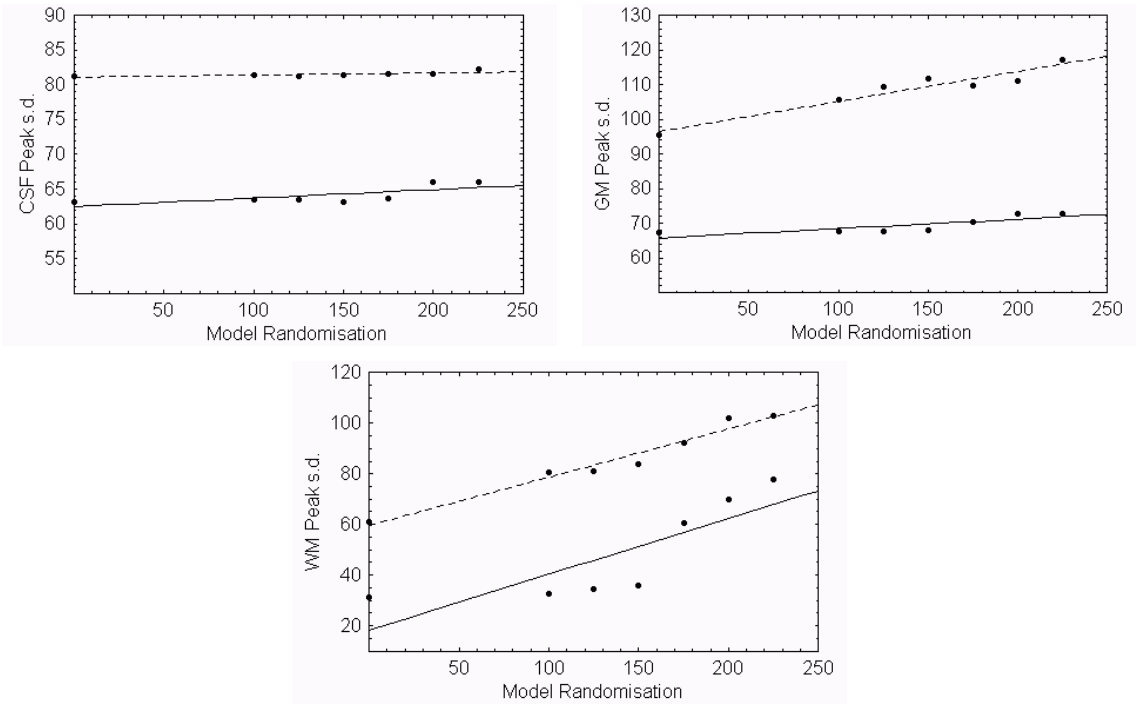

Figure 5: Standard deviations of the CSF (a), grey matter (b), and white matter (c) tissue means for fits to all 97 MR image volumes, plotted against the level of randomisation applied to the initial model. In each case, the upper curve shows the results from the EM-based segmentation without gradient, and the lower curve the results from the EM-based algorithm with gradient. The curves show linear fits to the data.

information is significant only at positions in the images representing boundaries between tissues: therefore, this result demonstrates the importance of the partial volume components on the stability of the tissue means.

Practical medical image analysis projects that include a segmentation component typically involve a large number of data sets, and the second experiment was designed to reflect this scenario. The variance of the mean tissue parameters of the final segmentation results in this experiment represented the convolution of two distributions: one representing the inherent inter-subject biological variability of the underlying parameters, and a second representing the additional variance introduced by the segmentation algorithm. As an aside, it should be noted that the result of the first and second experiments combined do not provide any quantitative estimate of the accuracy of the segmentation algorithms studied: the sum and difference of two numbers do not provide enough information to calculate those numbers. However, the number of data sets included in the second experiment (≈ 100) is representative of the typical number of subjects included in significant medical image analysis projects. Therefore the results, combined with error propagation, can be used to estimate the reduction in segmentation errors that would be gained through the inclusion of gradient information in the segmentation in such projects.

Acknowledgements

This work was funded by the MIAS (Medical Images and Signals) IRC under EPSRC grant no. GR/N14248/01 and the UK Medical Research Council Grant No. D2025/31. All software is freely available from our web site www.tina-vision.net.

References

[1] P A Bromiley, M Pokric, and N A Thacker. Computing covariances for mutual information coregistration. In *Proceedings MIUA'04*, pages 77–80, 2004.

[2] P A Bromiley, M Pokric, and N A Thacker. Emprical evaluation of covariance estimates for mutual information coregistration. In *Proceedings MICCAI'04*, pages 607–614, 2004.

[3] 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.

[4] W H Press, B P Flannery, S A Teukolsky, and W T Vetterling. *Numerical Recipes in C*. Cambridge University Press, New York, 2nd edition, 1992.

[5] 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.

[6] N A Thacker and M Pokric. Tina memo 2003-007: Noise filtering and testing for mr using a multi-dimensional partial volume model. Technical report, Imaging Science and Biomedical Engineering Division, University of Manchester, 2004.

[7] 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.

[8] 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.

[9] D . Williamson, N . Thacker, S R Williams, and M Pokric. Tina memo 2002-006: Partial volume tissue segmentation using grey-level gradient. Technical report, Imaging Science and Biomedical Engineering Division, University of Manchester, 2002.