Improving the Accuracy of MR Image Segmentation through the use of Local Gradient Information

P. A. Bromiley* and N.A. Thacker

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

Abstract. Segmentation is a core technology in medical image analysis, providing a route to tissue volume estimation that can be used in a wide range of applications, such as monitoring the progress of, or the effects of drug therapies on, tumour growth or the effects of atrophic diseases. In general, the more image information we can extract to use in segmentation, the more accurate the results will be. In previous work we have proposed a unified mathematical framework for incorporating local image gradient into feature-space based segmentation algorithms. In this paper we demonstrate, using simulated MR images of the normal brain, that the additional information present in the gradients can significantly improve segmentation accuracy.

1 Introduction

The physical processes underlying medical imaging equipment such as computed tomography (CT) and magnetic resonance imaging (MRI) result in the production of images in which the contrast between tissues is determined by their physical properties, such as X-ray attenuation or proton density. Accurate segmentation of images offers the opportunity to produce parametric images of tissue type (i.e. grey matter, white matter, tumour etc.) that are more relevant to clinical investigation. Once the images are segmented and tissue models obtained they can be used for extraction of tissue boundaries or quantitative estimation of volume.

One of the most significant artefacts associated with medical images in relation to segmentation accuracy is the partial volume (PV) effect, where image voxels contain a mixture of pure tissues. Such data cannot be analysed using a set of mutually exclusive, pure-tissue hypotheses; for example, [1] demonstrated that consistently misplacing the tissue boundaries in a segmentation of a 1mm isotropic MR brain image by a single voxel in each slice resulted in errors of approximately 30%, 40% and 60% on the estimated volumes of white matter, grey matter and cerebrospinal fluid (CSF) respectively. In common with most aspects of image segmentation, the development of probabilistic PV models for MR imaging can be tackled either in the image plane or in some feature space, most commonly the intensity histogram. The first approach was initiated by Choi et al. [2], assuming that the tissue mixing proportions varied smoothly across the image and introducing these spatial interactions using a Markov Random Field model. However, feature-space based approaches have a significant advantage in that they can be easily extended to utilise multi-dimensional data, simply by adding the intensities of additional images as extra dimensions of the feature space. This is particularly relevant in the case of MR imaging, where multiple images of the same subject with different contrast information for each tissue can be obtained using different pulse sequences. Additional features, such as local image gradient, can also be added to the feature space in the same way, further increasing the quantity of information available to the segmentation algorithm.

The development of feature-space based approaches to PV models for medical image segmentation was initiated by Santago and Gage [3], based on the observation that the physics of the image formation processes in a wide variety of medical imaging modalities allows PV distributions for paired tissue combinations to be modelled as a simple, linear process [4]. Relative fractions of different tissues therefore contribute proportionately to the intensity in a given voxel. In previous work [5–9] we extended the density model proposed by Santago and Gage to generate a unified mathematical framework incorporating both intensity and local image gradient, representing image intensities as vectors in a multi-dimensional feature space to allow simultaneous segmentation of multiple image volumes from the same subject. The parameters of the multi-dimensional probability density models are estimated using the Expectation-Maximisation (EM) algorithm. Bayes theory is then used to estimate the most probable proportions of each tissue within each voxel, rather than a hard classification to a single tissue type. In this paper we apply the proposed algorithm, both with and without the utilisation of local image gradients, to simulated MR images of the normal brain from Brainweb [10–13]. The results are compared to the tissue phantoms used to generate the simulated data, in order to quantitatively demonstrate the improvements in segmentation accuracy gained through the addition of gradient information to the feature space in segmentations of uni-spectral data.

*E-mail: paul.bromiley@man.ac.uk

2 Method

A complete description of the proposed algorithm is provided in our previous papers [5–9]; we provide only a brief overview here. The approach proposed by Santago and Gage is based upon two main assumptions. The first is that, in the absence of noise and artefacts such as intensity inhomogeneity, each tissue t has a well-defined signal intensity $\mathbf{M_t}$. The probability density distribution $d_t(\mathbf{g})$ describing the intensity $\mathbf{g}$ (which may be a vector of intensities from multiple image volumes of the same subject) of voxels containing pure tissues is therefore given by a delta function located at $\mathbf{M_t}$ convolved with some noise distribution. The physics of image formation in MR, and some other medical imaging modalities such as CT, is linear and so the intensity of PV voxels is given by the sum of the intensities of the pure tissues they contain, weighted by the proportions of each tissue in the voxel

$$\mathbf{g} = \sum_{i=1}^{N} q_i \mathbf{M_i} + \eta \quad where \quad \sum_{i=1}^{N} q_i = 1$$

where $\mathbf{M_i}$ is the signal intensity of tissue i, q_i is the volumetric proportion of tissue i in the voxel such that $0 < q_i < 1$, N is the number of tissues, and η is some noise distribution. The second assumption made in the Santago and Gage approach is that all values of q_i are equally likely. Limiting consideration of partial voluming to pairs of tissues t and r, the probability density distribution $d_{tr+rt}(\mathbf{g})$ describing the intensities of PV voxels is therefore a uniform distribution linking the intensities $\mathbf{M_t}$ and $\mathbf{M_r}$ of the pure tissues, again convolved with some noise distribution. Assuming the noise to be Gaussian gives

$$d_t(\mathbf{g}) = \alpha_t e^{-\frac{1}{2}(\mathbf{g}-\mathbf{M_t})^T C_t^{-1}(\mathbf{g}-\mathbf{M_t})} \quad and \quad d_{tr+rt}(\mathbf{g}) = \int_0^1 \beta_{tr+rt} e^{-\frac{1}{2}(\mathbf{g}-\mathbf{M_{tr+rt}})^T C_{tr+rt}^{-1}(\mathbf{g}-\mathbf{M_{tr+rt}})} dq_t$$

where $\mathbf{M_{tr+rt}} = q_t\mathbf{M_t} + (1-q_t)\mathbf{M_r}$ is the expected intensity of a PV voxel containing a proportion q_t of tissue t and $1-q_t$ of tissue r, C_{tr+rt} is the covariance matrix in PV voxels (estimated in the same way), C_t is the covariance matrix of tissue t, and α_t and β_{tr+rt} are constants that provide unit normalisation. In practice, we divide the PV distributions into two components $d_{tr}(\mathbf{g})$ and $d_{rt}(\mathbf{g})$, each describing the volumetric contribution of one of the pure tissues to the PV voxels i.e. $d_{tr}(\mathbf{g})$ describes the contribution of tissue t to PV voxels also containing tissue r. These are triangular distributions convolved with Gaussians, which sum to reproduce a uniform distribution convolved with a Gaussian.

We define the multi-dimensional local image gradient as

$$s = \sqrt{\sum_{k=1}^{n} \frac{\nabla_x I_k^2 + \nabla_y I_k^2}{\sigma_k^2}} - \lambda$$

where $\nabla_x I_k$ is the local gradient of image I_k in the x direction, n is the number of image volumes in the data, σ is the standard deviation of the image noise and λ is a fit parameter. The z component, i.e. the gradient in the inter-slice direction, is deliberately omitted so that the algorithm can be applied to single slices as well as image volumes. The probability density distribution describing the gradient at each voxel $\rho_n(s, \mathbf{g})$ must clearly be a function both of gradient and intensity. However, the algorithm can be considerably simplified if the explicit dependence on $\mathbf{g}$ is removed by using a power law dependent on an empirically determined shape parameter γ

$$\rho_n(s, \mathbf{g}) = \rho_n(s, a) = \frac{s^\gamma}{a^{\gamma+1}} \exp^{-\frac{s^2}{2a^2}} \quad where \quad a^2(q_t) = n(q_t a_t + (1-q_t)a_r)^2 + a_{tr}^2 w(q_t) \sum_{k=1}^{n} \frac{(M_{kt} - M_{kr})^2}{\sigma_k^2}$$

where a_t and a_{tr} form a matrix of scale parameters for each pure tissue and PV component. The power law is based on a set of logical constraints: the first term in the equation for a establishes a linear baseline between the pure tissue intensities. The gradient distribution must meet these intensities as $q_t \to 0$ and $q_t \to 1$, must peak at $q_t = 0.5$, and must vary smoothly with q_t i.e. in the most general case the distributions must be elliptical. This dependency is introduced by defining $w(q_t) = 1 - 4(q_t - 0.5)^2$.

Having defined probability density distributions for the intensity and local gradient features, Bayes theory can be used to determine the most probable generation mechanism n, either pure tissue or PV, for each voxel $(\mathbf{g}, s)$

$$P(n|\mathbf{g}, s) = \frac{f_n d_n(\mathbf{g}) \rho_n(s, \mathbf{g})}{f_0 + \sum_t f_t d_t(\mathbf{g}) \rho_t(s, \mathbf{g}) + \sum_t \sum_r f_{tr} d_{tr}(\mathbf{g}) \rho_{tr}(s, \mathbf{g})}$$

The f_n are prior probabilities describing the frequency of occurrence of each mechanism; f_0 is an outlier class that can be used to account for pathological or other unmodeled data generation mechanisms. The parameters of the

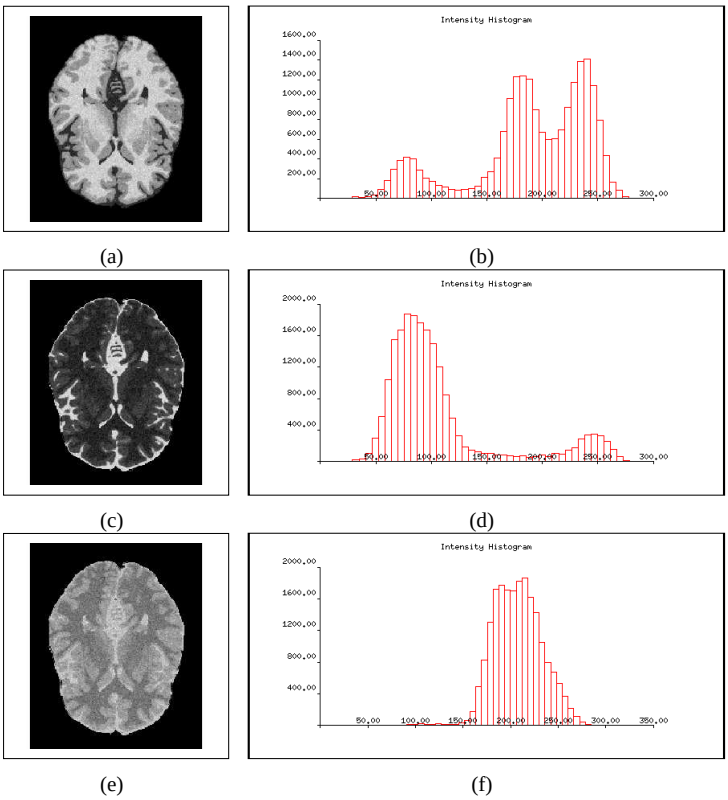


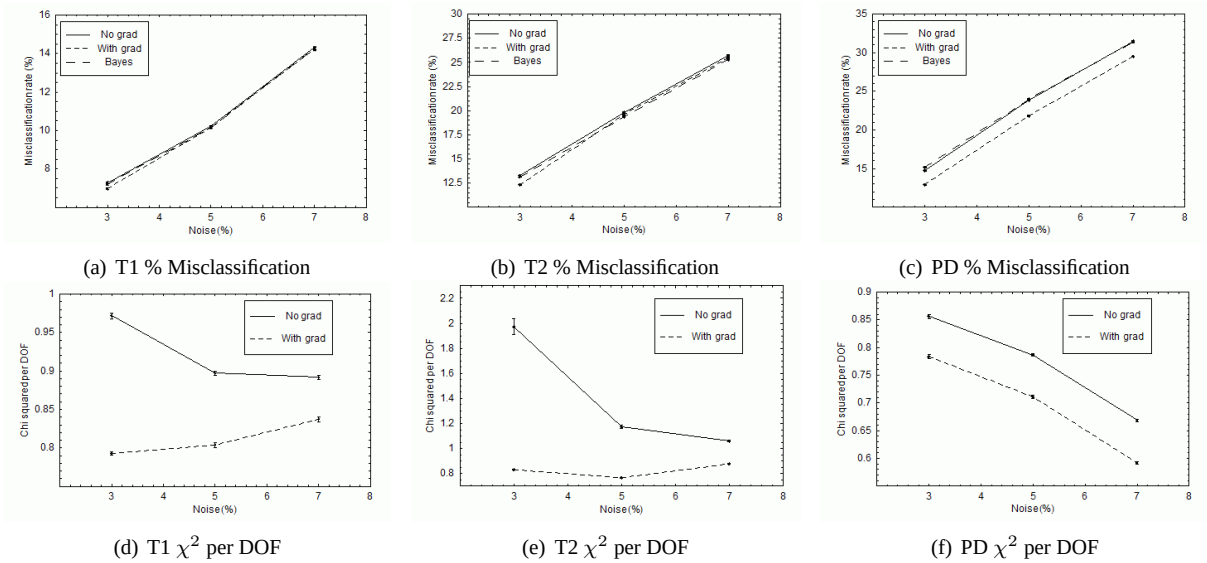
Figure 1. Example images and intensity histograms of the Brainweb T1 (a,b) T2 (c,d) and PD (e,f) images at 5% noise.

distributions ($\mathbf{M_n}$, C_n, f_n and a_n) can then be obtained through fitting the distributions to the image data in feature space using the EM algorithm. The result is a set of probability maps showing the volumetric contribution of each pure tissue to each voxel. In addition, since the algorithm is designed to operate on images in which the formation process is linear (i.e. the intensity of each voxel is given by the sum of the signal intensities of the pure tissues it contains, weighted by their volumetric proportions) and the fitting process provides estimates of the mean intensities of each pure tissue, noise-free reconstructions of the original images can be produced by multiplying the volumetric maps for each tissue with the estimated mean intensity of that tissue and summing the results. This process can be utilised for evaluation purposes, since it produces a spatial representation of the model that has been fitted to the data, which can be compared directly to the original images using standard statistical techniques such as the χ^2 metric.

2.1 Evaluation Methodology

The 1.0mm slice thickness Brainweb phantoms were downloaded, and nine structure-rich slices (142 to 150 inclusive) from an area of the brain expected to show significant partial voluming were chosen for analysis. Only the tissues within the brain (i.e. cerebro-spinal fluid (CSF), grey matter (GM) and white matter (WM)) were considered: the glial matter class was treated as white matter. The phantoms were scaled to lie between 0 and 1, such that they represented the estimated tissue volumes within each voxel. The noise-free, inhomogeneity-free, 1.0mm slice thickness Brainweb simulations of T1-, T2- and PD-weighted images were then downloaded. Pure CSF, WM and GM voxels were identified manually and their intensities measured, scaled such that the highest intensity tissue had a value of 250 grey levels. The scaled phantoms were then multiplied by these values and summed. Finally, the summed images were blurred by convolution with a Gaussian kernel of $\sigma = 0.8$ pixels in order to produce simulated MR images. 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.

A series of Monte-Carlo simulations was then performed, in which Gaussian noise was added to the simulated images at 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.1. A Gaussian noise distribution was used here, in contrast to the Rician noise present in Brainweb simulations. However, this was justified since only WM, GM and CSF were considered, and in these tissues the signal-to-noise



(a) T1 % Misclassification (b) T2 % Misclassification (c) PD % Misclassification

(d) T1 χ^2 per DOF (e) T2 χ^2 per DOF (f) PD χ^2 per DOF

Figure 2. Misclassification rates and χ^2 per DOF for segmentations of the Brainweb T1, T2 and PD images.

ratio never fell below 4.2; the difference between Gaussian and Rician noise is insignificant at SNR's above 3. 50 segmentations were performed for each set of images at each noise level and both with and without the use of gradient information in the feature space.

Two performance metrics were applied to the results. First, PV 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 [14]. However, since the algorithm was designed to measure tissue volumes accurately, and the additional information provided by including gradient in the feature space is most relevant to estimating tissue proportions in PV voxels (since these occur at tissue boundaries i.e. high gradient locations), simple classification-based measures obscure much of the increased accuracy of segmentations performed using gradient terms in the tissue model. Therefore, a second performance metric was also applied. Noise-free estimates of the original images were produced as described above, and a χ^2 metric was calculated by summing the squared differences between these and the original images, divided by the square of the added noise.

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 fit of an intensity-only model (i.e. without using gradient information).

3 Results

Fig. 2 shows 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 proposed 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 PV distributions. The inclusion of gradient information in the tissue model disambiguates the pure tissue and PV distributions in this case, since PV 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 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. Fig. 2 also shows the χ^2 between images reconstructed from the fitted tissue model and PV maps and the original image data. This is a more sensitive measure of algorithmic performance for the reasons outlined above.

Note that the absolute value of the χ^2 measure contains little information (other than the fact that it is close to unity in all cases, as expected), since the EM algorithm optimises a correlate of χ^2. However, relative values for segmentations with and without gradient terms can safely be compared, and demonstrate significant improvements in accuracy when gradients are used: the average improvement is 31.7%.

4 Conclusion

Segmentation algorithms form a core technology in medical image analysis, providing a route to tissue volume estimation that can be used in a wide range of circumstances, such as monitoring tumour volume changes or the progress of atrophic diseases, and measuring the effect of drug therapies on these processes. In the most general terms, it is clear that the more image information we can extract to use in segmentation, the more accurate the results will be. Increasing segmentation accuracy will therefore increase the statistical power available to measure volume change, potentially allowing size of subject groups in clinical trials to be reduced whilst retaining the same confidence limits on the conclusions.

The results presented here show that the inclusion of local gradient information in feature-space based segmentation algorithms can significantly improve the accuracy of segmentations. Since the high-gradient locations in images occur primarily at tissue boundaries, the gradient information is most relevant in PV voxels. Therefore, the improvements in accuracy are most marked in situations where PV distributions are obscured beneath overlapping pure tissue distributions. Gradient information disambiguates the pure tissue and PV data generation mechanisms in such circumstances, significantly increasing the accuracy of the segmentation result. The results also demonstrate the drawbacks of using voxel classification measures in segmentation evaluation: such measures ignore the majority of any error on the estimation of tissue proportions in PV voxels.

Acknowledgements

The authors would like to acknowledge Prof. Alan Jackson, Maja Pokric, Marietta Scott and Dave Williamson for their contributions to the early stages of the work. Part of the work was funded by the MIAS IRC (from Medical Images and Signals to Clinical Information Interdisciplinary Research Collaboration), under EPSRC GR/N14248/01, and UK Medical Research Council Grant No. D2025/31. The software is available at www.tina-vision.net.

References

1. W. J. Niessen, K. L. Vincken, J. Weickert et al. “Multiscale segmentation of three-dimensional MR brain images.” *International Journal of Computer Vision* **31**, pp. 185–202, 1999.
2. H. S. Choi, D. R. Haynor & Y. Kim. “Partial volume tissue classification of multichannel magnetic resonance images-a mixture model.” *IEEE Trans Med Imaging* **10**, pp. 395–407, 1991.
3. P. Santago & H. D. Gage. “Quantification of MR brain images by mixture density and partial volume modelling.” *IEEE Trans Med Imaging* **12**, pp. 566–574, 1993.
4. S. Webb. *The Physics of Medical Imaging*. Medical Science Series. Adam Hilge, Bristol, Philadelphia and New York, 1988.
5. D. C. Williamson, N. A. Thacker, S. R. Williams et al. “Partial volume tissue segmentation using grey-level gradient.” In *Proc. MIUA’02*, pp. 17–20. 2002.
6. P. A. Bromiley, N. A. Thacker, M. L. J. Scott et al. “Bayesian and non-Bayesian probabilistic models for medical image analysis.” *Image and Vision Computing* **21(10)**, pp. 851–864, 2003.
7. M. Pokrić, N. A. Thacker, M. L. J. Scott et al. “Multi-dimensional medical image segmentation with partial voluming.” In *Proc. MIUA’01*, pp. 77–80. 2001.
8. M. Pokrić, N. A. Thacker & A. Jackson. “The importance of partial voluming in multi-dimensional medical image segmentation.” In *Proc. MICCAI’01*, pp. 1293–1294. 2001.
9. N. A. Thacker, M. Pokrić & D. C. Williamson. “Noise filtering and testing illustrated using a multi-dimensional partial volume model of mr data.” In *Proc. BMVC*, pp. 909–919. Kingston, London, 2004.
10. C. A. Cocosco, V. Kollokian, R. K.-S. Kwan et al. “Brainweb: Online interface to a 3D MRI simulated brain database.” *NeuroImage* **5(4)**, pp. S425, 1997.
11. R. K.-S. Kwan, A. C. Evans & G. B. Pike. “MRI simulation-based evaluation of image-processing and classification methods.” *IEEE Trans Med Imaging* **18(11)**, pp. 1085–1097, 1999.
12. R. K.-S. Kwan, A. C. Evans & G. B. Pike. “An extensible MRI simulator for post-processing evaluation.” *Visualization in Biomedical Computing (VBC’96). Lecture Notes in Computer Science* **1131**, pp. 135–140, 1996.
13. D. L. Collins, A. P. Zijdenbos, V. Kollokian et al. “Design and construction of a realistic digital brain phantom.” *IEEE Trans Med Imaging* **17(3)**, pp. 463–468, 1998.
14. J. Tohka, E. Krestyannikov, I. D. Dinov et al. “Genetic algorithms for finite mixture model based voxel classification in neuroimaging.” *IEEE Trans. Med. Imag.* **26**, pp. 696–711, 2007.