Tina Memo No. 2009-002
Published in Annals of the BMVA Vol. 2008 no. 2 pp. 1-22, 2008.
http://www.bmva.org/annals/2008/2008-0002.pdf

Multi-dimensional Medical Image Segmentation with Partial Volume and Gradient Modelling

P.A. Bromiley and N.A. Thacker

Last updated
21 / 01 / 2009

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

Author's note: This memo is a preprint of a paper published in the Annals of the BMVA. The TINA medical image segmentation algorithm that it describes has been the result of a long algorithmic development project, described in TINA Memo nos. 2000-006, 2001-009, 2002-006, 2003-007, 2004-002, 2004-009, 2005-013, 2008-002, 2008-003 and 2008-009 (which provides derivations and proofs for some aspects of the feature space models used in this memo), and in many conference papers over the past decade. This paper summarises the contents of those memos, and represents the final report on this body of work. It is anticipated that future development of the algorithm will focus on the incorporation of spatial information into the feature space; preliminary work in this direction is described in TINA Memo no. 2007-005.

Multi-dimensional Medical Image Segmentation with Partial Volume and Gradient Modelling

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

Abstract

We present a new algorithm for the feature-space based segmentation of medical image volumes, based on a unified mathematical framework that incorporates both intensity and local gradient information. The algorithm addresses the problem of partial volume tissue estimation and is capable of using multiple image volumes, and the associated multi-dimensional image gradient, to increase tissue separability. Clustering is performed in the combined intensity and gradient histogram, followed by the use of Bayes theory to generate probability maps showing the most likely tissue volume fractions within each voxel, rather than a classification to a single tissue type. The approach also supports reconstruction of images from the estimates of volumetric voxel contents and the tissue model parameters.

Evaluation of the algorithm comprised three stages. First, objective measurements of segmentation accuracy, and the increase in accuracy when local gradient information was included in the feature space, were produced using simulated magnetic resonance (MR) images of the normal brain. Second, application to clinical MR data was demonstrated using an exemplar medical problem, the measurement of cerebrospinal fluid (CSF) volume in 70 normal volunteers, through comparison to a "bronze-standard" consisting of previously published measurements. Third, the accuracy of the multi-dimensional approach was demonstrated by assessing the errors on reconstructed images produced from the segmentation result. We conclude that the inclusion of gradient information in the feature space can result in significant improvements in segmentation accuracy compared to the use of intensity information alone.

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 image segmentation offers the opportunity to produce parametric images of tissue type that are more relevant to clinical investigation. The segmentation result can be used for extraction of tissue boundaries or quantitative estimation of volume. Derived 3D models of scanned anatomy can also be applied to pre-operative planning, surgical rehearsal and training [16].

One of the most significant imaging artefacts associated with tomographic biomedical imaging data, in relation to segmentation accuracy, is the partial volume effect. Such images are composed of three-dimensional data points (voxels) enclosing a finite region, which may contain a mixture of signals from several different tissues. Such data points are known as partial volume voxels, and they may make a significant contribution to typical biomedical images. For example, Laidlaw et al. [19], Noe and Gee [25] and our own work [29, 28] have all demonstrated that partial volume processes may affect as many as 40% of the voxels in MR images of the brain with thick-slice acquisitions. Such data cannot be analysed using a set of mutually exclusive, pure-tissue hypotheses: accurate interpretation requires that partial volume distributions are modelled. For example, Niessen et al. [23] 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 (WM), grey matter (GM) and CSF respectively. Fortunately, the physics of the image formation processes in a wide variety of medical imaging modalities allows partial volume distributions for paired tissue combinations to be modelled as a simple, linear process [39]. Tissues therefore contribute proportionately to the intensity in a given voxel.

Two main approaches to the development of probabilistic partial volume models for medical images exist, differing primarily in their assumptions about the spatial distribution of the tissue types. The first was initiated by Choi et

al. [9], and assumes that the tissue mixing proportions vary smoothly across the image. These spatial interactions were introduced using a Markov Random Field (MRF) model. In the original work [9], the authors assumed that the noise was tissue independent and that the mean pure-tissue intensities were known a-priori; they then obtained a maximum a-posteriori (MAP) partial volume segmentation by iteratively optimising the classification of each voxel based on its neighbours. Two heuristic approaches to updating the mean intensities of the pure tissues were also described. Nocera and Gee [24] extended the approach by applying a gradient-descent search algorithm to find the MAP segmentation, and allowing the mean intensities to vary smoothly across the images in order to account for MR inhomogeneities. Whilst the original work focused on multi-channel data, Pham and Prince [27] proposed a similar method for single-channel data, using a different MRF and an update rule for the mean pure tissue intensities based on a heuristic prior.

The second approach, initiated by Santago and Gage [31], dispenses with the assumption of spatial smoothness, and instead assumes a uniform prior probability distribution for tissue mixing proportions. This is based on three assumptions; that, in the absence of artefacts and noise, each pure tissue has a well-defined signal intensity, that in partial volume voxels the constituent tissues contribute proportionately to the intensity of the voxel (i.e. that the image formation process is linear), and that there is no correlation between voxel boundaries and tissue boundaries. The distributions in the intensity space generated by partial volume voxels therefore take the form of uniform distributions convolved with a distribution representing the noise process. In the original paper [31], only pairs or triplets of tissues with Gaussian noise characteristics were studied: a later extension to multiple pure tissues with Gaussian or Poisson noise was also published [32]. However, the authors focused only on estimating the total volume of each tissue within whole images. The same basic approach has been extended by other authors to classify, or to estimate the volumes of each tissue within, individual voxels. Laidlaw et al. [19] used the same model to fit the histogram of the entire image, and fitted the resulting model to individual voxels, optimising only the tissue proportions and local noise. The result was an estimate of the volume of each pure tissue within each voxel. Ruan et al. [30] replaced the Santago and Gage model with independent Gaussian distributions for each mixture class, on the basis that the convolution of a Gaussian with a uniform distribution linking the pure tissue distributions in the intensity space can be well approximated by a Gaussian when the pure tissue signal intensities are close together, and used an MRF prior to assign each voxel to a single tissue, thereby loosing all notion of partial voluming. This work was also specific to T1-weighted MR images. Shattuck et al. [33] used a sequence of low-level operations to fit the Santago and Gage model to T1-weighted MR brain images.

Other authors have attempted to extend the Santago and Gage approach to incorporate non-uniform prior distributions for the tissue mixing proportions, primarily focusing on distributions that peak at either extreme i.e. close to the pure tissue distributions. Van Leemput et al. [20] considered the partial voluming process as a down-sampling of the images, producing a non-uniform prior distribution with no functional constraint. In addition, they incorporated local structure information into the prior probability estimation using the MRF approach. This was compared to the uniform prior model, and showed some improvements in segmentation performance for widely separated pure tissue distributions, but the converse for overlapping distributions. This is possibly due to the difficulty in estimating the underlying partial volume prior distribution at points in the intensity histogram where it is obscured by the pure tissue distributions. A similar approach was adopted by Joshi and Brady [17]. However, these papers assume that the sensitivity to the underlying magnetisation is uniform across voxels in MR images, which may be an over-simplification in practice, particularly when slice gaps are present. Chiverton [8] investigated this topic in some detail, although again ignoring the possibility of anisotropies in sensitivity.

In this paper we extend the approach proposed by Santago and Gage [31], incorporating our earlier work on the subject as described by Pokric et al. [29] and Williamson et al. [41], to generate a unified framework incorporating both intensity and intensity gradient within the density model[1]. In low-contrast situations this approach effectively doubles the information available for partial volume estimation. In addition, it can be extended to utilise multispectral data simply by adding the intensity and gradient of each image or volume as separate dimensions of the feature space. The model parameters are estimated using a partial volume modification of the Expectation-Maximisation (EM) algorithm [11]. Finally, the most likely tissue volume fractions within each voxel, rather than a classification to a single tissue type, are estimated using Bayes theorem. The extension to include local image gradients as well as intensities can be performed using only the same assumptions underlying the standard Bayesian approach.

[1] Throughout this paper, the term gradient is used to refer to the image gradient i.e. the local first derivative of the image intensities, rather than any of the other definitions in common use in MR imaging.

2 Method

2.1 Multi-dimensional Intensity and Gradient Analysis

Consider a data set consisting of multiple, pre-registered images of the same region, where $\mathbf{g}$ is the intensity vector of a given voxel in the set of images and s is the local multi-dimensional gradient, the square-root of the sum-squared derivatives. The conditional probability that a voxel may be assigned to a specific tissue class t given its intensity and gradient is given by Bayes theorem as

$$P(t|\mathbf{g}, s) = \frac{P(t)P(\mathbf{g}, s|t)}{\sum_t P(t)P(\mathbf{g}, s|t)}$$

In this expression $P(t)$ are the prior probabilities[2] of the tissue class t, $P(\mathbf{g}, s|t)$ are the likelihoods of the intensity $\mathbf{g}$ and the derivative s given a tissue class t, so $P(\mathbf{g}, s)$ is the global likelihood of the instance of data $\mathbf{g}$,s. The gradient must be dependent upon the intensity as the partial volume process occurs at tissue boundaries, and so

$$P(\mathbf{g}, s|t) = P(\mathbf{g}|t)P(s|\mathbf{g}, t)$$

where $P(\mathbf{g}|t)$ is the expected distribution of intensities for each class and $P(s|\mathbf{g}, t)$ is the expected distribution of gradient values as a function of intensity for each class.

Partial volume voxels could be accounted for by including a specific class for each pair of tissues that share a common boundary, leading to a formulation well-suited for voxel classification. However, it is more useful in practice to estimate the volume of each tissue present in the images, which requires the estimation of the volumetric contribution of each tissue to each voxel. In order to facilitate this process, partial volume classes can be divided into pairs of distributions $P(\mathbf{g}, s|tr)$ and $P(\mathbf{g}, s|rt)$, each of which represents the volumetric contribution of one of the pure tissues to the partial volume voxels, as indicated by the order of the subscripts e.g. $P(\mathbf{g}, s|tr)$ is the volumetric contribution of tissue class t to partial volume voxels that also contain tissue class r. Adopting the assumption of a uniform distribution for partial volume voxels described by Santago and Gage [31], each partial volume class can be modelled using a pair of probability density distributions $d_{tr}(\mathbf{g})$ and $d_{rt}(\mathbf{g})$. Each of these takes the form of a triangular density distribution along the vector between the intensities of the pure tissues in the feature space, convolved with a multi-dimensional noise distribution. These distributions are described in more detail in the following section. Each pair of density distributions $d_{tr}(\mathbf{g})$ and $d_{rt}(\mathbf{g})$ sum to produce the overall uniform distribution for the partial volume class, again convolved with some noise distribution.

This formulation is equivalent to that suggested by other authors [31, 19]. However, in order to include local gradient information, terms describing the gradient density $\rho_{tr}(s, \mathbf{g})$ are required, which must also be functions of the intensity $\mathbf{g}$. Using the formulation

$$P(\mathbf{g}, s|n) \propto d_n(\mathbf{g})\rho_n(s, \mathbf{g})$$

the conditional probability of an intensity being due to a certain mechanism n (either a pure or mixture tissue component) can be calculated using Bayes theory, as

$$P(n|\mathbf{g}, s) = \frac{f_n d_n(\mathbf{g})\rho_n(s, \mathbf{g})}{f_0 + \sum_t f_t d_t(\mathbf{g})\rho_t(s, \mathbf{g}) + \sum_t \sum_r f_{tr} d_{tr}(\mathbf{g})\rho_{tr}(s, \mathbf{g})} \quad (1)$$

where f_n, f_0, f_t and f_{tr} are effectively priors expressing the frequencies of occurrence (in terms of numbers of voxels) of particular tissue types. Unknown tissues are accounted for in the Bayesian formulation by including a fixed extra term f_0 for infrequently occurring outlier data [12] in the summed likelihood term, which enables separation of pathological tissues.

It is possible to obtain the most probable volumetric measurement U_t for each tissue t given the observed data $\mathbf{g_v}$ in voxel v,

$$U_t(\mathbf{g_v}, s_v) = P(t|\mathbf{g_v}, s_v) + \sum_r P(tr|\mathbf{g_v}, s_v) \quad (2)$$

It is this step that necessitates the division of the partial volume processes into two triangular distributions, rather than working with class assignment probabilities. The approach is easily simplified to remove the gradient information by setting $\rho = 1$ in Eq. 1.

[2]We adopt the convention throughout that upper-case P represents a probability, whereas lower-case symbols represent probability density distributions: d is used for intensity distributions and ρ for gradient distributions.

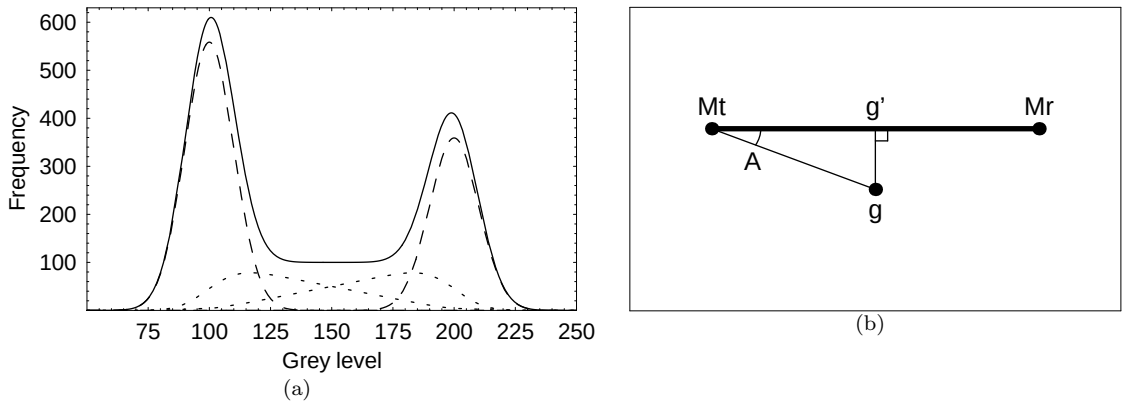


Figure 1: (a) An example partial volume model for two pure tissues. Pure tissues have Gaussian distributions (dashed line), whereas mixtures of tissues take form of triangular distributions convolved with a Gaussian (dotted lines). The model components are summed to give the overall distribution (solid line). (b) In multi-dimensional intensity spaces, noise can move the intensity off of the vector between the pure tissue intensities, and so it must be projected onto this vector in order to determine the proportions of each pure tissue.

2.2 Intensity Model

The formulation described above requires the assumption of noise distributions describing the inherent tissue variability and measurement noise. Assuming Gaussian distributions, each pure tissue t is described by

$$d_t(\mathbf{g}) = \alpha_t e^{-\frac{1}{2}(\mathbf{g}-\mathbf{M_t})^T C_t^{-1}(\mathbf{g}-\mathbf{M_t})}$$

where $\mathbf{M_t}$ is a vector describing the tissue mean intensity, C_t is the covariance matrix, and α_t is a constant that provides unit normalisation.

Consideration of partial volume voxels is limited to those containing mixtures of two pure tissues. The intensity distribution for each mixture class is modelled as a uniform distribution convolved with a Gaussian representing the noise, decomposed into two triangular distributions $d_{tr}(\mathbf{g})$ and $d_{rt}(\mathbf{g})$. An example for two well-separated tissues in a 1D intensity space is shown in Fig. 1(a). In the case of a multi-dimensional intensity space, the model will take the form of a complementary pair of triangular distributions lying along the vector between the pure tissue intensities, convolved with a multi-dimensional Gaussian noise distribution. The covariance matrix of this noise distribution is a function of position along the partial volume vector[3]

$$C_h = hC_t + (1-h)C_r$$

where C_t and C_r are the covariance matrices of voxels of pure tissues t and r and $0 \leq h \leq 1$ is the fractional distance along the partial volume vector. In order to obtain h the intensity of the partial volume voxel must be projected onto this vector. The procedure is illustrated in Fig. 1(b); h is given by the ratio $|(\mathbf{g_h} - \mathbf{M_t})|/|(\mathbf{M_r} - \mathbf{M_t})|$. For a normal projection this is given by $[(\mathbf{g} - \mathbf{M_t}).(\mathbf{M_r} - \mathbf{M_t})]/[(\mathbf{M_r} - \mathbf{M_t}).(\mathbf{M_r} - \mathbf{M_t})]$. However, in the general case the variances of the noise distribution will vary between the axes of the space. Therefore, in order to find the most probable projection, all measurements in the space must be weighted by the local covariance matrix, giving [2]

$$h = \frac{(\mathbf{g} - \mathbf{M_t})^T C_h^{-1}(\mathbf{M_r} - \mathbf{M_t})}{(\mathbf{M_r} - \mathbf{M_t})^T C_h^{-1}(\mathbf{M_r} - \mathbf{M_t})} \tag{3}$$

The estimates of h and C_h are therefore inter-dependent and must be obtained using an iterative process, initiated by assuming $h = 0.5$. Since C_h varies monotonically with h this process is stable and converges rapidly (typically within a few iterations).

Weighting distance measurements in the intensity space by the local covariance matrix transforms into a homoscedastic space in which the covariance matrix of the noise distribution is the identity matrix. This has the

[3]We consider the tissue-dependent noise case, where noise is added to the signal prior to partial volume averaging. The signal from a partial volume voxel is therefore given by a linear combination of Gaussian distributions, which is itself a Gaussian with

$$C_h = h^2 C_{th} + (1-h)^2 C_{rh}$$

where C_{th} and C_{rh} are the covariances of each pure tissue contribution. However, in the notation used here C_t and C_r are the covariance matrices of one-voxel-large samples of each pure tissue, not of the populations. Therefore, the reduction in the sample size of each pure tissue in a partial volume voxel must be taken into account i.e. $C_{th} = C_h/h$ and $C_{rh} = C_h/(1-h)$ [2].

useful consequence that the noise distribution can be decomposed into components parallel and perpendicular to the partial volume vector

$$d_{tr}(\mathbf{g}) = \beta_{tr} T_{tr}(h) e^{-\frac{z^2}{2}} \tag{4}$$

where β_{tr} provides unit normalisation and z is the magnitude of the vector from $\mathbf{g}$ to $\mathbf{g_h}$ in the variance-normalised space, given by

$$z^2 = (\mathbf{g} - \mathbf{M_t})^T C_h^{-1} (\mathbf{g} - \mathbf{M_t}) - h^2 [(\mathbf{M_r} - \mathbf{M_t})^T C_h^{-1} (\mathbf{M_r} - \mathbf{M_t})]$$

T_{tr} is the convolution of a triangular distribution normalised to 1/2 (since two complementary triangular distributions comprise each partial volume distribution) with a Gaussian distribution normalised to 1. The general form for a Gaussian with standard deviation σ and a triangular distribution with start and end points a and b, intercept c and gradient k is [2]

$$T_{tr}(x) = -\frac{kx + c}{2} \left[erf(\frac{x-b}{\sigma\sqrt{2}}) - erf(\frac{x-a}{\sigma\sqrt{2}}) \right] - \frac{k\sigma}{\sqrt{2\pi}} \left[e^{-\frac{(x-b)^2}{2\sigma^2}} - e^{-\frac{(x-a)^2}{2\sigma^2}} \right]$$

In this case $a = 0$, b is the length of the partial volume vector in the variance-normalised space

$$b = \sqrt{(\mathbf{M_r} - \mathbf{M_t})^T C_h^{-1} (\mathbf{M_r} - \mathbf{M_t})}$$

and $\sigma = 1$.

2.3 Multi-dimensional Gradient Model

In previous work [41] we demonstrated the inclusion of gradient information into the feature space for segmentations of single images. The mean and width of the gradient distribution were estimated from tissue contrast and image noise respectively. In this work, the use of local image gradients was extended to multi-dimensional data. However, rather than attempting to model the gradient distribution separately in each image, which would require the optimisation of a large number of model parameters and so be computationally expensive, a single multi-dimensional gradient distribution was used.

The multi-dimensional image gradient is estimated in the standard way, using

$$s = \sqrt{\sum_k^{N_I} \frac{\nabla_x I_k^2 + \nabla_y I_k^2}{\sigma_k^2}} - \lambda \tag{5}$$

where $\nabla_x I_k$ and $\nabla_y I_k$ are the local intensity gradients of image k in the x and y directions respectively, σ_k is standard deviation of noise in image k, N_I is the total number of images, and λ is a fit parameter. Each term in the sum is normalised to the image noise in order to take account of the intrinsic information content of the data. The absence of a z derivative term is deliberate. Standard acquisition practices commonly result in voxel dimensions that are larger in the through-plane direction (z) than within the imaging plane (x, y), and an inter-slice gap. These factors complicate any definition of a z derivative and confound any attempt at a theoretical description of the three-dimensional gradient distribution. In addition, avoiding the use of a z gradient allows an analysis of individual slices.

If the z component was included in the gradient calculation and the voxels were isotropic, then the multi-dimensional gradient distribution would be similar to a χ^2 for three degrees of freedom and non-zero mean [8]. The mean of the gradient distribution would increase linearly with the number of images considered, but the width of the distribution would increase only as the square-root. However, the presence of a finite z component that is ignored in the gradient calculation simplifies the model, as it ensures that the width of the distribution is largely determined by the mean i.e. the tissue contrast, rather than the noise. Therefore, for a randomly oriented boundary with fixed pure tissues on either side, the gradient distribution can be modelled using a single, scaled function, where the scale parameter is determined by the tissue contrast. The λ parameter ensures that the same distribution can be used to account for both pure tissue voxels, with zero mean gradient, and partial volume voxels with a finite mean gradient. In the limit of infinitely thin slices, the z component would be zero; the required distribution would then be composed of the root-mean-square of two squared quantities, each assumed to have independent Gaussian random noise, and so would be a Rician distribution [14]. In practice this will not be the case, but it was found empirically [37] that multiplication with a simple power law dependent on a parameter γ was sufficient to account for this effect, giving

$$\rho_n(s, \mathbf{g}) = \rho_n(s, a) = \frac{s^\gamma}{a^{\gamma+1}} \exp^{-\frac{s^2}{2a^2}} \tag{6}$$

where a is the scale parameter.

The scale parameter a can be derived purely on the basis of empirical observations and logical constraints. It was observed that the distributions in gradient vs. intensity projections of the feature space of MR images form approximately elliptical features between pairs of pure tissues, with the highest gradients occurring in voxels consisting of equal contributions from pairs of tissues i.e. at edges. This indicates that the scale parameter is linearly dependent on the fractional volumetric contribution from each tissue present in the voxel, as would be expected given the linear image formation process. Furthermore, the ends of the partial volume distributions must match the pure tissue distributions exactly. Therefore, the most general model for the scale parameter a consists of two terms

$$a^2(q) = N_I(qa_t + (1-q)a_r)^2 + a_{tr}^2 w(q) \sum_k^{N_I} \frac{(M_{kt} - M_{kr})^2}{\sigma_k^2} \tag{7}$$

where a_t and a_{tr} form a matrix of scale parameters for the pure tissues and partial volume combinations respectively. As before, M_{kt} is the mean intensity of tissue class t in image k where $k = 1...N_I$. The first term establishes a linear baseline between the two pure tissue distributions; q is the fractional volumetric contribution to the voxel from tissue class t determined from the multi-dimensional intensity model ($0 \leq d_{tr}(\mathbf{g})/(d_{tr}(\mathbf{g}) + d_{rt}(\mathbf{g})) \leq 1$). The second term introduces the shape of the distribution, scaled according to the difference in intensity across tissue boundaries. Defining $w(q) = 1 - 4(q - 0.5)^2$ (the equation for a half-ellipse scaled to lie between 0 and 1) enforces the required elliptical shape. Note that this selection of ρ gives identical values for both partial volumes tr and rt at a given $(\mathbf{g}, s)$.

2.4 Parameter Estimation

Parameters of the model can be iteratively estimated using a modified version of the Expectation Maximisation (EM) algorithm [11]. The presence of partial volume classes complicates the application of standard EM, and so these voxels are ignored when updating the means and covariances. The result is an approximation that does not constitute a proper EM algorithm in the exact sense of [11]. However, this strategy has been successfully applied in the past [38].

The expectation step involves computing the Bayes probability $P(n|\mathbf{g}, s)$ for every voxel based upon the current estimate of the parameters using Eq. 1. The maximisation step is then applied for the estimation of normalisation parameters for pure and mixture tissues (f_t, f_{tr}), mean ($\mathbf{M_t}$), and covariance matrices (C_t), which take the form

$$f'_t = \sum_v^V P(t|\mathbf{g}_v, s_v) \qquad f'_{tr} = f'_{rt} = \frac{1}{2}\sum_v^V P(tr|\mathbf{g}_v, s_v) + P(rt|\mathbf{g}_v, s_v)$$

$$\mathbf{M}'_t = \frac{1}{V}\sum_v^V P(t|\mathbf{g}_v, s_v)\mathbf{g}_v \qquad C'_t = \frac{1}{V}\sum_v^V P(t|\mathbf{g}_v, s_v)(\mathbf{g}_v - \mathbf{M_t}) \otimes (\mathbf{g}_v - \mathbf{M_t})^T$$

The estimated values for f_n correspond to the total volumetric estimate of the tissue process for the voxel. The use of the f_n as Bayesian priors in Eq. 1 is therefore sample dependent, although it remains consistent within any given sample. It may therefore be necessary to use one consistent value in order to maximise the repeatability of the measurement when attempting to quantify changes between two data sets [5]. Terms f_{tr} and f_{rt} are equal and sum to the total number of partial volume voxels generated by the two tissues t and r, as they must for any partition of voxels which is uncorrelated with imaging geometry. Prior knowledge of image content is incorporated into this process by constraining particular sets of prior frequencies f_n, due to impossible tissue combinations, to be zero.

The gradient parameters can be computed by rescaling the current estimate using the sample mean. The initial value of the a_n factor is set as a ratio of the mean value of the gradient data μ_{data} to the theoretical mean value μ_{model} calculated from the gradient model. The value of a_n for the pure t or mixture tissue component tr is then re-calculated during the Maximisation step using

$$\mu^t_{data} = \frac{\sum_v^V s_v P(t|\mathbf{g}_v, s_s)}{\sum_v^V P(t|\mathbf{g}_v, s_v)} \qquad \mu^{tr}_{data} = \frac{\sum_v^V s_v P(tr|\mathbf{g}_v, s_s) + s_v P(rt|\mathbf{g}_v, s_s)}{\sum_v^V P(tr|\mathbf{g}_v, s_v) P(rt|\mathbf{g}_v, s_v)}$$

$$\mu^t_{model} = \frac{\sum_v^V \hat{s}_v P(t|\mathbf{g}_v, s_v)}{\sum_v^V P(t|\mathbf{g}_v, s_v)} \qquad \mu^{tr}_{model} = \frac{\sum_v^V \hat{s}_v P(tr|\mathbf{g}_v, s_v) + \hat{s}_v P(rt|\mathbf{g}_v, s_v)}{\sum_v^V P(tr|\mathbf{g}_v, s_v) + P(rt|\mathbf{g}_v, s_v)}$$

where $\hat{s}_v$ is the first moment of the current estimate of the $\rho(s, a)$ distribution and takes the value $a(q)\sqrt{8/\pi}$ for $\gamma = 2$, but will more generally be of the form $a(q)\kappa$ (where κ is a constant) for any fixed distribution shape. As

the required parameter (a) is related to the first moment by a simple scale factor, the likelihood estimate of a can be obtained by scaling the likelihood estimates of the moments

$$a'_n = \frac{\mu^n_{data}}{\mu^n_{model}} a_n \tag{8}$$

These update equations generate values of partial volume gradient distribution parameters that satisfy $a_{tr} = a_{rt}$. This has the useful consequence that

$$q = d_{tr}(\mathbf{g})/((d_{tr}(\mathbf{g}) + d_{rt}(\mathbf{g})) = P(\mathbf{g}, s|tr)/(P(\mathbf{g}, s|tr) + P(\mathbf{g}, s|rt))$$

so that the necessary $a(q)$ term for $\hat{s}_v$ can be regenerated using Eq. 7 from the results of the expectation step.

2.5 Image Reconstruction

The proposed algorithm produces estimated tissue model parameters, including mean intensities for each pure tissue, and estimates of the volumetric contribution of each pure tissue to each voxel. Therefore, since a linear image formation process was assumed, the volumetric maps for each pure tissue can be multiplied by the fitted mean intensity for that tissue and summed to produce a noise-free estimate of the original image. The resulting image is a comprehensive representation of the model fitted to the data during the segmentation process. It can therefore be compared to the image data using standard statistical techniques such as the χ^2 metric in order to measure the goodness-of-fit of the model. It also represents a potential application of the algorithm as a noise-filtering technique.

This form of evaluation has two main advantages. First, it does not require gold-standard data, either in the form of manual segmentations (which can be prohibitively time-consuming to produce) or in the form of simulated images (which can be criticised as simplistic compared to clinical images, since they may not feature the whole range of potential imaging artefacts): the comparison can be performed directly between the segmentation result and the original data. Second, large amounts of data (typically ~ 100000 voxels in medical image volumes) are available to optimise the tissue model parameters, and thus high accuracy can be expected in the case where the model is a good fit to the data. The dominant source of error in the segmentation result will therefore be due to points at which the model fails to fit the data, rather than the errors on the model parameter estimation; segmentation accuracy can therefore be evaluated through testing the goodness-of-fit of the tissue model.

3 Evaluation Methodology

3.1 Evaluation using Simulated MR Data

Evaluation of the proposed algorithm was divided into three stages. In the first, simulated images from the Brainweb database [18] were used. Unispectral segmentations were performed on simulated T1, T2 and PD-weighted images, both with and without the inclusion of gradient information, and compared to the ground truth images used to generate the simulated data. The aim of this stage of the evaluation was to provide quantitative measures of the accuracy of the segmentation and to measure the improvements gained through the inclusion of gradient information.

In order to ensure that the pure tissue mean intensities were known exactly, as required for one of the performance metrics described below, the Brainweb images were resimulated. The noise-free, inhomogeneity-free, 1.0mm slice thickness T1-, T2, and PD-weighted Brainweb simulations were downloaded and the approximate intensities of pure CSF, GM and WM were identified from manually selected voxels. Next, the 1.0mm slice thickness fuzzy Brainweb phantoms were downloaded, and nine structure-rich slices (71 to 79 inclusive) from a region of the brain including parts of the basal ganglia, and so expected to show significant partial voluming, were chosen for analysis. Only the tissues within the brain (i.e. CSF, GM and WM) were considered, and the glial matter class was treated as WM. In accordance with the assumption of a linear image formation process, the phantom for each brain tissue was multiplied by the intensity of that tissue in the Brainweb simulated images, and the phantoms were then summed. Finally, the summed images were blurred by convolution with a Gaussian kernel of $\sigma = 0.8$ voxels 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 were then performed in which Gaussian noise was added to the re-simulated images with a standard deviation of 3%, 5% or 7% of the highest intensity in the images. Example images and

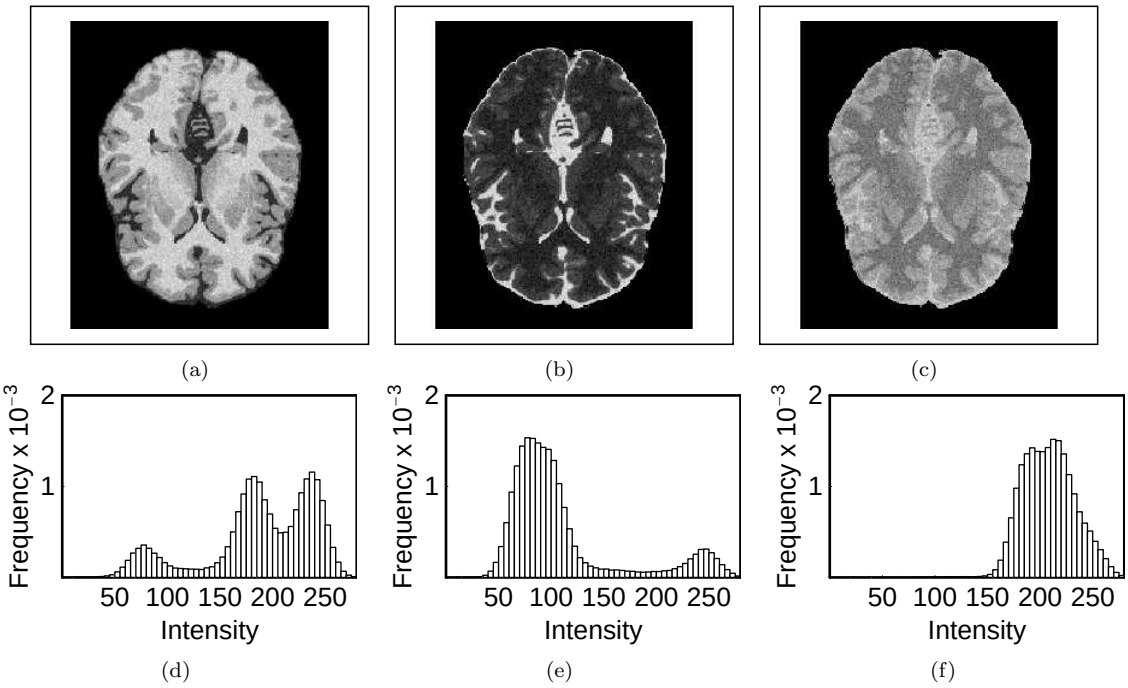


Figure 2: Example images and intensity histograms of the re-simulated Brainweb T1 (a,d) T2 (b,e) and PD (c,f) images. The noise level is 5%.

intensity histograms at 5% noise are shown in Fig. 2. A Gaussian noise distribution was used, in contrast to the Rician noise present in Brainweb simulations. This was justified as the difference between the two is insignificant at signal-to-noise ratios (SNR) greater than 3, and the SNR never fell below 4.2 in the tissues considered. Fifty segmentations were performed for each set of images, at each noise level, both with and without the use of gradient information.

Two performance measures 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 tissue phantoms, and the number of misclassified voxels 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 [38]. However, enforcing a hard classification to a single tissue type removes any notion of tissue volume estimation in partial volume voxels. Since the proposed algorithm was designed to measure these volumes accurately, and the additional information utilised by including gradient terms in the tissue model has a significant effect on partial volume voxels, simple classification-based measures do not reveal the increased accuracy of segmentations performed using gradient terms. Therefore, a second performance measure that takes full account of the accuracy of tissue volume estimation in each voxel was also applied. Noise-free estimates of the original images were produced as described above, and a χ^2 measure was calculated by summing the squared differences between these and the re-simulated images, divided by the square of the added noise. In order to provide a benchmark for the misclassification rate, the Bayes error for segmentations performed using only intensity information was also measured by counting the number of voxels in each 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.

3.2 Evaluation using Clinical MR Data

Evaluation on simulated data has the advantage of the availability of ground truth images to which the results can be compared in order to directly measure the number of misclassified voxels. However, the issue of clinical utility remains, since simulated image volumes may not exhibit all of the artefacts present in clinical MR data. Therefore, a second stage of evaluation was performed on clinical MR data. One of the primary uses of segmentation algorithms in a clinical setting is volume measurement. A volume measurement task of clinical relevance, measurement of the CSF volume, was identified. 70 normal volunteers (32 male and 38 female) ranging in age from 19 to 85 years with a mean age of 57±20 years were recruited. None of the subjects had vascular risk factors, neurologic disease,

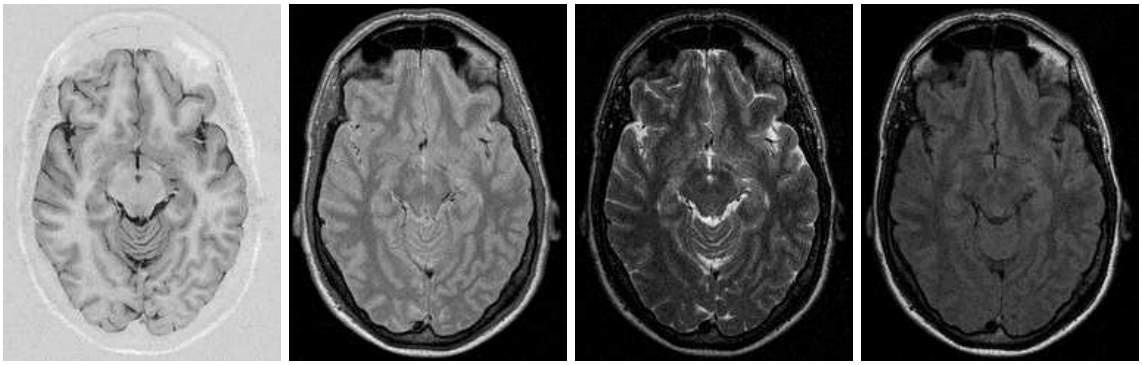
Reference	No. subjects	Definition of measurement space	Segmentation Method
Gur et al. [15]	69 T	Excludes cerebellum	T2/PD histogram fitting
Blatter et al. [1]	89 M 105 F	TIV	ANALYZE
Mueller et al. [22]	46 T	Excludes brainstem	REGION
Coffey et al. [10]	122 M 198 F	Excludes slices below midbrain	MedVision
Chan et al. [6]	10 T	TIV	MIDAS
Whitwell et al. [40]	55 T	Excludes slices below cerebellum	MIDAS
Good et al. [13]	265 M 200 F	TIV	SPM
Chard et al. [7]	13 M 14 F	Excludes slices containing cerebellum	SPM

Table 1: Details of the experimental method adopted by studies used in the CSF volume comparison, showing the number of subjects (M=male; F=female; T=total, where the number of each sex was not given), the definition of the measurement space (where TIV is indicated, the whole CSF pool inside the cranium was used), and the segmentation method.

or cognitive or psychiatric problems. The local ethics committee approved the research, and informed consent was given by the subjects. All subjects underwent MR imaging with a 1.5-T system (ACS-NT, with PowerTrack 6000 gradient subsystem; Philips Medical Systems, Hamburg, Germany) with a birdcage head coil receiver. Fast spin-echo inversion-recovery (IRTSE) images (repetition time, 6850 msec; echo time, 18 msec; inversion time, 300msec; echo train length, 9) were obtained in contiguous 3-mm thick sections throughout the brain, with an in-plane resolution of $0.89mm^2$ (matrix, 256×204, field of view, $230 \times 184mm$), and real image reconstruction was performed.

The proposed algorithm was then used to segment the images and produce volumetric maps of the CSF, using both intensity and gradient information. One image volume from the centre of the age range was then chosen, and a set of hand-drawn binary masks were prepared from this by a trained neuroanatomist. The masks eliminated non-CSF fluid spaces such as the eyes and sinuses. In addition, they defined a consistent inferior boundary to the CSF space by drawing a line in the midsaggital section parallel to the horizontal axis that passed through the junction of the calvarium and the tentorium cerebelli (the other boundaries of the CSF space were defined by the skull boundary). All image volumes were then registered to the chosen volume using a technique based on maximising the dot-product correlation between local edge structure (the square root of a summed squared gradient image), allowing the masks to be rotated into the coordinate system of each image volume. Finally, the CSF volume in each masked image volume was calculated. All measurements were then normalised to the total intra-cranial volume (TIV) (measured using the same technique) which has been shown [40] to be an effective normalisation for both inter-individual variations in head size and variation in voxel sizes in longitudinal studies.

Attempting to define a gold-standard against which comparison of the CSF volume measurements could be performed, for example by manual segmentation of the image volumes, would be prohibitively time-consuming. In addition, manual segmentations would be limited to voxel classification, rather than estimation of the tissue volume proportions within each voxel. The proposed algorithm was designed to estimate these proportions and so, as described in the previous section, evaluations based on voxel classification have limited utility in determining the accuracy of the algorithm. Therefore, an alternative methodology was used, in which a "bronze standard" consisting of CSF volume measurements from the literature was constructed. Eight papers published between 1991 and 2002, which quoted TIV and CSF volume measurements from MR images in normal subjects, were collected. These previous studies used a variety of MR pulse sequences, definitions of the measurement space and segmentation routines, including a variety of widely available automatic or semi-automatic software packages (REGION, ANALYZE, MIDAS, MedVision and SPM), summarised in Table 1. Most of the studies used 1.5T GE Signa MR scanners, with the exceptions of Coffey et al. [10], where some images were acquired on a 0.35T Toshiba scanner, and Good et al. [13], in which a 2T Siemens MAGNETOM Vision scanner was used. It should be noted that comparative evaluation based on such a bronze standard has a significant drawback in that each set of measurements in the comparison is derived from a different subject group. The presence of inter-group variations that are significant with respect to the random errors on the algorithms being compared would therefore reduce the statistical power of the technique to detect systematic errors in the algorithms. Furthermore, it cannot distinguish between inter-group variation in the quantity being measured and systematic error on the algorithms used. Therefore, a sufficient number of the algorithms must be free of systematic error to establish a clear consensus measurement. However, if this condition is met the approach has significant advantages; it can be performed rapidly, and allows the comparison of results from a wide variety of segmentation algorithms without the need to reimplement them all at a single site, thus avoiding concerns related to errors in reimplementation or sub-optimal use of unfamiliar, off-site software (e.g. issues related to the setting of free parameters in the algorithms).



(a) (b) (c) (d)

Figure 3: Examples of the MR data used: IRTSE (a), VE(PD) (b), VE (T2) (c) and FLAIR (d).

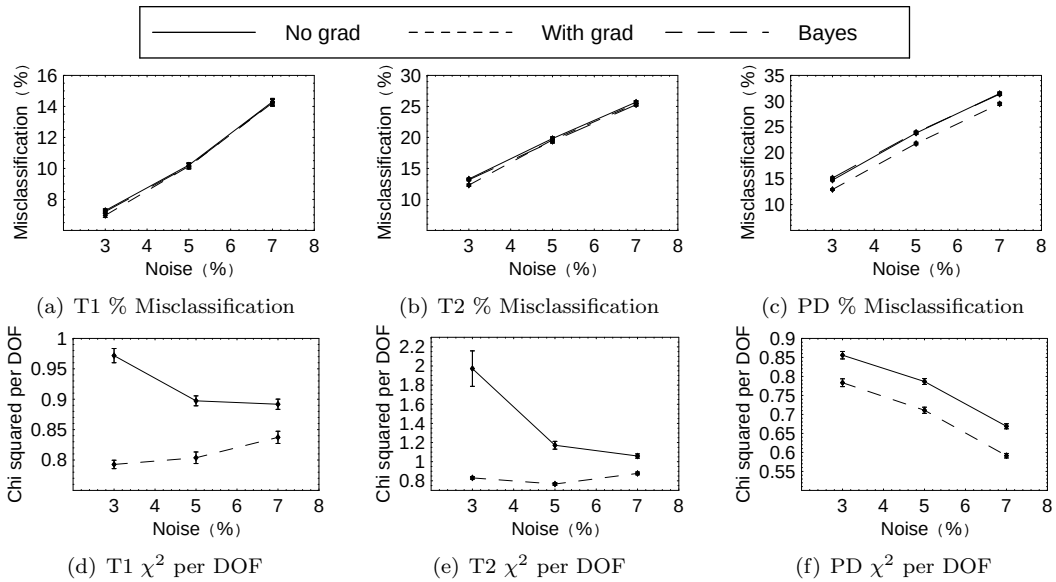
3.3 Noise Filtering using Multispectral Segmentation

The third stage of evaluation focused on the use of multispectral data. Large cohorts of multispectral MR data are not routinely available, limiting the range of evaluation methodologies that could be applied. However, since MR images typically contain of the order of 10000 voxels, the stability of the model fitting process is unlikely to be the major source of error. Instead, the accuracy of the segmentation result is likely to be limited by the goodness-of-fit of the model (e.g. unmodeled tissue classes). Therefore, an evaluation approach that provided a quantitative measure of the goodness-of-fit of the model was adopted. Four MR image volumes were obtained from one of the normal volunteers used in the previous evaluation stage, using a variety of pulse sequences (inversion recovery turbo spin echo (IRTSE), variable echo proton density (VE(PD)), variable echo T2 (VE(T2)), and fluid attenuated inversion recovery (FLAIR)). These pulse sequences were chosen for their good tissue separation and availability in a clinical environment; example images are shown in Fig. 3. A rigid registration technique (described in the previous section) was applied in order to align the images: since all images were acquired during single scanning sessions, any misalignment can be assumed to be due to patient motion, and thus rigid, to a good approximation[4]. Renormalised sinc interpolation with a 5×5 kernel was used to reslice the images [34] and they were segmented using the proposed technique.

Noise-free reconstruction, as described above, was then performed and the reconstructed images compared to the original image data in order to measure the goodness-of-fit of the model. Two performance measures were applied. First, a Monte-Carlo stability analysis [36], which measured the relative change in the output image intensities produced by the addition of a small amount of noise to the input images, was applied to estimate the fraction of noise remaining after filtering. Second, the number of reconstructed voxels whose intensity was modified by more than three standard deviations of the image noise was counted, after compensating for local field inhomogeneity using the algorithm described by Thacker et al. [35]. This measure quantified the number of voxels for which the tissue model was inappropriate i.e. the goodness-of-fit, and is referred to as the residual outlier measure (ROM). The local image noise was estimated independently of the Monte-Carlo stability analysis using a technique based upon the distribution of local derivatives [26]. In order to provide a benchmark for the evaluation, it was also applied to four conventional noise-filtering schemes. The first, tangential filtering [3], is the simplest possible version of anisotropic diffusion and applies averaging over three voxels (one central and two either side) along the normal to the direction of maximum local image gradient. In the absence of noise, the gradient along this normal is expected to be zero in any image composed of smooth, continuous regions. Since many medical image modalities produce images that conform to this behaviour, tangential smoothing is theoretically the least destructive (i.e. best edge-preserving) of the simple noise-filtering schemes (where simple is used in the sense that all voxels are treated equally). Gaussian filtering using a kernel with a standard deviation of 1 voxel and median filtering over the local neighbourhood of 9 voxels were also used. Finally, non-local means [21] was used to represent a state-of-the-art noise-filtering algorithm.

For all segmentations in the present study, the parameter estimation stage of the algorithm required an approximate tissue model to be used as an initialisation point for the optimisation. These models were prepared manually through visual inspection of histograms of sample images, in order to specify the number of distinguishable tissue

[4]Clearly, registration accuracy has a direct effect on the accuracy of subsequent partial volume segmentation. Therefore, a Monte-Carlo analysis of registration accuracy was performed. This indicated that the effects of registration error on partial volume estimates at the highest contrast tissue boundaries (around 10%) were less than the limits imposed by image noise (estimated as described by Olsen [26]). Therefore, registration accuracy was not a limiting factor for segmentation accuracy in this study.



(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 4: Voxel misclassification rates (a,b,c) and χ^2 per DOF (d,e,f) for segmentations of the re-simulated Brainweb T1 (a,d) T2 (b,e) and PD (c,f) images. The abscissa shows the image noise in all cases. 3σ error bars are shown.

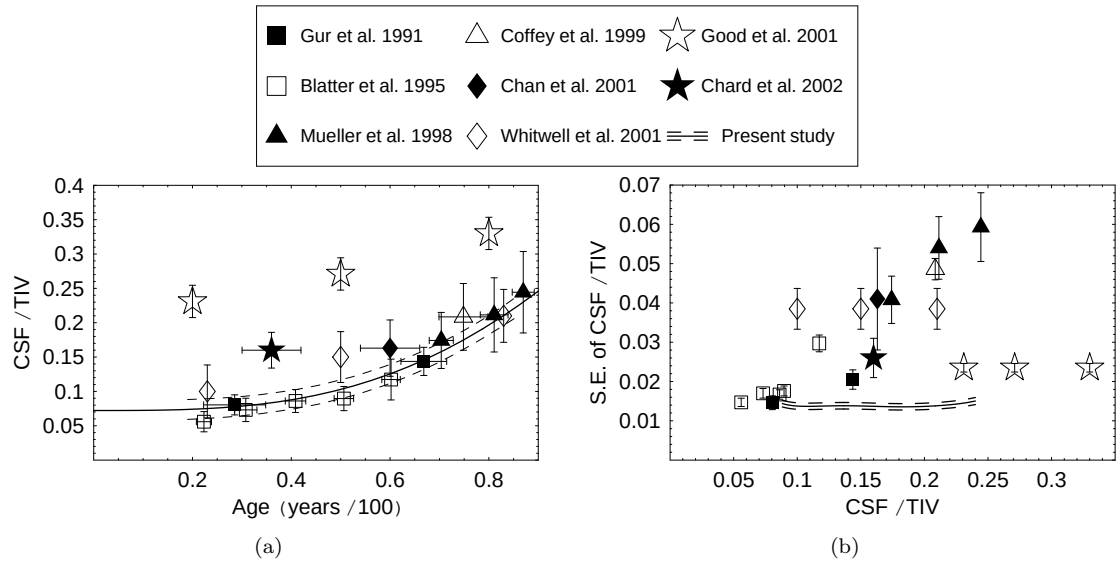
classes present and approximate means and standard deviations for each. In some cases, pairs of tissues were observed to have the same mean intensities, and so were represented using a single class. The allowable partial volume combinations were specified on the basis of prior anatomical knowledge. Six tissue classes were specified (air/bone, skin/muscle, fat, WM, GM and CSF) for segmentations of the clinical MR images and partial volume combinations were allowed between all pairs of classes except fat and skin/muscle with the tissues inside the skull (WM, GM and CSF).

4 Results

4.1 Evaluation using Simulated MR Data

Figure 4(a,b,c) 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. 2), obscuring the partial volume distributions. The inclusion of gradient information disambiguates partial volume voxels that occur at tissue boundaries i.e. positions of high gradient, but have the same intensity as pure tissue voxels, in this case. 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.

Figure 4(d,e,f) shows the χ^2 between images reconstructed from the fitted tissue model and partial volume maps and the original image data. This measure is more sensitive to errors in the estimation of tissue volumes in partial volume voxels than the misclassification rate, and so is more relevant when considering the differences introduced by the use of gradient terms (which primarily affect partial volume voxels occurring at tissue boundaries i.e. points of high intensity gradient). Note, however, that the EM algorithm optimises a correlate of the χ^2, and so this measure can only be applied comparatively between segmentations of the same data. The inclusion of the gradient terms significantly improves the quality of the segmentation result in all cases: the average improvement is 31.7%.



(a) (b)

Figure 5: Previous published TIV normalised CSF volume measurements (a). The standard errors on the TIV normalised CSF measurements (b).

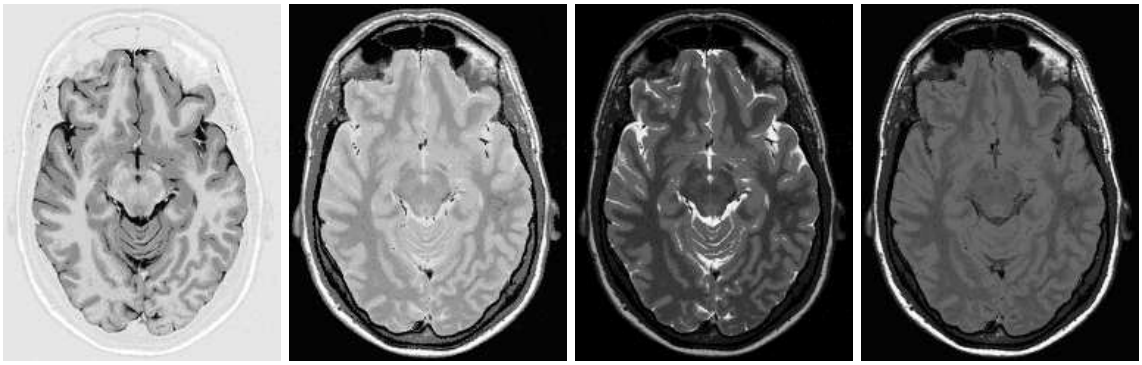
4.2 Evaluation using Clinical MR Data

The results of the comparison of TIV normalised CSF volume measurements are shown in Fig. 5(a). Points represent the published data; those with error bars on both the x and y directions were published numerically, whereas those with error bars only in the y direction were reconstructed from graphical representations. The solid curve shows a Weibull cumulative distribution function fit to the data from the present study (i.e. not to the previously published data). This is a flexible function that is widely used in areas such as reliability analysis: it is used here simply as a good empirical fit to the data [4]. The dashed curves represent the upper and lower 1σ error bounds. Apart from results generated using the SPM algorithm, there is no indication of significant differences between any of the previous studies and the study presented here. This indicates that any systematic error in the algorithm presented here is insignificant compared to biological variability for brain volume measurement tasks. The studies using the SPM algorithm, namely Good et al. [13] and Chard et al. [7], produced significantly higher CSF volume measurements than are supported by the consensus of the literature.

Fig. 5(b) shows the magnitude of the errors bars from Fig. 5(a) in the form of standard errors, plotted against the measurements themselves, in order to allow a more direct comparison of random error. The random errors achieved by the algorithm presented here are lower than those of the other techniques included in the study, whether considered in absolute terms or as percentage errors. Notably, since the algorithm is completely automatic and was applied to uni-spectral data in this study, it achieves lower random errors than the Gur et al. [15] and Blatter et al. [1] studies, which both used multispectral data and techniques involving significant amounts of manual intervention. Whilst the proposed algorithm is mathematically more sophisticated than the other algorithms included in the comparison, the number of parameters fitted is insignificant compared to the number of data in all cases, and so issues related to over-fitting need not be considered. Therefore, it is reasonable to compare the errors produced by each technique directly, since that will dictate the statistical power of the algorithm to detect inter-group volume changes when applied to clinical volume measurement tasks.

4.3 Noise Filtering using Multispectral Segmentation

Figure 6 shows the reconstructed (noise-free) multispectral MR data produced from the original images shown in Fig. 3. Figure 7 shows the ROM and Monte-Carlo stability of the reconstructed images. The Monte-Carlo stability provides a measure of the proportion of noise removed from the images by the regeneration process, whilst the ROM measures the number of voxel intensities modified by more than 3σ, where σ is the standard deviation of the image noise, measured using an independent technique. The ROM should therefore be equal to 0.27% of the number of voxels in the image if the model matches the data perfectly which, given the size of the images, equates to approximately 60 voxels. Tangential smoothing, which theoretically should be the least destructive form of simple noise removal, applies inappropriate smoothing to on average 2.7% of the voxels due to destabilisation of the local gradient calculation by the noise. Median filtering is twice as destructive to the image contents, without greatly



(a) (b) (c) (d)

Figure 6: Reconstructed MR data: IRTSE(a), VE (PD) (b) VE (T2) (c) and FLAIR (d)

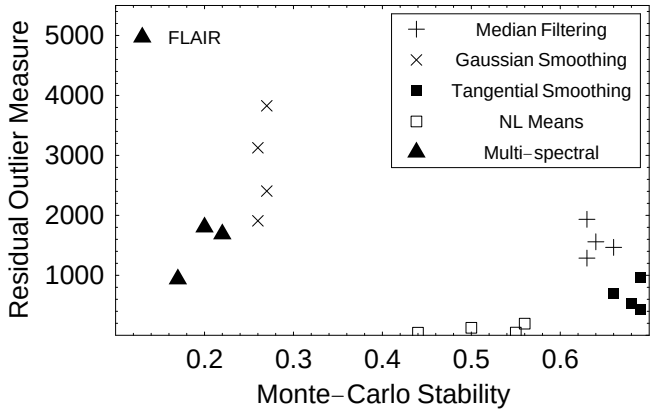


Figure 7: Residual outlier measure vs. Monte-Carlo stability.

improving noise removal characteristics, whilst Gaussian filtering removes significantly more noise at the expense of degrading the images significantly. Excluding the FLAIR image result, the multispectral image reconstruction is on average approximately twice as destructive as tangential smoothing: the intensity model fits the data well for all but around 6% of the voxels in the images. The ROM for the FLAIR image is higher as such images are susceptible to flow artefacts. The cerebral vasculature, and some voxels in the CSF compartments, therefore form extra tissue classes not included in the tissue model, leading to misinterpretation of such voxels. The non-local means algorithm represents a state-of-the-art noise filter, and removes approximately half of the noise whilst achieving an average ROM of ≈ 100, about 50% higher than the theoretical optimum. This is due to the incorporation of an explicit test of similarity; during the averaging process the voxels are weighted by their similarity to the voxel being filtered. The algorithm therefore avoids averaging over voxels drawn, for example, from different sides of a tissue boundary, and so preserves image structure. However, when the similarity test is applied to voxels with high levels of noise i.e. those in the tails of the intensity distributions, few voxels with similar intensities will be found. Such voxels will therefore be left largely unmodified due to a lack of data to incorporate into the averaging process. Non-local means therefore achieves a low ROM by definition, at the expense of applying a spatially varying amount of smoothing and producing an output with spatially varying noise.

5 Discussion and Conclusions

In previous work we have extended standard volumetric estimation techniques to multiple images. In this work a method to utilise not only the intensity information but also the local intensity gradient information in a MR image was described. The method can be considered as an alternative both to the assumption of local regional smoothness and to local resampling of the prior probabilities, which have previously been suggested by other authors [19]. Unlike these methods, local structural information is used directly in a manner that is quantitatively related to the image formation process.

Several evaluation methodologies were applied to the algorithm. In unispectral segmentation of simulated data using intensity information alone, the algorithm achieved almost optimal tissue classification. The addition of gradient information resulted in significant improvements in images where the pure tissue intensity distributions overlap to a significant extent, due to the ability of gradient information to disambiguate pure tissue and partial volume classes with similar intensities. The χ^2 measure indicated significant improvements in partial volume estimation in all cases when gradient information was included. Evaluation of unispectral segmentation on clinical images indicated that the algorithm showed no significant systematic error in the CSF volume measurement task, and achieved lower random errors than alternative techniques. It should be noted that this study did not evaluate the performance on repeated acquisitions from the same subject, since a lower random error limit is imposed by the intrinsic biological variability in CSF measurements, but it does indicate the expected performance of the algorithm in transverse volume measurement studies. The proposed algorithm is therefore capable of providing increased statistical power with which to detect volume changes or differentiate between subject groups in such studies. In addition, the comparison indicated that the SPM software, a de-facto standard in medical image analysis, produced significantly higher estimates of CSF volume than are supported by the consensus of the literature.

Evaluation of the multi-spectral application of the algorithm focused on measuring the goodness-of-fit of the tissue model, and indicated that the model fits the data well for all but approximately 6% of the voxels. In addition, this evaluation methodology demonstrates the use of the multispectral segmentation as a noise-filtering technique. In general multi-spectral filtering was shown to be no more destructive to image contents than median filtering but removed more image noise than Gaussian smoothing. Therefore, such reconstruction techniques may be a useful way of processing multiple MR acquisitions.

The main limitation of the algorithm presented here is the assumption, in common with previous authors (e.g. [31, 19]) of a uniform distribution for partial volume voxels. As described above, other authors have investigated non-uniform distributions with some success, and Chiverton [8] has investigated the incorporation of non-uniform partial volume distributions into the formulation described here. In addition, the EM update scheme used here assumes spatial independence in the noise. Finally, the region to be segmented must contain a sufficient quantity of pure tissues to allow model fitting; it may therefore not be applicable to segmentations limited to regions such as the basal ganglia, which consist primarily of partial volume voxels at the resolution of MR images.

6 Acknowledgements

We acknowledge the contributions of Prof. A. Jackson, Dr. M. Pokric, Dr. M. Scott and Dr. D. Williamson to the early stages of this work, and the the help of Profs. P. Rabbitt and C. Hutchinson with MR data collection. Part of this work was conducted during the course of a special MRC training fellowship and support was also received under the EPSRC/MRC funded MIAS IRC (from Medical Images and Signals to Clinical Information Interdisciplinary Research Collaboration), under EPSRC GR/N14248/01, and UK MRC Grant No. D2025/31. The software and associated technical notes are available at www.tina-vision.net.

References

[1] D D Blatter, E D Bigler, S D Gale, S C Johnson, C V Anderson, B M Burnett, N Parker, S Kurth, and S D Horn. Quantitative volumetric analysis of brain MR: Normative database spanning 5 decades of life. *AJNR Am J Neuroradiol*, 16:241–251, 1995.

[2] P A Bromiley. TINA Memo No. 2008-009: The TINA medical image segmentation algorithm: Mathematical derivations and proofs. Technical report, Imaging Sceince and Biomedical Engineering, School of Cancer and Imaging Sciences, University of Manchester, 2008. http://www.tina-vision.net/docs/memos/2008-009.pdf.

[3] P A Bromiley, N A Thacker, and P Courtney. Non-parametric image subtraction using grey level scattergrams. *Image and Vision Computing*, 20:609–617, 2002.

[4] P A Bromiley, N A Thacker, and A Jackson. Trends in brain volume change with normal ageing. In *Proc. MIUA'05*, pages 247–250, 2005.

[5] P A Bromiley, N A Thacker, M L J Scott, M Pokrić, A J Lacey, and T F Cootes. Bayesian and non-Bayesian probabilistic models for medical image analysis. *Image and Vision Computing*, 21(10):851–864, 2003.

[6] D Chan, N C Fox, R I Scahill, W R Crum, J L Whitwell, G Leschziner, A M Rossor, J M Stevens, L Cipolotti, and M N Rossor. Patterns of temporal lobe atrophy in Alzheimer's disease. *Ann Neurol*, 49:433–442, 2001.

[7] D T Chard, G J M Parker, C M B Griffin, A J Thompson, and D H Miller. The reproducibility and sensitivity of brain tissue volume measurements derived from an SPM-based segmentation methodology. *J Magn Reson Imaging*, 15:259–267, 2002.

[8] J P Chiverton. *Probabilistic Partial Volume Modelling of Biomedical Tomographic Image Data*. PhD thesis, Center for Vision, Speech and Signal Processing, University of Surrey, Guildford, Surrey GU2 7HX, U.K., Aug 2006.

[9] H S Choi, D R Haynor, and Y Kim. Partial volume tissue classification of multichannel magnetic resonance images-a mixture model. *IEEE Trans Med Imaging*, 10:395–407, 1991.

[10] C E Coffey, J A Saxton, G Ratcliff, R N Bryan, and J F Lucke. Relation of education to brain size in normal aging: Implications for the reserve hypothesis. *Neurology*, 53:189–196, 1999.

[11] A P Dempster, N M Laird, and D B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Society*, 39:1–38, 1977.

[12] K Fukenaga. *Introduction to Statistical Pattern Recognition*. Academic Press, San Diego, 2nd edition, 1990.

[13] C D Good, I S Johnsrude, J Ashburner, R N A Henson, K J Friston, and R S J Frackowiak. A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage*, 14:21–36, 2001.

[14] H Gudjbartson and S Patz. The Rician distribution of noisy MRI data. *Magn Reson Med*, 34(6):910–914, 1995.

[15] R C Gur, P D Mozley, S M Resnick, G L Gottlieb, M Kohn, R Zimmerman, G Herman, S Atlas, R Grossman, D Berretta, R Erwin, and R E Gur. Gender differences in age effect on brain atrophy measured by magnetic resonance imaging. *Proc Natl Acad Sci USA*, 88:2845–2849, 1991.

[16] N W John, N A Thacker, M Pokrić, and A Jackson. An integrated simulator for surgery of the petrous bone. In *Medicine Meets Virtual Reality 2001*, pages 218–224. IOS Press, 2001.

[17] N Joshi and J M Brady. A non-parametric model for partial volume segmentation of MR images. In *Proceedings BMVC'05*, pages 919–928, 2005.

[18] R K-S Kwan, A C Evans, and G B Pike. MRI simulation-based evaluation of image-processing and classification methods. *IEEE Trans Med Imaging*, 18(11):1085–1097, 1999.

[19] D H Laidlaw, K W Fleischer, and A H Barr. Partial-volume Bayesian classification of material mixtures in MR volume data using voxel histograms. *IEEE Trans Med Imaging*, 17(1):74–86, 1998.

[20] K Van Leemput, F Mayes, D Vandermuelen, and P Suetens. A unifying framework for partial volume segmentation of brain MR images. *IEEE Trans Med Imaging*, 22(1):105–119, 2003.

[21] J V Manjon, M Robles, and N Thacker. Multispectral MRI de-noising using non-local means. In *Proc. MIUA'07*, pages 41–45, Aberystwyth, Wales, 2007.

[22] E Mueller, M M Moore, D C R Kerr, G Sexton, R M Camicioli, D B Howieson, J F Quinn, and J A Kaye. Brain volume preserved in healthy elderly through the eleventh decade. *Neurology*, 51:1555–1562, 1998.

[23] W J Niessen, K L Vincken, J Weickert, B M Ter Haar Romeny, and M A Viergever. Multiscale segmentation of three-dimensional MR brain images. *International Journal of Computer Vision*, 31:185–202, 1999.

[24] L Nocera and G C Gee. Robust partial volume tissue classification of cerebral MR scans. In K M Hanson, editor, *Proc. SPIE Medical Imaging 1997: Volume 3034 of SPIE Proceedings*, pages 312–322, 1997.

[25] A Noe and J C Gee. Partial volume segmentation of cerebral MRI scans with mixture model clustering. In *Proc. IPMI'01*, pages 423–430, 2001.

[26] S I Olsen. Estimation of noise in images: An evaluation. *CVGIP: Graphical Models and Image Processing*, 55:319–323, 1993.

[27] D L Pham and J L Prince. Unsupervised partial volume estimation in single-channel image data. In *Proc. IEEE Workshop Mathematical Methods in Biomedical Image Analysis-MMBIA'00*, pages 170–177, 2000.

[28] M Pokrić, N A Thacker, and A Jackson. The importance of partial voluming in multi-dimensional medical image segmentation. In *Proc. MICCAI'01*, pages 1293–1294, 2001.

[29] M Pokrić, N A Thacker, M L J Scott, and A Jackson. Multi-dimensional medical image segmentation with partial voluming. In *Proc. MIUA'01*, pages 77–80, 2001.

[30] S Ruan, C Jaggi, J Xue, J Fadili, and D Bloyet. Brain tissue classification of magnetic resonance images using partial volume modelling. *IEEE Trans Med Imaging*, 19:1179–1187, 2000.

[31] P Santago and H D Gage. Quantification of MR brain images by mixture density and partial volume modelling. *IEEE Trans Med Imaging*, 12:566–574, 1993.

[32] P Santago and H D Gage. Statistical models of partial volume effect. *IEEE Trans Image Processing*, 4:1531–1540, 1995.

[33] D W Shattuck, S R Sandor-Leahy, K A Schaper, D A Rottenberg, and R M Leahy. Magnetic resonance image tissue classification using a partial volume model. *NeuroImage*, 13:856–876, 2001.

[34] N A Thacker, A Jackson, and D Moriarty. Improved quality of re-sliced MR images using renormalised sinc interpolation. *J Magn Reson Imaging*, 10:582–588, 1999.

[35] N A. Thacker, A J Lacey, and P A Bromiley. Validating MRI field homogeneity correction using image information measures. In *Proc. BMVC'02*, pages 626–635, 2002.

[36] N A Thacker, M Pokrić, and D C Williamson. Noise filtering and testing illustrated using a multi-dimensional partial volume model of MR data. In *Proc. BMVC*, pages 909–919, Kingston, London, 2004.

[37] N A Thacker, D C Williamson, and M Pokrić. Multi-dimensional MR image segmentation with gradient analysis. In *Proc. MIUA'05*, pages 23–26, Bristol, 19th-20th July, 2005.

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

[39] S Webb. *The Physics of Medical Imaging.* Medical Science Series. Adam Hilge, Bristol, Philadelphia and New York, 1988.

[40] J L Whitwell, W R Crum, H C Watt, and N C Fox. Normalisation of cerebral volumes by use of intracranial volume: Implications for longitudinal quantitative MR imaging. *Am J Neuroradiol*, 22:1483–1489, 2001.

[41] D C Williamson, N A Thacker, S R Williams, and M Pokrić. Partial volume tissue segmentation using grey-level gradient. In *Proc. MIUA'02*, pages 17–20, 2002.