Tina Memo No. 2004-009
Published in Annals of the BMVA

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

N.A. Thacker, P.A. Bromiley and D.C. Williamson

Last updated
16 / 3 / 2006

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

Author's note: a modified version of this paper was published in the Annals of the BMVA. A preprint is available as Tina Memo No. 2009-002.

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

N.A. Thacker[a], P.A. Bromiley[a] and D.C. Williamson[b]
a Division of Imaging Science and Biomedical Engineering,
School of Medicine, University of Manchester, M13 9PT, U.K.
neil.thacker@man.ac.uk
b Centre for Magnetic Resonance, Department of Chemistry,
University of York, YO10 5DD, U.K.

Abstract

We present a new algorithm for the segmentation of medical image volumes, which addresses the problem of partial volume tissue estimation, where a mixture of tissues combine to form the intensity value for a particular voxel. In addition, the algorithm is capable of using multiple image volumes, and the associated multi-dimensional image gradient, to increase tissue separability. It uses the Expectation-Maximisation (EM) algorithm to perform clustering in image intensity and gradient histograms. Bayes theory is used to generate probability maps for the most likely tissue volume fraction within each voxel, in contrast to previous approaches, which typically compute the most likely tissue class label.

Evaluation of the algorithm consisted of three stages, all of which used MR data sets of the normal human brain. First, the improvement in the model parameter stability gained through the inclusion of gradient information was evaluated. Second, the improved segmentation accuracy of multi-dimensional approaches was demonstrated by assessing the errors on reconstructed images produced from the segmentation result. Finally, the absolute accuracy of the segmentation when applied to an exemplar medical problem, the measurement of cerebrospinal fluid (CSF) volumes, was evaluated through comparison with a "bronze-standard" consisting of previous published measurements.

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 [40]. Derived 3D models of scanned anatomy can also be applied to pre-operative planning, surgical rehearsal and training [26].

The literature describes several approaches to the problem of medical image segmentation. These can be broadly divided into two types: those operating in the image plane, and those operating in feature space. Some level of manual intervention is also typically required. Most image-plane based approaches can be related in some sense to edge detection: for example, [14] directly used a combination of edge detection and manual tracing to segment structures within the brain from MR images. Several classes of algorithm combine edge detection with varying levels of a-priori knowledge of the expected shape of structures. For example, snake-based or active contour algorithms (e.g. [4]) combine edge detection with a relatively weak constraint of local edge smoothness, whereas active shape models (e.g. [17, 32, 46]) involve the construction of a comprehensive model of the expected shape of a structure, followed by location of that structure in the image. Alternatively, a definition of intensity similarity rather than intensity difference can be applied, as in region-growing approaches ([25]). Such algorithms require manual placement of a seed point within a region, followed by automatic identification of all connected voxels of similar intensity.

Feature-space approaches operate on some function of the image, rather than the image itself: an intensity histogram is the most common feature space used. Segmentation thus becomes the problem of identifying clusters in the intensity histogram. The simplest approach is binary thresholding at points of low data density in the histogram, which may be selected empirically as in the MIDAS software used in [10] and [44]. Again, region growing from manually selected seed points can also be applied, as in [24]. Split-and-merge algorithms (e.g. [31]) form

a related approach, but one which does not require the selection of seed points. However, these approaches are sensitive to partial volume effects and noise, which can create artificial connections and disconnections respectively. Classifier methods seek to avoid these problems: for example, the k-nearest-neighbour classifier (e.g. [5]) takes a set of manually segmented training data and identifies each voxel in the data to be segmented with the voxel in the training data having the most similar intensity, thus preserving the number of classes.

Clustering algorithms, such as that used in the SPM99 software (e.g. [21], [11]), essentially perform the same function as classifiers without the use of training data. Three types of clustering algorithm are in common use. The k-means or ISODATA algorithm [15] iteratively computes the mean of each class, then identifies each voxel with the class having the most similar mean, and is thus related to the split-and merge algorithm. The fuzzy c-means algorithm [2] generalises this approach to allow for the overlap of classes, using fuzzy set theory. The Expectation-Maximisation (EM) algorithm [15, 30] applies the clustering approach based on the assumption that the data follow a Gaussian mixture model.

One significant advantage of feature-space approaches is that they can easily be adapted to deal with multi-dimensional data, where several images of the same scene, acquired using different imaging modalities, are available. Such data is frequently encountered in medical image segmentation tasks: for example, both [5] and [24] present cerebral volume measurements based on multi-dimensional segmentations of proton density (PD) and T2-weighted MR image volumes. It is difficult to combine, for example, the results of edge detection across multiple images in a statistically well-founded way. In contrast, clustering approaches can be applied to feature spaces of arbitrary dimensionality. Therefore, in the simplest case, additional image volumes can be incorporated into the segmentation task by adding their intensities to the intensity histogram as an extra dimension.

Reasonable forced-choice segmentations can be achieved for many realistic medical data sets containing only a few tissues using only knowledge of the distributions of the grey-levels for each tissue type [16, 28]. However, such segmentations can be quite inaccurate and must often be supplemented by local smoothness assumptions, noise filtering or regional constraints. If more than three tissues are present it is necessary to examine the distribution of the grey levels in two or more registered images in order to achieve 10% or better volumetric estimation accuracy [35]. In addition, much work regarding probabilistic analysis of voxel contents uses the assumption that there is a negligible occurrence of pixels exhibiting partial volume effects i.e. containing more than one tissue. In fact, Laidlaw [28], Noe [34] and our own work [35, 36] have all demonstrated that the partial volume process is significant, especially with respect to MR images of the brain, with partial volume processes affecting as many as 40% of voxels in thick slice acquisitions. Such data cannot be analysed using a set of mutually exclusive, pure-tissue hypotheses: accurate interpretation of the data requires that partial volume distributions are modelled. Fortunately, the physics of the image formation processes in a wide variety of medical imaging modalities allow partial volume distributions for paired tissue combinations to be modelled as a simple, linear process [42]. Relative fractions of different tissues therefore contribute proportionately to the intensity in a given voxel.

In this paper, we define a computational approach to model partial volume distributions in multi-dimensional images produced by linear formation processes. The Expectation Maximisation (EM) algorithm [3, 18, 23] is utilised for the estimation of parameters of the multi-dimensional models. We have previously extended the EM approach for terms which describe both the pure and partial volume tissues [36]. A method for combined multi-image segmentation and field inhomogeneity correction within an EM framework has been suggested by Wells [43]. The conventional interpretation of this algorithm is as a system for estimating unseen variables: here it should simply be interpreted as an efficient way of fitting a density model to a multi-dimensional data set.

Further improvements in segmentation accuracy can be gained through extracting more information from the images under analysis, for example estimates of local image structure. Some previous work has attempted to modify the Bayesian formalism to account for local structure, such as boundaries, by using estimates of prior probabilities based upon a local resampling of data or Markov Random Field (MRF) formulations [29]. The freedom to take such a step is linked to the classic problem of identification of prior probabilities in Bayesian methods [9]. However, in the work presented here we take an alternative view that local information regarding image structure can be included directly within the density model. The use of Bayes theorem can be modified to include local image derivatives as well as grey levels using only the same assumptions underlying the standard approaches. In low-contrast situations this approach effectively doubles the information available for partial volume estimation.

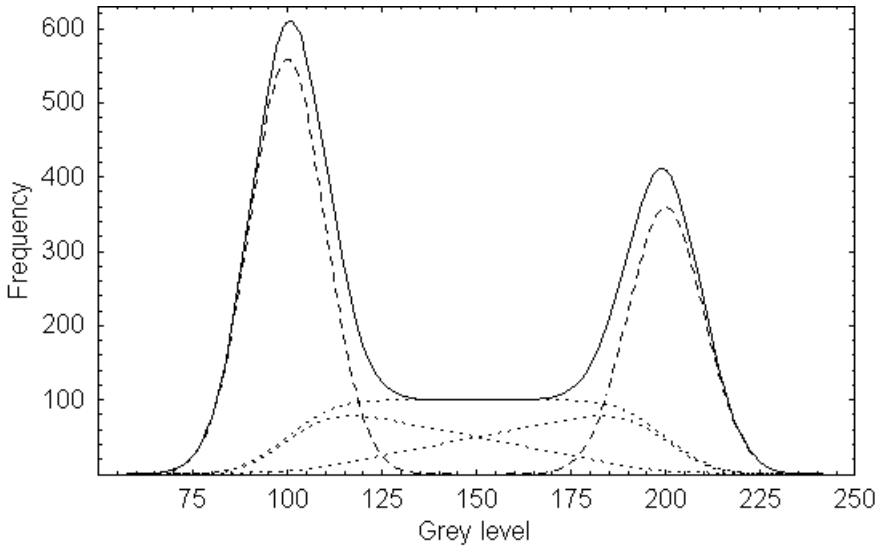


Figure 1: An example partial volume tissue model. Pure tissues have Gaussian distributions (dashed line), whereas mixtures of tissues take form of triangular distributions convolved with a Gaussian (dotted lines). The model components are summed to give the overall distribution (solid line).

2 Method

2.1 Multi-dimensional Grey Level and Slope Analysis

Consider a data set consisting of multiple, pre-registered images of the same region, where $\mathbf{g}$ is the vector of grey-levels of a given voxel in the set of images and s is the local multi-dimensional image gradient, the square-root of the sum-squared derivatives. The conditional probability that a pixel may be assigned to a specific tissue class t given its grey-level 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 of the tissue class t, $P(\mathbf{g},s|t)$ is the likelihood of the grey level $\mathbf{g}$ and the derivative s given a tissue class t and $P(\mathbf{g},s)$ is the global likelihood of the instance of data $\mathbf{g}$,s. Assuming independence of the grey level and slope probabilities it would be possible to write $P(\mathbf{g},s|t)$ as the product of two terms $P(\mathbf{g}|t)P(s|t)$. However, the slope value must be dependent upon the grey level, as the partial volume process is caused by tissue boundaries. Therefore, we have instead

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

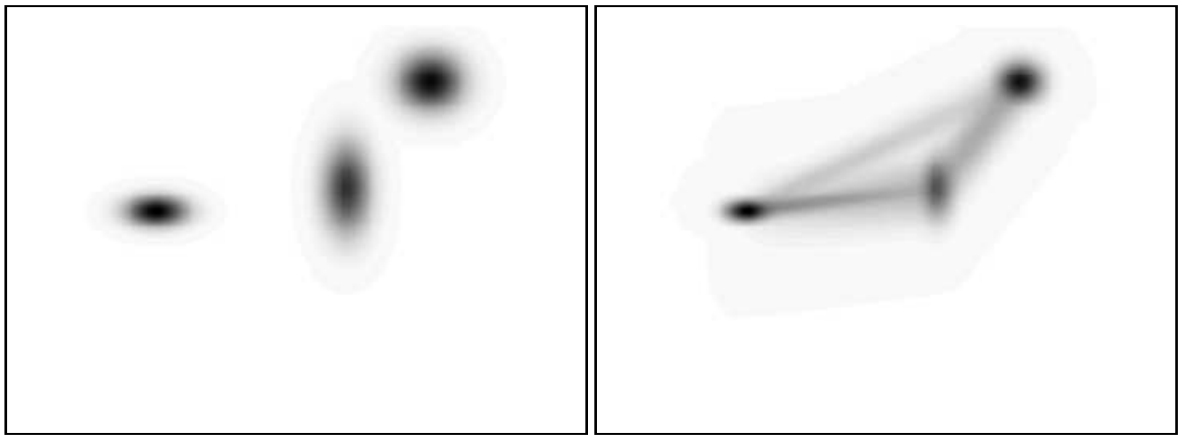
where $P(\mathbf{g}|t)$ is the expected distribution of grey levels for each class and $P(s|\mathbf{g},t)$ is the expected distribution of slope values as a function of grey level for each class.

Although partial volume voxels could be accounted for in this formulation through the inclusion of a specific class for each pair of tissues, the probability of classification of such a voxel is of little direct use: estimates of the most probable partial volume contribution from each tissue are required instead. Therefore, a model of the partial volume contributions for each tissue contributing to the mixture is needed, in the form of the two distributions $P(\mathbf{g},s|tr)$ and $P(\mathbf{g},s|rt)$. The ordering of the tissues in this notation is insignificant: $P(\mathbf{g},s|tr)$ can be interpreted as the volumetric contribution from tissue t in the presence of tissue r. Assuming a linear image formation process, these distributions $d_{tr}(\mathbf{g})$ can be estimated as the convolution of a triangular density along a line between the two tissue means in feature space with a multi-dimensional point spread function representing the image noise.

This formulation is equivalent to that suggested by other authors [28]. However, in order to include local slope information, terms describing the slope density $\rho_{tr}(s,\mathbf{g})$ are required, which must also be functions of the grey level $\mathbf{g}$. Using the formulation

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

the conditional probability of a grey level being due to a certain mechanism n (either a pure or mixture tissue



(a) (b)

Figure 2: An example of 2D distributions generated from the model for grey levels from two images using only pure tissue distributions (a) and combined distributions of pure tissues and partial volumes (b).

component) can be calculated using Bayes theory, as

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

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 [20] in the summed likelihood term, which enables separation of pathological tissues.

Through using this representation, rather than working with class assignment probabilities, it is possible to obtain the most probable volumetric measurement V_t for each tissue t given the observed data $\mathbf{g_v}$ in voxel v,

$$V_t(\mathbf{g_v}, s_v) \;=\; P(t|\mathbf{g_v}, s_v) + \sum_r P(tr|\mathbf{g_v}, s_v) \tag{2}$$

The approach is easily simplified to remove the slope information by simply setting $\rho = 1$ in 1.

2.2 Grey Level Model

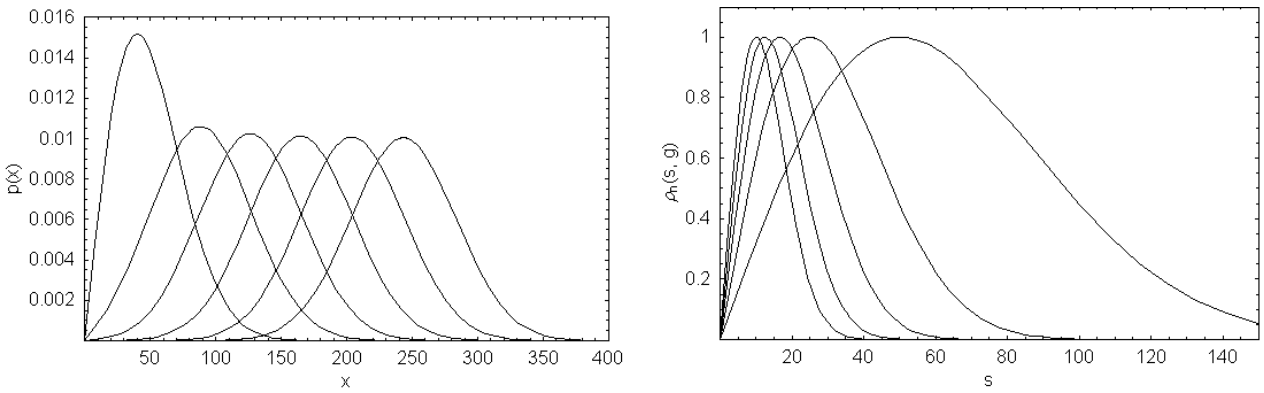
The formulation described above requires the assumption of distributions both for the pure tissues and for the point spread functions convolved with triangular distributions to describe the partial volume contributions. These distributions represent both the inherent tissue variability and the measurement noise. Assuming Gaussian distributions, each pure tissue t is described by

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

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

In [28] partial volume distributions were modelled as a uniform (1D) distribution convolved with a fixed point spread function. This model is valid for an equal prior probability of all possible partial volumes. In the approach described here, using an assumption of a linear image formation process, this partial volume distribution can be thought of as being composed of two triangular distributions convolved with a Gaussian, where $d_{tr}(\mathbf{g})$ is the local density estimate for tissue t generated by a partial voluming process with tissue r. This model can take a range of different forms, depending on the relative scales of the measurement noise and the difference between the pure tissue mean grey levels. Assuming that the latter dominates, multi-dimensional partial volume distributions can be modelled as the product of two terms: a 1D projection $T_{tr}(h)$ of the partial volume distribution onto the vector connecting the two pure tissue means $\mathbf{M_t}$ and $\mathbf{M_s}$, and the value of a Gaussian at the normal distance $N(\mathbf{g})$ of the data point $\mathbf{g}$ away from this line. Therefore

$$d_{tr}(\mathbf{g}) \;=\; \beta_{tr} T_{tr}(h) e^{-\frac{1}{2}N(\mathbf{g})^{\mathbf{T}}\mathbf{C_h}\mathbf{N}(\mathbf{g})}$$



(a) (b)

Figure 3: Rician distributions with varying mean (a) ρ distributions with $\gamma = 1$ for varying scale (b).

where h is a fractional distance between two centres of distribution $(0 < h < 1)$

$$h \;=\; \frac{(\mathbf{g} - \mathbf{M_s})^T C_h (\mathbf{M_t} - \mathbf{M_s})}{|(\mathbf{M_t} - \mathbf{M_s})^T C_h (\mathbf{M_t} - \mathbf{M_s})|}$$

C_h is an inverse covariance matrix

$$C_h \;=\; C_t h \;+\; C_s(1-h)$$

and β_{tr} is a constant which gives unit normalisation. As the definitions of h and C_h are dependent the first two steps of this process must be performed as an iterated closest point algorithm. Examples of such distributions for 2D image data are shown in Fig.2: the partial volume contributions are visible as elongated structures in feature space, linking the centres of the pure tissue distributions.

2.3 Multi-dimensional Slope Model

In previous work [45] we demonstrated the inclusion of slope information into the feature space for segmentation of single images. The mean of the slope distribution was estimated from the tissue contrast, whilst the width of the distribution was determined by the image noise. In this work, the use of local image slopes was extended to multi-dimensional data. However, rather than attempting to model the slope 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 slope distribution was used.

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

$$s \;=\; \sqrt{\sum_{k}^{N} \frac{\nabla_x I_k^2 + \nabla_y I_k^2}{\sigma_k^2}} \;-\; \lambda \tag{3}$$

where $\nabla_x I_k$ is the gradient component of the image k in x direction, $\nabla_y I_k$ is gradient component of the image k in y direction, σ_k is standard deviation of noise in the image k, N 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 slope distribution. In addition, avoiding the use of a z gradient allows an analysis of individual slices and avoids having to deal with edge effects.

If the z component were 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 [12]. 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 slope, and partial volume voxels with a finite mean slope. 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[1] [22], as shown in Fig./reffig:rice. In practice this will not be the case, but it was found empirically that multiplication with a simple power law dependent on a parameter γ was sufficient to account for this effect, giving

$$\rho_n(s, \mathbf{g}) \;=\; \rho_n(s, a) = \frac{s^\gamma}{a^{\gamma+1}} \exp^{-\frac{s^2}{2a^2}} \tag{4}$$

where a is the scale parameter. The shape of this distribution for $\gamma = 1$ is shown in 3(b).

A set of logical constraints can be applied to the scale parameter a. It was found [45] that the distributions in gradient vs. intensity projections of the feature space for MR images form approximately elliptical features between pairs of pure tissues, with the highest gradients occurring in voxels consisting of equal contributions from two tissues i.e. at edges. This would be expected of any linear image formation process, and indicates that the scale parameter is linearly dependent on the fractional volumetric contribution from each tissue class present in that voxel. Furthermore, the ends of the partial volume distributions must match exactly the pure tissue distributions. Therefore, a reasonable model for the scale parameter a in any given voxel is

$$a^2(q) = N(qa_t + (1-q)a_s)^2 + a_{tr}^2 w(q) \sum_k^N \frac{(m_{kt} - m_{ks})^2}{\sigma_k^2} \tag{5}$$

where a_t and a_{tr} form a matrix of scale parameters for the pure tissues and partial volume combinations respectively. 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 j determined from the grey-level multi-dimensional probability model $(0 < d_{tr}(\mathbf{g})/(d_{tr}(\mathbf{g}) + d_{rt}(\mathbf{g}))) < 1$. The required elliptical dependency upon q can then be introduced by defining $w(q) = 1 - 4(q - 0.5)^2$, such that $a(0) = a_t\sqrt{N}$ and $a(1) = a_s\sqrt{N}$, which are the pure tissue tissue distribution parameters. This selection of ρ gives identical values for both tissue partial volumes tr and rt at a given $\mathbf{g}, s$.

2.4 Parameter Estimation

Parameters of the model can be iteratively estimated using the Expectation Maximisation (EM) approach [18] [43], which is used to estimate the parameters by maximising the likelihood of the data distribution. The conventional interpretation of this algorithm is that it can be applied to estimating the parameters of Gaussian mixture models. However, as shown in Appendix 1, it has more general applicability and can be applied to distributions other than Gaussians provided that the locations and shapes of the mixture density functions are parameterised in terms of centroids and other moments.

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 1. The Maximisation step is then applied for the estimation of; normalisation parameters for pure and mixture tissues, (f_t , f_{tr}), mean ($\mathbf{G_t}$) and covariance matrices (C_t) which take the form

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

$$G'_t = \frac{1}{V}\sum_v^V P(t|\mathbf{g}_v, s_v)\mathbf{g}_v \qquad \textit{and} \qquad {C'}_t^{-1} \;=\; \frac{1}{V}\sum_v^V P(t|\mathbf{g}_v, s_v)(\mathbf{g}_v - \mathbf{G_t}) \otimes (\mathbf{g}_v - \mathbf{G_t})^T$$

The estimated values for f_n correspond to the total volumetric estimate of the tissue process for the voxel. Thus its interpretation as a Baysian prior is such that this value is at least consistent with the sample under analysis. Even so, this approach can produce problems when attempting to quantify changes between two data sets, when it might be necessary to use one consistent value in order to maximise repeatability of measurement [40]. Terms f_{tr} and f_{rt} are equal and sum to the total number of 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.

[1]The noise distribution in MR magnitude images takes the same form, as they are also composed of the root-mean-square of two quantities each typically having Gaussian random noise, namely the real and imaginary images.

The standard proof of convergence for the EM algorithm implies that the slope 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 slope data μ_{data} to the theoretical mean value μ_{model} calculated from the slope model. The value of a_n for the pure t or mixture tissue component tr is then re-calculated during the Maximisation step using

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

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

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. The general form of the update for the pure or partial volume slope parameters is then

$$a'_n = \frac{\mu^n_{data}}{\mu^n_{model}} a_n \tag{6}$$

These update equations generate values of partial volume slope distribution parameters which 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 5 from the results of the Expectation step. Although this process results in an exact likelihood estimate of the pure tissue parameters a_t it results in only a proportional under-correction of the partial volume terms. This would, however, appear to be enough to ensure rapid convergence.

3 Evaluation Methodology

Evaluation of the absolute accuracy of segmentation algorithms is typically performed through comparison with "gold-standard" segmentations. In the case of medical image segmentation, these are usually image data sets that have been segmented by hand by an expert observer. However, the production of such gold-standard data sets is time consuming, limited to labelling of the dominant tissue type in each voxel, and is particularly difficult for some clinical images, such MR images of the brain, due to the subjective nature of the delineation of many anatomical regions. Since the algorithm described here was intended to accurately model partial volume contributions to each voxel, and produced probability maps representing the most probable tissue volume fraction in each voxel rather than the most probable single tissue label, an evaluation methodology was adopted that avoided the need for gold-standard data. In addition, the method avoided reliance on simulated data sets, which ultimately could only prove consistency between the segmentation algorithm and the simulation process, and instead used in-vivo image data.

The bulk of the evaluation work presented here was based on MR images of the brain. 70 normal volunteers (32 male and 38 female) ranging in age from 19 to 85 years with a mean age of 57 $\pm$ 20 years were recruited. None of the subjects had vascular risk factors, neuro-logic disease, or cognitive or psychiatric problems. The local ethics committee approved the research, and informed consent was given for inclusion in the study by the subjects. All subjects underwent MR imaging with a 1.5-T system (ACS-NT, with PowerTrack 6000 gradient subsystem; Phillips Medical Systems, Hamburg, Germany) with a birdcage head coil receiver. Fast spin-echo inversion-recovery (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. In addition, in order to demonstrate segmentations of multi-dimensional data, several additional MR image types were obtained for one of the subjects: variable echo proton density and T2 (VE (PD) and VE (T2)), and fluid-attenuated inversion recovery (FLAIR) images. These pulse sequences were chosen for their good tissue separation and availability in a clinical environment. Finally, in order to illustrate application of the algorithm to tissues other than those present in the brain, one surface coil T2(PD) image of the leg was acquired and corrected for field inhomogeneity using the algorithm described in [41]; however, this was only used in qualitative evaluations.

The evaluation comprised three stages. The first was based upon the observation that good models of data behaviour should have better-constrained parameters than poor models. A study was therefore designed to assess the benefits of the inclusion of slope information in the feature space through the resulting improvement in tissue model parameter stability. Since the effect of this additional information will be greatest when the quantity of information available in the form of grey-level data is the least, these experiments were performed on single-dimensional data. Ten subjects were selected at random from the available data sets. The IRTSE images from each subject were used, and an initial, approximate set of model parameters was specified to act as a consistent starting point for the optimisation. Each data set was then segmented forty times, both with and without the inclusion of gradient information in the feature space, with random perturbations applied to the initial tissue mean grey level parameters prior to each segmentation. The standard deviations of the optimised grey matter, white matter and CSF mean grey-levels across the forty segmentations were calculated. The experiment was repeated with varying levels of initial randomisation from 50 and 200 grey-levels; the upper value of δ was selected to be the highest value that did not allow any two of the initial model tissue means to become equal. Finally, the standard deviations of the tissue mean grey levels after optimisation were averaged across the ten subjects, and the results for segmentations with and without gradient information were compared. In order to provide a baseline for this comparison, identical experiments were performed using a straw-man algorithm, based on the use of a conventional optimiser (downhill simplex) to perform a χ^2 based histogram fit for the partial volume model (i.e. without gradient information). Control parameters for this algorithm were set in accordance with guidelines given in [37].

The second stage of evaluation was based on the idea that, since the segmentation technique presented here provides a volumetric model description of voxel content, it should be capable of regenerating the original image content for each input image in the case of zero noise. The process is simply one of using the estimated model parameters to generate the expected grey level values, using the linear equations implicit in the segmentation. Thus multi-spectral segmentation techniques can be used as the basis for noise filtering. Four MR images (IRTSE, VE(T2), VE(PD), and FLAIR) of a single normal subject were used. The data was registered to an alignment accuracy better than a fraction of a voxel. The four images were analysed using the multi-dimensional partial volume analysis describing six tissue classes; air/bone, skin/muscle, fat, CSF, grey matter and white matter. Noise-free estimates of the original images were then generated from the optimised model, and two analyses were applied. The first was a Monte-Carlo stability analysis [39], in which the relative change in the output image grey-levels produced by the addition of a small amount of noise to the input images was evaluated. For small perturbations of input data the effects on output data can be estimated using a linearised approximation i.e. error propagation. In this limit the noise removal characteristics of an algorithm can be expressed as the proportion of remaining noise. The Monte-Carlo stability analysis therefore estimated the fraction of noise remaining after filtering. The second was a residual outlier measure (ROM), the number of reconstructed grey-level values which are modified by more than three standard deviations of the original image noise, after compensating for local field inhomogeneity. The local image noise was estimated using a technique based upon the distribution of local derivatives, and was thus estimated independently of the technique used in the Monte-Carlo stability analysis. The two analyses were complementary: the Monte-Carlo stability indicated the proportion of noise reduction on the assumption that the model used in the filtering was appropriate, and the ROM indicated the number of voxels for which the model was inappropriate. In order to provide a benchmark for the evaluation, it was also applied to three conventional noise filtering schemes. The first, tangential filtering [39], applies averaging over three pixels (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 noise filtering scheme. Gaussian filtering using a kernel with a standard deviation of 1 pixel and median filtering over the local neighbourhood of 9 pixels were also used.

The third stage of the evaluation was designed to test the absolute accuracy of the segmentation in volumetric estimation tasks. A task of clinical relevance, namely measuring the volume of cerebrospinal fluid (CSF) in the normal brain, was chosen as an exemplar. Rather than attempting to define a gold standard against which a comparison could be performed, a "bronze standard" consisting of CSF volume measurements from the literature was constructed. The CSF in the IRTSE images of the 70 normal volunteers described above was segmented, using both image grey-levels and gradients in the feature space. The volume was measured from CSF probability maps binarised at the 0.5 level. All measurements were then normalised to the total intra-cranial volume (TIV), measured using the same technique, which has been shown [44] to be an effective normalisation for both inter-individual variations in head size and variation in voxel sizes in longitudinal studies. A functional fit to the TIV normalised CSF volumes was then calculated based upon a Weibull cumulative distribution function, as described in [8].

Eight papers published between 1991 and 2002, which quoted TIV and CSF volume measurements from MR

Reference	No. subjects	Definition of measurement space	Segmentation
Gur et. al. 1991 [24]	69	Excludes cerebellum	T2/PD 2D histogram fitting
Blatter et. al. 1995 [5]	89 M 105 F	TIV	ANALYZE [13, 38]
Mueller et. al. 1998 [33]	46	Excludes brainstem	REGION [27]
Coffey et. al. 1999 [14]	122 M 198 F	Excludes slices below midbrain	MedVision
Chan et. al. 2001 [10]	10	TIV	MIDAS [19]
Whitwell et. al. 2001 [44]	55	Excludes slices below cerebellum	MIDAS [19]
Good. et. al. 2001 [21]	265 M 200 F	TIV	SPM99 [1]
Chard et. al. 2002 [11]	13 M 14 F	Excludes slices containing cerebellum	SPM99 [1]

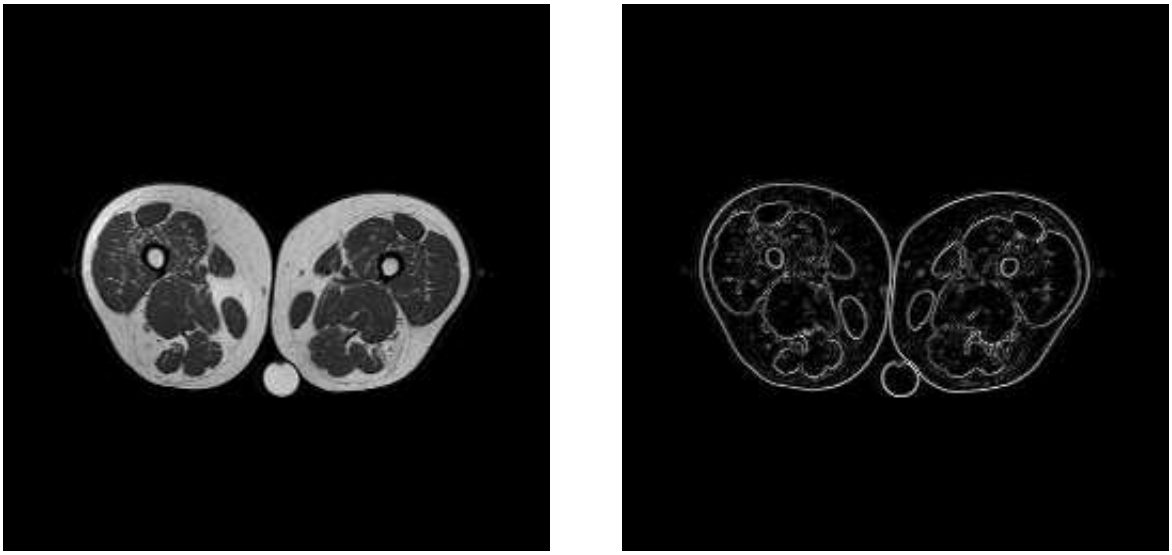
Table 1: Details of the experimental method adopted by studies used in the meta-study, showing the number of subjects included (M=male; F=female; T=total, where the number of each sex was not given), the definition of the measurement space (where TIV is indicated, the whole CSF pool inside the cranium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

images in normal subjects, were then collected. These previous studies used a variety of MR pulse sequences, definitions of the measurement space, and segmentation routines, summarised in Table 1. Most of the studies used 1.5T GE Signa MR scanners, with the exceptions of the present study, [14], where some images were acquired on a 0.35T Toshiba scanner, and [21], in which a 2T Siemens MAGNETOM Vision scanner was used. Of specific interest are the details of the segmentation routines. The in-house software used by [24] combined manual tracing of the skull boundary, followed by manual identification of clusters representing brain and CSF in 2D intensity histograms produced from proton-density (PD) and T2-weighted MR images. Local region growing was then applied to segment these tissues. A similar approach was used in [5], with manual identification of regions of CSF, grey-matter and white matter in the images, followed by k-nearest-neighbour classification of the pixels in 2D histograms produced from T1/T2 intermediate and T2-weighted images. The REGION software used in [33] features a recursive segmentation algorithm, where the user specifies points in bone, CSF, brain (grey and white matter) and high signal intensity areas in white matter. Each tissue type then recursively identified: bone is identified first, allowing skull stripping to leave the intracranial region, then brain and CSF are segmented. A largely manual technique was adopted in [14], with structures identified manually with the aid of printed normal brain and MR atlases, then segmented with a combination of edge detection and manual tracing. The MIDAS software was used to segment T1-weighted images in both [10] and [44]; segmentation was achieved by intensity thresholding at levels set empirically to fractions of the mean brain intensity or intracranial intensity. A fixed threshold of 60% of mean brain intensity for CSF segmentation to improve consistency. The SPM99 software was used in both [21] and [11], with the segmentation performed using cluster analysis on image intensities in T1-weighted MPRAGE and fast spoiled gradient recall (FSPOR) images respectively, plus a-priori knowledge of tissue locations. Intensity non-uniformity correction was also used in both studies.

4 Results

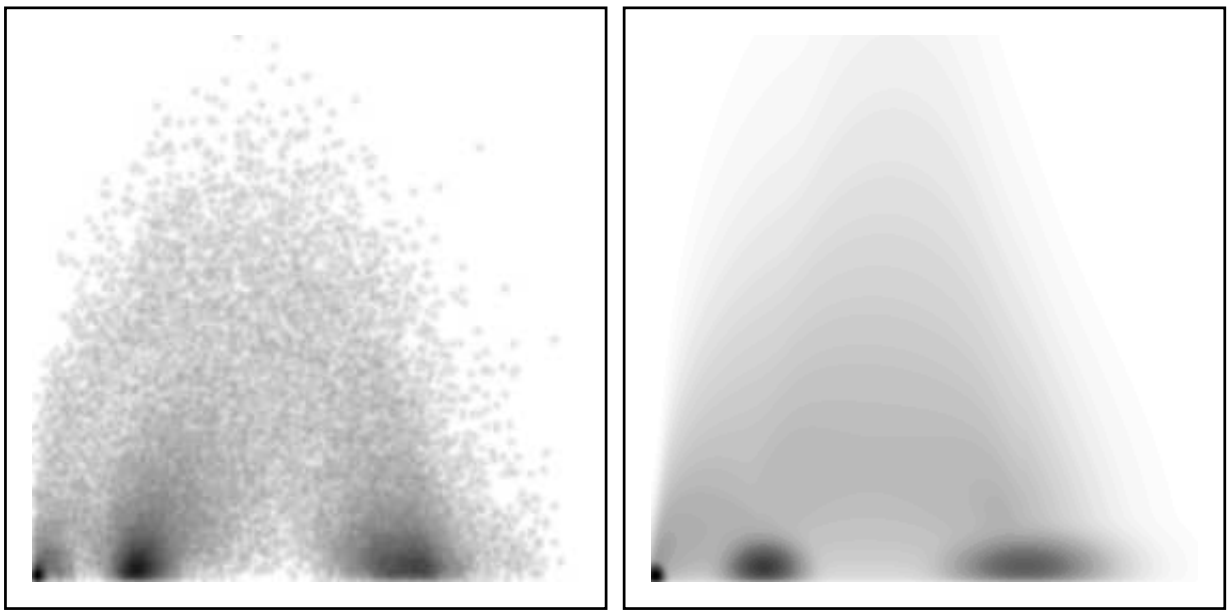
4.1 Single Image and Single Image plus Slope

Results obtained from MR images of a human leg and brain demonstrate qualitatively how the new segmentation method can remove commonly occurring artifacts in partial volume probability maps. Figure 4(a) shows an MR image of the leg. The segmentation algorithm was applied, without the use of gradient terms, for three tissue types: air/bone, muscle and fat. The most probable muscle volume fraction in each voxel computed using 2 is shown in 6(a). The slope image was then calculated 4(b) using 3, and the slope model parameters estimated using for the slope density distribution ρ with $\gamma = 1$ and $\lambda = 0$. A scatter plot of the resulting slope density distribution against grey level and the original data distribution are shown in 5(b) and 5(a). Pure tissue volumes are towards the bottom of these plots and the higher slope values are expected to correspond to partial volume processes. These distributions were compared quantitatively by plotting histogrammes of $s/a(q)$ for pure and partial volume voxel fractions, where each entry is weighted by the $P(n|\mathbf{g}, s)$, in order to confirm the selections for λ and γ. The volumetric analysis for muscle tissue based upon the combined slope and grey level image is shown in 6(b). Comparison of this result to the grey level only result illustrates the reduction in ambiguity between the muscle and edges around bone and skin, which generate grey level values similar to pure muscle.



(a) (b)

Figure 4: MR image of the leg (a) and its associated gradient image (b).

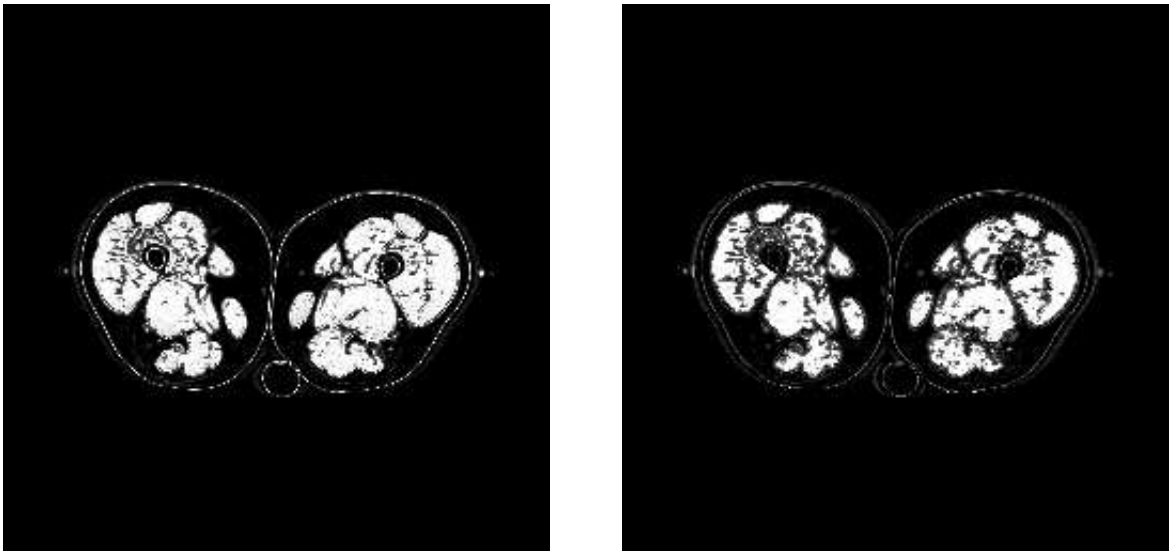


(a) (b)

Figure 5: Scatter plot of gradient vs grey level values for the MR leg image (a) and the optimised model (b).

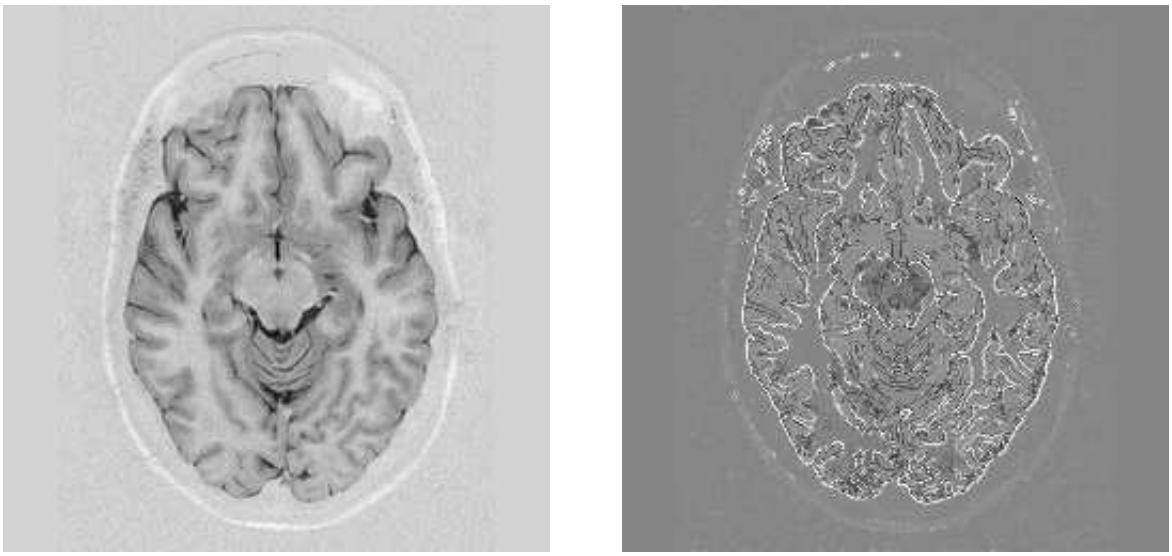
Similar results were obtained from analysis of the IRTSE brain image shown in Fig. 7(a). Segmentations were performed both with and without the inclusion of gradient information in the feature space, and volumetric probability maps of grey matter were produced. These maps were visually indistinguishable. However, the difference between the two 7(b) illustrates a systematic increase in grey matter probability when slope information is included, particularly at the grey matter/white matter boundary. The addition of gradient information reduces ambiguity between pure tissue and partial volume components of the model that have similar grey levels: grey matter is particularly affected as its mean grey level in this image lies between that of the other main tissue types present, white matter and CSF.

A more quantitative evaluation of the improvement in algorithmic performance gained through the use of gradient information, in terms of the stability of the tissue model parameters, is shown in Fig. 8. This shows the standard deviations of the final CSF, grey mater and white matter tissue means for 40 segmentations of single data set



(a) (b)

Figure 6: Tissue segmentation without (a) and with (b) the use of local image slope.



(a) (b)

Figure 7: Normal brain MR image (a) and volumetric differences (b) between segmentations with and without gradient terms.

at each of various levels of initial model randomisation, averaged over ten IRTSE images of the brain. Since the underlying data remained the same across all segmentations, any variation in the results reflects the ability of the segmentation algorithm to converge to a stable result. A clear difference between the three algorithms is demonstrated: the EM-based algorithm without gradient information is roughly an order of magnitude more stable than the straw-man algorithm, and the EM-based algorithm with gradient information is roughly an order of magnitude more stable than that without gradient information.

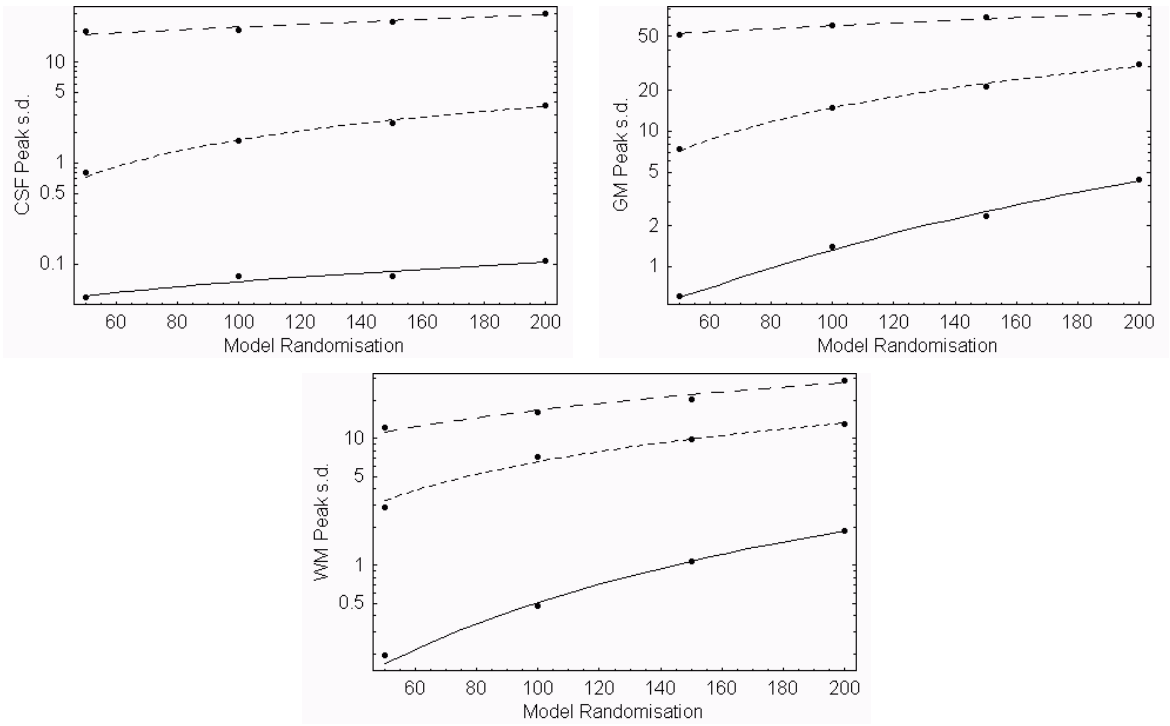


Figure 8: Standard deviations of the CSF (a), grey matter (b), and white matter (c) tissue means for 40 fits to single MR image volumes, plotted against the level of randomisation applies to the initial model. In each case, the upper curve shows the results from the χ^2 fit optimised using simplex, the middle curve the results from the EM-based segmentation without gradient, and the lower curve the results from the EM-based algorithm with gradient. The curves show linear fits to the data (note that the abscissa is plotted on a logarithmic scale).

	LNE	Median Filtering	Gaussian Smoothing	Tangential Smoothing	Multi-Spectral
IRTSE	58.76	0.64 (1559)	0.27 (2405)	0.66 (698)	0.22 (1689)
VE(PD)	64.06	0.66 (1466)	0.26 (3127)	0.68 (530)	0.20 (1804)
VE(T2)	58.2	0.63 (1287)	0.26 (1909)	0.69 (426)	0.17 (938)
FLAIR	52.4	0.63 (1934)	0.27 (3827)	0.69 (966)	0.13 (4971)

Table 2: Monte-Carlo estimate of fraction of remaining noise following filtering and data lying beyond 3σ of original value following filtering (brackets).

4.2 Multiple Image and Multiple Image plus Slope

Figure 9 shows the ROM and Monte-Carlo stability of noise-free images of the normal brain, regenerated from a tissue model produced from four images (IRTSE, VE(T2), VE(PD), and FLAIR). The same results are shown numerically in Table 2. 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 grey-level values that have been changed by more than 3σ, where σ is the original image noise, measured using an independent technique. The ROM should therefore be equal to 1.96% 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 650 voxels. Tangential smoothing, which theoretically should be the least destructive form of noise removal, achieves this limit and thus provides a baseline for the comparison. Excluding the FLAIR image result, the multispectral image reconstruction is on average approximately twice as destructive as tangential smoothing: this indicates that the grey-level model fits the data well for all but around 2% 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 therefore forms an extra tissue class that was not included in the tissue model, leading to misinterpretation of such voxels. In general multi-spectral filtering is no more destructive to image contents than median filtering but removes more image noise than Gaussian smoothing. Therefore, such image reconstruction techniques may be a useful way of processing multiple MR acquisitions.

A final qualitative analysis was applied to the multispectral data in order to demonstrate the effects of including

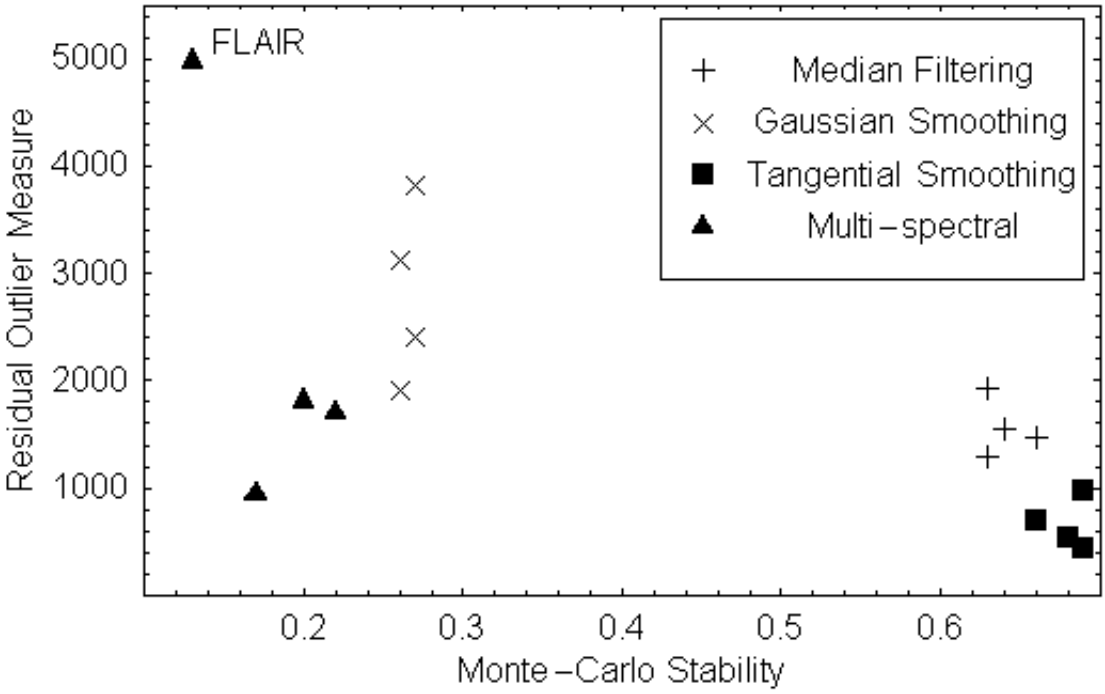
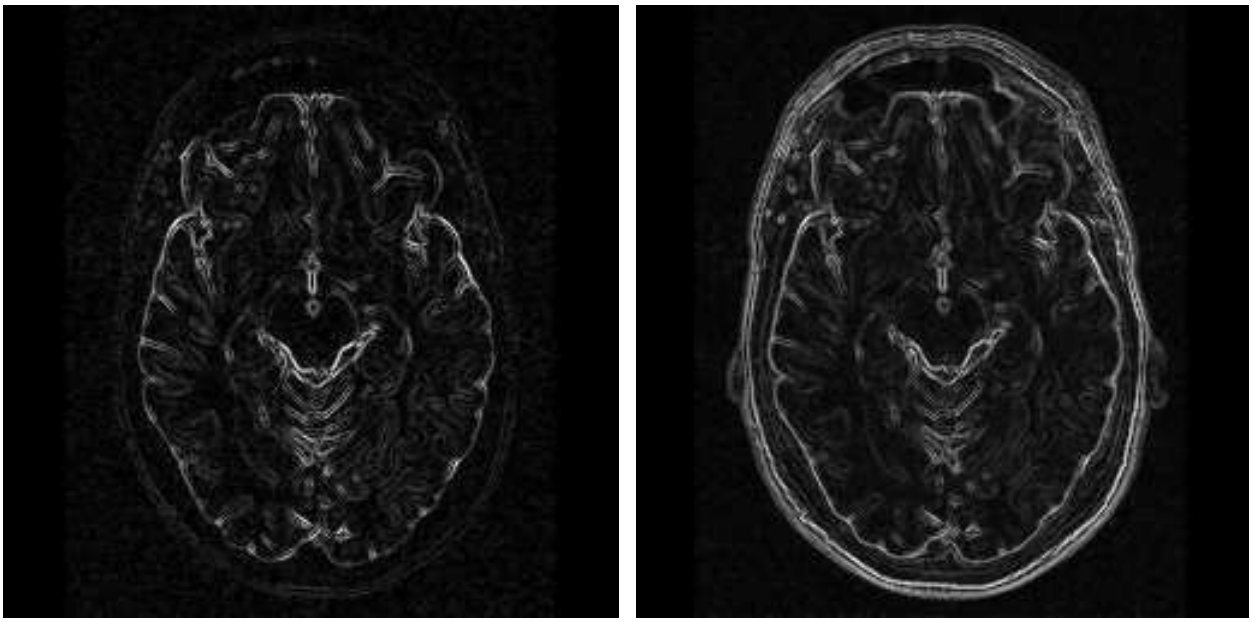


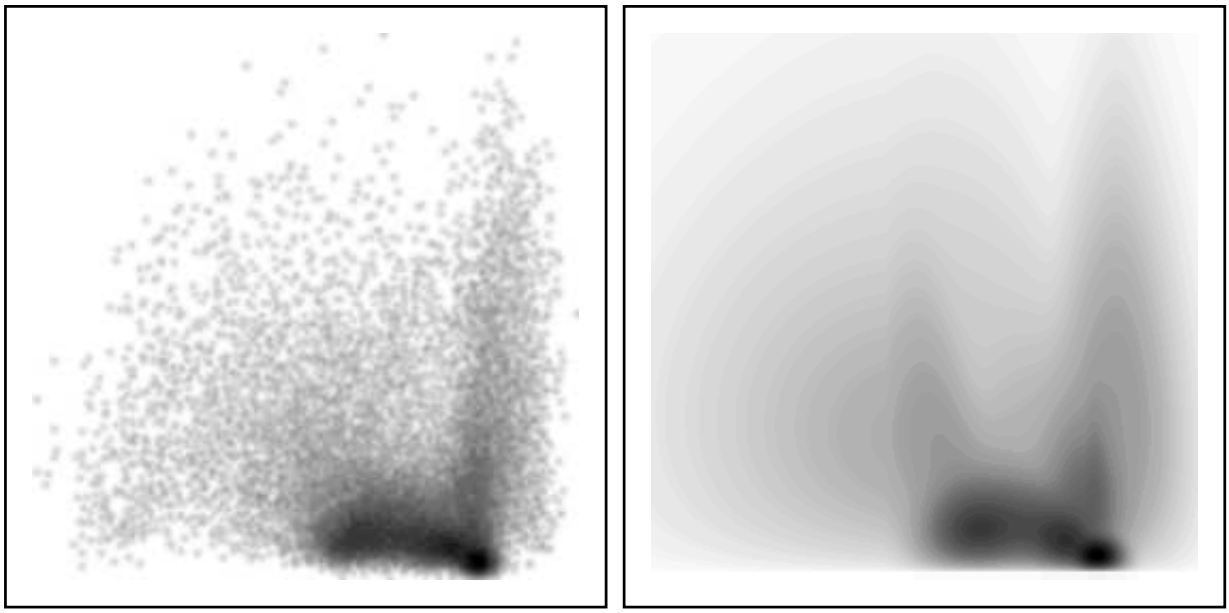
Figure 9: Residual outlier measure vs. Monte-Carlo stability.



(a) (b)

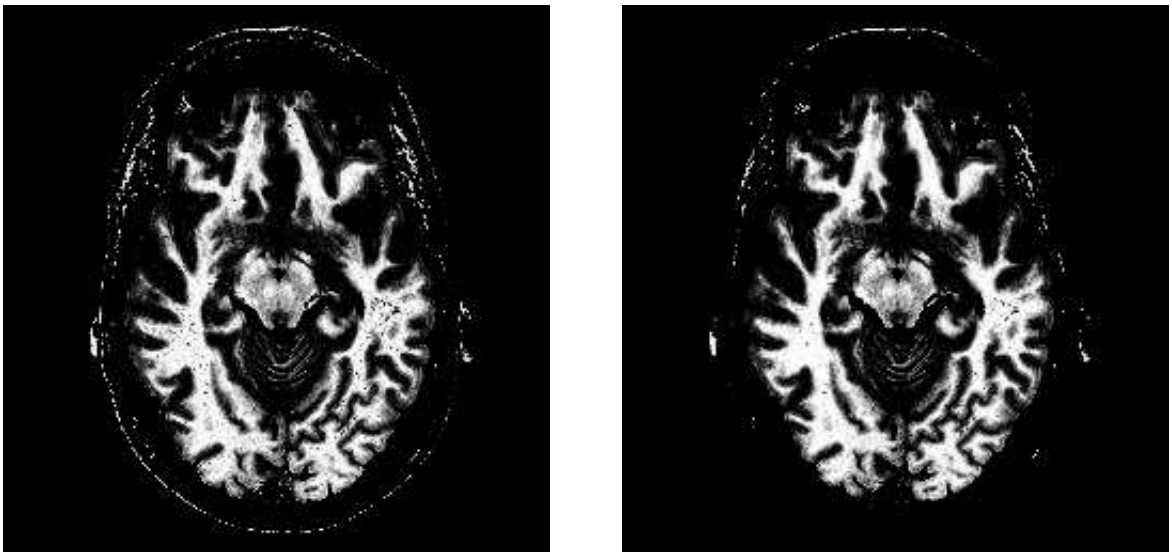
Figure 10: Gradient of single (a) and multiple (b) MR images of the brain.

both multi-dimensional data and gradient terms in the feature space. The gradient data was analysed using a ρ distribution with $\gamma = 2$ and $\lambda = 0.8$. This choice was validated using histogrammes of $s/a(q)$ for partial volume and pure tissue fractions. The distribution of multi-dimensional gradient s vs. the grey level for the image shown in 7(a) is shown for the data 11(a) and the model 11(b). In this case, no observable difference was noticed between the grey matter volume estimates (typically a change of less than 1%). This indicates that the effects of ambiguity have already been reduced by the use of multiple images to the extent that additional slope information is of no great value for grey matter volume estimation. However, the most probably volumetric white matter content of each voxel is shown in 12(b), illustrating improved disambiguation and stability for this tissue.



(a) (b)

Figure 11: Scatter plot of gradient vs grey level values (a) and the fitted density model (b).

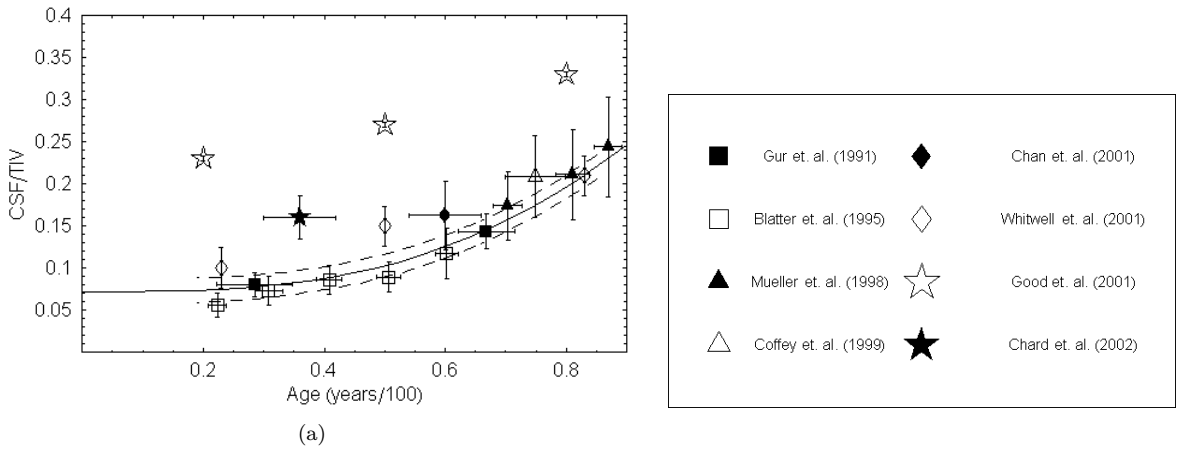


(a) (b)

Figure 12: Tissue segmentation without (a) and with (b) the use of local image slope.

4.3 Meta-study of TIV-normalised CSF Volume Measurements

The meta-study of TIV normalised CSF volume measurements provided an absolute measurement of the accuracy of the segmentation on a typical image analysis task of clinical relevance. The results are shown in Fig.13. Points represent the published data: the curves show the functional fit to the data collected in our own study (i.e. not to the points in the plot). Most of the data show a remarkable level of consistency with each other and with the functional fit, considering the variations in the experimental procedures. Indirectly, this implies that across these studies as a whole, the variations in volume measurements introduced by the different experimental procedures are comparable to or smaller than the inherent biological variability in total intracranial and CSF volumes. Some outliers are present, representing the results published in [21] and [11], both of which used SPM99 to perform the segmentation. Significantly higher CSF volumes were quoted by these authors than are supported by the rest of the literature.



(a)

Figure 13: Previous published TIV normalised CSF volume measurements. The solid curves show the functional fit to the data presented in this paper: the dashed curves represent the 1σ error bounds. The points show data from the literature. Points with error bars in both the x and y directions represent data published numerically, whilst points with error bars only in the y direction represent data read from graphs.

5 Conclusions

In previous work we have extended standard volumetric estimation techniques to multiple images. In this work a method to utilise not only the grey level information but also the local grey level slope information in a MR image was proposed. The method can be considered as an alternative to both the assumption of local regional smoothness or local resampling of the prior probabilities, which have previously been suggested by other authors [28]. Unlike these methods, local information is used directly in a manner which is quantitatively related to the image formation process. The algorithm also demonstrates the extension of the EM algorithm to fitting non-Gaussian models.

The inclusion of gradient information in the feature space disambiguates partial volume voxels from pure tissue voxels that have similar grey levels. It was shown to both significantly reduce artefacts arising through this ambiguity, and to consequently improve the stability of tissue model parameter estimates. The use of the multispectral segmentation as a noise-filtering technique was also demonstrated, indirectly confirming the accuracy of the partial volume tissue model. The absolute accuracy of the segmentation result was demonstrated through comparison with a clinically relevant bronze standard, consisting of published measurements of TIV normalised CSF volumes. This meta analysis shows that a partial volume based segmentation technique gives results which are consistent with a wide variety of manual and semi-automatic approaches. Such approaches inevitably apply a high degree of quality control and consistency checking. However, the magnitude of the disagreement with other literature, particularly in the case of the fully automatic techniques [21], is such that we are forced to conclude that these results may represent methodological flaws. This may be due to failings of a pure tissue (non-partial volume Gaussian) model, rather than accurate measurements of unexpectedly high CSF volumes. Finally, the effects of including gradient terms in the analysis of multispectral data were evaluated. The artefact reductions seen with single dimensional data were still present but, as might be expected in this data-rich scenario, were smaller.

6 Acknowledgements

Thanks are due to Prof. Alan Jackson, Maja Pokric and Marietta Scott, for their support and contributions to early stages of the work. Much of the brain MR data was made available by Pat Rabbit and supervision of data acquisition was overseen by Charles Hutchinson. 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 Medical Research Council Grant No. D2025/31. The software and associated technical notes are available from our web site www.tina-vision.net.

References

[1] J Ashburner and K Friston. Multimodal image coregistration and partitioning-a unified framework. *Neuroimage*, 6(3):209–217, 1997.

[2] J C Bezdek, L O Hall, and L P Clarke. Review of mr image segmentation techniques using pattern recognition. *Med. Phys.*, 20(4):1033–104, 1993.

[3] C M Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, 1995.

[4] A Blake and M Isard. *Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion*. Springer Verlag, Berlin and Heidelberg, 1998.

[5] 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. *American Journal of Neuroradiology*, 16:241–251, 1995.

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

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

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

[9] 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 Comput.*, 21(10):851–864, 2003.

[10] 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.

[11] 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. *JMRI*, 15:259–267, 2002.

[12] J Chiverton, K Wells, and D Johnson. A partial volume 3-d gradient magnitude model. In *Proc. MIUA*, pages 49–53, London, 2004.

[13] H E Cline, W E Lorensen, R Kinikis, and F Jolesz. Three-dimensional segmentation of mr images of the head using probability and connectivity. *Journal of Computer Assisted Tomography*, 14:1037–1045, 1990.

[14] 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 reverse hypothesis. *Neurology*, 53:189–196, 1999.

[15] G B Coleman and H C Andrews. Image segmentation by clustering. *Proc. IEEE*, 5:773–785, 1979.

[16] D L Collins, C J Holmes, and A C Evans. Automatic 3d model-based neuro-anatomical segmentation. *Human Brain Mapping*, 3(3):190–208, 1995.

[17] C Davatzikos and R N Bryan. Using a deformable surface model to obtain a shape representation of the cortex. *IEEE Trans. Med. Imag.*, 15:785–795, 1996.

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

[19] P A Freeborough, N C Fox, and R I Kitney. Interactive algorithms for the segmentation and quantitation of 3-d mri brain scans. *Comput. Methods Prog. Biomed*, 53:15–25, 1997.

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

[21] 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.

[22] H Gudjbartson and S Patz. The rician distribution of noisy mri data. *Magnetic Resonance in Medicine*, 34(6):910–914, 1995.

[23] R Guillemaud and J M Brady. Estimating the bias field of mr images. *IEEE Trans. Med. Imag.*, 16(3):238–251, 1997.

[24] 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.

[25] R M Haralick and L G Shapiro. Image segmentation techniques. *Comput. Vis. Graph. Image Proc.*, 29:100–132, 1985.

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

[27] J A Kaye, T Swihart, D B Howieson, A Dame, M M Moore, T Karnos, R M Camicioli, M Ball, B Oken, and G Sexton. Volume loss of the hippocampus and temporal lobe in healthy elderly persons destined to develop dementia. *Neurology*, 48:1297–1304, 1997.

[28] 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. Imag.*, 17(1):74–86, 1998.

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

[30] T Lei and W Sewchand. Statistical approach to x-ray ct imaging and its applications in image analysis ii: A new stochastic model-based image segmentation technique for x-ray ct image. *IEEE Trans. Med. Imag.*, 11(1):62=69, 1992.

[31] I N Manousakas, P E Undrill, G G Cameron, and T W Redpath. Split-and-merge segmentation of magnetic resonance medical images: performance evaluation and extension to three dimensions. *Comput. Biomed. Res.*, 31:393–412, 1998.

[32] T McInerney and D Terzopoulos. Medical image segmentation using topologically adaptable surfaces. *Lect. Notes Comput. Sci.*, 1205:23–32, 1997.

[33] 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.

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

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

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

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

[38] R A Robb. A software system for interactive and quantitative analysis for biomedical images. In K H Hohne, H Fuchs, and D M Pizer, editors, *3D imaging in medicine*, volume F60, pages 333–361. NATO ASI Series, 1990.

[39] 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.

[40] E A Vokurka, A Herwadkar, N A Thacker, R T Ramsden, and A Jackson. Using bayesian tissue classification to improve the accuracy of vestibular schwannoma volume and growth measurement. *AJNR*, 23:459–467, 2002.

[41] E A Vokurka, N A Thacker, and A Jackson. A fast model independant method for automatic correction of intensity non-uniformity in mri data. *JMRI*, 10(4):550–562, 1999.

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

[43] W M Wells, E L Grimson, and R Kikinis. Adaptive segmentation of mri data. *IEEE Trans. on Medical Imaging*, 15(4):429–442, 1996.

[44] 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. *American Journal of Neuroradiology*, 22:1483–1489, 2001.

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

[46] C Xu, D L Pham, J L Prince, M E Etemad, and D Yu. Reconstruction of the central layer of the human cerebral cortex from mr images. In *Proc. Int. Conf. Med. Image Comput. Comp. Assist. Interv., 1st*, pages 482–488, Cambridge, MA, 1998.

Appendix A

Consider the process of estimating the parameters for a distribution that models a sample of data measured as a set of n_i frequencies. The aim is to define the probability of observing a set of sample data $\mathbf{x}_i$ in a multi-dimensional (continuum) space based upon a specific probability density function $\lambda(\mathbf{x}_i)$. To do this we must define a finite set of sample cells ($\mathbf{X}_i$) and compute the probability of observing data within each and take the continuum limit. The correct statistical model for this is the Poisson distribution, so that the probability of observing a given number of samples within each discrete region within the non-zero portion of the probability density is

$$P(n_i, \lambda) \;=\; \frac{exp(-\lambda)\lambda_i^n}{n!}$$

The log probability for the model in terms of the data as the sum of independent terms for each quantity of observation is therefore $n_i = 0$, $n_i = 1$, $n_i = 2$...;

$$log(P(\lambda)) \;=\; \sum_i^{N_0} log\ P(0, \lambda(\mathbf{X_i})) \;+\; \sum_i^{N_1} log\ P(1, \lambda(\mathbf{X_i})) \;+\; \sum_i^{N_2} log\ P(2, \lambda(\mathbf{X_i})) \;+\; ...$$

The first term describes N_0 empty cells, the second N_1 cells containing one sample, the third N_2 cells containing two samples etc. This can be rewritten as

$$log(P(\lambda)) \;=\; \sum_i^{N_0+N_1+N_2+...} log\ P(0, \lambda(\mathbf{X_i})) \;+$$

$$\sum_i^{N_1} log\ P(1, \lambda(\mathbf{X_i})) - log\ P(0, \lambda(\mathbf{X_i})) \;+\; \sum_i^{N_2} log\ P(2, \lambda(\mathbf{X_i})) - log\ P(0, \lambda(\mathbf{X_i})) \;+\; ...$$

Substituting the appropriate Poisson terms gives

$$log(P(\lambda)) \;=\; \sum_i^{N_1} log(\lambda(\mathbf{X}_i)\ exp(-\lambda(\mathbf{X}_i))/1!) + (\lambda(\mathbf{X}_i))$$

$$+\; \sum_i^{N_2} log(\lambda^2(\mathbf{X})\ exp(-\lambda(\mathbf{X}_i))/2!) + (\lambda(\mathbf{X}_i)) \;+\; ... \;-\; \sum_i^{N_0+N_1+N_2+...} \lambda(\mathbf{X}_i)$$

which simplifies to

$$=\; \sum_i^{N_1+N_2+...} n_i log(\lambda(\mathbf{X}_i)) \;-\; \sum_i^{N_0+N_1+N_2+...} \lambda(\mathbf{X}_i) \;+\; k_1$$

where k_1 is a constant[2] defined by the sample of data and n_i is the quantity of data at $\mathbf{X}_i$. The second term, which corresponds to the probability of a zero entry in every location, is simply the integrated probability density, which can be made constant as part of the model definition: only the first term is data dependent. Therefore

$$log(P(\lambda)) \;=\; \sum_i^{N} n_i log(\lambda(\mathbf{X}_i)) \;-\; v \;+\; k_1$$

[2]Provided that the sampling process is not respecified, such as happens in some bootstrap likelihood techniques [6, 7].

where v is the integral of the density function and N is the number of non zero cells. Summing now over the M individual data entries j, the log probability is given by

$$log(P(\lambda)) \;=\; \sum_{j}^{M} log(\lambda(\mathbf{X}_j)) \;-\; v \;+\; k_1$$

This approach is also directly applicable to the continuous distributions $(\mathbf{x}_j)$

$$log\ L \;=\; \sum_{j}^{M} log(\lambda(\mathbf{x}_j)) \;-\; k_2 \int \lambda(\mathbf{x})\ d\mathbf{x} \tag{7}$$

where k_2 is an unknown constant which relates probability density $\lambda(\mathbf{x}_j)$ to discrete probability $\lambda(\mathbf{X}_j)$[3]. This measure, first described by Fermi, is known as extended maximum likelihood and forms the statistical basis of the EM algorithm. It extends standard maximum likelihood to situations where parameters of the density distributions, particularly the normalisation, are unknown. In the context of the use of an EM optimisation we can therefore estimate a new set of parameters by minimising the first term subject to the constraint of fixed normalisation of the density function, so that the arbitrary constant k_2 plays no role in the solution.

It is well known that the likelihood estimate of the mean of a Gaussian distribution is given by the mean of the sample. However, there is a more general solution which can be derived from the above likelihood. Differentiating $log\ L$ with respect to x_c (the translation parameter we get

$$\sum_{j} \frac{\partial \lambda(\mathbf{x}_j)}{\partial \mathbf{x}_c} / \lambda(\mathbf{x}_j) \;=\; 0 \qquad -(4)$$

We can then ask which class of definitions of $\mathbf{x}_c$ are consistent with solution of this expression corresponding to the mean, i.e.

$$k_3 \sum_{j} (\mathbf{x}_j \;-\; \mathbf{x}_c) \;=\; 0$$

where k_3 is a constant. This would imply

$$k_3(\mathbf{x} \;-\; \mathbf{x}_c) \;=\; \frac{\partial \lambda(\mathbf{x})}{\partial \mathbf{x}_c} / \lambda(\mathbf{x})$$

rewriting and integrating both sides with respect to $\mathbf{x}$ this gives

$$\int k_3(\mathbf{x} \;-\; \mathbf{x}_c)\lambda(\mathbf{x})\ d\mathbf{x} \;=\; \int \frac{\partial \lambda(\mathbf{x})}{\partial \mathbf{x}_c}\ d\mathbf{x}$$

The right hand term is zero for an integrable function, so the mean of a set of data $\mathbf{x}_j$ is consistent with a likelihood estimate of $\mathbf{x}_c$ when defined as the centroid of the density function

$$\mathbf{x}_c \;=\; \frac{\int \mathbf{x}\lambda(\mathbf{x})\ d\mathbf{x}}{\int \lambda(\mathbf{x})\ d\mathbf{x}}$$

for which of course the standard Gaussian result is simply a special case. This result can be trivially extended to higher order moments by changing the assumed parameter estimation process and repeating the other steps of the proof. Therefore, EM mixture modelling can be applied to distributions other than just Gaussians, provided that the location and shapes of the mixture density functions are parameterised in terms of centroids and higher moments.

[3]Strictly k_2 is the interval on $\mathbf{x}$ which specifies an independent sample and in order for it to take the form of a simple constant (and not be a function of $\mathbf{x}$, $k_2(\mathbf{x}_i)$), we must assume that this interval is the same everywhere within the measurement space i.e. we are in an equal variance domain.