

# Adaptive Partial Volume Classification of MRI Data

John P. Chiverton<sup>1</sup> and Kevin Wells<sup>2</sup>

<sup>1</sup>Department of Computer Science, University of Bristol, Woodland Road, Bristol, BS8 1UB, U.K.

<sup>2</sup>Centre for Vision, Speech and Signal Processing, Faculty of Engineering and Physical Sciences, University of Surrey, GU2 7XH, U.K.

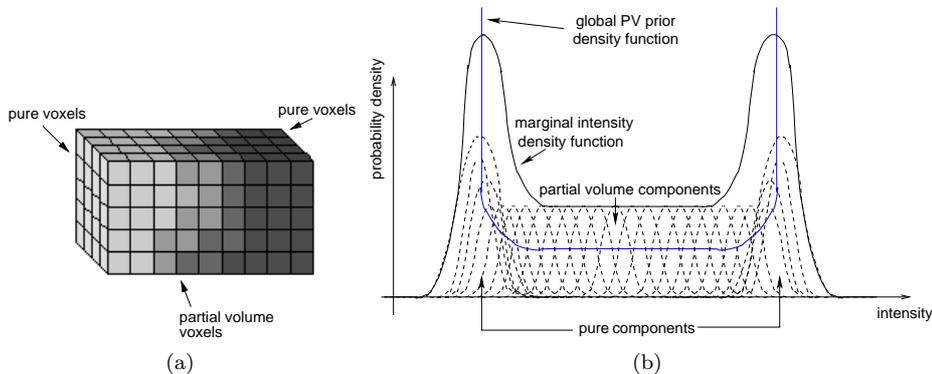
E-mail: <sup>1</sup>jpchiverton@theiet.org <sup>2</sup>k.wells@surrey.ac.uk

**Abstract.** Tomographic biomedical images are commonly affected by an imaging artifact known as the Partial Volume (PV) effect. The PV effect produces voxels composed of a mixture of tissues in anatomical Magnetic Resonance Imaging (MRI) data resulting in a continuity of these tissue classes. Anatomical MRI data typically consists of a number of contiguous regions of tissues or even contiguous regions of PV voxels. Furthermore discontinuities exist between the boundaries of these contiguous image regions. The work presented here probabilistically models the PV effect using spatial regularisation in the form of continuous Markov Random Fields (MRF)s to classify anatomical MRI brain data, simulated and real. A unique approach is used to adaptively control the amount of spatial regularisation imposed by the MRF. Spatially derived image gradient magnitude is used to identify the discontinuities between image regions of contiguous tissue voxels and PV voxels, imposing variable amounts of regularisation determined by simulation. Markov Chain Monte Carlo (MCMC) is used to simulate the posterior distribution of the probabilistic image model. Promising quantitative results are presented for PV classification of simulated and real MRI data of the human brain.

## 1. Introduction

The Partial Volume (PV) effect, (illustrated in figure 1) is an imaging artifact associated with many biomedical imaging modalities, including Magnetic Resonance Imaging (MRI). The accurate identification of regional voxel tissue content such as Grey Matter (GM) and White Matter (WM), is an important step for many clinical applications. Unfortunately the PV effect fundamentally limits the accuracy of any quantitative estimate of these tissues. Bayesian probabilistic modelling approaches can be used to model the PV effect. These models, particularly with the use of Markov Chain Monte Carlo (MCMC) simulation approaches can adapt to the data at hand.

Choi et al. proposed a probabilistic multi-channel PV image model, (Choi et al. 1991). Context was modelled via a continuous but homogeneous Markov Random Field (MRF), where constant spatial regularisation was assumed. This can be compared with the work of Wang et al. which is of particular interest here due to the use of an inhomogeneous MRF, (Wang et al. 2001). The amount of spatial regularisation was controlled with a locally defined entropy measure. Boundary regions were found to possess larger values of entropy in relation to non-boundary regions. Wang et al. described the pure and PV classes as distinct discrete components and also utilised a discrete MRF. However the PV effect results in voxels that can be considered



**Figure 1.** Illustration of the partial volume effect: (a) at the boundary between two pure components; and (b) a two tissue density function.

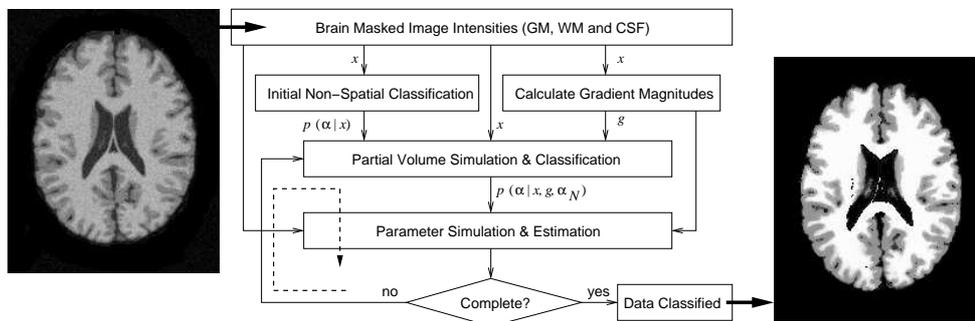
to be a mixture of different components, such as tissues. These mixed voxels therefore possess a signal that is continuous in nature which can be modelled probabilistically, as previously done by (Choi et al. 1991).

Woolrich et al. utilised continuous representations of the PV variables in combination with a continuous homogeneous MRF, where the amount of spatial regularisation was modelled via a prior density, (Woolrich et al. 2005). Unfortunately the prior density modelled the regularisation over the entire image data rather than on a local basis. Despite this, the Bayesian approach taken by Woolrich et al. enabled the effective amount of regularisation to adapt through out the imaging data. The continuous form of the PV effect is a result of the action of the Point Spread Function (PSF) during image acquisition. Van Leemput et al. overcame the complexity of modelling the PV effect by modelling the image data before and after the action of the PSF, (Leemput et al. 2003). This simplified the derivation of the Expectation-Maximisation (EM) optimisation that was used to estimate voxel labels and parameter values, but Van Leemput et al. did not incorporate a locally adaptive MRF.

Many authors have investigated PV image models that do not utilise a MRF, e.g. (Santago & Gage 1995, Vokurka et al. 2002). Some have included second order image information such as gradient magnitude that can aid identification of PV voxels in boundary regions, (Williamson et al. 2002, Chiverton & Wells 2004). These PV image models are often computationally efficient and relatively easy to implement but classification performance is limited at low image Contrast to Noise Ratios (CNR)s.

Pham in 2003 investigated gradient magnitude in fuzzy ‘C’ means clustering type algorithms to regulate the spatial continuity of PV voxel labels, (Pham 2003). Alternative work was presented in (Pham & Bazin 2004) which utilised spatial image intensity gradient magnitude. This model encouraged the classification of PV voxels along the boundary of pure tissue regions. However problems may occur for PV voxels arising from anatomical variability that are often found in MRI data.

The work presented here provides a novel methodology, utilising a locally adaptive Gradient-controlled Spatial Regulariser (GSR) in a Bayesian formulation for the quantification of classes in the presence of the PV effect. An MRF is used to model the underlying continuous PV voxel class membership with variable spatial regularisation, dependent upon spatially derived image gradient magnitude. Furthermore MCMC simulation of the posterior distribution enables the model to adapt to the imaging



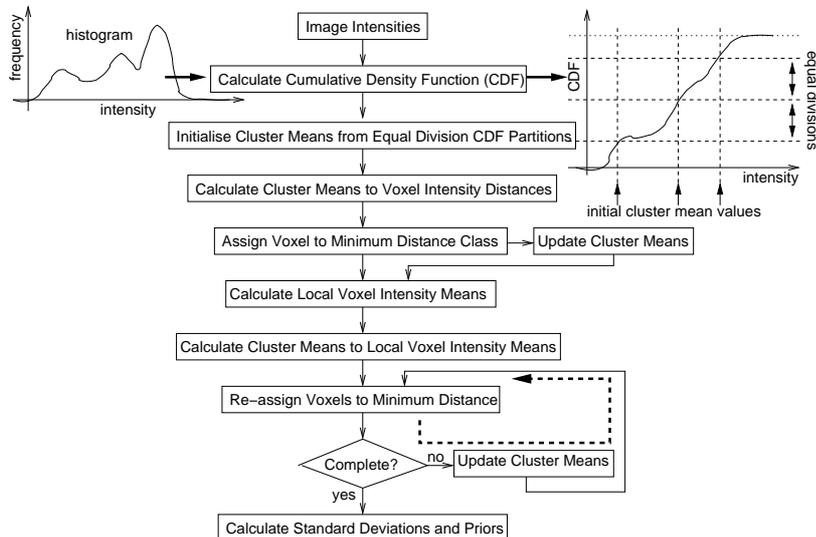
**Figure 2.** Illustration of the steps involved in the PV classification process.

data. Thus adaptive control of the amount of information drawn from neighbouring regions is provided, rather than attempting to impose strong prior constraints on particular anatomical configurations. This enables the classifier to adaptively identify contiguous regions of pure voxels and PV voxels and the boundaries (or discontinuities) between these contiguous regions purely from the image data.

## 2. Methodology

The algorithm proposed and evaluated here to classify PV affected imaging data is illustrated in figure 2. The first stage in the PV classification process is the initial non-spatial (histogram based) classification. This requires initial parameter estimates which are provided here via a clustering algorithm, the steps of which are illustrated in figure 3. The clustering algorithm assigns voxels to three individual classes corresponding to approximate discrete GM, WM and Cerebro-Spinal Fluid (CSF) classes. Initial cluster centres (means) are estimated by separating the vertical axis of the image data Cumulative Density Function (CDF) into a set of equal width partitions and then projecting the centres of these partitions onto the horizontal intensity axis. These initial means are then used to assign voxels to a particular class using a minimum distance rule. The means are iteratively updated and then the procedure is repeated using local voxel mean values, in place of the voxel intensities in the minimum distance rule. These local mean values are calculated from the neighbourhood of each voxel intensity. The cluster means are then again updated using the original voxel intensities. This latter stage reduces the effect of noise and is repeated until no voxel is re-assigned. Initial parameter estimates for the means, standard deviations and priors are then calculated from the intensities based on the cluster membership for each voxel.

After initialisation, the PV classifier utilises the initial parameter estimates to estimate the PV content of each voxel. Unfortunately it is unlikely that any algorithm or even human expert is able to accurately estimate the amount of each tissue in each voxel. This is because the image is affected by different sources of noise (in common with most real world signals). We can however make observations and assumptions regarding the image data. Firstly we can observe that the image data has intensities associated with each voxel. We can also observe that the finite spatial resolution of the image data results in individual voxels potentially corresponding to a mixture of tissues. This observation is intrinsic to the work presented here and is formally defined



**Figure 3.** Illustration of the steps involved in the clustering algorithm used to initialise the PV classification process.

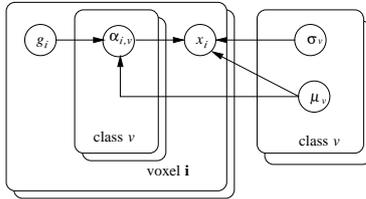
in section 2.1. A further observation can be made regarding the spatial nature of the image data. We would expect neighbouring voxels to usually possess similar quantities of tissues. However, this may not always be true, particular at the boundary between tissues. Therefore we also consider image gradient magnitude to help us identify boundaries between regions and therefore eventually enable us to control the amount of information drawn from neighbouring voxels used for regularisation via a MRF.

We can also make assumptions regarding the statistical properties of the image data such as the form of the noise. The image noise can be statistically described by Probability Density Functions (PDFs) which can also be used to directly model the image data. Our modelling approach is described in detail in section 2.2. We have sought to include the best possible observations into our model to enable us to accurately describe the probabilistic properties of the image data. As mentioned earlier, it is unlikely that an algorithm can provide a single accurate estimate of the content of every voxel. We therefore model the range of tissue content that each voxel is likely to contain via simulation of the PDF of the PV voxel content. The simulation process is discussed in section 2.3.

The initial parameter estimates provided by the clustering algorithm in figure 3 are then updated (see figure 2). The parameters of the system are also simulated here, where initial and preceding parameter estimates are used to initialise the simulation process. The statistical properties of the parameters are discussed in section 2.4. Convergence of the algorithm can be assessed by analysing the asymptotic parameter values. In actuality, we allow the algorithm to complete after a fixed number of iterations. This simplifies experimental assessment.

### 2.1. Image Model

Consider a volumetric data set,  $\Omega$ , consisting of individual voxels,  $\mathbf{i}$ , i.e.  $\Omega = \{\mathbf{i} = (w_1 \ w_2 \ w_3)^T\}$  where  $w_1, w_2, w_3$  are coordinates within the data volume.



**Figure 4.** Illustration of the chain of dependence between the image intensities,  $x$  and the gradient magnitudes,  $g$  via the PV mixtures,  $\alpha$ .

The volumetric data set also consists of noise affected intensities for each voxel,  $X = \{x_{\mathbf{i}} | \mathbf{i} \in \Omega\}$ . Every voxel in the imaging data is also associated with a PV mixture vector,  $\alpha_{\mathbf{i}}$ , which forms part of  $A = \{\alpha_{\mathbf{i}} | \mathbf{i} \in \Omega\}$ , where each element in the PV mixture vector,  $\alpha_{\mathbf{i},v}$  is indicative of the proportion of a particular tissue,  $v$  for voxel  $\mathbf{i}$ . The PV mixture vector for a particular voxel instance,  $\mathbf{i}$ , is therefore given by  $\alpha_{\mathbf{i}} = (\alpha_{\mathbf{i},1} \ \alpha_{\mathbf{i},2} \ \dots \ \alpha_{\mathbf{i},n})^T$ , where  $n$  is the number of tissues associated with this particular volumetric imaging data. A voxel,  $\mathbf{i}$  entirely composed of classification class  $v$  is described with  $\alpha_{\mathbf{i},v} = 1$ , similarly, the same voxel entirely devoid of classification class  $v$  is described with  $\alpha_{\mathbf{i},v} = 0$ . The PV mixture variables are thus subject to the constraints  $\sum_{v=1}^n \alpha_{\mathbf{i},v} = 1$  and  $0 \leq \alpha_{\mathbf{i},v} \leq 1$ .

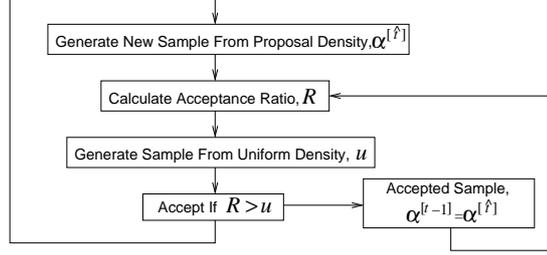
Also associated with each classification class is a vector of class specific parameters,  $\eta = (\eta_1 \ \eta_2 \ \dots \ \eta_m)^T$ , the form of which is dependent upon the choice of noise model, defined shortly. Each voxel is also considered to possess a gradient magnitude,  $g_{\mathbf{i}}$  indicative of the underlying changes in voxel intensity,  $G = \{g_{\mathbf{i}} | \mathbf{i} \in \Omega\}$ . These gradient magnitude variables provide a means for controlling the amount of spatial regularisation imposed by the PV classifier.  $G$  is estimated from the imaging data via convolution of appropriately designed gradient masks, such as the 2-D Sobel gradient masks extended to 3-D, see e.g. (Nikolaidis & Pitas 2001). The three 3-D gradient masks are specified for the three orthogonal dimensions of the image data which, when combined with each set of weight specifications, results in significant noise reduction. This reduces the effect of image noise on the resultant gradient magnitude estimates which act as proxies to the local underlying tissue variation.

## 2.2. Probabilistic Description of the PV Mixtures

Probabilistic models enable inter-dependencies between variables to be considered. The intensities  $X$  are considered here to be dependent on the PV mixtures  $A$ . Similarly the PV mixtures are considered here to be dependent on the gradient magnitudes  $G$ . This chain of dependence is modelled here using conditional independence so that the dependence of the gradient magnitudes on the intensities is incorporated through the PV mixtures, (illustrated in figure 4). Thus, the posterior density for  $A$  can be shown to be given by

$$p(A|X, G, \eta) = \frac{p(X|A, \eta)p(A|G, \eta)}{p(X|G, \eta)}. \quad (1)$$

The intensity distributions of voxels entirely composed of a single classification class, i.e.  $\alpha_{\mathbf{i},v} = 1$  are considered here to follow Gaussian distributions, valid for large enough Signal to Noise Ratios (SNRs), (Gudbjartsson & Patz 1995). Therefore the



**Figure 5.** Illustration of the steps involved in an Independent Metropolis Hastings (IMH) sampling process.

classification class parameters,  $\eta$  are the class mean vectors and standard deviation vectors, i.e.  $\eta \triangleq \mu, \sigma$ , where  $\mu = (\mu_1 \mu_2 \dots \mu_n)^T$  and  $\sigma = (\sigma_1 \sigma_2 \dots \sigma_n)^T$ .

For PV voxels, i.e.  $\alpha_{i,v} \in (0, 1)$  a mixture of the individual pure class densities has to be considered. An assumption of linear mixing, (Windham et al. 1988) for the PV components results in Gaussian distributed PV intensities with means  $\mu_{\alpha_i} = \sum_{v=1}^n \mu_v \alpha_{i,v}$ , and variances  $\sigma_{\alpha_i}^2 = \sum_{v=1}^n (\sigma_v \alpha_{i,v})^2$ .  $A$  is modelled here with a MRF (see Appendix A) and if we assume independently distributed image intensities then

$$p(A|X, G, \mu, \sigma) \propto \prod_{\forall i} p(x_i | \alpha_i, \mu, \sigma) p(\alpha_i | g_i, \alpha_{N_i}, \mu, \sigma). \quad (2)$$

where  $N_i = \{\mathbf{j} | \mathbf{j}$  are neighbours of voxel  $\mathbf{i}\}$ . MCMC is used here to simulate equation (2), which uses the ratio of posterior distributions. This removes the requirement to specify the form of the difficult to realise denominator in equation (1).

The posterior density,  $p(A|X, G, \mu, \sigma)$  describes volumetric PV imaging data with a MRF modelled in the form of a Gaussian density that is dependent on the gradient magnitudes (see Appendix A). This provides an approach to automatically apply variable amounts of spatial regularisation dependent on the local gradient magnitude for each voxel. This will be shortly seen (in section 3.1) to be a new way of spatially adapting and controlling the amount of spatial regularisation.

### 2.3. Markov Chain Monte Carlo Sampling

The posterior density is simulated via a Markov Chain Monte Carlo (MCMC) sampling process known as the Independent Metropolis-Hastings (IMH) sampler, illustrated in figure 5. MCMC approaches utilise the ratio of the posterior densities to determine whether a randomly generated sample value is to be accepted.. The IMH ratio is considered here on a voxel by voxel basis using a Gibbs sampler approach, hence (dropping  $\mathbf{i}$ ),

$$\mathcal{R} = \mathcal{R} \left( \alpha^{[\hat{t}]} \mid \alpha^{[t-1]} \right) = \frac{p \left( \alpha^{[\hat{t}]} \mid x, g, \alpha_N, \mu, \sigma \right) \lambda \left( \alpha^{[t-1]} \right)}{p \left( \alpha^{[t-1]} \mid x, g, \alpha_N, \mu, \sigma \right) \lambda \left( \alpha^{[\hat{t}]} \right)} \quad (3)$$

where  $\lambda()$  is the proposal density and efficient sampling schemes are used to generate random samples from this density.  $\alpha^{[t-1]}$  and  $\alpha^{[\hat{t}]}$  are the previously accepted PV mixture vector sample and the current proposal PV mixture vector sample respectively. The new PV mixture vector sample,  $\alpha^{[\hat{t}]}$  is accepted with probability

$\min\{\mathcal{R}, 1\}$ . The posterior density can be expanded using Bayes and conditional independence, and if  $\lambda() \cong p(\boldsymbol{\alpha})$  then (see Appendix B)

$$\mathcal{R} = \frac{p(x | \boldsymbol{\alpha}^{[t]}, \boldsymbol{\mu}, \boldsymbol{\sigma}) p(\boldsymbol{\alpha}^{[t]} | g, \boldsymbol{\alpha}_N, \boldsymbol{\mu})}{p(x | \boldsymbol{\alpha}^{[t-1]}, \boldsymbol{\mu}, \boldsymbol{\sigma}) p(\boldsymbol{\alpha}^{[t-1]} | g, \boldsymbol{\alpha}_N, \boldsymbol{\mu})}. \quad (4)$$

Thus, when new PV mixture vector samples,  $\boldsymbol{\alpha}^{[t]}$  are generated according to  $p(\boldsymbol{\alpha})$  their acceptance probability is given by equation (4). The PV mixture prior density,  $p(\boldsymbol{\alpha})$  is a global PV mixture prior density previously found to closely follow a Beta density form for two classes or a Dirichlet density for the multi-variate extension. The expectation of the PV mixing vector with respect to the posterior density is taken to estimate the PV mixture content for each voxel, i.e.  $\hat{\boldsymbol{\alpha}}_i = \mathbf{E}[\boldsymbol{\alpha}_i | p(\boldsymbol{\alpha}_i | x_i, g_i, \boldsymbol{\eta})]$ .

#### 2.4. Estimation of Means and Standard Deviations

The class mean vector,  $\boldsymbol{\mu}$  and class standard deviation vector,  $\boldsymbol{\sigma}$  are also estimated via simulation of their posterior densities and estimated via the expectation of the accepted samples. The posterior density for  $\boldsymbol{\mu}, \boldsymbol{\sigma}$  can be shown to be given by

$$p(\boldsymbol{\mu}, \boldsymbol{\sigma} | X, A) = \frac{p(X|A, \boldsymbol{\mu}, \boldsymbol{\sigma})p(\boldsymbol{\mu}, \boldsymbol{\sigma})}{p(X|A)}. \quad (5)$$

The parameters are assumed here to be independent, i.e.  $p(\boldsymbol{\mu}, \boldsymbol{\sigma}) \triangleq p(\boldsymbol{\mu})p(\boldsymbol{\sigma})$ . Suitable constraints are imposed on  $p(\boldsymbol{\mu})$  and  $p(\boldsymbol{\sigma})$  by appropriate selection of sampling distributions. Hence  $p(\boldsymbol{\sigma})$  is modelled here by the easily sampled multivariate Gamma density and  $p(\boldsymbol{\mu})$  is modelled by a multivariate normal distribution.

The PV classification process alternates estimation of class parameters with the estimation of the PV mixture values. Each step consists of a fixed number of accepted samples in the IMH sampling. The classification process is initialised here by the minimum distance clustering algorithm, discussed previously in section 2 and illustrated in figure 3.

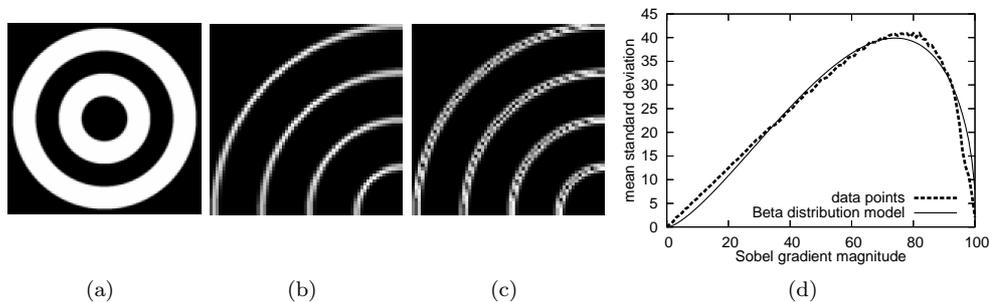
### 3. Experiments

#### 3.1. Adaptive Regularisation Model Validation

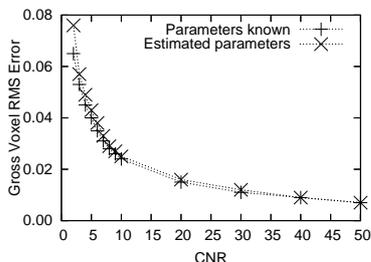
Recall that the PV mixture density,  $p(A|G, \boldsymbol{\mu}, \boldsymbol{\sigma})$  in equation (1) is dependent on the gradient magnitudes,  $G$ . The relationship between the regularising standard deviation  $\sigma_{N_i}$  and the per voxel gradient magnitude,  $g_i$  is therefore investigated. A PV imaging process was simulated by convolving high resolution volumetric data with a simulated Gaussian PSF and then sub-sampling to produce a lower resolution data volume with pure and PV voxels. The spatial gradient magnitude of this simulated PV data,  $G$  was then estimated using 3-D Sobel gradient magnitude masks, as outlined earlier. The simulated PV data is illustrated in figure 6 together with the gradient magnitudes,  $G$ , and the standard deviations,  $\sigma_{N_i}$  for each voxel neighbourhood,  $N_i$ .

The mean neighbourhood standard deviations for each gradient magnitude value were calculated, resulting in data points, also illustrated in figure 6. A Beta density function was fitted using the Nelder-Mead Simplex algorithm, (Nelder & Mead 1965) and found to provide a good fit in the form

$$\sigma_{N_i}(g, \boldsymbol{\mu}, \boldsymbol{\alpha}_{N_i}) = \frac{4.8\gamma_s}{\gamma_e(\boldsymbol{\mu}, \boldsymbol{\alpha}_{N_i})^{2.9}} g^{1.4} \sqrt{(\gamma_e(\boldsymbol{\mu}, \boldsymbol{\alpha}_{N_i}) - g)}, \quad (6)$$



**Figure 6.** Locally defined gradient magnitude versus standard deviation. (a) central slice from simulated concentric spheroid PV data. Cropped sections of (b) gradient magnitude,  $g_l$  and (c) standard deviation,  $\sigma_{N_l}$  data volumes. (d) mean standard deviation (from simulated data) as a function of Sobel gradient magnitude with Simplex fitted Beta density model (equation 6).



**Figure 7.** Voxel RMS errors obtained for GSR using known simulation parameters compared to GSR using estimated parameter values.

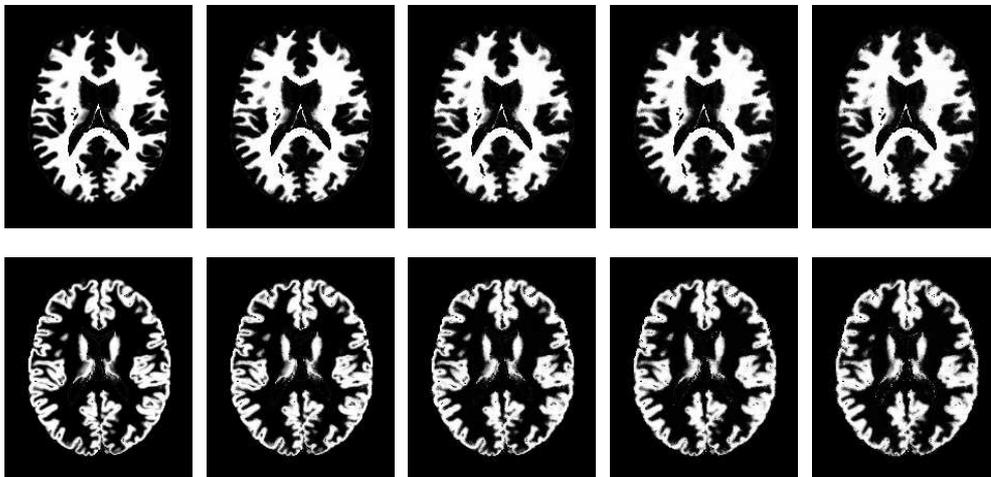
where  $\gamma_s$  and  $\gamma_e$  are an area scaling parameter and an extent parameter, respectively. This formulation thus provides a deterministic approach to estimating the optimal regularisation parameter value on a voxel by voxel basis.

### 3.2. Classification Errors with Estimated Parameter Values

Errors in the estimated parameter values may affect the classifier performance. Therefore some geometric PV affected imaging data was simulated (see figure 6a) with known noise parameters and classified using the MCMC estimated noise parameters. These results were then compared with classifier results using the known parameter values, see figure 7. Comparison was performed via a per voxel fraction Root Mean Square (RMS) error. The RMS error is used to quantify the discrepancies between the estimated mixture values for each tissue channel and the ground truth mixture values, in common with much of the PV literature, see e.g. (Laidlaw et al. 1998, Pham & Prince 2000, Shattuck et al. 2001).

### 3.3. Classification of McGill Simulated MRI Brain Data

The McGill simulated MRI brain phantom data (Cocosco et al. 1997, Collins et al. 1998, Kwan et al. 1996) is used here to assess the performance of the Gradient-controlled Spatial Regulariser (GSR) PV classifier. This data simulates the PV artifact under complex anatomical conditions associated with MRI of the human brain.



**Figure 8.** Slices from WM (top row) and GM (bottom row) classifier results for 1%, 3%, 5%, 7% and 9% noise in each column (using  $1\text{mm}^3$  voxels and 0% inhomogeneity field data sets). Note holes in WM classification are due to holes in brain mask.

Default 1%, 3%, 5%, 7% and 9% noise MRI T1 simulations together with the corresponding fuzzy tissue templates available from (BIC 2007) were selected. The selected data sets were limited to isotropic voxel sizes ( $1\text{mm}^3$ ) to simplify performance assessment. The GSR PV classifier was set to classify for three classes, corresponding to the CSF, GM and WM Central-Nervous System (CNS) components. The remaining components were automatically excluded via a GM, WM and CSF brain mask created from the fuzzy GM, WM and CSF tissue templates. In practice these components could be excluded via a skull-stripping algorithm, see e.g. (Chiverton et al. 2007).

*3.3.1. Results* Visual inspection of the output of the classifier enables qualitative analysis of the results of the classification process. Exemplar classified image slices are illustrated in figure 8 for each test volume at different levels of image noise (1%, 3%, 5%, 7% and 9%). Each of which demonstrate good agreement with the ground truth fuzzy tissue templates. Quantitative results for GSR are shown in table 1. Also shown are results from other authors using the same data.

Shattuck et al. tested brain segmentation techniques that modelled pure tissue components as Gaussian densities and PV components as a convolution of these Gaussian densities with a uniform density, (Shattuck et al. 2001). Two techniques were tested, one in which no spatial prior was modelled and the other in which a discrete spatial prior was used: Shattuck Maximum A Posteriori (SMAP) and Shattuck Maximum Likelihood (SML) respectively. Pham and Prince in 2000, developed and tested a number of classification techniques applied to the simulated brain data, (Pham & Prince 2000): Statistical Estimation of Tissue Spread 1 and 2 (SETS1 and SETS2), which relate the tissue proportion random variables to the spreading of the image data as a result of the image PSF. Also included were results for two other common techniques: Fuzzy ‘C’ Means (FCM); and a Gaussian based Generalised Expectation Maximisation Algorithm (AGEM). AGEM models only pure tissue types

**Table 1.** Voxel RMS errors for PV classifiers with simulated brain data, noise varying 0% to 9%.

Intensity inhomogeneity									
	0%			20%			40%		
<b>GSR</b>									
	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF
0%	0.064	0.096	0.058	0.121	0.149	0.071	0.194	0.224	0.097
1%	0.071	0.104	0.061	0.123	0.151	0.074	0.196	0.227	0.099
3%	0.094	0.127	0.076	0.139	0.170	0.085	0.206	0.239	0.106
5%	0.121	0.161	0.097	0.158	0.195	0.103	0.221	0.257	0.120
7%	0.143	0.191	0.118	0.179	0.223	0.123	0.237	0.279	0.138
9%	0.158	0.215	0.137	0.194	0.243	0.138	0.248	0.294	0.151
<b>SMAP</b> ( <i>CSF results not publicly available</i> )									
	WM	GM		WM	GM		WM	GM	
3%	0.088	0.137		0.084	0.139		0.102	0.159	
5%	0.129	0.178		0.122	0.172		0.125	0.176	
7%	0.165	0.215		0.160	0.211		0.159	0.220	
9%	0.359	0.384		0.184	0.240		0.187	0.245	
<b>SML</b> ( <i>CSF results not publicly available</i> )									
	WM	GM		WM	GM		WM	GM	
3%	0.096	0.144		0.090	0.145		0.105	0.163	
5%	0.158	0.206		0.146	0.196		0.142	0.194	
7%	0.253	0.299		0.249	0.297		0.214	0.277	
9%	0.353	0.420		0.304	0.366		0.293	0.362	
<i>0% intensity inhomogeneity, 3% image noise (other results not publicly available)</i>									
				WM	GM				
			<b>AGEM</b>	0.163	0.190				
			<b>FCM</b>	0.126	0.139				
			<b>SETS1</b>	0.111	0.152				
			<b>SETS2</b>	0.112	0.116				

(as Gaussian densities), but they also incorporate a discrete based spatial smoothness constraint. Their analyses were limited to the 3% noise simulated brain data, shown for comparison.

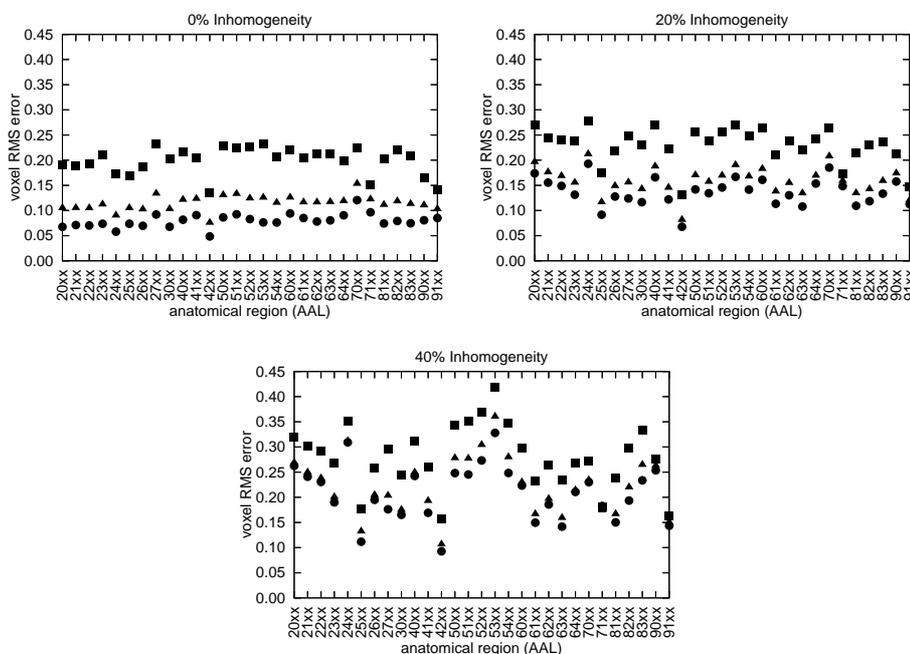
Different brain regions have different image properties which affect PV classifier performance. Therefore further RMS classification errors were calculated using a pre-registered Automated Anatomical Labelling (AAL) template developed in (Tzourio-Mazoyer et al. 2002) and distributed as part of the MRICro software, (Rorden & Brett 2000). These localised classification results (mean of GM and WM performance for each region) can be seen in figure 9 together with the corresponding anatomical labels for each region code in table 2.

### 3.4. Classification of 20 Normal MRI Brain Data Sets

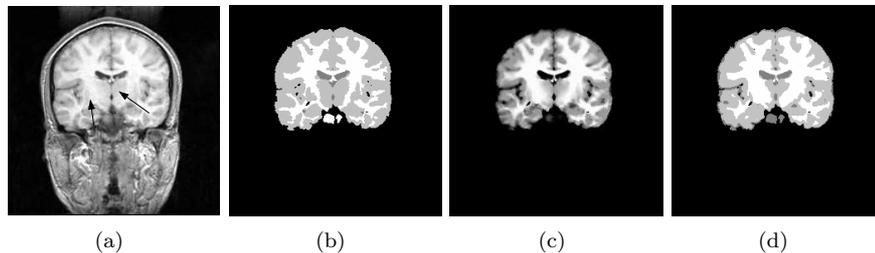
This section assesses the performance of the GSR PV classifier on twenty normal MRI brain data sets from the Center for Morphometric Analysis at Massachusetts General

**Table 2.** AAL region codes, (Tzourio-Mazoyer et al. 2002) for results in figure 9.

20xx: Precentral	40xx: Cingulum	62xx: Parietal-Inf
21xx: Frontal-[Sup/Sup-Orb]	41xx: Hippocampus/ParaHippocampal	63xx: Precuneus
22xx: Frontal-[Mid/Mid-Orb]	42xx: Amygdala	64xx: Paracentral-Lobule
23xx: Rolandic-Oper/ Frontal-Inf-[Opera/Tri/Orb]	50xx: Calcarine/Cuneus/Lingual	70xx: Caudate/Putamen/Thalamus
24xx: Supp-Motor-Area	51xx: Occipital-Sup	71xx: Thalamus
25xx: Olfactory	52xx: Occipital-Mid	81xx: Temporal-[Sup/Pole-Sup]
26xx: Frontal-[Sup-Medial/Mid-Orb]	53xx: Occipital-Inf	82xx: Temporal-[Mid/Pole-Mid]
27xx: Rectus	54xx: Fusiform	83xx: Temporal-Inf
30xx: Insula	60xx: Postcentral	90xx: Cerebellum
	61xx: Parietal-Sup	91xx: Vermis

**Figure 9.** Simulated brain data AAL region PV classifier RMS errors (mean of GM WM performance for each region), where  $\bullet$ ,  $\blacktriangle$ ,  $\blacksquare$ , correspond to 0,3,9% noise levels respectively. AAL region codes given in table 2.

Hospital, USA (IBSR 2006). Ground truth is available in the form of clinical manual segmentations. This therefore prevents direct PV assessment, particularly because we would like to perform quantitative comparison with existing techniques using a common discrete metric. Therefore we present assessment based on greatest voxel quantity of the output of the PV classifier. i.e. assign voxel  $i$  to class  $v_1$  if  $\alpha_{i,v_1} > \alpha_{i,v_2} \forall v_2 \neq v_1$ . The data sets were acquired along the coronal axis with voxel dimensions of  $1.0 \times 1.0 \times 3.0 \text{mm}^3$ . Linear interpolation was applied to determine isotropic voxel values for the purposes of the spatial regularisation and gradient magnitude calculations. The brain masked (CSF, GM and WM) image intensities for each image slice were



**Figure 10.** Exemplar coronal image slice: (a) from the 20 normal MRI brain data sets (IBSR 2006), with annotated Basal Ganglia; (b) corresponding ground truth image slice. Also shown (c) corresponding GM and WM PV classification result; and (d) discretised result.

normalised according to a volumetric mean value to remove an intensity inhomogeneity step artifact, similar to (Ibrahim et al. 2006).

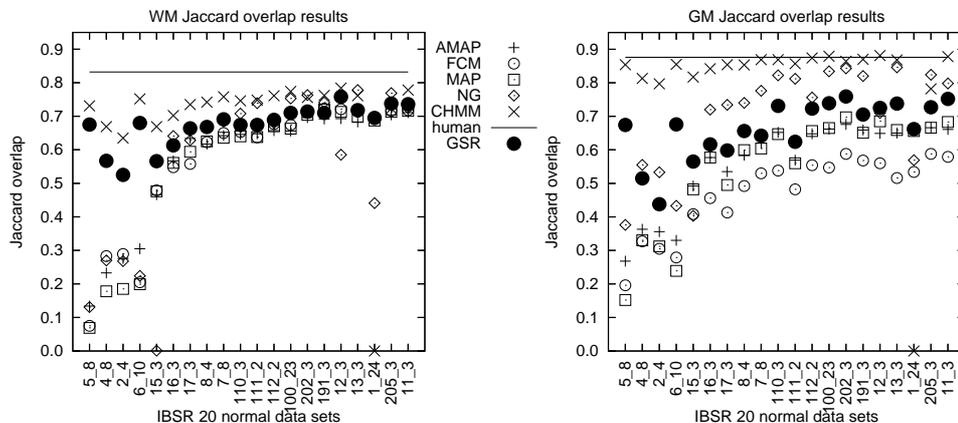
Classification performance was assessed with the Jaccard similarity metric. This enables comparison with other techniques that were also assessed with the Jaccard similarity metric for the 20 normal MRI Brain data sets with non-PV ground truth. The Jaccard similarity metric measures the amount of overlap i.e. agreement between the ground truth and the classifier output:  $J(\Theta_{GT}|\Theta_{est}) = |\Theta_{GT} \cap \Theta_{est}| / |\Theta_{GT} \cup \Theta_{est}|$ , where  $\Theta_{GT}$  is the set of voxels identified by the ground truth data as belonging to a particular classification class; and  $\Theta_{est}$  is the set of voxels experimentally classified as belonging to a particular classification class. This metric tends to 1.0 for very good segmentations and 0.0 for imperfect segmentations.

*3.4.1. Results* The GSR classifier and the discretisation process was applied to the data. An exemplar PV and discretised result can be seen in figure 10. An initial visual inspection appears to reveal overall good agreement between the ground truth.

The availability of results for other algorithms enables a quantitative comparison with the performance of GSR using clinical data. Rajapakse and Kruggel’s paper in 1998 developed two methods called adaptive-MAP (AMAP) and biased-MAP (BMAP). These methods used additional classification classes to describe the PV components (as was initially proposed in (Santago & Gage 1995)) but they also model the intensity inhomogeneities and include a MRF to improve spatial consistency. A comparison of results for the better performing technique (AMAP) together with representative results also presented in (Rajapakse & Kruggel 1998) including an FCM classifier and a conventional Maximum A Posteriori (MAP) classifier can be seen in figure 11. Also shown are more recent results from Noe and Gee, (NG):(Noe & Gee 2001); Continuous Hidden Markov Model (CHMM):(Ibrahim et al. 2006); and expert human reference metrics for inter-operator variability, (IBSR 2006).

## 4. Discussion

A novel spatially adaptive PV classifier framework has been described that spatially adapts the amount of spatial regularisation. This enables the PV classifier to adapt to the different anatomical configurations of the human brain. A completely automatic method of parameter estimation has been described. The performance of the classifier using estimated parameter values is very similar to the classifier using



**Figure 11.** Jaccard WM and GM similarity results for the 20 normals data sets, (IBSR 2006). Jaccard value of 1.0 represents a perfect segmentation.

actual simulation parameter values for a large range of Contrast to Noise Ratios (CNR)s on geometric spheres as demonstrated by the results in figure 7.

The GSR PV classifier has been evaluated on simulated brain data, (BIC 2007) for various levels of image intensity noise and intensity inhomogeneities. Classifier performance was found to be superior in comparison to other PV classifiers: SMAP and SML, (Shattuck et al. 2001) for the 0% intensity inhomogeneity data sets (0% to 9% image noise). However relative performance for GSR was worse for the 20% and 40% intensity inhomogeneity data sets. This is not surprising because the intensity inhomogeneity field was not explicitly modelled in GSR (unlike SMAP and SML). It is interesting to note that the performance for SMAP and SML for the 9% image noise and 0% intensity inhomogeneity data set were significantly worse than the SMAP and SML results for the 9% image noise 20% and 40% intensity inhomogeneity data sets. This suggests great care should be taken when incorporating this additional artifact into our future modelling process.

GM classification appears to be consistently worse in relation to the WM class for SMAP, SML and GSR classifiers on the simulated brain data. This appears to be intuitively consistent with the increased difficulty of correctly classifying GM voxels. GM intensities overlap both CSF and WM intensities and are therefore harder to classify. Furthermore, the cortex has a thin complex shape unlike the majority of the WM thus increasing the relative population and types of PV voxels. This complex topology further justifies the use of adaptive spatial regularisation: spatial correlative properties spatially vary and hence the relevance of neighbourhood information varies in these different regions.

The classifier performance has been evaluated for individual brain regions demonstrating differing classification results for different brain regions. For example, parts of the Cerebellum (90xx), Vermis (91xx), Olfactory (25xx) and Amygdala (42xx) have consistently good relative classification performance for varying levels of intensity inhomogeneity. Similarly, the individual parts of the Basal Ganglia (70xx - Caudate Nucleus, Putamen and Globus Pallidus) consistently have similar classification error. The Basal Ganglia are sub-cortical and may often be (relatively) least affected by

changes in the inhomogeneous intensity field. Regional performance evaluation was also undertaken for individual Brodmann areas (not shown here). This additional analysis highlighted differences in the relative ordering of some regions such as Brodmann region 8 which corresponds to part of the middle frontal Gyrus in the Cerebral Cortex. Very good relative performance was observed for 0% inhomogeneity data sets but very poor relative performance was observed for the 20% and 40% inhomogeneity data sets. Visual inspection of this region indicates the intensity inhomogeneity for this region is particularly significant geometrically because of the effect of the non-stationary mean intensity on the thin band of GM resulting in similar spatial and statistical properties in comparison to the surrounding WM. This regional performance analysis information will be useful to associate confidence levels to particular regions and where future classifier research could be concentrated to improve classifier performance.

Promising results for real MRI data of the human brain, (IBSR 2006) were obtained. Removal of the step change intensity inhomogeneity artifact via a pre-processing normalisation step was highly beneficial for classifier performance as such an artifact was not modelled in the GSR framework. However this pre-processing does not remove slowly varying intensity inhomogeneity fields.

The classifier framework was implemented in C++ and the execution time was found to vary from two hours to a number of days for a single data volume on a Pentium 4 3 gigahertz machine with 2 gigabyte of memory. This length and variability in execution time is due to the variable rate at which new samples are accepted in the IMH sampling process. New samples are more likely to be accepted for higher acceptance ratios, which is partly dependent here on the quality of the initialising parameters and other real world data effects such as the anatomical variability and departure from the probabilistic model such as intensity inhomogeneities. This resulted in two of the real MRI data sets (1\_24 and 13\_3) being run for fewer accepted samples due to the authors also imposing a maximum time limit of 24 hours in addition to the fixed number of accepted samples. Future research should include investigating ways of increasing the acceptance rate of new samples. This is an active area of research together with the analysis of MCMC long run behaviour. Furthermore the pure class parameters (mean intensities and standard deviations) were modelled separately from the PV gradient model and subsequently the simulations were performed separately. This was done for model convenience but may not help with the long term convergence.

There are also a number of further modelling assumptions. The conditional independence between the intensities and the gradient magnitudes uses the latent PV mixing variable,  $A$ . Conditional independence is not the same as independence. The intensities and gradient magnitudes are dependent on the same underlying variable,  $A$  which, together with the spatial information provides the causative underlying information that indicates the inter-related statistical properties that might otherwise have been included directly. The image intensities are treated here as independent from each other. The use of the PV mixture variable assists by providing an underlying MRF, but the covariation has been ignored. This is similarly true for the covariation of the PV mixture variables. Some correlative properties will have been included via the MRF, but the covariation matrix of a continuous random field is usually not considered diagonal, see e.g. (Cressie 1993).

The Jaccard similarity metric was used to quantify the performance of the discretised PV classifier results on real MRI data. This enabled comparison with other classifiers for this data. The Jaccard metric may not, however, be a fair performance

metric for a PV classifier as it requires discretisation of the PV estimates. Other classifiers, e.g. in (Rajapakse & Kruggel 1998) may not provide accurate PV estimates.

The approach described here is fully automatic and does not require an underlying anatomical prior. Such a template may help to increase performance for the data sets analysed here, but an anatomical prior would also have to be carefully modelled to allow for unusual anatomy. As already noted, intensity inhomogeneities could be included in a future model combined with a more rigorous Bayesian statistical modelling approach. These future developments may help to improve the already promising results.

## 5. Conclusions

A new probabilistic PV classifier that allows for locally adaptive spatial regularisation across volumetric imaging data has been presented. The formulation has been found to suitably adapt to both boundary and anatomical PV conditions encountered in simulated and real MRI data of the brain. Promising results and observations have been obtained for simulated brain data and real MRI data. Future work will seek to explicitly include further performance limiting imaging artifacts, such as intensity inhomogeneities into the GSR model framework.

## Acknowledgments

The authors gratefully acknowledge the Center for Morphometric Analysis at Massachusetts General Hospital and the Brain Imaging Center, Montreal Neurological Institute, McGill University for the availability of the data sets. This work was supported, in part, by the UK Engineering and Physical Sciences Research Council (EPSRC).

## Appendix A.

A MRF is usually implemented via a Gibbs distribution due to the Hammersley and Clifford theorem, (Clifford 1990), which shows that a Gibbs distribution is equivalent to a MRF. A Gibbs distribution uses an energy function,  $H(A)$  which measures the amount of disparity of the state of the system,  $A$ , which in this case is described by clique potentials,  $V_{i,j}$ , calculated over the PV mixture variables,  $\alpha_i$  and  $\alpha_j$ . Hence

$$H(A) = \sum_{i \in \Omega} \left( \sum_{j \in N_i} V_{i,j} \right) \quad (\text{A.1})$$

where  $V_{i,j} = \beta_i \sum_{v=1}^n (\alpha_{j,v} - \alpha_{i,v})^2$ .  $\beta_i$  controls the amount of regularisation, so that a small value of  $\beta_i$  reduces the influence of the neighbouring voxels' mixture values. The appropriateness of utilising a Gaussian density for this prior information was previously empirically validated in (Choi et al. 1991). The continuous form of the Gibbs distribution results in  $p(A|G, \boldsymbol{\mu}, \boldsymbol{\sigma})$ , which, assuming minimal covariation, can be shown to be given by

$$p(A|G, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \frac{1}{Z} \exp \left( - \sum_{i \in \Omega} \frac{1}{2\sigma_{N_i}^2} \sum_{v=1}^n (\alpha_{N_i,v} - \alpha_{i,v})^2 \right) \quad (\text{A.2})$$

where  $\beta_i \sum_{j \in N_i} (\alpha_{j,v} - \alpha_{i,v})^2 = \beta'_i (\alpha_{N_i,v} - \alpha_{i,v})^2$ . So that  $\alpha_{N_i,v}$  is the mean of the neighbouring PV mixture values for classification class  $v$  and the rescaled regularisation parameter,  $\beta'_i = 1/(2\sigma_{N_i}^2)$ . i.e. it is now in a variance form. So now a standard deviation,  $\sigma_{N_i}$  controls the amount of regularisation, where a large value of  $\sigma_{N_i}$  reduces the influence of the neighbouring voxels. The partition function,  $Z$  can be calculated based on the integral properties of this Gaussian density,  $Z \propto 1/(\prod_{v \in \mathcal{V}} \sigma_{N_i}^n)$ . Thus integration can be used to solve for the partition function in place of the discrete summations that often have to be avoided in many MRF models.

## Appendix B.

Each new PV mixture vector sample,  $\alpha^{[\hat{t}]}$  is accepted with probability  $\min\{\mathcal{R}, 1\}$ . The posterior density in equation (3) can be expanded using Bayes and conditional independence, yielding

$$p(\alpha|x, g, \mu, \sigma) = \frac{p(x|\alpha, \mu, \sigma)p(\alpha_N|\alpha, g, \mu)p(g|\alpha, \mu, \sigma)p(\alpha)}{p(x|g, \alpha_N, \mu, \sigma)p(g|\alpha_N, \mu, \sigma)p(\alpha_N)} \quad (\text{B.1})$$

and assume that  $p(g|\alpha, \mu, \sigma) \cong p(g|\alpha_N, \mu, \sigma)$  then

$$p(\alpha|x, g, \mu, \sigma) = \frac{p(x|\alpha, \mu, \sigma)p(\alpha_N|\alpha, g, \mu)p(\alpha)}{p(x|g, \alpha_N, \mu, \sigma)p(\alpha_N)}. \quad (\text{B.2})$$

Also recall the Gaussian form of the PV mixture prior density,  $p(\alpha|\alpha_N, \mu)$ . A Gaussian density possesses symmetry about the mean and in this case the mean is given by  $\alpha_N$ , so that  $p(\alpha|\alpha_N, \mu) \triangleq p(\alpha_N|\alpha, \mu)$ , thus

$$p(\alpha|x, g, \mu, \sigma) \triangleq \frac{p(x|\alpha, \mu, \sigma)p(\alpha|\alpha_N, g, \mu)p(\alpha)}{p(x|g, \alpha_N, \mu, \sigma)p(\alpha_N)}. \quad (\text{B.3})$$

The IMH ratio in equation (3) can therefore be written as

$$\mathcal{R} = \frac{p(x|\alpha^{[\hat{t}]}, \mu, \sigma)p(\alpha^{[\hat{t}]}|g, \alpha_N, \mu)p(\alpha^{[\hat{t}]})\lambda(\alpha^{[t-1]})}{p(x|\alpha^{[t-1]}, \mu, \sigma)p(\alpha^{[t-1]}|g, \alpha_N, \mu)p(\alpha^{[t-1]})\lambda(\alpha^{[\hat{t}]})}. \quad (\text{B.4})$$

An advantage of the IMH sampler is that when the proposal density is given by the prior density, defined here as  $\lambda(\alpha) \triangleq p(\alpha)$ , then the IMH ratio simplifies to

$$\mathcal{R} = \frac{p(x|\alpha^{[\hat{t}]}, \mu, \sigma)p(\alpha^{[\hat{t}]}|g, \alpha_N, \mu)}{p(x|\alpha^{[t-1]}, \mu, \sigma)p(\alpha^{[t-1]}|g, \alpha_N, \mu)}. \quad (\text{B.5})$$

## References

- BIC 2007, 'Simulated MRI Brain Data', "<http://www.bic.mni.mcgill.ca/brainweb>". Brain Imaging Center, Montreal Neurological Institute, McGill University.
- Chiverton J & Wells K 2004, Volumetric Partial Volume Quantification via a Statistical Model of 3-D Voxel Gradient Magnitude, in *IEEE NSS MIC*, Vol. 7, IEEE, Rome, Italy, pp. 4106–4110.
- Chiverton J, Wells K, Lewis E, Chen C, Podda B & Johnson D 2007, Statistical morphological skull stripping of adult and infant MRI data, *Comp. Biol. Med.* **37**(3), 342–357.
- Choi H, Haynor D & Kim Y 1991, Partial Volume Tissue Classification of Multichannel Magnetic Resonance Images-A Mixel Model, *IEEE Tr. Med. Imag.* **10**(3), 395–407.
- Clifford P 1990, Markov Random Fields in statistics, in G Grimmett & D Welsh, eds, *Disorder in physical systems*, pp. 19–32.
- Cocosco C, Kollokian V, Kwan R & Evans A 1997, Brainweb: Online interface to a 3D MRI simulated brain database, *NeuroImage* **5**(4, part 2/4), S425.

- Collins D, Zijdenbos A, Kollokian V, Sled J, Kabani N, Holmes C & Evans A 1998, Design and Construction of a Realistic Digital Brain Phantom, *IEEE Tr. Med. Imag.* **17(3)**, 463–468.
- Cressie N 1993, *Statistics for spatial data*, Wiley.
- Gudbjartsson H & Patz S 1995, The Rician Distribution of Noisy MRI Data, *Magn. Reson. Med.* **34(6)**, 910–914.
- Ibrahim M, John N, Kabuka M & Younis A 2006, Hidden Markov models-based 3D MRI brain segmentation, *Imag. Vis. Comp.* **24**, 1065–1079.
- IBSR 2006, <http://www.cma.mgh.harvard.edu/ibsr/>. Internet Brain Segmentation Repository, Center for Morphometric Analysis at Massachusetts General Hospital.
- Kwan R, Evans A & Pike G 1996, An extensible MRI simulator for post-processing evaluation, in *4th ICVBC*, Vol. LNCS 1131, Springer-Verlag, pp. 135–140.
- Laidlaw D, Fleischer K & Barr A 1998, Partial-Volume Bayesian Classification of Material Mixtures in MR Volume Data Using Voxel Histograms, *IEEE Tr. Med. Imag.* **17(1)**, 74–86.
- Leemput K V, Maes F, Vandermeulen D & Suetens P 2003, A Unifying Framework for Partial Volume Segmentation of Brain MR Images, *IEEE Tr. Med. Imag.* **22(1)**, 105–119.
- Nelder J & Mead R 1965, A Simplex method for function minimization, *Comp. Journal* **7**, 308–313.
- Nikolaidis N & Pitas I 2001, *3-D Image Processing Algorithms*, Wiley, New York, USA.
- Noe A & Gee J 2001, Partial volume segmentation of cerebral MRI scans with mixture model clustering, in *17th IPMI*, Springer, p. 423.
- Pham D 2003, Unsupervised tissue classification in medical images using edge-adaptive clustering, in *25th IEEE EMBC*, Vol. 1, Cancun, Mexico, pp. 634–637.
- Pham D & Bazin P 2004, Simultaneous Boundary and Partial Volume Estimation in Medical Images, in *MICCAI*, Vol. LNCS 3216, Saint Malo, France, pp. 111–113.
- Pham D & Prince J 2000, Unsupervised Partial Volume Estimation in Single-Channel Image Data, in *IEEE MMBIA*, SC, USA, pp. 170–177.
- Rajapakse J & Kruggel F 1998, Segmentation of MR images with intensity inhomogeneities, *Image Vis. Comp.* **16(3)**, 165–180.
- Rorden C & Brett M 2000, Stereotaxic display of brain lesions, *Behav. Neurol.* **12**, 191–200.
- Santago P & Gage H 1995, Statistical Models of Partial Volume Effect, *IEEE Tr. Imag. Proc.* **4(11)**, 1531–1540.
- Shattuck D, Sandor-Leahy S, Schaper K, Rottenberg D & Leahy R 2001, Magnetic Resonance Image Tissue Classification Using a Partial Volume Model, *Neuroimage* **13(5)**, 856–876.
- Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B & Joliot M 2002, Automated Anatomical Labeling of Activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain, *Neuroimage* **15**, 273–289.
- Vokurka E, Herwadkar A & Thacker N 2002, Using Bayesian Tissue Classification to Improve the Accuracy of Vestibular Schwannoma Volume and Growth Measurement, *Amer. J. Neurorad.* **23**, 459–467.
- Wang Y, Adali T, Xuan J & Szabo Z 2001, Magnetic Resonance Image Analysis by Information Theoretic Criteria and Stochastic Site Models, *IEEE Tr. Inf. Tech. Biom.* **5(2)**, 150–158.
- Williamson D, Thacker N, Williams S & Pokric M 2002, Partial Volume Tissue Segmentation using Grey-Level Gradient, in *MIUA*, BMVA, pp. 17–20.
- Windham J, Hagggar A, Hearshen D, Roebuck J & Reimann D 1988, A novel method for volume determination using MR image sequences, in *SMRM, 7th Annual Meeting, SF, USA*, p. 1081.
- Woolrich M, Behrens T, Beckmann C & Smith S 2005, Mixture Models with Adaptive Spatial Regularization for Segmentation, *IEEE Tr. Med. Imag.* **24(1)**, 1–11.