

Brain MRI Tissue Classification Based on Local Markov Random Fields ¹

Jussi Tohka^a, Ivo D. Dinov^{b,c}, David W. Shattuck^b, Arthur W. Toga^b

^aCorresponding author

Department of Signal Processing, Tampere University of Technology, P.O. Box 553,
FIN-33101, Finland.

Email address: jussi.tohka@tut.fi, fax: +358-3-31153087, tel.+358-3-31154960

^bLaboratory of Neuro Imaging, Department of Neurology, University of California Los Angeles
Medical School, Los Angeles, CA, USA

^cDepartment of Statistics, University of California Los Angeles, Los Angeles, CA, USA

Abstract

A new method for tissue classification of brain magnetic resonance images (MRI) of the brain is proposed. The method is based on local image models where each models the image content in a subset of the image domain. With this local modeling approach, the assumption that tissue types have the same characteristics over the brain needs not to be evoked. This is important because tissue type characteristics, such as T1 and T2 relaxation times and proton density, vary across the individual brain and the proposed method offers improved protection against intensity non-uniformity artifacts that can hamper automatic tissue classification methods in brain MRI. A framework in which local models for tissue intensities and Markov Random Field priors are combined into a global probabilistic image model is introduced. This global model will be an inhomogeneous Markov Random Field and it can be solved by standard algorithms such as iterative conditional modes. The division of the whole image domain into local brain regions possibly having different intensity statistics is realized via sub-volume probabilistic atlases. Finally, the parameters for the local intensity models are obtained without supervision by maximizing the weighted likelihood of a certain finite mixture model. For the maximization task, a novel genetic algorithm almost free of initialization dependency is applied. The algorithm is tested on both simulated and real brain MR images. The experiments confirm that the new method offers a useful improvement of the tissue classification accuracy when the basic tissue characteristics vary across the brain and the noise level of the images is reasonable. The method also offers better protection against intensity

¹The original version published as: Tohka, J., I.D. Dinov, D.W. Shattuck, and A.W. Toga. Brain MRI Tissue Classification Based on Local Markov Random Fields, Magnetic Resonance Imaging, 28(4): 557 - 573, 2010 <http://dx.doi.org/10.1016/j.mri.2009.12.0120>

non-uniformity artifact than the corresponding method based on a global (whole image) modeling scheme.

Key words:

segmentation ; magnetic resonance ; brain imaging ; sub volume probabilistic atlas ; classifier combination

1. Introduction

The automatic segmentation of three dimensional (3-D) magnetic resonance images (MRI) of the brain is increasingly important for neuroscientific studies. Studies concerning morphological measurements of different brain structures can benefit from the cost-effectiveness, speed and good reproducibility of automatic image segmentation [1]. The basic form of the 3-D brain MRI segmentation involves the classification of the voxels in the MR image into the three primary tissue types: white matter (WM), gray matter (GM) and cerebro-spinal fluid (CSF). This tissue classification has various important applications. For example, several cortical surface extraction algorithms depend on tissue classification [2, 3, 4, 5, 6, 7].

The research problems in tissue classification methodology include how to best cope with the intensity non-uniformity, partial volume effect, and scanner specificity of MRI as well as how to best apply the existing neuroanatomical knowledge and the spatial dependency of the labels of close-by image voxels to increase the quality of tissue classification. Recently, information and/or classifier fusion for the tissue classification, or more generally for MRI segmentation, have drawn an increasing amount of attention. For example, an interesting line of research has focused on the simultaneous registration and tissue classification [8, 9, 10, 11]. These approaches can be viewed as information fusion methods where the information from the brain atlases is combined with the information on the tissue composition of the individual image in a non-trivial manner. Similarly, label propagation techniques for the brain MRI segmentation can use information from multiple templates to better adapt to the anatomical variability of human brain leading to decision fusion approaches [12].

In this work, we propose a different kind of the information/classifier fusion approach targeted to the tissue classification problem. Below, we argue that the MR brain images are more accurately modeled locally than globally. We therefore divide the image domain into local subdomains that are described by their own local image model and derive a framework in which image models with local spatial support can be combined to yield a global, statistically based image model. Both local and global image models will be posed in a Markov Random Field (MRF) framework allowing the incorporation of the spatial information to

improve the quality of the voxel classification. We apply a Sub Volume Probabilistic Atlas (SVPA) to define the local support regions so that the division of the image domain into sub-domains is anatomically informed. The model-fusion framework is complemented by computational algorithms for solving the parameters of the model and performing the tissue classification in a completely automated manner. Particularly, we apply a global optimization approach based on a genetic algorithm for unsupervised estimation of the tissue type parameters for each local model [13]. We note that the approach presented here is in many senses complementary to the above referenced frameworks for joint image segmentation and registration. This is because, although we start from local image models, the final image model is a global generative model that could then be unified with a joint segmentation and registration approach.

As already mentioned, we argue that local image models can be more accurate than global image models. There are at least two important reasons for this. First, MR images contain low frequency intensity variations that are often described by the terms intensity non-uniformities or bias fields. In this work, we refer to this phenomenon as instrumentation non-uniformity. These low frequency variations cause non-stationarity of the parameters for the class conditional probability density functions (pdfs) of individual tissue classes. This has a negative effect on the quality of automatic tissue classifications that are based on the assumption that tissue classes have the same statistical properties throughout the image. Several methods for the correction of this low frequency noise in images have been proposed and a recent review can be found in [14]. The methods for the retrospective bias field correction typically assume that the bias field is multiplicative and spatially smooth. Some of them, including [15, 16, 17, 18] are stand-alone methods, while others perform the tissue classification jointly with non-uniformity correction [19, 20, 21]. With modern MR imaging devices, state-of-the-art methods for non-uniformity correction can be considered to be largely successful. However, these methods are not able to remove all of the low frequency noise in all parts of the image. In addition, there is evidence that the assumption of the spatial smoothness of the bias field may not be completely valid for T1 and T2 weighted scans [22]; furthermore Marroquin [23] presented a case against multiplicative nature of the bias fields. The framework we propose in this paper can reduce the unfavorable effects of unsuccessful intensity non-uniformity correction.

Second, the intrinsic properties of tissue types, such as T1 and T2 relaxation times and proton density (PD), vary across an individual brain. Since the MRI signal is a function of these intrinsic tissue characteristics, their variation across the brain causes variation in the MRI signal values as demonstrated in [24, 25]. This phenomenon - termed here as biological non-uniformity - has been confirmed in various studies. The variability of GM T1 relaxation time across the cortex is

well documented [9, 26, 27] as is the T1 relaxation time difference between cortical GM and various sub-cortical GM structures such as caudate, putamen, thalamus, amygdala and hippocampus [9, 28]. Differences in T2 between various GM and WM structures have been reported [27, 29]. There is evidence demonstrating a slight hemispheric asymmetry of PD as well as variability of PD in different WM structures [30]. All the above findings are about healthy adult populations with the exception of [29] where also schizophrenic subjects were studied. The biological non-uniformity can be even greater in pathology or in aging or d Since image intensities from a tissue class have changing statistical characteristics across the brain due to biological non-uniformity that may not be accounted for by methods designed to correct for instrumentation non-uniformity, the assumption of global models for tissue types has a negative effect on the quality of tissue classification and subsequent analysis.

Grabowski [31] and Kovacevic [32] have proposed tissue classification strategies based on similar arguments to those we present here, considering both biological and instrumentation non-uniformity. They first estimated parameters of a global mixture of Gaussians (MoG) based intensity model. Thereafter, the image domain was divided into smaller sub-domains whose intensities were modeled by individually by local MoGs. The parameters for these were estimated by a local optimization method starting from the global image parameter estimates. These previous approaches differ from our method in three important aspects. As opposed to our framework, these approaches did not include a spatial MRF prior and they did not lead to a generative, global image model. Moreover, partial volume effect (PVE) was not modeled in [32] and in [31] it was assumed that the PVE could be modeled using 50:50 mixtures, e.g., voxels containing 50% WM and 50% GM were included in the intensity model. That assumption is clearly overly simplistic.

Many methods for simultaneous instrumentation non-uniformity correction and tissue classification are similar to our method in that they consider tissue class parameters to be spatially varying over the image domain. These methods can include also an MRF prior for tissue classifications, and some of them incorporate statistically based modeling for PVE, e.g.,[33]. The general difference between these methods and our framework is that in these the spatial variation of tissue class parameters is strongly parametrized whereas our method is non-parametric in this respect. Note that we do not argue that our method should be used to completely replace existing non-uniformity correction methods, but rather to complement them. The adaptive K-means algorithm [34], modified and extended to MRI application in [35, 36] and the Adaptive MAP algorithm [37] are perhaps closest to our framework. The adaptive K-means algorithm accounts for local intensity variations by an iterative procedure involving averaging over a sliding window whose size decreases as the algorithm progresses. These vari-

ations may be tissue class specific. The algorithm also includes an MRF prior. The adaptive MAP algorithm can be considered as an extension of the adaptive K-means, where also the variance of the tissue class distribution is allowed to vary over the image. Both of these methods rely on the standard K-means clustering for their initialization, and as local optimization methods, their performance is dependent on this initialization. In contrast to our method, these methods also assume a single, global MRF prior and do not account for PVE. We will also provide experimental results clearly demonstrating the superiority of the proposed approach to the Adaptive MAP algorithm [37].

The organization of the paper is as follows. Section 2 describes the combination of the local MRFs defined on image sub-domains into a single global MRF-based image model. Section 3 describes the segmentation method rooted in the image model of Section 2. In Section 4, we present experimental validation of the method by comparing it quantitatively to other recent tissue classification schemes as well as the corresponding tissue classification method based on a traditional MRF-based image model. This validation is performed using both synthetic and real MR brain images. Section 5 summarizes the approach and discusses it in relation to the other approaches.

2. Theory: Combining local Markov Random Fields

2.1. Markov Random Fields, Gibbs distributions and image segmentation

Markov Random Fields [38] are widely applied for modeling the spatial dependency of nearby voxel tissue classes in tissue classification within brain MRI. The purpose of this sub-section is to briefly introduce the notation and give the required background information on inhomogeneous MRFs. It is not our intention to give a comprehensive review on MRFs within brain MRI segmentation.

The image voxels, or sites, are denoted by $s_i = (s_{ix}, s_{iy}, s_{iz}), i = 1, \dots, m$, where m is the number of voxels in the image. The set of all image voxels is called the image domain and denoted by S . The set of neighboring voxels adjacent to the voxel s_i is denoted by N_i . In this paper, the 26-neighborhood of the image voxels is assumed. This set of voxels together with the assumed adjacency relations forms a neighborhood system $(S, \{N_i\})$ [39, 38]. Every voxel s_i is associated with an observed intensity value $x_i \in \mathbb{R}$ and $X = \{x_i : i = 1, \dots, m\}$ is the observed image. We have a set of tissue classes $\mathcal{L} = \{l_k : k = 1, \dots, K\}$. A tissue classified image, or classification, is denoted by $C = \{c_i \in \mathcal{L} : i = 1, \dots, m\}$. Our task is then to find the best tissue class $c_i^* \in \mathcal{L}$ for each voxel given the observed image intensities. The best classification is denoted by $C^* = \{c_i^* : i = 1, \dots, m\}$.

The quality of a tissue classification is measured using its posterior probability given the observed image intensities. This leads to a maximum-a-posteriori

(MAP) optimization problem:

$$C^* = \arg \max_{C \in \mathcal{L}^m} P(C|X), \quad (1)$$

where $P(C|X)$ is the probability of the classification C given the image X . We assume that the probability of observing the intensity x_i is dependent only on the tissue class c_i and the site s_i . We write

$$P(C|X) = \frac{P(X|C)P(C)}{p(X)} = \frac{1}{Z} \left(\prod_{i=1}^m p(x_i|c_i, s_i) \right) P(C), \quad (2)$$

where $p(X|C)$ is the probability of X when the classification is C , $p(x_i|c_i, s_i)$ is the probability density of observing x_i when the class at s_i is c_i , $P(C)$ is the prior probability of the classified image C and $Z = p(X)$ is a normalization constant that is independent of C . By denoting the likelihood term by $p(x_i|c_i, s_i)$, we emphasize that the likelihood term can have a different parametric form for different voxels.

We define the prior probabilities by the Gibbs distributions:

$$P(C) \propto \exp(-U(C)); U(C) = \sum_Q U_Q(c_i : s_i \in Q) \quad (3)$$

where $U(\cdot)$ is termed the energy function and $Q \subset S$ denotes a clique (a single site or any subset of sites such that each site in the subset is contained in the neighborhood of every other location in the subset) in the graph $(S, \{N_i\})$. With certain assumptions (see [40]), the Gibbs distributions are equivalent to MRFs, and particularly, they satisfy the Markovian property of the MRFs stating that the prior probability of c_i depends only on $c_j, s_j \in N_i$. Denoting $C_{\partial(i)} = \{c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_m\}$ and $C_{N_i} = \{c_j : s_j \in N_i\}$, we can write

$$P(C) = P(c_i|s_i, C_{\partial(i)})P(C_{\partial(i)}) = P(c_i|s_i, C_{N_i})P(C_{\partial(i)}). \quad (4)$$

We emphasize that $P(c_i|s_i, C_{N_i})$ can be different for different s_i . These kinds of MRFs are termed inhomogeneous [38]. Strictly speaking, $P(C_{\partial(i)})$ should also be conditioned on sites, but we omit this to avoid unnecessarily confusing notation. By combining (2) and (4) we get [39]

$$P(c_i|s_i, X, C_{\partial(i)}) = \frac{1}{Z'} p(x_i|c_i, s_i) P(c_i|s_i, C_{N_i}), \quad (5)$$

where Z' is a normalization constant. The decompositions (4), (5) form a basis for widely applied techniques for solving the maximization problem in (1), namely iterated conditional modes (ICM) algorithm [39] and simulated annealing based techniques such as the Gibbs sampler [41]. Note that the decomposition (4)

does not directly help in generating a suitable $P(\cdot)$. If one took an arbitrary $P(\cdot|s_i, C_{Ni})$, for each i , then there might not exist a joint probability $P(\cdot) : \mathcal{L}^m \rightarrow \mathbb{R}$ generating these conditionals. Thus, MRFs are defined through Gibbs distributions and not through conditionals $P(c_i|s_i, C_{Ni})$.

In this paper, we consider second order MRFs for simplicity although the concepts can be extended to higher order MRFs as well. The focus on the second order MRFs is because tissue classification methods within MRI typically rely on second order MRFs. The energy of a second order Gibbs distribution is

$$U(C) = \sum_{s_i \in S} \left(G(c_i|s_i) + \frac{1}{2} \sum_{s_j \in N_i} H(c_i, c_j|s_i, s_j) \right), \quad (6)$$

where singleton potentials $G(\cdot|s_i)$ and doubleton potentials $H(\cdot, \cdot|s_i, s_j)$ can be freely selected. However, it is assumed that $H(c_i, c_j|s_i, s_j) = H(c_j, c_i|s_j, s_i)$ for all i, j . Note that above we sum over sites rather than interactions. Approximately, $G(\cdot|s_i)$ defines the probability for an individual tissue class occurring at the site s_i and $H(\cdot, \cdot|s_i, s_j)$ defines the probabilities of tissue classes at a site given the tissue classes at neighboring sites. With the energy given in (6) and using the relation $\frac{1}{2}(H(c_i, c_j|s_i, s_j) + H(c_j, c_i|s_j, s_i)) = H(c_i, c_j|s_i, s_j)$, the conditional is

$$P(c_i|C_{Ni}) = \frac{\exp\left(-G(c_i|s_i) - \sum_{s_j \in N_i} H(c_i, c_j|s_i, s_j)\right)}{\sum_{c_k \in \mathcal{L}} \exp\left(-G(c_k|s_k) - \sum_{s_j \in N_k} H(c_k, c_j|s_k, s_j)\right)}. \quad (7)$$

2.2. Sub volume probabilistic atlases

We apply Sub Volume Probabilistic Atlas (SVPA) to divide the image domain S into a set of sub-domains. It is important to note that, in our setting, the set of sub-domains does not coincide with the set of the tissue classes \mathcal{L} . SVPAs provide information on the probability that a certain voxel in the MR image in the stereotactic space is part of a certain brain structure [42]. The structures can include any brain regions having a clear anatomical interpretation. There can also exist multiple types of divisions of the image domain. Here, we need to distinguish between two of them. The simplest type of subdivisions - tissue probability maps (TPMs) - give the probability that a certain voxel is of a certain tissue type (CSF, GM or WM) and this has been used for automatic tissue classification in a variety of settings (see e.g. [10, 23, 43]). We do not use this information to divide the image domain into sub-domains, however it is possible to derive $G(c_i|s_i)$ terms in (6) and (7) based on this information. The second level of sub-divisions in SVPAs provides information whether a certain voxel belongs to certain sub-structure. We use this information to divide the image domain into sub-domains. In our experiments, we divide the brain into 12 local domains: cerebellum, a sub-cortical domain (consisting of basal ganglia,

thalamus, and ventricles), frontal lobe, occipital lobe, parietal lobe and temporal lobe, all divided into left and right hemispheres. This division is based on the ICBM (International Consortium for Brain Mapping) SVPA for young healthy adults [44]. For each voxel $s_i \in S$ and region (or domain) $R_b, b = 1, \dots, B$, the SVPA gives a probability $r_i(b) = P(s_i \in R_b)$ that the voxel i belongs to the region R_b . We assume that $\sum_{b=1}^B r_i(b) = 1$ for all i .

While the ICBM-SVPA divides GM into distinct sub-domains, more specifically into 20 cortical, sub-cortical or cerebellar GM structures, the ICBM-SVPA atlas does not provide similar divisions of WM and CSF. For example, in the ICBM-SVPA no distinction is made between WM in the temporal lobe and WM in the parietal lobe. Thus, we can extract $r_i(b)$ only for those voxels which are dominantly GM voxels in the atlas according to first-level (tissue type) subdivisions. For our purposes, it is imperative that we have $r_i(b)$ for every voxel, independent on its tissue type in the atlas, and we extend the ICBM-SVPA to cover also WM and CSF. In Appendix A, we present a simple procedure of extending GM SVPAs to atlases that divide WM and CSF into distinct brain regions. Obviously, this pragmatic solution takes away a part of the probabilistic interpretation of the extended atlas. However, this presents no real difficulties although some justifications are given based on the probabilistic interpretation of $r_i(b)$ as $P(s_i \in R_b)$.

2.3. Compound MRF based on local MRFs and SVPAs

It remains to be explained how to construct the inhomogeneous MRFs of Section 2.1 in practice. We use a set of homogeneous MRF-image models defined on local image sub-domains and construct a global MRF-based image model using these (we consider the likelihood term to be included within MRFs). The driving idea is that local MRFs describe the data better than a global MRF and hence allow more informed image segmentation. The sub-domains R_1, \dots, R_B are derived based on SVPAs and they must be overlapping, i.e. the SVPA must be probabilistic, as applying B distinct MRFs for the image segmentation would be likely to produce discontinuities in tissue classifications. It is better to combine these local MRFs into a single MRF and perform the segmentation with respect to this compound MRF. An additional benefit of the presented novel combination scheme is that it results in a global image model. We define *regional likelihood terms* $p_b(\cdot|\cdot)$ and the *regional prior terms* $P_b(\cdot)$ for $b = 1, \dots, B$. We interpret the regional likelihood via $p_b(x_i|c_i) = p(x_i|c_i, s_i \in R_b)$ and similarly for the regional prior. We combine $p_b(\cdot|\cdot)$ ($b = 1, \dots, B$) and $P_b(\cdot)$ ($b = 1, \dots, B$), respectively, into a compound likelihood term and a compound prior term whose regions of support cover the whole image domain. Note that the regional likelihoods do not depend on the site s_i .

2.3.1. Likelihood term

We write the *site likelihoods* in the likelihood term of Eq. (2) as

$$p(x_i|c_i, s_i) = \sum_{b=1}^B r_i(b) p_b(x_i|c_i). \quad (8)$$

Interpreting $r_i(b)$ as $P(s_i \in R_b)$, assuming that the events $s_i \in R_b, b = 1, \dots, B$ are mutually exclusive and assuming that $r_i(b)$ does not depend on c_i , (8) can be derived as

$$\begin{aligned} p(x_i|c_i, s_i) &= \sum_b p(x_i, s_i \in R_b|c_i) = \sum_b P(s_i \in R_b) p_b(x_i|c_i, s_i \in R_b) \\ &= \sum_{b=1}^B r_i(b) p_b(x_i|c_i). \end{aligned}$$

2.3.2. Prior term

The combination is not as straight-forward for the prior probability term as it was for the likelihood term. A mixture $\sum_{b=1}^B r_i(b) P_b(C)$ is dependent on the site s_i (via $r_i(b)$), and hence the compound prior term can not be derived this way. Averaging $\sum_{b=1}^B r_i(b) P_b(C)$ over the all sites is not a feasible solution either since this would take away the Markovian property (4).

Instead, we give the compound energy via regional energies and then derive the required conditionals based on this energy. We denote the regional G and H functions by G_b and H_b , respectively. Now the compound energy function is

$$\begin{aligned} U(C) &= - \sum_{s_i \in S} \left(\ln \left(\sum_{b=1}^B r_i(b) \exp(-G_b(c_i|s_i)) \right) \right. \\ &\quad \left. + \frac{1}{2} \sum_{s_j \in N_i} \ln \left(\sum_{b=1}^B (0.5r_j(b) + 0.5r_i(b)) \exp(-H_b(c_i, c_j|s_i, s_j)) \right) \right). \end{aligned} \quad (9)$$

The compound energy function defines a second order MRF based on pairwise interactions, which leads to efficient computations. This order preservation property relies on the separate combination of the 1st order and 2nd order terms. The second order term in Eq. (9) is weighted by $0.5r_j(b) + 0.5r_i(b)$ to preserve the symmetry of the 2nd order term in the energy function. This particular combination strategy is analogous to the combination of the local likelihood terms. We will show later on that it has a few interesting properties such as if all regional terms are equal then the energy in Eq. (9) reduces to the standard Gibbs energy in Eq. (6).

In this work, we assume that $H_b(c_i, c_j | s_i, s_j) = H(c_i, c_i)$ for all b . This assumption is made for practical purposes: usefulness of generating different 2nd order MRF terms for different regions would depend on the application of tissue classification and is beyond of the scope of the current work. However, starting from a more general model (9) than absolutely necessary demonstrates the extensibility of the general principles we propose. Now, because $H_b(c_i, c_j | s_i, s_j) = H(c_i, c_i)$ and $\sum_b r_i(b) = 1$ for all i , it follows that $\ln\left(\sum_{b=1}^B 0.5(r_i(b) + r_j(b)) \exp(-H(c_i, c_j))\right) = -H(c_i, c_j)$. Consequently, Eq. (9) simplifies to

$$U(C) = \sum_{s_i \in S} \left(-\ln\left(\sum_{b=1}^B r_i(b) \exp(-G_b(c_i | s_i))\right) + \frac{1}{2} \sum_{s_j \in N_i} H(c_i, c_j) \right). \quad (10)$$

This is the MRF used in this paper.

The conditionals of Eq. (10) are given by (see Eq. (7)):

$$P(c_i | s_i, C_{N_i}) \propto \exp\left(\ln\left(\sum_{b=1}^B r_i(b) \exp(-G_b(c_i | s_i))\right) - \sum_{s_j \in N_i} H(c_i, c_j)\right) \quad (11)$$

$$= \left(\sum_{b=1}^B r_i(b) \exp(-G_b(c_i | s_i))\right) \left(\exp\left(-\sum_{s_j \in N_i} H(c_i, c_j)\right)\right) \quad (12)$$

$$= \sum_{b=1}^B r_i(b) \exp\left(-G_b(c_i | s_i) - \sum_{s_j \in N_i} H(c_i, c_j)\right). \quad (13)$$

The fact that $H(c_i, c_j) = H(c_j, c_i)$ was used to simplify the conditionals.

The compound prior defined in Eq. (10) has two interesting properties:

1. If we set $G_b(c_i | s_i) \doteq G(c_i | s_i)$ for all b , i.e. we apply the same prior models for each region, Eq. (10) reduces to Eq. (6) with no site dependence in the H-functions. In other words, the compound MRF prior reduces to the ordinary 2nd order MRF prior. This was one of the reasons for choosing the above strategy for constructing the compound prior.
2. When we set $H(c_i, c_j) = 0$ for all c_i, c_j , i.e. the second order term in the MRF is ignored, the energy in Eq. (9) can be written as

$$U(C) = - \sum_{s_i \in S} \ln\left(\sum_{b=1}^B r_i(b) \exp(-G_b(c_i | s_i))\right).$$

We write $P_b^{1st}(c_i | s_i) \doteq \exp(-G_b(c_i | s_i))$ and assume that $\sum_{c_i \in \mathcal{L}} P_b^{1st}(c_i | s_i) =$

1. $P_b^{1st}(\cdot|\cdot)$ are interpreted as regional prior terms. This yields the prior

$$P(C) = \prod_{s_i \in S} \frac{\sum_{b=1}^B r_i(b) P_b^{1st}(c_i|s_i)}{\sum_{c_i \in \mathcal{L}} \sum_{b=1}^B r_i(b) P_b^{1st}(c_i|s_i)} = \prod_{s_i \in S} \sum_{b=1}^B r_i(b) P_b^{1st}(c_i|s_i).$$

Thus, if the higher order terms in the MRF are ignored, the combination strategy for priors is the mean rule in the terminology of the classifier combination. Also, this prior can be derived analogously to the derivation of the compound likelihood term.

2.3.3. Compound MRF and discussion

Substituting (3), (8), and (10) into (2) yields the posterior:

$$P(C|X) \propto \prod_{i=1}^m \left(\sum_{b=1}^B r_i(b) p_b(x_i|c_i) \right) \quad (14)$$

$$\times \exp \left(\sum_{s_i \in S} \ln \left(\sum_{b=1}^B r_i(b) \exp(-G_b(c_i|s_i)) \right) - \frac{1}{2} \sum_{s_j \in N_i} H(c_i, c_j) \right),$$

which we maximize with respect to C . Before we specify the likelihood and prior terms in our implementation, we discuss briefly why we have selected this particular strategy of combining local image models into a global image model among several possibilities. The combination strategy we have chosen has a few desirable properties: 1) it results in a global image model constructed from local models; 2) it contains a component for modeling spatial dependency of the tissue classes of neighboring voxels; and 3) it results in a computationally feasible segmentation algorithm suitable for a desktop computer. For example, an appealing way of generating a global posterior would be to proceed in a reverse order to what we have done and first combine regional priors and likelihood terms into regional posteriors and then combine the regional posteriors into a global posterior model. However, for this strategy, it is not immediately clear how to define the regional priors containing a component that models the spatial interaction between tissue classes of neighboring voxels. Further, in general, the partition functions of the MRFs that might be required for this reverse strategy are impossible to compute exactly [38]. Of course, one could assume a single prior model describing the whole image domain. However, according to property 1 of the Section 2.3.2, this would lead to the exactly the same posterior $P(C|X)$ as our scheme.

We still compare our combination scheme to an alternative scheme. In this alternative scheme, one would have $B \times K$ tissue classes (e.g. GM in occipital lobe) instead of K tissue classes and B sub-domains. The class of the voxel would

be the class maximizing the posterior probability among $B \times K$ classes and K class classification would be obtained by selecting the class among K classes corresponding to the maximizing class in $B \times K$ class classification (e.g. GM in occipital lobe would become GM). The priors for classes could be constructed by multiplying the priors for K classes with $r_i(b)$ s. This would be similar to the 'max-rule' in the classifier combination while our approach is more similar to 'mean rule' [45]. While this approach may seem simpler than the one proposed here, it has a few shortcomings which are both theoretical and practical. Theoretically, this might lead to problems of continuity in image segmentations if neighboring sub-domains have very different intensity characteristics. A practical difficulty arises if the likelihood values for each tissue class and voxel need to be stored for computational efficiency - this is the case for the algorithm introduced in the next section. Then, storage requirements would be B times higher than for our method which would be likely to complicate the segmentation of high resolution MR images on a desktop computer.

3. Methods

3.1. Algorithm overview

Our algorithm for tissue classification into CSF, GM, and WM using the compound MRF introduced in the previous section proceeds in two stages. First, the parameters for the image models (specified in Section 3.2) are estimated for each of B regions. Then, we substitute these estimated parameter values into (14) and solve the minimization problem with the Iterated Conditional Modes (ICM) algorithm [39]. The resulting algorithm is called SVPASEG.

Before any computations can be performed, the compound MRF must be fully specified. In Section 3.2, we specify the exact forms of the likelihood terms $p_b(\cdot|c_i)$ for all b and c_i , and give a method for the image parameter estimation. In Section 3.3, G_b (the 1st order terms of the local MRFs) and H (2nd order term of the local MRFs) functions of the prior term are specified. In Section 3.4, the complete algorithm is summarized.

3.2. Image model

3.2.1. Likelihood term

The same parametric image model, which follows the mixel model [46], is applied for all regions R_b . However, this model is assumed to have different parameter values for each region. Our image model does not model the instrumentation non-uniformity artifact within a region.

The images are composed of two kinds of voxels: those that contain only one type of tissue (pure voxels) and those that contain multiple types of tissues (partial volume (PV) voxels). The pure voxel classes are WM, GM, and CSF.

The PV classes are CSF/GM, WM/GM, and CSF/background. The WM/CSF class is not considered due to a very small number of WM/CSF voxels in normal human brain MRI. In each region R_b , it is assumed that the intensities of the pure voxels follow the normal density:

$$p_b(x|\mu_{kb}, \sigma_{kb}^2, l_k) = (2\pi\sigma_{kb}^2)^{-1/2} \exp - \frac{(x - \mu_{kb})^2}{2\sigma_{kb}^2} \quad (15)$$

where μ_{kb} is the mean and $\sigma_{kb}^2 > 0$ is the variance of the class l_k in the region R_b . The thermal noise in magnitude MR images is Rician, but with signal to noise ratios typical to anatomical MRI this Rician distribution can be well approximated by Gaussian distribution [47]. The pdfs for PV voxels are constructed by marginalization [48]. We assume that there are no more than two tissue types present in a voxel. The pdfs of PV voxels are dependent on the parameters of the appropriate pure voxel classes. For the mixture of tissue types l_u and l_v , the pdf of the PV class is

$$p_b(x|l_k, \theta_{ub}, \theta_{vb}) = \int_0^1 \frac{\exp\left(-\frac{(x-w\mu_{ub}-(1-w)\mu_{vb})^2}{2(w^2\sigma_{ub}^2+(1-w)^2\sigma_{vb}^2)}\right)}{\sqrt{(2\pi)(w^2\sigma_{ub}^2 + (1-w)^2\sigma_{vb}^2)}} dw, \quad (16)$$

where $\theta_{ub} = [\mu_{ub}, \sigma_{ub}^2]$. The derivation of Eq. (16) presented in [49] starts from the assumption that each voxel is divided between different tissue types which is slightly different from the starting assumption of [48]. To simplify later notation, we denote $\theta_{kb} = [\theta_{ub}, \theta_{vb}]$ when l_k is a mixture of l_u and l_v .

3.2.2. Parameter estimation

We formulate the estimation of the image parameters as finite mixture model (FMM) fitting problem. Rather than solving the image parameters μ_{kb}, σ_{kb}^2 for all regions simultaneously, we formulate a weighted likelihood (WL) approximation for solving them for each region separately. Let $\Theta_b = \{\alpha_{kb}, \mu_{kb}, \sigma_{kb}^2; k = 1, \dots, K\}$, where the mixing parameter α_{kb} represents the proportion of the tissue type l_k in the region b , denote the set of parameters to be estimated. The parameter estimation problem can now be posed as

$$\begin{aligned} \Theta_b^* &= \arg \max_{\Theta_b} \prod_{i=1}^m \left(\sum_{l_k \in \mathcal{L}} \alpha_{kb} p_b(x_i | \theta_{kb}, l_k) \right)^{r_i(b)} \\ &= \arg \max_{\Theta_b} \sum_{i=1}^m r_i(b) \ln \left(\sum_{l_k \in \mathcal{L}} \alpha_{kb} p_b(x_i | \theta_{kb}, l_k) \right) \end{aligned} \quad (17)$$

subject to $\sum_{k=1}^K \alpha_{kb} = 1$. The optimization problem (17) is solved by a novel real coded genetic algorithm (GA) [13]. This algorithm is global in the sense

that it is not sensitive to its initialization and it effectively avoids getting caught in weak local optima. This is important because the parameter estimation for FMMs often leads to complex optimization problems with many local optima. In this work, for even more accurate parameter estimates, we run the GA 10 times and select the parameters with the highest WL value as our estimate.

The tissue type parameters are estimated for each region separately rather than for all regions together partly because of computational considerations and partly because there could exist severe local optimum problems (see [50]) if trying to estimate all the parameters for all regions at once. The WL formulation is motivated by the idea that in the ICBM SVPA we have a number of voxels that belong to only one region R_b , i.e. $r_i(b) = 1$ and $r_i(b') = 0$ for $b' \neq b$. Therefore, in principle, it could be possible to estimate image parameters for each region by just sampling these voxels. However, this would lead to ignoring a large number of voxel intensities and complicate the estimation of α_{kb} . The concept of WL can be used to obviate these problems [51]. In the WL formulation, each data point is weighted according to its relevance to the estimation task. In our case, the relevance of x_i to the estimation of parameters for R_b is given by $r_i(b)$. It is interpreted that the basic sample to estimate parameters for R_b is $\{x_i : r_b(i) = 1\}$, and intensities x_i with $0 < r_i(b) < 1$ give complementary information to the estimation process. We refer to [51] for further details on the WL estimation.

The WL formulation departs from the assumptions made for the constructing of the compound MRF. However, as stated above, estimating the parameters for all regions simultaneously would be likely to produce many additional local optima. Indeed, an expectation-maximization (EM) type algorithm can be derived to solve the parameter estimation problem for all regions simultaneously using an FMM-formulation of the site likelihoods in Eq. (8). We have implemented such an algorithm and experimentally found that it was dependent on initial conditions to an extent that the robustness of the automatic segmentation would be compromised. Further, in Section 4.2, we demonstrate that the optimization of the WL objective (17) provides accurate parameter estimates for our purpose and thus this approximation serves well our purposes.

It is important for computational reasons that the WL function can be approximated by a Kullback-Leibler divergence similarly to the approximation of the ordinary likelihood presented in [13]. The only difference to the analysis presented in [13] is that the Parzen density estimate is replaced by a weighted version. This offers a considerable (up 10000 fold) speed-up for GA. A similar consideration is also valid for other parameter estimation algorithms.

3.3. Prior term

We introduce algorithms equipped with two different choices for the prior term, or more particularly, the G_b terms in the prior. SVPASEG-TPM is to be

used with images of poor quality or/and low resolution, and SVPASEG-PVE is to be used with high quality and resolution images.

- In the SVPASEG-TPM algorithm, if c_i is a pure tissue class (i.e. if $c_i \in \{CSF, GM, WM\}$), we set $G_b(c_i|s_i) = -\ln P_{atlas}(c_i|s_i)$, where $P_{atlas}(c_i|s_i)$ is the probability of a class c_i at the voxel s_i drawn from a tissue probability map (TPM, see section 2.2). $G_b(c_i) = \infty$ if c_i is a mixed tissue class. Hence, this algorithm cannot classify voxels to mixed tissue classes although the partial volume effect is accounted during the parameter estimation. We use the tissue probability maps provided in the ICBM SVPA that we used in the regional parcellation of the image domain.
- In the SVPASEG-PVE algorithm, we obtain G_b based on the estimated mixture model parameters. This is done by setting $G_b(c_i) = -\ln \alpha_{kb}^*$ when $c_i = l_k$, where α_{kb}^* is the WL estimate of α_{kb} . This choice is valid if the regions are relatively small and the estimates α_k^* are accurate. The small size of regions is required because then it can be assumed α_{kb} in the mixture model models the prior probability of the voxel s_i belonging to the class l_k .

The prior of the SVPASEG-TPM algorithm makes it more robust to noise and other image imperfections partly due to its use of TPMs and partly due to its ability to decide the pure tissue class for each voxel during the optimization of the MRF. SVPASEG-PVE algorithm requires an additional re-classification step for mixed voxels if the interest is in the segmentation into CSF, GM, and WM. We do not use any information besides the voxel intensity value for this re-classification. This makes it more vulnerable to image noise but, at the same time, this reduces the chances that the MRF term oversmooths the segmentation. Also, the estimates of the partial volume coefficients at each voxel are available based on this algorithm (see [49]). However, the focus of the current work is in crisp tissue classification.

We apply an extended Potts model similar to that suggested in [16] to model the functions $H(\cdot, \cdot)$. It is defined by

$$H(c_i, c_j) = \frac{\beta \rho_{ij}}{d(s_i, s_j)}, \quad (18)$$

where β is a user tunable parameter, $d(s_i, s_j)$ is the distance between the centers of voxels s_i and s_j , and

$$\rho_{ij} = \begin{cases} -1 & : c_i = c_j \\ 0 & : c_i \text{ and } c_j \text{ share a component} \\ 1 & : \text{otherwise} \end{cases} \quad (19)$$

We set $\beta = 0.05$ in all experiments unless otherwise mentioned. This value is based on our experience with this algorithm and our earlier work on MRI segmentation.

3.4. Segmentation algorithm

We now summarize the complete segmentation algorithm. We assume that the atlas consisting of probabilities $r_i(b)$, $i = 1, \dots, m$, $b = 1, \dots, B$ and the 2nd order term for the MRF are given. Pre-processing can include intensity non-uniformity correction, and in practice we assume that the images have been skull-stripped. Given that these preliminaries are met, the algorithm segmenting the brain image into WM, GM and CSF is as follows.

1. For each region R_b , $b = 1, \dots, B$, estimate parameters α_{kb} and the parameter vectors θ_{kb} , $k = 1, \dots, K$ by maximizing (17).
2. Find the MAP segmentation C^* by maximizing Eq. (14), where regional likelihood terms and priors are given in sections 3.2 and 3.3, respectively. The optimization is performed with the ICM algorithm [39].
3. For the SVPASEG-PVE, assign PV voxels into a pure tissue class. This is done by estimating the amount of each tissue type within a voxel as in [49]. The voxel is then assigned to the pure tissue type that is dominant in the voxel. This step is not needed for the SVPASEG-TPM algorithm as it does not allow the PV classes during the step 2.

While the ICM algorithm is a local optimization method, it has favorable properties compared to the global methods such as the Gibbs sampler in addition to computational efficacy: Besag [39] argued that because the MRFs are constructed based on local considerations, it can be better to solve the resulting MAP segmentations locally due to often unknown properties of the global MAP estimate. We note that a number of powerful optimization approaches for MRFs based, for example, on graph cuts and belief propagation have surfaced during recent years (c.f. [52] for an overview and an experimental comparison). However, as these algorithms have not been widely applied to MRI segmentation and the main purpose of this work is to introduce a novel MRF-based image model, we prefer a standard ICM-based optimization strategy.

4. Experiments and results

4.1. Performance metrics

We evaluated our segmentation methods with several datasets for which the ground-truth segmentations are available. Each of these datasets served a specific purpose for our evaluation. The experiments reported in Sections 4.2 and

4.3 depict the performance of SVPASEG in the presence of biological and instrumentation non-uniformity, respectively. The experiments reported in Section 4.4 depict the success of SVPASEG with MR images that correspond to the typical image quality in current neuroscience research.

The Jaccard coefficient² between the ground-truth and automatic segmentation was used as the performance measure. It was computed separately for each tissue class (CSF, GM, WM). The Jaccard coefficient between the sets U and V is defined as $|U \cap V|/|U \cup V|$ [53]. The value 1.0 corresponds to the perfect match and the value 0.0 corresponds to no match. The Dice coefficient defined as $2|U \cap V|/(|U| + |V|)$ is an alternative to the Jaccard coefficient. The Dice coefficient has also been applied in the evaluation of tissue classification methods. When it is necessary for comparison purposes, we report both Dice and Jaccard coefficient values. There exists a bijective correspondence between the Jaccard and Dice coefficients [16]. However, since this correspondence is non-linear, comparing the average of Jaccard coefficients to the average of Dice coefficients is not feasible.

4.2. Biological non-uniformity: Experiments with simulated data

4.2.1. Material

We simulated MR images while modeling the natural variation of T1 and T2 relaxation times across the brain to study the performance of our method when no instrumentation non-uniformity exists. The anatomical model was constructed based on the BrainWeb [54] and AAL atlases [55]. The BrainWeb atlas consists of a set of 3-D fuzzy tissue membership volumes, one for each tissue class, thus modeling the partial volume effect. To model the T1 and T2 variability, BrainWeb’s GM volume was further divided into 22 (12 cortical, 8 sub-cortical, and 2 cerebellar) membership volumes using the AAL atlas, which was created by manually segmenting a simulated T1-weighted BrainWeb image. Note that AAL atlas is different than the atlas we use for dividing the brain into local domains. Different T1 and T2 relaxation times, drawn from literature, were assigned for these 22 GM regions [26, 27, 29], thus modeling the biological non-uniformity. See Appendix B for details. Proton density (PD) values for different tissue types as well as T1 and T2 for other types of tissue than GM were the same as used in the BrainWeb simulations, see http://www.bic.mni.mcgill.ca/brainweb/tissue_mr_parameters.txt.

Given the T1, T2, PD values modeling the acquisition at 1.5T, the signal intensity I was generated using the steady-state equation for the spin-echo sequence with TR of 550 ms and TE of 15ms: $I = PD(1 - \exp(-TR/T1))\exp(-TE/T2)$.

²The Jaccard coefficient is also known as the Tanimoto coefficient in image processing literature.

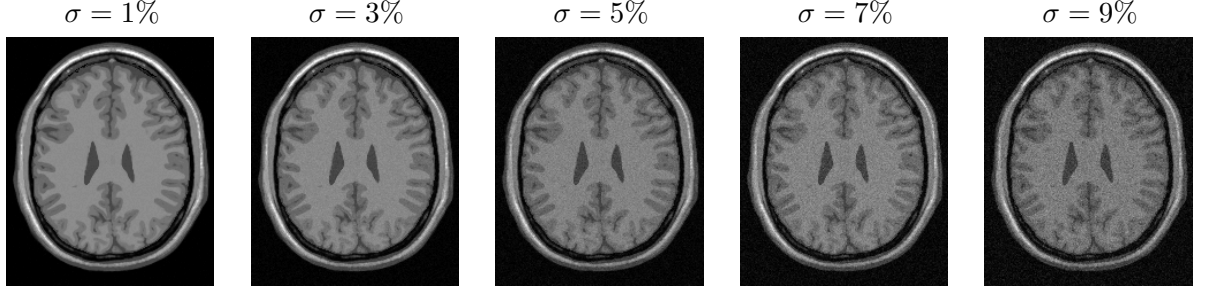


Figure 1: Examples of axial cross-sections of the simulated images.

The signal intensities were generated separately for 22 GM structures, WM, CSF, and extra-cerebral structures. The noise-free intensity of each voxel was then generated by multiplying the tissue membership value for each structure at the voxel by the signal intensity of the structure and summing these over all structures. Next, Rician noise was added to the images by adding white Gaussian noise to the real and imaginary parts of the signal and computing the complex modulus. The noise standard deviation was relative to the white matter intensity value. The noise levels were 1%, 3%, 5%, 7% and 9%. Examples of the simulated images are displayed in Fig. 1. Ten noise realizations were generated from each image. The image dimensions were $181 \times 217 \times 181$ with the voxel size of 1mm^3 .

4.2.2. Experimental settings

We compared our local MRFs based method (SVPASEG) to the corresponding tissue classification algorithm based on the homogeneous global model (MRFSEG) to investigate if the use of local MRFs improves the segmentation accuracy; MRFSEG was based on the homogeneous second order MRF with the image model parameters estimated by the same GA as used in the SVPASEG method. The MRFSEG algorithm can be formed from the SVPASEG algorithm by a modification of the atlas: in the global MRFSEG algorithm, the atlas has just one region (the brain) and $r_i(1) = 1$ for every voxel within the brain. We considered both MRFSEG-PVE and MRFSEG-TPM defined as their SVPASEG counterparts (see Section 3.3). For all algorithms, non-brain tissues were removed from consideration based on the ground truth anatomy. As the bias field was not simulated, the bias correction was not performed.

We segmented the images using the unified segmentation algorithm of the SPM5 software package ([10], <http://www.fil.ion.ucl.ac.uk/spm/software/spm5/>). This algorithm was chosen as the reference method because it is widely applied to MR brain image tissue classification. It incorporates tissue classification, nonlinear stereotactic registration, and non-uniformity correction into a single algorithm. We used two sets of parameters for the algorithm: the default

parameters (SPM5), and parameters that discourage the bias field correction (SPM5 no BFC, bias regularization was set to 10). The second parameter set was used because no instrumentation bias was simulated and thus it might be that bias correction would have negative effect to the results. We tested a few other parameter settings, but the default parameter settings appeared to be the best. The ground truth brain mask was used to prune the SPM5 segmentation, which incorporates the brain extraction into the same algorithm.

We subjected the resulting Jaccard coefficients to graphical analysis and statistical hypothesis testing. The hypothesis testing was performed using SOCR: Statistics Online Computational Resource <http://www.socr.ucla.edu/> [56]. All p-values we report were based on the (two paired sample) sign test. This test was selected because it is robust with respect to distributional assumptions and the Jaccard coefficient values cannot be assumed to be normally or even symmetrically distributed.

4.2.3. Results

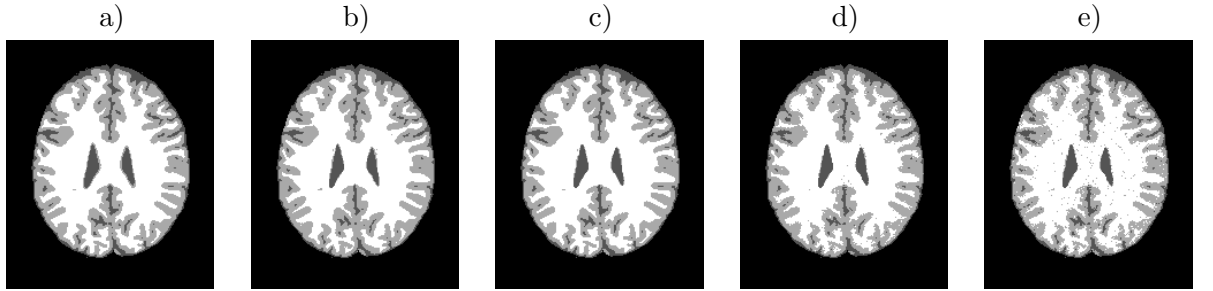


Figure 2: Examples of axial slices of segmentation results. a) Ground truth ; b) SVPASEG-TPM, $\sigma = 1\%$; c) SVPASEG-PVE, $\sigma = 1\%$; d) SVPASEG-TPM, $\sigma = 5\%$; e) SVPASEG-PVE, $\sigma = 5\%$.

The Jaccard coefficient values averaged over 10 noise realizations are presented in Table 1. Example segmentations are shown in Fig. 2. The results of the SVPASEG algorithms were better than those of the MRFSEG counterpart when the noise level was 1% (one sided $p < 10^{-7}$) or 3% (one sided $p < 10^{-5}$). The p-values were computed using the two paired sample sign test and pooling all noise realizations, tissue types, and algorithm types resulting in 60 pairs of Jaccard coefficients. When the noise level was 7% or 9%, the situation was reversed (for both noise levels $p < 10^{-9}$). This was not surprising since when the noise level increases, the biological variations in image intensities become smaller compared to the level of the instrumentation noise, and the increased complexity of the SVPASEG model (compared to that of MRFSEG) causes more errors in the parameter estimation. Also, the estimated noise variances were typically

σ		1%	3%	5%	7%	9%
	SVPASEG-PVE	0.9580	0.9279	0.8827	0.8101	0.7221
	SVPASEG-TPM	0.9175	0.9179	0.8910	0.8465	0.7887
CSF	MRFSEG-PVE	0.9603	0.9285	0.8940	0.8521	0.7379
	MRFSEG-TPM	0.8870	0.9189	0.8941	0.8577	0.8060
	SPM5	0.9487	0.9279	0.8921	0.8396	0.7803
	SPM5 no BFC	0.9491	0.9301	0.8906	0.8345	0.7714
	SVPASEG-PVE	0.9480	0.9044	0.8380	0.7588	0.6778
	SVPASEG-TPM	0.9113	0.9082	0.8710	0.8243	0.7763
GM	MRFSEG-PVE	0.9429	0.8976	0.8593	0.8082	0.7177
	MRFSEG-TPM	0.8895	0.9021	0.8674	0.8276	0.7832
	SPM5	0.9073	0.8993	0.8490	0.7798	0.7138
	SPM5 no BFC	0.9079	0.9017	0.8413	0.7632	0.6918
	SVPASEG-PVE	0.9530	0.9106	0.8453	0.7719	0.6869
	SVPASEG-TPM	0.9162	0.9145	0.8750	0.8250	0.7680
WM	MRFSEG-PVE	0.9459	0.9036	0.8646	0.8177	0.7525
	MRFSEG-TPM	0.8996	0.9060	0.8688	0.8287	0.7767
	SPM5	0.8977	0.9015	0.8528	0.7892	0.7373
	SPM5 no BFC	0.8986	0.9050	0.8458	0.7751	0.7222

Table 1: The Jaccard coefficients for the simulated images. Each coefficient is the average over 10 noise realizations.

lower with SVPASEG than with MRFSEG giving more weight to the prior term with MRFSEG. Increasing β would be likely to eliminate the difference between MRFSEG and SVPASEG with higher noise levels.

The performance of SVPASEG-PVE was superior to SVPASEG-TPM at the lowest noise level (this was evident based on the graphical, line-chart based analysis of the Jaccard coefficient values and no hypothesis test was required; see also Fig. 2 b) and c);). When the noise level was 3%, no considerable differences in the performance of the algorithms were noticed. When the noise level was higher than 3%, SVPASEG-TPM, which only considers PVE during the parameter estimation, was the best choice. Similar considerations were valid for MRFSEG algorithms. This was not surprising since the PV voxel reclassification step does not use spatial prior information and the inaccuracy of the fractional content estimation increases with the noise level. Thus, it can be expected that SVPASEG-PVE is more vulnerable to noise near the tissue type boundaries than SVPASEG-TPM (see Fig. 2 d) and e)). Conversely, when the noise level was low, the Potts model based MRF without PV classes slightly over-regularized the segmentation near the tissue type boundaries. This effect appeared to be avoided by the MRF including PV classes.

The Jaccard coefficients of SPM5 [10] are also listed in Table 1. We limit our discussion to the comparison of the SPM5 and SVPASEG-TPM because both use tissue probability maps as an aid to classification. SVPASEG-TPM outperformed both SPM5 methods (in both cases one-sided $p < 10^{-13}$, 150 pairs of samples). This serves as a clear indication that in certain settings using SVPASEG for tissue classification is advantageous.

We studied the accuracy of image parameters estimated by optimizing the WL objective. Based on the anatomical ground-truth, we computed the true image parameters (means and variances for each pure tissue class) and the correct proportions of each tissue type within each of the 12 brain regions with the procedure described in [49]. The averaged (over all domains, noise realizations, and tissue types) absolute error in the estimated tissue class means varied between 0.52 (1% noise) to 2.83 (9% noise) when the image intensities were in the range [0,255]. To illustrate the effect of these parameter estimation errors to the segmentation results, we compared SVPASEG-TPM equipped with true tissue parameters instead of the estimated ones. The average improvement in the Jaccard coefficient was only 0.002 (averaged over all tissue types, noise realizations and noise levels). Thus, we can conclude that the parameter estimation was sufficiently accurate for our purposes.

With these high resolution synthetic images, the time consumption of our C implementation of the SVPASEG algorithm was below 30 minutes for a single image. The time consumptions are reported for a IBM xSeries 335 with 3.0 GHz Intel Xeon processor running on Linux.

4.3. Instrumentation non-uniformity: The IBSR dataset

4.3.1. Material

The proposed tissue classification method was tested with the IBSR (Internet Brain Segmentation Repository) dataset. The IBSR dataset comprises 20 normal MR brain data sets and their semiautomatic segmentations. These are provided by the Center for Morphometric Analysis at Massachusetts General Hospital and are available at <http://www.cma.mgh.harvard.edu/ibsr/>. We call this dataset IBSR1 to distinguish between it and a more recent dataset, IBSR2, used in the next subsection. These images have been used to evaluate several tissue classification methods. This dataset is especially suitable for the evaluation of the capability of our method to reduce the non uniformity artifacts, since some of the images contain quite high levels of intensity non uniformity. The dimensions for the images were typically $256 \times 256 \times 62$. The voxel size was $1mm \times 1mm \times 3mm$.

4.3.2. Experimental settings

Preprocessing of the images consisted of correction of inter-slice intensity non-uniformity [57] and the brain extraction by the Brain Surface Extractor (BSE) [16]. The ICBM atlas was registered to the images with an affine transformation [58] and resampled to the image dimensions. We considered two sets of images: the first set was corrected for the 3D intensity non-uniformity by the N3 algorithm [15] and the second set images was not corrected for the 3D intensity non-uniformity. We will abbreviate the former set as N3 set and the latter set as No INUC set. Only SVPASEG-TPM and MRFSEG-TPM algorithms were considered due to the low image resolution. The SVPASEG-PVE and MRFSEG-PVE algorithms classify the PV voxels, numerous due to low resolution, without the use of spatial information that is expected to degrade the quality of segmentation results. In addition, intermediate PV classifications were not expected to be particularly useful with these low resolution images.

4.3.3. Results

The average Jaccard coefficient values are listed in Table 2. SVPASEG-TPM achieved the mean Jaccard coefficient values of 0.698 for GM and 0.685 for WM when the preprocessing included the N3 non-uniformity correction and 0.698 for GM and 0.675 for WM when the non-uniformity correction was skipped.

MRFSEG-TPM's mean Jaccard coefficients were 0.695 for GM and 0.673 for WM (with N3 non-uniformity correction). It is not reasonable to test MRFSEG without non-uniformity correction as a preprocessing step since this algorithm assumes that the tissue class intensities follow the same model across the brain. The Jaccard coefficients of individual images for MRFSEG and SVPASEG algorithms are visualized in Fig. 3. Especially with the images containing large intensity non-uniformity artifacts, the results with SVPASEG-TPM were better than with

Method	Mean Jaccard	
	GM	WM
MRFSEG-TPM N3	0.695	0.673
SVPASEG-TPM N3	0.698	0.685
SVPASEG-TPM No INUC	0.698	0.675
A-MAP	0.564	0.567
BrainSuite	0.595	0.664
PVE-TMCD	0.670	0.669
MPM-MAP	0.662	0.683
CGMM	0.68	0.66

Table 2: The average Jaccard similarity coefficients between automatic and manual segmentation results over the IBSR dataset for various algorithms. With SVPASEG algorithms, N3 refers to the results with N3 non-uniformity correction and No INUC refers to the results without non-uniformity correction as preprocessing step. See the text for references to competing methods. Note: The CGMM results are averages of 18 images, since the images 202_3 and 4_8 appeared to be missing from their experiments.

MRFSEG-TPM (see Fig. 3C for an example). Surprisingly, MRFSEG-TPM’s Jaccard coefficients for GM were in some cases better than SVPASEG-TPM’s Jaccard coefficients when both used N3-correction. This did not occur with WM Jaccard coefficients indicating that this lowering was probably due to differences in CSF/GM boundary between the methods (see Fig. 3D for an example). No notable differences between the GM Jaccard coefficient values of SVPASEG-TPM N3 and SVPASEG-TPM No Inuc were observed. For some images, WM Jaccard coefficients were lower for the No Inuc set indicating that the N3 correction was useful.

Several authors have evaluated their methods for the tissue classification using the IBSR1 dataset. The average Jaccard coefficient values of recent methods are provided in Table 2 along with the average Jaccard coefficient values of the MRFSEG and SVPASEG methods. The Jaccard coefficients included in Table are by *Adaptive MAP* (A-MAP, [37]), BrainSuite [16], MPM-MAP [23]; PVE-TMCD [49], and CGMM [59]. The average Jaccard coefficient values in Table 2 are as reported in the respective papers except for A-MAP which were obtained from <http://www.cma.mgh.harvard.edu/ibsr/>. The pre-processing choices (skull stripping, non-uniformity correction, whether they use TPMs) of the algorithms varied. The skull-stripping was implemented in the same way as here with the BrainSuite whereas for the other methods skull stripping method was different. For example, the PVE-TMCD results were based on the brain extraction using the ground-truth brain mask. The MPM-MAP used tissue probability maps to aid the tissue classification similarly to SVPASEG-TPM and MRFSEG-TPM whereas the other methods did not. SVPASEG-TPM produced

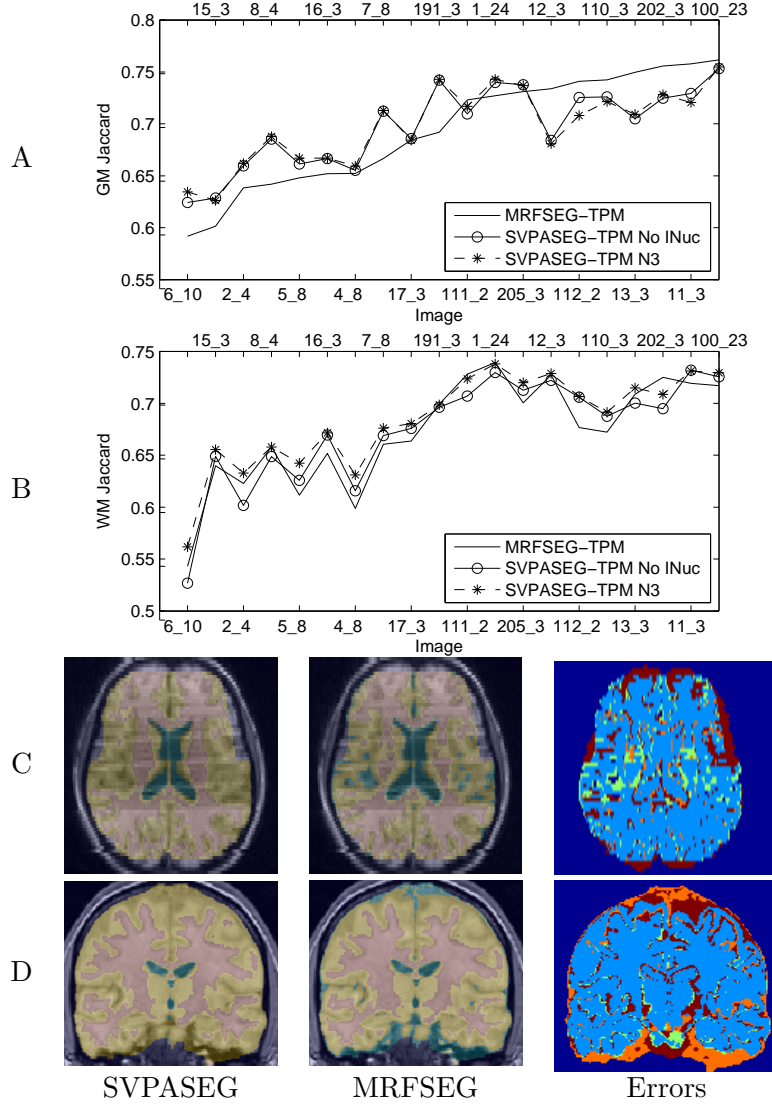


Figure 3: The GM (A) and WM (B) Jaccard coefficients of MRFSEG-TPM, SVPASEG-TPM (N3), and SVPASEG-TPM (No Inuc). The coefficients ordered so that the GM Jaccard coefficient for MRFSEG-TPM is increasing. Panel C: an example slice of the segmentation results of an image with a strong shading artifact (the axial slice 156 of the image 6_10). Panel D: An example slice where the MRFSEG's GM segmentation was better than that of SVPASEG (the coronal slice 30 of the image 12_3). C and D, from left, SVPASEG and MRFSEG segmentations (WM in red, GM in yellow) overlaid on the image slice and misclassified voxels compared to the manual segmentation. In the error images: light blue: SVPASEG and MRFSEG both correct; orange: SVPASEG incorrect, MRFSEG correct; green: SVPASEG correct, MRFSEG incorrect; red: both incorrect;

a higher average Jaccard coefficient values than any of the competing methods for both GM and WM when N3 non-uniformity correction was included in the pre-processing.

The tissue classification algorithms from widely used software packages SPM5 [10] and the FSL’s FIRST [60] were recently evaluated with these data [61]. The Dice coefficient was used as the performance measure. For GM (WM measures were not provided in [61]), the average Dice coefficients were 0.790 (SPM5) and 0.756 (FIRST) while for our methods these were 0.822 (SVPASEG-TPM and N3), 0.821 (SVPASEG-TPM No Inuc) and 0.818 (MRFSEG-TPM).

With the IBSR1 dataset, the processing time of our C implementation of the SVPASEG algorithm was less than 10 minutes for a single image (excluding preprocessing).

4.4. Anatomical variation: the IBSR dataset version 2.0

4.4.1. Material

The IBSR project has also made available a second dataset at <http://www.cma.mgh.harvard.edu/ibsr/>. This IBSR2 dataset consists of 18 T1-weighted MR brain images and their manual segmentations. The image size was $256 \times 256 \times 128$ with the voxel dimensions varying between $0.84 \times 0.84 \times 1.5$ mm³ and $1 \times 1 \times 1.5$ mm³. The images in the IBSR2 dataset are of better quality than the images in the original IBSR1 dataset and they correspond to the typical image quality in current neuroscience research. The 18 subjects represented a wide age range from 7 years to 71 years of age. Thus, considerable inter-subject differences in the anatomy existed. The images have been ‘positionally normalized’ into the Talairach orientation (rotation only) and non-uniformity corrected by the CMA ‘autoseg’ biasfield correction routines.

4.4.2. Experimental settings

The pre-processing included brain extraction by BSE and affine registration of the ICBM atlas to the images by MNI’s autoreg routines [58] followed by resampling of the atlas to image dimensions. No further bias field correction was done.

We evaluated the segmentation results using both Jaccard and Dice metrics. The Jaccard metric was used to provide a comparison to the other experiments in this work. The Dice metric was used for comparing our method to other recent tissue classification methods evaluated using this data and the Dice similarity metric.

4.4.3. Results

The average quantitative results are displayed in Table 3. The results of SVPASEG and MRFSEG are compared to the CGMM method [59], the EMS

method [20], and the Awate method [62]. The results of the CGMM and Awate methods were extracted from the respective papers and the results of the EMS method were extracted from [59].

The SVPASEG algorithms produced higher average Dice coefficients than the corresponding MRFSEG algorithms. This can be seen as evidence that the SVPASEG indeed complements the existing nonuniformity correction methods or copes better with the biological nonuniformity. We tried to apply N3 non-uniformity correction to the images, but this degraded the image quality and subsequently the segmentation results, sometimes considerably. The SVPASEG-PVE algorithm was the best of the algorithms, thus the PVE-model for the tissue classification appears to be advantageous for these images. The algorithms using the atlas prior (SVPASEG-TPM and MRFSEG-TPM) produced better segmentations of the GM and worse segmentations of the WM than the others. This was mainly because two reasons: the atlas prior resulted in less GM classified as CSF and more WM classified as GM. The first effect increased the Dice coefficient for GM while the second one decreased the Dice coefficient for WM.

The SVPASEG algorithms produced better average Dice coefficients than the competing methods with a single exception - SVPASEG-TPM's average Dice coefficient for WM was lower than the Awate's method. The increase in the average Dice coefficient for SVPASEG-PVE compared to competing methods was 3.6% - 6.4% for GM and 0.4% - 5.2% for WM.

method	Dice				Jaccard	
	GM mean	GM std	WM mean	WM std	GM mean	WM mean
SVPASEG-PVE	0.8362	0.0248	0.8906	0.0146	0.7192	0.8031
SVPASEG-TPM	0.8459	0.0174	0.8605	0.0137	0.7334	0.7554
MRFSEG-PVE	0.7967	0.0694	0.8683	0.0271	0.6669	0.7682
MRFSEG-TPM	0.8328	0.0243	0.8540	0.0253	0.7142	0.7460
Awate	0.8074	0.0426	0.8868	0.0179	n/a	n/a
CGMM	0.7890	0.0600	0.8470	0.0210	n/a	n/a
EMS	0.7860	0.0500	0.8570	0.0210	n/a	n/a

Table 3: The quantitative results with IBSR2 dataset. The abbreviation std stands for standard deviation. For the abbreviations of competing algorithms, see text. The Dice and Jaccard coefficients of MRFSEG and SVPASEG algorithms for each 18 images are available as supplementary material.

5. Discussion

We have introduced a novel framework for tissue classification in brain MRI founded on local, MRF-based image models. The driving idea was that modeling image content locally rather than globally can provide more accurate classification

results. There were two main reasons for this argument. First, natural biological variation of tissue types in different parts of the brain can cause problems for traditional tissue classification schemes where the whole image is assumed to have the same intensity characteristics after the intensity non-uniformity correction. Second, we hoped to cope better with large scale intensity non-uniformity artifacts that may not be completely eliminated by correction algorithms. To meet the aim of locally modeling image content, we introduced the concept of the local MRFs. These local MRFs were then combined to create an MRF-model describing the whole volume. Our framework included modeling of the partial volume effect and spatial dependency of the tissue classes of neighboring voxels. The regions of support for local image models were derived based on SVPA's and they were defined as overlapping fuzzy sets. Thus the resulting global MRF model is spatially continuous.

We evaluated our algorithm with respect to both of the reasons for designing it. The evaluation related to the biological variation was performed based on realistic synthetic images so that we could control the level of biological non-uniformity. The SVPASEG algorithm was shown to provide more accurate segmentation than the corresponding global image model based tissue classification algorithm (MRFSEG) when the noise level was reasonable. Obviously, it is not the case that SVPASEG would lead to improvements in the classification accuracy when the level of instrumentation noise significantly exceeds the level of biological non-uniformity and this was confirmed by the experiments. The evaluation against the instrumentation non-uniformity was performed based on the IBSR1 dataset. This dataset has two favorable features for this evaluation: many tissue classification algorithms have been evaluated using it and it contains images with strong non-uniformity artifacts. In this experiment, SVPASEG-TPM performed favorably compared to MRFSEG-TPM, and particularly, SVPASEG-TPM gave slightly better results than other recent algorithms evaluated using this dataset. We additionally performed an evaluation with IBSR2 dataset comprising of 18 high resolution MRIs of the subjects from a wide age range. The SVPASEG algorithms outperformed the MRFSEG algorithms and a few recent tissue classification methods with these data.

There exist several approaches for brain MRI tissue classification which combine an MRF modeling of the voxel labelling to a non-stationary intensity model accounting for the instrumentation non-uniformity. Many of these, for example [60] and [21], model the intensity non-uniformity as spatially smoothly varying multiplicative noise. The assumption about the multiplicative nature of intensity non-uniformity was criticized in [23] and the authors presented a segmentation algorithm (MPM-MAP) which does not evoke this assumption. However, the smoothness of the non-uniformity was assumed within a tissue class. Measurement noise was modeled with a single noise term which was independent on the

tissue-class. A non-parametric, MRF-based approach for tissue classification was suggested in [62] using an adaptive model of image neighborhoods that avoids the explicit modeling of the bias field. However, in [62] no particular attention was drawn to the biological non-uniformity. While these two algorithms [23, 62] are partly based on similar considerations as ours, there are important differences. First, we consider the intensity non-uniformity to arise both from the MR imaging methodology and from biological variations of the tissue composition across the human brain. The biological non-uniformity can violate the assumptions that the intensity non-uniformity is spatially smooth or that the noise level is constant all over the image. Also, there is a certain interplay between the instrumentation and biological non-uniformities in MRI [63]. Second, our approach models the partial volume effect in such a way that allows the estimation of the fractions of the tissue types in each voxel. It would be straightforward to generate the tissue fraction maps with SVPASEG-PVE although we concentrated on hard segmentation in this work. Third, our approach allows the use of different explicit MRF-prior models for different brain regions. This might be useful when tissue classification in a certain brain region is of interest. For example, WM/GM segmentation of the occipital cortex, required in vision research, has been reported to be especially difficult for automatic algorithms. A specific prior for accounting particular features of this brain region could be designed to alleviate these problems. We also evaluated SVPASEG with the same data as MPM-MAP [23] (IBSR1) and Awate’s method [62] (IBSR2). SVPASEG-TPM, which is the most relevant version of the algorithm to compare to MPM-MAP, produced higher average Jaccard coefficients than MPM-MAP. Both versions of SVPASEG produced a higher mean Dice coefficient (averaged over all subjects and both tissue types) than Awate’s method.

We have introduced two versions of the SVPASEG algorithm in this work. SVPASEG-PVE used partial volume (PV) classes during the tissue classification. This was noted to be advantageous with low noise high resolution images. The advantage probably followed from the ability of the algorithm to treat PV voxels (mostly existing at the tissue type boundaries) differently from pure tissue classes and decide their final (pure) class based on the intensity value only. As expected, when the noise level increased or the image quality otherwise degraded, including PV classes into the tissue classification became disadvantageous and SVPASEG-TPM, modeling PV effect only during the parameter estimation, outperformed SVPASEG-PVE. The SVPASEG-TPM algorithm also included an SPM-style tissue probability map prior, which is expected to provide more robustness to the segmentation when the images are of poor quality.

The parameter estimation for local intensity models was performed relying on FMMs and weighted likelihood formulation. These challenging optimization problems were solved by a novel GA [13]. The GA, as a global optimization

method, is not overly sensitive to its initialization and can provide accurate parameter estimates even when the optimization problem is difficult. The FMM based method was selected in part by its generality. For example, it would be possible to apply the SVPASEG method for tissue classification in brain MRI from other species in addition to human brain MRI. Indeed, algorithms for tissue classification of human brain MRI are not typically directly suitable for tissue classification in mouse brain MRI as we have demonstrated [13]. We experimentally confirmed with the synthetic MR images that the classification accuracy improves only minimally if the true values of the tissue type parameters were used instead of the parameter values estimated by the GA. This confirmation served two related purposes. It demonstrated that the heuristic weighted likelihood formulation leads to accurate tissue-class parameter estimates, and more importantly, it appears that no major improvements to the tissue classification accuracy can be gained by an alternate parameter estimation approach. However, if the interest lies in partial volume estimation (estimation of tissue fraction maps) instead of the tissue classification, we have previously found that application specific parameter estimation techniques offer an advantage over general techniques [64]. A minor disadvantage of the specialized parameter estimation techniques such as those based on outlier detection [64, 49] is that the estimation of the number of voxels of a particular tissue type, required for the SVPASEG-PVE algorithm, would not be possible.

In summary, we have proposed a framework that makes less assumptions about the sources and nature of intensity non-uniformity within brain MRI than many of the existing methods for tissue classification and we have supplied computational tools to support this framework. This tissue classification method can be combined with an existing non-uniformity correction method if that is considered to be beneficial. We have supplied experimental evidence about the increased segmentation accuracy of the general framework proposed in several settings. Dividing the brain into 12 sub-domains, as we have done, may not be optimal. It is obvious that the performance can be improved if a specific division is utilized. A more precise sub-division based on an analysis of the biological and instrumentation intensity non-uniformities and their combined effects to tissue classification could bring further improvements to the tissue classification results. However, more precise sub-divisions might be dependent on a particular pulse sequence or scanner and advantages they offer would be more dependent on the accuracy of stereotactic registration than with the coarse sub-division used here.

Appendix A: Extending gray matter SVPA to white matter and CSF regions

We describe the procedure used to extend to gray matter ICBM SVPA to cover also white matter and CSF regions. Denote the set of voxels in the brain by BM and the set of voxels belonging to a GM region (in the ICBM-SVPA) with probability greater than zero by GM . Our purpose is generate a SVPA $\{r_i(b), s_i \in BM, b = 1, \dots, B\}$ covering whole brain given a partial SVPA $\{r_i^*(b), s_i \in GM, b = 1, \dots, B\}$ such as ICBM SVPA. First, we create an initial SVPA by setting

$$r_i^0(b) = r_i^*(b) \quad \text{if} \quad s_i \in GM; r_i^0(b) = 0 \quad \text{otherwise.}$$

Given this initialization we then proceed by evolving the fields $\{r_i^0(b)\}$ by the 3D heat equation for all $b = 1, \dots, B$ separately. Note that this amounts to iterated Gaussian smoothing [65]. The evolution is terminated when all $r_i^t(b) > 0$ for some b . Then, we obtain our SVPA by setting

$$r_i(b) = r_i^*(b) \quad \text{if} \quad s_i \in GM; r_i(b) = \frac{r_i^t(b)}{\sum_{b=1}^B r_i^t(b)} \quad \text{otherwise.}$$

Appendix B: T1 and T2 values for simulated images

T1 and T2 relaxation times in ms used in the simulation are provided in table 4. T1 values reported in [27] at 3.0T were converted to corresponding values at 1.5T using the relationship $T1 = a(f^q)$, where f is the Larmor frequency of the scanner, and a and q are constants [66]. The constants were solved assuming that the T1 values at frontal lobe should be equal in [26] and [27]. T2 values should theoretically be independent on the field strength. However, considerable differences in T2 values between [29] at 1.5T and [27] at 3.0T were observed; these were corrected for in the same way as T1 relaxation times.

Acknowledgments

Thanks to Grace Liang for her help in typing. The SVPASEG and MRFSEG algorithms are available at <http://www.loni.ucla.edu/Software>.

This work is funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 RR021813. Information on the National Centers for Biomedical Computing can be obtained from <http://nihroadmap.nih.gov/bioinformatics>. This work was supported by research grants P41 RR013642 and R01 MH071940 (funded by the NIMH). J. Tohka's work was supported by the Academy of Finland (application number 129657, Finnish Programme for Centres of Excellence in Research, 2006-2011) and the University Alliance Finland Cluster of Excellence STATCORE.

Region	hemisphere					
	left		right		medial	
	T1	T2	T1	T2	T1	T2
Frontal GM	1108	84	1085	84	1097	84
Parietal GM	1162	86	1117	86	1140	86
Occipital GM	1064	101	1064	101	1064	101
Temporal GM	1124	90	1124	89	1124	90
Insula	1124	78	1124	78	N/A	N/A
Caudate	966	73	1080	74	N/A	N/A
Putamen	930	68	953	69	N/A	N/A
Thalamus	827	71	850	72	N/A	N/A
Cerebellar GM	1097	84	1097	84	N/A	N/A
WM	641	66	641	66	N/A	N/A
CSF	2569	329	2569	329	N/A	N/A

Table 4: T1 and T2 relaxation times used in the simulation in ms. The values for the medial structures were the averages of relaxation times of the corresponding left and right structures. N/A means that there was no medial structure.

References

- [1] A. Toga, J. Mazziotta (Eds.), Brain Mapping: The Methods, 2nd Edition, Academic Press, San Diego, 2002.
- [2] A. Dale, B. Fischl, M. Sereno, Cortical surface-based analysis. I. segmentation and surface reconstruction, *NeuroImage* 9 (1999) 179 – 194.
- [3] D. MacDonald, N. Kabani, D. Avis, A. Evans, Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI, *Neuroimage* 12 (3) (2000) 340 – 356.
- [4] X. Zeng, L. Staib, R. Schultz, J. Duncan, Segmentation and measurement of the cortex from 3-D MR images using coupled-surfaces propagation, *IEEE Trans Med Imaging* 18 (10) (1999) 927 – 937.
- [5] C. Xu, D. Pham, M. Rettman, D. Yu, J. Prince, Reconstruction of the human cerebral cortex from magnetic resonance images, *IEEE Trans Med Imaging* 18 (6) (1999) 467 – 480.
- [6] D. Shattuck, R. Leahy, BrainSuite: An automated cortical surface identification tool, *Med Image Anal* 6 (2002) 129 – 142.
- [7] J. S. Kim, V. Singh, J. K. Lee, J. Lerch, Y. Ad-Dab’bagh, D. MacDonald, J. M. Lee, S. I. Kim, A. C. Evans, Automated 3-d extraction and evaluation of the inner and outer cortical surfaces using a laplacian map and partial volume effect classification., *Neuroimage* 27 (1) (2005) 210–221.
- [8] S. K. Warfield, M. Kaus, F. A. Jolesz, R. Kikinis, Adaptive, template moderated, spatially varying statistical classification., *Med Image Anal* 4 (1) (2000) 43–55.
- [9] B. Fischl, D. H. Salat, A. J. van der Kouwe, N. Makris, F. Segonne, B. T. Quinn, A. M. Dale, Sequence-independent segmentation of magnetic resonance images, *NeuroImage* 23 (2004) S69 – S84.
- [10] J. Ashburner, K. Friston, Unified segmentation, *NeuroImage* 26 (2005) 839 – 851.
- [11] K. Pohl, J. Fisher, W. Grimson, R. Kikinis, W. Wells, A bayesian model for joint segmentation and registration, *Neuroimage* 31 (2006) 228 – 239.
- [12] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, A. Hammers, Automatic anatom-

- ical brain mri segmentation combining label propagation and decision fusion., *Neuroimage* 33 (1) (2006) 115–126.
- [13] J. Tohka, E. Krestyannikov, I. D. Dinov, A. MacKenzie-Graham, D. W. Shattuck, U. Ruotsalainen, A. W. Toga, Genetic algorithms for finite mixture model based voxel classification in neuroimaging, *IEEE Trans Med Imaging* 26 (5) (2007) 696 – 711.
 - [14] B. Belaroussi, J. Milles, S. Carme, Y. Zhu, H. Benoit-Cattin, Intensity non-uniformity correction in mri: Existing methods and their validation, *Med Image Anal* 10 (2006) 234 – 246.
 - [15] J. G. Sled, A. P. Zijdenbos, A. C. Evans, A non-parametric method for automatic correction of intensity non-uniformity in MRI data, *IEEE Trans Med Imaging* 17 (1) (1998) 87 – 97.
 - [16] D. W. Shattuck, S. R. Sandor-Leahy, K. A. Schaper, D. A. Rottenberg, R. M. Leahy, Magnetic resonance image tissue classification using a partial volume model, *Neuroimage* 13 (5) (2001) 856 – 876.
 - [17] B. Likar, M. A. Viergever, F. Pernus, Retrospective correction of mr intensity inhomogeneity by information minimization., *IEEE Trans Med Imaging* 20 (12) (2001) 1398–1410.
 - [18] J. V. Manjon, J. J. Lull, J. Carbonell-Caballero, G. Garcia-Marti, L. Marti-Bonmati, M. Robles, A nonparametric mri inhomogeneity correction method., *Med Image Anal* 11 (4) (2007) 336–345.
 - [19] W. Wells III, W. Grimson, R. Kikinis, F. A. Jolesz, Adaptive segmentation of MRI data, *IEEE Trans Med Imaging* 15 (4) (1996) 429 – 442.
 - [20] K. Van Leemput, F. Maes, D. Vandermeulen, P. Suetens, Automated model-based tissue classification of MR images of the brain, *IEEE Trans Med Imaging* 18 (10) (1999) 897 – 908.
 - [21] K. Van Leemput, F. Maes, D. Vandermeulen, P. Suetens, Automated model-based bias field correction of mr images of the brain, *IEEE Trans Med Imaging* 18 (10) (1999) 885 – 896.
 - [22] J. Sled, G. Pike, Understanding intensity non-uniformity in mri., in: *Medical Image Computing and Computer Assisted Intervention (MICCAI98)*, Lecture notes in computer science 1496, 1998, pp. 615 – 622.
 - [23] J. Marroquin, B. Vemuri, S. Botello, F. Calderon, A. Fernandez-Bouzas, An accurate and efficient bayesian method for automatic segmentation of brain MRI, *IEEE Trans Med Imaging* 21 (8) (2002) 934 – 945.
 - [24] B. Fischl, D. Salat, E. Busa, M. Albert, M. Dietrich, C. Haselgrove, A. van der Kouwe, R. Killany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, A. Dale, Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain, *Neuron* 33 (2002) 341 – 355.
 - [25] U. Amato, M. Larobina, A. Antoniadis, B. Alfano, Segmentation of magnetic resonance brain images through discriminant analysis, *J Neurosci Methods* 131 (2003) 65 – 74.
 - [26] R. Steen, W. Reddick, R. Ogg, More than meets the eye: significant regional heterogeneity in human cortical T₁, *Magn Reson Imaging* 18 (2000) 361 – 368.
 - [27] J. P. Wansapura, S. K. Holland, R. S. Dunn, W. S. Ball, Nmr relaxation times in the human brain at 3.0 tesla., *J Magn Reson Imaging* 9 (4) (1999) 531–538.
 - [28] S. Cho, D. Jones, W. Reddick, R. Ogg, R. Steen, Estabilishing norms for age-related changes in proton t_1 of human brain tissue in vivo, *Magn Reson Imaging* 15 (1997) 1133 – 1143.
 - [29] T. Supprian, E. Hofmann, M. Warmuth-Metz, E. Franzek, T. Becker, MRI t2 relaxation times of brain regions in schizophrenic patients and control subjects., *Psychiatry Res* 75 (3) (1997) 173–182.
 - [30] J. B. M. Warntjes, O. Dahlqvist, P. Lundberg, Novel method for rapid, simultaneous t₁, t₂, and proton density quantification., *Magn Reson Med* 57 (3) (2007) 528–537.
 - [31] T. J. Grabowski, R. J. Frank, N. R. Szumski, C. K. Brown, H. Damasio, Validation of

- partial tissue segmentation of single-channel magnetic resonance images of the brain., *Neuroimage* 12 (6) (2000) 640–656.
- [32] N. Kovacevic, N. J. Lobaugh, M. J. Bronskill, B. Levine, A. Feinstein, S. E. Black, A robust method for extraction and automatic segmentation of brain images., *Neuroimage* 17 (3) (2002) 1087–1100.
 - [33] L. Nocera, J. Gee, Robust partial volume tissue classification of cerebral MRI scans, in: *SPIE Medical Imaging 1997: Image Processing*, 1997, pp. 312 – 322.
 - [34] T. Pappas, An adaptive clustering algorithm for image segmentation, *IEEE Transactions on Signal Processing* 40 (1992) 901–914.
 - [35] M. Yan, J. Karp, Segmentation of 3D brain MR using an adaptive k-means clustering algorithm, in: *1994 Nuclear Science Symposium and Medical Imaging Conference*, 1994, pp. 1529 – 1533.
 - [36] M. Yan, J. Karp, An adaptive bayesian approach to three-dimensional MR brain segmentation, in: *Proceedings of Information Processing in Medical Imaging (IPMI95)*, 1995, pp. 201 – 213.
 - [37] J. C. Rajapakse, J. N. Giedd, J. L. Rapoport, Statistical approach to segmentation of single-channel cerebral MR images., *IEEE Trans Med Imaging* 16 (2) (1997) 176–186.
 - [38] Z. Li, *Markov random fields in computer vision*, Springer Verlag, 1995.
 - [39] J. Besag, On the statistical analysis of dirty pictures, *J R Stat Soc, Ser B* 48 (3) (1986) 259 – 302.
 - [40] M. Kaiser, N. Cressie, The construction of multivariate distributions from markov random fields, *J Multivariate Analysis* 73 (2000) 199 – 220.
 - [41] S. Geman, D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesin restoration of images, *IEEE Trans Patt Anal Mach Intell* 6 (6) (1984) 721 – 741.
 - [42] M. Mega, I. Dinov, P. Thompson, M. Manese, C. Lindshield, J. Moussai, N. Tran, K. Olsen, J. Felix, C. Zoumalan, R. Woods, A. Toga, J. Mazziotta, Automated brain tissue assessment in the elderly and demented population: Construction and validation of a sub-volume probabilistic brain atlas, *NeuroImage* 26 (4) (2005) 1009 – 1018.
 - [43] M. Kamber, R. Shinghal, D. L. Collins, G. S. Francis, A. C. Evans, Model-based 3-D segmentation of multiple sclerosis lesions in magnetic resonance brain images, *IEEE Transactions in Medical Imaging* 14 (3) (1995) 442–453.
 - [44] J. Mazziotta, A.W. Toga, A.C. Evans, et al, A four-dimensional probabilistic atlas of the human brain, *J Am Med Inform Assoc* 8 (2001) 401 – 430.
 - [45] J. Kittler, Combining classifiers: A theoretical framework, *Pattern Analysis and Applications* 1 (1998) 18 – 27.
 - [46] H. S. Choi, D. R. Haynor, Y. Kim, Partial volume tissue classification of multichannel magnetic resonance images - a mixel model, *IEEE Trans Med Imag* 10 (3) (1991) 395 – 407.
 - [47] H. Gudbjartsson, S. Patz, The rician distribution of noisy mri data., *Magn Reson Med* 34 (6) (1995) 910–914.
 - [48] P. Santago, H. D. Gage, Statistical models of partial volume effect, *IEEE Trans Image Process* 4 (11) (1995) 1531 – 40.
 - [49] J. Tohka, A. Zijdenbos, A. Evans, Fast and robust parameter estimation for statistical partial volume models in brain MRI, *NeuroImage* 23 (1) (2004) 84 – 97.
 - [50] G. McLachlan, D. Peel, *Finite Mixture Models*, Wiley, New York, 2000.
 - [51] F. Hu, J. Zidek, The weighted likelihood, *The Canadian Journal of Statistics* 30 (2002) 347 – 371.
 - [52] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, C. Rother, A comparative study of energy minimization methods for markov random fields with smoothness-based priors, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30 (2008) 1068 – 1080.

- [53] D. Jackson, K. Somers, H. Harvey, Similarity coefficients: Measures of co-occurrence and association or simply measures of occurrence?, *The American Naturalist* 133 (3) (1989) 436 – 453.
- [54] D. L. Collins, A. Zijdenbos, V. Kollokian, J. Sled, N. Kabani, C. Holmes, A. C. Evans, Design and construction of a realistic digital brain phantom, *IEEE Trans Med Imaging* 17 (3) (1998) 463 – 468.
- [55] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, M. Joliot, Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain., *Neuroimage* 15 (1) (2002) 273–289.
- [56] I. Dinov, Statistics online computational resource, *Journal of Statistical Software* 16 (2006) 1 – 16.
- [57] A. Zijdenbos, B. M. Dawant, R. A. Margolin, Intensity correction and its effect on measurement variability in the computer-aided analysis of MRI, in: *Proc. of the 9th International Symposium and Exhibition on Computer Assisted Radiology(CAR)*, 1995, pp. 216 – 221.
- [58] D. L. Collins, P. Neelin, T. M. Peters, A. C. Evans, Automatic 3d inter-subject registration of mr volumetric data in standardized talairach space, *J Comput Assist Tomogr* 18 (2) (1994) 192 – 205.
- [59] H. Greenspan, A. Ruf, J. Goldberger, Constrained gaussian mixture model framework for automatic segmentation of mr brain images., *IEEE Trans Med Imaging* 25 (9) (2006) 1233–1245.
- [60] Y. Zhang, M. Brady, S. Smith, Segmentation of brain mr images through a hidden random markov field model and the expectation-maximization algorithm, *IEEE Trans Med Imaging* 20 (1) (2001) 45 – 57.
- [61] O. Tsang, A. Gholipour, N. Kehtarnavaz, K. Gopinath, R. Briggs, I. Panahi, Comparison of tissue segmentation algorithms in neuroimage analysis software tools, in: *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, 2008, pp. 3924–3928.
- [62] S. P. Awate, T. Tasdizen, N. Foster, R. T. Whitaker, Adaptive markov modeling for mutual-information-based, unsupervised mri brain-tissue classification., *Med Image Anal* 10 (5) (2006) 726–739.
- [63] U. Vovk, F. Pernus, B. Likar, A review of methods for correction of intensity inhomogeneity in mri., *IEEE Trans Med Imaging* 26 (3) (2007) 405–421.
- [64] J. Manjon, J. Tohka, G. Garcia-Marti, J. Carbonell-Caballero, J. Lull, L. Marti-Bonmati, M. Robles, Robust MRI brain tissue parameter estimation by multi-stage outlier detection, *Magn Reson Med* 59 (2008) 866 – 873.
- [65] J. Koenderink, The structure of images, *Biol Cybern* 50 (1984) 363 – 370.
- [66] P. A. Bottomley, T. H. Foster, R. E. Argersinger, L. M. Pfeifer, A review of normal tissue hydrogen nmr relaxation times and relaxation mechanisms from 1-100 mhz: dependence on tissue type, nmr frequency, temperature, species, excision, and age., *Med Phys* 11 (4) (1984) 425–448.