A joint Bayesian framework for MR brain scan tissue and structure segmentation based on distributed Markovian agents

Benoit Scherrer[a,c,d], Florence Forbes[b,1], Catherine Garbay[c,d], Michel Dojat[a,d]

[a] *INSERM, U836, Grenoble, France*

[b] *INRIA Grenoble Rhône-Alpes, France*

[c] *Laboratoire d'Informatique de Grenoble, France*

[d] *Universit Joseph Fourier, Institut des Neurosciences, Grenoble, France*

Abstract

In most approaches, tissue and subcortical structure segmentations of MR brain scans are handled globally over the entire brain volume through two relatively independent sequential steps. We propose a fully Bayesian joint model that integrates within a multi-agent framework local tissue and structure segmentations and local intensity distribution modelling. It is based on the specification of three conditional Markov Random Field (MRF) models. The first two encode cooperations between tissue and structure segmentations and integrate *a priori* anatomical knowledge. The third model specifies a Markovian spatial prior over the model parameters that enables local estimations while ensuring their consistency, handling this way nonuniformity of intensity without any bias field modelling. The complete joint model provides then a sound theoretical framework for carrying out tissue and structure segmentations by distributing a set of local agents that estimate cooperatively local MRF models. The evaluation, using a previously affine-registered atlas of 17 structures, was performed using both phantoms and real 3T brain scans. It shows good results and in particular robustness to nonuniformity and noise with a low computational cost. The innovative coupling of agent-based and Markov-centered designs appears as a robust, fast and promising approach to MR brain scan segmentation.

Key words:

1. Introduction

Difficulties in automatic MR brain scan segmentation arise from various sources. The nonuniformity of image intensity results in spatial intensity variations within each tissue, which is a major obstacle to an accurate automatic tissue segmentation. The automatic segmentation of subcortical structures is a

challenging task as well. It cannot be performed based only on intensity distributions and requires the introduction of *a priori* knowledge. Most of the proposed approaches share two main characteristics. First, tissue and subcortical structure segmentations are considered as two successive tasks and treated relatively independently although they are clearly linked: a structure is composed of a specific tissue, and knowledge about structure locations provides valuable information about local intensity distributions. Second, tissue models are estimated globally through the entire volume and then suffer from imperfections at a local level. Alternative local procedures exist but are either used as a preprocessing step (Shattuck et al., 2001a) or use redundant information to ensure consistency of local models (Rajapakse et al., 1997). Recently, good results have been reported using an innovative local and cooperative approach (Scherrer et al., 2007, 2009c). The approach is implemented using a multi-agent framework. It performs tissue and subcortical structure segmentation by distributing through the volume a set of local agents that compute local Markov Random Field (MRF) models which better reflect local intensity distributions. Local MRF models are used alternatively for tissue and structure segmentations and agents cooperate with other agents in their neighborhood for model refinement. Although satisfying in practice, these tissue and structure MRF's do not correspond to a valid joint probabilistic model and are not compatible in that sense. As a consequence, important issues such as convergence or other theoretical properties of the resulting local procedure cannot be addressed. In addition, in (Scherrer et al., 2009c), cooperation mechanisms between local agents are somewhat arbitrary and independent of the MRF models themselves. In this paper, we aim at filling in the gap between an efficient distributed system of agents and a joint modelling accounting for their cooperative processing in a formal manner. Markov models with the concept of conditional independence, whereby each variable is related locally (conditionally) to only a few other variables, are good candidates to complement the symbolic level of the agent-based cooperation with the numerical level inherent to the targeted applications.

Following these considerations, we propose a fully Bayesian framework in which we define a joint model that links local tissue and structure segmentations but also the model parameters. It follows that both types of cooperations, between tissues and structures and between local models, are deduced from the joint model and optimal in that sense. Our model, originally introduced in (Scherrer et al., 2008) and described in details in this chapter, has the following main features: 1) cooperative segmentation of both tissues and structures is encoded via a joint probabilistic model specified through conditional MRF models which capture the relations between tissues and structures. This model specifications also integrate external *a priori* knowledge in a natural way; 2) intensity nonuniformity is handled by using a specific parametrization of tissue intensity distributions which induces local estimations on subvolumes of the entire volume; 3) global consistency between local estimations is automatically ensured by using a MRF spatial prior for the intensity distributions parameters. Estimation within our framework is defined as a maximum *a posteriori* (MAP) estimation problem and is carried out by adopting an instance of the

Expectation Maximization (EM) algorithm (Byrne and Gunawardana, 2005). We show that such a setting can adapt well to our conditional models formulation and simplifies into alternating and cooperative estimation procedures for standard Hidden MRF models that can be implemented efficiently via a two agent-layer architecture.

The chapter is organized as follows. In Section 2, we explain the motivation in coupling agent-based and Markov-centered designs. In Section 3, we introduce the probabilistic setting and inference framework. The joint tissue and structure model is described in more details in Section 4. An appropriate estimation procedure is proposed in Section 5. Experimental results are reported in Section 6 and a discussion ends the chapter.

2. Distributed cooperative Markovian agents

While Markov modelling has largely been used in the domain of MRI segmentation, agent-based approaches have seldom been considered. Agents are autonomous entities sharing a common environment and working in a cooperative way to achieve a common goal. They are usually provided with limited perception abilities and local knowledge. Some advantages of multi-agent systems are among others (Shariatpanahi et al., 2006) their ability: to handle knowledge from different domains, to design reliable systems able to recover from agents with low performance and wrong knowledge, to focus spatially and semantically on relevant knowledge, to cooperate and share tasks between agents in various domains, and to reduce computation time through distributed and asynchronous implementation.

Previous work has shown the potential of multi-agent approaches for MRI segmentation along two main directions : first of all as a way to cope with grey level heterogeneity and bias field effects, by enabling the development of local and situated processing styles (Richard et al., 2007) and secondly, as a way to support the cooperation between various processing styles and information types, namely tissue and structure information (Germond et al., 2000; Scherrer et al., 2009a). Stated differently, multi-agent modelling may be seen as a robust approach to identify the main lines along which to distribute complex processing issues, and support the design of situated cooperative systems at the symbolic level as illustrated in Figure 1. A designing approach has been furthermore proposed (Scherrer et al., 2009a) to benefit from both Markov-centered and agent-based modelling toward robust MRI processing systems. In the course of this work, however, we pointed out the gap existing between the symbolical level of the agent-based cooperation and the numerical level of Markov optimization, this gap resulting in a difficulty to ground formally the proposed design. In particular the mutual dependencies between the Markov variables (tissue and structure segmentations on one hand, local intensity models on the other hand) were handled through rather *ad hoc* cooperation mechanisms.

The point here is then to start from the observation that Markov graphical modelling may be used to visualize the dependencies between local intensity models, and tissue and structure segmentations, as shown in Figure 2. According to this Figure, the observed intensities $\mathbf{y}$ are seen as reflecting both tissue-dependent and structure dependent information. Tissues and structures are considered as mutually dependent, since structures are composed of tissues. Also, the observed variations of appearance in tissues and structures reflect the spatial dependency of the tissue model parameters to be computed. Figure 2 then illustrates the hierarchical organization of the variables under consideration and its adequation with the agent hierarchy at a symbolic level. In the following section, we show that this hierarchical decomposition can be expressed in terms of a coherent systems of probability distributions for which inference can be carried out. Regarding implementation, we adopt subsequently the two agent-layer architecture, as illustrated in Figure 3, where tissue and structure agents cooperate through shared information including tissue intensity models, anatomical atlas, tissue and structure segmentations.

3. Hierarchical analysis using the EM algorithm

Hierarchical modelling is, in essence, based on the simple fact from probability that the joint distribution of a collection of random variables can be decomposed into a series of conditional models. That is, if $\mathbf{Y}$, $\mathbf{Z}$, θ are random variables, then we write the joint distribution in terms of a factorization such as $p(\mathbf{y}, \mathbf{z}, \theta) = p(\mathbf{y}|\mathbf{z}, \theta)p(\mathbf{z}|\theta)p(\theta)$. The strength of hierarchical approaches is that they are based on the specification of coherently linked system of conditional models. The key elements of such models can be considered in three stages, the data stage, process stage and parameter stage. In each stage, complicated dependence structure is mitigated by conditioning. For example, the data stage can incorporate measurement errors as well as multiple datasets. The process and parameter stages can allow spatial interactions as well as the direct inclusion of scientific knowledge. These modelling capabilities are especially relevant to tackle the task of MRI brain scan segmentation. In image segmentation problems, the question of interest is to recover an unknown image $\mathbf{z} \in \mathcal{Z}$, interpreted as a classification into a finite number K of labels, from an image $\mathbf{y}$ of observed intensity values. This classification usually requires values for a vector parameter $\theta \in \underline{\Theta}$ considered in a Bayesian setting as a random variable. The idea is to approach the problem by breaking it into the three primary stages mentioned above. The first data stage is concerned with the observational process or data model $p(\mathbf{y}|\mathbf{z}, \theta)$, which specifies the distribution of the data $\mathbf{y}$ given the process of interest and relevant parameters. The second stage then describes the process model $p(\mathbf{z}|\theta)$, conditional on usually other parameters still denoted by θ for simplicity. Finally, the last stage accounts for the uncertainty in the parameters through a distribution $p(\theta)$. In applications, each of these stages may have multiple sub-stages. For example, if spatial interactions are to be taken into account, it might be modelled as a product of several conditional distributions suggested by neighborhood

relationships. Similar decompositions are possible in the parameter stage.

Ultimately, we are interested in the distribution of the process and parameters updated by the data, that is the so-called posterior distribution $p(\mathbf{z}, \theta \mid \mathbf{y})$. Due to generally too complex dependencies, it is difficult to extract parameters θ from the observed data $\mathbf{y}$ without explicit knowledge of the unknown true segmentation $\mathbf{z}$. This problem is greatly simplified when the solution is determined within an EM framework. The EM algorithm (McLachlan and Krishnan, 1996) is a general technique for finding maximum likelihood solutions in the presence of missing data. It consists in two steps usually described as the E-step in which the expectation of the so-called complete log-likelihood is computed and the M-step in which this expectation is maximized over θ. An equivalent way to define EM is the following. Let $\mathcal{D}$ be the set of all probability distributions on $\mathcal{Z}$. As discussed in (Byrne and Gunawardana, 2005), EM can be viewed as an alternating maximization procedure of a function F defined, for any probability distribution $q \in \mathcal{D}$, by

$$F(q, \theta) \quad = \quad \sum_{\mathbf{z} \in \mathcal{Z}} \ln p(\mathbf{y}, \mathbf{z} \mid \theta)\, q(\mathbf{z}) + I[q], \tag{1}$$

where $I[q] = -E_q[\log q(\mathbf{Z})]$ is the entropy of q (E_q denotes the expectation with regard to q and we use capital letters to indicate random variables while their realizations are denoted with small letters). When prior knowledge on the parameters is available, the Bayesian setting consists in replacing the maximum likelihood estimation by a maximum a posteriori (MAP) estimation of θ using the prior knowledge encoded in distribution $p(\theta)$. The maximum likelihood estimate of θ *ie.* $\hat{\theta} = \arg\max_{\theta \in \underline{\Theta}} p(\mathbf{y}|\theta)$ is replaced by $\hat{\theta} = \arg\max_{\theta \in \underline{\Theta}} p(\theta|\mathbf{y})$. The EM algorithm can also be used to maximize the posterior distribution. Indeed, the likelihood $p(\mathbf{y}|\theta)$ and $F(q, \theta)$ are linked through $\log p(\mathbf{y}|\theta) = F(q, \theta) + KL(q, p)$ where $KL(q, p)$ is the Kullback-Leibler divergence between q and the conditional distribution $p(\mathbf{z}|\mathbf{y}, \theta)$ and is non-negative,

$$KL(q, p) = \sum_{\mathbf{z} \in \mathcal{Z}} q(\mathbf{z}) \log \left(\frac{q(\mathbf{z})}{p(\mathbf{z}|\mathbf{y}, \theta)} \right) .$$

Using the equality $\log p(\theta|\mathbf{y}) = \log p(\mathbf{y}|\theta) + \log p(\theta) - \log p(\mathbf{y})$ it follows $\log p(\theta|\mathbf{y}) = F(q, \theta) + KL(q, p) + \log p(\theta) - \log p(\mathbf{y})$ from which, we get a lower bound $\mathcal{L}(q, \theta)$ on $\log p(\theta|\mathbf{y})$ given by $\mathcal{L}(q, \theta) = F(q, \theta) + \log p(\theta) - \log p(\mathbf{y})$. Maximizing this lower bound alternatively over q and θ leads to a sequence $\{q^{(r)}, \theta^{(r)}\}_{r \in \mathbb{N}}$ satisfying $\mathcal{L}(q^{(r+1)}, \theta^{(r+1)}) \geq \mathcal{L}(q^{(r)}, \theta^{(r)})$. The maximization over q corresponds to the standard E-step and leads to $q^{(r)}(\mathbf{z}) = p(\mathbf{z}|\mathbf{y}, \theta^{(r)})$. It follows that $\mathcal{L}(q^{(r)}, \theta^{(r)}) = \log p(\theta^{(r)}|\mathbf{y})$ which means that the lower bound reaches the objective function in $\theta^{(r)}$ and that the sequence $\{\theta^{(r)}\}_{r \in \mathbb{N}}$ increases $p(\theta|\mathbf{y})$ at each step. It then appears that when considering our MAP problem, we replace (see *eg.* (Gelman et al., 2004)) the function $F(q, \theta)$ by $F(q, \theta) + \log p(\theta)$. The corresponding alternating procedure is: starting from a current value $\theta^{(r)} \in \underline{\Theta}$, set alternatively

$$q^{(r)} = \arg\max_{q \in \mathcal{D}} \; F(q, \theta^{(r)}) = \arg\max_{q \in \mathcal{D}} \sum_{\mathbf{z} \in \mathcal{Z}} \log p(\mathbf{z}|\mathbf{y}, \theta^{(r)})\, q(\mathbf{z}) + I[q], \tag{2}$$

and

$$\begin{aligned}\theta^{(r+1)} &= \arg\max_{\theta\in\underline{\Theta}} \; F(q^{(r)},\theta) + \log p(\theta) \\ &= \arg\max_{\theta\in\underline{\Theta}} \; \sum_{\mathbf{z}\in\mathcal{Z}} \log p(\mathbf{y},\mathbf{z},\theta)\; q^{(r)}(\mathbf{z}) + \log p(\theta) \\ &= \arg\max_{\theta\in\underline{\Theta}} \; \sum_{\mathbf{z}\in\mathcal{Z}} \log p(\theta|\mathbf{y},\mathbf{z})\; q^{(r)}(\mathbf{z}) \; . \end{aligned} \tag{3}$$

The last equality in (2) comes from $p(\mathbf{y},\mathbf{z}|\theta) = p(\mathbf{z}|\mathbf{y},\theta)p(\mathbf{y}|\theta)$ and the fact that $p(\mathbf{y}|\theta)$ does not depend on $\mathbf{z}$. The last equality in (3) comes from $p(\mathbf{y},\mathbf{z}|\theta) = p(\theta|\mathbf{y},\mathbf{z})\; p(\mathbf{y},\mathbf{z})/p(\theta)$ and the fact that $p(\mathbf{y},\mathbf{z})$ does not depend on θ. The optimization with respect to q gives rise to the same E-step as for the standard EM algorithm, because q only appears in $F(q,\theta)$. It can be shown (eg. (Gelman et al., 2004) p.319) that EM converges to a local mode of the posterior density except in some very special cases. This EM framework appears as a reasonable framework for inference. In addition, it appears in (2) and (3) that inference can be described in terms of the conditional models $p(\mathbf{z}|\mathbf{y},\theta)$ and $p(\theta|\mathbf{y},\mathbf{z})$. In the following section, we show how to define our joint model so as to take advantage of these considerations.

4. A Bayesian model for robust joint tissue and structure segmentations

In this section, we describes the Bayesian framework that enables us to model the relationships between the unknown linked tissue and structure labels, the observed MR image data and the tissue intensity distributions parameters.

We consider a finite set V of N voxels on a regular 3D grid. We denote by $\mathbf{y} = \{y_1,\ldots,y_N\}$ the intensity values observed respectively at each voxel and by $\mathbf{t} = \{t_1,\ldots,t_N\}$ the hidden tissue classes. The t_i's take their values in $\{e_1,e_2,e_3\}$ where e_k is a 3-dimensional binary vector whose k^{th} component is 1, all other components being 0. In addition, we consider L subcortical structures and denote by $\mathbf{s} = \{s_1,\ldots,s_N\}$ the hidden structure classes at each voxel. Similarly, the s_i's take their values in $\{e'_1,\ldots,e'_L,e'_{L+1}\}$ where e'_{L+1} corresponds to an additional background class. As parameters θ, we consider the parameters describing the intensity distributions for the $K = 3$ tissue classes. They are denoted by $\theta = \{\theta_i^k, i\in V, k = 1\ldots K\}$. We write for all $k = 1\ldots K$, $\theta^k = \{\theta_i^k, i\in V\}$ and for all $i\in V$, $\theta_i = {}^t(\theta_i^k, k = 1\ldots K)$ (^{t} means transpose). Note that we describe here the most general setting in which the intensity distributions can depend on voxel i and vary with its location. Standard approaches usually consider that intensity distributions are Gaussian distributions for which the parameters depend only on the tissue class. Although the Bayesian approach makes the general case possible, in practice we consider θ_i^k's equal for all voxels i in some prescribed regions. More specifically, our local approach consists in

dividing the volume V into a partition of subvolumes and consider the θ_i^k constant over each subvolume (see Section 4.2).

To explicitly take into account the fact that tissue and structure classes are related, a *generative* approach would be to define a complete probabilistic model, namely $p(\mathbf{y}, \mathbf{t}, \mathbf{s}, \theta)$. To define such a joint probability is equivalent to define the two probability distributions $p(\mathbf{y})$ and $p(\mathbf{t}, \mathbf{s}, \theta|\mathbf{y})$. However, in this work, we rather adopt a *discriminative* approach in which a conditional model $p(\mathbf{t}, \mathbf{s}, \theta|\mathbf{y})$ is constructed from the observations and labels but the marginal $p(\mathbf{y})$ is not modelled explicitly. In a segmentation context, the full generative model is not particularly relevant to the task of inferring the class labels. This appears clearly in equations (2) and (3) where the relevant distributions are conditional. In addition, it has been observed that conditional approaches tend to be more robust than generative models (Lafferty et al., 2001; Minka, 2005). Therefore, we focus on $p(\mathbf{t}, \mathbf{s}, \theta|\mathbf{y})$ as the quantity of interest. It is fully specified when the two conditional distributions $p(\mathbf{t}, \mathbf{s}|\mathbf{y}, \theta)$ and $p(\theta|\mathbf{y}, \mathbf{t}, \mathbf{s})$ are defined. The following subsections 4.1 and 4.2 specify respectively these two distributions.

4.1. A conditional model for tissues and structures

The distribution $p(\mathbf{t}, \mathbf{s}|\mathbf{y}, \theta)$ can be in turn specified by defining $p(\mathbf{t}|\mathbf{s}, \mathbf{y}, \theta)$ and $p(\mathbf{s}|\mathbf{t}, \mathbf{y}, \theta)$. The advantage of the later conditional models is that they can capture in an explicit way the effect of tissue segmentation on structure segmentation and vice versa. Note that on a computational point of view there is no need at this stage to describe explicitly the joint model that can be quite complex. In what follows, notation ${}^t xx'$ denotes the scalar product between two vectors x and x'. Notation $U_{ij}^T(t_i, t_j; \eta_T)$ and $U_{ij}^S(s_i, s_j; \eta_S)$ denotes pairwise potential functions with interaction parameters η_T and η_S. Simple examples for $U_{ij}^T(t_i, t_j; \eta_T)$ and $U_{ij}^S(s_i, s_j; \eta_S)$ are provided by adopting a Potts model which corresponds to

$$U_{ij}^T(t_i, t_j; \eta_T) = \eta_T \, {}^t t_i t_j \quad \text{and} \quad U_{ij}^S(s_i, s_j; \eta_S) = \eta_S \, {}^t s_i s_j \; .$$

Such a model captures, within each label set $\mathbf{t}$ and $\mathbf{s}$, interactions between neighboring voxels. It implies spatial interaction within each label set.

Structure conditional Tissue model. We define $p(\mathbf{t}|\mathbf{s}, \mathbf{y}, \theta)$ as a Markov Random Field in $\mathbf{t}$ with the following energy function,

$$H_{T|S,Y,\Theta}(\mathbf{t}|\mathbf{s}, \mathbf{y}, \theta) \quad = \quad \sum_{i \in V} \left({}^t t_i \gamma_i(s_i) + \sum_{j \in \mathcal{N}(i)} U_{ij}^T(t_i, t_j; \eta_T) + \log g_T(y_i; {}^t \theta_i t_i) \right), \tag{4}$$

where $\mathcal{N}(i)$ denotes the voxels neighboring i, $g_T(y_i; {}^t\theta_i t_i)$ is the Gaussian distribution with parameters θ_i^k if $t_i = e_k$ and the external field γ_i depends on s_i and is defined by $\gamma_i(s_i) = e_{T^{s_i}}$ if $s_i \in \{e'_1, \ldots, e'_L\}$ and

$\gamma_i(s_i) = \mathbf{0}$ otherwise, with T^{s_i} denoting the tissue of structure s_i and $\mathbf{0}$ the 3-dimensional null vector. The rationale for choosing such an external field, is that depending on the structure present at voxel i and given by the value of s_i, the tissue corresponding to this structure is more likely at voxel i while the two others tissues are penalized by a smaller contribution to the energy through a smaller external field value. When i is a background voxel, the external field does not favor a particular tissue. The Gaussian parameters $\theta_i^k = \{\mu_i^k, \lambda_i^k\}$ are respectively the mean and precision which is the inverse of the variance. We use similar notation such as $\mu = \{\mu_i^k, i \in V, k = 1 \dots K\}$ and $\mu^k = \{\mu_i^k, i \in V\}$, etc.

Tissue conditional structure model. *A priori* knowledge on structures is incorporated through a field $f = \{f_i, i \in V\}$ where $f_i = {}^t(f_i(e'_l), l = 1 \dots L+1)$ and $f_i(e'_l)$ represents some prior probability that voxel i belongs to structure l , as provided by a registered probabilistic atlas. We then define $p(\mathbf{s}|\mathbf{t}, \mathbf{y}, \theta)$ as a Markov Random Field in $\mathbf{s}$ with the following energy function,

$$H_{S|T,Y,\Theta}(\mathbf{s}|\mathbf{t}, \mathbf{y}, \theta) = \sum_{i \in V} \left({}^t s_i \log f_i + \sum_{j \in \mathcal{N}(i)} U_{ij}^S(s_i, s_j; \eta_S) + \log g_S(y_i|t_i, s_i, \theta_i) \right) \quad (5)$$

where $g_S(y_i|t_i, s_i, \theta_i)$ is defined as follows,

$$g_S(y_i|t_i, s_i, \theta) = [g_T(y_i; {}^t\theta_i e_{T^{s_i}})\ f_i(s_i)]^{w(s_i)}\ [g_T(y_i; {}^t\theta_i t_i)\ f_i(e'_{L+1})]^{(1-w(s_i))} \quad (6)$$

where $w(s_i)$ is a weight dealing with the possible conflict between values of t_i and s_i. For simplicity we set $w(s_i) = 0$ if $s_i = e'_{L+1}$ and $w(s_i) = 1$ otherwise but considering more general weights could be an interesting refinement. Other parameters in (4) and (5) include interaction parameters η_T and η_S which are considered here as hyperparameters to be specified (see Section 6).

4.2. A conditional model for intensity distribution parameters

To ensure spatial consistency between the parameter values, we define also $p(\theta|\mathbf{y}, \mathbf{t}, \mathbf{s})$ as a MRF. In practice however, in our general setting which allows different values θ_i at each i, there are too many parameters and estimating them accurately is not possible. As regards estimation then, we adopt a local approach as in Scherrer et al. (2009c). The idea is to consider the parameters as constant over subvolumes of the entire volume. Let $\mathcal{C}$ be a regular cubic partionning of the volume V in a number of non-overlapping subvolumes $\{V_c, c \in \mathcal{C}\}$. We assume that for all $c \in \mathcal{C}$ and all $i \in V_c$, $\theta_i = \theta_c$ and consider a pairwise MRF on $\mathcal{C}$ with energy function denoted by $H_\Theta^{\mathcal{C}}(\theta)$ where by extension θ denotes the set of distinct values $\theta = \{\theta_c, c \in \mathcal{C}\}$. Outside the issue of estimating θ in the M-step, having parameters θ_i's depending on i is not a problem. For the E-steps, we go back to this general setting using an interpolation step specified in Section 5.2. It follows that $p(\theta|\mathbf{y}, \mathbf{t}, \mathbf{s})$ is defined as a MRF with the following energy function,

$$H_{\Theta|Y,T,S}(\theta|\mathbf{y}, \mathbf{t}, \mathbf{s}) = H_\Theta^{\mathcal{C}}(\theta) + \sum_{c \in \mathcal{C}} \log \prod_{i \in V_c} g_S(y_i|t_i, s_i, \theta_c),$$

where $g_S(y_i|t_i, s_i, \theta_c)$ is the expression in (6). The specific form of the Markov prior on θ is specified in Section 5.

5. Estimation by Generalized Alternating Minimization

The particularity of our segmentation task is to include two label sets of interest, $\mathbf{t}$ and $\mathbf{s}$ which are linked and that we would like to estimate cooperatively using one to gain information on the other. We denote respectively by $\mathcal{T}$ and $\mathcal{S}$ the spaces in which $\mathbf{t}$ and $\mathbf{s}$ take their values. Denoting $\mathbf{z} = (\mathbf{t}, \mathbf{s})$, we apply the EM framework introduced in Section 3 to find a MAP estimate $\hat{\theta}$ of θ using the procedure given by (2) and (3) and then generate $\mathbf{t}$ and $\mathbf{s}$ that maximize the conditional distribution $p(\mathbf{t}, \mathbf{s}|\mathbf{y}, \hat{\theta})$. Note that this is however not equivalent to maximizing over $\mathbf{t}, \mathbf{s}$ and θ the posterior distribution $p(\mathbf{t}, \mathbf{s}, \theta|\mathbf{y})$. Indeed $p(\mathbf{t}, \mathbf{s}, \theta \mid \mathbf{y}) = p(\mathbf{t}, \mathbf{s}|\mathbf{y}, \theta)\ p(\theta|\mathbf{y})$ and in the EM setting, θ is found by maximizing the second factor only.

However, solving the optimization (2) over the set $\mathcal{D}$ of probability distributions $q_{(T,S)}$ on $\mathcal{T} \times \mathcal{S}$ leads for the optimal $q^{(r)}_{(T,S)}$ to $p(\mathbf{t}, \mathbf{s}|\mathbf{y}, \theta^{(r)})$ which is intractable for the joint model defined in Section 4. We therefore propose an EM variant appropriate to our cooperative context and in which the E-step is not performed exactly. The optimization (2) is solved instead over a restricted class of probability distributions $\tilde{\mathcal{D}}$ which is chosen as the set of distributions that factorize as $q_{(T,S)}(\mathbf{t}, \mathbf{s}) = q_T(\mathbf{t})\ q_S(\mathbf{s})$ where q_T (resp. q_S) belongs to the set $\mathcal{D}_T$ (resp. $\mathcal{D}_S$) of probability distributions on $\mathcal{T}$ (resp. on $\mathcal{S}$). This variant is usually referred to as Variational EM (Jordan et al., 1999). It follows that the E-step becomes an approximate E-step,

$$(q_T^{(r)}, q_S^{(r)}) = \arg \max_{(q_T, q_S)} F(q_T q_S, \theta^{(r)})\ .$$

This step can be further generalized by decomposing it into two stages. At iteration r, with current estimates denoted by $q_T^{(r-1)}, q_S^{(r-1)}$ and $\theta^{(r)}$, we consider the following updating,

E-T-step: $q_T^{(r)} = \arg \max_{q_T \in \mathcal{D}_T} F(q_T\ q_S^{(r-1)}, \theta^{(r)})$

E-S-step: $q_S^{(r)} = \arg \max_{q_S \in \mathcal{D}_S} F(q_T^{(r)}\ q_S, \theta^{(r)})$.

The effect of these iterations is to generate sequences of paired distributions and parameters $\{q_T^{(r)}, q_S^{(r)}, \theta^{(r)}\}_{r \in \mathbb{N}}$ that satisfy $F(q_T^{(r+1)} q_S^{(r+1)}, \theta^{(r+1)}) \geq F(q_T^{(r)} q_S^{(r)}, \theta^{(r)})$. This variant falls in the modified Generalized Alternating Minimization (GAM) procedures family for which convergence results are available (Byrne and Gunawardana, 2005).

We then derive two equivalent expressions of F when q factorizes as in $\tilde{\mathcal{D}}$. Expression (1) of F can be rewritten as $F(q,\theta) = E_q[\log p(\mathbf{T}|\mathbf{S},\mathbf{y},\theta)] + E_q[\log p(\mathbf{S},\mathbf{y}|\theta)] + I[q]$. Then,

$$\begin{aligned} F(q_T\ q_S,\theta) &= E_{q_T}[E_{q_S}[\log p(\mathbf{T}|\mathbf{S},\mathbf{y},\theta)]] + E_{q_S}[\log p(\mathbf{S},\mathbf{y}|\theta)] + I[q_T\ q_S] \\ &= E_{q_T}[E_{q_S}[\log p(\mathbf{T}|\mathbf{S},\mathbf{y},\theta)]] + I[q_T] + G[q_S]\ , \end{aligned}$$

where $G[q_S] = E_{q_S}[\log p(\mathbf{S},\mathbf{y}|\theta)] + I[q_S]$ is an expression that does not depend on q_T. Using the symmetry in $\mathbf{T}$ and $\mathbf{S}$, it is easy to show that similarly,

$$\begin{aligned} F(q_T\ q_S,\theta) &= E_{q_S}[E_{q_T}[\log p(\mathbf{S}|\mathbf{T},\mathbf{y},\theta)]] + E_{q_T}[\log p(\mathbf{T},\mathbf{y}|\theta)] + I[q_T\ q_S] \\ &= E_{q_S}[E_{q_T}[\log p(\mathbf{S}|\mathbf{T},\mathbf{y},\theta)]] + I[q_S] + G'[q_T]\ , \end{aligned}$$

where $G'[q_T] = E_{q_T}[\log p(\mathbf{T},\mathbf{y}|\theta)] + I[q_T]$ is an expression that does not depend on q_S. It follows that the E-T and E-S steps reduce to,

$$\textbf{E-T-step: } q_T^{(r)} = \arg\max_{q_T\in\mathcal{D}_T} E_{q_T}[E_{q_S^{(r-1)}}[\log p(\mathbf{T}|\mathbf{S},\mathbf{y},\theta^{(r)})]] + I[q_T] \tag{7}$$

$$\textbf{E-S-step: } q_S^{(r)} = \arg\max_{q_S\in\mathcal{D}_S} E_{q_S}[E_{q_T^{(r)}}[\log p(\mathbf{S}|\mathbf{T},\mathbf{y},\theta^{(r)})]] + I[q_S] \tag{8}$$

and the **M-step**

$$\theta^{(r+1)} = \arg\max_{\theta\in\underline{\Theta}}\ E_{q_T^{(r)}q_S^{(r)}}[\log p(\theta|\mathbf{y},\mathbf{T},\mathbf{S})]\ . \tag{9}$$

More generally, we adopt in addition, an incremental EM approach (Byrne and Gunawardana, 2005) which allows re-estimation of the parameters (here θ) to be performed based only on a sub-part of the hidden variables. This means that we incorporate an M-step (9) in between the updating of q_T and q_S. Similarly, hyperparameters could be updated there too.

It appears in equations (7), (8) and (9) that for inference the specification of the three conditional distributions $p(\mathbf{t}|\mathbf{s},\mathbf{y},\theta)$, $p(\mathbf{s}|\mathbf{t},\mathbf{y},\theta)$ and $p(\theta|\mathbf{t},\mathbf{s},\mathbf{y})$ is necessary and sufficient.

5.1. Structure and tissue conditional E-steps

Then, steps E-T and E-S have to be further specified by computing the expectations with regards to $q_S^{(r-1)}$ and $q_T^{(r)}$. Using the structure conditional model definition (4), it comes,

$$\begin{aligned} E_{q_S^{(r-1)}}[\log p(\mathbf{T}|\mathbf{S},\mathbf{y},\theta^{(r)})] &= -E_{q_S^{(r-1)}}[\log W_T] \\ &\quad + \sum_{i\in V} {}^tT_i E_{q_S^{(r-1)}}[\gamma_i(S_i)] + \sum_{j\in\mathcal{N}(i)} U_{ij}^T(T_i,T_j;\eta_T) + \log g_T(y_i;{}^t\theta_i^{(r)}T_i)\ , \end{aligned}$$

where W_T is a normalizing constant that does not depend on $\mathbf{T}$ and can be omitted in the maximization of step E-T. The external field term leads to

$$
\begin{aligned}
E_{q_S^{(r-1)}}[\gamma_i(S_i)] &= \sum_{l=1}^{L} e_{T^l}\ q_{S_i}^{(r-1)}(e'_l) \\
&= {}^t\Big(\sum_{l\ st.T^l=1} q_{S_i}^{(r-1)}(e'_l),\ \sum_{l\ st.T^l=2} q_{S_i}^{(r-1)}(e'_l),\ \sum_{l\ st.T^l=3} q_{S_i}^{(r-1)}(e'_l)\Big)\ .
\end{aligned}
$$

The k^{th} $(k = 1 \dots 3)$ component of the above vector represents the probability that voxel i belongs to a structure whose tissue class is k. The stronger this probability the more a priori favored is tissue k. Eventually, we notice that step E-T is equivalent to the E-step one would get when applying EM to a standard Hidden MRF over $\mathbf{t}$ with Gaussian class distributions and an external field parameter fixed to values based on the current structure segmentation. To solve this step, then, various inference techniques for Hidden MRF's can be applied. In this paper, we adopt Mean field like algorithms (Celeux et al., 2003) used in (Scherrer et al., 2009c) for MRI brain scans. This class of algorithms has the advantage to turn the initial intractable model into a model equivalent to a system of independent variables for which the exact EM can be carried out. Following a mean field principle, when spatial interactions are defined via Potts models, these algorithms are based on the approximation of $U_{ij}^T(t_i, t_j; \eta_T)$ by $\tilde{U}_{ij}^T(t_i, t_j; \eta_T) = \eta_T\ {}^t t_i \tilde{t}_j$ where $\tilde{\mathbf{t}}$ is a particular configuration of $\mathbf{T}$ which is updated at each iteration according to a specific scheme. We refer to Celeux et al. (2003) for details on three possible schemes to update $\tilde{\mathbf{t}}$.

Similarly, using definitions (5) and (6),

$$
\begin{aligned}
E_{q_T^{(r)}}[\log p(\mathbf{S}|\mathbf{T}, \mathbf{y}, \theta^{(r)})] &= -E_{q_T^{(r)}}[\log W_S] \\
&\quad + \sum_{i \in V} {}^t S_i f_i + \sum_{j \in \mathcal{N}(i)} U_{ij}^S(S_i, S_j; \eta_S) + E_{q_T^{(r)}}[\log g_S(y_i|T_i, S_i, \theta_i^{(r)}]
\end{aligned}
$$

where the normalizing constant W_S does not depend on S and can be omitted in the maximization in step E-S. The last term can be further computed,

$$
E_{q_T^{(r)}}[\log g_S(y_i|T_i, S_i, \theta_i^{(r)})] = \log g'_S(y_i|S_i, \theta_i^{(r)}),
$$

where

$$
g'_S(y_i|s_i, \theta_i) = [g_T(y_i; {}^t\theta_i e_{T^{s_i}})\ f_i(s_i)]^{w(s_i)}\ [(\prod_{k=1}^{3} g_T(y_i; \theta_i^k)^{q_{T_i}^{(r)}(e_k)}) f_i(e'_{L+1})]^{(1-w(s_i))}\ .
$$

In this later expression, the product corresponds to a Gaussian distribution with mean $\sum_{k=1}^{3} \mu_i^k\ \lambda_i^k\ q_{T_i}^{(r)}(e_k) / \sum_{k=1}^{3} \lambda_i^k\ q_{T_i}^{(r)}(e_k)$ and precision $\sum_{k=1}^{3} \lambda_i^k\ q_{T_i}^{(r)}(e_k)$.

It follows that step E-S can be seen as the E-step for a standard Hidden MRF with class distributions defined by g'_S and an external field incorporating prior structure knowledge through f. As already mentioned, it can be solved using techniques such as those described in Celeux et al. (2003).

5.2. Updating the tissue intensity distribution parameters

As mentioned in Section 4.2, we now consider that the θ_i's are constant over subvolumes of a given partition of the entire volume. The MRF prior on $\theta = \{\theta_c, c \in \mathcal{C}\}$ is $p(\theta) \propto \exp(H_\Theta^{\mathcal{C}}(\theta))$ and (9) can be written as,

$$\theta^{(r+1)} = \arg\max_{\theta \in \underline{\Theta}} p(\theta) \prod_{i \in V} \prod_{k=1}^{K} g_T(y_i; \theta_i^k)^{a_{ik}} = \arg\max_{\theta \in \underline{\Theta}} p(\theta) \prod_{c \in \mathcal{C}} \prod_{k=1}^{K} \prod_{i \in V_c} g_T(y_i; \theta_c^k)^{a_{ik}} ,$$

where $a_{ik} = q_{T_i}^{(r)}(e_k) q_{S_i}^{(r)}(e'_{L+1}) + \sum_{l \ st.T^l = e_k} q_{S_i}^{(r)}(e_l)$. The second term in a_{ik} is the probability that voxel i belongs to one of the structures made of tissue k. The a_{ik}'s sum to one (over k) and a_{ik} can be interpreted as the probability for voxel i to belong to the tissue class k when both tissue and structure segmentations information are combined. Using the additional natural assumption that $p(\theta) = \prod_{k=1}^{K} p(\theta^k)$, it is equivalent to solve for each $k = 1 \dots K$,

$$\theta^{k\,(r+1)} = \arg\max_{\theta^k \in \underline{\Theta}^k} p(\theta^k) \prod_{c \in \mathcal{C}} \prod_{i \in V_c} g_T(y_i; \theta_c^k)^{a_{ik}}. \tag{10}$$

However, when $p(\theta^k)$ is chosen as a Markov field, the exact maximization (10) is still intractable. We therefore replace $p(\theta^k)$ by a product form given by its *modal-field* approximation (Celeux et al., 2003). This is actually equivalent to use the ICM (Besag, 1986) algorithm. Assuming a current estimation $\theta^{k\,(\nu)}$ of θ^k at iteration ν, we consider in turn,

$$\forall c \in \mathcal{C}, \quad \theta_c^{k\,(\nu+1)} = \arg\max_{\theta_c^k \in \underline{\Theta}^k} p(\theta_c^k \mid \theta_{\mathcal{N}(c)}^{k\,(\nu)}) \prod_{i \in V_c} g_T(y_i; \theta_c^k)^{a_{ik}} , \tag{11}$$

where $\mathcal{N}(c)$ denotes the indices of the subvolumes that are neighbors of subvolume c and $\theta_{\mathcal{N}(c)}^k = \{\theta_{c'}^k, c' \in \mathcal{N}(c)\}$. At convergence, the obtained values give the updated estimation $\theta^{k\,(r+1)}$.

The particular form (11) above guides the specification of the prior for θ. Indeed, Bayesian analysis indicates that a natural choice for $p(\theta_c^k \mid \theta_{\mathcal{N}(c)}^k)$ has to be among conjugate or semi-conjugate priors for the Gaussian distribution $g_T(y_i; \theta_c^k)$. We choose to consider here the latter case. In addition, we assume that the Markovian dependence applies only to the mean parameters and consider that $p(\theta_c^k \mid \theta_{\mathcal{N}(c)}^k) = p(\mu_c^k \mid \mu_{\mathcal{N}(c)}^k)\, p(\lambda_c^k)$, with $p(\mu_c^k \mid \mu_{\mathcal{N}(c)}^k)$ set to a Gaussian distribution with mean $m_c^k + \sum_{c' \in \mathcal{N}(c)} \eta_{cc'}^k (\mu_{c'}^k - m_{c'}^k)$ and precision λ_c^{0k}, and $p(\lambda_c^k)$ set to a Gamma distribution with shape parameter α_c^k and scale parameter b_c^k. The quantities $\{m_c^k, \lambda_c^{0k}, \alpha_c^k, b_c^k, c \in \mathcal{C}\}$ and $\{\eta_{cc'}^k, c' \in \mathcal{N}(c)\}$ are hyperparameters to be specified. For this choice, we get valid joint Markov models for the μ^k's (and therefore for the θ^k's) which are known as auto-normal (Besag, 1974) models. Whereas for the standard Normal-Gamma conjugate prior the resulting conditional densities fail in defining a proper joint model and caution must be exercised.

Standard Bayesian computations lead to a decomposition of (11) into two maximizations: for μ_k^c, the product in (11) has a Gaussian form and the mode is given by its mean. For λ_k^c, the product turns into a Gamma distribution and its mode is given by the ratio of its shape parameter over its scale parameter. After some straightforward algebra, we get the following updating formulas:

$$\mu_c^{(\nu+1)\,k} = \frac{\lambda_c^{(\nu)\,k} \sum_{i\in V_c} a_{ik} y_i + \lambda_c^{0k} \left(m_c^k + \sum_{c'\in\mathcal{N}(c)} \eta_{cc'}^k (\mu_{c'}^{(\nu)\,k} - m_{c'}^k)\right)}{\lambda_c^{(\nu)\,k} \sum_{i\in V_c} a_{ik} + \lambda_c^{0k}} \tag{12}$$

$$\text{and} \quad \lambda_c^{(\nu+1)\,k} = \frac{\alpha_c^k + \sum_{i\in V_c} a_{ik}/2 - 1}{b_c^k + 1/2[\sum_{i\in V_c} a_{ik}(y_i - \mu_c^{(\nu+1)\,k})^2]} \tag{13}$$

In these equations, quantities similar to the ones computed in standard EM for the mean and variance parameters appear weighted with other terms due to neighbors information. Namely, standard EM on voxels of V_c would estimate μ_c^k as $\sum_{i\in V_c} a_{ik} y_i / \sum_{i\in V_c} a_{ik}$ and λ_c^k as $\sum_{i\in V_c} a_{ik} / \sum_{i\in V_c} a_{ik}(y_i - \mu_c^k)^2$. In that sense formulas (12) and (13) intrinsically encode cooperation between local models.

From these parameters values constant over subvolumes we compute parameter values per voxel by using cubic splines interpolation between θ_c and $\theta_{c'}$ for all $c' \in \mathcal{N}(c)$. We go back this way to our general setting which has the advantage to ensure smooth variation between neighboring subvolumes and to intrinsically handle nonuniformity of intensity inside each subvolume.

The key-point emphasized by these last derivations of our E and M steps is that it is possible to go from a joint cooperative model to an alternating procedure in which each step reduces to an intuitive well identified task. The goal of the above developments was to propose a well based strategy to reach such derivations. When cooperation exists, intuition is that it should be possible to specify stages where each variable of interest is considered in turn but in a way that uses the other variables current information. Interpretation is easier because in each such stage the central part is played by one of the variable at a time. Inference is facilitated because each step can be recast into a well identified (Hidden MRF) setting for which a number of estimation techniques are available.

6. Results

We choose not to estimate the parameters η_T and η_S but fixed them to the inverse of a decreasing temperature as proposed in Besag (1986). In expressions (12) and (13), we considered a general case but it is natural and common to simplify the derivations by setting the m_c^k's to zero and $\eta_{cc'}^k$ to $|\mathcal{N}(c)|^{-1}$ where $|\mathcal{N}(c)|$ is the number of subvolumes in $\mathcal{N}(c)$. This means that the distribution $p(\mu_c^k | \mu_{\mathcal{N}(c)}^k)$ is a Gaussian centered at $\sum_{c'\in\mathcal{N}(c)} \mu_{c'}^k / |\mathcal{N}(c)|$ and therefore that all neighbors c' of c act with the same weight. The precision parameters λ_c^{0k} is set to $N_c \lambda_g^k$ where λ_g^k is a rough precision estimation for class k obtained for instance using some standard EM algorithm run globally on the entire volume and N_c is

the number of voxels in c that accounts for the effect of the sample size on precisions. The α_c^k's were set to $|\mathcal{N}(c)|$ and b_c^k to $|\mathcal{N}(c)|/\lambda_g^k$ so that the mean of the corresponding Gamma distribution is λ_g^k and the shape parameter α_c^k somewhat accounts for the contribution of the $|\mathcal{N}(c)|$ neighbors. Then, the size of subvolumes is set to $20 \times 20 \times 20$ voxels. The subvolume size is a mildly sensitive parameter. In practice, subvolume sizes from $20 \times 20 \times 20$ to $30 \times 30 \times 30$ give similar good results on high resolution images (1 mm^3). On low resolution images, a size of $25 \times 25 \times 25$ may be preferred.

Evaluation was then performed following the two main aspects of our model. The first aspect is the partionning of the global clustering task into a set of local clustering tasks using local MRF models. The advantage of our approach is that, in addition, a way to ensure consistency between all these local models is dictated by the model itself. The second aspect is the cooperative setting which is relevant when two global clustering tasks are considered simultaneously. It follows that we first assessed the performance of our model considering the local aspect only. We compared (Section 6.1) the results obtained with our method, restricted to tissue segmentation only, with other recent or state-of-the-art methods for tissue segmentation. We then illustrate more of the modelling ability of our approach by showing results for the joint tissue and structure segmentation (Section 6.2). As in Ashburner and Friston (2005); Shattuck et al. (2001b); Van Leemput et al. (1999), for a quantitative evaluation, we used the Dice similarity metric (Dice, 1945) measuring the overlap between a segmentation result and the gold standard.

6.1. **A local method for segmenting tissues**

We first carried out tissue segmentation only (FBM-T) and compare the results with LOCUS-T (Scherrer et al., 2009c), FAST (Zhang et al., 2001) from FSL and SPM5 (Ashburner and Friston, 2005) on both BrainWeb (Collins et al., 1998) phantoms (Table 1) and real 3T brain scans (Figure 4). Our method shows very satisfying robustness to noise and intensity nonuniformity. On BrainWeb images, it is better than SPM5 and comparable to LOCUS-T and FAST, for a low computational time. The mean Dice metric over all eight experiments and for all tissues is 86% for SPM5, 88% for FAST and 89% for LOCUS-T and FBM-T. The mean computation times for the full 3-D segmentation were 4min for LOCUS-T and FBM-T, 8min for FAST and more than 10min for SPM5. On real 3T scans, LOCUS-T and SPM5 also give in general satisfying results.

6.2. **Joint tissue and structure segmentation**

We then evaluated the performance of the joint tissue and structure segmentation (FBM-TS). We introduced *a priori* knowledge based on the Harvard-Oxford subcortical probabilistic atlas[1]. Figures 5 and 6 show an evaluation on real 3T brain scans, using FLIRT (Jenkinson and Smith, 2001)[2] to affine-register the atlas. In Figures 5 and 6, the gain obtained with tissue and structure cooperation is particularly clear for the putamens and thalamus.

We also computed via STAPLE (Warfield et al., 2004) a structure gold standard using three manual expert segmentations of BrainWeb images. We considered the left caudate, left putamen and left thalamus which are of special interest in various neuroanatomical studies and conpared the results with LOCUS-TS (Scherrer et al., 2009c) and FreeSurfer (see Table **??**). FBM-TS lower results for the caudate were due to a bad registration of the atlas in this region. For the putamen and thalamus the improvement is respectively of 14.7% and 20.3%. Due to high computational times,only the 5% noise, 40% nonuniformity image was considered with Freesurfer. For this image, results obtained for FBM-TS were respectively 74%, 84%, 91% for caudate, putamen and thalamus. We then considered 18 images from the IBSR v2 database. The mean Dice metric for the 9 right structures (17 were segmented) is reported in Figure 7.

7. Discussion

The results obtained with our approach are very satisfying and compare favorably with other existing methods. The strength of our fully Bayesian joint model is to be based on the specification of a coherently linked system of conditional models for which we make full use of modern statistics to ensure tractability. The tissue and structure models are linked conditional MRF's that capture several level of interactions. They incorporate 1) spatial dependencies between voxels for robustness to noise, 2) relationships between tissue and structure labels for cooperative aspects and 3) *a priori* anatomical information via the MRF external field parameters for consistency with expert knowledge. Besides, the addition of a conditional MRF model on the intensity distribution parameters allow to handle local estimations for robustness to nonuniformities. In this setting, the whole consistent treatment of MR brain scans is made possible using the framework of Generalized Alternating Minimization (GAM) procedures that generalize the standard EM framework. Another advantage of this approach is that it is made of steps that are easy to interpret and could be enriched with additional information. In particular, results currently highly depend on the atlas registration step which could be introduced in our framework as in Pohl et al. (2006). A step

[1] http://www.fmrib.ox.ac.uk/fsl/

[2] http://www.fmrib.ox.ac.uk/fsl/flirt/

in this direction is proposed in Scherrer et al. (2009b). Also a different kind of prior knowledge could be considered such as the fuzzy spatial relations used in Scherrer et al. (2009c). Other on going work relates to the interpolation step we added to increase robustness to nonuniformities at a voxel level. We believe this stage could be generalized and incorporated in the model by considering successively various degrees of locality, mimicking a multi resolution approach and refining from coarse partitions of the entire volume to finer ones. Also considering more general weights w, to deal with possible conflicts between tissue and structure labels, is possible in our framework and would be an interesting refinement. Eventually, our choice of prior for the intensity distribution parameters was guided by the need to define appropriate conditional specifications $p(\theta_c^k|\theta_{\mathcal{N}(c)}^{k(\nu)})$ in (11) that lead to a valid Markov model for the θ^k's. Nevertheless, incompatible conditional specifications can still be used for inference, *eg.* in a Gibbs sampler or ICM algorithm with some valid justification (see Heckerman et al. (2000) or the discussion in Arnold et al. (2001)). In applications, one may found that having a joint distribution is less important than incorporating information from other variables such as typical interactions. In that sense, conditional modelling allows enormous flexibility in dealing with practical problems. However, it is not clear when incompatibility of conditional distributions is an issue in practice and the theoretical properties of the procedures in this case are largely unknown and should be investigated.

In terms of algorithmic efficiency, our agent-based approach enables us to by-pass computationally intensive implementations usually inherent to MRF models. It results very competitive computational times unusual for such structured models.

References

Arnold, B. C., Castillo, E., Sarabia, J. M., 2001. Conditionally specified distributions: an introduction. Statistical Science 16 (3), 249–274.

Ashburner, J., Friston, K. J., 2005. Unified Segmentation. NeuroImage 26, 839–851.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. J. Roy. Statist. Soc. Ser. B 36 (2), 192–236.

Besag, J., 1986. On the statistical analysis of dirty pictures. J. Roy. Statist. Soc. Ser. B 48 (3), 259–302.

Byrne, W., Gunawardana, A., 2005. Convergence theorems of Generalized Alternating Minimization Procedures. J. Machine Learning Research 6, 2049–2073.

Celeux, G., Forbes, F., Peyrard, N., 2003. EM procedures using mean field-like approximations for Markov model-based image segmentation. Pat. Rec. 36 (1), 131–144.

Ciofolo, C., Barillot, C., 2006. Shape analysis and fuzzy control for 3D competitive segmentation of brain structures with level sets. In: Europ. Conf. Comp. Vision (ECCV). pp. 458–470.

Collins, D. L., Zijdenbos, A. P., Kollokian, V., Sled, J. G., Kabani, N. J., Holmes, C. J., Evans, A. C., 1998. Design and construction of a realistic digital brain phantom. IEEE Trans. Med. Imag. 17 (3), 463–468.

Dice, L. R., 1945. Measures of the amount of ecologic association between species. Ecology 26, 297–302.

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 2004. Bayesian Data Analysis. Chapman & Hall, 2nd edition.

Germond, L., Dojat, M., Taylor, C., Garbay, C., 2000. A cooperative framework for segmentation of MRI brain scans. Artificial Intelligence in Medicine 20, 77–94.

Heckerman, D., Chickering, D. M., Meek, C., Rounthwaite, R., Kadie, C., 2000. Dependency networks for inference, collaborative filtering and data visualization. J. Machine Learning Research 1, 49–75.

Jenkinson, M., Smith, S. M., 2001. A global optimisation method for robust affine registration of brain images. Medical Image Analysis 5 (2), 143–156.

Jordan, M., Ghahramani, Z., Jaakkola, T., Saul, L., 1999. An introduction to variational methods for graphical models. In: Jordan, M. (Ed.), Learning in Graphical Models. pp. 105–162.

Lafferty, J., McCallum, A., Peirera, F., 2001. Conditional Random Fields: Probabilistic models for segmenting and labelling sequence data. In: 18th Inter. Conf. on Machine Learning.

McLachlan, G., Krishnan, T., 1996. The EM Algorithm and Extensions. Wiley.

Minka, T., 2005. Discriminative models not discriminative training. Tech. Report MSR-TR-2005-144, Microsoft Research.

Pohl, K., Fisher, J., Grimson, E., Kikinis, R., Wells, W., 2006. A Bayesian model for joint segmentation and registration. NeuroImage 31 (1), 228–239.

Rajapakse, J. C., Giedd, J. N., Rapoport, J. L., 1997. Statistical approach to segmentation of single-channel cerebral MR images. IEEE Trans. Med. Imag. 16 (2), 176–186.

Richard, N., Dojat, M., Garbay, C., 2007. Distributed markovian segmentation: Application to MR brain scans. Pattern Recognition 40 (12), 3467–3480.

Scherrer, B., Dojat, M., Forbes, F., Garbay, C., 2007. LOCUS: LOcal Cooperative Unified Segmentation of MRI brain scans. In: MICCAI 2007, LNCS, Springer-Verlag, New York City. pp. 1066–74.

Scherrer, B., Dojat, M., Forbes, F., Garbay, C., 2009a. Agentification of Markov model based segmentation: Application to magnetic resonance brain scans. Artificial Intelligence in Medicine 46, 81–95.

Scherrer, B., Forbes, F., Garbay, C., Dojat, M., 2008. Fully Bayesian Joint Model for MR Brain Scan Tissue and Structure Segmentation. In: MICCAI'08, LNCS, Springer-Verlag, New York City. pp. 1066–74.

Scherrer, B., Forbes, F., Garbay, C., Dojat, M., 2009b. A Conditional Random Field approach for coupling local registration with robust tissue and structure segmentation. In: MICCAI 2009, LNCS, Springer-Verlag, New York City. pp. 540–48.

Scherrer, B., Forbes, F., Garbay, C., Dojat, M., 2009c. Distributed Local MRF Models for Tissue and Structure Brain Segmentation. IEEE Trans. Med. Imag. 28, 1296–1307.

Shariatpanahi, H. F., Batmanghelich, N., Kermani, A. R. M., Ahmadabadi, M. N., Soltanian-Zadeh, H., 2006. Distributed behavior-based multi-agent system for automatic segmentation of brain MR images. In: International Joint Conference on Neural Networks, IJCNN'06.

Shattuck, D. W., Sandor-Leahy, S. R., Schaper, K. A., Rottenberg, D. A., Leahy, R. M., 2001a. Magnetic resonance image tissue classification using a partial volume model. NeuroImage 13 (5), 856–876.

Shattuck, D. W., Sandor-Leahy, S. R., Schaper, K. A., Rottenberg, D. A., Leahy, R. M., 2001b. Magnetic resonance image tissue classification using a partial volume model. NeuroImage 13 (5), 856–876.

Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999. Automated model-based bias field correction in MR images of the brain. IEEE Trans. Med. Imag. 18 (10), 885–896.

Warfield, S. K., Zou, K. H., Wells, W. M., 2004. Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. IEEE Trans. Med. Imag. 23 (7), 903–921.

Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the Expectation-Maximisation algorithm. IEEE Trans. Med. Imag. 20 (1), 45–47.

	CSF	GM	WM	M.C.T
FBM-T	79.9 %	91.6 %	93.6 %	$\approx$ 4min
LOCUS-T	79.8 %	91.8 %	93.7 %	$\approx$ 4min
SPM5	79.5 %	89.2 %	90.4 %	$\approx$ 12min
FAST	79.6 %	91.3 %	94.1 %	$\approx$ 8min

Table 1: FBM-T. Mean Dice metric and mean computational time (M.C.T) values on BrainWeb over 8 experiments for different values of noise (3%, 5%, 7%, 9%) and nonuniformity (20%, 40%).

	Caudate	Putamen	Thalamus	M.C.T
FBM-TS	73.7 %	84.7 %	91.3 %	≈25min
Locus-TS	84 %	70 %	71 %	≈15 min
Freesurfer	88 %	86 %	92 %	≈15 h.

Table 2: FBM-TS. Mean Dice metric and mean computational time (M.C.T) values on BrainWeb over 8 experiments for different values of noise (3%, 5%, 7%, 9%) and nonuniformity (20%, 40%). Note that for Freesurfer only 5% noise and 40% nonuniformity was considered.

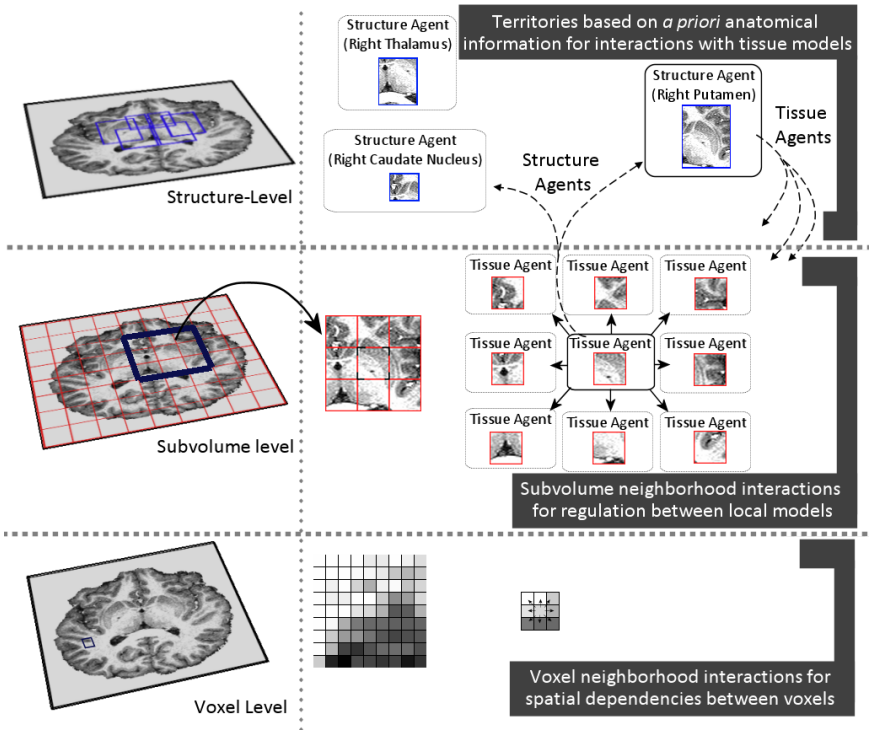


Figure 1: Symbolic multi-level design of our approach.

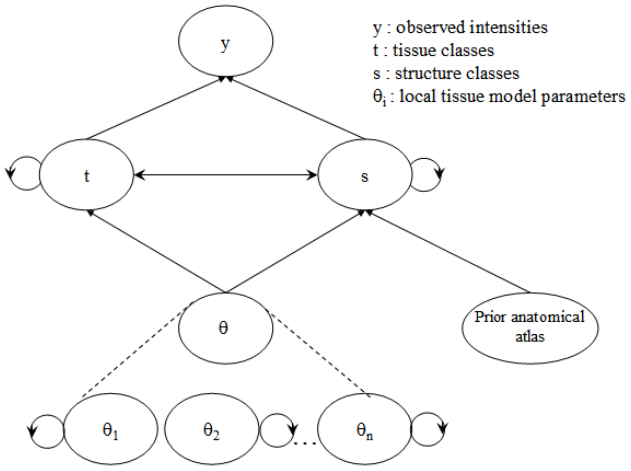


Figure 2: Graphical model illustrating the joint dependencies between local intensity models, and tissue and structure segmentations.

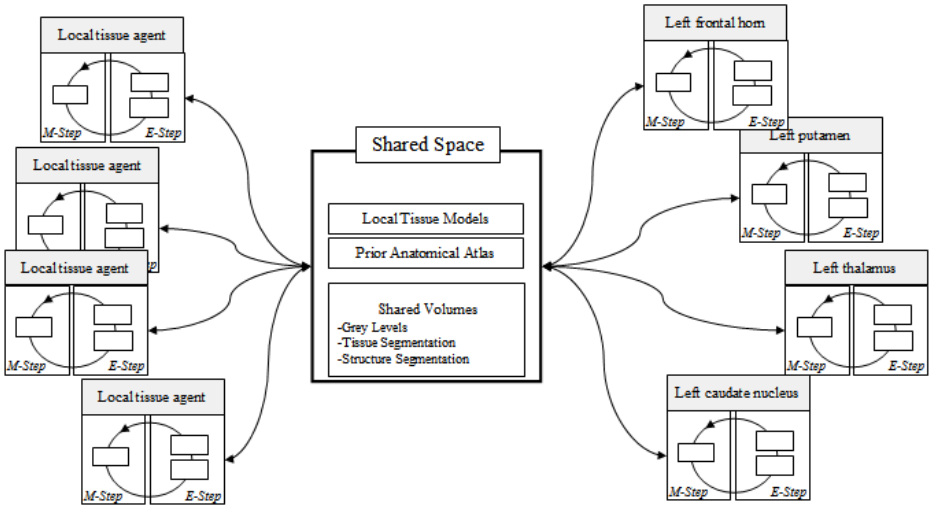
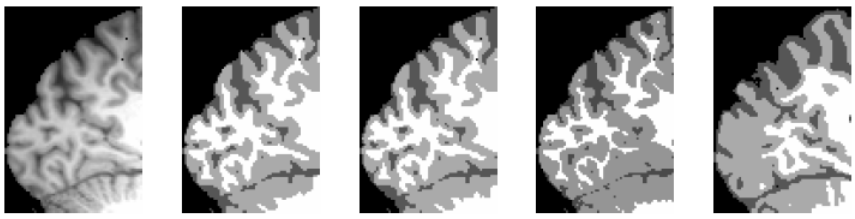


Figure 3: Implementation using a two agent-layer architecture and a shared space for communication.



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

Figure 4: FBM-T. Segmentations respectively by FBM-T (b), LOCUS-T (c), SPM5 (d) and FAST (e) of a highly nonuniform real 3T image (a).

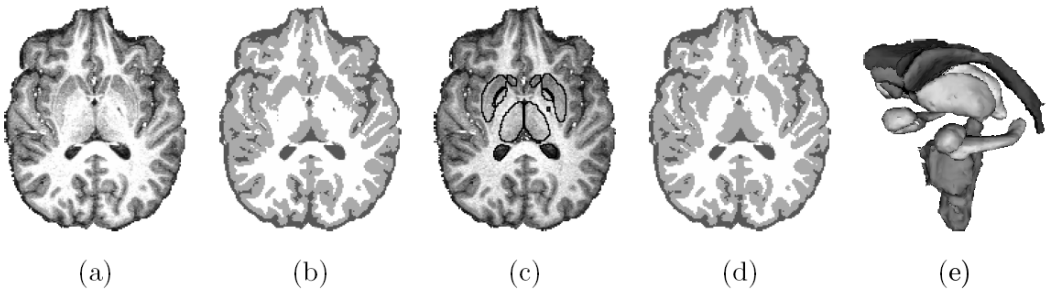


Figure 5: Evaluation of FBM-TS on a real 3T brain scan (a). For comparison the tissue segmentation obtained with FBM-T is shown in (b). Images (c) and (d): structure segmentation by FBM-TS and corresponding improved tissue segmentation. Image (e): 3-D reconstruction of the 17 segmented structures: the two lateral ventricules, caudates, accumbens, putamens, thalamus, pallidums, hippocampus, amygdalas and the brain stem. The computational time was $< 15min$ after the registration step.

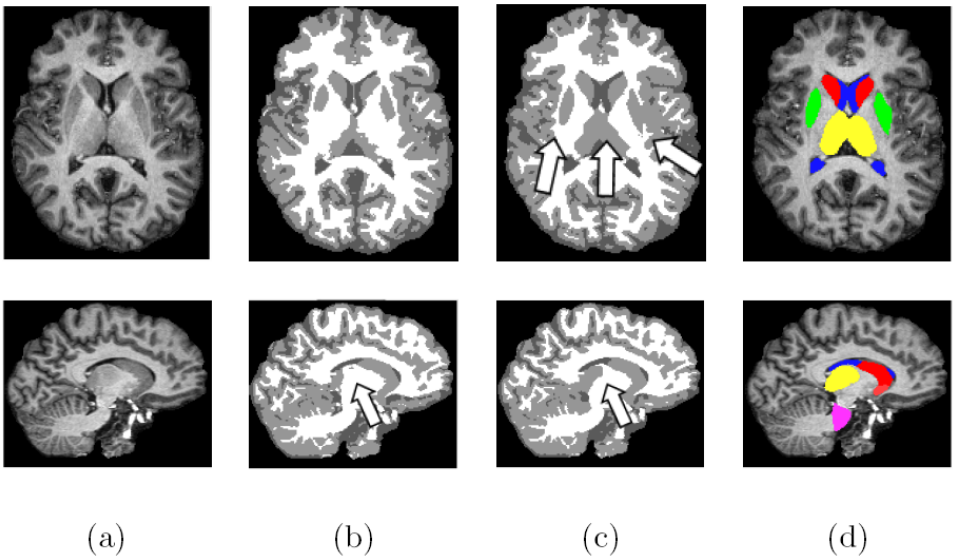


Figure 6: Evaluation of FBM-TS on a real 3T brain scan (a). For comparison the tissue segmentation obtained with FBM-T is shown in (b). The tissue segmentation obtained with FBM-TS is given in (c). Major differences between tissue segmentations (images (b) and (c)) are pointed out using arrows. Image (d) shows the structure segmentation with FBM-TS.

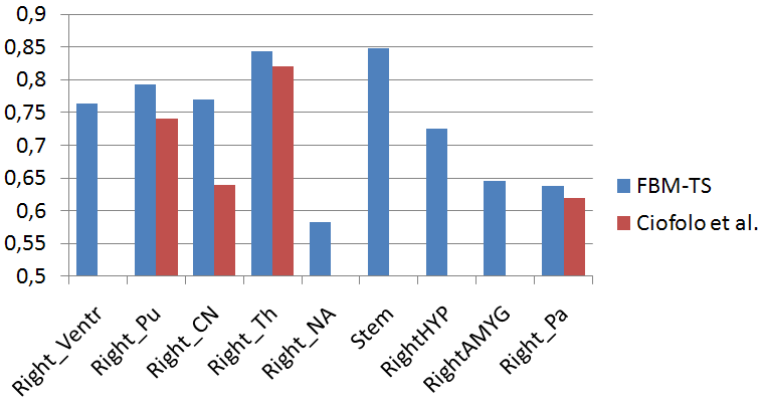


Figure 7: Evaluation of FBM-TS on IBSR v2 (9 right structures) and comparison with Ciofolo and Barillot (2006). Y axis = Mean Dice metric