HAL
open science

Fully Bayesian joint model for MR brain scan tissue and structure segmentation

Benoit Scherrer, Florence Forbes, Catherine Garbay, Michel Dojat

To cite this version:

Benoit Scherrer, Florence Forbes, Catherine Garbay, Michel Dojat. Fully Bayesian joint model for MR brain scan tissue and structure segmentation. 11th International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008, Sep 2008, New York, NY, United States. pp.1066-1074, 10.1007/978-3-540-85990-1_128 . inserm-00356883

HAL Id: inserm-00356883

https://inserm.hal.science/inserm-00356883v1

Submitted on 28 Jan 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Fully Bayesian Joint Model for MR Brain Scan Tissue and Structure Segmentation

B. Scherrer[1,3,4], F. Forbes[2,4], C. Garbay[3,4], M. Dojat[1,4]

[1] INSERM, U836, Grenoble, F-38043, France
[2] INRIA, MISTIS, Grenoble, France
[3] CNRS, MAGMA, Grenoble, France
[4] Université Joseph Fourier, 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 local tissue and structure segmentations and local intensity distributions. 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 modelization. The complete joint model provides a sound theoretical framework for carrying out tissue and structure segmentation by distributing a set of local and cooperative MRF models. The evaluation, using a previously affine-registred atlas of 17 structures and performed on both phantoms and real 3T brain scans, shows good results.

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 structures locations provides valuable information about local intensity distribution for a given tissue. 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 [1] or use redundant information to ensure consistency of local models [2]. Recently, good results are reported using an innovative local and cooperative approach

[3]. It performs tissue and subcortical structure segmentation by distributing through the volume a set of local Markov Random Field (MRF) models which better reflect local intensity distributions. Local MRF models are used alternatively for tissue and structure segmentations. 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 [3], cooperation mechanisms between local models are somewhat arbitrary and independent of the MRF models themselves. Our contribution is then to propose a fully Bayesian framework in which we define a joint model that links local tissue and structure segmentations but also the model parameters so 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 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 [4]. 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.

2 A fully Bayesian joint model for tissues and structures

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 will 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 will write (^{t} means transpose) for all $k = 1, \ldots K$, $\theta^k = (\theta_i^k, i \in V)$ and for all $i \in V$, $\theta_i = {}^t(\theta_i^k, k = 1, \ldots K)$. 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 will consider θ_i^k's equal for all voxels i in some prescribed regions (see below).

To explicitly take into account the fact that tissue and structure classes are related, we 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 (see [5] for the rational of such an approach). The distribution $p(\mathbf{t}, \mathbf{s}, \theta|\mathbf{y})$ 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 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. In what follows, notation ${}^t xx'$ denotes the scalar product between two vectors x and x'.

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,

$$H_{T|S,Y,\theta}(\mathbf{t}|\mathbf{s}, \mathbf{y}, \theta) = \sum_{i \in V} \left({}^t t_i \gamma_i(s_i) + \sum_{j \in \mathcal{N}(i)} V_{ij}^T(t_i, t_j; \eta_T) + \log g_T(y_i; {}^t\theta_i t_i) \right), \quad (1)$$

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 rational 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 $\mu = (\mu_i^k, i \in V, k = 1 \ldots K)$ 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 $f_i = {}^t(f_i(e'_l), l = 1 \ldots L+1)$ and $f_i(e'_l)$ represents some prior probability that voxel i belongs to structure l , 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,

$$H_{S|T,Y,\theta}(\mathbf{s}|\mathbf{t}, \mathbf{y}, \theta) = \sum_{i \in V} \left({}^t s_i \log f_i + \sum_{j \in \mathcal{N}(i)} V_{ij}^S(s_i, s_j; \eta_S) + \log g_S(y_i|t_i, s_i, \theta_i) \right) \quad (2)$$

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

$$g_S(y_i|t_i, s_i, \theta) = [g_T(y_i; {}^t\theta_i e_{T^{s_i}})\ f_i(s_i)]^{w(s_i)}\ [g_T(y_i; {}^t\theta_i t_i)\ f_i(e'_{L+1})]^{(1-w(s_i))} \quad (3)$$

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. Other parameters in (1) and (2) include interaction parameters η_T and η_S which are considered here as hyperparameters to be specified (see Section 4).

Conditional model for the parameter θ. To ensure spatial consistency between the parameter values, we define $p(\theta|\mathbf{y}, \mathbf{t}, \mathbf{s})$ also as a MRF. In practice however, in the general setting of Section 2, there are too many parameters and

estimating them accurately is not possible. Our local approach consists then in considering a regular cubic partioning $\mathcal{C}$ of 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 $\theta = \{\theta_c, c \in \mathcal{C}\}$. 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 (3). The specific form of the Markov prior on θ is specified in Section 3.

3 Estimation by Generalized Alternating Minimization

Ultimately, we are interested in maximizing over $\mathbf{t}, \mathbf{s}$ and θ the posterior distribution $p(\mathbf{t}, \mathbf{s}, \theta \mid \mathbf{y})$. This problem is greatly simplified when the solution is determined within an EM algorithm framework. Let us denote $\mathbf{z} = (\mathbf{t}, \mathbf{s})$ and let $\mathcal{D}$ be the set of all probability distributions on $\mathbf{z}$. As discussed in [4], EM can be viewed as an alternating maximization procedure of a function F defined, for any probability distribution $q \in \mathcal{D}$, by $F(q, \theta) = \sum_{\mathbf{z} \in \mathcal{Z}} \ln p(\mathbf{y}, \mathbf{z} \mid \theta)\ q(\mathbf{z}) + I[q]$, where $I[q] = -E_q[\log q(\mathbf{Z})]$ is the entropy of q (E_q denotes the expectation with regard to q). When considering our MAP problem, we can replace (see *eg.* [6]) the function $F(q, \theta)$ by $F(q, \theta) + \log p(\theta)$. It comes the following alternating procedure: starting from a current value $(q^{(r)}, \theta^{(r)}) \in \mathcal{D} \times \Theta$, set alternatively

$$q^{(r+1)} = \arg\max_{q \in \mathcal{D}}\ F(q, \theta^{(r)}) = \arg\max_{q \in \mathcal{D}} \sum_{\mathbf{z} \in \mathcal{Z}} \log p(\mathbf{z}|\mathbf{y}, \theta^{(r)})\ q(\mathbf{z}) + I[q] \tag{4}$$

$$\theta^{(r+1)} = \arg\max_{\theta \in \Theta}\ F(q^{(r+1)}, \theta) + \log p(\theta)\ =\ \arg\max_{\theta \in \Theta} \sum_{\mathbf{z} \in \mathcal{Z}} \log p(\theta|\mathbf{y}, \mathbf{z})\ q^{(r+1)}(\mathbf{z})\ . \tag{5}$$

The last equalities in (4) and (5) come from straightforward probabilistic rules and show that inference can be described in terms of the conditional models $p(\mathbf{z}|\mathbf{y}, \theta)$ and $p(\theta|\mathbf{y}, \mathbf{z})$. However, solving the optimization (4) over the set $\mathcal{D}$ of probability distributions $q_{(T,S)}$ on $(\mathbf{T}, \mathbf{S})$ leads for the optimal $q_{(T,S)}$ to $p(\mathbf{t}, \mathbf{s}|\mathbf{y}, \theta^{(r)})$ which is intractable in our model. We therefore propose an EM variant in which the E-step is not performed exactly. This variant falls in the Generalized Alternating Minimization (GAM) procedures family for which convergence results are available [4]. The optimization (4) 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 $\mathbf{T}$ (resp. on $\mathbf{S}$). This variant leads to a two stage E-step. Using two equivalent expressions of F when q factorizes as in $\tilde{\mathcal{D}}$, we can show that these two steps reduce to,

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

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

More generally, we can adopt in addition, an Incremental EM approach [4] which allows re-estimation of the parameters (here the θ) to be performed based on only a sub-part of the hidden variables. This mean that we can incorporate an M-step (5) in between the updating of q_T and q_S. Similarly, hyperparameters could be updated there too. Steps E-T and E-S can be further specified by computing the expectations with regards to $q_S^{(r)}$ and $q_T^{(r+1)}$. Using the structure conditional model definition (1) and omitting the terms that do not depend on T, it appears that step E-T is equivalent to the E-step one would get when applying EM to a standard Hidden MRF in $\mathbf{t}$. More specifically, the external field term comes from $E_{q_S^{(r)}}[\gamma_i(S_i)] = \sum_{l=1}^{L} e_{T^l}\ q_{S_i}^{(r)}(e'_l)$, which is a 3-component vector whose k^{th} ($k = 1, \ldots, 3$) component 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. The data term is $\sum_{i \in V} \log g_T(y_i; {}^t\theta_i^{(r)} t_i)$. To solve this step, then, various inference techniques for Hidden MRF's can be applied. In this paper, we adopt Mean field like algorithms as in [3]. Similarly, using definitions (2) and (3) and omitting terms that do not depend on S, step E-S can be seen as the E-step for a standard Hidden MRF in $\mathbf{s}$ with an external field incorporating prior structure knowledge through f and class distributions given by g'_S which comes from the computation of $E_{q_T^{(r+1)}}[\log g_S(y_i|T_i, S_i, \theta_i^{(r)})]$. More specifically,

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

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

We now turn to the resolution of step (5), for θ_i's constant over subvolumes of a given partition $\mathcal{C}$. Denoting by $p(\theta)$ the MRF prior on $\theta = \{\theta_c, c \in \mathcal{C}\}$, *ie.* $p(\theta) \propto \exp(H_\theta^{\mathcal{C}}(\theta))$ and using the additional natural assumption that $p(\theta) = \prod_{k=1}^{K} p(\theta^k)$, (5) is equivalent to

$$\forall k = 1 \ldots K, \quad \theta^{k\ (r+1)} = \arg\max_{\theta^k \in \Theta^k} p(\theta^k) \prod_{c \in \mathcal{C}} \prod_{i \in V_c} g_T(y_i; \theta_c^k)^{a_{ik}}. \tag{6}$$

where $a_{ik} = q_{T_i}(e_k) q_{S_i}(e'_{L+1}) + \sum_{l\ st. T^l = e_k} q_{S_i}(e_l)$. 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. When $p(\theta^k)$ is chosen as a Markov field, the exact maximization (6) is still intractable and we use the following ICM [7] algorithm: for a current estimation of θ^k at iteration ν, we consider in turn,

$$\forall c \in \mathcal{C}, \quad \theta_c^{k\ (\nu+1)} = \arg\max_{\theta_c^k \in \Theta^k} p(\theta_c^k \mid \theta_{\mathcal{N}(c)}^{k\ (\nu)}) \prod_{i \in V_c} g_T(y_i; \theta_c^k)^{a_{ik}}\ , \tag{7}$$

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 (7) above actually dictates 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 $g_T(y_i;\theta_c^k)$ [6]. 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 [8] 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. After some straightforward algebra, we get the following updating formulas:

$$\mu_c^{(\nu+1)\,k} = \frac{\lambda_c^{(\nu)\,k} \sum_{i\in V_c} a_{ik} y_i + \lambda_c^{0k}\,(m_c^k + \sum_{c'\in\mathcal{N}(c)} \eta_{cc'}^k(\mu_{c'}^{(\nu)\,k} - m_{c'}^k))}{\lambda_c^{(\nu)\,k} \sum_{i\in V_c} a_{ik} + \lambda_c^{0k}} \quad (8)$$

$$\text{and} \quad \lambda_c^{(\nu+1)\,k} = \frac{\alpha_c^k + \sum_{i\in V_c} a_{ik}/2 - 1}{b_c^k + 1/2[\sum_{i\in V_c} a_{ik}(y_i - \mu_c^{(\nu+1)\,k})^2]} \quad (9)$$

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 (8) and (9) 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.

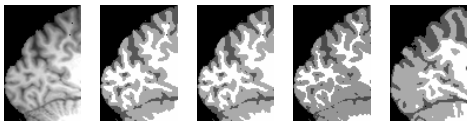
4 Results

We chose not to estimate the parameters η_T and η_S but fixed them to the inverse of a decreasing temperature as proposed in [7]. In expressions (8) and (9), we set 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 $\mathcal{N}(c)$. 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 $|\mathcal{N}(c)|/\lambda_g^k$, and the size of subvolumes to 20^3 voxels. We first carried out tissue segmentation only (FBM-T) and compare the results with LOCUS-T [3], SPM5 and FAST on both BrainWeb phantoms and real 3T brain scans (Figure 1). 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. On real 3T scans, LOCUS-T and SPM5 also give in general satisfying results. For the joint tissue and structure model (FBM-TS) we introduced *a priori* knowledge based on the Harvard-Oxford subcortical probabilistic atlas. Figure 2 shows an evaluation on a real 3T brain scan, using FLIRT to affine-register the atlas. In this Figure, the gain obtained with tissue and structure cooperations is particularly clear for the putamens and thalamus. We also computed via STAPLE 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. The mean Dice metric over 8 experiments (phantoms with 3%, 5%, 7%, 9% of noise, with 20% or 40% of nonuniformity) was 73.7% for the caudate, 84.7% for the putamen and 91.3% for the thalamus. The computational time was less than 25min including the registration step. For comparison, LOCUS-TS [3], which uses a priori fuzzy spatial relations, led respectively to 84%, 70% and 71% in 15min. 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%. Freesurfer led respectively to 88%, 86%, 90% on the 5% noise, 40% nonuniformity image (resp. 74%, 84%, 91% for FBM-TS) with a computational time larger than 15 hours for 37 structures.

	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

(a)



(b) (c) (d) (e) (f)

Fig. 1. FBM-T. Table (a): 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%). Images (c), (d), (e), (f): segmentations respectively by FBM-T, LOCUS-T, SPM5 and FAST of a highly nonuniform real 3T image (b).

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

Fig. 2. Real 3T brain scan (a). (b): tissue segmentation by FBM-T. (c) and (d): structure segmentation by FBM-TS and corresponding improved tissue segmentation. (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.

5 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 [9]. Also a different kind of prior knowledge could be considered such as the fuzzy spatial relations used in [3]. 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 multiresolution 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.

References

1. Shattuck, D., et al: Magnetic resonance image tissue classification using a partial volume model. NeuroImage **13**(5) (2001) 856–876
2. Rajapakse, J.C., Giedd, J.N., Rapoport, J.L.: Statistical approach to segmentation of single-channel cerebral MR images. IEEE TMI **16**(2) (1997) 176–186
3. Scherrer, B., Dojat, M., Forbes, F., Garbay, C.: LOCUS: LOcal Cooperative Unified Segmentation of MRI brain scans. In: MICCAI 2007, Brisbane, Australia
4. Byrne, W., Gunawardana, A.: Convergence theorems of Generalized Alternating Minimization Procedures. J. Machine Learning Research **6** (2005) 2049–2073
5. Kumar, S., Hebert, M.: Discriminative random fields. Int. J. Comput. Vision **68**(2) (2006) 179–201
6. Gelman, A., et al: Bayesian Data Analysis. Chapman & Hall, 2nd edition (2004)
7. Besag, J.: On the statistical analysis of dirty pictures. J. Roy. Statist. Soc. Ser. B **48**(3) (1986) 259–302
8. Besag, J.: Spatial interaction and the statistical analysis of lattice systems. J. Roy. Statist. Soc. Ser. B **36**(2) (1974) 192–236
9. Pohl, K.M., Fisher, J., Grimson, W.E.L., Kikinis, R., Wells, W.M.: A Bayesian model for joint segmentation and registration. NeuroImage **31**(1) (2006) 228–239