A Weighted Multi-Sequence Markov Model For Brain Lesion Segmentation

Florence Forbes, Senan Doyle, Daniel Garcia-Lorenzo, Christian Barillot, Michel Dojat

A Weighted Multi-Sequence Markov Model For Brain Lesion Segmentation

F. Forbes, S. Doyle
INRIA Grenoble Rhône-Alpes,
LJK
Mistis team
Montbonnot, France

D. Garcia-Lorenzo, C. Barillot
INRIA Rennes Bretagne Atlantique
Visages team
Rennes, France

M. Dojat
INSERM
GIN
Grenoble, France

Abstract

We propose a technique for fusing the output of multiple Magnetic Resonance (MR) sequences to robustly and accurately segment brain lesions. It is based on an augmented multi-sequence hidden Markov model that includes additional weight variables to account for the relative importance and control the impact of each sequence. The augmented framework has the advantage of allowing 1) the incorporation of expert knowledge on the *a priori* relevant information content of each sequence and 2) a weighting scheme which is modified adaptively according to the data and the segmentation task under consideration. The model, applied to the detection of multiple sclerosis and stroke lesions shows promising results.

1 Introduction

Magnetic Resonance (MR) brain scans consist of 3D data *volumes*, composed of *voxels* (volume elements). The segmentation of such data sets into their constituent tissues is a fundamental task in a number of applications. A healthy brain is generally segmented into three tissues: cephalo spinal fluid, grey matter and white matter. Statistical based approaches usually aim to model probability distributions of voxel intensities with the idea that such distributions are tissue-dependent.

The delineation and quantification of brain lesions is critical to establishing patient prognosis, and for charting the development of pathology over time. Typically, this is performed manually by a medical expert, however automatic methods have been proposed (see [20] for review) to alleviate the tedious, time consuming and subjective nature of manual delineation.

Automatic brain image segmentation remains a challenging task due to the presence of various artifacts such as noise or intensity nonuniformities. The latter induce spatial intensity variations within each tissue which is a major obstacle to an accurate tissue segmentation.

Automated or semi-automated brain lesion detection methods can be classified according to their use of multiple sequences, *a priori* knowledge about the structure of normal brain, tissue segmentation models, and whether or not specific lesion types are targeted. A common feature is that most methods are based on the initial identification of *candidate regions* for lesions. In most approaches, normal brain tissue *a priori* maps are used to help identify regions where the damaged brain differs, and the lesion is identified as an outlier.

Existing methods frequently avail of complementary information from multiple sequences. For example, lesion voxels may appear atypical in one modality and normal in another. This is well known and implicitly used by neuroradiologists when examining data. Within a mathematical framework, multiple sequences enable the superior estimation of tissue classes in a higher dimensional space.

For multiple MRI volumes, intensity distributions are commonly modelled as multi-dimensional Gaussian distributions. This provides a way to combine the multiple sequences in a single segmentation task but with all the sequences having equal importance.

However, given that the information content and discriminative power to detect lesions varies between different MR sequences, the question remains as to how to best combine the multiple channels. Depending on

Appearing in Proceedings of the 13^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2010, Chia Laguna Resort, Sardinia, Italy. Volume 9 of JMLR: W&CP 9.

the task at hand, it might be beneficial to weight the various sequences differently.

In this paper, rather than trying to detect lesion voxels as outliers from a normal tissue model, we adopt an incorporation strategy whose goal is to identify lesion voxels as an additional fourth component. Such an explicit modelling of the lesions is usually avoided. It is difficult for at least two reasons: 1) most lesions have a widely varying and inhomogeneous appearance (*eg.* tumors or stroke lesions) and 2) lesion sizes can be small (*eg.* multiple sclerosis lesions). In a standard tissue segmentation approach, both reasons usually prevent accurate model parameter estimation resulting in bad lesion delineation. Our approach aims to make this estimation possible by modifying the segmentation model with an additional weight field. We propose to modify the tissue segmentation model so that lesion voxels become inliers for the modified model and can be identified as a genuine model component. Compared to *robust estimation* approaches (*eg.* [22]) that consist of down-weighting the effect of outliers on the main model estimation, we aim to increase the weight of candidate lesion voxels to overcome the problem of under-representation of the lesion class.

We introduce additional weight variables in the segmentation model and then solve the issue of prescribing values for these weights by developing an appropriate estimation framework. This has the advantage to avoid the specification of *ad-hoc* weight values and to allow the incorporation of expert knowledge through a weight distribution. We provide an estimation procedure based on a variational Expectation Maximization (EM) algorithm to produce the corresponding segmentation. Furthermore, in the absence of explicit expert knowledge, we show how the weight distribution can be specified to guide the model toward lesion identification. Experiments on artificial and real lesions of various sizes are reported to demonstrate the good performance of our approach.

2 A weighted multi-sequence Markov model

Combining sequences is a data fusion issue which, in a probabilistic setting, naturally becomes an issue of combining probabilistic distributions. This relates to the so-called *pooling* of distributions in the statistical literature [12]. Examples include *linear* and *logarithmic* pooling. The former corresponds to a mixture of distributions, while the latter consists of combining the distributions into a product where each component is raised to a power. This power is viewed as a weight. In this work we will consider logarithmic pooling for it appears that it is more appropriate to our segmentation framework. Note however that the link to pooling although interesting is only mention for information and that our approach could be presented without referring to such aspects.

We consider a finite set V of N voxels on a regular 3D grid. We denote by $\mathbf{y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$ the intensity values observed respectively at each voxel.

Each $\mathbf{y}_i = \{y_{i1}, \ldots, y_{iM}\}$ is itself a vector of M intensity values corresponding to M different MR sequences. Our goal is to assign each voxel i to one of K classes considering the observed features data $\mathbf{y}$. For brain tissue segmentation, we consider in general 3 tissues plus some possible additional classes to account for lesions in pathological data. We denote the hidden classes by $\mathbf{z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$, and the set in which $\mathbf{z}$ takes its values by $\mathcal{Z}$.

Typically, the $\mathbf{z}_i$'s take their values in $\{1 \ldots K\}$. We consider nonnegative weights $\omega = \{\omega_i, i \in V\}$ in a state space denoted by $\mathcal{W}$ and with $\omega_i = \{\omega_{i1}, \ldots, \omega_{iM}\}$. In our general setting the weights are sequence and voxel-specific. The rationale is that relevant information is not usually uniformly localized so that the weights cannot be fixed equally for all the voxels in a given sequence but may depend on the location in the brain.

Spatial dependencies between voxels are then introduced through Markov Random Field (MRF) modelling.

The segmentation task is recast into a missing data framework in which $\mathbf{y}$ are observations and $\mathbf{z}$ are missing variables. Their joint distribution $p(\mathbf{y}, \mathbf{z}|\omega; \psi)$ is governed by the weights $\omega \in \mathcal{W}$ and parameters $\psi \in \underline{\Psi}$, which are both unknown and need to be estimated within the segmentation procedure. A prior distribution $p(\omega)$ is defined on the weights.

Taking advantage of the fact that Bayesian inference does not differentiate between missing data and random parameters, we propose a framework in which the weights ω are viewed as additional missing variables. Denoting the parameters by $\psi = \{\beta, \phi\}$, we assume that the joint distribution $p(\mathbf{y}, \mathbf{z}, \omega; \psi)$ is a MRF with the following energy function:

$$H(\mathbf{y}, \mathbf{z}, \omega; \psi) = H_{\mathbf{Z}}(\mathbf{z}; \beta) + H_W(\omega) + \sum_{i \in V} \log g(\mathbf{y}_i|\mathbf{z}_i, \omega_i; \phi) \quad (1)$$

where the energy term $H_W(\omega)$ involving only ω does not depend on ψ and the $g(\mathbf{y}_i|\mathbf{z}_i, \omega_i; \phi)$s are probability density functions of $\mathbf{y}_i$. The three terms in this energy are further specified below.

Data term. The data term $\sum_{i \in V} \log g(\mathbf{y}_i|\mathbf{z}_i, \omega_i; \phi)$ in (1) corresponds to the modelling of tissue dependent

intensity distributions. We consider M-dimensional Gaussian distributions with diagonal covariance matrices. For each class k, ${}^t(\mu_{k1}, \ldots, \mu_{kM})$ is the mean vector and $\{s_{k1}, \ldots, s_{kM}\}$ the covariance matrix components. When $\mathbf{z}_i = k$, then $\mathcal{G}(y_{im}; \mu_{\mathbf{z}_i m}, s_{\mathbf{z}_i m})$ represents the Gaussian distribution with mean μ_{km} and variance s_{km}. The whole set of Gaussian parameters is denoted by $\phi = \{\mu_{km}, s_{km}, k = 1 \ldots K, m = 1 \ldots M\}$. Our data term is then defined by setting

$$g(\mathbf{y}_i|\mathbf{z}_i, \omega_i; \phi) \quad = \quad \prod_{m=1}^{M} \mathcal{G}(y_{im}; \mu_{\mathbf{z}_i m}, \frac{s_{\mathbf{z}_i m}}{\omega_{im}}) \; ,$$

which is proportional to $\prod_{m=1}^{M} \mathcal{G}(y_{im}; \mu_{\mathbf{z}_i m}, s_{\mathbf{z}_i m})^{\omega_{im}}$. This corresponds to a *modified* logarithmic pooling [12] of the M distributions $p(\mathbf{z}_i|y_{im}, \omega_{im}; \psi)$ and $p(\mathbf{z}_i; \beta)$.

Intuitively, the impact of a larger ω_{im} is to give more importance to the intensity value y_{im} in the model. Typically an integer ω_{im} greater than one would correspond to increase ω_{im} times the number of voxels with intensity value y_{im}. When the weights are all one, a standard multivariate Gaussian case is recovered.

Missing label term. The missing data term $H_{\mathbf{Z}}(\mathbf{z}; \beta)$ in (1) is set to a standard Potts model, with external field ξ and spatial interaction parameter η, and whose energy is

$$H_{\mathbf{Z}}(\mathbf{z}; \beta) = \sum_{i \in V} (\xi_{i\mathbf{z}_i} + \sum_{j \in \mathcal{N}(i)} \eta \, \langle \mathbf{z}_i, \mathbf{z}_j \rangle),$$

where $\mathcal{N}(i)$ denotes the voxels neighboring i and $\langle \mathbf{z}_i, \mathbf{z}_j \rangle$ is 1 when $\mathbf{z}_i = \mathbf{z}_j$ and 0 otherwise. Parameter $\beta = \{\xi, \eta\}$ with $\xi = \{{}^t(\xi_{i1} \ldots \xi_{iK}), i \in V\}$ being a set of real-valued K-dimensional vectors and η a real positive value.

Missing weight term. The weights are assumed independent from parameters ψ and independent across modalities. The simplest choice is to define a prior $p(\omega) = \prod_{m=1}^{M} \prod_{i \in V} p(\omega_{im})$ where each $p(\omega_{im})$ is a Gamma distribution with hyperparameters α_{im} (shape) and γ_{im} (inverse scale). Thus

$$H_W(\omega) = \sum_{m=1}^{M} \sum_{i \in V} ((\alpha_{im} - 1) \log \omega_{im} - \gamma_{im} \, \omega_{im}).$$

In practice, the set of hyperparameters is fixed so that the modes of each prior $p(\omega_{im})$ are located at some *expert weights* $\{\omega_{im}^{exp}, m = 1 \ldots M, i \in V\}$ accounting for some external knowledge, if available. Formally, we set $\alpha_{im} = \gamma_{im} \, \omega_{im}^{exp} + 1$ to achieve this. The expert weights can be chosen according to the specific task. For example, when voxels with typical lesion intensities are not numerous enough to attract a model component, increasing the expert weight for some of them will help in biasing the model toward the identification of a lesion class.

Note that we also investigated the use of Dirichlet distributions for the weights adding the constraints that they should sum to the sample size N in each modality. However, there were no real theoretical reasons to do so and it required less stable numerical computation. This is due to the fact that the Dirichlet distribution is not a conjugate distribution in our setting. In addition, results were not improved compared to the simpler independent Gamma case.

3 Estimation by Variational EM

We propose to use an Expectation-Maximization (EM) [7] framework to deal with the missing label and weight data.

Let $\mathcal{D}$ be the set of all probability distributions on $\mathcal{Z} \times \mathcal{W}$. EM can be viewed [16] as an alternating maximization procedure of a function F such that for any $q \in \mathcal{D}$, $F(q, \psi) = E_q[\log p(\mathbf{y}, \mathbf{Z}, W \; ; \; \psi)] + I[q]$ where $I[q] = -E_q[\log q(\mathbf{Z}, W)]$ is the entropy of q, and E_q denotes the expectation with respect to q. Capital letters indicate random variables, and lower case their realisations. Denoting current parameter values by $\psi^{(r)}$, the corresponding alternating procedure assigns:

$$\begin{aligned} \textbf{E-step: } & q^{(r)} = \arg\max_{q \in \mathcal{D}} \; F(q, \psi^{(r)}) \\ \textbf{M-step: } & \psi^{(r+1)} = \arg\max_{\psi \in \underline{\Psi}} \; F(q^{(r)}, \psi) \end{aligned} \tag{2}$$

However, the optimization (2) leads to $q^{(r)}(\mathbf{z}, \omega) = p(\mathbf{z}, \omega|\mathbf{y}; \psi^{(r)})$ which is intractable for non trivial Markov models.

We therefore propose to use an EM variant in which the E-step is instead solved over a restricted class of probability distributions, $\tilde{\mathcal{D}}$, chosen as the set of distributions that factorize as $q(\mathbf{z}, \omega) = q_Z(\mathbf{z}) \; q_W(\omega)$ where $q_Z \in \mathcal{D}_Z$ and $q_W \in \mathcal{D}_W$, the sets of probability distributions on $\mathcal{Z}$ and $\mathcal{W}$ respectively.

The fact that the weights can be equivalently considered as missing variables or random parameters induces some similarity between our Variational EM variant and the Variational Bayesian EM algorithm presented in [2, 13]. Our framework differs slightly. In contrast to these latter papers, our observations are not *i.i.d.* and condition (2) in Section 3 of [2] is not satisfied. However, these differences are not significant. More importantly, our missing data presentation offers the possibility to deal with extra parameters (the Gaussian means and variances in our setting) for which no prior information is available. This is done

in a maximum likelihood manner and avoids the use of non-informative priors that could be problematic (difficulties with non informative priors are listed in [11] p. 64-65). As a consequence, the variational Bayesian M-step of [2] is transferred into our E-step while our M-step has no equivalent in the formulation of [2].

It follows then that the E-step becomes an approximate E-step that can be further decomposed into two stages using two equivalent expressions of F when q factorizes in $\tilde{\mathcal{D}}$. At iteration (r), with current estimates denoted by $q_W^{(r-1)}$ and $\psi^{(r)}$, the updating becomes

$$\begin{aligned} \textbf{E-Z-step:}\ q_Z^{(r)} &= \arg\max_{q_Z \in \mathcal{D}_Z} F(q_W^{(r-1)}\, q_Z; \psi^{(r)}) \\ \textbf{E-W-step:}\ q_W^{(r)} &= \arg\max_{q_W \in \mathcal{D}_W} F(q_W\, q_Z^{(r)}; \psi^{(r)})\ . \end{aligned}$$

These expressions can be written in terms of a Kullback-Liebler divergence so that it is not necessary to use calculus of variations to take functional derivatives with respect to q_Z and q_W. From the Kullback-Liebler divergence properties, the solutions of the E-Z and E-W steps satisfy:

$$\textbf{E-Z:}\ q_Z^{(r)} \propto \exp\left(E_{q_W^{(r-1)}}[\log p(\mathbf{z}|\mathbf{y}, \mathbf{W}; \psi^{(r)}]\right) \quad (3)$$

$$\textbf{E-W:}\ q_W^{(r)} \propto \exp\left(E_{q_Z^{(r)}}[\log p(\omega|\mathbf{y}, \mathbf{Z}; \psi^{(r)})]\right)\ . \quad (4)$$

The corresponding **M-step** is

$$\textbf{M:}\ \psi^{(r+1)} = \arg\max_{\psi \in \Psi} E_{q_Z^{(r)} q_W^{(r)}}[\log p(\mathbf{y}, \mathbf{Z}, W; \psi)]\ . \quad (5)$$

It follows from the Markovianity of the joint distribution (1) that any conditional distribution is also Markovian. In particular, $p(\mathbf{z}|\mathbf{y}, \omega; \psi)$ is Markovian with energy

$$H(\mathbf{z}|\mathbf{y}, \omega; \psi) = H_{\mathbf{Z}}(\mathbf{z}; \beta) + \sum_{i \in V} \log g(\mathbf{y}_i|\mathbf{z}_i, \omega_i; \phi),$$

omitting terms independent of $\mathbf{z}$. This latter expression is linear in ω, so that the solution of (3) is

$$q_Z^{(r)}(\mathbf{z}) = p(\mathbf{z}|\mathbf{y}, E_{q_W^{(r-1)}}[W]; \psi^{(r)}).$$

For most models with dependencies, the Markov probability $p(\mathbf{z}|\mathbf{y}, \omega; \psi)$ is intractable, but a number of approximation techniques are available. In particular, we use a mean-field like algorithm as described in [4]. Denoting $E_{q_W}^{(r-1)}[W_{im}]$ by $\bar{\omega}_{im}^{(r-1)}$, at each iteration, $q_Z^{(r)}(\mathbf{z})$ is approximated by

$$\tilde{q}_Z^{(r)}(\mathbf{z}) \propto \prod_{i \in V} \prod_{m=1}^{M} \mathcal{G}(y_{im}; \mu_{\mathbf{z}_i m}^{(r)}, \frac{s_{\mathbf{z}_i m}^{(r)}}{\bar{\omega}_{im}^{(r-1)}}) \quad \times \quad p(\mathbf{z}_i|\tilde{\mathbf{z}}_{\mathcal{N}(i)}^{(r)}; \beta^{(r)}), \quad (6)$$

where $\tilde{\mathbf{z}}^{(r)}$ is a particular configuration of $\mathbf{Z}$ updated at each iteration according to a specific scheme and $p(\mathbf{z}_i|\tilde{\mathbf{z}}_{\mathcal{N}(i)}^{(r)}; \beta^{(r)}) \propto \exp(\xi_{i\mathbf{z}_i}^{(r)} + \sum_{j \in \mathcal{N}(i)} \eta^{(r)} \langle \mathbf{z}_i, \tilde{\mathbf{z}}_j^{(r)} \rangle)$ (see [4] for details).

Of the three different schemes available in [4], all correspond to a product approximation (6), however only the mean field scheme can be seen as a standard variational approximation.

Similarly, for step (4), the conditional distribution $p(\omega|\mathbf{y}, \mathbf{z}; \psi)$ is Markovian with energy

$$H(\omega|\mathbf{y}, \mathbf{z}; \psi) = H_W(\omega) + \sum_{i \in V} \log g(\mathbf{y}_i|\mathbf{z}_i, \omega_i; \phi).$$

The choice of independent Gamma distributions for $H_W(\omega)$, has the advantage of producing a product of independent *conjugate* Gamma distributions for $q_W^{(r)}(\omega)$. The expectation $E_{q_W^{(r)}}[W_{im}]$ denoted by $\bar{\omega}_{im}^{(r)}$ becomes:

$$\bar{\omega}_{im}^{(r)} = \frac{\alpha_{im} + \frac{1}{2}}{\gamma_{im} + \frac{1}{2} \sum_{k=1}^{K} \delta(y_{im}, \mu_{km}^{(r)}, s_{km}^{(r)})\ q_{Z_i}^{(r)}(k)}, \quad (7)$$

where $\delta(y, \mu, s) = (y - \mu)^2/s$ is the squared Mahalanobis distance between y and μ (when the variance is s) and $q_{Z_i}^{(r)}(k)$ stands for $q_Z^{(r)}(\mathbf{Z}_i = k)$.

For a given sequence m, expression (7) shows that the expected weight $\bar{\omega}_{im}^{(r)}$ at voxel i is lower when $\bar{\delta}(y_{im}) = \sum_{k=1}^{K} \delta(y_{im}, \mu_{km}^{(r)}, s_{km}^{(r)})\ q_{Z_i}^{(r)}(k)$ is higher. The quantity denoted by $\bar{\delta}(y_{im})$ can be interpreted as the expectation, with regard to q_{Z_i} of the squared Mahalanobis distance between y_{im} and the mean of its tissue class. A large $\bar{\delta}(y_{im})$ is typical of a model outlier while a small $\bar{\delta}(y_{im})$ corresponds to an inlier. The value of $\bar{\omega}_{im}^{(r)}$ also depends on the expert weight through $\alpha_{im} = \gamma_{im}\omega_{im}^{exp} + 1$. The higher ω_{im}^{exp} the higher $\bar{\omega}_{im}^{(r)}$ is. It appears then that the value of the expected weight as given by (7) is a balance between the atypicality of the voxel and the expert weighting for this voxel.

The M-step (5) can be divided into two independent steps leading respectively to $\beta^{(r+1)}$ and $\phi^{(r+1)}$. The maximization over β corresponds to the M-step obtained for a standard Hidden MRF model and can be solved using a mean field like approximation as in [4]. For the updating of ϕ, the maximization leads straightforwardly to

$$\begin{aligned} \mu_{km}^{(r+1)} &= \frac{\sum_{i=1}^{N} q_{Z_i}^{(r)}(k)\ \bar{\omega}_{im}^{(r)}\ y_{im}}{\sum_{i=1}^{N} q_{Z_i}^{(r)}(k)\ \bar{\omega}_{im}^{(r)}}, \\ s_{km}^{(r+1)} &= \frac{\sum_{i=1}^{N} q_{Z_i}^{(r)}(k)\ \bar{\omega}_{im}^{(r)}\ (y_{im} - \mu_{km}^{(r+1)})^2}{\sum_{i=1}^{N} q_{Z_i}^{(r)}(k)}\ , \end{aligned}$$

where it appears that voxels with small expected weights have small impact on the Gaussian parameter values. As justified in [15] par.7.5.3, we can then replace the divisor $\sum_{i=1}^{N} q_{Z_i}^{(r)}(k)$ in the $s_{km}^{(r+1)}$ formula by $\sum_{i=1}^{N} q_{Z_i}^{(r)}(k)\, \bar{\omega}_{im}^{(r)}$.

Related models. It appears that equation (7) and the Gaussian parameters updating formulas have forms similar to the equations derived in the case of K-component mixtures of M-dimensional t-distributions [15]. However, our model cannot be reduced to a standard mixture model due to the non *i.i.d.* nature of the observed data. The class memberships $\mathbf{z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$ are not independent but linked trough a Markov distribution and more importantly, our weight distributions depend on the MR sequence m but also on the voxel location i, this latter dependency being crucial for lesion detection.

4 Lesion segmentation procedure

The weight field depends on some expert weights to be fixed so as to incorporate expert knowledge. In practice, such expert knowledge may not be available or at least difficult to translate into weight values. Therefore, following the intuition that lesion voxels should be weighted more to be appropriately identified, we propose to set the expert weights to a value $\omega_{\mathcal{L}}$ greater than 1 for all voxels in a region $\mathcal{L}$ while the others are weighted 1. To determine region $\mathcal{L}$, we can make use of any method that defines *candidate lesions*, *eg.* manual seeding by an expert in a semi-supervised framework or any outlier detection technique (*eg.* [15] 7.3 or [22]). In our framework, we propose to apply our algorithm with $K = 3$, considering only the three normal tissue classes (with all ω_{im}^{exp} and γ_{im} set to 1). In this preliminary step, the ξ parameters in the MRF prior are set to $\xi_{ik} = \log f_{ik}$ where f_{ik} is the normalized value given by a normal tissue atlas. We used the ICBM452 probabilistic atlas from The International Consortium for Brain Mapping[1]. The interaction parameter η is estimated as specified in [4] and using a stochastic gradient descent. In such a three-class model, lesions voxels are likely to appear as outliers and then assigned a low weight. Identifying these outliers can then be done directly by thresholding the estimated weights as given by (7) or equivalently the expected Mahalanobis distances $\bar{\delta}(y_{im})$. In our experiments, the threshold for the weights is found using a chi-squared percentile to be specified below (see [15] 7.3 for justification). This percentile choice determines a threshold by indicating how large $\bar{\delta}(y_{im})$ has to be in order for y_{im} to be classified as outlier. The region $\mathcal{L}$ is deduced by thresholding the more informative weight map (according to medical expertise, T1, Flair or DW in our experiments). We note that the selected voxels are not just outliers with respect to the tissue model, but also with respect to the MRF and prior atlas model. The candidate region is refined using additional intensity constraints, as in [22, 9, 10]. For example for MS, lesions which are not in white matter and too small lesions (typically less than 3mm) are removed.

The lesion segmentation can then be carried out using our model with $K = 4$ classes. In this second step, we propose to set $\gamma_{im}s$ according to: $\gamma_{im} = \gamma_{\mathcal{L}}$ for all $i \in \mathcal{L}$ and $\gamma_{im} = \gamma_{\bar{\mathcal{L}}}$ for all $i \notin \mathcal{L}$, where $\gamma_{\mathcal{L}}$ and $\gamma_{\bar{\mathcal{L}}}$ are values to be specified. The γ_{im} parameters are related to the variance of the weight distributions and therefore express the confidence in prior expert knowledge. A high γ_{im} induces a higher impact of the prior in the estimation process. In practice, this means that a high γ_{im} will constrain successive weight estimations to the vicinity of the initial expert weight ω_{im}^{exp}. We propose setting $\gamma_{\bar{\mathcal{L}}}$ to a high value (*e.g.* $\gamma_{\bar{\mathcal{L}}} = 1000$) so as to express our a priori trust in the estimation of the normal brain tissue classes from the preliminary first step. The $\gamma_{\mathcal{L}}$ value is then set to a lower value to allow the corresponding average weights $\bar{\omega}_{im}$ to vary more. As regards the external field parameters ξ, they are considered constant over voxels so that ξ reduces to a single vector ${}^t(\xi_i \ldots \xi_K)$, that is estimated. The interaction parameter η is also estimated. Then, to carry out the lesion segmentation algorithm with $K = 4$, an initial classification is required. It is computed, from the result of the preceeding 3-class segmentation, by assigning all the voxels in $\mathcal{L}$ to the lesion class and by using the final segmentation obtained with the 3-class model to initialize the normal tissue classes. It follows that our method requires the choice of three values: the chi-squared percentile, the weight value $\omega_{\mathcal{L}}$ and the $\gamma_{\mathcal{L}}$ value that expresses confidence in the prior. Using simulated data, we first perform a number of experiments to assess the sensitivity of our method to these parameter values. Following this analysis, not reported here, we propose to set the chi-squared percentile as follows, so as to limit the number of false positives in $\mathcal{L}$. For different levels of the chi-squared percentile varying from 99% to 99.999%, we compute the percentage of voxels in the corresponding $\mathcal{L}$. We then select the region $\mathcal{L}$ for which this percentage is closest to 0.4%. If this percentage is greater than 0.4%, the expert weight $\omega_{\mathcal{L}}$ is fixed to $\omega_{\mathcal{L}} = 2$ and it is set to $\omega_{\mathcal{L}} = 10$ otherwise. Large values of $\omega_{\mathcal{L}}$ make the lesion class more representative and handles the possibility of very small lesions, while a small $\omega_{\mathcal{L}}$ ensures that the weighting of a large candidate lesion region does not affect the estimation of other classes. We set $\gamma_{\mathcal{L}} = 10$ to allow some flexibility in the weight estimation.

[1] http://www.loni.ucla.edu/Atlases/

5 Results

Results on simulated data. The simulated data we consider are 1 mm^3 phantoms from the Brain Web database[2] [5]. The database contains simulated images with multiple sclerosis (MS) lesions for varying levels of nonuniformity and noise. We use three co-registered modalities (T1-weighted, T2-weighted and PD sequences), with three types of lesions (mild, moderate and severe). The noise level is set successively to 3%, 5%, 7% and 9%. The nonuniformity level is set to 0% and 40%. Note that in practice real data sets rarely show noise and inhomogeneity levels exceeding 9% and 20%. As in [1, 21, 23] we perform a quantitative evaluation using the Dice similarity coefficient (DSC) [8]. This coefficient measures the overlap between a segmentation result and the gold standard. Denoting by TP_k the number of true positives for class k, FP_k the number of false positives and FN_k the number of false negatives the DSC is given by: $d_k = \frac{2\mathrm{TP}_k}{2\mathrm{TP}_k+\mathrm{FN}_k+\mathrm{FP}_k}$ and d_k takes its value in $[0, 1]$ where 1 represents the perfect agreement. The obtained Dice similarity coefficients (DSC) [8] are reported in Table 1. To our knowledge, very few other BrainWeb MS results exist in the literature. A recent paper [10] provides results for all lesion loads, three non-uniformity levels and the above levels of noise. Another paper [17] reports results for the three lesion loads but only in the 3% noise 0% nonuniformity case while [9] reports DSC values for all above levels of noise but only in the moderate lesion case. Note however, that the results in [9] cannot be directly compared to ours as a number of slices were removed (61 of 181 were kept) before processing, which tends to produce DSC's overestimated by a few percent. For the 0% non-uniformity and 3%, 5%,7% and 9% noise cases reduced to 61 slices, [9] obtained respectively 77%, 77%, 75% and 73%. We then apply the method described in [22] using the EMS software package[3]. We use the default setting of a 3D polynomial of order 4 with a Mahalanobis threshold of 3. The results from these methods and the gain in DSC values using ours are reported in Table 1. These values show the good performance of our method in the 0% inhomogeneity case. However, it appears that the segmentation quality decreases as the noise level increases. For comparison, when applying our method on data previously denoised with the Non Local Means method [6], we observe significantly higher DSC's especially for 7% and 9% noise levels and mild lesions. For a high inhomogeneity level (40%), we observe satisfying results considering that our method is not explicitly designed to account for such distortions in contrast to the methods in [22] and [10] which include some bias field modelling. The performance of our method however can be significantly impacted especially in the mild lesion cases. Although a common solution then is to pre-process the data for inhomogeneity intensity correction, dealing with nonuniformities is important when delineating lesions and we mention in Section 6 directions for improvement of our model in future work.

Real data sets. We use co-registered T1-weighted, T2-weighted and Flair sequences (voxel size $1mm^2 \times 3mm$) from five patients with MS for which the lesions were manually delineated by an expert. As an illustration, we show also how the method can be applied on a stroke data set for which three co-registered sequences (T2, DW and Flair) were available (voxel size $1mm^2 \times 5mm$) . In all cases, the images are preprocessed for denoising [6] and inhomogeneity intensity correction [14].

For the MS data, the results using our method and the EMS package [22] are reported in Table 2. They correspond to an average DSC of 60%+/-16 for our method and of 55%+/-8 for EMS. The scores with our method vary from 82% to 45%. For the poorest scores, we observe a high number of false positives that heavily impact on the DSC as the lesions are small (patients 3 and 5). They also suggest that a stronger Markov interaction parameter could be more appropriate for these data sets. However, the percentage of lesion voxels does not seem to explain satisfyingly our scores variations. The latter are more likely due to poor contrast between lesions and normal tissues, and variations in the locations and shapes of the lesions. Further investigations are needed. As an illustration of the different possible aspects of MS lesions, the segmentations for Patient 1 and 3 are shown in Figures 1 and 2. Our method gives better results than the EMS software in 3 cases and equivalent ones (2% loss) in the 2 others. Comparison with results in [10] suggests that our method performs satisfyingly considering the difficulty of the task. Indeed the authors in [10] obtain, for three sequences, T1-w, T2-w and PD, an average DSC of 63%+/-17 over 10 data sets and report an average DSC of 59%+/-16 using EMS. The average gain over EMS in [10] is then equivalent to our which is promising considering that the lesion loads in our data sets are on average smaller than those in [10]. For additional comparison, for three sequences, T1-w, T2-w and PD, the authors in [22] report a best DSC of 45% over 20 data sets. In [18], using PD, T2 and Flair images, an average DSC of 78%+/-12 over 23 data sets is obtained with a method that benefits from human intervention in addition to parameter tuning.

The result obtained for the stroke example is shown in Figure 3 (b) (the DSC is 63%). To report other results

[2] http://www.bic.mni.mcgill.ca/brainweb/

[3] http://www.medicalimagecomputing.com/EMS

Method	3%	5%	7%	9%
Mild lesions (0.02% of the voxels)				
AWEM	**72 (+5)**	55 (-15)	**39 (+5)**	**22 (+18)**
[10]	67	**70**	34	0
[22]	56	33	13	4
[17]	52	NA	NA	NA
Moderate lesions (0.18% of the voxels)				
AWEM	**86 (+7)**	80 (-1)	**77 (+18)**	**73 (+36)**
[10]	72	**81**	59	29
[22]	79	69	52	37
[17]	63	NA	NA	NA
Severe lesions (0.52% of the voxels)				
AWEM	**93 (+8)**	88 (0)	**78 (+6)**	**74 (+33)**
[10]	79	**88**	72	41
[22]	85	72	56	41
[17]	82	NA	NA	NA

0% non-uniformities

-

Method	3%	5%	7%	9%
Mild lesions (0.02% of the voxels)				
AWEM	50 (-25)	0 (-65)	0 (-20)	0 (-30)
[10]	**75**	**65**	**20**	**30**
[22]	58	27	13	6
Moderate lesions (0.18% of the voxels)				
AWEM	64 (-12)	66 (-10)	66 (-1)	0 (-48)
[10]	75	**76**	**67**	**48**
[22]	**76**	64	47	31
Severe lesions (0.52% of the voxels)				
AWEM	**88 (+2)**	**84 (+1)**	**80 (+6)**	**67 (+9)**
[10]	75	83	74	58
[22]	86	74	62	45

40% non-uniformities

Table 1: DSC results (%) on MS Brain Web simulated data, for various lesion sizes, noise and non-uniformity levels. Comparison of different methods: AWEM for our Adaptive Weighted EM, Garcia-Lorenzo & al's method [10], Van Leemput & al.'s method [22] and Rousseau & al.'s method [17]. The corresponding gain/loss over the best comparable results is given in parenthesis. NA means not available.

on stroke lesions, using single T1 sequences for eight real cases, an average DSC of 64%+/-10 is found in [20].

6 Discussion

We propose an adaptive weighting scheme for multiple MR sequences for Brain lesion segmentation. It uses a tissue segmentation model and is not specific to a lesion type. Our approach differs from the mainstream approaches in that lesion voxels are not considered solely as outliers to be detected. The general idea is to first identify voxels that would not be well represented by a normal tissue model in order to use them to bias, via the weight prior specification, the model toward the identification of a lesion class. The weight prior, and in particular the expert weights, require selection.

	LL	EMS	AWEM
Patient1	0.42	62	**82 (+20)**
Patient2	1.71	54	**56 (+2)**
Patient3	0.29	**47**	45 (-2)
Patient4	1.59	65	**72 (+7)**
Patient5	0.31	**47**	45 (-2)
Average		55 +/-8	60 +/-16

Table 2: Lesion load or percentage of lesion voxels (LL), DSC results (%) for Van Leemput & al.'s method [22] (EMS) and for our Adaptive Weighted EM (AWEM), for 5 patients with MS.

We propose an automatic selection method and future work includes investigating other settings, particularily in relation to targeting specific lesion types. More generally, robustness to intensity nonuniformity could be provided by incorporating the local estimation principle of [19]. Another interesting generalization would be to use full covariance matrices for the Gaussian intensity distributions to handle possible strong correlations between the different sequences. They would result in more complex estimation formulas but this could be an important refinement for applications using temporal multi-sequence data (*eg.* follow-up of lesions). As a generalization of Gaussian mixtures, our model has larger modelling capabilities. It is entirely based on a mathematical framework in which each step is theoretically well-founded. Its ability to provide good results, when application related expertise is difficult to formalize, is particularly promising for medical applications. Therefore, it has advantages over other methods that include ad-hoc processing while being open to incorporation of more task dependent information.

References

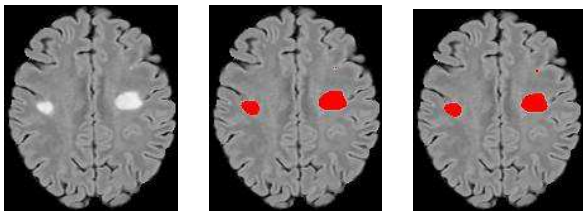
[1] J. Ashburner and K. J. Friston. Unified Segmentation. *NeuroImage*, 26:839–851, 2005.

[2] M. Beal and Z. Ghahramani. *The variational Bayesian EM Algorithm for incomplete data: with application to scoring graphical model structures.* Bayesian Statistics. Oxford University Press, 2003.

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

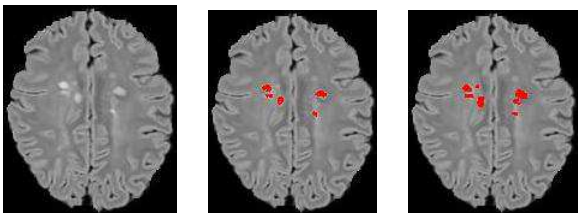
[4] G. Celeux, F. Forbes, and N. Peyrard. EM procedures using mean field-like approximations for Markov model-based image segmentation. *Pat. Rec.*, 36(1):131–144, 2003.

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



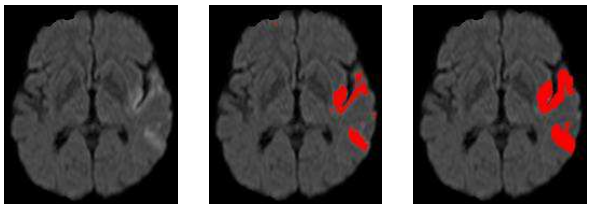
(a) (b) (c)

Figure 1: Real MS data, patient 1. (a): Flair image. (b): identified lesions with our approach (DSC 82%). (c): ground truth .



(a) (b) (c)

Figure 2: Real MS data, patient 3. (a): Flair image. (b): identified lesions with our approach (DSC 45%). (c): ground truth .



(a) (b) (c)

Figure 3: Real stroke data. (a): DW image. (b): identified lesions with our approach (DSC 63%). (c): ground truth.

[6] P. Coupe, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot. An Optimized Blockwise Non Local Means Denoising Filter for 3D Magnetic Resonance Images. *IEEE Transactions on Medical Imaging*, 27(4):425–441, April 2008.

[7] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 39:1–38, 1977.

[8] L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26:297–302, 1945.

[9] O. Freifeld, H. Greenspan, and J. Goldberger. Lesion detection in noisy MR brain images using constrained GMM and active contours. In *IEEE ISBI*, pages 596–599, 2007.

[10] D. Garcia-Lorenzo, L. Lecoeur, D.L. Arnold, D. L. Collins, and C. Barillot. Multiple Sclerosis lesion segmentation using an automatic multimodal graph cuts. In *MICCAI*, pages 584–591, 2009.

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

[12] C. Genest, K. J. McConway, and M. J. Schervish. Characterization of externally Bayesian pooling operators. *Ann. Statist.*, 14:487–501, 1986.

[13] Z. Ghahramani and M. Beal. Propagation algorithms for variational Bayesian learning. In *Advances in Neural Information Processing Systems. MIT Press*, 2001.

[14] J-F. Mangin. Entropy minimization for automatic correction of intensity nonuniformity. *Mathematical Methods in Biomedical Image Analysis, IEEE Workshop on*, 0:162, 2000.

[15] G.J. McLachlan and D. Peel. *Finite Mixture Models*. Wiley, 2000.

[16] R.M. Neal and G.E. Hinton. A view of the EM algorithm that justifies incremental, sparse and other variants. In Jordan, editor, *Learning in Graphical Models*, pages 355–368. 1998.

[17] F. Rousseau, F. Blanc, J. de Seze, L. Rumbac, and J.P. Armspach. An a contrario approach for outliers segmentation: application to multiple sclerosis in MRI. In *IEEE ISBI*, pages 9–12, 2008.

[18] B. R. Sajja, S. Datta, R. He, M. Mehta, R. K. Gupta, J. S. Wolinsky, and P. A. Narayana. Unified approach for multiple sclerosis lesion segmentation on brain MRI. *Ann. Biomed. Eng.*, 34(1):142–151, 2006.

[19] B. Scherrer, F. Forbes, C. Garbay, and M. Dojat. Fully Bayesian Joint Model for MR Brain Scan Tissue and Structure Segmentation. In *MICCAI*, pages 1066–1074, 2008.

[20] M.L. Seghier, A. Ramlackhansingh, J. Crinion, A.P. Leff, and C. J. Price. Lesion identification using unified segmentation-normalisation models and fuzzy clustering. *Neuroimage*, 41:1253–1266, 2008.

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

[22] K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE trans. Med. Ima.*, 20(8):677–688, 2001.

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