Adaptive weighted fusion of multiple MR sequences for brain lesion segmentation

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

To cite this version:

HAL Id: inserm-00723807
https://www.hal.inserm.fr/inserm-00723807
Submitted on 14 Aug 2012

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.
ADAPTIVE WEIGHTED FUSION OF MULTIPLE MR SEQUENCES 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 a Bayesian multi-sequence Markov model that includes weight parameters to account for the relative importance and control the impact of each sequence. The Bayesian 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.

Index Terms— MRI, segmentation, brain lesion, Bayesian model, MRF, variational EM

1. INTRODUCTION

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 [13] for review) to alleviate the tedious, time consuming and subjective nature of manual delineation. Automated or semi-automated brain lesion detection methods can be classified according to their segmentation task under consideration. The three terms in this energy are further specified below.

In this paper, rather than trying to detect lesion voxels as outliers from a normal tissue model, we adopt an incorporation strategy. 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 (e.g. [14]) 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 weight parameters in the segmentation model and then solve the issue of prescribing values for these weights by developing a Bayesian 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 prior 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 prior 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

We consider a set \( V \) of \( N \) voxels and denote by \( y = \{y_1, \ldots, y_N\} \) the intensity values observed at each voxel. Each \( y_i = \{y_{i1}, \ldots, y_{iM}\} \) is itself a vector of \( M \) intensity values corresponding to \( M \) different MR sequences. For brain tissue segmentation, we denote by \( z = \{z_1, \ldots, z_N\} \) the hidden classes (generally 3 tissues plus some possible additional classes to account for lesions) and by \( Z \) the set in which \( z \) takes its values. Typically, the \( z_i \)'s take their values in \( \{1, \ldots, K\} \). We consider nonnegative weights \( \omega = \{\omega_i, i \in V\} \) in a state space denoted by \( \mathcal{V} \) 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. Denoting \( \psi = \{\beta, \phi\} \) additional parameters, we assume that the joint distribution \( p(y, z, \omega; \psi) \) is a MRF with the following energy function:

\[
H(y, z, \omega; \psi) = H_z(z; \beta) + H_w(\omega) + \sum_{i \in V} \log g(y_i | z_i, \omega_i; \phi)
\]  

(1)

where the energy term \( H_w(\omega) \) involving only \( \omega \) does not depend on \( \psi \) and the \( g(y_i | z_i, \omega_i; \phi) \)s are probability density functions of \( y_i \).

The three terms in this energy are further specified below.

Data term. The data term \( \sum_{i \in V} \log g(y_i | z_i, \omega_i; \phi) \) in (1) corresponds to the modelling of tissue dependent intensity distributions. We consider M-dimensional Gaussian distributions with diagonal covari-
ance matrices. For each class \( k \), \( \{ \mu_{k,1}, \ldots, \mu_{k,M} \} \) is the mean vector and \( \{ s_{1,k}, \ldots, s_{M,k} \} \) the covariance matrix components. When \( z_i = k \), then \( G(y_{im}; \mu_{km}, s_{km}) \) represents the Gaussian distribution with mean \( \mu_{km} \) 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

\[
g(y; x, \phi) = \prod_{m=1}^{M} G(y_{im}; \mu_{km}, s_{km})
\]

which is proportional to \( \prod_{m=1}^{M} G(y_{im}; \mu_{km}, s_{km})^{w_{im}} \). Intuitively, the impact of a larger \( \omega_{im} \) is to give more importance to the intensity value \( y_{im} \) in the model. Typically an integer \( \omega_{im} \) greater than one would correspond to increase \( \omega_{im} \) times the number of voxels with intensity value \( y_{im} \). When the weights are all one, a standard multivariate Gaussian case is recovered.

**Missing data term.** The missing data term \( H_2(z; \beta) \) in (1) is set to a standard Potts model, with external field \( \xi \) and spatial interaction parameter \( \eta \), whose energy is

\[
H_2(z; \beta) = \sum_{i \in V} (\xi(x_i, \sum_{j \in N(i)} \eta(x_i, x_j)))
\]

where \( N(i) \) denotes the voxels neighboring \( i \) and \( (x_i, x_j) \) is 1 when \( x_i = x_j \) and 0 otherwise. \( \beta = \{ \xi, \eta \} \) with \( \xi = \{ \xi_1, \ldots, \xi_K \} \), \( t \in V \) being a set of real-valued \( K \)-dimensional vectors and \( \eta \) a real positive value.

**Parameter prior term.** The weights are assumed independent from the other parameters \( \psi \) 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 \( \alpha_{im} \) and \( \gamma_{im} \) (shape and inverse scale, respectively). This corresponds to

\[
H_3(\omega) = \sum_{m=1}^{M} \sum_{i \in V} (\alpha_{im} - 1) \log \omega_{im} - \gamma_{im} \omega_{im}
\]

In practice, the set of hyperparameters is fixed so that the modes of each prior \( \omega_{im} \) located at some expert weights \( \{ \omega_{exp,m}, m = 1 \ldots M, i \in V \} \) accounting for some external knowledge, if available. Formally, we set \( \alpha_{im} = \gamma_{im} \omega_{exp,m} + 1 \). 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.

3. **ESTIMATION BY VARIATIONAL BAYESIAN EM**

The joint distribution of the observations \( y \) and the missing variables \( z \) is governed by the weights \( \omega \in \mathcal{W} \) and parameters \( \psi \in \Psi \), which are both unknown and need to be estimated within the segmentation procedure. Our framework must then provide point estimates for the \( \psi \) while accounting for prior knowledge on \( \omega \). A natural solution is to adopt an Expectation-Maximization (EM) framework in which the weights \( \omega \) are considered as additional missing variables, but this results in an intractable E-step for our model. We therefore propose to use an EM variant similar to the Variational Bayesian EM presented in [1]. Our framework varies slightly from the case of conjugate exponential models described in [1] and more importantly, our presentation offers the possibility to deal with extra parameters (the Gaussian means and variances in our setting) for which no prior information is available. As a consequence, the variational Bayesian M-step of [1] is transferred into our E-step while our M-step has no equivalent in the formulation of [1]. It follows that our E-step becomes an approximate E-step which can be further decomposed into two stages, respectively E-Z and E-W in which the goal is to update probability distributions \( q_Z \) and \( q_W \) in turn on \( Z \) and \( W \) respectively. At iteration \( r \), with \( q_Z^{(r-1)} \) and \( q_W^{(r-1)} \) representing current estimations, the E-step decomposes into:

**E-Z:**

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

**E-W:**

\[
E(W) \quad q_W^{(r)}(w) \propto \exp \left( E[q_Z^{(r)}] \log p(w, Z; \psi^{(r)}) \right)
\]

The corresponding **M-step** is

\[
M: \quad \psi^{(r+1)} = \arg \max_{\psi \in \Psi} E[q_W^{(r)}] \log p(y, W, \psi)
\]

Then step (2) leads to

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

which remains intractable but for which a number of approximation techniques are available. In particular, we use a mean-field like algorithm as described in [3]. Among the three different schemes available in [3], all correspond to a product approximation but only one (the so-called mean field scheme) can be seen as a standard variational approximation.

In step (3), the choice of independent Gamma priors for \( H_2(\omega) \), has the advantage of producing a product of independent conjugate Gamma distributions for \( q_Z^{(r)}(w) \). The expectation \( E[q_W^{(r)}] W_{im} \) denoted by \( \omega_{im}^{(r)} \) becomes:

\[
\omega_{im}^{(r)} = \frac{\alpha_{im} + \frac{1}{2}}{\gamma_{im} + \frac{1}{2} \sum_{k=1}^{K} \delta(y_{im}, s_{km}) q_{Z_{im}}^{(r)}(k)}
\]

where \( \delta(y, s) = (y - s)^2/s \) is the squared Mahalanobis distance between \( y \) and \( s \) (when the variance is \( s \) and \( q_{Z_{im}}^{(r)}(k) \) stands for \( q_{Z_{im}}^{(r)}(Z_i = k) \). For a given sequence \( m \), expression (5) shows that the expected weight \( \omega_{im}^{(r)} \) at voxel \( i \) is lower when \( \delta(y_{im}) = \sum_{k=1}^{K} \delta(y_{im}, s_{km}) q_{Z_{im}}^{(r)}(k) \) is higher. The quantity denoted by \( \delta(y_{im}) \) can be interpreted as the expectation, with regard to \( q_{Z_{im}}^{(r)} \) of the Mahalanobis distance between \( y_{im} \) and the mean of its tissue class. A large \( \delta(y_{im}) \) is typical of a model outlier while a small \( \delta(y_{im}) \) corresponds to an inlier. The value of \( \omega_{im}^{(r)} \) also depends on the expert weight through \( \alpha_{im} = \gamma_{im} \omega_{exp,m} + 1 \). The higher \( \omega_{im}^{(r)} \) the higher \( \omega_{exp,m} \) is. It appears then that the value of the expected weight as given by (5) is a balance between the atypicality of the voxel and the expert weighting for this voxel.

The M-step (4) can be divided into two independent steps leading respectively to \( \beta^{(r+1)} \) and \( \gamma^{(r+1)} \). The maximization over \( \beta \) corresponds to the M-step obtained for a standard Hidden MRF model and can be solved using a mean field like approximation as in [3]. For the updating of \( \gamma \), the maximization leads straightforwardly to

\[
H_{km}^{(r+1)} = \sum_{i=1}^{N} q_{Z_{im}}^{(r)}(k) \omega_{im}^{(r+1)} y_{im}
\]

and

\[
H_{km}^{(s_{km})} = \sum_{i=1}^{N} q_{Z_{im}}^{(r)}(k) \omega_{im}^{(r+1)} \left( y_{im} - \mu_{km}^{(r+1)} \right)^2
\]

where it appears that voxels with small expected weights have small impact on the Gaussian parameter values. As justified in [9] par.7.5.3, we can then replace the divisor \( \sum_{i=1}^{N} q_{Z_{im}}^{(r)}(k) \) in the \( s_{km}^{(r+1)} \) formula by

\[
\sum_{i=1}^{N} q_{Z_{im}}^{(r)}(k) \omega_{im}^{(r+1)}
\]
4. RESULTS

Lesion segmentation procedure. Following the intuition that lesion voxels should be weighted more to be appropriately identified, we propose to set the expert weights to a value $\omega_L$ greater than 1 for all voxels in a region $L$ while the others are weighted 1. To determine region $L$, we propose to apply our algorithm with $K = 3$, considering only the three normal tissue classes (with all $\omega_{im}$ and $\gamma_{im}$ set to 1). In this preliminary step, the $\xi$ 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. The interaction parameter $\eta$ is also fixed as specified below. 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 (5) or equivalently the expected Mahalanobis distances $\delta(y_{im})$. In our experiments, the threshold for the weights is found using a chi-squared percentile (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 [14, 6, 7].

The lesion segmentation can then be carried out using our model with $K = 4$ classes. In this second step, we propose to set $\gamma_{im}$ according to $\gamma_{im} = \gamma_L$ for all $i \in V$ and $\gamma_{im} = \gamma_C$ for all $i \notin V$, where $\gamma_L$ and $\gamma_C$ are values to be specified. The $\gamma_{im}$ parameters are related to the variance of the weight priors and therefore express the confidence in prior expert knowledge. A high $\gamma_{im}$ reduces a priori trust in the estimation of lesion segments only in the moderate lesion case. Note however, that the results in [6] 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 3%, 5%, 7% and 9% noise cases reduced to 61 slices, [6] obtained respectively 77%, 75%, and 73%. We then apply the method described in [14] using the EMS software package. 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. 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 [4], we observe significantly higher DSC's especially for 7% and 9% noise levels and mild lesions.

Real data sets. We use co-registered T1-weighted, T2-weighted and Flair sequences (voxel size $1\text{mm}^3 \times 3\text{mm}$) from three patients with MS for which the lesions were manually delineated by an expert. As an illustration, we show 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 $1\text{mm}^3 \times 5\text{mm}$). In all cases, the images are pre-processed for denoising and inhomogeneity intensity correction [8]. For the MS data, we obtain DSC values of 82% (patient 1), 56% (patient 2) and 45% (patient 3) using our method and 57%, 54% and 46% using the EMS package [14]. The segmentation for patient 3 is shown in Figure 1. For comparison, for three sequences, T1-w, T2-w and PD, the authors in [14] report a best DSC of 45% over 20 data sets. In [11], using PD, T2 and Flair images, an average DSC of 78%+/-12 over 23 data sets is obtained with a method that requires human intervention in addition to parameter tuning. The result obtained for the stroke example is shown in Figure 2 (b) (the DSC is 63%). To report other results on stroke lesions, using single T1 sequences for eight real cases, an average DSC of 64%+/-10 is found in [13].

5. 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.

---

1http://www.loni.ucla.edu/Atlases/Atlas_Detail.jsp?atlas_id=6

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

3http://www.medicalimagecomputing.com/EMS
### Table 1. DSC results (%) on MS Brain Web simulated data, for various lesion sizes and noise levels. Comparison of different methods: AWEM for our Adaptive Weighted EM, Garcia-Lorenzo & al.’s method [7], Van Leemput & al.’s method [14] and Rousseau & al.’s method [10]. The corresponding gain/loss over the best comparable results is given in parenthesis. NA means/loss over the best comparable results is given in parenthesis. NA means not available.

<table>
<thead>
<tr>
<th>Method</th>
<th>3%</th>
<th>5%</th>
<th>7%</th>
<th>9%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mild lesions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AWEM</td>
<td>68 (+1)</td>
<td>49 (-21)</td>
<td>36 (+2)</td>
<td>12 (+8)</td>
</tr>
<tr>
<td>[7]</td>
<td>67</td>
<td>70</td>
<td>34</td>
<td>0</td>
</tr>
<tr>
<td>[14]</td>
<td>56</td>
<td>33</td>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>[10]</td>
<td>52</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Moderate lesions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AWEM</td>
<td>86 (+7)</td>
<td>80 (-1)</td>
<td>73 (+14)</td>
<td>64 (+27)</td>
</tr>
<tr>
<td>[7]</td>
<td>72</td>
<td>81</td>
<td>59</td>
<td>29</td>
</tr>
<tr>
<td>[14]</td>
<td>79</td>
<td>69</td>
<td>52</td>
<td>37</td>
</tr>
<tr>
<td>[10]</td>
<td>63</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Severe lesions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AWEM</td>
<td>92 (+7)</td>
<td>86 (-2)</td>
<td>78 (+6)</td>
<td>68 (+27)</td>
</tr>
<tr>
<td>[7]</td>
<td>79</td>
<td>88</td>
<td>72</td>
<td>41</td>
</tr>
<tr>
<td>[14]</td>
<td>85</td>
<td>72</td>
<td>56</td>
<td>41</td>
</tr>
<tr>
<td>[10]</td>
<td>82</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Fig. 1. Real MS data, patient 3. (a): Flair image. (b): identified lesions with our approach (DSC 45%). (c): ground truth.

Fig. 2. Real stroke data. (a): DW image. (b): identified lesions with our approach (DSC 63%). (c): ground truth.

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. We propose an automatic selection method and future work includes investigating other settings, particularly in relation to targeting specific lesion types. More generally, robustness to intensity nonuniformity could be provided by incorporating the local estimation principle of [12]. 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 multisquence 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.

### 6. REFERENCES


