Adapted MRF Segmentation of Multiple Sclerosis Lesions Using Local Contextual Information

Nagesh Subbanna[1]
nagesh@cim.mcgill.ca
Simon Francis[2]
simonjfrancis@gmail.com
Doina Precup[1]
dprecup@cs.mcgill.ca
Louis Collins[1]
louis@bic.mni.mcgill.ca
Doug Arnold[12]
doug@mrs.mni.mcgill.ca
Tal Arbel[1]
arbel@cim.mcgill.ca

[1] Centre for Intelligent Machines, Department of Electrical and Computer Engineering, McGill University, Montreal, Quebec, Canada
[2] NeuroRx Research, Montreal, Quebec, Canada

Abstract

We present a fully automated technique to segment lesions from multimodal brain MRIs of patients with Multiple Sclerosis. We describe an adapted Markov Random Field that uses intensity at every voxel, its neighbourhood intensity difference information and neighbouring voxel class information to infer voxel labels at every voxel. We test our technique on 25 real, clinical MS volumes evaluated by five experts. Our method outperforms two state of the art methods: one an outlier based MRF technique and the other a hybrid Bayesian-MRF technique both qualitatively and according to the Dice similarity coefficients and the number of present negative lesions.

1 Introduction

Multiple Sclerosis (MS) is an inflammatory, demyelinating disease of the Central Nervous System (CNS). Current clinical practice involves manual labelling of lesion voxels by experts. However, manual lesion segmentation has proven to be cumbersome, laborious and most importantly, inconsistent, due to significant differences in lesion segmentation by different experts (inter-rater variability) and by the same experts at different points of time (intra rater variability). Fast, consistent automatic segmentation by machines constitutes a significant advantage. The problem of segmenting MS lesions is particularly difficult as tissue class intensities vary due to noise, geometric distortions, distance from the magnet, magnetisation inhomogeneities, etc. Further, it is hard to determine the class of the voxel based solely on the tissue intensity at each voxel, since the models for different tissue classes, particularly

lesions, show considerable overlap. Finally, the shapes, volumes and sizes of lesions vary greatly across brains based on the disease stage and type, making it harder to model lesions.

Many techniques have attempted to model healthy brain tissue classes, based on their intensity profiles and define lesions as the outliers of those class models. For example, van Leemput et. al., [11] model the healthy tissues as multivariate Gaussians, Ait-Ali et. al., [5], Souplet et. al, [4] use trimmed likelihood estimators (TLE) to model healthy brain tissues, and Freifeld et. al., [6] model brain tissues with constrained Gaussian mixture model (CGMM). However, in real, clinical data, lesion distributions overlap significantly with the healthy brain tissues, making outlier based models suspect. Topological features and template matching techniques like the k-Nearest Neighbours technique [9] have also been employed to detect lesions based on similarities to a set of training volumes. However, template matching is hard, given the diversity of brain features across subjects.

Here, we devise a Bayesian technique that models all tissue classes, including lesions during training. To leverage neighbourhood information, we model local relations using Markov Random Fields (MRFs). MRFs have been previously employed by [11] and [8] for MS lesion segmentation. In addition to exploiting local contextual voxel labels as in [11], our technique differs from the previous work in that we also model the neighbourhood intensity differences for each set of classes, thus avoiding a dependency on absolute voxel intensities, and resulting in more accurate voxel class predictions. We also propose to exhaustively evaluate all cliques in our neighbourhood, instead of limiting ourselves to only the smaller cliques [11] or selected cliques [7]. As such, we hope to capture all the relations in the neighbourhood, since these relations enable us to mimic the behaviour of tissue classes which occur in groups. Finally, as proposed in [7], in order to alleviate the problem of intensity variation across the brain, we divide the brain into regions and treat each region as a separate MRF. We train our models on real clinical data and test our classification scheme against multi-expert evaluated clinical data using a five fold cross validation technique. We compare our results against state of the art techniques like those employed in [11], and [7] using Dice similarity coefficients, false positive and false negative lesions as metrics for comparison.

2 MRF based Lesion Segmentation

Our goal is to classify each voxel of the test volume into one of the M classes. Our approach is divided into two parts. In the first part, distributions needed for MRF models are estimated from pre-labeled training volumes. Next, volumes are classified. Given the K MRI image modalities, the intensity of voxel v_i is a K dimensional vector $\mathbf{I_i} = (I_i^0, \ldots, I_i^{K-1})$ with the elements corresponding to the voxel intensities in the K modalities. Let the label of each voxel v_i be modelled as a random variable f_i, $f_i \in \{0, \ldots, M-1\}$, and $v_i \in \mathscr{S}$ where $\mathscr{S}$ is a collection of sites, with each site being a voxel. Given the intensities $\mathbf{I}_i$ and $\mathbf{I}_{N_i}$ of the voxel v_i, and its neighbours v_{N_i} respectively, we need to find the class f_i of the voxel. This is given by $P(f_i \mid \mathbf{I}_i, \mathbf{I}_{N_i}) =$

$$
\begin{aligned}
\sum_{f_{N_i}=0}^{M-1} P(f_i, f_{N_i} \mid \mathbf{I}_i, \mathbf{I}_{N_i}) &\propto \sum_{j=0}^{M-1} P(\mathbf{I}_{N_i}, \mathbf{I}_i \mid f_i, f_{N_i}) P(f_i, f_{N_i}) [\text{Since } P(\mathbf{I}_{N_i}, \mathbf{I}_i) \text{is same } \forall f_i)]. \\
&\propto \sum_{f_{N_i}=0}^{M-1} P(\mathbf{I}_{N_i} \mid \mathbf{I}_i, f_i, f_{N_i}) P(\mathbf{I}_i \mid f_{N_i}, f_i) P(f_{N_i} \mid f_i) P(f_i) \\
&\approx P(\mathbf{I}_i \mid f_i) P(f_i) \sum_{f_{N_i}=0}^{M-1} P(\Delta \mathbf{I}_{N_i} \mid f_i, f_{N_i}) P(f_{N_i} \mid f_i), \qquad (1)
\end{aligned}
$$

where f_{N_i} are the classes of neighbours of v_i, $\Delta\mathbf{I}_{N_i} = I_{N_i} - I_i$. We assume that $P(\Delta\mathbf{I}_{N_i} \mid f_i, f_{N_i}, \mathbf{I}_i) = P(\Delta\mathbf{I}_{N_i} \mid f_i, f_{N_i})$ and $P(\mathbf{I}_i \mid f_i) = P(\mathbf{I}_i \mid f_{N_i}, f_i)$. We need to compute $P(\mathbf{f} \mid \mathbf{I})$ where N is the number of sites and $\mathbf{I}$ is the intensity at all voxels across all modalities. According to Hammersley-Clifford theorem [1], this MRF is equivalent to a Gibbs distribution, whose form is given by $P(\mathbf{f}) = \frac{1}{Z} exp(U(\mathbf{f}))$, where $Z = \sum_{f \in \mathbf{F}} exp(U(\mathbf{f}))$, and $U(\mathbf{f} \mid \mathbf{I})$ is given by

$$U(\mathbf{f} \mid \mathbf{I}) = \sum_{i=0}^{N-1} \Bigg[ln(P(f_i)) + (\mathbf{I}_i - \mu_{f_i})^T \Sigma_{f_i}^{-1} (\mathbf{I}_i - \mu_{f_i}) + \sum_{N_i} (\Delta\mathbf{I}_{N_i} - \mu_{f_{N_i}, f_i})^T \Sigma_{f_{N_i}, f_i}^{-1} (\Delta\mathbf{I}_{N_i} - \mu_{f_{N_i}, f_i}) + \alpha m(f_{N_i}, f_i) \Bigg], \quad (2)$$

where $m(f_{N_i}, f_i) = 1$ if $f_{N_i} = f_i$, and 0 otherwise, and α is the weighting coefficient vector. In this case, we model both the $P(\Delta\mathbf{I}_{N_i} \mid f_i, f_{N_i})$ and $P(\mathbf{I}_i \mid f_i)$ using the training data. The neighbourhood and cliques are chosen appropriately to compute the energy $m(f_{N_i}, f_i)$ and the prior probability $P(f_i)$ is obtained from a statistical prior atlas. To maximise $P(\mathbf{f} \mid \mathbf{I})$, we minimise $U(\mathbf{f} \mid \mathbf{I})$, where $\mathbf{f}_{min} = \text{argmin}_{f \in \mathscr{F}} U(\mathbf{f} \mid \mathbf{I})$.

2.1 Implementation

Our database comprises of 25 volumes containing both relapsing-remitting (RR) and secondary progressive (SP) MS patients. A leave-5-out strategy was employed for the training and testing, such that 20 volumes were used for training and 5 for testing each time. The volumes had a lesion load varying from $5cc$ to $19cc$, and tested on 5 volumes. All volumes were labelled by 5 experts and their consensus (silver standard) was taken in determining a lesion.

We divide the brain into the frontal, temporal, occipital, parietal, central and posterior fossa using the method suggested in [3] and build the distributions for the various tissue classes in the different regions separately, considering each area a different MRF. In our implementation, we chose $M = 6$, with the 6 classes being Background (Bk), White Matter (WM), Grey Matter (GM), Cerebral Spinal Fluid (CSF), $T1$-hypointense lesions ($T1_{les}$) and $T2$-hyperintense lesions ($T2_{les}$). The division of lesions into $T1$ and $T2$ lesions is anatomically justified in [9]. The distribution of lesions, on the whole, is bimodal. However, dividing the lesions into $T1$ and $T2$ lesions allows us to choose normal distribution for the two classes [7]. In the training phase, given the expert labelled set of 20 images, $\mathbf{I}^{(k)}$, $k \in 0, \ldots, 19$, the intensities of the voxels belonging to the six classes for all modalities were fit to multivariate Gaussians, with corresponding means μ and covariances Σ. Further, class intensity differences $P(\Delta\mathbf{I}_{N_i} \mid f_i, f_{N_i}) \, \forall f_{N_i}, f_i$ in the neighbourhood were fit multivariate Gaussians.

In our MRF, we choose the 8 neighbourhood of v_i in plane and the two corresponding voxels in the slices above and below. We evaluate all the cliques in the neighbourhood (which happen to be $1, 2, 3,$ and 4 voxels in size). Typically, MRF segmentation approaches evaluate only single and pairwise cliques [8] or selective cliques [7], which do not fully comprise the relationships in the neighbourhood. The larger cliques have an important role in eliminating errors due to local inhomogeneities and eliminating false positives. This comes at the price of slightly greater computational time.

We initialise our MRF using the result of the Bayesian classifier evaluating $P(f_i \mid \mathbf{I}_i) = P(\mathbf{I}_i \mid f_i) P(f_i)$ in the various regions of the brain. Indeed, this is nothing more than picking the minimum energy component of the voxel energy, i.e., $min_{f_i \in 0, \ldots, M-1} ln(P(f_i)) + (\mathbf{I}_i - \mu_{f_i})^T \Sigma_{f_i}^{-1} (\mathbf{I}_i - \mu_{f_i})$, in eq. 2. However, any reasonable classifier that gives us a good estimate

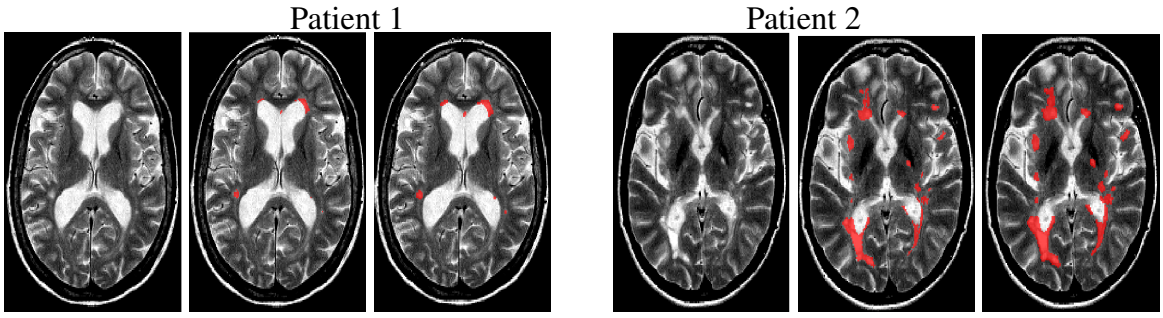

Figure 1: (a) (b) (c) (a) (b) (c)
We have the results of the algorithm on two patient cases. We have (a) Central Slice of T2-weighted MRI of MS patients (b) the lesions labelled by experts and (c) lesions labelled by our algorithm. Patient 1 has a low lesion load, and algorithm localises lesions accurately with $\kappa = 0.58$. Patient 2, has a heavy lesion load along with juxta-cortical lesions, but the algorithm detects both with a $\kappa = 0.8$.

Patient	1	2	3	4	5	Mean
κ (van Leemput [11])	0.47 ± 0.08	0.53 ± 0.13	0.38 ± 0.16	0.5 ± 0.06	0.48 ± 0.06	0.47 ± 0.11
κ (Harmouche [7])	0.55 ± 0.08	0.64 ± 0.07	0.49 ± 0.12	0.53 ± 0.06	0.61 ± 0.08	0.57 ± 0.1
κ (our technique)	0.62 ± 0.09	0.75 ± 0.06	0.59 ± 0.14	0.56 ± 0.05	0.69 ± 0.09	0.64 ± 0.11

Table 1: The means of the 5 κ values of each fold in 5 fold cross validation. The κ values are computed by comparing our results to the experts' consensus (silver standard).

of the solution can be used to initialise the MRF. Once initialised, ICM [2] (which finds local extrema) can be employed to find the global solution as well.

3 Results

The qualitative results of our algorithm are seen in Fig. 1, where it can be seen to be working well on two different patients: one with small lesions and the other with large lesions. Both the small and large lesions, as well as peri-ventricular and juxta cortical lesions are correctly detected, and our results are in agreement with the experts' labels (Fig. 1).

The most commonly employed quantitative measure is the Dice Similarity Coefficient, $\kappa = \frac{2(A \bigcap B)}{A+B}$, where A and B are the set of voxels denoted as lesion by the algorithm and the expert respectively. The quantitative results of the algorithm are shown in Table 1, where the mean of each of the 5 sets of 5 test volumes apiece has been tabulated. The performance of our algorithm has been compared against the performance of two other state of the art algorithms. Our algorithm's performance is superior to that of existing techniques in κ terms (Tab. 1). The maximum and minimum κs in the test set were 0.81 and 0.35.

The average number of false positive and false negative lesions for each of the 5 folds are shown in Fig. 2. A lesion is defined as a set of at least 3 contiguous voxels in its cubic neighbourhood, marked as the lesion class [10]. As can be seen in Fig. 2, there is a large improvement in terms of false negatives compared to [7] and [11], while the number of false positives is comparable in all three methods. Most of our false positives occur at the bottom of the temporal lobe, where there are artefacts in volume acquisition, and even experts disagree about lesions. The elimination of false negatives is the main benefit of our technique.

4 Conclusion

In this paper, we have demonstrated the superiority of our proposed MRF tissue classification technique compared to two other state of the art techniques, in the classification of MS

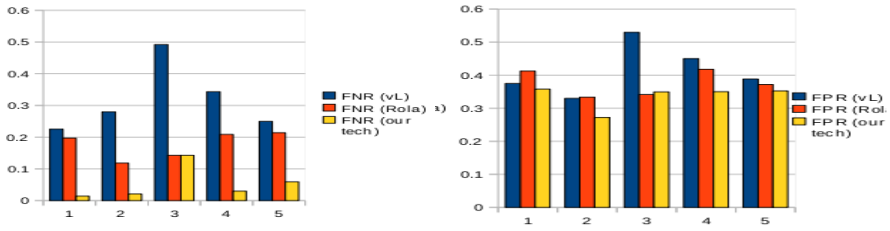

Figure 2: Comparison of the false negative rates (left) and false positive rates (right) of the three algorithms compared, for each fold in the 5 fold cross validation tabulated in Table 1. In 3 of the 5 sets, there are almost no false negatives in our case, while the number of false positives is comparable.

lesions in real clinical data. Our approach outperforms the others in terms of Dice similarity coefficients and the number of false negatives, which are almost completely eliminated. Future work will be focused on reducing the number of false positives.

References

[1] J. Besag. Spatial interaction and statistical analysis of lattice systems. *J. Royal Stat. Soc., Ser. B*, 36(2):192–236, 1974.

[2] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley and Sons, 2000.

[3] D. L. Collins et. al. Automatic model based neuro-anatomical segmentation. *Human Brain Mapping*, 3:190–208, 1995.

[4] J. Souplet et. al. An automatic segmentation of T2-FLAIR multiple sclerosis lesions. *Midas Journal, IJ-2008 MICCAI Workshop, MS Lesion Segmentation*, 2008.

[5] L. Ait-Ali et. al. STREM: A robust multidimensional parametric method to segment MS lesions in MRI. *MICCAI 2005*, pages 409–415, 2005.

[6] O. Freifeld et. al. MS lesion detection using constrained GMM and curve evolution. *International Journal of Biomedical Imaging*, 2009.

[7] R. Harmouche et. al. Bayesian MS lesion classification modelling regional and local spatial information. *ICPR 2006*, pages 984–987, 2006.

[8] R. Khayati et. al. Fully automatic segmentation of multiple sclerosis lesions in brain MR FLAIR images using adaptive mixtures method and markov random field model. *Computers in Biology and Medicine*, 38:379–390, 2008.

[9] Y. Wu et. al. Automated segmentation of multiple sclerosis subtypes with multichannel MRI. *Neuroimage*, 32:1205–1215, 2006.

[10] S. J. Francis. Automatic lesion identification in MRI of MS patients. *MSc Thesis, Department of Neurology, McGill University*, 2004.

[11] K. van Leemput et. al. Automated segmentation of sultiple sclerosis lesions by model outlier detection. *IEEE Trans. Medical Imaging*, 20(8):677–688, 2001.