

Multiple Sclerosis lesion segmentation using an automatic multimodal Graph Cuts

D. García-Lorenzo^{1,2,3*}, J. Lecoœur^{1,2,3}, D.L. Arnold⁴,
D.L. Collins⁴, and C. Barillot^{1,2,3}

¹ INRIA, VisAGeS Unit/Project, IRISA, Rennes, France

² University of Rennes I, CNRS IRISA, Rennes, France

³ INSERM, U746 Unit/Project, IRISA, Rennes, France

⁴ Montreal Neurological Institute, McGill University, Montreal, Canada

Abstract. Graph Cuts have been shown as a powerful interactive segmentation technique in several medical domains. We propose to automate the Graph Cuts in order to automatically segment Multiple Sclerosis (MS) lesions in MRI. We replace the manual interaction with a robust EM-based approach in order to discriminate between MS lesions and the Normal Appearing Brain Tissues (NABT). Evaluation is performed in synthetic and real images showing good agreement between the automatic segmentation and the target segmentation. We compare our algorithm with the state of the art techniques and with several manual segmentations. An advantage of our algorithm over previously published ones is the possibility to semi-automatically improve the segmentation due to the Graph Cuts interactive feature.

1 Introduction

Multiple Sclerosis (MS) is a chronic demyelinating disease that affects the central nervous system. Magnetic Resonance Imaging (MRI) has been proven as a very useful technique to study MS.

Manual segmentation of MS lesions is often performed in clinical trials. However, it is a lengthy process which shows high intra- and inter-rater variability [1]. Automatic segmentation methods have been developed to deal with these limitations [2–4], but in clinical studies these segmentations are often revised by an expert and manually edited when necessary.

Graph Cuts [5] is a recently developed technique for interactive segmentation which has been successfully employed in different medical domains such as organ segmentation [6], healthy brain MRI [7], and pathological brain MRI [8]. It is based on both regional and contour information and always reaches the global minimum of its cost function [5]. The user has to select some seeds in both the “object” and the “background” in order to perform the segmentation. The result can be refined interactively by adding new seeds.

* We are thankful to ARSEP (Association pour la Recherche en Sclérose en Plaques) and UEB (Université Européenne de Bretagne) for funding.

In this paper, we propose to automate the Graph Cuts algorithm to segment MS lesions using several MR sequences. The initialization for Graph Cuts is given by a Finite Gaussian Mixture Model (FGMM) estimated by robust version of the Expectation-Maximization (EM) algorithm [9]. An advantage of our method over other automated methods is the possibility for an expert to easily refine the segmentation as the original semi-automatic Graph Cuts.

2 Method

In the following sections, we group all the MR sequences to a unique multi-dimensional image of dimension m equal to the number of sequences. In our case $m = 3$: T1-w, T2-w and PD-w. We assume that all the MR sequences of the same patient are previously registered in the same space, intensity inhomogeneity correction has been performed and the brain has been extracted.

We explain the general framework of the Graph Cuts, and the choices required for boundary information (the Spectral Gradient) and for regional information (an EM-based approach).

2.1 Graph Cuts

The segmentation problem can be described by a flow graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ which represents the image [5, 6]. In the graph \mathcal{G} , each voxel of the image corresponds to a node. The node set \mathcal{V} also contains two particular nodes called terminal nodes - also known as “source” and “sink” - which respectively represents the classes “object” and “background”. Nodes are connected with undirected edges \mathcal{E} .

We define \mathcal{P} as the set containing all the nodes p of the brain and \mathcal{N} as the set containing all the connection between two nodes $\{p, q\}$. Segmentation is represented by V , where V_p can be either “object” or “background”. The energy $E(V)$ is then minimize by the Graph Cuts:

$$E(V) = \alpha \cdot \sum_{p \in \mathcal{P}} R_p(V_p) + \sum_{\substack{\{p,q\} \in \mathcal{N} \\ V_p \neq V_q}} B_{\{p,q\}} \quad (1)$$

The term $R_p(\cdot)$, referred to as the regional term, expresses how the voxel p fits into given models of the object and background. In the graph, this relation is expressed by the connection of all the nodes to the source and sink nodes, called t-links, with weights, W_{SO}^p and W_{SI}^p respectively.

The boundary term $B_{\{p,q\}}$ reflects the similarity of the voxels p and q . Neighboring nodes are connected in the graph, n-links, with weight $B_{\{p,q\}}$. Their values are close to zero when the existence of a contour between p and q is very likely and large otherwise. The coefficient α is used to adjust the importance of the region and boundary terms.

This graph representation enables us to employ the Boykov-Kolmogorov graph minimization method [5] which is able to efficiently find the global minimum of the energy function (1).

2.2 The Spectral Gradient

The boundary weights $B_{\{p,q\}}$ are usually computed by an edge detection technique such as local intensity gradient. In our case, we choose the Spectral Gradient previously employed for multi-sequence MRI brain segmentation [8].

The objective is to consider our three MR sequences as an RGB color image and use an invariant color-edge detector [8]. This detector is based on a physical property of color, the spectral intensity e , and its derivatives with respect to the light wavelength λ , e_λ and $e_{\lambda\lambda}$, that are simple to compute [10].

For the graph, the detector is discretized, giving the following boundary term [8]:

$$B_{\{p,q\}} = \exp\left(-\frac{(\varepsilon(p) - \varepsilon(q))^2 + (\varepsilon_\lambda(p) - \varepsilon_\lambda(q))^2}{2\sigma^2}\right) \cdot \frac{1}{dist(p,q)} \quad (2)$$

where σ is a smoothing parameter (in our experiments $\sigma = 1$), and

$$\varepsilon = \frac{1}{e} \cdot \frac{\partial e}{\partial \lambda} = \frac{e_\lambda}{e} \quad \text{and} \quad \varepsilon_\lambda = \frac{\partial \varepsilon}{\partial \lambda} = \frac{e \cdot e_{\lambda\lambda} - e_\lambda^2}{e^2} \quad (3)$$

2.3 Automatic Calculation of the Sink and the Source

In semi-automatic frameworks, the weights $R_p(\mathcal{B})$ and $R_p(\mathcal{O})$ for the voxel p are usually defined this way:

$$R_p(\mathcal{B}) = \begin{cases} \infty & \text{if } p \in \mathcal{B} \\ 0 & \text{if } p \in \mathcal{O} \\ \alpha \cdot -\ln W_p(\mathcal{O}) & \text{elsewhere} \end{cases} \quad R_p(\mathcal{O}) = \begin{cases} 0 & \text{if } p \in \mathcal{B} \\ \infty & \text{if } p \in \mathcal{O} \\ \alpha \cdot -\ln W_p(\mathcal{B}) & \text{elsewhere} \end{cases} \quad (4)$$

The point sets \mathcal{B} and \mathcal{O} are the seeds of the background and the object respectively. The probability $P(I_p|\mathcal{B})$ reflects how the intensity vector I_p of voxel p fits into the intensity model estimated from \mathcal{B} . Most authors assume seeds follow a Gaussian distribution[7, 8, 5].

Our objective is to eliminate the dependence on the selection of seeds to create a fully automated method. Therefore we need to remove all the dependencies to the point sets \mathcal{B} and \mathcal{O} . We propose to replace $P(I_p|\mathcal{B})$ and $P(I_p|\mathcal{O})$ in (4) with weights W_{SI} and W_{SO} described below. We keep the rest of (4) the same to allow posterior interaction of an expert to refine the segmentation if not satisfied with the automatic one.

Although the nature of the MRI noise is usually Rician [11], brain image intensities in MRI have been successfully modeled as a 3-class Multivariate FGMM in healthy subjects [12, 13] and in MS patients [2, 4]. These three classes correspond to Cerebrospinal Fluid (CSF), Grey Matter (GM) and White Matter (WM). On the contrary, MS lesions are usually considered not as a new class but as outliers from this 3-class NABT model [2-4].

First, we estimate the NABT model using a robust EM [9] which computes the trimmed likelihood in order to avoid outliers and has been successfully used in MS patients [4, 14]. This algorithm has a parameter h , the rejection ratio, which adjusts the number of voxels that are not taken into account in the M-step in order to be robust to outliers. In our experiments h is fixed to 10% of the voxels, this value is large enough to avoid errors in estimation due to the MS lesions, veins, or registration and brain extraction errors.

Once the NABT model is estimated, we compute the Mahalanobis distance of each voxel p to each class i .

$$d_{Mahalanobis}^i(p) = \sqrt{(p - \mu_i)^T \Sigma_i^{-1} (p - \mu_i)}. \quad (5)$$

where μ_i is the mean vector and Σ_i is the covariance matrix of the class i .

The Mahalanobis distance follows a χ_m^2 distribution with m degrees of freedom when the data follow an m -dimension Gaussian law. This characteristic allows to obtain the p -value for each voxel, the probability the voxel does not fit into the model. We only consider the class with the lowest p -value for each voxel. Voxels with high p -value are more likely to be outliers than the others.

Sink: NABT voxels Voxels that follow the model should have a weight with a high value. Therefore we assign

$$W_p(\mathcal{B}) = 1.0 - p\text{-value} \quad (6)$$

Source: MS lesions Outliers, voxels with high p -value, are normally MS lesions but they can also occur due to veins or registration and skull-stripping errors. To select MS lesions from other outliers we apply a priori knowledge about the intensity of the lesions.

MS lesions are usually described as “hyperintense” compared to the WM in T2-w and PD-w images, we choose a fuzzy logic approach to model this experts’ knowledge. Figure 2 describes the fuzzy function associated to “hyperintense” using the previously computed NABT model information. For each sequence, T2-w and PD-w, we obtain the fuzzy weights, W_{PD-w} and W_{T2-w} . We merge this information with the p -value using the fuzzy AND operator:

$$W_p(\mathcal{O}) = \text{AND}\{W_{PD-w}, W_{T2-w}, p\text{-value}\} \quad (7)$$

One post-processing step is performed after Graph Cuts. As many false positives occur due to artifacts in the external CSF, all lesions detected neighboring the brain border are removed from the segmentation.

3 Evaluation

We test our algorithm (GC) in two different situations. At first, we use synthetic images to evaluate the impact of different levels of noise and inhomogeneity. Then MR images of ten MS patients are segmented and compared with the manual segmentation delineated by several experts.

The evaluation measure is the Dice Similarity Coefficient (DSC) [15], widely employed for image segmentation evaluation. DSC values range from 0.0 to 1.0 where a value higher than 0.7 is usually considered good agreement.

3.1 Data

Synthetic Images Brainweb MRI simulator [16] allows the creation of synthetic MR volumes (T1-w, T2-w and PD-w) of an MS patient controlling some acquisition parameters as slice thickness, noise and field inhomogeneity. Three different phantoms are available with different lesion loads: mild, moderate and severe.

We perform the automatic segmentation varying the noise level (1%, 3%, 5%, 7% and 9%) and the intensity inhomogeneity (0%, 20% and 40%) for the three available phantoms. The ground truth is available, the evaluation is done comparing the results of our algorithm with this ground truth using the DSC. In these synthetic images, no errors due to registration or brain extraction exist, which means that the total number of outliers will be significantly reduced. For this reason we set h , the parameter of the robust EM, to 5% instead of 10%.

Real Images Real images in our experiments consist of MR volumes (T1-w, T2-w and PD-w) of 10 patients with MS. All the images undergo the same preprocessing workflow: intensity inhomogeneity correction [17], intra-subject multimodal registration [18] and brain extraction [19].

Each patient image is manually segmented by 5 experts. A silver standard is built considering the consensus of 3 out of 5 experts to define a lesion voxel. Our algorithm and the EMS software [2] are compared with the silver standard using DSC. To have an idea of the common agreement among experts we compute the average inter-rater DSC (IRDSC). For each pair of experts we compute the DSC and for each patient, all the DSC values are averaged to the IRDSC. The common agreement is good if IRDSC is above 0.7.

STAPLE algorithm [20] was designed to study the performance of different raters when the ground truth is not available. The idea is to compute the sensibility and specificity of each rater while estimating the most probable true segmentation. We use the ITK¹ implementation to compute STAPLE among the 5 raters and our method. We did not include EMS algorithm because it is also an EM-based approach and it could bias the STAPLE results.

3.2 Results

Synthetic Images Fig. 1 sums up the results for the three phantoms and variations in noise and inhomogeneity. The performance of the algorithm decreases significantly for high levels of noise (7% and 9%) and for low levels (1%), showing DSC values over 0.7 for 3% and 5%. On the contrary, the intensity inhomogeneity

¹ <http://www.itk.org>

slightly affects the performance of the segmentation. Mild phantom (left figure) shows lower DSC scores, this is reasonable because errors affect more the DSC when the target region is smaller.

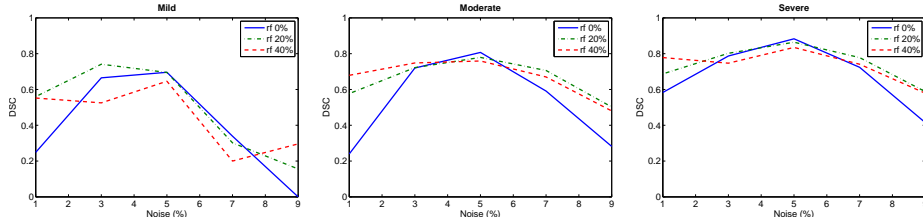


Fig. 1. From left to right: DSC values from mild, moderate and severe phantoms for different values of noise and inhomogeneity (rf).

	TLL(ml)	IRDSC	EMS	GC
Patient 1	33.9	0.77	0.67	0.75
Patient 2	2.8	0.70	0.61	0.75
Patient 3	1.4	0.60	0.42	0.38
Patient 4	1.0	0.61	0.41	0.46
Patient 5	20.0	0.81	0.80	0.86
Patient 6	2.7	0.51	0.48	0.40
Patient 7	6.0	0.69	0.72	0.71
Patient 8	6.2	0.61	0.69	0.72
Patient 9	1.2	0.64	0.35	0.50
Patient 10	47.7	0.81	0.71	0.73
Average		0.68	0.59	0.63

Table 1. Total lesion load (TLL), the inter-rater DSC (IRDSC), the DSC for EMS software and the DSC for our method (GC).

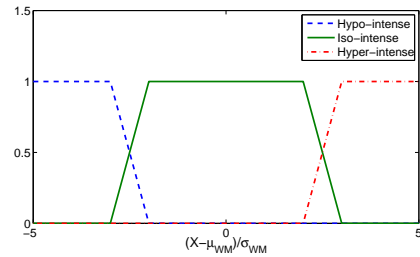


Fig. 2. Fuzzy membership functions for: hypo-intensity, hyper-intensity and iso-intensity. For each voxel we can assign its fuzzy value using these functions.

Real Images Results are shown in Table 1 and Figure 4. Globally we conclude that our algorithm performs a better segmentation than EMS. Our algorithm obtains better scores for 7 out of 10 patients compared to EMS. Patients (3, 4, 6 and 9) with low lesion loads ($< 2.7ml$) obtain poor scores for both algorithms as well as low experts agreement ($IRDSC < 0.65$). This can be partially explain because DSC decreases more a for small TLL. For the rest of the patients our DSC is always higher than 0.7 showing a good agreement with the target segmentation.

As described before, STAPLE gives the specificity and sensibility for each expert as presented in Figure 3. We can observe that the sensibility of our Graph Cuts algorithm is higher than three experts but with a slightly lower specificity.

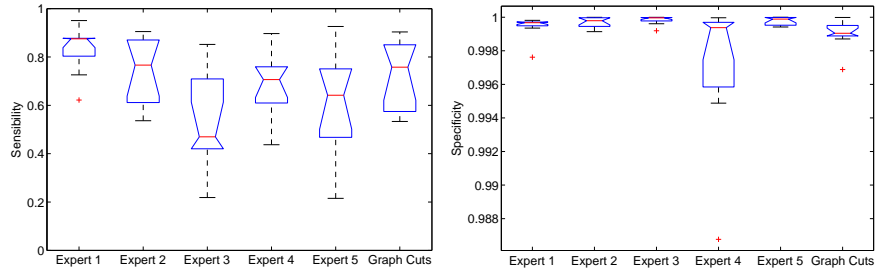


Fig. 3. Sensibility (left) and Specificity (right) boxplot for the different experts and the automatic segmentation. Specificity values are very high because the size of the lesions is very small compared to the size of the brain. However, small variations of specificity greatly modifies the DSC value.

4 Discussion

In this paper, we have proposed an automatic algorithm for the MS lesion segmentation for multi-sequence MRI. Experiments with synthetic images have shown good agreement with the ground truth for all levels of inhomogeneity and standard levels of noise. For very noisy images, a noise reduction preprocessing step might be necessary. The evaluation in real images demonstrates a good agreement of our segmentation with a group of experts while shows a better segmentation than state of the art algorithm EMS.

Fully automated methods are often revised by an expert in clinical trial to verify their validity and edited when necessary. In Figure 5, we can observe an example of semi-automatic edition of our automatic segmentation. When a lesion is missed, a user can add a seed, in this case a source seed, and the Graph Cuts is recomputed in few seconds.

References

1. Grimaud, J., Lai, M., Thorpe, J., Adeleine, P., Wang, L., Barker, G.J., Plummer, D.L., Tofts, P.S., McDonald, W.I., Miller, D.H.: Quantification of MRI lesion load in multiple sclerosis: a comparison of three computer-assisted techniques. *Magn Reson Imaging* **14**(5) (1996) 495–505
2. Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P.: Automated segmentation of multiple sclerosis lesions by model outlier detection. *Medical Imaging, IEEE Transactions on* **20**(8) (Aug. 2001) 677–688

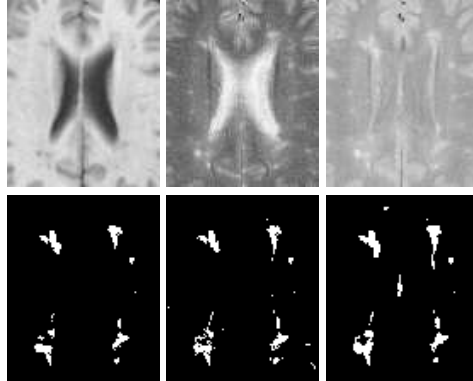


Fig. 4. Top, from left to right: T1-w, T2-w and PD-w images of patient 8. Bottom, from left to right: Consensus, EMS and Graph Cuts segmentations.

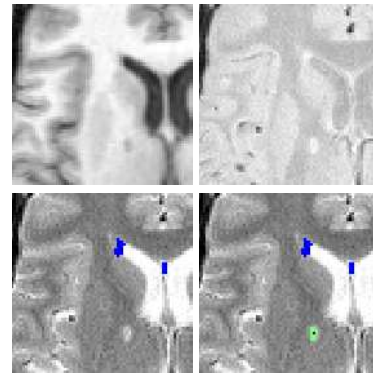


Fig. 5. Top, left to right: T1-w, and PD-w images of patient 4. Bottom, left to right: automatic solution and semi-automatic solution (Red: source seed, Green: Graph Cut solution, Blue: automatic segmentation).

3. Dugas-Phocion, G., Gonzalez, M., Lebrun, C., Chanalet, S., Bensa, C., Malandain, G., Ayache, N.: Hierarchical segmentation of multiple sclerosis lesions in multi-sequence MRI. In: Biomedical Imaging: Macro to Nano, 2004. IEEE International Symposium on. Volume 1. (April 2004) 157–160
4. Aït-Ali, L.S., Prima, S., Hellier, P., Carsin, B., Edan, G., Barillot, C.: STREM: a robust multidimensional parametric method to segment MS lesions in MRI. *Med Image Comput Assist Interv Int Conf Med Image Comput Assist Interv* **8**(Pt 1) (2005) 409–416
5. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient N-D images segmentation. *International Journal of Computer Vision* **70**(2) (November 2006) 109–131
6. Boykov, Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In: *International Conference on Computer Vision*. Volume 1. (July 2001) 105–112
7. Song, Z., Tustison, N., avants, B., Gee, J.: adaptive graph cuts with tissue priors for brain MRI segmentation. In: *International Symposium Biomedical Imaging*. (2006) 762–765
8. Lecoqeur, J., Morissey, S., Ferré, J.C., Arnold, D., Collins, D., Barillot, C.: Multiple sclerosis lesions segmentation using spectral gradient and graph cuts. In: *Proceedings of MICCAI workshop on Medical Image Analysis on Multiple Sclerosis (validation and methodological issues)*. (September 2008) 92–103
9. Neykov, N., Filzmoser, P., Dimova, R., Neytchev, P.: Robust fitting of mixtures using the trimmed likelihood estimator. *Computational Statistics & Data Analysis* **52**(1) (September 2007) 299–308
10. Geusebroek, J.M., van den Boomgaard, R., Smeulders, A., Dev, A.: Color and space : The spatial structure of color images. In: *IEEE European Conference*

- on Computer Vision. Volume 1842 of Lecture Notes in Computer Science. (2000) 331–341
11. Dietrich, O., Raya, J.G., Reeder, S.B., Ingrisch, M., Reiser, M.F., Schoenberg, S.O.: Influence of multichannel combination, parallel imaging and other reconstruction techniques on mri noise characteristics. *Magnetic Resonance Imaging* **26**(6) (2008) 754 – 762
 12. Schroeter, P., Vesin, J.M., Langenberger, T., Meuli, R.: Robust parameter estimation of intensity distributions for brain magnetic resonance images. *IEEE Trans Med Imaging* **17**(2) (Apr 1998) 172–186
 13. Wells, W.M., I., Grimson, W., Kikinis, R., Jolesz, F.: Adaptive segmentation of MRI data. *Medical Imaging, IEEE Transactions on* **15**(4) (Aug. 1996) 429–442
 14. García-Lorenzo, D., Prima, S., Collins, D.L., Arnold, D.L., Morrissey, S.P., Barillot, C.: Combining Robust Expectation Maximization and Mean Shift algorithms for Multiple Sclerosis Brain Segmentation. In: *MICCAI Workshop in Medical Image Analysis for Multiple Sclerosis (MIAMS)*, New York, USA (September 2008) 82–91
 15. Zijdenbos, A., Dawant, B., Margolin, R., Palmer, A.: Morphometric analysis of white matter lesions in mr images: method and validation. *Medical Imaging, IEEE Transactions on* **13**(4) (Dec 1994) 716–724
 16. Collins, D., Zijdenbos, A., Kollokian, V., Sled, J., Kabani, N., Holmes, C., Evans, A.: Design and construction of a realistic digital brain phantom. *Medical Imaging, IEEE Transactions on* **17**(3) (Jun 1998) 463–468
 17. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging* **17**(1) (Feb 1998) 87–97
 18. Collins, D., Neelin, P., Peters, T.M., Evans, A.C.: Automatic 3D Intersubject Registration of MR Volumetric Data in Standardized Talairach Space. *Journal of Computer Assisted Tomography* **18** (March 1994) 192–205
 19. Smith, S.M.: Fast robust automated brain extraction. *Hum Brain Mapp* **17**(3) (November 2002) 143–155
 20. Warfield, S., Zou, K., Wells, W.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *Medical Imaging, IEEE Transactions on* **23**(7) (July 2004) 903–921