Optimized supervised segmentation of MS lesions from multispectral MRIs

Jérémy Lecoeur, Jean-Christophe Ferré, Christian Barillot

Optimized supervised segmentation of MS lesions from multispectral MRIs

Jérémy Lecoeur[1,2,3], Jean-Christophe Ferré[1,2,3,4] and Christian 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] Department of Neuroradiology, Pontchaillou University Hospital, Rennes, France

Abstract. We present an optimized supervised segmentation method from multispectral MRIs. As MR images do not behave as natural images, using a spectral gradient based on a psycho-visual paradigm is sub-optimal. Therefore, we propose to create an optimized spectral gradient using multi-modalities MRIs. To that purpose, the algorithm learns the optimized parameters of the spectral gradient based on ground truth which are either phantoms or manual delineations of an expert. Using Dice Similarity Coefficient as a cost function for an optimization algorithm, we were able to compute an optimized gradient and to utilize it in order to segment MRIs with the same kind of modalities. Results show that the optimized gradient matrices perform significantly better segmentations and that the supervized learning of an optimized matrix is a good way to enhance the segmentation method.

1 Introduction

It is now clinical routine to acquire different MR sequences at the same time for Multiple Sclerosis (MS). Therefore, in order to access the relevant and complementary information in the data, these multimodal sequences should be taken into account in order to achieve a better segmentation.

MS lesion segmentation methods have been developed [1–6] in order to replace the use of manual delineations that have proven to have high inter- and intra-variability [7]. However, the main inconvenience of those methods is that they do not allow an iterative refinement of the results.

The semi-automatic Graph Cuts method [8] is able to efficiently and successfully give an optimal solution for the joint use of regional and border information in a similar way that Markov Random Fields work.

In another field of application, Ali's work [9] uses multimodal data in a graph cut but it is restricted to two dimensional images. Thus, it can not insure the spatial consistency of the solution for 3D images. In our previous method [10, 11], we proposed to use the Graph Cuts method combined with a spectral gradient technique [12] in order to segment 3 MRI sequences like it was a color image. The spectral gradient is based on a psycho-visual paradigm and, as such, can be sub-optimal for MRIs.

In this paper, we propose a supervised learning technique of an optimal spectral gradient. This optimized gradient will then be used in a graph cut scheme to perform MS lesions segmentation. An evaluation on clinical and phantom data is provided.

2 Methods

Using a spectral gradient, based on a psycho-visual context, can be sub-optimal for the kind of image used. We proposed to learn a new spectral space that optimizes the separability of the sequences used and to utilized this new space on other MRIs.

In the following sections, we explain the original Spectral Gradient, the graph cut paradigm used to segment the volumes, and the search for the new optimized multispectral gradient.

2.1 Spectral Gradient: short and simple

In our previous approach[10, 11], we proposed to use the spectral gradient, first introduced by Geusebroek et al. [12], as border information. We built a RGB-color-like image by merging three different MR modalities into a single volume.

It has been proven [12] that the spectral gradient operator can be computed from the reflected spectrum of a surface (e) and its first and second order derivative with respect to the wavelength (e_λ, $e_{\lambda\lambda}$). Those three terms, seen as a column vector E, are very well approximated by simply multiplying the RGB values (seen as a column vector $\mathcal{V}$) by a projection matrix $\mathcal{M}$:

$$E = \mathcal{M} \cdot \mathcal{V} \tag{1}$$

Finally, using this multispectral space, the detection of all color edges can be performed with :

$$\aleph = \sqrt{(\partial_x \varepsilon)^2 + (\partial_y \varepsilon)^2 + (\partial_z \varepsilon)^2 + (\partial_x \varepsilon_\lambda)^2 + (\partial_y \varepsilon_\lambda)^2 + (\partial_z \varepsilon_\lambda)^2} \tag{2}$$

and :

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

2.2 Spectral Gradient based Graph Cuts

According to the scheme by Boykov et al. [13, 8], the segmentation problem is described by a directional flow graph which represents the image. The node set is defined by two particular nodes called terminal nodes - also known as "source" and "sink" - which respectively represents the classes "object" and "background", the other nodes being the 3D volume voxels, and directed weighted edges connecting the nodes.

Let the set $\mathcal{P}$ contain all the voxels p of the image, the set $\mathcal{N}$ be all the pairs $\{p,q\}$ of the neighboring elements of $\mathcal{P}$ and $V = (V_1, V_2, ..., V_{|\mathcal{P}|})$ be a binary vector where each V_p can be one of the two labels "object" or "background". Therefore, the vector V defines a segmentation. The point sets $\mathcal{B}$ and $\mathcal{O}$ are the seeds of the background and object, respectively. The energy we want to minimize by the graph cut has the form given by:

$$E(\mathcal{V}) = \alpha \cdot \sum_{p \in \mathcal{P}} R_p(V_p) + \sum_{\substack{\{p,q\} \in \mathcal{N} \\ V_p \neq V_q}} B_{\{p,q\}} \tag{4}$$

The term $Rp(\cdot)$, commonly referred as the regional term, expresses how the voxel p fits into given models of the object and background. It is encoded in the graph through the *t-links*, the edges connected the nodes to the source and sink nodes.

Two cases are to be considered: the weight W_{so} of the t-link involving the "source" node and the weight W_{si} of the t-link involving the "sink" node which can be computed as follows:

$$W_{so} = \begin{cases} 0 & \text{if } p \in \mathcal{B} \\ \infty & \text{if } p \in \mathcal{O} \\ -\ln P(\Psi_p|\mathcal{B}) & \text{elsewhere} \end{cases} \qquad W_{si} = \begin{cases} \infty & \text{if } p \in \mathcal{B} \\ 0 & \text{if } p \in \mathcal{O} \\ -\ln P(\Psi_p|\mathcal{O}) & \text{elsewhere} \end{cases} \tag{5}$$

where Ψ_p is a 3-component vector containing the intensities of the voxel p in each modality, $P(\Psi_p|\mathcal{O})$ is the probability for the voxel p to be a member of the object and $P(\Psi_p|\mathcal{B})$ is the probability for the voxel p to be a member of the background. These probabilities being computed from the sets of seeds with Gaussian Mixture Model assumption.

The term $B_{\{p,q\}}$, known as the boundary term, reflects the similarity of the voxels p and q. Hence, the *n-links*, connecting the neighboring voxels, are large when p and q are similar and close to zero otherwise.

To compute those n-links, we use an *ad-hoc* function with a discretization of the spectral gradient:

$$B_{\{p,q\}} \propto \exp\left(-\frac{(\varepsilon(p) - \varepsilon(q))^2 + (\varepsilon_\lambda(p) - \varepsilon_\lambda(q))^2}{2\sigma^2}\right) \cdot \frac{1}{dist(p,q)} \tag{6}$$

where ε and ε_λ are the quantities defined in equation 3.

2.3 Multispectral Space Optimization

The spectral gradient, defined by Geusebroek in [12], is based on a psycho-visual paradigm, which may be problematic since MR images do not behave as natural images. Therefore, the color spectral gradient cannot be assumed to be optimal for the spectral gradient decomposition of MR images. To alleviate this

limitation, we propose to create an optimized multispectral space that differs from (e, e_λ, $e_{\lambda\lambda}$). This enhances the separability of the complementary MR modalities.

The learning step of the algorithm is based on data with ground truth (either phantoms or manual delineation by an expert).

The algorithm learns the optimized parameters of a new squared 3 x 3 matrix $\widetilde{\mathcal{M}}$ to replace the matrix in equation 1. In this new scheme, the input vector is not the RGB values of a "color MRI" but the different modalities.

As a similarity measure for the optimization, we use the Dice Similarity Coefficient (DSC) which is well suited to compare two segmentations. The DSC is computed as follows:

$$DSC = 2 \cdot \frac{Card(Res \cap GT)}{Card(Res) + Card(GT)} \tag{7}$$

Res being the segmentation result and GT the ground truth.

For the given sets of seeds $\mathcal{B}$ and $\mathcal{O}$ and the ground truth (the MS lesions delineated by an expert), we can compute a DSC score for the segmentation obtained from our Graph Cut algorithm. Based on this measure, we want to find the matrix that maximizes this DSC. In other words:

$$\widetilde{\mathcal{M}} = \arg\max_{\mathcal{M}} DSC = \arg\max_{\mathcal{M}} f(\mathcal{M}) \tag{8}$$

For the optimization, we use Powell's NEWUOA [14, 15] because it has been shown as very effective to reach the optimal matrix even with different initializations, either for the parameters of the matrix or the sets of source and sink seeds needed by the Graph Cut algorithm. This optimization algorithm uses a local quadratic approximation of the cost function (here, the cost function is the DSC) and is able to handle the search of nine parameters at a time.

The method is relatively straightforward. The first step is to develop an optimal matrix from a given combination of MRIs modalities. The second step uses this optimal matrix in order to segment MRIs with the same kind of modalities.

In order to evaluate this method, we have run the optimization algorithm on each of the n subjects of the different studies, hence computing n optimal matrices for each study. Then, we used each matrix for the other $n-1$ subjects, ergo we obtained $n \times n$ DSC scores and we compared those n vectors of scores with an ANalysis Of VAriance (ANOVA) in order to evaluate the statistical significance of the obtained results with respect to the classical color spectral gradient.

3 Material

We have run experiments on two sets of modalities. First, on synthetic data with multiple sclerosis lesions obtained from BrainWeb [16] with three different cases presenting respectively a mild, moderate or severe lesion load. We built the input vector $\mathcal{V}$ from simulated T1-w, T2-w and PD-w sequences. The three

images belong to the same subject, are registered in the same space and are consist of 217 slices of 181 x 181 isometric 1 mm voxels with 3% noise (relative to the brightest tissue in the image) and 20% inhomogeneity.

Secondly, we have validated our approach on real data with multiple sclerosis for which we possess a ground truth, consisting of manual segmentations from an expert. We used 14 subjects with T1-w, T2-w and FLAIR sequences, all three registered to the same space [17]. The data were acquired on a 3T Philips ACHIEVA with 3D 1mm isometric T1-w, 2D 3mm axial slice thickness Dual Echo (T2-w & PD) and 2D 3mm slice thickness FLAIR. Intensity inhomogeneity correction has been performed [18] and the noise has been removed [19]. The volumes present various lesion loads.

4 Results

The main idea is to improve the segmentation of pathological data. Therefore, we present data with pathologies only. We have worked on both synthetic and clinical data.

The overall optimization time for a single subject is about 500 seconds on a laptop (Dual Core at 2.16 GHz and 2GB of RAM running Linux), whatever the combination of sequences.

4.1 Synthetic data with multiple sclerosis

Using the proposed scheme, we were able, for each of the three subjects i, to compute a matrix $\widetilde{\mathcal{M}}_i$ that optimizes the DSC. We want to show that the "optimal" matrices are statistically equivalent with respect to the expected optimal result. To that purpose, we used each matrix for the other two subjects and computed the DSC of the obtained segmentation.

Subject	1	2	3
Lesion load	Mild	Moderate	Severe
$\mathcal{M}$	0.7787	0.7602	0.7804
$\widetilde{\mathcal{M}}_1$	**0.8612**	0.8417	0.8301
$\widetilde{\mathcal{M}}_2$	0.8514	**0.8852**	0.8528
$\widetilde{\mathcal{M}}_3$	0.8316	0.8224	**0.8903**

Table 1. DSC scores obtained with $\widetilde{\mathcal{M}}_{i=1}^{3}$ (optimal matrix of subject i) are better than those obtained with $\mathcal{M}$ (original color spectral gradient matrix). Numbers in bold are the optimized DSC for each subject

Table 1 shows that each optimal matrix allows a better segmentation of the multiple sclerosis lesions than the standard spectral gradient. The average computation time for a segmentation is 71.5 seconds per subject.

4.2 Clinical data with multiple sclerosis

Using the same experiment protocol as for the synthetic data, we computed an optimized matrix for each of our 14 subjects. An expert manually delineated the lesions and we used these segmentations as ground truth. Then, we used each matrix for every other subject, hence obtaining 196 DSC scores for the optimized spectral space. Once again, we've compared these values to those obtained when running the graph cut algorithm with the original color SG matrix.

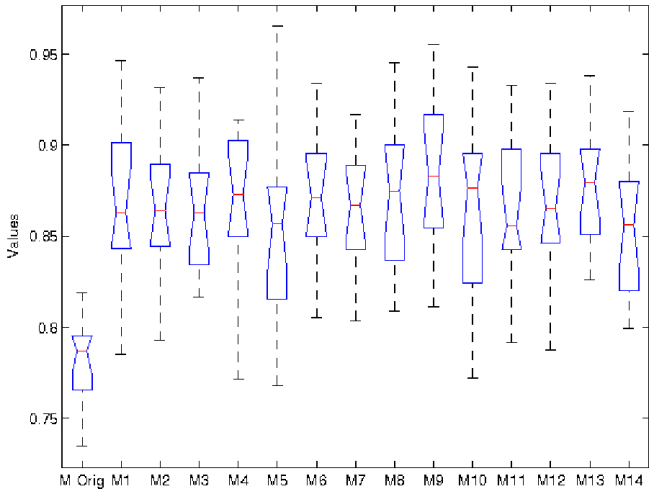

Fig. 1. Comparison of the performance of the different matrices for clinical data

Fig. 1 presents those results and it shows that the optimal matrices usually achieve better segmentation than the original one. The average computation time for a segmentation is 69.5 seconds per subject.

We can see on fig. 2 and 3 that the optimized version of the algorithm corrects most of the misclassified pixels, either false positive or false negative.

4.3 Statistical Significance of the results

In order to quantify the difference between the results obtained in the synthetic data experiment, we ran an ANOVA on the data with the null hypothesis being

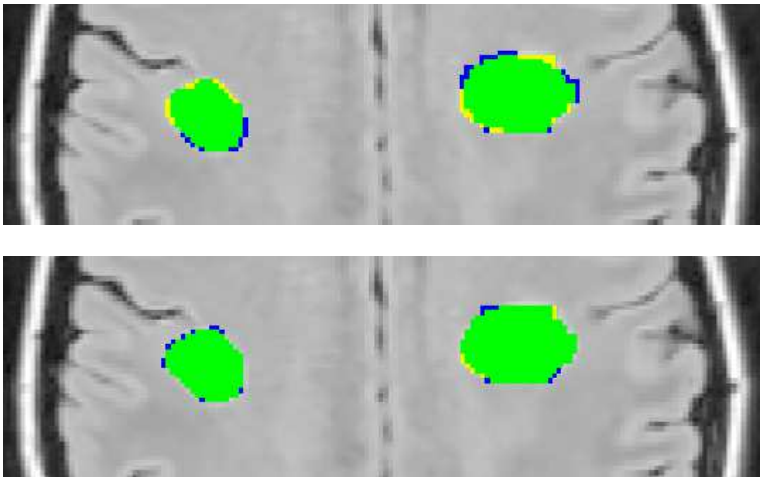

Fig. 2. Top image : segmentation result of clinical data with the classical algorithm. Bottom image : segmentation result of the same data with the optimized algorithm. Green = correctly classified ; Blue = false positive ; Yellow = false negative

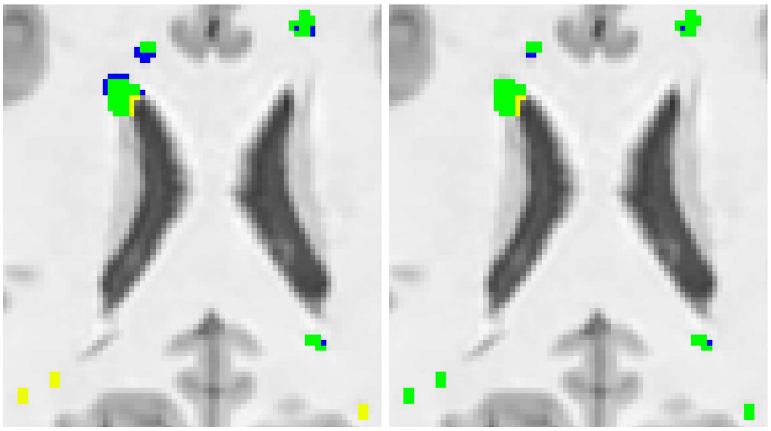

Fig. 3. Left image : segmentation result of clinical data with the classical algorithm. Right image : segmentation result of the same data with the optimized algorithm. Green = correctly classified ; Blue = false positive ; Yellow = false negative

that the results are homogeneous. The *p-value* of the four groups together is $8.3.10^{-3}$ ($F = 8.084$) which means that there is a statistical difference between these groups, whereas the *p-value* of the three groups using optimized matrices is 0.75 ($F = 0.3080$). This shows that the results obtained for the optimized matrices are significantly better compared to those for the original color spec-

tral gradient matrix. This also demonstrates that the optimized matrices form an homogeneous group. Thus, we can find the optimal decomposition from a learning set of data and apply this to new subjects without deviation from the optimal range.

Again in the clinical data experiment, to quantify the difference between the results, we ran an ANOVA on the data. The *p-value* of the fifteen groups of DSC is 2.10^{-7} ($F = 4.724$), showing that there is a statistical difference between the groups, whereas the *p-value* of the fourteen groups of results obtained with optimized matrices is 0.89 ($F = 0.5368$).

Once again, the optimized matrices perform statistically better segmentation than the original matrix. We then used a Leave One Out Cross Validation (LOOCV) utilizing these different matrices and test data sets. The homogeneity of the results of the optimized matrices shows that it is possible to use one volume with ground truth as a training sample and perform accurate and precise segmentation by using its optimized matrix on different subjects with the same sequences.

5 Conclusion and future work

In this paper, we have proposed a supervised technique for optimization of the intermediate spectral space for the computation of a multi-spectral gradient in a graph cut paradigm. Experiments showed that this optimization allows supervised learning of an optimal matrix that can be used to efficiently and successfully segment the structures of interest for subjects with the same kind of sequences that were used to compute the optimal matrix. We also showed that this scheme works on different combinations of sequences.

Once the optimal matrix is computed, the overall computation load for the segmentation is rather low (roughly 70 seconds on a laptop). This allows the use of a semi-automatic segmentation paradigm in order to refined, if necessary, the initial results. Indeed, adding new source and sink seeds would modify the cut in the graph and allows the false negative (or false positive) parts of the segmentation to be retrieved in the new iteration of the segmentation within seconds.

As shown in the paper, the proposed method allows this segmentation scheme to start with various initialization states, since the computed matrices always lead to similar segmentation performances. Moreover, this capability allows the extension of the initial algorithm to multiple sequences with more than 3 modalities. Hence, the matrix $\mathcal{M}$, which was a square 3 x 3 matrix, could be generalized to a n x 3 matrix where the new elements can be initialized almost randomly (e.g. with 0 values). Additional experiments are ongoing to exhibit this capability. It will be also possible to include multiple modalities by using the reduction of modalities algorithm. This algorithm will reduce the multiple modalities to the three images with the most information.

References

1. Aït-Ali, L., Prima, S., Hellier, P., Carsin, B., Edan, G., Barillot, C.: STREM: a robust multidimensional parametric method to segment MS lesions in MRI. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. (October 2005) 409–416
2. Bricq, S.: Segmentation d'images IRM anatomiques par inférence bayésienne multimodale et détection de lésions. PhD thesis, Université Louis Pasteur de Strasbourg (2008)
3. Freifeld, O., Greenspan, H., Goldberger, J.: Lesion detection in noisy MR brain images using constrained GMM and active contours. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro. (2007) 596–599
4. García-Lorenzo, D., Prima, S., L. Collins, D., L. Arnold, D., Barillot, C., P. Morrissey, S.: Multiple Sclerosis (MS) Brain segmentation combining robust Expectation Maximization (EM) and Mean Shift (MeS). In: Annual Meeting of the European Society for Magnetic Resonance in Medicine and Biology, Valencia, Spain (October 2008)
5. Rousseau, F., Blanc, F., de Sèze, J., Rumbac, L., Armspach, J.: An a contratio approach for outliers segmentation : application to multiple sclerosis in MRI. In: IEEE International Symposium on Biomedical Imaging. (2008) 9–12
6. Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P.: Automated segmentation of multiple sclerosis lesions by model outlier detection. IEEE Transactions on Medical Imaging **20**(8) (2001) 677–688
7. Grimaud, J., Lai, M., Thorpe, J., Adeleine, P., Wang, L., Barker, G., Plummer, D., Toft, P., McDonald, W., Miller, D.: Quantification of MRI lesion load in multiple sclerosis: a comparison of three computer-assisted techniques. MR Imaging **14**(5) (1996) 495–505
8. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient N-D images segmentation. International Journal of Computer Vision **70**(2) (November 2006) 109–131
9. Ali, A., Farag, A.: Graph cut based segmentation of multimodal images. In: IEEE International Symposium on Signal Processing and Information Technology. (2007) 1036–1041
10. Lecoeur, J., Morissey, S.P., 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)
11. Lecoeur, J., Ferré, J.C., Collins, D.L., Morrissey, S., Barillot, C.: Multi Channel MRI Segmentation With Graph Cuts Using Spectral Gradient And Multidimensional Gaussian Mixture Model. In: Proceedings of SPIE Medical Imaging 2009: Image Processing. Volume 7259. (February 2009)
12. Geusebroek, J.M., Dev, A., van den Boomgaard, R., Smeulders, A., Cornelissen, F., Geerts, H.: Color invariant edge detection. In: Scale-Space Theories in Computer Vision. Volume 1252 of Lecture Notes in Computer Science. (1999) 459–464
13. 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
14. Powell, M.J.D.: The NEWUOA software for unconstrained optimization without derivatives. In: Workshop On Large Scale Nonlinear Optimization. Volume 83 of Nonconvex Optimization and Its Applications. (2004)

15. Wiest-Daesslé, N., Prima, S., Morrissey, S., Barillot, C.: Validation of a new optimisation algorithm for registration tasks in medical imaging. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro. (April 2007)
16. Collins, D.L., Zijdenbos, A.P., Kollokian, V., Sled, J.G., Kabani, N.J., Holmes, C.J., Evans, A.C.: Design and construction of a realistic digital brain phantom. IEEE Transactions on Medical Imaging **17**(3) (June 1998) 463–468
17. Collins, D.L., 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
18. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in mri data. IEEE Transations on Medical Imaging **17**(1) (February 1998) 87–97
19. Coupé, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C.: An Optimized Blockwise Non Local Means Denoising Filter for 3D Magnetic Resonance Images. IEEE Transactions on Medical Imaging **27**(4) (2008) 425–441