

STREM: A Robust Multidimensional Parametric Method to Segment MS Lesions in MRI

L.S. Ait-Ali¹, S. Prima¹, P. Hellier¹, B. Carsin², G. Edan², and C. Barillot¹

¹ IRISA Campus Universitaire Beaulieu 35042 Rennes Cedex, France

² CHRU Pontchaillou 2 rue Henri Le Guilloux 35000 Rennes, France
laure.ait-ali@irisa.fr

Abstract. We propose to segment Multiple Sclerosis (MS) lesions over-time in multidimensional Magnetic Resonance (MR) sequences. We use a robust algorithm that allows the segmentation of the abnormalities using the whole time series simultaneously and we propose an original rejection scheme for outliers. We validate our method using the BrainWeb simulator. To conclude, promising preliminary results on longitudinal multi-sequences of clinical data are shown.

1 Introduction

Multiple Sclerosis (MS) is an inflammatory autoimmune disorder of the central nervous system. MS is a disease with a high variability in symptoms, clinical course, neuro-radiological appearance of the lesions, response to therapy and pathological mechanisms [1].

Nowadays, MRI is widely used for disease diagnosis, patient follow-up, validation of therapies, and more generally for the understanding of MS. For instance, Gadolinium (Gd) enhanced areas in T1-weighted (T1-w) images and hyper-intense areas in T2-weighted (T2-w) and PD-weighted (PD-w) images are often considered as markers of clinical relapses, whereas brain atrophy (evaluated in T1-w or T2-w images) and hypo-intense areas in T1-w images are often seen as markers of the neuroaxonal loss [2]. These measures complement and enrich the clinical observations based on scales such as the expanded disability status scale (EDSS) or the MS functional composite (MSFC) whose drawbacks are numerous. Clinical observations are inherently subjective, show poor inter- and intra-rater reliability, have limited sensitivity, etc. These drawbacks are particularly critical when dealing with hundreds of subjects as in phase III clinical trials. On the other side, such large-scale studies, composed of multiple subjects, image modalities, time points, and acquisition centers, require automated image processing pipelines for the efficient computation of MR markers.

Many statistical approaches have been developed for the automatic segmentation of brain structures in MR data. A particularly convenient framework consists in modeling the distribution of the intensities as a finite mixture of Gaussians. Such a Gaussian Mixture Model (GMM) allows the modelization of the image intensities with a reduced number of parameters. These parameters can then be estimated within a well-defined statistical framework based on the

Maximum Likelihood Estimator (MLE) using optimization methods such as the Expectation-Maximization (EM) algorithm. Such an approach is unfortunately unadapted to pathological cases characterized by abnormal intensities in MRI. Some authors have proposed modifications of this classical approach to handle outlying intensities caused by MS lesions. Schroeter *et al.* [3] add an uniform component into the GMM to model lesions intensities. Van Leemput *et al.* [4] introduce a weight reflecting the degree of typicality of each brain voxel, the latter being consider as an outlier if its Mahalanobis distance with respect to each mixture component is greater than a threshold. Dugas-Phocion *et al.* use the Mahalanobis distance directly within the EM [5]. All these “robust” EM algorithms are computed for multi-sequences but only for one time point. On the other hand, few methods have been proposed to study the MS evolution over time [6, 7, 8, 9] using statistical tests or deformation fields. Shahar *et al.* [10] propose a combination of these two approaches for the spatio-temporal segmentation of lesions. They perform both lesions segmentation at one time point and lesions evolution over time from 2D T2-w images using a multifeature EM algorithm.

We propose an original multidimensional (multiple sequences and time points) robust algorithm using $3D + t$ MR data to segment MS lesions over time in a standardized clinical protocol. Our method consists in originally modeling the multidimensional problem. The MLE is replaced by a robust estimator, the Trimmed Likelihood Estimator (TLE). We propose an iterative scheme to compute this estimator. In the last step, we refine the segmentation using both the Mahalanobis distance and prior information coming from clinical knowledge on lesion appearance across sequences. The method is described in Section 2, and the experimental results are presented and discussed in Section 3. Finally, we conclude and address future work in Section 4.

2 Method

We first detail a classical algorithm for the segmentation of multi-sequence MR data (see Section 2.1). Then, we present several adaptations for the robust segmentation of MS lesions in multidimensional MR data (see Section 2.2).

2.1 A Classical Parametric Multi-sequence Segmentation Method

Modelization stage. MRI noise is known to follow a Rician distribution, which can be fairly approximated by a Gaussian distribution [11]. Assuming that the distribution of intensities within each brain structure is also well approximated by a Gaussian law, it is common to consider that the image intensities are generated by a Gaussian Mixture Model (GMM). When considering m different sequences simultaneously, each voxel i , $i \in \{1, \dots, n\}$, is described by a m -D intensity vector $\mathbf{y}_i = (y_i^1, \dots, y_i^m)$, which can be modeled by a m -D GMM defined by:

$$p(\mathbf{y}_i; \theta) = \sum_{j=1}^k \alpha_j (2\pi)^{-m/2} |\Sigma_j|^{-1/2} \cdot \exp\left(-\frac{1}{2}(\mathbf{y}_i - \mu_j)^T \Sigma_j^{-1}(\mathbf{y}_i - \mu_j)\right), \quad (1)$$

k is the number of mixture components and $\theta = (\alpha, \mu, \Sigma)$ is the hyper-parameter of the GMM where: (i) $(\alpha_j)_{j \in \{1, \dots, k\}}$ are the mixture proportions, (ii) $(\mu_j)_{j \in \{1, \dots, k\}}$ are the mean vectors, (iii) $(\Sigma_j)_{j \in \{1, \dots, k\}}$ are the covariance matrices.

In MRI, three structures have relatively homogeneous intensities in the intracranial cavity, and constitute the components of the GMM: White Matter (WM), Grey Matter (GM), Cerebro-Spinal Fluid (CSF). The aim is to give a label x_i to each voxel vector \mathbf{y}_i , with $x_i \in \{c_1, \dots, c_k\}$ and c_j being one of the three GMM components. The output $X = (x_1, \dots, x_n)$ is the desired image segmentation of the MR data $Y = (\mathbf{y}_1, \dots, \mathbf{y}_n)$.

Estimation stage. The computation is based on the ML estimation of the GMM hyper-parameter: $\hat{\theta} = \arg \max_{\theta} \prod_{i=1}^n p(\mathbf{y}_i; \theta)$. It can be seen as an *incomplete data problem*, since neither the voxels memberships x_i nor the Gaussian mixture parameters are known but only the voxel intensities y_i .

Optimization stage. The EM algorithm [12] is a well established method to tackle this type of problems. It consists in iterating two steps until convergence: labelization of the image based on a prior knowledge of the unknown parameters (Expectation step) and estimation of the mixture proportions and class parameters based on this labelization (Maximization step).

Practically, this EM algorithm is not directly ran on the m MR images, but rather on a m -D joint histogram built from the m images, which is computationally much more efficient.

Classification stage. The final probabilities $P(x_i = c_j | Y; \hat{\theta})$ are a natural output of the EM algorithm and are used to compute the segmentation X .

2.2 The Proposed Spatio-Temporal Robust EM: STREM

Spatio-temporal model. For the patient follow-up, the longitudinal study is essential. In this context, we propose to perform a multi-sequence segmentation with time points $(1, \dots, t)$ with a robust estimation approach. The whole dataset can then be named $Y = \{Y^{(1)}, \dots, Y^{(t)}\}$, $Y^{(l)} = (\mathbf{y}_1^{(l)}, \dots, \mathbf{y}_n^{(l)})$ being the multi-sequence MR dataset at time point l . Once the images are preprocessed (*cf.* Section 3), a m -D joint histogram is computed for each time-point and these histograms are merged into a global m -D joint histogram containing the information from all sequences and all time-points. As all time points are merged, we omit the l index for the vectors \mathbf{y}_i in the following paragraphs.

Trimmed Likelihood Estimator (TLE). In order to make the classification less sensitive to noise, registration error, partial volume effects (PVE) and last but not least pathological abnormalities (*e.g.* MS lesions) which do not fit well the GMM model, we propose to replace the MLE by the TLE proposed by Neykov *et al.* [13].

The TLE is defined by: $\tilde{\theta} = \arg \max_{\theta} \prod_{i=1}^h p(\mathbf{y}_{\nu(i)}; \theta)$, where $\nu(i)$ are the permutations of $i = \{1, \dots, n\}$ such as $r_{\nu(i)} \leq r_{\nu(i+1)}$, $r_i = -\log p(\mathbf{y}_i; \theta)$ being the residuals. The TLE and the MLE are related to each other in the same way as the LTS (Least Trimmed Square) and the LS (Least Square) criteria.

Optimization. We propose an iterative scheme to estimate $\tilde{\theta}$. We name this scheme the STREM algorithm. This scheme converges at least to a local minimum of the criterion [13]. It can be summarized in the following way:

1. computation of the MLE on the whole data set using an EM giving a first estimation of the hyper-parameter $\hat{\theta}$;
2. sorting of the residuals : $r_i = -\log p(\mathbf{y}_i; \hat{\theta})$;
3. computation of the MLE on the voxels with the h smallest residuals ($h > 50\%$ of data), giving a new estimation $\hat{\theta}'$ (still using an EM);
4. back to step 2 until convergence.

Classification. Once healthy tissue model parameters are robustly estimated using STREM, we extract MS lesions as outliers of this model using the Mahalanobis distance (STREM-1), then we refine the segmentation using MS *a priori* information (STREM-2). We term the lesions class c_{k+1} .

1. For each vector \mathbf{y}_i in the joint histogram, the Mahalanobis distance

$$d_{i,j} = (\mathbf{y}_i - \mu_j)^T \Sigma_j^{-1} (\mathbf{y}_i - \mu_j)$$

between each class c_j is computed and gives a measure of how \mathbf{y}_i fits the model. These distances follow a chi-square distribution with m degree of freedom (χ_m^2). For a vector \mathbf{y}_i , if the Mahalanobis distance between each class j is greater than the critical value of χ_m^2 distribution for a given p-value, then the vector is considered as an outlier and belongs to c_{k+1} .

2. This potential lesions set still includes partial volume effects, noise or vessels. To discriminate real lesions from these false positives, we bring some *a priori* knowledge about MS and its manifestations in MRI. This *a priori* knowledge is summarized in Tab.1, where -1 , $+1$ and 0 denote respectively hyposignal, hypersignal and normally appearing with respect to the WM. Among the outliers extracted using the Mahalanobis distance those which follow the rules in Tab.1 belong to c_{k+1} .

Finally, the segmented images $X = (X^{(1)}, \dots, X^{(t)})$, $X^{(l)} = (x_i^{(l)}, x_i^{(l)} \in \{c_1, \dots, c_{k+1}\})$ being the segmentation at time point l , are constructed.

The advantage of such a multidimensional analysis is twofold: from a computational point of view, it increases the signal-to-noise ratio of the dataset and thus gives a more precise segmentation; from a clinical point of view, we use the information from various sequences, and we study the temporal variations.

Table 1. *A priori* knowledge about MS. Inflammatory MS lesions are hyper-intense with respect to the WM in T2-w and PD-w images. Young (4 to 6 weeks) lesions are hyper-intense in Gd T1-w images. Necrosis MS lesions are hypo-intense in T1-w images.

| | T1-w | Gd T1-w | T2-w | PD-w |
|-----------------------------|------|---------|------|------|
| young inflammatory lesions | 0 | +1 | +1 | +1 |
| others inflammatory lesions | 0 | 0 | +1 | +1 |
| necrosis | -1 | 0 | +1 | +1 |

3 Validation and Clinical Experimentation

Before applying the STREM algorithm, we extract the intra-cranial contents using a level set method [14]. Then, a bias field correction is done to reduce the radio-frequency inhomogeneities [15]. Finally, the various sequences and time points are rigidly registered [16], and global intensity changes between consecutive scans are corrected [17].

We first quantitatively validate the method using the BrainWeb database. Then, we show preliminary results on real data.

3.1 Quantitative Validation Using BrainWeb

BrainWeb is used for validation purposes. We simulate T1-w, T2-w and PD-w synthetic 3D MR images [18] with a 1mm slice thickness and 3% noise level. We simulate a set of healthy images (set A) and a set of MS images (set B).

The Dice Similarity Coefficient (DSC) is used to compare segmentations. Given two targets R_1 and R_2 , the DSC is defined by:

$$DSC(R_1, R_2) = 2 \cdot \text{card}(R_1 \cap R_2) / \text{card}(R_1 + R_2).$$

First, we prove the impact of using multiple sequences and, second, the impact of using longitudinal data. The third paragraph shows the improvement brought by the two previously proposed refinements (STREM-1 and -2).

Impact of multi-sequence analysis. We previously stressed the importance of such a multi-sequence analysis from a clinical point of view in MS study. Here, we focus on the computational improvement of using multiple sequences even for the segmentation of healthy tissues (set A). We empirically show that using several sequences improves the segmentation doing three experiments:

- (i) mono-sequence STREM on T2-w giving X_{T2} and on PD-w giving X_{PD} ;
- (ii) multi-sequence STREM on T2-w and PD-w giving $X_{T2,PD}$;
- (iii) multi-sequence STREM on the whole set A giving $X_{T1,T2,PD}$.

The obtained DSCs are compared and show that (i) < (ii) < (iii) (see Tab. 2). Values for CSF are not mentioned because it is not well segmented using only the PD-w sequence. We use $h = 96\%$.

Impact of the longitudinal analysis. The longitudinal analysis allows to detect both MS lesions and their evolution. Our method treats in a unified manner the segmentation and tracking of MS lesions. That is to say, the segmentation at a given point is used to allow better tracking of lesions over time, and, conversely, all the information over time is used to allow better segmentation of lesions at a given time point.

As shown in the previous section, using three sequences jointly (T1-w, T2-w and PD-w) improves the segmentation. In this second experiment, we simulate a longitudinal dataset with two time points. The first time point is the set A, $Y_A = Y_{T1,T2,PD}$. The second time point is the set B, $Y_B = Y_{T1_{MS},T2_{MS},PD_{MS}}$. The following experiments are done:

- (i) multi-sequence STREM-1 on Y_A giving $X_A (= X_{T_1, T_2, PD})$;
- (ii) multi-sequence STREM-1 on Y_B giving X_B ;
- (iii) multidimensional STREM-1 on $Y = (Y_A, Y_B)$ giving $X = (X_{A/B}, X_{B/A})$.

MS lesions are rejected by STREM-1: $h = 96\%$ and $p_value = 0.025$.

We compare the DSCs between multi-sequence segmentation and multidimensional segmentation for Y_A and Y_B . The multidimensional (multi-sequence and longitudinal) analysis improves the segmentation (see Tab. 3).

Table 2. Set A. The DSCs increase when using several sequences simultaneously for the hyper-parameter estimation.

| | WM | GM |
|--------------------|------|------|
| X_{T_2} | 0.86 | 0.84 |
| X_{PD} | 0.88 | 0.83 |
| $X_{T_2, PD}$ | 0.93 | 0.88 |
| $X_{T_1, T_2, PD}$ | 0.95 | 0.89 |

Table 3. Set A and B. The DSCs increase with the number of time points, particularly for all tissues in set A and MS lesions.

| | WM | GM | CSF | lesions |
|-----------|------|------|------|---------|
| X_A | 0.95 | 0.89 | 0.56 | - |
| $X_{A/B}$ | 0.96 | 0.95 | 0.65 | - |
| X_B | 0.95 | 0.90 | 0.59 | 0.48 |
| $X_{B/A}$ | 0.95 | 0.90 | 0.58 | 0.75 |

Impact of refinements. We perform the three following experiments:

- (i) STREM on $Y = (Y_A, Y_B)$ with $h = 96\%$;
- (ii) STREM-1 on Y with $h = 96\%$ and $p_value = 0.025$;
- (iii) STREM-2 on Y with $h = 96\%$ and $p_value = 0.025$.

Table 4. DSC measures for $X_{B/A}$. The DSC of MS lesions increases with the refinements.

| | WM | GM | CSF | lesions |
|---------|------|------|------|---------|
| STREM | 0.95 | 0.90 | 0.58 | 0.58 |
| STREM-1 | 0.95 | 0.90 | 0.58 | 0.75 |
| STREM-2 | 0.95 | 0.90 | 0.58 | 0.76 |

For STREM, we consider as outliers the vectors that have the bigger residuals. STREM-1 performs better than STREM in terms of DSC because it is more specific. STREM-2 is still more specific than STREM-1 without being less sensitive and thus performs better. The increment is not obvious on this data because there are only few outliers which are not MS lesions (*cf.* Tab.4).

3.2 Results on a Real Multidimensional Data Set

To be more complete, we present a robust segmentation on a multidimensional clinical data set. This multidimensional MR dataset was acquired on a GE 1.5T scanner and corresponds to a patient follow-up every three months during one year. Each acquisition is composed of three sequences: PD-w (TR = 2740 ms, TE

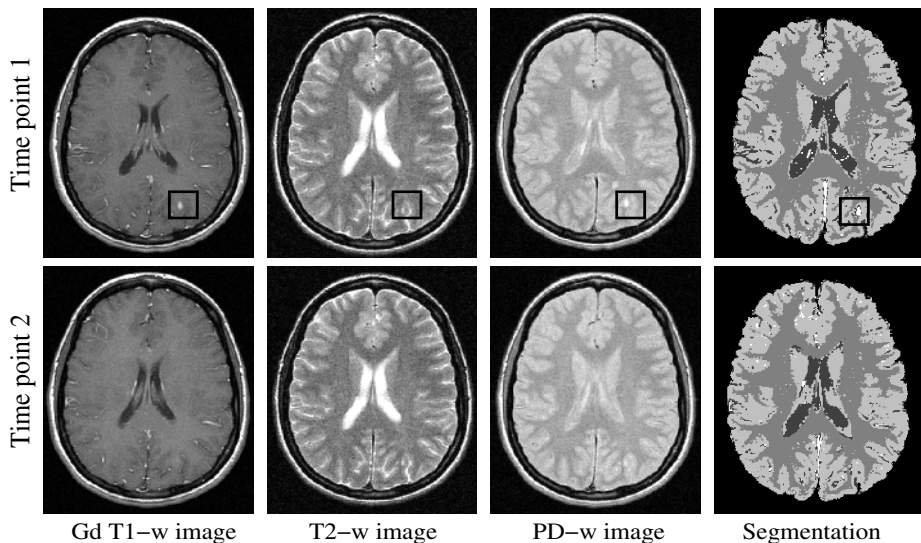


Fig. 1. STREM-1 on clinical data set ($h = 95\%$ and $p\text{-value} = 0.01$). A young lesion appears at time point 1 and disappears at time point 2. Preliminary results for multidimensional segmentation show the detection of a small and evolving lesion.

= 22.9 ms), T2-w (TR = 2740 ms, TE = 91.7 ms), and T1-w images after gadolinium injections (TR = 460 ms, TE = 20.0 ms), with a slice thickness of 3mm.

The STREM-1 is ran on the three sequences and the five time points simultaneously. Results for the first two time points are presented (there is no more evolution after). The refinement based on lesion intensity rules is not yet computed which explains the false positives (localized in the CSF). Even considering its small size, a lesion is segmented at time point 1 (see Fig. 1). This result is very promising because even very small structures are recovered. In addition, these small and evolutive lesions correspond to early stage of the pathology where MR lacks specificity. This encouraging result should even be improved by applied the lesion intensity rules to make the segmentation even more specific.

4 Conclusion

We proposed a consistent robust multidimensional iterative scheme to segment MS lesions over time. The advantages of this method are numerous. Using several sequences and time points increases the signal to noise ratio and thus improves the hyper-parameter estimation. From a clinical point of view, such a multi-sequence approach allows to take into account all types of MS lesions, while using the longitudinal data allows to handle the segmentation of static and evolving MS lesions in an unified manner. Our robust estimation scheme, coupled with refinements steps using Mahalanobis distances and a priori knowledge allows to differentiate MS lesions from other voxels showing abnormal intensities.

In the future, we will validate our algorithm on larger clinical databases, by comparing our classifications of lesions with ones manually segmented by an expert according to their type (young, inflammatory, necrosis) and their evolution characteristics over time.

References

1. C. Lucchinetti *et al.* Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. *Annals of Neurology*, 47(6):707–717, June 2000.
2. D.H. Miller. Biomarkers and surrogate outcomes in neurodegenerative disease: lessons from multiple sclerosis. *NeuroRx*, 1(2):284–294, April 2004.
3. P. Schroeter *et al.* Robust Parameter Estimation of Intensity Distributions for Brain Magnetic Resonance Images. *IEEE TMI*, 17(2):172–186, 1998.
4. K. Van Leemput *et al.* Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE TMI*, 20(8):677–688, 2001.
5. G. Dugas-Phocion *et al.* Hierarchical segmentation of multiple sclerosis lesions in multi-sequence MRI. In *ISBI'04*, Arlington, USA, April 2004.
6. D. Welti *et al.* Spatio-temporal segmentation of active multiple sclerosis lesions in serial MRI data. In *IPMI'2001*, Davis, USA, June 2001.
7. M. Bosc *et al.* Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution. *NeuroImage*, 20(2):643–656, October 2003.
8. S. Prima *et al.* Multivariate statistics for detection of MS activity in serial multimodal MR images. In *MICCAI'2003*, Montreal, Canada, November 2003.
9. D. Rey *et al.* Automatic detection and segmentation of evolving processes in 3D medical images: Application to multiple sclerosis. *Medical Image Analysis*, 6(2):163–179, June 2002.
10. A. Shahar *et al.* A probabilistic framework for the detection and tracking in time of multiple sclerosis lesions. In *ISBI'04*, Arlington, USA, April 2004.
11. J. Sijbers *et al.* Maximum Likelihood estimation of Rician distribution parameters. *IEEE TMI*, 17(3):357–361, 1998.
12. A. Dempster *et al.* Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38, 1977.
13. N. Neykov *et al.* Mixture of GLMs and the trimmed likelihood methodology. In *Compstat'04*, pages 1585–1592, 2004.
14. C. Baillard *et al.* Segmentation of brain 3D MR images using level sets and dense registration. *Medical Image Analysis*, 5(3):185–194, 2001.
15. S. Prima *et al.* Maximum likelihood estimation of the bias field in MR brain images: investigating different modelings of the imaging process. In *MICCAI'2001*, Utrecht, The Netherlands, October 2001.
16. F. Maes *et al.* Multimodality image registration by maximization of mutual information. *IEEE TMI*, 16(2):187–198, April 1997.
17. A. Guimond *et al.* Three-Dimensional Multimodal Brain Warping using the Demons Algorithm and Adaptive Intensity Corrections. *IEEE TMI*, 20(1):58–69, 2001.
18. C.A. Cocosco *et al.* BrainWeb: Online Interface to a 3D MRI Simulated Brain Database. *HBM'97*, 5(4), 1997.