Medical Image Analysis 35 (2017) 446–457

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Automated tissue segmentation of MR brain images in the presence of white matter lesions

Sergi Valverde[a,*], Arnau Oliver[a], Eloy Roura[a], Sandra González-Villà[a], Deborah Pareto[b], Joan C. Vilanova[c], Lluís Ramió-Torrentà[d], Àlex Rovira[b], Xavier Lladó[a]

[a] Research institute of Computer Vision and Robotics, University of Girona, Ed. P-IV, Campus Montilivi, 17071 Girona, Spain
[b] Magnetic Resonance Unit, Dept of Radiology, Vall d'Hebron University Hospital, Spain
[c] Girona Magnetic Resonance Center, Spain
[d] Multiple Sclerosis and Neuroimmunology Unit, Dr. Josep Trueta University Hospital, Spain

ARTICLE INFO

Article history:
Received 3 March 2016
Revised 27 August 2016
Accepted 29 August 2016
Available online 30 August 2016

Keywords:
Brain
MRI
Multiple sclerosis
Automatic tissue segmentation

ABSTRACT

Over the last few years, the increasing interest in brain tissue volume measurements on clinical settings has led to the development of a wide number of automated tissue segmentation methods. However, white matter lesions are known to reduce the performance of automated tissue segmentation methods, which requires manual annotation of the lesions and refilling them before segmentation, which is tedious and time-consuming. Here, we propose a new, fully automated T1-w/FLAIR tissue segmentation approach designed to deal with images in the presence of WM lesions. This approach integrates a robust partial volume tissue segmentation with WM outlier rejection and filling, combining intensity and probabilistic and morphological prior maps. We evaluate the performance of this method on the MRBrainS13 tissue segmentation challenge database, which contains images with vascular WM lesions, and also on a set of Multiple Sclerosis (MS) patient images. On both databases, we validate the performance of our method with other state-of-the-art techniques. On the MRBrainS13 data, the presented approach was at the time of submission the best ranked unsupervised intensity model method of the challenge (7th position) and clearly outperformed the other unsupervised pipelines such as *FAST* and *SPM12*. On MS data, the differences in tissue segmentation between the images segmented with our method and the same images where manual expert annotations were used to refill lesions on T1-w images before segmentation were lower or similar to the best state-of-the-art pipeline incorporating automated lesion segmentation and filling. Our results show that the proposed pipeline achieved very competitive results on both vascular and MS lesions. A public version of this approach is available to download for the neuro-imaging community.

1. Introduction

Brain tissue volume based on Magnetic Resonance Imaging (MRI) is increasingly being used in clinical settings to assess brain volume in different neurological diseases such as Multiple Sclerosis (MS) (Giorgio and De Stefano, 2013). In MS, several studies have analyzed the histopathological changes in patients with respect to the progress of the disease, showing that the percentage of change in brain volume tends to correlate with worsening conditions (Pérez-Miralles et al., 2013; Sormani et al., 2014). However, manual segmentation of brain tissue is both challenging and time-consuming because of the large number of MRI slices for each patient that make up the three-dimensional information, and the inherent intra/inter-observer variability of manually segmented scans (Cabezas et al., 2011). The development of automated MS tissue segmentation methods that can segment large quantities of MRI data, do not suffer from intra/inter-observer variability and specific changes of the brain such as MS associated lesions and brain atrophy, is still an active research field (Klauschen et al., 2009; de Bresser et al., 2011; Valverde et al., 2015a; Mendrik et al., 2015).

There are various brain tissue segmentation methods that have been used in MS so far. General purpose intensity based methods combining intensity and a priori statistical anatomic information such as FAST (Zhang et al., 2001) or SPM (Ashburner and Friston, 2005) are widely used nowadays. However, tissue abnormalities found in MS image patients such as White Matter (WM) lesions reduce the accuracy of these techniques (Chard et al., 2010; Battaglini et al., 2012). This causes an overestimation of

* Corresponding author. Fax: +34 972 418976.
E-mail address: svalverde@eia.udg.edu (S. Valverde).

http://dx.doi.org/10.1016/j.media.2016.08.014
1361-8415/

Gray Matter (GM) tissue not only by the effect of hypointense WM lesion voxels classified as GM, but also by the effect of these lesion voxels on normal-appearing tissue (Valverde et al., 2015b). In these cases, methods that in-paint lesions on the T1-weighed image (T1-w) may be used to reduce the effects of the WM lesions on tissue segmentation (Chard et al., 2010; Battaglini et al., 2012; Valverde et al., 2014). These methods refill WM lesions with signal intensities of the normal-appearing WM before tissue segmentation However, MS lesions have to be delineated manually first, which may be a tedious, challenging and time-consuming task depending on the characteristics of the image (Lladó et al., 2012).

Regarding this issue, a wide number of automated lesion segmentation techniques have been proposed during the last years (Lladó et al., 2012; García-Lorenzo et al., 2013). Most of them integrate other imaging modalities such as T2-weighted, Proton Density (PD) and Fluid Attenuated Inverse Recovery (FLAIR), as these modalities present a high contrast between tissue and lesions (Lladó et al., 2012). More recent techniques include supervised learning classifiers based on a spatial decision forest (Geremia et al., 2011), statistical methods (Sweeney et al., 2013), patch-based models (Guizard et al., 2015) or adaptive dictionary learning methods (Deshpande et al., 2015). Furthermore, different unsupervised learning techniques make use of probabilistic models to separate WM lesions from normal-appearing tissue by considering lesions as an outlier class (Harmouche et al., 2015; Tomas-Fernandez and Warfield, 2015; Jain et al., 2015). Also, other unsupervised techniques make use of the signal intensity of lesions on FLAIR to threshold regions with similar intensity to WM lesions. These methods incorporate afterwards various post-processing steps to automatically classify these regions as either WM lesions or normal-appearing tissue (Schmidt et al., 2012; Roura et al., 2015). In contrast, there are fewer studies that have focused on tissue segmentation of MS images containing lesions. Those include non-supervised techniques combining intensity, anatomical and morphological maps (Nakamura and Fisher, 2009; Shiee et al., 2010), or supervised methods such as statistical classifiers (Datta and Narayana, 2013), atlas based nearest-neighbor methods (De Boer et al., 2009) and sparse dictionary learning approaches (Roy et al., 2015).

The increasing amount of published studies regarding automated WM lesion segmentation may be due to the particular need of a quantitative analysis of focal MS lesions in individual and temporal studies (Lladó et al., 2012). Recent studies in MS (Chard et al., 2010; Gelineau-Morel et al., 2012; Ceccarelli et al., 2012; Pérez-Miralles et al., 2013; Popescu et al., 2014; Magon et al., 2014; Valverde et al., 2015b) indicate a certain tendency to the use of widely validated segmentation tools such as Siena (Smith et al., 2002), FAST (Zhang et al., 2001) or SPM (Ashburner and Friston, 2005) in combination with automated lesion segmentation and/or lesion-filling approaches. However, their application in clinical practice is still not generalized (Giorgio and De Stefano, 2013).

In this paper, we present the Multiple Sclerosis SEGmentation pipeline (*MSSEG*). This pipeline is a multi-channel method designed to segment GM, WM and cerebro-spinal fluid (CSF) tissues in images of MS patients. This method was motivated by our previous analysis of the effects of tissue segmentation on MS images (Valverde et al., 2015b), the role of lesion-filling (Valverde et al., 2014), and its combination with automated lesion segmentation on tissue segmentation (Valverde et al., 2015b). Similar to the work of Nakamura and Fisher (2009) and Shiee et al. (2010), our approach uses a combination of intensity and anatomical and morphological prior maps to guide the tissue segmentation. However, here the tissue segmentation is based on a robust, partial volume segmentation where WM outliers are estimated and refilled before the segmentation using a multi-channel post-processing algorithm. This post-processing algorithm was partially inspired by the MS lesion segmentation algorithm proposed by Roura et al. (2015). Nevertheless, here we integrate multi-channel support, partial volume segmentation, spatial context, and prior anatomical and morphological atlases. We perform a quantitative and qualitative evaluation of our approach with general and MS data: we first evaluate the accuracy when using the MRBrainS13 challenge database, which is composed of images containing vascular WM lesions and includes manual tissue annotations, allowing to compare the performance of our method with different state-of-the-art techniques that also competed in the challenge. We also evaluate the performance of our proposal on a set of MS patient images with different lesion burdens. On this data, we analyze the differences in the performance of our approach when using only T1-w or when using both T1-w and FLAIR modalities. We compare it with other state of the art pipelines that include automated lesion segmentation and filling processes before tissue segmentation. The *MSSEG* method is currently available for downloading at our research group webpage (http://atc.udg.edu/nic/msseg).

2. Methods

The proposed brain tissue segmentation method is composed of five different steps: registration of a statistical atlas into the T1-w space (Section 2.2), tissue estimation (Section 2.3), detection and re-assignation of lesion candidates to T1-w (Section 2.4), tissue re-estimation, and partial volume re-assignation of tissue maps into CSF, GM and WM (Section 2.5). The overall schema of the pipeline is depicted in Fig. 1. We describe each step in detail in the following subsections.

2.1. Notation

To describe our approach, we employ the following notations. T and F denote the input images T1-w and FLAIR, respectively. P^c denotes a probabilistic tissue atlas of a particular class $c = \{csf, \ldots, wm\}$. S^{st} denotes a morphological brain atlas of a particular parcellated structure st. For each of the above images, T_j, F_j, P_j^c and S_j^{st} denote an observation at a voxel $j \in \Omega$, Ω being the image domain.

2.2. Tissue prior registration

The MNI-ICBM 152 2009a Nonlinear T1-w average structural template image[1] was first affine registered to the native T1-w image space. Affine registration was based on a block matching approach (Ourselin et al., 2002), followed by a non-rigid registration using a fast free-form deformation method (Modat et al., 2010). Nifty Reg package[2] was used in both registration processes. Transformation parameters obtained were then used to resample the available MNI CSF, GM and WM tissue priors to the T1-w space. The resampled probabilistic tissue maps P^{csf}, P^{gm} and P^{wm} were extended to build intermediate partial volumes P^{csfgm} as ($P^{csf} \geq 0.4 \cap P^{gm} \geq 0.4$) and P^{gmwm} as ($P^{gm} \geq 0.4 \cap P^{wm} \geq 0.4$) and taking the mean value of the two input atlases.

Moreover, a brain structural atlas of the cortical GM (S^{CORTEX}), ventricles (S^{VENT}), basal ganglia (S^{BASAL}) and brainstem ($S^{BRAINSTEM}$) was also resampled to the native T1-w space with the same transformation parameters used for the tissue probability maps. The initial morphological atlas had been built a priori by automatically segmenting the MNI-ICBM 152 T1-w image into each of these brain structures using the hierarchical algorithm proposed by

[1] http://www.bic.mni.mcgill.ca/ServicesAtlases/ICBM152NLin2009.
[2] http://cmictig.cs.ucl.ac.uk/wiki/index.php/NiftyReg.

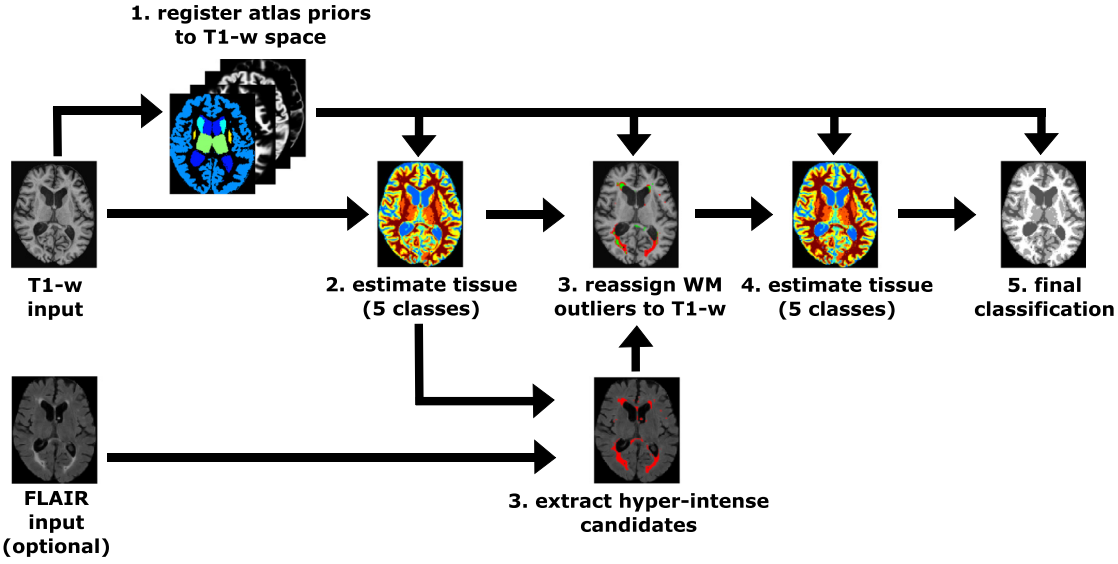

Fig. 1. The proposed method *MSSEG* consists of five different steps: (1) Three statistical *a-priori* tissue atlases (CSF, GM and WM) and a brain structure atlas are first registered into the patient space (Section 2.2) and then used to (2) guide the tissue segmentation of the input T1-w image (Section 2.3). (3) Then, the same output segmentation is employed to detect and refill WM outliers as normal-appearing WM based on the registered *a-priori* and hyper-intense FLAIR maps if available (Section 2.4). The voxel intensities of candidate regions on T1-w are then refilled with normal-appearance WM intensities and (4) the tissue is re-estimated again using the process described in (Section 2.3). (5) Finally, intermediate volume maps are reassigned into CSF, GM and WM using both neighbor and spatial prior information (Section 2.5).

Pohl et al. (2007) on EMSegmenter,[3] and manually adjusted by an expert technician afterwards.

2.3. Tissue estimation

Brain tissue was estimated following a robust fuzzy clustering approach similar to the one proposed in Pham (2001), as this method provides a fairly robust behavior including spatial context information, applicability to multichannel data, the ability to model uncertainty within the data and a straightforward implementation (Pham, 2001). Similar to previous approaches (Shattuck et al., 2001; Bach-Cuadra et al., 2005), additional mixed-based classes (CSFGM and GMWM) were added between adjacent pure classes GM, WM and CSF in order to encode the distribution of partial volume voxels, i.e. voxels that contain not only a single tissue but rather a mixture of two or more adjacent tissues types. The spatial penalizing weights were also extended by incorporating the probabilistic tissue priors in the segmentation process, similar to Shiee et al. (2010). Hence, we modified the objective function proposed by Pham (2001) in order to incorporate also prior-atlas information as follows:

$$J_{MSSEG} = \sum_{j\in\Omega}\sum_{k=1}^{C} u_{jk}^{q}||T_j - v_k||^2 + \frac{\beta}{2}\sum_{j\in\Omega}\sum_{k=1}^{C} u_{jk}^{q}\sum_{l\in N_j^w}\sum_{m\in M_k}^{C} u_{lm}^{q} + \frac{\gamma}{2}\sum_{j\in\Omega}\sum_{k=1}^{C} u_{jk}^{q}\sum_{l\in N_j^w}\sum_{m\in M_k}^{C} P_l^m \quad (1)$$

where $\{k \in C \mid C = \{csf, csfgm, gm, gmwm, wm\}\}$, u_{jk} denotes the membership probability of each voxel j for a particular class, v_k are the cluster signal intensity centers of each class, N_j^w is the set of two-dimensional (2D) $(2w+1)^2$ or three-dimensional (3D) $(2w+1)^3$ neighbors centered on the voxel j, and $M_k = \{1,\ldots,C\}\backslash\{k\}$. This approach depends on four parameters to adjust the membership functions: the weighting parameter q that controls the degree of fuzziness, the spatial constraint parameter β that controls the amount of neighbor information added, the prior belief parameter γ used to control the amount of prior atlas information about each tissue, and finally the window radius of neighbors w.

Similar to the work of Pham (2001), the objective function was minimized using the iterative Algorithm 1 that evaluated the centroids and membership functions that satisfy a zero gradient condition with respect to the objective. In our approach, we also adapted the initial formulation proposed by Pham (2001) to incorporate prior-atlas information. u_{jk} satisfied the condition to be at local minimum with respect to Eq. 1 when:

Initial centroids v_k were estimated for each class C by taking the mean signal intensity of the voxels on the T1-w image with prior-tissue probability $P_j^k \geq 0.5$. The membership function u_{jk} was also adapted to incorporate prior-atlas information and computed as follows:

$$u_{jk} = \frac{(||T_j - v_k||^2 + \beta\sum_{l\in N_j}\sum_{m\in M_k} u_{lm}^{q} + \gamma\sum_{l\in N_j}\sum_{m\in M_k} P_l^m)^{-1/(q-1)}}{\sum_{i=1}^{C}((||T_j - v_i||^2 + \beta\sum_{l\in N_j}\sum_{m\in M_i} u_{lm}^{q} + \gamma\sum_{l\in N_j}\sum_{m\in M_k} P_l^m)^{-1/(q-1)}} \quad (2)$$

The five class tissue segmentation mask SEG_j was computed by assigning to each voxel the class with the maximum membership as follows:

$$SEG_j = \arg\max_k u_{jk} \quad \forall j \in \Omega \quad (3)$$

Algorithm 1 Tissue estimation.

1: Obtain the initial estimates of the centroids for each class $k = \{1,\ldots,C\}$:
$$v_k = \frac{1}{n}\sum_{j\in\Omega}(T_j^k \mid P_j^k \geq 0.5)\ n = |(T_j^k | P_j^k \geq 0.5)|$$
2: Compute the membership functions u_{jk}
3: Compute the new centroids:
$$v_k = \frac{\sum_{j\in\Omega} u_{jk}^{q} T_j}{\sum_{j\in\Omega} u_{jk}^{q}}\ k = \{1,\ldots,C\}$$
4: Repeat steps 2 and 3 until convergence

[3] https://www.slicer.org/slicerWiki/index.php/EMSegmenter-Overview.

The parameters q, γ and w can be tuned manually to increase the performance of the method, but were set to default values $q = 2$, $\gamma = 0.1$ and $w = 1$ with 2D that worked well in the majority of cases. In contrast, the β parameter depends on the brightness of the image, the deviation of the signal intensities of voxel class members with respect to their centroid value, and image noise (Pham, 2001). Hence, choosing a proper value for the β parameter is important to obtain an optimal or near-optimal performance. In our implementation, the β parameter was automated by fitting a function of the optimal empirical selection of the parameter with respect to different levels of noise. This particular function was increasing the weight of β with images with high level of noise by allowing more spatial constraints, reducing the effect of noise in tissue segmentation. In contrast, β was very low with images with low level of noise. To do this, we iteratively estimated the sub-optimal β parameter of 10 images of the Brainweb dataset[4] that included different noise level (1–9%) and ground-truth annotations. For each image, we also computed the noise level using the Fast Noise Variance method proposed by Immerkær (1996). Then, the correspondent β parameters and noise levels were used to fit a polynomial function to interpolate the β parameter. For all the evaluated images in this paper, we have automatically approximated the β as a function $B(x)$ of their noise level x as $B(x) = 0.0011x^4 - 0.0015x^3 + 0.0074x^2 - 0.001x + 0.05$.

2.4. WM outlier estimation and filling

By adding the intermediate tissue classes CSFGM and GMWM, WM lesion regions tended to be entirely classified as a single class out of WM either as GMWM, GM or CSFGM. This allowed us to detect lesion regions as WM outliers by analyzing all the local regions not initially segmented as WM based on their prior probability and spatial connection to WM.

First, different binary segmentation masks M^c were computed for each of the classes $c = \{gmwm, gm, csfgm\}$. For each mask, all 2D regions of connected components were computed using a flood-fill algorithm with a 4-connected neighborhood. We define the set of all 4-connected p regions given an input binary image as follows:

$$R_p^T \leftarrow \oplus(M_j^c, n), \quad p = \{1, \ldots, |R_p^T|\}$$

where the operator $\oplus$ refers to the connected components function and n is the number of connected neighbors.

Optionally, a map of hyperintense region candidates was computed on the FLAIR image following the same strategy shown in Roura et al. (2015). The binary mask M^{GM} was first used to compute the intensity distribution on the FLAIR image, where GM is typically hyperintense with respect to CSF and WM, and WM lesions are considered hyperintense outliers to GM. The mean and standard deviation of the GM distribution was computed using the full-width at half maximum (FWHM) of the main peak of a generated histogram. Then, an initial map of hyperintense regions voxels, H^{FLAIR}, was determined by thresholding the FLAIR image F as follows:

$$H_j^{FLAIR} = \begin{cases} 1 & \text{if } F_j > \mu + \alpha\sigma \\ 0 & otherwise \end{cases} \qquad \forall j \in \Omega \tag{4}$$

where μ and σ were the mean and standard deviation respectively, of the GM distribution as computed using the FWHM, and α was a weighting parameter that scaled the minimum signal intensity of outliers. The binary mask H^{FLAIR} was then used to group the candidate voxels into connected regions using the same method proposed before:

$$R_t^F \leftarrow \oplus(H^{FLAIR}, n), \quad t = \{1, \ldots, |R_t^F|\}$$

[4] http://brainweb.bic.mni.mcgill.ca/brainweb/.

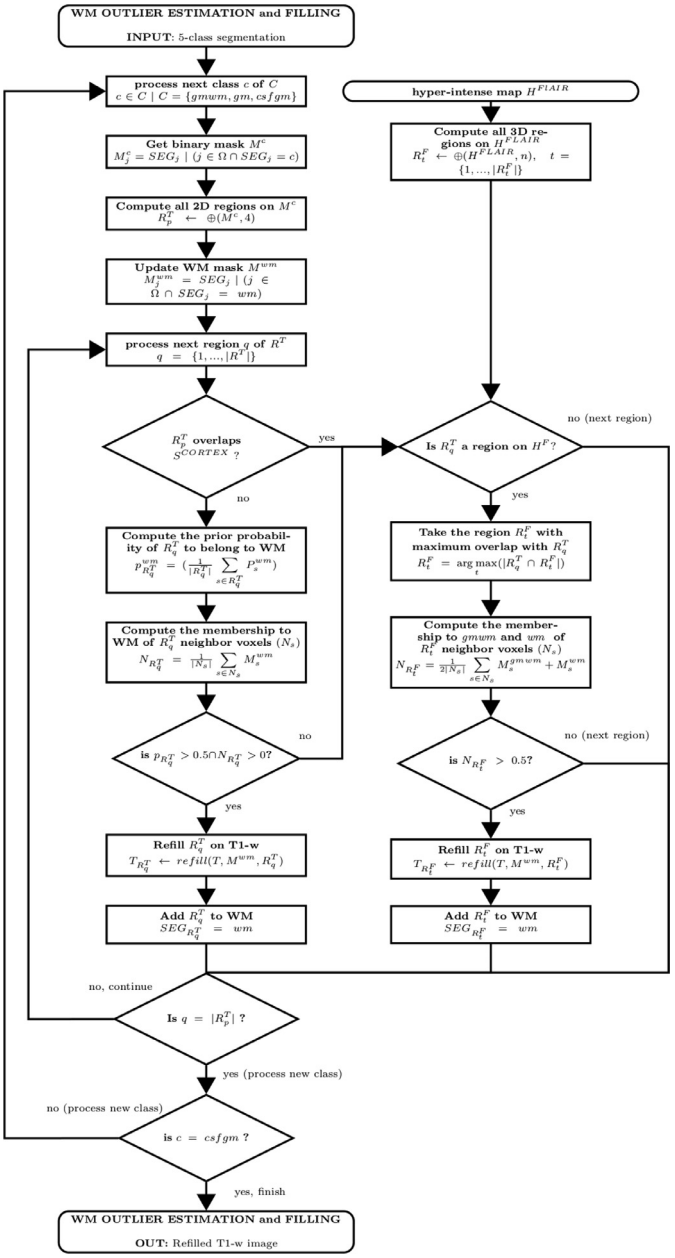

Fig. 2. Proposed algorithm to estimate and refill outlier candidate regions into T1-w. The algorithm takes the 5 class T1-w segmentation and the hyper-intensity map H^{FLAIR} if available as inputs. Connected regions of voxels with similar intensities are filtered based on their spatial location probability in tissue and morphological prior atlases. Selected regions are then refilled in the original T1-w image.

where n was set to 3D connected elements ($n = 6$) in order to reduce the amount of 2D false positive regions such as hyperintense sub-arachnoid tissue.

Given the computed binary masks for each tissue M^{gmwm}, M^{gm} and M^{csfgm}, the map of hyperintense voxels on FLAIR H^{FLAIR}, and its connected components R_t^F, we used a two-step iterative algorithm that first estimated the regions with a high probability of belonging to WM and refilled them into normal-appearing WM afterwards following the approach proposed by Valverde et al. (2014). Fig. 2 shows in detail each step of the algorithm. Regions overlapping cortical GM on the brain structure atlas S_s^{CORTEX} were only processed if a matched region was also hyperintense in FLAIR, in order to reduce the amount of false positive regions such as isolated cortical GM segmented regions. Regions not touching cortical GM were filtered based

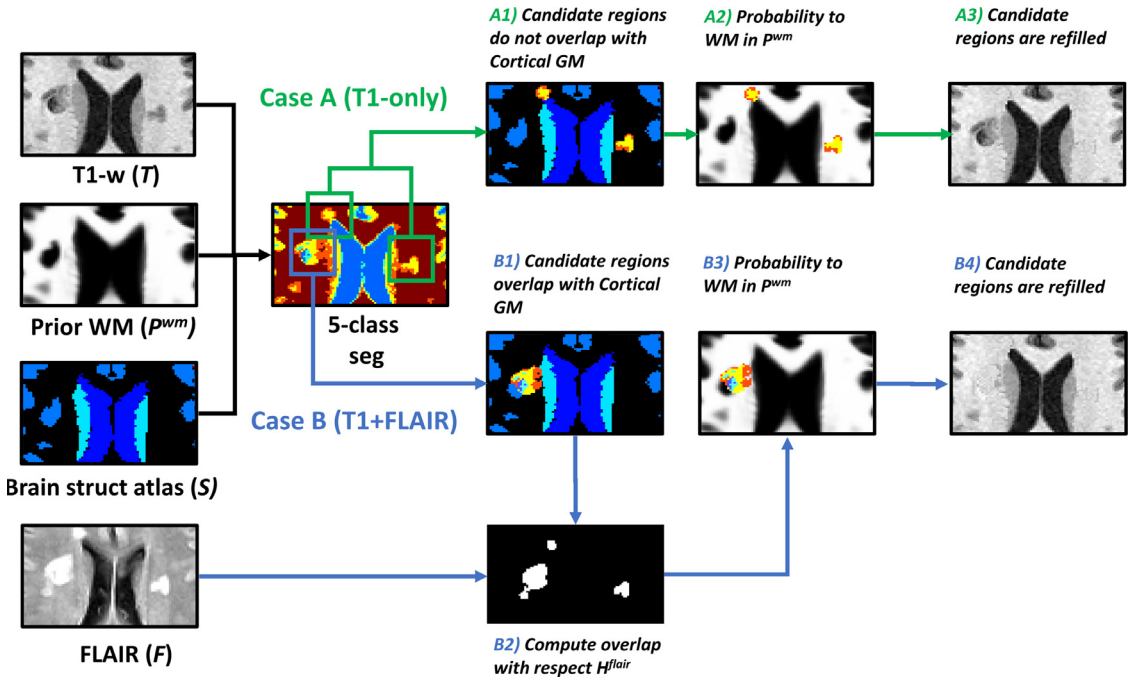

Fig. 3. Graphical explanation of the WM outlier algorithm for two different cases when only using T1-w (Case A) and when using T1-w and FLAIR (Case B). Given the initial 5-class tissue segmentation as explained in Section 2.3, all candidate regions were estimated and refilled using the iterative algorithm proposed in Section 2.4. When only using T1-w (Case A), candidate regions were processed when they did not overlap with cortical GM in the brain structural atlas S (A1). For each of these regions, the probability of belonging to WM was computed based on the prior P^{wm} atlas (A2), and these regions were refilled if the majority of the voxels had a probability of belonging to $WM \geq 50\%$ (A3). In contrast, when using both T1-w and FLAIR (Case B), candidate regions that were not previously processed because of the overlap with cortical GM, were then evaluated if it also existed an overlap with respect to the FLAIR H^{FLAIR} map (B2). The rest of the processes (B3 and B4) were similar to the ones of the case A.

on their prior probability of belonging to WM and their distance to surrounding WM. Then, matched regions where half of their surrounding neighbors were actually classified as *GMWM* or *WM* were refilled in the T1-w image with normal appearing WM signal intensities. Note that classes were visited from *GMWM* to *CSFGM* in order to add new belief of the actual WM and use it to filter the next region. R_p^T regions were computed in 2D in order to reduce the impact of the registration differences between the brain structural atlas and the T1-w image. By choosing R_p^T regions in 2D, we limited the impact of the registration process to a single slice, reducing the number of false negative estimated regions.

If the FLAIR modality is not available, H^{FLAIR} is automatically set to zero, disabling the evaluation of R^H regions and henceforth forcing the method to evaluate the next T1 region R_p^T. If FLAIR is used, all the regions that were discarded in the first part of the algorithm or overlapped the cortex, were filtered according to their spatial attributes in the FLAIR image. Each discarded region R_q^T in the segmented mask *SEG* was matched with a particular region in FLAIR R_t^H based on their overlap $(R_t^H \mid t = \arg\max_t(|R_t^H \cap R_q^T|))$. Then, matched regions where half of their surrounding neighbors were actually classified as *GMWM* or *WM* were refilled in the T1-w image with normal appearing WM signal intensities. Fig. 3 depicts a graphical schema of the different processes followed by the WM outlier algorithm when using only T1-w and T1-w + FLAIR, respectively. In all cases, we referred to the neighboring voxels of a region N_S as the neighbors with one voxel of distance from the region's boundaries.

2.5. Partial volume maps

Once the WM outliers were refilled in T1-w image as normal-appearing WM, the resulting refilled image was used to segment the brain tissue following the same method described in Section 2.3. Afterwards, partial volume maps (*CSFGM*) and (*GMWM*) were reassigned to each of the three main classes CSF, GM and WM following a region-wise approach.

Local 2D regions with similar intensities that were classified as *CSFGM* and *GMWM* were estimated using the same connected component algorithm described before. The structural brain atlas S was then used to reassign regions where at least half of their voxels overlapped with certain structures as follows:

$$SEG_{R_p} = \begin{cases} CSF & \text{if } \left(\frac{1}{|R_p|}\sum_{s\in R_p} S_s^{VENT}\right) > 0.5 \\ GM & \text{if } \left(\frac{1}{2|R_p|}\sum_{s\in R_p} S_s^{CORTEX} + S_s^{BASAL}\right) > 0.5 \\ WM & \text{if } \left(\frac{1}{|R_p|}\sum_{s\in R_p} S_s^{BRAINSTEM}\right) > 0.5 \end{cases} \quad (5)$$

for all the regions $p = \{1, \ldots, |R_p|\}$. The voxels not reassigned previously were reclassified by adding them to the surrounding pure class with the most similar intensity as follows:

$$SEG_j = \arg\min_c \left| T_j - \frac{1}{|N_j|}\sum_{s=1}^{|N_j|}(T_s \mid SEG_s = c) \right| \quad (6)$$

for pure classes $c = \{csf, gm, wm\}$ and partial volume voxels $j = \{\forall j \in \Omega \mid SEG_j = csfgm \cup gmwm\}$. The radius for neighbor voxels was set to 6 in two dimensions. Fig. 4 depicts the partial volume re-assignation process for one particular T1-w image.

2.6. Implementation details

The proposed pipeline was entirely developed in MATLAB (v2014a, The Mathworks Inc, US), except for the registration process that was run using the available NiftyReg package (Ourselin et al., 2002; Modat et al., 2010). The method was configured to run either in CPU or GPU. Experiments were carried

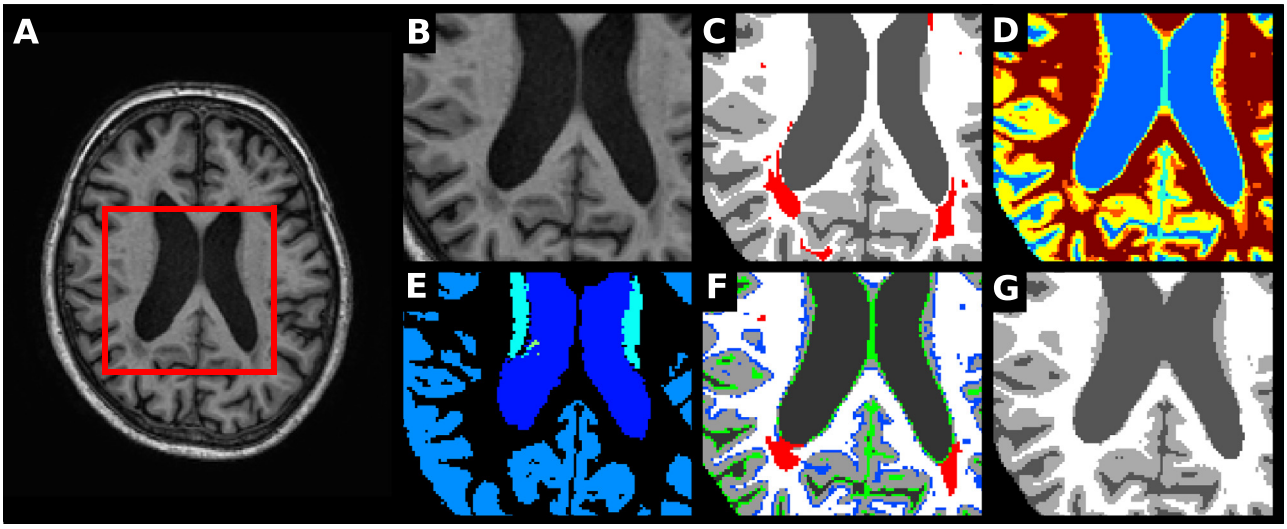

Fig. 4. Partial volume assignment during tissue segmentation (Section 2.5). A) Original T1-w image. B) Detailed view of the T1-w image. C) Tissue ground-truth with WM lesions highlighted in red. D) Initial 5 class tissue segmentation where *csf, csfgm, gm, gmwm* and *wm* tissues are depicted in blue, cyan, yellow, orange and red, respectively (see Section 2.3). E) Morphological brain tissue atlas with parcellated GM regions and ventricles. F) Partial volume *csfgm* and *gmwm* regions (depicted in green and blue, respectively). Previously estimated lesion candidates are depicted in red (see Section 2.4). 2D regions with half of their voxels overlapping with the morphological atlas are reassigned to the correspondent tissue. The rest of the voxels are reassigned to the neighboring pure class with the most similar signal intensity. G) Final tissue segmentation with partial tissue volume maps re-estimated as CSF (dark gray), GM (light gray) and WM (white). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

out on a GNU/Linux machine with a single Intel core i7 processor at 3.4 Ghz (Intel Corp, US), and a NVIDIA K40 with 12GB of RAM (NVIDIA, US). The average execution time for the proposed method including registration and tissue segmentation was 8 min running on the CPU core. Execution time on the GPU was approximately 2 min, reducing the execution time on the CPU processor by four times.

3. Results

3.1. MRBrainS database

3.1.1. Data

The available public MRBrainS 2013 database[5] consisted of 20 scans with varying degrees of brain atrophy and white matter lesions. These scans were acquired on a 3.0T Philips Achieva MR scanner at the University Medical Center Utrecht (The Netherlands) with the following sequences: 3D T1-w (TR: 7.9 ms, TE: 4.5 ms), T1-Inverse Recovery (TR: 4416 ms, TE: 15 ms, and TI: 400 ms), and T2-weighted/FLAIR (TR: 11000 ms, TE: 125 ms, and TI: 2800 ms). Each of the scans was co-registered (Klein et al., 2010) and intensity-corrected (Ashburner and Friston, 2005) before releasing the data. The T1, T1-IR, and T2/FLAIR voxel size was ($0.96 \times 0.96 \times 3.00$ mm^3) after registration (Mendrik et al., 2015).

Three experts manually delineated each of the 20 scans into CSF, GM and WM and these annotations were used as the reference standard for the evaluation framework (Mendrik et al., 2015). Extended manual annotation containing various brain structures and white matter lesions for 5 scans were provided for training while the remaining 15 scans were blind and had to be skull-stripped and segmented into CSF, GM and WM by participating teams.

3.1.2. Evaluation:

The segmentation results had to be submitted online for external evaluation based on the following scores for the CSF, GM and WM tissues (c):

- Dice similarity coefficient (DSC_c) (Dice, 1945) between the manual tissue segmentation (GT_C) and the computed segmentation (SEG_C) masks:

$$DSC_c = \frac{2\ |SEG_c \cap GT_c|}{|SEG_c| + |GT_c|} \times 100 \tag{7}$$

- The modified Hausdorff distance (95th percentile) (Huttenlocher et al., 1993) between the manual tissue segmentation (GT_C) points p' and the computed segmentation points p in (SEG_C) masks:

$$h_c^{95} = \max_{p \in\ SEG_c} \min_{p' \in\ GT_c} |p - p'| \tag{8}$$

- The absolute difference in tissue volume (AVD_c) between manual tissue segmentation (GT_C) and the computed segmentation (SEG_C) masks:

$$AVD_c = \left\| \frac{|SEG_c| - |GT_c|}{|GT_c|} \right\| \tag{9}$$

In order to evaluate the performance of our method, we submitted two different segmentation sets; one using only the T1-w sequences, and the other using both T1-w and FLAIR images. We validated the performance by comparing our scores with those of other submitted segmentation pipelines.

3.1.3. Experiment details

The skull stripping of the input images was performed using a similar approach to other methods participating in the challenge (Jog et al., 2013; Opbroek et al., 2013; Rajchl et al., 2015). The 5 training images were non-rigidly registered to the image space of each of the T1-w (Modat et al., 2010), and the brainmask was generated by a simple voting of the registered masks. Afterwards, each mask was refined in the T1-IR image by thresholding hyperintense voxels.

All the parameters of our tissue segmentation method were set to default values ($q = 2, \gamma = 0.1, w = 1$). The β parameter was computed automatically as described in Section 2.3. The α parameter that scaled the minimum signal intensity on the H^F mask was set to $\alpha = 3$.

[5] Available for downloading at: http://mrbrains13.isi.uu.nl/.

3.1.4. Results

Table 1 shows the mean DSC_c, h_c^{95} and AVD_c scores obtained by our proposed method. We compare our scores with other non-supervised strategies that also participated in the challenge, such as FAST, SPM12, or VBM12,[6] and also with respect to the best ranked method proposed by Stollenga and Byeon (2015). The overall ranking of the methods also included the combined brain (GM + WM) and intracranial (CSF + GM + WM) volumes, which are not shown in the table for simplicity.[7] At the time of submitting our results in the online application, our approach, using both T1 and FLAIR (*MSSEG T1+FLAIR*), was the best unsupervised intensity model method in the challenge 7th position overall 31 participants), and its performance was very competitive in comparison with several supervised methods that were explicitly trained for the challenge. When using only the T1-w modality (*MSSEG only T1*), our method was ranked in the 10th position, but still clearly out-performed *FAST* (21th position), and *SPM12* using FLAIR+T1 (17th position), the T1-IR modality (18th position), and the T1-w modality (20th position).

Fig. 5 illustrates qualitatively the different steps performed by our approach. After registering the probabilistic atlases into the subject space (Fig. 5 panels E–H), tissue was estimated from the T1-w into 5 different classes (Fig. 5 panel I). Then, WM outliers were estimated in the T1-w image by analyzing all the regions not initially segmented as WM with a high probability of belonging to WM based on its spatial local probability and prior tissue information (red regions depicted in Fig. 5 panel J). If FLAIR was also provided, lesion candidates were analyzed as well based on their signal intensity in the FLAIR image and their spatial local probability of belonging to WM (green regions depicted in Fig. 5 panel J). Afterwards, the lesion candidate regions were refilled in the T1-w image with signal intensities similar to the WM, the refilled T1-w image was re-estimated again and CSFGM and GMWM volumes were reassigned to the three main classes (Fig. 5 panels K and L).

Table 1
Segmentation results on the 15 test images of the MRBrainS challenge. Mean DSC_c, h_c^{95} and AVD_c scores for CSF, GM and WM tissue are shown for our proposed method when using only the T1-w modality (*MSSEG only T1*) and when using the T1-w and FLAIR modalities (*MSSEG T1+FLAIR*). The obtained values are compared with the best approach at the time of writing this paper (Stollenga and Byeon, 2015), and also with other unsupervised intensity models techniques that also participated in the challenge such as *FAST, SPM12* and *VBM12*.

	DSC_c		
Method	CSF	GM	WM
Best method	83.72 ± 2.63	84.82 ± 1.37	88.33 ± 0.89
FAST only T1	69.95 ± 2.81	78.66 ± 2.24	85.98 ± 2.58
SPM12 only T1	70.69 ± 3.75	80.34 ± 2.37	85.58 ± 1.73
SPM12 T1+FLAIR	74.03 ± 3.42	81.17 ± 2.24	86.03 ± 1.48
SPM12 T1–iR	78.25 ± 3.78	79.41 ± 2.15	83.54 ± 2.14
VBM12	74.56 ± 2.70	82.29 ± 1.49	87.95 ± 1.71
MSSEG only T1 (T1–IR skull)	80.18 ± 2.67	82.06 ± 1.68	87.05 ± 1.46
MSSEG T1+FLAIR (T1–IR skull)	80.16 ± 2.67	82.20 ± 1.60	87.33 ± 1.35
	h_c^{95}		
Method	CSF	GM	WM
Best method	2.14 ± 0.36	1.70 ± 0.01	2.08 ± 0.33
FAST only T1	3.41 ± 0.25	4.35 ± 1.13	3.65 ± 0.85
SPM12 only T1	5.34 ± 1.47	2.93 ± 0.25	3.06 ± 0.08
SPM12 T1+FLAIR	4.59 ± 0.61	2.90 ± 0.15	3.00 ± 0.07
SPM12 T1–iR	4.01 ± 0.63	3.01 ± 0.29	3.60 ± 0.27
VBM12	3.03 ± 0.14	3.20 ± 0.32	2.32 ± 0.42
MSSEG only T1 (T1–IR skull)	2.81 ± 0.21	3.33 ± 0.21	2.91 ± 0.42
MSSEG T1+FLAIR (T1–IR skull)	2.81 ± 0.21	3.18 ± 0.15	2.88 ± 0.39
	AVD_c		
Method	CSF	GM	WM
Best method	7.09 ± 4.01	6.77 ± 3.28	7.05 ± 5.22
FAST only T1	11.83 ± 10.38	8.65 ± 6.34	11.47 ± 6.24
SPM12 only T1	23.24 ± 16.04	6.95 ± 6.57	5.99 ± 3.95
SPM12 T1+FLAIR	10.07 ± 4.86	10.59 ± 8.34	5.21 ± 3.88
SPM12 T1–iR	10.47 ± 5.69	7.23 ± 6.50	6.34 ± 4.61
VBM12	6.80 ± 4.57	5.91 ± 3.91	6.06 ± 4.42
MSSEG only T1 (T1–IR skull)	7.18 ± 3.33	6.15 ± 3.51	6.20 ± 5.45
MSSEG T1+FLAIR (T1–IR skull)	7.21 ± 3.31	5.99 ± 3.43	5.95 ± 5.44

3.2. MS database

3.2.1. Data

This non-public database of images was composed of 24 images of clinically isolated syndrome (CIS) patients acquired with a 3T Siemens MR scanner (Trio Tim, Siemens, Germany) with a 12-channel phased-array head coil (data from Hospital Vall D'Hebron, Barcelona, Spain). The following pulse sequences were obtained: 1) transverse proton density and T2-weighted fast spin-echo (TR = 2500 ms, TE = 16-91 ms, voxel size = 0.78 × 0.78 × 3 mm^3); 2) transverse fast T2-FLAIR (TR = 9000 ms, TE = 93 ms, TI = 2500 ms, flip angle = 120°, voxel size = 0.49 × 0.49 × 3 mm^3); and 3) sagittal 3D T1 magnetization prepared rapid gradient-echo (MPRAGE) (TR = 2300 ms, TE = 2 ms; flip angle = 9°; voxel size = 1 × 1 × 1.2 mm^3). For each scan, T1-w and FLAIR images were first skull-stripped using BET (Smith, 2002) and then intensity-corrected using the N3 method (Sled et al., 1998). Finally, FLAIR images were co-registered into the T1-w space and then re-aligned into the MNI space using SPM12 co-registration tools with the normalized mutual information as the objective function and tri-linear interpolation with no wrapping (Ashburner and Friston, 2005). White matter lesion masks were semi-automatically delineated from FLAIR using JIM software[8] by a trained technician at the hospital center. The mean lesion volume was 4.30 ± 4.84 ml (range 0.1–18.3 ml).

3.2.2. Evaluation

Expert manual annotations of tissues were not available for this database. As validated in previous MS studies (Battaglini et al., 2012; Valverde et al., 2015b,c), manual annotated lesions masks were used to refill WM lesions in the original T1-w scans with signal intensities similar to normal-appearing WM using the SLF lesion filling method (Valverde et al., 2014). Then, both the original and the refilled images were segmented into CSF, GM and WM tissues using our proposed approach. The performance of our tissue segmentation method was evaluated by computing the absolute difference in tissue volume (AVD_c) between the images segmented containing lesions and the same images where WM lesions were refilled before the tissue segmentation:

$$AVD_c = \left\| \frac{|SEG_c| - |GT_c^{fill}|}{|GT_c^{fill}|} \right\| \times 100 \quad (10)$$

where SEG_c refers to the output segmentation masks of the segmented images containing lesions, and GT_c^{fill} refers to the output tissue segmentation masks of images where lesions were filled before segmentation and considered as ground-truth.

Several works (DellOglio et al., 2014; Valverde et al., 2015c) have already shown that part of the actual error in tissue segmentation may be partially masked by opposite directions in the differences in total and normal-appearing tissue. For instance, at a certain lesion load, WM lesion voxels that have been misclassified as GM may have an impact in normal-appearing GM voxels, moving the intensity boundary threshold between GM and WM intensities. This cause that GM voxels with signal intensities in this

[6] http://www.neuro.uni-jena.de/.
[7] Overall ranking of methods for all the measurements can be consulted at http://mrbrains13.isi.uu.nl/results.php.
[8] Xinapse Systems, http://www.xinapse.com/home.php.

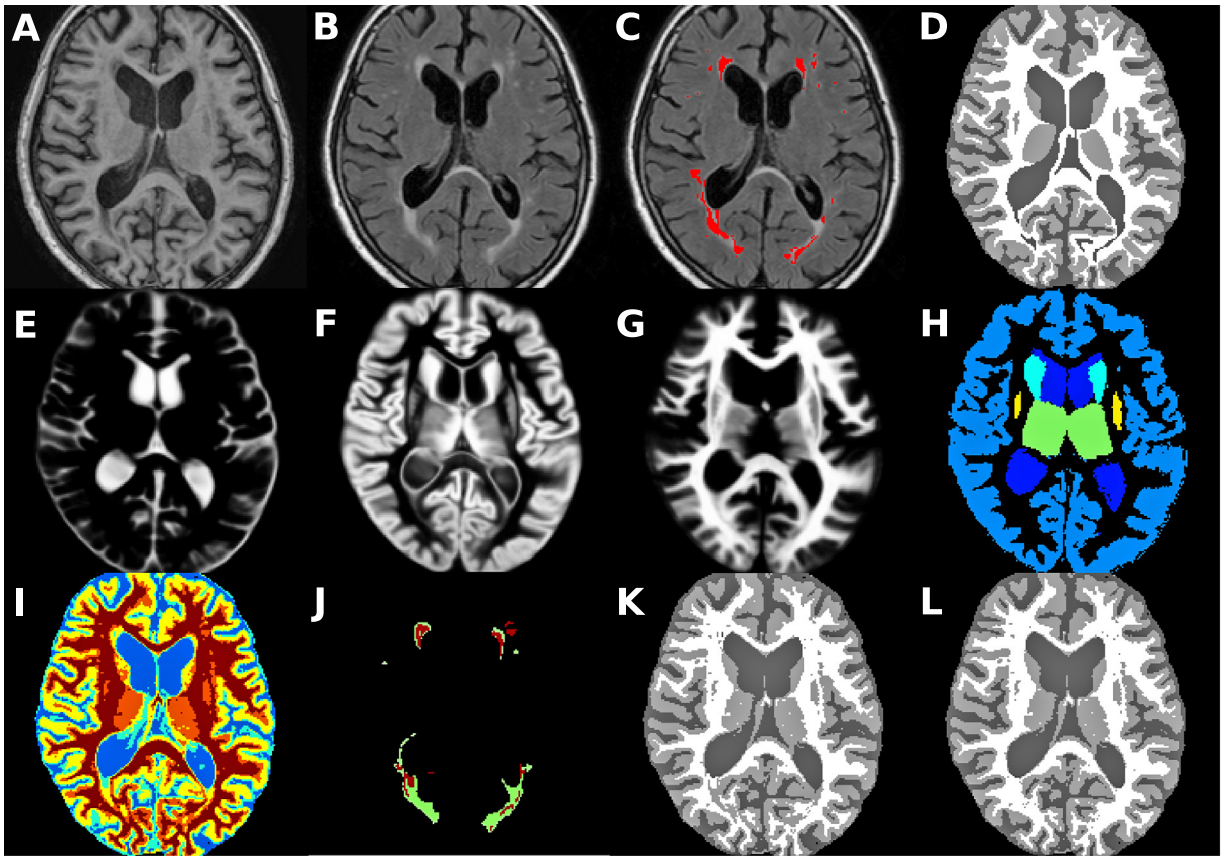

Fig. 5. Automated tissue segmentation of the *MSSEG* method on the *second subject* of the training set of the MRBrainS13 database. A) Original T1-w image. B) Original FLAIR image. C) FLAIR image with manual annotated WM lesions depicted in red. D) Provided ground-truth for training purposes. Registered CSF, GM and WM prior atlas to the subject space (E, F and G, respectively). H) Morphological brain structural atlas registered to the subject space. I) First partial volume segmentation with *csf* depicted in blue, *csfgm* in cyan, *gm* in yellow, *gmwm* in orange and *wm* in red. J) Obtained WM outliers extracted from either T1-w (depicted in red) and FLAIR (depicted in green). K) Final tissue segmentation using only the T1-w image, with *CSF* depicted in dark gray, *GM* in light gray and *WM* in white. L) Final tissue segmentation when using both T1 and the FLAIR images. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

boundary may be misclassified as WM, cancelling totally or partly the actual differences in WM tissue. In order to add an additional measure estimator of the actual error in tissue segmentation that will not biased by these differences, we also compared, for each tissue, the percentage of misclassified voxels PMC_c between the original SEG_c and the expert filled GT_c^{fill} masks:

$$PMC_c = \frac{|\overline{SEG_c} \cap GT_c^{fill}|}{|GT_c^{fill}|} \times 100 \tag{11}$$

In order to analyze the benefits of using FLAIR images in the proposed approach, we evaluated the performance of our method without estimating and refilling WM lesions before segmentation, and also estimating and refilling them using only a T1-w image first, and then using both T1-w and FLAIR. We then validated it with two other automated pipelines widely used in brain tissue segmentation, FAST (Zhang et al., 2001) (version FSL 5.0) and SPM (Ashburner and Friston, 2005) (version SPM12 rev 6225), using either original images or after estimating lesions using the automated approach SLS proposed by Roura et al. (2015). With images where lesions were automatically segmented, estimated lesion masks were then filled with the same SLF method (Valverde et al., 2014) before tissue segmentation. Similar to our approach, we considered the tissue segmentation masks of the experts refilled T1-w images segmented with FAST and SPM12 as the ground-truth for each method. Table 2 summarizes each of the evaluated pipelines and the corresponding process followed to segment the MS images.

3.2.3. Statistical significance

The statistical significance of the performance between methods was computed by running a series of permutation tests (Menke and Martinez, 2004; Klein et al., 2009; Diez et al., 2014) between the differences in the scores obtained by each method. Permutation tests select random subsets of independent subjects of the dataset, and for each of possible pair of methods, perform all possible permutations of their values in the corresponding subset, counting the number of times that the differences of one method are significant with respect to the other with ($p \leq 0.05$). After repeating this process over a number of iterations S, the mean and standard deviation (μ, σ) of the fraction of times when each method produced significant p-values is calculated over all the iterations. With this approach, methods with higher means indicate a higher significance of their reported values. The methods were then ranked into three different levels according to the difference between the mean score of the best method $\mu_o \pm \sigma_o$ and the distance with respect to the mean scores of the rest of the methods. Hence, Rank 1 contained methods with mean scores of $(\mu_o - \sigma_o, \mu_o]$, Rank 2 contained those with mean scores of $(\mu_o - 2\sigma_o, \mu_o - \sigma_o]$ and Rank 3 those in the interval $(\mu_o - 3\sigma_o, \mu_o - 2\sigma_o]$ (Klein et al., 2009; Diez et al., 2014; Valverde et al., 2015a). For all the tests, we set the number of comparisons between each pair of methods to $S = 1000$.

3.2.4. Experiment details

The BET skull-stripping process was optimized as proposed by Popescu et al. (2012) without removing CSF from the brainmask.

Table 2
Summary of evaluated pipelines and processes used on the MS data. On the *FAST only T1, SPM12 only T1* pipelines, images were segmented containing lesions without any prior automated lesion segmentation. On the *FAST + SLS* and *SPM12 + SLS* pipelines, WM lesions were automatically segmented using the SLS approach (Roura et al., 2015) and estimated lesion masks were afterwards filled using the SLF method (Valverde et al., 2014). On our proposed pipeline, when using *MSSEG no filling* images were segmented with MSSEG without estimating and refilling WM lesions before segmentation, while when using *MSSEG only T1* and *MSSEG T1 + FLAIR*, lesion segmentation and filling was part of the same segmentation method. Manual lesion annotations were used to refill T1–w images on *FAST GT, SPM12 GT* and *MSSEG GT* pipelines before segmenting the images using *FAST, SPM12* and *MSSEG*, respectively. AVD_c and PMC_c scores were then computed between pipelines 1 vs 3, 2 vs 3, 4 vs 6, 5 vs 6, 7 vs 10, 8 vs 10, and 9 vs 10.

Pipeline	Modality	Lesion seg.	Lesion filling	Tissue seg.
1. FAST only T1	T1	*none*	*none*	FAST
2. FAST + SLS	T1, FLAIR	SLS (FLAIR)	SLF	FAST
3. FAST GT	T1	*expert manual*	SLF	FAST
4. SPM12 only T1	T1	*none*	*none*	SPM12
5. SPM12 + SLS	T1, FLAIR	SLS (FLAIR)	SLF	SPM12
6. SPM12 GT	T1	*expert manual*	SLF	SPM12
7. MSSEG no filling	T1	*none*	*none*	MSSEG
8. MSSEG only T1	T1	*internal*	*internal*	MSSEG
9. MSSEG T1 + FLAIR	T1, FLAIR	*internal*	*internal*	MSSEG
10. MSSEG GT	T1	*expert manual*	SLF	MSSEG

Table 3
Mean % of absolute difference in the CSF, GM and WM volume between the 24 3T tissue masks where expert annotations were refilled before segmentation and the same images segmented including white matter lesions. For each method, the reported values are the mean and standard deviation $\mu \pm \sigma$ for the (a) AVD_c and (b) PMC_c scores obtained along the entire database.

(a) Differences in AVD_c

Method	Dif CSF (%)	Dif GM (%)	Dif WM (%)
FAST only T1	0.07 ± 0.13	0.33 ± 0.45	0.42 ± 0.56
FAST + SLS	0.04 ± 0.07	0.08 ± 0.12	0.11 ± 0.16
SPM12 only T1	0.31 ± 0.46	0.27 ± 0.45	0.56 ± 0.69
SPM12 + SLS	0.22 ± 0.22	0.13 ± 0.23	0.20 ± 0.32
MSSEG no filling	0.19 ± 0.31	0.36 ± 0.45	0.68 ± 0.87
MSSEG only T1	0.13 ± 0.20	0.21 ± 0.26	0.42 ± 0.54
MSSEG T1+FLAIR	0.04 ± 0.05	0.06 ± 0.05	0.13 ± 0.11

(b) Differences in PMC_c

Method	CSF (%)	GM (%)	WM (%)
FAST only T1	0.08 ± 0.11	0.09 ± 0.12	0.53 ± 0.69
FAST + SLS	0.06 ± 0.06	0.14 ± 0.16	0.25 ± 0.30
SPM12 only T1	0.16 ± 0.32	0.25 ± 0.33	0.73 ± 0.86
SPM12 + SLS	0.22 ± 0.31	0.24 ± 0.29	0.41 ± 0.43
MSSEG no filling	0.02 ± 0.02	0.04 ± 0.01	0.70 ± 0.89
MSSEG only T1	0.02 ± 0.03	0.03 ± 0.05	0.46 ± 0.58
MSSEG T1+FLAIR	0.04 ± 0.04	0.14 ± 0.13	0.27 ± 0.29

N3 was run with optimized parameters by reducing the smoothing distance parameter to 30–50 mm (Boyes et al., 2008; Zheng et al., 2009).

The SLF lesion filling method was run with default parameters in all experiments. In the FAST and SPM12 images, where we estimated lesion masks automatically, the lesion segmentation method SLS was optimized for 3.0T data identically as shown in Roura et al. (2015).

All the parameters of our proposed method were fixed to default values ($q = 2, \gamma = 0.1, w = 1$) as done in the MRBrainS13 database. The β parameter was computed automatically. The α parameter that scaled the minimum signal intensity on the H^F mask was set again to $\alpha = 3$.

3.2.5. Results

Table 3(a) depicts the mean % of absolute differences in the CSF, GM and WM volume (AVD_c) for each of the methods evaluated. Table 4(a) shows the final ranking in the significance permutation tests for obtained AVD_c values. In general, automated lesion segmentation and filling reduced the impact of WM lesions in tissue segmentation, and the differences in tissue volume were significantly lower when the methods used the FLAIR image to estimate the WM outliers. Both the FAST+SLS and MSSEG T1+FLAIR were first ranked in all tissues, with differences in tissue volume below 0.15% with respect images where expert manual annotated lesions were filled before segmentation. In contrast, absolute differences in GM and WM volume were significantly higher in all the pipelines that did not incorporate automated lesion segmentation and filling, and these pipelines were ranked in the second and third groups.

Table 3(b) depicts the mean % of misclassified CSF, GM and WM voxels (PMC_c) for each method. Table 4 shows the final ranking of each of the evaluated methods based on their % of misclassified tissue. In general, the % of error in the pipelines that did not incorporate the FLAIR sequence to estimate WM lesions was concentrated in WM, as the performance of these pipelines was significantly lower (Rank 2 and 3) when compared with MSSEG T1+FLAIR and FAST+SLS, that yielded again significantly lower % of misclassified WM voxels when compared with the rest of pipelines.

3.2.6. WM outlier rejection

Finally, we evaluated the performance of the proposed WM outlier rejection algorithm with respect to the other pipelines by comparing the percentages of lesion detection ratios and lesion segmentation. At the baseline, the state-of-the-art method SLS proposed by Roura et al. (2015) detected the 36% of true lesions (percentage of true positive incomes). This percentage decreased to 33% when the *MSSEG only T1* was used. In contrast, the number of true lesions detected increased up to 41% when the *MSSEG+FLAIR* was used. The number of regions detected that were not true lesions (percentage of false positive incomes) was lower in the former method, as it was optimized to reduce the number of positive incomes (Roura et al., 2015).

Furthermore, we evaluated the performance of each method segmenting WM lesions as WM. Fig. 6 shows the % of absolute difference in WM lesion volumes for each of the evaluated pipelines and their correspondent GT_{fill_c} images.

4. Discussion

In this paper we have presented a new, automated brain tissue segmentation pipeline for images containing lesions. The proposed approach combines multi-channel intensities, anatomical

Table 4
Permutation tests results for evaluated methods on the 3T MS database. (a) Final rank based on the absolute % difference in CSF, GM and WM volume between methods. (b) Final rank based on % of misclassified CSF, GM and WM voxels between methods. Reported values are mean and standard deviation (μ_o, σ_o) of the fraction of times when each method produces significant p-values ($p \leq 0.05$). Positive values indicate that in average, the method out-performed the other methods in pair-wise significant tests. Negative values indicate the contrary. Rank 1: $(\mu_o - \sigma_o,\ \mu_o]$, Rank 2: $(\mu_o - 2\sigma_o,\ \mu_o - \sigma]$, Rank 3 $(\mu_o - 3\sigma_o,\ \mu_o - 2\sigma_o]$. All permutation tests were run with 1000 random iterations.

(a) Evaluated methods ranked by the absolute % of CSF, GM and WM volume of 3T data.

Rank	Method (CSF)	$\mu \pm \sigma$	Method (GM)	$\mu \pm \sigma$	Method (WM)	$\mu \pm \sigma$
Rank 1	**MSSEG T1+FLAIR**	0.57 ± 0.53	**MSSEG T1+FLAIR**	0.57 ± 0.53	FAST + SLS	0.71 ± 0.49
	FAST only T1	0.57 ± 0.53	FAST + SLS	0.57 ± 0.53	**MSSEG T1+FLAIR**	0.57 ± 0.53
	FAST + SLS	0.57 ± 0.53	SPM12 + SLS	0.43 ± 0.52	SPM12 + SLS	0.25 ± 0.72
Rank 2	**MSSEG only T1**	−0.24 ± 0.79	**MSSEG only T1**	0 ± 0.82	FAST only T1	−0.07 ± 0.73
	SPM12 + SLS	−0.43 ± 0.53	SPM12 only T1	−0.43 ± 0.53	**MSSEG only T1**	−0.2 5± 0.72
	SPM12 only T1	−0.48 ± 0.5				
Rank 3	**MSSEG no filling**	−0.57 ± 0.53	FAST only T1	−0.57 ± 0.53	SPM12 only T1	−0.5 ± 0.51
			MSSEG no filling	−0.57 ± 0.53	**MSSEG no filling**	−0.71 ± 0.49

(b) Evaluated methods ranked by the absolute % of misclassified CSF, GM and WM of 3T data.

Rank	Method (CSF)	$\mu \pm \sigma$	Method (GM)	$\mu \pm \sigma$	Method (WM)	$\mu \pm \sigma$
Rank 1	**MSSEG only T1**	0.86 ± 0.38	**MSSEG only T1**	0.86 ± 0.38	**MSSEG T1+FLAIR**	0.71 ± 0.49
	MSSEG no filling	0.57 ± 0.78	**MSSEG no filling**	0.57 ± 0.79	FAST + SLS	0.71 ± 0.49
Rank 2	**MSSEG T1 + FLAIR**	0.24 ± 0.8	FAST only T1	0.29 ± 0.95	SPM12 + SLS	9 ± 0.82
					MSSEG only T1	0 ± 0.82
					FAST only T1	0 ± 0.82
Rank 3	FAST + SLS	−0.14 ± 0.89	FAST + SLS	−0.29 ± 0.76	**MSSEG no filling**	−0.69 ± 0.54
	SPM12 only T1	−0.38 ± 0.49	**MSSEG T1+FLAIR**	−0.29 ± 0.75	SPM12 only T1	−0.74 ± 0.44
	FAST only T1	−0.43 ± 0.79	SPM12 only T1	−0.43 ± 0.53		
	SPM12 + SLS	−0.71 ± 0.49	SPM12 + SLS	−0.71 ± 0.49		

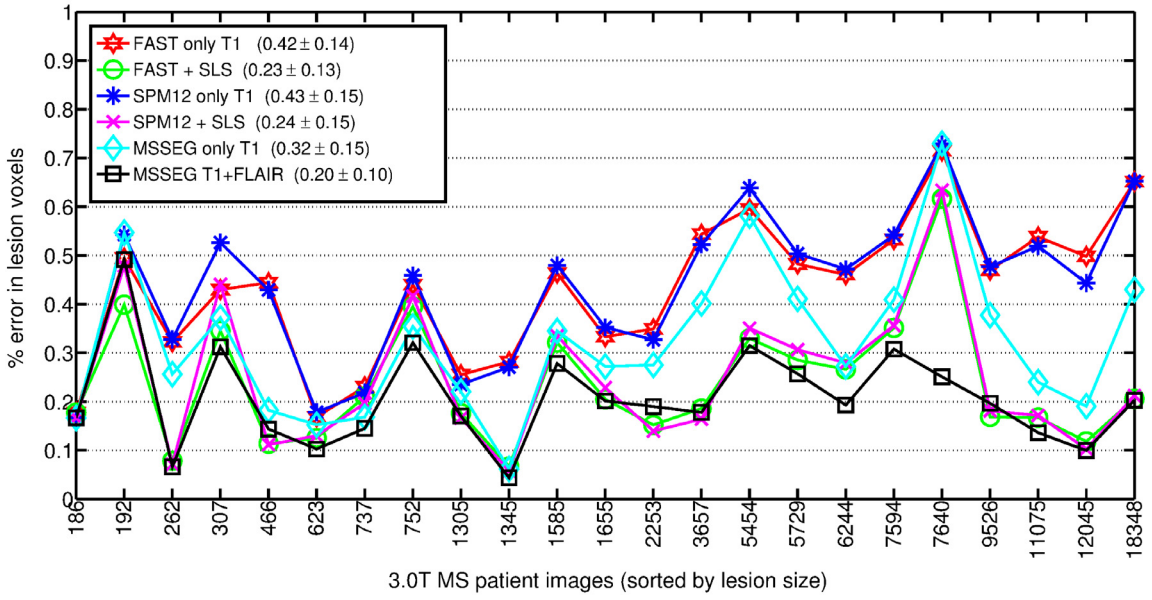

Fig. 6. % of absolute difference in WM lesion volume for each of the pipelines evaluated with the 3T MS database. Figure legends also show the mean and standard deviation $\mu \pm \sigma$ for the entire set of images. Images are sorted by lesion size (number of lesion voxels).

and morphological prior maps at different levels to estimate brain tissue in the presence of WM lesions. The current method integrates a WM outlier estimation and refilling algorithm that is applied intermediately in order to reduce the effect of WM lesions on tissue segmentation. As shown by the results, the proposed technique yields competitive and consistent results in both general and MS specific databases without parameter tweaking. Furthermore, although we did not explicitly analyze the execution times of each of the evaluated algorithms, the proposed method takes advantage of new affordable processors such as GPUs. These processors reduce up to four times the execution time to register and segment tissue when compared with general purpose CPUs.

Although the method has been designed to deal with images containing MS lesions, the performance of the method was also competitive in other images containing vascular WM lesions as those of the MRBrainS13 dataset, where lesions resemble those of the MS. This permitted us to evaluate the efficacy of our method and to validate it with other state-of-the-art tissue segmentation methods. All the challenge's participant methods were evaluated by comparing their obtained tissue masks with respect to manual expert annotations of tissues and WM lesions, which provided

a quantitative measure of the performance of the methods. The overall results showed that supervised methods obtained the best results in the challenge, taking advantage of the inherent capability to fit the database characteristics. At the time of writing this paper, *MSSEG T1+FLAIR* was ranked in 7th position out of 31 participants, being the best non-supervised strategy followed by the VBM12 approach. As shown by the differences in each of the scores obtained, the FLAIR modality appeared useful to improve the performance of the method when compared with the *MSSEG T1*, which was ranked 10th. The performance of the *MSSEG* was superior with all tissues when compared to general purpose methods such as *FAST* (ranked 21th) and *SPM12* (best ranked 17th), even if they used both image modalities. However, the final method ranking should be taken with care, given the differences in the skull-stripping processes between methods. Differences in the boundaries of the estimated skull masks may be behind the remarkable differences in CSF among methods, altering also the intra-cranial cavity measurements and consequently the overall score of each of the methods.

In MS data, the performance of our method was similar or better to the best pipeline incorporating a state-of-the-art method for lesion segmentation and filling, validating its overall capability to reduce the effects of WM lesions on tissue segmentation. The *MSSEG T1+FLAIR* and *FAST+SLS* were ranked in the first group of methods with error differences in tissue volume below 0.15% in all the tissues. All the pipelines using only T1-w without lesion correction and filling (*FAST/SPM only T1* and*MSSEG no filling*) showed a similar or lower % of misclassified CSF and GM voxels than those using both FLAIR and T1-w. In contrast, the % of misclassified WM voxels (Table 4(a)) and the differences in the reassigned lesion volume (Fig. 6) were significantly higher on the former, showing that these methods tended to overestimate GM and underestimate WM caused by the effect of WM lesions. In this aspect, our results are consistent with previous studies that also analyzed the effects of WM lesions on tissue segmentation (Battaglini et al., 2012; Gelineau-Morel et al., 2012; Valverde et al., 2015b,c).

Differences in the AVD_c between the *MSSEG T1+FLAIR* and the *MSSEG only T1* on the MRBrains13 data were similar to those reported in MS data, showing that, in general, the inclusion of the FLAIR modality reduced the overall error in tissue volume on all the analyzed databases. On MS data, the % of misclassified CSF and GM voxels was significantly lower on the *MSSEG only T1*, but significantly higher in WM, evidencing that *MSSEG only T1* tended to overestimate WM, while the error in the *MSSEG T1+FLAIR* was similar in both GM and WM. In addition, the results show that the % difference in the total WM and lesion volume was significantly lower on the *MSSEG T1+FLAIR* in comparison with the *MSSEG only T1*. Hence, we would recommend using both the T1-w and FLAIR modalities when possible. However, the accuracy of the *MSSEG only T1* pipeline was still superior to the *FAST* and *SPM12* when compared with the ground-truth annotations of the MRBrainS13 database. This suggests that at least with the available data, the improvement in tissue segmentation was not only caused by the addition of the FLAIR modality, but also by the combination of intensity and the anatomical and morphological priors.

This study, however, has some limitations. The lack of a database consisting of MS images with manual annotations on the tissue, limits our analysis to the differences in the tissue volume with respect to images where expert lesion annotations were lesion filled before tissue segmentation. However, the previous analysis in prior studies proved to be effective in evaluating the effects of WM lesions on tissue segmentation (Battaglini et al., 2012; Valverde et al., 2015b,c). Furthermore, the mean lesion sizes of the MS cohorts do not allow us to investigate better the performance of the proposed method in the presence of images with higher lesion loads. As a future work, we believe that an additional study on MS with manual annotated tissue masks and higher lesion loads would be helpful not only to analyze the benefits of the proposed algorithm with MS images, but also to investigate the benefits of adding other image channels such as T2 or PD. Furthermore, although the method was designed for cross-sectional data, we are sensible to the fact that the current approach could be benefited by the possibility of evaluating longitudinal changes in the tissue volume.

5. Conclusion

In this paper, we have proposed the Multiple Sclerosis SEGmentation pipeline (*MSSEG*), a new MRI brain tissue segmentation method designed to deal with images containing lesions. Our proposed approach incorporates a robust partial volume tissue segmentation with outlier rejection and filling, combining intensity and probabilistic and morphological prior maps in a novel-way. When combining T1-w and FLAIR modalities, our method has shown very competitive results with the MRBrainS13 database, ranked at the time of submission in 7th position out of 31 participant strategies and being the best non-supervised intensity model approach so far. With MS data, differences in the tissue volume were lower or similar to the best available pipeline composed of the FAST and a state-of-the-art method for lesion segmentation and filling. In all the experiments, the inclusion of the FLAIR modality into the proposed method reduced the effect of WM lesions on the tissue segmentation, which suggests that this modality should be used when available. In conclusion, our results show that, at least with the presented data, the *MSSEG* improves the measurement of brain tissue volume in images containing WM lesions. The proposed method is currently available to download at the authors' webpage (http://atc.udg.edu/nic/msseg). We strongly believe that the neuro-image community can benefit from its use in future settings.

Acknowledgments

S. Valverde holds a FI-DGR2013 grant from the Generalitat de Catalunya. E. Roura holds a BR-UdG2013 grant from the University of Girona. This work has been partially supported by "La Fundació la Marató de TV3", by Retos de Investigacin TIN2014-55710-R, and by the MPC UdG 2016/022 grant from the University of Girona. The authors gratefully acknowledge the support of the NVIDIA Corporation with their donation of the Tesla K40 GPU used in this research.

References

Ashburner, J., Friston, K.J., 2005. Unified segmentation. NeuroImage 26, 839–851.
Bach-Cuadra, M., Cammoun, L., Butz, T., Cuisenaire, O., Thiran, J.P., 2005. Comparison and validation of tissue modelization and statistical classification methods in t1-weighted mr brain images. IEEE Trans. Med. Imaging 24 (12), 1548–1565.
Battaglini, M., Jenkinson, M., Stefano, N.D., De Stefano, N., 2012. Evaluating and reducing the impact of white matter lesions on brain volume measurements. Hum. Brain Mapp. 33 (9), 2062–2071.
Boyes, R.G., Gunter, J.L., Frost, C., Janke, A.L., Yeatman, T., Hill, D.L.G., Bernstein, M.A., Thompson, P.M., Weiner, M.W., Schuff, N., Alexander, G.E., Killiany, R.J., DeCarli, C., Jack, C.R., Fox, N.C., 2008. Intensity non-uniformity correction using N3 on 3-T scanners with multichannel phased array coils. NeuroImage 39, 1752–1762.
de Bresser, J., Portegies, M.P., Leemans, A., Biessels, G.J., Kappelle, L.J., Viergever, M.A., Bresser, J.D., Jan, G., 2011. A comparison of MR based segmentation methods for measuring brain atrophy progression. NeuroImage 54 (2), 760–768.
Cabezas, M., Oliver, A., Lladó, X., Freixenet, J., Bach Cuadra, M., 2011. A review of atlas-based segmentation for magnetic resonance brain images. Comput. Methods Programs Biomed. 104 (3), e158–e177.
Ceccarelli, A., Jackson, J.S., Tauhid, S., Arora, A., Gorky, J., Dell'Oglio, E., Bakshi, A., Chitnis, T., Khoury, S.J., Weiner, H.L., Guttmann, C.R.G., Bakshi, R., Neema, M., 2012. The impact of lesion in-painting and registration methods on voxel-based morphometry in detecting regional cerebral gray matter atrophy in multiple sclerosis. Am. J. Neuroradiol. 33 (8), 1579–1585.

Chard, D.T., Jackson, J.S., Miller, D.H., Wheeler-Kingshott, C.A.M., 2010. Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. J. Magn. Reson. Imaging 32, 223–228.
Datta, S., Narayana, P.A., 2013. A comprehensive approach to the segmentation of multichannel three-dimensional MR brain images in multiple sclerosis. NeuroImage: Clinical 2, 184–196.
De Boer, R., Vrooman, H.A., van der Lijn, F., Vernooij, M.W., Ikram, M.A., van der Lugt, A., Breteler, M.M.B., Niessen, W.J., 2009. White matter lesion extension to automatic brain tissue segmentation on MRI. NeuroImage 45 (4), 1151–1161.
DellOglio, E., Ceccarelli, A., Glanz, B.I., Healy, B.C., Tauhid, S., Arora, A., Saravanan, N., Bruha, M.J., Vartanian, A.V., Dupuy, S.L., Benedict, R.H., Bakshi, R., Neema, M., 2014. Quantification of global cerebral atrophy in multiple sclerosis from 3T MRI using SPM: the role of misclassification errors. J. Neuroimaging 25(2), 191–199.
Deshpande, H., Maurel, P., Barillot, C., 2015. Classification of multiple sclerosis lesions using adaptive dictionary learning. Comput. Med. Imaging Graph 46, 2–10.
Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology 26 (3), 297–302.
Diez, Y., Oliver, A., Cabezas, M., Valverde, S., Martí, R., Vilanova, J.C., Ramió-Torrentà, L., Rovira, À., Lladó, X., 2014. Intensity based methods for brain MRI longitudinal registration. a study on multiple sclerosis patients. Neuroinformatics 12 (3), 365–379.
García-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D.L., Collins, D.L., 2013. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. Med. Image Anal. 17 (1), 1–18.
Gelineau-Morel, R., Tomassini, V., Jenkinson, M., Johansen-berg, H., Matthews, P.M., Palace, J., 2012. The effect of hypointense white matter lesions on automated gray matter segmentation in multiple sclerosis. Hum. Brain Mapp. 2814 (June 2011), 2802–2814.
Geremia, E., Clatz, O., Menze, B.H., Konukoglu, E., Criminisi, A., Ayache, N., 2011. Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. NeuroImage 57 (2), 378–390.
Giorgio, A., De Stefano, N., 2013. Clinical use of brain volumetry. J. Magn. Reson. Imaging 37 (1), 1–14.
Guizard, N., Coupé, P., Fonov, V.S., Manjón, J.V., Arnold, D.L., Collins, D.L., 2015. Rotation-invariant multi-contrast non-local means for MS lesion segmentation. NeuroImage 8, 376–389.
Harmouche, R., Subbanna, N.K., Collins, D.L., Arnold, D.L., Arbel, T., 2015. Based on modeling regional intensity variability and local neighborhood information. IEEE Trans. Biomed. Eng. 62 (5), 1281–1292.
Huttenlocher, D., Klanderman, G., Rucklidge, W., 1993. Comparing images using the hausdorff distance. IEEE Trans. Pattern Anal. Mach. Intell. 15 (9), 850–863.
Immerkær, J., 1996. Fast noise variance estimation. Comput. Vision Image Understanding 64 (2), 300–302.
Jain, S., Sima, D.M., Ribbens, A., Cambron, M., Maertens, A., Van Hecke, W., De Mey, J., Barkhof, F., Steenwijk, M.D., Daams, M., Maes, F., Van Huffel, S., Vrenken, H., Smeets, D., 2015. Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images. NeuroImage: Clinical 8, 367–375.
Jog, A., Roy, S., Prince, J., Carass, A., 2013. Mr brain segmentation using decision trees. In: Proceedings of the MICCAI Workshops, Challenge on MR Brain Image Segmentation (MRBrainS 13), The Midas Journal.
Klauschen, F., Goldman, A., Barra, V., Meyer-Lindenberg, A., Lundervold, A., 2009. Evaluation of automated brain MR image segmentation and volumetry methods. Hum. brain Mapp. 30 (4), 1310–1327.
Klein, A., Ardekani, B., Andersson, J., Ashburner, J., Avants, B., Chiang, M., Christensen, G., Collins, D., Gee, J., Hellier, P., et al., 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage 46 (3), 786–802.
Klein, S., Staring, M., Murphy, K., Viergever, M.a., Pluim, J.P.W., 2010. Elastix: a toolbox for intensity-based medical image registration. IEEE Trans. Med. Imaging 29 (1), 196–205.
Lladó, X., Oliver, A., Cabezas, M., Freixenet, J., Vilanova, J.C., Quiles, A., Valls, L., Ramió-Torrentà, L., Rovira, A., 2012. Segmentation of multiple sclerosis lesions in brain MRI: a review of automated approaches. Inf. Sci. 186 (1), 164–185.
Magon, S., Gaetano, L., Chakravarty, M.M., Lerch, J.P., Naegelin, Y., Stippich, C., Kappos, L., Radue, E.-W., Sprenger, T., 2014. White matter lesion filling improves the accuracy of cortical thickness measurements in multiple sclerosis patients: a longitudinal study. BMC Neurosci. 15 (1), 106.
Mendrik, A.M., Vincken, K.L., Kuijf, H.J., Breeuwer, M., Bouvy, W.H., De Bresser, J., Alansary, A., De Bruijne, M., Carass, A., El-Baz, A., Jog, A., Katyal, R., Khan, A.R., Van Der Lijn, F., Mahmood, Q., Mukherjee, R., Van Opbroek, A., Paneri, S., Pereira, S., Persson, M., Rajchl, M., Sarikaya, D., Smedby, Ö., Silva, C.A., Vrooman, H.A., Vyas, S., Wang, C., Zhao, L., Biessels, G.J., Viergever, M.A., 2015. MRBrainS challenge: online evaluation framework for brain image segmentation in 3T MRI scans. Comput. Intell. Neurosci. 2015.
Menke, J., Martinez, T., 2004. Using permutations instead of student's t distribution for p-values in paired-difference algorithm comparisons. In: Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on, vol. 2, pp. 1331–1335 vol.2.
Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S., 2010. Fast free-form deformation using graphics processing units. Comput. Methods Programs Biomed. 98 (3), 278–284.
Nakamura, K., Fisher, E., 2009. Segmentation of brain magnetic resonance images for measurement of gray matter atrophy in multiple sclerosis patients. NeuroImage 44 (3), 769–776.
Opbroek, A.V., der Lijn, F.V., Bruijne, M.D., 2013. Automated brain-tissue segmentation by multi-feature svm classification. In: Proceedings of the MICCAI Workshops, Challenge on MR Brain Image Segmentation (MRBrainS 13), The Midas Journal.
Ourselin, S., Stefanescu, R., Pennec, X., 2002. Robust registration of multi-modal images: Towards real-time clinical applications. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI'02. Springer, pp. 140–147.
Pérez-Miralles, F., Sastre-Garriga, J., Tintoré, M., Arrambide, G., Nos, C., Perkal, H., Río, J., Edo, M.C., Horga, A., Castilló, J., Auger, C., Huerga, E., Rovira, A., Montalban, X., 2013. Clinical impact of early brain atrophy in clinically isolated syndromes. Mult. Scler. 19 (14), 1878–1886.
Pham, D.L., 2001. Spatial models for fuzzy clustering. Comput. Vision Image Understanding 297, 285–297.
Pohl, K.M., Bouix, S., Nakamura, M., Rohlfing, T., McCarley, R.W., Kikinis, R., Grimson, W.E.L., Shenton, M.E., Wells, W.M., 2007. A hierarchical algorithm for MR brain image parcellation. IEEE Trans. Med. Imaging 26 (9), 1201–1212.
Popescu, V., Battaglini, M., Hoogstrate, W.S., Verfaillie, S.C.J., Sluimer, I.C., van Schijndel, R.A., van Dijk, B.W., Cover, K.S., Knol, D.L., Jenkinson, M., Barkhof, F., de Stefano, N., Vrenken, H., Barkhof, F., Montalban, X., Fazekas, F., Filippi, M., Frederiksen, J., Kappos, L., Miller, D., Palace, J., Polman, C., Rocca, M., Rovira, A., Yousry, T., 2012. Optimizing parameter choice for FSL-Brain Extraction Tool (BET) on 3D T1 images in multiple sclerosis. NeuroImage 61, 1484–1494.
Popescu, V., Ran, N.C.G., Barkhof, F., Chard, D.T., Wheeler-Kingshott, C., Vrenken, H., 2014. Accurate GM atrophy quantification in MS using lesion-filling with co-registered 2D lesion masks. NeuroImage: Clinical 4, 366–373.
Rajchl, M., Baxter, J.S., McLeod, A.J., Yuan, J., Qiu, W., Peters, T.M., Khan, A.R., 2015. Hierarchical max-flow segmentation framework for multi-atlas segmentation with Kohonen self-organizing map based Gaussian mixture modeling. Med. Image Anal. 000, 1–12.
Roura, E., Oliver, A., Cabezas, M., Valverde, S., Pareto, D., Vilanova, J.C., Ramió-Torrentà, L., Rovira, À., Lladó, X., 2015. A toolbox for multiple sclerosis lesion segmentation. Neuroradiology 57 (10), 1031–1043.
Roy, S., He, Q., Sweeney, E., Carass, A., Reich, D.S., Prince, J.L., Pham, D.L., 2015. Subject-specific sparse dictionary learning for atlas-based brain MRI segmentation.. IEEE J. Biomed. Heal. Inf. 19 (5), 1598–1609.
Schmidt, P., Gaser, C., Arsic, M., Buck, D., Förschler, A., Berthele, A., Hoshi, M., llg, R., Schmid, V.J., Zimmer, C., Hemmer, B., Mühlau, M., 2012. An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis. NeuroImage 59 (4), 3774–3783.
Shattuck, D.W., Sandor-Leahy, S.R., Schaper, K.A., Rottenberg, D.A., Leahy, R.M., 2001. Magnetic resonance image tissue classification using a partial volume model. Neuroimage 13 (5), 856–876.
Shiee, N., Bazin, P.-L., Ozturk, A., Reich, D.S., Calabresi, P.A., Pham, D.L., 2010. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions.. NeuroImage 49 (2), 1524–1535.
Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17 (1), 87–97.
Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17, 143–155.
Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P., Federico, A., De Stefano, N., 2002. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. NeuroImage 17 (1), 479–489.
Sormani, M.P., Arnold, D.L., De Stefano, N., 2014. Treatment effect on brain atrophy correlates with treatment effect on disability in multiple sclerosis.. Ann. neurology 75 (1), 43–49.
Stollenga, M., Byeon, W., 2015. [M] Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation. In: Neural Information Processing Systems Conference.
Sweeney, E.M., Shinohara, R.T., Shiee, N., Mateen, F.J., Chudgar, A.A., Cuzzocreo, J.L., Calabresi, P.A., Pham, D.L., Reich, D.S., Crainiceanu, C.M., 2013. OASIS is automated statistical inference for segmentation, with applications to multiple sclerosis lesion segmentation in MRI.. NeuroImage: Clinical 2, 402–413.
Tomas-Fernandez, X., Warfield, S.K., 2015. A model of population and subject (MOPS) intensities with application to multiple sclerosis lesion segmentation. IEEE transactions on medical imaging 0062 (c), 1–15.
Valverde, S., Oliver, A., Cabezas, M., Roura, E., Lladó, X., 2015. Comparison of 10 brain tissue segmentation methods using revisited IBSR annotations. J. Magn. Reson. Imaging 41 (1), 93–101.
Valverde, S., Oliver, A., Dez, Y., Cabezas, M., Vilanova, J., Rami-Torrentà, L., Rovira, A., Lladó, X., 2015. Evaluating the effects of white matter multiple sclerosis lesions on the volume estimation of 6 brain tissue segmentation methods. Am. J. Neuroradiol. 36(6), 1109–1115.
Valverde, S., Oliver, A., Lladó, X., 2014. A white matter lesion-filling approach to improve brain tissue volume measurements. NeuroImage: Clinical 6, 86–92.
Valverde, S., Oliver, A., Roura, E., Pareto, D., Vilanova, J.C., Ramió-Torrentà, L., Sastre-Garriga, J., Montalban, X., Rovira, A., Lladó, X., 2015. Quantifying brain tissue volume in multiple sclerosis with automated lesion segmentation and filling. NeuroImage: Clinical 9, 640–647.
Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging 20, 45–57.
Zheng, W., Chee, M.W.L., Zagorodnov, V., 2009. Improvement of brain segmentation accuracy by optimizing non-uniformity correction using N3. NeuroImage 48, 73–83.