Computers in Biology and Medicine 73 (2016) 10–23

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Automatic iterative segmentation of multiple sclerosis lesions using Student's *t* mixture models and probabilistic anatomical atlases in FLAIR images

Paulo G.L. Freire *, Ricardo J. Ferrari

Departamento de Computação, Universidade Federal de São Carlos, Rod. Washington Luís, Km 235, 13565-905 São Carlos, SP, Brazil

ARTICLE INFO

Article history:
Received 4 January 2016
Received in revised form
16 March 2016
Accepted 29 March 2016

Keywords:
Image segmentation
Multiple sclerosis
Mixture models
Student's *t* distribution
Probabilistic anatomical atlases

ABSTRACT

Multiple sclerosis (MS) is a demyelinating autoimmune disease that attacks the central nervous system (CNS) and affects more than 2 million people worldwide. The segmentation of MS lesions in magnetic resonance imaging (MRI) is a very important task to assess how a patient is responding to treatment and how the disease is progressing. Computational approaches have been proposed over the years to segment MS lesions and reduce the amount of time spent on manual delineation and inter- and intra-rater variability and bias. However, fully-automatic segmentation of MS lesions still remains an open problem. In this work, we propose an iterative approach using Student's *t* mixture models and probabilistic anatomical atlases to automatically segment MS lesions in Fluid Attenuated Inversion Recovery (FLAIR) images. Our technique resembles a refinement approach by iteratively segmenting brain tissues into smaller classes until MS lesions are grouped as the most hyperintense one. To validate our technique we used 21 clinical images from the 2015 Longitudinal Multiple Sclerosis Lesion Segmentation Challenge dataset. Evaluation using Dice Similarity Coefficient (DSC), True Positive Ratio (TPR), False Positive Ratio (FPR), Volume Difference (VD) and Pearson's *r* coefficient shows that our technique has a good spatial and volumetric agreement with raters' manual delineations. Also, a comparison between our proposal and the state-of-the-art shows that our technique is comparable and, in some cases, better than some approaches, thus being a viable alternative for automatic MS lesion segmentation in MRI.

1. Introduction

Multiple sclerosis (MS) is a demyelinating autoimmune disease that attacks the central nervous system (CNS) and affects more than 2 million people worldwide [1]. It is mainly characterized by the presence of white matter (WM) lesions [2], which are visible on magnetic resonance imaging (MRI) scans and appear hyperintense on T2-weighted and Fluid Attenuated Inversion Recovery (FLAIR) images. Segmentation of MS lesions is usually done by a radiologist, who has to visually assess and manually delineate them when measuring total lesion volume. Since MRI brain scans are usually volumetric, the manual delineation procedure is done in a slice-by-slice manner, which is time consuming and suffers from large intra- and inter-rater variability and bias [3].

Clinical trials have shown that lesion volumes are meaningful outcomes for assessing disease burden in multiple sclerosis (MS) [4], and thus accurately measuring such volumes is of considerable interest in clinical practice [5]. In this scenario, an automated technique to segment MS lesions and measure their volumes would reduce the time needed from the rater and decrease the observer dependency as well. However, automatic segmentation of MS lesions is challenging, since a number of variables such as partial volume effect (PVE), bias field, acquisition parameters and different scanner magnetic field strengths may directly influence the segmentation outcome. There have been many proposals over the years regarding automatic MS lesion segmentation, but no single one appears to be widely used at the present time. The proposed techniques range from a myriad of approaches such as dictionary learning [6], logistic regression [7,8], patch-based [9], random decision forest [10] and mixture models [11,12].

In this paper, we propose a new fully-automatic technique for MS lesions' segmentation in MR FLAIR images using an iterative approach based on Student's *t* mixture models and probabilistic anatomical atlases. Our technique resembles a refinement approach by iteratively segmenting brain tissue classes into

* Corresponding author.
E-mail address: paulo.freire@dc.ufscar.br (P.G.L. Freire).
URL: http://www.bipgroup.dc.ufscar.br/ (R.J. Ferrari).

http://dx.doi.org/10.1016/j.compbiomed.2016.03.025
0010-4825/

subclasses until MS lesions are grouped as the most hyperintense one. Since our segmentation technique is intensity-based, we use probabilistic anatomical atlases to constrain the iterative process to the WM region, thus avoiding misclassification of voxels that have similar intensities to lesion voxels but are outside the WM tissue class.

To validate our technique we used 21 images from the 2015 Longitudinal Multiple Sclerosis Lesion Segmentation Challenge (also known as 2015 ISBI Longitudinal Challenge) and assessed its performance using the Dice Similarity Coefficient (DSC), True Positive Ratio (TPR), False Positive Ratio (FPR), Volume Difference (VD) and Pearson's r coefficient on our results and rater manual delineations. It is important to note that each case in the dataset had 2 lesion delineations, each one done by a different rater, and we compared our results to both of them. Also, in some cases the manual delineations did not have a good agreement rate between themselves, which confirms the inter-variability and bias between observers.

We also compared our findings with proposals that took part in the 2015 ISBI Longitudinal Challenge. Since they used the same database as we did, a direct comparison was possible. We briefly describe the works that presented quantitative results in the following.

In [13], an automatic hierarchical framework for the segmentation of healthy tissues and lesions in brain MRI was proposed. The authors used a Markov Random Field segmentation framework that leveraged spatial prior probabilities for 9 healthy tissues through multi-atlas fusion and then used a random forest classifier to provide region level lesion refinement.

In [14], lesions were segmented using a fast patch matching approach, which was extended to multimodal data. To do so, the authors registered all available modalities to a common space and stacked them to form a 4D volume of multimodal intensities. Patches were defined and used to segment MS lesions.

An approach using random forest and local context intensity features was proposed in [15] to segment MS lesions. The authors extracted features from the images such as voxel intensity values (before and after image smoothing) and local histogram features and trained a random forest with supervised learning to segment MS lesions.

In [16], the authors proposed a 3D convolutional neural network (CNN) using a voxel-wise classifier with multi-channel 3D patches of MRI volumes as input. For each ground truth, a CNN was trained and the final segmentation was obtained by combining the probability outputs of these CNNs.

Finally, in [17] the authors used an estimation of spatially global within-the-subject intensity distribution and a spatially local intensity distribution derived from a healthy reference population to segment MS lesions. Using this approach, the authors aimed to distinguish locations in the brain with abnormal intensity levels when compared to the expected value at the same location in a healthy population.

All the works previously described are intensity-based methods and tried to use spatial information or patches to improve the segmentation results. The DSC, TPR and FPR values for these works are summarized in Table 10. These three metrics were the ones used by all authors, allowing a direct comparison of our approach to theirs. Along with DSC, TPR and FPR, we also used VD and Pearson's r coefficient metrics when comparing our proposal results to the raters' ground truths.

This paper is divided in the following manner. In Section 2 we explain the methodology of our work, including details about the dataset, metrics used to evaluate our model, preprocessing, our segmentation technique and post-processing stages. Results and discussions are presented in Section 3, where we compare results from our technique to each rater delineation and other proposals and also analyze how well the manual delineations for each image agree with each other. Finally, Section 4 concludes our paper.

2. Methodology

This section provides information about the dataset used in this work, along with the description of the metrics used to evaluate our results. It also presents information regarding our segmentation technique and pre- and post-processing steps.

2.1. Dataset

2.1.1. Clinical images

The dataset used to validate our technique consisted of longitudinal images from 5 patients obtained from the 2015 Longitudinal MS Lesion Segmentation Challenge[1] conducted at the 2015 International Symposium on Biomedical Imaging in New York, NY, April 16–19. Each longitudinal dataset included T1-, T2-, PD-weighted and FLAIR MR images with 4–5 time points acquired on a 3T MR scanner. Every longitudinal dataset had two manual lesion delineations provided by rater 1 and rater 2. Considering all longitudinal dataset cases, 21 time points were provided in total. T1-weighted images had approximately 1 mm^3 voxel resolution, while the other weighted images had a resolution of 1 mm^2 in plane with 3 mm thickness. To minimize the dependency of the results on registration performance and brain extraction, all images were already rigidly registered to the baseline T1-weighted image with automatically computed skull stripping masks. After registration, image dimensions were $181 \times 217 \times 181$ for every image.

2.1.2. Probabilistic anatomical atlases

Three probabilistic anatomical atlases, corresponding to gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF), were used to provide spatial information to our algorithm. They were obtained from the ICBM[2] project [18]. Their spatial resolution was $1 \times 1 \times 1$ mm and their dimensions were $256 \times 256 \times 256$.

2.2. Metrics

To evaluate our technique, we used the Dice Similarity Coefficient (DSC) [19], True Positive Rate (TPR), False Positive Rate (FPR) Volume Difference (VD) and Pearson's r coefficient.

The DSC is defined as the ratio between the number of voxels where both the automatic and rater reference segmentation (ground truth) agree (true positives) and the sum of the total number of voxels labeled as lesion by both methods (manual and automatic). Also, according to Bartko [20], DSC values of 0.7 or higher suggest good agreement between two delineations. The TPR, FPR and VD metrics were calculated taking into account only the lesion voxels.

The Pearson's r coefficient [21] was used to assess the volumetric correlation between our segmentation and the ground truths from the raters. It is given by

$$r = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}}, \quad (1)$$

where n is the number of time points, x_i and y_i are the absolute volumes of the ground truth and the automatic segmentation for a particular time point and $\bar{x}$ and $\bar{y}$ are their respective means. Pearson's r coefficient values lie inside the interval $[+1, -1]$.

[1] http://iacl.ece.jhu.edu/MSChallenge

[2] http://www.bic.mni.mcgill.ca/ServicesAtlases/ICBM152NLin2009