

A Supervised Approach for Multiple Sclerosis Lesion Segmentation Using Context Features and an Outlier Map

Mariano Cabezas, Arnau Oliver, Jordi Freixenet, and Xavier Lladó

Department of Computer Architecture and Technology, University of Girona. Spain
{mcabezas,aoliver,jordif,llado}@eia.udg.edu

Abstract. Automatic multiple sclerosis (MS) lesion segmentation in magnetic resonance imaging (MRI) is a challenging task due to the small size of the lesions, its heterogeneous shape and distribution, overlapping tissue intensity distributions, and the inherent artifacts of MRI. In this paper we propose a pipeline for MS lesion segmentation that combines prior knowledge and contextual information into a boosting classifier. The prior knowledge is introduced in terms of atlas distribution of the main brain tissues while the contextual information is based on a large set of features describing the spatial context in the lesion neighbourhood. Besides, we investigate the inclusion of a probability map describing the likelihood of a voxel to be an outlier, i.e. not being part of any healthy tissue. The experimental results, performed using a set of 30 MRI volumes of MS patients with very different lesion load, shows the feasibility of our approach. Besides, the results demonstrate the benefits of taking the outlier map into account.

1 Introduction

Magnetic resonance imaging (MRI) of the brain has been widely used during the last years in clinical practice. Due to its high soft contrast tissue, this image modality is the best choice on several research topics related to the brain such as disease diagnosis and patient follow-up, therapy evaluation or human brain mapping. Accurate and robust brain tissue and lesion segmentation from MRI is a key issue in many applications of medical image analysis and, particularly, in the study of multiple sclerosis (MS). Manual tracing by an expert of the three brain tissue types, white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF) as well as white matter lesions (WML) is far too time consuming as the data involved in studies is large. On the other hand, automated and reliable tissue classification is a demanding task as the intensity representation of the data does not allow a clear delimitation of the different tissue types present in a natural MRI. This is due to the partial volume (PV) effect (presence of more than one brain tissue type in a voxel), image noise and intensity non-uniformities caused by the inhomogeneities in the magnetic field of the MR scanner.

In [5] the state of the art on MS lesion segmentation was analysed, concluding that the use of a priori knowledge can improve the segmentation of brain tissues.

Furthermore, one of the most common strategies relying on prior knowledge consists on registering a template (called atlas) with tissue priors to aid the segmentation [1]. Probabilistic atlases define a soft segmentation of the brain providing a probability for each voxel to belong to a certain structure or tissue. These soft segmentations can be then used on probabilistic approaches to detect abnormal tissue types as outliers, i.e. multiple sclerosis lesions [5]. However, due to the variability in terms of shape, size and location of lesions, there has not been much works relying on probabilistic maps for abnormal tissue types. A close approximation could be to learn an outlier map from image intensities to help the MS segmentation as proposed on the work of Schmidt et al. [9].

Other common supervised approaches rely on the use of a classification algorithm. These algorithms involve a first stage in which a model is estimated on training data composed by a set of features and their corresponding ground truth, and a second stage in which the model is tested on a new dataset to provide the desired classification. Most classification algorithms differentiate between lesion and not lesion (object and background), however, other algorithms can use multiple labels for each tissue and produce a probabilistic map as a result. These probabilistic maps are usually the result of using an ensemble of classifiers such as random forests [3] or boosting [6]. These strategies combine simple discriminative classifiers that focus on a small subset of a rich feature pool to classify a training set. One of the advantages of this specialisation is the implicit feature selection applied to the dataset. These strategies give a stronger impact to the most discriminating features while underscoring or discarding the less discriminating ones. Notice that having a rich pool of features is important when using ensemble classifiers. These features can comprise image intensities, spatial locations, atlas probabilities or more complex features such as meta-features that exploit contextual information [3]. These approaches have been extensively used in other domains such as object detection [10].

In this paper, following the idea of the work of Geremia et al. [3], we propose a novel classification approach which aims to segment MS lesions on MRI. Our approach is based on an ensemble classifier of linear discriminants with a rich feature pool that combines multi-channel image intensities, probability maps, contextual meta-features and an outlier map based on image intensities. This map is an extension of the approach introduced by Schmidt et al. [9] in order to deal with multiple channels. The experimental evaluation is done with real data from two hospitals.

2 MS Lesion Segmentation Pipeline

2.1 Preprocessing

Skull Stripping. The presence of non-brain tissues in the MRI scans affects intensity distributions on the image. This is also inherent to the capture process but it is not clear how the probability density function of each main tissue (GM, WM and CSF) is altered by those external intensities. However, segmentation results are usually improved when those voxels are masked out. We apply here

the BET tool from the FSL toolbox [4], which is freely available and widely used on brain image analysis, to perform the skull stripping process.

Bias Correction and Noise Reduction. As any capture system, signals are corrupted by noise, often assumed additive and following a Gaussian distribution. Inhomogeneities on the magnetic field also cause a smooth inhomogeneity field across the MRI [12]. This artifact is also known as bias field. Even though those two issues can not be avoided on acquisition, they can be reduced in order to increase the signal to noise ratio. In our approach we first use a gradient anisotropic diffusion algorithm to reduce the image noise. Afterwards, we use the N4 algorithm [11] to correct the intensity inhomogeneities.

Registration. A registration step is needed in order to align the atlas with the new volume. When dealing with intersubject applications such as atlas matching, nonrigid algorithms can explain better local anatomical variations between the template and the subject brain. Notice this is a situation in which rigid methods fail to reproduce. In our proposal, we have implemented one of the most common registration methods based on a multi-resolution approach [8]. We start with a pre-registration using affine transformations to globally align the atlas template and the patient's scan, followed by a multi-resolution b-spline transformation.

2.2 Outlier Map

Lesions are usually seen as hyperintense regions on PD-w, T2-w images and FLAIR. However, lesions may present a wide range of grey level values on T1-w images (ranging from normal appearing WM to CSF). Although FLAIR provides a good contrast between tissues and WML, it also presents some disadvantages and artifacts that may cause misclassifications and false positive (FP) detections [7]. Moreover, lesions should also appear on PD-w and T2-w images, while most FLAIR artifacts should not be present. Therefore, we decided to extend the map presented by Schmidt et al. [9] to a multidimensional outlier map using PD-w and T2-w images to reduce the FP presence (see Fig. 1). The original approach consisted on assigning to each voxel the distance to a previously estimated tissue distribution for WM, GM and CSF (since WML have a higher intensity than all 3 tissues) when the value was positive and 0 otherwise. We propose to extend this approach to multi-channel using the Mahalanobis distance as follows:

$$O_k(x) = \begin{cases} d(x, \theta_k) & \text{if } \forall s : I^s(x) > \mu_k^s \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where x represents the voxel coordinates, $s = \{T_2, PD, FLAIR\}$, $k \in \{CSF, GM, WM\}$, $\theta_k = \{\mu_k, \Sigma_k\}$ and $d(x, \theta_k)$ is computed using the following equation:

$$d(x, \theta_k) = \sqrt{(I(x) - \mu_k) \Sigma_k^{-1} (I(x) - \mu_k)^t} \quad (2)$$

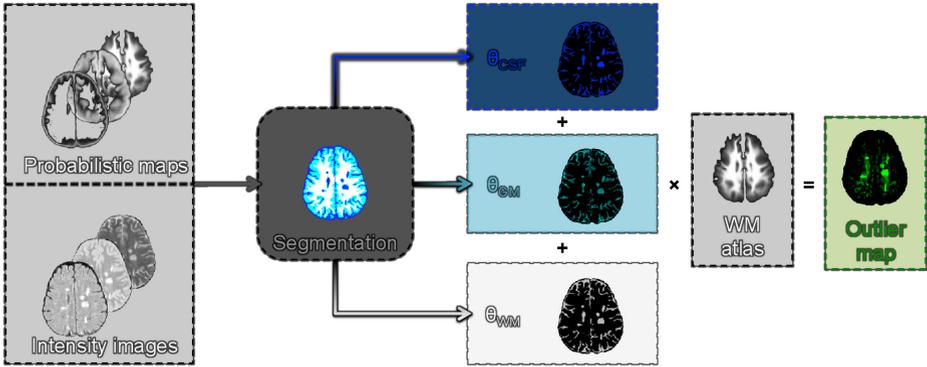


Fig. 1. Scheme of the outlier map computation. θ_k represents the mean and covariance matrix of each tissue k used to compute the Mahalanobis distance to obtain the outlier maps for that tissue.

Afterwards, we sum together the three tissue maps and weight them using the WM probabilistic atlas to reduce the effect of outliers outside the tissue. Finally, we relax this map using neighbouring information for each voxel.

2.3 Context Meta-Features

Even though the outlier map can be used to detect lesions as hyperintense outliers, due to image artifacts other regions may present similar properties. These false positive regions are usually located at unlikely places for WML (such as the cortex). Contextual information or spatial constraints can, therefore, aid automated strategies to differentiate between real lesions and artifacts.

To introduce this information we propose to use meta-features similar to the ones presented by Geremia et al. [3]. These meta-features (see Fig. 2) provide a comparison between the voxel on channel C_1 (which can be either a probabilistic map or an intensity image) to the mean of two different regions (R_1 and R_2) on channel C_2 as follows:

$$\lambda(x) = C_1(x) - \overline{C_2}(R_1) - \overline{C_2}(R_2) \quad (3)$$

where $C_1, C_2 \in \{I^D, \pi_k, O_k\}$, I^D is the image $D \in \{T_1, PD, T_2, FLAIR\}$ and π_k is the probability atlas for tissue k .

The three-dimensional regions R_1 and R_2 are randomly sampled (in size, shape and location) in a neighbouring cube around the voxel x , and the mean for them is efficiently computed using integral volume processing.

2.4 Boosting Classifier

At this point we have a rich pool of features characterised by 4 intensity channels (T1-w, PD-w, T2-w and FLAIR), 3 probabilistic atlases (CSF, GM and WM),

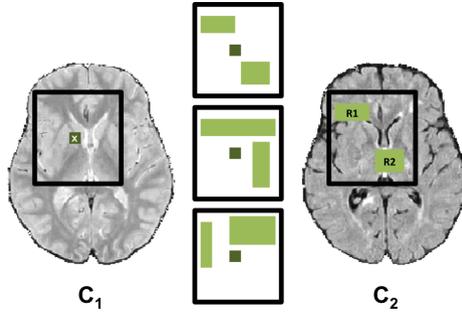


Fig. 2. Scheme of the meta-features. The 3D regions R_1 and R_2 are randomly sampled inside a fixed neighbourhood around the central voxel x .

an outlier map, a set of randomly sampled context features and a set of training cases. From these training cases we randomly sample a set of positive examples pertaining to WML and negative samples pertaining to any tissue with a ratio of 1 positive example for 3 negative ones (to account for the 3 possible tissues). These samples are the ones used to train a classifier. We decide here to use the Gentleboost algorithm [2], which is based on the simple idea that the sum of weak classifiers can produce a strong classifier. In the Gentleboost algorithm, the weak classifiers (h_t) are regression stumps with one of the features, so at each round t (of a total of N_{rounds}) the feature with less error is selected. The weak classifier used is:

$$h_t(p) = a\delta(p_i > th) + b \quad (4)$$

where th is a threshold that determines if pattern x belongs to the object class, x_i is the i th dimension of x , and a and b are parameters selected to minimise the error of the classifier (a is the regression slope and b the offset). At each round the training data weights (z) are updated, increasing in the following round the possibility of classifying correctly the previous incorrectly classified points:

$$z_{t+1} = z_t e^{y \cdot h_t(p)} \quad (5)$$

Hence, when testing a new data, the final (strong) classifier is computed using the weak classifier created at each round of the boosting. Therefore, the testing data is classified according to the sign of the sum of weak classifiers:

$$H(p) = \sum_t^{N_{rounds}} h_t(p) \quad (6)$$

The absolute value of $H(p)$ shows the confidence of the classified data.

3 Results

The dataset contains 30 cases and it comprises images from two different hospitals: 15 cases were acquired using a 1.5T Siemens Symphony Quantum scanner

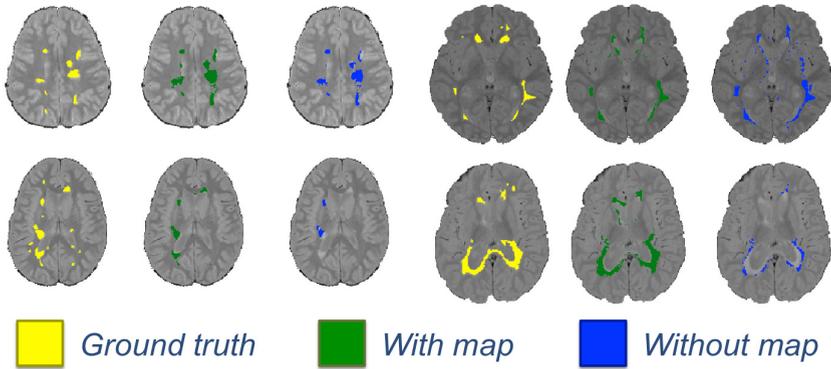


Fig. 3. Two different slices for each hospital with the ground truth, the segmentation using the lesion map and the segmentation without the lesion map

(Hospital Vall d’Hebron - 1XX) and the other 15 were obtained using a 1.5T Philips Intera scanner (Hospital Josep Trueta - 2XX). The scanning protocol included T1-w, T2-w, PD-w and FLAIR images for each patient with $1 \times 1 \times 3$ mm dimensions. All cases were manually annotated by an expert radiologist. The database contained a high variability in terms of lesion load per patient and also per hospital.

To evaluate the approach, we performed 2 different experiments: one using only image intensities, atlas probabilities and 800 context features, and a second one where we extended the feature pool by adding the outlier map as a feature and as part of the context features computation. For each experiment we created 5 different feature sets (with random R_1 and R_2 regions) that were used on a leave-one-out strategy per hospital.

During the training step, for each training set (containing 14 images from the same hospital) we randomly sampled 5 different sets of positive and negative examples. These samples were then used to train the classifier. Finally this classifier was used on the testing image to perform the segmentation. Therefore, for each patient we obtained 50 different classification maps (25 with lesion map and 25 without) that were thresholded to obtain the final segmentation. To obtain the threshold, we tested the classifier with the training images and computed the optimal threshold for each one according to the Dice similarity coefficient (DSC) value. Afterwards, the mean of those thresholds was used on the new patient’s output. Fig. 3 shows some qualitative results of segmentations for one patient of each hospital. For comparative purposes, we also include, in our analysis, the results obtained by the optimal threshold.

Fig. 4.a) shows a comparison of the segmentation results in terms of DSC between training the classifier with the outlier map or without it. In this plot, the images are ordered in terms of lesion load in ascending order. Therefore, the last image is the one with the highest lesion load. In average, when using the lesion map, DSC values were higher (with the exception of the cases with a low

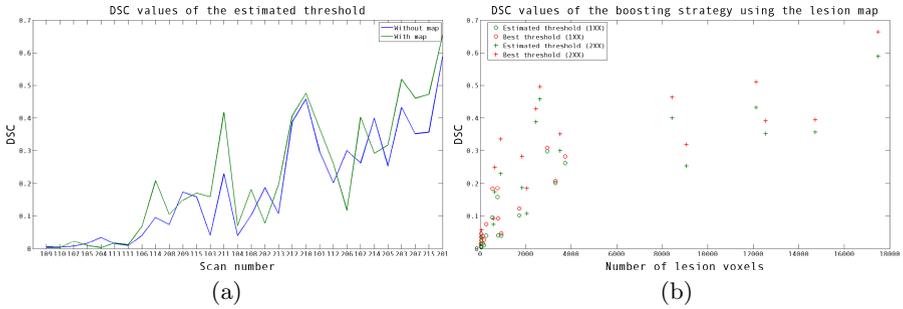


Fig. 4. (a) Comparison between using or not using the outlier map as a feature and (b) DSC values when using the lesion map according to the lesion volume

lesion load), pointing to a better training and classification process when this feature is used. Moreover, the first feature selected by the classifier (and thus the most discriminative one) was always the outlier map, meaning that the map is a good initial estimate of the WML.

In Fig. 4.b) we can see a better comparison between the total lesion load and obtained DSC values when using the outlier map. Results are presented for the estimated final threshold and the optimal threshold in terms of DSC. In this plot, three sets of results can be differentiated in terms of lesion load. The first group comprises the cases with the lowest lesion load (less than 100 voxels), the second group comprises the cases with what we consider a medium lesion load (between 100 and 5000 voxels) and the last group comprises the cases with a high lesion load (more than 5000 voxels). We can clearly see how the first group obtained lowest results due mostly to false positives and the difficulty of segmenting small lesions, while the second group showed a linear correlation between the lesion load and the DSC value. Finally, we can see that the last group (the one with patients with big lesions) obtained fairly good results in terms of DSC (with a mean of 0.5), mainly due to the higher number of true positive voxels. This proves that the higher the lesion load, the easier the segmentation. It also demonstrates how the DSC measure is biased to size when both the automatic and manual masks are small.

Regarding the lesion detection performance, using the outlier map we obtained mean true positive fraction values of 53% and 63% for hospital Vall d’Hebron and Josep Trueta respectively, in contrast to a mean true positive fraction of 49% and 51% when the outlier map is not used. The true positive fractions obtained with the outlier map are higher than those recently reported by Geremia et al. [3] and in accordance with the current state of the art within the automatic MS lesion segmentation.

4 Conclusions

A boosting classification approach to segment MS lesions on brain MRI have been presented in this paper. The classification has been done using a rich pool

of features which include intensity features, contextual features and an extended version of the outlier map introduced in [9] to deal with multi-channel images. The experimental results obtained with two datasets from two different hospitals have shown a better segmentation when using the extended outlier map as a feature in conjunction with the classical and contextual features. Further work will intend to improve the specificity of the approach, being this a common issue in all the automated approaches, as well as to optimise the approach for 3T images, to be tested with the 2008 MS Challenge database.

Acknowledgments. This work has been funded by the grants PI09/91018, CTM2011-29691-C02-02, CEM-Cat 2011 and TIN2012-37171-C02-01. M. Cabezas holds a FI grant.

References

1. Cabezas, M., Oliver, A., et al.: A review of atlas-based segmentation for magnetic resonance brain images. *Comput. Meth. Prog. Biomed.* 104(3), e158–e177 (2011)
2. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting. *Ann. Stat.* 38(2), 337–374 (2000)
3. Geremia, E., Clatz, O., et al.: Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. *NeuroImage* 57(2), 378–390 (2011)
4. Jenkinson, M., Beckmann, C.F., et al.: FSL. *NeuroImage* 62(2), 782–790 (2012)
5. Lladó, X., Oliver, A., et al.: Segmentation of multiple sclerosis lesions in brain MRI: A review of automated approaches. *Inform. Sciences* 186(1), 164–185 (2012)
6. Morra, J., Tu, Z., et al.: Automatic segmentation of MS lesions using a contextual model for the MICCAI grand challenge. In: *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, pp. 1–11 (2008)
7. Okuda, T., Korogi, Y., et al.: Brain lesion: when should fluid-attenuated inversion recovery sequences be used in MR evaluation? *Radiology* 212(3), 793–798 (1999)
8. Rueckert, D., Sonoda, L., et al.: Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans. Med. Imag.* 18(8), 712–721 (1999)
9. Schmidt, P., Gaser, C., et al.: An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis. *NeuroImage* 59(4), 3774–3783 (2012)
10. Shotton, J., Winn, J., et al.: Textonboost: multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *Int. J. Comp. Vis.* 81(1), 2–23 (2009)
11. Tustison, N., Avants, B., et al.: N4ITK: Improved N3 bias correction. *IEEE Trans. Med. Imag.* 29(6), 1310–1320 (2010)
12. Vovk, U., Pernus, F., Likar, B.: A review of methods for correction of intensity inhomogeneity in MRI. *IEEE Trans. Med. Imag.* 26(3), 405–421 (2007)