



Multi-atlas Parcellation in the Presence of Lesions: Application to Multiple Sclerosis

Sandra González-Vilà^{1,2(✉)}, Yuankai Huo², Arnau Oliver¹, Xavier Lladó¹,
and Bennett A. Landman²

¹ Institute of Computer Vision and Robotics, University of Girona, Ed. P-IV,
Campus Montilivi, 17003 Girona, Spain

sgonzalez@eia.udg.edu

² Electrical Engineering, Vanderbilt University, Nashville, TN 37235, USA

Abstract. Intensity-based multi-atlas strategies have shown leading performance in segmenting healthy subjects, but when lesions are present, the abnormal lesion intensities affect the fusion result. Here, we propose a reformulated statistical fusion approach for multi-atlas segmentation that is applicable to both healthy and injured brains. This method avoids the interference of lesion intensities on the segmentation by incorporating two *a priori* masks to the Non-Local STAPLE statistical framework. First, we extend the theory to include a lesion mask, which improves the voxel correspondence between the target and the atlases. Second, we extend the theory to include a known label mask, that forces the label decision in case it is beforehand known and enables seamless integration of manual edits. We evaluate our method with simulated and MS patient images and compare our results with those of other state-of-the-art multi-atlas strategies: Majority vote, Non-local STAPLE, Non-local Spatial STAPLE and Joint Label Fusion. Quantitative and qualitative results demonstrate the improvement in the lesion areas.

Keywords: Brain parcellation · Segmentation · Multiple sclerosis

1 Introduction

Brain parcellation has become an essential tool for understanding neurological structural-functional associations at a millimeter scale. The resulting voxelwise tissue classifications are integral to identifying structural regions for connectomics, functional activations, quantitative/metabolical changes, diffusion connectivity, etcetera. These techniques require reproducible segmentations; however, manual delineation is time-consuming, exhibits poor reproducibility, and is subject to inter- and intra- operator variability. For these reasons, automatic brain parcellation has been widely studied [1–3]. Several automatic strategies have been proposed in the literature to segment brain structures, such as deformable, learning-based, region-based, etc. [4–6]; however, most of these

methods are structure-specific and do not allow segmentation of the whole brain. In contrast, atlas-based strategies provide a whole parcellation when the atlases used have all the structures labelled.

In multi-atlas segmentation, a collection of atlases is registered to the target image and their labels are propagated and fused in the target image space, obtaining the final segmentation. Label fusion strategies based on intensities [7–10] have been demonstrated to be robust and provide good performance when dealing with healthy subjects. However, as most other state-of-the-art methods, they are designed to segment healthy subjects and their performance tends to be affected when segmenting brains hindered by tumors and lesions, for instance, as a result of multiple sclerosis (MS) [12].

Herein, we propose a novel statistical fusion algorithm that reformulates the non-local STAPLE (NLS) [8] statistical framework to handle (anatomical) MRI visible lesions. As in NLS, our method models the registered atlases as collections of volumetric patches with intensity and label information. To complement the non-local criteria, we introduce lesion mask information to resolve the imperfect correspondences between the healthy atlases and the lesioned target derived from inaccurate registrations. Additionally, a second mask is integrated into the estimation process, which forces the voxel label assignment in case it is known beforehand. For instance, this modification is useful when segmenting brains with tumors for which sub-regions are known. Together, these innovations enable inclusion of masks of abnormal anatomy and manually provided *edits* within modern statistical fusion approaches. We derive the theoretical basis governing our method and demonstrate segmentation improvement with respect to other multi-atlas strategies on the state of the art on both simulated and MS images.

2 Theory

Consider a target gray-level image (with lesions) represented as a vector $I \in \mathbb{R}^{N \times 1}$. Let $T \in \mathcal{L}^{N \times 1}$ be the latent representation of the true target segmentation, where $\mathcal{L} = \{0, \dots, L - 1\}$ is the set of possible labels which can be assigned to a concrete voxel. Let $M \in \{0, 1\}^{N \times 1}$ be a binary lesion mask indicating whether a given voxel i of the target image contains or is part of a lesion and $K \in \{0, 1\}^{N \times 1}$ a second mask specifying if for a given voxel i of the target image, the true label is known, hence $M_i = p(I_i \in \textit{lesion})$ and $K_i = p(T_i = T_k \in \mathcal{L}^{N \times 1})$. Note that both masks are optional and can be neglected if all voxels in the mask are set to 0. Consider a set R of registered healthy atlases with associated gray level images, $A \in \mathbb{R}^{N \times R}$, and propagated label decisions, $D \in \mathcal{L}^{N \times R}$. Let $\theta \in [0, 1]^{R \times N \times L \times L}$ be the performance level parameters of the raters (registered atlases), defined voxel-wise. Each element of θ , $\theta_{jis's}$, represents the probability that rater j observes label s' given that the true label is s at a given voxel i and the corresponding voxel i^* on the associated atlas—i.e., $\theta_{jis's} = p(D_{i^*j} = s', A_j|T_i = s, I_i, M_i, K_i, \theta_{jis's})$, where i^* is the voxel on atlas j that corresponds to the target voxel i .

2.1 Non-local Correspondence Model

Non-local STAPLE (NLS) [8] incorporates the concept of patch-based non-local correspondence based on the image intensities of both the target image I and the registered atlases A to the STAPLE framework. Although this concept has proven useful for matching healthy tissues to account for registration accuracy, we cannot rely on intensity similarities between the target lesion areas and the healthy atlases to rectify registration errors. Therefore, we assume that voxel correspondence inside the lesions cannot be further improved based on intensity and, hence, enforce the non-local weighting ($\alpha_{ji'i}$) between voxel i in the target image at voxel i' on the j th atlas as follows:

$$\alpha_{ji'i} = \left(\frac{1}{Z_\alpha} \exp \left(- \frac{\|\wp_{M_i} \circ (\wp(A_{i'j}) - \wp(I_i))\|_2^2}{2 \cdot \sigma_i^2 \cdot \|\wp_{M_i}\|} \right) \exp \left(- \frac{\varepsilon_{i'i}^2}{2 \cdot \sigma_d^2} \right) \right) \cdot (1 - M_i) + \delta(i' = i) \cdot M_i \quad (1)$$

where $\wp(\cdot)$ is the set of intensities in the patch neighborhood of a given intensity location. In this definition, $\wp_{M_i} = \wp(1 - M_i)$ is the masking term that excludes lesion voxels from the patch calculation and enforces the same patch neighborhood size/shape in both the atlas and the target, $\|\wp_{M_i} \circ (\wp(A_{i'j}) - \wp(I_i))\|_2^2$ is the L2-norm between the atlas patch centered at i' and the target patch centered at i , $\varepsilon_{i'i}^2$ is the Euclidean distance in physical space between i and i' , σ_i and σ_d are the standard deviations of the intensity and distance weights, and Z_α is a partition function that enforces the constraint that $\sum_{i' \in \mathcal{N}(i)} \alpha_{ji'i} = 1$, where $\mathcal{N}(i)$ is the set of voxels in the search neighborhood of a given target voxel. $\delta(i' = i)$ is the Dirac delta function, and $\|\wp_{M_i}\|$ is the number of voxels in the patch neighborhood.

2.2 The Algorithm

If the exact voxel correspondences between the target and the atlases (non-local model) were known, the lesion mask, and the target and atlas intensity relationships could be ignored and the spatial STAPLE [11] definition of θ could be used.

$$\begin{aligned} \theta_{jis's} &\equiv p(D_{i^*j} = s', A_j | T_i = s, I_i, M_i, K_i, \theta_{jis's}) \\ &= p(D_{i^*j} = s' | T_i = s, M_i, K_i, \theta_{jis's}) \end{aligned} \quad (2)$$

However, this correspondence is not known and we have to learn it with the model defined in Sect. 2.1. Note that using this model we can approximate the relationship by taking the expected value of $p(D_{i^*j} = s', A_j | T_i = s, I_i, M_i, K_i, \theta_{jis's})$ across the raters. Using an assumption of conditional independence between the labels, lesion mask and intensity, we approximate the density function as:

$$\begin{aligned} p(D_{i^*j} = s', A_j | T_i = s, I_i, M_i, K_i, \theta_{jis's}) &\approx E[p(D_j, A_j | T_i = s, I_i, M_i, K_i, \theta_{jis})] \\ &= E[p(D_j | T_i = s, M_i, K_i, \theta_{jis}) \cdot p(A_j | I_i, M_i)] \\ &= \sum_{i' \in \mathcal{N}(i)} p(D_{i^*j} = s' | T_i = s, M_i, K_i, \theta_{jis's}) \cdot p(A_{i'j} | I_i, M_i) = \sum_{i' \in \mathcal{N}(i)} \theta_{jis's} \cdot \alpha_{ji'i} \end{aligned} \quad (3)$$

E-step. Let $W \in \mathbb{R}^{L \times N}$, where $W_{si}^{(t)}$ represents the probability that the true label associated with voxel i is s at iteration t of the algorithm given the provided information and the performance level parameters.

$$W_{si}^{(t)} \equiv p\left(T_i = s | D, A, I, M, K, \theta^{(t)}\right) \quad (4)$$

Using Bayes' rule to separate the prior label probability ($p(T_i = s)$) and assuming independence among the raters, we can rewrite this equation as follows:

$$W_{si}^{(t)} \equiv \frac{(1-K_i) \cdot (p(T_i=s) \cdot \prod_j p(D_{i^*j}=s', A_j | T_i=s, I_i, M_i, K_i, \theta_{jis's}^{(t)})) + K_i \cdot \delta(s'=s)}{(1-K_i) \cdot (\sum_n p(T_i=n) \cdot \prod_j p(D_{i^*j}=s', A_j | T_i=n, I_i, M_i, K_i, \theta_{jis's}^{(t)})) + K_i \cdot \delta(s'=s)} \quad (5)$$

where $\delta(s' = s)$ is the Dirac delta function (probability that the known label for voxel i of the truth segmentation is s). Using the non-local correspondence model and the approximated density function, we obtain:

$$W_{si}^{(t)} \equiv \frac{(1-K_i) \cdot (p(T_i=s) \cdot \prod_j \sum_{i' \in \mathcal{N}(i)} \theta_{jis's}^{(t)} \cdot \alpha_{ji'i'}) + K_i \cdot \delta(s'=s)}{(1-K_i) \cdot (\sum_n p(T_i=n) \cdot \prod_j \sum_{i' \in \mathcal{N}(i)} \theta_{jis's}^{(t)} \cdot \alpha_{ji'i'}) + K_i \cdot \delta(s'=s)} \quad (6)$$

M-step. In this step, the calculated $W_{si}^{(t)}$ is used to update $\theta_{ji}^{(t+1)}$ by maximizing the expectation of the complete data log likelihood. As the complete data log likelihood is not observable, it is replaced by its conditional expectation given the observable data D, A, I, M, K using the current estimate θ .

$$\begin{aligned} \theta_{ji}^{(t+1)} &= \arg \max_{\theta_{ji}} \sum_{i' \in \mathcal{B}_i} E \left[\ln \left(p \left(D_j, A_j | T_{i'}, I_{i'}, M_{i'}, K_{i'}, \theta_{ji} | D, A, I, M, K, \theta^{(t)} \right) \right) \right] \\ &= \arg \max_{\theta_{ji}} \sum_{i' \in \mathcal{B}_i} \sum_s p \left(T_{i'} = s | D, A, I, M, K, \theta^{(t)} \right) \cdot \ln \left(p \left(D_j, A_j | T_{i'}, I_{i'}, M_{i'}, K_{i'}, \theta_{ji} \right) \right) \\ &= \arg \max_{\theta_{ji}} \sum_{i' \in \mathcal{B}_i} \sum_s W_{si'}^{(t)} \cdot \ln \left(p \left(D_{i^*j} = s', A_j | T_{i'}, I_{i'}, M_{i'}, K_{i'}, \theta_{ji} \right) \right) \\ &= \arg \max_{\theta_{ji}} \sum_{i' \in \mathcal{B}_i} \sum_s W_{si'}^{(t)} \cdot \ln \left(\sum_{i'' \in \mathcal{N}(i'): D_{i''j}=s'} \theta_{jis's} \cdot \alpha_{ji''i'} \right) \end{aligned} \quad (7)$$

As each row of θ must sum one to be a valid probability mass function, we can maximize the performance level parameters for each element by using a Lagrange multiplier (λ) to formulate the constrained optimization problem.

$$0 = \frac{\delta}{\delta \theta_{jis's}} \left[\sum_{i' \in \mathcal{B}_i} \sum_s W_{si'}^{(t)} \cdot \ln \left(\sum_{i'' \in \mathcal{N}(i'): D_{i''j}=s'} \theta_{jis's} \cdot \alpha_{ji''i'} \right) + \lambda \sum_{s'} \theta_{jis's}^{(t+1)} \right] \quad (8)$$

By solving this equation, we obtain

$$\theta_{jis's}^{(t+1)} = \frac{\sum_{i' \in \mathcal{B}_i} \left(\sum_{i'' \in \mathcal{N}(i'): D_{i''j}=s'} \alpha_{ji''i'} \right) \cdot W_{si'}^{(t)}}{\sum_{i' \in \mathcal{B}_i} W_{si'}^{(t)}} \quad (9)$$

2.3 Initialization and Priors

The voxel-wise prior $p(T_i = s)$ was initialized using the weak log-odds majority vote, as in NLSS. The performance parameters, $\theta_{jis's}$, were initialized assuming each atlas has high performance as: 1, if $s = s'$; 0.95, if $s \neq s'$; and $\frac{0.05}{L-1}$, otherwise. The search neighborhood $\mathcal{N}(\cdot)$ was set to $7 \times 7 \times 7$, patch $\wp(\cdot)$ dimensions to $5 \times 5 \times 5$ and σ_i and σ_d were set to 0.25 and 1.5, respectively. Algorithm convergence was detected when the average change in the diagonal elements of θ was below 10^{-4} .

3 Experiments and Results

The atlases used in our experiments were taken from the MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling database [13]. This database consists of 35 T1-w MR images, obtained from the OASIS¹ project and labeled by Neuromorphometrics, Inc.², and includes labels for the whole brain. PCA atlas selection was performed and only the 15 most similar atlases were used for segmentation. All images were histogram normalized and N4 bias field corrected before registration. All pair-wise registrations were performed using an initial affine registration (niftyreg³) followed by a non-rigid (ANTs⁴) procedure. In all the registrations performed, the lesions were masked-out to avoid their intensities to interfere in the similarity metric calculation.

As benchmarks, we compare the proposed algorithm to majority vote (MV) [14], non-local STAPLE (NLS) [8], non-local Spatial STAPLE (NLSS) [9] and Joint Label Fusion (JLF) [10]. For a fair comparison, all the parameters that NLS and NLSS share with our algorithm were set to the same values. Also, JLF was executed with the same patch and neighborhood size.

3.1 Simulated Lesions

Evaluating the performance of segmentation algorithms on real lesioned images is not an easy task since there is a lack of public databases with ground truth for both lesions and structures. For this reason, in the first experiment, we simulated two sets of artificially lesioned images: (1) 10 with uniform intensity lesions, to test the proposed theory and, (2) 15 with lesion shapes, intensities and locations obtained from an in-house MS patient database, to simulate realistic cases. All the lesions were generated on random subjects from the MICCAI 2012 database. The lesion load of the generated images ranged from [33.49–119.74] mm³ in the first cohort and from [3.16–26.96] mm³ in the second one.

¹ <http://www.oasis-brains.org/>.

² <http://neuromorphometrics.com/>.

³ <http://cmictig.cs.ucl.ac.uk/wiki/index.php/NiftyReg>.

⁴ <https://sourceforge.net/projects/advants/>.

We evaluated the segmentation results quantitatively using a global Dice Similarity Coefficient (DSC) across all the structures as the main measure. As the lesion intensities not only necessarily affect the lesion area segmentation itself, but also the surrounding tissues, two measures were calculated: (1) DSC inside the lesion mask, and (2) DSC inside a mask that included three voxels of the contour. Note that $\mathcal{N}(\cdot)$ was set to $7 \times 7 \times 7$.

Figure 1(A) shows that, inside the lesion mask, our method performed significantly better than all the intensity-based strategies (JLF, NLS and NLSS) in both cohorts. However, the performance was similar to that of MV. This is due to the fact that we cannot trust the intensities inside the lesions, and we can only rely on an accurate registration (same as MV does). On the other hand, when the performance was analyzed around the lesion areas, our proposal was the one that provided the best results (similar to MV in the first cohort and to JLF in the second one). This behavior is depicted in Fig. 1(B), where JLF (b and h) misclassifies several structures inside the lesion areas, whereas in NLSS (c and i) the segmentation is being also affected in the surrounding structures.

For the evaluation of the manual edits (K mask integration), we segmented the first dataset again, this time feeding the algorithm with the same lesion mask for both M and K. The results showed, as expected, a DSC of 1 inside the lesion areas (M/K mask), whereas the mean DSC around the lesions was 0.7901 ± 0.0463 , very similar to that of the first execution (0.7919 ± 0.0457), conserving a similar effect on the tissues surrounding the lesions.

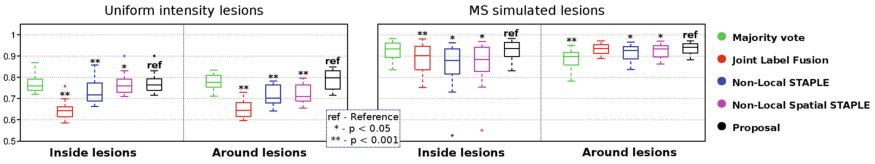
3.2 MSSeg 2016 Challenge

For the second experiment, we qualitatively compared the fusion results obtained by the analyzed algorithms on a MS patient database (MSSeg 2016 challenge⁵).

Figure 2 shows the segmentations obtained with all the analyzed multi-atlas strategies. As we can observe from Fig. 2(a), MS lesions are hypo-intense in T1-w modality, which makes its intensity profile similar of that of the gray matter (GM) and even sometimes similar to the cerebro-spinal fluid (CSF) which may affect the results of intensity-based algorithms. The lesions shown in Fig. 2(b), should be classified as white matter, however, the intensity-based algorithms of the state of the art, Fig. 2(f–h), tend to misclassify those regions as GM or CSF, whereas our method, Fig. 2(c), shows better classification results in those areas. When our method is fed with a K mask, Fig. 2(d), the lesion surrounding voxels remain practically the same as when the K mask is not used, Fig. 2(c), whereas the segmentation result inside the lesions agrees entirely with the labels imposed by this mask, as seen in Sect. 3.1.

⁵ <https://portal.fli-iam.irisa.fr/msseg-challenge/overview>.

A. Global DSC



B. Qualitative Comparison

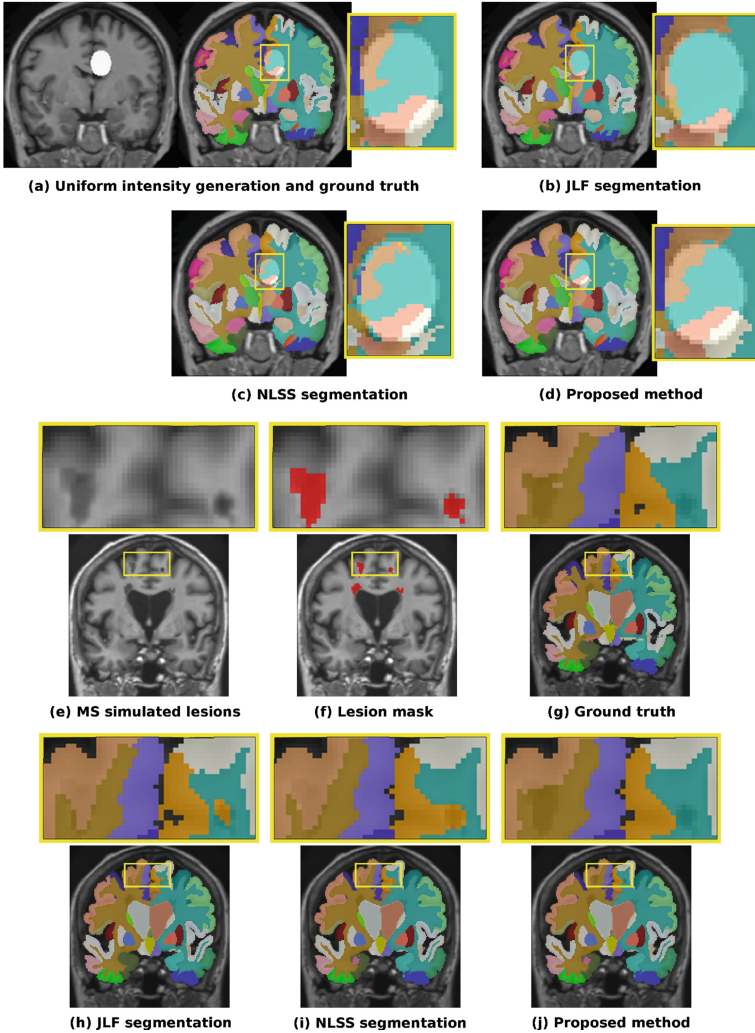


Fig. 1. (A) Global DSC and (B) qualitative segmentation results of analyzed multi-atlas strategies on both simulated databases: (a–d) uniform intensity lesions, and (e–j) MS simulated lesions.

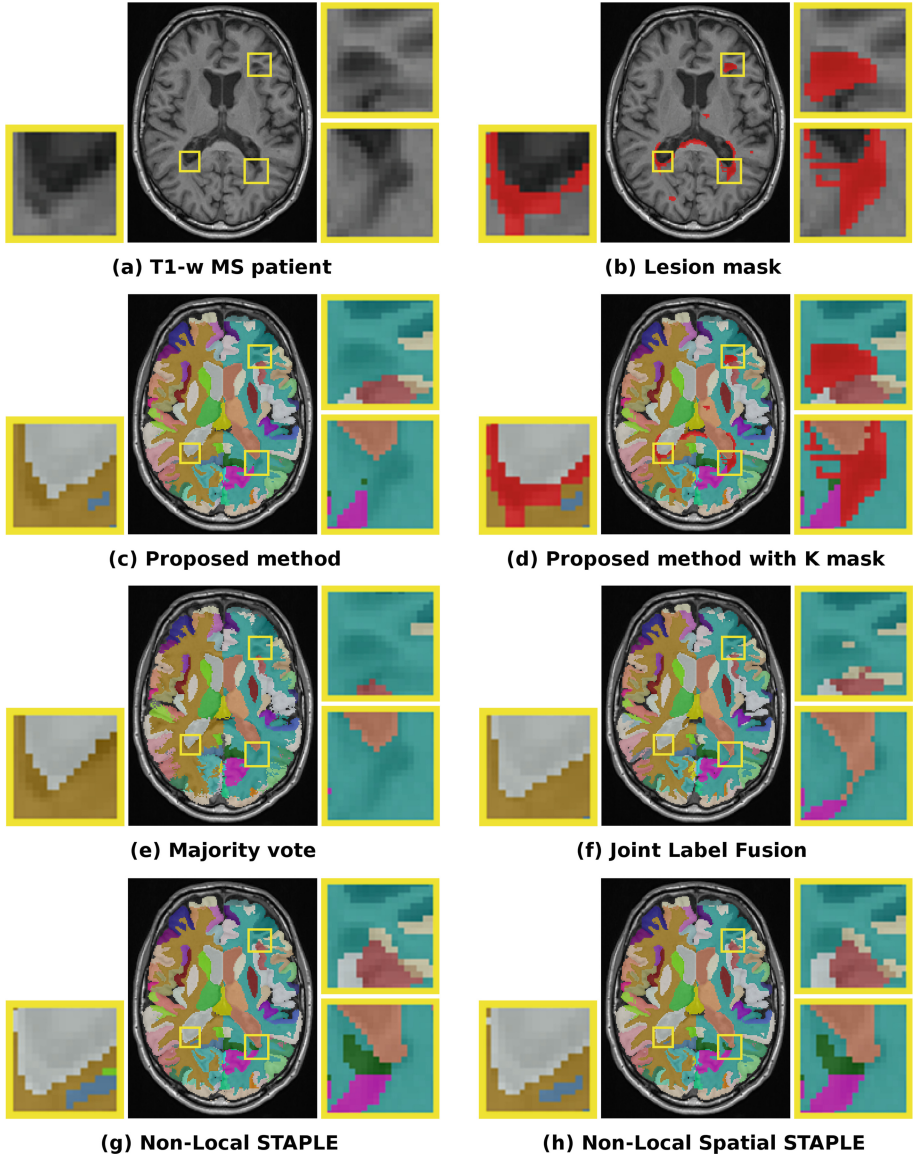


Fig. 2. Segmentation results of the analyzed multi-atlas strategies for the image 01038PAGU of the MICCAI2016 Challenge database.

4 Discussion

Accurate structural volume measurements are important in MS, since the atrophy of some structures such as the deep GM is relevant to the disease progression. However, we have shown that multi-atlas strategies based on intensities, which

achieve good segmentation results on healthy subjects, are affected by lesions, and therefore corrupting real measures.

Herein, we have presented the theory to modify the non-local STAPLE framework to deal with MRI visible lesions. The experiments performed show that our proposal outperforms the state-of-the-art multi-atlas strategies in the lesion areas for both simulated and MS patient images.

Over-performance of MV compared to the state-of-the-art intensity-based strategies was observed on the experiments performed on the uniform intensity lesions database around the lesion areas. This behavior could be due to the fact that the other strategies are patch-based. These strategies consider mean patch differences to calculate the correspondences, hence the bright voxels of the lesions could bias the mean intensity, finding wrong atlas correspondences. Even though, this is an extreme case to test the proposed theory, it shows, combined with the over-performance of MV inside the lesion areas, the effect of the lesion intensities on the segmentation.

In this work, we have only focused on the segmentation performance of the lesion areas, since those are the ones concerning the proposed reformulation. Nonetheless, as these areas are better segmented with our strategy, the average whole brain segmentation performance slightly increases compared to the non-local STAPLE variants. This small improvement is due to lesions are small compared to the whole brain volume. For this reason, we believe that extending our theory to other methods of the literature, such as JLF, would be beneficial in terms of segmentation accuracy of the lesion areas but also of the whole brain.

Acknowledgements. The authors would like to thank Jose Bernal for helpful discussion. Sandra González-Villà holds a UdG-BRGR2015 grant. This work has been supported by “La Fundació la Marató de TV3”, by Retos de Investigación TIN2014-55710-R and TIN2015-73563-JIN, by DPI2017-86696-R and by UdG mobility grant MOB17. This research was supported by NSF CAREER 1452485, NIH R01-EB017230, and the National Center for Advancing Translational Sciences, Grant 2 UL1 TR000445-06. We appreciate the NIH S10 Shared Instrumentation Grant 1S10OD020154-01 (Smith), Vanderbilt IDEAS grant (Holly-Bockelmann, Walker, Meliler, Palmeri, Weller), and ACCRE’s Big Data TIPs grant from Vanderbilt University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, National Science Foundation, or other sponsors.

References

1. Iglesias, J.E., Sabuncu, M.R.: Multi-atlas segmentation of biomedical images: a survey. *Med. Image Anal.* **24**(1), 205–219 (2015)
2. Li, W., Wang, G., Fidon, L., Ourselin, S., Cardoso, M.J., Vercauteren, T.: On the compactness, efficiency, and representation of 3D convolutional networks: brain parcellation as a pretext task. In: Niethammer, M., et al. (eds.) *IPMI 2017*. LNCS, vol. 10265, pp. 348–360. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_28
3. Wachinger, C., Reuter, M., Klein, T.: DeepNAT: deep convolutional neural network for segmenting neuroanatomy. *NeuroImage* **170**, 434–445 (2018)

4. González-Villà, S., Oliver, A., Valverde, S., Wang, L., Zwiggelaar, R., Lladó, X.: A review on brain structures segmentation in magnetic resonance imaging. *Artif. Intell. Med.* **73**, 45–69 (2016)
5. Kushibar, K., et al.: Automated sub-cortical brain structure segmentation combining spatial and deep convolutional features. *Med. Image Anal.* **48**, 177–186 (2018)
6. Dolz, J., Desrosiers, C., Ayed, I.B.: 3D fully convolutional networks for subcortical segmentation in MRI: a large-scale study. *NeuroImage* **170**, 456–470 (2018)
7. Coupé, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. *NeuroImage* **54**(2), 940–954 (2011)
8. Asman, A.J., Landman, B.A.: Non-local statistical label fusion for multi-atlas segmentation. *Med. Image Anal.* **17**(2), 194–208 (2013)
9. Huo, Y., Asman, A.J., Plassard, A.J., Landman, B.A.: Simultaneous total intracranial volume and posterior fossa volume estimation using multi-atlas label fusion. *Hum. Brain Mapp.* **38**(2), 599–616 (2017)
10. Wang, H., Suh, J.W., Das, S.R., Pluta, J.B., Craige, C., Yushkevich, P.A.: Multi-atlas segmentation with joint label fusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(3), 611–623 (2013)
11. Asman, A.J., Landman, B.A.: Formulating spatially varying performance in the statistical fusion framework. *IEEE Trans. Med. Imaging* **31**(6), 1326–1336 (2012)
12. González-Villà, S., et al.: Evaluating the effect of multiple sclerosis lesions on automatic brain structure segmentation. *NeuroImage Clin.* **15**, 228–238 (2017)
13. Landman, B., Warfield, S.: MICCAI 2012 workshop on multi-atlas labeling. In: MICCAI Grand Challenge and Workshop on Multi-Atlas Labeling, Nice, France (2012)
14. Sabuncu, M., Yeo, B., Van Leemput, K., Fischl, B., Golland, P.: A generative model for image segmentation based on label fusion. *IEEE Trans. Med. Imaging* **29**(10), 1714–1729 (2010)