Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2014 February 15; 9034: . doi:10.1117/12.2043917.

Example Based Lesion Segmentation

Snehashis Roy^{*}, Qing He^{*}, Aaron Carass[†], Amod Jog[†], Jennifer L. Cuzzocreo[§], Daniel S. Reich[‡], Jerry Prince[†], and Dzung Pham^{*}

^{*}Center for Neuroscience and Regenerative Medicine, Henry M. Jackson Foundation for the Advancement of Military Medicine, USA

[†]Department of Electrical and Computer Engineering, The Johns Hopkins University, USA

[§]Department of Radiology, The Johns Hopkins School of Medicine, USA

[‡]Translational Neuroradiology Unit, Neuroimmunology Branch, National Institute of Neurological Disorders and Stroke, USA

Abstract

Automatic and accurate detection of white matter lesions is a significant step toward understanding the progression of many diseases, like Alzheimer's disease or multiple sclerosis. Multi-modal MR images are often used to segment T_2 white matter lesions that can represent regions of demyelination or ischemia. Some automated lesion segmentation methods describe the lesion intensities using generative models, and then classify the lesions with some combination of heuristics and cost minimization. In contrast, we propose a patch-based method, in which lesions are found using examples from an atlas containing multi-modal MR images and corresponding manual delineations of lesions. Patches from subject MR images are matched to patches from the atlas and lesion memberships are found based on patch similarity weights. We experiment on 43 subjects with MS, whose scans show various levels of lesion-load. We demonstrate significant improvement in Dice coefficient and total lesion volume compared to a state of the art model-based lesion segmentation method, indicating more accurate delineation of lesions.

Keywords

magnetic resonance imaging; MRI; lesion segmentation; MS; patches

1. INTRODUCTION

With Alzheimer's disease, multiple sclerosis (MS) and cerebral small vessel ischemic diseases, lesions often occur in cerebral white matter (WM). These lesions are usually hypointense in 3D gradient-echo T_1 -weighted images and hyperintense in T_2 -weighted images. Lesion volume has been shown to be correlated with the state and progression of MS.¹ Although manual segmentation of lesions are considered to be the "gold standard", fast and accurate automatic segmentations are essential to quantitatively assess the lesion load more robustly, as manual segmentations are time-consuming and are subject to both

inter-rater and intra-rater variations, especially when more than one image modality is used to guide manual delineation.

FLAIR (fluid attenuated inversion recovery) is often the preferred modality for segmenting white matter (WM) lesions since they show up as hyper-intense compared to other soft tissues (Fig. 1). In T_1 -w MPRAGE (magnetization prepared rapid gradient echo) or T_2 -w scans, lesion intensities are very similar to CSF (cerebrospinal fluid). Many model-based algorithms take MPRAGE and FLAIR images as inputs and model the lesion intensities as outliers in the intensity distributions,^{2,3} since lesion volumes are usually small compared to total WM volumes. Lesion-TOADS⁴ segments lesions using a multi-modal fuzzy C-means (FCM) based criteria. Recently it has been shown that many example based methods outperform their model-based counterparts, in segmentation⁵⁻⁸ and skull-stripping.⁹ In this paper, we improve upon a model-based method Lesion-TOADS⁴ and demonstrate that our example based lesion segmentation outperforms it significantly.

We use T_1 -w MPRAGE and FLAIR scans to find lesions using examples from atlases. Patches from a subject MPRAGE and FLAIR scans are matched to similar patches from atlas T_1 -w and FLAIR scans. The corresponding hard segmentation labels (Fig. 1) of the atlas are weighted by the similarity criteria accordingly,¹⁰ thereby creating a fuzzy lesion membership. The lesion memberships are thresholded to obtain a hard segmentation. Although this idea of patch matching has been explored previously for the purpose of image synthesis,¹⁰⁻¹³ atrophy detection,^{14,15} and tissue segmentation,⁶ the novelty of our method lies in fuzzy lesion segmentation using a sparse dictionary of WM lesions.

2. METHOD

We define the “*atlas*” as a triplet of co-registered T_1 -w MPRAGE, FLAIR, and the manual binary lesion segmentation images (e.g., Fig. 1). The “*subject*” is a co-registered pair of T_1 -w MPRAGE and FLAIR images. The atlas MPRAGE and FLAIR (both skull-stripped), having the same resolution as the subject, are generally not registered to the subject. Both subject and atlas MPRAGE and FLAIRs are normalized such that their modes of WM intensities are at unity. The modes are automatically obtained from the corresponding image histograms using kernel density estimator.¹⁶

At each voxel of an image, $3D\ p \times q \times r$ patches are stacked into 1D vectors of size $d \times 1$, $d = pqr$. Since atlas images are co-registered, the pair of patches centered at the same voxel of the atlas MPRAGE and FLAIR are concatenated to form a $2d \times 1$ vector, denoted $\mathbf{a}_1(i) \in \mathbb{R}^{2d \times 1}$, $i = 1, \dots, M$. For simplicity of notation, we will interchangeably denote “patch” to identify $2d \times 1$ vectors also. The subject MPRAGE and FLAIR are also similarly decomposed into patches, denoted $\mathbf{b}_1(j) \in \mathbb{R}^{2d \times 1}$, $j = 1, \dots, N$. M and N denote the number of non-zero voxels in the atlas and the subject, respectively. For typical 1mm^3 images, $M, N \sim 10^7$. For the i^{th} atlas patch, the corresponding patch in the lesion segmentation image is denoted by $\mathbf{a}_2(i) \in \mathbb{R}^{d \times 1}$. The elements of $\mathbf{a}_2(i)$ are $\{0, 1\}$.

We assume that for the j^{th} subject patch $\mathbf{b}_1(j)$, a small number of similar atlas patches can always be found from a rich collection of atlas patches, whose convex combination is $\mathbf{b}_1(j)$.

It is unlikely that a single atlas patch $\mathbf{a}_1(j)$ will match $\mathbf{b}_1(j)$ exactly, but a convex combination of a few atlas patches from a large dictionary of atlas patches is more likely to approximate $\mathbf{b}_1(j)$ closely. The idea can be succinctly written as

$$\mathbf{b}_1(j) \approx A_1(j)\mathbf{x}_j, \text{ for some } \mathbf{x}_j \in \mathbb{R}^{L \times 1}, \|\mathbf{x}_j\|_0 \ll L, \mathbf{x}_j \geq 0. \quad (1)$$

$A_1(j)$ is a dictionary containing atlas patches ($\mathbf{a}_1(j)$ s) similar to $\mathbf{b}_1(j)$. Instead of using all M atlas patches for every subject patch, we use a kd-tree to choose L similar atlas patches (from the larger set of M patches) for every single subject patch,¹⁷ such that $A_1(j) \in \mathbb{R}^{2d \times L}$. L is empirically chosen to be 100 based on speed and memory limitation. The variable $\mathbf{x}_j$ is a $L \times 1$ sparse weight vector indicating that only a few patches out of the L patches in $A_1(j)$ are used to reconstruct $\mathbf{b}_1(j)$. The non-negativity constraint on the weights $\mathbf{x}_j$ enforces the “texture” of the chosen atlas patches to match with that of the subject patch $\mathbf{b}_1(j)$. This idea of sparse weight reconstruction has been explored earlier in relation to atlas based super-resolution¹⁸ and image synthesis.^{10,11} Because of speed limitations, it has been suggested that similar patches should be chosen from a windowed neighborhood of the voxel in the atlas,⁶ after the atlas is registered to the subject. However, we choose atlas patches irrespective of their spatial location, since the anatomy might vary widely between the subject and atlas. For example, if the subject has large ventricles and the atlas does not, the size of the search window should vary according to the size of the structure,^{6,9} and the accuracy of the registration used. We avoid this problem at the cost of a nearest-neighbor search using a kd-tree irrespective of the spatial locations of the atlas patches.

Every atlas patch $\mathbf{a}_1(j)$ has a corresponding lesion label patch $\mathbf{a}_2(j)$. Thus, once the sparse weights $\mathbf{x}_j$ are found, the lesion labels are combined accordingly to form the lesion membership for the j^{th} subject patch as

$$\hat{\mathbf{b}}_2(j) = A_2(j) \frac{\mathbf{x}_j}{\|\mathbf{x}_j\|_1}, \quad (2)$$

where $A_2(j) \in \mathbb{R}^{d \times L}$ is a wide matrix that includes lesion segmentation patches as columns corresponding to $A_1(j)$. The lesion membership $\hat{\mathbf{b}}_2(j)$ is normalized by $\|\mathbf{x}_j\|_1$ so that it ranges between [0, 1]. It is advantageous to generate a continuous valued variable (instead of a hard segmentation) that could have a probabilistic interpretation to allow tuning the threshold. Also, because the boundaries of lesions are often vague, this provides a measure of confidence.

To solve the ℓ_0 problem in Eqn. 1 efficiently, it can be reformulated as a ℓ_1 minimization, which gives the same optimal solution as the ℓ_0 problem, if $\|\mathbf{x}_j\|_0$ is sufficiently small.²⁰ Eqn. 1 is rewritten as

$$\hat{\mathbf{x}}_j = \arg \min_{\mathbf{x} \geq 0} \left[\|\mathbf{b}_1(j) - A_1 \mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right], \quad (3)$$

where λ is a weight penalizing the sparsity of $\hat{\mathbf{x}}_j$. Here we note that for uniqueness of the solution of Eqn. 3, both $\mathbf{b}_1(j)$ and each column of A_1 are to be normalized so that their ℓ_2 norm is unity.¹⁸ All patches are normalized by the maximum of their ℓ_2 norms.¹⁰ We solve for $\hat{\mathbf{x}}_j$ using SparseLab.²⁰ In all our experiments, we use overlapping $3 \times 3 \times 3$ patches ($d=27$) and empirically set $\lambda = 0.02$. Once $\hat{\mathbf{x}}_j$ is found from Eqn. 3, the corresponding lesion membership $\hat{\mathbf{b}}_2(j)$ is obtained from Eqn. 2 and only the value at the center voxel is used.

3. RESULTS

We use a pool of 47 MS subjects, all having lesions. Manual segmentations are available on all of them. Imaging protocols and manual segmentation details can be found in.⁴ To quantitatively evaluate the segmentation accuracy with Lesion-TOADS, we used the following measures: Dice coefficient, total lesion volume, sensitivity (true positive rate) and specificity (false negative rate). Four randomly chosen subjects are used as atlases, and the remaining 43 are used as subjects. For every subject, we generate four lesion memberships (one for each of the atlases) and average them for robustness. Fig. 2 shows a diagram of the methodology having a subject, two out of four atlases, corresponding fuzzy lesion memberships and the average membership based on four atlases. The average membership is then thresholded to obtain a hard segmentation. Estimation of the threshold is explained later. Since there is no spatial information while choosing patches from the atlas, there are false positives occurring mostly in the anterior part of the brain, cortex and cerebellum, where the FLAIR scans contain some hyper-intensities as image artifacts. As we are interested only in white matter lesions, the memberships are corrected by masking with WM masks obtained from TOADS.¹⁹ An example is shown in Fig. 3, where a subject MPRAGE and FLAIR are shown in the top row. The corresponding lesion membership (Fig. 3 (e)) is corrected by the WM mask (Fig. 3 (d)) to obtain the hard segmentation (Fig. 3 (f)). The manual segmentation is also shown in Fig. 3 (c).

The hard segmentation threshold is chosen based on a leave-one-out cross-validation on the four atlases. For every atlas image, the same methodology is applied to generate three memberships, using the remaining three atlas images. Figs. 4 (a) and Figs. 4 (b) show MPRAGE and FLAIR scans from an atlas, with its three memberships obtained using three other atlas images. The corresponding manual segmentation is also shown in Fig. 4 (c). Average Dice coefficient (averaged over four atlases, each having three memberships) between the manual segmentation (e.g., Fig. 4(c)) and the hard segmentation (e.g., Fig. 4(g)) are plotted against the threshold in Fig. 5 (a), which shows a threshold of 0.3 to be optimal. We used this threshold on the remaining 43 subjects.

Instead of using one atlas, memberships are averaged over four atlases to improve robustness in the final segmentation. Fig. 5 (b) shows the Dice coefficients obtained from each of the individual atlases, and the one from their average. The subjects are sorted in ascending order

based on their Dice from the average membership. After averaging, Dice coefficients improve significantly ($p < 0.05$). Dice numbers are prone to vary widely if the target structure size (i.e., lesions) is small.

Figs. 6 (a) and Figs. 6 (b) shows the plot of manually segmented lesion volumes vs. volumes obtained from our method as well as from Lesion-TOADS for the 43 subjects. A solid blue line represents a linear fit of the points and a black line represents the line of unit slope. Ideally, one would expect for all the points to lie on the black line. The linear fit in our case, is much closer to the unit-slope line (slope = 0.95) than Lesion-TOADS (slope = 0.47), indicating that the lesion volumes derived using our method are significantly closer to the truth. R^2 for the linear fit is 0.85 for our method compared to 0.44 for Lesion-TOADS, indicating less variability of the lesion volumes. Fig. 7 shows MPRAGE, FLAIR, Lesion-TOADS, and manual delineations of lesions as well as our segmentation on four subjects. When the lesion load is high (Fig. 7 bottom row), Lesion-TOADS performs comparably with our example based method. But when the lesion load is small (Fig. 7 top three rows), Lesion-TOADS tends to over-estimate the lesions, also evident from Fig. 6 (b). We hypothesize that Lesion-TOADS overestimates lesion load in these cases because the model-based method tries to find a single threshold from the joint intensity distribution of MPRAGE and FLAIR to identify lesions. In a model-based method, if the intensity of a voxel is below a certain threshold in FLAIR, it cannot be segmented as a lesion. In contrast, our example based method and others^{6,10} look for similar patches (incorporating the neighborhood information via 3D patches as well) from the dictionary and fuse the labels as delineated by human raters. Thus even if there is a patch with intensities below such a threshold (obtained from a model-based method), it can have high lesion membership, if there are similar examples in the dictionary.

Significant improvement is obtained for Dice, shown in Fig. 6 (c). The median Dice coefficient increases from 0.373 to 0.431 and median sensitivity increases from 0.464 to 0.483. A non-parametric test reveals that the Dice increases significantly ($p < 0.05$). The sensitivity and specificity is statistically similar ($p = 0.45$ and $p = 0.12$). Note that the average Dice score is biased by subjects that have very low lesion volume and thus the Dice score becomes very low for any false positive.

4. DISCUSSION

In this paper, we present a non-parametric example based method to segment lesions using MPRAGE and FLAIR scans. We do not assume any intensity models and rely only on patch intensity information from atlases and transfer segmentation labels based on similarity between atlas and subject patches. Although we used only T_1 -w MPRAGE and FLAIR scans to segment lesions, it can be easily extended to include T_2 -w or proton-density images.

The primary advantage of our non-parametric method compared to many model-based methods (e.g., Lesion-TOADS) is that it is not necessary to determine the amount of lesion load a-priori. In model-based methods, the amount of outliers in joint intensity distribution (of MPRAGE and FLAIR) is often governed by a user input. Since LesionTOADS is run on all the subjects with same parameter for the amount of outlier, the model forces over-

estimation of lesions when they are few, thus artificially increasing the lesion-load when the true lesion-load is small. Also our method is more resistant to imaging artifacts, shown in Fig. 7 top row. The FLAIR scan contains some imaging artifacts (red arrow), which is correctly classified as non-lesion voxels in our example based method, while being classified as lesions in LesionTOADS.

Acknowledgments

The data were acquired under grants from the National MS Society and EMD Serono to Dr. Peter Calabresi. This research was supported in part by the Intramural Research Program of the National Institute of Neurological Disorders and Stroke (NINDS) and NIH/NIBIB under grants 1R01EB017743 and 1R21EB012765.

References

1. Fisher E, Lee J-C, Rudick KNRA. Gray matter atrophy in multiple sclerosis: a longitudinal study. *Annals Neurol*. 2008; 64(3):255–265.
2. Leemput KV, Maes F, Vandermeulen D, Colchester A, Suetens P. Automated Segmentation of Multiple Sclerosis Lesions by Model Outlier Detection. *IEEE Trans on Med Imag*. 2001; 20(8): 677–688.
3. Ait-Ali LS, Prima S, Hellier P, Carsin B, Edan G, Barillot C. STREM: A Robust Multidimensional Parametric Method to Segment MS Lesions in MRI. *Med Image Comp and Comp Asst Intervention (MICCAI)*. 2005:409–416.
4. Shiee N, Bazin P-L, Ozturk A, Reich DS, Calabresi PA, Pham DL. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. *NeuroImage*. 2010; 49(2):1524–1535. [PubMed: 19766196]
5. Sweeney EM, Shinohara RT, Shiee N, Mateen FJ, Chudgar AA, Cuzzocreo JL, Calabresi PA, Pham DL, Reich DS, Crainiceanu CM. OASIS is Automated Statistical Inference for Segmentation, with applications to multiple sclerosis lesion segmentation in MRI. *NeuroImage Clinical*. 2013; 2:402–413. [PubMed: 24179794]
6. Coupé P, Manjon JV, Pruessner J, Robles M, Fonov V, Collins DL. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage*. 2011; 54(2): 940–954. [PubMed: 20851199]
7. Jog A, Roy S, Prince JL, Carass A. MR Brain Segmentation using Decision Trees. *MICCAI Grand Challenge on MR Brain Image Segmentation*. 2013
8. Lao Z, Shen D, Liu D, Jawad AF, Melhem ER, Launer LJ, Bryan RN, Davatzikos C. Computer-Assisted Segmentation of White Matter Lesions in 3D MR Images Using Support Vector Machine. *Academic Radiology*. 2008; 15(3):300–313. [PubMed: 18280928]
9. Eskildsen SF, Coupé P, Fonov V, Manjon JV, Leung KK, Guizard N, Wassef SN, Ostergaard LR, Collins DL. and The Alzheimer's Disease Neuroimaging Initiative. BEaST: Brain extraction based on nonlocal segmentation technique. *NeuroImage*. 2012; 59(3):2362–2373. [PubMed: 21945694]
10. Roy S, Carass A, Prince JL. Magnetic Resonance Image Example-Based Contrast Synthesis. *IEEE Trans on Med Imag*. 2013; 32(12):2348–2363.
11. Roy S, Carass A, Prince JL. A Compressed Sensing Approach For MR Tissue Contrast Synthesis. *Inf Proc in Med Imaging (IPMI)*. 2011:371–383.
12. Hertzmann, A.; Jacobs, CE.; Oliver, N.; Curless, B.; Salesin, DH. Image analogies. *Proc. of 28th Ann. Conf. on Comp. Graphics and Interactive Techniques SIGGRAPH '01*; 2001. p. 327-340.
13. Roy S, Carass A, Shiee N, Pham DL, Prince JL. MR contrast synthesis for lesion segmentation. *IEEE Intl Symp on Biomed Imaging*. 2010:932–935.
14. Ye DH, Zikic D, Glocker B, Criminisi A, Konukoglu E. Modality Propagation: Coherent Synthesis of Subject-Specific Scans with Data-Driven Regularization. *Med Image Comp and Comp Asst Intervention (MICCAI)*. 2013:606–613.
15. Wang H, Yushkevich PA. Multi-atlas Segmentation without Registration: A Supervoxel-Based Approach. *Med Image Comp and Comp Asst Intervention (MICCAI)*. 2013; 8151:535–542.

16. Christensen JD. Normalization of brain magnetic resonance images using histogram even-order derivative analysis. *Magn Resonance Imaging*. 2003; 21(7):817–820.
17. Mairal, J.; Bach, F.; Ponce, J.; Sapiro, G.; Zisserman, A. Non-local sparse models for image restoration. *IEEE Intl. Conf. on Comp. Vision*; 2009. p. 2272-2279.
18. Yang J, Wright J, Huang TS, Ma Y. Image Super-Resolution Via Sparse Representation. *IEEE Trans Imag Proc*. Nov.2010 19:2861–2873.
19. Bazin PL, Pham DL. Topology-preserving tissue classification of magnetic resonance brain images. *IEEE Trans on Medical Imaging*. Apr.2007 26:487–496. [PubMed: 17427736]
20. Donoho DL, Tsaig Y, Drori I, Starck J-L. Sparse Solution of Underdetermined Systems of Linear Equations by Stagewise Orthogonal Matching Pursuit. *IEEE Trans on Inf Theory*. 2012; 58(2): 1094–1121.

Figure 1.

A T_1 -w MPRAGE, FLAIR, and manual lesion segmentation are shown. Lesions are usually hypo-intense in MPRAGE, resembling CSF. They are hyper-intense in FLAIR, which is the preferred modality for WM lesion segmentation.

Figure 2.

Flow diagram of the method: the left column shows subject MPRAGE and FLAIR scans and the middle columns show two out of four atlases (MPRAGE, FLAIR, and binary lesion segmentation). Columns on the right show memberships generated from each of the atlases and their average, as well as the LesionTOADS and manual segmentation.

Figure 3.

Top row shows a subject MPRAGE, FLAIR, and manual segmentation images. Bottom row shows the corresponding WM mask (obtained from TOADS¹⁹), an uncorrected membership, and a hard segmentation corrected by masking the membership.

Figure 4.

Shown are a subject's (a) MPRAGE, (b) FLAIR, and (c) manual segmentation, the (d–f) memberships generated using the three other atlases, and the (g) thresholded hard segmentation.

Figure 5.

(a) Average Dice coefficient from a leave-one-out cross-validation of four atlas images are plotted against membership threshold. (b) Dice coefficients using the optimal threshold (0.3) are plotted for 43 subjects (red line), along with individual Dice coefficients from each of the four memberships.

Figure 6.

A plot of manually segmented lesion volumes vs. **(a)** the volumes obtained from our example based method and **(b)** Lesion-TOADS are shown for 43 subjects. Dice coefficient, sensitivity and specificity compared with manual segmentations are shown in **(c)**–**(e)**.

Figure 7.

T_1 -w MPRAGE, FLAIR, LesionTOADS, manual and our lesion segmentations are shown for four subjects.