MS Lesion Segmentation
Using FLAIR MRI Only

Jesse Knight, April Khademi

Image Analysis in Medicine Lab, School of Engineering, University of Guelph
`jknigh04@uoguelph.ca`

Abstract. There have been many efforts to automate segmentation of MS lesions in brain MRI, since human delineation is time consuming and error prone. However, most existing methods require multiple coregistered MR sequences, tissue priors or parametric models, and are rarely validated on multi-scanner image databases. In this work, a fast, FLAIR-only lesion segmentation algorithm is proposed, that does not use tissue priors or parametric models. The method uses an edge-based model of partial volume averaging to estimate fuzzy membership profiles of tissue classes. Results are further refined using an upstream image standardization pipeline, and downstream post processing. Lesion segmentation performance is measured on 15 volumes from three different scanners, demonstrating the robustness of the approach.

1 Introduction

Multiple Sclerosis (MS) is a demyelinating autoimmune disease affecting the central nervous system. MR imaging of MS lesions plays an integral role in disease diagnosis, monitoring, and research [1]. Since manual segmentation of MS lesions is subjective, unreliable, and time consuming, there has been significant interest in automating this task [2]. Existing approaches have performed classification based on voxel values from multiple coregistered MR sequences, neighborhood voxel values, normalized coordinates, and tissue prior images [2].

Despite the large number of proposed methods, very few have been validated on images from different scanners, which reflects the challenging variability in clinical images [2]. For instance, the works by Souplet et al. [3], Garcia-Lorenzo et al. [4], and Wang et al. [5] fit parametric models to the graylevel distributions of each tissue class, and use images from a single scanner for validation. However, it has been shown that graylevel distributions are highly specific to the acquiring scanner [6]. Similarly, in the works by Wu et al. [7], Steenwijk et al. [8], and Samaille et al. [9], K-Nearest Neighbour classification is used with features from multiple MR sequences; yet without normalization of graylevels across scanners, these approaches are unlikely to generalize. There are also additional challenges associated with using multiple MRI modalities or tissue priors, including the need for accurate registration, additional scan time, limited retrospective image availability, and increased segmentation model complexity.

Here we present an edge-based segmentation method, based on previous work [6, 10] and apply it to MS lesion segmentation. Our method overcomes many of the challenges outlined above, since it does not rely on parametric graylevel distribution models, it employs only the most lesion-discriminative MRI sequence, Fluid Attenuation Inversion Recover (FLAIR), and it does not require registration to any additional templates or tissue priors. We also demonstrate reliability of the method on images from three different MR scanners.

2   Methods

2.1   Image Data & Manual Segmentations

This work uses the 15 raw FLAIR images provided in the unprocessed training dataset of the MICCAI 2016 MSSEG Challenge; the brain is isolated in each image, however, using the brain masks provided in the pre-processed dataset. The binary consensus segmentations also provided, being the fusion of 7 expert manual tracings using the LOP-STAPLE method [11], are used as the ground truth for performance evaluation.

2.2   Edge-Based Lesion Segmentation

The proposed image processing pipeline is summarized in Fig. 1. The method takes a single FLAIR image, denoted $Y(x)$ in 3D space, $x = \{x_1, x_2, x_3\} \in \mathbb{Z}^3$. The pipeline consists of image preprocessing, an edge-based fuzzy classification model, thresholding, and post processing. Output lesion masks are then compared to the consensus segmentations using quantitative performance measures.
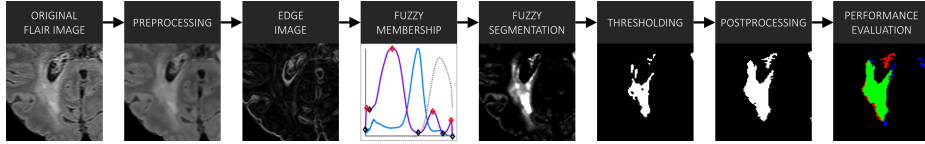

Fig. 1. Image processing pipeine

Preprocessing Preprocessing is employed to correct intensity inhomogeneity and reduce random noise in the source FLAIR images. First, intensity inhomogeneity (bias field) is corrected using the Segment feature in the SPM12 toolbox [12]. While this tool also produces tissue probability images, these are not needed for lesion segmentation and are discarded. A 3D Gaussian low pass filter, with $\sigma = 0.5$ mm, is then used to minimize image noise; it was found that this filter outperformed median, anisotropic diffusion, and bilateral filters.

Fuzzy Classification Next, we perform fuzzy classification using an edge-based model [6, 10]. This model assumes that the graylevels of each tissue class (CSF,

Brain, Lesion) are distributed along a unique range, and that graylevels in between these ranges represent voxels subject to partial volume averaging (PVA, a mixture of two tissue types). For a PVA-affected voxel, comprising a mixture of classes c_j and c_k, with graylevels y_j and y_k, respectively, and where α_{jk} indicates the mixing proportion of class c_j, the graylevel $y_{jk}(x)$ is modelled as [13],

$$y_{jk}(x) = \alpha_{jk}(x) \times y_j(x) + (1 - \alpha_{jk}(x)) \times y_k(x)$$

We take the gradient of this equation to solve for $\alpha_{jk}(x)$, in order to quantify the exact amount of each tissue present in each voxel. However, we observe that the solution in the spatial domain is intractable, since the integration bounds x_0 and x_N are undefined [10],

$$\alpha'_{jk}(x) = \frac{y'_{jk}(x)}{y_j - y_k} \quad \rightarrow \quad \alpha_{jk}(x) = \frac{1}{y_j - y_k} \int_{x_0}^{x_N} y'_{jk}(x) dx$$

To circumvent this, all voxels in the image volume are considered simultaneously in the graylevel (global) domain: $\alpha'(x) \rightarrow \alpha'(y)$, where the integration becomes feasible. The global estimate of $\alpha'(y)$ is computed using the conditional expectation operator, given an initial spatial estimate of edge strength, which is proportional to $\alpha'(x)$.

First, the gradient magnitude image, $G(x)$, is computed using the centered difference kernel, $D = [-1, 0, +1]$, applied in all three dimensions.

$$G(x) = |\nabla_x Y(x)|$$
$$= \left[\left(\frac{dY(x)}{dx_1} \right)^2 + \left(\frac{dY(x)}{dx_2} \right)^2 + \left(\frac{dY(x)}{dx_3} \right)^2 \right]^{\frac{1}{2}}$$

For a robust estimate of the spatial domain edge strength, histogram equalization is performed to ensure that edge magnitudes are distributed consistently and without outliers, yielding $\alpha'(x)$,

$$\alpha'(x) = CDF_G(G(x))$$

Next, the distribution of edge strength in the graylevel domain, $\alpha'(y)$, is estimated as the expected value of edge in the image for each graylevel. The expectation is computed robustly using a 2-bin Gaussian kernel density estimate with bins $p(\alpha'(x) = 1 \mid y)$ and $p(\alpha'(x) = 0 \mid y)$, and $\sigma = 1/6$ (see [10] for derivation),

$$\alpha'(y) = \mathbb{E}\{\alpha'(x) \mid y\}$$
$$= \sum_{a \in 0,1} a \cdot p(\alpha'(x) = a \mid y)$$
$$= p(\alpha'(x) = 1 \mid y)$$

Postprocessing of $\alpha'(y)$ consists of smoothing, and trimming to remove the top 0.02% of brain voxels. This reliably yields a distribution like that shown in Fig.

2, top, dark purple, in which local minima (black diamonds) represent graylevels corresponding to pure tissue classes — minimal edge information — and maxima (red diamonds) represent class transitions — edgy regions. Before integration to solve for $\alpha(y)$, we refine: $\alpha'(y) \rightarrow \tilde{\alpha}'(y)$ to yield a more robust estimate of pure tissue and PVA ranges: extrema are identified using peak detection, and locally normalized to $\{0,1\}$, and the magnitude is squared, as shown in Fig. 2, top, light purple. This $\tilde{\alpha}'(y)$ is then used to find the partial volume fraction $\alpha(y)$
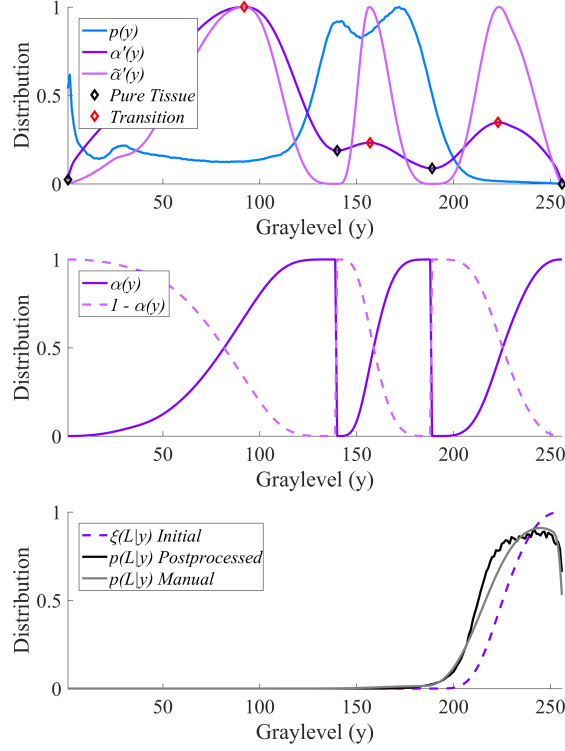

Fig. 2. Graylevel distributions for one FLAIR volume; top - histogram, edge distribution, and extrema; middle - partial volume fraction from piecewise integration of the edge distribution; bottom - distributions of lesion: initial fuzzy estimate, postprocessed, and manual estimate.

for each pair of adjacent tissue classes, since the desired integration can now be performed between pure tissue graylevels y_j, y_k (local minima),

$$\alpha_{jk}(y) = \frac{\sum_{y_k}^{y} \tilde{\alpha}'(y)dy}{\sum_{y_k}^{y_j} \tilde{\alpha}'(y)dy}, \quad y \in [y_k, y_j]$$

Finally, the partial volume fraction profiles for each pair of adjacent tissue classes, $\alpha_{jk}(y)$ (Fig. 2, middle) are combined in smooth piecewise segments to give the fuzzy membership function for each tissue class, $\xi(c_k \mid y)$ (e.g. Fig. 2, bottom),

$$\xi(c_k \mid y) = \begin{cases} 0, & y \in [0, y_j) \\ \alpha_{jk}(y), & y \in [y_j, y_k) \\ 1, & y = y_k \\ 1 - \alpha_{jk}(y), & y \in (y_k, y_\ell] \\ 0, & y \in (y_\ell, y_{max}] \end{cases}$$

This smooth characterization of $\xi(c_k \mid y)$ differs from previous works [6, 10], where graylevel ranges corresponding to pure tissues and PVA were estimated using piecewise *linear* functions. This new formulation ensures that graylevels with minimum edge correspond to pure classes, and that maximum fuzzy membership slope (class transition) occurs at the maximum edge content $\alpha(y)$ between pure tissue classes, reflecting the edge-based estimate of PVA distribution.

Post Processing The initial fuzzy classification is thresholded at a level $\tau \in \{0, 1\}$ to give a binary segmentation image, denoted A. Three false positive reduction (FPR) strategies are then appended to the pipeline to refine A based on prior knowledge. The FPR exclusion criteria include:

- Minimum lesion volume (using 6-connectedness, MLV, mm^3)
- Minimum distance from the brain edge (3D, DBE, mm)
- Minimum distance from the brain midline (2D, DBM, mm)

False negatives are also reduced using region growing (RG), in an attempt to mimic the experts' inclusion of marginally hyperintense white matter (also called dirty-appearing white matter) which surrounds lesion cores. During each of N_{RG} region growing iterations, each neighboring voxel at location x is added to the segmentation A if the graylevel is acceptably close to an adaptive threshold: the qth quantile of the graylevels in the original segmentation:

$$\mathbf{if}: \left| y(x^*) - Q_{Y(x \in A)}(q) \right| < \Psi, \quad \mathbf{then}: A \leftarrow A \cup x^*$$

The fuzzy threshold τ, FPR constraints, and RG parameters were all optimized using a semi-guided simplex gradient ascent with respect to group DSC.

2.3   Performance Analysis

Performance of the algorithm for lesion segmentation is evaluated using the overlap metrics recommended by the MSSEG Challenge. These compare automated segmentation image $A(x)$ to the ground truth segmentation image $T(x)$:

- Dice Similarity Coefficient: $2\frac{A \cap T}{A + T}$
- Positive Predictive Value: $\frac{A \cap T}{A}$
- True Positive Rate: $\frac{A \cap T}{T}$

Additionally, volume agreement with expert lesion load (LL) is illustrated using a Bland-Altman plot, and analysis of covariance (ANCOVA) is used to test for significant impacts of scanner and LL on DSC.

3   Results

This section presents segmentation performance data, following initial fuzzy segmentation, and postprocessing using the optimized parameters, as well as factors influencing performance. The mean DSC for all 15 training cases was 0.60 (Table 1), but for the 10 cases with LL > 5 ml, it increased to 0.70. Accordingly, analysis of covariance revealed that DSC was significantly correlated with LL, independent of scanner ($p = 0.013$), while performance was not significantly impacted by scanner when controlling for LL ($p = 0.354$).

Table 1. Performance metrics by case

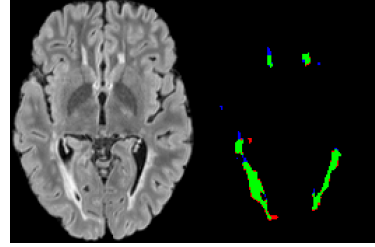| LL (ml)	Case #	DSC	PPV	TPR
72	1	0.81	0.80	0.82
49	7	0.78	0.84	0.74
43	3	0.78	0.82	0.75
39	13	0.61	0.83	0.49
30	5	0.66	0.68	0.66
29	15	0.71	0.85	0.62
22	2	0.63	0.72	0.56
13	10	0.64	0.99	0.48
12	11	0.69	0.88	0.57
6	14	0.68	0.92	0.54
5	9	0.26	1.00	0.15
4	8	0.40	0.92	0.26
2	12	0.32	0.25	0.44
2	6	0.34	0.98	0.21
1	4	0.61	0.55	0.69
Mean		0.60	0.80	0.53
Mean (LL>5)		0.70	0.83	0.62

Fig. 3. Sample segmentation; green: true positives, blue: false negatives, red: false positives

An example segmentation is shown in Fig. 3, where it can be seen that false positives arise mostly along the lesion borders, while false negatives are attributable to small, ambiguous lesions near the gray matter. The Bland Altman plots (Fig. 4), demonstrate a potential bias of the method towards underestimation of LL (mean difference of 3.6 ml), and a small positive correlation between LL and undersegmentation (slope of 0.97). This result is corroborated by higher mean PPV (0.80) than TPR (0.53), over all cases, indicating higher specificity than sensitivity.

4   Discussion

The edge-based segmentation algorithm performs well on the training data provided, given the large intensity inhomogeneity and low resolution in the MRI volumes. We observed that the ground truth distribution of lesion versus graylevel,
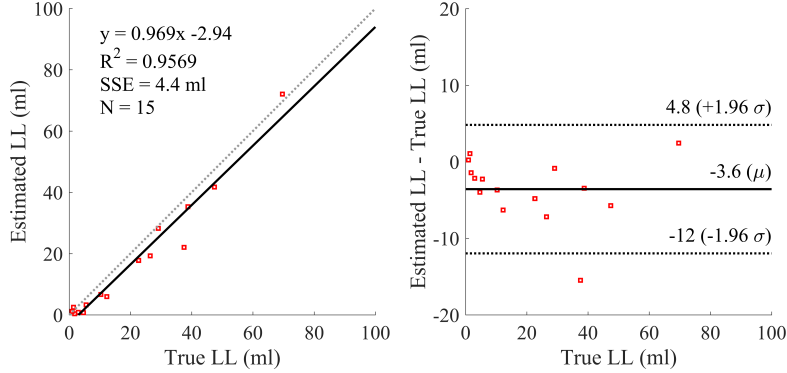
Fig. 4. Bland-Altman plots showing volume agreement with expert segmentations

$p(L \mid y)$ (Fig. 2, bottom), using the binary consensus segmentations, rarely reaches a value of 1, and has long, gradual slopes towards the maximum. This likely arises from the ambiguous segmentation of so-called dirty appearing white matter, and it means that even optimal thresholding results would incur significant false positive and false negative fractions. Pre- and post-processing techniques both help to reduce these effects, and it is worth noting that our preprocessing pipeline resulted in better separability of classes by graylevel than the preprocessed images provided for the competition.

As with most MS lesion segmentation algorithms, the proposed method is challenged by small lesion loads. However, performance was not impacted by scanner or the associated voxel resolution, demonstrating the robustness of this efficient nonparametric approach to modeling FLAIR image variability. Many of the false positives occur either in the gray matter (due to bias field) or as CSF pulsation artifacts, so future works could investigate the use of tissue priors, improved inhomogeneity correction, or multiscale estimation of edge content. However, in hoping to avoid expensive and error-prone image registrations in our approach, we are investigating primarily multiscale or context-based edge characterization, and an iterative bias-field correction with lesion masking.

5   Conclusion

In conclusion, we present a fast, robust segmentation algorithm capable of delineating MS lesions in FLAIR MRI alone. The method achieves good performance (DSC = 0.60) on a challenging database of 15 images from 3 different scanners, and even better performance (DSC = 0.70) on lesion loads greater than 5 ml. Unlike many segmentation tools, the algorithm does not rely on parametric models of tissue class graylevel distributions, and consequently demonstrates no change in performance with different scanners. Additionally, the method requires only one MRI sequence, FLAIR, and no additional tissue priors or training data, thereby avoiding additional MR scan time, limited image availability, and the need for expensive and imperfect registration.

References

1. Polman, C.H., Reingold, S.C., Banwell, B., Clanet, M., Cohen, J.A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., Lublin, F.D., Montalban, X., O'Connor, P., Sandberg-Wollheim, M., Thompson, A.J., Waubant, E., Weinshenker, B., Wolinsky, J.S.: Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Annals of Neurology **69**(2) (2011) 292–302
2. Garcia-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D.L., Collins, D.L.: Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. Medical Image Analysis **17**(1) (2013) 1–18
3. Souplet, J., Lebrun, C., Ayache, N., Malandain, G.: An Automatic Segmentation of T2-FLAIR Multiple Sclerosis Lesions. MICCAI Grand Challenge Workshop: Multiple Sclerosis Lesion Segmentation Challenge (2008) 1–11
4. García-Lorenzo, D., Prima, S., Arnold, D.L., Collins, D.L., Barillot, C.: Trimmed-likelihood estimation for focal lesions and tissue segmentation in multisequence MRI for multiple sclerosis. IEEE Transactions on Medical Imaging **30**(8) (2011) 1455–67
5. Wang, R., Li, C., Wang, J., Wei, X., Li, Y., Zhu, Y., Zhang, S.: Automatic segmentation and volumetric quantification of white matter hyperintensities on fluid-attenuated inversion recovery images using the extreme value distribution. Neuroradiology **57**(3) (2015) 307–320
6. Khademi, A., Venetsanopoulos, A., Moody, A.R.: Robust white matter lesion segmentation in FLAIR MRI. IEEE Transactions on Bio-Medical Engineering **59**(3) (2012) 860–871
7. Wu, Y., Warfield, S.K., Tan, I.L., Wells, W.M., Meier, D.S., van Schijndel, R.A., Barkhof, F., Guttmann, C.R.G.: Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI. NeuroImage **32**(3) (2006) 1205–1215
8. Steenwijk, M.D., Pouwels, P.J.W., Daams, M., van Dalen, J.W., Caan, M.W.A., Richard, E., Barkhof, F., Vrenken, H.: Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs). NeuroImage. Clinical **3** (2013) 462–9
9. Samaille, T., Fillon, L., Cuingnet, R., Jouvent, E., Chabriat, H., Dormont, D., Colliot, O., Chupin, M.: Contrast-based fully automatic segmentation of white matter hyperintensities: method and validation. PloS one **7**(11) (2012) e48953
10. Khademi, A., Venetsanopoulos, A., Moody, A.R.: Generalized method for partial volume estimation and tissue segmentation in cerebral magnetic resonance images. Journal of Medical Imaging **1**(1) (2014) 14002
11. Akhondi-Asl, A., Hoyte, L., Lockhart, M.E., Warfield, S.K.: A logarithmic opinion pool based STAPLE algorithm for the fusion of segmentations with associated reliability weights. IEEE Transactions on Medical Imaging **33**(10) (2014) 1997–2009
12. Ashburner, J., Friston, K.J.: Unified segmentation. NeuroImage **26**(3) (2005) 839–851
13. Santago, P., Gage, H.D.: Quantification of MR brain images by mixture density and partial volume modeling. IEEE Transactions on Medical Imaging **12**(3) (1993) 566–574