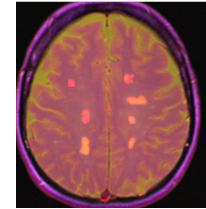


# Automated quantification of MS lesions in MRI: a validation study

Edward Ashton, Chihiro Takahashi, Michel Berg\*, Andrew Goodman\*, Saara Totterman  
Dept. of Radiology; \*Dept. of Neurology, University of Rochester Medical Center



**ABSTRACT:** Two novel methods for automated quantification of total lesion burden in multiple sclerosis patients using multi-spectral magnetic resonance imaging are examined. The first method, geometrically constrained region growth, requires user specification of lesion location. The second, directed multi-spectral segmentation, requires only the location of a single exemplar lesion. The performances of these methods are compared to manual tracing using three parameters: speed, precision, and accuracy. Both methods are shown to provide significant improvement over manual tracing in terms of processing time, inter- and intra-operator coefficients of variation, and global accuracy using both phantoms and clinical data.

**BACKGROUND:** Accurate quantification of brain lesion count and volume in multiple sclerosis patients is a vital tool for evaluation of disease progression and patient response to therapy. Currently used methods for obtaining these data points are largely manual and subjective, and are therefore both error-prone and subject to inter- and intra-operator variability. In addition, manual tracing of white matter lesions requires as much as 60 minutes of expert intervention time per case.

Other research groups, including Brunetti *et al.*<sup>1</sup>, Rovaris *et al.*<sup>2</sup>, Dastidar *et al.*<sup>3</sup>, and Ballester *et al.*<sup>4</sup>, and Van Leemput *et al.*<sup>5</sup> have developed automated or semi-automated approaches to this problem. The results reported in these studies serve as useful benchmarks for the work presented here.

**MEASUREMENT TECHNIQUES:** One manual and two automated measurement techniques were evaluated in this study. All measurements were carried out using software provided by VirtualScopics, a biomedical image analysis firm based in Rochester, NY.

The first automated method examined was geometrically constrained region growth (GEORG)<sup>6,7</sup>. This technique requires a user to place a "seed" within each lesion in the volume using a single mouse click. The seed region then expands into neighboring voxels provided that two constraints are satisfied: the spectral signature of the neighboring voxel must have a high probability of falling within the statistical distribution defined by all current included voxels, and inclusion of the neighboring voxel must not cause the shape of the included region to deviate excessively from the *a priori* regional shape model. The operation of this algorithm is illustrated in Fig. 1.

The second automated method, directed multi-spectral segmentation (DMSS), requires a user to identify one lesion within the volume. In this study this exemplar lesion was identified using GEORG, reducing user interaction to a single mouse click. Given an exemplar, DMSS then identifies the statistical characteristics of the apparent normal tissue using an adaptive multivariate Bayesian classifier<sup>8,9</sup>. Finally, an iterated conditional modes (ICM)<sup>10</sup>-based classification pass is used, using the background statistics supplied by the Bayesian classifier and the target statistics supplied by the exemplar. The operation of this algorithm is illustrated in Fig. 2.

**EXPERIMENTAL PROCEDURE:** The experiments involved in this study were intended to assess the performance of the two algorithms under consideration with respect to manual tracing and to the studies cited above in terms of three parameters: speed, precision, and accuracy. Two data sets, one clinical and one synthetic, were used to assess required processing time for each algorithm, inter- and intra-operator measurement variability, and global accuracy.

The multispectral MRI data sets shown in Fig. 3 were used to evaluate required operator time and reproducibility for both manual and automated processing, as well as to establish correlation between manual and automated results. First, each of the three clinical baseline studies was evaluated using manual tracing, GEORG and DMSS. Ten separate measurements of lesion burden in each data set were made by a single operator using both GEORG and DMSS. In addition, ten manual measurements of each data set were made by a trained neuroradiologist. Each of the six revisit scans was then evaluated using manual tracing as well as both automated algorithms. These experiments served to establish intra-operator variability for all three processes.

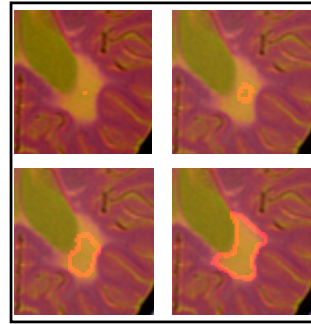


Fig. 1: Identification of lesion boundaries using GEORG, from initial seed to stable solution.

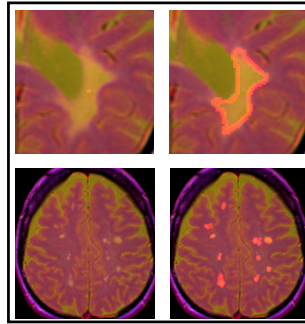


Fig. 2: (Top) Identification of an exemplar lesion on slice 16 using GEORG. (Bottom) Slice 21 before and after DMSS.

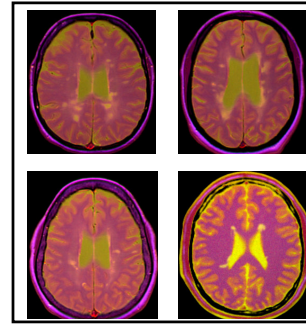


Fig. 3: Single slices from baseline studies for three clinical data sets and (bottom right) one phantom data set obtained from MBIC. Each data set consists of T1, T2, and proton density weighted studies.

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	Mean	S.D.	C.V.
<b>Patient 1</b>													
Manual	20.0	21.2	20.8	20.6	20.2	19.1	21.0	20.4	19.2	19.2	20.3	0.7	3.4%
GEORG	19.5	19.5	19.6	19.7	19.3	19.1	19.3	19.2	19.2	19.5	19.4	0.2	1.0%
DMSS	19.4	19.5	19.4	19.4	19.4	20.4	19.4	19.4	19.5	19.4	19.5	0.3	1.5%
<b>Patient 2</b>													
Manual	26.8	26.5	22.5	23.1	24.3	24.1	26.0	26.8	24.9	27.7	25.3	1.7	6.7%
GEORG	22.1	21.9	22.0	22.1	21.9	21.8	21.7	21.7	21.7	21.8	21.9	0.2	0.9%
DMSS	21.4	22.2	22.0	22.0	22.2	22.3	22.1	22.1	22.2	22.2	22.1	0.2	0.9%
<b>Patient 3</b>													
Manual	9.6	10.5	10.6	9.2	10.4	10.4	10.1	8.0	10.1	8.9	9.8	0.8	8.2%
GEORG	8.5	8.5	8.3	8.3	8.3	8.0	8.0	8.0	8.0	8.1	8.2	0.2	2.4%
DMSS	9.7	9.8	9.8	9.7	9.4	9.8	9.7	9.8	9.7	9.2	9.7	0.2	2.1%

Table 1: Results of intra-operator variability study 1. This table shows results for 10 measurements of each patient's total lesion burden in cubic centimeters, along with mean value, standard deviation, and coefficient of variation for each measurement technique.

	Rep. 1	Rep. 2	Rep. 3	Rep. 4
<b>Patient 1</b>				
Manual	20.0	29.3	34.9	N/A
GEORG	19.4	19.4	20.2	N/A
DMSS	19.4	19.6	19.1	N/A
<b>Patient 2</b>				
Manual	26.8	34.8	N/A	N/A
GEORG	22.1	20.7	N/A	N/A
DMSS	21.4	20.0	N/A	N/A
<b>Patient 3</b>				
Manual	9.6	9.8	8.4	9.4
GEORG	8.5	8.0	8.2	7.6
DMSS	9.7	10.5	11.3	8.1

Table 2: Results of repeated scan study. This data also shows good agreement between measurements of studies temporally separated by several days. Note also that the variability of the manual measurements has increased greatly with respect to that of the automated measurements.

	Noise 3 Dist. 0	Noise 3 Dist. 20	Noise 5 Dist. 20
Manual	10.2	8.5	8.7
GEORG	3.2	3.2	3.1
DMSS	4.5	3.4	4.1
Ground Truth	3.5	3.5	3.5

Table 4: Results of phantom experiments. Note that in addition to showing significantly better precision, automated results are much closer to ground truth than those obtained via manual tracing. Poor manual results are primarily due to the large number of very small and difficult to trace lesions present in this data set.

	Obs. 1	Obs. 2	Obs. 3	Obs. 4	Mean	S.D.	C.V.
<b>Patient 1</b>							
Manual	27.0	18.8	20.6	18.4	21.2	4.0	18.9%
GEORG	19.5	19.6	19.7	19.3	19.5	0.2	1.0%
DMSS	20.0	22.6	19.4	20.0	20.5	1.4	6.8%
<b>Patient 2</b>							
Manual	22.8	25.3	27.3	21.0	24.1	2.8	11.5%
GEORG	22.1	22.2	21.8	21.8	22.0	0.2	0.9%
DMSS	20.9	22.5	21.4	19.5	21.1	1.2	5.9%
<b>Patient 3</b>							
Manual	9.9	9.7	10.8	6.7	9.3	1.8	19.2%
GEORG	8.5	8.4	7.7	8.1	8.2	0.4	4.9%
DMSS	9.7	10.1	9.7	9.4	9.7	0.3	3%

Table 3: Results of inter-operator variability trials. As expected, the automated methods show good correlation with each other and with manual results. The automated methods also show significantly lower coefficients of variation.

In order to establish inter-operator variability, each baseline scan was then evaluated by four trained operators using GEORG and DMSS, and by four expert observers using manual tracing. Mean results for all three techniques were compared, in order to establish the level of correlation between manual and automated measurements.

Phantom data sets obtained from MBIC<sup>6,7,8,9</sup> were used to determine some estimate of the global accuracy of both manual tracing and the automated measurement methods. Because these data sets contained various levels of noise and distortion they also allowed an examination of the sensitivity of both manual and automated measurement to these parameters. The results of these measurements were compared to the provided ground truth map in order to estimate global accuracy. Also, results were examined relative to one another in order to determine sensitivity to noise and distortion in the data.

**RESULTS:** Results of the first intra-operator variability trials, measuring precision over repeated measurements of the same data sets, are given in Table 1. These results indicate that both automated techniques provide substantial improvement over manual tracing in terms of precision. In addition, they show significant correlation between automated and manual results, and between the two automated techniques.

Our second intra-operator variability study involved measurement of lesion burden in repeat scans taken of each of the three patients over a period of 4 – 8 days. The presumption is that total lesion burden should not have changed significantly over this time period. Results of this experiment are given in Table 2.

Inter-operator variability was evaluated by measuring lesion burden in each of the initial scans using four different operators for each algorithm. Manual tracing was carried out by qualified experts. Automated processing was carried out by trained observers. Results of this experiment are given in Table 3.

Our final experiment involved data taken from the MBIC brain database. This experiment was designed to assess global accuracy, as the MBIC data was thoroughly ground-truthed. Experimental results were generated for data sets with noise levels of 3% and 5%, and distortion levels of 0% and 20%. Results of this experiment are given in Table 4.

## REFERENCES

1. A. Brunetti, A. Lacobini, M. Quarantelli, M. Tedeschi, E. Ciancibello, A. Corvelli, and M. Salvatore, "Automated segmentation and measurement of global white matter lesion volume in patients with multiple sclerosis," *J. of MRI* 12(6), pp. 799 – 807, 2000.
2. M. Rovaris, M. Hagbin, R. van Schijndel, M. Sonmani, M. Rodegher, G. Comi, and M. Filippi, "Sensitivity and reproducibility of volume charge measurements of different brain portions on magnetic resonance imaging in patients with multiple sclerosis," *J. of Neurology* 249(2), pp. 960 – 965, 2000.
3. P. Dastidar, T. Heikonen, T. Lehtimäki, M. Ukkonen, J. Peltola, T. Ehtä, E. Laasonen, and I. Eloranta, "Volumes of brain atrophy and plaques correlated with neurological disability in secondary progressive multiple sclerosis," *J. of the Neurological Sciences* 164(1), pp. 38 – 42, 1999.
4. M. Ballester, A. Zsombok, and M. Brady, "Segmentation and measurement of brain structures in MRI including confidence bounds," *Medical Image Analysis* 4(3), pp. 189 – 200, 2000.
5. E. Van Leemput, F. Maes, D. Vandermeulen, A. Collignon, and P. Sussner, "Automated segmentation of multiple sclerosis lesions by model outlier detection," *IEEE Trans. Medical Imaging* 20(8), pp. 677 – 688, 2001.
6. C. Cozziso, V. Kolbinka, R. Kwan, and A. Evans, "BrainWeb: Online interface to a 3D MRI simulated brain database," *NeuroImage* 5(4), 1997.
7. R. Kwan, A. Evans, and G. Pike, "MRI simulation-based evaluation of image-processing and classification methods," *IEEE Trans. Medical Imaging* 18(11), pp. 1085 – 97, 1999.
8. R. Kwan, A. Evans and G. Pike, "An extensible MRI simulator for post-processing evaluation," *Visualization in Biomedical Computing '96*, pp. 135 – 140, 1996.
9. D. Collins, A. Zijdenbos, V. Kolbinka, J. Sidé, N. Kabani, C. Holmes, and A. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Trans. Medical Imaging* 17(10), pp. 463 – 468, 1998.
10. E. Ashton, S. Totterman, C. Takahashi, J. Tanna-Pena, and R. Parker, "Automated measurement of structures in CT and MR imagery: A validation study," *Proc. 14<sup>th</sup> IEEE Symposium on Computer-Based Medical Systems*, pp. 300 – 306, 2001.
11. E. Ashton, K. Parker, M. Berg, and C. Chen, "A novel volumetric feature extraction technique, with applications to MR images," *IEEE Trans. Medical Imaging* 16(4), pp. 365 – 371, 1997.
12. I. Carlbom, D. Terzopoulos, and K. Harris, "Computer assisted registration, segmentation and 3D reconstruction from images of neuronal tissue sections," *IEEE Trans. Medical Imaging* 13, pp. 351 – 362, 1994.
13. D. Taylor and W. Buxton, "Image segmentation using globally optimum growth in three dimensions with an adaptive feature set," *Visualization in Biomedical Computing*, pp. 98 – 107, 1994.
14. E. Ashton, "Detection of subpixel anomalies in multispectral infrared imagery using an adaptive Bayesian classifier," *IEEE Trans. Geoscience and Remote Sensing* 36(2), pp. 306 – 317, 1998.
15. E. Ashton, "Multithreshold solution for automated multispectral target detection," *Optical Engineering* 38(4), pp. 717 – 724, 1999.
16. J. Benag, "On the statistical analysis of dirty pictures," *J. Royal Stat. Society* 48, pp. 279 – 302, 1986.
17. C. Gutman, R. Kikinis, M. Anderson, M. Jahn, S. Warfield, R. Killany, H. Weisler, P. Jolesz, "Quantitative follow-up of patients with multiple sclerosis using MRI reproducibility," *J. of MRI* 9(4), pp. 509 – 518, 1999.