

Segmentation of Progressive Multifocal Leukoencephalopathy Lesions in FLAIR MRI

Laurent Itti, PhD (1), Linda Chang, MD (2), and Thomas Ernst, PhD (2)

University of Southern California, Los Angeles, California

Brookhaven National Laboratory, Upton, New York

Short title: PML lesion segmentation

Correspondence: Laurent Itti, University of Southern California, Department of Computer Science, Hedco Neuroscience Building, Room 30A, 3641 Watt Way, Los Angeles, CA 90089-2520, USA. Email: itti@pollux.usc.edu - WWW: <http://iLab.usc.edu> - Tel: +1 213 740-3527 - Fax: +1 213 740-5687.

Key words: White matter lesion, progressive multifocal leukoencephalopathy (PML), segmentation, MRI, FLAIR.

Acknowledgment: Support for L.C. by NIH (DA 00280).

Abstract:

Volumes of white matter lesions were determined using FLAIR MRI in 23 AIDS patients with progressive multifocal leukoencephalopathy. Manual outlining was compared to an automated method based on region growing and adaptive thresholding. Lesion volumes from the two methods correlated well (61 lesions, $r=0.99$, $p < 10^{-4}$), although the volumes differed substantially ($12.8 \pm 13.7\%$, mean \pm S.D). Interscan intrasubject reproducibility was better for the automated than the manual method ($2.9 \pm 3.2\%$ vs. $12.4 \pm 16.2\%$ volume difference, $p=0.02$).

Accurate and reproducible segmentation of white matter lesions is necessary to obtain reliable quantitative assessment of disease progression [1]. While the total volume of lesion (the lesion load) is only an approximate measure of clinical disease severity, high reproducibility in measuring the lesion load is essential for longitudinal monitoring of disease progression. In this study, we evaluate the reproducibility of segmentation of progressive multifocal leukoencephalopathy (PML) lesions in patients with AIDS. Reliable segmentation will be particularly important as pharmacological treatments become available for PML, and must be assured before the efficacy of a given treatment for reducing lesion load can be concluded. However, two major sources of quantitation error may affect reproducibility: (1) The expert’s subjectivity when interactive outlining methods are employed, and (2) technical imperfections such as poor lesion contrast, magnetic field inhomogeneities and partial volume effects [2], which vary with patient positioning.

[Figure 1 about here]

We evaluated quantitation errors from both subjective and technical sources, using the Fluid-Attenuated Inversion-Recovery (FLAIR) MR sequence for its high contrast between lesions and both normal tissue and cerebrospinal fluid (CSF) [3, 4]. Manual outlining was compared to an automated segmentation algorithm; furthermore, the reproducibility of both methods was assessed using pairs of scans acquired consecutively with different patient positions (figure 1).

Methods. *Patients and data acquisition:* Twenty-three male patients with one to four PML lesions (44 distinct lesions) were imaged on a 1.5T GE Signa MR scanner (General Electric Medical Systems, Milwaukee, WI, USA) using a FLAIR sequence (TE=140, TR=10000, TI=2200, $0.9375 \times 0.9375 \times 5$ mm³ voxels). For 9 patients (17 lesions), repeat scans were acquired within 30 minutes after patient repositioning (with orientation differences of more than 30°; figure 1). Datasets were processed on a DEC Alpha workstation (Digital Equipment Corporation, MA, USA) using customized software.

Manual Segmentation: The manual method consisted of outlining the edges of the lesions on all slices, using a mouse pointer. Dedicated software allowed an experienced operator to

draw and edit polygonal lines on the magnified FLAIR images. All of the displaying and drawing guidelines proposed by Filippi *et al.* [5] were applied.

Automated segmentation: Lesions appeared as hyperintense regions in the white matter [6, 7] surrounded by relatively uniform, lower intensity normal tissue. From a manually selected starting point (seed), a lesion was extracted by three-dimensional (3D) flooding into neighboring volume elements (voxels) with intensities higher than a given threshold t . The flooding process is a 3D extension of the “bucket paint” algorithms present in most graphics manipulation software. It consists of growing the region, from the seed point, into adjacent voxels on the same slice as well as on adjacent slices (Figure 2). The region grows only into those neighboring voxels whose intensity is above t , and the process is recursively applied until all voxels above t that are connected to the initial seed point have been flooded. Because of its 3D nature, this algorithm requires the operator to only seed each lesion on one slice where it is visible, as the algorithm will spread into adjacent slices as required.

Subjective and technical segmentation variabilities were reduced through the automatic and adaptive determination of the appropriate threshold for each lesion: Starting from the intensity at the seed point, approximated by its closest multiple of a constant a ($a = 5$ intensity units in our implementation), the threshold was progressively decreased by the discrete amount a . For every threshold value t , flooded lesion volume and its ratio r to the volume obtained with the previous threshold ($t + a$) were computed. The algorithm ceased to calculate for the next smaller t value when the ratio r exploded (i.e., r was greater than a constant $b = 6$), as flooding began to spread into normal brain tissue or large extents of the extracranial structures. The constant b was determined empirically under two competing constraints: with small values for b , the algorithm would only detect lesions with uniform intensity (which typically is not the case), while using larger values for b presents the risk that flooding would extend into normal white matter when the transition between lesion and white matter is smooth. The final threshold for segmentation was the value of t at which the explosion occurred, plus an increment d ($d = 7$ in our implementation). d determined the tolerance of the algorithm with respect to partial volume effects at lesion boundaries; with

smaller values, more voxels with partial lesion volume were included in the segmentation. The value $d = 7$ was determined empirically.

[Figure 2 about here]

By construction, this algorithm yielded exactly reproducible segmentation from a given seed point. In addition, segmentation proved largely independent of the choice of the seed location (figure 1). Although manual editing of the segmentation results was possible, it was not used in this study.

[Figure 3 about here]

Results. Manual drawing correlated well with automated extraction (figure 3.a) for all 61 lesions (17 of which were from the repeat scans with different orientations). However, lesion volumes from the two methods differed substantially (up to 59% in absolute value for the smallest lesions, and $12.8 \pm 13.7\%$ on average; figure 3.b). Better agreement was found for larger lesions; for instance, the difference was $6.8 \pm 4\%$ for lesions larger than 9cc.

The automated method proved more reproducible than the manual method with respect to different patient orientation (figure 3.c,d,e). For the 17 lesions evaluated from two consecutive scans, good correlations were found between the volumes measured in the first and second scan, both with manual drawing ($r=0.989$, $p < 10^{-4}$) and with the automated method ($r=0.999$, $p < 10^{-4}$). However, paired t-tests showed significantly different relative volume differences in the two orientations with the manual versus the automated method ($p=0.02$). This is due to the smaller volume differences for the automated method compared to manual drawing. For all 17 lesions, the volume difference was $12.4 \pm 16.2\%$ for the manual method, and only $2.9 \pm 3.2\%$ for the automated method. For the 14 larger lesions with volumes above 2cc, average and maximum volume differences were 11.6% and 47.5% for the manual method, and only 1.8% and 5.4% for the automatic method.

Discussion. We found substantial discrepancies between the manual and automated methods, especially for the smaller lesions. The major sources of discrepancies were (figure 4):

[Figure 4 about here]

a) Imaging artifacts. The FLAIR sequence presents obvious advantages over the regular

T2-weighted sequences [4, 9, 8]. However, it suffers from the presence of artifactual hyperintensities at the interface between tissues and surrounding CSF [10]. These artifacts are easily misclassified by automated methods, while human observers have less difficulty identifying them. Improved FLAIR sequences may however improve the performance of the automated method [4].

b) Uncertain three-dimensional (3D) shape coherence. Even when adjacent slices are available, human observers experience difficulties in identifying the 3D shape of a lesion. The most common manual misclassification in this study was the omission of small isolated regions, which were disconnected from the main the lesion in the slice plane, but were connected to the lesion in an adjacent slice. The 3D automated method did not suffer from this problem.

c) Shape irregularities. The manual segmentation results typically had smoother shape than the automated results. When lesion boundaries were uncertain, the observer approximated them with a straight line segment. This was particularly true for larger lesions, which required drawing of numerous long polygonal contours. While this source of subjective variability is less problematic in diseases with smaller and smooth lesions, it was more apparent with the larger, highly irregularly-shaped PML lesions. Containing no smoothness constraint, the automated algorithm yielded more objective delineation of lesion boundaries.

d) Inconsistent drawing rules. Finally, the major source of interscan variability was the inconsistency of manual drawing rules. Although the operator always tried to include the same amount of partial volume around each lesion, manually drawn outlines were more conservative in some regions than in others. It could be argued that this constitutes an advantage for the manual method, which is guided by expert knowledge of certain technical imaging irregularities. For sequential scans however, accurate absolute volume quantitation is less important than interscan intrasubject reproducibility. With this respect, the automated method was superior to the manual method for longitudinal studies.

Overall, the automated algorithm extracted lesions rapidly and with high reproducibility. The particular technique proposed here has the advantage of being largely independent of the manually chosen seed point (figure 1), while such dependence has been pointed out as

an important weakness of automated procedures [5]. Although it still is possible to obtain a small number of different segmentation outcomes with our algorithm, based on different seed locations, these outcomes are usually so dissimilar that only one is acceptable.

In principle, the algorithm can be applied directly to any type of lesion and any imaging sequence, under the condition that lesions should appear significantly more intense than surrounding normal tissue. Standard T2-weighted sequences would hence be appropriate for the detection of small lesions deep into the white matter, but would pose problems if lesions are adjacent to the ventricles or other fluid-filled spaces (which appear as hyperintense in T2-weighted imaging), as the flooding algorithm would spread from the lesion into the ventricles. The algorithm is also directly applicable to the detection of abnormalities seen in Diffusion-Weighted Imaging (DWI), especially when those are small or when their irregular shape renders them difficult to manually delineate.

Conclusion. PML lesions are often large and irregularly shaped; hence, they pose different quantitation challenges from many other white matter diseases. We found that three of the four major sources of quantitation variability, uncertain 3D shape coherence, shape irregularities and inconsistent drawing rules, could be minimized by using an automated segmentation procedure. The fourth source of error, imaging artifacts with the FLAIR sequence, was more easily identified by manual drawing. However, with improved FLAIR sequences [4], this problem may no longer exist. Our study demonstrates that an automated approach, coupled with careful inspection and possible interactive editing, is more reliable and efficient than manual drawing. Therefore, automated segmentation of lesion volume provides an objective measure for monitoring disease progression.

References

- [1] Johnston L, Atkins MS, Mackiewich B, Anderson M. Segmentation of Multiple-Sclerosis Lesions in Intensity Corrected Multispectral MRI. *IEEE Trans Med Ima* 1996;15(2):154-69
- [2] Zijdenbos AP, Dawant BM. Brain Segmentation and White Matter Lesion Detection in MR Images. *Critic Rev Biomed Eng* 1994;22(5-6):401-65
- [3] Filippi M, Yousry T, Baratti C, *et al.* Quantitative assessment of MRI lesion load in multiple sclerosis. A comparison of conventional spin-echo with fast fluid-attenuated inversion recovery. *Brain* 1996;119(Pt 4):1349-1355
- [4] Hajnal JV, Oatridge A, Murdoch J, Kasuboski L, Bydder GM, Failure of FLAIR sequences to suppress CSF in the Posterior Fossa: Diagnosis of a Problem and Solution with adiabatic inversion pulses. In *Proc. 6th Annual Meeting of the International Society for Magnetic Resonance in Medicine*, p. 1350, Sydney, April 18–24, 1998
- [5] Filippi M, Gawne-Cain ML, Gasterini C, *et al.* Effect of training and different measurement strategies on the reproducibility of brain MRI lesion load measurements in multiple sclerosis. *Neurology* 1998;50(1):238-244
- [6] Mark AS, Atlas SW. Progressive multifocal leukoencephalopathy in patients with AIDS: appearance on MR images. *Radiology* 1989;173(2):517-520
- [7] Trotot PM, Vazeux R, Yamashita HK, *et al.* MRI pattern of progressive multifocal leukoencephalopathy (PML) in AIDS. Pathological correlations. *J Neuroradiol* 1990;17(4):233-254
- [8] Bastianello S, Bozzao A, Paolillo A, *et al.* Fast spin-echo and fast fluid-attenuated inversion-recovery versus conventional spin-echo sequences for MR quantification of multiple sclerosis lesions. *Am J Neuroradiol* 1997;18(4):699-704
- [9] Yousry TA, Filippi M, Becker C, Horsfield MA, Voltz R. Comparison of MR pulse sequences in the detection of multiple sclerosis lesions. *Am J Neuroradiol* 1997;18(5):959-963
- [10] Gawne-Cain ML, Silver NC, Moseley IF, Miller DH. Fast FLAIR of the brain: the range of appearances in normal subjects and its application to quantification of white-matter disease. *Neuroradiology* 1997;39(4):243-249

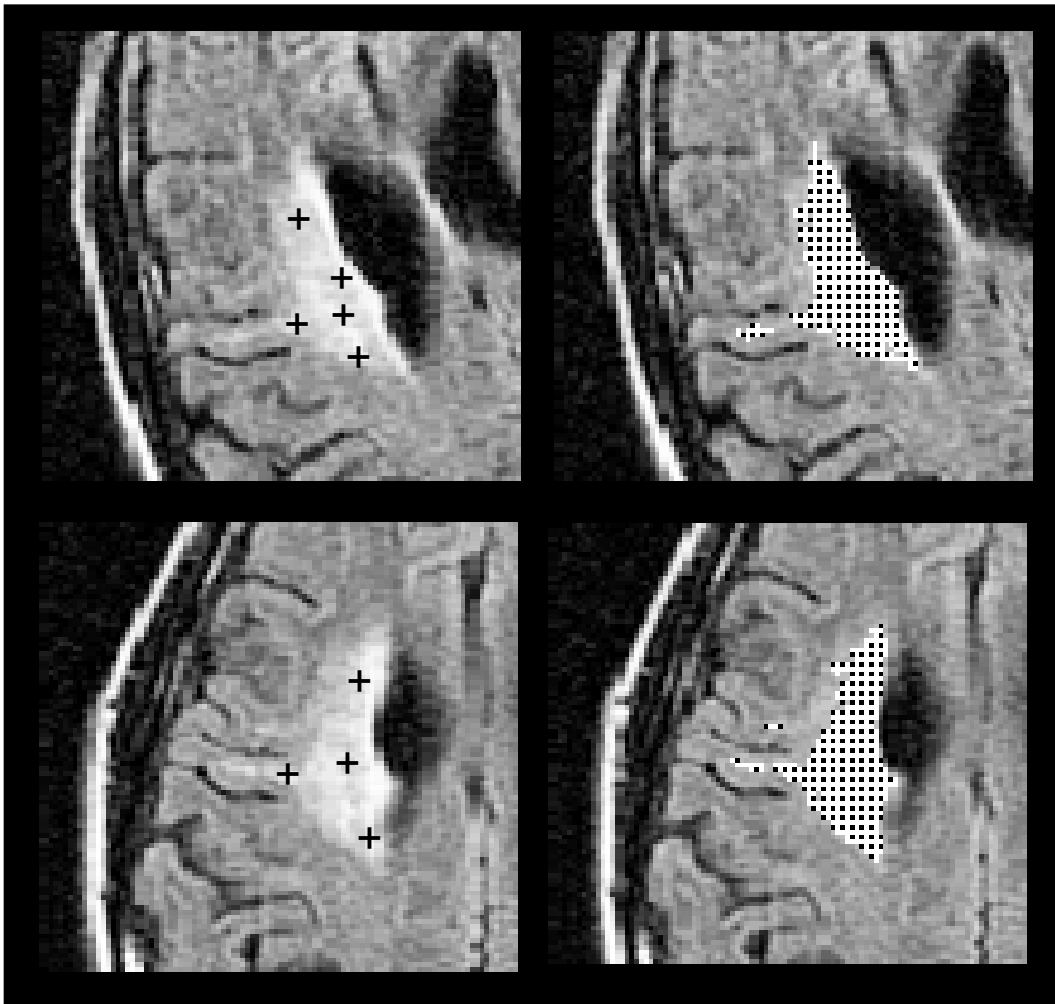


Figure 1: Automated extraction of a medium-sized lesion in two scans acquired consecutively with two head orientations (top: First scan; bottom: Second scan). Crosses on the left indicate several manually-selected seed points, which, given individually as input to the automated algorithm, all yielded the segmented result shown on the right.

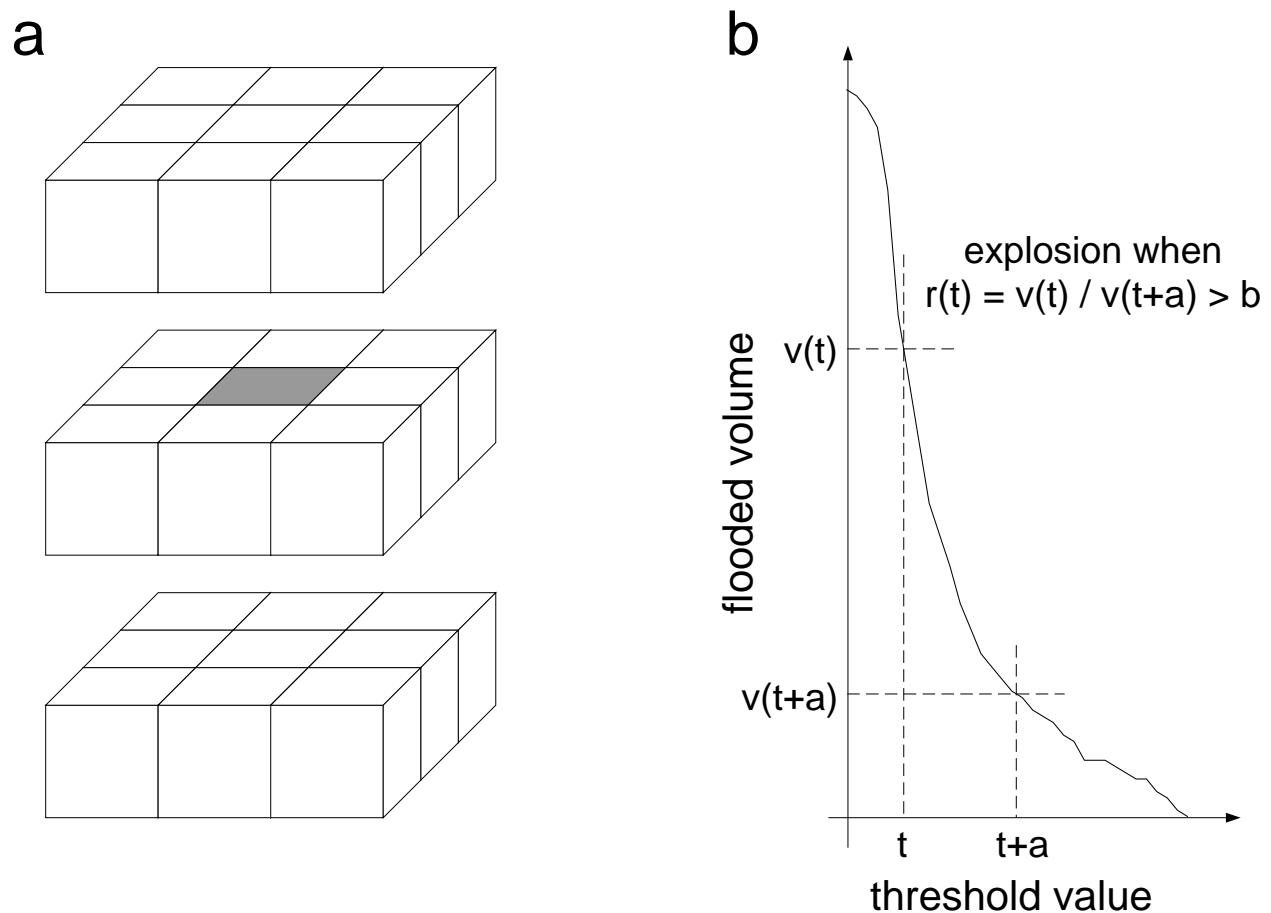


Figure 2: **(a)** Neighborhood used during flooding: the region will grow into volume elements (voxels) with intensities above a threshold t and adjacent to a voxel already known to belong to the lesion (shown in gray), on the same slice as well as on the two adjacent slices. This local flooding is applied recursively until no neighbors above t can be found. **(b)** The threshold t is adaptively determined for each lesion, by starting from the intensity value at the manually selected seed point, and progressively decreasing the threshold by discrete amounts a , until the ratio of flooded lesion volumes obtained for t and $t + a$ becomes greater than a given constant b . This typically occurs as the lesion volume explodes when the threshold becomes sufficiently low as to include voxels in the normal white matter.

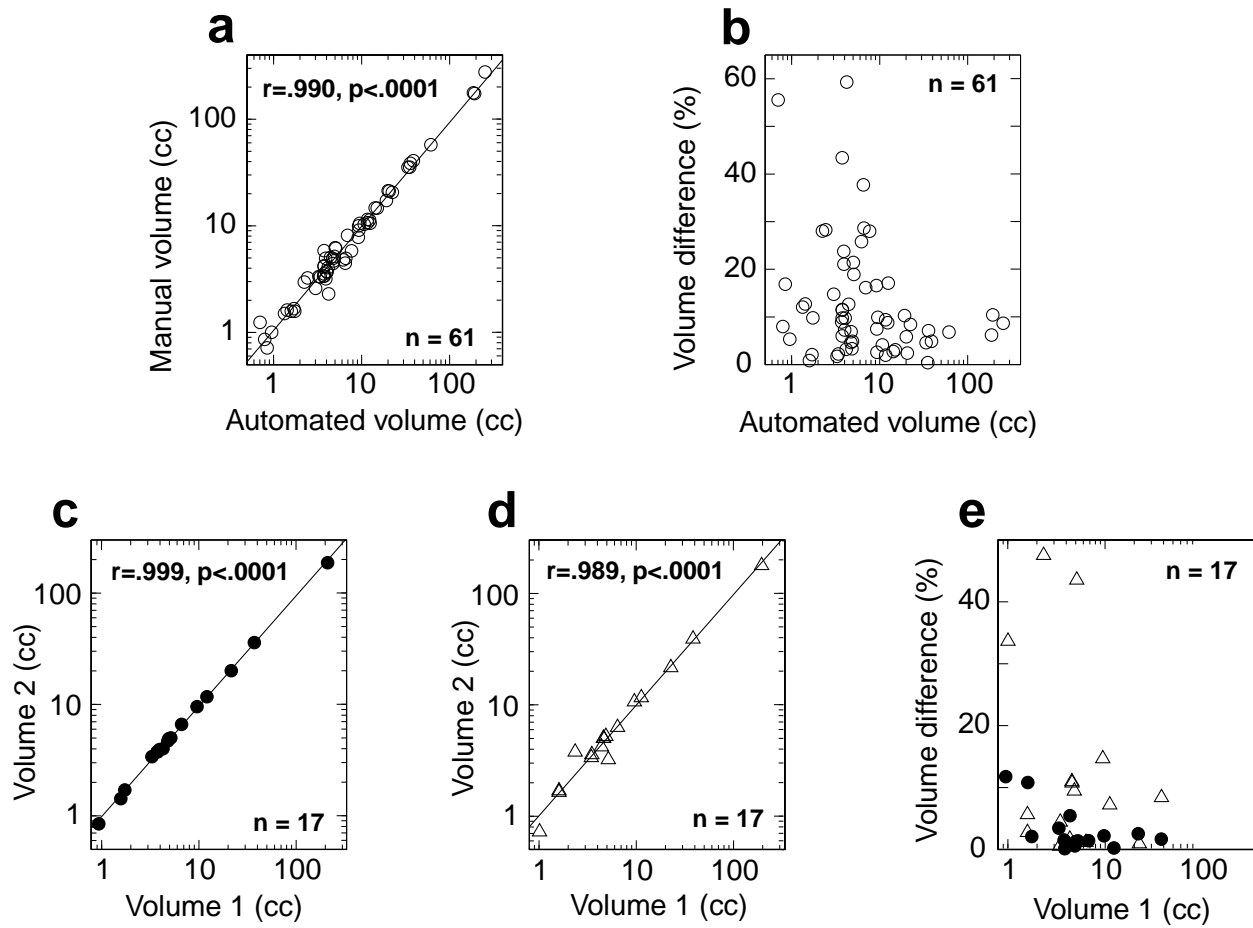


Figure 3: Results of the analysis. (a): Lesion volume computed with the automated algorithm correlated well with the volume found with the manual outlining method. (b): However, the difference between manual and automated volumes was often large, particularly for small lesions (for which an error of only a few pixels can be significant). (c)–(e): Experiments with pairs of repositioned scans. Correlation of computed lesion volumes between scans 1 and 2 was excellent for the automated method (c), and also very good for the manual method (d). Volume differences were, however, much lower with the automated method (e; black circles) than with the manual method (e; triangles).

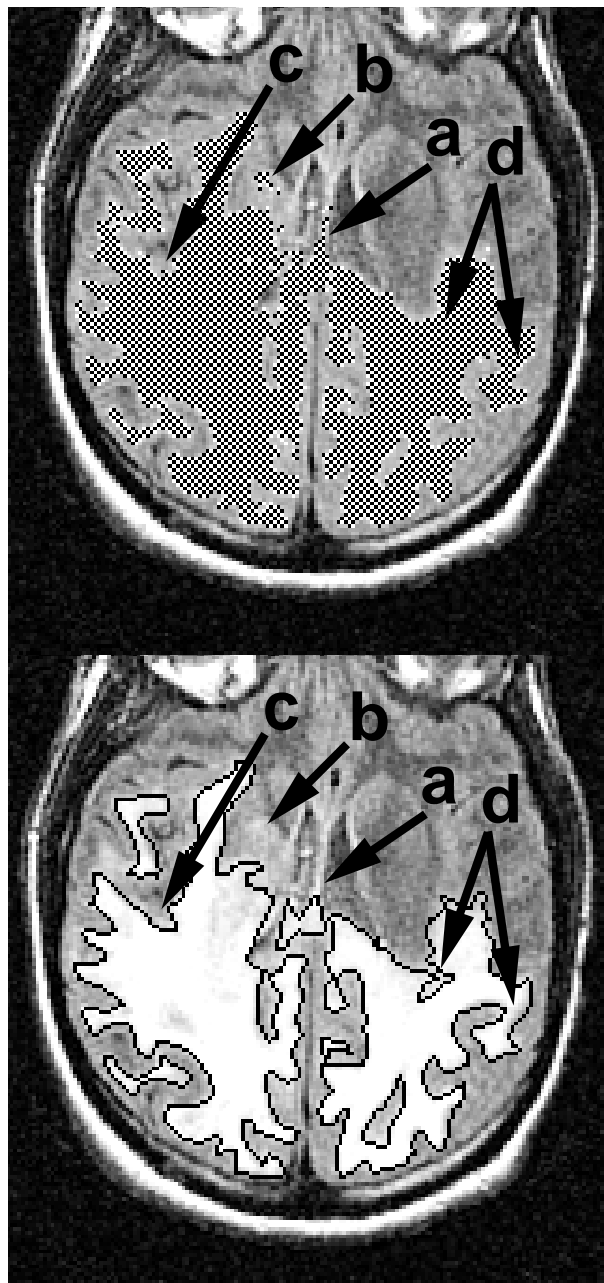


Figure 4: Major sources of discrepancies between the automated (top) and manual (bottom) methods: **(a)** Imaging artifacts: Hyperintensities at the brain/fluid interface are ambiguous for the automated algorithm; **(b)** 3D shape coherence: Here the human observer omitted a small island connected to the main body of the lesion in another slice; **(c)** Small shape irregularities: The manual drawing smoothed out the exact shape of the lesion, which was correctly followed by the automated algorithm; and **(d)** Inconsistent drawing rules in the manual method, more conservative in some regions (left arrow) than others (right arrow).