ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 32 (2006) 1205 – 1215

Technical Note

Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI

Ying Wu,[a,1] Simon K. Warfield,[a,b] I. Leng Tan,[c] William M. Wells III,[d,e] Dominik S. Meier,[a] Ronald A. van Schijndel,[f] Frederik Barkhof,[f] and Charles R.G. Guttmann[a,*]

[a] *Center for Neurological Imaging, Departments of Radiology and Neurology, Brigham and Women's Hospital, Harvard Medical School, 221 Longwood Avenue RF394A, Boston, MA 02115, USA*
[b] *Computational Radiology Laboratory, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA*
[c] *Department of Radiology, Slotervaart Hospital, 1066 EC Amsterdam, The Netherlands*
[d] *Neuroimage Informatics at SPL and Harvard Imaging Center, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA*
[e] *Artificial Intelligence Laboratory, MIT, Boston, MA 02138, USA*
[f] *Diagnostic Radiology, VU Medical Center, PO Box 7057, 1007 MB, Amsterdam, The Netherlands*

Received 28 November 2005; revised 14 March 2006; accepted 5 April 2006
Available online 22 June 2006

Purpose: **To automatically segment multiple sclerosis (MS) lesions into three subtypes (i.e., enhancing lesions, T1 "black holes", T2 hyperintense lesions).**
Materials and methods: **Proton density-, T2- and contrast-enhanced T1-weighted brain images of 12 MR scans were pre-processed through intracranial cavity (IC) extraction, inhomogeneity correction and intensity normalization. Intensity-based statistical *k*-nearest neighbor (*k*-NN) classification was combined with template-driven segmentation and partial volume artifact correction (TDS$^+$) for segmentation of MS lesions subtypes and brain tissue compartments. Operator-supervised tissue sampling and parameter calibration were performed on 2 randomly selected scans and were applied automatically to the remaining 10 scans. Results from this three-channel TDS$^+$ (3ch-TDS$^+$) were compared to those from a previously validated two-channel TDS$^+$ (2ch-TDS$^+$) method. The results of both the 3ch-TDS$^+$ and 2ch-TDS$^+$ were also compared to manual segmentation performed by experts.**
Results: **Intra-class correlation coefficients (ICC) of 3ch-TDS$^+$ for all three subtypes of lesions were higher (ICC between 0.95 and 0.96) than that of 2ch-TDS$^+$ for T2 lesions (ICC = 0.82). The 3ch-TDS$^+$ also identified the three lesion subtypes with high specificity (98.7–99.9%) and accuracy (98.5–99.9%). Sensitivity of 3ch-TDS$^+$ for T2 lesions was 16% higher than with 2ch-TDS$^+$. Enhancing lesions were segmented with the best sensitivity (81.9%). "Black holes" were segmented with the least sensitivity (62.3%).**
Conclusion: **3ch-TDS$^+$ is a promising method for automated segmentation of MS lesion subtypes.**

Introduction

Multiple sclerosis (MS) lesions undergo a complex evolution, characterized by initial breakdown of the blood–brain barrier, accompanied by demyelination, inflammation, axonal damage and, in later stages, by reparatory processes such as astrocytosis and limited remyelination (Guttmann et al., 1995; Bruck et al., 2002; Barkhof and van Walderveen, 1999). Although the specificity of MRI has not permitted complete discrimination of these pathologic components, three main groups of lesions can be identified with conventional MR imaging. These include acute lesions demonstrating blood–brain barrier leakage on contrast-enhanced MR imaging (enhancing lesions), chronic severely damaged lesions that are hypointense, so-called "black holes" on T1-weighted MR images (T1WI) and hyperintense T2 lesions (T2 lesions) on T2-weighted MRI (T2WI).

Classification of MS lesions into these subtypes has achieved wide acceptance and shown good clinical utility in clinical trials (Barkhof et al., 2001; Paty and Li, 2001). The presence of enhancing lesions is currently the most sensitive index of disease activity in MS (Miller et al., 1993). Changes in both number and volume of T2 and enhancing lesions are taken as outcome measurements in drug trials (Molyneux et al., 1998). "Black holes" have also been used as surrogate markers of destructive pathology in clinical studies (Filippi et al., 2001). "Black holes" on T1WI have been correlated with axonal loss in postmortem MRI studies (Barkhof et al., 1998; Van Walderveen et al., 1998) as well as with clinical severity measures in a clinical setting (Truyen et al., 1996).

Over the years, automated methods of T2 lesion segmentation have been studied intensively and have become established methods for the analysis of disease expression on MRI in the

* Corresponding author. Fax: +1 617 2645154.
E-mail address: guttmann@bwh.harvard.edu (C.R.G. Guttmann).
[1] Current address: Center for Advanced Imaging, Department of Radiology, Evanston Northwestern Healthcare and Department of Radiology, Feinberg School of Medicine, Northwestern University, 2650 Ridge Avenue, Evanston, IL 60201, USA.
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2006.04.211

context of clinical, cognitive and immunological aspects of MS (Warfield et al., 2000; Kikinis et al., 1999; Guttmann et al., 1999; Zijdenbos et al., 2002; Van Leemput et al., 2001; Weiner et al., 2000; Wei et al., 2002). In contrast, only few technical reports have described computer-assisted segmentation methods targeting enhancing lesions (Samarasekera et al., 1997; Bedell and Narayana, 1998; He and Narayana, 2002). In addition, manual delineation and human-supervised semi-automated segmentation, such as local thresholding (Rovaris et al., 1997; Filippi et al., 1998) and seed growing (Filippi et al., 2001), have been used for the segmentation of "black holes", and these methods may subject to operator bias introduced by human interaction.

In previous work, we developed and validated automated template-driven segmentation and combined it with a heuristic partial volume correction algorithm (TDS^+) (Warfield et al., 2000; Wei et al., 2002) to identify and outline T2 lesions. In this study, we have developed an automated three-channel TDS (3ch-TDS^+) MRI segmentation pipeline for the identification of MS lesion subtypes. We compare the new 3ch-TDS^+ method's sensitivity, specificity and accuracy for identifying and measuring T2 lesion burden to that derived using 2ch-TDS^+. Furthermore, we assess the sensitivity, specificity and accuracy of 3ch-TDS^+ for Gd-enhancing lesions and T1 "black holes" with respect to manual segmentation.

Materials and methods

MRI acquisition

Twelve MRI scans were obtained from six patients with clinically defined relapsing-remitting (RR) MS at two time points with duration of approximately 5.6 months. The patients were between 27 and 52 years old (mean age 40.5 years) and had EDSS scores between 1 and 6 (mean EDSS score was 2.58 at the first and 3.92 at the second scan). Disease duration was between 5.4 and 11.1 years.

Whole-brain MR imaging was obtained on a 1.5-T MR system (Siemens, Erlangen Germany). After patients were given an intravenous bolus injection of 0.2 mmol/kg of gadolinium-DTPA, PDWI and T2WI were acquired with a conventional dual spin-echo sequence (TR/TE1/TE2: 2500 ms/30 ms/80 ms). Post-contrast T1-weighted spin-echo images (contrast T1WI) TR/TE 625 ms/12 ms were acquired after a 20-min post-injection delay. The head was imaged with 46 contiguous 3-mm-thick axial sections. The nominal voxel size was $0.9765 \times 0.9765 \times 3$ mm^3. Consent was obtained after the nature of the procedure had been fully explained.

Image processing

MRI data were transferred to a Sun Ultra 80 workstation (Sun Microsystems, Inc.). The combined strategies of injection of contrast agent before the scanning of dual-echo and contrast T1WI and careful head constraint limited the likelihood of head movement between the two sequences. We did not observe head displacement between dual-echo and contrast T1WI in this group of subjects. PDWI, T2WI and contrast T1WI were natively well aligned. No post-acquisition image registration was applied. The image processing procedures are summarized by the flow chart in Fig. 1 and described in detail hereafter.

Intracranial cavity (IC) identification

Masks of the IC were generated automatically from the PDWI and T2WI. This IC extraction procedure combined non-parametric intensity-based statistical (Parzen window) segmentation and automated morphological operations (Kikinis et al., 1992).

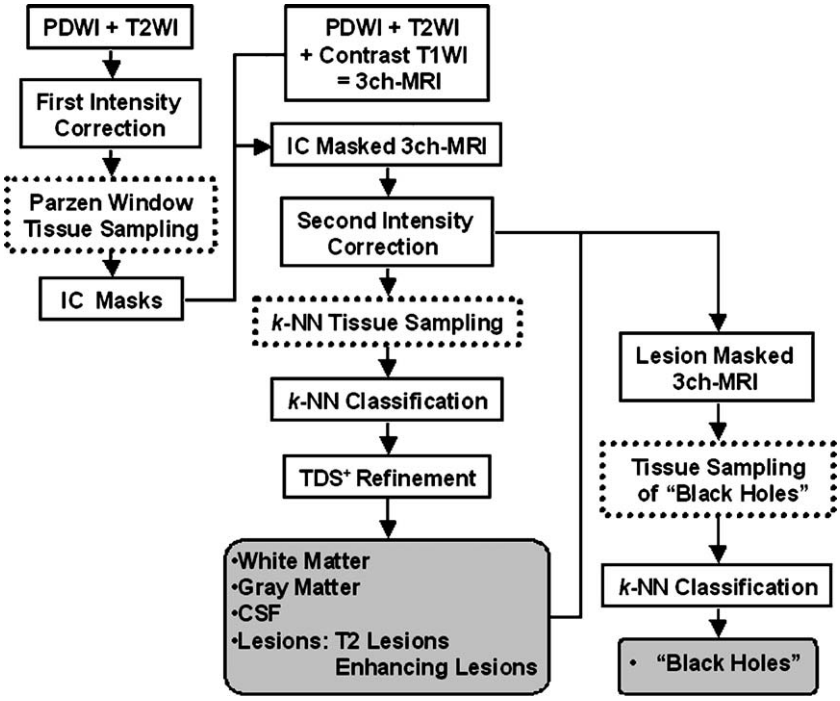

Fig. 1. Flowchart for the automated three-channel segmentation. Dashed boxes indicate procedures performed only once and then used for all the subjects. Abbreviations: 3ch-MRI = three-channel MRI consisted of proton density-weighted image, T2-weighted image and post-contrast T1-weighted image; IC = intracranial cavity mask extraction; TDS^+ = template-driven segmentation plus partial volume artifact correction.

Preprocessing using EM segmentation (Wells et al., 1996) for inhomogeneity correction before the IC extraction procedure (Wu et al., 2001) obviated the need for per-case expert-supervised Parzen window tissue sampling and significantly improved the automation of the IC identification for this group of brain images. After the human-supervised Parzen window tissue sampling was calibrated one time on two test scans, the Parzen window classifier was saved for automated segmentation of ICs on the remaining 10 scans. Occasionally, minimal operator manual editing was conducted on the resulting IC masks.

The IC masks were superimposed onto the PDWI, T2WI and contrast T1WI MR images to exclude extracranial tissues, skull and large vessels on the brain surface.

Intensity correction

After the IC masking procedure described in the previous section, EM segmentation was applied to the IC portion of the PDWI, T2WI and contrast T1WI for inhomogeneity correction and intensity normalization. The EM segmenter has a component that compensates for intra- and inter-scan intensity inhomogeneities and normalizes the observed scan intensities (Wells et al., 1996; Wu et al., 2001). Studies have shown the usefulness of applying inhomogeneity correction exclusively to the IC portion of MR images (Johnston et al., 1996; Mackiewich, 1995).

Statistical intensity-based k-nearest neighbor segmentation (k-NN)

Non-parametric *k*-nearest neighbor (*k*-NN) segmentation was chosen for the initial image intensity-based statistical classification of the intensity corrected and IC masked three-channel image data into tissue classes. The *k*-NN rule is effective for multichannel MR data and particularly suitable for this three-channel segmentation. Previous studies also show that post-probability *k*-NN segmentation is more accurate and stable than segmentation approaches based on a priori statistical assumptions (Clarke et al., 1993; Vinitski et al., 1997). Decoupling the image intensity correction step from the statistical classification of multichannel data (as opposed to using the EM approach of simultaneously estimating tissue class assignment and bias field) has the advantage of reducing the complexity of the interactive calibration procedure by limiting the number of parameters (and their permutations) to be set for each step.

The *k*-NN segmentation was developed based on Friedman's *k*-NN algorithm (Friedman et al., 1975; Warfield, 1996). Two major procedures were involved in this process: the one time operator-supervised tissue sampling and the automated classification. Two randomly chosen "calibration" scans were used to build a master classifier and tune the algorithm's parameters. The stored master classifier and parameters were then applied for the automated segmentation of the remaining 10 scans. An experienced operator sampled approximately 20 pixels for each of seven categories (Class 1: T1 hyperintense enhancing lesions; Class 2: T1 hypointense, cerebrospinal fluid (CSF)-like "black holes", note the "black holes" with similar signal intensity as gray matter and slightly lower than white matter are not included in Class 2 during this step; Class 3: T1 isointense but T2WI and PDWI hyperintense lesions ("other T2 lesions"); Class 4: normal appearing white matter; Class 5: gray matter; Class 6: CSF; Class 7: background) by simultaneously viewing spatially corresponding images from the three different contrast channels of the two calibration scans. All of the sample points were used to construct a master classifier on the basis of the *k*-NN rules. The operator was able to interactively select *k* value and the tissue training prototypes to maximize the classification accuracy on the basis of classification results. After the expert was satisfied with the classification on the two calibration scans, the classifier was saved as a master classifier. A *k* value of 3 was estimated and used throughout the 10 studied scans. The master classifier was used as a three-dimensional (three image contrasts) table to assign each voxel within the IC to one of the seven possible categories based on Friedman's *k*-NN algorithm in the 10 scans without further expert intervention.

To ensure proper comparison, the identical tissue sample points were engaged in both 2ch-TDS$^+$ and the 3ch-TDS$^+$ processing, the master classifier for 2ch-TDS$^+$ was derived from the existing master classifier for 3ch-TDS$^+$ by removing the intensity vector from the third channel, and keeping only the intensity vectors from PDWI and T2WI.

Template-driven segmentation and partial volume artifact correction (TDS$^+$)

Significant overlap in the signal intensity distribution of different tissue or lesion classes is present in the feature space defined by the contrast mechanisms of the three image channels. This overlap in feature space leads to ambiguities in assigning tissue classes and consequent misclassifications of pixels when using a signal-based statistical classifier such as *k*-NN.

TDS$^+$ (Warfield et al., 1995; Guttmann et al., 2000; Wei et al., 2002) was adapted and applied to correct misclassifications on *k*-NN segmented images and thereby improve the classification of MS lesion subtypes by providing *a priori* anatomical probabilities (Figs. 2 and 3). TDS$^+$ employs a deformable digital anatomical atlas to extract white matter masks for each individual brain using non-linear registration. Based on the assumption that white matter lesions (T2 lesions, enhancing lesions, or T1 "Black Holes") are only within white matter regions, misclassified lesions outside the white matter masks were relabeled; within white matter masks, abnormal white matter areas misclassified as gray matter by *k*-NN were also relabeled as lesions (Warfield et al., 1995) (Figs. 2 and 3). The method (Warfield et al., 2000) makes use of individualization of an anatomical atlas to compensate for the MRI signal intensity overlap of different tissue types. The atlas provides spatial context (a priori probabilities of tissue class assignment), which enables improved classification when different tissue classes have similar or overlapping MRI signal intensities but different locations in space. The amount of overlap of signal intensity can be increased by decreased SNR. To perform well in the presence of signal intensity overlap, we devised the joint signal and spatial context probability density function estimation procedure. The action of the probability density function estimation procedure converges to the optimal estimate of tissue class distribution accounting for both signal intensity and spatial location. Partial volume effects near ventricles and subarachnoid CSF were corrected relying on morphological operators (Wei et al., 2002).

Refining "black holes" segmentation

"Black holes" (Class 2) identified by the above-described procedure do not include areas of the white matter that are hypointense with respect to healthy white matter but isointense with respect to grey matter. Therefore, in a second step the classification of "black holes" is refined to include subtly hypointense signal. A

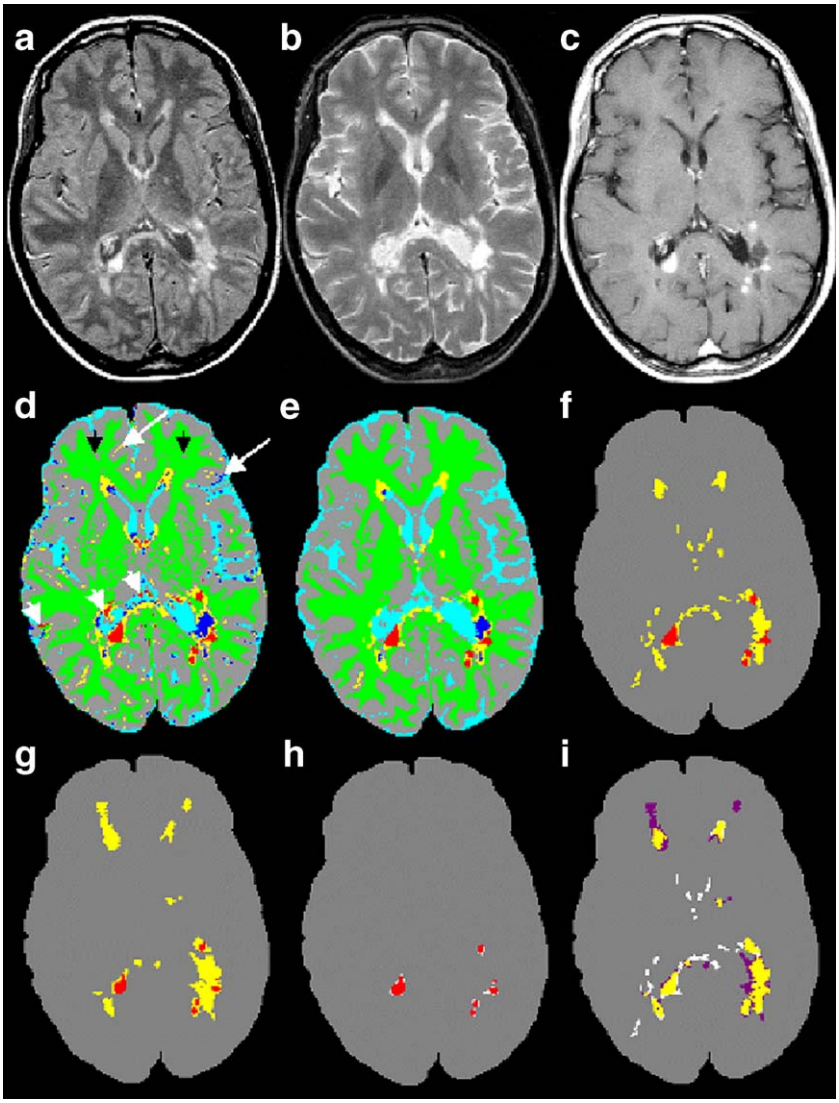

Fig. 2. Segmentation and validation of T2 and Gd-enhancing lesions. One axial section of an MS patient's brain and segmentation results are shown: proton density-weighted image (a), T2-weighted image (b), contrast-enhanced T1-weighted image (c), tissue classification label maps after *k*-NN segmentation (d) and after TDS$^+$ (e). Tissue classes in panels d and e are color coded as follows: red, enhancing lesion; dark blue, "black holes"; yellow, T2 lesion; light blue, CSF; gray, gray matter; green, white matter; and black, background. Three types of misclassification are shown: false-positive enhancing lesions, for example, due to choroid plexus and other vascular enhancement (short white arrows), false-negative misclassifications of subtle lesions as grey matter inside the white matter (short black arrows), false-positive misclassification of grey matter as white matter "black holes" on the brain surface (long white arrows). Note that misclassifications have been largely eliminated after 3ch-TDS$^+$ compared to the results using *k*-NN alone. Identification of enhancing (red) and non-enhancing T2 lesions by 3ch-TDS$^+$ (f) and manual outlining by an expert (g) were compared using overlap analysis for enhancing lesions (h) and T2-lesions (i). Color code in panel h: red, true-positive enhancing lesions; white, false-positive enhancing lesions; grey, true-negative enhancing lesions. Color code in panel i: yellow, true-positive total T2 lesions; white, false-positive total T2 lesions; grey, true-negative total T2 lesions; purple, false-negative total T2 lesions.

more sensitive *k*-NN classifier is obtained by adding training points from mildly T1-hypointense WM regions and is selectively applied to lesion classes (Classes 1, 2 and 3). One master classifier for this second segmentation step was generated and stored for *k*-NN segmentation of all the subjects. Fig. 3e shows the refined "black holes" segmentation.

Statistical methods

Volumetric as well as overlap analyses were performed to compare overall lesion volumes identified using either 3ch-TDS$^+$ or 2ch-TDS$^+$ to expert outlines performed on proton-density-weighted MR images covering the entire brain. Similarly, volumetric analysis was used to compare automated and manual lesion burden estimates for lesion subtypes and overlap analysis was used to assess sensitivity, specificity and accuracy of 3ch-TDS$^+$ for contrast-enhanced lesions and T1 "black holes" outlined by experts on contrast T1WI (see Figs. 2 and 3). True-positive (TP), true-negative (TN), false-positive (FP) and false-negative (FN) pixels were assessed with a spatial-overlap method implemented in-house using MATLAB 12.1 (MathWorks). TP pixels were defined as the overlapping area of the results of 3ch-TDS$^+$ or 2ch-TDS$^+$ and expert outlines; TN pixels were defined as all IC pixels not outlined as lesions by the experts; FP pixels were those detected by 3ch-TDS$^+$ or 2ch-TDS$^+$, respectively, but not by human experts; FN pixels were those identified by the experts as

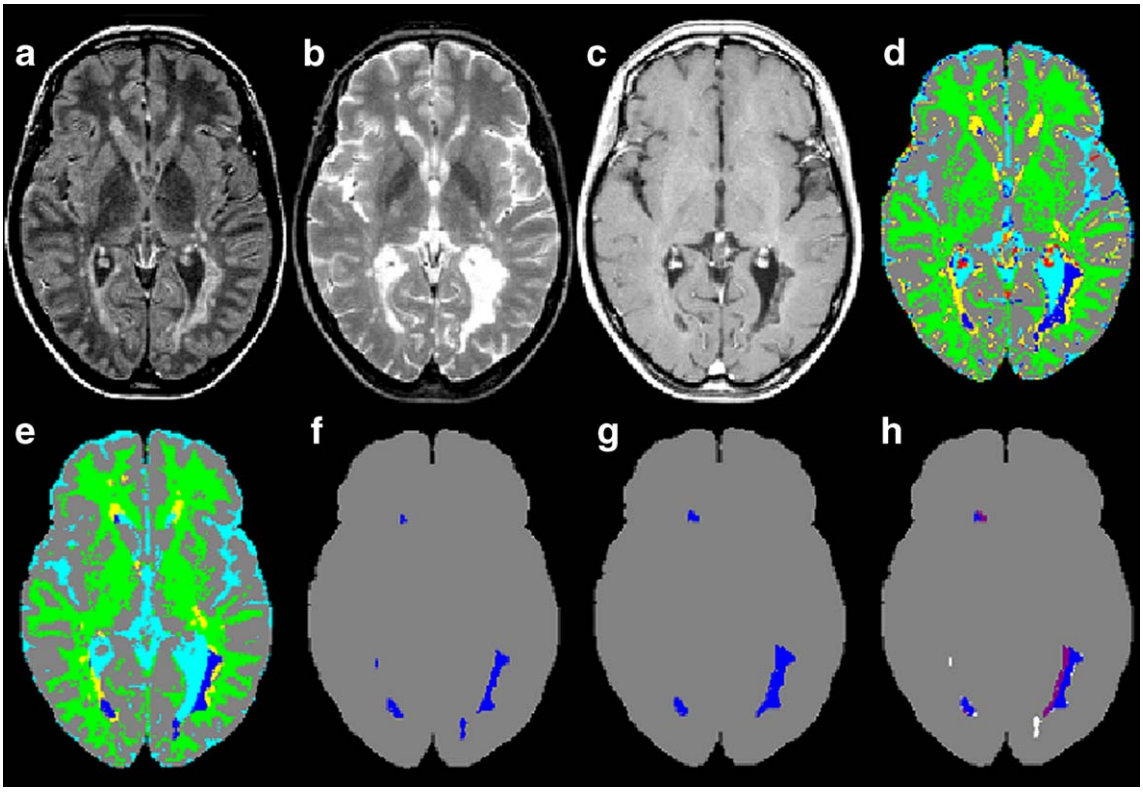

Fig. 3. Segmentation and validation of "black holes" identified by 3ch-TDS$^+$. "Black holes" appear hyperintense on both proton density-weighted (a) and T2-weighted images (b), while they appear hypointense with a spectrum of intensities below that of normal appearing white matter on the contrast-enhanced T1-weighted image (c). Results of *k*-NN segmentation (d) and TDS$^+$ (e) are shown with following color coding: red, enhancing lesions; dark blue, "black holes"; yellow, other T2 lesions that are not enhancing or "black holes" on post-contrast T1-weighted images; light blue, CSF; gray, gray matter; green, white matter; and black, background. Results of 3ch-TDS$^+$ of "black holes" (dark blue in panel f) are compared with manual segmentation of "black holes" by an expert (dark blue in panel g) using overlap analysis (h). Color coding for the overlap analysis in panel h: dark blue, true positive; white, false positive; purple, false negative; gray, true negative.

lesions but not by 3ch-TDS$^+$ or 2ch-TDS$^+$. Overlap analysis accounts not only for volumetric, but also for spatial accuracy. The sensitivity, specificity and accuracy of enhancing lesion, "black holes" and total T2 lesion identification using 3ch-TDS$^+$ or total T2 lesion identification for 2ch-TDS$^+$ was calculated from TP, TN, FP and FN, as customary (Zou et al., 2002; Anbeek et al., 2004): Sensitivity = True-positive fraction = TP / (TP + FN); specificity = True-negative fraction = TN / (TN + FP); Accuracy = Percentage agreement = (TN + TP) / (TN + TP + FN + FP).

The relationship between volumetry derived through automated segmentation (2ch-TDS$^+$ and 3ch-TDS$^+$) and manual segmentation was studied with correlation analysis (R^2 correlation coefficients and intra-class correlation coefficients). The volumetric comparison of lesion volume between the automated segmentation and the gold standard were performed to evaluate the correlations using scatter plots (Matlab 6.5.1, MathWorks) and to assess the agreement and reliability using intra-class correlation coefficient (ICC) (SPSS 12.0; Chicago, IL) as previously reported by others (Admiraal-Behloul et al., 2005). Similar analysis was performed to compare volumetric change detection between automated segmentation methods and expert outlining. Pairs of scans were available to assess T2 lesion volume change over time (average inter-scan interval: 5.6 months) with 2ch-TDS$^+$, 3ch-TDS$^+$ and manual outlining. The relationship between lesion volumes and measures of sensitivity, specificity and accuracy was evaluated with scatter plots and linear regression analysis.

For an objective evaluation, the manual outlining was performed at another institute where the experts were blinded to the 3ch-TDS$^+$ and the 2ch-TDS$^+$. T2 lesions were traced on nine scans with PDWI; enhancing lesions and "black holes" were outlined on contrast T1WI in four and six scans, respectively. A total of 772 T2 lesions (confluent lesions were considered a single lesion), 51 contrast-enhancing lesions and 120 T1 "black holes" were identified and manually outlined.

Results

Figs. 2 and 3 illustrate typical results of the 3ch-TDS$^+$ applied to T2WI, PDWI and contrast T1WI and compare them to the results of manual outlining by multiple experts. *k*-NN segmentation captured most of the lesions; however, three types of misclassification were evident: misclassification of choroid plexus and other enhancing vascular structures as enhancing lesions, misclassifications of subtle signal abnormalities of the white matter as grey matter and misclassification of pixels on the cortical surface as white matter lesions, in particular as T1 "black holes" (Figs. 2d and 3d). A significant portion of the misclassifications were corrected for both the enhancing lesions (Fig. 2e) and the black holes (Fig. 3e) by applying the TDS$^+$ strategy. 3ch-TDS$^+$

showed a high degree of similarity to the manual segmentation performed by experts (Figs. 2f, g, and 3f, g).

Figs. 4 and 5 provide a side-by-side comparison of 2ch-TDS$^+$ and 3ch-TDS$^+$. The addition of contrast T1WI as a third channel appeared to improve the accuracy of detection of white matter lesions and reduce false-positive misclassifications of cortical regions as white matter lesions (Fig. 4). Notably, 3ch-TDS$^+$ also improved the classification of gray matter, white matter and CSF at the upper convexity of the brain, an area that is error prone for 2ch-TDS$^+$ (Fig. 5).

Significant positive correlations were found between lesion volume of the automated segmentations and the manual segmentation (Fig. 6). We found high agreement (as measured by ICC) between the manual segmentation and both the 2ch-TDS$^+$ and 3ch-TDS$^+$ (ICC: 0.82–0.96). The volumetric agreement between automated and manual segmentations was higher for all lesion subtypes using 3ch-TDS$^+$ (ICCs all above 0.95) than that for T2 lesion load using 2ch-TDS$^+$ (ICC = 0.82). Among the three subtypes, the segmentation of T2 lesions using 3ch-TDS$^+$ had the best agreement to the manual segmentation (ICC = 0.96).

The overlap analysis, a more stringent assessment of the quality of assessment of both lesion size and special location, showed that 3ch-TDS$^+$ identified the three lesion subtypes with high specificity (98.7–99.9%) and accuracy (98.5–99.9%) with respect to the gold standard. It also indicated an improved sensitivity of 16% for segmentation of T2 lesions compared to 2ch-TDS$^+$ (see Table 1). Positive correlations (Fig. 7) were found between the magnitude of T2 lesion load in brain white matter and the sensitivity and accuracy of the T2 lesion segmentations using 2ch-TDS$^+$ [significant for accuracy and trending towards significance (P = 0.06) for sensitivity] and 3ch-TDS$^+$ (significant for both sensitivity and accuracy) (Table 2). These findings are consistent with previous studies (Anbeek et al., 2004; Admiraal-Behloul et al., 2005) suggesting that better sensitivity and accuracy were associated with bigger T2 lesion loads. The specificity and accuracy in the identification of "black holes" were also associated with the lesion volume of "black holes".

Because segmentation methods are often used in longitudinal studies, we also assessed the agreement between estimates of lesion volume change by automated segmentation compared with the manual "gold standard". The correlation and the agreement to the gold standard of lesion volume change over time were higher with 3ch-TDS$^+$ (R^2 = 0.715; ICC = 0.644) than with 2ch-TDS$^+$ (R^2 = 0.257; ICC = 0.499) (Fig. 8).

Discussion

We have developed an algorithm that provides quantitative volumetric measurements of three types of MS lesions, as well as segmentation of gray matter, white matter and CSF simultaneously. Due to the complex misclassifications encountered, the studies of enhancing lesions are very limited (Samarasekera et al., 1997; Bedell and Narayana, 1998; He and Narayana, 2002). A fuzzy connectivity algorithm has been used to remove enhancing blood vessels (Samarasekera et al., 1997), and T2 lesion segmentation has been utilized to constrain the enhancing lesion segmentation (Bedell and Narayana, 1998; He and Narayana, 2002). A particular MRI sequence has also been used to suppress enhancement within vessels (Bedell and Narayana, 1998). Seed-growing, thresholding-based techniques (Filippi et al., 2001) have been used for segmentation of "black holes", which however is susceptible to variations by seed selection and threshold settings (Filippi et al., 1996). This investigation utilized a novel TDS$^+$ strategy making use of a deformable digital brain atlas to identify the white matter tissue compartment and eliminate the confounding misclassifications.

TDS$^+$ identified three types of MS lesions on intensity normalized MR images obtained with an MRI acquisition protocol, which included standard dual-echo and contrast T1WI, but was designed to maximize the conspicuity of enhancing lesions by injecting double the standard dose of Gd-DTPA and lengthening the interval between contrast agent injection and post contrast T1WI to 20 min. Currently, clinically used protocols often include a shorter post-injection delay of 5 min. While TDS$^+$ is adaptable to a multitude of MRI acquisition protocols, the specific impact of lesser contrast-to-noise between enhancing lesions and non-enhancing WM needs to be assessed and validated on a protocol by protocol basis. Although MRI acquisition and analysis are traditionally discussed as independent steps in the morphometric assessment of healthy and pathological brain structures, they are intimately intertwined and constitute an integrated measurement system. Variations in either acquisition or analysis procedures (including MRI scanner and protocol variablilities in multicenter studies and clinical treatment trials) are likely to modify sensitivity, accuracy and specificity with respect to structures of interest and require careful ad hoc validation. Nevertheless, the presented results combined with our previous work with other MRI acquisition protocols on multiple MRI scanners (Guttmann et al., 1999; Wei et al., 2002) suggest that TDS$^+$ is a highly adaptable analytical approach suitable for use in multicenter clinical trials,

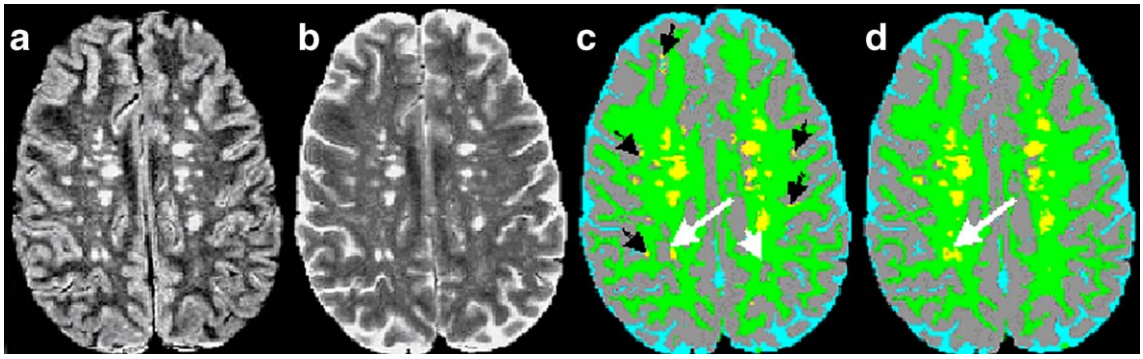

Fig. 4. T2 lesion segmentation with 2ch-TDS$^+$ and 3ch-TDS$^+$. Proton density-weighted image (a) and T2-weighted image (b) at the level of the centrum semiovale showing multiple hyperintense oval shaped plaque-like and punctate lesions in the white matter of an MS patient. 2ch-TDS$^+$ (c) and 3ch-TDS$^+$ (d) of the same section. The results of 3ch-TDS$^+$ demonstrate less false-negative (white arrows) and false-positive (black arrows) T2 lesions.

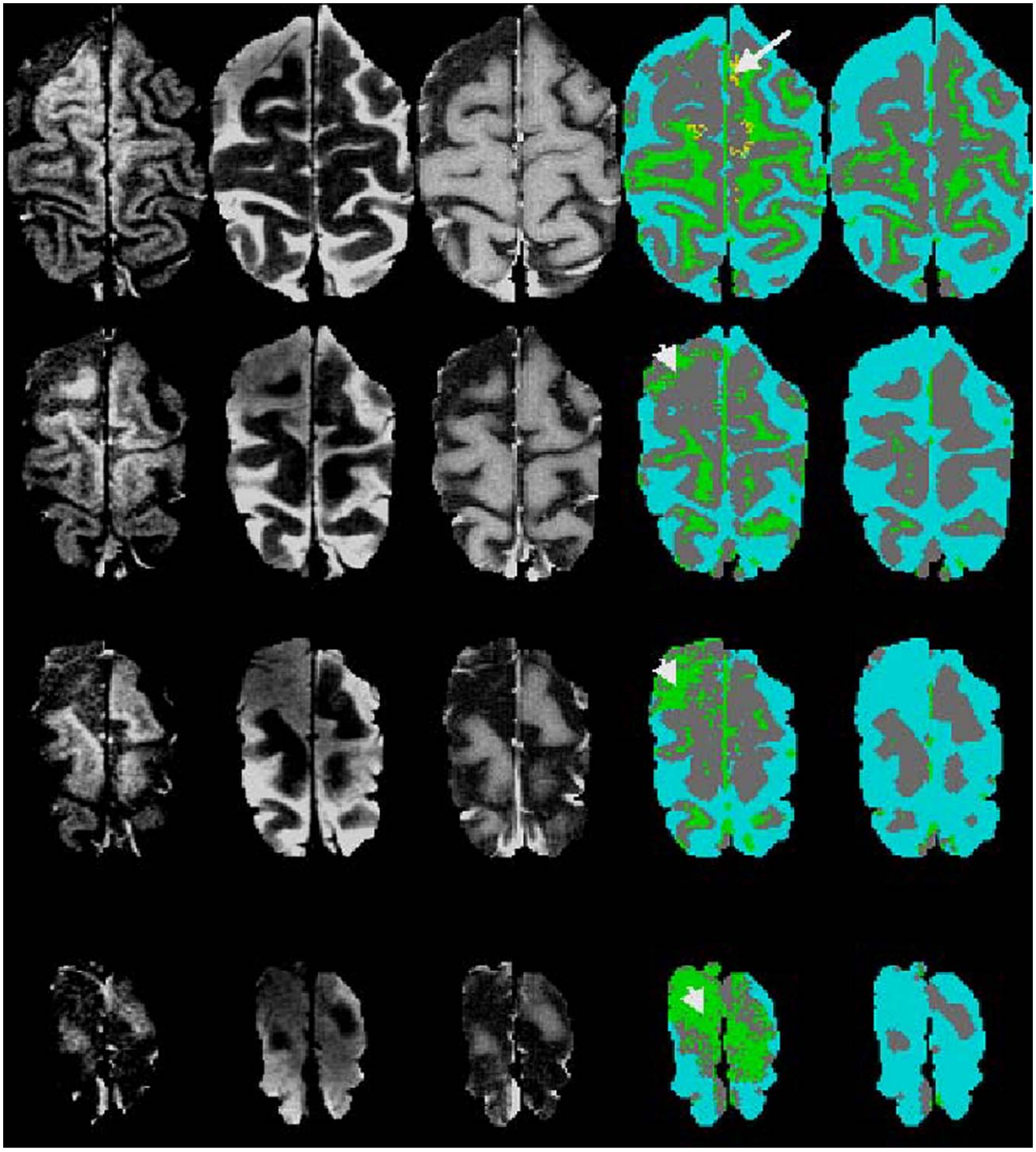

Fig. 5. Improved segmentation of the brain's upper convexity with 3ch-TDS$^+$. Four consecutive axial sections of the upper convexity of the brain are shown: proton density-weighted images (column one), T2-weighted images (column two) and contrast-enhanced T1-weighted images (column three), as well as results of 2ch-TDS$^+$ (fourth column) and 3ch-TDS$^+$ (fifth column) are shown. Misclassified regions of gray matter (long arrow) and CSF (short arrows) observed in 2ch-TDS$^+$ were segmented more accurately by 3ch-TDS$^+$.

provided adequate statistical modeling of potential sources of variability such as differing SNR and CNR from scanner to scanner.

The EM component of the TDS$^+$ pipeline has previously been shown to significantly diminish scan-to-scan variability of segmentation results, and normalized ratios of lesions to total white matter appear to confer added robustness when comparing results obtained with a same MRI acquisition protocol on different scanners (Guttmann et al., 1999). The addition of anatomical knowledge through the template-driven component of TDS$^+$ has been previously shown to improve the accuracy of lesion detection compared with expert outlining (Wei et al., 2002).

Our current TDS$^+$ procedure does not include a spatial model of the ventricles. Partial volume artifacts at the interface of the

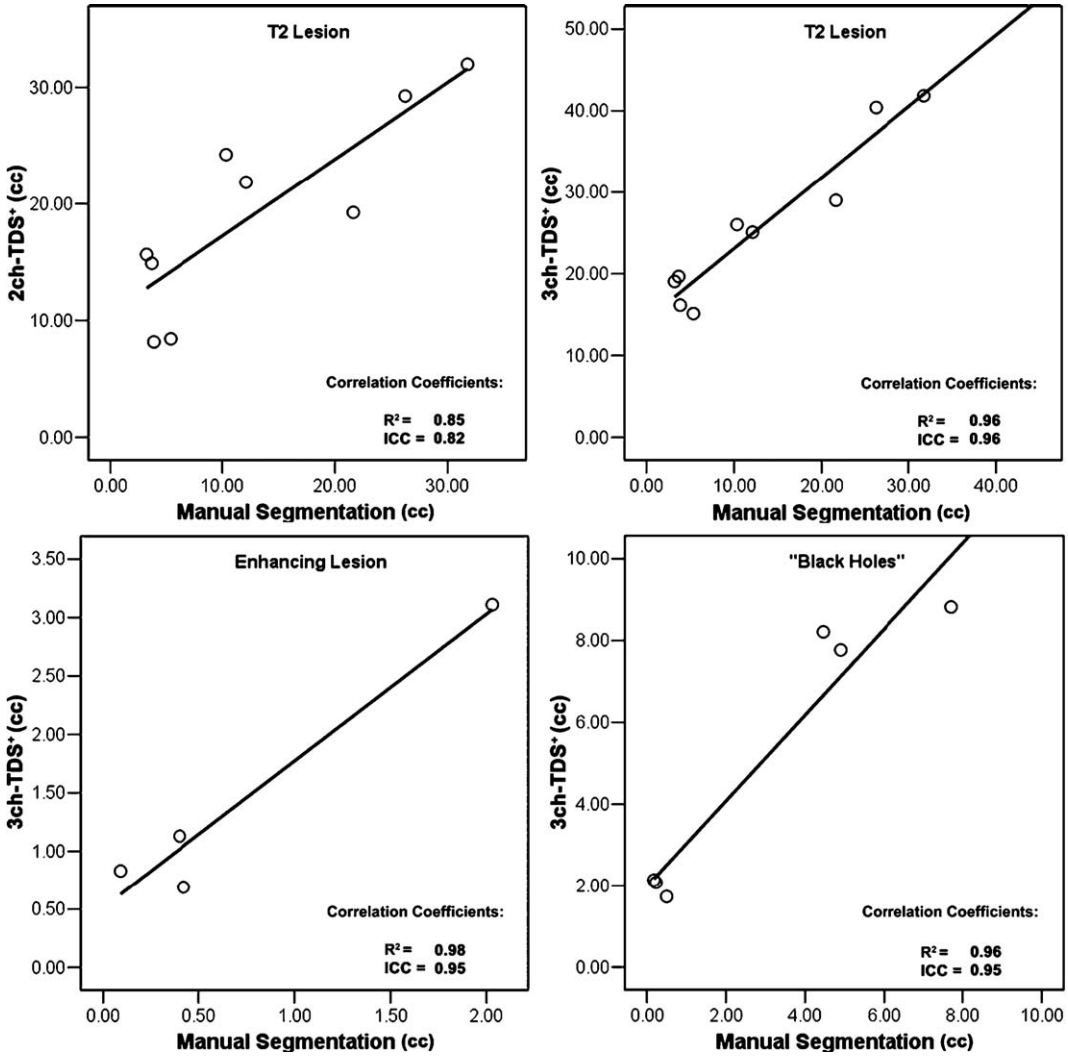

Fig. 6. Scatter plots of lesion volume measurements illustrate the strong positive correlation between manual and automated segmentation (2ch-TDS$^+$ and 3ch-TDS$^+$). Compared to the 2ch-TDS$^+$, the 3ch-TDS$^+$ has stronger correlation and higher agreement with manual segmentation.

lateral ventricles and white matter masks can introduce misclassifications of enhanced choroid plexus as enhancing lesions even after TDS refinement. In this study, we made use of morphologic operators and connectivity principles to reduce this misclassification. In the future, we expect that the integration of explicit anatomical models of ventricles may reduce this source of misclassification.

The reproducibility of 3ch-TDS$^+$ was not studied due to the impracticality of administering contrast agent twice within a short interval. Studies of other automated methods have shown that inter-scan variability within patients is mainly due to resampling errors from repositioning (Guttmann et al., 1999; Gawne-Cain et al., 1996). In a scan–rescan validation study, 2ch-TDS$^+$ demonstrated good reproducibility with inter-scan coefficients of variation for T2 lesion volume averaging 4.98% in 20 MS patients that had undergone MRI exams twice within 30 min (Wei et al., 2002, 2004). The accuracy of pairwise change in T2 lesion volume over time with respect to the results of expert outlining was higher for 3ch-TDS$^+$ than for 2ch-TDS$^+$ (Fig. 8). Systematic measurement errors tend to cancel out when subtracting repeated measures from an individual patient. This finding, suggests that 3ch-TDS$^+$ is more robust on repeated measures, than 2ch-TDS$^+$. Although we were not able to perform scan–rescan experiments within 30 min using contrast-enhanced MRI (the contrast agent from the first scan would not be completely cleared from the system within such a short interval and would confound the second MRI assessment), we expect that resampling errors, which are main contributors to scan–rescan discrepancy, would be comparable for enhancing lesions and "black holes" (Guttmann et al., 1999; Wei et al., 2002, 2004). The preliminary finding of improved accuracy of change detection using 3ch-TDS$^+$ vs. 2ch-TDS$^+$ suggests that 3ch-TDS$^+$ is better suited for longitudinal studies, such as clinical trials of novel treatments. Future studies of reproducibility could be performed using a

Table 1
Overlap analysis of automated segmentation with respect to manual segmentation

Segmentation methods	Type of lesions	Sensitivity (%)	Specificity (%)	Accuracy (%)
2ch-TDS$^+$	T2 lesions	60.0	99.1	98.8
3ch-TDS$^+$	T2 lesions	70.0	98.7	98.5
	Enhancing lesions	75.2	99.9	99.9
	"Black holes"	62.3	99.7	99.7

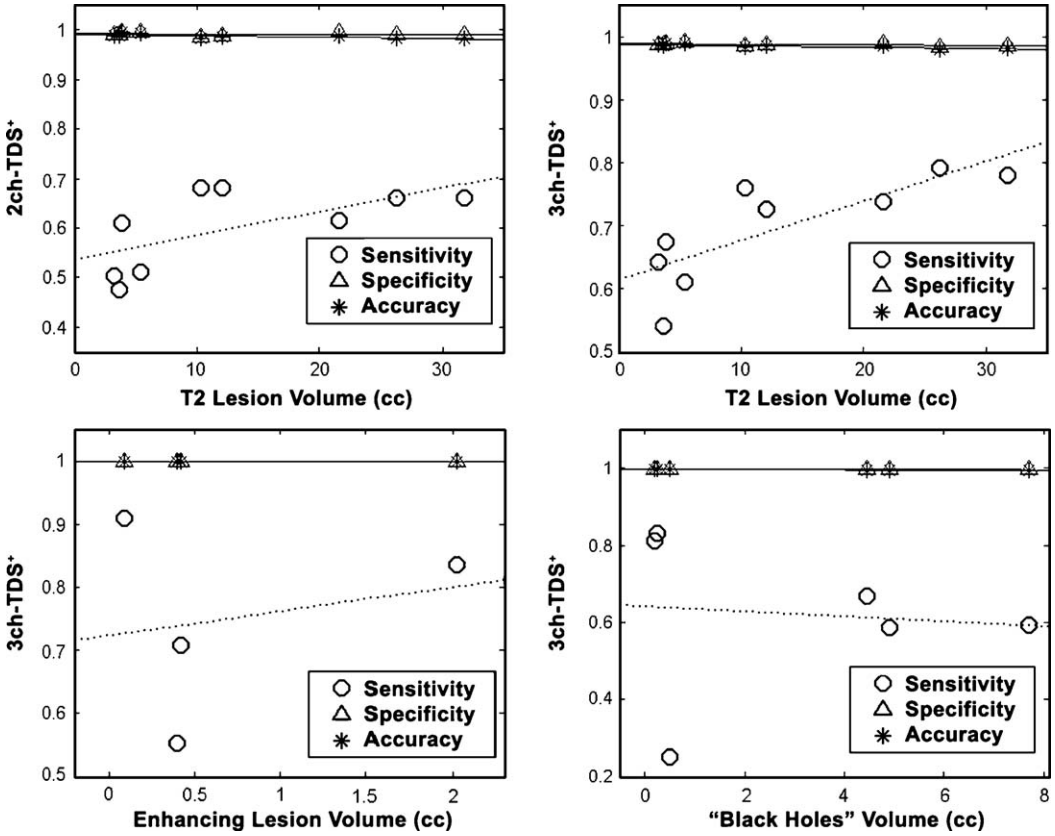

Fig. 7. Relationship between sensitivity, specificity and accuracy of automated segmentation and lesion burden. The scatter plots show higher sensitivity of automated segmentation above T2 lesion load of 10 cc. Lines represent results of linear regression analysis.

modified protocol omitting contrast enhancement. However, the differences in image contrast between enhanced and unenhanced T1WI are likely to have an important impact on the performance of 3ch-TDS$^+$, so that results with one protocol are not undisputably comparable with results from the other. A second option would be to wait 48 h between scans to ensure adequate clearance of the contrast agent between the two measures, even though there is a realistic chance that real biological change (e.g., new enhancing lesions or change in strength of enhancement of existing lesions) would occur in that interval.

It should be noted that there is no true gold standard for studying segmentation of brain lesions. We used expert manual segmentation results as the practical gold standard in order to assess the accuracy of MS lesion segmentation, similarly to what has been done in previous work (Wei et al., 2002; Ashton et al., 2003; Anbeek et al., 2004; Admiraal-Behloul et al., 2005).

Table 2
Lesion load dependency of sensitivity, specificity and accuracy of automated segmentation

Segmentation methods	Type of lesions	R^2 SE	R^2 SP	R^2 AC
2ch-TDS$^+$	T2 lesions	0.63	0.08	0.72*
3ch-TDS$^+$	T2 lesions	0.80*	0.53	0.87**
	Enhancing lesions	0.21	0.84	0.90
	"Black holes"	0.10	0.96**	0.96**

R^2 = R^2 correlation coefficient; SE = sensitivity; SP = specificity; AC = accuracy.
* $P < 0.05$.
** $P < 0.01$.

Our evaluation indicated a high volumetric agreement between the automated 3ch-TDS$^+$, 2ch-TDS$^+$ and the manual segmentation. We also used a strict pixel-by-pixel accuracy estimation that takes into account not only volumetric agreement between the automated segmentation and manual segmentation but also their spatial overlap. Our results indicate that, compared with 2ch-TDS$^+$, 3ch-TDS$^+$ has higher sensitivity for detecting T2 lesions while maintaining similar accuracy and specificity. We noticed that lesion volume estimates were systematically higher in the 3ch-TDS$^+$ than the manual segmentation. This systematic bias is likely due to different thresholds set in the two different segmentation approaches because the calibration of the automated approach was set by a different expert than those outlining the lesions in order to ensure objective validation. Consensus between the criteria used by the tissue training process and the manual segmentation, or use of the experts' manual segmentation as training points can further reduce the systematic variability (Anbeek et al., 2004). Sensitivity, specificity and accuracy, as well as volumetric measures of agreement for the lesion subtypes (enhancing lesions and "black holes"), were higher for 3ch-TDS$^+$ than corresponding metrics for T2 lesions using 2ch-TDS$^+$. We therefore conclude that overall 3ch-TDS$^+$ is superior to 2ch-TDS$^+$.

Studies have shown that the sensitivity of segmentation algorithms can vary as a function of lesion load and have ascribed this relationship to small errors (in absolute volume) having a relatively larger effect with respect to sensitivity when the total lesion load the error refers to is also small. We also found that lesion load had an impact on the sensitivity of both the 2ch-TDS$^+$

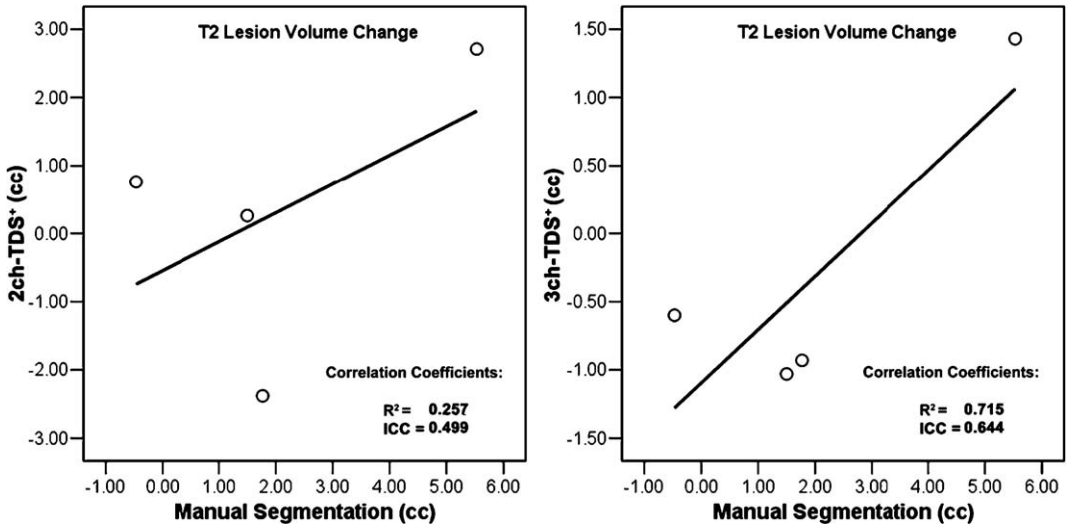

Fig. 8. Accuracy of T2 lesion change estimates. Compared to the 2ch-TDS$^+$, the scatter plots show the improved correlation between the 3ch-TDS$^+$ and manual segmentation for T2 lesion volume change over time on four MS patients scanned twice (average interval = 5.6 months).

and the 3ch-TDS$^+$. It appeared that sensitivity increased with increasing lesion load to stabilize above a threshold around 10 cc. This finding is particularly important when performing power calculations and determining overall study design for different phenotypes and stages of MS that have different average lesion loads. According to our findings, the presented methods would likely demonstrate better sensitivity in secondary progressive or advanced relapsing-remitting patients, which typically present higher lesion loads, than in patients with early MS or clinically isolated syndromes (CIS). In this initial investigation, the average sensitivity, specificity and accuracy were based on MR scans from a limited number of MS patients with a broad range of EDSS scores and lesion load. Larger studies would be needed to better characterize lesion load and distribution variables impacting the sensitivity of segmentation methods.

Pre- or post-contrast T1WI have been used to identify "black holes" (Miller et al., 1998; Van Waesberghe et al., 1998). We use post-contrast T1WI to identify "black holes" because some "black holes" on pre-contrast T1WI are presented as enhancing lesions on contrast T1WI (Van Waesberghe et al., 1998). Chronic non-enhancing "black holes" are more representative of chronic degenerative damage, as opposed to transient edematous changes (Van Walderveen et al., 1995). "Black holes" and "other T2 lesions" are hyperintense on T2WI, and most of enhancing lesions are hyperintense on T2WI. In rare cases, enhancing lesions are not hyperintense on T2WI. We derived the volume of T2 lesions in the 3ch-TDS$^+$ by combining enhancing lesions, "black holes" and "other T2 lesions", which may also include the non-T2 hyperintense enhancing lesions, yet this fraction should be minimal.

In conclusion, we have developed a 3ch-TDS$^+$ that can be used for additional automated segmentation of subtypes of MS lesions with improved sensitivity and high specificity and accuracy.

Acknowledgments

The authors are grateful to Mariann Polgar-Turcsanyi and Marlieke De Vos for their assistance. The authors thank Kelly H. Zou for consultations on statistical issues. This work was supported in parts by a grant from the U.S. National Multiple Sclerosis Society (RG 3574-A-1); a grant form the Dutch Foundation for MS Research (Voorschoten, The Netherlands); grants from the National Institutes of Health (P41 RR13218-01; R01 NS35142; AGO9675; P01 AG0495316; and 2R44MH57200-002) and a grant from the Foundation for Neurological Diseases.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2006.04.211.

References

Admiraal-Behloul, F., van den Heuvel, D.M., Olofsen, H., van Osch, M.J., van der Grond, J., van Buchem, M.A., Reiber, J.H., 2005. Fully automatic segmentation of white matter hyperintensities in MR images of the elderly. NeuroImage 28, 607–617.

Anbeek, P., Vincken, K.L., van Osch, M.J., Bisschops, R.H., van der Grond, J., 2004. Probabilistic segmentation of white matter lesions in MR imaging. NeuroImage 21, 1037–1044.

Ashton, E.A., Takahashi, C., Berg, M.J., Goodman, A., Totterman, S., Ekholm, S., 2003. Accuracy and reproducibility of manual and semiautomated quantification of MS lesions by MRI. J. Magn. Reson. Imaging 17, 300–308.

Barkhof, F., van Walderveen, M., 1999. Characterization of tissue damage in multiple sclerosis by nuclear magnetic resonance. Philos. Trans. R. Soc. London, Ser. B Biol. Sci. 354, 1675–1686.

Barkhof, F., McGowan, J.C., van Waesberghe, J.H., Grossman, R.I., 1998. Hypointense multiple sclerosis lesions on T1-weighted spin echo magnetic resonance images: their contribution in understanding multiple sclerosis evolution. J. Neurol., Neurosurg. Psychiatry 64, S77–S79.

Barkhof, F., van Waesberghe, J.H., Filippi, M., et al., 2001. T(1) hypointense lesions in secondary progressive multiple sclerosis: effect of interferon beta-1b treatment. Brain 124, 1396–1402.

Bedell, B.J., Narayana, P.A., 1998. Automatic segmentation of gadolinium-enhanced multiple sclerosis lesions. Magn. Reson. Med. 39, 935–940.

Bruck, W., Lucchinetti, C., Lassmann, H., 2002. The pathology of primary progressive multiple sclerosis. Mult. Scler. 8, 93–97.

Clarke, L.P., Velthuizen, R.P., Phuphanich, S., Schellenberg, J.D., Arrington,

J.A., Silbiger, M., 1993. MRI: stability of three supervised segmentation techniques. Magn. Reson. Imaging 11, 95–106.
Filippi, M., Rovaris, M., Campi, A., Pereira, C., Comi, G., 1996. Semi-automated thresholding technique for measuring lesion volumes in multiple sclerosis: effects of the change of the threshold on the computed lesion loads. Acta Neurol. Scand. 93, 30–34.
Filippi, M., Horsfield, M.A., Hajnal, J.V., et al., 1998. Quantitative assessment of magnetic resonance imaging lesion load in multiple sclerosis. J. Neurol., Neurosurg. Psychiatry 64, S88–S93.
Filippi, M., Rovaris, M., Rocca, M.A., Sormani, M.P., Wolinsky, J.S., Comi, G., 2001. Glatiramer acetate reduces the proportion of new MS lesions evolving into "black holes". Neurology 57, 731–733.
Friedman, J.H., Baskett, F., Shustek, L.J., 1975. An algorithm for finding nearest neighbors. IEEE Trans. Comput. C-24, 1000–1006.
Gawne-Cain, M.L., Webb, S., Tofts, P., Miller, D.H., 1996. Lesion volume measurement in multiple sclerosis: how important is accurate repositioning? J. Magn. Reson. Imaging 6, 705–713.
Guttmann, C.R., Ahn, S.S., Hsu, L., Kikinis, R., Jolesz, F.A., 1995. The evolution of multiple sclerosis lesions on serial MR. Am. J. Neuroradiol. 16, 1481–1491.
Guttmann, C.R., Kikinis, R., Anderson, M.C., et al., 1999. Quantitative follow-up of patients with multiple sclerosis using MRI: reproducibility. J. Magn. Reson. Imaging 9, 509–518.
Guttmann, C.R., Benson, R., Warfield, S.K., et al., 2000. White matter abnormalities in mobility-impaired older persons. Neurology 54, 1277–1283.
He, R., Narayana, P.A., 2002. Automatic delineation of Gd enhancements on magnetic resonance images in multiple sclerosis. Med. Phys. 29, 1536–1546.
Johnston, B., Atkins, M.S., Mackiewich, B., Anderson, M., 1996. Segmentation of multiple sclerosis lesions in intensity corrected multispectral MRI. IEEE Trans. Med. Imaging 15, 154–169.
Kikinis, R., Shenton, M.E., Gerig, G., 1992. Routine quantitative analysis of brain and cerebrospinal fluid spaces with MR imaging. J. Magn. Reson. Imaging 2, 619–629.
Kikinis, R., Guttmann, C.R., Metcalf, D., et al., 1999. Quantitative follow-up of patients with multiple sclerosis using MRI: technical aspects. J. Magn. Reson. Imaging 9, 519–530.
Mackiewich, B., 1995. Intracranial boundary detection and radio frequency correction in magnetic resonance images. Master's thesis, Simon Fraser Univ. Computer Science Dept., Burnaby, British Columbia.
Miller, D.H., Barkhof, F., Nauta, J.J., 1993. Gadolinium enhancement increases the sensitivity of MRI in detecting disease activity in multiple sclerosis. Brain 116, 1077–1094.
Miller, D.H., Grossman, R.I., Reingold, S.C., McFarland, H.F., 1998. The role of magnetic resonance techniques in understanding and managing multiple sclerosis. Brain 121, 3–24.
Molyneux, P.D., Filippi, M., Barkhof, F., et al., 1998. Correlations between monthly enhanced MRI lesion rate and changes in T2 lesion volume in multiple sclerosis. Ann. Neurol. 43, 332–339.
Paty, D.W., Li, D.K., 2001. Interferon beta-lb is effective in relapsing-remitting multiple sclerosis: II. MRI analysis results of a multicenter, randomized, double-blind, placebo-controlled trial (1993 [classical article]). Neurology 57, S10–S15.
Rovaris, M., Filippi, M., Calori, G., et al., 1997. Intra-observer reproducibility in measuring new putative MR markers of demyelination and axonal loss in multiple sclerosis: a comparison with conventional T2-weighted images. J. Neurol. 244, 266–270.
Samarasekera, S., Udupa, J.K., Miki, Y., Wei, L., Grossman, R.I., 1997. A new computer-assisted method for the quantification of enhancing lesions in multiple sclerosis. J. Comput. Assist. Tomogr. 21, 145–151.
Truyen, L., van Waesberghe, J.H., van Walderveen, M.A., et al., 1996. Accumulation of hypointense lesions ("black holes") on T1 spin-echo MRI correlates with disease progression in multiple sclerosis. Neurology 47, 1469–1476.
Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P., 2001. Automated segmentation of multiple sclerosis lesions by model outlier detection. IEEE Trans. Med. Imaging 20, 677–688.
Van Waesberghe, J.H., van Walderveen, M.A., Castelijns, J.A., et al., 1998. Patterns of lesion development in multiple sclerosis: longitudinal observations with T1-weighted spin-echo and magnetization transfer MR. Am. J. Neuroradiol. 19, 675–683.
Van Walderveen, M.A., Barkhof, F., Hommes, O.R., et al., 1995. Correlating MRI and clinical disease activity in multiple sclerosis: relevance of hypointense lesions on short-TR/short-TE (T1-weighted) spin-echo images. Neurology 45, 1684–1690.
Van Walderveen, M.A., Kamphorst, W., Scheltens, P., et al., 1998. Histopathologic correlate of hypointense lesions on T1-weighted spin-echo MRI in multiple sclerosis. Neurology 50, 1282–1288.
Vinitski, S., Gonzalez, C., Mohamed, F.B., Iwanaga, T., Kamil Khalili, K., Mack, J., 1997. Improved intracranial lesion characterization by tissue segmentation based on a 3D feature map. Med. Risk Manag. 37, 457–469.
Warfield, S.K., 1996. Fast *k*-NN classification for multichannel image data. Pattern Recogn. Lett. 17, 713–721.
Warfield, S., Dengler, J., Zaers, J., et al., 1995. Automatic identification of gray matter structures from MRI to improve the segmentation of white matter lesions. J. Image Guid. Surg. 1, 326–338.
Warfield, S.K., Kaus, M., Jolesz, F.A., Kikinis, R., 2000. Adaptive, template moderated, spatially varying statistical classification. Med. Image Anal. 4, 43–55.
Wei, X., Warfield, S.K., Zou, K.H., et al., 2002. Quantitative analysis of MRI signal abnormalities of brain white matter with high reproducibility and accuracy. J. Magn. Reson. Imaging 15, 203–209.
Wei, X., Guttmann, C.R., Warfield, S.K., Eliasziw, M., Mitchell, J.R., 2004. Has your patient's multiple sclerosis lesion burden or brain atrophy actually changed? Mult. Scler. 10, 402–406.
Weiner, H.L., Guttmann, C.R., Khoury, S.J., et al., 2000. Serial magnetic resonance imaging in multiple sclerosis: correlation with attacks, disability, and disease stage. J. Neuroimmunol. 104, 164–173.
Wells, W.M., Grimson, W.E.L., Kikinis, R., Jolesz, F., 1996. Adaptive segmentation of MRI data. IEEE Trans. Med. 15, 429–443.
Wu, Y., Wells, W.M., Warfield, S.K., et al., 2001. Effect of gain field inhomogeneity correction on automated segmentation of the intracranial cavity. Proceedings of the 10th Annual Meeting of ISMRM, Glasgow, p. 805.
Zijdenbos, A.P., Forghani, R., Evans, A.C., 2002. Automatic "pipeline" analysis of 3-D MRI data for clinical trials: application to multiple sclerosis. IEEE Trans. Med. Imaging 21, 1280–1291.
Zou, K.H., Wells, W.M. III, Kaus, M.R., Kikinis, R., Jolesz, F.A., Warfield, S.K., 2002. Statistical validation of automated probabilistic segmentation against composite latent expert ground truth in MR imaging of brain tumors. MICCAI 2002, 5th International Conference, Sept. 2002, Tokyo, Japan. Springer-Verlag, Berlin, pp. 315–322.