Segmentation of Multiple Sclerosis Lesions in Intensity Corrected Multispectral MRI

B. Johnston, M. S. Atkins,* B. Mackiewich, *Member, IEEE*, and M. Anderson

Abstract— To segment brain tissues in magnetic resonance images of the brain, we have implemented a stochastic relaxation method which utilizes partial volume analysis for every brain voxel, and operates on fully three–dimensional (3-D) data. However, there are still problems with automatically or semi-automatically segmenting thick magnetic resonance (MR) slices, particularly when trying to segment the small lesions present in MR images of multiple sclerosis patients. To improve lesion segmentation we have extended our method of stochastic relaxation by both pre- and post-processing the MR images. The preprocessing step involves image enhancement using homomorphic filtering to correct for nonhomogeneities in the coil and magnet. Because approximately 95% of all multiple sclerosis lesions occur in the white matter of the brain, the post-processing step involves application of morphological processing and thresholding techniques to the intermediate segmentation in order to develop a mask image containing only white matter and Multiple Sclerosis (MS) lesion. This white/lesion masked image is then segmented by again applying our stochastic relaxation technique. The process has been applied to multispectral MRI scans of multiple sclerosis patients and the results compare favorably to manual segmentations of the same scans obtained independently by radiology health professionals.

I. INTRODUCTION

MAGNETIC Resonance Imaging (MRI) is a noninvasive method for imaging internal tissues and organs. The problems of image segmentation (the identification and quantitation of tissues and organs) of MRI images have been described extensively in the literature and many algorithms have been developed in an attempt to solve the problems. One of the goals of segmentation of MRI images is to determine the volumes of organs, tissues and lesions present in a given patient. These volumes, and the changes in these volumes over time, may aid in the diagnosis, prognosis and treatment planning of patients under investigation.

Manual tracing of structural boundaries is the most basic of all segmentation techniques and is still quite commonly used in clinical and research settings [1]–[3]. These techniques

suffer from operator bias and the results are rarely reproducible between operators. Automatic methods of image segmentation have been investigated extensively and include edge detection methods [4]–[6], thresholding and clustering [7]–[12], and region growing based methods [13], [14]. Extensions to the fundamental techniques of image segmentation have appeared recently and include techniques which employ mathematical morphology [15], pyramidal or hierarchical methods [16]–[18] and model based approaches [19]–[22]. Edge detection methods suffer from heavy computational complexity and sensitivity to noise and image artifacts. Thresholding, region growing and morphological methods usually require subjective operator selection of input parameters. The hierarchical and model driven approaches have achieved reasonable success in recent years. However, the model based techniques, which incorporate external knowledge about structure and physiology, are limited by the requirement of the often costly and difficult development of a tissue model. The hierarchical methods have yet to be fully explored in MRI.

We have developed a hybrid of several approaches to image segmentation in order to achieve good segmentations of tissues from MRI brain images. In this paper we describe extensions to our segmentation method based on a model-based method of image segmentation which applies a variation of stochastic relaxation in three dimensions [23].

One of the major difficulties we encounter when applying segmentation techniques to MRI images is the nonuniform intensity levels seen in tissues of identical composition. This nonuniformity is caused by variations in the sensitivities of the instrumentation used in obtaining MRI images. We have enhanced our algorithm so that it accounts for the nonuniformities in signal and exploits the benefits of several different segmentation techniques including automatic thresholding, mathematical morphology and statistical analysis. We have applied the algorithm to segment the lesions seen in dual echo MRI data sets of Multiple Sclerosis patients.

Multiple Sclerosis: Multiple Sclerosis (MS) is a disease of the central nervous system (CNS), i.e., the brain and spinal cord [24]. It is a debilitating and progressive disease which may result in a variety of symptoms from blurred vision to severe muscle weakness and degradation, depending on the area of the CNS which is affected. Multiple Sclerosis is caused by a breakdown in the myelin sheath, a soft, white, fatty material which insulates the neurons of the CNS and provides for the rapid transmission of nerve impulses along the nerve fibers of the brain and spinal cord. MS lesions go through various hist-pathological stages and are heterogeneous
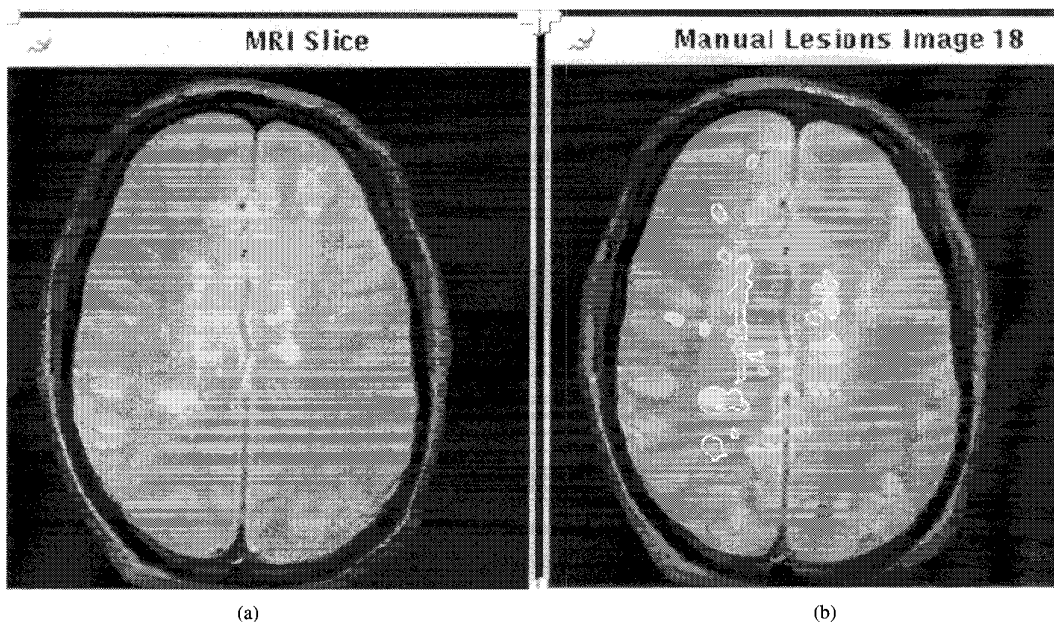
(a)   (b)

Fig. 1. (a) MRI image with (b) MS lesions manually outlined by radiologist.

in their presentation. Some chronic lesions may contain "scar" tissue composed of astrocytic fibers. MRI detects not only these chronic astrocytic lesions but also the acute edematous patches, with varying amounts of demyelination and inflammatory cell infiltration, typical of newer lesions. These lesional components (referred to as MS plaques) are also readily observable post-mortem. The inherent heterogeneity of lesions, which is reflected on MR images, causes some of the difficulties encountered while attempting to separate lesions from healthy brain tissue and cerebro-spinal fluid compartments using segmentation approaches based on signal intensity characteristics alone.

Fig. 1 shows MS plaques in a typical slice from a proton density (PD) weighted MRI image. Fig. 1(a) shows the image itself while Fig. 1(b) has been overlaid with hand-drawn regions of interest (ROI's) indicating the presence of MS lesions. Note how the lesions are detected almost invariably in the white matter of the brain.

Observation of the characteristic plaques over time in MS patients shows how the demyelination of nerves is a reversible and recurrent process in these patients. In a majority of patients, over time with and without treatment, MS lesions grow and shrink and some lesions may show confluence with others. It is important to be able to quantitatively assess the numbers and sizes of MS lesions in patients undergoing therapy for the disease, and several large scale clinical trials are under way to determine which quantitative measures obtained from the MR images are most useful for assessing the impact of drugs on the disease.

The hand-drawn outlines of Fig. 1(b) are representative of the typical manner in which lesions are quantified in many centres. Much recent research in MS using MRI has been directed toward automating the quantitation of MS lesions in the brain [12], [20], [23], [25]–[29]. One of the goals of our research is to provide physicians with the ability to accurately track the progress of MS patients undergoing drug therapy [30]. The following discussion presents the approach we have taken and compares the results we obtained with that of manually segmented data.

II. METHODS

A. Overview

Fig. 2 illustrates the system we developed to segment MS lesions from MRI volumes of part of a patient's brain, typically acquired in a single double-echo sequence (PD and T2-weighted). Following acquisition of the multispectral MRI data, the images are scaled, the nonbrain tissue is masked out and the data is corrected for radio frequency (RF) inhomogeneities. Each spectral image volume is then segmented independently using a modified iterated conditional modes (ICM) algorithm in 3-D [23], [31]. The result is a partial volume solution indicating for each voxel a percentage composition of each of several tissue types. Following this step the segmentations of each spectrum are then combined and normalized. The white matter and lesion segmentations are then extracted and summed, and morphological processing and automatic thresholding are applied in order to obtain a white matter/lesion mask. This mask is then applied to each of the RF-corrected data sets, yielding images containing only white matter and lesion. The ICM algorithm is then re-applied to the masked images, and the results are combined and normalized once again. As a final step in the segmentation small lesions are filtered out and the result is then ready for comparison with manual segmentations of the same data.

B. Acquisition

Five MRI volumes have been acquired according to the specifications found in Table I. We used a single dual echo MRI sequence (PD-weighted and T2-weighted), each consisting of between 21 and 27 256 × 256 slices acquired in the
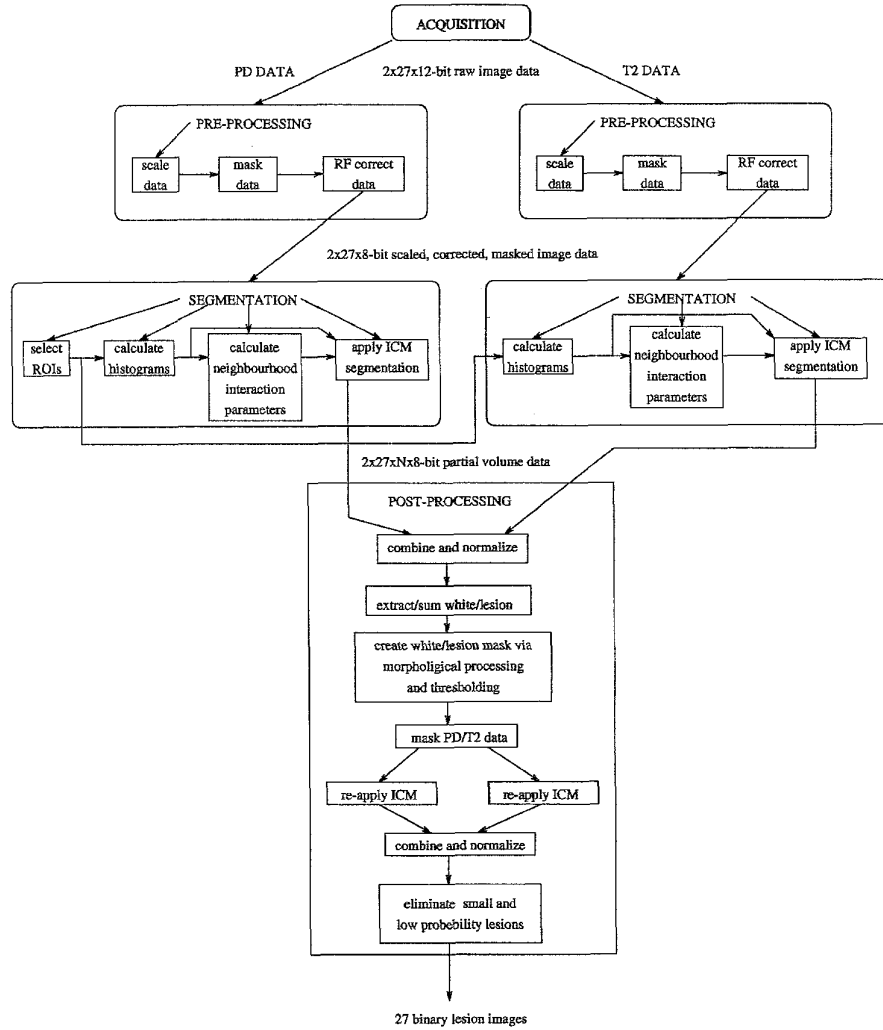
Fig. 2.   Overview of segmentation process. N is the number of tissues, including lesion, deemed to be present in the volume.

TABLE I
MRI Data Acquisition Specifications from a Single Double-Echo Scan

Parameter	PD Weighted	T2 Weighted
Sequence Type	Spin Echo	Spin Echo
Field of View	200mm²	200mm²
Repetition Time	2000ms	2000ms
Echo Time	35ms	70ms
Pixel size	0.78mm²	0.78mm²
Slice thickness	5.0mm	5.0mm
Mode of Acquisition	Interleaved	Interleaved
Acquisition plane	Axial	Axial

axial plane on a General Electric 1.5 Tesla MRI scanner. The slices are contiguous. The echo times were chosen so that the cerebrospinal fluid (CSF) was low intensity (dark) on the first echo (the PD scan) and high intensity (bright) on the

second echo. These sequences were chosen to maximize lesion detection and facilitate manual lesion tracing. In MRI data set 1 each voxel measures 0.78 mm × 0.78 mm × 5.0 mm in the $x, y,$ and z planes, respectively; in MRI data sets 2 to 5, each voxel measures 0.82 mm × 0.82 mm × 5.00 mm.

C. Preprocessing

1) Scaling: The images first have to be scaled and the nonbrain tissue masked out. The original 12-b data has been scaled to 8 b according to the maximum intensity M, obtained in the entire volume. For any original voxel intensity i, the scaled value i' is given by

$$i' = i/M * 255.0 + 0.5.$$

2) Masking: The tissues surrounding the brain, e.g., meninges, skin, fat, bone, share intensity profiles with all the tissues of interest that we are attempting to extract. The task of automatically or semiautomatically masking or otherwise extracting the noninteresting brain tissues has been
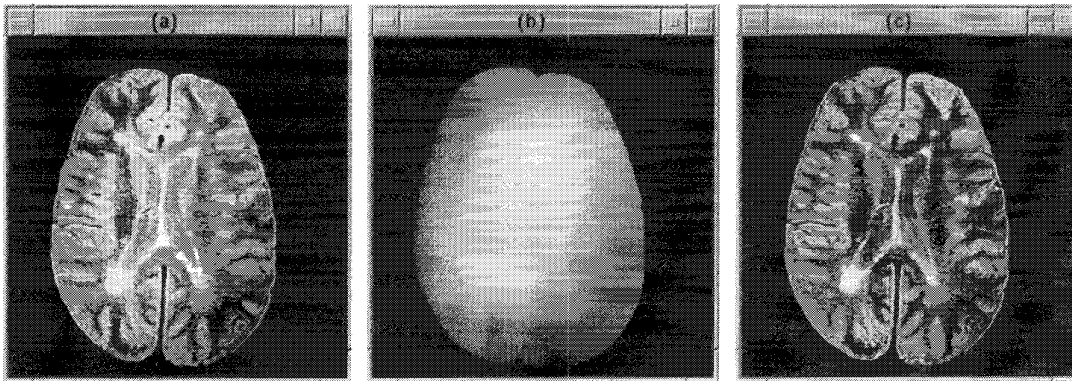
Fig. 3.   PD-weighted image data of slice 16 of MRI data set 2 (a) before and (b) after (c) RF-correction. (b) shows the detected RF inhomogeneity, $(a)/(c)$.

investigated by other authors [20], [26], [32]–[34]. We have developed a novel method of automatically masking the brain which works on all our data sets without any user interaction [35].

   3) RF-Correction: The nonuniformities in MRI signal over tissues of similar composition present a major deterrent to the solution of the segmentation problem [20], [28], [33], [34]. Both inter-planar and intra-planar variations in tissue intensity occur. These variations have a pronounced effect on algorithms which use histogram or cluster analysis, especially where the algorithms are trained by obtaining sample regions of interest (ROI's) from a limited number of sites.

   Axel et al. use MRI "*phantoms*" for intensity correction [36]. A phantom is the MR image produced by scanning a homogeneous object, such as cylinder of a silicon solution. The phantom profiles the unwanted intensity variation and can therefore be used to reduce that variation. Because we do not have access to the scanners from which our MR images were scanned, we cannot use phantoms for correction. For situations when phantom data is not available, in the same paper, Axel et al. suggest low pass filtering the original MR image to approximate the phantom. Others have implemented similar approaches [29], [37]–[41]. We have found that this approach undesirably reduces contrast between brain tissues and produces artifacts in the "corrected" volumes.

   Wells et al. combine statistical segmentation and intensity correction in a two-step iterative approach called expectation maximization (EM) [28]. Although apparently effective in the presence of RF inhomogeneity, the EM algorithm does not account for partial volumes. Our algorithm, therefore, potentially offers superior segmentation once RF inhomogeneities have been reduced.

   Dawant et al. correct intra-planar variations by fitting a surface to interactively defined points [42]. Since all of these points must correspond to similar tissues, the user must possess some anatomical knowledge, or have access to a segmented image. The inter-slice intensity variation is estimated by comparing the intensities of similar tissue voxels in adjacent slices. Although proven effective, this method is not automatic and requires a priori knowledge of tissue types.

   We apply a method of RF-correction based on an equivalent of homomorphic filtering which is fast, simple, and effective [35], [43]. In our method, the average intensity of brain voxels,

defined by the automatically generated brain mask, is first assigned to all voxels outside the brain. This masking operation eliminates strong edges such as the intracranial cavity. Next, the following nonlinear filter is applied to the masked volume [44]

$$I'(x, y, z) = \exp\{\ln(I(x, y, z)) - \mathrm{lpf}[\ln(I(x, y, z))]\} \quad (1)$$

where $I(x, y, z)$ is the masked volume and "lpf" denote a low-pass filter operation. Because $I(x, y, z)$ contains no strong edges, $\exp\{\mathrm{lpf}[\ln(I(x, y, z))]\}$ approximates the intensity variation due to RF inhomogeneity. Finally, $I'(x, y, z)$ is normalized (to the range, [0, 255], in this case) producing the corrected volume, $I_c(x, y, z)$

$$I_c(x, y, z) = \frac{I'(x, y, z) - \min[I'(x, y, z)]}{\max[I'(x, y, z)] - \min[I'(x, y, z)]} \cdot 255. \quad (2)$$

   Fig. 3 illustrates the effect of RF-correction on PD-weighted MRI data. Note the subtle darkening in the lower right quadrant and the bright interior region of the raw data in Fig. 3(a). The result of applying RF-correction to Fig. 3(a) is given in Fig. 3(c). Fig. 3(b) contains the correction factor, given in (1), by which the raw image is divided. The gradual darkening of the image seen in Fig. 3(a) is noticeably reduced in Fig. 3(c). We have quantitatively examined this inhomogeneity by measuring the variation in mean brain tissue intensities in each dimension. Fig. 4 shows how the gradual intensity variation due to RF inhomogeneity is reduced by application of RF-correction using a 32-sample correction. In all cases studied, the brain tissue intensity variance was reduced by at least 33%, as seen in Table II. For both PD and T2-weighted images, the y-axis correction has the largest effect (60%–70%). However, high frequency variations, representing contrast between brain tissue types, remain intact, as can be seen in Fig. 4.

D. Segmentation

   Following scaling, masking, and correction the data is now ready to be segmented using our variation of ICM first used by Besag [45] to restore noisy and corrupted images. Using this method an image is assumed to represent a locally dependent Markov random field, exploiting the assumption that voxels in a given local neighborhood of the image have similar intensity. For a detailed description of our modified
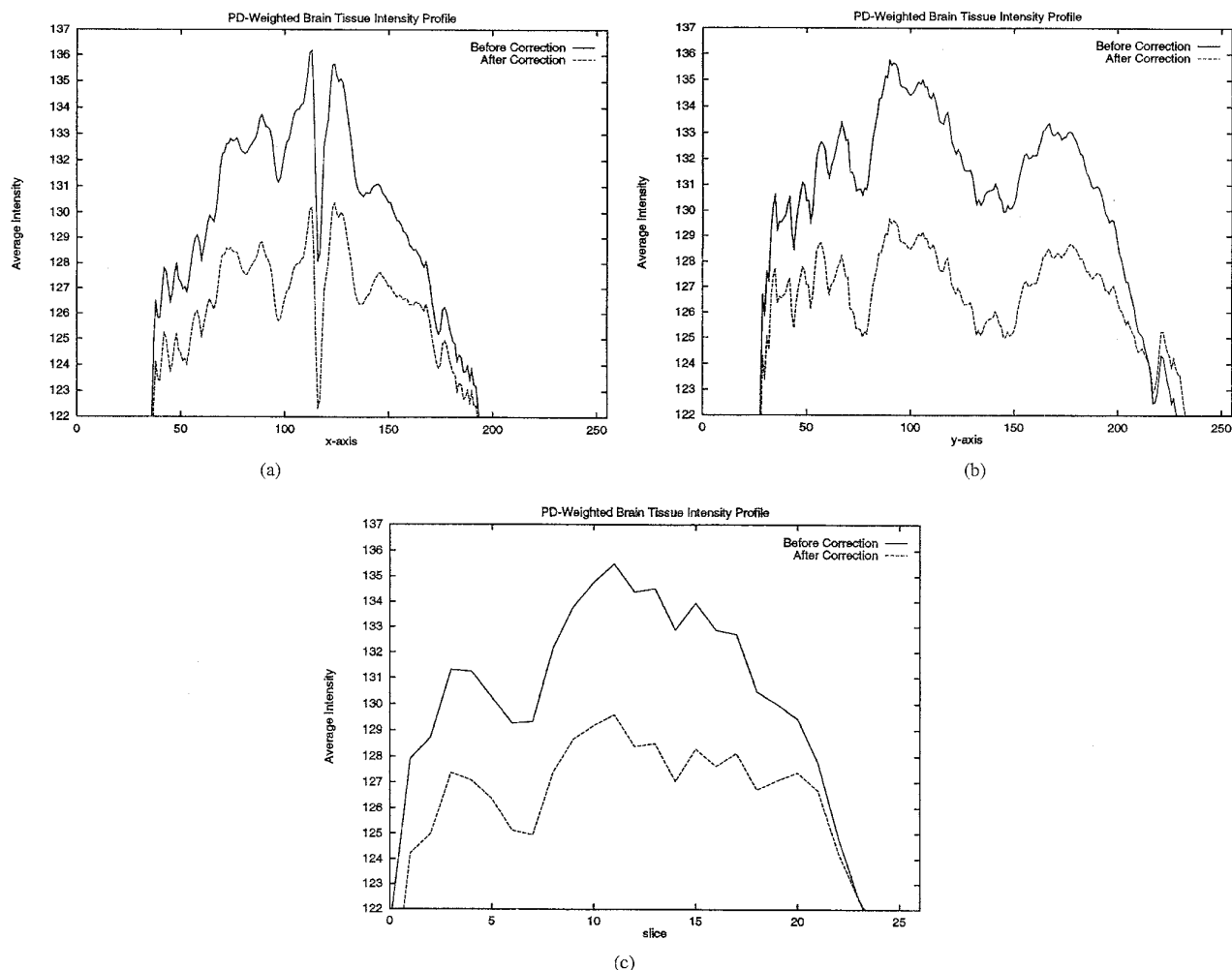
Fig. 4. Average brain tissue intensity variation in each dimension before and after RF correction. The uncorrected and corrected images from which the graphs were calculated have equivalent dynamic ranges. (a) x-axis, (b) y-axis, and (c) slice.

TABLE II
RF-CORRECTION—EACH ROW SHOWS THE BRAIN TISSUE INTENSITY
VARIANCES FOR RAW (UNCORRECTED) DATA AND RF-CORRECTED DATA

Profile		Raw Data	Corrected	Reduction (%)
PD	x	39.3	26.0	33.8
	y	44.8	17.5	60.9
	z	31.8	13.9	56.4
T2	x	28.8	15.2	47.3
	y	64.9	18.9	70.8
	z	6.05	3.78	37.6

ICM algorithm see [23], [31], and [46]. The problem is to maximize the probability of a candidate segmentation, given the observed data, a neighbor interaction model, and some *a priori* knowledge about the various tissues under consideration. This prior knowledge comes from selecting a minimum number of ROI's containing "pure" samples

of each tissue of interest. These ROI's are then used to compute histogram profiles of each tissue type in the volume. Without any RF-correction, we found that tissue samples were needed from almost every slice before a reasonable segmentation could be achieved, as will be shown in Section IIIon "Quantitative Analysis". The number of ROI's required to give an accurate profile of each tissue is reduced, however, by the RF-correction procedures outlined above so that ROI's from one or two slices will usually suffice to achieve a reasonable segmentation of the surrounding four slices, which corresponds to a 40-cm-wide band within the head. This is shown in the Section III. The histograms so obtained are used to determine an initial segmentation which in turn is used to develop a set of neighborhood interaction parameters, indicating the strength of interaction between different tissue types.

A final segmentation is achieved iteratively through successive applications of the ICM algorithm to each voxel in the volume until convergence is achieved, usually after five to eight iterations of the algorithm. There are a number of methods of measuring convergence of the algorithm, usually
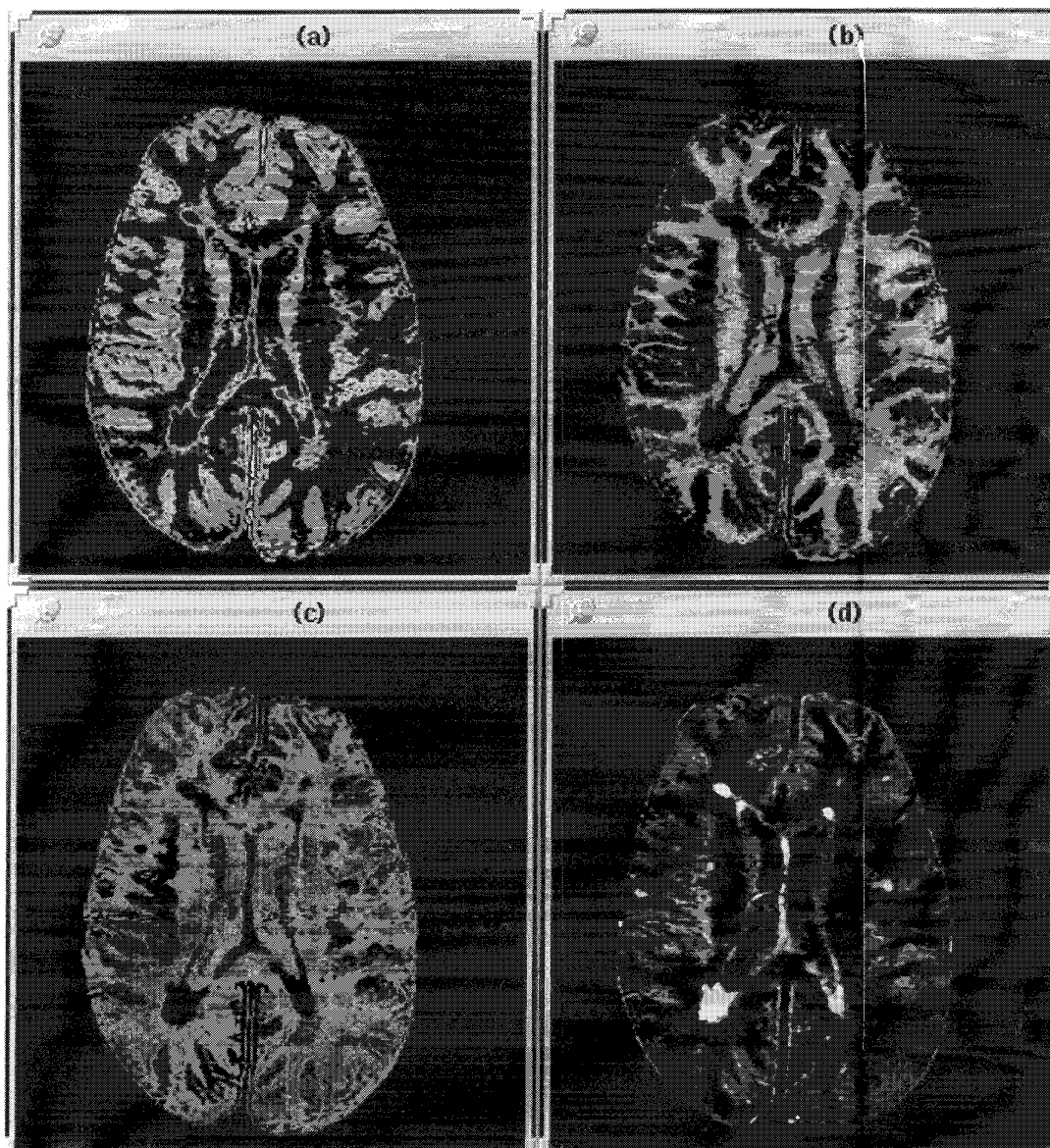
Fig. 5. Partial Volume Segmentation of slice 16 from the PD-weighted scan. (a) gray matter, (b) white matter, (c) CSF, (d) lesion.

done by placing a limit on the number of voxels which *change* from one iteration to the next. With a partial volume solution to the problem, the number of voxels which *change* between iterations is not readily measurable. However we have adopted a method which tracks the change in $p(\hat{x}_{ij})$ between iterations, resulting in a vector of values for each $\hat{x}_i$, indicating the change in each tissue. A scalar measure of the change in voxel i, on iteration m, denoted $|c_{i_m}|$, is obtained by calculating the 2-Norm of this vector such that

$$|c_{i_m}| = \sqrt{\sum_k (p(\hat{x}_{ik})_m) - p(\hat{x}_{ik})_{(m-1)})^2}.$$

The user may then assess changes between successive iterations and apply suitable threshold criteria on the number of changes required per iteration, in order to allow convergence to be achieved within five to eight iterations. In practice we

have found that five iterations are always sufficient for the segmentations. The computation time for one iteration over a 27-slice 256×256 scan is around five min on a Sun SPARC 2 workstation.

The resulting segmentation constitutes a maximum probability of the classification of each voxel given the original data and the calculated neighborhood interaction model. Following convergence, a partial volume estimate is obtained, where, for each voxel, a set of probabilities is obtained indicating the percentage composition of each tissue in the voxel. We carried out the segmentation step using our 3-D version of the ICM algorithm on each of the PD-weighted and T2-weighted sequences independently, as we found it useful to view the segmentations of the PD and T2 images separately (see Section III). Furthermore, the four-tissue partial volume segmentation of a 27-slice 8-b MRI scan occupies seven
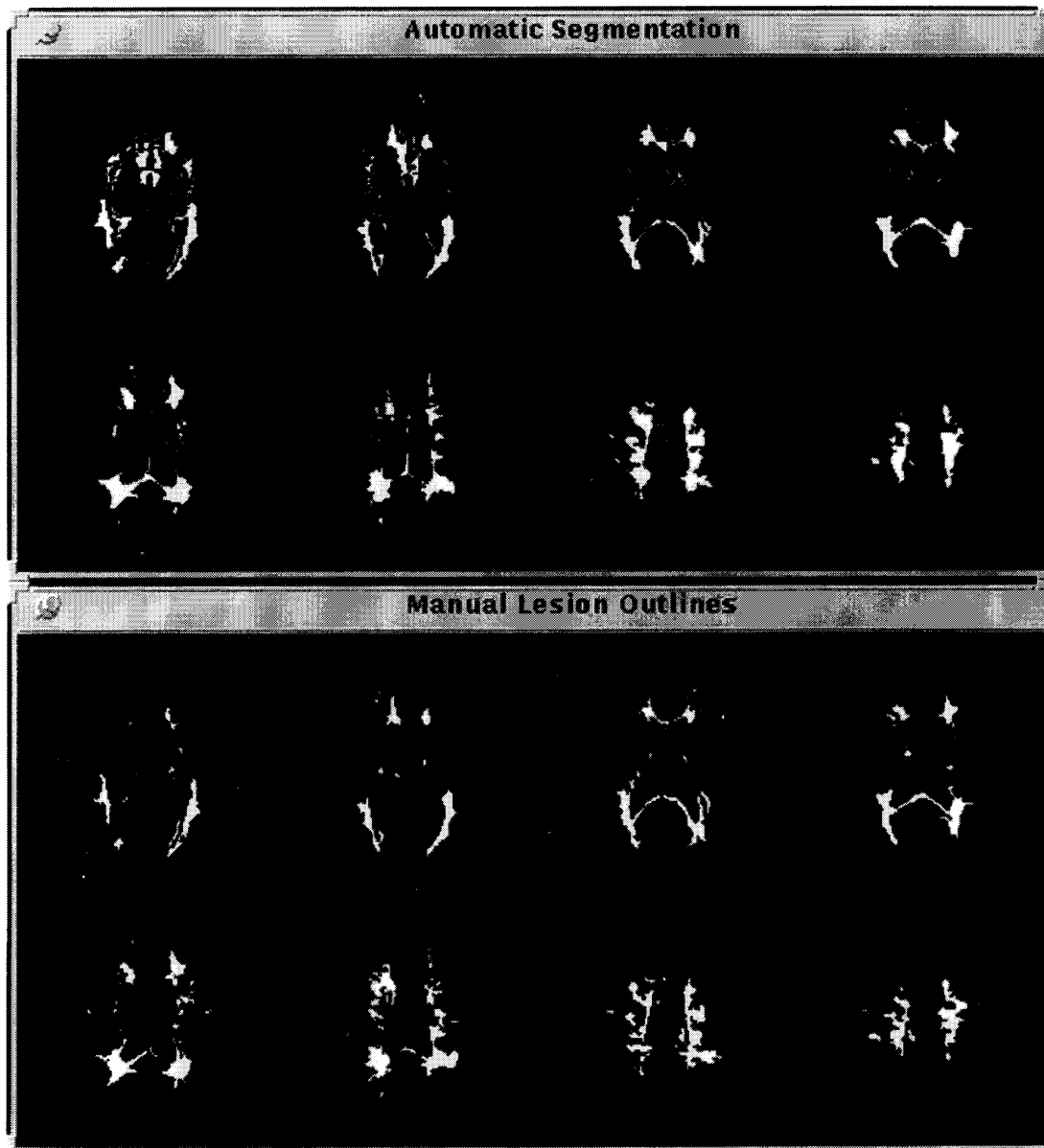
Fig. 6.   Binary Lesion Segmentation of central slices of another PD-weighted data set compared with the manual segmentations.

megabytes and we encountered severe computer performance problems when we attempted to use the multispectral data sets during segmentation, probably due to the computer memory requirements.

Fig. 5 shows an example of the output of this step for the PD-weighted sequence from slice 16 of the study. The output consists of a set of N images for each slice in the volume, where N is the number of tissues present in the volume (in this case $N = 4$). The intensity of each voxel indicates the probability that that voxel is of the tissue type represented in the image. There is an equivalent set of images obtained from the T2 sequence.

Fig. 6 shows how good our ICM algorithm is at segmentation, using just PD data from another data set. The partial volume segmentation has been thresholded, so that all values between 250 and 255 (representing high probabilities of lesion)

have been set to one and all lower probabilities have been set to zero. The manual segmentation is shown below for comparison. Only the central slices are shown, as there are only a few small lesions in the outer slices.

E. Post-Processing

1) Introduction: Almost 95% of all MS lesions occur in the white matter of the brain [24], [20]. In both the PD-weighted and T2-weighted scans, healthy white matter presents as a dark tissue and MS lesion as very bright. Intensity profiles taken across the diameter of many MS lesions show a gradual falling off of intensity near the border of the lesion (here we are temporarily ignoring the complex histology of some MS lesions which show a darker center to the lesion).

Fig. 7 illustrates this phenomenon. A model of a slice from a typical PD-weighted scan is shown indicating the

Fig. 7.   Model of typical MS lesion occurring in the white matter of the brain. The intensity of lesion voxels decreases with distance from the centre of the voxel, traveling through an area characteristic of gray matter (see Fig. 8).
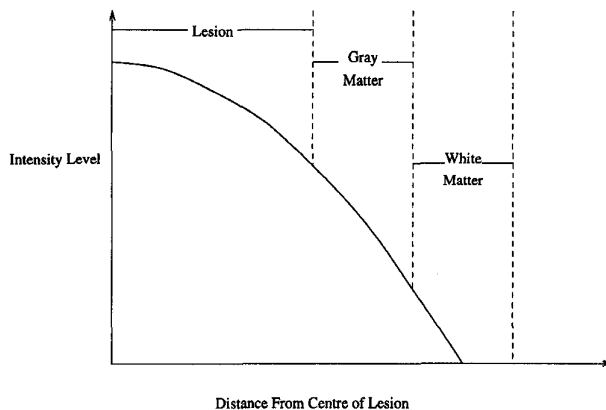

Fig. 8.   1-D intensity profile of typical MS lesion, as shown in Fig. 7, measured from the centre of the lesion and extending into the white matter. The profile travels through an area characteristic of gray matter.
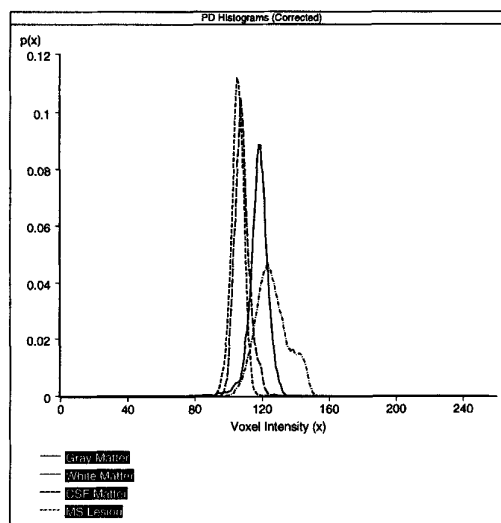

Fig. 9.   Tissue profiles obtained from ROI's selected from an RF-corrected PD-weighted scan.

presence of an MS lesion in the white matter of the brain. Fig. 8 shows the one-dimensional (1-D) intensity profile of the lesion measured from the centre of the lesion through its border and into the white matter. As the distance along the profile to the white matter decreases the intensity profile travels through an area which is characteristic of gray matter. This phenomenon is probably due to the partial volume effect: voxels near the borders of tissues contain contributions from two or more tissue types. The ultimate intensity of the voxel is an average value over all the tissue types it covers. This partial volume effect causes confusion of lesion with gray matter and results in significant mis-classification of lesion near its border. Furthermore, the histology of MS lesions is complex, so further partial volume effects may be observed within the lesions themselves, as well as at all tissue boundaries.

As is evident from Fig. 5, the result of the segmentation to this point is not adequate for accurate quantification of MS lesion. One of the reasons for the inaccuracy is the significant sharing of intensity profiles obtained from different tissues. This overlap is present in both the PD and T2-weighted data sets.

Figs. 9 and 10 illustrate the histogram profiles obtained for gray matter, white matter, CSF and MS lesion taken from the PD-weighted and T2-weighted scans, respectively. The histograms have been normalized and low-pass filtered to

reduce noise. In the PD histograms, overlap between white matter and CSF, and between gray matter and lesion is especially noticeable. The T2 histograms show considerable sharing of intensities between lesion and CSF, and between lesion and gray matter.

The overlap of white matter with CSF in the PD profile is the reason for the error visible in Fig. 5(b) and (c). The effect of the gray matter/lesion overlap is noticeable in Fig. 5(a) and (d). Because of this significant overlap of histogram profiles and the resulting errors introduced in the ICM segmentation it is necessary to post-process the images.

We apply two levels of post-processing to the PD/T2 segmentations. In the first step the two segmentations that have been achieved independently to this point are combined. This combination reduces, but does not eliminate, the confusion caused by the overlapping histograms. The second phase of post-processing exploits knowledge about the nature of MS lesions which occur almost exclusively in the white matter of the brain. We obtain a white matter/lesion mask and then segment the lesions from the masked image.
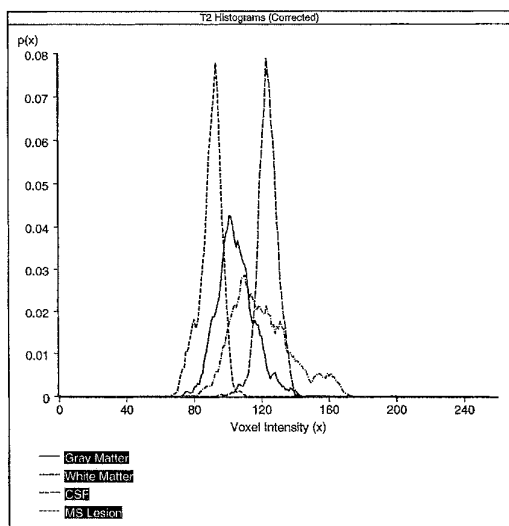
Fig. 10. Tissue profiles obtained from ROI's selected from an RF-corrected T2-weighted scan.

TABLE III

CONFUSION MATRIX—ILLUSTRATING THE CONFUSION OF GRAY MATTER WITH MS LESION. EACH ROW SHOWS THE PERCENTAGE OF EACH TISSUE WHICH HAS BEEN CLASSIFIED AS EVERY OTHER TISSUE (DATA IS FOR SLICES 11-20 ONLY)

Tissue	Percent of Manual Seg Classified by ICMSeg			
	Gray	White	CSF	Lesion
Gray	65.0(%)	7.9	6.4	20.7
White	3.9	95.9	0.0	0.2
CSF	10.2	0.5	82.9	6.4
Lesion	41.6	5.9	2.0	50.5

2) Partial Volume Product: As a first step in resolving the problem of histogram overlap, we will combine the results of the PD and T2 segmentations obtained by our ICM algorithm. We accomplish this by 1) multiplying the respective probabilities for each tissue type and 2) normalizing the result so that the total probability for any given voxel adds to one. We performed the combination by multiplying the probabilities obtained from segmenting the PD and T2 scans separately because it gave better resulting segmentations then either scan alone [31]. However, other methods of combining the partial volume segmentations may yield results which are better still, and we will investigate this in future. Other researchers such as Zijdenbos [27] report that using a third image such as a separately acquired T1-weighted image gives even better results; unfortunately we were unable to obtain a third matching data set as our data sets were already acquired as part of a clinical trial.

The output from this step is a single set of N images which reflect in some way the tissue partial volume and for simplicity, we assume that the resulting combined set of images represents the partial volume of each tissue. Fig. 11 shows the result of combining the result shown in Fig. 5 (PD segmentation) with the equivalent result from the T2 sequence.

Although an improvement over the separate segmentations, the results given in Fig. 11 still show a significant amount of confusion of gray matter with MS lesion. We have compared this segmentation with sample manual segmentations of all tissue types and produced a confusion matrix to quantify the extent of misclassification. Table III illustrates the confusion of gray matter with MS lesion.

The rows in the confusion matrix indicate each tissue which has been segmented *manually*. Each row adds to 100%. The columns indicate the percentage of that tissue classified by the ICM algorithm (after combination of the PD/T2 segmentations). For example, for the manually classified White Matter our algorithm classified 3.9% as gray matter, 95.9% as white matter, none as CSF and 0.2% as lesion. Note that about 50%

of the "true" manually outlined lesion has been classified as lesion; these are the true positive lesion segmentations. Note also the confusion of lesion with gray matter illustrated in rows labeled *Gray* and *Lesion*. In effect 20% of the manually classified gray matter has been labeled incorrectly as lesion, and remarkably 41% of manually classified lesion has been labeled by the ICM algorithm as gray matter.

A further quantitative measure of the accuracy of the automatic segmentation methods, the *similarity index,* is described in Section, III.

This confusion of lesion and gray matter provides the motivation for post-processing the segmentation.

3) White Matter/Lesion Mask: Although the results from the previous step are encouraging, they fall short of what is needed to accurately quantify the extent of MS disease state. It can be seen from Fig. 11 that the areas surrounding the borders of some of the MS lesions have been incorrectly classified as gray matter. This is a result of the nature of the intensity profile of the MS lesion and the dependence of the ICM algorithm on the initial histograms obtained previously. Although our ICM algorithm generates a partial volume solution, it does not fully compensate for the partial volume effect around the lesion borders. As well, it is difficult to validate a partial volume segmentation since the only bench mark for comparison is a manual segmentation, obtained by physically outlining lesions on all slices in the volume. The output of this manual outlining is a binary segmentation for each slice. Further post-processing of the ICM segmentation will provide us with a binary solution suitable for comparison with the manual segmentation.

As mentioned previously, MS lesions occur almost exclusively in the white matter of the brain. If we can obtain a mask of white matter, including lesion, exclusive of gray matter, CSF, and background, it should be much easier to separate lesions from the white matter. The intensity profiles of these tissues are sufficiently contrasted to enable separation of the tissues by a second application of the ICM algorithm to the white matter mask. Therefore, our next step in the post-processing of the segmentation is to combine the results obtained for white matter and lesion in order to obtain a white matter/lesion mask, to be followed by another application of ICM. The result is binarized by an automatic thresholding step which reduces the number of low probability lesion areas. Following the binarization phase lesions of small size are eliminated from the image.
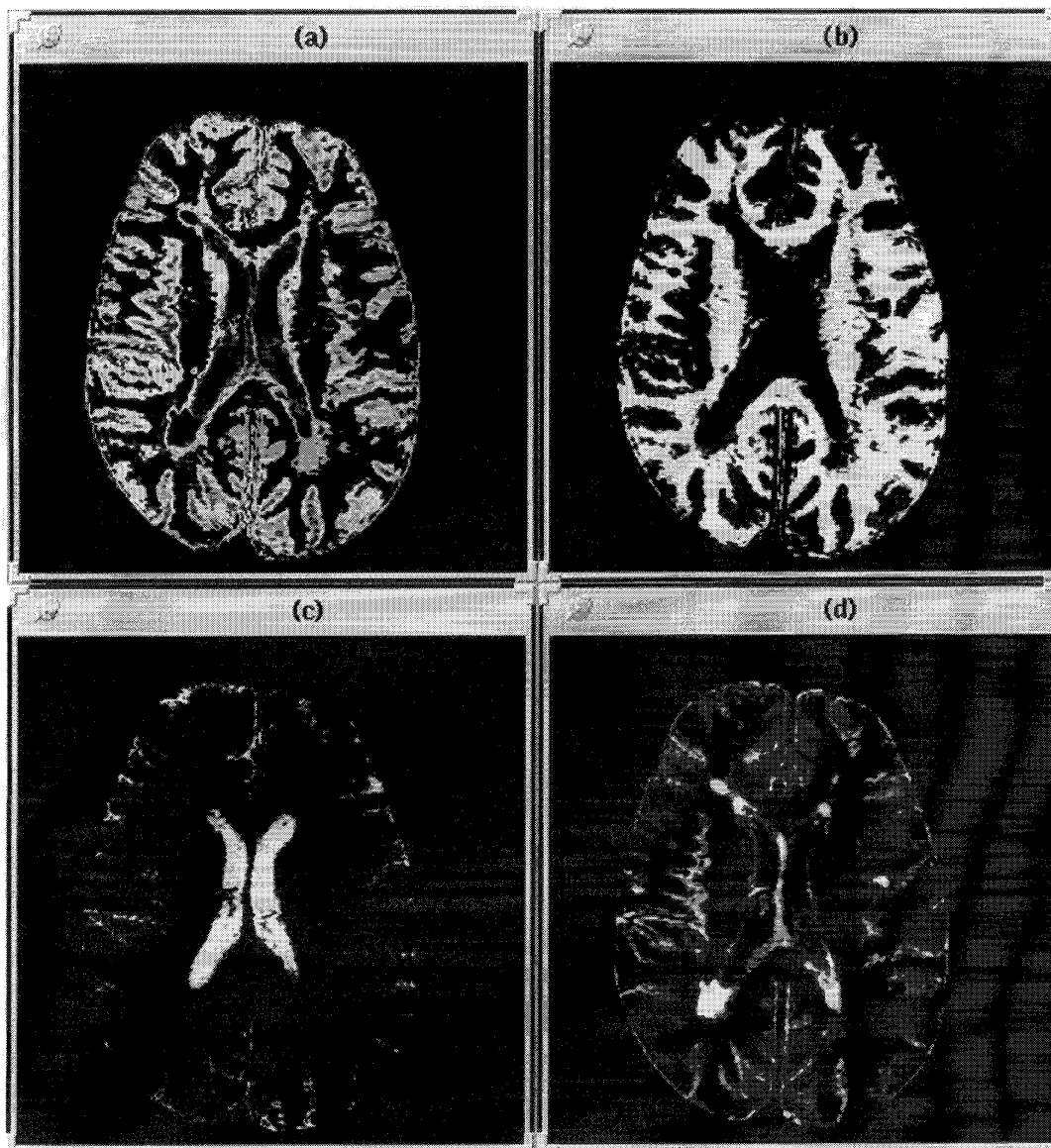
Fig. 11. Segmentation of slice 16 from combination of the PD (Fig. 5) and T2 (not shown) segmentations: (a) gray matter, (b) white matter, (c) CSF, (d) lesion.

The first step in developing the white matter/lesion mask is to sum the partial volume images for white matter and lesion, as shown in Fig. 11.

Fig. 12 shows the result of this addition. The intensity of voxels in the resulting sum image is guaranteed not to exceed the maximum since the images combined are partial volume images. Note the dark areas surrounding lesion borders in the sum image, most of which are areas previously classified as gray matter. It is these areas which must be *filled in* in order to accurately obtain a white matter mask. There are also some other areas of low probability for both lesion and white matter, probably the result of noise, which must also be eliminated from the sum image.

We now apply a series of morphological filtering and optimum thresholding to reduce this erroneous noise and to expand the lesions to merge with the white matter in which they are contained. To reduce the tiny islands and eliminate the

low probability areas we first apply a morphological opening process using a 3×3 diamond-shaped filter. In order to expand the lesions to merge with the surrounding white matter we then apply a morphological closing using a 5×5 circular kernel. The size of the closing filter is increased because the 3×3 kernel is too small to close many of the openings in most images. Following application of the morphological operators there are still several larger areas of low probability which needed to be eliminated. We applied an automatic thresholding technique in order to remove these areas.

The automatic thresholding technique we use was first introduced by Tsai et al. and employs a technique called moment-preserving thresholding (MPT) [47]. In MPT the object is to preserve the first k'th moments of an image and to find the threshold values which maintain the moments in the segmentation. Let $I(i,j)$ denote the intensity of sample points
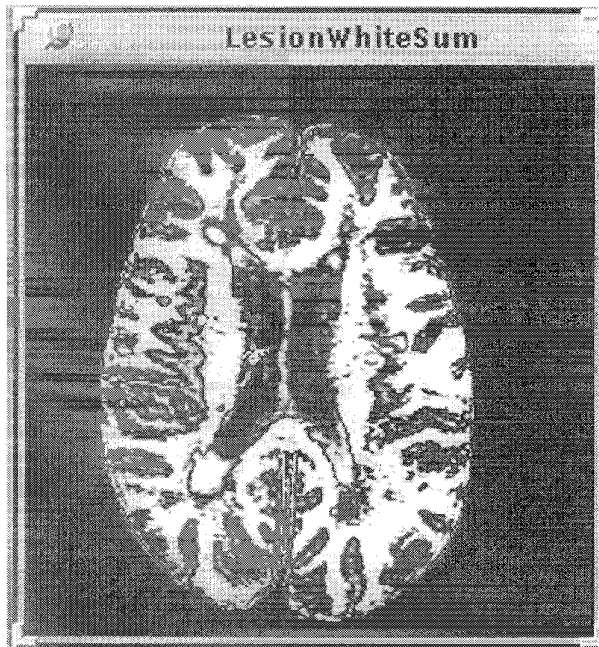
Fig. 12. Sum of lesion and white matter segmentations from Fig. 11.

in the image, i.e., the set of intensities of the indexed pairs

$$(i,j) \quad i = 1, 2, \cdots, M, \quad j = 1, 2, \cdots, N$$

where M and N are the number of pixels in the x and y directions, respectively.

The kth moment of an image, m_k, is defined to be

$$m_k = \frac{\sum_i \sum_j I^k(i,j)}{MN}.$$

Thus the zeroth moment of an image is one, and the first moment is the average gray level present in the image. These moments are also obtainable from the histogram of the image. Preservation of moments is motivated by the assumption that the original image is simply a blurred version of the true segmentation. We use a value of $k = 3$, maintaining the first three moments of the image, in order to obtain a threshold which best separates white matter/lesion from background or low probability areas. The MPT method has been tested and compared to other thresholding methods by Lee *et al.* [48] and Sahoo *et al.* [49].

Fig. 13 shows the mask obtained following application of the morphological operations and MPT thresholding to the white matter/lesion sum of Fig. 12. This mask is now applied to the original RF-corrected image to create an image consisting of white matter and lesion only. This is done for both the PD and T2 data sets.

4) Re-Application of ICM: Now that we have an image containing only white matter and lesion, we are again ready to apply the ICM algorithm. The original histograms of white matter and lesion are again used to drive the algorithm. Because of the presence of only two tissue types, the computational time and space requirements are significantly reduced. Following application of ICM to each of the two
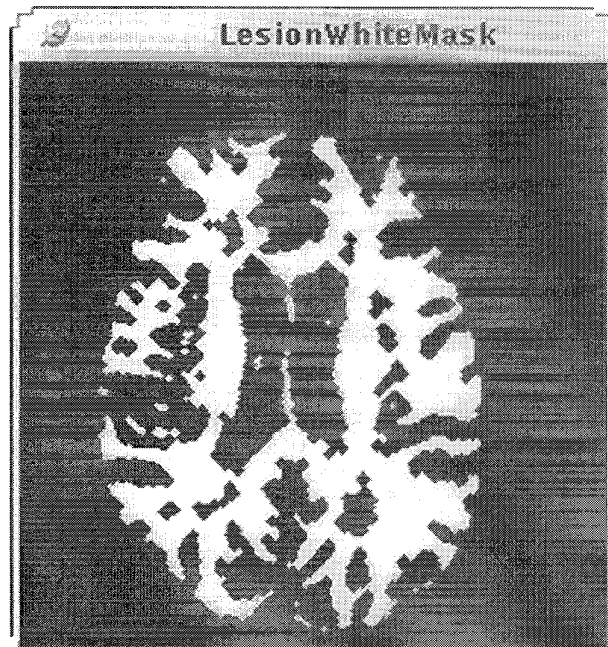

Fig. 13. Mask used to extract lesion and white matter from the original data of Fig. 3(c).

echo sequences the outputs are again combined and normalized as before. The result of this step for slice 16 of the lesion segmentation is given in Fig. 14(a). Comparing this slice with that of Fig. 11 illustrates the improvement in the segmentation that has resulted from application of the white matter mask.

There are still areas of Fig. 14(a) which are erroneous and likely the result of noise and/or errors in the white matter mask. The final step in the segmentation process will eliminate the low probability lesions and remove the small lesions.

5) Removal of Small/Low Probability Lesions: In order to eliminate the low probability lesions from the segmentation, we next apply MPT thresholding to the results of the previous step. Following this we remove small lesions from the segmentation by a series of basic image processing steps. For each lesion in the segmentation a set of run-length encodings are produced which define the lesion. These run-length encodings are used to determine the size of each lesion. Any lesion size found to be less than an input bound (in this case 15 pixels) is eliminated from consideration. The minimum lesion size was arrived at by examining the sizes of manually drawn lesions. We tried to match our minimum size with that of the human observer. The resulting lesion segmentation for slice 16 is given in Fig. 14(c). This final segmentation compares well with manually segmented lesions from the same study. Fig. 14(e) shows the manual segmentation of slice 16 from the study.

Fig. 14 illustrates the success of the segmentation on slice number 16 from the study. Fig. 14(b) and (d) shows the number of false positive and negative lesions resulting from our segmentation before and after elimination of small and low probability lesions. These results compare favorably with the manual segmentation of the same slice shown in Fig. 14(e).
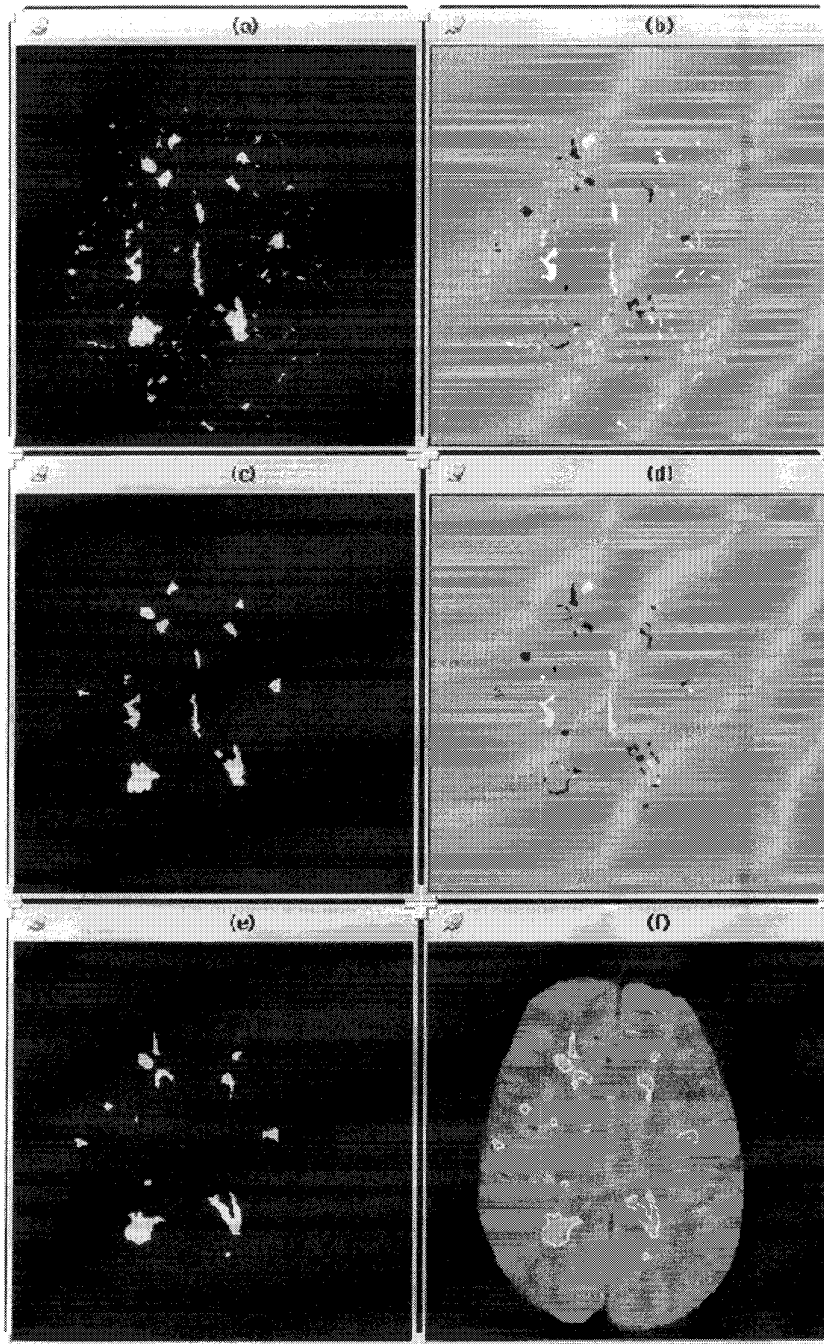
Fig. 14. Our segmentation versus manual segmentation (a) Segmentation of slice 16 of the study before elimination of small lesions. (b) False negative (dark) and false positive (light) lesions of Fig. 14(a) compared with manual segmentation [Fig. 14(e)]. (c) Segmentation after elimination of small lesions. (d) False negative (dark) and false positive (light) lesions of Fig. 14(c) compared with manual segmentation [Fig. 14(e)]. (f) Original RF-corrected PD data of slice 16 with manually outlined lesions overlayed in white.

Techniques for postprocessing image segmentations using white matter masks have also been reported by Zijdenbos et al. in [27], although they focus on removing false positive lesions based on their location within the brain. Our postprocessing could be extended to incorporate such spatial information, so that classified lesions in particular areas could be eliminated as being unlikely to be true positive lesions.

III. QUANTITATIVE ANALYSIS

To quantify these results, we use the *similarity index* described by Zijdenbos [27], derived from a reliability measure known as the kappa statistic, used by Vannier for analysing MR images [50]. Consider a binary segmentation as a set A containing the pixels considered to belong to the classification. The similarity of two segmentations A_1 and A_2 is given by a
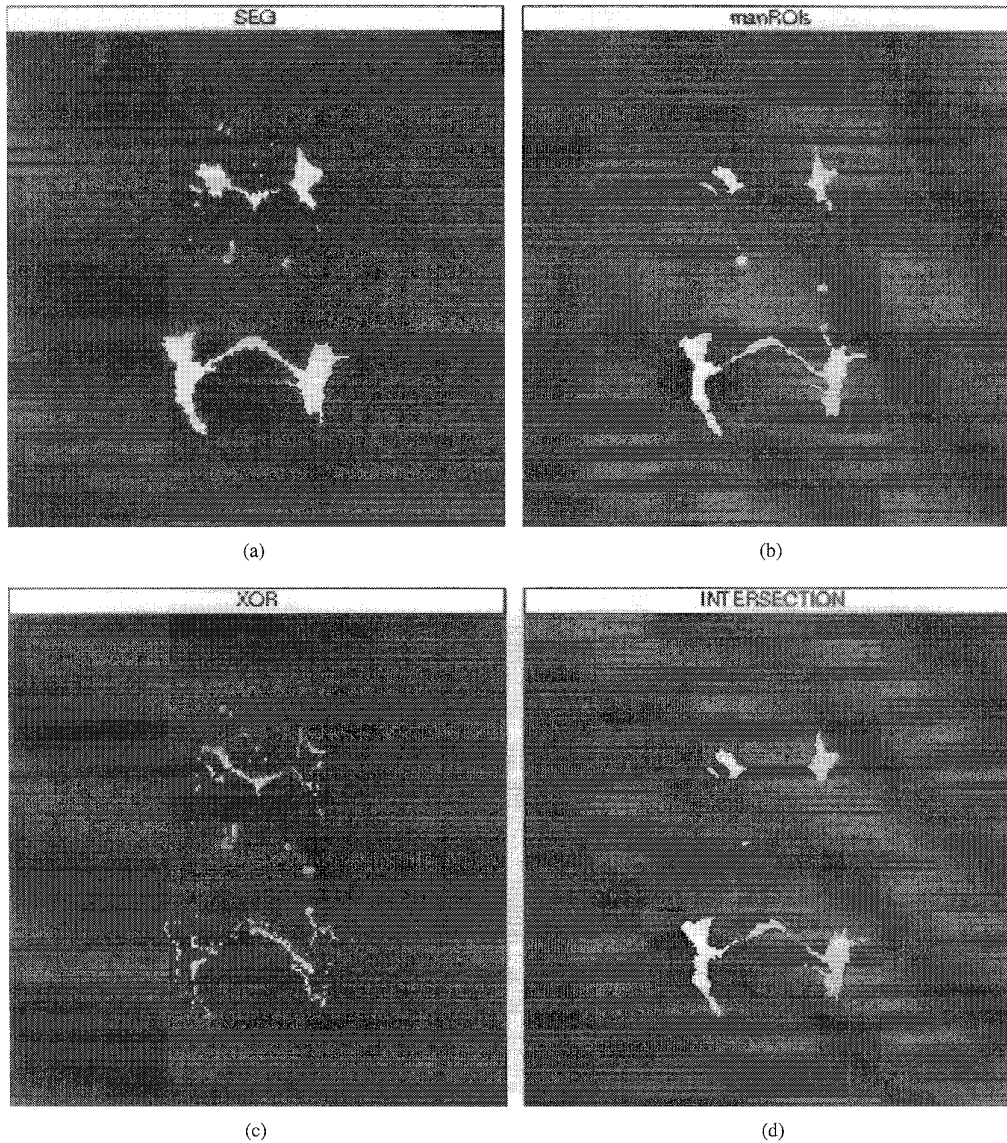
Fig. 15.   Similarity between the (a) semiautomatic segmentation (SEG) with the (b) manual segmentation (man ROI's). (c) The XOR figure illustrates the false positives and the false negatives, and (d) the INTERSECTION figure shows the correctly labeled lesion pixels. The similarity index for this slice is 0.80.

real number $S \in \{0 \cdots 1\}$ defined by

$$S = 2 \, \frac{|A_1 \cap A_2|}{|A_1| + |A_2|}.$$

Because the similarity index is the ratio of twice the common area of the segmentations to the sum of the sizes of the individual areas, it is sensitive to both size and location. Therefore, two equally sized regions that overlap each other with half of their area results in $S = \frac{1}{2}$ and a region completely overlapping a smaller one of half its size yields $S = \frac{2}{3}$. This reflects the intuitive feeling that two regions, of which one fully encompasses the other, are more similar than two partially overlapping regions.

Now Zijdenbos states that although $S > 0.7$ indicates excellent agreement, the absolute value of S is difficult to interpret; however, as an example, the similarity index of the

two images in Fig. 14(c) and (e) is 0.65. Further, the similarity index of the two segmentations in Fig. 15 is 0.80.

The similarity measure can be applied to any binary segmentation, so we typically apply our MPT thresholding technique to the semi-automatic grey-scale lesion segmentations to obtain the similarity measure of the semi-automatic lesion segmentation compared with the manual lesion segmentation.

Fig. 16 shows the effects of our RF correction technique. The final semi-automatic segmentation of lesions using the raw (uncorrected) data provide poor overlap with the manual segmentations for slices 11–13 and 20, and somewhat reasonable overlap (0.4–0.5) for the other slices. After the RF correction has been applied, the similarity index improves for every slice except slice 17, hence showing the efficacy of the RF correction technique. We presume that the filtering which occurs during RF correction can, on occasion, have a deletrious effect on segmentation, and indeed, we observe that the best
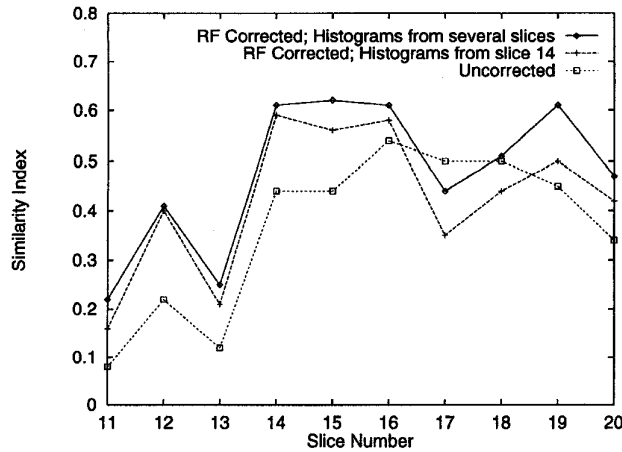
Fig. 16. Similarity indices for the central slices of the semi-automatic segmentation of the raw (uncorrected) data, the segmentation of the RF-corrected data using histograms from slice 14 only, and the segmentation of the RF-corrected data using histograms from several slices with respect to the manual lesion segmentation.

Fig. 18. Similarity indices for the central slices of the semi-automatic lesion segmentation using just PD-data, with respect to the manual lesion segmentation.

improves the similarity index further, (except for slices 17 and 18). There is a fairly good overlap with the manual segmentations but there are a large number of false positive lesions which lower the weighted similarity measure over all the central slices to 0.36.

After postprocessing, however, the false positives are much reduced so the similarity measures are much improved. The weighted average over all the central slices is 0.505, which may be considered to represent a 50% recording of true positives, and another 50% of a combination of false positives and false negatives. This is well within the range observed by Vannier and Zijdenbos, although the number of false positives are still a concern. We noted that the larger the lesions, the better the similarity index, as there tends to be a constant small number of false positive lesion pixels and when there are only a few gold standard lesion pixels, these false positives lower the similarity measure further than when there are a large number of gold standard pixels.

The false positives arise because of the confusion of lesion with grey matter, which arises because of the high standard deviation of the intensity histograms for these tissues. This is due in part to the difficulty in isolating pure tissue samples in such thick slices (5 mm thick), to the residual RF inhomogeneity which is not corrected, and to the age of the scanner used. Data sets from our new MR scanner show much less RF inhomogeneity and less variation within the tissue types, so we would expect these data to show an even higher similarity measure for lesion segmentation, and initial results indicate that similarity measures of greater than 90% will be obtainable with these data sets. However, manually outlined lesions from the new scanner are not yet available for us.

We applied our ICM segmentation technique to other PD data sets from the same scanner with the same acquisition parameters, with even better results, as shown in Fig. 18. If the T2 data for these sets were available, we would expect that the segmentations would improve further, as we noted in Fig. 17 that the final segmentations after post-processing were much better than the initial segmentation from the PD data alone.
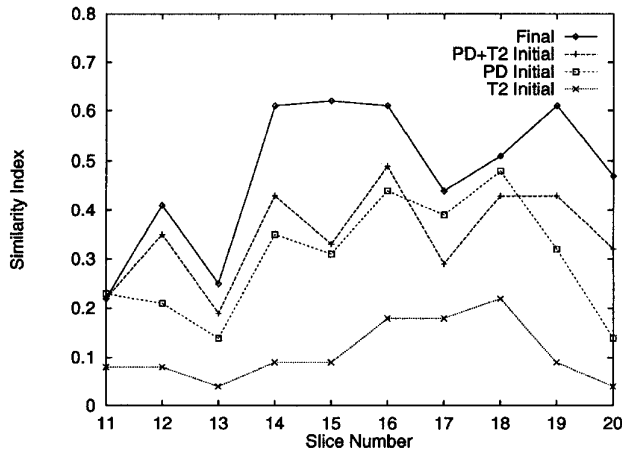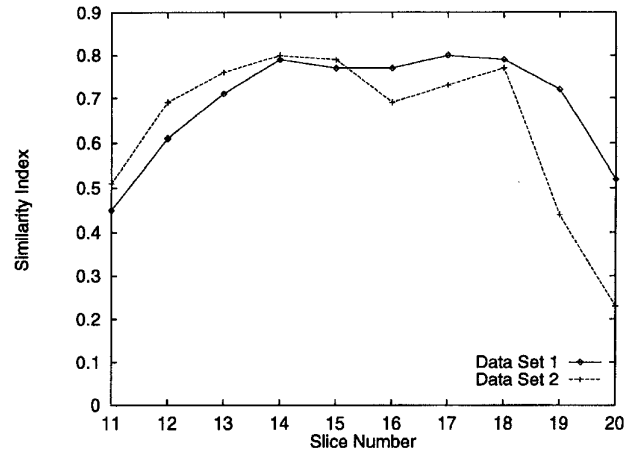

Fig. 17. Similarity indices for the central slices of the initial segmentations using just the initial T2 and PD data, the segmentation of the combined initial PD and T2 segmentations, and the final, post-processed lesion segmentation with respect to the manual lesion segmentation.

segmentations (and similarity measures) arise when the filter width is between 16 and 32 pixels wide.

To show whether the RF correction worked properly, we used histogram data from one slice only (slice 14) to be used in the segmentation algorithm. The final segmentation is strikingly similar to the segmentation using several slices for obtaining the histogram data, showing that the RF-correction appears to have worked well.

Fig. 17 illustrates the effects of post-processing after the initial segmentation. We applied an automatic thresholding technique to remove areas of low probability to the initial segmentation data shown in Fig. 11 to obtain a binary initial lesion classification suitable for the similarity measure. We then used the similarity index to illustrate the effects of our postprocessing methods with the same manually outlined "gold standard" lesion segmentations. Fig. 17 shows that the initial lesion segmentations using the T2 data is very poor, but is better using the PD data. Combining the PD and T2 data

IV. CONCLUSION AND FUTURE WORK

We have developed a semi-automatic system for segmentation of MRI brain images which provides an acceptable segmentation of the white matter and lesions, based on minimal manual intervention. We have outlined the brain with a new fully automatic method before applying an RF correction technique. In each patient scan, only a few ROI's need to be manually drawn for each of the four main tissue types (lesion, grey matter, white matter, and CSF). Furthermore, because of our RF correction technique, these tissues could be selected from the same slice, thus our method is suitable for almost fully automatic segmentation. Our method utilizes partial volume analysis for every brain voxel, and does not require any distortion or resampling.

Future work will include improving the preprocessing steps of image scaling, image enhancements and RF inhomogeneity corrections. We hope to develop a graphical user interface to guide the radiologist through the stages of the segmentation algorithm, so that the system can be used for speedy segmentation of large numbers of images. We are also actively investigating the effects of data compression on the image segmentations. Finally, we will use the segmentations to aid in analysis of data from follow-up exams, where the lesions are evolving over time.

ACKNOWLEDGMENT

REFERENCES

[1] R. W. Albright Jr. and E. K. Fram, "Microcomputer-based technique for 3-D reconstruction and volume measurement of computed tomographic images—Part i: Phantom studies," *Investigat. Radiol.,* vol. 23, no. 12, pp. 881–885, Dec. 1988.

[2] P. Lundin and G. Pedersen. "Volume of pituitary macroadenomas: Assessment by MRI," *J. Comput. Assist. Tomogr.,* vol. 16, no. 4, pp. 519–528, July/Aug. 1992.

[3] P. Maeder, A. Wirsen, M Bajc, W. Schalen, H. Sjoholm, H. Skeidsvoll, S. Cronoqvist, and D. H. Ingvar, "Volumes of chromis traumatic frontal brain lesions measured by MR imaging and CBF tomography," *Acta Radiologica,* vol. 32, pp. 271–278, 1991.

[4] D. Williams, P. Bland, L.Liu, L. Farjo, I. R. Francis, and C. R. Meyer, "Liver-tumour boundary detection: Human observer versus computer edge detection," *Investigat. Radiol.,* 24, no. 10, pp. 768–775, Oct. 1989.

[5] J. K. Udupa, S. Samarasekera, and W. A. Barrett. "Boundary detection via dynamic programming," in *Proc. SPIE: Visualization in Biomed. Comput. 1992,* Chapel Hill, NJ, 1992, vol. 1808, pp. 33–39.

[6] D. T. Long, M. A. King, and B. C. Penney, "2-D versus 3-D edge detection as a basis for volume quantitation in SPECT," *Inform. Processing Med. Imag.,* pp. 457–471, 1991.

[7] M. Joliot and B. M. Mazoyer, "Three-dimensional segmentation and interpolation of magnetic resonance brain images," *IEEE Trans. Med. Imag.,* 12, no. 2, pp. 269–277, June 1993.

[8] H. Suzuki and J. Toriwaki. "Automatic segmentation of head MRI images by knowledge guided thresholding," *Computerized Med. Imag. Graphics,* 15, no. 4, pp. 233–240, Jul–Aug 1991.

[9] F. Lachmann and C. Barillot, "Brain tissue classification from MRI data by means of texture analysis," in *Proc. SPIE: Med. Imag. VI: Image Processing,* Chapel Hil', NJ, 1992, vol. 1652, pp. 72–83.

[10] R. E. Gur, P. D. Mozley, S. M. Resnick, D. Shtasel, M. Kohn, R. Zimmerman, G. Herman, S. Atlas, R. Grossman, R. Erwin, and R. C. Gur, "Magnetic resonance imaging in schizophrenia: I. volumetric analysis of brain and cerebrospinal fluid," *Archives Gen. Psych.,* vol. 48, pp. 407–412, May 1991.

[11] M. I. Kohn, N. K. Tanna, G. T. Herman, S. M. Resnick, P. D. Mosley, R. E. Gur, A. Alavi, R. A. Zimmerman, and R. C. Gur, "Analysis of brain and cerebrospinal fluid volumes with {MR} imaging—Part i: Methods, reliability and validation," *Radiol.,* 178, no. 1, pp. 115–122, Jan. 1991.

[12] J. R. Mitchell, S. J. Karlik, D. H. Lee, and A. Fenster, "Automated detection and quantification of multiple sclerosis lesions in MR volumes of the brain," in *Proc. SPIE: Medical Imaging VI: Image Processing,* Chapel Hill, NJ, 1992, vol. 1652, pp. 99–106.

[13] T. Sandor, D. Metcalf, and Y. Kim, "Segmentation of brain CT images using the concept of region growing," *Int. J. Biomed. Comput.,* vol. 29, pp. 143–147, 1991.

[14] R. R. Stringham, W. A. Barrett, and D. C. Taylor , "Probabilistic segmentation using edge detection and region growing," in *Proc. SPIE: Visualization Biomed. Comput. 1992,* Chapel Hill, NJ, 1992, vol. 1808, pp. 40–51.

[15] K. H. Hohne and W. H. Hanson, "Interactive 3-D segmentation of MRI and CT volumes using morphological operations," *J. Comput. Assist. Tomogr.,* vol. 16, no. 2, pp. 285–294, 1992.

[16] N. J. Mankivich, L. I. Rudin, G. Koepfler, J. M. Morel, and S. Osher, "Application of a new pyramidal segmentation algorithm to medical images," in *Proc. SPIE: Medical Imaging VI: Image Processing,* Chapel Hill, NJ, 1992, vol. 1652, pp. 23–30.

[17] C. Mathieu, I. E. Magnin, and C. Baldy-Porcher, "Optimal stochastic pyramid: Segmentation of MRI data," in *Proc. SPIE: Medical Imaging VI: Image Processing,* Chapel Hill, NJ, 1992, vol. 1652, pp. 14–22.

[18] A. Rosenfeld, "Computer vision: A source of models for biological visual processes," *IEEE Trans. Biomed. Eng.,* 36, no. 1, pp. 93–96, 1989.

[19] L. Arata, A. Dhawan, J. Broderick, and M. Gaskill, "Three dimensional model-guided segmentation and analysis of medical images," in *Proc. SPIE: Medical Imaging VI: Image Processing,* Chapel Hill, NJ, 1992, vol. 1652, pp. 253–259.

[20] M. Kamber, D. L. Collins, R. Shinghal, G. S. Francis, and A. C. Evans, "Model-based 3-D segmentation of multiple sclerosis lesions in dual-echo MRI data," in *Proc. SPIE: Visualization Biomed. Comput. 1992,* Chapel Hill, NJ, 1992, vol. 1808, pp. 590–600.

[21] H. S. Choi, D. R. Haynor, and Y. Kim, "Partial volume tissue classification of multichannel magnetic resonance images–a mixel model," *IEEE Trans. Med. Imag.,* 10, no. 3, pp. 395–407, Sept. 1991.

[22] K. R. Smith and L. A. Kendrick "Bayesian computer vision methods for improved tumor localization and delineation," in *Proc. IEEE Med. Imag. Conf.,* Santa Fe, NM, 1991, pp. 2140–2144.

[23] B. Johnston, M. S. Atkins, and K. S. Booth, "Partial volume segmentation in 3-D of lesions and tissues in magnetic resonance images," in *Proc. SPIE-Medical Imaging 1994,* Bellingham, WA, 1994. SPIE–The Int. Soc. Opt. Eng., vol. 2167, pp. 28–39.

[24] L. J. Rosner and S. Ross, *Multiple Sclerosis.* New York: Simon and Schuster, 1992.

[25] G. J. Ettinger *et al.* "Automatic registration for multiple sclerosis change detection," in *Proc. IEEE Workshop Biomed. Image Anal.,* Los Alamitos, CA, June 1994, IEEE Comput. Soc. Press, pp. 297–306..

[26] F. Pannizzo, M. J. B. Stallmeyer, J. Friedman, R. J. Jennis, J. Zabriskie, C. Pland, R. Zimmerman, J. P. Whalen, and P. T. Cahill, "Quantitative MRI studies for assessment of multiple sclerosis," *Magn. Resonance Med.,* vol. 24, pp. 90–99, 1992.

[27] Alex P. Zijdenbos, Benoit M. Dawant, Richard A. Margolin, and Andrew C. Palmer, "Morphometric analysis of white matter lesions in MR images: Method and validation," *IEEE Trans. Med. Imag.,* vol. 13, no. 4, pp. 716–724, Dec. 1994.

[28] W. M. Wells III, W. E. L. Grimson, R. Kikinis, and F. A. Jolesz, "In-vito intensity correction and segmentation of magnetic resonance image data," in *AAAI 1994 Spring Symp. Ser.,* AAAI, Mar. 1994.

[29] M. Kamber, R. Shinghal, D. L. Collins, G. S. Francis, and A. C. Evans, "Model-based 3-D segmentation of multiple sclerosis lesions in magnetic resonance brain images," *IEEE Trans. Med. Imag.,* 1995.

[30] D. W. Patey and D. K. B. Li, "Interferon beta-1b is effective in relapsing-remitting multiple sclerosis—Part ii: MRI analysis results of a multicenter, randomized, double-blind, placebo-controlled trial," *Neurol.,* 43:662–667, Apr. 1993.

[31] B. Johnston, "3-D multispectral stochastic image segmentation," Master's thesis, The Univ. of British Columbia, Computer Science Dept., Vancouver, British Columbia, Canada,, Jan. 1994.

[32] R. Acharya and Y. Ma, "Segmentation algorithms for cranial magnetic resonance images," in *Proc. SPIE: Medical Imaging VI: Image Processing*, Chapel Hill, NJ, 1992, vol. 1652, pp. 50–61.

[33] M. E. Brummer, R. M. Mersereau, R. L. Eisner, and R. R. J. Lewine, "Automatic detection of brain contours in MRI data sets," *IEEE Trans. Med. Imag.*, vol. 12, no. 2, pp. 153–166, June 1993.

[34] H. E. Cline, W. E. Lorensen, R. Kikinis, and F. Jolesz, "Three dimensional segmentation of MR images of the head using probability and connectivity," *J. Comput. Assist. Tomogr.*, vol. 14, no. 6, pp. 1037–1045, 1990.

[35] B. Mackiewich, "Intracranial boundary detection and radio frequency correction in magnetic resonance images," Master's thesis, Simon Fraser Univ., Computer Science Dept., Burnaby, British Columbia, Aug. 1995.

[36] L. Axel, J. Costantini, and J. Listerud, "Intensity correction in surface-coil MR imaging," *Amer. J. Roentgenol.*, vol. 148, pp. 418–420, Feb. 1987.

[37] J. Haselgrove and M. Prammer, "An algorithm for compensation of surface-coil images for sensitivity of the surface coil," *Magn. Resonance Imag.*, vol. 4, no. 6, pp. 469–472, 1986.

[38] W. W. Brey and P. A. Narayana, "Correction for intensity falloff in surface coil magnetic resonance imaging," *Med. Phys.*, vol. 15, no. 2, pp. 241–245, Mar./Apr. 1988.

[39] D. T. Glennon, L. P. Clarke, R. P. Velthuizen, and F. A. Jolesz, "MR image nonuniformity correction techniques," in *Proc. 13th Annu. Meet. IEEE Eng. Med. Biol. Soc.*, IEEE EMBS, Orlando, Florida, Oct./Nov. 1991 pp. 89–90.

[40] K. O. Lim and A. Phefferbaum, "Segmentation of MR brain images into cerebrospinal fluid spaces, white and gray matter," *J. Comput. Assist. Tomogr.*, 13, no. 4, pp. 588–593, July/Aug. 1989.

[41] P. A. Narayana, W. W. Brey, M. V. Kulkarni, and C. L. Sievenpiper, "Compensation for surface coil sensitivity variation in magnetic resonance imaging," *Magn. Resonance Imag.*, vol. 6, no. 3, pp. 271–274, 1988.

[42] B. M. Dawant, A. P. Zijdenbos, and R. A. Marolin, "Correction of intensity variations in MR images for computer-aided tissue classification," *IEEE Trans. Med. Imag.*, vol. 12, no. 4, pp. 770–781, Dec. 1993.

[43] R. C. Gonzalez and R. E. Woods, *Digital Image Processing.* Reading, MA: Addison-Wesley, 1992.

[44] R. B. Lufkin, T. Sharpless, B. Flannigan, and W. Hanafee, "Dynamic-range compression in surface-coil MRI," *Amer. J. Roentgenol.*, vol. 147, pp. 379–382, Aug. 1986.

[45] J. Besag, "On the statistical analysis of dirty pictures," *J. Royal Statistical Soc.*, vol. 48, no. 3, pp. 259–302, 1986.

[46] B. Johnston, M. S. Atkins, and K. S. Booth, "Three-dimensional partial volume brain tissue segmentation of multispectral magnetic resonance images using stochastic relaxation," in *Non-Linear Image Processinfg VI*, vol. 2180, E. R. Dougherty, Ed. Bellingham, WA: SPIE–The International Soc. for Optical Engineering, pp. 268–279, 1994.

[47] W. Tsai, "Moment-preserving thresholding: A new approach," *Comput. Vision, Graphics Image Processing*, vol. 29, pp. 377–393, 1985.

[48] S. U. Lee and S. Y. Chung, "A comparative performance study of several global thresholding techniques for segmentation," *Comput. Vision, Graphics Image Processing*, vol. 52, pp. 171–190, 1990.

[49] P. K. Sahoo, S. Soltani, and K. C. Wong, "A survey of thresholding techniques," *Comput. Vision, Graphics, Image Processing*, vol. 41, pp. 233–360, 1988.

[50] M. W. Vannier, T. K. Pilgrim, C. M. Speidel, L. R. Neumann, D. L. Rickman, and L. D. Schertz, "Validation of magnetic resonance imaging (MRI) multispectral tissue classification," *Comput. Med. Imag. Graphics*, vol. 15, no. 4, pp. 217–223, 1991.