NeuroImage **17,** 1087–1100 (2002)
doi:10.1006/nimg.2002.1221

A Robust Method for Extraction and Automatic Segmentation of Brain Images

N. Kovacevic,*,[1] N. J. Lobaugh,*,†,[2] M. J. Bronskill,*,† B. Levine,*,‡ A. Feinstein,*,† and S. E. Black*,†

*Sunnybrook and Women's College Health Sciences Centre, †University of Toronto, and ‡Rotman Research Institute of Baycrest Centre for Geriatric Care, Toronto, Ontario, Canada

Received September 13, 2001

A new protocol is introduced for brain extraction and automatic tissue segmentation of MR images. For the brain extraction algorithm, proton density and T2-weighted images are used to generate a brain mask encompassing the full intracranial cavity. Segmentation of brain tissues into gray matter (GM), white matter (WM), and cerebral spinal fluid (CSF) is accomplished on a T1-weighted image after applying the brain mask. The fully automatic segmentation algorithm is histogram-based and uses the Expectation Maximization algorithm to model a four-Gaussian mixture for both global and local histograms. The means of the local Gaussians for GM, WM, and CSF are used to set local thresholds for tissue classification. Reproducibility of the extraction procedure was excellent, with average variation in intracranial capacity (TIC) of 0.13 and 0.66% TIC in 12 healthy normal and 33 Alzheimer brains, respectively. Repeatability of the segmentation algorithm, tested on healthy normal images, indicated scan–rescan differences in global tissue volumes of less than 0.30% TIC. Reproducibility at the regional level was established by comparing segmentation results within the 12 major Talairach subdivisions. Accuracy of the algorithm was tested on a digital brain phantom, and errors were less than 1% of the phantom volume. Maximal Type I and Type II classification errors were low, ranging between 2.2 and 4.3% of phantom volume. The algorithm was also insensitive to variation in parameter initialization values. The protocol is robust, fast, and its success in segmenting normal as well as diseased brains makes it an attractive clinical application.

INTRODUCTION

One challenge in the development of intensity-based algorithms for segmentation of brain MR images is that local signal intensity is influenced by system factors such as scanner and head coil properties (e.g., RF inhomogeneity, Sled *et al.*, 1998; Clarke *et al.*, 1995), as well as biological factors such as age (Cho *et al.*, 1997; Imon *et al.*, 1998; Miot *et al.*, 1995; Steen *et al.*, 1997) and disease status (Steen *et al.*, 1999; Thatcher *et al.*, 1997; Laakso *et al.*, 1996; Hirai *et al.*, 1996). One impact of biological factors can be seen clearly in Fig. 1, where representative T1-weighted intensity histograms for healthy young and elderly and from a patient with probable Alzheimer's disease (AD) are presented. The normal elderly and AD histograms have large cerebral spinal fluid (CSF) compartments compared to that seen in the young normal histogram. Additionally, in the diseased brain, the contrast between gray and white matter is considerably reduced, and the two histogram peaks have merged.

We present here a fully automatic, single-channel, histogram-based approach to discrete segmentation of brain tissue. The procedure is applicable to a wide variety of brains because it is robust against variations in tissue intensities and proportional tissue volumes. The approach presented here has been optimized for T1-weighted sequences, where reasonable contrast is obtained between the three main tissue classes in brain (gray matter (GM), white matter (WM), and CSF) with fairly short image acquisition times.

Other single-channel segmentation approaches require user input or expertise to various degrees, and can be grouped into two categories based on aspects of their segmentation strategies. One class of algorithms uses directly the intensity data contained in the image, and the other uses an ideal prior segmentation to guide the assignment of voxels in the image being segmented. Examples of approaches from the first category include algorithms that fit Gaussian or polynomial models to the data (Grabowski *et al.*, 2000; Schnack *et al.*, 2001; Rajapakse *et al.*, 1996), sometimes combined with Markov random field models (e.g., Rajapakse *et al.*, 1997; Ruan *et al.*, 2000; Van Leemput *et al.*, 1999; Held *et al.*, 1997). Atlas- or model-based methods fall into the second category and typically

[1] Present address: Hospital for Sick Children, Toronto, ON, Canada.
[2] To whom correspondence should be addressed at Imaging Research and Cognitive Neurology, 2075 Bayview Avenue, Room S604, Toronto, ON M4N 3M5, Canada. Fax: 416-480-5714. E-mail: nlobaugh@sten.sunnybrook.utoronto.ca.

1053-8119/02 $35.00

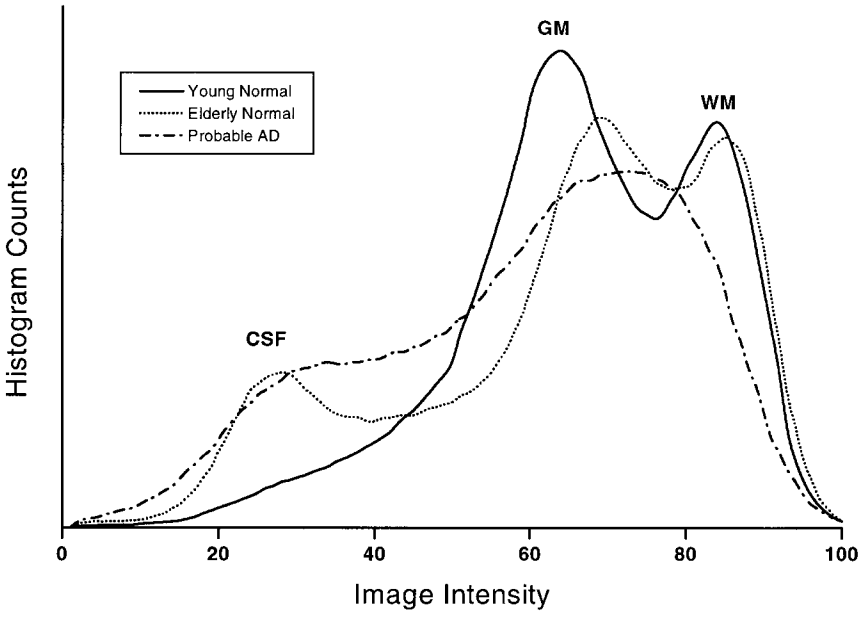

FIG. 1. Global brain histograms of young normal, elderly normal, and probable AD. Note the increased amplitude of the CSF peak in the elderly and the decreased separation between the gray and the white matter peaks in AD. Images have been masked to exclude nonbrain tissues.

require images to be coregistered to a segmented template image (Van Leemput *et al.*, 1999; Ashburner and Friston, 1997, 2000).

The recent work of Grabowski *et al.* (2000) is an excellent example of a Gaussian model approach. The user initializes the peak intensity for GM and WM, and the algorithm derives the peak intensity for the CSF compartment based on an operator-defined region of interest (ROI). Five Gaussians representing three pure tissue classes (GM, WM, CSF) and two partial volume compartments (GM/WM, CSF/GM) are fitted at a local level and are used to generate either discrete or continuous segmentations. The primary evaluation of the algorithm was based on discrete segmentations, and it performed well in comparison with human experts and a digital phantom. Although it requires user input, it is quite robust against minor variations in the initial parameters.

Gaussian fitting algorithms can be sensitive to noise in the image, which produces speckled regions in the final segmentation (e.g., speckles of gray matter within white matter). To reduce the impact of this noise, and at the same time enforce tissue smoothness and piecewise contiguity, some authors (e.g., Rajapakse *et al.*, 1997; Ruan *et al.*, 2000; Held *et al.*, 1997; Zhang *et al.*, 2001; Van Leemput *et al.*, 2001) have suggested the use of Markov random field models in the segmentation process. This approach starts from the assumption that tissues are relatively homogeneous in most brain regions. For example, major white-matter pathways would not be expected to contain the gray speckles frequently seen on T1-weighted images. While the specifics of the approaches vary, the common feature of Markov random field segmentations is that small, disconnected volumes of speckle are reassigned to the value of the surrounding tissue, restoring the expected spatial correlation among voxels. In young normal brains, this approach maintains sharp tissue boundaries, while smoothing over small irregularities within tissue types. It is known that T1 relaxation in the brain varies regionally (Steen *et al.*, 1997) and is affected by age. The local discontinuities and changes in signal variability seen in T1-weighted images may thus reflect important biological processes in diseased and elderly brains. To date, assessments of single-channel segmentations using the Markov random field approach have used simulations (Zhang *et al.*, 2001), phantoms (Rajapakse *et al.*, 1997), or young brains (Rajapakse *et al.*, 1997; Zhang *et al.*, 2001; Held *et al.*, 1997) or are restricted to a single brain slice with specific features (Van Leemput *et al.*, 2001). It is not clear that the Markov random field approach would be appropriate in populations whose MR properties vary substantially from those presented above, and we return to this issue in the Discussion.

Atlas- or template-based segmentation methods provide information about the spatial locations of voxels belonging to various tissue classes, but require the image of interest to be coregistered to the template (e.g., Ashburner and Friston, 1997). Both the choice of coregistration algorithm and the degree to which the target image matches the template segmentation will

contribute to the accuracy of the final segmentation. Segmentation errors due to miscoregistration are likely to be small when the images are similar. Separate templates would be required in situations of brain pathology, where anatomy or T1 relaxation can deviate substantially from the template (e.g., Thompson *et al.*, 2001).

The fundamental assumption of our approach to tissue segmentation is that a MRI head scan consists of four distinct tissue classes. These include one nonbrain class consisting of extracerebral tissues such as skull, fat, muscle, and dura and three cerebral tissue classes, namely, gray matter, white matter, and CSF. The objective was to devise a robust procedure that could be used to obtain consistent, and plausible volumetric estimates of brain tissues irrespective of underlying pathology.

METHODS

MR protocols and participants. All methods described in this paper were developed on images produced on a 1.5 T-Signa scanner (GE Medical Systems). These images include a T1-weighted acquisition (axial 3D SPGR with a 5-ms TE, 35-ms TR, 1 NEX, 35° flip angle, 22 × 16.5 cm FOV, 0.859 × 0.859 mm in-plane resolution, and 1.2- to 1.4-mm slice thickness), a proton-density (PD), and a T2-weighted acquisition (interleaved axial spin echo with TEs of 30 and 80 ms, 3-s TR, 0.5 NEX, 22 × 22 cm FOV, 0.859 × 0.859-mm in-plane resolution, and 3-mm slice thickness). These images were acquired in the same imaging session from healthy young and elderly volunteers, or from patients with probable Alzheimer's disease (AD) who were participating in an ongoing research project at Sunnybrook & Women's. Informed consent was obtained, and this research was approved by the local ethics review board.

Image processing. The full procedure consists of two steps. In the first step (brain extraction), all nonbrain matter is removed, leaving only brain voxels. For the purposes of this report, "brain" refers to all intracranial voxels (i.e., GM, WM, and CSF) and "nonbrain" refers to the remaining voxels in the volume. In the second step (T1 segmentation), all brain voxels are classified into one of three discrete tissue types: CSF, GM, or WM.

Brain extraction. The use of T1-weighted images to create the brain mask was considered and rejected because most currently available algorithms have been designed to produce extractions that highlight the cortical surface and thus do not include a substantial portion of subdural CSF. Our procedure for brain extraction is semiautomatic and uses the PD and T2 images. First, an automatic segmentation algorithm classifies all voxels into either brain or nonbrain matter. Because some extracerebral structures will be incorrectly classified as brain matter, an automatic spatial connectivity algorithm is applied to refine the classification. The final manual step involves a small amount of editing of the output.

A typical two-dimensional PD/T2 histogram is shown in Fig. 2. The automatic segmentation algorithm first defines the upper limit of the intensity range for each of the two images (l_{PD}, l_{T2}). This limit was set at the point where the one-dimensional histogram counts fall below 50 voxels (~100 mm^3). One characteristic of PD/T2 histograms is a lightly populated region (Fig. 2, arrows) that defines a natural division between brain (*tissue_1*) and nonbrain voxels (*tissue_0*). The next step was to define automatically an optimal cutoff curve within this region to isolate the brain voxels, and the use of lines, hyperbolas, and ellipses were considered for this purpose. In actuality, the most important factor was the distance from the histogram origin, and all three methods could be optimized to produce equivalently robust results. For the present application, an ellipse was implemented, as seen in Fig. 2. The cutoff ellipse is defined by:

$$\frac{x^2}{l_{T2}^2} + \frac{y^2}{l_{PD}^2} = 0.4.$$

The value of 0.4 was chosen by inspecting the histograms from a number of PD/T2 pairs that were all normalized to the same intensity ranges. Figure 3 shows 2D histograms for five young controls (yellow) and five AD patients (red), all normalized to an intensity range of 800 for PD and 600 for T2. To highlight that the region is consistent across subjects, and to make the figure more readable, only two contours are plotted for each histogram, reflecting lightly populated (thin lines) and densely populated (thick lines) regions of the histograms. Because of the considerable lesion load in the five AD brains, the histograms are not well aligned at the interface between GM and CSF. All 10 histograms align quite well at the interface of *tissue_1* and *tissue_0*, however, enabling the same cutoff curve to be used. Furthermore, the general region of the cutoff ellipse is lightly populated within the histogram (thin histogram contours), which significantly adds to the robustness of the cutoff.

Results from the application of this protocol are shown in Fig. 4A. It is clear that *tissue_1* contains more than just brain in that pieces of scalp and eyes have been included. To remedy this, we take advantage of the fact that a band of *tissue_0* (e.g., air and dura) surrounds the entire brain. This allows successful extraction of brain matter alone on most slices, using a 2D-connectivity algorithm based on a seed placed automatically in the middle of the brain tissue. This

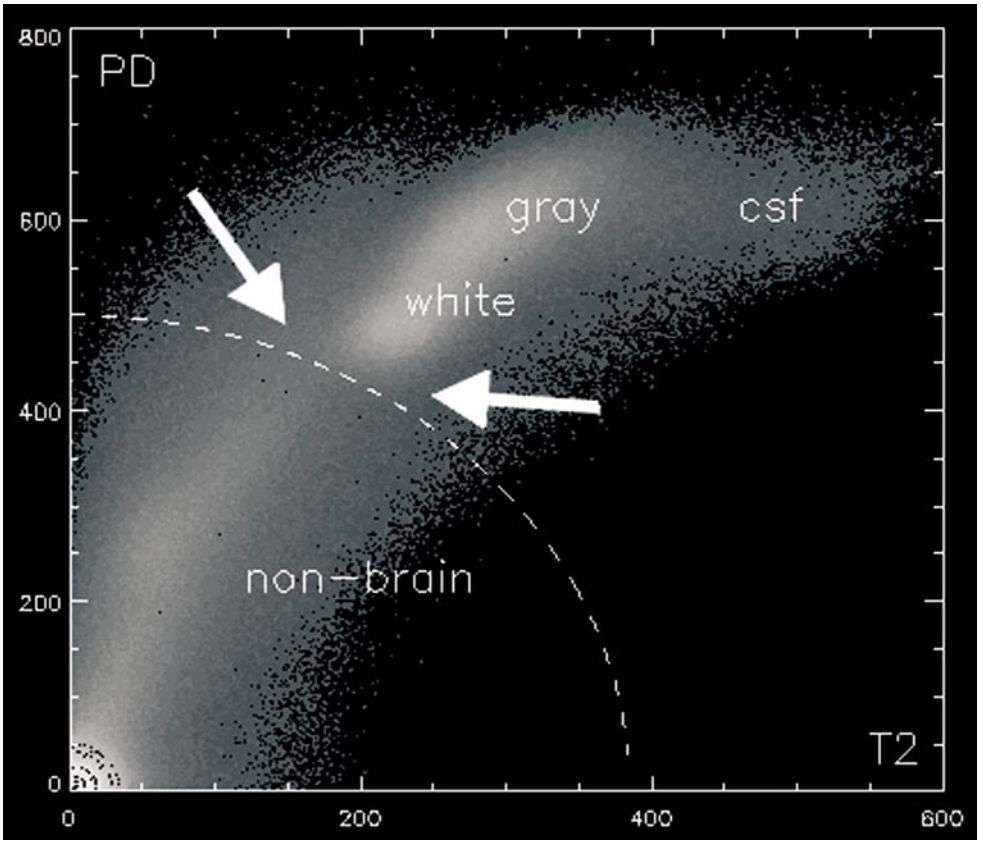

FIG. 2. Two-dimensional histogram based on T2 (*x*-axis) and PD (*y*-axis) image intensity. Brighter regions represent higher histogram counts. Arrows indicate the natural break in the histogram separating brain from nonbrain tissues. Dotted line represents the cutoff ellipse used to isolate brain from nonbrain tissue.

seed-growing step is accurate on all but a few slices, where eyes or other extracerebral tissues will remain, as can be seen in Fig. 4B. For such circumstances, the image is edited manually to correct these errors (Fig. 4C).

The manual intervention needed is not very complicated, because the errors made by the automatic seed growing are quite blatant. Consequently, training for the manual editing is minimal and the need for time-consuming inter- and intrarater comparisons is reduced. The final output of the brain extraction procedure is a binary mask to prepare the T1 images for segmentation. The full brain extraction protocol, including manual editing, requires approximately 10 min on a 256 MHz Sun Spare workstation.

Segmentation. Each PD image was coregistered to the T1-weighted image using a rigid body transformation (Automatic Image Registration, v. 3.07, Woods *et al.*, 1998). The resulting transformation matrix was used to move the brain mask into the T1 acquisition space, and all nonbrain voxels were set to zero. The subsequent T1 histogram smoothly increases from zero toward the CSF peak (Fig. 5). Should improper brain extraction occur, the inclusion of nonbrain tissues (e.g., dura) will cause the unwanted side effect of a small hump on the histogram near zero (Fig. 5, arrow).

The T1 segmentation algorithm unfolds in two steps. First, the normalized global histogram of the preprocessed T1 image is modeled using a Gaussian mixture. Model parameters are derived automatically, eliminating the need for user input. If the model parameters indicate an improper brain extraction (i.e., presence of a peak below the CSF peak, Fig. 5), a warning is issued. In the second step, the image is subdivided into small local regions. The corresponding local histogram modeling is initialized using parameters from the global histogram. The localized modeling approach deals with issues of image RF nonuniformity and intrinsic tissue heterogeneity, similar to Grabowski *et al.* (2000) and Rajapakse *et al.* (1996, 1997). The calculated model parameters are then used to define the local intensity thresholds for the three tissue types.

The Gaussian mixture model was fitted to intensity histograms (normalized such that the area under the curve equals 1) using the Expectation Maximization (EM) algorithm (McLachlan and Krishnan, 1997), as outlined by Alder (2001). The EM algorithm is appropriate for automatic fitting of T1 histograms because it

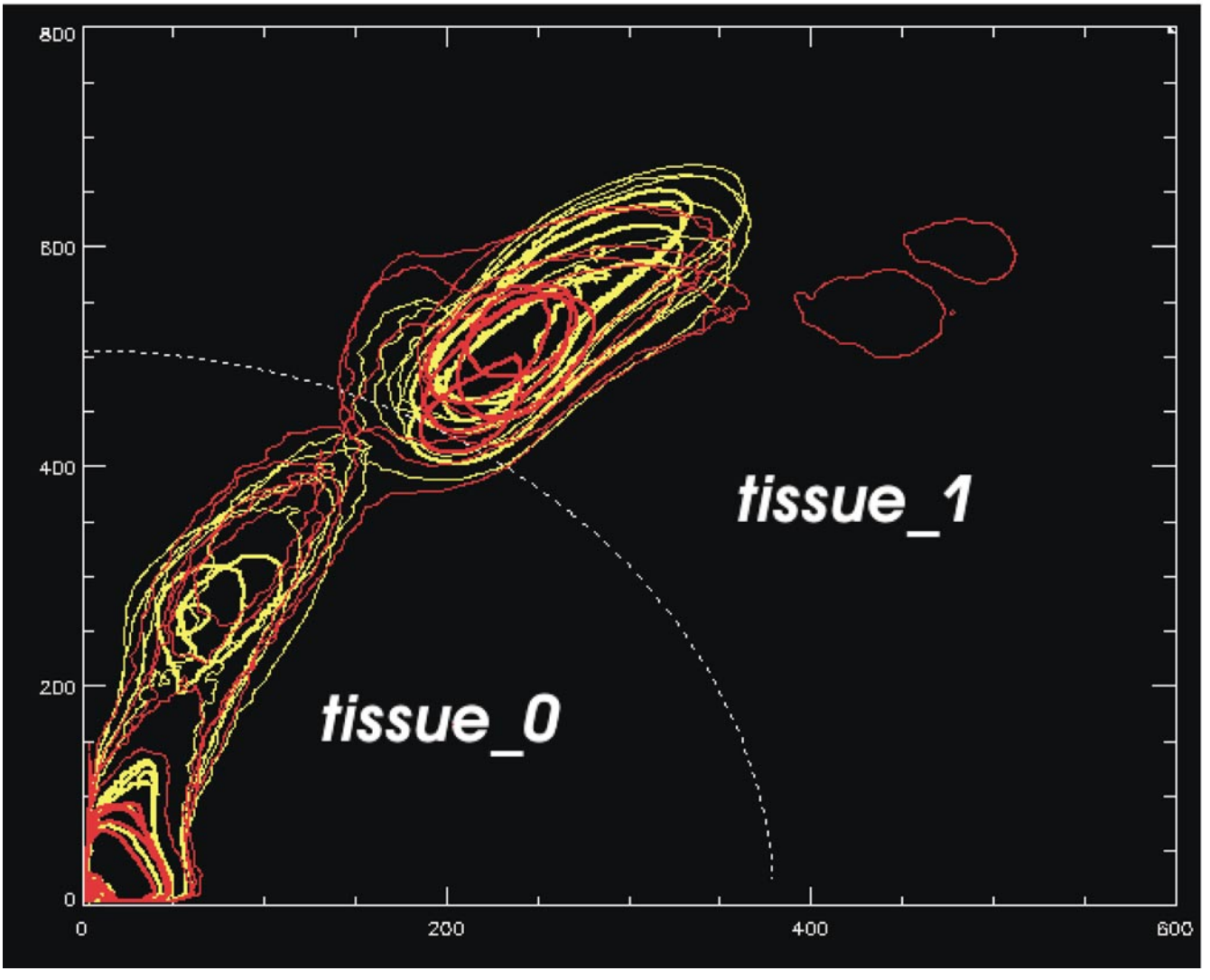

FIG. 3. Contour plots of 10 2D histograms normalized to the same intensity range. Thin contour lines indicate the histogram level representing ~200 voxels; thick lines indicate ~500 voxels. Yellow lines indicate healthy young subjects and red lines indicate AD patients.

is quite insensitive to initialization parameters and has a high rate of convergence (typically less than 100 iterations). The algorithm also adjusts well to large biological differences in relative tissue amounts. For example, in a young normal person, roughly 10% of the total intracranial capacity (TIC) is CSF, while for an older demented person, that percentage can be as high as 35%. This flexibility is crucial for fitting the local normalized histograms, where the relative amounts of the modeled tissue types vary greatly from one brain region to another.

Global histogram modeling. A careful investigation of fitting the global histogram was made with 3-, 4-, and 5-Gaussian mixture models. The choice of a 4-Gaussian mixture model to fit these histograms was the only rational compromise between goodness-of-fit and biological plausibility of the underlying models. Examples of the results are shown in Fig. 6, where a fits from a young normal (left), an elderly normal (centre), and an Alzheimer's patient (right) are presented for 3-(A), 4- (B), and 5-Gaussian fits (C). With our T1 protocol, the contrast between CSF and GM is high, leading to a low goodness-of-fit using the 3-Gaussian mixture model (Fig. 6A) for all three types of images (mean square error = 5.7×10^7, $n = 10$ images). The mean square errors for 4- and 5-Gaussian models were better (1.6×10^7 and 7.6×10^6, respectively), forcing a choice between these two models.

In the 4-Gaussian model (Fig. 6B), there was reasonable correspondence in all three image types between the histogram peaks and the modeled peaks for the three main tissue classes (CSF, GM, WM). This sug-

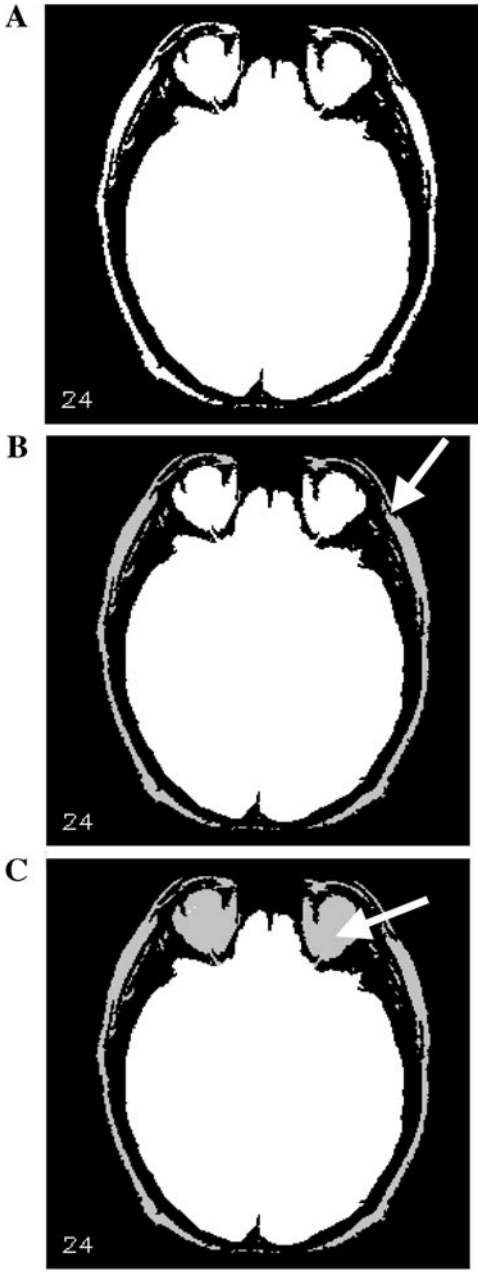

FIG. 4. (A) Initial output from the brain extraction algorithm. White represents *tissue_1* and black represents *tissue_0*. (B) After the slice-by-slice seed-growing algorithm is applied, most of the extracerebral tissue (gray) is removed (arrow). (C) The final stage of brain extraction requires manual intervention (arrow).

gests the modeled peaks are consistently reflecting the MR properties of the tissues. On the other hand, in the 5-Gaussian model (Fig. 6C), the modeled peaks shift away from their respective histogram peaks. The magnitude of this shift was completely dependent on the size and location of the mixture classes, which varied unsystematically across images. This variation in the behavior of the mixture classes also produced implausible effects in the tissue compartments of interest. In the young brain, the proportions of tissues assigned to each of the five classes appeared reasonable, based on visual inspection. However, all cases, there was a major impact on the relationship of the CSF and CSF/GM partial volume compartments. In the 5-Gaussian model, the proportion of the histogram modeled as CSF was reduced by as much as 60%, with a corresponding doubling of the CSF/GM compartment (proportions for each class are listed below the corresponding peaks in Figs. 6B and 6C). More importantly, the partial volume compartments in the three brains were not modeled in the same way. In the young normal, the GM/WM compartment was small, relative to the GM and WM compartments, as would be expected. In the elderly normal, the two mixture classes were given approximately equal weight with the GM and WM classes, while in the Alzheimer's brain they had equal weights with the WM class. Given that the integrity of the underlying tissue is not known, arbitrarily constraining the model to increase its consistency was not an acceptable option. Because the 5-Gaussian model shifted the location of the mean tissue intensities, and unsystematically overestimated the two partial volume compartments in the elderly and diseased brains, the 4-Gaussian model was selected to maintain the robustness of the algorithm.

Global model fitting using EM algorithm. As an iterative algorithm, EM requires reasonable initial estimates for the model parameters. One new feature of our procedure is that all Gaussian means (m's) are initialized automatically, based on a proportion of the intensity range.

To implement automatic initialization of the EM algorithm, it was important to identify the consistencies among images varying in intensity ranges. This was done by normalizing the images as follows. First, a number of different scan histograms were rescaled to the same intensity range. The upper limit of the intensity range, l_{T1}, was defined as the point where the T1 histogram counts fell below 0.02% of the TIC. An example is shown in Fig. 7, where 10 histograms are plotted, all normalized to an intensity range of $l_{T1} = 100$. As can be seen in the figure, the rescaled histograms exhibit strikingly consistent locations for the peak intensities.

Based on the consistency seen across the normalized histograms, $0.25*l_{T1}$, $0.35*l_{T1}$, $0.67*l_{T1}$, and $0.83*l_{T1}$ were chosen as initial values for means of CSF, CSF/GM partial volume mixture, GM, and WM, respectively. The weight parameters (ws) were initialized by "average" biological amounts of 0.15, 0.05, 0.45, and 0.35 for CSF, the CSF/GM partial volume mixture, GM, and WM, respectively. Tissue standard deviations are consistent in any good imaging system; thus, the initial values were based on the average standard deviations over a few scans (SDs = 7, 3, 12, and 12 for CSF, CSF/GM, GM, and WM, respectively).

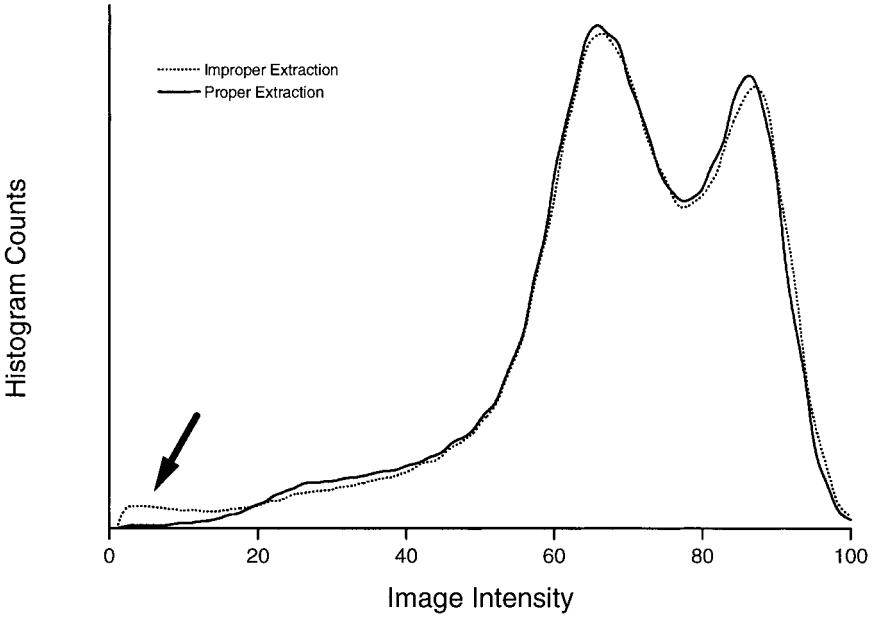

FIG. 5. Impact of improper separation of brain from nonbrain tissues on the T1 histogram (young normal). Large number of voxels in the low intensity range (arrow) are seen when dura and other nonbrain tissues are included as part of the brain mask. As these voxels negatively affect histogram fitting and segmentation results, a warning is issued if they are present.

After automatic estimation of the initial parameters, the 4-Gaussian model is fitted to the normalized global histogram (H) as follows:

$$H(x) = \sum_{i=1}^{4} w_i * \frac{1}{\sqrt{2\pi} * s_i} * e^{-(x-m_i)^2/2s_i^2},$$

where w_is represent weights, m_is represent means, s_is represent standard deviations of 4-Gaussians, and x is voxel intensity. The total degrees of freedom is 11 because

$$\sum_{i=1}^{4} w_i = 1.$$

Local model fitting and segmentation. The 12 global parameter values are used to initialize the fitting of local histograms. Localization is achieved by restricting the fit to volumes (boxes) of 48 × 48 × 30 voxels. These dimensions achieve a balance between size and amount of data within each volume. On one hand, smaller volumes produce better inhomogeneity correction. On the other, there needs to be sufficient representation of each tissue class for the 4-Gaussian mixture model to be applicable.

From the 12 model parameters, only the means for CSF, GM, and WM are used to calculate intensity cutoffs for the segmentation. The mean was chosen because it is less affected by factors such as final box size (see below) and proportions of tissues in the box, both of which are determined by the underlying anatomy. The cutoffs between CSF and GM (c_{cg}) and the cutoff between GM and WM (c_{gw}) are defined as:

$$c_{\mathrm{cg}} = (m_{\mathrm{c}} + m_{\mathrm{g}})/2$$

$$c_{\mathrm{gw}} = (m_{\mathrm{g}} + m_{\mathrm{w}})/2.$$

Given a voxel with intensity x, it is segmented as CSF if $0 < x < c_{\mathrm{cg}}$, as GM if $c_{\mathrm{cg}} \leq x < c_{\mathrm{gw}}$, and as WM if $c_{\mathrm{gw}} \leq x$.

To produce the segmentation, the algorithm tessellates the entire brain image into small boxes, 16 × 16 × 10 voxels in size. Each small box forms the central core of the larger local box (48 × 48 × 30 voxels) that is used to estimate model parameters, but only the small box is segmented. For any two adjacent central core boxes, two-thirds of the voxels contributing to the parameter estimates are shared; consequently, the cutoff values vary from one core to the next in a smooth and seamless fashion. This box size works well on images with slices 1.2–1.4 mm thick, and the box size is adjusted accordingly if the segmentation is implemented on images with isotropic voxels.

Local boxes in areas of the brain that are complicated in shape (e.g., cerebellum) may contain too few data points for reliable convergence of the EM algorithm. If this occurs, the local box is increased in size (in all three directions) until it contains at least 10,000 brain voxels. Again, the enlarged box is used for model

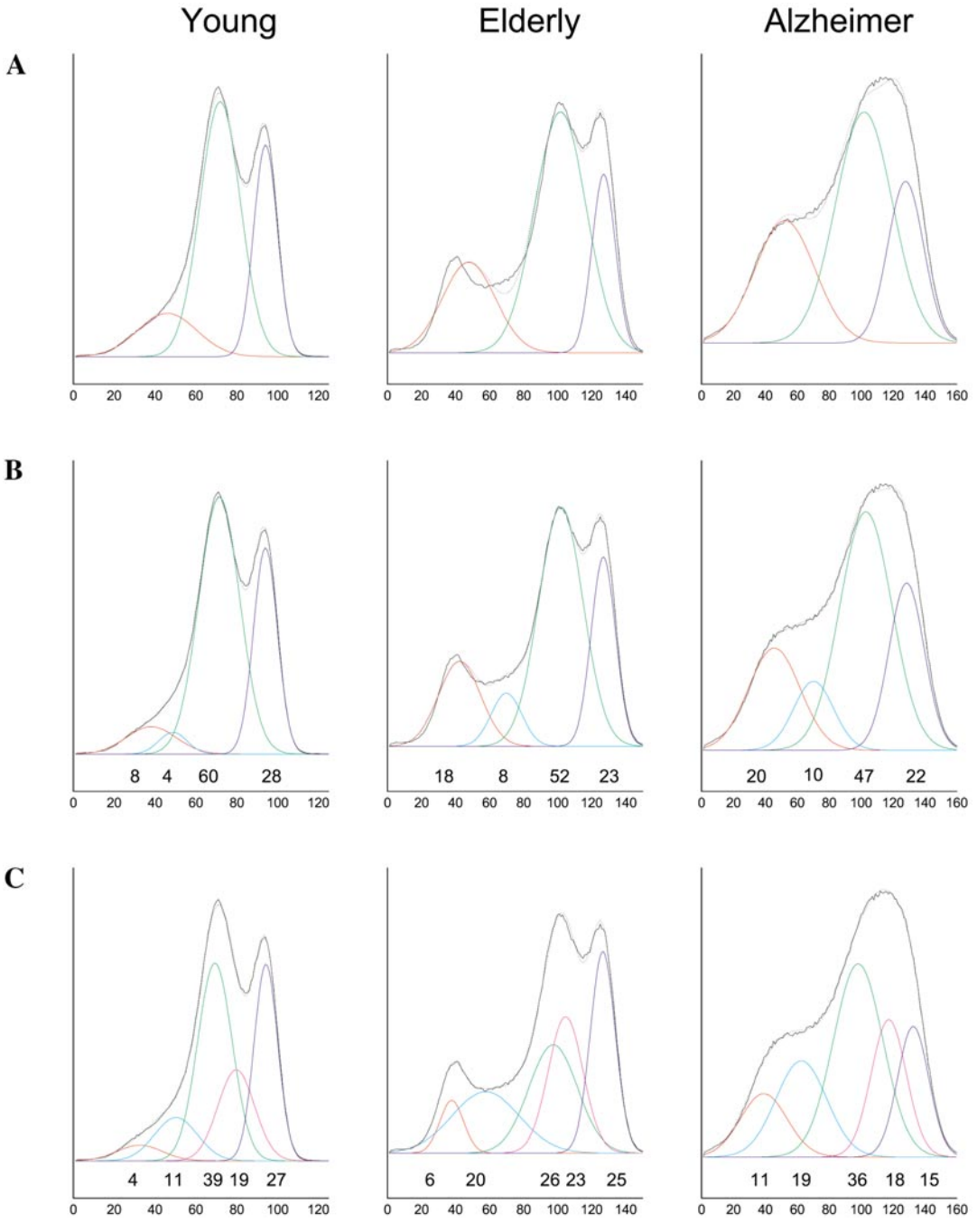

FIG. 6. Global scan histograms fitted with a mixture of (A) 3, (B) 4, and (C) 5 Gaussians. Left, young normal; center, elderly normal; right, probable AD patient. In B and C, numbers below each Gaussian indicate the proportion (weight) for each compartment.

fitting, but only the central core of the initial box is segmented. The T1 segmentation algorithm takes about 2 min on a 256 MHz Sun workstation.

RESULTS AND VALIDATION

Visual inspection. As part of the development of the protocol, a set of 10 images of varying image quality and resulting segmentations was inspected for plausibility by an expert neuroradiologist. In all cases, the brain extraction and segmentation results were judged acceptable, irrespective of image quality. However, as subjective impressions do not provide sufficient quantitative assessment, more stringent analyses were applied and are described below.

Reproducibility of brain extraction. Because reliable segmentation requires consistent separation of

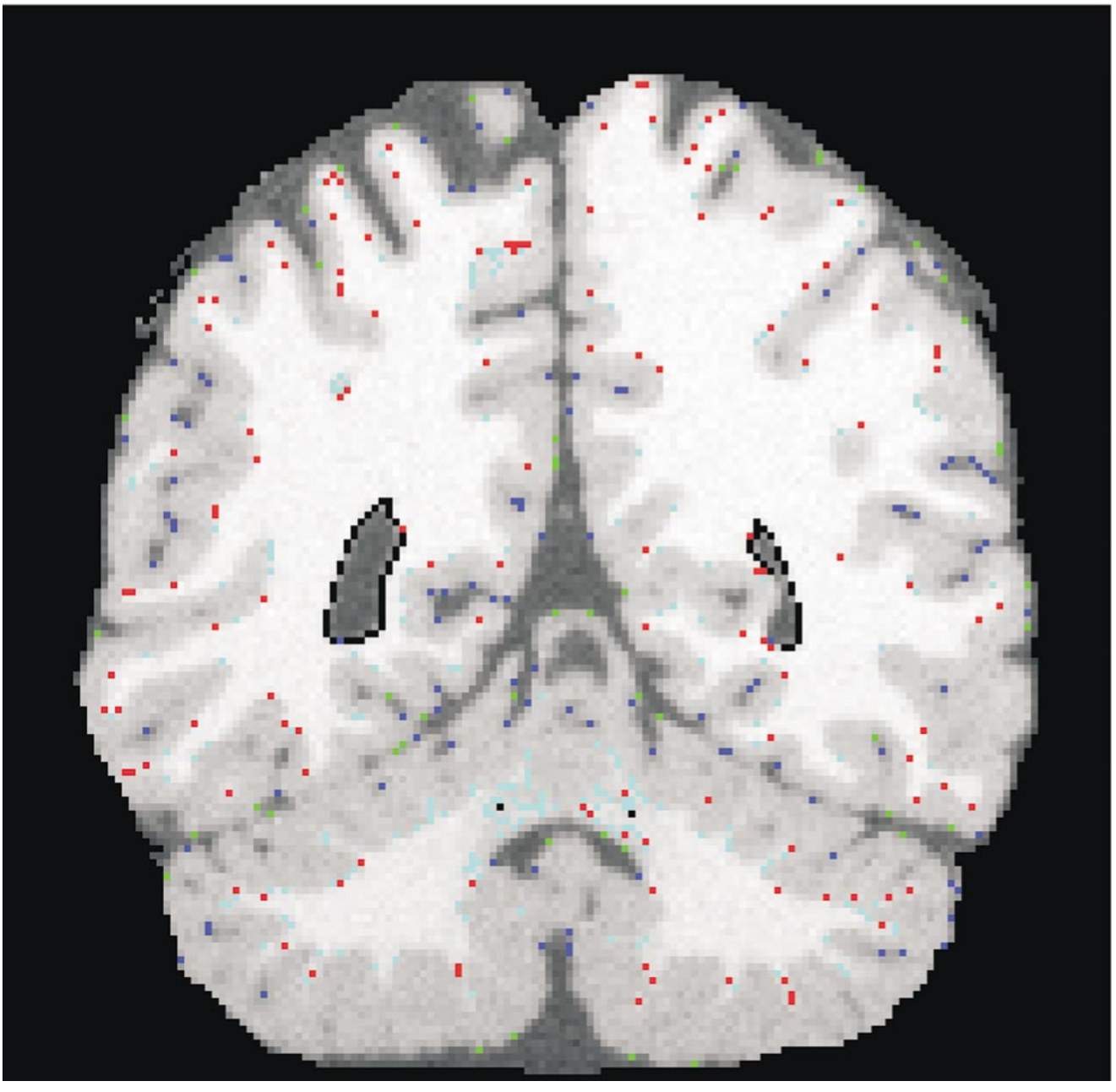

FIG. 8. Representative misclassified voxels in the MNI digital brain phantom. Blue, CSF misclassified as GM; green, GM misclassified as CSF; cyan, GM misclassified as WM; red, WM misclassified as GM.

brain from nonbrain tissues, short-term reproducibility of the brain extraction procedure was measured on images from eight young controls (20–30 years) and four elderly controls (>50 years). Each subject was scanned twice, with interscan intervals of 15 min to 1 month. One motivation for developing this procedure was to use it in images from patients. Thus, long-term reproducibility was tested in a set of 33 AD patients (50–83 years), with scans 1 year apart. TIC measurements based on the binary brain masks were obtained from each of the two scans and compared. Intraclass correlation coefficients indicated good correspondence in both groups across the two scans (ICCs = 0.999 and 0.998 for controls and AD, respectively). For the short-term assessment, the average error was 0.13% TIC (± 0.09% SD), which is equivalent to 1.84 ± 1.38 mL. For the AD patients, the average long-term error was 0.66% TIC (± 0.63% SD), or 9.44 ± 9.05 mL. The results here were similar to, or better than, results based on scans from young adults in other studies (Lemieux *et al.*, 1999, 0.40% TIC; Suckling *et al.*, 1999, 0.57% TIC, 8.3 ml; Alfano *et al.*, 1998, 16.3 ml

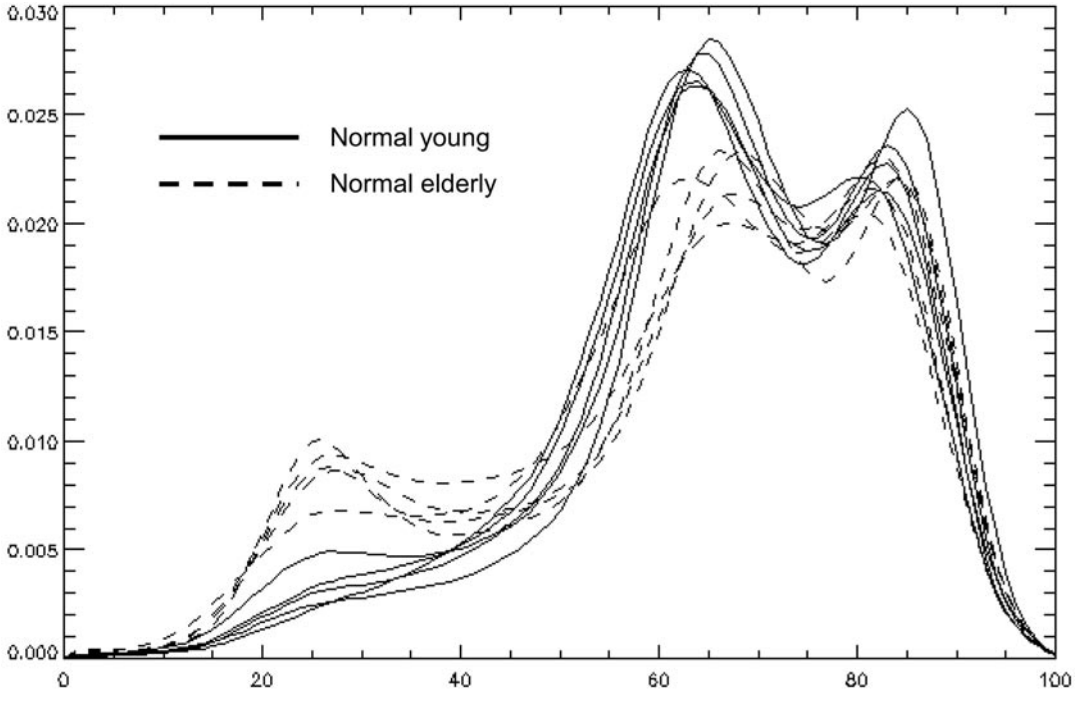

FIG. 7. Histograms from 10 images rescaled to an intensity range of 100. Solid lines represent histograms from 5 normal young brains; dotted lines represent histograms from 5 normal elderly brains.

SD). The choice to define the mask based on the lightly populated region between brain and nonbrain thus proves to be robust against intensity variation, within and across scans. These results indicate excellent short-term reproducibility in healthy normals and acceptable long-term reproducibility in elderly AD patients.

Reproducibility of segmentations. Reproducibility of the whole procedure (brain extraction plus segmentation) was tested on the repeated scans from the young and elderly normal controls. Measurements of total volumes for CSF, GM, and WM were compared. These results are summarized in Table 1, which shows volumetric differences of tissue estimates between two scans, averaged across all 12 subjects. As can be seen from the table, the maximal differences were less than 1% of TIC in all tissue classes.

TABLE 1

Segmentation Reliability: Scan–Rescan Absolute Differences in Global Brain Tissue Volumes

	CSF		GM		WM	
	cm^3	%TIC	cm^3	%TIC	cm^3	%TIC
Mean*	3.95	0.30	3.98	0.29	2.49	0.21
SD	2.38	0.20	1.52	0.11	3.05	0.26
Max	7.35	0.66	6.35	0.44	9.40	0.87
Min	0.11	0.01	0.41	0.04	0.17	0.05

* $n = 12$.

To test the reliability of segmentation results within local regions, a regional analysis was made on the set of repeated images from five of the normal controls. The Talairach proportional grid system (Talairach and Tournoux, 1988) was applied to each brain. This system produces 12 large ROIs (6 per hemisphere), defined by a reference system incorporating planes passing horizontally and vertically through the anterior and posterior commissures, and sagittally along midline. The anterior–posterior, superior–inferior, and lateral extents of the brain determine the planes that define the external boundaries of the 12 ROIs. Within each ROI, the obtained volumes for CSF, GM, and WM were ratio-adjusted for ROI tissue volume (range, 55–190 cm^3). Across all ROIs, average adjusted tissue volumes on the first scan for CSF, GM, and WM were 14.2, 51.0, and 34.8%, respectively. The scan–rescan results for this sample indicated no differences for any tissue compartment ($F(1,4) > 0.10$). The mean absolute scan–rescan differences in proportional tissue volume were 2.5, 0.8, and 1.3% for CSF, GM, and WM, respectively. Thus, this method appears to provide stable estimates of regional and global tissue volumes.

Accuracy of segmentations. The accuracy of the T1 segmentation algorithm was tested using a simulated MRI digital brain phantom (Kwan *et al.*, 1999), obtained from the Brain Imaging Center at the Montreal

TABLE 2

Comparison of Segmentation Algorithm Volumes with Digital Phantom Volumes

Tissue type	MNI phantom	Algorithm
CSF	372,107 (19.1)	367,195 (18.8)
GM	902,924 (46.3)	920,891 (47.2)
WM	674,765 (34.6)	661,710 (33.9)

Note. Data in mm^3 (percentage of phantom value).

Neurological Institute (MNI, www.bic.mni.mcgill.ca/brainweb). The simulated image contained 3% added noise and 20% RF nonuniformity. The actual noise level in our images is typically under 2.5% and RF nonuniformity is typically less than 20%. Before applying the segmentation algorithm, all nonbrain matter in the simulated image was removed, using the probabilities provided in the MNI fuzzy model. Since our segmentation protocol classifies only three tissue types, voxels from the tiny "glial" compartment, less than 6 cm^3, were not included. Most of those voxels form a single layer at the borders of the ventricles. They have been assigned to GM by others using this phantom (e.g., Ruan *et al.*, 2000), and our protocol would assign them to GM as well.

A comparison of total tissue volumes derived from the segmentation algorithm with "true" tissue volumes derived from the phantom is presented in Table 2. The percentages of misclassified voxels, based on true tissue volume, are shown in Table 3. The agreement was excellent, with errors for all tissue classes less than 1% of the total volume in the phantom. The errors were primarily located along tissue interfaces, as would be expected. This is demonstrated in Fig. 8, where misclassified voxels are indicated on a coronal slice of the digital phantom. Any segmentation based on grayscale intensity cutoffs is susceptible to image noise, which will have the greatest impact along tissue boundaries. Since noise filtering was not applied, it is likely that the misclassified voxels are largely introduced by the noise in the image. This is supported by the relative symmetry of the multiclass Type I and Type II errors (Table 3; Zhang, 1996).

Robustness. A robust model-fitting algorithm that has the ability to adjust to large differences in regional weight parameters is essential. To test this, 1000 initial parameter values were randomly generated and the variation in the resulting cutoff values was measured. The parameters were varied simultaneously and independently by drawing from independent normal distributions. Relative to initial parameter values, weight parameters were varied to ±30%, mean parameters were varied to ±10 grayscale units, and standard deviations were varied to ±3.0 grayscale units. This large magnitude of erratic variation is unlikely to occur for actual scans (see Fig. 7 for variability in normalized global histograms); nevertheless, the resulting variation for the two cutoffs was quite small. The mean absolute deviations for c_{cg} and c_{gw} were 0.51 and 0.33 grayscale levels, respectively. Given that the cutoff values are rounded to the nearest integer, the typical result was that the cutoff values were not changed. The maximal effect would be 1 intensity unit, introducing minimal error on total tissue volume.

DISCUSSION

The protocol is a fast, robust, and reliable technique for obtaining single-channel tissue segmentations of brain images. Three features of the protocol distinguish it from others in the literature. Perhaps most importantly, it can be used to segment images that vary widely in the shape of their global histograms. Second, we provide a method for automatically initializing the parameters for the Gaussian fitting algorithm, which results in a fully automatic segmentation protocol. In this present application, the EM algorithm proved to be insensitive to variation in initial parameter estimates. This is a considerable advantage over some other approaches (e.g., Grabowski *et al.*, 2000; Rajapakse *et al.*, 1996) that require the user to select initial values for means (peak intensities, tissue ROIs) and that are often quite sensitive to initialization values. Finally, the amount of manual editing required to obtain the binary brain mask appears to be less than other methods (e.g., ATOMIA, Ruan *et al.*, 2000). Because the initial mask is quite good, manual editing is restricted to obvious errors such as the inclusion of eyes, resulting in a protocol with high reproducibility in healthy and diseased brains.

Accuracy of segmentation protocols is difficult to assess; there is no gold standard with which to compare results. As manual segmentations can be quite unreliable (e.g., Pham *et al.*, 2000) and are time consuming,

TABLE 3

Percentage of Voxels Misclassified by the Algorithm and Error Probabilities

True classification	Algorithm misclassification[a]			Type I error[b]	Type II error[c]
	CSF	Gray	White		
CSF	—	3.7	0.0	3.7	2.2
Gray	0.9	—	2.1	3.0	4.3
White	0.0	3.8	—	3.8	2.8

[a] Expressed as the proportion of the true tissue type.
[b] Probability a voxel from the tissue class is identified as belonging to another class.
[c] Probability a voxel from other tissue classes are identified as belonging to the class.

the accuracy of the procedure was tested using the MNI digital phantom. The results were as good as or better than other published protocols that assessed accuracy using the phantom (Grabowski *et al.*, 2000; Ruan *et al.*, 2000; Van Leemput *et al.*, 1999) and that also obtained reasonable concordance with manual tracing (Grabowski *et al.*, 2000; Van Leemput *et al.*, 1999). Ruan *et al.* (2000) used a 5-Gaussian fitting algorithm followed by a Markov random field model. Their coefficient of total agreement at the same noise (3%) and RF inhomogeneity (20%) levels used here was 0.91 while ours was 0.97. The results here also compared favorably with the test of Van Leemput *et al.* (1999), which used 3% noise and no RF inhomogeneity. Their overlap metrics (described in Zijdenbos *et al.*, 1994) for GM and WM were approximately 0.94 and 0.93, respectively. In the presence of 20% RF inhomogeneity, our overlap matrices were 0.96 and 0.97, respectively. Grabowski *et al.* (2000) used the same noise and inhomogeneity values as were used here. In terms of proportion of misclassified voxels, the present algorithm was more accurate in the GM and WM compartments (3.0% vs 3.75% and 3.8% vs 4.86% misclassified voxels, respectively), but not in the CSF compartment (3.7% vs 2.77% misclassified voxels). Not surprisingly, voxels at tissue interfaces were most susceptible to misclassification. Visual inspection of the phantom results suggested these misclassifications were distributed evenly across the image, and error analysis (Table 3) confirmed this impression.

An extensive assessment of the robustness and reliability of the procedure was also made. The algorithm used for model fitting was quite stable, thus allowing automatic parameter initialization. The brain extraction procedure was robust against differences across scan intervals of 15 min to 1 year. The segmentation algorithm was also robust at both the global and regional level. Global scan–rescan differences were smaller than 1% of TIC, and local differences were less than 0.15% of TIC.

Other approaches to segmentation. Many recent segmentation protocols have incorporated explicit segmentation of partial volume mixture voxels and/or Markov random fields (Ruan *et al.*, 2000; Van Leemput *et al.*, 2001). These are interesting advances in tissue classification, but certain assumptions currently required for these techniques are not valid across the range of images we are segmenting, and their application would introduce more error than would be acceptable. The two assumptions of interest are, first, that the boundaries between tissue types are discrete in most of the brain and, second, that local regional homogeneity is the norm. We address each below.

The limited resolution and short acquisition times in typical MR images force a noisy undersampling of local tissue boundaries, resulting in partial volume voxels at tissue interfaces. When tissue boundaries are relatively discrete, as is typical in young brains, the behavior of this partial volume band can be simulated and incorporated into the segmentation scheme to produce partial volume tissue classes (Van Leemput *et al.*, 2001). In Alzheimer's disease, and even in normal aging, the boundary between areas that can be clearly defined as GM and WM appears to widen. This results in an increased number of voxels falling into the "par-

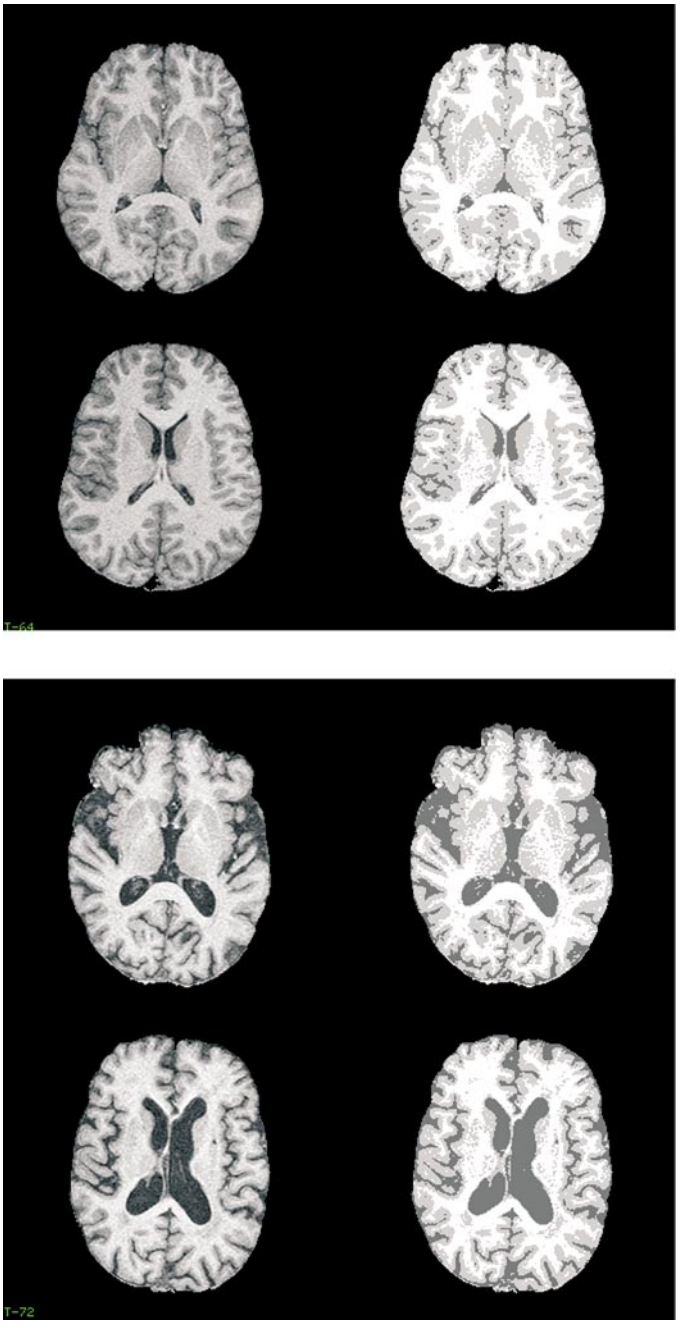

FIG. 9. Segmentation results for young normal brain (top) and normal elderly brain (bottom). White, white matter; light gray, gray matter; dark gray, CSF.

tial volume" intensity range, which are located some distance from the tissue interface. This reflects biological tissue changes rather than the effects of undersampling. The focus was to develop a robust method to identify tissues typically identified by experts as CSF and gray and white matter and we did not attempt to model this variance. While we take advantage of partial volume mixtures in the fitting of the Gaussians, we and others (Ruan *et al.*, 2000; Grabowski *et al.*, 2000) have used partial volume mixture modeling solely to aid in the accurate classification of CSF and gray and white matter, rather than to classify explicitly partial volume voxels (e.g., Van Leemput *et al.*, 2001; Laidlaw *et al.*, 1998).

The assumption of local tissue homogeneity is central to the application of Markov random field models in tissue segmentation. In most instances, tissue class is established, in part, on voxel intensities in local neighborhoods (e.g., Laidlaw *et al.*, 1998), which requires that tissue intensities be well-separated. In elderly normal and Alzheimer's brains, this requirement is violated in two ways. First is the appearance of speckling within white matter, and second is the general decrease in contrast between gray and white matter. The biological source of speckle is not fully understood, although white matter pallor and other changes have been observed in postmortem specimens (Grafton *et al.*, 1991; Kitajima *et al.*, 1999). In our algorithm, speckle in white matter is generally segmented as gray matter or CSF (see Fig. 9), especially in more severely diseased brains. We are evaluating the characteristics of the signal intensities of voxels in white-matter speckle to determine whether this feature truly reflects a fourth tissue compartment that can be reasonably modeled and separately segmented.

SUMMARY

Our protocol is similar to other methods using Gaussian models, but has certain features that make it more attractive. The algorithm provides a fast, iterative fit to a 4-Gaussian mixture with no constraints. The fully automatic segmentation was possible due to the implementation of a novel, robust initialization procedure. The protocol presented here produced accurate and stable solutions in a wide spectrum of brain images. Thus, this protocol is applicable in patient populations (Levine *et al.*, 2002) as well as in healthy controls.

ACKNOWLEDGMENTS

The authors thank C. Rockel, P. Roy, R. Swartz, and F. Q. Gao for their assistance in testing and validating the protocol. We thank G. Sela for his input in the initial stages of algorithm development. This work was supported by grants from the Canadian Institutes for Health Research awarded to A.F., M.J.B., and S.E.B.; from the Canadian Neurotrauma Research Program awarded to B.L.; and by financial support from General Electric Medical Systems, Canada. The C++ source code for the segmentation protocol is available by contacting the corresponding author.

REFERENCES

Alder, M. 2001. *An Introduction to Pattern Recognition*. http://www.HeavenForBooks.com.

Alfano, B., Quarantelli, M., Brunetti, A., Larobina, M., Covelli, E. M., Tedeschi, E., and Salvatore, M. 1998. Reproducibility of intracranial volume measurement by unsupervised multispectral brain segmentation. *Magn. Reson. Med.* **39:** 497–499.

Ashburner, J., and Friston, K. 1997. Multimodal image coregistration and partitioning—A unified framework. *NeuroImage* **6:** 209–217.

Ashburner, J., and Friston, K. J. 2000. Voxel-based morphometry—The methods. *NeuroImage* **11:** 805–821.

Cho, S., Jones, D., Reddick, W. E., Ogg, R. J., and Steen, R. G. 1997. Establishing norms for age-related changes in proton T1 of human brain tissue in vivo. *Magnetic Resonance Imaging* **15:** 1133–1143.

Clarke, L. P., Velthuizen, R. P., Camacho, M. A., Heine, J. J., Vaidyanathan, M., Hall, L. O., Thatcher, R. W., and Silbiger, M. L. 1995. MRI segmentation: methods and applications. *Magnetic Resonance Imaging* **13:** 343–368.

Grabowski, T. J., Frank, R. J., Szumski, N. R., Brown, C. K., and Damasio, H. 2000. Validation of partial tissue segmentation of single-channel magnetic resonance images of the brain. *NeuroImage* **12:** 640–656.

Grafton, S. T., Sumi, S. M., Stimac, G. K., Alvord, E. C., Jr., Shaw, C. M., and Nochlin, D. 1991. Comparison of postmortem magnetic resonance imaging and neuropathologic findings in the cerebral white matter. *Arch. Neurol.* **48:** 293–298.

Held, K., Kops, E. R., Krause, B. J., Wells, W. M., 3rd, Kikinis, R., and Muller-Gartner, H. W. 1997. Markov random field segmentation of brain MR images. *IEEE Trans. Med. Imaging* **16:** 878–886.

Hirai, T., Korogi, Y., Sakamoto, Y., Hamatake, S., Ikushima, I., and Takahashi, M. 1996. T2 shortening in the motor cortex: effect of aging and cerebrovascular diseases. *Radiology* **199:** 799–803.

Imon, Y., Yamaguchi, S., Katayama, S., Oka, M., Murata, Y., Kajima, T., Yamamura, Y., and Nakamura, S. 1998. A decrease in cerebral cortex intensity on T2-weighted with ageing images of normal subjects. *Neuroradiology* **40:** 76–80.

Kitajima, M., Korogi, Y., Okuda, T., Shiraishi, S., Ikeda, O., Morishita, S., Takahashi, M., and Eto, K. 1999. Hyperintensities of the optic radiation on T2-weighted MR images of elderly subjects. *Am. J. Neuroradiol.* **20:** 1009–1014.

Kwan, R. K., Evans, A. C., and Pike, G. B. 1999. MRI simulation-based evaluation of image-processing and classification methods. *IEEE Trans. Med. Imaging* **18:** 1085–97.

Laakso, M. P., Partanen, K., Soininen, H., Lehtovirta, M., Hallikainen, M., Hanninen, T., Helkala, E. L., Vainio, P., and Riekkinen, P. J., Sr. 1996. MR T2 relaxometry in Alzheimer's disease and age-associated memory impairment. *Neurobiol. Aging* **17:** 535–540.

Laidlaw, D. H., Fleischer, K. W., and Barr, A. H. 1998. Partial-volume bayesian classification of material mixtures in MR volume data using voxel histograms. *IEEE Trans. Med. Imaging* **17:** 74–86.

Lemieux, L., Hagemann, G., Krakow, K., and Woermann, F. G. 1999. Fast, accurate, and reproducible automatic segmentation of the brain in T1-weighted volume MRI data. *Magnetic Resonance Med.* **42:** 127–135.

Levine, B., Katz, D. I., Dade, L., and Black, S. E. 2002. New approaches to the assessment of frontal damage and executive defi-

cits in traumatic brain injury. In *Principles of Frontal Lobe Function* (D. T. Stuss and R. Knight, Eds.), pp. 448–465. Oxford Univ. Press, New York.

McLachlan, G. J., and Krishnan, T. 1997. *The EM Algorithm and Extensions.* Wiley, New York.

Miot, E., Hoffschir, D., Poncy, J. L., Masse, R., Le Pape, A., and Akoka, S. 1995. Magnetic resonance imaging in vivo monitoring of T2 relaxation time: Quantitative assessment of primate brain maturation. *J. Med. Primatol.* **24:** 87–93.

Pham, D. L., Xu, C., and Prince, J. L. 2000. Current methods in medical image segmentation. *Annu. Rev. Biomed. Eng.* **2:** 315–337.

Rajapakse, J. C., Giedd, J. N., DeCarli, C., Snell, J. W., McLaughlin, A., Vauss, Y. C., Krain, A. L., Hamburger, S., and Rapoport, J. L. 1996. A technique for single-channel MR brain tissue segmentation: Application to a pediatric sample. *Magnetic Resonance Imaging* **14:** 1053–65.

Rajapakse, J. C., Giedd, J. N., and Rapoport, J. L. 1997. Statistical approach to segmentation of single-channel cerebral MR images. *IEEE Trans. Med. Imaging* **16:** 176–186.

Ruan, S., Jaggi, C., Xue, J., Fadili, J., and Bloyet, D. 2000. Brain tissue classification of magnetic resonance images using partial volume modeling. *IEEE Trans. Med. Imaging* **19:** 1179–1187.

Schnack, H. G., Hulshoff Pol, H. E., Baare, W. F., Staal, W. G., Viergever, M. A., and Kahn, R. S. 2001. Automated separation of gray and white matter from MR images of the human brain. *NeuroImage* **13:** 230–237.

Sled, J. G., Zijdenbos, A. P., and Evans, A. C. 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* **17:** 87–97.

Steen, R. G., Langston, J. W., Ogg, R. J., Xiong, X., Ye, Z., and Wang, W. C. 1999. Diffuse T1 reduction in gray matter of sickle cell disease patients: Evidence of selective vulnerability to damage? *Magnetic Resonance Imaging* **17:** 503–515.

Steen, R. G., Ogg, R. J., Reddick, W. E., and Kingsley, P. B. 1997. Age-related changes in the pediatric brain: Quantitative MR evidence of maturational changes during adolescence. *Am. J. Neuroradiol.* **18:** 819–828.

Suckling, J., Brammer, M. J., Lingford-Hughes, A., and Bullmore, E. T. 1999. Removal of extracerebral tissues in dual-echo magnetic resonance images via linear scale-space features. *Magnetic Resonance Imaging* **17:** 247–256.

Talairach, J., and Tournoux, P. 1988. *Co-Planar Stereotaxic Atlas of the Human Brain.* Thieme Medical, New York.

Thatcher, R. W., Camacho, M., Salazar, A., Linden, C., Biver, C., and Clarke, L. 1997. Quantitative MRI of the gray-white matter distribution in traumatic brain injury. *J. Neurotrauma* **14:** 1–14.

Thompson, P. M., Mega, M. S., Woods, R. P., Zoumalan, C. I., Lindshield, C. J., Blanton, R. E., Moussai, J., Holmes, C. J., Cummings, J. L., and Toga, A. W. 2001. Cortical change in Alzheimer's disease detected with a disease-specific population-based brain atlas. *Cereb. Cortex* **11:** 1–16.

Van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. 1999. Automated model-based tissue classification of MR images of the brain. *IEEE Trans. Med. Imaging* **18:** 897–908.

Van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. 2001. A statistical framework for partial volume segmentation. *Lecture Notes Comput. Sci.* **2208:** 204–212.

Woods, R. P., Grafton, S. T., Holmes, C. J., Cherry, S. R., and Mazziotta, J. C. 1998. Automated image registration. I. General methods and intrasubject, intramodality validation. *J. Comput. Assist. Tomogr.* **22:** 139–152.

Zhang, Y., Brady, M., and Smith, S. 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* **20:** 45–57.

Zhang, Y. J. 1996. A survey on evaluation methods for image segmentation. *Pattern Recognition* **29:** 1335–1346.

Zijdenbos, A. P., Dawant, B. M., Margolin, R. A., and Palmer, A. C. 1994. Morphometric analysis of white matter lesions in MR images: Method and validation. *IEEE Trans. Med. Imaging* **13:** 716–724.