NeuroImage **13,** 856–876 (2001)
doi:10.1006/nimg.2000.0730, available online at http://www.idealibrary.com on IDEAL®

Magnetic Resonance Image Tissue Classification Using a Partial Volume Model

David W. Shattuck,* Stephanie R. Sandor-Leahy,† Kirt A. Schaper,‡
David A. Rottenberg,‡'§ and Richard M. Leahy*

**Signal and Image Processing Institute, University of Southern California, Los Angeles, California 90089; †TRW, Inc., Redondo Beach, California 90278; ‡PET Imaging Center, VA Medical Center, Minneapolis, Minnesota 55417; and §Department of Neurology and Department of Radiology, University of Minnesota, Minneapolis, Minnesota 55455*

Received May 26, 2000

We describe a sequence of low-level operations to isolate and classify brain tissue within T1-weighted magnetic resonance images (MRI). Our method first removes nonbrain tissue using a combination of anisotropic diffusion filtering, edge detection, and mathematical morphology. We compensate for image nonuniformities due to magnetic field inhomogeneities by fitting a tricubic B-spline gain field to local estimates of the image nonuniformity spaced throughout the MRI volume. The local estimates are computed by fitting a partial volume tissue measurement model to histograms of neighborhoods about each estimate point. The measurement model uses mean tissue intensity and noise variance values computed from the global image and a multiplicative bias parameter that is estimated for each region during the histogram fit. Voxels in the intensity-normalized image are then classified into six tissue types using a maximum *a posteriori* classifier. This classifier combines the partial volume tissue measurement model with a Gibbs prior that models the spatial properties of the brain. We validate each stage of our algorithm on real and phantom data. Using data from the 20 normal MRI brain data sets of the Internet Brain Segmentation Repository, our method achieved average κ indices of κ = 0.746 ± 0.114 for gray matter (GM) and κ = 0.798 ± 0.089 for white matter (WM) compared to expert labeled data. Our method achieved average κ indices κ = 0.893 ± 0.041 for GM and κ = 0.928 ± 0.039 for WM compared to the ground truth labeling on 12 volumes from the Montreal Neurological Institute's BrainWeb phantom.

INTRODUCTION

Accurate classification of magnetic resonance images according to tissue type at the voxel level is important in many neuroimaging applications. Changes in the composition of gray matter (GM), white matter (WM), or cerebrospinal fluid (CSF) in the whole volume or within specific regions can be used to characterize physiological processes and disease entities (Guttmann *et al.*, 1998) or to characterize disease severity (Heindel *et al.*, 1994). Additionally, fractional content images can be used to correct for effects of cerebral atrophy in quantitative PET and MR spectroscopy image volumes (Koepp *et al.*, 1997; Ibanez *et al.*, 1998). Low-level tissue classifications also provide anatomical information about structures such as the cortical surface, which is useful for visualization and analysis of neuroimaging data (Thompson *et al.*, 1998) and as a constraint for localizing functional activation from magnetoencephalography and electroencephalography data (Dale and Sereno, 1993; Baillet *et al.*, 1999). While MRI can provide relatively high-resolution neuroanatomical details, the identification of tissue information is hindered by several factors. These include measurement noise, partial volume effects, and image nonuniformity due to magnetic field inhomogeneities. These artifacts confound characterization of tissue types based solely on individual voxel intensities.

The problem of segmenting MRI has been extensively addressed by many researchers. A common approach to structural segmentation is the use of atlas registration-based techniques such as those of Bajcsy *et al.* (1983), Bajcsy and Kovacic (1989), Miller *et al.* (1993), Christensen *et al.* (1996), Christensen (1999), and Davatzikos (1997). In these methods a prelabeled atlas volume is matched to a subject volume. Once registration is achieved, the labels are transfered to the subject and structures in the subject volume are identified. While these techniques can provide very good matches with subcortical structures, they do not typically match the cortex well due to intersubject variability of cortical sulci and gyri. These methods can also be very computationally intensive. Atlas-based segmentation techniques have been combined with low-level tissue classification by Collins *et al.*, (1999) greatly improving tissue classification rates in the cortex.

1053-8119/01 $35.00

In contrast to atlas-based approaches are low-level techniques in which the tissue content of each individual voxel in the image is identified. A variety of methods have been developed for low-level tissue classification, most of which are focused on one or more of the three tasks addressed in this paper. The first task is skull stripping, in which nonbrain tissue is removed from the MRI. The second task is compensation for image nonuniformity. The third task is the actual labeling of the individual voxels.

In this paper we present a three-stage method that addresses the problem of isolating the brain from the calvarium in T1-weighted MRI and classifying its tissues. The stages of the method have been developed to be minimally interactive and operate in reasonable time on widely available desktop computer hardware. We begin by applying a skull-stripping routine that uses anisotropic diffusion filtering, edge detection, and mathematical morphology. We then compensate for nonuniformity in the stripped brain using a method that fits a tricubic B-spline to local estimates of image gain variation. The local estimates are computed by fitting a partial volume tissue measurement model to histograms of local regions of the image. The measurement model uses mean tissue intensity and noise variance values computed from the global image and a multiplicative bias parameter that is estimated for each region during the histogram fit. Each voxel in the intensity-normalized image is then labeled using a maximum *a posteriori* (MAP) classifier that combines the tissue measurement model with a Gibbs prior that models the spatial properties of brain tissue. We validate each stage of our approach using real and phantom MRI data. We first review the related literature on each of the three steps in our algorithm.

Skull Stripping

The first task we address is the analysis of the MRI volume to identify brain and nonbrain voxels. Our work is concerned with the predominant tissues of the brain: GM, WM, and CSF. The measured signal intensities of these tissues can overlap with those of other tissues in the head, such as skin, bone, muscle, fat, and dura. This complicates reliable identification of tissue regions and properties. The difficulties faced by later stages in our method are lessened by clearly identifying what is brain tissue and what is not.

Skull stripping is often performed using a sequence of mathematical morphological operations following an initial separation of the brain from other tissues of the head. Bomans *et al.* (1990) apply a Marr-Hildreth edge detector, followed by manual labeling of connected components within each slice of the volume and a morphological closing operation applied to improve the surface definition. Brummer *et al.* (1993) segment each slice of the brain using thresholds determined from estimated gray and white matter intensities, followed by a sequence of connected component and morphology operations. Sandor and Leahy address the task of brain extraction using anisotropic diffusion filtering, Marr-Hildreth edge detection, and morphological operations (Sandor, 1994; Sandor and Leahy, 1997); the method we use is a refinement of this procedure. Kapur *et al.* (1996) use an initial classification found by the expectation-maximization (E-M)-based classification method by Wells *et al.* (1996) described below, followed by morphological processing to isolate the brain. This extraction is then refined if necessary using a deformable surface. Kruggel and von Cramon (1999) have produced a "brain peeling" procedure that performs an initial Lee filtering to smooth nonuniformities, followed by a brain delineation procedure that combines distance-transforms and edge gradients and finally a sequence of morphological steps to clean the initial brain mask. Other examples of skull stripping techniques include the use of deformable templates, as described by Dale *et al.* (1999).

Image Nonuniformity

In an ideal MRI acquisition system, the signal intensity measured at each voxel will vary through-out the volume depending only on the tissues present at that voxel location. In practice, MRI exhibits nonuniform tissue intensities caused by inhomogeneities in the magnetic fields, magnetic susceptibility variations in the subject, and other factors (Sled and Pike, 1998). As a result, tissue labels cannot be reliably assigned to voxels based solely on individual voxel intensity. Fortunately, nonuniformity in MRI is typically characterized by a slowly varying gain field, also called a bias field (Wells *et al.*, 1996), and several methods have been developed to compensate for this artifact.

Nonuniformity correction is often applied prior to tissue classification. Dawant *et al.* (1993) identify white matter reference points within the MRI volume and normalize the image intensities relative to these; a similar technique is used by Dale *et al.* (1999). DeCarli *et al.* (1996) compute the difference between local median and global median values in order to estimate intensity variation. Techniques that perform bias correction prior to tissue classification also include homomorphic unsharp masking and other filtering techniques (Brinkmann *et al.*, 1998). Sled *et al.* (1998) address the problem of nonuniformity using an approach that estimates a gain field to sharpen the histogram of the MRI; this field is kept smooth using a cubic B-spline. This method has been publicly released under the name nonparametric nonuniform intensity normalization

(N3) (available from http://www.bic.mni.mcgill.ca/software/N3/).

A number of iterative methods correct for nonuniformity by alternating between tissue classification and nonuniformity estimation. Wells *et al.* (1996) developed an E-M-based method in which they perform a statistical classification of the image, then low-pass filter the residual difference between a reconstructed image and the measured image to provide an estimate of the image bias. This bias is removed from the image and the process is repeated. Guillemaud and Brady (1997) take a similar approach using an additional tissue class and report improved results. Van Leemput *et al.* (1999a) use the same classification step as Wells, but estimate the bias field by fitting a fourth-order polynomial to the bias field residual. Yan and Karp (1995) account for the inhomogeneity by incorporating a cubic B-spline into a MAP classifier. MAP labeling approaches combined with bias estimation are also described by Nocera and Gee (1997) and Rajapakse and Kruggel (1998). In each of these iterative methods an initial MAP classification is used to reconstruct an estimated image, and the residual between this image and the measured image is smoothed using a tricubic B-spline to represent the nonuniformity; this procedure is then iterated. In a similar manner, Pham and Prince (1999) incorporate a gain field into an adaptive fuzzy C-means algorithm to produce a spatially adaptive method for classification that compensates for image nonuniformity.

Tissue Labeling

We define tissue classification to be the process of assigning a label to each voxel in the volume that identifies what types of tissues are present in that voxel. In addition to measurement noise and image nonuniformity, the classification task is made more difficult by the finite resolution of the scanning hardware. In a high-resolution anatomical MRI, the image is sampled at a spacing on the order of 1 mm in each dimension. The complexity of anatomical structures results in numerous voxels that contain multiple tissue types. The intensity measured at such a voxel will be a weighted average of the intensities of the tissues within it; this effect is known as partial volume averaging.

Mensuration of brain tissue volumes with modeling for partial volume effects is useful for PET metabolic and ligand–receptor studies. Early work on computing tissue proportions in X-ray computed tomography appears in Thaler *et al.* (1978). Choi *et al.* (1991) developed a generalized partial volume model in the form of a Markov random field allowing any number of tissues and independent measurements. Rusinek *et al.* (1991) developed a partial volume estimation method for two-channel MR measurements obtained using specialized pulse sequences. Bonar *et al.* (1993) used graphical analysis of the 2-D feature space formed by two-channel measurements to assign tissue proportions in a three-compartment model. Santago and Gage (1993) developed a partial volume model for analyzing the tissue content of an entire volume to determine relative tissue contribution. Laidlaw *et al.* (1998) classified MRI using this partial volume model, adjusting it to a region about each voxel to identify the fractional content of that particular voxel. Warfield *et al.* (1999) developed a technique focused on fractional segmentation of white matter into damaged and nondamaged tissue, after preprocessing by the nonuniformity correction method described in Wells *et al.* (1996) and the use of a spatial template for initial white matter and gray matter segmentation. Numerous algorithms have been described that use Gibbs priors as statistical models for the labeled brain volume. Labeling is performed by combining the prior with a likelihood for the data conditioned on the labels and computing a MAP estimate (Leahy *et al.*, 1991; Choi *et al.*, 1991; Yan and Karp, 1995; Kapur *et al.*, 1996; Nocera and Gee, 1997; Rajapakse and Kruggel, 1998; Van Leemput *et al.*, 1999b). Pham and Prince (1999) proposed an adaptive fuzzy C-means method in which the tissue fraction at each voxel is the membership function for that voxel belonging to a particular class. This is one of several statistical classification and segmentation techniques that compensate for nonuniformity during the classification process (Yan and Karp, 1995; Wells *et al.*, 1996; Nocera and Gee, 1997; Guillemaud and Brady, 1997; Van Leemput *et al.*, 1999b, as described in the previous section. There are also numerous techniques for associating voxels with tissue classes that involve labeling without an inhomogeneity correction, including methods based on neural-network models (Özkan *et al.*, 1993) and graph theoretics (Wu and Leahy, 1991).

The tissue classification method used by Dale *et al.* (1999) is intended for identifying the cerebral cortex, hence only white matter tissue voxels are identified. The cortical surface identification approach of Zeng *et al.* (1999) uses a level-set method to find coupled surfaces representing the interior and exterior boundary of the cerebral cortex. Neither of these methods labels areas of the brain other than cortex.

METHODS

Skull Stripping

The tissue model we use for nonuniformity estimation and voxel classification allows for CSF, GM, and

WM. Our classification problem is greatly simplified if we strip skull and scalp tissue from the MRI volume prior to nonuniformity correction and tissue classification. Skull stripping is also a useful preprocessing tool for image registration techniques such as AIR (Woods *et al.*, 1998). Our procedure consists of three steps. First, the MRI is processed with an anisotropic diffusion filter to smooth nonessential gradients. We then apply a Marr–Hildreth edge detector to the filtered image to identify important anatomical boundaries. The objects defined by these boundaries are identified and refined using a sequence of morphological and connected component operations.

Anisotropic Diffusion Filtering

Our method finds anatomical boundaries that separate the brain from the dura, skull, and other tissues. In some cases, particularly with MRI data that have low signal-to-noise ratios, these boundaries will be obscured by noise, or their edges will be indistinguishable from other edges in the image. For this reason we apply an anisotropic diffusion filter that smoothes noisy regions in the image while respecting edge boundaries.

Anisotropic diffusion filtering was proposed as an image processing method by Perona and Malik (1990). Gerig *et al.* (1992) applied this 2-D technique to slices of MRI data; it has also been extended to 3-D (Mackiewich, 1995). The filtered image is modeled as the solution to the anisotropic diffusion equation

$$\frac{\partial I}{\partial t} = \nabla\cdot(c(\boldsymbol{p}, t)\nabla I) = c(\boldsymbol{p}, t)\nabla^2 I + \nabla c \cdot \nabla I, \qquad (1)$$

where $\boldsymbol{p}$ is a point in R^3, ∇ and ∇^2 represent the gradient and Laplacian operators with respect to spatial variables, and $\nabla\cdot$ indicates the divergence operator (Perona and Malik, 1990). If c has a constant value, then Eq. (1) is simply the isotropic heat diffusion equation (John, 1982) whose solution is the original image convolved with a Gaussian kernel with variance that increases linearly with time. This convolution will blur both strong and weak edges.

Instead of this, Perona and Malik proposed the use of diffusion coefficients based on a measure of edge strength. The diffusion coefficient $c(\boldsymbol{p}, t)$ then adaptively controls the diffusion strength, smoothing the image within a moderately continuous region while not smoothing across sharp discontinuities. Perona and Malik demonstrated that using the gradient of image intensity as an estimate of edge strength produces excellent results. We use the function

$$c(\boldsymbol{p}, t) = g(\|\nabla I(\boldsymbol{p}, t)\|) = e^{-\|\nabla I(\boldsymbol{p},t)\|^2/\kappa_d^2}, \qquad (2)$$

where κ_d is the diffusion constant. Using Eq. (2) gives preference to high-contrast edges over low-contrast ones. The diffusion Eq. (1) is discretized onto a 3-D lattice resulting in an update equation for each voxel in the image,

$$i_k^{(n+1)} = i_k^{(n)} + \tau_0 \sum_{j\in N_k} c(j, n)(i_{r(j,k)}^{(n)} - i_k^{(n)}), \qquad (3)$$

where k is a 3-D spatial index, N_k is the set of 6 neighbors nearest k, i_k^n is the intensity of the voxel indexed by k at the nth time step, $r(j, k)$ is the voxel adjacent to k opposite j, and τ_0 is the size of the time step used to discretize the system. To ensure stability, τ_0 is selected to be $0 \le \tau_0 \le \frac{1}{7}$ (Mackiewich, 1995); we use a time step of $\frac{1}{8}$. All intensity values in the volume are updated at each iteration.

We use two parameters to specify the anisotropic diffusion filter used in our skull stripping technique. The first is the diffusion parameter, κ_d, and the second is the number of iterations, possibly none, applied to the image. Selection of these parameters is done empirically, though in practice we have found that a single set of parameter settings can often be found for a specific MRI acquisition system and protocol.

Edge Detection

We locate the anatomical boundaries in MRI brain volumes using the Marr–Hildreth edge detector (Marr and Hildreth, 1980). This detector has a low computational cost and produces closed contours. Other edge finding approaches, such as the Canny and Deriche method (Deriche, 1987), could also be used to produce edge segments from the image. However, in certain regions these segments may be too dense to be formed into anatomical boundaries using a simple edge linking procedure.

The Marr-Hildreth edge detector is based on a low-pass filtering step with a symmetric Gaussian kernel, followed by the localization of zero-crossings in the Laplacian of the filtered image. The Marr–Hildreth operator, extended to three dimensions, is defined as

$$C(k) = \nabla^2(I(k)*g_\sigma(\boldsymbol{p})), \qquad (4)$$

where C is the output contour image, I is an input image, *is the convolution operator, g_σ is a Gaussian kernel with variance σ^2,

$$g_\sigma(\boldsymbol{p}) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\|\boldsymbol{p}\|^2/2\sigma^2}, \qquad (5)$$

$\boldsymbol{p}$ is a point in the 3-D volume, and ∇^2 is the Laplacian operator (Bomans *et al.*, 1990). By finding the voxels in the contour image, C, where zero-crossings occur, a binary image E is produced that separates the image into edge-differentiated segments. The σ parameter in

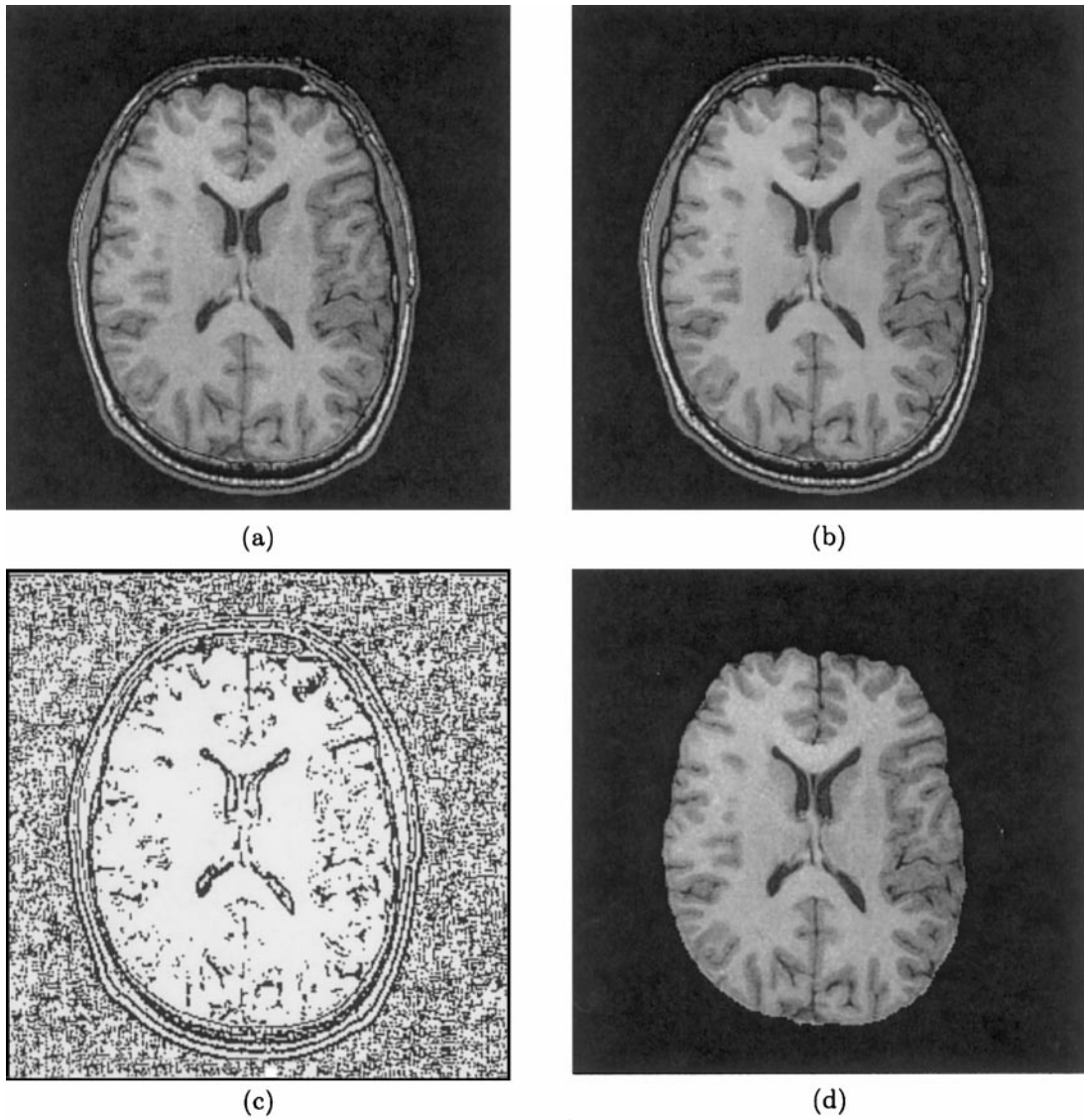

FIG. 1. Skull-stripping stages. (a) A slice from the initial volume. (b) The same slice after anisotropic diffusion filtering (three iterations with diffusion parameter 15). (c) The edge map following application of the Marr–Hildreth operator (edge kernel $\sigma = 0.6$). (d) The same slice from the extracted brain, following morphological processing of the edge map.

Eq. (5) provides the third parameter for the algorithm. Increasing the value of σ makes the blurring kernel wider, and only very strong edges remain in the image. Small values for σ produce narrow filters, resulting in more edges in the image.

Using the Laplacian as opposed to the second derivative in the direction of the gradient implies the well-known poor localization performance of this detector for rounded edges with large curvature. Spurious edges may also appear with the Marr–Hildreth operator due to the closed nature of zero-crossing contours.

We have found that at a resolution appropriate for extracting the brain surface, this processor may detect sulcal boundaries corresponding to white–gray matter transitions rather than gray–CSF transitions; similar results are described in Bomans *et al.* (1990). Thus, the cortical surface that we extract during edge detection has a tendency to move between the outer cortical boundary on the gyri and the inner cortical surface in the sulcal folds. This may be overcome during morphological processing.

Morphological Processing

The edge detection step produces a binary volume $E(k)$, where k is a 3-D spatial index describing a point in Z^3, the 3-D integer space. $E(k)$ represents edge voxels by values of 0 while nonedge voxels are represented by values of 1; an example image is shown in Fig. 1c, in which edge voxels are black and nonedge voxels are

white. We denote the set of all nonedge voxels as $\boldsymbol{X} = \{x_k: E(k) = 1\}$. The task of the morphological processor is to select the voxels corresponding to the brain tissue from the original image based on the voxels in $\boldsymbol{X}$.

The edge detector will produce a distinct boundary between the brain and the surrounding tissues in most areas of the MRI volume. The brain should be well separated from the rest of the cerebral tissue by edge voxels in $E(k)$. However, there are often meninges and blood vessels that are not distinguished from the brain due to noise, low contrast between brain and meninges, or true anatomical continuity. The set $\boldsymbol{X}$ will most likely contain voxels that link brain regions to extraneous surrounding structures, such as dura or skin. Fortunately, the connective tissues are characteristically thin with typical widths of 2 mm or less. Connections of this nature are broken by performing a morphological erosion on $\boldsymbol{X}$ using a rhombus structuring element of size 1, which is a 3-D cross with a width of 3 voxels. The first morphological step in the algorithm is described by

$$\boldsymbol{X}_E = \boldsymbol{X} \ominus R1, \tag{6}$$

where $\boldsymbol{X}$ is the input edge set, $\boldsymbol{X}_E$ is the eroded volume, $\ominus$ is the erosion operator, and $R1$ is the size 1 rhombus structuring element. This operation will delete narrow connections without globally damaging or distorting the image. For MRI resolutions on the order of 1 mm, regions with a width of 2 mm in any direction will be deleted.

The eroded set X_E will consist of several connected components. In this context, we use connectedness in the 6-neighbor sense, in which two voxels are connected if and only if they share a common face. The erosion step Eq. (6) will eliminate a number of small structures in the volume and provide increased separation of larger anatomical structures. We assume that upon completion of this step the largest connected region centered in the volume consists entirely of brain tissue. Additional tests on the mean intensity of connected component regions in the MRI volume help ensure that we select the brain and not the background. At this point in the algorithm, a human operator can also intercede should the algorithm select the wrong connected component. We assume that the selected region contains nearly all of the brain voxels. We describe the selection step as

$$\boldsymbol{X}_I = S_{CC}(\boldsymbol{X}_E), \tag{7}$$

where $S_{CC}(\cdot)$ is our procedure for selecting the largest connected component that is centered in the volume.

The set X_I will be slightly smaller than the actual brain volume due to the initial erosion step. For this reason, we dilate X_I using the $R1$ structuring element to create a set X_D that will cover nearly the entire brain region. This is described by the equation

$$\boldsymbol{X}_D = \boldsymbol{X}_I \oplus R1, \tag{8}$$

where $\oplus$ is the dilation operator.

Due to imperfections in the edge boundaries identified by the Marr–Hildreth edge detector, $\boldsymbol{X}_D$ may contain pits in its surface or even small holes within the volume. A closing routine, consisting of a dilation followed by an erosion, will fill small pits in the surface and close off some holes that occur within the volume. However, the closing routine is often insufficient to fill larger holes inside the volume. We apply a special closing operator, which we denote as $\odot$. This operator includes a routine to fill background cavities in the brain volume between the dilation and the erosion steps. Thus any regions of background voxels that are connected components lying completely within the dilated brain volume will be filled. We use an octagonal morphological element, $O4$, that has a diameter of 9 voxels. $O4$ approximates a sphere, and closing with this element will remove any surface pits and fill any holes that have diameter of 9 voxels or less. We apply this operation to X_D to produce our closed volume, Ω:

$$\Omega = \boldsymbol{X}_D \odot O4. \tag{9}$$

The morphological steps of the algorithm are summarized in Table 1.

The brain surface extraction output may be further modified for certain neuroimaging needs. By taking the intersection of the edge map with Ω, we can re-introduce edge detail onto the smooth brain mask; we term this image the high-detail brain mask. By computing the boundary of the high-detail brain mask, we obtain a representation of the outer cerebral cortex. This representation may, in fact, wander between the outer cortical gray and white matter due to poor edge localization. A better way to obtain the cortical surface is to complete the process of tissue classification. The sequence of skull-stripping operations is illustrated in Fig. 1.

TABLE 1

Summary of Morphological Operations for Brain Extraction

Operation	Description
$\boldsymbol{X}_E = \boldsymbol{X} \ominus R1$	Erosion to separate brain
$\boldsymbol{X}_I = S_{CC}(\boldsymbol{X}_E)$	Selection of brain
$\boldsymbol{X}_D = \boldsymbol{X}_I \oplus R1$	Dilation to restore brain
$\Omega = \boldsymbol{X}_D \odot O4$	Closing

Note. $\boldsymbol{X}$ is the initial set obtained from edge detection; Ω is the final set representing the smooth brain mask.

Image Nonuniformity Compensation

Image Model

Inhomogeneity in the magnetic fields used during image acquisition and magnetic susceptibility variations in the scanned subjects cause intensity nonuniformities in MRI that prevent characterization of voxel tissue content based solely on image intensity. As a result, segmentation and quantitative studies of MRI require compensation for these nonuniformities. We model the nonuniform image gain as a spatially slowly varying multiplicative bias field. We estimate the variations in the bias field by fitting a parametric tissue measurement model to the histograms of small neighborhoods within the stripped MRI volume. The measurement model uses tissue class means and noise variance estimated from the global image, as well as parameters for the local multiplicative bias and the probability of each tissue type within the neighborhood. The local parameters are estimated during the histogram fitting procedure. We smooth and interpolate these values using a regularized tricubic B-spline, which provides our estimate of the nonuniformity field.

Santago and Gage (1993) proposed a tissue measurement model for the purpose of the quantification of brain tissue in an MRI based on its intensity histogram. We extend this model to include a spatially variant multiplicative bias term, b_k, which describes the nonuniformity effect at the kth voxel in Ω. Our model for the measurement process is

$$x_k = b_k y_k + \eta_k \qquad k \in \Omega, \tag{10}$$

where x_k is the measured value, b_k is the nonuniformity present at the voxel site, y_k is the value that would be measured in the absence of noise or bias, and η_k is additive spatially white Gaussian noise. The corresponding probability density function is

$$p(x_k|b_k, y_k) = g(x_k; b_k y_k, \sigma) \qquad k \in \Omega, \tag{11}$$

where $g(x; \mu, \sigma)$ is a Gaussian density function with mean μ and variance σ^2. This represents a measurement process with a nonstationary mean, governed by the tissue present within a particular voxel in the image and the bias present at that site. We assume that b_k varies slowly across the volume.

The value for y_k is determined by the type of tissue present in the kth voxel. We assume that only CSF, GM, and WM remain in the image after skull stripping. Due to the finite resolution of the scanning system, we further assume that the tissue combination in each voxel is restricted to the set $\Gamma = \{CSF, GM, WM, CSF/GM, GM/WM, CSF/other\}$, where the last three classes represent partial volume voxels. We assume that CSF is sufficiently separated from WM that voxels will not contain both of these materials. In practice, some additional tissues such as blood vessels and dura may also remain in the volume. These typically comprise a very small percentage of the brain volume and will not drastically affect estimation of the bias field. The use of a CSF/other class accounts for boundary voxels that occur between the sulcal CSF and the surrounding tissues, which we assume to be darker than CSF if they remain in the volume after skull stripping. Brain Surface Extractor should remove most nonbrain tissue with intensities brighter than CSF.

For voxels composed of a single tissue type (pure voxels), the characteristic intensity of the pure type is assumed to be $y_k = \mu_{\lambda_k}$, where $\lambda_k \in \Gamma$ is a label describing the tissue types present in the kth voxel. Thus our measurement model for pure tissues (i.e., $\lambda_k \in \{CSF, GM, WM\}$) is

$$p(x_k|b_k, \lambda_k) = g(x_k; b_k \mu_{\lambda_k}, \sigma). \tag{12}$$

The pure tissue voxels have nonstationary mean values that vary multiplicatively from a global mean value, $y_k = \mu_{\gamma_k}$, according to the bias b_k.

For mixed tissue types, we assume that the ideal intensity y_k at a particular voxel is a linear combination of the ideal intensities of two pure types,

$$y_k = \alpha_k \mu_A + (1 - \alpha_k)\mu_B, \tag{13}$$

where y_k is composed of tissue types A and B, which have characteristic intensities μ_A and μ_B, respectively, and $\alpha_k \in (0, 1)$ describes the fractional content of the kth voxel. The measurement model for each mixed tissue voxel, where $\lambda_k \in \{CSF/GM, GM/WM, CSF/other\}$, is then

$$p(x_k|b_k, \alpha_k, \lambda_k) = g(x_k; b_k(\alpha_k \mu_A + (1 - \alpha_k)\mu_B), \sigma), \tag{14}$$

where μ_A and μ_B are the characteristic values of the two tissue types that constitute the voxel.

Instead of computing α at each mixed-voxel location, we follow the method of Santago and Gage (1993) and assume α to be a random variable uniformly distributed between 0 and 1 since the boundary between tissue types will occur arbitrarily within a mixed voxel. We then marginalize α_k from Eq. (14) to obtain our measurement model for mixed types,

$$p(x_k|b_k, \lambda_k) = \int_0^1 g(x; b_k(\alpha \mu_A + (1 - \alpha)\mu_B), \sigma) d\alpha. \tag{15}$$

Because the bias changes very slowly throughout the image, we approximate it as being constant within a

small region $\Omega_m \subset \Omega$ of the image. By making this assumption, we can represent the measurement model for voxels in a particular region as

$$p(x_k|\theta_m) = \sum_{\gamma \in \Gamma} p(\gamma)p(x_k|b, \gamma) \qquad \forall\, k \in \Omega_m, \quad (16)$$

where $p(\gamma)$ is the probability of each tissue type occurring within the region and θ is a collection of parameters describing our model for the region Ω_m. Specifically,

$$\theta_m = [bp(CSF)\ p(GM)\ p(WM)\ p(CO)\ p(CG)\ p(GW)]^T, \quad (17)$$

where b is the bias for the region, and $p(\gamma)$, $\gamma \in \Gamma$, are the relative probabilities for each type of tissue occurring in the region. The normalized intensity histogram of the region is then described by Eq. (16).

There are several parameters that must be specified for Eq. (16). The first of these are global properties: the mean tissue values for the pure classes, μ_{CSF}, μ_{GM}, and μ_{WM}, and the noise variance σ^2. The second set of parameters, θ_m, is governed by the region of the image being analyzed: the bias within the region, b, and the tissue class probabilities, $p(\gamma)$. By fitting our model to the data of a given region we arrive at estimates of its tissue composition and nonuniformity relative to the rest of the image.

Initialization

We determine the global tissue mean values and noise variance by analyzing the intensity histogram of the stripped brain image. We bracket the histogram into three regions based on quantiles, chosen empirically to suit histograms from T1-weighted MRI. The region of the histogram between the 0.01 quantile and the 0.09 quantile is assumed to contain most of the CSF, the region between the 0.25 and the 0.55 quantiles is assumed to contain most of the GM, and the region between the 0.65 and the 0.95 quantiles is assumed to contain most of the WM. Centroids are computed for each region; these provide the estimates for the ideal values for CSF, GM, and WM. These values were chosen empirically based on examination of histograms from T1-weighted volumes. In preliminary use of our bias correction algorithm, this method of tissue mean estimation was more robust than K-means clustering (Duda and Hart, 1973), which proved unsuccessful for initialization on volumes with high levels of nonuniformity.

We assume any intensity greater than the WM mean, μ_{WM}, is due to noise and compute the noise variance as the sample variance relative to μ_{WM} of the intensity values that are greater than μ_{WM}. To make this estimate more robust, we perform this calculation on the lower 0.99 quantile of the histogram.

Computing Local Bias Estimates

We use Eq. (16) to estimate the bias on a lattice of M points spaced uniformly throughout the image. We denote this spacing as d_s, or the sampling distance. We compute an intensity histogram $h_m[n]$ at each lattice point $m \in \{1, 2, \ldots, M\}$ on a cubic region centered about that point. The size of this cube is d_h, or the histogram radius. As d_h becomes larger, we obtain more sample points for the region histogram, but at the expense of implicitly assuming a smoother bias field. We match the normalized histogram $h_m[n]$ to a sampled version of Eq. (16), $p(n|\theta_m)$, where $n \in \{0, 1, \ldots, N - 1\}$ is the set of allowed image intensities, by minimizing the cost function

$$e(\theta_m) = \sum_{n=0}^{N-1} (h_m[n] - p(n|\theta_m))^2, \quad (18)$$

which is the squared difference between our parametric model and the actual data. This cost function is used rather than maximum likelihood or mutual information criteria, as we achieved better results using this metric. We expect that this is due to the nonconvexity of the cost functions.

We minimize Eq. (18) using steepest descent with an Armijo line search to determine step size (Bertsekas, 1995). The seven parameters for each block are computed while fitting the model to the histogram, though only the bias parameter b will be used in later stages of the bias correction algorithm. Neither the model $p(x_k/\theta_m)$ nor its gradient with respect to θ can be computed in closed form due to the integration over the mixture parameter α for mixed voxels. Fortunately, both can be rewritten in terms of the Gaussian error function and computed using standard numerical methods.

Outlier Rejection

Small errors in the local bias estimates are smoothed by a spline, assuming that the control point spacing, d_c, is significantly larger than the bias estimate spacing, d_s, or that the bending energy coefficient is significantly large. However, the spline may still be significantly distorted by large errors in the local estimates. In earlier versions of the method we observed problems in regions of MRI with large populations of partial volume voxels, such as the cerebellum. In these locations, our fitting procedure would often obtain poor estimates for $p(\gamma)$ and b, attributing the GM/WM mixture voxels to either GM or WM. This would result in a bias parameter that shifted the dominant peak in the histogram to the GM or WM mean. To make our method more robust to poor local nonuniformity esti-

mates such as these, we apply a four-step outlier rejection strategy.

The first requirement is that the block being examined contains enough voxels to form a valid histogram. For this, we check that a significant fraction of voxels are brain (one-eighth of the number of voxels in the block). Second, the bias estimates must be within a predetermined range $[b_{min}, b_{max}]$. In our implementation, this range is specified by the user, with a default of [0.8, 1.2].

The third step examines the value of the cost function Eq. (18) for each remaining estimate. To devise an outlier rejection strategy, we applied known nonuniformity fields to phantom data and real data that were observed to have small nonuniformity artifacts. We examined the distribution of Eq. (18), which we observed to be log-normal. Thus, if the log of Eq. (18) for a particular estimate is more than 1 standard deviation above the mean value for the log of the cost function throughout the volume, then we deem that estimate to be an outlier.

The final step for outlier rejection relies on the smoothly varying property of the bias field. At each estimate site, we compute a roughness function as the mean squared difference between the estimate and its 6 nearest neighbors that have survived the previous outlier rejection steps. We devised this rejection strategy also based on the study of the distribution of errors for phantom and real data with known applied fields. As in the third step, we observed these errors to follow a log-normal distribution. We use the log of the roughness function so that we can model the errors as normally distributed. We thus establish a maximum allowed log-roughness term as the mean plus 1 standard deviation of the log-roughness values for the estimate points. If the log-roughness value at an estimate site is greater than the threshold, the estimate is deemed an outlier and replaced by the mean value of its neighbors. We iterate this procedure which typically converges after a few steps.

Estimation of the Complete Nonuniformity Field

Because the bias field varies slowly throughout the image, we may sample its values at a coarse scale relative to the dimensions of the volume. We use a tricubic B-spline to smooth and interpolate our sampled bias values throughout the volume. The tricubic B-spline provides a continuous function in R^3 with continuous second derivatives. The shape of the spline is governed by a set of control vertices that are spaced uniformly throughout the image. Increasing the distance between the control vertices decreases the amount of spatial variation allowed in the spline. If the input volume has anisotropically sampled voxels, we use an appropriately scaled distance along each axis. The control vertices are spaced at an interval d_c throughout the volume. Each control vertex governs the amplitude of a basis function that has compact support. These basis functions are summed to form a 3-D volume of real numbers. Explicit definitions of splines and spline basis functions are provided in Bartels *et al.* (1987).

We represent the spline on a discrete volume in compact matrix notation as

$$\boldsymbol{s} = \boldsymbol{Q}\boldsymbol{v}, \tag{19}$$

where $\boldsymbol{Q}$ is a sparse matrix whose rows are samples of the tricubic B-spline basis functions, $\boldsymbol{v}$ is the vector of control points values, and $\boldsymbol{s}$ is a vector of spline values for the brain volume.

We collect the robust bias estimates produced in the previous section into a vector $\boldsymbol{b}$ and define the error metric

$$e_s(\boldsymbol{v}) = \frac{1}{N_e}\|\boldsymbol{b} - \boldsymbol{Q}\boldsymbol{v}\|^2 + \lambda_s \frac{1}{N_c} E_{bending}(\boldsymbol{v}), \tag{20}$$

where $E_{bending}(\boldsymbol{v})$ is the bending energy of the spline produced by the set of control vertices $\boldsymbol{v}$ and N_e and N_c are the numbers of estimate points and spline control vertices, respectively. We define the bending energy functional as

$$E_{bending}(\boldsymbol{v}) = \int_D \sum_{|\alpha|=2} \left| \frac{\partial^2 S(\boldsymbol{v})}{\partial \boldsymbol{p}^{\alpha}} \right|^2 d\boldsymbol{p}, \tag{21}$$

which is the sum of the squared second-order partial derivatives integrated over D, the domain of the volume. This integral reduces to a simple matrix product,

$$E_{bending}(\boldsymbol{v}) = \boldsymbol{v}^T \boldsymbol{M} \boldsymbol{v}, \tag{22}$$

where $\boldsymbol{M}$ is an $N_c \times N_c$ matrix. The entries of $\boldsymbol{M}$ are energy coefficients determined from the spline basis functions. Since the spline bases have compact support, $\boldsymbol{M}$ will be sparse, making computation of the bending energy and its derivative very tractable. The final form for our error function is thus

$$e_s(\boldsymbol{v}) = \frac{1}{N_e}\|\boldsymbol{b} - \boldsymbol{Q}\boldsymbol{v}\|^2 + \frac{\lambda_s}{N_c} \boldsymbol{v}^T \boldsymbol{M} \boldsymbol{v}. \tag{23}$$

We use the conjugate gradient method (Bertsekas, 1995) to fit the spline to our data $\boldsymbol{b}$ by adjusting the control vertices $\boldsymbol{v}$. We then compute the spline values for all points in the brain; this is our estimate of the image nonuniformity. We divide the initial stripped brain image by the multiplicative bias field estimate at each point in the brain to achieve our intensity nor-

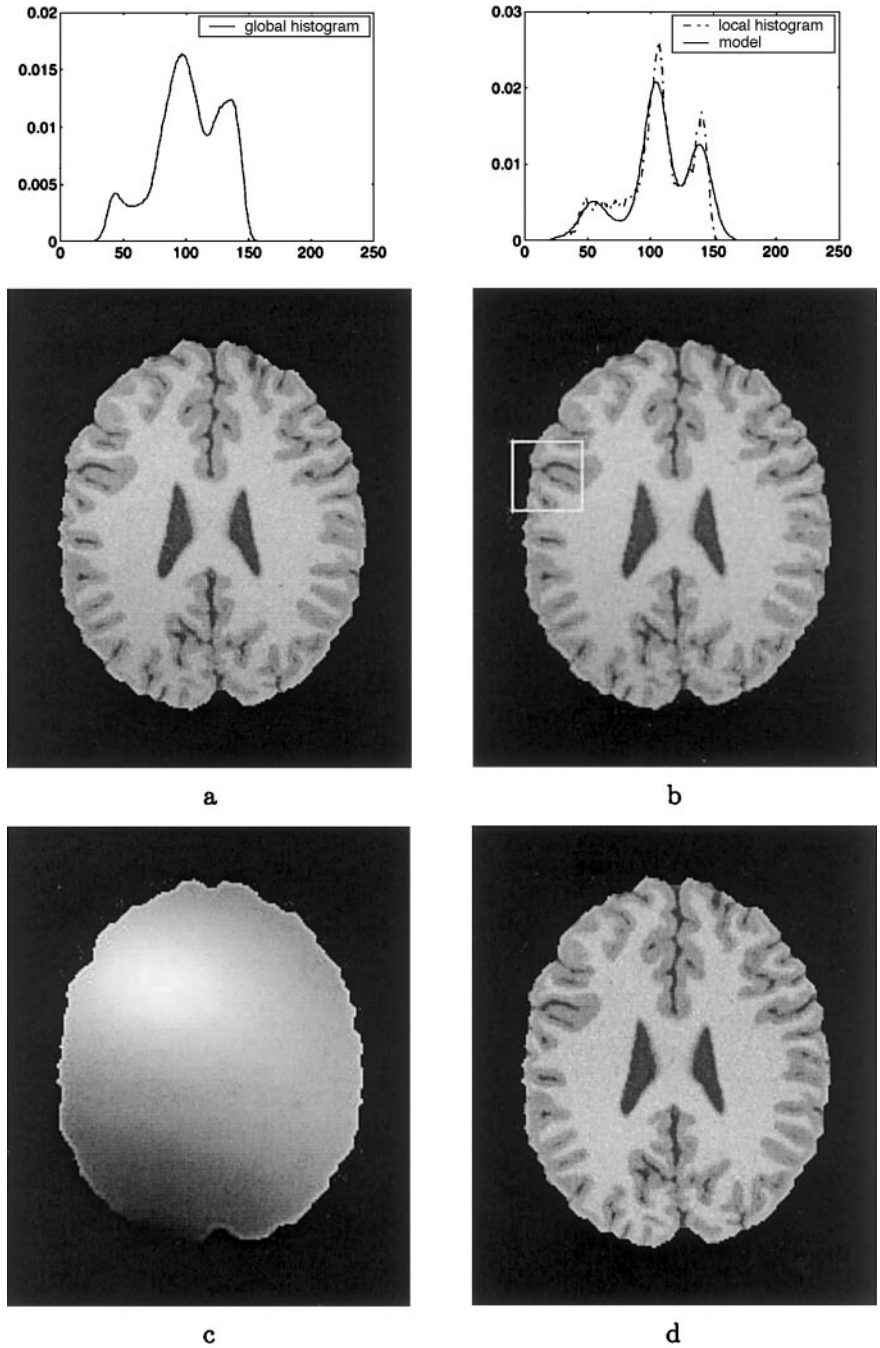

FIG. 2. Bias field correction stages. (a) The histogram of the skull stripped volume is analyzed to estimate the noise variance and mean values for GM, WM, and CSF. (b) A parametric measurement model is fit to the histograms of local blocks to estimate the bias on a lattice of points throughout the image. (c) These points are smoothed and interpolated using a tricubic B-spline with a stiffness constraint. (d) The skull-stripped image is divided by the B-spline values to form the intensity-normalized image.

malized image. Figure 2 illustrates the bias field correction procedure on the BrainWeb phantom.

There are four spatial parameters that must be specified for our algorithm. These are the histogram radius, d_h; the bias estimate sampling distance, d_s; the spline control point spacing, d_c; and the stiffness multiplier, λ_s. Also, the user must specify the range of the accepted bias estimates, $[b_{min}, b_{max}]$. Because the nature of the intensity nonuniformities is not known in advance, these parameters may be adjusted to suite the particular image or set of images being processed. On machines and acquisition sequences producing fields that

change more quickly, smaller distances and a small stiffness penalty are used, as they allow for more rapid variations in the bias field.

Partial Volume Classification

The tissue classification problem is greatly simplified by compensation for the nonuniformity in the MRI. However, noise is still present in the system and not all of the nonuniformity may have been removed. In this section we describe our technique for labeling the tissue content of each voxel in the image.

Image Model

We use the same image measurement model as before under the assumption that no bias is present in the image. Setting $b_k = 1$ everywhere, we get simply

$$x_k = y_k + \eta_k \qquad k \in \Omega, \tag{24}$$

where again x_k is the measurement at the kth voxel, y_k is the true value of the voxel, and η_k is the Gaussian white noise. Our model allows for six tissue types to be present in the volume: CSF, GM, WM, CSF/GM (CG), GM/WM (GW), and CSF/other (CO). The three pure types are modeled by

$$p(x_k|\gamma) = g(x_k;\, \mu_\gamma,\, \sigma) \qquad \gamma \in \{CSF,\, GM,\, WM\}, \tag{25}$$

while mixed-tissue measurements are modeled as

$$p(x_k|\gamma) = \int_0^1 g(x_k;\, \alpha\mu_A + (1-\alpha)\mu_B,\, \sigma)\, d\alpha \qquad \gamma \in \{CG,\, GW,\, CO\}. \tag{26}$$

We use a collection of labels, $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_{|\Omega|}\}$, where $\lambda_k \in \Gamma$, to describe the tissue present at each voxel in the brain image. Because the noise is assumed spatially independent we can write the likelihood for all measurements in the volume as

$$p(X|\Lambda) = \prod_{k\in\Omega} p(x_k|\lambda_k). \tag{27}$$

We could combine Eq. (27) with the same tissue-class prior as in Eq. (16) where the probabilities are based on estimates of the fractions of each tissue-class within the brain. However, this does not introduce any information about the local continuity of tissues within the brain. The fact that the brain is made up of contiguous regions of GM, WM, and CSF, with partial-volume regions in between, means that the probability of occurrence of each tissue-type for a particular voxel is influenced by the tissue-types of its neighbors. This local continuity is included in our method through the use of a Markov prior which specifically models local spatial interaction. Ideally we should have used the same model in Eq. (16); however, this is impractical since the parameter estimation in Eq. (18) would have required marginalization of the posterior density over all possible label configurations. Conversely, for the purpose of labeling each voxel, marginalization is not required and the use of the spatial interaction prior clearly improves segmentation results.

We use a simple Potts model,

$$p(\Lambda) = \frac{1}{Z}\exp[-\beta \sum_k \sum_{j\in N_k} \delta(\lambda_k,\, \lambda_j)], \tag{28}$$

where Λ is the labeled image, Z is a scaling constant to ensure that we have a proper density function, and β controls the strength of the prior. N_k is the D18 neighborhood (neighbors share an edge or face) about the kth voxel. The δ terms govern the likelihood of different tissue labels being neighbors, hence we set $\delta(\lambda_k, \lambda_j)$ to -2 if labels k and j are identical, -1 if they have a common tissue type, and 1 if they have no common tissues. These scores are scaled according to the inverse of the distance of voxel k to voxel j. In this way, the model penalizes configurations of voxels that are not likely to occur in the brain, e.g., white matter directly adjacent to CSF, while encouraging more likely types, e.g., white matter next to a partial volume mixture of white matter and gray matter.

Classifier Formulation

We use Bayes' formula to create a MAP classifier, which maximizes

$$p(\Lambda|X) = \frac{p(X|\Lambda)p(\Lambda)}{p(X)}. \tag{29}$$

Note that $p(X)$ is independent of Λ and needs not be specified to maximize Eq. (29). We then find the labeling Λ^* that maximizes this function:

$$\Lambda^* = \arg\max_\Lambda p(\Lambda|X) = \arg\max_\Lambda [\sum_{k\in\Omega} \log p(x_k|\lambda_k) - \beta \sum_k \sum_{j\in N_k} \delta(\lambda_k,\, \lambda_j)]. \tag{30}$$

The estimation task is to determine the optimal Λ in Eq. (30). We use the iterated conditional modes (ICM) algorithm developed by Besag (1986) to search for an optimal labeling. We initialize our ICM iterations using a maximum likelihood (ML) classification. Because of the spatial independence of the noise, the ML clas-

sification is computed simply by computing $p(x_k|\gamma)$ for each tissue class and selecting the label giving the maximum value at each voxel in the image.

Once the initial labeling has been selected via ML classification we begin iteratively updating the labeling. For each voxel in turn we select

$$\lambda_k^{n+1} = \arg\max_{\gamma\in\Gamma} [\log p(x_k|\gamma) - 2\beta \sum_{j\in N_k} \delta(\lambda_j^n, \gamma)], \quad (31)$$

where λ_k^n is the kth label at the nth update. We update each voxel in turn during the iteration and repeat this procedure until no labels change between iterations. An example of the classification method is shown in Fig. 3.

Fractional Content

Following tissue classification, we assign to each voxel a tissue weighting triple for CSF, GM, and WM based on its tissue label and its intensity in the MRI. Voxels assigned pure tissue labels are set to 1 for their corresponding tissue type and 0 for the other types. Voxels assigned mixed-tissue labels are assigned according to their MRI intensity, x_k, and the means of the two pure tissue types of which they are composed, μ_A and μ_B, according to the formula

$$f_A = U\left(\frac{\mu_B - x_k}{\mu_B - \mu_A}\right), \quad (32)$$

where $U(\cdot)$ is a soft limiter restricting the range of the fractional content to [0, 1]. In some cases, our classifier will overestimate the mean value for CSF; this will result in a significant portion of CSF voxels being assigned as partial volume CSF/other. For this reason, we assign to each CSF/other voxel the fractional content corresponding to a pure CSF voxel.

Parameter Initialization

The model presented in this section has parameters for the tissue class means and noise variance, as well as for the tissue prior strength. We estimate the tissue class means, μ_{CSF}, μ_{GM}, and μ_{WM}, and the noise variance σ^2, from the histogram of the nonuniformity-corrected brain region. A K-means clustering algorithm provides an initial set of tissue means. A nearest neighbor rule is used to determine class labelings for each intensity. The modes of the sets of GM and WM are then computed to serve as estimates of the mean values for these tissues. In our experience, CSF peaks do not reliably appear in histograms; we estimate the CSF mean by taking the mean value of the K-means clustered voxels grouped as CSF. The noise variance is estimated in the same manner as in the initialization described under Image Nonuniformity Compensation.

The one parameter we leave for the user to vary is the β term that controls the strength of the prior. We have found that on most data sets using $\beta = 0.05$ works well. However, as signal to noise decreases, the quality of the spatial information degrades and the prior has a tendency to oversmooth regions. With such data β can be decreased in order to obtain a better classification.

RESULTS

Implementation

We implemented each of the three methods using the C++ computer programming language. The skull stripping method was implemented with both a command line and an X-motif-based graphical user interface called Brain Surface Extractor (BSE). BSE is publicly available at http://neuroimage.usc.edu/bse/. We call our tissue classification method the Partial Volume Classifier (PVC) and our intensity normalization method the Bias Field Corrector (BFC).

Validation on Real and Phantom Data

Assessing the performance of skull stripping, image nonuniformity correction, and tissue classification using real data is difficult because the ground truth is not known. For this reason we tested our skull-stripping and tissue segmentation methods on data from the Internet Brain Segmentation Repository. The 20 normal MRI brain data sets and their manual segmentations were provided by the Center for Morphometric Analysis (CMA) at Massachusetts General Hospital and are available at http://neuro-www.mgh.harvard.edu/cma/ibsr/. The data sets provided by CMA were selected because they have been used in published studies and have various levels of artifacts. A few of these volumes have low contrast and relatively large intensity gradients, and the performance of the tested algorithms is poor on these. The volumes have slice dimensions of 256×256, with resolution of 1×1 mm^2. Interslice distance is 3 mm, with the number of slices for each volume between 60 and 65. CMA provides expert-labeled volumes of their data, including both brain extraction masks and tissue labelings of each brain into CSF, GM, and WM. Additionally, reference values for several classification techniques are provided based on work done by Rajapakse and Kruggel (1998).

Testing nonuniformity on real data is particularly problematic, as humans cannot always see the shading effects. For this reason, we tested our methods on the BrainWeb phantom (available from http://www.bic.mni.mcgill.ca/brainweb/) produced by the McConnell Brain Imaging Centre at the Montreal Neurological Institute (MNI) (Collins *et al.*, 1998). The BrainWeb phantom was generated from a ground truth image.

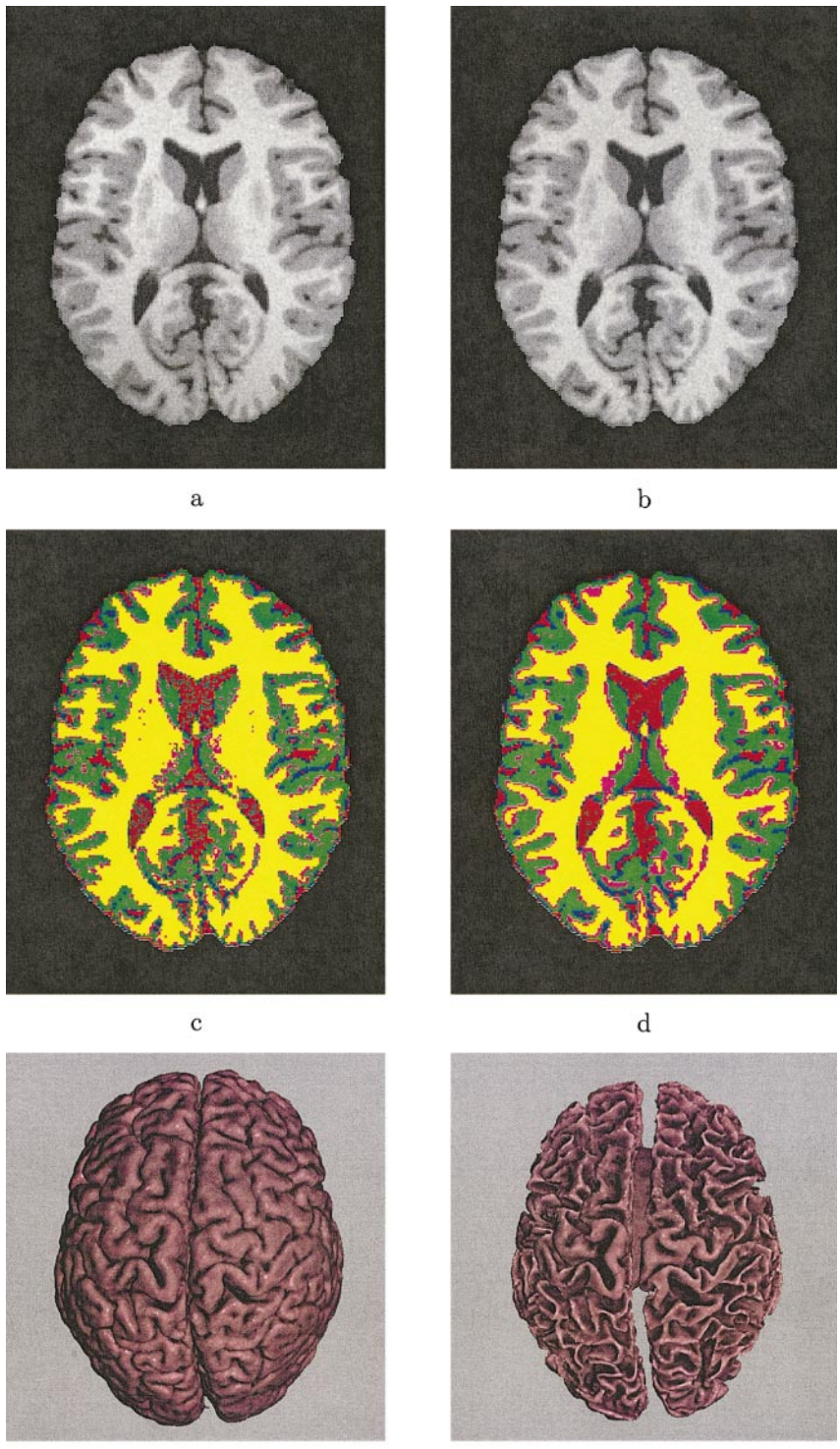

FIG. 3. Partial volume labeling stages. (a) A slice from the skull-stripped brain volume. (b) The same slice after nonuniformity correction. (c) A maximum likelihood labeling initializes the iterated conditional modes algorithm. (d) The spatial prior used in the classifier encourages a thin band of partial volume voxels to occur between pure tissue regions, while discouraging speckle noise. (e and f) This labeling can be used to identify the (e) GM/CSF and (f) GM/WM cortical boundaries.

This image was generated from 27 scans of a single subject, which were intensity normalized using N3 and averaged into an isotropic image space. The composite image was then automatically classified with refinement by an expert. The resulting fuzzy classification was used to reconstruct a ground truth image repre-

senting the intensities expected in the absence of noise or intensity nonuniformity. The BrainWeb phantom provides several simulated MRI acquisitions of this phantom, including RF nonuniformities of 0, 20, or 40% and noise levels of 0, 3, 5, 7, or 9%. Each level of nonuniformity represents a different scaling of the same bias field, hence this validation tests our bias correction method on a single shape. We tested each of our methods on each combination of these parameters for the phantom.

Brain Surface Extractor

BSE has been in public release for a few years and has performed well on a large number of images. For example, over 300 brains have been identified using BSE at UCLA's Laboratory of NeuroImaging. We have observed BSE to work well on various parts of the brain, including the cerebellum, brain stem, and top of the head. In some cases, small notches in the surface may result from poor edge localization by the Marr–Hildreth edge detector. Dura may also be included in the brain mask if the edge detector cannot find a clear boundary between it and the brain. Often these problems can be corrected by adjusting the parameters of the anisotropic diffusion filter or the edge detector. In cases in which the parameters cannot be tuned to eliminate such problems, some manual editing of the brain mask may be necessary.

IBSR data. We studied the performance of BSE on the data from CMA's IBSR. CMA provides brain masks for 20 volumes. We computed the Jaccard similarity between the brain mask BSE identifies and the brain masks provided by CMA. This metric, also termed the Tanimoto coefficient, measures the similarity of two sets as the ratio of the size of their intersection divided by the size of their union,

$$J(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|}, \tag{33}$$

and ranges from 0 for sets that have no common elements to 1 for sets that are identical.

Another metric often used for comparing set similarity is the Dice coefficient (Dice, 1945), defined as

$$\kappa(S_1, S_2) = \frac{2|S_1 \cap S_2|}{|S_1| + |S_2|}. \tag{34}$$

This metric appears frequently in the literature (Collins *et al.*, 1999; Van Leemput *et al.*, 1999b). Zijdenbos *et al.* (1994) showed that this is a special case of the κ index, appropriate for assessment of image segmentation. It can be rewritten as

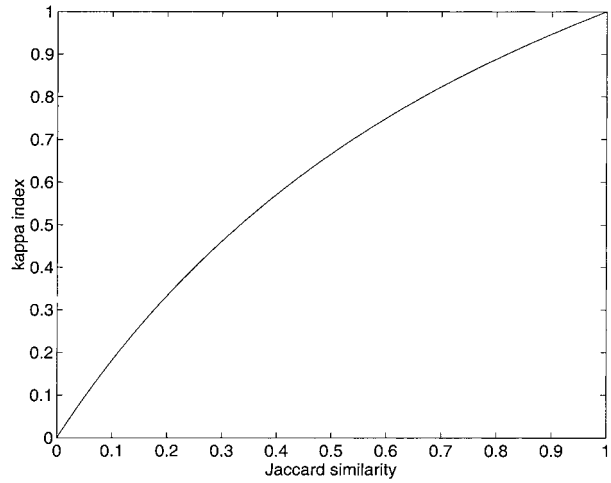

FIG. 4. Comparison of the Jaccard similarity metric and the κ index.

$$\kappa(S_1, S_2) = \frac{|S_1 \cap S_2|}{\frac{1}{2}(|S_1| + |S_2|)} = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2| - \frac{1}{2}(|S_1 \backslash S_2| + |S_2 \backslash S_1|)}, \tag{35}$$

which shows that this metric is always larger than the Jaccard metric, except at 0 and 1 where they are equal. This case of the κ index is related to the Jaccard metric (Shattuck, 2000) by the function

$$\kappa = \frac{2J}{J+1}. \tag{36}$$

Figure 4 shows a plot of this relationship. Both metrics agree that 0 means the two sets are completely dissimilar and that 1 means the two sets are identical. The metrics are consistent for the purpose of comparison, as an increase in the Jaccard metric also means an increase in the κ index. For reference, a Jaccard score of 0.900 corresponds to a κ index of 0.947, while a Jaccard score of 0.800 corresponds to a κ index of 0.889. We also computed κ indices to assess our method's performance.

While it is important that our mask closely matches the expert extraction, we also need to be concerned with what type of tissue is being excluded or included in the brain mask. Excluding all of the sulcal CSF is not a problem for identifying GM and WM structures. However, we do want all GM and WM voxels to be included in the brain mask. We also want nothing but CSF, GM, or WM to be included. For this reason, we computed a false positive rate as the number of voxels in the included mask that are none of these three types, divided by the number of CSF, GM, and WM

voxels in the expert labeling. We compute as a false negative rate the number of GM or WM voxels excluded from the brain mask, divided by the number of GM and WM voxels in the expert labeling.

It should be noted that the severe anisotropy of the voxels in the CMA volumes ($1 \times 1 \times 3$ mm^3) is much higher than BSE expects. Nevertheless, BSE was able to perform a reasonable brain extraction on most of these volumes. The average Jaccard similarity between the reference skull-stripped brains and BSE's skull-stripped brains was 0.910 ± 0.013; the average κ index was 0.953. The average false negative and false positive rates for BSE on these data were 2.0 and 8.0%, respectively. This is in keeping with our goal for BSE, which is to be more conservative about what tissues it removes. Also, since the labeling of the data provided by CMA does not label all CSF in the volume, some of the false positive rate may be due, in fact, to unlabeled CSF. However, the fact that only 2.0% of the GM or WM is excluded from the brain mask means that nearly all of the brain is present for later processing.

BrainWeb phantom. We applied BSE to the various brain phantoms obtained from the Montreal Neurological Institute. We adjusted the strength of the diffusion filter for each brain to ensure that the whole brain was included. In practice, we were able to keep the diffusion parameter constant with nonuniformity level and needed only to vary it as the noise increased. The other parameters remained the same: three iterations of anisotropic diffusion filtering with an image-specific diffusion parameter and an edge detection kernel of $\sigma = 0.6$. Also, because the voxels are higher resolution than typical clinical volumes we increased the size of the erosion structuring element to 2 for the initial erosion step.

We computed the false positive and false negative rates for each volume; results are tabulated in Table 2. On average, BSE excluded only 1.29% of the GM and WM of the phantom, while containing only 1.18% extraneous tissue. These rates are more in keeping with our visual assessment of BSE's performance on high-resolution data. It should also be noted that we can decrease the false negative rate at the expense of the false positive rate by dilating the brain mask to form a more generous set.

Since no reference "stripped" brain is provided by MNI, we compared the similarity of each brain extraction with the extraction that had the best FP and FN rates, which was the phantom with 3% noise and 40% bias. These metrics are also shown in Table 2. Even as the noise and bias increased, the similarity of the final brain masks averaged 0.971 ± 0.017 ($\kappa = 0.985 \pm 0.034$), emphasizing BSE's ability to cope with varying artifact levels.

Repeat studies—human subject data. We applied BSE to three repeat T1 images of a single subject. The first volume was stripped using BSE (three iterations, diffusion constant of 25, $\sigma = 0.6$). The second and third volumes were aligned to the first using the Automated Image Registration program provided by Woods (Woods *et al.*, 1998). The aligned volumes were then skull stripped using BSE (three iterations, diffusion constant of 15, $\sigma = 0.6$). The need for a reduced diffusion constant may be because of the filtering effect that occurs during alignment. We observed an average similarity of 0.963 ($\kappa = 0.981$) between the repeated extractions.

TABLE 2

BSE Results on BrainWeb Phantom

Noise	RF	False positive rate	False negative rate	Similarity	κ index
0%	0%	2.28%	0.44%	0.969	0.984
3%	0%	0.98%	0.73%	0.985	0.993
	20%	1.47%	0.65%	0.976	0.988
	40%	0.98%	0.69%	1.000	1.000
5%	0%	1.09%	1.24%	0.975	0.987
	20%	1.17%	0.86%	0.978	0.989
	40%	0.99%	1.44%	0.974	0.987
7%	0%	1.02%	1.04%	0.977	0.988
	20%	1.10%	0.93%	0.977	0.989
	40%	1.56%	0.69%	0.972	0.986
9%	0%	1.07%	1.03%	0.975	0.987
	20%	0.55%	6.23%	0.917	0.957
	40%	1.13%	0.77%	0.976	0.988
Average		**1.18%**	**1.29%**	**0.973**	**0.986**

Note. False positive rates (nonbrain tissue included in mask), false negative rates (GM or WM excluded from mask), and agreement (both Jaccard and κ index) to mask found for 3% noise, 40% RF phantom. RF, or radiofrequency, refers to the level of intensity nonuniformity applied the image.

Bias Field Corrector

Since no nonuniformity ground truth images exist for the IBSR data set, we validated our bias correction method on the MNI phantom. However, since bias correction is a preprocessor to tissue classification, BFC's performance is also reflected in the performance of the tissue classification routine.

We examined the performance of BFC compared to N3 at correcting biased and noisy data from the BrainWeb phantom database. We assessed the performance of each method by measuring how closely its intensity correction images matched the ground truth image. The database did not provide the gain fields used to simulate the nonuniformity in the phantom images, thus we were unable to assess the similarity of the recovered fields with the ground truth field.

To examine the bias correction results consistently, we generated a single mask using BSE to compare each image. The stripped brains were then processed by BFC using a single set of parameters ($d_h = 8$ mm, $d_s =$

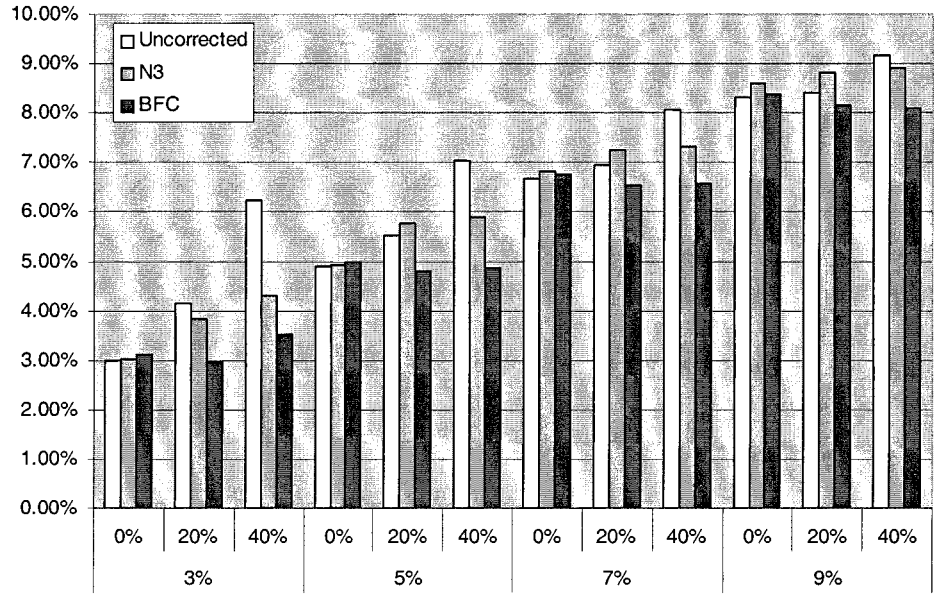

FIG. 5. Normalized RMS error metrics as a percentage of WM ground truth intensity calculated between the ground truth image and the uncorrected phantom image, the N3-corrected image, and the BFC-corrected image.

16 mm, $d_c = 64$ mm, $\lambda_s = 0.1$) for all volumes. We also processed each stripped brain volume with N3 using its default parameters.

Comparison of intensity-corrected images with a ground truth requires that we allow for affine translations of the image intensities because global scalings or translations of the image intensity will not affect intensity-based classification techniques. For this reason we compute an affine transformation that minimizes the RMS difference between the ground truth image and intensity corrected images being compared. This Procrustean metric allows for all possible scales and translations of a subject image's intensity and is simply

$$e_{RMS}(\boldsymbol{X}, \hat{\boldsymbol{X}}) = \min_{a,b} \sqrt{\frac{1}{|\Omega|} \sum_{k \in \Omega} (x_k - (a\widehat{x_k} + b))^2}, \quad (37)$$

where $\boldsymbol{X}$ and $\hat{\boldsymbol{X}}$ are the ground truth and intensity-corrected images, respectively, and Ω is the set of voxels in the brain or other region of interest. With the use of the affine transformation defined by a and b, the RMS difference between the original image and the transformed image provides a fair comparison of performance between different bias-correction methods. These results are shown in Fig. 5 for both algorithms and each available level of noise and bias. In the cases in which noise was applied with no bias, the normalized RMS difference metric shows that N3 and BFC both left the phantom volumes relatively unchanged. N3 performed slightly better on the 3 and 5% noise phantoms, while BFC performed slightly better on the 7 and 9% cases. In every case with simulated bias fields, the BFC-corrected image was closer to the original than the corresponding N3-corrected image. In most cases the RMS difference was very near to that of the unbiased image with the same level of noise, signifying that we have removed most of the variation attributable to inhomogeneity effects.

Partial Volume Classifier

Internet Brain Segmentation Repository classification. We applied PVC after skull stripping by BSE and intensity normalization by BFC to the 20 normal brain volumes from the IBSR. We performed skull stripping using one of three parameter settings. Bias correction was performed using the same settings for all 20 brains. The selection of the tissue prior weighting for tissue classification was performed manually, using one of three settings. Settings were varied only when the results were clearly unacceptable. No manual editing of the brain volume or labels was performed. It is possible that more tuning of the parameters to the individual data would have improved the results.

To analyze the performance of our method we used the Jaccard similarity metric. CMA provides reference similarity results for several methods described by Rajapakse and Kruggel (1998) that have been tested using the IBSR data sets. The results of each method are averaged over the 20 volumes. Since the IBSR data are classified only into pure tissue types, we converted our tissue classification result into a three-class (CSF, GM, and WM) labeling. We reassigned each mixed voxel by setting its label to the pure tissue having the greatest fraction. We then computed the Jaccard similarity for

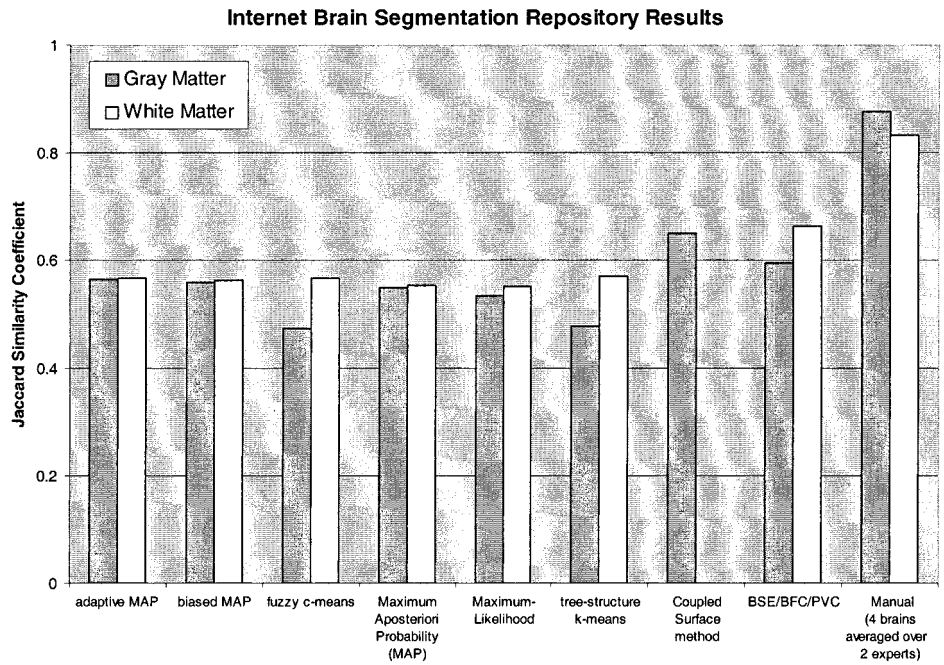

FIG. 6. Comparison of BSE/BFC/PVC to other methods on the IBSR data set (WM values are not provided by Zeng *et al.*, 1999, for the coupled surfaces method).

our GM and WM sets compared to the expert labelings. The average scores over all 20 volumes for GM and WM for the reference methods, our method, and the GM measure provided in Zeng *et al.* (1999) are shown in Fig. 6 (WM metrics are not provided in Zeng *et al.*, 1999). Also shown are reference metrics for interoperator variability, 0.876 for GM and 0.832 for WM, proposed by CMA based on an interoperator variability comparison of two experts segmenting four brains. The brains used in that comparison were not from this data set and do not necessarily represent the similarity that would be achieved by experts segmenting the 20 volumes we examined. The best average performance of the six reference methods is 0.564 for GM and 0.571 for WM. Our method produced average similarity measures of 0.595 ± 0.130 for GM and 0.664 ± 0.107 for WM. These values equate to κ indices of 0.746 ± 0.114 for GM and 0.798 ± 0.089 for WM. The GM similarity measure for the coupled surface result on the whole brain is 0.657, which outperforms our method (Zeng *et al.*, 1999). While none of the methods reach the target performance suggested by CMA, all of these methods should achieve better performance on more recently acquired data, with voxel dimensions that are less anisotropic. Our results on the BrainWeb phantom support this.

BrainWeb phantom classification. We used the BFC software on each of the available BrainWeb phantoms after skull stripping with BSE. We then computed labelings and fractional tissue values for each image using the PVC software. We generated three-class labelings for both the ground truth labeling and our labeled results by reassigning each voxel with the pure tissue label having the largest fraction for that voxel. We then computed the Jaccard similarity coefficient for the GM and WM sets for each volume compared to the ground truth; these results are shown in Fig. 7. The average Jaccard scores for these 12 volumes were 0.808 ± 0.063 for GM and 0.867 ± 0.063 for WM. These numbers are approaching interoperator variability levels suggested by CMA. The average κ indices were $\kappa = 0.893 \pm 0.041$ for GM and $\kappa = 0.928 \pm 0.039$ for WM.

Our method performed well on most volumes from this data set, and Fig. 7 demonstrates the robustness of our method on these data in the presence of varying levels of the applied nonuniformity. Our method produced relatively poor results for one volume, the 9% noise, 0% RF phantom. In this case, examination of the labeled volume revealed that the MAP classifier had oversmoothed the labeling. By reducing the prior to $\beta = 0.01$, we were able to improve the Jaccard metric for this classification to 0.735 for GM and 0.818 for WM, which improved the average scores slightly to 0.816 ± 0.044 and 0.878 ± 0.033 for GM and WM, respectively.

Partial volume fraction results. To assess our method's performance at computing partial volume fractions, we compared our fractional images to the ground truth fractional volumes provided by MNI for their BrainWeb phantom. We computed the RMS error between two fractional volumes X and Y as

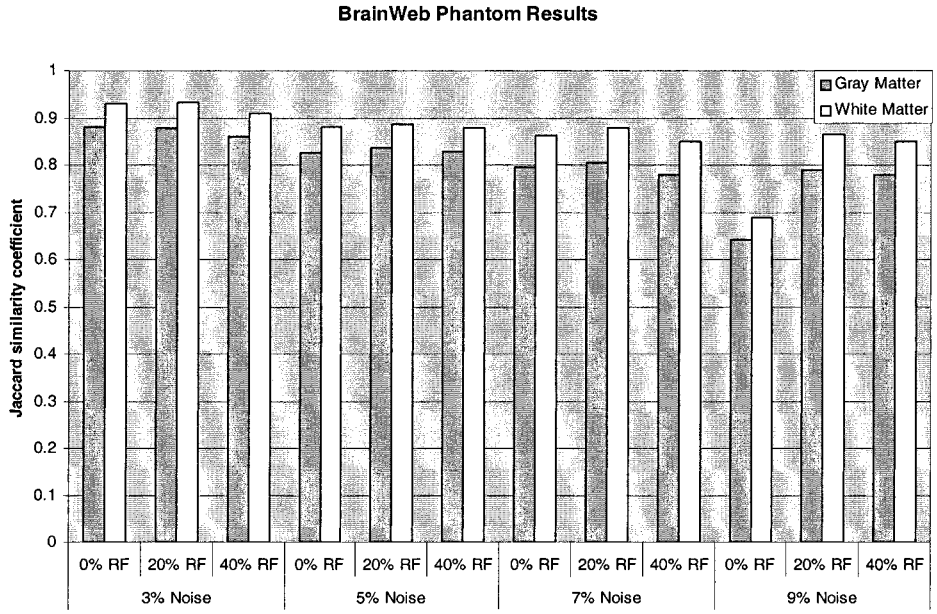

FIG. 7. Similarity metrics for BSE/BFC/PVC processed phantom volumes compared to the ground truth tissue labelings.

$$e_{frac}(X, Y) = \sqrt{\frac{1}{|\Omega|} \sum_{k \in \Omega} |y_k - x_k|^2}, \qquad (38)$$

where Ω is the region of the brain mask. These results are provided in Table 3 for the MAP and ML classified results. In every case, we saw an improvement in this metric comparing the MAP results to the ML results.

Pham and Prince (1999) studied the performance of several tissue classification methods using the BrainWeb phantom. They applied each technique to phantoms with 3% noise and 0, 20, and 40% RF nonuniformity fields, then computed the mean-squared error (MSE) for the GM fractional result of each method on each simulated image. The best result on the 0% RF volume was 0.0194 as classified by the fuzzy C-means (FCM) method. The truncated-multigrid FCM algorithm had the best performance on the 20 and 40% RF volumes, with MSE metrics of 0.0214 and 0.0244, respectively. Our method produced MSE results for these three images of 0.0188, 0.0192, and 0.0252. It should be noted that a direct comparison cannot be made since we use a different brain mask, and therefore a different region for the MSE computation, from what was used by Pham and Prince. Nevertheless, these metrics do suggest that our algorithm would be competitive with the best of these methods. Figure 8 shows images of the fractional volume computed for a real MR volume.

TABLE 3

MNI BrainWeb Phantom Tissue Fraction Results

	Fractional content RMS error											
	3% noise			5% noise			7% noise			9% noise		
	0% RF	20% RF	40%	0%	20%	40%	0%	20%	40%	0%	20%	40%
	Maximum a posteriori classification											
WM	0.088	0.084	0.102	0.129	0.122	0.125	0.165	0.160	0.159	0.359	0.184	0.187
GM	0.137	0.139	0.159	0.178	0.172	0.176	0.215	0.211	0.220	0.384	0.240	0.245
	Maximum likelihood classification											
WM	0.096	0.090	0.105	0.158	0.146	0.142	0.253	0.249	0.214	0.353	0.304	0.293
GM	0.144	0.145	0.163	0.206	0.196	0.194	0.299	0.297	0.277	0.420	0.366	0.362

Note. RMS error of the fractional voxel content computed by BSE/BFC/PVC for the MNI BrainWeb phantom.

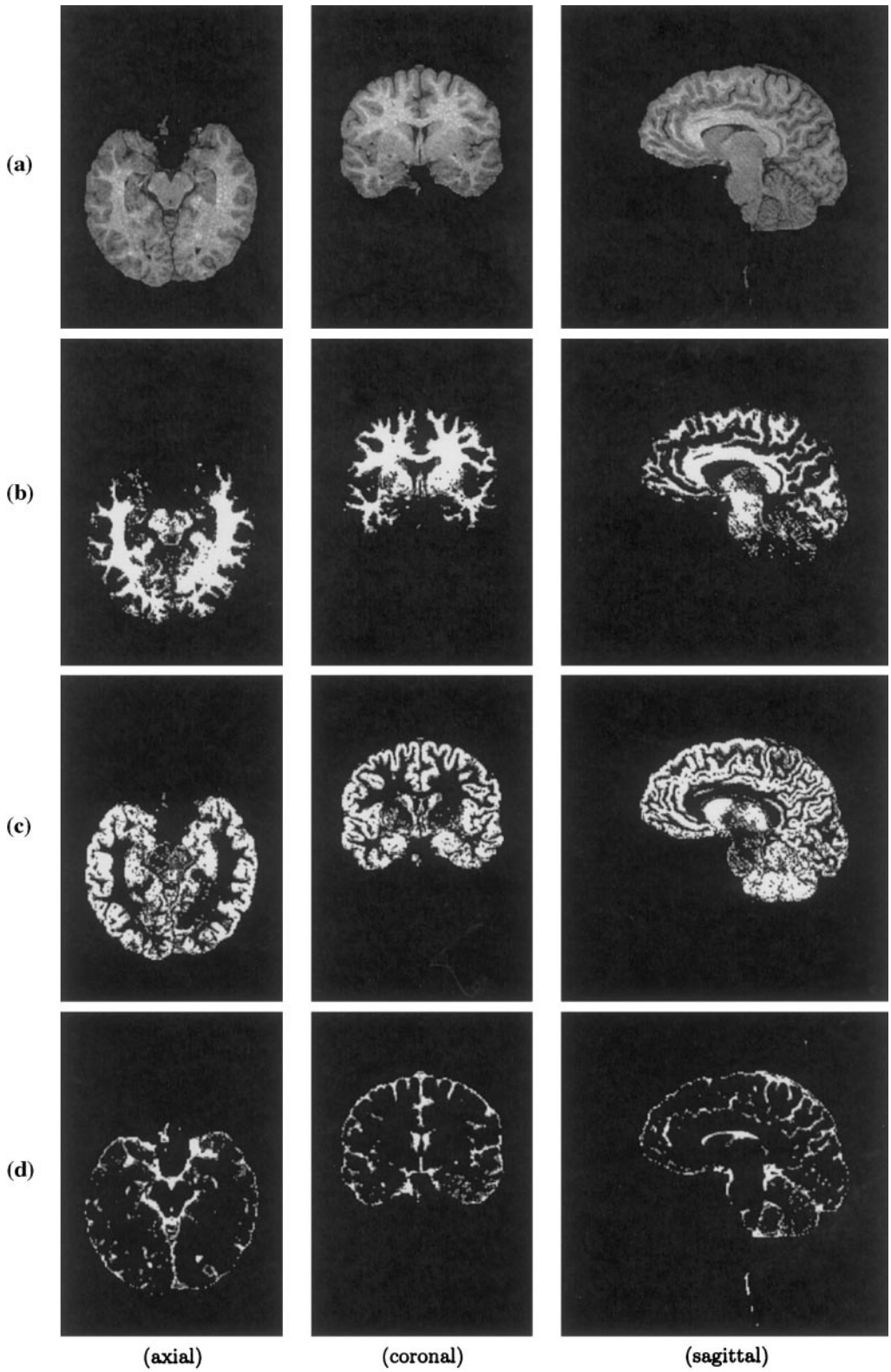

FIG. 8. Partial volume fraction images. (a) Nonuniformity-corrected skull-stripped MRI, (b) WM volume, (c) GM volume, (d) CSF volume.

Computation Time

Using the similarity measures on real and phantom data, we have shown the methods described in this paper to be accurate. Additionally, our methods have been implemented in a computationally efficient manner. On an Intel 933-MHz Pentium III Xeon processor with 256 MB of RAM, the entire process of skull stripping, bias correcting, and classifying a brain typically takes less than 5 min. For the IBSR collection of data, the average computation times for each brain were 5.1 s for BSE, 51.8 s for BFC, and 3.4 s for PVC. The BrainWeb phantom represents a much more typical volume with image dimensions of $181 \times 217 \times 181$. On average, BSE took 10 s for skull stripping, BFC took 2 min 20 s, and PVC took 25 s, for a total processing time of less than 3 min per volume on these data.

CONCLUSION

We have presented a three-stage sequence of techniques for identifying and classifying the brain tissues within a T1-weighted MRI of the human head. The method makes use of low-level methods and provides information at the voxel level regarding the tissue content of the image. We validated our method on phantom and real human data and demonstrated that it outperformed several published methods.

ACKNOWLEDGMENT

This work was supported in part by the National Institute of Mental Health Grants RO1-MH53213 and P50-MH57180.

REFERENCES

Baillet, S., Mosher, J., Leahy, R., and Shattuck, D. 1999. Brainstorm: A Matlab toolbox for the processing of MEG and EEG signals. *NeuroImage* **9:**S246.

Bajcsy, R., and Kovacic, S. 1989. Multi-resolution elastic matching. *Comput. Vision Graph. Image Processing* **46:**1–21.

Bajcsy, R., Lieberson, R., and Reivich, M. 1983. A computerized system for elastic matching of deformed radiographic images to idealized atlas images. *J. Comput. Assisted Tomogr.* **7:**618–625.

Bartels, R., Beatty, J., and Barsky, B. 1987. *An Introduction to Splines for Use in Computer Graphics and Geometric Modeling.* Kaufmann, Los Altos, CA.

Bertsekas, D. 1995. *Nonlinear Programming.* Athena Scientific, Boston.

Besag, J. 1986. On the statistical analysis of dirty pictures. *J. R. Stat. Soc. B* **48:**259–302.

Bomans, M., Hohne, K., Tiene, U., and Riemer, M. 1990. 3-D segmentation of MR images of the head for 3-D display. *IEEE Trans. Med. Imag.* **9:**177–183.

Bonar, D. C., Schaper, K. A., Anderson, J. R., Rottenberg, D. A., and Strother, S. C. 1993. Graphical analysis of MR feature space for measurement of CSF, gray-matter, and white-matter volumes. *J. Comput. Assisted Tomogr.* **17:**461–470.

Brinkmann, B. H., Manduca, A., and Robb, R. A. 1998. Optimized homomorphic unsharp masking for MR grayscale inhomogeneity correction. *IEEE Trans. Med. Imag.* **17:**161–171.

Brummer, M. E., Merseau, R. M., Eisner, R. L., and Lewine, R. R. J. 1993. Automatic detection of brain contours in MRI data sets. *IEEE Trans. Med. Imag.* **12:**153–166.

Choi, H. S., Haynor, D. R., and Kim, Y. 1991. Partial volume tissue classification of multi-channel magnetic resonance images—A mixel model. *IEEE Trans. Med. Imag.* **10:**395–407.

Christensen, G. E. 1999. Consistent linear-elastic transformations for image matching. In *Lecture Notes in Computer Science,* Vol. 1613, *Proceedings of the 16th International Conference on Information Processing in Medical Imaging* (A. Kuba, M. Sámal, and A. Todd-Pokropek, Eds.), pp. 224–237. Springer-Verlag, Berlin/Heidelberg.

Christensen, G. E., Rabbitt, R. D., and Miller, M. I. 1996. Deformable templates using large deformation kinematics. *IEEE Trans. Image Processing* **5:**402–417.

Collins, D., Zijdenbos, A., Kollokian, V., Sled, J., Kabani, N., Holmes, C., and Evans, A. 1998. Design and construction of a realistic digital brain phantom. *IEEE Trans. Med. Imag.* **17:**463–468.

Collins, D. L., Zijdenbos, A. P., Baaré, W. F., and Evans, A. C. 1999. ANIMAL+INSECT: Improved cortical structure segmentation. In *Lecture Notes in Computer Science,* Vol. 1613, *Proceedings of the 16th International Conference on Information Processing in Medical Imaging* (A. Kuba, M. Sámal, and A. Todd-Pokropek, Eds.), pp. 210–223. Springer-Verlag, Berlin/Heidelberg.

Dale, A. M., Fischl, B., and Sereno, M. I. 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage* **9:**179–194.

Dale, A. M., and Sereno, M. I. 1993. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *J. Cognit. Neurosci.* **5:**162–176.

Davatzikos, C. 1997. Spatial transformation and registration of brain images using elastically deformable models. *Comput. Vision Image Understand.* **66:**207–222.

Dawant, B. M., Zijdenbos, A. P., and Margolin, R. A. 1993. Correction of intensity variations in MR images for computer-aided tissue classification. *IEEE Trans. Med. Imag.* **12:**770–781.

DeCarli, C., Murphy, D. G. M., Teichberg, D., Campbell, G., and Sobering, G. S. 1996. Local histogram correction of MRI spatially dependent image pixel intensity nonuniformity. *J. Magn. Reson. Imag.* **6:**519–528.

Deriche, R. 1987. Using Canny's criteria to derive a recursively implemented optimal edge detector. *Int. J. Comput. Vision* **1:**167–187.

Dice, L. 1945. Measures of the amount of ecologic association between species. *Ecology* **26:**297–302.

Duda, R. O., and Hart, P. E. 1973. *Pattern Classification and Scene Analysis.* Wiley, New York.

Gerig, G., Kubler, O., Kikinis, R., and Jolesz, F. 1992. Nonlinear anisotropic filtering of MRI data. *IEEE Trans. Med. Imag.* **11:**221–232.

Guillemaud, R., and Brady, M. 1997. Estimating the bias field of MR images. *IEEE Trans. Med. Imag.* **16:**238–251.

Guttmann, C., Jolesz, F. A., Kikinis, R., Killiany, R., Moss, M., Sandor, T., and Albert, M. 1998. White matter changes with normal aging. *Neurology* **50:**972–978.

Heindel, W. C., Jernigan, T. L., Archibald, S. L., Achim, C. L., Masliah, E., and Wiley, C. A. 1994. The relationship of quantitative brain magnetic resonance imaging measures to neuropathologic indexes of human immunodeficiency virus infection. *Arch. Neurol.* **51:**1129–1135.

Ibanez, V., Pietrini, P., Alexander, G. E., Furey, M., Teichberg, D., Rajapakse, J., Rapoport, S., Schapiro, M., and Horowitz, B. 1998. Regional glucose metabolic abnormalities are not the result of atrophy in Alzheimer's disease. *Neurology* **50:**1585–1593.

John, F. 1982. *Partial Differential Equations,* 4th ed. Springer-Verlag, New York.

Kapur, T., Grimson, W. E. L., Wells, W. M., and Kikinis, R. 1996. Segmentation of brain tissue from magnetic resonance images. *Med. Image Anal.* **1:**109–127.

Koepp, M. J., Richardson, M. P., Labbe, C., Brooks, D. J., Cunningham, V. J., Ashburner, J., Paesschen, W. V., Resevz, T., and Duncan, J. S. 1997. ^{11}C-Flumazenil PET, volumetric MRI, and quantitative pathology in mesial temporal lobe epilepsy. *Neurology* **49:**764–773.

Kruggel, F., and von Cramon, D. 1999. Alignment of magnetic-resonance brain datasets with the stereotactical coordinate system. *Med. Image Anal.* **3:**175–185.

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. Imag.* **17:**74–86.

Leahy, R., Hebert, T., and Lee, R. 1991. Applications of Markov random fields in medical imaging. In *Proceedings of the 11th International Conference on Information Processing in Medical Imaging* (D. A. Ortendahl and J. Llacer, Eds.), pp. 1–14. Wiley–Liss, New York.

Mackiewich, B. 1995. *Intracranial Boundary Detection and Radio Frequency Correction in Magnetic Resonance Images.* Simon Fraser Univ., Burnaby, British Columbia, Canada. [Ph.D. thesis]

Marr, D., and Hildreth, E. 1980. Theory of edge detection. *Proc. R. Soc. London* **207**(B):187–217.

Miller, M., Amit, Y., Christensen, G. E., and Grenander, U. 1993. Mathematical textbook of deformable neuroanatomies. *Proc. Nat. Acad. Sci. USA* **90:**11944–11948.

Nocera, L., and Gee, J. C. 1997. Robust partial-volume tissue classification of cerebral MRI scans. In *SPIE Medical Imaging 1997* (K. M. Hanson, Ed.), Vol. 3034, pp. 312–322. SPIE, Bellingham, WA.

Özkan, M., Dawant, B. M., and Maciunas, R. J. 1993. Neural-network-based segmentation of multi-model medical images: A comparative and prospective study. *IEEE Trans. Med. Imag.* **12:**534–544.

Perona, P., and Malik, J. 1990. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **12:**629–639.

Pham, D., and Prince, J. 1999. Adaptive fuzzy segmentation of magnetic resonance images. *IEEE Trans. Med. Imag.* **18:**737–752.

Rajapakse, J. C., and Kruggel, F. 1998. Segmentation of MR images with intensity inhomogeneities. *Image Vision Comput.* **16:**165–180.

Rusinek, H., de Leon, M. J., George, A. E., Stylopoulos, L. A., Chandra, R., Smith, G., Rand, T., Mourino, M., and Kowalski, H. 1991. Alzheimer disease: Measuring loss of cerebral gray matter with MR imaging. *Radiology* **178:**109–114.

Sandor, S., and Leahy, R. 1997. Surface-based labeling of cortical anatomy using a deformable database. *IEEE Trans. Med. Imag.* **16:**41–54.

Sandor, S. R. 1994. *Atlas of Guided Deformable Models for Automatic Anatomic Labeling of Magnetic Resonance Brain Images.* Univ. of Southern California, Los Angeles. [Ph.D. thesis]

Santago, P., and Gage, H. D. 1993. Quantification of MR brain images by mixture density and partial volume modeling. *IEEE Trans. Med. Imag.* **12:**566–574.

Shattuck, D. W. 2000. *Automated Segmentation and Analysis of Human Cerebral Cortex Imagery.* Univ. of Southern California, Los Angeles. [Ph.D. thesis]

Sled, J. G., and Pike, G. B. 1998. Understanding intensity non-uniformity in MRI. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI '98, Lecture Notes in Computer Science* (W. M. Wells, A. C. F. Colchester, and S. Delp, Eds.), pp. 614–622. Springer-Verlag, Berlin/Heidelberg.

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. Imag.* **17:**87–97.

Thaler, H., Ferber, P., and Rottenberg, D. A. 1978. A statistical method for determining proportions of gray matter, white matter, and CSF using computed tomography. *Neuroradiology* **16:**133–135.

Thompson, P. M., MacDonald, D., Mega, M. S., Holmes, C. J., Evans, A. C., and Toga, A. W. 1998. Detection and mapping of abnormal brain structure with a probabilistic atlas of cortical surfaces. *J. Comput. Assisted Tomogr.* **21:**567–581.

Van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. 1999a. Automated model-based bias field correction of MR images of the brain. *IEEE Trans. Med. Imag.* **18:**885–896.

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

Warfield, S. K., Westin, C.-F., Guttmann, C. R. G., Albert, M., Jolesz, F. A., and Kikinis, R. 1999. Fractional segmentation of white matter. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI '99, Lecture Notes in Computer Science* (C. Taylor and A. Colchester, Eds.), pp. 928–935. Springer-Verlag, Berlin/Heidelberg.

Wells, W. M., Grimson, W. E. L., Kikinis, R., and Jolesz, F. A. 1996. Adaptive segmentation of MRI data. *IEEE Trans. Med. Imag.* **15:**429–442.

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. Assisted Tomogr.* **22:**141–154.

Wu, Z., and Leahy, R. 1991. A graph theoretic approach to tissue classification in MR images. In *Proceedings of the SPIE,* Vol. 1450, *Biomedical Image Processing II.* (A. C. Bovik and V. Howard, Eds.), pp. 120–132. SPIE, Bellingham, WA.

Yan, M. X. H., and Karp, J. S. 1995. An adaptive Bayesian approach to three-dimensional MR brain segmentation. In *Information Processing in Medical Imaging* (Y. Bizais, C. Barillot, and R. D. Paola, Eds.), pp. 201–213. Kluwer Academic, Dodrecht.

Zeng, X., Staib, L. H., Schultz, R. T., and Duncan, J. S. 1999. Segmentation and measurement of the cortex from 3D MR images using coupled surfaces propagation. *IEEE Trans. Med. Imag.* **18:** 100–111.

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