During the summer of 2000, Albert Montillo developed the algorithm and implementation. He demonstrated for the first time, that the approach worked well on many brains in a test set containing healthy controls and Alzheimer's subjects. This led to extensive testing on Alzheimer's (AD) subjects with numerous clinical collaborators, which is a focus of this publication. This paper has been cited >1800 times. This code forms the basis of open source subcortical segmentation module in FreeSurfer and related FDA approved version in NeuroQuant product. Success of subcortical approach inspired a closely related, highly popular cortical segmentation approach on the sphere

Neuron, Vol. 33, 341–355, January 31, 2002, Copyright ©2002 by Cell Press

Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain

Neurotechnique

Bruce Fischl,[1] David H. Salat,[1] Evelina Busa,[1] Marilyn Albert,[2,3] Megan Dieterich,[5] Christian Haselgrove,[5] Andre van der Kouwe,[1] Ron Killiany,[4] David Kennedy,[5] Shuna Klaveness,[5] Albert Montillo,[6] Nikos Makris,[5] Bruce Rosen,[1] and Anders M. Dale[1,7]

[1]Massachusetts General Hospital
Nuclear Magnetic Resonance Center
Rm. 2328, Building 149
13th Street
Charlestown, Massachusetts 02129

[2]Department of Neurology
Massachusetts General Hospital
Harvard Medical School
55 Fruit Street, VBK 901
Boston, Massachusetts 02114

[3]Department of Psychiatry, CNY-9
Massachusetts General Hospital
Boston, Massachusetts 02114

[4]Department of Anatomy and Neurobiology
Boston University School of Medicine
715 Albany Street
Boston, Massachusetts 02118

[5]Center for Morphometric Analysis
Neuroscience Center, MGH-East
Building 149, 13th Street
Charlestown, Massachusetts 02129

[6]Computer Science Department
University of Pennsylvania
111 Towne Building
220 South 33rd Street
Philadelphia, Pennsylvania 19104

Summary

We present a technique for automatically assigning a neuroanatomical label to each voxel in an MRI volume based on probabilistic information automatically estimated from a manually labeled training set. In contrast to existing segmentation procedures that only label a small number of tissue classes, the current method assigns one of 37 labels to each voxel, including left and right caudate, putamen, pallidum, thalamus, lateral ventricles, hippocampus, and amygdala. The classification technique employs a registration procedure that is robust to anatomical variability, including the ventricular enlargement typically associated with neurological diseases and aging. The technique is shown to be comparable in accuracy to manual labeling, and of sufficient sensitivity to robustly detect changes in the volume of noncortical structures that presage the onset of probable Alzheimer's disease.

Introduction

Neurodegenerative disorders, psychiatric disorders, and healthy aging are all frequently associated with structural changes in the brain. These changes can cause alterations in the imaging properties of brain tissue, as well as changes in morphometric properties of brain structures. Morphometric changes may include variations in the volume or shape of subcortical regions, as well as alterations in the thickness, area, and folding pattern of the cortex. While surface-based analyses that depend on models of the position and orientation of the cortical ribbon can provide an accurate assessment of cortical variability, volumetric techniques are required to detect changes in noncortical structures. For example, changes in ventricular or hippocampal volume are frequently associated with a variety of diseases (e.g., Puri et al., 1999; Killiany et al., 2000; Wolf et al., 2001). This type of analysis has commonly been accomplished by a having a trained anatomist or technician manually label some or all of the structures in the brain, a procedure that can take up to a week for high-resolution images. Here, we use the results of the manual labeling using the validated techniques of the Center for Morphometric Analysis (Caviness et al., 1989; Kennedy et al., 1989; Goldstein et al., 1999; Seidman et al., 1999) to automatically extract the information required for automating the segmentation procedure. The automated segmentation procedure requires approximately 30 min on current workstation hardware (e.g., 1 GHz Pentium III). The additional capability to run multiple processes in parallel enables the labeling of thousands of brains per day.

Typically, manual labeling of brain structures is accomplished using a variety of information including image intensities, global position within the brain, position relative to neighboring brain structures, as well as anatomical landmarks. The challenge in labeling brain structures based on MRI image intensities alone is illustrated in Figure 1, which shows the intensity histograms of nine different neuroanatomical structures defined by a manual segmentation procedure based on a typical T_1-weighted MRI image. Examining this figure, it is apparent why no global classification scheme can successfully distinguish structures from each other based only on intensity information—there is far too much overlap between the class distributions (even cortical gray matter and white matter overlap by more than 12%). While adding additional MRI sequences with differing contrast properties or different imaging modalities entirely can help separate the class distributions, spatial information is still required to disambiguate the classification problem.

The use of spatial information to aid in classification is facilitated by the construction of a *probabilistic atlas* (Collins et al., 1994; Fox et al., 1994; Mazziotta et al., 1995; Thompson et al., 1997). In this type of atlas, information regarding the statistical properties of anatomical structures is stored in a space in which coordinates have anatomical meaning as opposed to the somewhat arbitrary coordinates in a raw image, which are dependent on the position, orientation, and shape of a subject's head in the MR scanner. Spatial information can aid in classification in several ways: (1) the number of possible anatomical classes at a given global position in

[7]Correspondence: dale@nmr.mgh.harvard.edu

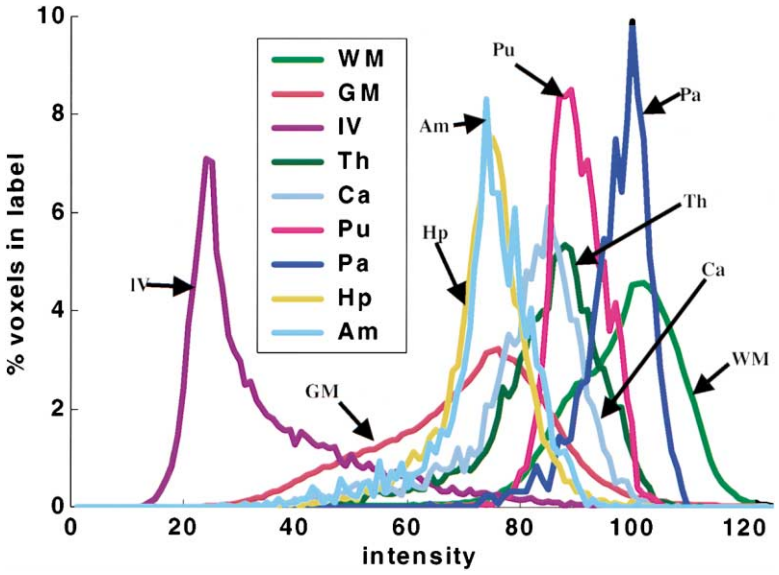

Figure 1. Intensity Histograms for White Matter (WM), Cortical Gray Matter (GM), Lateral Ventricle (IV), Thalamus (Th), Caudate (Ca), Putamen (Pu), Pallidum (Pa), Hippocampus (Hp), and Amygdala (Am)

the brain as specified by an atlas coordinate is typically relatively small (note that we will use *tissue class* to mean the type of tissue within a voxel [e.g., gray matter] and *anatomical class*, or just *class*, to indicate the neuroanatomical label assigned to a voxel [e.g., caudate] throughout this manuscript); (2) neuroanatomical structures occur in a characteristic spatial pattern relative to one another (e.g., the amygdala is anterior and superior to the hippocampus); and (3) many tissue classes have spatially heterogeneous MRI imaging properties that vary in a spatially predictable fashion. This latter result has been quantified for some structures using MR relaxometry, which reveals that different regions of white matter have significantly different T_1 properties (Cho et al., 1997). Furthermore, even within cortical gray matter, significant variations have been reported in the intrinsic tissue parameters. For example, frontal cortex has been found to have an average T_1 that is 20% longer than that found in motor and somatosensory cortex (Steen et al., 2000). Thus, it is clear that information regarding global position in the brain could aid in the segmentation process, to account for within-structure variability in the intrinsic tissue parameters, as well as to indicate the prior probability (the probability before observing the data) of a structure occurring at a given location independent of intensity information.

The problem of differentiating multiple gray matter structures presents additional challenges beyond those encountered by classification schemes designed to label only gray, white, and CSF, as is typically done (Wells et al., 1996; Held et al., 1997; Teo et al., 1997; Kapur et al., 1998; Dale et al., 1999; Suckling et al., 1999; Ballester et al., 2000; Germond et al., 2000; H. Tang et al., 2000; Zhang et al., 2001). However, accounting for the heterogeneity of the tissue properties of cortical and subcortical gray matter structures can simplify the classification procedure by reducing within-class variability, as subcortical structures such as the thalamus, the putamen, and the caudate and cortex all have significantly different T_1 properties (as implied by Figure 1). Compiling statistics separately for subcortical structures, as opposed to collapsing all gray matter and white matter into two classes, prevents the broadening of the underlying distributions that would otherwise occur, and facilitates a more accurate segmentation. Thus, the current procedure obviates the need to pool information across structures *or* across space by maintaining class statistics (e.g., means and variances of the MRI intensities of a given neuroanatomical structure) on a per-location per-class basis throughout an atlas space.

Local spatial relationships between labeled structures have been encoded by modeling the labeled image using *Markov random fields* (MRFs) in a variety of image processing contexts (Geman and Geman, 1984). In the MRF approach, the probability of a label at a given voxel is computed not just in terms of the intensities and prior probabilities at that voxel, but also as a function of the labels in a neighborhood around the voxel in question. In the context of segmenting MR images, *isotropic* (all directions are equal) and *stationary* (the probabilities are the same for all spatial location) MRFs have been used to provide smoothness constraints on a given segmentation (Held et al., 1997; Kapur et al., 1998; Zhang et al., 2001). In this way, the prior probability of a label is computed by examining how likely the label is given the labels of its neighbors, regardless of the direction of the neighbor, or the position within the brain. While this type of approach can obviate the need for prefiltering of the images, it does not provide for the use of information regarding the spatial relationships of classes to one another. For example, as can be seen from Figure 1, the amygdala and the hippocampus are close to indistinguishable using only intensity information. However, they occur in an anatomically stereotypical relationship to one another, with the amygdala residing anterior and superior to the hippocampus. Encoding this type of information requires relaxing the spatial isotropy constraint of the standard MRF, and tabulating statistics for each spatial direction separately. This allows the separate calculation, for example, of the probability that a voxel labeled hippocampus will have its inferior neighbor labeled as amygdala, providing a strong set of constraints on the space of allowable segmentations.

Thus, we resolve the inherent ambiguity of the class intensity distributions in a number of ways. The first is through the use of a space-varying classification procedure. That is, class statistics (e.g., means and covariance matrices) are tabulated regionally throughout an atlas space, using an optimal linear alignment procedure to register each brain with an average. Further, prior probabilities are computed via a frequency histogram in the atlas space, allowing the calculation of the probability that a given anatomical class occurs at a given atlas location. Finally, the prior probability of a given spatial arrangement of anatomical labels is incorporated into the final segmentation procedure. These priors are also computed from the training set for each point in the atlas by modeling the segmentation as an anisotropic nonstationary Markov random field, resulting in a procedure that even using a low-dimensional linear transform (i.e., a transformation of coordinates **y** into coordinates **x** of the form $\mathbf{x} = \mathrm{M}\,\mathbf{y}+\mathbf{b}$ where M is a 3×3 matrix and **b** is a vector of translations resulting in 12 total parameters) is comparable in terms of accuracy to manual labeling.

Results and Discussion

Comparison with Manual Segmentation

In order to validate the automated segmentation procedure, we compared the automated results with those of manually labeling the same datasets. The test-retest reliability of the manual segmentation procedure itself was assessed in a separate study, in which five users, experienced in manual segmentation, each labeled a single test image. Each of these labelings was then compared to every other using two criteria for quantifying inter-rater reliability, as suggested by Collins et al. (1995): percent overlap and percent volume difference (note that we normalize by the volume of the average of the automatic and manually labeling, as opposed to using the manual labeling).

Given two different labelings of a structure, denoted by L_1 and L_2, and a function $V(L)$, which takes a label and returns its volume, the percent volume overlap is given by:

$$O(L_1,L_2) = \frac{V(L_1 \cap L_2)}{(V(L_1) + V(L_2))/2} \times 100\% \qquad (1)$$

For identical labelings, $O(L_1,L_2)$ achieves its maximum value of 100, with decreasing values indicating less perfect overlap. Note that the overlap between two different labelings will be reduced by slight shifts in the spatial location of one label with respect to another. Given that many neuroanatomical studies are only interested in quantifying volumetric changes in structures, Collins et al. (1995) also define another metric, which is insensitive to spatial shift, and only quantifies volume difference between two labelings:

$$D(L_1,L_2) = \frac{|V(L_1) - V(L_2)|}{(V(L_1) + V(L_2))/2} \times 100\% \qquad (2)$$

For labels with identical volume, $D(L_1,L_2)$ achieves its optimal value of 0, with increasing values indicating a greater volume difference between the two labelings. Note that greater values of $D(L_1,L_2)$ lead directly to reduced statistical power to detect subtle volumetric changes in subcortical structures.

The results of quantifying $O(L_1,L_2)$ and $D(L_1,L_2)$ for the inter-rater reliability study as well as for the comparison of manual with automated techniques are given in Figure 2. In this study, magnetic resonance imaging (MRI) scans from seven healthy young subjects were manually segmented, and a separate atlas was built for each volume, using a standard jackknifing procedure in which the remaining 6 volumes were used to estimate the class statistics and prior densities. Each volume was then registered and segmented using the atlas constructed without it. The results of this labeling were then compared to the manual labeling of the same volume in order to compute the volume overlap and volume difference between the automatic and manual labelings. The inter-rater reliability of the manual labeling was computed in a similar manner, with a single volume being segmented by five separate experts, as described above. The results of these studies are given in Figure 2, which illustrates volume overlap (top) and volume difference (bottom) for the manual (light bars) and automated techniques (dark bars). The error bars are standard errors of the mean. As can be seen, the agreement between the automated and manual labelings is comparable to that obtained by comparing the labelings of different experts. Finally, for these same datasets, we computed the volume of the same set of structures as Figure 2 for both the manual and automated labelings. As shown in Figure 3, there is no discernable bias in the automated volume measurements, which are statistically indistinguishable from the manually computed volumes.

Detection of Volumetric Changes in Mild Alzheimer's Disease

In order to assess the ability of the segmentation procedure to reveal subtle structural differences associated with disease, MRI scans from 134 subjects were registered and labeled using the automated methods outlined in this paper. The labeling was based on an atlas generated from 12 manually labeled datasets that had been acquired using the same MR protocol as the larger subject group. Structure volumes were corrected for total brain size by dividing each structure by the volume of all brain labels (note that this is a nonstandard means for accounting for the variability of intra-cranial volume (ICV). However, for the current purposes, it renders the procedure self-contained, as it does not depend on a separate tool for measuring ICV. For any detailed morphometric study using these tools, we anticipate the need to correct for ICV). Subjects in this study were categorized into four groups based on Clinical Dementia Rating (CDR; Hughes et al., 1982) and follow up evaluations. Normal controls (n = 25, 9 men and 16 women, mean age 72.1; Mini-Mental State Examination (MMSE) 29.3) consisted of subjects with normal cognition (CDR = 0.0) on first evaluation (baseline), and remained cognitively intact for up to three years of annual follow up evaluations. Another subpopulation of subjects (n = 92) met CDR criteria for questionable AD (CDR = 0.5) at baseline. This group was further subdivided based on follow up examinations into two categories. Converters

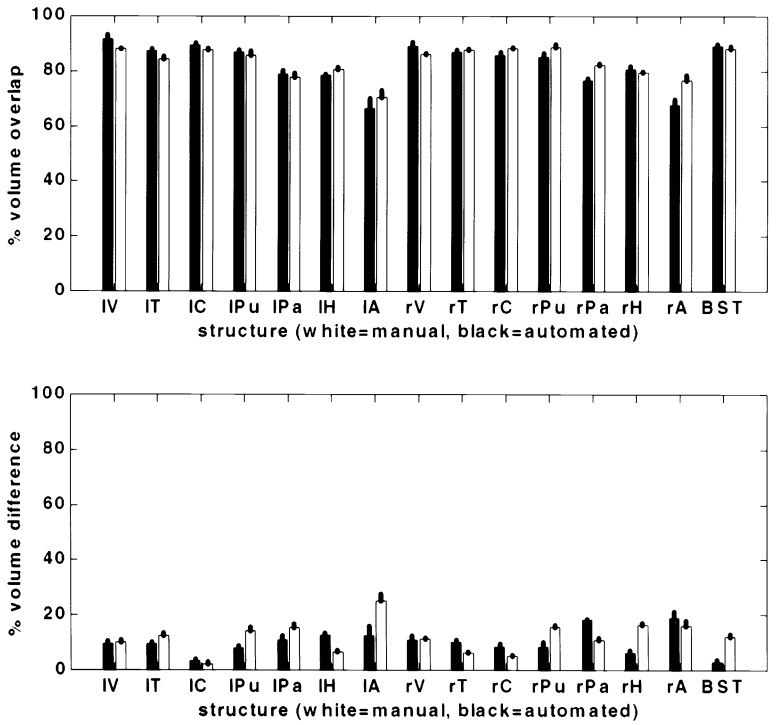

Figure 2. Comparison of Inter-Rater Reliability for Manual Segmentation (White) with the Reliability of the Automated with the Manual Measures (Dark)

Top: percent volume overlap for various structures using the two techniques. Bottom: percent volume difference for the same structures. Key: lV—left lateral ventricle, lT—left thalamus, lC—left caudate, lPu—left putamen, lPa—left pallidum, lH—left hippocampus, lA—left amygdala, rV—right lateral ventricle, rT—right thalamus, rC—right caudate, rPu—right putamen, rPa—right pallidum, rH—right hippocampus, rA—right amygdala, BST—brainstem.

(n = 21, 9 men and 12 women; mean age 74.2; MMSE 28.7) consisted of subjects who met CDR criteria for questionable AD (CDR = 0.5) at baseline, and progressed to meet NINCDS/ADRDA (National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer's Disease and Related Disorders Association) criteria for probable AD. Questionables (n = 71, 29 men and 42 women; mean age 72; MMSE 29.3) made up the remainder of this group, and were defined as those subjects with mild memory problems who did not progress to meet clinical requirements for probable AD over the 3 years of the study. Mild AD (n = 17, 7 men and 10 women; mean age 67.1) consisted of patients who met NINCDS/ADRDA criteria for probable AD at the time MR scans were collected. These patients were mildly impaired (CDR = 1; MMSE = 24.3) and clinically examined to exclude medical causes known to produce dementia. All subjects were free of significant underlying medical, neurological, or psychiatric illness (based on standard laboratory tests and a clinical evaluation).

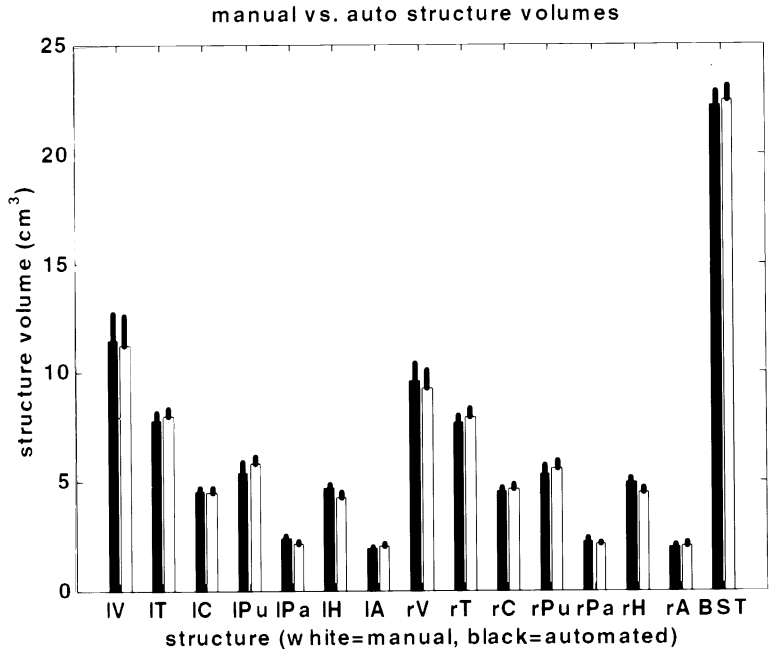

Figure 3. Comparison of Structure Volumes Computing Using Manual Labeling (White) and Automated Segmentation (Dark)

The error bars are standard errors of the mean. Key: lV—left lateral ventricle, lT—left thalamus, lC—left caudate, lPu—left putamen, lPa—left pallidum, lH—left hippocampus, lA—left amygdala, rV—right lateral ventricle, rT—right thalamus, rC—right caudate, rPu—right putamen, rPa—right pallidum, rH—right hippocampus, rA—right amygdala, BST—brainstem.

Groups did not differ in years of education. Detailed descriptions of the recruitment procedures and criteria for subject recruitment have been published elsewhere (Johnson et al., 1998; Daly et al., 2000; Killiany et al., 2000).

A sample of the results of this study is shown below, summarized as the volumetric difference between seven different brain structures: the lateral ventricles, the third and fourth ventricles (Figure 4, top right), the temporal horn of the lateral ventricles (Figure 4, middle left), the hippocampus (Figure 4, middle right) the amygdala (Figure 4, bottom left), and the thalamus (Figure 4, bottom right). Table 1 lists the statistical significances of the volumetric differences between the groups. The statistical significance of the volumetric differences between the different groups was computed using a random effects model (using a two-tailed t test with unequal variance). Significances are given if they are below the 0.05 level (comparisons below this threshold are listed as NS for not significant). Note that no correction for multiple comparisons has been performed. If no prior hypothesis existed, one would need to correct for the number of comparisons made, and the p values given in this section would be overly liberal. While further work is required in order to analyze these results (as well as a number of other affected brain structures), it is apparent that even the subtle structural changes that presage the onset of Alzheimer's disease are clearly distinguishable using these automated techniques. The great majority of these findings would be expected with a sufficiently sensitive measurement technique, given the literature on AD, and therefore support the use of these automated procedures in discriminating patients with specific disorders from a heterogeneous group of patients with similar clinical presentations (e.g., discriminating converters from questionable AD in this study).

Normal Controls versus AD

Significant dilation of the lateral (Murphy et al., 1993; Forstl et al., 1995; Pedersen et al., 1999) and third (Powell et al., 1993; Forstl et al., 1995; Pedersen et al., 1999) ventricles has been reported in prior studies of AD. The lateral ventricles are expanded with mild AD (MMSE $>$ 20) (Murphy et al., 1993), supporting our findings that lateral ventricular volume is significantly increased early in the disease process (converters in the present study). The results are in agreement with numerous prior studies showing hippocampal degeneration in imaging (Jack et al., 1992; Killiany et al., 1993) and histological (Ball et al., 1985) studies of AD as well as with the finding that amygdala volume could be reduced to a greater extent than hippocampal volume with early AD (Mizuno et al., 2000). Volumetric reduction of the thalamus has also been reported with AD (Jernigan et al., 1991). Still, no prior study has measured all of these structures in the same participants. Thus, the automated segmentation procedure allows a comprehensive analysis of the relative degeneration among numerous structures in the brain.

Converters versus Questionable AD

The results also agree with prior studies demonstrating that volumetric measurements are useful in distinguishing converters from nonconverters in participants with questionable AD (Killiany et al., 2000). Regions that differed in volume between the questionable group and the converters include the hippocampus, amygdala, and the third, fourth, and lateral ventricles. These findings support prior studies that show hippocampal volume reduction in confirmed preclinical AD (Kaye et al., 1997; Jack et al., 1999) but did not support a reduction in hippocampal volume with questionable AD (De Toledo-Morrell et al., 2000; Wolf et al., 2001), as there were no differences in hippocampal volume between the questionable group and healthy control subjects. Still, participants with questionable AD in this study likely differed from those examined in prior studies. This is because participants with questionable AD who progressed to develop probable AD within three years of scanning were a distinct group in this study (converters). Thus, it is possible that the questionable AD group included participants with disorders other than AD, contributing to the lack of differences. Prior studies have shown equivalent amygdala volume in healthy subjects and subjects with early AD (Killiany et al., 1993). Although it is clear that the amygdala degenerates with AD, there is controversy over the temporal progression of this degeneration in the broader literature that is possibly due to differential volumetric methods, subject selection, and statistical power. Importantly, the current study examined a large number of clinically screened subjects, supporting confidence in our findings of early involvement of the amygdala. Thus, the results presented suggest that these automated procedures could be useful in the discrimination of healthy aging from prodromal AD as described in prior studies (Killiany et al., 2000).

The results presented here suggest a progression of degeneration in the brain with AD. There is an enlargement of all ventricular regions and a decline in amygdala and hippocampal volume in the early stages of the disease. In addition, the results suggest that degenerative changes in the hippocampus continue with disease progression as the AD group had significantly less hippocampal volume than converters. Still, it is important to note that as with neuropathological staging of AD (Braak and Braak, 1991), there is likely great intersubject variation in both anatomy and pathology that confounds the use of these measures in the staging of disease progression. Thus, these techniques represent a unique possibility to indirectly examine the neuropathological staging of disease progression in a large number of clinically characterized patients. Additionally, although the four groups studied are useful for understanding the degenerative progression of AD, it is possible that differences observed in cross-sectional study do not represent a progression of the disease process but could be the result of a cohort effect among groups (e.g., the AD group had smaller hippocampal volumes *prior* to development of clinical symptoms). Thus, it is important to follow these studies with similar longitudinal studies to track degenerative change across time in each individual subject. Finally, the results described only represent a small number of labels assigned to the cortex in this study and must be considered within the larger context of neural degeneration. Future work will model the progression of the disease throughout the brain using all labels assigned in the automated segmentation procedure and sensitive clinical measures of disease progression and cognitive decline.

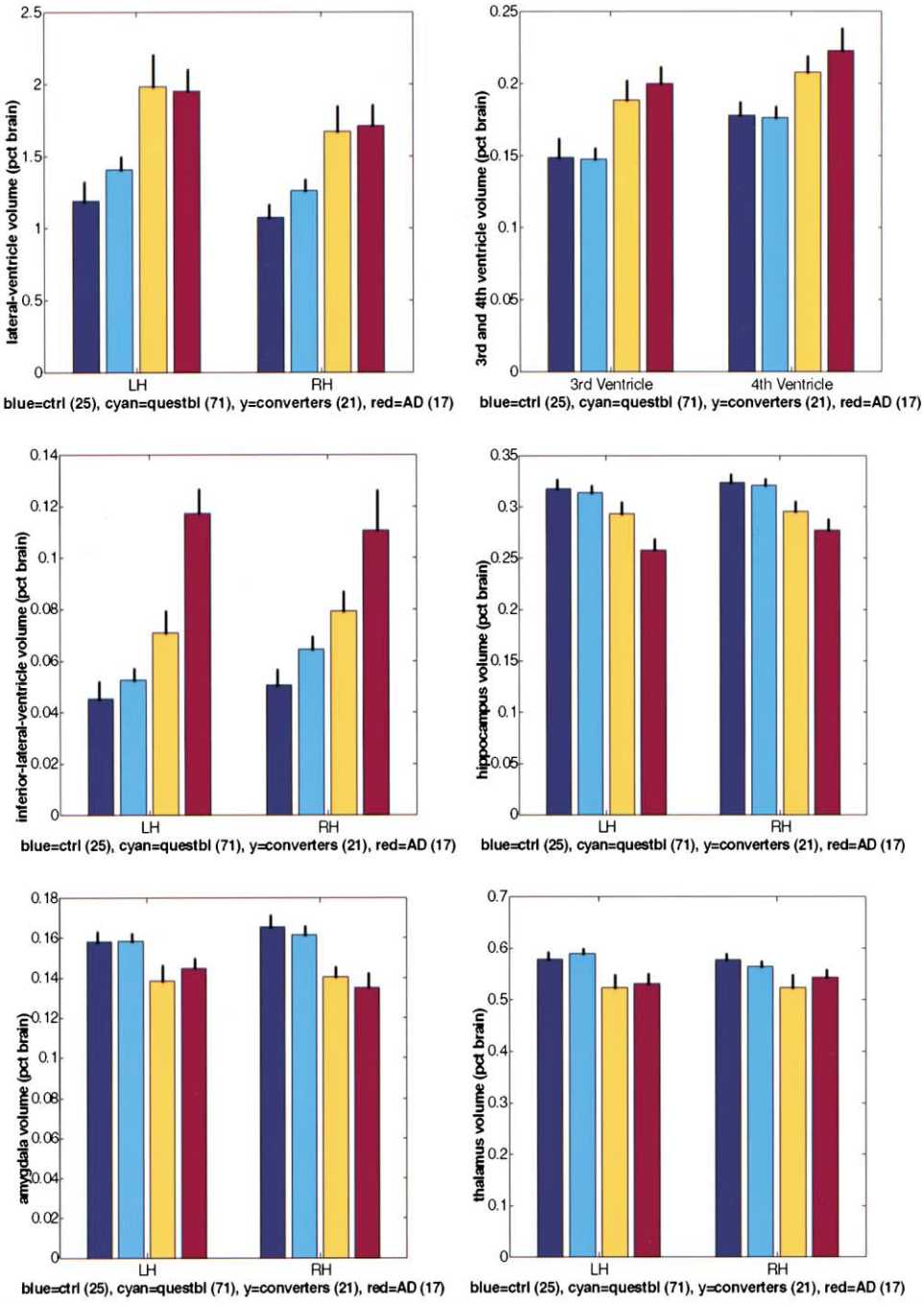

Figure 4. Comparison of the Volume of Various Structures in Normal Controls (Blue), Questionables (Cyan), Converters (yellow), and Patients with Mild AD (red)

Top left: lateral ventricles (LH = left hemisphere, RH = right hemisphere). Top right: third and fourth ventricles. Middle left: inferior lateral ventricles. Middle right: hippocampus. Bottom left: amygdala. Bottom right: thalamus.

Limitations and Future Improvements

The morphological properties of subcortical structures are potentially valuable markers of a variety of disorders, including schizophrenia (Goldstein et al., 1999; Puri et al., 1999; Seidman et al., 1999), Alzheimer's disease (Luxenberg et al., 1987; Laakso et al., 1995; Lehtovirta et al., 1995; Albert, 1996; Double et al., 1996; Frisoni et al., 1996; Convit et al., 1997; Jack et al., 1997, 1999, 2000; Kaye et al., 1997; Killiany et al., 2000; Small et al., 2000), Huntington's disease (Halliday et al., 1998; Vonsattel and DiFiglia, 1998), and other conditions (Caviness et al., 1992, 1996a; Breiter et al., 1994; Double et al., 1996; Jenike et al., 1996; Kaye et al., 1997; Makris et al., 1997; O'Sullivan et al., 1997; Wolf et al., 2001). While manual methods exist for assessing this type of change, the process of manually labeling an entire high-

Table 1. Statistical Significances of Structure Volume Differences

Comparison	Significance (Lateral Ventricle)	Significance (3rd and 4th Ventricle)	Significance (Inferior Lateral Ventricle)	Significance (Hippocampus)	Significance (Amygdala)	Significance (Thalamus)
Control versus AD (left hemisphere)	$p < 3.3 \times 10^{-4}$	$p < 5.0 \times 10^{-3}$ (3rd ventricle)	$p < 8.5 \times 10^{-8}$	$p < 3.2 \times 10^{-5}$	$p < 4.6e^{-2}$	$p < 3.5 \times 10^{-2}$
Control versus AD (right hemisphere)	$p < 4.3 \times 10^{-4}$	$p < 1.2 \times 10^{-2}$ (4th ventricle)	7.1×10^{-4}	$p < 5.1 \times 10^{-4}$	$p < 1.3 \times 10^{-3}$	NS
Control versus converter (left hemisphere)	$p < 3.0 \times 10^{-3}$	$p < 3.2 \times 10^{-2}$ (3rd ventricle)	$p < 1.6 \times 10^{-2}$	NS	$p < 3.1 \times 10^{-2}$	$p < 4.0 \times 10^{-2}$
Control versus converter (right hemisphere)	$p < 3.4 \times 10^{-4}$	4.0×10^{-2} (4th ventricle)	$p < 3.1 \times 10^{-3}$	$p < 1.8 \times 10^{-2}$	$p < 1.4 \times 10^{-3}$	$p < 3.9 \times 10^{-2}$
Converter versus AD (left hemisphere)	NS	NS (3rd ventricle)	$p < 4.9 \times 10^{-4}$	$p < 2.1 \times 10^{-2}$	NS	NS
Converter versus AD (right hemisphere)	NS	NS (4th ventricle)	NS	NS	NS	NS
Converter versus questionable (left hemisphere)	1.6×10^{-2}	5.9×10^{-3} (3rd ventricle)	$p < 4.6 \times 10^{-2}$	NS	1.8×10^{-2}	$p < 9.3 \times 10^{-3}$
Converter versus questionable (right hemisphere)	3.0×10^{-2}	1.7×10^{-2} (4th ventricle)	NS	$p < 1.9 \times 10^{-2}$	7.0×10^{-4}	NS

Shown are statistical comparisons of the volumes fo the lateral ventricles (column 2), the 3rd and 4th ventricles (column 3), the inferior lateral ventricles (column 4), the hippocampus (column 5), the amygdala (column 6), and the thalamus (column 7) in normal controls versus AD (rows 1 and 2), controls versus converters (rows 3 and 4), converters versus AD (rows 5 and 6), and converters versus questionables (rows 7 and 8). The table entries are p values for a t-test of the significance of the volumetric differences. Alternate rows are the left and right hemisphere except in column 3, in which alternate rows represent comparisons for the 3rd and the 4th ventricle.

resolution structural MR volume requires on the order of a week for a trained neuroanatomist or technician. This makes the routine analysis of large patient and control populations untenable. Further, manual labeling procedures typically generate a labeling that is more consistent when viewed in one slice direction than in others, or in noncardinal directions. Finally, manual labeling procedures do not generalize well to the use of multi-spectral inputs.

The automated method described in this paper for assigning a neuroanatomical label to every voxel in the brain has been shown to be comparable in terms of accuracy to a previously validated method of manual segmentation. The accurate labeling of a large number of structures is enabled through the use of both global and local spatial information. The global information is encoded by distributing classifiers throughout an atlas volume and maintaining class statistics on a per-class, per-location basis, allowing the classifiers to be robust to variations in the contrast properties of an anatomical class over space. Local information is incorporated into the classification procedure by modeling the segmentation as a nonstationary anisotropic Markov random field. In contrast to earlier work, in which isotropic Markov random fields have been employed in order to yield a smoother segmentation, the introduction of anisotropy and nonstationarity into the segmentation model allows the spatial relationships of anatomical classes to one another to be incorporated into the segmentation procedure in a principled fashion. The incorporation of high-dimensional registration techniques (Bajcsy et al., 1983; Bookstein, 1989; Miller et al., 1993; Gee et al., 1994; Vannier et al., 1994; Christensen et al., 1996; Ashburner et al., 1997; Collins and Evans, 1997; Woods et al., 1998; Thompson et al., 2000) should further improve the accuracy of the labeling.

The automatic labeling procedure can also be used to automatically define regions of interest (ROIs) for use in functional imaging studies. Specifically, this will allow one to generate average time courses by structure, or even parts of structures, facilitating, for example, the comparison of the response of the caudate to that of the putamen, or anterior hippocampus to posterior hippocampus. Furthermore, having access to voxel labels should help MR relaxometry analysis, in which intrinsic tissue parameters are inferred from a set of MR images, a procedure that is extremely sensitive to partial volume effects, as the models of image formation used in the parameter estimation rarely allow more than one tissue type to occur in a voxel. Explicit models of the anatomical classes would permit the parameter estimates to be computed using voxels that do not border a different tissue type, avoiding partial voluming.

Although the results presented in this paper only make use of single-valued (T_1-weighted) images, the derivations of the algorithms are vector based, and hence the incorporation of multi-spectral data is straightforward. Basing the classification on images acquired with multiple types of scan sequences (e.g., T_1, T_2, proton density, diffusion) should increase the accuracy of the segmentation. Such multi-spectral features could also include derived variables such as image gradients or Laplacians.

The incorporation of an explicit forward model for MRI signal intensities as a function of intrinsic tissue parameters as well as pulse sequence parameters, as described in section 0, potentially makes the segmentation largely invariant to details of the image acquisition procedure. This includes invariance to factors such as scanner model, software version, and scan protocol. This is particularly important when comparing data across sites, as in multi-center clinical trials, or within a site across time in longitudinal studies, where it is

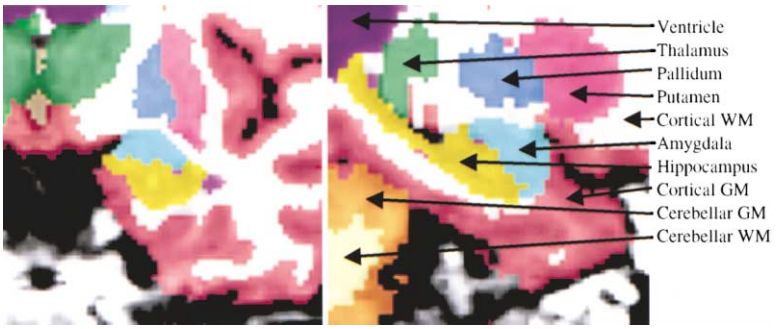

Figure 5. Manual Segmentation Results in the Temporal Lobe

Left: coronal view, right: sagittal view.

impractical or undesirable to maintain the same acquisition protocol for the duration of the study. The invariance to the details of the image acquisition results from basing the classification on intrinsic properties of the underlying tissue, as opposed to the somewhat arbitrary image intensities obtained using a particular scan protocol. Ultimately, such an approach would greatly enhance the value and feasibility of constructing large-scale, multi-site medical imaging databases.

The automated nature of the methods described here, in contrast with existing manual or semi-automated techniques, allows for their routine application in large-scale studies. Having access to this type of detailed morphometric information for large populations including various disorders as well as a spectrum of normal controls should facilitate the characterization of the anatomical signatures associated with specific disorders. Ultimately, this may provide a more accurate and sensitive tool for early diagnosis of brain disorders.

Experimental Procedures

Problem Statement

The problem of automatically labeling brain structures from neuroimaging data can be naturally phrased within the framework of Bayesian parameter estimation theory. In this approach, one can relate the probability of a segmentation W given the observed (potentially multispectral) image $\mathbf{I}$ to the probability $p(\mathbf{I}|W)$ of the image occurring given a certain segmentation, together with the prior probability of the segmentation $p(W)$:

$$p(W|\mathbf{I}) \propto p(\mathbf{I}|W)\, p(W) \tag{3}$$

The primary advantage of the Bayesian approach is that it allows for the explicit incorporation of prior information via the $p(W)$ term in Equation 3. In order to render the problem more tractable in the face of the large degree of overlap in the class distributions shown in Figure 1, both the priors on W and the conditional probability of observing the image given the classification $p(\mathbf{I}|W)$ can be expressed within an atlas space, allowing them to vary as a function of position within the brain (hence making them nonstationary). The advantage of using an atlas space is that coordinates in the atlas have more anatomical meaning than the native coordinate system of the image (Bajcsy et al., 1983; Bookstein, 1989; Miller et al., 1993; Gee et al., 1994; Vannier et al., 1994; Christensen et al., 1996; Ashburner et al., 1997; Collins and Evans, 1997; Woods et al., 1998; Fischl et al., 1999; Thompson et al., 2000). Classifiers can then be distributed throughout the atlas, allowing each one to focus on only the small number of classes that may occur within the region for which the classifier is responsible. The number of classes that occur within a region of space is then directly related to the accuracy of the atlas coordinate system. That is, $P(W(\mathbf{r}) = c)$ will be 0 for all but a few values of c at each atlas location $\mathbf{r}$. In practice, if the classifiers are reasonably dense in the atlas space, then the number of classes at each location is typically relatively small. For example, using a linear transform to map native MR coordinates into atlas coordinates, the number of classes at a given location is rarely greater than 4, and in fact averages less than 3 within the brain. In this way, the intractable problem of classifying each voxel into one of 40 or so labels with similar intensity distributions is decomposed into a set of tractable problems of classifying the voxels in each region of the image into only a small number of labels.

The definition of the atlas requires the calculation of a function $f(\mathbf{r})$, which takes native image coordinate as input, and returns the coordinate of the corresponding point in the atlas. For f to be useful in this context, the coordinates it returns should be related to the anatomical location of $\mathbf{r}$. This type of mapping therefore provides the ability to meaningfully relate coordinates across subjects. In the most general case, we wish to maximize the joint probability of both the segmentation W and the atlas function f:

$$p(W, f|\mathbf{I}) \propto p(\mathbf{I}|W, f)\, p(W \mid f)\, p(f) \tag{4}$$

The terms $p(\mathbf{I}|W,f)$ and $p(W \mid f)$ in Equation 4 provide a natural means for incorporating atlas information into the segmentation procedure. The first term encodes the relationship between the class label at each atlas location and the predicted image intensities. Using the atlas space, we can allow the class statistics to vary as a function of location, allowing the within-class variations in tissue properties that are known to exist in the human brain (Cho et al., 1997; Steen et al., 2000) to be captured in a natural manner. The second term allows the expression of prior information regarding the spatial structure of the anatomical classes. Finally, the term $p(f)$ provides a means for constraining the space of allowable atlas functions (e.g., continuity, differentiability, and invertibility).

Note that the atlas information expressed by $p(\mathbf{I}|W,f)$ depends on the details of the image acquisition protocol or scanner type. In order to reduce this dependence, information about the relationship between the image intensities and the acquisition parameters α and tissue properties β can be explicitly incorporated into this term. That is, $p(\mathbf{I}|W,f)$ can be factored as follows:

$$p(\mathbf{I}|W,f) = p(\mathbf{I}(\alpha)|\beta)\, p(\beta|W,f) \tag{5}$$

The intrinsic tissue parameters β (e.g., T_1, T_2, proton density) can be estimated using MR relaxometry techniques (e.g., Rosen et al., 1984; Jackson et al., 1993; Cho et al., 1997; Ogg and Steen, 1998; Steen et al., 2000). The probability of observing the image given the tissue parameters ($p(\mathbf{I}(\alpha)|\beta)$ in Equation 5) can then be estimated using a forward model for image formation based on the Bloch equations (Bloch, 1946). The term $p(\beta|W,f)$ can be computed from the manual labeling if the tissue parameters have been estimated on the manually labeled subjects. The conjunction of these two techniques—that of storing class conditional densities in terms of intrinsic tissue parameters as opposed to the somewhat arbitrary image intensities, together with a physics-based forward model of image formation—should allow for the construction of a segmentation procedure that is applicable across a wide variety of acquisition types. In the following, we will treat the more specific case in which class statistics are computed in terms of image intensities.

Atlas Construction

In general, two different approaches have been taken to the construction of anatomical atlases from neuroimaging data. The first is

to use an individual as a template, and either manually or automatically estimate a transformation that aligns each new dataset with the individual template (e.g., Talairach et al., 1967; Talairach and Tournoux, 1988; Van Essen and Drury, 1997). The alternative is to compile a probabilistic atlas based on the anatomy of a large number of subjects (e.g., Collins et al., 1994; Fox et al., 1994; Mazziotta et al., 1995; Thompson et al., 1997). Each of these approaches has strengths and weaknesses. The former allows one to represent anatomical structures at as fine a scale as the neuroimaging technology allows, but the atlas is then biased by the idiosyncrasies of the individual anatomy chosen as the template. The latter technique resolves this problem by averaging across anatomies, thus only retaining that which is common in the majority of subjects. Nevertheless, the cross-subject averaging removes potentially useful information in the atlas. Here, we preserve the advantages of each technique by using a group of subjects to construct an atlas that retains information about each anatomical class at each point in space. Given the atlas function f, and a group of N manually segmented subjects using the technique described in Kennedy et al. (1989), Rademacher et al. (1992), Filipek et al. (1994), and Caviness et al. (1996b), we first estimate the prior probability of anatomical class c occurring at each atlas location, independent of all other locations:

$$p(W(\mathbf{r}) = c) \approx \frac{\#\text{ of times class } c \text{ occurred at location } f(\mathbf{r})}{\#\text{ of voxels that map to } \mathbf{r} \text{ in the training set}} \quad (6)$$

Next, at each atlas location r, we model the intensity distribution of each class as a Gaussian, the parameters of which are again computed from the manually labeled training set:

$$\mu_c(\mathbf{r}) = \frac{1}{M}\sum_{i=1}^{M} \mathbf{I}_i(f(\mathbf{r})), \quad (7)$$

where I_i are the set of M (again potentially multi-spectral) images for which label c occurs at location $f(\mathbf{r})$ in the corresponding manually labeled image S_i (i.e., $S_i(f(r))=c$). The covariance matrix for class c at location r in the atlas is then given by:

$$\Sigma_c(\mathbf{r}) = \frac{1}{M-1}\sum_{i=1}^{M} (\mathbf{I}_i(f(\mathbf{r})) - \mu_c(\mathbf{r}))(\mathbf{I}_i(f(\mathbf{r})) - \mu_c(\mathbf{r}))^T \quad (8)$$

Thus, image intensity information is maintained separately for each anatomical class at every location in the atlas, obviating the need to average intensity information across classes. Finally, we also estimate the pairwise probability that anatomical class c_2 is the neighbor at $\mathbf{r}_i$ when anatomical class c_1 is the label at $\mathbf{r}$, for $\mathbf{r}_i \in N(\mathbf{r})$, where $N(\mathbf{r})$ is a neighborhood function of $\mathbf{r}$.

$$p(W(\mathbf{r}) = c_1 \mid W(\mathbf{r}_i) = c_2 \mathrm{m}, \mathbf{r}_i) \approx \left(\frac{\#\text{ of times class } c_2 \text{ occured at location } \mathbf{r}_i \text{ when class } c_1 \text{ occured at location } \mathbf{r}}{\#\text{ of times class } c_1 \text{ occured at location } \mathbf{r}} \right) \quad (9)$$

As before, this information is stored separately for each atlas location. It is important to note that the probabilities are stored separately for each pair of classes as well as for each neighborhood location $\mathbf{r}_i$. While it may seem that this would lead to combinatorial explosion and intractable memory requirements, in practice the space is sparse as relatively few configurations of anatomical labels actually occur. An example of the manual labeling that is the basis for the atlas is given in Figure 5, which displays coronal (left) and sagittal (right) views of an individual subject. The atlas information is currently stored at a spatial scale of 4 mm, a limit that is essentially imposed by available memory. While this is significantly less than the resolution of standard structural MRI images (on the order of 1 mm^3), it is sufficient to represent the information required for low-dimensional transforms such as the optimal linear transformation presented in the next section. Typical memory requirements for this resolution are on the order of 200 MB of RAM.

Optimal Linear Transform

The problem of computing the atlas function f in Equation 4 is known as the registration or correspondence problem. The goal of this procedure is to take a set of images and determine a function for each one that aligns anatomically corresponding locations across subjects. That is, one wishes in general to align hippocampus with hippocampus, amygdala with amygdala, etc. Unfortunately, direct information regarding the anatomical label of voxels in the input volumes is in general unavailable to the registration procedure. Thus, procedures have been developed in order to align the intensity images with the assumption that if locations with similar intensities are aligned everywhere in the brain, then anatomical correspondence will follow. The standard atlas used for registration purposes in the neuroimaging community is that of Talairach and Tournoux (Talairach et al., 1967; Talairach and Tournoux, 1988). The means for computing the alignment parameters necessary in order to associate atlas coordinates with those of each individual are varied and range from low-dimensional linear transforms to fluid deformations with tens of millions of degrees of freedom (Fox et al., 1984; Woods et al., 1992; Collins et al., 1994; Ashburner et al., 1997). The most common of these finds the linear transformation L (usually between 9 and 12 parameters) that minimizes the mean squared difference between a template image T, typically constructed by averaging multiple brains that have been manually aligned with the Talairach atlas (Collins et al., 1994), and an individual image I:

$$L = \arg\min_{L'} \iiint (T(\mathbf{r}) - I(L'\mathbf{r}))^2 d\mathbf{r} \quad (10)$$

This formulation assumes that the predicted image value at each location in the atlas can be well represented by a single mean. This assumption is clearly inaccurate; each image is made up of a variety of tissue types (e.g., gray matter, white matter, CSF, fat, etc.). Due to the large degree of variability in individual anatomy, most locations in the brain, as defined by a linear transform, may contain different tissue types in different subjects. This problem can be mitigated using high-dimensional nonlinear warps, which can achieve better alignment of like anatomical classes across subjects. Nevertheless, some residual variability is inevitable, in particular in cases in which lesions change the topology of the underlying anatomical structures.

For example, substantial ventricular enlargement is common in AD. Attempting to align an image from an AD patient with an atlas generated from a population of healthy subjects can result in large misalignments due to this neuroanatomical variability. One means of addressing this problem would be to include examples of subjects with various pathologies into the atlas. However, such an approach fails to resolve the underlying problem, as the resulting average would be representative of neither case. For example, ventricular locations in the AD patient containing CSF, which is quite dark on T_1-weighted MRI images, align with white matter regions in the normal subjects. Averaging the dark CSF with the bright white matter would result in an average tissue intensity in the target $T(\mathbf{r})$ that is closer to gray matter than either white matter or CSF—a tissue class that *never* occurs in these regions.

Since the ultimate goal of the registration procedure is to bring structures into alignment across subjects, it seems reasonable to seek a transformation that maximizes the probability of the segmentation, given the observed image. However, since both the segmentation and the alignment function are unknown, this would necessitate the maximization of the joint probability of f (or L) and W, as given by Equation 4, resulting in the maximum a posteriori (MAP) estimate of W and L. Techniques for simultaneously solving for two such parameters are well known in the machine vision and numerical optimization literature (Dempster et al., 1977; Wells et al., 1996; Zhang et al., 2001). Standard computer vision techniques frequently employ structures such as Gaussian pyramids (Burt and Adelson, 1983) in order to reduce execution time as well minimize sensitivity to local maxima in error functionals of this type. In this approach, the image is smoothed with successively narrower Gaussian blurring kernels, with the minimization (or maximization) of the error functional at each level being initialized with the final results from the previous (more blurred) level. However, averaging across anatomical classes that inevitably results from substantial spatial blurring is precisely what we are setting out to avoid. Thus, instead of using these techniques, we take a somewhat different approach to construct a means for finding the globally optimal L.

The difficulty in finding a good solution for Equation 7 stems from the dual ambiguity of trying to solve for two sets of mutually

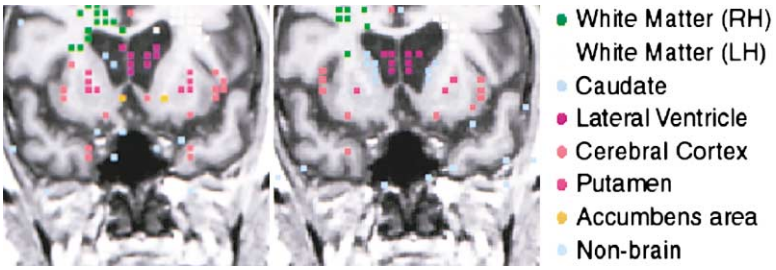

Figure 6. Atlas Samples before (Left) and after (Right) Optimization

Note that the sample size has been scaled up for visualization purposes.

dependent parameters. If a good alignment function were available, then the segmentation could be estimated. Conversely, if the segmentation were known, then an optimal atlas function could be obtained using it. In order to resolve this dilemma, we note two important points. The first is that there are many atlas regions in which the prior probability of an anatomical class occurring (given by Equation 6) approaches one. Thus, in these regions, the segmentation problem is eliminated, as only one anatomical class ever occurs there across normal and pathological populations. The second observation is that Equation 10 is dramatically overdetermined for linear or even nonlinear transforms with hundreds or thousands of parameters. These two facts suggest a solution to the alignment problem: instead of using the millions of voxels in an input image and attempting to find a MAP estimate of their classes together with an alignment transformation L, we select a much smaller subset of atlas location and find the alignment that maximizes the likelihood of these samples. Specifically, we choose a set of samples that fulfill two criteria. The first is that each sample must be the most probable class at that atlas location. Second, the prior probability for the sample must be close to the maximum prior probability for that class across the entire atlas ($p(W(\mathbf{r}) = c) > k \max(p(W(\mathbf{r}') = c))$, $\forall \mathbf{r}'$, where $k * .9$) Automatically selecting these points typically reduces the atlas size from several hundred thousand to several thousand. This then makes a global search of the parameter space for L tractable, obviating the problem of local maxima. Formally, we assume a classification W and find the 12 parameter affine transform L that maximizes:

$$L = \arg\max_{L'} = p(L'|W,\mathbf{I}) \propto p(\mathbf{I}|L',W)\, p(L') \qquad (11)$$

Equation 11 is maximized assuming $p(\mathbf{I}|L',W)$ is normally distributed with parameters computed using the atlas Equations 7 and 8, using an iterative global search along each of the rotation, scale, and translation axes, followed by a Davidson-Fletcher-Powell (DFP) numerical maximization using the gradient of (11) (Press et al., 1994). After this procedure has converged, we then remove the second constraint above and use the highest prior class at all locations in the atlas to guide a final DFP minimization. This final step is helpful in aligning the borders of the brain where priors are typically low. An example of this procedure is given in Figure 6, which shows a coronal slice through a T_1-weighted volume, with the atlas samples before (left) and after (right) maximization of (11). Note that the global search of the parameter space obviates the need to estimate the prior density $p(L')$, which has been shown to help registration procedures avoid local minima (Ashburner et al., 1997).

The accuracy of the registration procedure can be assessed by examining the number of anatomical classes that occur at each atlas location. In the limit of perfect registration, assuming the source and destination have the same topology (which may not be the case), all voxels would have only a single anatomical class ever occur at that atlas location. As the registration become less accurate, the number of anatomical classes occurring within an atlas voxel grows. A histogram showing the distribution of number of anatomical classes per voxel for an atlas generated from 7 manually labeled and linearly aligned volumes is given in Figure 7. As can be seen, the mean and mode are both approximately 3 classes per voxel.

Segmentation

Given the atlas and a linear alignment function L mapping an individual set of images $\mathbf{I}$ into the atlas, we now turn to the problem of assigning class labels to each voxel. Segmentation of MR images into labeled regions is a rich area of research in the image processing and biomedical engineering communities. Approaches to this difficult problem include fuzzy clustering (Suckling et al., 1999; Xu et al., 1999), deformable surfaces (Davatzikos and Bryan, 1996; Ghanei et al., 1998; Germond et al., 2000; MacDonald et al., 2000), region growing (H. Tang et al., 2000), model-based segmentation (Dale et al., 1999), level sets (Zeng et al., 1999), atlas-based segmentation (Collins and Evans, 1997; Sandor and Leahy, 1997), Gaussian mixture modeling (Wells et al., 1996; Teo et al., 1997; Kapur et al., 1998), nonparametric (Warfield, 1996; Held et al., 1997) and neural network classifiers (Wang et al., 1998; Magnotta et al., 1999). The vast majority of this work labels only a few classes, such as gray matter, white matter, non-brain, and CSF (Wells et al., 1996; Held et al., 1997; Teo et al., 1997; Kapur et al., 1998; Dale et al., 1999; Suckling et al., 1999; Ballester et al., 2000; Germond et al., 2000; H. Tang et al., 2000; Zhang et al., 2001). Partial-volume effects have been modeled by extending the number of classes to include voxels that contain more than one label (Wang et al., 1998; Ruan et al., 2000). In addition, low spatial frequency artifacts that frequently occur in magnetic resonance imaging have been detected and removed as part of the segmentation process using the expectation maximization algorithm (Wells et al., 1996; Held et al., 1997; Kapur et al., 1998; Ballester et al., 2000; Zhang et al., 2001). Prior use of Markov random fields has been limited to the stationary, spatially isotropic case, which essentially acts as a smoothness constraint on the segmentation (Leahy and Yan, 1991; Held et al., 1997; Kapur et al., 1998; Zhang et al., 2001). Some research has extended the small number of classes by applying atlas-based information to the output of the classification in order to label a few subcortical structures (Collins and Evans, 1997; Magnotta et al., 1999).

In the current work, we take a Bayesian approach, as this allows us to incorporate prior information that is necessary for the segmentation procedure. The prior information takes two forms. The first

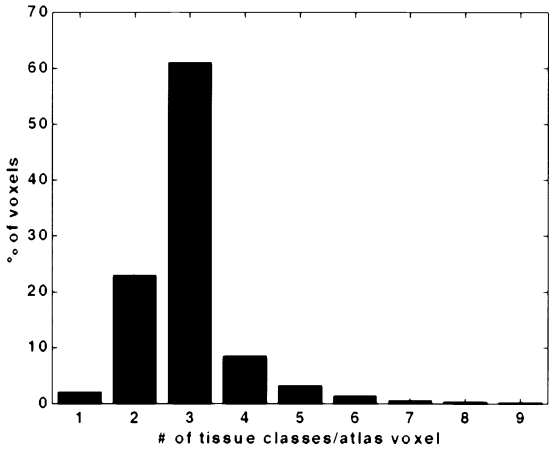

Figure 7. Histogram of the Number of Anatomical Classes That Occur at Each Voxel in an Atlas Made up of Seven Manually Labeled Volumes Aligned Using the Optimal Linear Transform Described Above

Figure 8. Relabeling Using the ICM Algorithm
From left to right: 0, 2, 4, 5, and 6 iterations.

makes use of the global spatial information provided by L and the atlas in order to express the probability that a given anatomical class occurs at a particular location in the atlas, independent of other information. The second encodes the local spatial relationship between anatomical classes, allowing information such "posterior amygdala is frequently superior to anterior hippocampus, but never inferior to it" to be automatically detected and incorporated into the segmentation.

Formally, we compute the maximum a posteriori (MAP) estimate of the segmentation W given an input image I, and the linear transform L computed as described above. The MAP estimate can be expressed as maximizing $p(W|I,L)$, the probability distribution of the segmentation given the observed image intensities. Using Bayes rule, we can relate this to the product of the probability of observing the image with the prior probability of a given spatial configuration of labels $p(W)$:

$$p(W|\mathbf{I},L) \propto p(\mathbf{I}|W,L)\, p(W) \tag{12}$$

Assuming the noise at each voxel is independent from all other voxels in the image, we can rewrite $p(I|W,L)$ as the product of the distribution at each voxel over the image domain R:

$$p(\mathbf{I}|W,L) = \prod_{\mathbf{r}\in R} p(\mathbf{I}(L\mathbf{r})|W(\mathbf{r})) \tag{13}$$

Note that in the more general case in which the spatial correlation structure of the noise is constant across space, the equality in (13) should be replaced by proportionality (Worsley et al., 1992; Thompson et al., 1996), which does not affect any of the subsequent derivations. The intensity distribution of each class at each location in the atlas is modeled as a Gaussian, the mean vector $\mu_c(\mathbf{r})$ and covariance matrix $\Sigma_c(\mathbf{r})$ of which are computed using Equations 7 and 8. The probability of observing the image intensity at $I(L\mathbf{r})$ is then expressed as:

$$p(\mathbf{I}(L\mathbf{r})|W(\mathbf{r}) = c) = \frac{1}{|\Sigma_c(\mathbf{r})|^{1/2}\sqrt{2\pi}}$$
$$\times \exp(-0.5(\mathbf{I}(L\mathbf{r}) - \mu_c(\mathbf{r}))^T \Sigma_c(\mathbf{r})^{-1}(\mathbf{I}(L\mathbf{r}) - \mu_c(\mathbf{r})) \tag{14}$$

All that remains is to find an expression for the prior probability of a given classification W. Here we assume that the spatial distribution of labels can be well approximated by an anisotropic nonstationary Markov random field. This allows one to encode prior information about the relationship between labels as a function of location within the brain (i.e., nonstationary), as well as with local direction (i.e., anisotropic). Formally, the Markov assumption can be expressed as:

$$p(W(\mathbf{r})|W(R - \{\mathbf{r}\})) = p(W(\mathbf{r})|W(\mathbf{r}_1),W(\mathbf{r}_2),\ldots,W(\mathbf{r}_K)),\mathbf{r}_i \in N(\mathbf{r}) \tag{15}$$

That is, the prior probability of a label at a given voxel $\mathbf{r}$ is only influenced by the labels within some neighborhood of $\mathbf{r}$. The locality restriction imposed by the Markov model permits the probability of the entire segmentation to be written in terms of neighborhood or *clique* potentials $V_c(W)$ via the Hammersley-Clifford theorem (Besag, 1974). That is, the probability $p(W)$ can be equivalently characterized by a Gibbs distribution:

$$p(W) = \frac{1}{Z} e^{-U(W)}, \tag{16}$$

where Z is a normalizing constant and will be dropped in the following, and $U(W)$ is an energy function that can be written in the form:

$$U(W) = \sum_c V_c(W) \tag{17}$$

The clique potentials $V_c(W)$ encode the energy associated with a certain configuration of labels within the c^{th} clique. Choosing $V_c(W)$ to be $-log(p(W(\mathbf{r})|W(\mathbf{r}_1),W(\mathbf{r}_2)\ldots W(\mathbf{r}_K))$, where $\mathbf{r}$ is the central voxel of the c^{th} clique, allows us to write the probability of the entire segmentation as the product of the probability of the label at each voxel, given its neighborhood:

$$p(W) = \prod_{\mathbf{r}\in R} p(W(\mathbf{r})|W(\mathbf{r}_1),W(\mathbf{r}_2),\ldots,W(\mathbf{r}_K)),\mathbf{r}_i \in N(\mathbf{r}) \tag{18}$$

Using Bayes rule, we can rewrite this as:

$$p(W) \propto \prod_{\mathbf{r}\in R} p(W(\mathbf{r}))\, p(W(\mathbf{r}_1),W(\mathbf{r}_2),\ldots,W(\mathbf{r}_K)|W(\mathbf{r})),\mathbf{r}_i \in N(\mathbf{r}) \tag{19}$$

Equation 19 allows the probability of a given label to be modulated by any configuration of neighboring labels. While this would be extremely useful, it is unfortunately not computationally tractable to implement, as one would need to compute separate prior probabilities for every combination of neighboring labels that occur. Instead, we make the simplifying assumption that only the first order conditional dependence is important. That is, that the dependence of a label on its neighbors can be expressed as the product of the probability given each of the neighbors:

$$p(W(\mathbf{r})|W(\mathbf{r}_1),W(\mathbf{r}_2),\ldots,W(\mathbf{r}_K)) = \prod_{\mathbf{r}_i\in N(\mathbf{r})} p(W(\mathbf{r})|W(\mathbf{r}_i),\mathbf{r}_i), \tag{20}$$

where again we have explicitly included the dependence on the neighbor location $\mathbf{r}_i$ to emphasize that the probability densities are maintained separately for each neighbor position in $N(\mathbf{r})$. Using this assumption, we arrive at an expression for the prior probability of the full segmentation:

$$p(W) \propto \prod_{\mathbf{r}\in N} p(W(\mathbf{r}))\prod_{i=1}^{K} p(W(\mathbf{r}_i)|W(\mathbf{r}),\mathbf{r}_i) \tag{21}$$

Equation 21 allows two types of prior information to be incorporated into the segmentation procedure. The approximate location a neuroanatomical structure may occupy within the brain is given by $p(W(\mathbf{r}))$, which is computed and stored in the atlas using Equation 6. The local relationship between anatomical classes is encoded in $p(W(\mathbf{r}_i)|W(\mathbf{r}),\mathbf{r}_i)$ using Equation 9. We currently let the neighborhood function $N(\mathbf{r})$ include the 6 voxels in the positive and negative cardinal directions at each location in the atlas space. This allows the segmentation procedure to automatically extract and apply information such as "if a voxel is labeled hippocampus then the probability of the voxel inferior to it being labeled amygdala is low."

Directly computing the global MAP estimate of W in Equation 12 using the Markov model of Equation 21 is computationally intractable. Instead, we employ the iterated conditional modes (ICM) algorithm proposed by Besag (1986). In this approach, the segmentation is initialized with the MAP estimate assuming $p(W(\mathbf{r}_i)|W(\mathbf{r}),\mathbf{r}_i)$ is uniform, as no label has yet been assigned to each voxel. The segmentation is then sequentially updated at each location by computing the label $W(\mathbf{r})$ that maximizes the conditional posterior probability $p(W(\mathbf{r})|W(\mathbf{r}_i),I,\mathbf{r}_i)$:

$$W(\mathbf{r}) = \arg\max_{c} p(W(\mathbf{r}) = c | W(\mathbf{r}_i), \mathbf{I}(L\mathbf{r}), \mathbf{r}_i) =$$
$$p(\mathbf{I}(L\mathbf{r}) | W(\mathbf{r}) = c)\, p(W(\mathbf{r}) = c) \prod_{i=1}^{K} p(W(\mathbf{r}_i) | W(\mathbf{r}) = c, \mathbf{r}_i) \quad (22)$$

Equation 22 is then iteratively applied until no voxels are changed. Snapshots of the evolution of this procedure on a coronal slice are given in Figure 8. As can be seen, convergence is rapid, resulting in a procedure that requires approximately 10 min on a 1 GHz Pentium III. A small amount of pre- and postprocessing is carried out on the results of the labeling, as detailed in the next two sections.

Preprocessing

Before applying the iterative relabeling based on Markov random field modeling of the segmentation as outlined above, a few preprocessing steps are performed. Note that all changes made during this processing are provisional, and may be corrected by the ICM relabeling. Due to the extreme thinness of the base of the parahippocampal gyrus that separates cortical gray matter from hippocampus, a few rule-based changes are made in an effort to keep hippocampal labels superior to the parahippocampal gyrus. This can be thought of as incorporating a set of combinatorial Gibbs priors into the segmentation procedure. Unfortunately, a direct incorporation of all pairwise tissue occurrences is intractable due to combinatorial explosion. Specifically, we relabel all voxels that have been labeled as hippocampus that have cortical white matter within 2 mm both left and right or both anterior and posterior of them as cortical gray matter. Note that this certainly results in the removal of some accurately labeled voxels. Nevertheless, this is not a cause for concern, as the Gibbs priors typically provide a sufficiently strong set of constraints in order to change the label of these voxels back to hippocampus. The intent is simply to remove incorrect labels inferior to parahippocampal white matter, as otherwise islands of incorrect hippocampal labels may survive the relabeling in these regions. These rules can be seen as incorporating higher order Markov relationships into the classification procedure, that is, expanding the neighborhood size and allowing the probability of one label to be dependent on the joint occurrence of two or more other labels.

Postprocessing

The postprocessing is carried out immediately following the complete convergence of the ICM relabeling and consists of two steps. The first is a maximum likelihood relabeling that reflects the inherent uncertainty of the coordinate system. Specifically, the neighbor of each label is reclassified using a maximum likelihood classifier (i.e., ignoring prior information). The second step is to build connected components representations of all labels, and to reclassify connected components whose volume is less than 10% of the total volume for that label as the most likely label of all neighboring voxels. This removes small "islands" of labels that are disconnected from the main body. The inclusion of larger neighborhood sizes into the Markov random field model would obviate the need for this type of procedure, but is also unfortunately not tractable in terms of memory requirements.

Appendix

Labels for whole brain segmentation
Left/Right Cerebral White Matter
Left/Right Cerebral Cortex
Left/Right Lateral Ventricle
Left/Right Inferior Lateral Ventricle
Left/Right Cerebellum White Matter
Left/Right Cerebellum Cortex
Left/Right Thalamus
Left/Right Caudate
Left/Right Putamen
Left/Right Pallidum
Left/Right Hippocampus
Left/Right Amygdala
Left/Right Lesion
Left/Right Accumbens area
Left/Right Ventral Diencephalon
Left/Right vessel (non-specific).
Third Ventricle
Fourth Ventricle
Brain Stem
Cerebrospinal Fluid
Unknown (not brain)

Acknowledgments

This Human Brain Project/Neuroinformatics research is funded jointly by the National Institute of Neurological Disorders and Stroke, the National Institute of Mental Health, and the National Cancer Institute (R01-NS39581, R01-NS34189). Further support was provided by the National Center for Research Resources (P41-RR14075 and R01-RR13609). We thank Maureen Glessner for extensive testing of the proposed algorithms. We also thank Sean Marrett for helpful comments on the manuscript.

Received July 9, 2001; revised December 6, 2001.

References

Albert, M.S. (1996). Cognitive and neurobiologic markers of early Alzheimer disease. Proc. Natl. Acad. Sci. USA *93*, 13547–13551.

Ashburner, J., Neelin, P., Collins, D.L., Evans, A.C., and Friston, K.J. (1997). Incorporating prior knowledge into image registration. Neuroimage *6*, 344–352.

Bajcsy, R., Lieberson, R., and Reivich, M. (1983). A computerized system for the elastic matching of deformed radiographic images to idealized atlas images. J. Comput. Assist. Tomogr. *5*, 618–625.

Ball, M., Fisman, M., Hachinski, V., Blume, W., Fox, A., Kral, V., Kirshen, A., Fox, H., and Merskey, H. (1985). A new definition of Alzheimer's disease: a hippocampal dementia. Lancet *1*, 14–16.

Ballester, M.A.G., Zisserman, A., and Brady, M. (2000). Segmentation and measurement of brain structures in MRI including confidence bounds. Med. Image Anal. *4*, 189–200.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). J. R. Stat. Soc. [Ser B] *36*, 192–326.

Besag, J. (1986). On the statistical analysis of dirty pictures (with discussion). J. R. Stat. Soc. [Ser B] *48*, 259–302.

Bloch, F. (1946). Nuclear Induction. Phys. Rev. *70*, 460–474.

Bookstein, F.L. (1989). Principal warps: thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence *11*, 567–585.

Braak, H., and Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. Acta Neuropathol. (Berl.) *82*, 239–259.

Breiter, H.C., Filipek, P.A., Kennedy, D.N., Baer, L., Pitcher, D.A., Olivares, M.J., Renshaw, P.F., and Caviness, V.S., Jr. (1994). Retrocallosal white matter abnormalities in patients with obsessive-compulsive disorder. Arch. Gen. Psychiatry *51*, 663–664.

Burt, P.J., and Adelson, E.H. (1983). The Laplacian pyramid as a compact image code. IEEE Transactions on Communications *9*, 532–540.

Caviness, V.S., Filipek, P.A., and Kennedy, D.N. (1989). Magnetic resonance technology in human brain science: Blueprint for a program based upon morphometry. Brain Dev. *11*, 1–13.

Caviness, V.S., Jr., Filipek, P.A., and Kennedy, D.N. (1992). Quantitative magnetic resonance imaging and studies of degenerative diseases of the developing human brain. Brain Dev. Suppl. *14*, S80–S85.

Caviness, V.S., Jr., Kennedy, D.N., Richelme, C., Rademacher, J., and Filipek, P.A. (1996a). The human brain age 7-11 years: a volumetric analysis based on magnetic resonance images. Cereb. Cortex *6*, 726–736.

Caviness, V.S., Jr., Meyer, J., Makris, N., and Kennedy, D.N. (1996b). MRI-based topographic parcellation of the human neocortex: an anatomically specified method with estimate of reliability. J. Cogn. Neurosci. *8*, 566–587.

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. Magn. Reson. Imaging *15*, 1133–1143.

Christensen, G.E., Rabbitt, R.D., and Miller, M.I. (1996). Deformable

templates using large deformation kinematics. IEEE Transactions on Image Processing *5*, 1435–1447.

Collins, D., and Evans, A. (1997). Animal: validation and applications of non-linear registration-based segmentation. International Journal of Pattern Recognition and Artificial Intelligence *11*, 1271–1294.

Collins, D.L., Neelin, P., Peters, T.M., and Evans, A.C. (1994). Data in standardized talairach space. J. Comput. Assist. Tomogr. *18*, 192–205.

Collins, D., Dai, W., Peters, T., and Evans, A. (1995). Automatic 3D model-based neuroanatomical segmentation. Hum. Brain Mapp. *3*, 190–208.

Convit, A., De Leon, M.J., Tarshish, C., De Santi, S., Tsui, W., Rusinek, H., and George, A. (1997). Specific hippocampal volume reductions in individuals at risk for Alzheimer's disease. Neurobiol. Aging *18*, 131–138.

Dale, A.M., Fischl, B., and Sereno, M.I. (1999). Cortical surface-based analysis I: segmentation and surface reconstruction. Neuroimage *9*, 179–194.

Daly, E., Zaitchik, D., Copeland, M., Schmahmann, J., Gunther, J., and Albert, M. (2000). Predicting conversion to Alzheimer disease using standardized clinical information. Arch. Neurol. *57*, 675–680.

Davatzikos, C., and Bryan, R.N. (1996). Using a deformable surface model to obtain a shape representation of the cortex. IEEE Transactions on Medical Imaging *15*, 785–795.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. [Ser B] *39*, 1–38.

De Toledo-Morrell, L., Goncharova, I., Dickerson, B., Wilson, R.S., and Bennett, D.A. (2000). From healthy aging to early Alzheimer's disease: in vivo detection of entorhinal cortex atrophy. Ann. N Y Acad. Sci. *911*, 240–253.

Double, K., Halliday, G., Dril, J., Harasty, J., Cullen, K., Brooks, W.S., Creasey, H., and Broe, G.A. (1996). Topography of brain atrophy during normal aging and Alzheimer's disease. Neurobiol. Aging *17*, 513–521.

Filipek, P.A., Richelme, C., Kennedy, D.N., and Caviness, V.S., Jr. (1994). The young adult human brain: an MRI-based morphometric analysis. Cereb. Cortex *4*, 344–360.

Fischl, B., Sereno, M.I., Tootell, R.B.H., and Dale, A.M. (1999). High-resolution inter-subject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. *8*, 272–284.

Forstl, H., Zerfass, R., Geiger-Kabisch, C., Sattel, H., Besthorn, C., and Hentschel, F. (1995). Brain atrophy in normal aging and Alzheimer's disease. Volumetric discrimination and clinical correlations. Br. J. Psychiatry *167*, 739–746.

Fox, P.T., Perlmutter, J.S., and Raichle, M.E. (1984). Stereotactic method for determining anatomical localization in physiological brain images. J. Cereb. Blood Flow Metab. *4*, 634.

Fox, P.T., Mikiten, S., Davis, G., and Lancaster, J.L. (1994). BrainMap: A Database of Human Functional Brain Mapping. Functional Neuroimaging. R.W. Thatcher, M. Hallett, T. Zeffiro, E.R. John, and M. Huerta, eds. (San Diego, CA: Academic Press), pp. 95–106.

Frisoni, G.B., Beltramello, A., Weiss, C., Geroldi, C., Bianchetti, A., and Trabucchi, M. (1996). Linear measures of atrophy in mild Alzheimer disease. Am. J. Neuroradiol. *17*, 913–923.

Gee, J.C., Haynor, D.R., Reivich, M., and Bajcsy, R. (1994). Finite element approach to warping of brain images. Medical Imaging 1994: Image Processing. M.H. Loew, ed. (Bellingham: SPIE).

Geman, S., and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence *6*, 721–741.

Germond, L., Dojat, M., Taylor, C., and Garbay, C. (2000). A cooperative framework for segmentation of MRI brain scans. Artif. Intell. Med. *20*, 77–93.

Ghanei, A., Soltanian-Zadeh, H., and Windham, J.P. (1998). A 3D deformable surface model for segmentation of objects from volumetric data in medical images. Comput. Biol. Med. *28*, 239–253.

Goldstein, J.M., Goodman, J.M., Seidman, L.J., Kennedy, D.N., Makris, N., Lee, H., Tourville, J., Caviness, V.S., Jr., Faraone, S.V., and Tsuang, M.T. (1999). Cortical abnormalities in schizophrenia identified by structural magnetic resonance imaging. Arch. Gen. Psychiatry *56*, 537–547.

H. Tang, E.X., Wu, Q., Ma, Y., Gallagher, D., Perera, G.M., and Zhuang, T. (2000). MRI brain image segmentation by multi-resolution edge detection and region selection. Computerized Medical Imaging and Graphics *24*, 349–357.

Halliday, G.M., McRitchie, D.A., Macdonald, V., Double, K.L., Trent, R.J., and McCusker, E. (1998). Regional specificity of brain atrophy in Huntington's disease. Exp. Neurol. *154*, 663–672.

Held, K., Kops, E.R., Krause, B.J., W., W.M., III, Kikinis, R., and Muller-Gartner, H.-W. (1997). Markov random field segmentation of brain MR images. IEEE Transactions on Medical Imaging *16*, 876–886.

Hughes, C., Berg, L., Danziger, W., Coben, L., and Martin, R. (1982). A new clinical scale for the staging of dementia. Br. J. Psychiatry *140*, 566–572.

Jack, C.R., Jr., Petersen, R., O'Brien, P., and Tangalos, E. (1992). MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. Neurology *42*, 183–188.

Jack, C.R., Petersen, R.C., Xu, Y.C., Waring, S.C., O'Brien, P.C., Tangalos, E.G., Smith, G.E., Ivnik, R.J., and Kokmen, E. (1997). Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. Neurology *49*, 786–790.

Jack, C.R., Jr., Petersen, R.C., Xu, Y.C., O'Brien, P.C., Smith, G.E., Ivnik, R.J., Boeve, B.F., Waring, S.C., Tangalos, E.G., and Kokmen, E. (1999). Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. Neurology *52*, 1397–1403.

Jack, C.R., Jr., Petersen, R.C., Xu, Y., O'Brien, P.C., Smith, G.E., Ivnik, R.J., Boeve, B.F., Tangalos, E.G., and Kokmen, E. (2000). Rates of hippocampal atrophy correlate with change in clinical status in aging and AD. Neurology *55*, 484–489.

Jackson, G., Connelly, A., Duncan, J., Grunewald, R., and Gadian, D. (1993). Detection of hippocampal pathology in intractable partial epilepsy: Increased sensitivity with quantitative magnetic resonance T2 relaxometry. Neurology *43*, 1793–1799.

Jenike, M.A., Breiter, H.C., Baer, L., Kennedy, D.N., Savage, C.R., Olivares, M.J., O'Sullivan, R.L., Shera, D.M., Rauch, S.L., Keuthen, N., et al. (1996). Cerebral structural abnormalities in obsessive-compulsive disorder. A quantitative morphometric magnetic resonance imaging study. Arch. Gen. Psychiatry *53*, 625–632.

Jernigan, T.L., Salmon, D.P., Butters, N., and Hesselink, J.R. (1991). Cerebral structure on MRI, Part II: Specific changes in Alzheimer's and Huntington's diseases. Biol. Psychiatry *29*, 68–81.

Johnson, K., Jones, K., Holman, B., Becker, J., Spiers, P., Satlin, A., and Albert, M. (1998). Preclinical prediction of Alzheimer's disease using SPECT. Neurology *50*, 1563–1571.

Kapur, T., Grimson, W.E.L., Wells, W.M., and Kikinis, R. (1998). Enhanced Spatial Priors for Segmentation of Magnetic Resonance Imagery. Medical Image Computing and Computer-Assisted Intervention, Massachusetts Institute of Technology, Cambridge, MA .

Kaye, J.A., Swihart, T., Howieson, D., Dame, A., Moore, M.M., Karnos, T., Camicioli, R., Ball, M., Oken, B., and Sexton, G. (1997). Volume loss of the hippocampus and temporal lobe in healthy elderly persons destined to develop dementia. Neurology *48*, 1297–1304.

Kennedy, D.N., Filipek, P.A., and Caviness, V.S. (1989). Anatomic segmentation and volumetric calculations in nuclear magnetic resonance imaging. IEEE Transactions on Medical Imaging *8*, 1–7.

Killiany, R., Moss, M., Albert, M., Sandor, T., Tieman, J., and Jolesz, F. (1993). Temporal lobe regions on magnetic resonance imaging identify patients with early Alzheimer's disease. Arch. Neurol. *50*, 949–954.

Killiany, R.J., Gomez-Isla, T., Moss, M., Kikinis, R., Sandor, T., Jolesz, F., Tanzi, R., Jones, K., Hyman, B.T., and Albert, M.S. (2000). The use of structural MRI to predict who will get Alzheimer's disease. Ann. Neurol. *47*, 430–439.

Laakso, M.P., Soininen, H., Partanen, K., Helkala, E.L., Hartikainen, P., Vainio, P., Hallikainen, M., Hanninen, T., and Riekkinen, P.J., Sr. (1995). Volumes of hippocampus, amygdala and frontal lobes in the

MRI-based diagnosis of early Alzheimer's disease: correlation with memory functions. J. Neural Transm. Park. Dis. Dement. Sect. *9*, 73–86.

Leahy, R.M., and Yan, X. (1991). Incorporation of Anatomical MR Data for Improved Functional Imaging with PET. Information Processing in Medical Imaging, 12th international conference, Wye, UK.

Lehtovirta, M., Laakso, M.P., Soininen, H., Helisalmi, S., Mannermaa, A., Helkala, E.L., Partanen, K., Ryynanen, M., Vainio, P., Hartikainen, P., et al. (1995). Volumes of hippocampus, amygdala and frontal lobe in Alzheimer patients with different apolipoprotein E genotypes. Neuroscience *67*, 65–72.

Luxenberg, J.S., Haxby, J.V., Creasey, H., Sundaram, M., and Rapoport, S.I. (1987). Rate of ventricular enlargement in dementia of the Alzheimer type correlates with rate of neuropsychological deterioration. Neurology *37*, 1135–1140.

MacDonald, D., Kabani, N., Avis, D., and Evans, A.C. (2000). Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. Neuroimage *12*, 340–356.

Magnotta, V.A., Heckel, D., Andreasen, N.C., Cizadlo, T., Corson, P.W., Erhardt, J.C., and Yuh, W.T.C. (1999). Measurement of brain structures with artificial neural networks: two- and three-dimensional applications. Radiology *211*, 781–790.

Makris, N., Worth, A.J., Sorensen, A.G., Papadimitriou, G.M., Wu, O., Reese, T.G., Wedeen, V.J., Davis, T.L., Stakes, J.W., Caviness, V.S., et al. (1997). Morphometry of in vivo human white matter association pathways with diffusion-weighted magnetic resonance imaging. Ann. Neurol. *42*, 951–962.

Mazziotta, J.C., Toga, A.W., Evans, A.C., Fox, P., and Lancaster, J. (1995). A probabilistic atlas of the human brain: theory and rationale for its development. Neuroimage *2*, 297–318.

Miller, M.I., Christensen, G.E., Amit, Y., and Grenander, U. (1993). A mathematical textbook of deformable neuroanatomies. Proc. Natl. Acad. Sci. USA *90*, 11944–11948.

Mizuno, K., Wakai, M., Takeda, A., and Sobue, G.S.F.-ó. (2000). Medial temporal atrophy and memory impairment in early stage of Alzheimer's disease: an MRI volumetric and memory assessment study. J. Neurol. Sci. *173*, 18–24.

Murphy, D., DeCarli, C., Daly, E., Gillette, J., McIntosh, A., Haxby, J., Teichberg, D., Schapiro, M., Rapoport, S., and Horwitz, B. (1993). Volumetric magnetic resonance imaging in men with dementia of the Alzheimer type: correlations with disease severity. Biol. Psychiatry *34*, 612–621.

Ogg, R.J., and Steen, R.G. (1998). Age-related changes in brain T1 are correlated with iron concentration. Magn. Reson. Med. *40*, 749–753.

O'Sullivan, R.L., Rauch, S.L., Breiter, H.C., Grachev, I.D., Baer, L., Kennedy, D.N., Keuthen, N.J., Savage, C.R., Manzo, P.A., Caviness, V.S., et al. (1997). Reduced basal ganglia volumes in trichotillomania measured via morphometric magnetic resonance imaging. Biol. Psychiatry *42*, 39–45.

Pedersen, N., Miller, B., Loebach, W.J., Vallo, J., Toga, A., Knutson, N., Mehringer, C., Small, G., and Gatz, M. (1999). Neuroimaging findings in twins discordant for Alzheimer's disease. Dement. Geriatr. Cogn. Disord. *10*, 51–58.

Powell, A., Mezrich, R., Coyne, A., Loesberg, A., and Keller, I. (1993). Convex third ventricle: a possible sign for dementia using MRI. J. Geriatr. Psychiatry Neurol. *6*, 217–221.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (1994). Numerical Recipes in C (Cambridge, MA: Cambridge University Press).

Puri, B.K., Richardson, A.J., Oatridge, A., Hajnal, J.V., and Saeed, N. (1999). Cerebral ventricular asymmetry in schizophrenia: a high resolution 3D magnetic resonance imaging study. Int. J. Psychophysiol. *34*, 207–211.

Rademacher, J., Galaburda, A.M., Kennedy, D.N., Filipek, P.A., and Caviness, V.S. (1992). Human cerebral cortex: localization, parcellation and morphometry with magnetic resonance imaging. J. Cogn. Neurosci. *4*, 352–374.

Rosen, B.R., Pykett, I.L., and Brady, T.J. (1984). Spin lattice relaxation time measurements in two-dimensional nuclear magnetic resonance imaging: Corrections for plane selection and pulse sequence. J. Comput. Assist. Tomogr. *8*, 195–199.

Ruan, S., Jaggi, C., Xue, J., Fadili, J., and Bloyet, D. (2000). Brain tissue classification of magnetic resonance images using partial volume modeling. IEEE Transactions on Medical Imaging *19*, 1179–1187.

Sandor, S., and Leahy, R. (1997). Surface-based labeling of cortical anatomy using a deformable atlas. IEEE Transactions on Medical Imaging *16*, 41–54.

Seidman, L.J., Faraone, S.V., Goldstein, J.M., Goodman, J.M., Kremen, W.S., Toomey, R., Tourville, J., Kennedy, D., Makris, N., Caviness, V.S., et al. (1999). Thalamic and amygdala-hippocampal volume reductions in first-degree relatives of patients with schizophrenia: an MRI-based morphometric analysis. Biol. Psychiatry *46*, 941–954.

Small, S.A., Nava, A.S., Perera, G.M., Delapaz, R., and Stern, Y. (2000). Evaluating the function of hippocampal subregions with high-resolution MRI in Alzheimer's disease and aging. Microsc. Res. Tech. *51*, 101–108.

Steen, R.G., Reddick, W.E., and Ogg, R.J. (2000). More than meets the eye: significant regional heterogeneity in human cortical T1. Magn. Reson. Imaging *18*, 361–368.

Suckling, J., Sigmundsson, T., Greenwood, K., and Bullmore, E.T. (1999). A modified fuzzy clustering algorithm for operator independent brain tissue classification. Magn. Reson. Imaging *17*, 1065–1076.

Talairach, J., Szikla, G., Tournoux, P., Prosalentis, A., Bordas-Ferrier, M., Covello, L., Iacob, M., and Mempel, E. (1967). Atlas d'Anatomie Stereotaxique du Telencephale (Paris: Masson).

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

Teo, P.C., Sapiro, G., and Wandell, B.A. (1997). Creating connected representations of cortical gray matter for functional MRI visualization. IEEE Transactions on Medical Imaging *16*, 852–863.

Thompson, P.M., Schwartz, C., and Toga, A.W. (1996). High-resolution random mesh algorithms for creating a probabilistic 3D surface atlas of the human brain. Neuroimage *3*, 19–34.

Thompson, P.M., MacDonald, D., Mega, M.S., Holmes, C.J., Evans, A.C., and Toga, A.W. (1997). Detection and mapping of abnormal brain structure with a probabilistic atlas of cortical surfaces. J. Comput. Assist. Tomogr. *21*, 567–581.

Thompson, P., Woods, R., Mega, M., and Toga, A. (2000). Mathematical/computational challenges in creating deformable and probabilistic atlases of the human brain. Hum. Brain Mapp. *9*, 81–92.

Van Essen, D.C., and Drury, H.A. (1997). Structural and functional analyses of human cerebral cortex using a surface-based atlas. J. Neurosci. *17*, 7079–7102.

Vannier, M.W., Miller, M.I., and Grenander, U. (1994). Modeling and data structures for registration of a brain atlas of mutimodality images. Functional Neuroimaging - Technical Foundations. R.W. Thatcher, M. Hallett, E.R. John, and E. Huerta, eds. (New York: Academic Press), pp. 217–221.

Vonsattel, J.P., and DiFiglia, M. (1998). Huntington disease. J. Neuropathol. Exp. Neurol. *57*, 369–384.

Wang, Y., Adah, T., Kung, S.-Y., and Szabo, Z. (1998). Quantification and segmentation of brain tissues from MR images: a probabilistic neural network approach. IEEE Transactions on Image Processing *7*, 1165–1181.

Warfield, S. (1996). Fast k-NN classification for multichannel image data. Pattern Recognition Letters *17*, 713–721.

Wells, W., Grimson, W., Kikinis, R., and Jolesz, F. (1996). Adaptive segmentation of MRI data. IEEE Transactions on Medical Imaging *15*, 429–442.

Wolf, H., Grunwald, M., Kruggel, F., Riedel-Heller, S.G., Angerhofer, S., Hojjatoleslami, A., Hensel, A., Arendt, T., and Gertz, H.-J. (2001). Hippocampal volume discriminates between normal cognition; questionable and mild dementia in the elderly. Neurobiol. Aging *22*, 177–186.

Woods, R.P., Cherry, S.R., and Mazziotta, J.C. (1992). Rapid automated algorithm for aligning and reslicing PET images. J. Comput. Assist. Tomogr. *16*, 620–633.

Woods, R.P., Grafton, S.T., Holmes, C.J., Cherry, S.R., and Mazziotta, J.C. (1998). Automated image registration: II. Inter-subject validation of linear and nonlinear models. J. Comput. Assist. Tomogr. *22*, 155–165.

Worsley, K.J., Evans, A.C., Marrett, S., and Neelin, P. (1992). A three dimensional statistical analysis for CBF activation studies in human brain. J. Cereb. Blood Flow Metab. *12*, 900–918.

Xu, C., Pham, D.L., Rettmann, M.E., Yu, D.N., and Prince, J.L. (1999). Reconstruction of the human cerebral cortex from magnetic resonance images. IEEE Transactions on Medical Imaging *18*, 467–480.

Zeng, X., Staib, L.H., Schultz, R.T., and Duncan, J.S. (1999). Segmentation and measurement of the cortex from 3-d MR images using coupled-surfaces propagation. IEEE Transactions on Medical Imaging *18*, 927–937.

Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of Brain MR Images through hidden markov random field model and the expectation maximization algorithm. IEEE Transactions on Medical Imaging *20*, 45–57.