NEUROIMAGE **6,** 13–25 (1997)
ARTICLE NO. NI970274

An Automated Registration Algorithm for Measuring MRI Subcortical Brain Structures[1,2]

Dan V. Iosifescu,* Martha E. Shenton,* Simon K. Warfield,† Ron Kikinis,‡ Joachim Dengler,§ Ferenc A. Jolesz,‡ and Robert W. McCarley*

**Clinical Neuroscience Division, Laboratory of Neuroscience, Department of Psychiatry, Harvard Medical School, Brockton VAMC, Brockton, Massachusetts 02401; ‡ Surgical Planning Laboratory, Department of Radiology, MRI Division, Brigham and Women's Hospital, Boston, Massachusetts 02115; †School of Computer Science and Engineering, The University of New South Wales, Sydney, Australia 2052; and §German Cancer Research Institute, Heidelberg, Germany 69120*

Received November 26, 1996

An automated registration algorithm was used to elastically match an anatomical magnetic resonance (MR) atlas onto individual brain MR images. Our goal was to evaluate the accuracy of this procedure for measuring the volume of MRI brain structures. We applied two successive algorithms to a series of 28 MR brain images, from 14 schizophrenia patients and 14 normal controls. First, we used an automated segmentation program to differentiate between white matter, cortical and subcortical gray matter, and cerebrospinal fluid. Next, we elastically deformed the atlas segmentation to fit the subject's brain, by matching the white matter and subcortical gray matter surfaces. To assess the accuracy of these measurements, we compared, on all 28 images, 11 brain structures, measured with elastic matching, with the same structures traced manually on MRI scans. The similarity between the measurements (the relative difference between the manual and the automated volume) was 97% for whole white matter, 92% for whole gray matter, and on average 89% for subcortical structures. The relative spatial overlap between the manual and the automated volumes was 97% for whole white matter, 92% for whole gray matter, and on average 75% for subcortical structures. For all pairs of structures rendered with the automated and the manual method, Pearson correlations were between $r = 0.78$ and $r = 0.98$ ($P < 0.01$, $N = 28$), except for globus pallidus, where $r = 0.55$ (left) and $r = 0.44$ (right) ($P < 0.01$, $N = 28$). In the schizophrenia group, compared to the controls, we found a 16.7% increase in MRI volume for the basal ganglia (i.e., caudate nucleus, putamen, and globus pallidus), but no difference in total gray/white matter volume or in thalamic MR volume. This finding reproduces previously reported results, obtained in the same patient population with manually drawn structures, and suggests the utility/efficacy of our automated registration algorithm over more labor-intensive manual tracings.

INTRODUCTION

The quantitative analysis of magnetic resonance imaging (MRI) data to examine anatomical brain structures is a fundamental component in the assessment of structural brain abnormalities (e.g., Shenton *et al.*, 1992), in mapping functional activation onto human anatomy (e.g., Dhawan *et al.*, 1992), and in computer-assisted neurosurgery (e.g., Kikinis *et al.*, 1996). Such analyses are important in structural MRI studies of schizophrenia, dementia, and other brain diseases, where very high precision in measurement is required, due to small, subtle volumetric differences between brain regions in patients and in normal control subjects. Moreover, some of those brain disorders (e.g., schizophrenia) involve volumetric changes in several brain regions simultaneously and an ideal study should be able to address all those changes in the same patient population. However, the method currently used for analyzing those differences involves a laborious manual tracing of the contours of anatomical structures derived from MRI scans. Using manual tracing, the resulting time requirements make it practically impossible for a

[1] Presented in part at the 81st Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, Illinois, November 1995; at the Annual Convention of the Society of Biological Psychiatry, New York, New York, May 1996; and at the Annual Meeting of the American Psychiatric Association, New York, New York, May 1996.

[2] This research was supported by the Medical Research Service and Brockton Schizophrenia Center of the Department of Veteran Affairs and a Senior Investigator Award from the National Alliance for Research in Schizophrenia and Depression (Dr. McCarley); by grants R01-40799 (Dr. McCarley), 1K02 MH-01110-01, and 1R29 MH-50747-01 (Dr. Shenton) from the National Institute for Mental Health; by the Commonwealth of Massachusetts Research Center (Dr. McCarley); by the Scottish Rite Foundation (Dr. Shenton); and by an Australian Postgraduate Award (S. Warfield).

1053-8119/97 $25.00

research team to measure more than a limited number of brain structures in a patient population and in a matched control group. An automated procedure would thus greatly increase both the number of regions as well as the number of individuals who could be investigated in any one study.

An automated technique for quantitative MRI analysis would have to be able to differentiate between structures with similar intensity on the MRI scans. One solution would be to use an anatomical atlas as a template, in order to provide anatomical information not obvious from MR contrast alone. Several digital atlases have, in fact, been developed in the past decade in order to solve this problem, one of the best known being the atlas established by Höhne and coworkers (e.g., Höhne *et al.*, 1992).

In the current study, we have used an MR brain atlas developed in our laboratory (Kikinis *et al.*, 1996; Shenton *et al.*, 1995), in order to evaluate, and compare, our automated registration algorithm for the volumetric measurement of MRI brain structures, to the same measurements performed using manual tracing. Accordingly, we took information from the MR brain atlas, based on one control subject, and projected it into other MRI scans by applying an elastic match (i.e., warping the atlas into the shape of the new brain image). The global registration technique that we used to match our anatomical MR atlas onto new MR images is based on work by Dengler and coworkers (e.g., Dengler *et al.*, 1988; Schmidt *et al.*, 1989; Warfield *et al.*, 1995). This technique builds on the theory of elastic membranes, and it is similar to Grenander's approach (Grenander and Miller, 1994).

The elastic membrane model can be intuitively understood as the deformations occurring when a set of points on the membrane is stretched. In this context, the atlas brain can be compared to a rubber brain, which is stretched and compressed nonlinearly in order to match the contours of the new brain. At the end of the registration, all the structures previously defined for the atlas brain are also defined for the new brain image. The general hypothesis underlying this method is that the topology of cerebral structures remains an invariant, which means the differences between individuals may be considered variations in the shape of a common underlying plan (the atlas).

As we report in this study, we were able to measure, with high accuracy, the volumes of 11 brain structures using elastic matching. (These brain structures included total brain volume, total gray matter, total white matter, left and right thalamus, left and right caudate, left and right putamen, and left and right globus pallidus.) This technique also allows for the study of larger groups of patients, due to the remarkable increase in speed over manual tracing. Therefore, using elastic matching, as we will describe in this study, can increase the possibility of highlighting subtle morphological changes associated with schizophrenia and other chronic brain disease.

METHODS

Subjects

Fourteen chronic schizophrenia patients, from among the patients of the Brockton VA Hospital, and 14 normal controls were selected and matched for age, gender, handedness, and parental socioeconomic status. The inclusion criteria for all subjects were: (1) age 17–55, (2) right handed, (3) English as primary language, (4) verbal IQ above 75, (5) no history of electroconvulsive therapy, and (6) no history of alcohol or drug dependence, and no abuse in the last 5 years. The exclusion criteria for normal controls were no personal or familial history of mental disorder (for further subject characteristics see Shenton *et al.*, 1992).

MR Image Processing

MRI Acquisition

The MR images were acquired on a 1.5-T General Electric Signa System (GE Medical Systems, Milwaukee). Sagittal localizer images were acquired, followed by three-dimensional Fourier-transform spoiled-gradient-recalled acquisition (SPGR), with a repetition time of 35 ms, an echo time of 5 ms with one repetition, a flip angle of 45°, a field of view of 24 cm, and a matrix of 256 by 256 (192 phase-encoding steps, with zero filling) by 124. The data were stored and analyzed as 124 coronal slices of 1.5-mm thickness. Voxel dimensions were $0.9375 \times 0.9375 \times 1.5$ mm. To reduce flow-related artifact from cerebrospinal fluid (CSF) and blood, presaturation of a slab inferior to the head was performed. Following MR acquisition, we used a noise-filtering program (Gerig *et al.*, 1992). This program and all image-processing methods described or implemented on SPARC workstations (Sun Microsystems, Mountain View, CA) and have been successfully used for a number of studies in our laboratory (e.g., Shenton *et al.*, 1992; Hokama *et al.*, 1995; Wible *et al.*, 1995).

Manual Region of Interest (ROI) Definition by Experts

We used the measurements for basal ganglia volume published by Hokama and coworkers (1995). These measurements included left and right caudate nucleus, left and right putamen, and left and right globus pallidus. For thalamus (left and right) measures, we used the values obtained by Portas *et al.* (1997, submitted for publication). Both studies were done on the same set of MR brain images that we used for the automated method. (Measures of whole brain volume, as well as gray and white matter volume, were computed using automated segmentation algorithms—see

below—whereas the smaller ROIs were traced manually by raters.) The intrarater reliability in the basal ganglia study (in which the editing was done by a single rater), measured as the percentage of error, was 0.5 and 0.5% for globus pallidus, 2.0 and 3.2% for putamen, and 4.3 and 4.4% for caudate. In the study of thalamic volume by Portas and coworkers, the interrater reliability (mean intraclass r) was $r = 0.910$, based on three brain images, edited by three raters. For the intrarater reliability, the percentage difference was 4.55%. We therefore considered the results from both studies to be a reasonable standard for comparison purposes.

Automated Segmentation

A comprehensive description and validation of the tissue classification procedure we used has been published by Wells and coworkers (1996). In the first series of segmentations, we used the classification algorithm to segment SPGR voxels for the entire brain into white matter, gray matter, and CSF. In a second series, we segmented those same voxels into cortical gray matter, subcortical gray matter, white matter, and CSF. The segmentation algorithm allowed the operator to choose representative pixels for each tissue class. Those points were chosen in 1 brain image and then we used the same values for the other 27 brain images. We defined subcortical gray matter by choosing approximately 20 representative voxels in the best defined regions from the center of the thalamus and putamen, over one or two MR slices. The automated program was then able to classify the rest of the voxels. We also segmented the skin, which was used in the initial registration stage (see below, under Linear Registration).

This fully automated algorithm used a priori knowledge of tissue properties and intensity inhomogeneities in order to correct for those intensity differences in the MRI data (Wells *et al.*, 1996). Only the segmented tissue class data were used for the elastic matching. Working with segmented data improved the robustness of the elastic matching algorithm to noise effects and it increased the speed of the matching.

There was no relationship between the spatial location of the chosen training points and the accuracy of the segmentation or of the subsequent elastic match. At the segmentation level, the tissue classification was voxel independent and the spatial location of the training points has no effect on the recovered classification (see Wells *et al.*, 1996). The key criterion in the training step was to model the distribution of signal intensities. Only the variation of signal intensities made any difference. Therefore, selecting training points with identical signal intensity but at different spatial locations had no effect on the statistical model for the distribution of tissue classes, and resulted in no changes in the classification. Training points with different signal intensities led to a different statistical model for the distribution of tissue classes. This, in turn, may or may not have led to a different classification, depending on whether feature space boundaries were altered.

Definition of Subcortical Gray Matter

For the second series of segmentations, the automatic segmentation algorithm misclassified some of the cortical gray matter as subcortical gray matter, due to the similarity in MRI signal intensity. In order to partially eliminate the misclassified subcortical gray matter from the subsequent match, we outlined a rectangular volume (seen as a rectangle in Fig. 1). Its borders were defined to include the entire subcortical areas. It also included some cortical areas (e.g., insular cortex), since we tried to keep to a minimum the amount of manual editing involved. The rectangular volume was manually outlined on the atlas image and it was then linearly registered into each new patient image. Within this rectangular space we included all tissue classified by the segmentation algorithm as subcortical gray matter as a target for the matching program. We used as the posterior border the most posterior coronal MRI slice in which the temporal horns were separated from the main body of the lateral ventricles. The anterior border was defined as the most anterior MRI slice containing a part of the lateral ventricles. The inferior border was taken as a horizontal line drawn through the inferior tip of the third ventricle and, more anteriorly, the superior tip of the inferior interhemispheric sulcus. The superior border was taken as the roof of the lateral ventricles. The lateral borders were chosen as the gray matter of the insular lobe on both sides. In Fig. 1, we present the outlining of those borders on an actual MR image. The operation was important in order to target the elastic matching algorithm toward a specific brain area and to eliminate the misclassified tissue from the match. We want to emphasize, however, that those borders did not contain anatomical information, nor did they represent anatomical landmarks which the computer was supposed to match, as in the method of Evans and coworkers (Evans *et al.*, 1991).

Linear Registration

The first step toward matching the atlas brain image onto a new brain image (patient) was to use a linear registration program to correct for the differences in size, rotation, and translation between the two brain images (Ettinger *et al.*, 1994). The linear registration performs a gross alignment of the two three-dimensional (3D) data sets, through a combination of energy-minimization registration techniques. The bases of the registration were the 3D surfaces extracted from MR images. In the first step, the system extracted the 3D surface points corresponding to the skin in the segmented images from the atlas brain and from the

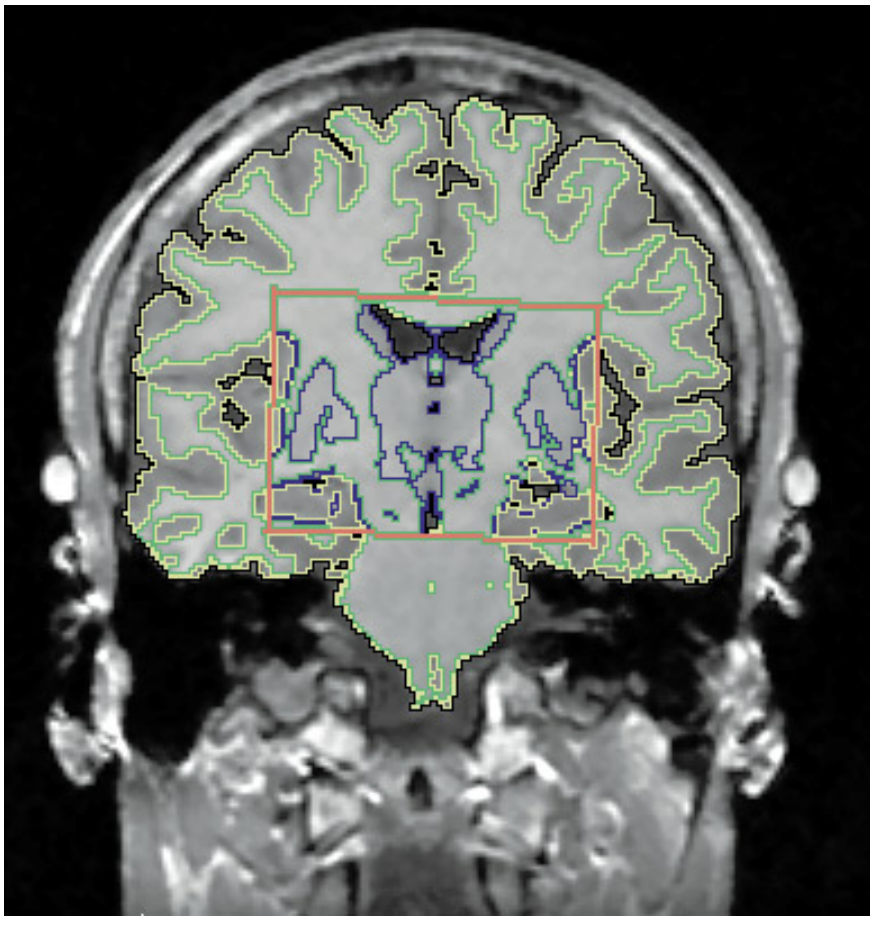

FIG. 1. An example of the segmentation into cortical white matter (yellow borders), white matter (green borders), and subcortical gray matter (blue borders). The area within the orange rectangle was segmented into cortical and subcortical gray matter, whereas the area outside the rectangle was segmented using a single tissue class for the gray matter.

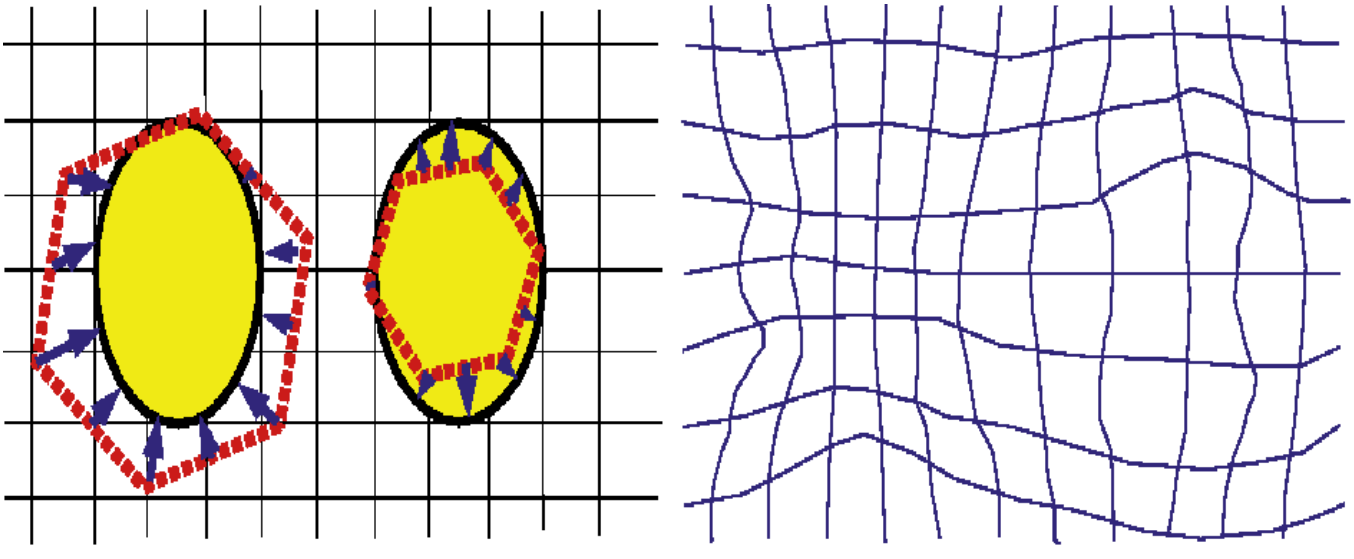

FIG. 2. Schematically, an "atlas" image (the red hexagon) is mapped onto a new "patient" image (yellow oval). A simple uniform global displacement (translation, rotation, or scaling) does not work. Instead, we used a procedure that "warped" the atlas image onto the patient image. This warping resulted in a "vector field" (blue arrows). The nonuniform displacement of each pixel is then represented in the right field, by means of the deformation of the rectangular grid (the elastic membrane).

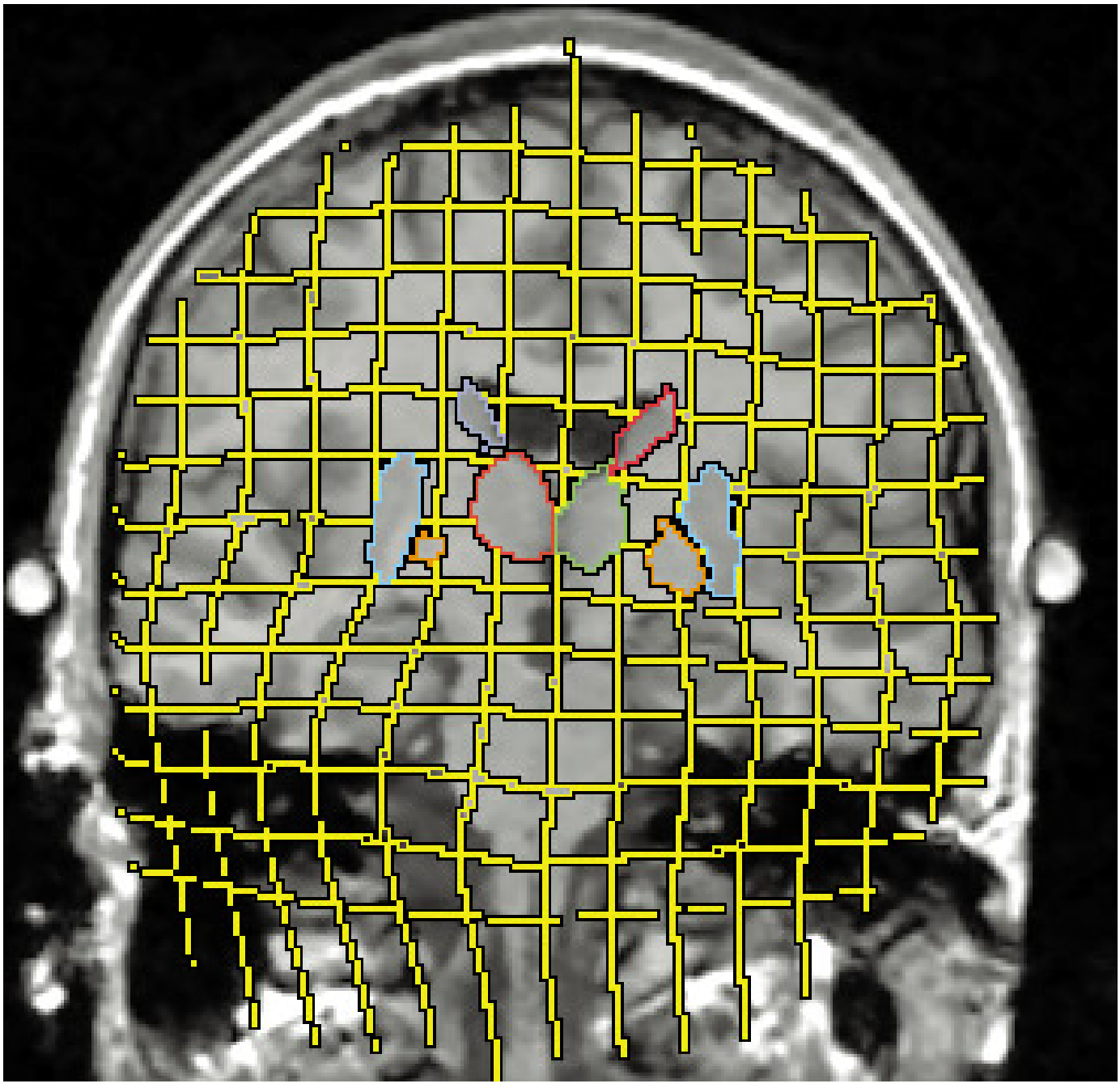

FIG. 3. An actual example for the elastic matching. The background is the MR scan of the patient. We outlined the contours of the basal ganglia, taken from the warped atlas. The grid (yellow) was rectangular in the atlas coordinates, and its deformation illustrates the warping process. Please compare with the schematic presentation in Fig. 2.

patient brain. The subsequent linear transform was computed on those skin surfaces and then applied to the atlas brain. The final product was an atlas brain image linearly registered onto the patient brain image.

Elastic Matching

We summarize here some technical details of Dengler's regularization procedure (Dengler *et al.*, 1988; Schmidt and Dengler, 1989), which we used to map 3D MR images onto each other. (For a more detailed discussion on the mathematical model underlying the algorithm, please see the Appendix). The goal of the elastic matching algorithm was to find a 3D deformation vector field which transformed the source data set (segmented atlas) so that it matched the target data set (segmented patient image). The deformation field maxi-

mized the local similarity between the two data sets. In Fig. 2, we present a schematic example to illustrate the principle of the warping process and the elastic deformation of the MR image.

The elastic matching is volume based, with the forces driving the atlas and patient together being derived from the mismatch of the boundaries of the structures. The elasticity constraint is a global 3D regularization that applies to every voxel in the data set. A mapping between two MRI images required defining the full vector field (i.e., displacement of each voxel); this was an "ill-posed" problem since displacement information was present only where there was structure in the images, and the number of voxels accounting for structure was much less than the total voxel number. This occurred with the dimensional increase or decrease in the regions of interest. For example, schizophrenics had increased basal ganglia volume compared with normal subjects. In this case, the mapping of a larger to a smaller volume involved a "many-to-fewer" mapping of voxels. Thus a region of, say, 100 voxels was mapped onto an 80-voxel region. An intuitive way to understand Dengler's (Dengler *et al.*, 1988; Schmidt and Dengler, 1989) solution is to consider it as involving a rather detailed mapping of regions where there is important information (e.g., edges), while interpolating in regions where there is little information (e.g., large isodense gray matter zones).

The segmented MRI data were interpolated to form a data set with isotropic voxel size. The finest spatial scale of the matching was set by selecting a Gaussian filter sampled to a size of $9 \times 9 \times 9$ voxels at this resolution. This same filter was applied through a multiresolution pyramid, so that at each level of the pyramid a fixed-width window was used, but this corresponded to different spatial scales. In the current implementation we used a three-level pyramid.

The algorithm was completely automated; the only user input was the selection of the segmented tissue classes as variables in the matching procedure. The values assigned depended, in general, on the stage of the registration and were empirically determined. We used only the skin surface for the initial linear registration, then the white matter and the subcortical gray matter for the elastic matching. The matching algorithm determines correspondences between segmented data sets. We currently use a sequence of matches of binary masks to achieve the overall alignment. The sequence of binary mask matches is designed to capture different types of local shape deformation at each stage. The first match involves a mask formed from the white matter. This match captures local differences in the white matter volumes. To further refine this alignment for the subcortical gray matter structures, we performed a match of the gray matter structures inside a mask of the subcortical gray matter region.

The use of a binary mask essentially equalizes the image gradients, so that every edge has the same gradient and inside homogeneous regions no gradients arise. The multiresolution pyramid makes use of low-pass-filtered (Gaussian-windowed) segmentations and so in the actual matching process no true step edges are present; all edges have been smoothed by the low-pass-filtering operation.

We further improved the matching results by using the output of one matching cycle as an input (atlas) for a new matching cycle (*iterative cycles*). In the elastic membrane model, at the end of the match cycle we had a tense and deformed membrane. By removing the tensions and applying a new match cycle, the membrane was further deformed toward the model. Four iterative matching cycles were applied for each new patient image.

Data Analysis

We evaluated the *accuracy of each match* by comparing, for each case, the volumes of 11 brain structures, measured both manually and with elastic matching. We computed a coefficient of similarity between the measurements of MRI volumes, done with the two methods. This coefficient was defined, for each structure, as

$$S = 1 - \frac{|\text{manual volume} - \text{automat. volume}|}{\text{manual volume}}.$$

Thus the accuracy was defined relative to the "true" volume, measured manually. Using the absolute value of the difference between the two measurements we penalized in the same way either a larger or a smaller value of the matched volume. For example, if the manually measured volume had 1000 voxels, the coefficient of similarity was the same (0.5), whether the matched volume had 500 or 1500 voxels. Furthermore, we used the absolute value of the difference, so that when added together the differences will not result in an artificially high similarity. The coefficient of similarity for each structure was averaged over all 28 cases as

$$S = 1 - \frac{1}{N} \Sigma \frac{|\text{manual volume} - \text{automat. volume}|}{\text{manual volume}}.$$

We also measured the *degree of spatial accuracy* of the automated algorithm by computing, for each separate structure, the relative spatial overlap between the manual and the automated volumes:

$$\text{RO} = \frac{\text{intersection of manual volume and automat. volume}}{\text{manual volume}}.$$

Two-tailed tests of significance were used for the Pearson and the Spearman correlations among all 28 cases, which assessed the correlation between the manual measurements and the results of each cycle of iterative elastic matching. The correlations were computed separately for the 11 neuroanatomical structures. We also computed the total volume of the basal ganglia in each case, by adding the volumes of the left and right caudate nucleus putamen, and globus pallidus.

RESULTS

The first set of measurements was computed on the series of segmentation including the subcortical gray matter. We present the *accuracy of the volume measurements* for all 11 brain structures, computed as the similarity between the automated and the manual method (see Methods for the definition of the similarity coefficient). This is followed by a review of the correlation between manual and automated measurements. Next, we present the *spatial accuracy* of the automated algorithm, computed as the percentage of overlap between structures edited with the two methods (i.e., automated and manual). We further present the similarity of the measurements on the series of segmentations not including a separate tissue class for subcortical gray matter. The entire white and gray matters were used there as features for the match. We also present the accuracy of the volumetric measurements for cortical brain structures. Finally, we present a comparison for the 11 brain structures' volumes between the schizophrenia patients and the normal controls.

For an actual example, Fig. 3 presents the automated outlining of the basal ganglia on an actual MR coronal slice. The deformation of the yellow grid illustrates the warping process. Figure 4 shows a comparison, on one slice (i.e., in two dimensions), between the manual edited basal ganglia and thalamus (on the left) and the automatic registration of the basal ganglia and thalamus by elastic matching (on the right). Figure 5 represents the 3D displacement of the basal ganglia in the atlas in the warping process.

1. The Similarity between Automated and Manual Measurements, When Using the Subcortical Gray and White Matter Surfaces as Match Features

We computed the similarity coefficients between automated and manual measurements, using the formula presented under Methods. The elastic match was done using the subcortical gray and white matter surfaces as match features. (See Methods for the operational definition of the subcortical gray matter. The rationale for using subcortical gray matter surface as the primary feature of the match was to obtain a better result for subcortical structures.) The first column in Table 1 lists the mean similarity coefficients for whole brain volume and the 10 other brain structures, averaged over 28 cases. The improvement of the results of the first match during subsequent iterative cycles is highlighted in Table 2 by presenting the results for all four iterations of the elastic match. After four iterative warping cycles, the similarity coefficients were 97% for whole brain volume, 97% for whole white matter, 92% for whole gray matter, 96% for thalamus (both sides), and 91% for basal ganglia (93% for putamen, 91% for caudate nucleus, and 76% for globus pallidus).

2. The Correlations between Pairs of Automated and Manual Measurements When Using Subcortical Gray Matter and White Matter as Match Features

Two-tailed tests of significance were used for the Pearson and the Spearman correlations between the manual measurements and the results of each of the four cycles of iterative elastic matching. We averaged the results for all 28 brain images. Columns 3 and 4 in Table 1 present the Pearson and the Spearman correlations, computed separately for the 11 neuroanatomical structures. After four iterative warping cycles, the Pearson correlations between the two data sets were $r = 0.94$ for the whole brain, $r = 0.98$ for the white matter, $r = 0.79$ for the gray matter, $r = 0.90$ for thalamus, and $r = 0.79$ for basal ganglia ($r = 0.78$ for the caudate nucleus, $r = 0.85$ for putamen, and $r = 0.50$ for globus pallidus). All these r values were significant at $P < 0.01$.

3. The Spatial Overlap between Automatically Edited and Manually Edited Structures, When Using Subcortical Gray Matter and White Matter as Match Features

We measured the degree of spatial accuracy of the automated algorithm by computing, for each separate structure, the relative overlap between the manual and the automated volumes. The relative overlap was computed as the intersection of the two volumes divided by the manually edited volume. Column 5 in Table 1 presents the mean overlap for each of the 11 structures, averaged over the 28 brain images. The mean overlap was 96% for whole brain, 97% for total white matter, 92% for total gray matter, 86% for thalamus, 79% for putamen, 78% for caudate nucleus, and 59% for globus pallidus.

4. The Similarity between Automated and Manual Measurements, When Using the Subcortical Gray and White Matter Surfaces as Match Features

We also measured the volumes of the same structures using the gray and white matter surfaces as match features for the warping algorithm. Our hypothesis was that we would obtain a worse result for subcortical structures, since in this case the algorithm

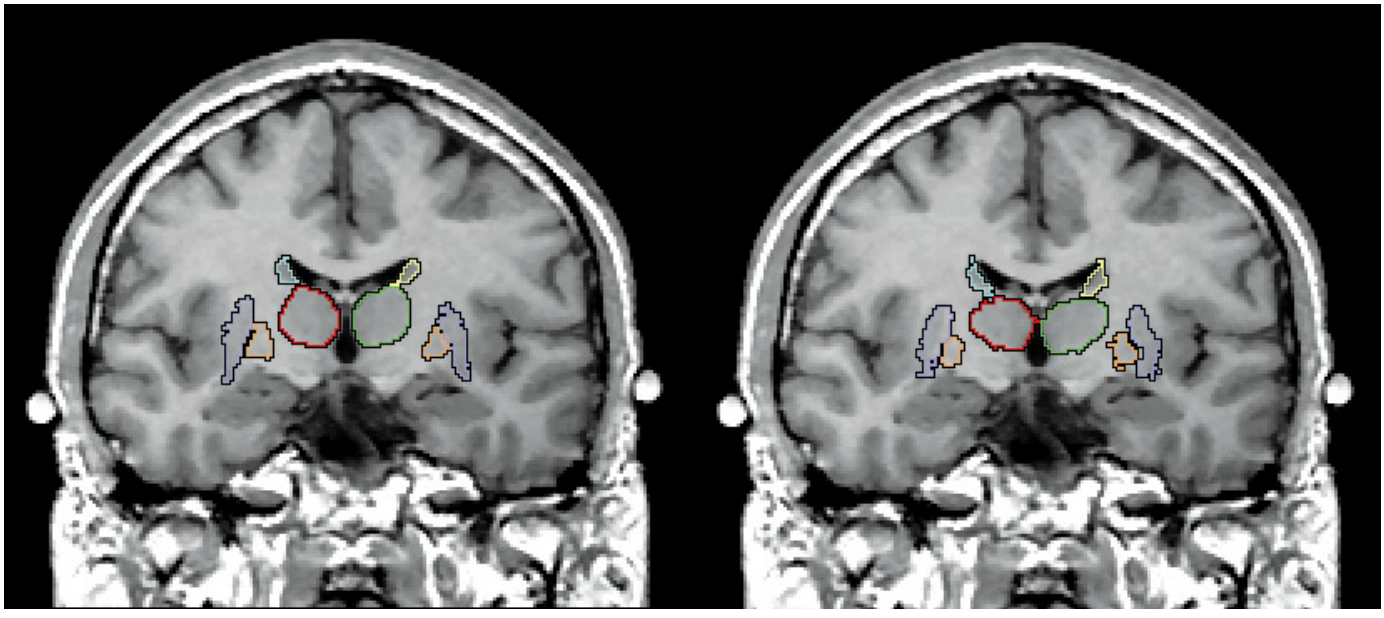

FIG. 4. The only gold standards we had for the elastic matching algorithm were the manually edited basal ganglia and thalamus. On the left is one MRI slice showing the manually edited basal ganglia and thalamus. On the right is the same slice but here we see the results of the basal ganglia and thalamus depicted using the elastic matching program.

would not be able to recognize precisely their position in the new brain image and would instead approximate this position. We also expected the results for the total cortical gray matter to improve, since the algorithm would specifically take into account cortical gray matter boundaries. Column 6 in Table 1 presents the mean similarity coefficients for the same 11 brain structures, averaged over 28 cases. With this method, the similarity coefficients were 87% for thalamus and 89% for basal ganglia (89% for putamen, 90% for caudate nucleus, and 82% for globus pallidus). Thus, if we compare these findings to No. 1 above, the results were as predicted: using total gray matter instead of subcortical gray matter as the matching feature resulted in a

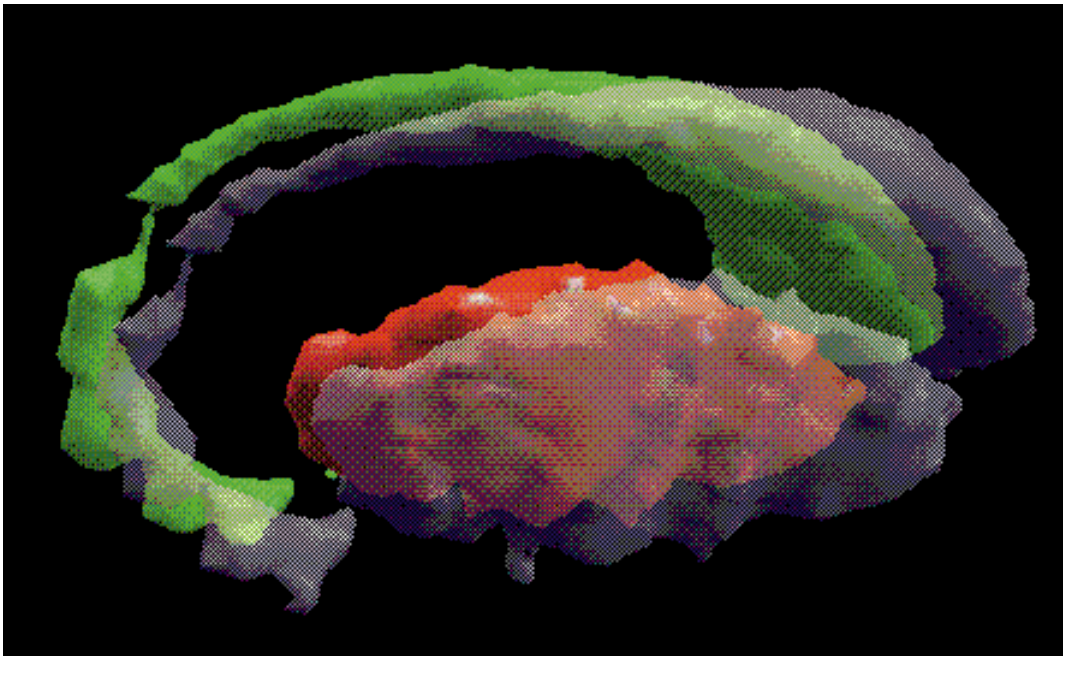

FIG. 5. A 3D representation of the right basal ganglia, taken from the atlas (caudate nucleus, green; putamen, red). Here, we superimposed (in gray) the final shapes and positions of the basal ganglia after being warped onto a new brain image.

TABLE 1

Comparison between Manual and Automated Measurements of Brain Volumes

Structure	Similarities (%)	Pearson r ($P < 0.01$)	Spearman r ($P < 0.01$)	Spatial overlap (%)	Similarities (%) (W/G matter match)
Total gray matter	91.5 ± 5.4	0.79	0.78	92.3 ± 6.9	95.2 ± 3.2
Total white matter	97.0 ± 1.5	0.98	0.96	97.2 ± 1.8	96.9 ± 1.5
Right thalamus	95.3 ± 4.2	0.88	0.83	84.9 ± 2.7	88.0 ± 9.3
Left thalamus	96.3 ± 3.0	0.91	0.93	85.6 ± 2.6	87.0 ± 11.0
Right caudate	90.9 ± 7.3	0.79	0.81	77.1 ± 4.3	90.7 ± 9.0
Left caudate	90.3 ± 7.4	0.78	0.75	78.5 ± 4.1	89.0 ± 12.3
Right putamen	92.8 ± 5.9	0.83	0.85	77.4 ± 3.9	87.6 ± 10.6
Left putamen	92.7 ± 5.6	0.87	0.84	79.9 ± 3.3	90.6 ± 9.3
Right globus pallidus	76.4 ± 13.4	0.44	0.47	59.2 ± 7.7	82.4 ± 11.5
Left globus pallidus	76.1 ± 15.2	0.55	0.50	58.9 ± 10.3	81.5 ± 11.6
Total (whole) brain	97.0 ± 2.3	0.94	0.94	95.9 ± 2.5	98.1 ± 2.0

lower accuracy of the volume measurements for the subcortical gray matter structures (see Tables 1 and 2). The only apparent exception was the left and right globus pallidus, but there the differences between the two methods were not statistically significant, due to the large standard error.

5. The Similarity of the Measurements for Cortical Gray Matter Structures

Our study focused primarily on subcortical brain structures, since we assumed that cortical gyri, which present significant variability even among normal brains, would be less reliably measured with elastic matching. The automated computer algorithm assumes the neuroanatomical variability among subjects to be a topological invariant. But cerebral gyri frequently split in two in some subjects, whereas they remain one single structure in others, differences which cannot be taken into account by the automated registration in its present form. We tested this assumption by measuring 20 brain cortical and subcortical brain structures in one brain image. We used the first series of segmentations of the brain images (white matter, gray matter, and CSF). We found a mean similarity coefficient of 80% for cortical gyri (91% for the medial frontal gyri, 83% for orbital gyri, 64% for postcentral gyri, 89% for the superior parietal gyri, and 83% for the inferior parietal gyri). For the same brain image, the mean similarity coefficient was 90% for subcortical structures (87% for thalami, 86% for the caudate nucleus, 95% for putamen, and 78% for globus pallidus) and 98% for total gray and white matter volumes. Please refer to Shenton *et al.* (1992) and Wible *et al.* (1995) for the definitions of the cortical gyri and boundaries.

6. The Comparison of the Brain Structures' Volumes between the Patient Group and the Normal Controls

We compared subcortical brain structures measured with elastic matching between schizophrenia patients and the normal control group (Table 3). We found no significant difference in total brain volume, total gray matter, total white matter, or thalamic MR volume between the two groups. These findings are in concordance with previous results obtained with manually measured volumes in the same patient population (Hokama *et al.*, 1995; Portas and coworkers, submitted for publication). We found, however, a 16.7% increase in MR volume for the basal ganglia in schizophrenics compared with normal controls (9.8% increase in caudate nucleus, 20.3% in putamen, and 28.2% in globus pallidus). All differences measured in basal ganglia structures were statistically significant (Pearson $P < 0.02$), with the exception of the left caudate nucleus ($P = 0.08$). But this difference also becomes significant ($P = 0.03$) if we compare the total volume of the two caudate nuclei on both sides, as was reported by Hokama *et al.* (1995). These data are in accordance

TABLE 2

Similarity Coefficients (Subcortical Gray Matter Used as Primary Match Feature)

Structure	First match (%)	Second match (%)	Third match (%)	Fourth match (%)
Total gray matter	88.6 ± 6.7	89.7 ± 6.2	90.8 ± 5.8	91.5 ± 5.4
Total white matter	94.5 ± 2.5	95.9 ± 1.6	96.6 ± 1.5	97.0 ± 1.5
Right thalamus	89.7 ± 7.7	95.0 ± 3.2	93.8 ± 3.8	95.3 ± 4.2
Left thalamus	89.4 ± 10.1	94.6 ± 4.1	94.6 ± 3.4	96.3 ± 3.0
Right caudate	71.0 ± 28.5	88.6 ± 10.0	90.7 ± 7.8	90.9 ± 7.3
Left caudate	70.7 ± 29.5	86.0 ± 10.3	90.9 ± 7.9	90.3 ± 7.4
Right putamen	85.2 ± 12.1	92.0 ± 5.6	92.1 ± 4.8	92.8 ± 5.9
Left putamen	83.6 ± 18.3	92.9 ± 5.6	92.7 ± 5.3	92.7 ± 5.6
Right globus pallidus	78.6 ± 12.5	77.1 ± 15.3	76.2 ± 12.7	76.4 ± 13.4
Left globus pallidus	72.2 ± 10.7	71.6 ± 13.3	74.7 ± 15.9	76.1 ± 15.2
Total (whole) brain	96.5 ± 2.4	96.6 ± 2.5	96.9 ± 2.3	97.0 ± 2.3

TABLE 3

Comparison between Brain Structures' Volumes in Schizophrenia Patients and Normal Controls

Structure	NC (ml)	SZ (ml)	Mean difference (ml) (%)	Significance (*P*) of the difference
Total gray matter	721.74 ± 12.68	736.27 ± 11.48	+14.53 (+2.01%)	0.403
Total white matter	584.20 ± 12.90	589.51 ± 12.60	+14.31 (+2.45%)	0.435
Right thalamus	7.17 ± 0.25	7.29 ± 0.24	+0.11 (1.59%)	0.743
Left thalamus	6.97 ± 0.18	7.24 ± 0.20	+0.27 (+3.80%)	0.337
Right caudate	4.75 ± 0.15	5.31 ± 0.16	+0.56 (**+11.87%**)	**0.015**
Left caudate	4.97 ± 0.14	5.34 ± 0.16	+0.38 (+7.59%)	0.080
Right putamen	4.02 ± 0.12	4.96 ± 0.14	+0.94 (**+23.31%**)	**0.001**
Left putamen	4.46 ± 0.14	5.19 ± 0.14	+0.74 (**+16.52%**)	**0.001**
Right globus pallidus	0.87 ± 0.04	1.08 ± 0.05	+0.20 (**+24.10%**)	**0.003**
Left globus pallidus	0.82 ± 0.05	1.09 ± 0.06	+0.27 (**+32.44%**)	**0.001**
Total (whole) brain	1305.94 ± 24.64	1334.78 ± 22.99	+28.94 (+2.22%)	0.400

Note. SZ, schizophrenia patients; NC, normal controls. Significant differences shown in bold.

with the results of Hokama and coworkers' study (1995), where manual definitions of the basal ganglia were done in the same patient population.

DISCUSSION

Using elastic matching, we were able to measure the volumes of several brain structures with very high accuracy. The best results were obtained in matching large and regularly shaped objects. Thus, the best matched structure was the thalamus, relatively large and egg-shaped. The volume measurements of the thalamus were reproduced with an accuracy of 96%, and the structure was defined 90% within the boundaries traced manually. This result is similar to previous studies (e.g., Dann *et al.*, 1989; Kikinis *et al.*, 1992; Gee *et al.*, 1993; Haller *et al.*, 1995), which have demonstrated greater similarity in results and reliability for regularly shaped objects. Any error of the matching algorithm had a larger impact on smaller structures, since we measured the accuracy of the match relative to the structure's volume. Therefore, the relatively small globus pallidus was matched with an accuracy of 80% and overlapped 50% with the manually traced structure.

The shape also played an important role in the accuracy of the match. The putamen and the caudate nucleus were roughly similar in volume, but we obtained better results for the putamen, a compact structure, whose volume was measured with an accuracy of 92%, and the structure overlapped 82% with the manually traced contours. On the other hand, the caudate nucleus, an irregular, tail-shaped structure, had a 90% accuracy in volume measurement but intersected 70%, as opposed to 82% for putamen, with the manually traced contours.

A key finding derived from this validation study was that the specific volumetric differences between schizophrenic brains and the normal controls were defined in exactly the same way by the automated algorithm and by manual tracing: (1) There were no significant differences in total brain volume, total gray or white matter volume, or thalamic volume. (2) For the basal ganglia, there was a 16.7% increase in volume for the schizophrenic group, with statistically significant enlargement of all basal ganglia structures. These results correlate with previous studies in the literature and, more specifically, with the measurements reported by Hokama *et al.* (1995), based on the same patient population. In our opinion, the ability of the elastic matching algorithm to recognize and measure subtle volumetric differences induced by pathology is a powerful argument for the reliability of this method. This finding also suggests that the small errors introduced by the automated algorithm are systematic and therefore do not affect the comparison between groups.

The automated technique has several major advantages over human manual tracing of anatomical structures. First, the automated technique is rapid and efficient. This means that the measurements of multiple structures in one brain, which would have involved weeks of manual editing, can now be completed in hours. Second, the method is very reliable, in that the results are 100% reproducible in a second assessment, due to the method being completely automated. This result has to be compared with the inherent differences between successive assessments done by a human rater, and the even larger differences between different raters. As we determined in this study, these advantages of the automated method are accompanied by excellent accuracy in measurement.

The technique also has some limitations, at least in its present form. The most important limitation is the lower accuracy in measuring cortical structures, due to the high normal variability of those structures (though keep in mind that the measurement of these areas was,

nonetheless, highly correlated with the manual measurements of these same regions). Further refining of the matching algorithm (Warfield *et al.,* 1995) will allow for better measurements of the cortical structures. Another limitation is related to the size of the structures to be investigated. Currently, very small structures, such as globus pallidus (volume 1 ml) tend to be recognized and measured with lower accuracy. Those structures, due to their small size, can be more accurately traced by the manual rater. Therefore, the advantages of using an automated elastic matching algorithm are less clear for very small structures.

The main similarities between Dengler's algorithm (e.g., Dengler *et al.,* 1988; Schmidt and Dengler, 1989), which we used, and the techniques used by other research teams include: (1) the use of nonlinear registration techniques for deforming (warping) the images to be matched and (2) the use of a pyramidal approach, whereby the match proceeds in stages from lower to higher levels of resolution.

Our method is different from the techniques used by other research groups, however, in several respects:

(1) We used a digital atlas derived from ROIs which were manually drawn on a normal subject MRI brain (Shenton *et al.,* 1995; Kikinis *et al.,* 1996), rather than an abstract (probabilistic) atlas (Collins *et al.,* 1992) or interpolated data from an anatomical atlas (e.g., Dann *et al.,* 1989; Gee *et al.,* 1993).

(2) Our registration (warping) procedure used only automated nonlinear matching, with an automated linear registration as a first step, rather than operator-defined global linear transformations (scale, rotation, and translation) and operator-defined landmarks for elastic match, followed by operator-performed local adjustments (e.g., Evans *et al.,* 1991).

(3) Other research teams used linear transforms (e.g., Andreasen *et al.,* 1996) or elastic transforms (e.g., Collins *et al.,* 1994) to register MRI data into standard Talairach coordinates. While valid in the way these groups have reported their work, this approach is very different conceptually from ours. We registered the atlas into the patient coordinates. The registration of all images into a standard set of coordinates is very well suited for functional activation studies. For morphological studies, however, we think that matching the atlas into the patient coordinates allows for a better recognition of the individual variability of each brain.

(4) We used segmented images as the basis of the elastic match, and not the unsegmented MR data (e.g., Haller *et al.,* 1995; Christensen *et al.,* 1993; Miller *et al.,* 1993). It is true that segmenting the MRI data involves a degree of postprocessing and some loss of information. Therefore, it is theoretically valid to use unsegmented, raw MR data more directly related to the actual brain characteristics. We selected segmented (histological) MR data to use as a basis for the automatic registration for several reasons. The noise and the qualitative differences between the MR scans make it problematic for the automated registration to rely exclusively on the raw MRI data. Both noise and qualitative differences are unrelated to displacement, and therefore cannot be compensated by registration, but both can be compensated during the segmentation process. Thus, the advantage of using segmented data was that it minimized several confounding factors in the computation of the elastic transform. In our method, the imaging system noise is "filtered" by the segmentation algorithm and it is thus minimized as confounding factor in the computation of the elastic transform. Therefore, the final accuracy of the warping depends on the accuracy of the segmentation (see Wells *et al.,* 1996, for a comprehensive description and validation of the segmentation algorithm).

Also, in order to compensate for inhomogeneities and noise present in unsegmented MR data, more sophisticated algorithms are required, which in turn are more computationally demanding, even for the match of very small areas. Thus, there are serious practical limits for the size of the brain areas which can be matched. In their reliability study, Haller and coworkers (1995) used a fluid dynamic system to match a small area (the size of hippocampus) of a single 2D slice. Using segmented MRI data, we were able to match the entire 3D volume of the brain images, with an important increase in speed (minutes versus hours) and comparable accuracy. The entire process of automatic segmentation, linear registration, and four cycles of elastic matching requires less than 90 min of computer time on a SPARC workstation (including about 15 min on a Thinking Machine parallel computer, for the segmentation), as opposed to several hours for the method of Haller and coworkers (1997).

Although the fluid dynamic system has a greater computational complexity, the aims of the two algorithms (to project an anatomical model onto patient data) are the same. Hence it is important to note that, while both these methods are able to match patient anatomy, the algorithm presented here has significantly less computational complexity and requires significantly less computational resources to use.

(5) There was a weighting of the relative importance of features in each step of the matching according to their utility in matching. For example, in the initial match, the white matter features were weighted most heavily since the peaks and valleys of white matter that border gray matter were extremely "feature rich." In later stages the voxels classified as subcortical gray matter were weighted most heavily, since this specifically "targeted" the warping process on the features of the thalami and basal ganglia. For example, the mea-

surements for subcortical structures (i.e., the thalamus, the basal ganglia) were much improved (on average the accuracy increased by 10%) when the subcortical gray matter, and not cortical gray matter, was used as a primary feature for the match. As expected, the result of the match was more accurate when we provided the matching algorithm with more information, or more features specifically related to the structures we tried to match.

Compared with elastic matching algorithms developed by other research teams (e.g., Evans *et al.*, 1991; Dann *et al.*, 1989; Gee *et al.*, 1993; Haller *et al.*, 1995; Bajcsy and Kovacic, 1989), the algorithm we used presents similar or higher accuracy and sensitivity to small local differences. It was difficult, however, to make direct comparisons between our data and those from other research teams developing algorithms for elastic matching (e.g., Dhawan *et al.*, 1992), since these other researchers did not report detailed measurements for validation purposes. Also, other research teams (e.g., Andreasen *et al.*, 1996; Collins *et al.*, 1994) have registered MR brain images into standard Talairach coordinates. As we already explained, there was an important conceptual difference between our approach and this method. Therefore, we think that a comparison of the results is not meaningful.

In conclusion, we obtained high-precision volume measurements and position recognition for different brain structures in the MRI scans. The differences between the measurements done manually by human raters and those done with the automated algorithm are very small, comparable to the differences accepted between different human investigators. Important advantages of the elastic matching algorithm include accuracy and reproducibility. Moreover, the automated algorithm recognized and measured accurately specific basal ganglia volume differences between schizophrenic brains and normal controls. Further refining of the matching algorithm (Warfield *et al.*, 1996) will allow better measurements of the cortical structures. Therefore, the automated method used for elastic matching will prove to be a reliable and efficient tool for measuring complex anatomical brain structures and for recognizing subtle brain abnormalities associated with brain disease.

APPENDIX

Mathematical Model for Elastic Matching

The goal of the elastic matching algorithm is to find a 3D deformation vector field *u(x)* which transforms the source data set so that it matches the target data set. Let $g_1(x)$ be the source data set (atlas) and $g_2(x)$ be the target data set (patient). The deformation field u should maximize the local similarity of $g_1(x - u)$ and $g_2(x)$. This can be expressed as the problem of finding the deformation $u(x)$ that minimizes the squared differences between two corresponding image patches:

$$E[u(x)] = \int w(x)[g_2(x) - g_1(x - u)]^2 \, dx.$$

Here $w(x)$ represents a window function that defines a local region in each data set and is typically chosen to have a Gaussian shape, and E can be regarded as the energy of the deformation field or the error of the deformation.

A first-order Taylor series approximation for the value of image g_1 near x for a locally constant deformation u allows the difference between the target and the source image values to be written as

$$g_2(x) - g_1(x - u) = g_2(x) - g_1(x) + u(x)^T \nabla g_1(x).$$

So the functional to be minimized can be written

$$E[u(x)] = \int w[(g_2 - g_1)^2 + (u^T \nabla g_1)^2 + 2(g_2 - g_1)u^T \nabla g_1]dx.$$

Differentiating this expression with respect to u and solving for 0 leads to

$$\int w[(g_2 - g_1)\nabla g_1 + \nabla g_1 u^T \nabla g_1]dx = 0,$$

which can be written as

$$Qu = -F,$$

where

$$Q = \int w(x)\nabla g_1(x)\nabla g_1^T(x)dx$$

and

$$F = \int w(x)\nabla g_1(x)[g_2(x) - g_1(x)]dx.$$

Equation (1) is the estimation part of the functional. In order to regularize the equation a 3D elastic membrane smoothness constraint is introduced, leading to the model for the elastic matching

$$-\nabla^2 u + Qu = -F.$$

Solution of the Model

A FEM discretization of the model leads to a system of equations which are efficiently solved with a nested multigrid algorithm (e.g., Dengler *et al.*, 1988; Schmidt and Dengler 1989). The source and target data volumes are converted into a multiresolution pyramid. At the coarsest resolution the deformation field is assumed to

be 0 and the system of equations is solved. The resulting deformation field is then interpolated (with linear interpolation) to the next higher resolution where it is used as the initial deformation field and the system of equations is solved again. This procedure is repeated up to the finest resolution level.

The model is also valid for features calculated from the source and target data sets. Directly using the gray scale values of the image data can lead to problems caused by sensitivity to noise and contrast changes in the data. Instead the data are first classified with a robust multispectral statistical classification algorithm (Warfield *et al.*, 1995) and the tissue class data are used for matching. This improves the robustness of the algorithm to noise and makes possible an increase in the speed of the matching.

REFERENCES

Andreasen, N. C., Rajarethinam, R., Cizadlo, T., Arndt, S., Swayze, V. W., Flashman, L. A., O'Leary, D. S., Ehrhardt, J. C., and Yuh, W. T. C. 1996. Automatic atlas-based volume estimation of human brain regions from MR images. *J. Comput. Assisted Tomogr.* **20**(1):98–106.

Bajcsy, R., and Kovacic, S. 1989. Multiresolution elastic matching. *Comp. Vision Graphics Image Proc.* **46:**1–21.

Christensen, G. E., Rabbitt, R. D., and Miller, M. I. 1993. A deformable neuroanatomy textbook based on viscous fluid mechanics. *Proceedings of the 27th Annual Conference on Information Science and Systems,* Baltimore, pp 211–216.

Collins, D. L., Peters, T. M., Dai, W., and Evans, A. C. 1992. Model based segmentation of individual brain structures from MRI data. *SPIE Visualizat. Biomed. Comput.* **1808:**10–23.

Collins, D. L., Neelin, P., Peters, T. M., and Evans, A. C. 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J. Comput. Assisted Tomogr.* **18**(2): 192–205.

Dann, R., Hoford, J., Kovacic, S., Reivich, M., and Bajcsy, R. 1989. Evaluation of elastic matching systems for anatomic (CT, MRI) and functional (PET) cerebral images. *J. Comput. Assisted Tomogr.* **13**(4):603–611.

Dengler, J., Schmidt, M. 1988. The dynamic pyramid—A model for motion analysis with controlled continuity. *Int. J. Pattern Recognit. Artif. Intell.* **2:**275–286.

Dhawan, A. P., and Arata, L. 1992. Knowledge-based multi-modality three-dimensional image analysis of the brain. *Am. J. Physiol. Imaging* **7**(3–4):210–219.

Ettinger, G. J., Grimson, W. E. L., Lozano-Perez, T., Wells, W. M., White, S. J., and Kikinis, R. 1994. Automatic registration for multiple sclerosis change detection. *Proc. IEEE Biomed. Image Anal.* 297–306.

Evans, A. C., Dai, W., Collins, L., Neelin, P., and Marett, S. 1991. Warping of a computerized 3D atlas to match brain image volumes for quantitative neuroanatomical and functional analysis. *Image Processing* **1445:**236–246.

Gee, J. C., Reivich, M., and Bajcsy, R. 1993. Elastically deforming 3D atlas to match anatomical brain images. *J. Comput. Assisted Tomogr.* **17:**225–236.

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

Grenander, U., and Miller, M. I. 1994. Representation of knowledge in complex systems. *J. R. Stat. Soc. B* **56**(3).

Haller, J. W., Christensen, G. E., Miller, M. I., Joshi, S. C., Gado, M., Csernansky, J. G., and Vannier, M. W. 1995. Comparison of automated and manual segmentations of hippocampus MR images. *Proc. SPIE Med. Imaging 1995 Image Processing* **2434:**206–215.

Haller, J. W., Banerjee, A., Christensen, G. E., Gado, M., Joshi, S., Miller, M. I., Sheline, Y., Vannier, M. W., and Csernansky, J. G. 1997. Three-dimensional hippocampal MR morphometry with high-dimensional transformation of a neuroanatomic atlas. *Radiology* **202:**504–510.

Höhne, K. H., Bomans, M., Riemer, M., Schubert, R., Tiede, U., and Lierse, W. A. 1992. 3D anatomical atlas based on a volume model. *IEEE Comput. Graphics Appl.* **12:**72–78.

Hokama, H. H., Shenton, M. E., Nestor, P. G., Kikinis, R., Levitt, J. J., Metcalf, D., Wible, C. G., O'Donnell, B. F., Jolesz, F. A., and McCarley, R. W. 1995. Caudate, putamen and globus pallidus volume in schizophrenia: A quantitative MRI study. *Psychiatry Res. Neuroimaging* **61:**209–229.

Kikinis, R., Shenton, M. E., Gerig, G., Martin, J., Anderson, M., Metcalf, D., Guttmann, C. R. G., McCarley, R. W., Lorensen, W., Cline, H., and Jolesz, F. A. 1992. Routine quantitative analysis of brain and cerebrospinal fluid spaces with MR imaging. *J. Magn. Reson. Imaging* **2:**619–629.

Kikinis, R., Gleason, L., and Moriarty, T. M. 1996. Computer-assisted interactive three-dimensional planning for neurosurgical procedures. *Neurosurgery* **38**(4):640–649.

Kikinis, R., Shenton, M. E., Iosifescu, D. V., McCarley, R. W., Saiviroonporn, P., Hokama, H. H., Robatino, A., Metcalf, D., Wible, C. G., Portas, C. M., Donnino, R. M., and Jolesz, F. A. 1996. A digital brain atlas for surgical planning, model driven segmentation, and teaching. *IEEE Visualizat. Comput. Graphics* **2**(3):232–241.

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

Schmidt, M., and Dengler, J. 1989. Adapting multi-grid methods to the class of elliptic partial differential equation appearing in the estimation of displacement vector fields. In *Recent Issues in Pattern Analysis and Recognition, Lecture Notes in Computer Science* (V. Cantoni, R. Creutzburg, S. Levialdi, and H. Wolf, Eds.), pp. 266–274. Springer, Berlin/Heidelberg/New York/Tokyo.

Shenton, M. E., Kikinis, R., Jolesz, F. A., Pollak, S. D., LeMay, M., Wible, C. G., Hokama, H., Martin, J., Metcalf, D., Coleman, M., and McCarley, R. W. 1992. Abnormalities of the left temporal lobe and thought disorder in schizophrenia: A quantitative magnetic resonance imaging study. *N. Engl. J. Med.* **327:**604–612.

Shenton, M. E., Kikinis, R., McCarley, R. W., Saiviroonporn, P., Hokama, H. H., Robation, A., Metcalf, D., Wible, C. G., Portas, C. M., Iosifescu, D. V., Donnino, R., Goldstein, J. M., and Jolesz, F. A. 1995. Harvard brain atlas: A teaching and visualisation tool. *Proc. IEEE Biomed. Image Anal.* 10–17.

Warfield, S., Dengler, J., Zaers, J., Guttmann, C. R. G., Wells, W. M., Ettinger, G. J., Hiller, J., and Kikinis, R. 1996. Automatic identification of gray matter structures from MRI to improve the segmentation of white matter lesions. *J. Image Guided Surg.* **1**(6):326–338.

Wells, W. M., Grimson, W. E. L., Kikinis, R., and Jolesz, F. A. 1996. Statistical intensity correction and segmentation of MRI data. *IEEE TMI* **15**(4):429–443.

Wible, C. G., Shenton, M. E., Hokama, H., Kikinis, R., Jolesz, F. A., Metcalf, D., and McCarley, R. W. 1995. Prefrontal cortex and schizophrenia: A quantitative magnetic resonance imaging study. *Arch. Gen. Psychiatry* **52:**279–288.