Automatic 3-D Segmentation of Internal Structures of the Head in MR Images Using a Combination of Similarity and Free-Form Transformations: Part I, Methodology and Validation on Normal Subjects

B. M. Dawant,* *Member, IEEE*, S. L. Hartmann, *Student Member, IEEE*, J.-P. Thirion, F. Maes, D. Vandermeulen, and P. Demaerel

***Abstract*—The study presented in this paper tests the hypothesis that the combination of a global similarity transformation and local free-form deformations can be used for the accurate segmentation of internal structures in MR images of the brain. To quantitatively evaluate our approach, the entire brain, the cerebellum, and the head of the caudate have been segmented manually by two raters on one of the volumes (the reference volume) and mapped back onto all the other volumes, using the computed transformations. The contours so obtained have been compared to contours drawn manually around the structures of interest in each individual brain. Manual delineation was performed twice by the same two raters to test inter- and intrarater variability. For the brain and the cerebellum, results indicate that for each rater, contours obtained manually and contours obtained automatically by deforming his own atlas are virtually indistinguishable. Furthermore, contours obtained manually by one rater and contours obtained automatically by deforming this rater's own atlas are more similar than contours obtained manually by two raters. For the caudate, manual intra- and interrater similarity indexes remain slightly better than manual versus automatic indexes, mainly because of the spatial resolution of the images used in this study. Qualitative results also suggest that this method can be used for the segmentation of more complex structures, such as the hippocampus.**

***Index Terms*—Atlas-based segmentation, deformation, registration.**

I. Introduction

THE quantitative analysis of anatomical structures and substructures in MR images has recently become of great interest for the study of epilepsy, schizophrenia, and alcoholism. These studies require analyzing volumes and shapes of structures such as the hippocampus, the cerebellum, or the caudate, measuring the whole brain volume, evaluating brain atrophy, or quantitating the asymmetry between the two hemispheres. The large amount of data to be analyzed makes manual analysis of these images impractical, thus calling for automatic segmentation methods. Because of the lack of clearly defined edges, segmenting these structures remains a challenging task that will not be accomplished by designing algorithms that rely solely on information present in the image, but that also use *a priori* information. Methods such as deformable models or active contours and shapes [1] that can capture statistical information about the shape of structures of interest is a partial answer to this problem, but the initialization of these algorithms prior to deformation remains difficult. Another approach is to view segmentation as a registration task. The basic tenet of these techniques is that a transformation can be found that registers one image volume (called the reference or the atlas) in which structures of interest have been labeled to the volume to be segmented. If such a transformation can be computed, regions labeled in the atlas can simply be projected onto the volume of interest. The key to these approaches is thus to design methods capable of computing the transformation between the atlas and other image volumes in a reliable and accurate way.

In this paper, we distinguish between local and global transformations. Global transformations are defined as transformations that can be expressed with a few parameters, such as rigid body transformations (three rotation angles and three translation vectors) or similarity transformations (rigid body plus anisotropic scaling). Among the techniques proposed to compute these transformations, voxel-based methods have received a great deal of attention since the pioneering work of Woods [2]. In this approach, the main assumption is that each gray-level value in one image volume corresponds to one gray-level value in the other volume. Based on this assumption, Woods proposed a measure in which the variance of the gray-level ratios between corresponding pairs of voxels is minimized. Expanding on this idea, Hill *et al.* [3] proposed a series of features extracted from the two-dimensional (2-D) gray-

Manuscript received March 12, 1999; revised September 26, 1999. Part of this paper was presented at the 1998 SPIE Medical Imaging Conference. The work of B. M. Dawant was supported in part by KULeuven under Senior Fellowship F/96/40. The work of D. Vandermeulen was supported in part by the FWO-Vlaanderen under Grant 1.5.397.97. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was M. Vannier. *Asterisk indicates corresponding author*.

*B. M. Dawant is with the Department of Electrical and Computer Engineering, P.O. Box 1662-B, Vanderbilt University, Nashville, TN 37235, USA (e-mail: dawant@vuse.vanderbilt.edu).

S. L. Hartmann is with the Department of Biomedical Engineering, Vanderbilt University, Nashville, TN 37235 USA.

J.-P. Thirion is with Focus Medical, Sophia-Antipolis, France.

F. Maes and D. Vandermeulen are with the Department of Electrical Engineering—ESAT, Catholic University of Leuven, Belgium.

P. Demaerel is with the Department of Radiology, Catholic University of Leuven, Belgium.

Publisher Item Identifier S 0278-0062(99)09629-9.

level histogram (scatter plot) following the observation that the dispersion of the 2-D gray-value histogram of the images to be registered increases as the misregistration increases. More recently, a new measure of scatter-plot dispersion, called mutual information, has been proposed independently by Collignon [4], [5] and Wells [6]. This measure is robust and permits the automatic computation of the similarity transformation. In a large intersite comparison study involving intermodality intrapatient registration, voxel-based methods and, in particular, methods using mutual information, were shown to be the most robust and accurate techniques for rigid body registration [7]. Although no such study has been performed to evaluate the adequacy of this approach for similarity transformations, our experience indicates that it can also be used to compute these nine degrees of freedom transformations. However, similarity transformations are insufficient to take into account local and subtle differences between brains. Thus, brains registered with these transformations appear to be globally registered, but locally the registration remains inaccurate. This problem can be solved by computing local rather than global transformations and a number of methods of varying complexity have been proposed over the years to address this problem. Collins [8] has used a multiresolution approach in which the overall nonlinear transformation is composed of a set of local linear deformations obtained by maximizing the correlation of intensity and gradient features in the images to be registered. Bajcsy [9], [10] used an elastic model approach. Algorithms based on viscous fluid models were put forth by Christensen [11] and Bro-Nielsen [12]. Recently, Thirion [13] developed an approach which trades the rigor of physical modeling for simplicity of implementation and speed of execution.

Validation of these methods is difficult because of the lack of accepted and established gold standards. The aforementioned study designed to evaluate rigid-body transformations relied upon a gold standard provided by bone implantable markers. Transformations obtained with various registration algorithms were compared to results obtained with the markers. No such standard has been agreed upon for transformations more complex than rigid body. Yet, the testing and validation of these methods is of critical importance to understand their respective merits and demerits and to guide researchers in the field faced with the task of choosing a method for a particular application. In this study we have tested the hypothesis that a combination of a global similarity transformation and of a local free-form transformation computed using the idea of demons, put forth by Thirion, can be used for the automatic and accurate three-dimensional (3-D) segmentation of internal structures in MR images of the brain. The global transformation is computed using a mutual information based method and accounts for large scale and orientation differences. The local transformations permit the accommodation of small local anatomical differences between individual brains. Both these methods are fully automatic and have been applied to nine image volumes. Validation has been performed both visually and quantitatively. To test the method visually we have generated images that show that structures, such as the central sulcus or the hippocampus, are well registered. To test the method quantitatively, we have compared contours obtained manually and automatically for a number of structures ranging from the head of the caudate to the whole brain. Manual delineation was performed by two raters and we have computed indexes of similarity between contours obtained manually and contours obtained automatically.

II. Methods

A. Data

Nine magnetization prepared rapid acquisition with gradient echo (MP-RAGE) image volumes (four females and five males) were acquired sagittally with a Siemens 1.5T MR scanner and the following acquisition parameters: TR 9.7 ms, TE 4 ms, Flip angle 12°, slab thickness 160 mm, effective slice thickness 1 mm, pixel size $1 \times 1\text{mm}^2$. Among the nine volumes, one was chosen as the reference volume or atlas and the other ones were used to test the algorithm. The atlas volume was chosen at random.

B. Similarity Transformation

The similarity transformations (three rotation angles, three translation vectors, and three scaling factors) have been computed using the mutual information criterion previously mentioned and the MIRIT software developed at the Catholic University of Leuven [14]. First, this algorithm was used to register the atlas to another image volume already transformed into the Talairach space, obtained from the Montreal Neurological Institute, courtesy of Dr. Alan Evans. This volume consists of $217 \times 181 \times 181$ isotropic 1 mm^3 voxels. This step placed our own atlas in the Talairach space. Each of the remaining eight volumes were then registered to our atlas and reformatted using a trilinear interpolation method. After this series of steps, all brain volumes used in this study are in Talairach space where measurements are made.

C. Free-Form Transformation

The free-from transformation method used in this work is the one proposed by Thirion and it is based on the concept of demons [13]. Although the concept of demons provides a framework, these can be implemented and designed in a number of ways. In this application, we have used the instantaneous optical flow equation as presented in [15]. The hypothesis is that the intensity of points in the images is preserved under motion i.e., $I(x(t), y(t), z(t)) = \text{constant}$. Differentiating this equation leads to

$$\frac{\partial i}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial i}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial i}{\partial z}\frac{\partial z}{\partial t} = -\frac{\partial i}{\partial t}. \tag{1}$$

In our case, we consider the two volumes to be registered as two time frames f and g and we are looking for a displacement $\vec{v}$ that brings the two volumes in local correspondence. Thus, we assume that f and g are separated by one unit of time. Therefore, $\partial i/\partial t = f - g$ and $\vec{v} = (dx/dt, dy/dt, dz/dt)$ is the instantaneous velocity from g to f. Using this model, $\vec{v} \cdot \vec{\nabla} f = f - g$. Since this equation is not sufficient to compute $\vec{v}$ locally, it is usually determined using regularization

techniques. An alternative is to use the projection of the velocity vector on the direction of the spatial gradient, which leads to the following expression for the local displacement vector:

$$\vec{v} = \frac{(g-f)\vec{\nabla} f}{\vec{\nabla} f^2}. \tag{2}$$

Although is it possible to use (2) to compute the local displacement, this equation is unstable when the gradient norm is small. To address this problem, we compute the local displacement vector as

$$\vec{v} = \frac{(g-f)\vec{\nabla} f}{\vec{\nabla} f^2 + (g-f)^2}. \tag{3}$$

The algorithm is thus iterative and it is applied in a multiscale way. The matching is first computed on coarse downsampled images, then successively to images with a finer spatial resolution. This strategy presents several advantages. It speeds up the computations, improves the convergence properties of the algorithm, and it uses the fact that, for human anatomy, macroscopic features are, in general, more stable than microscopic features. In the implementation used in this work, two image pyramids are derived from the images to be registered, up to a predetermined scale. A number of iterations of the algorithm are applied to the images at the coarsest scale and the results obtained at this scale serve as initial conditions for the next one, until the finer scale is reached. At each iteration, a displacement vector is computed for each pixel, thus generating a deformation field. The final deformation applied to the images can be totally free form, i.e., each pixel could be moved by the computed displacement vector. For anatomic images, this approach is unrealistic. Voxels located close to each other should be displaced by comparable amounts. This type of constraint can be imposed by smoothing the instantaneous deformation field with, for instance, a Gaussian filter with standard deviation σ. The larger σ, the less deformable the images. Furthermore, an additional mechanism is used to ensure a one-to-one correspondence between the two images to be matched. This is done by computing both a direct and a reverse deformation field, which are maintained compatible through iterations in a way similar to the one proposed by Burr [16]. In our experience, this greatly increases the robustness of the algorithm. It also has the great advantage of providing both a forward (i.e, from the volume to the atlas) and a reverse (i.e., from the atlas to the volume) transformation. Applying the algorithms to sets of images thus requires only the selection of a number of parameters: the number of levels in the image pyramids; the number of iterations at each scale; and the standard deviation of the smoothing filter, which determines the rigidity of the transformation. In this application, these parameters have been chosen once as follows: four levels in the image pyramids (32 $\times$ 32 $\times$ 32; 64 $\times$ 64 $\times$ 64; 128 $\times$ 128 $\times$ 128; and 217 $\times$ 181 $\times$ 181 voxels, respectively), 256 iterations of the algorithm were applied at the highest level, 128, 64, and 32 iterations were applied at levels 3, 2, and 1. The standard deviation of the Gaussian smoothing filter was set to one at every level. The entire registration procedure including both the computation of the similarity and of the free from transformations takes approximately 90 min. on a Sun Ultra 1.

D. Contour Delineation

Quantitative evaluation of segmentation techniques is notoriously difficult, mainly because of the lack of a gold standard. Here we have used manual delineation as the reference method and we have used the following protocol. Two human raters: one experienced radiologist (P.D.) and one nonexpert (B.D.) delineated contours for three structures (the whole brain, the cerebellum, and the head of the left caudate) on the atlas. We have limited our quantitative study to the head of the caudate because the spatial resolution of the images we have used makes the precise delineation of the tail of the caudate difficult. These contours were subsequently used to create three binary volumes, one for each structure. Using the free-form transformations computed between the atlas and each of the remaining eight volumes, these binary volumes were mapped onto each of the volumes in the test set using shape-based interpolation [17].

To quantitatively compare contours obtained with the automatic method described above and contours obtained manually by a trained observer, three slices were chosen for each structure and each of the remaining volumes as follows. For each volume and for each structure, the range of images in which the structure is visible has been manually determined. Three numbers were generated by a random number generator with a uniform distribution within this range. Contours were subsequently drawn manually on the slices whose number matched the random values. This resulted in 24 contours (8 volumes $\times$ 3 slices) per structure for a total of 72 contours (24 contours $\times$ 3 structures) per rater. Each rater performed this procedure twice to test intrarater variability. This evaluation strategy was followed because it would have been too time consuming to draw the entire structures for each of the volumes in the test set. Contours obtained manually on the selected slices were then compared with each other and to contours obtained for the same slices with the automatic technique. Contours were quantitatively compared using a similarity index defined as follows:

$$S = \frac{2N(C1 \cap C2)}{N(C1) + N(C2)} \tag{4}$$

with $N(\cdot)$ the number of pixels included in a region, and $C1$ and $C2$ the manual and automatic contours, respectively. This index ranges from zero to one, with zero indicating no overlap and one indicating a perfect agreement between two contours. It is related to a reliability measure known as the kappa statistic [18] and it is sensitive to both differences in size and in location of the two contours being compared. In the case of the cerebellum, we were interested in measuring the volume of brain parenchyma rather than measuring the volume of its envelope. To do so, a thresholding operation was applied before the similarity indexes were computed to exclude CSF from the regions encircled by the contour. To determine the threshold, each slice was examined visually and

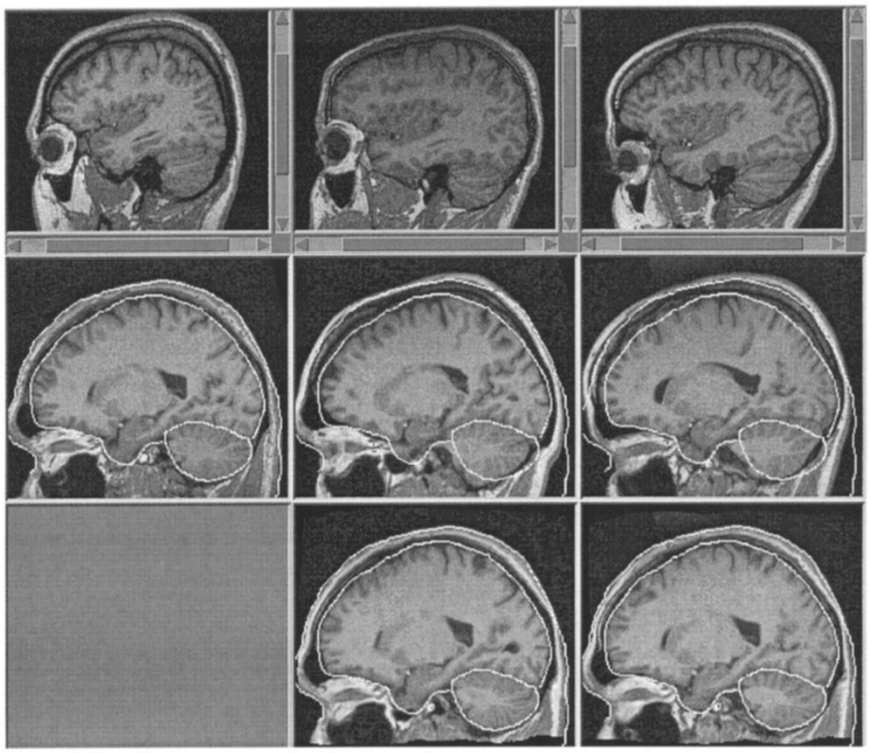

Fig. 1. Top row: the slice with the same index in three of the nine volumes used in this study prior to registration and reformatting. Middle row: the same three volumes in Talairach space using a similarity transformation. Bottom row: the two rightmost volumes have been registered and reformatted to the left one using the free-form transformation (see text for additional explanations).

a threshold was selected (in this study no interslice intensity correction was attempted). The same threshold was applied for the manual and the automatic segmentations.

III. Results

A. Qualitative Evaluation

In this section, several figures illustrating the type of results we have obtained are presented. The top row of Fig. 1 shows the slice with the same index (e.g., slice 80) in three of the nine original volumes before any type of registration or reformatting. It illustrates the range of head sizes and shapes that were included in the experiment. The left panel on the second row shows one slice in our atlas in Talairach space. The other two panels on this row show the corresponding slice in two other volumes registered and reformatted to our atlas using a similarity transformation. Contours of the head, the brain, and the cerebellum have been drawn on the leftmost image and copied on the other two. If a similarity transformation was able to account for all the differences between volumes, these contours should also precisely encircle the corresponding structure in each of the other volumes. Clearly this is not the case. The third row shows the same two volumes registered and reformatted to the atlas using the free-form transformation. As opposed to the situation illustrated on the panels in the second row, the contours now accurately encircle the structures of interest in each of the volumes, thus showing the ability of the free-form transformation to bring these volumes into local correspondence.

Fig. 2 further illustrates the difference between the similarity and the free-form transformation. The leftmost panels on the top and bottom rows are the same and are renderings of the cortical surface of our atlas. The other panels show renderings of two other volumes registered to the atlas using the similarity transformation (top row) and using the free-form transformation (bottom row). The lateral sulcus has been outlined on the atlas and copied on the other renderings. Observe the considerable improvement in the realignment of the lateral sulcus when the free-form transformation is used. Observe also that the free-form deformation algorithm maintains the integrity of the cortical surface and preserves topological differences between volumes.

Fig. 3 illustrates the type of segmentation results that have been obtained for the caudate. From left to right, this figure shows a sagittal, coronal, and transverse view of one of the volumes included in the study. Overlaid in white are the contours that have been obtained automatically by deforming our atlas.

Fig. 4 compares manual (white) and automatic (black) caudate contours for one representative slice in four of our volumes. Similarity indexes obtained for these four images are, from left to right: 0.80, 0.89, 0.87, and 0.89.

Fig. 5 illustrates the performance of the algorithm on more complex structures such as the hippocampus. The top four panels show a portion of a slice in four of the subject volumes in which the hippocampus is visible. The corresponding bottom panels show the same slice in the volumes obtained by registering and reformatting the atlas to the corresponding subject volume. Contours overlaid in white

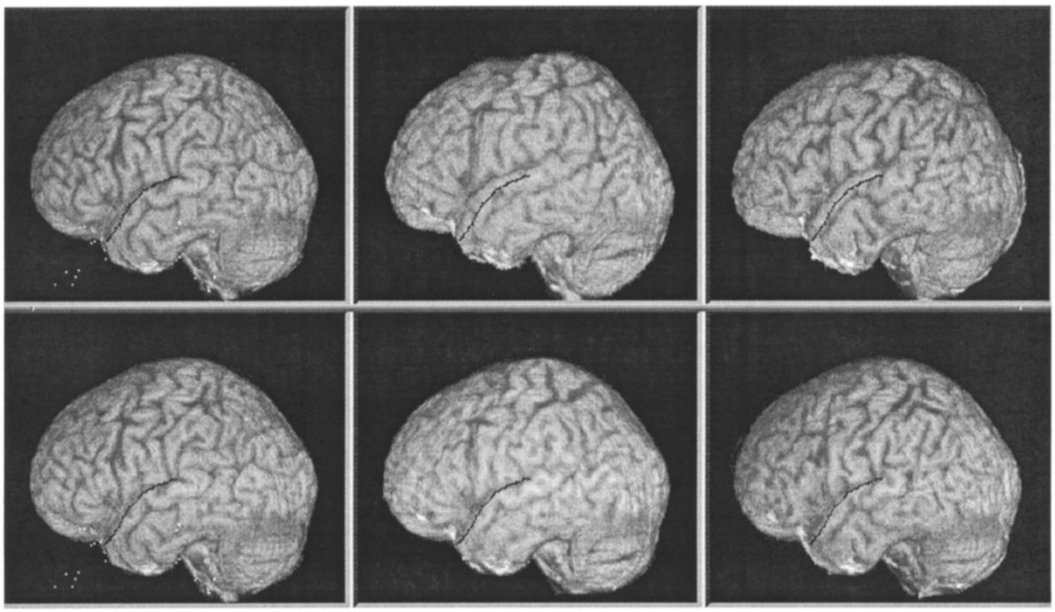

Fig. 2. Top row: rendering of the cortex for the atlas used in this study (left) and two other volumes registered to the atlas using a similarity transformation. Bottom row: rendering of the cortex for the atlas used in this study (left) and two other volumes registered to the atlas using a similarity transformation. The lateral sulcus has been drawn on the atlas image and copied on all the panels.

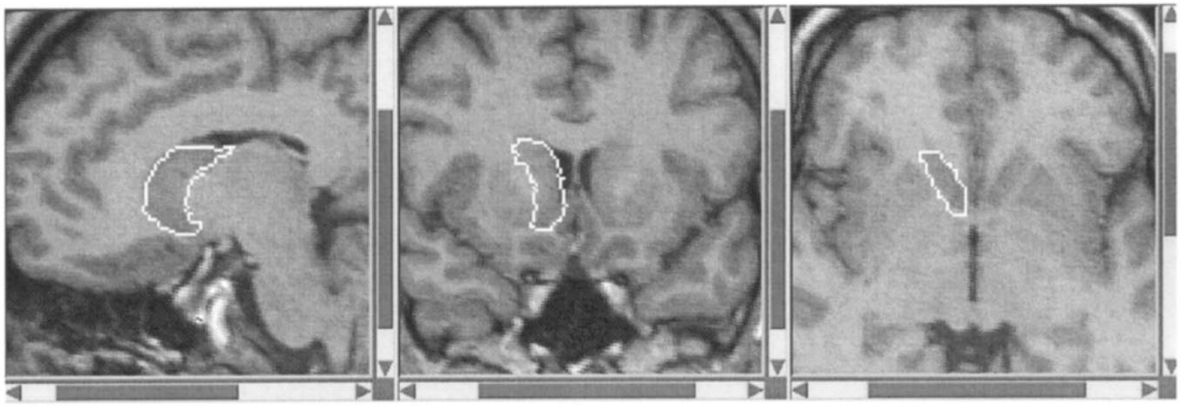

Fig. 3. From left to right; sagittal, coronal, and transverse views of caudate contours obtained automatically by deforming the atlas.

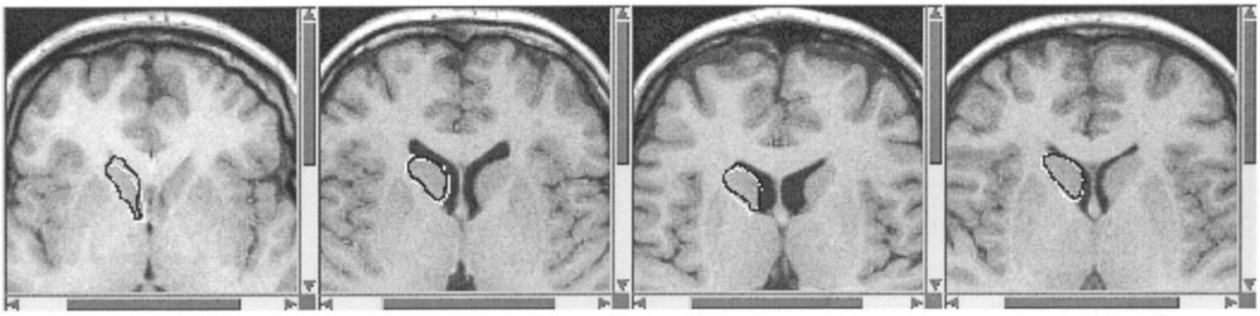

Fig. 4. Comparison of manual (white) and automatic (black) caudate contours for one representative slice in four volumes.

on the top panels are the hippocampus contours obtained automatically by deforming the atlas.

B. Quantitative Evaluation

The contours we have obtained for this study can be used to evaluate a number of measures related to intrarater and interrater variability, as well as differences between manual and automatic methods. Among all the possible combinations, we have selected the following: manual intrarater variability; manual interrater variability; and differences between manual and automatic methods. In the ensuing discussion, the following convention has been used to identify an individual rater

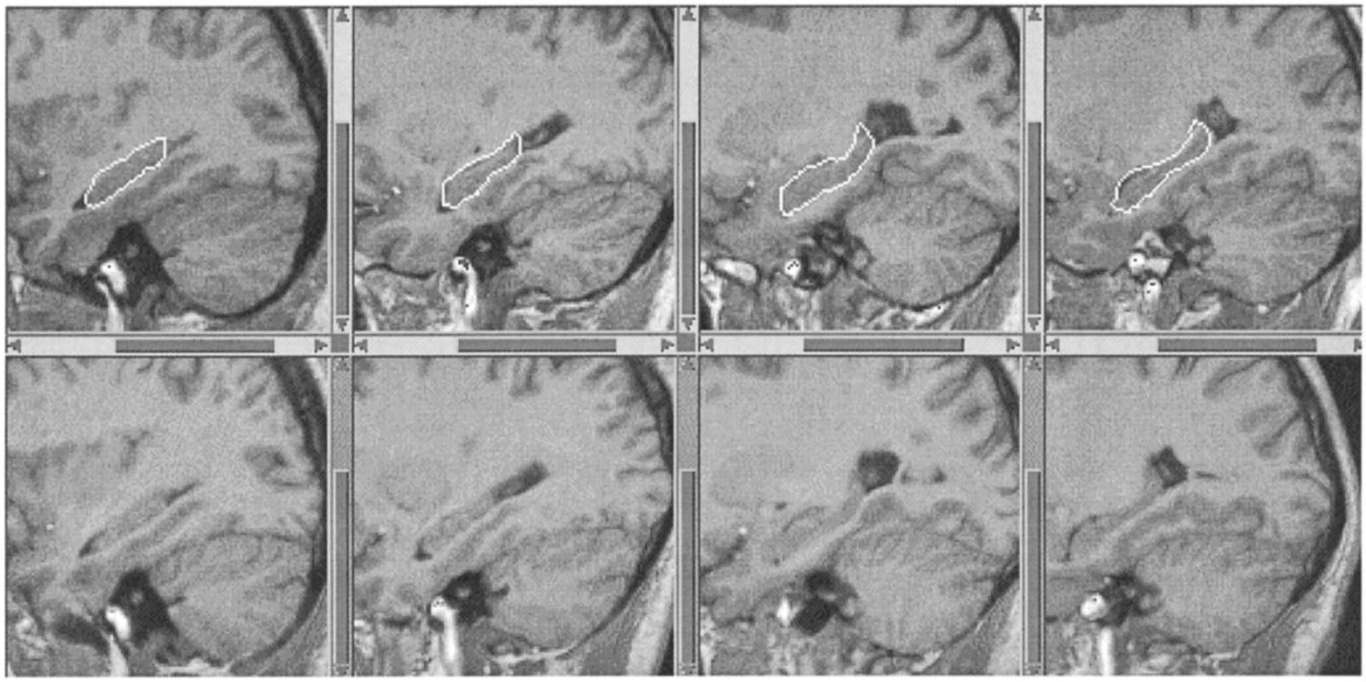

Fig. 5. Illustration of the algorithm performance on more complex structures. Top row: portion of a slice showing the hippocampus in four volumes. Bottom row: same slice obtained by registering and reformatting the atlas to each of the corresponding volumes. Contours overlaid in white are hippocampus contours obtained automatically by deforming the hippocampus delineated on the atlas.

TABLE I
INTRARATER SIMILARITY INDEXES FOR THE BRAIN, THE CEREBELLUM, AND THE CAUDATE. EACH ENTRY REPORTS THE MEAN VALUE AND, IN PARENTHESIS, THE MINIMUM, MAXIMUM, AND STANDARD DEVIATION OF 24 CONTOURS

Manual Intra-rater		
	R1	R2
Brain	0.97 (0.94, 0.98, 0.001)	0.97 (0.96, 0.98, 0.007)
Cerebellum	0.97 (0.94, 0.99, 0.011)	0.99 (0.96, 0.99, 0.008)
Caudate	0.88 (0.78, 0.94,0.036)	0.9 (0.86, 0.94, 0.02)

TABLE II
INTERRATER SIMILARITY INDEXES FOR THE BRAIN, THE CEREBELLUM, AND THE CAUDATE. EACH ENTRY REPORTS THE MEAN VALUE AND, IN PARENTHESIS, THE MINIMUM, MAXIMUM, AND STANDARD DEVIATION OF 24 CONTOURS

Manual Inter-rater		
Brain		
	R11	R12
R21	0.95 (0.85, 0.97, 0.02)	0.95 (0.89, 0.97, 0.018)
R22	0.96 (0.87, 0.97, 0.023)	0.95 (0.91, 0.97, 0.015)
Cerebellum		
	R11	R12
R21	0.97 (0.94, 0.99, 0.012)	0.97 (0.93, 0.99, 0.013)
R22	0.97 (0.95, 0.99, 0.011)	0.97 (0.94, 0.00, 0.01)
Caudate		
	R11	R12
R21	0.89 (0.78, 0.93, 0.038)	0.89 (0.82, 0.95, 0.034)
R22	0.89 (0.79, 0.93, 0.034)	0.89 (0.81, 0.94, 0.04)

and the sets of contours he has drawn. R_{ij}, with $i = \{1, 2\}$ the rater identity, and $j = \{1, 2\}$, the index of the delineation. For instance, $R12$ indicates the set of contours drawn by rater 1 the second time he drew these contours. Atlas1 and Atlas2 are the atlases created by the two raters (one each).

1) Manual Intrarater Variability: Table I reports the indexes we have computed to measure intrarater variability for each of the three structures we have studied. These were obtained by computing indexes of similarity between contours drawn on the eight test volumes and averaging them structure by structure. Individual similarity indexes were computed between the first and the second manual delineation by a particular rater. Each entry in this table reports the mean and, in parenthesis, the minima, maxima, and standard deviations of 24 (eight image volumes, three slices per image volume) indexes of similarity. $R1$ relates to rater 1, $R2$ relates to rater 2.

2) Manual Interrater Variability: Table II reports the numbers we have computed to evaluate manual inter-rater variability for the head, the cerebellum, and the caudate. Again, each entry in these tables reports averages, minima, maxima, and standard deviations of 24 similarity indexes.

3) Indexes of Similarity Between Manual and Automatic Delineation: Table III reports indexes of similarity between manual and automatic delineation for each structure. In this table, Atlas1 and Atlas2 indicate contours obtained automatically by deforming the atlas created by raters 1 and 2, respectively. For instance, the entry {Atlas1, R11} measures the similarity between contours obtained automatically using the atlas of rater 1 with the first set of contours manually delineated by rater 1. Similarly, entry {Atlas1, R22} measures the similarity between contours obtained automatically using the atlas of rater 1 and the second set of contours drawn by rater 2.

TABLE III
SIMILARITY INDEXES COMPUTED BETWEEN MANUAL CONTOURS AND CONTOURS OBTAINED AUTOMATICALLY FOR THE BRAIN, THE CEREBELLUM, AND THE CAUDATE. EACH ENTRY REPORTS THE MEAN VALUE AND, IN PARENTHESIS, THE MINIMUM, MAXIMUM, AND STANDARD DEVIATION OF 24 CONTOURS

Manual vs. automatic

Brain

	R11	R12	R21	R22
Atlas 1	0.96 (0.92, 0.97, 0.011)	0.96 (0.88, 0.97, 0.017)	0.94 (0.83, 0.97, 0.03)	0.95 (0.84, 0.97, 0.02)
Atlas 2	0.95 (0.89, 0.97, 0.18)	0.95 (0.89, 0.97, 0.018)	0.96 (0.88, 0.97, 0.018	0.96 (0.89, 0.98, 0.018)

Cerebellum

	R11	R12	R21	R22
Atlas 1	0.97 (0.94, 0.99, 0.011)	0.97 (0.94, 0.99, 0.011)	0.97 (0.94, 0.99, 0.011	0.97 (0.95, 0.99, 0.01)
Atlas 2	0.97 (0.94, 0.99, 0.12)	0.97 (0.94, 0.99, 0.010)	0.98 (0.96, 0.99, 0.007	0.98 (0.97, 0.99, 0.006)

Caudate

	R11	R12	R21	R22
Atlas 1	0.84 (0.70, 0.93, 0.06)	0.85 (0.69, 0.92, 0.06)	0.84 (0.73, 0.94, 0.05)	0.86 (0.74, 0.93, 0.05)
Atlas 2	0.83 (0.7, 0.92, 0.06)	0.85 (0.73, 0.93, 0.05)	0.84 (0.74, 0.94, 0.05)	0.86 (0.75, 0.94, 0.05)

IV. DISCUSSION

A number of conclusions can be drawn from the indexes of similarity presented in this manuscript. First of all, for the whole head and the cerebellum, contours obtained manually and automatically are virtually indistinguishable. Manual intrarater similarity indexes are slightly higher than manual versus automatic similarity indexes for the head and are the same for the cerebellum (for the brain the average indexes are 0.97 for manual versus manual and 0.96 for manual versus automatic). Indexes are higher for the cerebellum because of the threshold that has been applied. Slight differences between contours are eliminated by this operation. More importantly, however, these results reveal that manual interrater similarity indexes are slightly lower than manual versus automatic similarity indexes obtained between a rater and his own atlas for the brain and cerebellum. This finding is important for large longitudinal studies. In these studies, contours and volumes need to be determined for a large number of data sets. It is thus highly unlikely that a single human expert could perform this time consuming and tedious task by him/herself. It is more probable that the task will be distributed among a team of raters. The numbers we have obtained show that more consistent results would be obtained if one relied upon a single atlas and its deformation rather than on a team of raters. Thus, results obtained with the data set used herein indicate that for structures such as the head and the cerebellum, an automatic method would be preferable to a team of human raters. For the head of the caudate, the same claim cannot be made. Manual intra- and interrater similarity indexes remain slightly better than the manual versus automatic indexes. The average interrater index is 0.89, while the average index of similarity between the manual contours of a rater with contours obtained automatically from his own atlas is 0.85. Even though, for this structure, one cannot claim that the automatic method is superior to a team of raters, the differences we report are extremely small. In fact, these differences can be imputed to the spatial resolution of the images, rather than to the inaccuracy of the deformation algorithm. Indeed, the head of the caudate is a small structure that appears ellipsoidal on the transverse images on which it has been delineated. Typical values for the long and short axes are 15 and 8 pixels, leading to an area of 377 pixels. Suppose the delineation error is modeled as an error in the value of the short and long axes. A similarity index of 0.89 corresponds to an error of one pixel in the length of the long and short axes (this corresponds approximately to a contour that is drawn half-a-pixel inside the true contour). An index of similarity of 0.84 corresponds to an error of 1.5 pixels in the length of the long and short axes, thus adding approximately another $\frac{1}{4}$ of a pixel to the previous error. Following this model, the average marginal error introduced by the automatic technique is thus a mere $\frac{1}{4}$ pixel along the contour of the structure. Because of the spatial resolution of the image, errors in the contours would be introduced even if the deformation algorithm was perfectly accurate. Indeed, recall how automatic contours are obtained. First, contours are delineated on a slice-by-slice basis on the atlas. From these slices, a 3-D binary volume is created which is projected onto each of the other volumes using a 3-D transformation. But, the shape of the head of the caudate changes considerably from slice to slice and its 3-D representation based on a stack of contours is thus inaccurate. It is this inaccurate 3-D representation that is reformatted and projected onto each of the other volumes. As discussed earlier, shape-based interpolation has been used to minimize errors caused by this process, but it cannot be eliminated completely. Inter- and intrarater similarity indexes are computed differently. A slice is selected in one volume and contours are simply drawn by each rater on this slice. Intra- and interrater variability measured in this way only reflects the difficulty to delineate this structure due to the in-plane spatial resolution. When comparing manual to automatic contours, errors due to the spatial resolution in the third dimension are also included. Possible solutions to this problem include increasing the spatial resolution of the images or improving on the method used to create the 3-D volume in the atlas. For instance, splines could be used to construct better models of the structures of interest. These models could then be deformed and projected onto volumes to be segmented.

V. Conclusion

The results presented in this manuscript indicate that a combination of global and local transformations can be used to segment fully automatically MR images of the head. Our experience also indicates that the global transformation is important for the success of the overall process. Without a good starting point, the free-form transformation can lead to disappointing results. The only manual intervention currently required is the selection of the intensity threshold used to eliminate CSF from the cerebellar contours. Automating this step is not very difficult but it was not implemented when we performed the validation study. Quantitative results indicate that an automatic approach is as good as a manual approach for structures the size of the cerebellum or the brain. For smaller structures with simple shapes such as the head of the caudate, the spatial resolution of the images limits the attainable accuracy, a problem that can be addressed by acquiring data volumes with a smaller voxel size. Qualitative results obtained so far also indicate that the method we propose can be used to segment structures with more complex shapes such as the hippocampus. Further quantitative evaluations need to be performed to assess the attainable accuracy with this type of structure and the method we propose will have to be compared to semi-automatic methods such as the one used in [19]. In this work, segmentation is performed in two steps. First, 16 landmarks are manually identified around the hippocampus to compute an affine transformation between the atlas and the volume to segment. Next, subvolumes that contain the hippocampus in both volumes are extracted, and a viscous fluid transformation is computed on these subvolumes to warp the atlas onto the other volume.

Although a side-by-side comparison of the various methods that have been proposed for nonrigid registration is highly desirable, it is currently difficult to achieve because of the lack of agreed upon standard and because these algorithms are, in general, not in the public domain. Based on published results, the method described herein has the advantage of speed. For instance, the method proposed in [8] requires 5–6 h on an SGI Origin 200 for downsampled volumes and the method proposed in [11] takes up to 9 h on a MASPAR parallel computer (as mentioned, ours takes about 90 min for a full resolution volume).

Acknowledgment

B. M. Dawant is grateful to Prof. Suetens and Prof. Marchal who have provided him with all the support required to conduct this research while on sabbatical in the Medical Image Computing Laboratory (Radiology-ESAT) of the Catholic University of Leuven.

References

[1] T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam, "The use of active shape models for locating structures in medical images," *Image and Vision Computing*, 1994, vol. 12, no. 6, pp. 355–366.

[2] R. Woods, J. Maziotta, and S. Cherry, "MRI-PET registration with automated algorithm," *J. Comp. Assisted Tomogr.*, 1993, vol 17, no. 2, pp. 536–546.

[3] D. Hill, C. Studholme, and D. Hawkes, "Voxel similarity measures for automated image registration," In *Proc. Third Conf. Visualization Biomedical Computing*, 1994, pp. 205–216.

[4] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marshal, "Automated multimodality image registration using information theory," in *Information Processing in Medical Imaging*, Y. Bizais, C. H. Barillot, and R. Di Paloa, Eds. Doordrecht, The Netherlands: Kluwer, 1995, pp. 263–274.

[5] F. Maes, A. Collignon, D. Vanderneulen, G. Mar.al, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, 1997.

[6] W. M. Wells, III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Med. Image Anal.*, vol. 1, no. 1, pp. 35–51, 1996,

[7] J. West, J. M. Fitzpatrick, Y. Wang, and B. M. Dawant *et al.* "Comparison and evaluation of retrospective intermodality image registration techniques," *J. Comp. Assisted Tomogr.*, 1997, vol. 21, no. 4, pp. 554–566.

[8] D. L. Collins, P. Neelin, T. M. Peters, and A. C. Evans, "Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space," *J Comp. Assisted Tomogr.*, vol. 18, no. 2, pp. 192–205, 1994.

[9] R. Bajcsy and S. Kovacic, "Multiresolution elastic matching," *Comp. Vision, Graph. Image Processing*, vol. 46, pp. no. 1, 1–21, 1989.

[10] J. C. Gee, M. Reivich, and R. Bajcsy, "Elastically deforming 3D atlas to match anatomical brain images," *J. Comp. Assisted Tomogr.*, vol. 17, no. 2, pp. 225–236 1993.

[11] G. E. Christensen, M. I. Miller, and M. Vannier, "3D brain mapping using a deformable neuroanatomy," *Phys. Med. Biol.*, vol. 39, no. 3, pp. 609–618, 1994.

[12] M. Bro-Nielsen and C. Gramkow, "Fast fluid registration of medical images" in *Proc Visualization Biomedical Computing Conf.* (Lecture Notes in Computer Science) K. H. Hohne and R. Kikinis, Eds. Berlin, Germany: Springer-Verlag, 1996, vol. 1131, pp. 267–276.

[13] J. P. Thirion, "Non-rigid matching using demons," Image matching as a diffusion process: An analogy with Maxwell's demons, *Med. Image Anal.*, vol. 2, no. 3, pp. 243–260, 1998.

[14] F. Maes, A. Collignon, D. Vandermeulen, and P. Suetens, "Multimodality image registration using information theory (MIRIT)," Version 97/08 K. U. Leuven, ESAT, Leuven, Belgium, Int. Rep. KUL/ESAT/MI2/9708, Sept. 1997.

[15] B. P. K. Horn and B. G. Schunck, "Determining optical flow," *Artific. Intell.*, vol. 17, no. 2, pp. 185–203, 1981.

[16] D. J. Burr, "A dynamic model for image registration," *Comp. Graph. Image Processing*, vol. 15, pp. 102–112, no. 2, 1981.

[17] S. P. Raya and J. K. Udupa, "Shape-based interpolation of multidimensional objects," *IEEE Trans. Med. Imag.*, no. 1, vol. 9, pp. 32–42, 1990.

[18] A. P. Zijdenbos, B. M. Dawant, and R. Margolin, "Morphometric analysis of white matter lesions in MR images: Method and validation," *IEEE Trans. Med. Imag.*, vol. 13, pp. 716–724, Dec. 1994.

[19] J. W. Haller, A. Banerjee, and G. E. Chistensen *et al.* "Three-dimensional hippocampal MR morphometric with high-dimenstional transformation of a neuroanatomic atlas," *Radiology*, vol. 202, no. 2, pp. 504–510, 1997.