Computer Vision and Image Understanding **77,** 192–210 (2000)
doi:10.1006/cviu.1999.0815, available online at http://www.idealibrary.com on IDEAL®

Average Brain Models: A Convergence Study

Alexandre Guimond

Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, Montréal, Québec H3C 3JC, Canada; and INRIA Sophia Antipolis, Projet Epidaure, France

Jean Meunier

Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, Montréal, Québec H3C 3J7, Canada

and

Jean-Philippe Thirion

INRIA Sophia Antipolis, Projet Epidaure, France

Received December 22, 1998; accepted September 14, 1999

We present a completely automatic method to build stable average anatomical models of the human brain using a set of magnetic resonance (MR) images. The models computed present two important characteristics: an average intensity and an average shape, both in a single image. We provide results showing convergence toward the centroid of the image set used for the computation of the model. In particular, the RMS distances between the model and the MR images contained in the set stabilize in a range of 2.88 to 3.36 mm from a range of 4.62 to 5.51 mm initially after only one iteration. As for the influence of the reference image chosen for the model construction, this is minimal with differences of about 1.0 mm, from approximately 3.5 mm initially. These results ensure the usefulness of our approach. © 2000 Academic Press

1. INTRODUCTION

An important tool used to diagnose abnormal anatomical variations are medical atlases [1]. Traditional ones, such as those by Talairach and Tournoux [2] or Schaltenbrand and Wahren [3], are presented in textbooks, but computerized atlases comprising information in a more practical and quantitative manner are becoming available [4–16]. They usually include information obtained from a set of subjects, as opposed to a single individual in most paper atlases, making them more representative of a population. For example, the

1077-3142/00 $35.00

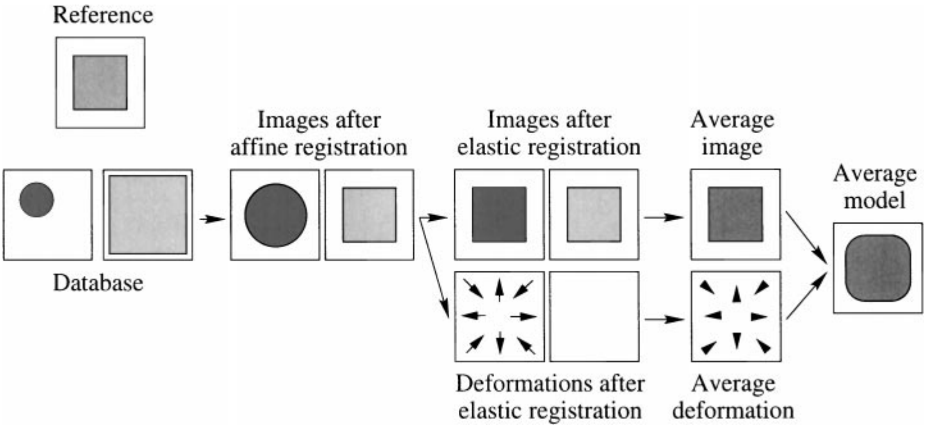


FIG. 1. Average model construction method.

Montreal Neurological Institute (MNI) used 305 normal subjects to build an atlas comprising intensity variations after affine registration in the stereotactic space defined by Talairach and Tournoux [8]. These methods also enable the calculation of normal shape variations, such as in the works of Gee *et al.* [17], which present a statistical framework for the construction of upgradable statistical atlases, and Thompson and Toga [18], which presents a probabilistic atlas of the human brain based on random vector field transformations.

The following work aims to develop and validate a concept drafted in a previous paper [19] to build an average model of the human brain using a set of magnetic resonance (MR) images obtained from normal subjects. This model has two important characteristics: average tissue intensity and average tissue shape up to an affine transformation. We intend to demonstrate that the model construction converges toward the centroid of the MR image set.

As depicted in Fig. 1, our method can be summarized in the following manner. Affine registration between all the images of the set and a reference image corrects for positioning and global shape differences due to translation, rotation, scaling, and shearing, as well as global linear intensity variations due to acquisition parameters or preprocessing. These are variations that are not of concern for our study. Elastic registration is then used to evaluate residual variations due to pure morphological differences and produce images having the same shape as the reference. Averaging the residual deformations, which are the inverses of the transformations that map each image of the set to the reference image, and the locally registered images yields an average deformation and an average intensity image, respectively. The average deformation is then applied to the average image to produce the model. It presents in a single image an average intensity and shape modulo an affine transformation corresponding to the affine characteristics of the reference image.

Although similar in terminology, the average shape and the average intensity characteristics have different purposes. The average shape represents an average of *morphological variations*. This is generally a well-understood concept. On the other hand, the average intensity is meant to average the *response of corresponding tissues to the acquisition parameters*. Also, assuming Gaussian noise in each MR scan, the average intensity increases the signal to noise ratio in the resulting model compared with individual scans.

The main contribution of this paper is the description of a fully automatic technique to obtain an *average intensity and shape* image, producing the *average model* M, and to show that this model, up to an affine transformation, is stable with respect to the choice of the initial reference image and repeated applications of the algorithm (iterations).

The most similar work regarding average intensity atlases is that of Bookstein [20] who created from nine MR scans a two-dimensional image representing the average intensity of the mid-sagittal plane. Thirteen manually identified landmarks in the mid-sagittal plane of each scan where matched with a reference image using the thin-plate spline interpolant [21]. The nine resampled images where then averaged to result into a morphometric average atlas. Our method differs primarily by two aspects. First, as suggested by Bookstein [20], we make full use of the three-dimensionality of the scans to compute a three-dimensional average image. Second, our registration method is automatic and computes a dense deformation field instead of an interpolated function based on 13 landmarks. This deformation identifies for each voxel of the reference the corresponding positions in the other scans. Within this process, every voxel of the reference can be thought of as a landmark automatically determined in the other scans.

The work of the MNI group [8], where 305 three-dimensional MR scans were registered using translations, rotations, and scalings and averaged to build a statistical neuroanatomical model, also relates to our work. We enrich this idea by proceeding further in using a less constrained type of deformation after the affine match to accommodate for local shape variations.

In the same vein, Woods *et al.* [13] describe a method that finds from a set of images a common space that preserves the average orientation, size, and affine shape of the group by registering all possible pairs of images it contains. Averaging the images after affine mapping to this common space produces an average intensity brain atlas in the average affine space. Their method is computationally very intensive as it requires $n(n-1)/2$ registrations, n being the number of subjects in the group. The basic difference between this approach and the one by the MNI group is that this one finds the average affine space, whereas the MNI method uses Talairach space.

The average shape concept is most similar to the beautiful work of the Brown and Washington groups [14, 22], who have put together a framework in which the construction of a template from a set of anatomies is proven to minimize the energy of the deformations required to map it onto all the elements of that set. Our work complements theirs in that we provide quantitative measurements confirming their formulation, although we do not compute small deformations as is required for their proof. It should be noted that our resulting model also includes average intensity information and that our respective groups use different registration methods.

Le Briquer and Gee [12] have also developed a method that provides, for a given group of subjects, the mean shape and the modes of principal variation along with their amplitude. Their approach is set in a statistical framework and aims at deriving a shape model. Our method differs in that we aim for the analysis of local information rather than global patterns.

The work presented here also relates to the methodologies of Subsol *et al.* [16], Bookstein [23], and Kendall [24], who compute average shapes modulo similarity or affine transformations. We have not tried to strictly follow the theory developed in their works. Our intention was to conform to the idea of making an abstraction of differences between images due to first-order transformations and analyze residual variations. Our main contribution resides in the characteristics used to build the average shape, that is, the image intensities instead of landmarks or crestlines. Again, this enables the computation of dense deformation fields representing variations everywhere in the MR scan, as opposed to interpolating transformations found using landmarks, lines, or surfaces. We believe this technique may find less accurate matches in the close surroundings of the landmarks, but that it provides better overall registration.

As will be shown, compared to these previous efforts, our method provides clearer images with higher contrasts and sharper definitions of tissue boundaries. Most importantly, we provide numbers showing the convergence of the model toward the centroid of the image set.

The remaining sections of this paper are organized in the following manner. First, we detail the method used to construct the average model. We then present results showing the convergence of the method toward an average intensity and shape model and show the effect of the choice of reference image. We conclude with a discussion on future research tracks.

2. METHODOLOGY

2.1. Registration

The work that follows assumes that each point in one image has a corresponding equivalent in the others. It also assumes that a matching method which is able to find these correspondences and is capable of providing a vector field representing those relationships is available. In theory, neither of these conditions is realized. That is, at a microscopic scale, there is not a one-to-one relationship between the brain cells of two individuals, and assuming there is, to this day, no algorithm has been able to find it. In practice however, deforming one brain so its shape matches that of another is conceivable and many algorithms realizing this process have been developed [21, 4, 25–27, 17, 13]. The procedure used in the following work is the demons method [28] using a complete grid of demons. We briefly detail it here and refer the reader to the original article for more information.

2.1.1. Evaluating shape differences. When applied to MR images, the demons algorithm can be considered as an optical flow variant [29]. From this point of view, the 3D images to be registered are considered as a time sequence represented by $I(x, t)$ where $x = (x_1, x_2, x_3)$ is a voxel position in the image and t is time. It computes forces by constraining the brightness of brain structures to be constant in time so that

$$\frac{dI(x,t)}{dt} = 0.$$

This leads us to the basic optical flow formulation (see [29] for details about the derivation),

$$v = -\nabla_x I(x,t)\frac{\partial I(x,t)/\partial t}{\|\nabla_x I(x,t)\|^2},$$

which is the movement component in the direction of the brightness gradient $\nabla_x I(x,t) = ((\partial I(x,t)/\partial x_1), (\partial I(x,t)/\partial x_2), (\partial I(x,t)/\partial x_3))$. For numerical stability reasons, when $\nabla_x I(x,t)$ is close to zero, the denominator of the above formula is modified to result in the basic displacement formulation for the demons algorithm using a complete grid of demons,

$$v = -\nabla_x I(x,t)\frac{\partial I(x,t)/\partial t}{\|\nabla_x I(x,t)\|^2 + |\partial I(x,t)/\partial t|^2}.$$

When $\|\nabla_x I(x,t)\| = 0$ no displacement is computed.

As with all optical flow formulations based on differential techniques, the problem here resides in finding the components of the movement in the directions orthogonal to the gradient. Many regularization methods have been proposed [30], each with their strengths and weaknesses. The one proposed by Thirion is to apply a Gaussian filter to each of the three components of v. This provides a smooth displacement field in a time efficient way. It is interesting to know that Bro-Nielsen and Gramkow [31] have shown that regularizing the deformation field using a Gaussian filter approximates linear elasticity.

The method is iterative and makes use of a multiscale scheme which resolved the problem of finding large deformations, a common problem with optical flow techniques and a basic assumption in the formulation and implementation of the derivative filters.

2.1.2. Relaxing the intensity constraint. We mentioned that the registration algorithm assumes the same intensity for corresponding brain structures in the images to be registered. For all sorts of reasons, such as acquisition parameters or preprocessing, this may not be the case. To relax this constraint, a linear intensity correction is evaluated at each iteration of the registration procedure. It is obtained by finding the line that best fits the joint histogram of the two images (see Fig. 2). This line is obtained using linear regression and outlier

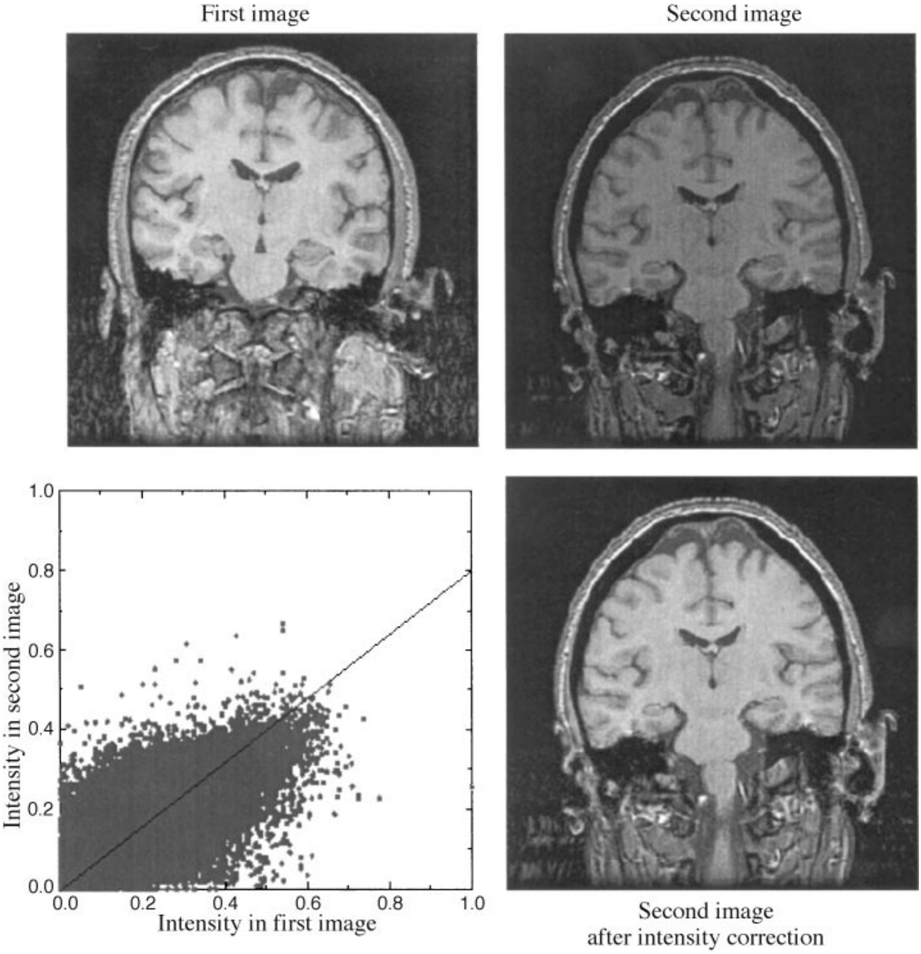


FIG. 2. Intensity correction method.

rejection. From experience, we know that such an intensity correction provides images in which boundary definitions are clearer and better matched.

2.1.3. Relevance of the resulting transformation. In the case of intersubject nonrigid registration, quantifying the accuracy of a method is difficult. One could deform an image I using a known deformation D into I', register I with I', and compare the result of the registration with D, but this comparison is biased by the way D is generated. For example, since the demons algorithm produces a smooth deformation field, if the vectors of D were to be generated randomly, the method is expected to perform poorly. Another method is to place landmarks in the images to register and evaluate differences between landmarks after registration (see, for example, [13]). No such study has been performed using the demons algorithm.

An approach which is a generalization of the previous one is to compare manual and automatic segmentation using segmentation propagation, the manual result serving as ground truth (see, for example, [32]). We believe this approach may be better suited to evaluate the quality of high-dimensional transformations such as the ones obtained using the previously described algorithm (typically $3 \times 200^3 = 24 \times 10^6$ degrees of freedom). Such a study has been performed previously by Dawant *et al.* [33] for the demons algorithm. To summarize their work, contours of different brain structures, large and small, have been segmented manually on nine MR images. One of these images was mapped to all eight other images. Then, manual and propagated segmentations were compared. Their similarity index is defined as two times the area encircled by both contours divided by the sum of the areas encircled by each contour. This index ranges from zero to one, with zero indicating zero overlap and one indicating a perfect agreement between two contours. It is sensitive to both displacement and differences in shape and it is thus preferable to a simple area comparison. The average similarity indices between the manual and automatic segmentations was 0.96, 0.97, and 0.845 for the whole head, the cerebellum, and the head of the caudate, respectively. Compared with intrarater results of 0.97, 0.97, and 0.88, the difference in the mean similarity indices between two manual delineations and between the manual delineations and the automatic segmentation method are statistically significant for the whole head and the caudate but not for the cerebellum. The authors put forth though that similarity indices over 0.85 correspond to contours that are virtually indistinguishable and that a more relevant comparison should be performed using interrater manual results which are known to have more discrepancies.

It should be pointed out that the demons algorithm does not explicitly track the transformation's Jacobian to make sure its determinant is positive, so that in theory it is possible to obtain a singular transformation. In our experience with MR data, this does not occur when using a sigma of 1 voxel to define the Gaussian filter for the smoothing operation. Also, the algorithm matches intensities and a *global* intensity correction is made over the whole image. Hence, the transformed image is not an exact duplicate of the target. This is due to the smoothness constraint applied to the displacement field which establishes a compromise between intensity resemblance and uniform local deformations at each iteration and thus in the final result.

2.2. Average Model Construction

The average model construction needs as input a reference image I_R and a set S of N images $I_1, \ldots, I_N$ representing the group of subjects under consideration. The method can be divided into six steps as follows:

1. The first step is the evaluation of global shape and intensity differences between the reference and each image of the set. Elastic registration between I_R and I_i provides vector fields D_i giving for each voxel x_R of I_R the corresponding anatomical location x_i in I_i as well as an intensity transformation IT_i. An affine transformation A_i that best approximates, in a least squares sense, the corresponding D_i is computed. Since we have correspondences between anatomical points of the I_i and I_R that have the form $x_i = D_i(x_R)$, we compute the A_i by minimizing the distance $\sum_x \|x - A_i^{-1}(D_i(x))\|^2$ (see, for example, [34] for a closed form), where the summation is performed on the voxel positions in I_R corresponding to cerebral tissues[1].

2. In the second step, residual variations due to pure morphological differences are calculated. Elastic registration is performed between I_R and each I_i using the corresponding A_i and IT_i as initial transformation estimates. This provides the resulting matched images I_i' as well as the residual vector fields R_i.

3. The third step averages the I_i', producing a mean intensity image $\overline{I}$ with the shape of I_R.

4. The fourth step aims to produce the deformation presenting the shape variations between I_R and the average shape of the set elements after correction of affine differences. Since the residual deformations R_i are all defined in the same anatomical space, that of I_R, calculating their vectorwise average $\overline{R}(x) = 1/N \sum_i^N R_i(x)$ will provide the desired deformation.

5. The fifth and final step consists of applying this average residual deformation to the average intensity image to obtain an average intensity and shape image representing the anatomical average model M. To avoid the cumbersome computation of the inverse of a vector field, we use forward resampling [36] with mathematical morphology [37] to guarantee continuity.

Considering numerical errors due to the fact that automatic registration methods usually perform better when images are closer to each other, all these steps may be repeated by replacing I_R with M, thus constructing a model with a reference image closer to the centroid of our set. Intuitively, this should reduce the mean registration error and provide a new model M' closer to the theoretical solution.

In the next section we will further study this convergence with respect to the choice of the reference image I_R and the number of iterations needed to achieve convergence.

3. RESULTS

The method is tested by computing four models using two reference images I_{R_1} and I_{R_2} (see Figs. 3a and 3b) and two image sets S_1 and S_2, each composed of five images (see Table 1).

The 3D MR protocol provides coronal images obtained using a 1.5 Tesla SIGNA (General Electric, Milwaukee, WI) whole body MR imaging system. One hundred and twenty four coronal T1-weighted images were obtained using a spoiled gradient echo (SPGR) pulse sequence (TE = 9 ms, TR = 34 ms, flip angle = 45°). Two NEX acquisitions took 27 min and 52 s. The field of view (FOV) of the images was 20 cm and each image refers to a

---

[1] These positions are obtained using an automatic method for brain segmentation similar to that of Brummer *et al.* [35]. From this point forward, all summations over x are assumed to be on the voxel positions obtained using this algorithm.

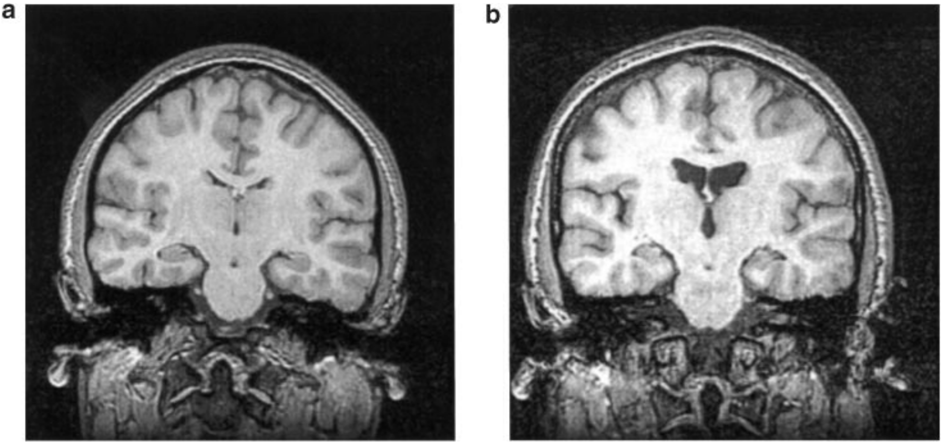


FIG. 3. Coronal slices from the two reference images, (a) I_{R_1} and (b) I_{R_2}.

contiguous section of tissue of 1.6 mm thickness. The two acquisitions, as opposed to one, gave increased contrast between gray and white matter and therefore better definition of structure boundaries. The images showed no evidence of movement or chemical shift artifacts, and partial volumizing effects were minimal. The acquisition time was well tolerated by all subjects. The $256 \times 256 \times 124$ voxels of size $0.78 \text{ mm} \times 0.78 \text{ mm} \times 1.6 \text{ mm}$ were trilinearly interpolated to $200 \times 200 \times 198$ to give cubic voxels of 1 mm side.

We analyze our results with regards to two factors. First, the iteration process is investigated to see if convergence is achieved, and if so, the speed of the convergence rate. Second, we study the effect of changing the reference image. If the model is a true average of the image set, changing the reference should produce an identical model up to an affine transformation defined by the affine difference between references.

In our evaluation procedure, three metrics are used. The first determines the average distance (AD) from an image I to the elements of a set S,

$$\mathrm{AD}(I, S) = \sqrt{\frac{1}{n}\sum_{x}\frac{1}{N}\sum_{i=1}^{N}\|x - R_i(x)\|^2},$$

where R_i is the residual deformation from I to the ith element of S, n is the number of voxels characterizing cerebral tissues and N represents the number of elements in S.

TABLE 1
References and Image Sets Used to Build the Different Models

Model	Reference	Image set
M_{11}	I_{R_1}	S_1
M_{21}	I_{R_2}	S_1
M_{12}	I_{R_1}	S_2
M_{22}	I_{R_2}	S_2

The second is the root mean square norm (RMSN) which supplies information regarding the shape variation expressed by a deformation field D,

$$\mathrm{RMSN}(D) = \sqrt{\frac{1}{n}\sum_{x} \|x - D(x)\|^2},$$

where n is the number of voxels characterizing cerebral tissues in the reference from which D was obtained.

The third provides a measure of brightness disparity between two images, I_i and I_j. It is the normalized intensity difference (NID) of the images' intensities at corresponding locations,

$$\mathrm{NID}(I_i, I_j) = \sqrt{\frac{\sum_x (I_i(x) - I_j(x))^2}{\sum_x (I_i(x))^2}}.$$

An easy way to interpret this formula is to note that if $I_j = I_i$, $\mathrm{NID}(I_i, I_j) = 0$; if $I_j = 2I_i$, $\mathrm{NID}(I_i, I_j) = 1$; if $I_j = \frac{1}{2}I_i$, $\mathrm{NID}(I_i, I_j) = 0.5$; and so on.

3.1. *Effect of Iterating*

To evaluate the effect of iterating, we construct the four models repeating the process five times using the result of the previous iteration as the reference image. We will designate the model M_{jk}, obtained with reference image I_j and set S_k, computed at the ith iteration by $M_{jk}^{(i)}$. For convenience, $M_{jk}^{(0)}$ will be identified to the average intensity image having the shape of I_j. This represents a sort of iteration after applying only the first three steps described in Section 2.2.

Four measures were computed:

$\mathrm{AD}(M_{jk}^{(i)}, S_k)$: The average distance from the reference of the current iteration to all the elements of the set.

$\mathrm{RMSN}(\overline{R}_{jk}^{(i)})$: The shape variation expressed by the residual deformation field $\overline{R}_{jk}^{(i)}$ when $M_{jk}^{(i)}$ is used as the reference.

$\mathrm{RMSN}(D_{jk}^{(i)})$: The shape difference between models computed at successive iterations. $D_{jk}^{(i)}$ is the deformation obtained by registering $M_{jk}^{(i)}$ with $M_{jk}^{(i+1)}$.

$\mathrm{NID}(M_{jk}^{(i)}, M_{jk}^{(i+1)})$: The brightness disparity between models obtained at successive iterations.

If the models computed tend toward the centroid of the image set, the first measure should diminish. This process is depicted in Fig. 4a: as the model evolves toward the center (dotted line), the average distance to the image set elements decreases. The second and third measures, representing the shape evolution of the model (see Fig. 4b), should tend toward zero. Finally, the fourth value should also decrease to zero since it represents the brightness differences between successive models.

The results of these calculations on the four models are presented in Fig. 5. Note that the iterations range up to 4 and not 5 since we compare models computed at iterations i and $i + 1$. We remind the reader that models $M_{jk}^{(0)}$, that is, models before the first iteration, characterize only average intensities and not average shapes.

From Fig. 5a, we know the average distance from the references to the image set elements is between 4.62 and 5.51 mm and reduces to a range of 2.88 to 3.36 mm. Note that the

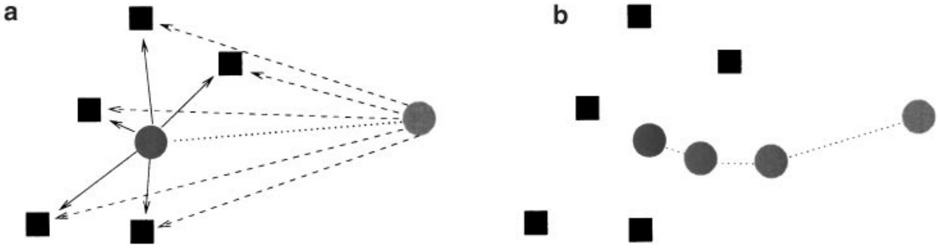


FIG. 4. Evolution of the model (circles) toward the center of the image set (squares). (a) The average distance of the model to the image set element decreases. (b) The distance between successive models diminishes.

average distances for the models built using the same image set are very close (0.09 mm for both models), while the distance for the models built using different image sets is much higher (in the range of 0.30 to 0.48 mm, depending on which models are compared). This is easily explainable by the fact that different small image sets will tend to have different average distances from their centroid. Compared to these values, the variation between successive models (see Figs. 5b and 5c), which is in the range of 0.41 to 0.52 mm, seems minor. Figure 5d presents numbers showing the brightness difference between successive models diminishes rapidly to almost 0, increasing our belief that models do not evolve significantly after the first iteration.

3.2. *Effect of the Reference*

If the models computed are equal up to an affine transformation, changing the reference image should produce a model identical to the previous one after removing their affine differences. To verify this characteristic, we performed an *affine* registration between models built using the same image set. $M_{21}^{(i)}$ is registered with $M_{11}^{(i)}$ to provide the image $M'^{(i)}_{21}$ and $M_{22}^{(i)}$ with $M_{12}^{(i)}$ to result in $M'^{(i)}_{22}$.

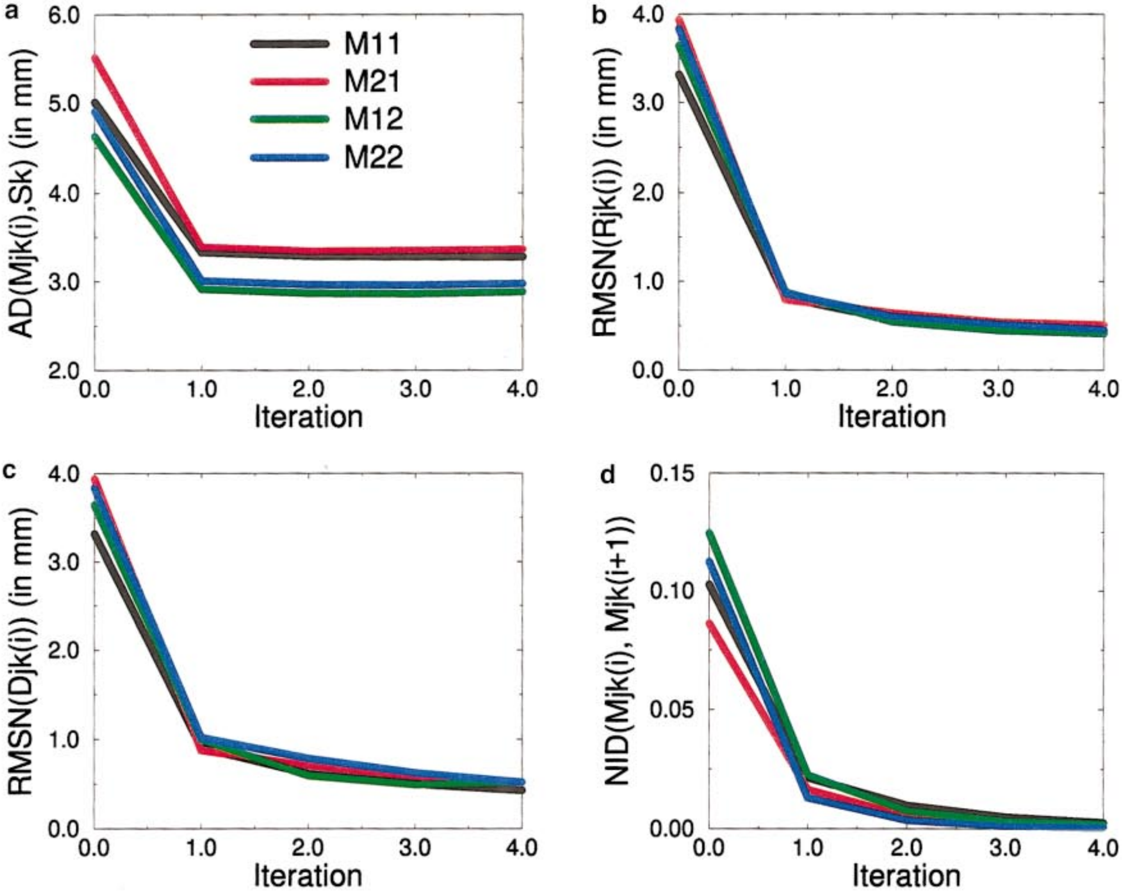
Two measures were used:

RMSN($D_k^{(i)}$): The shape variation from $M_{1k}^{(i)}$ to $M'^{(i)}_{2k}$. $D_k^{(i)}$ is the deformation obtained by registering the two images.

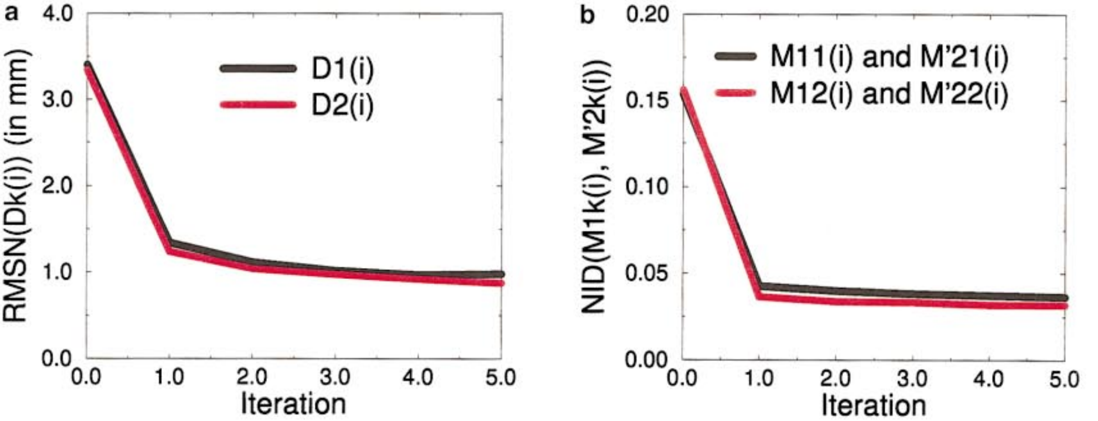
NID($M_{1k}^{(i)}$, $M'^{(i)}_{2k}$): The brightness disparity between the two models.

Results are shown in Figs. 6a and 6b, respectively. We note that shape variation between the models reduces from about 3.4 to 0.9 mm. This last value is close to the difference between successive models which we know from Figs. 5b and 5c to be approximately 0.4 mm. The brightness disparity also diminishes rapidly and does not change drastically after the first iteration. From these results, one can see that the models built using different reference images are very close, but are not identical. We believe the differences are due to errors in our registration method and to the resampling procedure applied to the models to put them in the same affine space, which tends to smooth the image. This last artifact is not present when comparing models obtained from successive iterations as we did for the results of Fig. 5d. Still, the average distance between models is less than the resolution of the image, and both the average distance and the normalized intensity difference evolve only slightly after the first iteration.

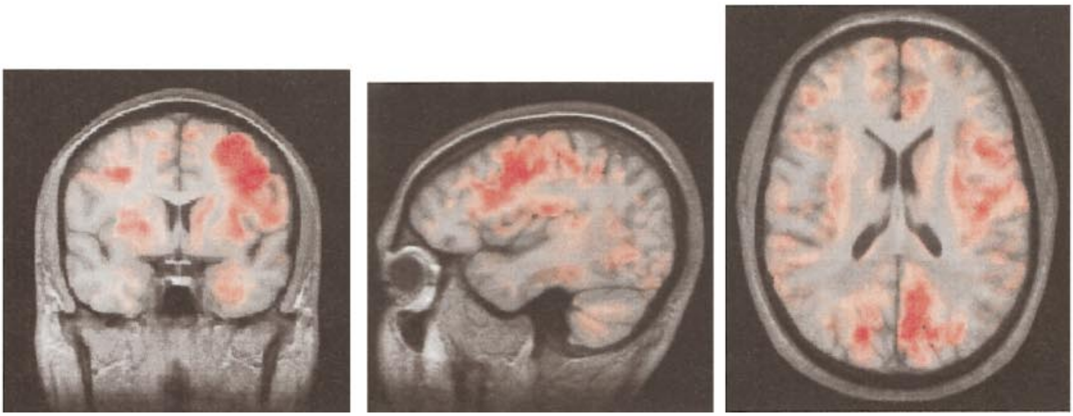
Figures 7 to 10 present some results obtained during this process. In Fig. 7 the average intensity images corresponding to the reference images of Fig. 3 are presented. They basically have the same shape as their respective reference image and their intensities corresponds


a
AD(Mjk(i),Sk) (in mm)
6.0
5.0
4.0
3.0
2.0
0.0
1.0
2.0
3.0
4.0
Iteration
M11
M21
M12
M22
b
RMSN(Rjk(i)) (in mm)
4.0
3.0
2.0
1.0
0.0
0.0
1.0
2.0
3.0
4.0
Iteration
c
RMSN(Djk(i)) (in mm)
4.0
3.0
2.0
1.0
0.0
0.0
1.0
2.0
3.0
4.0
Iteration
d
NID(Mjk(i), Mjk(i+1))
0.15
0.10
0.05
0.00
0.0
1.0
2.0
3.0
4.0
Iteration

5


a
RMSN(Dk(i)) (in mm)
4.0
3.0
2.0
1.0
0.0
0.0
1.0
2.0
3.0
4.0
5.0
Iteration
D1(i)
D2(i)
b
NID(M1k(i), M'2k(i))
0.20
0.15
0.10
0.05
0.00
0.0
1.0
2.0
3.0
4.0
5.0
Iteration
M11(i) and M'21(i)
M12(i) and M'22(i)

6



12

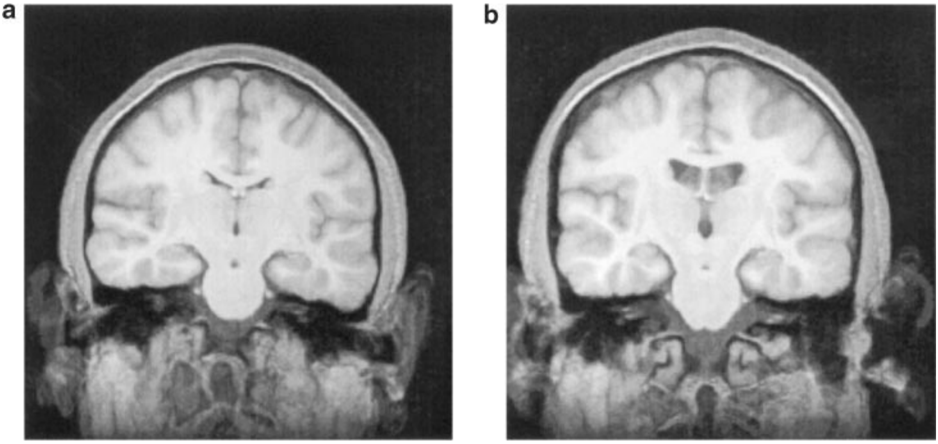


FIG. 7. Average intensity images built using reference images I_{R_1} (Fig. 3a) and I_{R_2} (Fig. 3b) and the same image set S_1. They were obtained during the first iteration of the model construction method. Note how the signal to noise ratio is increased while contrast is preserved. (a) Average intensity image $\bar{I}$ built using I_{R_1} and S_1 during the first iteration. (b) Average intensity image $\bar{I}$ built using I_{R_2} and S_1 during first iteration.

to the average of the image set S_1. Note how the signal to noise ratio is increased while contrast is preserved. These average intensity images are deformed using the corresponding average residual deformations to provide the average models of Fig. 8. The average model $M_{21}^{(1)}$ is then registered with $M_{11}^{(1)}$ using an affine transformation to set it in the same affine space. This result is presented in Figs. 9a and 9b. The same procedure has been performed for all five iterations for both image sets S_1 and S_2. The resulting models for the fifth iteration using image set S_1 are also shown in Figs. 9c and 9d. In Fig. 10, slices are taken where there is more variability in the cortex area and thus where our registration method finds correspondences in which we have less confidence. As can be seen, in these regions the model image is less clear and the contrast between gray and white matter is less pronounced.

4. DISCUSSION

Figure 5 presents numbers showing that our method constructs average models well representing the average intensity and shape of our image sets. In particular, Fig. 5a shows that

---

FIG. 5. Impact of the iteration process when computing the models. Note that the iterations range up to 4 and not 5 since we compare models computed at iterations i and $i+1$. We remind the reader that "models" $M_{jk}^{(0)}$, that is, models before the first iteration, characterize only average intensities and not average shapes. (a) Average distance to the reference of the current iteration. (b) Shape variation of the reference for the current iteration. (c) Shape difference between models computed at successive iterations. (d) Brightness disparity between models computed at successive iterations.

FIG. 6. Influence of the reference on the model computed. (a) Shape variation between $M_{1k}^{(i)}$ and $M'^{(i)}_{2k}$. (b) Brightness disparity between $M_{1k}^{(i)}$ and $M'^{(i)}_{2k}$.

FIG. 12. (a) Coronal, (b) sagittal (left hemisphere), and (c) axial sections of the probabilistic atlas. Shape variability is superimposed in red (see text for more details). Images are shown using the radiological convention, i.e., patient's left side is on the right side and vice-versa.

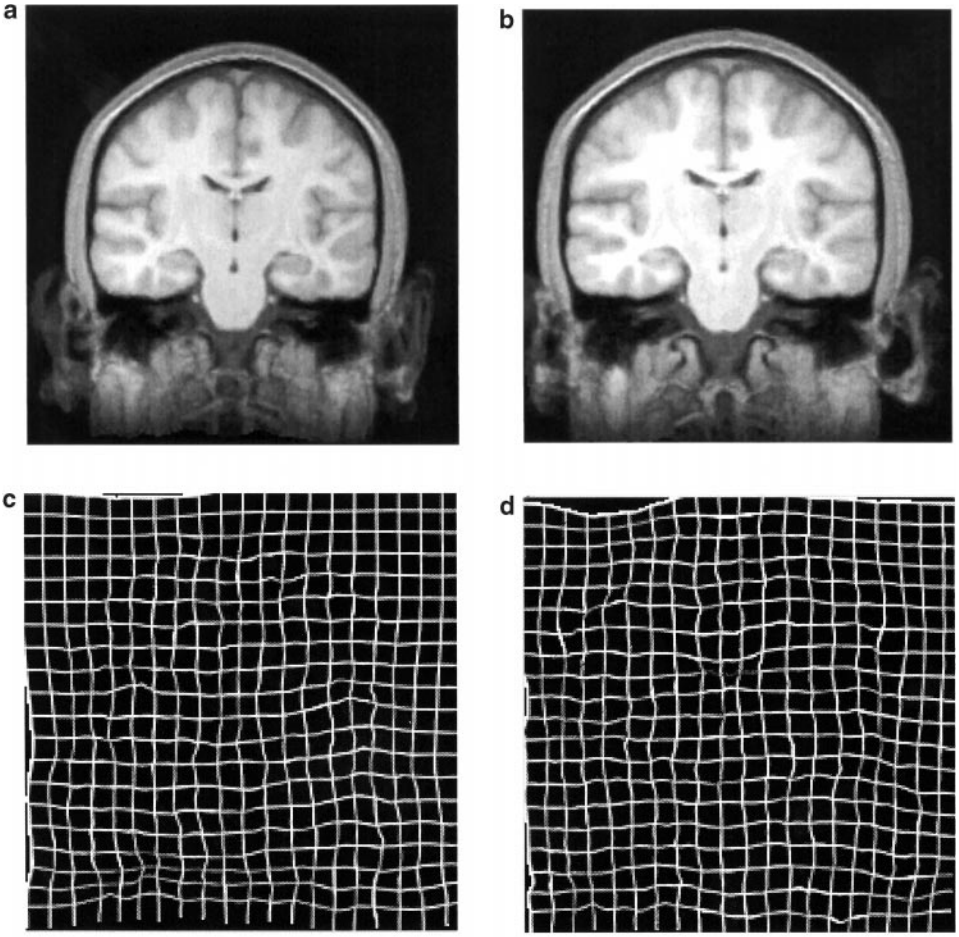


FIG. 8. Average models $M_{11}^{(1)}$ (a) and $M_{21}^{(1)}$ (b) computed using reference images I_{R_1} (Fig. 3a) and I_{R_2} (Fig. 3b) and the same image set S_1. They were build from the average intensity images of Figs. 7a and 7b respectively. (c) and (d) are the corresponding grids deformed by the average residual deformation $\bar{R}_{11}^{(0)}$ and $\bar{R}_{21}^{(0)}$, respectively.

the average distance from one image to the set elements varies between 4.62 and 5.51 mm. This distance reduces and stays between 2.88 and 3.36 mm after the first iteration. Figures 5b and 5c illustrate a minor shape evolution of the models at each iteration. Furthermore, Fig. 5d allows us to claim that the visual aspect of the models changes only minimally. This leads us to the conclusion that the models constructed are different, but equivalent from a practical point of view. That is, we believe the model slowly wanders around the optimum solution. Their intensity difference is practically null, and their shapes, although different, all have the same average distance to the other elements of the set. Hence, we believe one or two iterations are sufficient to build representative average models.

The reader should appreciate the low noise, high contrast, and high quality of the models produced in Figs. 8–10. They present models built using different reference images and the same image set. The ventricular shape bias introduced using I_{R_2} is minimal if not null. Also, Fig. 6 shows that the models built using different references seem to converge toward the same solution. Their shape difference of about 0.9 mm presented in Fig. 6a is low compared to the average distance between the models and the set elements, which is in the range of

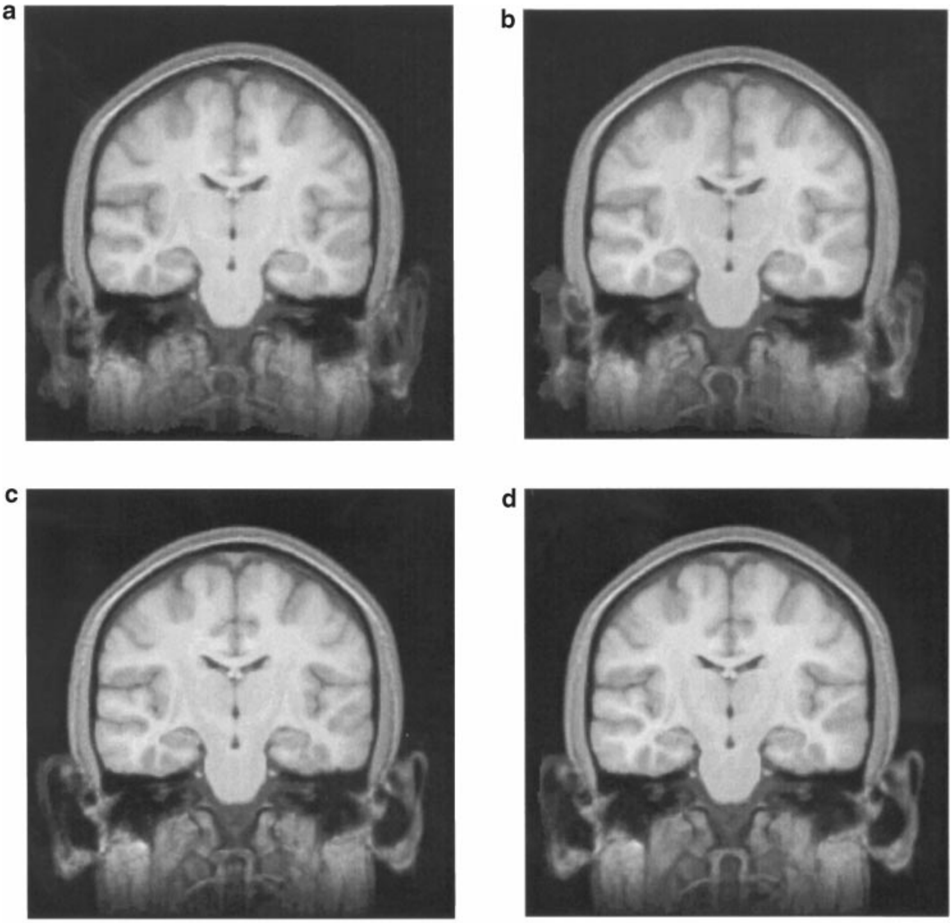


FIG. 9. Average models computed using reference images I_{R_1} (Fig. 3a) and I_{R_2} (Fig. 3b) and the same image set S_1. A coronal slice of (a) $M_{11}^{(1)}$, (b) $M'^{(1)}_{21}$, (c) $M^{(5)}_{11}$, and (d) $M'^{(5)}_{21}$. (b) and (d) have been transformed into the same affine space as (a) and (c), respectively, and can therefore be compared directly.

2.88 to 3.36 mm and just over the distance between successive average models which varies from 0.30 to 0.48 mm. Figure 6b also presents a low disparity between the different models' intensities.

If familiar with the work of Bookstein [20] or the MNI group [8], the reader will appreciate the high contrast and visual quality of the images produced, although due in part to the smaller number of images used. To better visually appreciate the gain in using high-dimensional volumetric maps ($200^2 \times 198 \times 3 = 2.376 \times 10^7$ degrees of freedom) instead of affine transformations (12 degrees of freedom) during registration, Fig. 11 presents corresponding slices of the model M_{11} built using the method presented in this paper and the one built restraining registration to affine transformations. Note that Figs. 11a–11c, obtained using the method described above, provide higher contrasts and sharper definitions of tissue boundaries than Figs. 11d–11f, which were obtained using affine registration only.

In the event that bad matches occur during registration, the model would reflect this by having larger variances in deformation and/or intensity. The image obtained and the corresponding statistics regarding deformations and intensities would not reflect normal

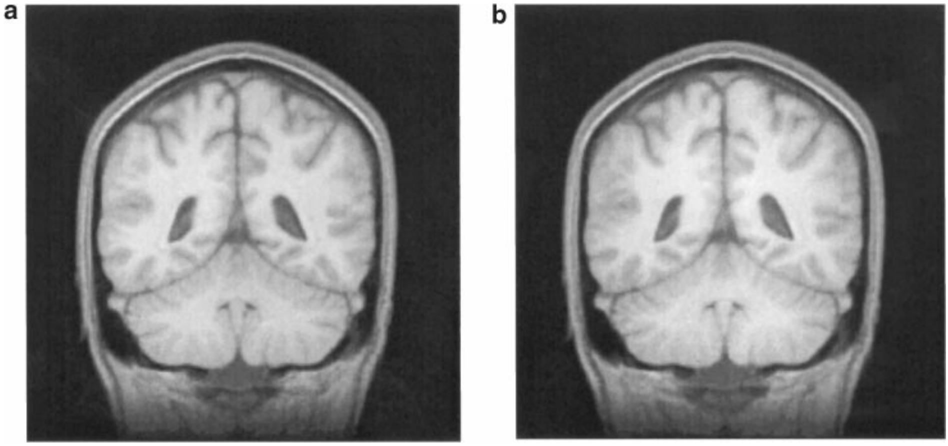


FIG. 10. Average models computed using reference images I_{R_1} (Fig. 3a) and I_{R_2} (Fig. 3b) and the same image set S_1. (a) A coronal slice of (a) $M_{11}^{(5)}$ and (b) $M'^{(5)}_{21}$. (b) has been transformed into the same affine space as Fig. 9c and can therefore be compared directly. Compared with Fig. 9, these slices are taken where there is more variability in the cortex area and thus where our registration method finds correspondences in which we have less confidence. As can be seen, in these regions the model image is less clear and the contrast between gray and white matter is less pronounced.

variations, but to a certain extent reflect normal variations *within our registration framework*. It is our belief that the models obtained in this paper are intricately linked to our modeling of the registration problem. Though cross-validation of registration techniques' results reports mutually coherent matches [38], average variations of 3 to 4 mm have been observed in this

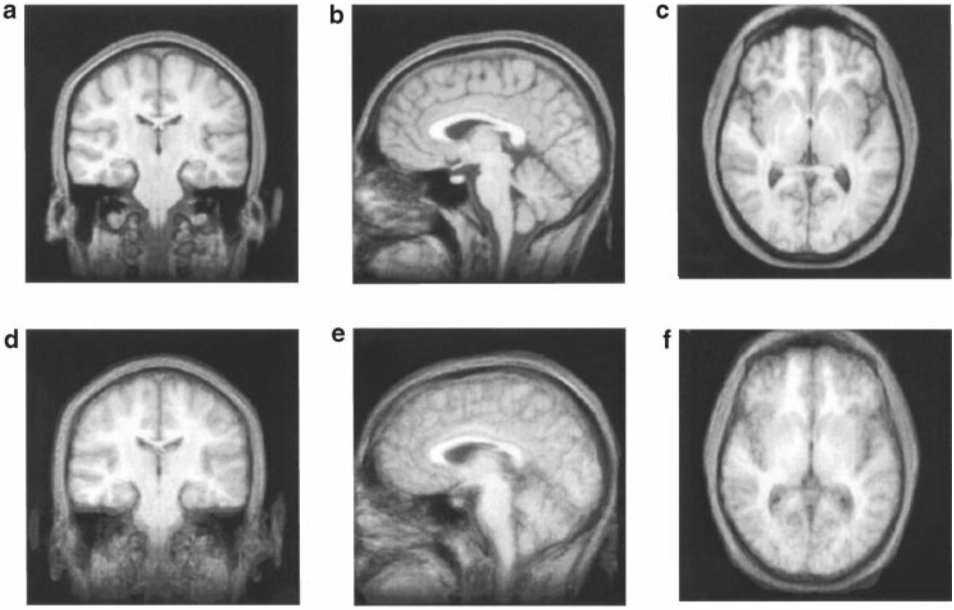


FIG. 11. Corresponding slices of M_{11} using affine registration ((d), (e) , (f)) instead of elastic registration ((a), (b), (c)). (a) Coronal slice, elastic registration, (b) sagittal slice, elastic registration, (c) axial slice, elastic registration, (d) coronal slice, affine registration, (e) sagittal slice, affine registration, (f) axial slice, affine registration.

study. These figures are too large to give significance to our models' statistics using other registration methods.

Although beyond the scope of this article, we present in Fig. 12 preliminary results of a probabilistic atlas built using the information gathered during the construction of M_{11}. This figure presents slices of the average model on which information about shape variability is superimposed in red. This variability is computed as follows: we calculate the voxelwise covariance matrix of the residual deformation fields R_i obtained from the registration of each element of S_1 with M_{11}. The determinant, or the product of the eigenvalues, of each of these matrices is computed. The resulting values can be interpreted as probabilistic volumes (see [39] for more detailed geometric interpretations of this measure). Hence, for a given voxel, a large volume corresponds to a large variability of the positions of corresponding voxels found in S_1. This volume has been normalized and converted to a red intensity value added to the MR image.

We remind the reader that this probabilistic atlas was built using only five normal subjects and hence does not reflect the extent of normal variability that is to be found in a larger population. Nonetheless, we point out that high variability was found in the cortical region, which is known to present large shape variations. There also seems to be more variability in the left hemisphere than in the right.

We are presently working on the evaluation of the number of subjects needed to fully represent the extent of variations in a normal population. Further interpretations of the corresponding probabilistic atlas will follow this work.

5. CONCLUSION

We have presented a completely automatic method to build average anatomical models of the human brain using a set of MR images. To this end, brain shape variations between subjects were identified. Differences due to linear transformations were excluded, resulting in the quantification of pure morphological differences. The result is an average intensity and shape image representative of the characteristics of the image set elements used for the construction. Furthermore, we have shown that this model corresponds to the centroid of the image set and does not depend on the reference image used for its construction. This stability is very important to ensure the efficiency and usefulness of our method. The coupling of such a high quality model with statistical information regarding normal deformations, such as the one presented in Fig. 12, in the work of Thompson and Toga [18], or of the Brown and Washington groups [14, 22], could enrich the significance of statistical tests by adding intensity information, useful, for example, in detecting gliosis in T2 MR images, and would supply an important tool in the analysis of normal anatomy.

6. APPENDIX

Nomenclature

A_i	Affine component of D_i
D	Deformation field resulting from elastic registration between any two images I_i and I_j
D_i	Deformation field resulting from elastic registration between a reference image I_R and an image I_i

$D_{jk}^{(i)}$	Deformation obtained by registering $M_{jk}^{(i)}$ with $M_{jk}^{(i+1)}$
I	Any image
I_i	A particular image in a set S
I_R	Any reference image
I_{R_i}	A particular reference image
I'	Image I after application of the deformation D
I_i'	Image I_i after application of the deformation D_i
$\bar{I}$	A voxelwise average intensity image
M	Any average model
M_{jk}	Average model obtained with reference image I_j and set S_k
$M_{jk}^{(i)}$	M_{jk} at the ith iteration
N	Number of images in a set S
n	Number of voxels characterizing cerebral tissues in an image
R_i	Residual component of D_i
$\bar{R}$	A vectorwise average deformation field
$\bar{R}_{jk}^{(i)}$	$\bar{R}$ when $M_{jk}^{(i)}$ is used as the reference
S	Any set of images
S_k	A particular set of images
x_R	A voxel position in a reference image I_R
x_i	A voxel position in an image I_i
AD	Average distance
NID	Normalized intensity difference
RMSN	Root mean square norm

ACKNOWLEDGMENTS

Dr. Neil Roberts from the Magnetic Resonance & Image Analysis Research Center (MARIARC) of Liverpool University provided the MR images used in this study. Part of this work was funded by the Natural Sciences and Engineering and Research Council of Canada (NSERC) and the Ministère de l'Éducation du Québec (MEQ).

REFERENCES

1. J. C. Mazziotta, A. W. Toga, A. Evans, P. Fox, and J. Lancaster, A probabilistic atlas of the human brain: Theory and rationale for its development, *Neuroimage* **2**, 1995, 89–101.
2. J. Talairach and P. Tournoux, *Co-Planar Stereotaxic Atlas of the Human Brain*, Thieme Medical, New York, 1988.
3. G. Schaltenbrand and W. Wahren, *Atlas of Stereotaxy of the Human Brain*, Thieme, Stuttgart, 1977.
4. R. Bajcsy and S. Kovacic, Multiresolution elastic matching, *Comput. Vision Graphics Image Process.* **46**, 1989, 1–21.
5. T. Greitz, C. Bohm, S. Holte, and L. Eriksson, A computerized brain atlas: Construction, anatomical content, and some applications, *J. Comput. Assisted Tomography* **15**, 1991, 26–38.
6. D. Lemoine, C. Barillot, B. Gibaud, and E. Pasqualini, An anatomical-based 3D registration system of multimodality and atlas data in neurosurgery, in *Lecture Notes in Computer Science*, Vol. 511, pp. 154–164, Springer-Verlag, Berlin, 1991.
7. K. H. Höhne, M. Bomans, M. Riemer, R. Schubert, U. Tiede, and W. Lierse, A 3D anatomical atlas based on a volume model, *IEEE Comput. Graphics Appl.* **12**(4), July 1992, 72–78.
8. A. C. Evans, M. Kamber, D. L. Collins, and D. Macdonald, An MRI-based probabilistic atlas of neuroanatomy, in *Magnetic Resonance Scanning and Epilepsy* (S. Shorvon, D. Fish, F. Andermann, G. M. Bydder, and

H. Stefan, Eds.), NATO ASI Series A, Life Sciences, Vol. 264, pp. 263–274. Plenum Press, New York, 1994.

9. F. L. Bookstein, *Functional Neuroimaging*, Chap. 10, Academic Press, San Diego, 1994.
10. G. Christensen, M. I. Miller, and M. W. Vannier, A 3D deformable magnetic resonance textbook based on elasticity, in *Spring Symposium: Applications of Computer Vision in Medical Image Processing, Stanford, CA, March 1994*, American Association for Artificial Intelligence.
11. R. Kikinis, M. E. Shenton, D. V. Iosifescu, R. W. McCarley, P. Saiviroonporn, H. H. Hokama, A. Robatino, D. Metcal, C. G. Wible, C. M. Portas, R. M. Donnino, and F. A. Jolesz, A digital brain atlas for surgical planning, model-driven segmentation, and teaching, *IEEE Trans. Visualiz. Comput. Graphics* **2**(3), 1996, 232–241.
12. L. Le Briquer and J. C. Gee, Design of a statistical model of brain shape, in *Proceedings of the Information Processing in Medical Imaging Conference (IPMI'97), Vermont, 1997* (J. S. Duncan and G. R. Gindi, Eds.), Springer-Verlag, Berlin/New York, 1997.
13. R. P. Woods, S. T. Grafton, J. D. G. Watson, N. L. Sicotte, and J. C. Mazziotta, Automated image registration: Ii. intersubject validation of linear and nonlinear models, *J. Comput. Assisted Tomography* **22**, 1998, 153–165.
14. U. Grenander and M. I. Miller, Computational anatomy: An emerging discipline, *Quart. Appl. Math.* **56**(4), December 1998, 617–694.
15. P. M. Thompson, D MacDonald, M. S. Mega, C. J. Holmes, A. C. Evans, and A. W. Toga, Detection and mapping of abnormal brain structure with a probabilistic atlas of cortical surfaces, *J. Comput. Assisted Tomography* **21**(4), 1998, 567–581.
16. G. Subsol, J.-P. Thirion, and N. Ayache, A scheme for automatically building three-dimensional morphometric anatomical atlases: Application to a skull atlas, *Medical Image Anal.* **2**, 1998, 37–60.
17. J. C. Gee, D. R. Haynor, L. Le Briquer, and R. K. Bajcsy, Advances in elastic matching theory and its implementation, in *Conference on Computer Vision, Virtual Reality and Robotics in Medecine & Medical Robotics and Computer Aided Surgery (CVRMed-MRCAS'97), Heidelberg 1997* (P. Cinquin, R. Kikinis, and S. Lavallee, Eds.), Springer-Verlag, Berlin/New York, 1997.
18. P. M. Thompson and A. W. Toga, Detection, visualization and animation of abnormal anatomic structure with a deformable probabilistic brain atlas based on random vector field transformations, *Medical Image Anal.* **1**(4), 1997, 271–294.
19. A. Guimond, G. Subsol, and J.-P. Thirion, Automatic MRI database exploration and applications, *Internat. J. Pattern Recognition Artificial Intelligence* **11**(8), 1997, 1345–1365.
20. F. L. Bookstein, Thin-plate splines and the atlas problem for biomedical images, in *Proceedings of the Information Processing in Medical Imaging Conference, Wye, United Kingdom, July 1991* (A. C. F. Colchester and C. J. Hawkes, Eds.), Lecture Notes in Computer Science, Vol. 511, pp. 326–342, Springer-Verlag, Berlin, 1991.
21. F. L. Bookstein, Principal warps: Thin-plate splines and the decomposition of deformations, *IEEE Trans. Pattern Anal. Mach. Intelligence* **11**(6), 1989, 567–585.
22. M. I. Miller *et al.*, Statistical methods in computational anatomy, *Statist. Methods Medical Res.* **6**, 1997, 267–299.
23. F. L. Bookstein, Shape and the information in medical images: A decade of the morphometric synthesis, *Comput. Vision Image Understanding* **66**, 1997, 97–118.
24. D. G. Kendall, A survey of the statistical theory of shape, *Statist. Sci.* **4**(2), 1989, 87–120.
25. G. Christensen, R. D. Rabbitt, and M. I. Miller, 3D brain mapping using a deformable neuroanatomy, *Phys. Medicine Biol.* **39**, 1994, 608–618.
26. D. L. Collins, T. M. Peters, and A. C. Evans, An automated 3D non-linear image deformation procedure for determination of gross morphometric variability in human brain, in *Visualisation in Biomedical Computing, Rochester, October 1994* (R. A. Robb, Ed.), SPIE Proceedings, Vol. 2359, pp. 180–190.
27. P. Thompson and A. W. Toga, A surface-based technique for warping 3-dimensional images of the brain, *IEEE Trans. Medical Imaging* **15**, 1996, 1–16.
28. J.-P. Thirion, Image matching as a diffusion process: an analogy with Maxwell's demons, *Medical Image Anal.* **2**(3), 1998, 243–260.

29. B. K. P. Horn and B. G. Schunck, Determining optical flow, *Artificial Intelligence* **17**, 1981, 185–203.
30. J. L. Barron, D. J. Fleet, and S. S. Beauchemin, Performance of optical flow techniques, *Internat. J. Comput. Vision* **12**, 1994, 43–77.
31. M. Bro-Nielsen and C. Gramkow, Fast fluid registration of medical images, in *Proceedings of the 4th International Conference Visualisation in Biomedical Computing (VBC'96), Hamburg, Germany, September 22–25, 1996* (K. H. Höhne and R. Kikinis, Eds.), Lecture Notes in Computer Science, Vol. 1131, pp. 267–276, Springer-Verlag, Berlin, 1996.
32. J. W. Haller, A. Banerjee, G. E. Christensen, M. Gado, S. C. Joshi, M. I. Miller, Y. Sherline, M. W. Vannier, and J. G. Csernansky, Three-dimensional hippocampal MR morphometry with high-dimensional transformation of a neuroanatomic atlas, *Radiology* **202**, 1997, 504–510.
33. B. M. Dawant, J.-P. Thirion, F. Maes, D. Vandermeulen, and P. Demaerel, Automatic 3D segmentation of internal structures of the head in MR images using a combination of similarity and free form transformations, in *Medical Imaging 1998: Image Processing (MI'98) San Diego, February 23–26, 1998* (K. M. Hanson, Ed.), SPIE Proceedings, Vol. 3338, pp. 545–554, *International Society for Optical Engineering*.
34. O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, MIT Press, Cambridge, MA, 1993.
35. M. E. Brummer, R. M. Mersereau, R. L. Eisner, and R. R. J. Lewine, Automatic detection of brain contours in MRI data sets, *IEEE Trans. Medical Imaging* **12**(2), 1993, 153–166.
36. A. Guimond, J. Meunier, and J.-P. Thirion, *Average brain models: A convergence study*, Technical Report 3731, Institut National de Recherche en Informatique et en Automatique, Sophia-Antipolis, France, July 1999. [Available at http://www.inria.fr/RRRT/RR3731.html]
37. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, Reading, MA, 1992.
38. J.-P. Thirion, G. Subsol, and D. Dean, Cross validation of three inter-parients matching methods, in *Visualisation in Biomedical Computing (VBC'96), Hamburg, Germany, September 1996* (K. H. Höhne and R. Kikinis, Eds.), Lecture Notes in Computer Science, Vol. 1131, pp. 327–336. Springer-Verlag, Berlin, 1996.
39. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed., Chap. 2, pp. 11–50, Academic Press, London, 1990.