Robust Multiresolution Alignment of MRI Brain Volumes

Oscar Nestares and David J. Heeger*

An algorithm for the automatic alignment of MRI volumes of the human brain was developed, based on techniques adopted from the computer vision literature for image motion estimation. Most image registration techniques rely on the assumption that corresponding voxels in the two volumes have equal intensity, which is not true for MRI volumes acquired with different coils and/or pulse sequences. Intensity normalization and contrast equalization were used to minimize the differences between the intensities of the two volumes. However, these preprocessing steps do not correct perfectly for the image differences when using different coils and/or pulse sequences. Hence, the alignment algorithm relies on robust estimation, which automatically ignores voxels where the intensities are sufficiently different in the two volumes. A multiresolution pyramid implementation enables the algorithm to estimate large displacements. The resulting algorithm is used routinely to align MRI volumes acquired using different protocols (3D SPGR and 2D fast spin echo) and different coils (surface and head) to subvoxel accuracy (better than 1 mm). Magn Reson Med 43: 705–715, 2000. © 2000 Wiley-Liss, Inc.

Key words: MRI brain imaging; registration; multiresolution; robust estimation

Among the large number of applications of MRI, some of them require scanning the brain of the same subject repeatedly on different days, using possibly different acquisition protocols and methods. Two slices from a typical example are shown in Fig. 1. This example is from a research study using functional MRI (fMRI) to measure brain activity in visual cortex. In these experiments, a high-resolution volume anatomy of the subject's brain was acquired in one scanning session using a head coil and a 3D SPGR pulse sequence (Fig. 1b). Then the subject participated in a series of fMRI experiments across several separate scanning sessions on different days. In these scanning sessions we used a surface coil placed at the back of the head near the visual cortex of the brain. During each fMRI scanning session, a set of structural images were acquired in the same slices and at the same resolution as the functional images, using the surface coil and a fast spin echo pulse sequence (Fig. 1a). The slice orientation and voxel size of these volumes were completely different from those of the initial high-resolution volume. Even so, to analyze the fMRI data it was critical for all of the images from multiple scanning sessions to be coregistered.

The difficulty of the alignment problem can be appreciated by inspecting the two images in Fig. 1. We can find some features (e.g., marked with the white arrows) that have a similar appearance in the two images. However,

because they were acquired using different coils and different pulse sequences, other features (e.g., marked with the black arrows) are very different in the two images. In this particular example, there are particularly dramatic differences in the intensity levels of the scalp and the skull. There is also a severe gradient in intensity with distance from the back of the head in the images acquired with the surface coil (Fig. 1a).

Because manual alignment of volumes is a tedious, time-consuming, and unreliable task, it is desirable to have automatic alignment tools. Moreover, precise automatic alignment is valuable not only in the example research study described above, but in many situations (both research and clinical) involving repeated MRI acquisitions from the same subject.

There are several methods in the literature for the automatic alignment of volumes from different imaging modalities (1–6) (for a comparison between several methods, see (7)), most of which have been developed ad hoc for this difficult problem. On the other hand, image registration and estimation of motion in image sequences has been a very active research topic in the computer vision literature. Since these problems are closely related with volume alignment, it is possible to adapt standard, efficient techniques for image registration to the alignment of MRI brain volumes.

Most image registration or motion estimation techniques rely on the assumption that corresponding voxels in the two volumes have equal intensity, often referred to as the intensity conservation assumption. This assumption does not hold for MRI volumes acquired with different coils and/or pulse sequences (e.g., as illustrated in Fig. 1). The images can be preprocessed to minimize the intensity differences between the two volumes. After preprocessing, the intensities will be approximately conserved for most of the voxels. Intensity correction has been, and continues to be, an active area of research. None of the existing methods can correct perfectly for all intensity differences, especially when using different pulse sequences. In fact, it is not clear that it will ever be possible to correct perfectly for these intensity differences.

In this article, we propose an algorithm for the automatic alignment of MRI volumes that relies on robust estimation to compute the alignment parameters. The robust estimation procedure automatically ignores those voxels (outliers) where the intensities are sufficiently different in the two volumes. The resulting method is capable of accurate alignment (better than 1 mm), even when a large percentage of the voxels (more than 20%) violate the intensity conservation assumption. The resulting procedures are used routinely in our laboratory in the analysis of fMRI data.
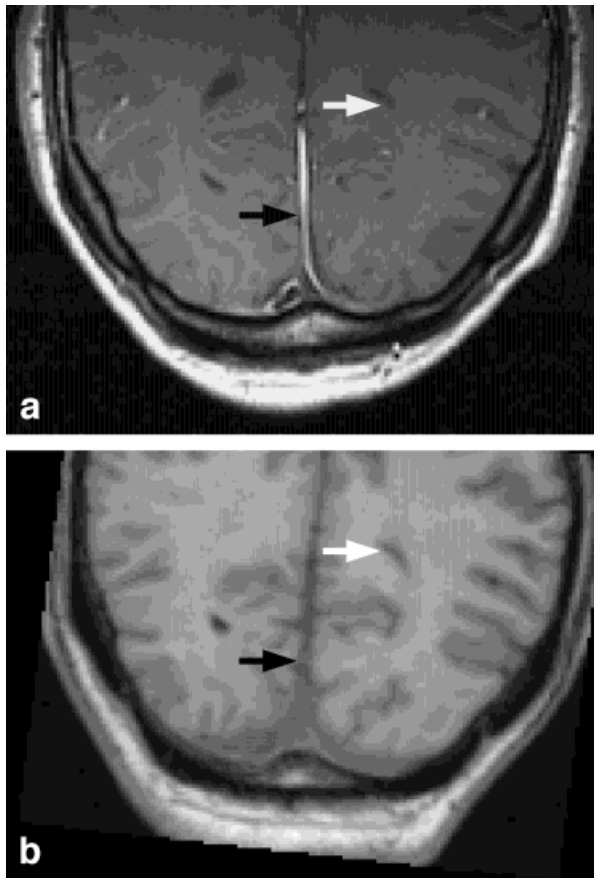
FIG. 1. Two typical slices of the volumes to be aligned: (**a**) acquired with a surface coil; (**b**) acquired with a head coil.

METHODS

We first review gradient-based displacement estimation for the special case of rigid-body motion. Second, we describe the robust estimation methods used to estimate the rigid-body alignment parameters, which are crucial for accurate alignment of volumes acquired with different pulse sequences. Third, we explain how multiresolution techniques can be applied to estimate large displacements. Preprocessing of the volumes, described next, is necessary to remove intensity gradients introduced by surface coils, and to reduce the intensity differences between the volumes. This section ends with some details about the implementation of the algorithm.

Gradient-Based Image Registration

Image motion estimation and registration has been an active research topic in the computer vision literature over the past two decades (8–13). Many of these methods are based on the so-called gradient constraint. There are different ways to derive the gradient constraint, but the main assumptions are that the intensities are shifted (locally translated) from one image to the next, and that the shifted intensity values are conserved, i.e.,

$$f_1(x, y, z) = f_2(x + d_x, y + d_y, z + d_z) \quad [1]$$

where $f_1(x, y, z)$ and $f_2(x, y, z)$ are the intensities of the two volumes as a function of space, and $\mathbf{d} = (d_x, d_y, d_z)^T$ is the local displacement. Gradient methods assume that the volume intensity function is well approximated by a first-order Taylor series expansion,

$$f_2(x + d_x, y + d_y, z + d_z) \approx f_2(x, y, z) + d_x f_x(x, y, z)$$
$$+ d_y f_y(x, y, z) + d_z f_z(x, y, z) \quad [2]$$

where f_x, f_y and f_z are the partial derivatives along the axis specified by the subscript. Ignoring second- and higher-order terms in the Taylor series, and substituting the resulting approximation into Eq. [1], we obtain:

$$d_x f_x(x, y, z) + d_y f_y(x, y, z) + d_z f_z(x, y, z)$$
$$= -\Delta f(x, y, z) \quad [3]$$

where $\Delta f(x, y, z) = f_2(x, y, z) - f_1(x, y, z)$. This equation relates the displacement at one location in the volume to the spatial derivatives of intensity at that same location. Equation [3] is called the motion gradient constraint.

It is impossible to recover the displacement $\mathbf{d}$, given just the gradient constraint at a single voxel because Eq. [3] offers only one linear constraint to solve for the three unknown components for the displacement of that voxel. Further assumptions or measurements are required to constrain the solution. One attractive possibility, which is of direct application in our alignment problem, is to consider a global model for the displacement with a small number of parameters that simultaneously characterize the displacements of all voxels (11,13). Because the volumes to be aligned come from the same subject, a rigid body displacement model with six parameters can be used. Three of the parameters correspond to a translation and the other three correspond to a rotation, and the final change of coordinates can be expressed as a rotation matrix plus a translation vector. For small rotations, the 3D displacement at each location can be approximated as:

$$\mathbf{d}(\mathbf{x}) = \begin{pmatrix} d_x(\mathbf{x}) \\ d_y(\mathbf{x}) \\ d_z(\mathbf{x}) \end{pmatrix} = \omega \times \mathbf{x} + \mathbf{t}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 & z & -y \\ 0 & 1 & 0 & -z & 0 & x \\ 0 & 0 & 1 & y & -x & 0 \end{pmatrix} \begin{pmatrix} t_x \\ t_y \\ t_z \\ \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} \quad [4]$$

where we have made explicit the dependence of the motion $\mathbf{d}$ on the spatial location $\mathbf{x} = (x, y, z)^T$, and where the unknowns are a rotation $\omega = (\omega_x, \omega_y, \omega_z)^T$, and a translation $\mathbf{t} = (t_x, t_y, t_z)^T$.

Introducing the rigid body model of Eq. [4] into the gradient constraint (Eq. [3]), and considering all the voxels simultaneously, we obtain the following over-determined linear system:

$$\begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ f_x^i & f_y^i & f_z^i & (f_z^i y^i - f_y^i z^i) & (f_x^i z^i - f_z^i x^i) & (f_y^i x^i - f_x^i y^i) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \tag{5}$$

$$\underbrace{}_{\mathbf{A}}$$

$$\underbrace{\begin{pmatrix} t_x \\ t_y \\ t_z \\ \omega_x \\ \omega_y \\ \omega_z \end{pmatrix}}_{\mathbf{p}} = - \underbrace{\begin{pmatrix} \vdots \\ \Delta f^i \\ \vdots \end{pmatrix}}_{\mathbf{b}}$$

where the superscript *i* indexes all the voxels of the volume. This equation can be expressed in matrix form as $\mathbf{Ap} = \mathbf{b}$, where $\mathbf{p}$ are the unknown rigid body displacement parameters. This overdetermined system can be solved using conventional least-squares, but in our case it is more accurate to use robust regression techniques, which is necessary to avoid the influence of voxels that violate the intensity conservation assumption due to the different acquisition protocols.

Robust Estimation of the Model Parameters

The estimation of the motion parameters requires solving the following linear system:

$$\mathbf{a}_i^T \mathbf{p} = b_i, \ i = 1, \dots, N \tag{6}$$

where $\mathbf{a}_i$ is the *i*th row of matrix $\mathbf{A}$ in Eq. [5], $\mathbf{p}$ is the vector of unknown displacement parameters, $b_i = f_1(x_i, y_i, z_i) - f_2(x_i, y_i, z_i)$ is the *i*th component of vector $\mathbf{b}$ in Eq. [5], and N is the number of voxels. N is a large number (typically tens of thousands), resulting in a heavily overconstrained system of equations. The least-squares solution of the system in Eq. [6] minimizes the sum, over all voxels, of the squared residual errors, defined as $r_i^2 = (\mathbf{a}_i^T \mathbf{p} - b_i)^2$.

However, the least-squares solution is very sensitive to isolated points having high residual errors, or outliers, which have a large influence on the error functional. In principle, one really bad voxel with a very large residual, r_i, can corrupt the solution. Because the volumes to be aligned come from different pulse sequences, there are voxels for which the intensity conservation is violated, thus having large residual errors. Therefore, the least-squares solution will be corrupted, and alternative error functions have to be used.

Robust statistical estimation techniques are designed to minimize the influence of outliers (14). One class of robust estimators, called the M-estimators, minimize the sum of a function of the residuals $\rho(r_i)$ which does not grow indefinitely (in contrast to the squared function used by least-squares), but that rather saturates for large residual errors. Therefore, outliers have a limited influence on the final
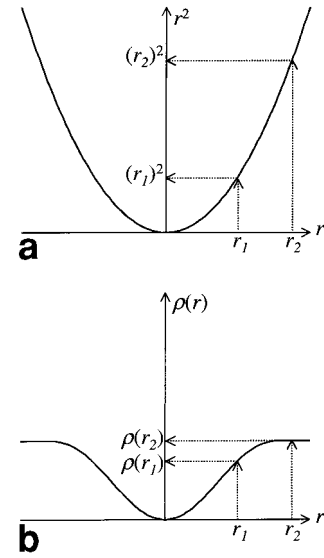

FIG. 2. Graphs of the residual error functions for: (**a**) the least squares estimator; and (**b**) a robust M-estimator. The point marked as r_1 corresponds to a typical residual value, while the point marked as r_2 corresponds to an outlier.

solution; this is illustrated in Fig. 2. The curve in Fig. 2a represents the square of the residual error, while Fig. 2b corresponds to one particular choice of a robust function of the residual error $\rho(r)$. The point labeled as r_1 corresponds to a small residual error, and the two functions in Fig. 2a and b have similar values. The point r_2, which corresponds to a large residual error, has a large squared error (Fig. 2a), and therefore will have a large influence on the estimate. On the contrary, the same point r_2 has a limited influence for the robust error function, because this function saturates for high residual error values (Fig. 2b).

One problem with the robust estimators is that the resulting function to be minimized is nonconvex, and thus there is no guarantee of convergence to an absolute minimum. Therefore, it is necessary to start with an initial condition close to the absolute minimum. In our case, the use of multiresolution and iterative refinement (described in the next section) contribute to the convergence of the algorithm because many of the local minima are removed at coarser scales. The accurate estimates obtained in our experimental results confirm that the algorithm is almost always converging to the global minimum, or at least to a nearby local minimum.

There are different choices for the robust error function $\rho(r)$ in the literature (14). We chose to use Beaton and Tukey's biweight function (15), illustrated in Fig. 2b. The resulting error functional can be minimized using different techniques (16). Our implementation uses iterative reweighted least-squares (15,16), where the problem is solved at each iteration using the very efficient weighted least-squares estimator, and the weights are updated at each iteration, such that those points displaying greater residual errors are given less weight in the next iteration. More detail about the biweight function and the iterative reweighted least-squares implementation are given in Appendix A.

One important feature of the robust estimator is that it allows us to use very simple and efficient intensity and contrast correction operations. The robust estimator provides accurate estimates of the alignment parameters even when the preprocessing steps fail to correct perfectly for the image differences caused by different coils and/or pulse sequences.

Multiresolution and Warping

The Taylor series approximation in Eq. [2] and the infinitesimal rotation approximation in Eq. [4], taken together, require that the displacements must be small, typically less than four voxels. However, typical alignment problems have displacements that are much greater than this.

A standard extension of the above technique is to use an iterative, multiresolution algorithm. The multiresolution representation of the images is computed by recursive low-pass filtering and subsampling (17). The displacements are smaller in the subsampled volumes; for example, a displacement of eight voxels in the original volumes is reduced to four voxels after one stage of filtering and subsampling. The algorithm then proceeds to use to a coarse-to-fine strategy for estimating the displacements (11,13). The alignment parameters are initially estimated at the coarsest (most subsampled) scale where the displacements are small. This coarse estimate is used to warp the volumes into approximate alignment at the next finer scale. Then the process is repeated to iteratively refine the estimate of the alignment parameters up to the finest scale (i.e., the original volumes).

We also apply the algorithm (estimating the alignment parameters and warping the volumes into alignment) iteratively within each level of resolution, to refine further the accuracy of the estimates.

The final transformation is obtained by concatenating together the transformations using matrix multiplication. At each iteration the six parameters in Eq. [5] are estimated. The homogeneous coordinate transformation corresponding to the i-th iteration is then represented in matrix form:

$$\mathbf{M}_i = \begin{pmatrix} r_{11_i} & r_{12_i} & r_{13_i} & t_{x_i} \\ r_{21_i} & r_{22_i} & r_{23_i} & t_{y_i} \\ r_{31_i} & r_{32_i} & r_{33_i} & t_{z_i} \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad [7]$$

where the column on the right is the estimated translation, and the upper-left 3×3 submatrix is a rotation matrix that corresponds to the estimated rotation, $(\omega_{x_i}, \omega_{y_i}, \omega_{z_i})$, as shown in Table 1. The global transformation is computed by multiplying together the homogeneous coordinate transformation matrices corresponding to each iteration:

$$\mathbf{M} = \prod_i \mathbf{M}_i \qquad [8]$$

where the subscript i indexes over all the iterations. The step of transforming from the small motion approximation to the homogeneous transformation representation is absolutely critical. First, the small motion approximation

Table 1
Components of the Rotation Matrix Obtained From the Estimated Rotation Ignoring the Infinitesimal Angle Approximation (see Eq. [6])

r_{11_i}	$\cos(\omega_{y_i})\cos(\omega_{z_i})$
r_{12_i}	$-\cos(\omega_{y_i})\sin(\omega_{z_i})$
r_{13_i}	$\sin(\omega_{y_i})$
r_{21_i}	$\sin(\omega_{x_i})\sin(\omega_{y_i})\cos(\omega_{z_i}) + \cos(\omega_{x_i})\sin(\omega_{z_i})$
r_{22_i}	$-\sin(\omega_{x_i})\sin(\omega_{y_i})\sin(\omega_{z_i}) + \cos(\omega_{x_i})\cos(\omega_{z_i})$
r_{23_i}	$-\sin(\omega_{x_i})\cos(\omega_{y_i})$
r_{31_i}	$-\cos(\omega_{x_i})\sin(\omega_{y_i})\cos(\omega_{z_i}) + \sin(\omega_{x_i})\sin(\omega_{z_i})$
r_{32_i}	$\cos(\omega_{x_i})\sin(\omega_{y_i})\sin(\omega_{z_i}) + \sin(\omega_{x_i})\cos(\omega_{z_i})$
r_{33_i}	$\cos(\omega_{x_i})\cos(\omega_{y_i})$

cannot represent any possible rotation. Second, concatenating the small motion approximations from successive iterations could produce a result that does not correspond to a rotation. The homogeneous transform matrices, on the other hand, can represent any possible rotations and when they are multiplied together, the result is guaranteed to be a rotation plus translation.

Once the final transformation $\mathbf{M}$ has been estimated, the volumes can be warped into alignment using standard resampling techniques (18).

Intensity Preprocessing

The most severe violation of intensity conservation in our volumes (although this is a quite general problem) is due to the different coils used during the two acquisitions. Specifically, one of the volumes was acquired with a surface coil, which results in a dramatic intensity gradient as a function of the distance to the coil (see Fig. 1a). This intensity gradient has to be corrected before attempting to estimate the alignment parameters. There are several methods in the literature for correcting intensity inhomogeneities in MR images (19–21). Although these procedures are quite effective, they do not correct perfectly for the image differences.

We chose to use a very simple and efficient strategy for intensity correction. The intensity correction procedure is described in detail in Appendix B. Briefly, we first divide the intensity at each voxel by the local mean intensity of the nearby voxels. We then equalize the contrasts of the two volumes by fitting models to the intensity histograms. Although it does not correct perfectly for the differences in the two volumes, we have found that this procedure is more than sufficient when used in conjunction with the robust estimator outlined above. More sophisticated techniques (like those cited above) could be used, but they are not necessary because the robust estimator provides accurate estimates of the alignment parameters even when a large percentage of the voxels are outliers.

Implementation Details

The final algorithm is presented in Fig. 3. There are two main loops, one corresponding to the different resolution levels in the multiresolution approach, and another loop for iteratively refining the estimates within each level of resolution.

Preprocess Vol1 and Vol2

Obtain Multiresolution Gaussian Pyramid of Vol1 and Vol2

From scale = Coarse to Fine

    repeat

        Warp Vol2(scale) with the current motion estimate

        Estimate Motion from Vol1(scale)→Warped Vol2(scale)

    until not significant update in the motion estimate

End

(**Preprocess** = Intensity normalization and constrast equalization)

(**Estimate Motion** = Robust M-estimator of the over-determined linear system in Eq. 5)

FIG. 3. Pseudocode corresponding to the main steps of the algorithm.

We implemented the coarse-to-fine strategy using a multiresolution spatial Gaussian pyramid. This pyramid is built by successively filtering the volume with a Gaussian spatial filter, and downsampling by a factor of 2. The multiresolution algorithm has the additional advantage that the coarse scale calculations are computationally efficient because they are performed over a much smaller number of voxels. We used a 5-tap cubic-B-Spline filter, shown in Table 2, as an approximate Gaussian filter to build the pyramid. In our case, we applied the multiresolution only to the two inplane spatial dimensions; we typically do not have enough samples in the z dimension because we collect fMRI data (and the corresponding structural images) with only 4–16 slices, covering a small fraction of the brain.

Solving the system of equations in Eq. [5] relies on knowing the partial derivatives of the volume with respect to x, y, and z. Computing the derivatives of a discretized intensity signal requires some regularization. This can be done by prefiltering the volume with a smoothing filter, and then differentiating. Since filtering and differentiation are linear operations, this is equivalent to filtering the signal directly with the derivative of the prefilter (22). Table 2 shows the 5-tap pre-filter and derivative filters used for the estimation of the spatial gradient (22). Each partial derivative is estimated by separable convolution of the volume with the derivative filter along one axis, and by the prefilter along the other two axes.

The iterative solution at each resolution level, as well as the coarse to fine implementation, requires warping one of the volumes given the current estimate of the alignment parameters. We implemented the warping operation using backwards tri-linear interpolation (18). The transformed

voxels lying out of bounds were ignored during the estimation process.

The different voxel sizes between the two volumes were corrected by resampling the volumes using tri-linear interpolation.

Before starting the automatic alignment procedure it is necessary to provide an initial estimate of the alignment as a starting point. This was done manually using interactive software, a process that typically takes only a minute or two. The remaining steps of the algorithm are fully automatic, taking between 1–3 min running on a Pentium III 450Mhz under Matlab (there is an implementation of the algorithm (Matlab 5.0 and higher) available at http://white.stanford.edu). This is a huge advantage compared with the typical time of a full manual alignment, performed interactively by marking corresponding points, which takes about 45 min.

RESULTS

Figure 4a shows four of eight slices from a low resolution (fast spin echo) volume. Figure 4b shows a region from the high resolution (3D SPGR) volume that was reinterpolated and cropped according to the initial manual alignment. These images have already been preprocessed using the operations described in Appendix B. The images in Fig. 1 correspond to the rightmost slices in Fig. 4, but before applying the preprocessing.

The images were acquired using a standard clinical GE 1.5T Signa scanner. The high-resolution volume (Fig. 4b) was acquired with a field of view covering the entire head, using a custom designed head coil, and a 3D SPGR pulse sequence: 33.3 ms TR (repetition time); minimum TE (echo time); 40° FA (flip angle); and a voxel size of

Table 2

Convolution Masks Used as Prefilter and Derivative Filter for the Estimation of the Gradients (First and Second Row, Respectively), and Low-Pass Filter Used to Build the Gaussian Pyramid (Third Row)

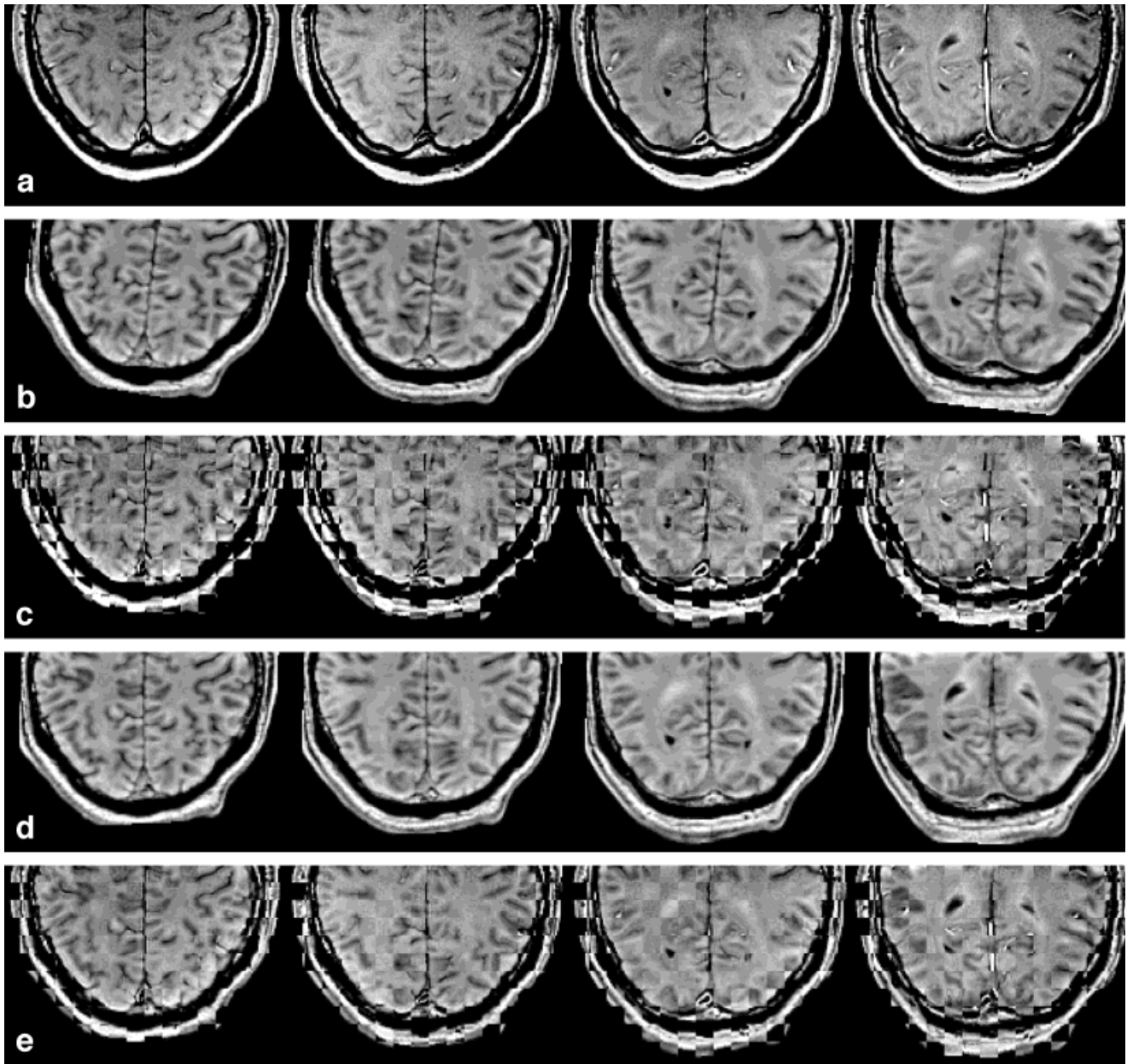|					
Prefilter	0.03504	0.24878	0.43234	0.24878	0.03504
Derivative filter	0.10689	0.28461	0	−0.28461	−0.10689
Gaussian pyramid	0.0625	0.25	0.375	0.25	0.0625

FIG. 4. Alignment results for a typical case. **a:** Four of eight slices from the low-resolution volume acquired with a surface coil. **b:** Slices from the high-resolution volume acquired with a head coil, reinterpolated using the initial manual alignment. **c:** Mosaic formed by alternating blocks of **a** and **b**, following a checkerboard pattern. **d:** High-resolution volume, reinterpolated using the final alignment estimate. **e:** Mosaic formed from **a** and **d**.

0.94 mm $\times$ 0.94 mm $\times$ 1.2 mm. The low-resolution partial volume (Fig. 4a) was acquired using a custom-designed dual surface coil placed at the back of the head. We used a T_1-weighted spin echo pulse sequence: 500 ms TR; minimum TE; 90° FA; eight adjacent, obliquely oriented slices selected with the most ventral slice positioned along the boundary between the occipital lobe and the cerebellum; and a voxel size of 1.02 mm $\times$ 1.02 mm $\times$ 4 mm.

The initial alignment is far from perfect, as shown in the mosaic of Fig. 4c, formed by taking small alternate blocks from each of the volumes, following a checkerboard pattern. The final alignment matrix, which transforms the

coordinates from one volume to the other, was estimated using the automatic method as the following:

$$\mathbf{M} = \begin{pmatrix} 0.9938 & 0.1096 & -0.0169 & -4.8152 \\ -0.1098 & 0.9939 & -0.0122 & 9.6834 \\ 0.0155 & 0.0140 & 0.9998 & -2.1708 \\ 0 & 0 & 0 & 1.0000 \end{pmatrix} \quad [9]$$

where the estimated translation is represented by the vector on the right and the estimated rotation is represented by the 3 $\times$ 3 block in the upper left (see Eq. [7]). The modulus of the estimated translation vector is 11.0 voxels,

and the magnitude of the rotation vector is 6.4°. The maximum estimated displacement (taking into account both the translation and the rotation) is 11.8 voxels, and the mean displacement is 6.0 voxels. Figure 4d shows the region of the high-resolution volume reinterpolated using the estimated alignment matrix. This alignment is almost perfect, as can be appreciated in the mosaic in Fig. 4e, which combines the intensity corrected versions of the two aligned volumes. Although the volumes were intensity corrected, there are still some local differences in intensity and contrast. Even so, you can track the pattern of sulci and gyri from one block to the next; the robust estimation resulted in a good alignment in spite of the intensity and contrast differences. Laborious attempts to do the alignment by hand always yielded results that were not as good.

Multiresolution permits us to estimate large translations and rotations, as is illustrated in the example of Fig. 5. These images were acquired using a GE 3T scanner. The low-resolution partial volume (one slice shown in Fig. 5a) was acquired with a custom head coil and a 2D SPGR pulse sequence: minimum TE; 100 ms TR; 45° FA; 0.94 mm × 0.94 mm × 2.6 mm voxel size. The full, high-resolution volume (one slice shown in Fig. 5b, interpolated according to the initial manual alignment) was acquired using the same head coil but using a 3D SPGR pulse sequence: minimum TE; 15° FA; 0.94 mm × 0.94 mm × 1.2 mm voxel size. The initial alignment is far from the true position, as is shown in the mosaic in Fig. 5c formed from Fig. 5a and b. The slice in Fig. 5d has been interpolated from the high-resolution volume using the estimated rotation and translation. The magnitude of the estimated translation is 37.9 voxels, and the estimated rotation angle is 14.1°. Despite the large initial displacement (maximum and mean displacements of 37.9 and 19.4 voxels, respectively), the final alignment is highly accurate, as can be appreciated in the mosaic of Fig. 5e, formed from Fig. 5a and d.

Note again that there are dramatic differences in the image intensities near the skull, even after applying the intensity correction preprocessing procedures. These outliers were largely ignored by the robust estimator, allowing for an accurate alignment of the brain, without having to first perform a complicated segmentation procedure to remove the scalp and skull.

Comparison With Other Registration Methods

We performed a quantitative comparison with two other methods that are both widely used for volume registration by the fMRI research community: 1) automated image registration, AIR, by Woods et al. (1,6); and 2) statistical parametric mapping, SPM, by Ashburner et al. (4,5). The SPM package provides two different algorithms. The first SPM algorithm minimizes the sum of squared differences of voxel intensities, and is recommended for use when the two volumes are acquired with the same imaging modality. The second SPM algorithm is recommended for use when the two volumes are acquired with different imaging modalities. This algorithm proceeds in three steps: 1) coarse registration with a template that is specific for each imaging modality, 2) segmentation of different tissues
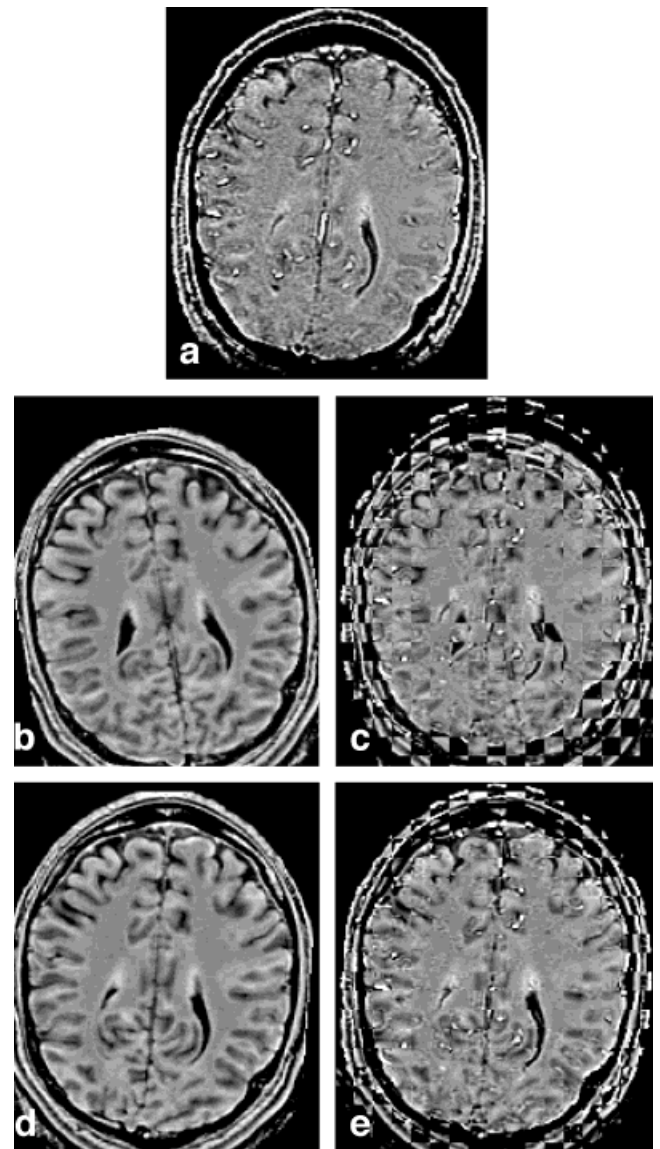

FIG. 5. Example of a case with a very large displacement, both translation and rotation (same format as Fig. 4, except that only one of the eight slices is shown).

based on a priori information, and 3) matching of the resulting segments. We tested both SPM algorithms, choosing the T1 template for the first coarse alignment. We tested AIR using the ratio image uniformity cost function.

We used synthetic datasets for which the true registration is known, because it is very difficult to perform a quantitative comparison on real datasets for which the correct registration is not known. The synthetic datasets were generated from a high-resolution volume (3D SPGR, 1.5 T, pulse sequence parameters listed above). For each trial of the simulations, we interpolated a pair of subvolumes, with either 8 or 16 slices, from the full high-resolution volume. The pose (slice orientation and position) of one of the two subvolumes was chosen to be similar to that in Fig. 4. The second subvolume was chosen to have the same pose on average, but randomized so that the displacement between the two volumes had a mean transla-

Table 3

Comparison of the Mean Alignment Error in Voxels (± Standard Deviation) for the Robust Estimator, the Least Squares Estimator, AIR, and SPM Using the Same Modality Algorithm (SPM1) and the Different Modality Algorithm (SPM2)

	2.5%	5%	10%	20%	Time (sec)
Robust	0.12 ± 0.03	0.14 ± 0.04	0.15 ± 0.05	0.77 ± 0.34	78
Least Squares	0.43 ± 0.17	0.62 ± 0.25	1.02 ± 0.4	1.65 ± 0.52	65
AIR	0.66 ± 0.28	0.70 ± 0.27	1.04 ± 0.37	1.73 ± 0.64	39
SPM1	0.62 ± 0.24	0.94 ± 0.38	1.73 ± 0.73	1.80 ± 0.77	11
SPM2	2.43 ± 1.26	3.24 ± 1.55	7.66 ± 3.29	10.77 ± 5.44	162

Data are shown for different percentages of pixels violating the intensity conservation assumption. Last column presents the mean execution time for each method (in seconds).

tion of 2.0 voxels, and mean rotation angle of 0.36°. Noise was added to the first subvolume to imitate the differences that one might expect between real volumes acquired at different times with different pulse sequences. Specifically, a certain number of voxels were selected as outliers; the image intensities at these voxels were set to random values, thereby grossly violating the intensity conservation assumption. The intensities at other voxels were corrupted with additive Gaussian white noise, with a power equal to $\frac{1}{16}$ the power of the original images. We performed a number of simulations with different random initial poses, different numbers of outliers, and different realizations of the additive Gaussian white noise. To quantify the accuracy of the estimated registration parameters, we computed the modulus of the error between the original coordinates and the coordinates transformed by the estimated rigid body transformation, averaged across all of the voxels in the brain.

Table 3 compares the mean error (in units of voxels), along with the corresponding standard deviation (in units of voxels), for each of the registration algorithms: 1) the proposed robust estimator, 2) a least-squares estimator (same as our proposed algorithm but with a quadratic function of the residual error), 3) AIR, 4) SPM1, and 5) SPM2. For all cases when there are more than 2.5% outliers, the robust method consistently produced the most accurate results. When there are no outliers, all methods (except SPM2) provided similar accuracy. However, even a small number of outliers caused a noticeable increase in the error for both AIR and SPM. The use of a robust estimator is critical to achieve accurate estimates in cases where the intensities are not perfectly matched in the two volumes (compare the first and second rows of Table 3). The accuracy of the robust method in our simulations was always better than 1 voxel. Since the voxel size is about 1 mm for our high-resolution volumes, we conclude that the mean error of the alignment is less than 1 mm.

We also tried to align real volumes similar to those in Figs. 4 and 5 using AIR and SPM, both with and without intensity normalization. Both SPM1 and SPM2 failed to align the volumes. The results using AIR were always worse than the results obtained using our robust method, although the differences were subtle.

Finally, we also measured the average execution time during the simulations (on a Pentium III, 450 MHz). Although the execution time is relatively long for our method in comparison to AIR and SPM1 (Table 3), this is primarily because it is currently implemented in Matlab. An efficient implementation of the robust method in the C program-

ming language could reduce the execution time to levels comparable with those achieved by AIR.

CONCLUSIONS

Automatic alignment of MRI volumes is a valuable tool to avoid time-consuming manual alignments, both in basic research and in clinical applications. Several approaches have been proposed to perform this alignment automatically, but they do not take advantage of the techniques reported in the extensive computer vision literature on image motion estimation and registration.

We adopted standard techniques from the computer vision literature to develop an algorithm for aligning MRI volumes. The basis of the algorithm is the gradient constraint that relates the displacement of a voxel with the spatial derivatives of the intensity at that voxel. The other components of the algorithm (robust estimation, multiresolution, intensity normalization, and contrast equalization) are all critical for obtaining accurate alignments, given that the volumes have been obtained with different protocols and coils. The algorithm has been used routinely for several months, and has never failed to produce a satisfactory alignment.

Robust estimation and multiresolution are tools that are generally applicable to a wide variety of signal and image processing problems. These tools could also be applied to improve the performance of other approaches to MRI volume registration, including AIR and SPM.

ACKNOWLEDGMENTS

APPENDIX A: ROBUST ESTIMATION

This appendix provides more details about the robust estimator used in our implementation. Given the following system of linear equations,

$$\mathbf{a}_i^T \mathbf{p} = b_i, \ i = 1, \ldots, N \qquad [10]$$

robust M-estimators minimize a sum, $\Sigma_i \rho(r_i)$, where $r_i = \mathbf{a}_i^T \mathbf{p} - b_i$ are the residuals, and where the function $\rho(r)$ saturates for large residual errors (see the graph of Fig. 2b). Among the different choices for the robust error function (14) we chose Beaton and Tukey's biweight function (15):

$$\rho(r) = \begin{cases} (C_B^2/2)(1 - (1 - (r/C_B)^2)^3) & \text{if } |r| \le C_B \\ C_B^2/2 & \text{otherwise} \end{cases} \quad [11]$$

where C_B is a parameter that controls the saturation point. This kind of estimator can be implemented very efficiently using iterative reweighted least squares (15,16), where the weights are updated at each iteration such that those points displaying greater residual errors are given less weight in the next iteration. The estimated parameters are refined at each iteration as follows:

$$\mathbf{p}_{j+1} = \mathbf{p}_j - \left[\mathbf{A}^T \mathbf{W} \left(\frac{\mathbf{A}\mathbf{p}_j - \mathbf{b}}{\sigma} \right) \mathbf{A} \right]^{-1}$$
$$\times \mathbf{A}^T \mathbf{W} \left(\frac{\mathbf{A}\mathbf{p}_j - \mathbf{b}}{\sigma} \right) (\mathbf{A}\mathbf{p}_j - \mathbf{b}) \quad [12]$$

where $\mathbf{p}_j$ are the estimated parameters at iteration j, $\mathbf{A}$ is the matrix formed by $\mathbf{a}_i^T$, $\mathbf{b}$ is the column matrix formed by b_i, $\mathbf{W}(\mathbf{r})$ is a diagonal matrix whose diagonal elements are the weights given by:

$$w(r_i) = \begin{cases} (1 - (r_i/C_B)^2)^2 & \text{if } |r_i| \le C_B \\ 0 & \text{otherwise} \end{cases} \quad [13]$$

and where σ is a scale parameter that can be estimated from the residuals (16):

$$\sigma = 1.48 \text{med} \{|r_i - \text{med} \{r_i\}|\} \quad [14]$$

where med{·} is the median value, which is taken over all the voxels, $i = 1, \ldots, N$. The iterations stop when the total squared error at iteration j, defined as:

$$E_j^2 = \frac{\sum_{i=1}^{N} w_i r_i^2}{\sum_{i=1}^{N} w_i}, \quad [15]$$

does not show a significant reduction with respect to the error in the previous iteration, or when a maximum number of iterations is reached.

We empirically adjusted the value of $C_B = 2.5$ in Eq. [11]. This value is more restrictive for outliers than the value of $C_B = 4.685$ suggested in (16) for Gaussian noise. In any case, we checked that the choice of this parameter is not critical.

APPENDIX B: INTENSITY CORRECTION

In this appendix we describe in more detail the preprocessing operations towards the intensity correction that we used for the results presented in this article.

The intensity inhomogeneities in MRI images can be modeled as a multiplicative gain field that varies slowly in space (19,23). A basic version of intensity normalization simply divides the volume intensity by an estimate of the smooth gain field. This gain field can be estimated using a polynomial model in the logarithm of the intensity, or through low-pass filtering. We tried both estimates, finding that a low-pass filtered version (in x, y and z) of the volume is sufficient for our purposes. The corrected volume intensity $f_n(x, y, z)$ is then computed as:

$$f_n(x, y, z) = \frac{f(x, y, z)}{f(x, y, z) * g(x, y, z)} \quad [16]$$

where $f(x, y, z)$ is the original volume intensity, $g(x, y, z)$ is a low-pass filter, and $*$ stands for convolution. However, this expression can be unstable in regions where the denominator, $f(x, y, z) * g(x, y, z)$, is very small due to an absence of signal, thereby amplifying the noise. For this reason we used a Wiener-like normalization, which takes into account the noise, so that when the normalizing bias field is smaller than the noise level, the noise is not amplified. The final expression for the corrected intensity volume is:

$$f_n(x, y, z) = \frac{f(x, y, z)(f(x, y, z) * g(x, y, z))}{(f(x, y, z) * g(x, y, z))^2 + \sigma_n^2} \quad [17]$$

where σ_n^2 is an estimate of the noise variance.

We applied the cubic-B-spline filter in Table 2 to each spatial axis two times to estimate the gain field. The mean noise level σ_n^2 in Eq. [17] has been estimated as the local variance of the intensity values (computed within each local block of $7 \times 7 \times 7$ voxels), plus a constant term equal to half the average of all of the local variance values. This empirically adjusted value is not an accurate estimate of the noise in the MRI images, but it has been chosen to ensure the stability of the normalization in very low intensity regions.

After removing the intensity gradients, the two volumes still differ from one another. This is illustrated in Fig. 6a, which plots sample histograms for the two intensity normalized volumes. The peak and variance of the main lobe, located around 1, is slightly different for the two volumes. These differences also violate the intensity conservation assumption, and must be corrected before estimating the alignment parameters.

The simplest strategy might be to subtract the mean value from each volume and divide by its standard deviation. However, the mean and variance of each volume depends on the extent of the regions within the volumes that correspond to skin and skull and the regions outside the head. For example, if an acquisition covers a very wide field of view (outside the head) then it will contain many more voxels with low intensities that will bias the mean and variance.

The central lobes in each of the histograms in Fig. 6a, which include most of the voxels, correspond to the white and gray matter. However, there is also a narrow peak near 0, corresponding to voxels that before normalization were
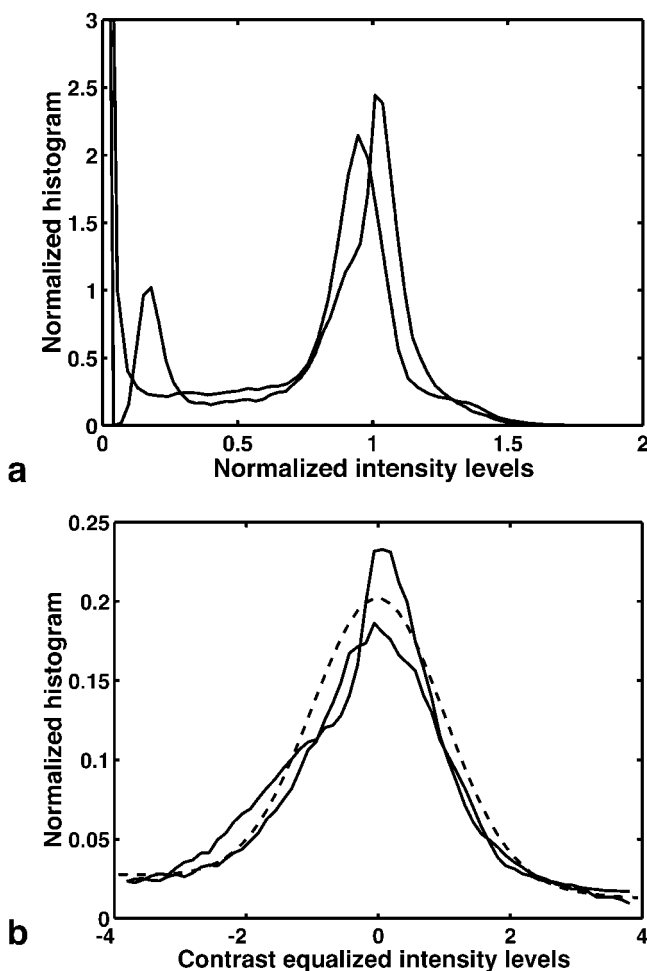
a:

b:

FIG. 6. **a:** Sample histograms of the two volumes, after intensity normalization. **b:** Solid curve, sample histograms after contrast equalization. Dashed curve, best fit (sum of two Gaussians) model of the histograms.

in a region of low intensity (outside the head or cerebrospinal fluid). There is also a residual, broad lobe, ranging from low to very high intensities, which corresponds to regions in the skin and the skull.

Our goal is to equalize the contrast of the regions corresponding to the gray and white matter. To avoid the influence of the other regions, we modeled the intensity histograms as the sum of two Gaussian densities, one for midrange values corresponding the gray and white matter of the brain, and the other one for the broad distribution of very low to very high values. Then the mean, standard deviation, and peak value of these two Gaussian functions are estimated as those that minimize the following robust measure of the error between the sample histogram and the model:

$$E(\mu_1, \sigma_1, M_1, \mu_2, \sigma_2, M_2)$$

$$= \sum_x -\frac{1}{1 + (h(x) - g_1(x) - g_2(x))^2} \quad [18]$$

where $g_i = M_i/(\sqrt{2\pi}\sigma_i)\exp(-(x - \mu_i)^2/(2\sigma_i^2))$ is a Gaussian function, $h(x)$ is the sample histogram, and the summation is performed over all the values x where the sample histogram has been computed. We minimized the expression in Eq. [18] using a simplex method, and then used the mean and variance of the central (main lobe) Gaussian to contrast equalize the two volumes.

Figure 6b plots the sample histograms within ±4 standard deviations (solid curves) after contrast equalization, along with the corresponding shifted and scaled sum of Gaussians fit (dashed line). Using the robust model fit, the narrow peak close to 0 was effectively ignored in estimating the mean and variance of the central lobe of each histogram. This allowed us to perform the contrast equalization of the gray and white matter regions without first performing a difficult and computationally expensive image segmentation.

REFERENCES

1. Woods RP, Cherry SR, Mazziotta JC. Rapid automated algorithm for aligning and reslicing PET images. J Comput Assist Tomogr 1992;16:620–633.
2. Kumar R, Dana K, Anandan P, Okamoto N, Bergen J, Hemler P, Sumanaweera TS, van den Elsen PA, Adler J. Frameless registration of MR and CT 3D volumetric data sets. In: Proceedings of the Second IEEE Workshop on Applications of Computer Vision, Sarasota, FL, Dec. 1994. p 240–248.
3. Wells WM III, Viola P, Kikinis R. Multi-modal volume registration by maximization of mutual information. In: Proceedings of 2nd International Symposium on Medical Robotics and Computer Assisted Surgery, Baltimore, MD, Nov. 1995. p 55–62.
4. Ashburner J, Neelin P, Collins DL, Evans AC, Friston KJ. Incorporating prior knowledge into image registration. NeuroImage 1997;6:344–352.
5. Ashburner J, Friston KJ. Multimodal image coregistration and partitioning—a unified framework. NeuroImage 1997;6:209–217.
6. Woods RP, Grafton ST, Holmes CJ, Cherry SR, Mazziotta JC. Automated image registration. I. General methods and intrasubject, intramodality validation. J Comput Assist Tomogr 1998;22:139–152.
7. West J, Fitzpatrick M, Wang MY, Dawant BM, Maurer CR Jr, Kessler RM, Maciunas RJ, Barillot C, Lemoine D, Collignon A, Maes F, Suetens P, Vandermeulen D, van den Elsen PA, Napel S, Sumanaweera TS, Harkness B, Hemler PF, Hill DLG, Hawkes DJ, Studholme C, Maintz JBA, Viergever MA, Malandain G, Pennec X, Noz ME, Maguire GQ Jr, Pollack M, Pelizzari CA, Robb RA, Hanson D, Woods RP. Comparison and evaluation of retrospective intermodality brain image registration techniques. J Comput Assist Tomogr 1997;21:554–566.
8. Fennema CL, Thompson WB. Velocity determination in scenes containing several moving objects. Comput Vision Graphics Image Process 1979;9:301–315.
9. Horn BKP, Schunk BG. Determining optical flow. Arti Intell 1981;17:185–203.
10. Lucas BD, Kanade T. An iterative image registration technique with an application to stereo vision. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence, Vancouver, 1981. p 674–679.
11. Bergen JR, Anandan P, Hanna K, Hingorani R. Hierarchical model-based motion estimation. In: Proceedings of Second European Conf. on Comp. Vis., Santa Margherita Ligure, Italy, Springer-Verlag, 1992. p 237–252.
12. Barron J, Fleet DJ, Beauchemin SS. Performance of optical flow techniques. Int J Comput Vision 1994;12:43–77.
13. Black MJ, Anandan P. The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. Comput Vision Image Understand 1996;63:75–104.
14. Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA. Robust statistics: the approach based on influence functions. New York: John Wiley & Sons; 1986.
15. Beaton AE, Tukey JW. The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. Technometrics 1974;16:147–185.
16. Holland PW, Welsch RE. Robust regression using iteratively reweighted least-squares. Commun Stat Theory Methods 1977;A6:813–827.

17. Burt P, Adelson EH. The Laplacian pyramid as a compact image code. IEEE Trans Commun 1983;31:532–540.
18. Wolberg G. Digital image warping. Los Alamitos, CA: IEEE Computer Society Press; 1994.
19. Wells WM III, Grimson WEL, Kikinis R, Jolesz FA. Statistical intensity correction and segmentation of MRI data. In: Proceedings of the SPIE, vol. 2359, Rochester, MN, 1994. p 13–24.
20. Meyer CR, Bland PH, Pipe J. Retrospective correction of intensity inhomogeneities in MRI. IEEE Transact Med Imaging 1995;14:36–41.
21. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 1998;17:87–97.
22. Simoncelli EP. Design of multi-dimensional derivative filters. In: Proceedings of International Conference on Image Processing, vol. I, Austin, TX, 1994. p 790–793.
23. Moyher SE, Vigneron DB, Nelson SJ. Surface coil MR imaging of the human brain with an analytic reception profile correction. J Magn Reson Imaging 1995;5:139–144.