ELSEVIER

Journal of Neuroscience Methods 142 (2005) 67–76

JOURNAL OF NEUROSCIENCE METHODS

www.elsevier.com/locate/jneumeth

Quantitative comparison of algorithms for inter-subject registration of 3D volumetric brain MRI scans

Babak A. Ardekani[a,c,*], Stephen Guckemus[a], Alvin Bachman[a], Matthew J. Hoptman[b,c], Michelle Wojtaszek[a], Jay Nierenberg[a,c]

[a] *Center for Advanced Brain Imaging, Nathan Kline Institute for Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA*
[b] *Clinical Research Division, Nathan Kline Institute for Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA*
[c] *Department of Psychiatry, New York University School of Medicine, USA*

Received 13 April 2004; received in revised form 20 July 2004; accepted 22 July 2004

Abstract

The objective of inter-subject registration of three-dimensional volumetric brain scans is to reduce the anatomical variability between the images scanned from different individuals. This is a necessary step in many different applications such as voxelwise group analysis of imaging data obtained from different individuals. In this paper, the ability of three different image registration algorithms in reducing inter-subject anatomical variability is quantitatively compared using a set of common high-resolution volumetric magnetic resonance imaging scans from 17 subjects. The algorithms are from the *automatic image registration* (AIR; version 5), the *statistical parametric mapping* (SPM99), and the *automatic registration toolbox* (ART) packages. The latter includes the implementation of a non-linear image registration algorithm, details of which are presented in this paper. The accuracy of registration is quantified in terms of two independent measures: (1) post-registration spatial dispersion of sets of homologous landmarks manually identified on images before or after registration; and (2) voxelwise image standard deviation maps computed within the set of images registered by each algorithm. Both measures showed that the ART algorithm is clearly superior to both AIR and SPM99 in reducing inter-subject anatomical variability. The spatial dispersion measure was found to be more sensitive when the landmarks were placed after image registration. The standard deviation measure was found sensitive to intensity normalization or the method of image interpolation.

Keywords: MRI; Brain; Image registration; Spatial normalization

1. Introduction

An important methodological consideration for analysis of human brain imaging data is *inter-subject registration* or *spatial normalization* of images acquired from different individuals. The aim of inter-subject registration is to reduce the anatomical variability in three-dimensional (3D) volumetric brain scans obtained from different subjects. For example, inter-subject registration allows voxelwise group analysis of functional magnetic resonance imaging (fMRI) data (Svensen et al., 2002; Zeffiro et al., 1997), and studies of brain white matter using diffusion tensor imaging (Ardekani et al., 2003; Jones et al., 2002). Another class of applications of inter-subject registration can be categorized as automatic 'image understanding', where higher order information (tissue type, locations of anatomical landmarks, structural boundaries, specific sulci/gyri, etc.) that are known about a *template* image are obtained about a *subject* or *test* image after the subject image is registered to the template by non-linear spatial transformation or deformation (Collins et al., 1995; Marroquin et al., 2002; Webb et al., 1999). An additional important application is the quantification of small changes in volume observed in anatomical structures over time (Holden et al., 2002; Rey et al., 2002) that can be used in diagnosis and evaluation of disease progression and treatment.

* Corresponding author. Tel.: +1 845 398 5471; fax: +1 845 398 5472.
E-mail address: ardekani@nki.rfmh.org (B.A. Ardekani).

0165-0270/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jneumeth.2004.07.014

Spatial registration is specified in terms of a 3D transformation or displacement field $\boldsymbol{w}$: $\Re^3 \rightarrow \Re^3$ which is applied to the subject image, $I_s(\boldsymbol{r})$, to obtain the spatially normalized or warped image $I_w(\boldsymbol{r}) = I_s(\boldsymbol{r} + \boldsymbol{w}(\boldsymbol{r}))$. The main objective of most inter-subject image registration algorithms is to find a displacement field $\boldsymbol{w}$ such that the warped image $I_w(\boldsymbol{r})$ is as 'similar' as possible to a template image, $I_t(\boldsymbol{r})$. For a given pair of subject and template images I_s and I_t, algorithms mainly differ in their approach to modeling and estimation of $\boldsymbol{w}$. In the SPM99 software package (Friston et al., 1995), the displacement field $\boldsymbol{w}$ is modeled by a finite orthogonal series with trigonometric basis functions. The algorithm computes a displacement field by estimating the coefficients of the series using an iterative linearized least-squares approach. In the automated image registration (AIR) software package (Woods et al., 1998), the displacement field is modeled as a polynomial, and the algorithm computes the displacement field by estimating the polynomial coefficients using non-linear least-squares optimization. Other methods of inter-subject registration model the displacement field as the displacement field in an elastic object (Bajcsy and Kovačič, 1989) or a viscous fluid (Christensen et al., 1997; Christensen and Johnson, 2001) reacting to internal forces proportional to 'image mismatch' between template and subject volumes. These methods are governed by partial differential equations that model the physical phenomenon. In several other methods, the displacement fields are estimated as non-parametric vector fields subject to regularity constraints (Collins et al., 1995; Kjems et al., 1999; Kosugi et al., 1993).

Although many methods have been proposed for spatial normalization, there are few studies comparing the performance of various algorithms on a common set of real MRI data. At least two important questions can be raised: (1) how do the algorithms compare in terms of their ability to reduce anatomic differences between subjects? And (2) how does image registration accuracy impact the final analysis results (e.g., activation maps in fMRI)? The present paper is concerned with the first question. The latter question has been previously addressed primarily in the context of positron emission tomography (PET) functional imaging studies (Crivello et al., 2002; Kjems et al., 1999; Senda et al., 1998). Overall, little difference has been found between the functional activation maps obtained by processing PET activation data using different inter-subject registration methods. Crivello et al. (2002) attributed this finding to the limited spatial resolution of PET and the inherent functional variability across subjects. However, this conclusion may not be automatically extended to higher resolution fMRI studies or group analysis of diffusion tensor imaging data. Recently, Ardekani et al. (2004) studied the impact of inter-subject registration on group analysis of fMRI data. They showed that increased accuracy of inter-subject registration in removing anatomic variability between subjects results in significant increases in the *sensitivity* of activation detection and the *reproducibility* of activation maps. Thus, at least in fMRI studies, it is important for researchers to understand the relative accuracy of inter-subject registration tools available to them.

The first objective of the present paper was to present details of our implementation of an inter-subject registration algorithm included in our *automatic registration toolbox* (ART). This algorithm is a non-parametric method similar to those proposed by Collins et al. (1995), Kjems et al. (1999), and Kosugi et al. (1993) with some new components. The second objective was to quantitatively compare the performance of three different inter-subject registration programs: (1) SPM99; (2) AIR; and (3) ART. The third aim was to evaluate different independent criteria for assessment of registration accuracy. One measure is the post-registration *spatial dispersion* of sets of homologous landmarks located manually on the images. Another criterion compares the sample standard deviation (S.D.) of voxel intensity maps computed from registered image sets. The precise mathematical definitions of these measures are presented in the following section.

2. Materials and methods

2.1. Implementation of ART inter-subject registration

Without loss of generality, we assume that the template image $I_t(\boldsymbol{r})$ and the subject image $I_s(\boldsymbol{r})$ are of the same matrix and voxel dimensions. If not, the subject image is resized and interpolated so that its voxel and matrix dimensions match those of the template image. The objective of ART is to find a displacement vector $\boldsymbol{w}(\boldsymbol{r}) = (u_x(\boldsymbol{r}), u_y(\boldsymbol{r}), u_z(\boldsymbol{r}))$ at each voxel $\boldsymbol{r}$. To achieve this, each voxel is visited in a raster scan fashion. Let $\Omega_{\boldsymbol{r}}$ be a neighborhood around and including voxel $\boldsymbol{r}$. The *template feature vector* at voxel $\boldsymbol{r}$, denoted by $\boldsymbol{f}^t_r$, is defined to be comprised of elements $\{I_t(\boldsymbol{v})\colon \boldsymbol{v} \in \Omega_{\boldsymbol{r}}\}$ of the template image. Similarly, the *subject feature vector* at voxel $\boldsymbol{r}$, denoted by $\boldsymbol{f}^s_r$, is constructed from the voxel values $\{I_s(\boldsymbol{v})\colon \boldsymbol{v} \in \Omega_{\boldsymbol{r}}\}$ of the subject image. We define the similarity between two arbitrary vectors $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ of the same dimension to be:

$$S(\boldsymbol{w}_1, \boldsymbol{w}_2) = \frac{\boldsymbol{w}_1^{\mathrm{T}} \boldsymbol{H} \boldsymbol{w}_2}{\sqrt{\boldsymbol{w}_2^{\mathrm{T}} \boldsymbol{H} \boldsymbol{w}_2}} \tag{1}$$

where H is an idempotent ($H^2 = H$) symmetric centering matrix defined so that it removes the mean of the vector it premultiplies. Note that Eq. (1) is asymmetric in the sense that $S(\boldsymbol{w}_1, \boldsymbol{w}_2) \neq S(\boldsymbol{w}_2, \boldsymbol{w}_1)$. This is of no practical significance but will save some computation, as will become clear below. Next consider a *search neighborhood*, $\psi_{\boldsymbol{r}}$, around and including voxel $\boldsymbol{r}$. Let voxel $\boldsymbol{q} \in \psi_{\boldsymbol{r}}$ be the voxel in this neighborhood at which the similarity measure $S(\boldsymbol{f}^t_r, \boldsymbol{f}^s_q)$ is maximum, that is:

$$S(\boldsymbol{f}^t_r, \boldsymbol{f}^s_q) = \max_{v \in \Psi_r} S(\boldsymbol{f}^t_r, \boldsymbol{f}^s_v) \tag{2}$$

Our initial estimate of the displacement field at voxel $\boldsymbol{r}$ is $\boldsymbol{w}(\boldsymbol{r}) = \boldsymbol{q} - \boldsymbol{r}$. Note that Eq. (1) can be made symmetric by

dividing the right-hand side by $\sqrt{\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{H}\boldsymbol{w}_1}$. This does not, however, change the voxel $\boldsymbol{q}$ at which S attains its maximum and will only increase the computational burden.

Described above is the essence of the registration algorithm implemented in ART. However, there still remain a number of very important details that need to be addressed. First, the registration algorithm can benefit from an initial linear registration, either a six-parameter rigid-body, or a 12-parameter affine transformation. In ART, there is an option for performing an initial rigid-body registration using the method described by Ardekani et al. (1995).

Next is the issue of speed. The speed of the algorithm depends on several factors including the number of image voxels, size of the feature neighborhood $\Omega_{\boldsymbol{r}}$, and size of the search neighborhood $\psi_{\boldsymbol{r}}$. In ART, the neighborhoods $\Omega_{\boldsymbol{r}}$ and $\psi_{\boldsymbol{r}}$ are cubic and centered on voxel $\boldsymbol{r}$. Initially, the search neighborhood $\psi_{\boldsymbol{r}}$ must be large enough to enable the algorithm to find large scale displacements. The search neighborhood is shrunk iteratively and the computations repeated for finding finer scale displacements. In practice, this is achieved through a multiresolution approach. In other words, the image is processed in multiple resolution levels using scale-space theory. The scale-space of a 3D image $I(\boldsymbol{r})$ is a 4D image $I(\boldsymbol{r}, c)$ that satisfies the isotropic diffusion partial differential equation:

$$\frac{\partial I(\boldsymbol{r}, c)}{\partial c} = \nabla^2 I(\boldsymbol{r}, c) \tag{3}$$

with the initial condition $I(\boldsymbol{r}, 0) = I(\boldsymbol{r})$. The solution to Eq. (3) is the scale-space of the image and is given by $I(\boldsymbol{r}, c) = G_c(\boldsymbol{r}) * I(\boldsymbol{r}, 0)$, where $G_c(\boldsymbol{r})$ is the Gaussian function with variance $2c$ and '*' denotes the convolution operation (Lindeberg, 1994). Therefore, the initial image is low-pass filtered with Gaussian kernels of various widths to form the scale-space. In ART, the core registration process that is described above is repeated in scale-space starting from coarse resolutions towards the finest resolution at $c = 0$. At each step, the displacement field found is applied to the subject image before starting the next iteration. The number of iterations is an option to the program. At coarse resolution levels, the algorithm is fast because: (1) the image is represented by fewer voxels with a proportionally larger voxel size; and (2) the search neighborhood $\psi_{\boldsymbol{r}}$ can be large but comprised of few voxels because of the larger voxel size. In order to use the displacement field obtained at a lower resolution level as the starting point for the next stage, the field components (u_x, u_y, u_z) need to be interpolated to match the voxel size at the higher resolution level. In the present method, this is achieved by using a fast digital filter implementation of the cubic spline interpolation method (Unser, 1999). Overall, the multiresolution approach improves the speed and robustness of the algorithm.

An additional strategy for increasing the computational efficiency of the algorithm was implemented by computing the template image gradient L_2 norm, $\|\nabla I_{\mathrm{t}}(\boldsymbol{r})\|$, at the highest resolution level and only updating the displacement field in those voxels where the gradient norm is in a certain upper percentile of the gradient magnitude histogram of the template image. The percentile level can be specified as an optional argument to the ART program. This is only done at the iteration corresponding to the finest resolution level that is the most computationally intensive iteration. The philosophy behind this approach is that voxels with low spatial derivatives $\|\nabla I_{\mathrm{t}}(\boldsymbol{r})\|$ are located in featureless 'planes' of the image and contain little information for guiding the algorithm and may even result in errors. The gradient vector $\nabla I_{\mathrm{t}}(\boldsymbol{r})$ is calculated by approximating the template image using cubic splines (Unser, 1999) enabling us to easily compute the necessary partial derivatives.

The next point to consider when implementing the algorithm is the regularization of the displacement field. Intuitively speaking, regularization is necessary to ensure that points that are close to each other in the subject image $I_{\mathrm{s}}(\boldsymbol{r})$ remain close in the warped image $I_{\mathrm{w}}(\boldsymbol{r})$. Kjems et al. (1999) reviewed several methods of displacement field regularization. In the present paper, we use simple Gaussian low-pass filtering of the displacement field that is obtained at the end of each iteration in the multiresolution algorithm. This method was also applied by Kosugi et al. (1993) and has worked well in our experiments.

Another very important issue that needs to be addressed when solving for the displacement field $\boldsymbol{w} = (u_x, u_y, u_z)$ is that the algorithm must ensure that the resulting non-linear transformation is a *homeomorphism*, that is, a continuous mapping between two spaces that has an inverse which is also continuous. Since at each multiresolution iteration, the displacement field is interpolated using a cubic spline fit, we can easily compute the partial derivatives of the displacement field components (u_x, u_y, u_z) with respect to the x, y, and z spatial coordinates using the cubic spline coefficients. This allows us to compute the Jacobian determinant of the deformation field $\boldsymbol{r} + \boldsymbol{w}(\boldsymbol{r})$ at every voxel. We ensure that the Jacobian determinant is strictly positive at every voxel. This, together with the fact that the displacement field is kept to zero at all air voxels surrounding the head, can be shown to be a necessary and sufficient condition for the transformation to be homeomorphic (Kaplan, 1973). If the requirement of a positive Jacobian everywhere in the image is not met, we incrementally increase the width of the smoothing Gaussian kernel and repeat the procedure until the condition is met. Smoothing is guaranteed to have the desired effect because at the limit of infinite width smoothing kernel, the displacement field $\boldsymbol{w}(\boldsymbol{r})$ approaches a constant and, therefore, the Jacobian determinant of the deformation $\boldsymbol{r} + \boldsymbol{w}(\boldsymbol{r})$ approaches 1.

Finally, an advantage of methods such as AIR and SPM99 that represent the displacement fields in terms of basis functions is that the displacement fields can be represented and stored very efficiently using only a few parameters corresponding to the coefficients of the basis functions. We take advantage of this fact by approximating the components

of the displacement fields (u_x, u_y, u_z) by a truncated Fourier–Legendre series as follows:

$$u(\boldsymbol{r}) = \sum_{n,m,q=0}^{M} c_{nmq} P_n(x) P_m(y) P_q(z) \quad (4)$$

where P_n denotes a Legendre polynomial of degree n (Kaplan, 1973). The coefficients c_{nmq} of the series are efficiently computed using the orthogonality property of Legendre polynomials. The coefficients are stored and can be recalled later to synthesize the displacement field.

2.2. MRI scans

Three dimensional magnetization prepared rapid acquisition gradient echo (MP-RAGE) volumetric MRI scans were conducted on 17 healthy individuals (eight males, nine females, mean (S.D.) age = 31 (8.85) years) using a 3.0 Tesla Surrey Medical Imaging Systems (SMIS) scanner. The acquisition parameters were: matrix size = 256 × 256 × 128; voxel size = 0.976 mm × 0.976 mm × 1.3 mm; $\alpha = 18^\circ$; TE = 4 ms; TI = 1100 ms; and TR = 3000 ms. The data were part of a larger fMRI study approved by the local Institutional Review Board. Informed consent was obtained from all subjects.

2.3. Preprocessing

The extra cranial regions were automatically deleted from the 3D MRI volumes using the brain extraction tool (BET) software (Smith, 2002). In cases where BET did not completely remove the non-brain regions, manual editing was performed using the MEDx software package (Sensor Systems, Inc., Sterling, Virginia).

2.4. Registration

The seventeen subjects were ranked according to the volume of their intra-cranial content. The MRI scan of the subject with the median intra-cranial volume was selected as the template image set. The remaining 16 subjects were registered to the template volume using the nonlinear fifth order polynomial registration algorithm of AIR software package, the spatial normalization algorithm of SPM99, and the nonlinear image registration algorithm in ART. Thus, we obtained 48 spatially normalized volumes (16 for each algorithm). In all cases, the registered volumes were obtained using linear interpolation.

Implementation of inter-subject registration in ART was described above. Algorithms used in AIR and SPM99 are summarized by Friston et al. (1995) and Woods et al. (1998), respectively. Both AIR and SPM99 use the least-squared image difference cost function as a measure of registration accuracy while ART uses the *cross-correlation* criterion given in Eq. (1). The fifth order polynomial registration of AIR software has 168 parameters. The degrees of freedom used when applying the SPM99 software was the default of 7 × 8 × 7 = 392 basis functions. The effective degrees of freedom in ART can be considered to be $(M + 1)^3$, where M is the order of the truncated Fourier–Legendre series given in Eq. (4). In the present paper, $M = 14$, therefore, the degrees of freedom of ART was 3375.

2.5. Dispersion of homologous landmarks post-registration

Let $L = \{\boldsymbol{q}_i\}$ ($i = 1, 2, \ldots, N$) represent a set of N landmarks identified on each of $N = 16$ original images *before* registration. After spatial normalization, each point is mapped to a new location $\boldsymbol{r}_i$ resulting in a new set of points $L' = \{\boldsymbol{r}_i\}$. A measure of dispersion of this set can be defined as the root mean square distance from the mean landmark position within set L' given by:

$$C_{L'} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \|\boldsymbol{r}_i - \bar{\boldsymbol{r}}\|^2} \quad (5)$$

Twenty landmarks on each of the 16 original unregistered image volumes were manually located by an operator trained in neuroanatomy. These landmarks included the left and right frontal poles, the left and right temporal poles, the superior aspect of the anterior commissure (AC), the superior aspect of the posterior commissure (PC), the most anterior point of the corpus callosum (CC), the most posterior point of the CC, the left and right frontal horns, the left and right occipital poles, right and left anterior and posterior aspects of the putamen, right and left anterior aspect of the caudate nucleus, and the right and left inferior aspect of the lateral extension of the fourth ventricle. Landmarks were located using software written in IDL (Research Systems, Inc., Boulder, Colorado). Thus, we obtained coordinates for the 20 different landmarks in each of the 16 volumes. Following landmark selection, the 16 original volumes were warped using the deformation fields obtained from AIR, SPM99, and ART programs. Thus, we obtain 48 deformed image volumes. The new locations of each of the original 20 landmarks were located on the 48 spatially normalized volumes. The dispersion of each set of homologous landmarks was calculated according to Eq. (5), resulting in a 20 × 3 (landmarks × methods) table, which was subsequently analyzed using a variety of statistical tests. It should be emphasized that only one set of landmarks are manually located on each subject's original image volume.

2.6. Sample standard deviation maps

Let the N registered volumes obtained by using a given algorithm be represented by sets of voxel intensities $\{v_{ij}\}$, where $i = 1, 2, \ldots, N$ is the subject index, $j = 1, 2, \ldots, V$ is the voxel index, and V is the total number of voxels in a selected subset of image voxels. Then at each voxel j, a

sample standard deviation S.D.$_j$ may be computed as:

$$\text{S.D.}_j = \sqrt{\sum_{i=1}^{N} \frac{(v_{ij} - \bar{v}_j)^2}{N-1}} \quad (6)$$

where $\bar{v}_j$is the mean voxel intensity at voxel j in the N volumes. Let two different methods be denoted by 'A' and 'B' with their corresponding voxelwise sample standard deviations as S.D.$_{\text{A}j}$ and S.D.$_{\text{B}j}$. The sample S.D.-map calculated from a set of registered volumes reflects the degree of success of the registration algorithm in reducing the inter-subject anatomical variability. Under the null hypothesis that the two methods have equal registration ability, each algorithm is expected to have a smaller S.D. in about half of the voxels. This hypothesis can be tested using a sign test. Let s be a random variable representing the number of voxels out of V with S.D.$_{\text{A}j}$ < S.D.$_{\text{B}j}$. The probability that $s > s_0$ is given by:

$$\Pr(s > s_0) = 1 - \sum_{s=0}^{s_0} \frac{V!}{s!(V-s)!}(0.5)^V \quad (7)$$

This probability measure can be used to test the null hypothesis that two algorithms reduce anatomical variability to the same extent.

3. Results and discussion

The 16 MP-RAGE volumes were registered to the template volume using ART, AIR, and SPM99, resulting in a total of 48 spatially normalized volumes (16 per method). The 16 registered volumes obtained from applying each algorithm were averaged. Selected slices from the resulting three average volumes are shown in Fig. 1. Qualitatively, it can be clearly seen that the average registered volume corresponding to ART (first column) has higher resolution than those corresponding to AIR (second column) and SPM99 (third column). The gyral patterns can be more clearly identified in the average images processed by ART. Fig. 2 shows the S.D.-maps corresponding to the averages in Fig. 1. These maps are displayed using the same gray-level window so that their intensity can be directly compared. Qualitatively, it can be seen that ART has lower variance in general, particularly in regions around the ventricles. It is also noted that the three methods have more difficulty in registering the parietal lobe region between subjects as compared to other cortical regions. Visually, there does not appear to be a significant difference between the AIR (second column) and SPM99 (third column) S.D.-maps.

Quantitative comparisons were conducted using the S.D.-maps. One thousand voxels (V = 1000) were randomly selected across the volume, while ensuring that no two voxels were neighbors. Neighborhood definition was based on eight-connectively, that is, voxels that share at least a vertex were considered neighbors. The scatter plots for the three comparisons are shown in Fig. 3. In these plots, the scattered points lie mainly below the $y = x$ line, which indicates advantage of the method plotted as the ordinate. When comparing AIR to ART using this approach, in 715 voxels S.D.$_{\text{ART}}$ < S.D.$_{\text{AIR}}$ ($P \approx 0$; two-tailed). When comparing SPM99 to ART, in 761 voxels S.D.$_{\text{ART}}$ < S.D.$_{\text{SPM}}$ ($P \approx 0$; two-tailed). Both results indicate that the images registered using ART are significantly less variable that those obtained using either SPM99 or AIR. When comparing SPM99 to AIR, in 553 voxels S.D.$_{\text{AIR}}$ < S.D.$_{\text{SPM}}$ (P = 0.0007; two-tailed). Therefore, the variance in images registered by AIR is significantly less than the variance in images registered using SPM99. However, one must be careful when interpreting this latter result. The reason is that the AIR program automatically scales the images in a way that tends to reduce the variance. Therefore, we cannot rely on this method for quantitative comparison of AIR. In contrast, SPM99 and ART do not change the image intensity and only apply spatial deformation to the original image.

Table 1
Post-registration dispersion of landmarks (mm), with landmarks placed before registration

Landmark	Method		
	ART	AIR	SPM99
Frontal Pole—L	2.64	5.37	3.55
Frontal Pole—R	2.48	4.94	2.72
Temporal Pole—L	2.41	3.31	2.60
Temporal Pole—R	2.88	3.78	3.74
Anterior Commissure	1.06	1.20	1.63
Posterior Commissure	1.35	1.73	2.07
Corpus Callosum—A	1.77	1.58	1.65
Corpus Callosum—P	1.85	2.28	2.04
Frontal Horn—L	1.69	1.78	1.67
Frontal Horn—R	1.93	1.52	1.62
Occipital Pole—L	2.49	3.70	2.94
Occipital Pole—R	2.34	3.99	2.75
Putamen—R A	1.98	2.07	1.91
Putamen—R P	2.22	2.42	1.92
Putamen—L A	1.65	2.04	1.91
Putamen—L P	1.76	2.09	2.09
Caudate Nucl.—R A	1.91	1.81	1.72
Caudate Nucl.—L A	2.38	1.98	2.46
Lat. 4th Ven.—R I	1.09	1.22	1.50
Lat. 4th Ven.—L I	1.26	1.31	1.55
Mean	1.96	2.51	2.20
S.D.	0.52	1.24	0.66

Note: L, left; R, right; A, anterior; P, posterior; I, inferior; Lat. 4th Ven., lateral extension of the 4th ventricle.

The dispersion of homologous landmarks post-registration is shown in Table 1. The values for ART in the first column are mostly lower than those of AIR and SPM. Paired t-tests were applied to assess the statistical significance of the differences in registration accuracy. Comparison of ART and SPM99 yielded t = −3.07 (P = 0.006), meaning that the dispersions were significantly smaller in ART as compared to SPM99. For ART versus AIR, t = −2.83 (P = 0.011), again indicating greater accuracy of ART over AIR. Comparing AIR and SPM99 we found t = 1.87 (P = 0.077) indicating no significant difference. It can be seen that the dispersion of homologous

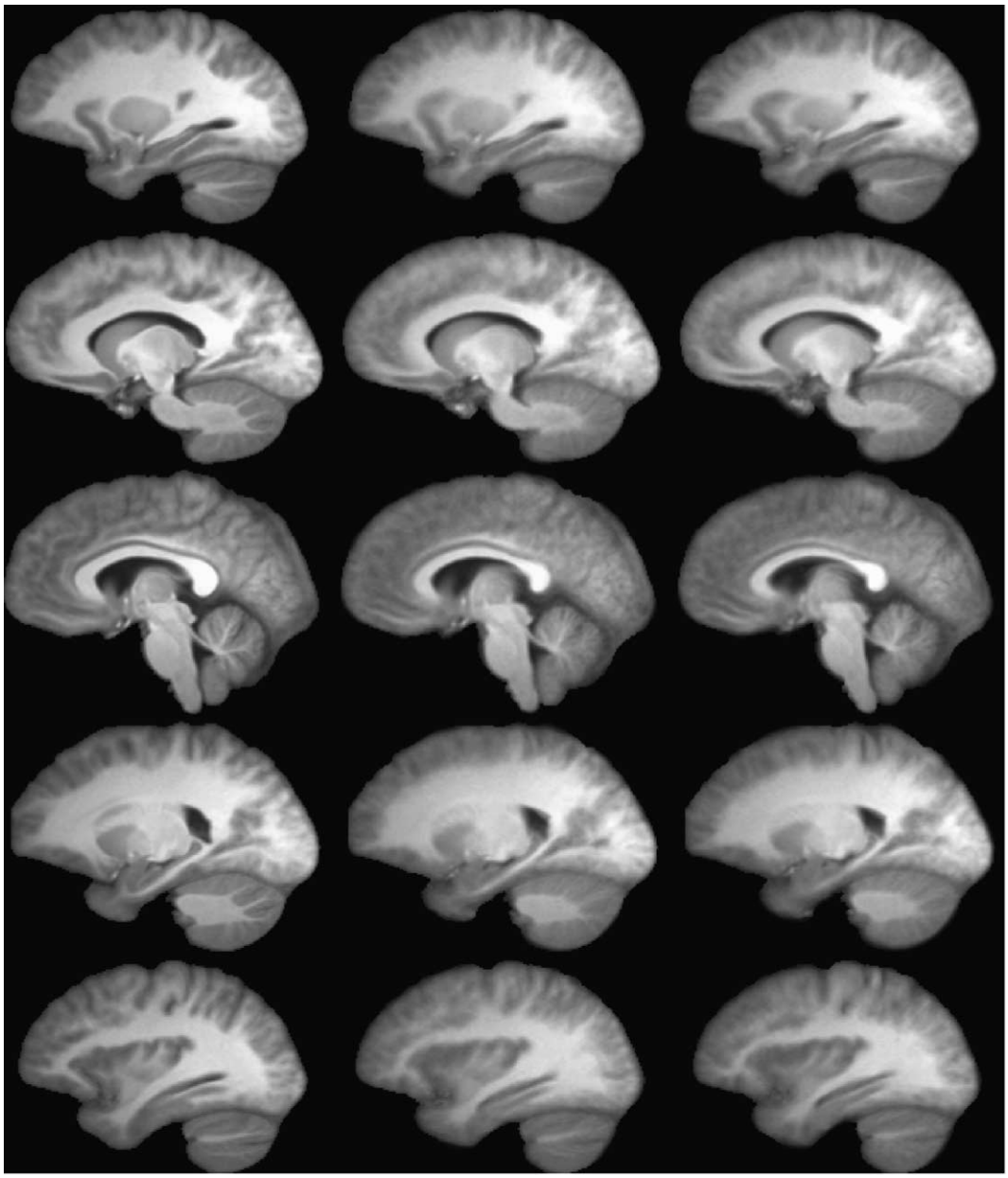

Fig. 1. Selected slices from the average of 16 volumes registered using ART (first column), AIR (second column), and SPM99 (third column). The images obtained by averaging the volumes registered using ART clearly have a higher resolution indicating that a greater reduction of inter-subject anatomical variability is achieved by ART as compared to AIR or SPM99.

landmarks post-registration is less sensitive than S.D.-maps in indicating differences between algorithms. But it should be noted that the S.D.-maps are quite sensitive to any changes in image intensities that an algorithm may apply. For example, since AIR applies an arbitrary scaling factor to the images, the S.D.-maps of AIR cannot be directly compared with those of SPM99 and ART. Interpolation methods also significantly affect this measure and it must be ensured that the same interpolation method is used in reconstructing the spatially normalized images obtained from different algorithms, otherwise the S.D.-maps cannot be used to compare the accuracy of the methods.

The dispersions of homologous landmarks post-registration can been considered to be due to two factors. Firstly, it can be due to 'errors' by the operator in selecting the landmarks. Secondly, it can be due to inability of the algorithm to accurately match the landmarks together. Our objective in using this criterion is to quantify and compare the latter factor between different registration methods. The presence of the first factor, however, diminishes the sensitivity of this approach. To increase the sensitivity, we must reduce the landmark selection error as much as possible. One approach is to identify the landmarks *after* the images have been registered (Grachev et al., 1999). It

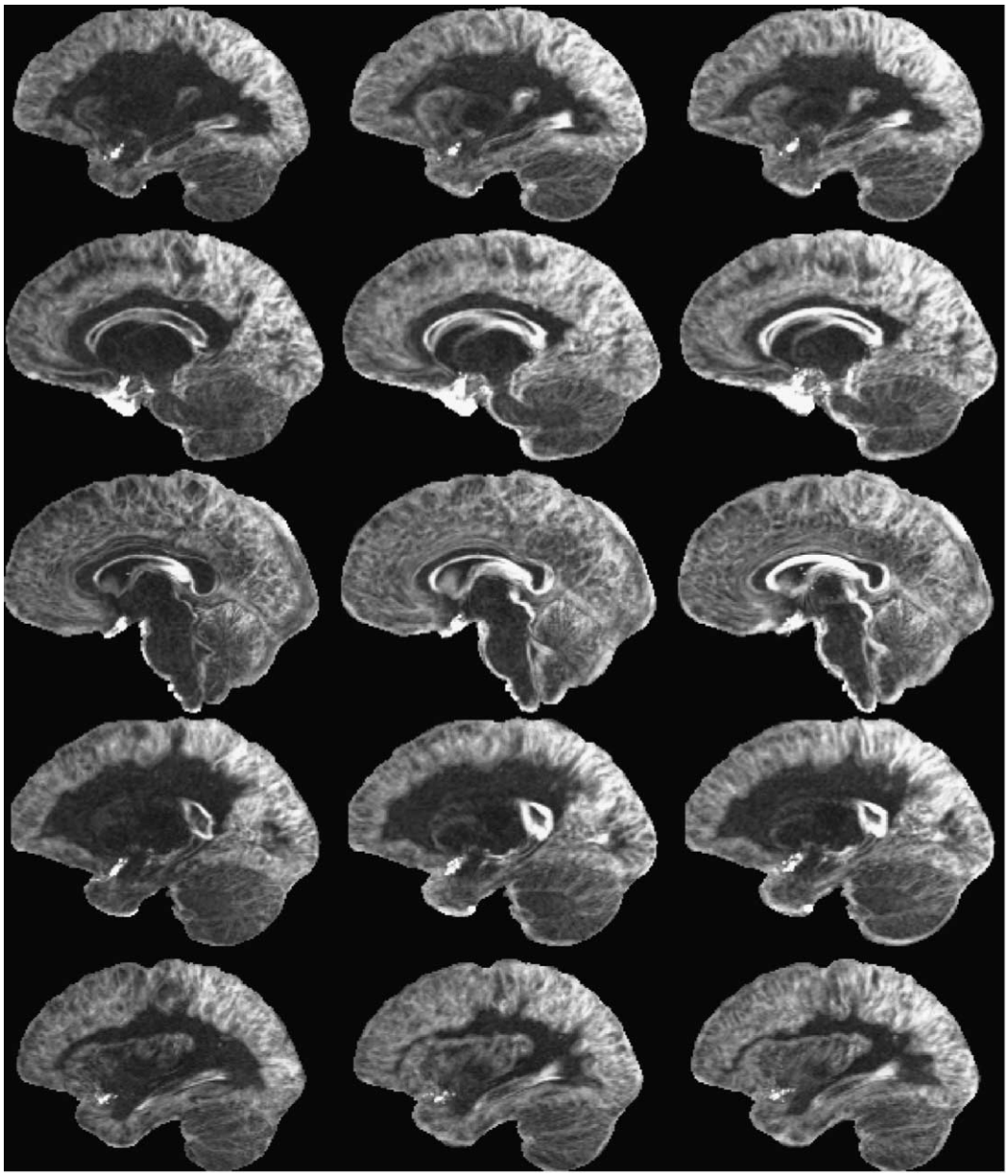

Fig. 2. Voxelwise image standard deviation maps of the slices shown in Fig. 1.

has been our experience that landmarks can be placed more consistently post-registration. However, the drawback is that the number of landmarks that have to be placed manually increase by a factor of M, where M is the number of methods being compared. For example, if we are comparing M methods using Q landmarks on data from N subject, we only need to place $Q \times N$ landmarks if the landmarks are placed before image registration, as compared to having to place $Q \times N \times M$ landmarks after registration. We also applied this second method and found that indeed the sensitivity of the comparison increases dramatically. The resulting dispersions are shown in Table 2. Paired t-tests were applied, yielding values of $t = -8.05$ ($P < 10^{-4}$) for ART versus SPM, and $t = -8.32$ ($P < 10^{-4}$) for ART versus AIR, indicating that the differences between ART and each of the other programs is significant. Comparing AIR and SPM we found $t = 0.45$ ($P \approx 0.66$) indicating no significant difference. It can be seen that placement of landmarks after registration generally reduces the dispersion measures and makes this criterion significantly more sensitive to the differences between registration methods.

Many of the adjustable parameters of the ART algorithm can be provided as optional inputs to the program. In the results presented in this paper, we used the following parameters: (1) the order of the Fourier–Legendre series M was 14; and (2) we used four levels of resolution starting with voxels sizes that were 8, 4, 2, and 1 times the original voxel size. The variance parameter c for the Gaussian kernel in $I(\boldsymbol{r}, c)$

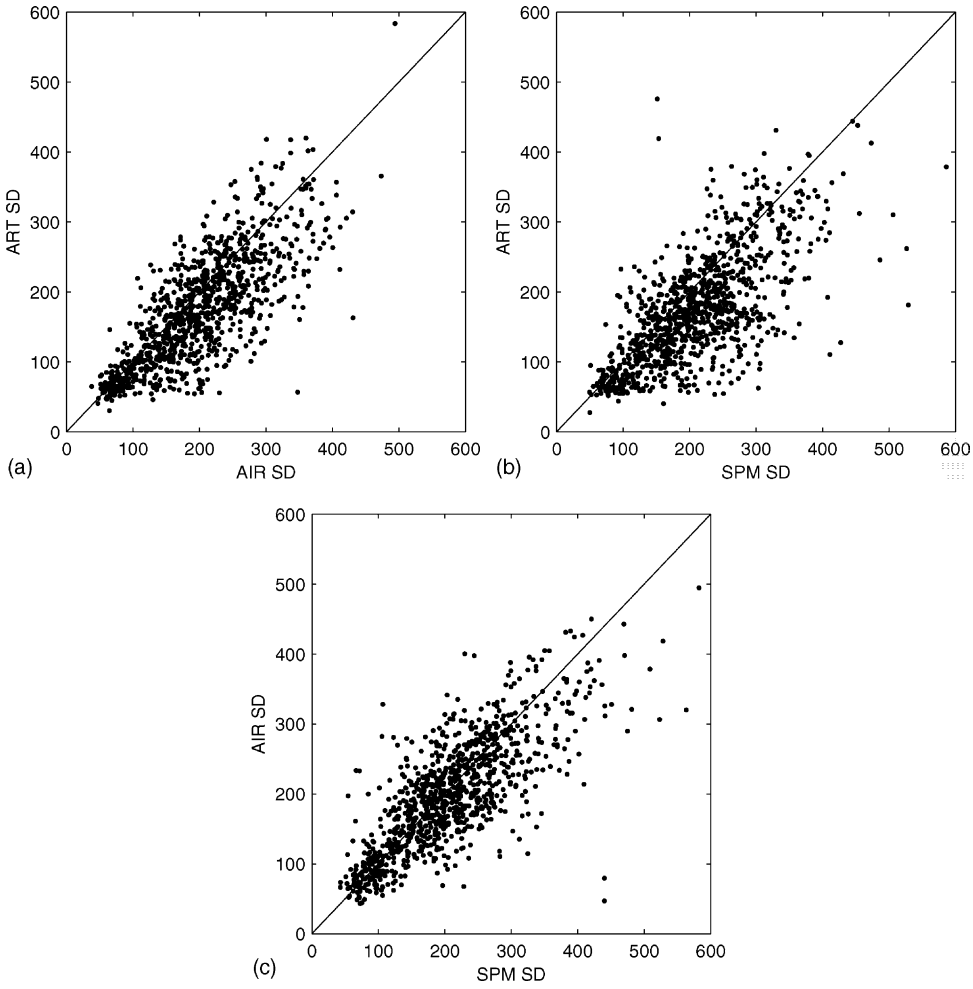

Fig. 3. Scatter plots of pairs of image standard deviation values measured at 1000 randomly selected voxels comparing: (a) ART vs. AIR, (b) ART vs. SPM, and (c) AIR vs. SPM. A greater number of points below the line $y = x$ indicates a greater reduction of inter-subject anatomical variability achieved by the registration method labeled as the ordinate.

$= G_c(\boldsymbol{r}) * I(\boldsymbol{r}, 0)$ was computed using the method described by Ardekani et al. (1997). (3) The feature and search neighborhoods $\Omega_{\boldsymbol{r}}$ and $\psi_{\boldsymbol{r}}$ were both cubic with size 5 voxels × 5 voxels × 5 voxels and remained fixed for all resolution levels. (4) At the highest resolution level ($c = 0$), only 50% of the displacement field voxels were updated, those voxels, for which the L_2 norm of the intensity gradient $\|\nabla I_t(\boldsymbol{r})\|$ was in the upper 50th percentile.

It was noted that the different algorithms completed the registration task at noticeably different speeds. To examine this, a test was run in which one set of data was analyzed on the same machine (800 MHz Pentium III Dell Precision 420 SuSE 7.3 Linux workstation) three times using the different algorithms. SPM99 was most noticeably the fastest, completing the task in 4 min and 22 s. The ART package completed the task in 23 min and 56 s, while the AIR package took 32 min and 49 s to complete the task. Although the processing time for every image in this study was not recorded (due to the variance of processing power on different computers used to process the data), this example is consistent with the observed trend in processing times (with SPM99 finishing significantly faster than the other two packages, and AIR taking the most processing time).

It was also observed that the AIR package has difficulty dealing with situations where the template and subject images have a significant initial misalignment. In such circumstances, the output image produced by AIR was highly distorted while the other two algorithms seemed able to account for the misalignment and functioned normally. In such cases, the AIR algorithm was aided by an initial rigid-body registration of the template and subject images. In addition, the removal of non-brain regions before starting the registration process was critical to the success of the AIR algorithm. Without this preprocessing step, AIR simply failed in most cases. ART and SPM99 were more robust in this respect, but still benefited from the initial brain extraction step using BET (Smith, 2002).

The results of this paper showed that ART is significantly more accurate as compared to SPM99 and AIR in the registration of multiple high-resolution anatomical images to the

Table 2
Post-registration dispersion of landmarks (mm), with landmarks placed after registration

Landmark	Method		
	ART	AIR	SPM99
Frontal Pole—L	1.548	2.062	1.972
Frontal Pole—R	1.686	2.542	2.683
Temporal Pole—L	1.998	3.533	3.442
Temporal Pole—R	2.540	3.621	2.928
Anterior Commissure	0.699	1.332	2.065
Posterior Commissure	0.834	1.426	1.997
Corpus Callosum—A	1.009	1.848	1.963
Corpus Callosum—P	1.233	2.014	2.025
Frontal Horn—L	1.482	1.552	1.584
Frontal Horn—R	1.546	2.042	1.822
Occipital Pole—L	1.369	1.845	1.885
Occipital Pole—R	1.565	2.222	2.393
Putamen—R A	1.714	2.273	2.126
Putamen—R P	1.410	2.394	2.231
Putamen—L A	1.663	2.155	2.044
Putamen—L P	1.291	2.281	2.124
Caudate Nucl.—R A	1.796	1.837	1.971
Caudate Nucl.—L A	1.209	1.692	1.763
Lat. 4th Ven.—R I	0.817	1.425	1.373
Lat. 4th Ven.—L I	1.067	1.317	1.611
Mean	1.424	2.071	2.100
S.D.	0.437	0.628	0.474

Note: L, left; R, right; A, anterior; P, posterior; I, inferior; Lat. 4th Ven., lateral extension of the 4th ventricle.

same template. This result should not be surprising given the fact that ART enjoys a much higher degree of freedom (∼3375 d.f.) as compared to the default SPM99 method (392 d.f.) and the 5th degree polynomial method of AIR (168 d.f.). Thus, the greater ability of ART in reducing anatomical variability between subjects can be attributed to its greater degrees of freedom, since the essence of the algorithms used by the three methods are to a large extent similar. However, the particular non-parametric implementation of ART allows incorporating a much larger number of degrees of freedom while keeping the computational cost manageable. Of course, in SPM99 and AIR, it is possible to use a larger number of basis functions than what is used in this paper. If our above assertion is correct, using a larger number of basis functions in these algorithms should improve the registration results. In fact, in a preliminary study, we used a set of $11 \times 13 \times 11$ basis functions in SPM99 and found that the registration results improved over our current SPM99 method, which uses $7 \times 8 \times 7$ basis functions but did not reach the quality of ART registration. However, the nature of the optimization problem in SPM99 and AIR is such that increasing the number of degrees excessively increases the computation cost.

Another issue that arises is whether it is desirable to make all images "look alike". This is certainly a goal for fMRI and voxelwise diffusion tensor analyses, where analysis of homologous voxels is critical in inferring group differences, but is explicitly not the goal of voxel based morphometry (VBM) (Ashburner and Friston, 2000). In this latter approach, images are matched up to a certain point and then residual differences are taken to represent critical comparisons between groups. Nonetheless, even in these cases the accuracy of registration is critical, as the VBM algorithm entails an interaction between the spatial normalization and voxelwise comparisons (Bookstein, 2001).

Finally, as pointed out in Section 1, an important question is whether the degree of improvement in registration accuracy, as shown in Tables 1 and 2 and Fig. 1, is such that there is an appreciable impact on the final group analysis results of MRI data. Ardekani et al. (2004) addressed this question in the case of fMRI group analysis and showed that increased accuracy of inter-subject registration in removing anatomic variability between subjects results in significant increases in the *sensitivity* of activation detection and the *reproducibility* of activation maps. The results suggest that at least in the visual oddball stimulation paradigm studied in that paper, a significant portion of variability in activation across subjects was due to anatomical variability as opposed to variability in functional anatomy, and this variability was reduced when the quality of inter-subject registration improved.

4. Conclusions

Details of implementation of a non-parametric method (ART) for estimating displacement fields for inter-subject registration of high-resolution volumetric MRI images were presented. The ability of ART in reducing anatomical variability between subjects was compared to the registration accuracy of two other popular registration programs: AIR and SPM99. Homologous landmark sets manually identified on the images before registration were significantly less dispersed in ART after registration, as compared to their dispersion after registration by AIR or SPM99. There was no significant difference between AIR and SPM99 with respect to this measure. It was found that the post-registration spatial dispersion of homologous landmarks measure becomes significantly more sensitive to the differences between algorithms when the landmarks are placed after image registration, as opposed to before. Sets of images registered by ART also showed significantly less voxelwise intensity variance as compared to the other two methods. Overall, results indicate that non-parametric inter-subject registration algorithms with high degrees of freedom are able to reduce inter-subject anatomic variability to a greater extent as compared to lower dimensional parametric methods such as AIR and SPM99, while keeping the computational cost within acceptable limits.

Acknowledgement

This research was supported by Biomedical Engineering Research Grant RG-00-0350 from the Whitaker Foundation to BAA. MJH gratefully acknowledges the support of NIH R01 MH64783 and a NARSAD Young Investigator award.

References

Ardekani BA, Bachman AH, Strother SC, Fujibayashi Y, Yonekura Y. Impact of inter-subject image registration on group analysis of fMRI data. In: Iida H, Shah NJ, Hayashi T, Watabe H, editors. Quantitation in biomedical imaging with PET and MRI. Elsevier International Congress Series, vol. 1265C. 2004. pp. 49–59.

Ardekani BA, Braun M, Hutton BF, Kanno I, Iida H. A fully automatic multimodality image registration algorithm. J Comput Assist Tomogr 1995;19:615–23.

Ardekani BA, Kershaw J, Braun M, Kanno I. Automatic detection of the mid-sagittal plane in 3D brain images. IEEE Trans Med Imaging 1997;16:947–52.

Ardekani BA, Nierenberg J, Hoptman MJ, Javitt DC, Lim KO. MRI study of white matter diffusion anisotropy in schizophrenia. Neuroreport 2003;14:2025–9.

Ashburner J, Friston KJ. Voxel based morphometry—the methods. Hum Brain Mapp 2000;7:254–66.

Bajcsy R, Kovačič S. Multiresolution elastic matching. Comput Vis Graph Im Proc 1989;46:1–21.

Bookstein FL. "Voxel-based morphometry" should not be used with imperfectly registered images. NeuroImage 2001;14:1452–62.

Christensen GE, Johnson HJ. Consistent image registration. IEEE Trans Med Imag 2001;20:568–82.

Christensen GE, Joshi SC, Miller MI. Volumetric transformation of brain anatomy. IEEE Trans Med Imag 1997;16:864–77.

Collins DL, Holmes CJ, Peters TM, Evans AC. Automatic 3D model-based neuroanatomical segmentation. Hum Brain Mapp 1995;3:190–209.

Crivello F, Schormann T, Tzourio-Mazoyer N, Roland PE, Zilles K, Mazoyer BM. Comparison of spatial normalization procedures and their impact on functional maps. Hum Brain Mapp 2002;16:228–50.

Friston KJ, Ashburner J, Poline JB, Frith CD, Heather JD, Frackowiak RSJ. Spatial registration and normalisation of images. Hum Brain Mapp 1995;2:165–89.

Grachev ID, Berdichevsky D, Rauch SL, Heckers S, Kennedy DN, Caviness VS, Alpert NM. A method for assessing the accuracy of intersubject registration of the human brain using anatomic landmarks. Neuroimage 1999;9:250–68.

Holden M, Schnabel JA, Hill DL. Quantification of small cerebral ventricular volume changes in treated growth hormone patients using nonrigid registration. IEEE Trans Med Imaging 2002;21:1292–301.

Jones DK, Griffin LD, Alexander DC, Catani M, Horsfield MA, Howard R, Williams SC. Spatial normalization and averaging of diffusion tensor MRI data sets. Neuroimage 2002;17:592–617.

Kaplan W. Advanced Calculus. Reading, MA: Addison-Wesley; 1973.

Kjems U, Strother SC, Anderson J, Law I, Hansen LK. Enhancing the multivariate signal of [15O] water PET studies with a new nonlinear neuroanatomical registration algorithm. IEEE Trans Med Imag 1999;18:306–19.

Kosugi Y, Sase M, Kuwatani H, Kinoshita N, Momose T, Nishikawa J, Watanabe T. Neural network mapping for nonlinear stereotactic normalization of brain MR images. J Comput Assist Tomogr 1993;17:455–60.

Lindeberg T. Scale-space Theory in Computer Vision. Boston: Kluwer Academic Publishers; 1994.

Marroquin JL, Vemuri BC, Botello S, Calderon F, Fernandez-Bouzas A. An accurate and efficient Bayesian method for automatic segmentation of brain MRI. IEEE Trans Med Imaging 2002;21:934–45.

Rey D, Subsol G, Delingette H, Ayache N. Automatic detection and segmentation of evolving processes in 3D medical images: application to multiple sclerosis. Med Image Anal 2002;6:163–79.

Senda M, Ishii K, Oda K, Sadato N, Kawashima R, Sugiura M, Kanno I, Ardekani B, Minoshima S, Tatsumi I. Influence of ANOVA design and anatomical standardization on statistical mapping for PET activation. Neuroimage 1998;8:283–301.

Smith SM. Fast robust automated brain extraction. Hum Brain Mapp 2002;17:143–55.

Svensen M, Kruggel F, Benali H. ICA of fMRI group study data. Neuroimage 2002;16:551–63.

Unser M. Splines: a perfect fit for signal and image processing. IEEE Sig Proc Mag 1999;16:22–38.

Webb J, Guimond A, Eldridge P, Chadwick D, Meunier J, Thirion JP, Roberts N. Automatic detection of hippocampal atrophy on magnetic resonance images. Magn Reson Imaging 1999;17:1149–61.

Woods RP, Grafton ST, Watson JD, Sicotte NL, Mazziotta JC. Automated image registration. II. Intersubject validation of linear and nonlinear models. J Comput Assist Tomogr 1998;22:153–65.

Zeffiro TA, Eden GF, Woods RP, VanMeter JW. Intersubject analysis of fMRI data using spatial normalization. Adv Exp Med Biol 1997;413:235–40.