NeuroImage **17,** 825–841 (2002)
doi:10.1006/nimg.2002.1132

Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images

Mark Jenkinson,* Peter Bannister,*,† Michael Brady,† and Stephen Smith*

**Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, John Radcliffe Hospital, Headington, Oxford OX3 9DU; and †Medical Vision Laboratory, Department of Engineering Science, University of Oxford, Parks Road, Oxford OXI 3PJ, United Kingdom*

Received September 19, 2001

Linear registration and motion correction are important components of structural and functional brain image analysis. Most modern methods optimize some intensity-based cost function to determine the best registration. To date, little attention has been focused on the optimization method itself, even though the success of most registration methods hinges on the quality of this optimization. This paper examines the optimization process in detail and demonstrates that the commonly used multiresolution local optimization methods can, and do, get trapped in local minima. To address this problem, two approaches are taken: (1) to apodize the cost function and (2) to employ a novel hybrid global–local optimization method. This new optimization method is specifically designed for registering whole brain images. It substantially reduces the likelihood of producing misregistrations due to being trapped by local minima. The increased robustness of the method, compared to other commonly used methods, is demonstrated by a consistency test. In addition, the accuracy of the registration is demonstrated by a series of experiments with motion correction. These motion correction experiments also investigate how the results are affected by different cost functions and interpolation methods. © 2002 Elsevier Science (USA)

Key Words: **accuracy; affine transformation; global optimization; motion correction; multimodal registration; multiresolution search; robustness.**

INTRODUCTION

Geometric registration and motion correction are important stages in the analysis of functional brain imaging studies. Consequently, it is important that these stages perform robustly and accurately. Furthermore, for large imaging studies it is desirable that they be fully automated.

There has been a considerable amount of research into registration and motion correction of brain images, and many different methods have been proposed (Maintz and Viergever, 1998). Most methods in common usage are based on the mathematical framework of optimizing an intensity-based cost function. However, although much work has concentrated on how the choice of cost function affects registration performance, there has been far less examination of the effect of the optimization method. Moreover, when optimization methods are discussed, global methods are often ignored and local methods compared purely on the basis of speed (Maes *et al.*, 1999).

One of the most common and serious problems for registration methods is the presence of local minima in the cost function. These cause local optimization methods to "get stuck" and hence to fail to find the desired the global minimum. Most registration methods attempt to solve this problem by incorporating a local optimization strategy within a multiresolution framework. Such a multiresolution framework, which typically involves starting with low-resolution images (containing only gross features) and working progressively through to higher resolutions, aims to avoid the local minima "traps." As we show later, this simple multiresolution approach is not always sufficient for avoiding local minima, and by using more sophisticated optimization methods, the chances of becoming "trapped" in these local minima can be substantially reduced.

Two types of local minima commonly occur for the cost functions used in image registration: large-scale basins and small-scale dips. The first type, the large-scale basin, is responsible for large misregistrations since the local minimum is often far from the global minimum. The second type, small-scale dips, can cause the optimization to get stuck at any stage and so are responsible for large misregistrations at low resolutions and small misregistrations at high resolutions.

We propose two methods for dealing with the local minima problem. These are cost function apodization, which reduces or eliminates small-scale dips, and a hybrid global–local optimization technique which utilizes prior knowledge about brain registration to create an optimization technique that combines the speed of local optimization with the robustness of global optimization.

1053-8119/02 $35.00

The following sections of this paper are background theory, methods (including both cost function apodization and the hybrid optimization method), results, and discussion. The results section contains a number of experiments on real, whole brain images which demonstrate the effectiveness of the registration in two different settings: (1) structural image registration (intermodal/intersubject) of an anatomical image to a standard template; and (2) functional image motion correction (intramodal/intrasubject) which registers each image in a time-series to a particular example image from that time-series. The first case is examined using a robustness study (as accuracy is hard to define for intersubject registration, and robustness is a more important issue in this context), while the second case is examined using an accuracy study (as, in this context, it is accuracy that is more important). In each case real brain image data are used. Comparisons with some commonly used methods are also included (in both cases) which demonstrate the superior robustness and accuracy which can be obtained using this approach.

MATERIALS

Registration

The registration problem studied here is to find the best geometric alignment of two (volumetric brain) images. Call the two images the reference (Y) and floating (X) images. More precisely, the registration problem seeks that transformation which, when applied to the floating image, maximizes the "similarity" between this transformed floating image and the reference image.

A standard, and common, way of formulating this as a mathematical problem is to construct a cost function which quantifies the dissimilarity between two images and then search for the transformation (T^*) which gives the minimum cost. In mathematical notation this is

$$T^* = \arg \min_{T \in S_T} C(Y, T(X)), \quad (1)$$

where S_T is the space of allowable transformations, $C(I_1, I_2)$ is the cost function, and $T(X)$ represents the image X after it has been transformed by the transformation T. In this paper we shall only consider linear registration so that S_T is either the set of all affine transformations or some subset of this (such as the set of all rigid-body transformations).

Cost Function

Many different cost functions have been proposed for image registration problems. Some use geometrically defined features, found within the image, to quantify the (dis)similarity, while others work directly with the intensity values in the images. A large comparative study of different registration methods (West *et al.*, 1997) indicated that intensity-based cost functions are more accurate and reliable than the geometrically based ones. Consequently, most recent registration methods have used intensity-based cost functions, and these are the ones which will be discussed in this paper.

Intensity-based cost functions can be divided naturally into two categories: those suitable for intramodal problems and those suitable for intermodal problems. In the former category the most commonly used cost functions are *least squares* (LS) and *normalized correlation* (NC). For the latter, and more difficult, category the most commonly used functions are *mutual information* (MI), *normalized mutual information* (NMI), *woods* (W), and *correlation ratio* (CR). These functions are defined mathematically in Table 1 (see (Jenkinson and Smith, 2001) for more information).

Interpolation. In addition to a pair of images and a

TABLE 1

Mathematical Definitions of the Most Commonly Used Intensity-Based Cost Functions: least squares (LS); normalized correlation (NC); Woods (W); correlation ratio (CR); mutual information (MI); and normalized mutual information (NMI)

Cost function	Definition	Minimum	Maximum
C^{LS}	$\Sigma(Y - X)^2$	0	∞
C^{NC}	$\frac{\Sigma(X \cdot Y)}{\sqrt{\Sigma X^2}\sqrt{\Sigma Y^2}}$	-1	1
C^{W}	$\Sigma_k \frac{n_k}{N} \frac{\sqrt{\mathrm{Var}(Y_k)}}{\mu(Y_k)}$	0	∞
C^{CR}	$\frac{1}{Var(Y)} \Sigma_k \frac{n_k}{N} \mathrm{Var}(Y_k)$	0	1
C^{MI}	$H(X, Y) - H(X) - H(Y)$	$-\infty$	0
C^{NMI}	$\frac{H(X,Y)}{H(X)+H(Y)}$	0	1

Note. The notation is as follows: quantities X and Y denote images, each represented as a set of intensities; $\mu(A)$ is the mean of set A; $\mathrm{Var}(A)$ is the variance of the set A; Y_k is the kth iso-set defined as the set of intensities in image Y at positions where the intensity in X is in the kth intensity bin; n_k is the number of elements in the set Y_k such that $N = \Sigma_k n_k$; $H(X,Y) = -\Sigma_{ij} p_{ij} \log p_{ij}$ is the standard entropy definition where p_{ij} represents the probability estimated using the (i, j) joint histogram bin, and similarly for the marginals, $H(X)$ and $H(Y)$. Note that the sums in the first two rows are taken over all corresponding voxels.

particular transformation, the cost function requires that a method of interpolation be defined, that is, some method of calculating what the intensity is in the floating image at points in between the original voxel (or grid) locations. This is necessary in order to know the intensity at corresponding points in the images after the geometrical transformation has been applied to the floating image.

Interpolation methods that are commonly used are trilinear (also called linear or, in 2D, bilinear), nearest neighbor, sinc (of various kernel sizes and with or without various windowing functions; e.g., Blackman), spline, and Fourier. The choice of method has some impact on cost function smoothness, although all interpolation methods except nearest neighbor are continuous. However, the choice of method becomes most critical for motion correction as the transformed image intensities are needed for later statistical analysis.

Optimization

Once a cost function has been chosen it is necessary to search for the transformation which will yield the minimum cost value. To do this, an optimization method is used which searches through the parameter space of allowable transformations. Note that rigid-body transformations are specified by 6 parameters (3 rotations and 3 translations) while affine transformations are specified by 12 parameters. Consequently, even for linear transformation, the optimization takes place in a high dimensional space; R^n, where $6 \leq n \leq 12$.

While the problem specified in Eq. (1) is a global optimization, quite often local optimization methods are employed as they are simpler and faster.[1] However, this can result in the method returning a transformation that corresponds to a local minimum of the cost function, rather than the desired global minimum. Such cases often appear as misregistrations, of varying severity, and are a major cause of registration failure.

Unfortunately, there are very few global optimization methods that are suitable for a 3D brain image registration problem. This is because, in terms of operations, the cost function is expensive to evaluate and most global optimization methods require a great many evaluations leading to unacceptable execution times (e.g., days).

Multiresolution techniques. To both speed up the optimization process and avoid local minima, most currently used registration methods employ some form of multiresolution optimization. That is, a sequence of image pairs, at progressively larger spatial scales, is created from the initial pair of images: (I^r, I^f). The images at larger scales are subsampled versions (often with preblurring) of the original high-resolution images and so contain fewer voxels which means that evaluating the cost function requires less computation. In addition, as only gross features of the images remain at these large scales, it is hoped that there will be fewer local minima for the optimization to get stuck in.

Motion Correction

In functional brain imaging a series of brain images is acquired. The time elapsed between each acquisition is usually a few seconds or less. Due to the small acquisition times required, these images usually have poor resolution. Furthermore, as the imaging parameters are tuned to highlight physiological changes (e.g., blood oxygenation), the images often have poor anatomical contrast.

Extracting functional information from such a series of images is done by applying statistical time-series analysis, which assumes that the location of a given voxel within the brain *does not change over time*. However, there is usually some degree of subject motion within the scanner, especially when the scanning takes a long time or when clinical patients are involved. Therefore, in order to render the data fit for statistical analysis this motion must be estimated and corrected for. This is the task of motion correction methods and it is essentially a multiple-image registration task.

Normally, motion correction methods deal with the registration task by selecting a reference image from within the series and registering each image in turn to this fixed reference. As all images are of the same subject, using the same imaging parameters, it can be classified as an intrasubject, intramodal registration problem. Therefore, a rigid-body transformation space and intramodal cost function can be used. Furthermore, as the values in the corrected images are important for later statistical analysis, the choice of interpolation method for the transformation of the images is of particular importance (Hajnal *et al.,* 1995a,b).

METHODS

Apodization of the Cost Functions

As seen in Fig. 1, the local behavior of the cost function shows small discontinuities as the transformation parameters are varied smoothly. This creates local minima traps for the optimization method. Since all interpolation methods are continuous (except nearest neighbor, which is consequently seldom used) the discontinuities are not due to the type of interpolation used. The cause of these discontinuities is the changing amount of overlap of the reference and floating image.

[1] A *global* method searches the entire range of possible parameters for the most optimal cost function value, while *a local* method simply starts somewhere in parameter space and moves about locally trying to find an optimum, stopping when there is no better nearby set of parameters.

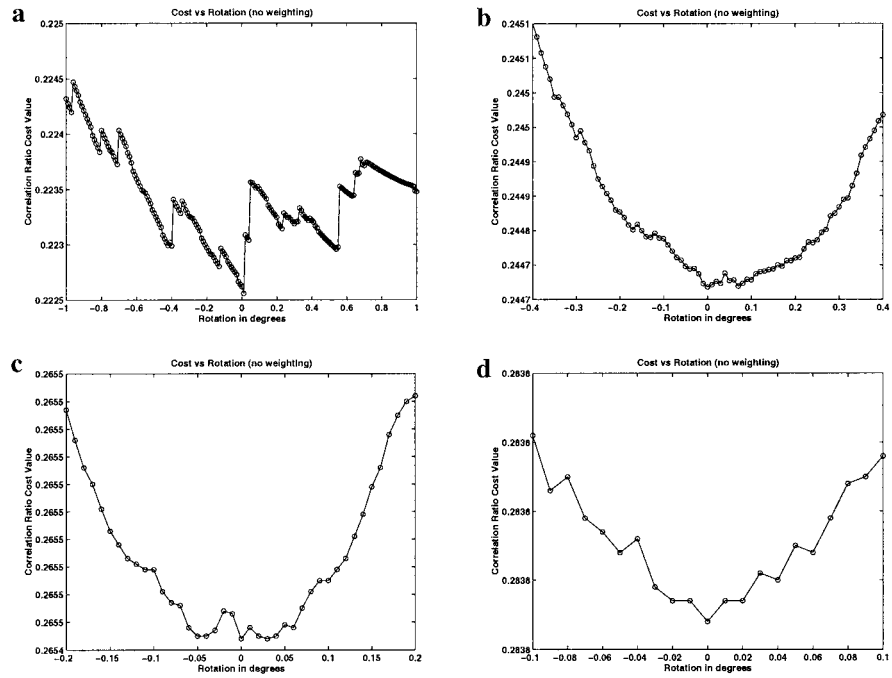

FIG. 1. Illustration of local discontinuities in cost function plots.

There are two different ways of treating values outside the field of view (FOV) of an image:

1. Treat all values outside the FOV as zero.
2. Do all calculations strictly in the overlapping region.

The first method is undesirable as it creates artificial intensity boundaries when the object is not wholly contained within the FOV. However, in the second method the number of points counted in the overlapping region varies, not just the expressions involving intensities. Therefore, in the second case both the numerator and the denominator of the cost functions (except least squares) will change discontinuously as the amount of overlap changes.

The discontinuities exist because the images are discrete sets of voxels. In particular, the reference image defines a fixed set of voxel locations over which the cost function is calculated. Then, for a given transformation, the floating image intensities at these locations are calculated using interpolation. A reference image voxel location is only counted when it is valid, that is, within the overlapping region such that it maps to a location inside the FOV of the floating image. When the edge of the FOV of the floating image crosses a reference voxel location, the location suddenly goes from being inside the overlapping region to outside, causing a discontinuous change in the number of valid locations, as shown in Fig. 2.

We aim to apodize the cost function by removing these discontinuities. To do this, our approach has been to introduce a geometric apodization that deweights the contributions of locations that are near the edge of the overlapping region. The weighting is chosen so that the contribution of such locations drops continuously until it reaches zero at the edge of the overlapping region. Any continuous weighting function could be used but for simplicity and computational efficiency we choose one that is linear.

For instance, consider a 2D example of a reference location that maps to a point inside the overlapping region, where the distance from the nearest edges of the floating image FOV is d_X and d_Y units, as shown in Fig. 2. In each dimension, if this value is less than some threshold D, then the influence of that point is weighted by a weight $w = d/D$. In higher dimensions, the product of the weighting functions in each dimension is used. That is, $w(d_X, d_Y, d_Z) = w(d_X)w(d_Y)w(d_Z)$.

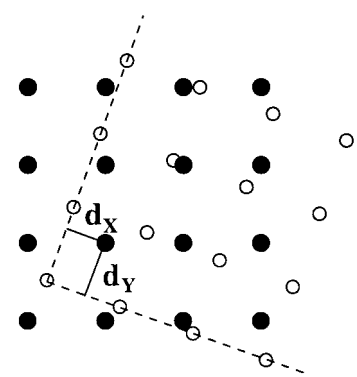

FIG. 2. FOV change.

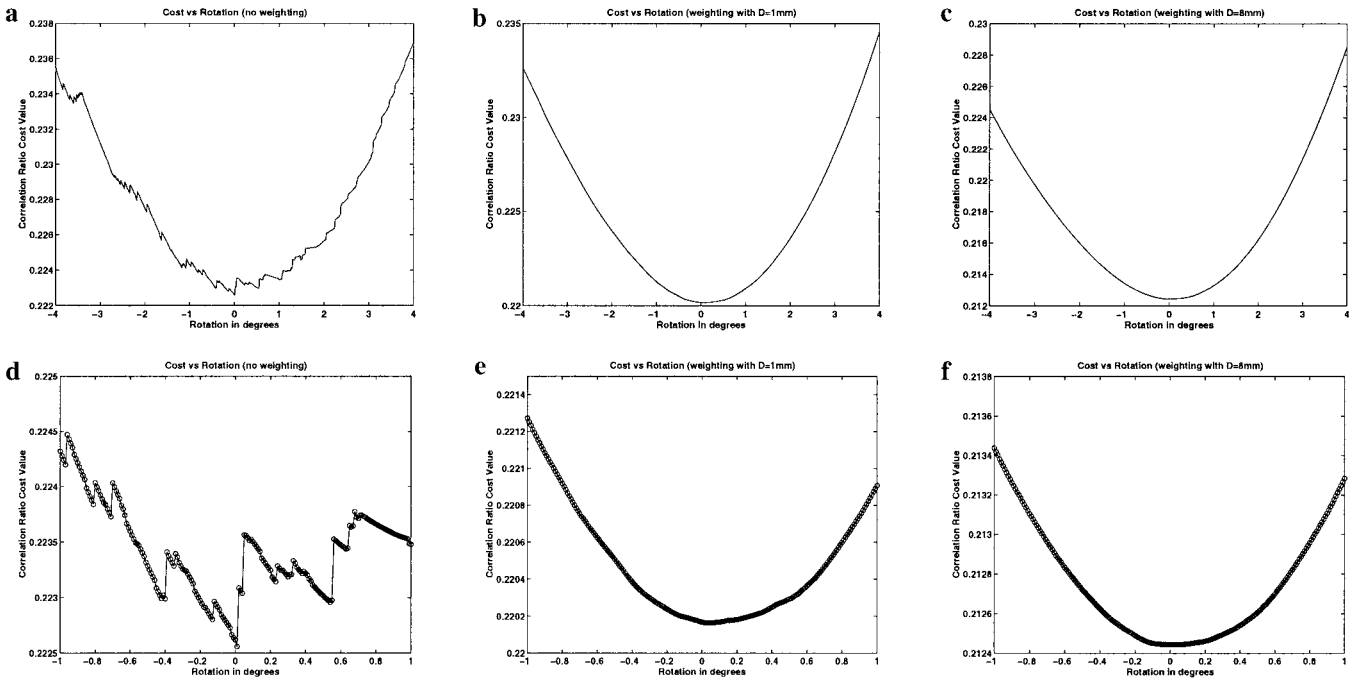

FIG. 3. Illustration of cost function apodization.

The weighting is applied to all terms involving that location's intensity as well as to the number of locations in the region. For example, consider the nth moment of an iso-set:

$$M_n(Y_k) = \frac{1}{N_k} \sum_{j|X_j \epsilon I_k} (Y_j)^n \tag{2}$$

$$N_k = \sum_{j|X_j \epsilon I_k} 1, \tag{3}$$

where M_n is the nth moment, j is a voxel index, X and Y represent the reference and floating images, respectively, and I_k denotes the kth intensity bin. With general weighting this becomes

$$M_n(Y_k) = \frac{1}{N_k} \sum_{j|X_j \epsilon I_k} w_j (Y_j)^n \tag{4}$$

$$N_k = \sum_{j|X_j \epsilon I_k} w_j, \tag{5}$$

where w_j is the weight of the location j, which is 0 outside the overlapping region, d/D for $d < D$ or 1 for $d > D$ inside the overlapping region.

This weighting scheme can be simply and efficiently applied to any of the non-entropy-based cost functions (i.e., LS, NC, W and CR). It depends on one parameter—the threshold distance D— which can be varied to increase the amount of apodization. When $D = 0$ there is no apodization, while increasing D creates smoother and smoother cost functions, although the cost function will be continuous for *any* nonzero value of D. Also note that making D larger than the voxel spacing is permitted and just has a greater smoothing effect, as shown in Fig. 3.

Joint histogram apodization. A more general weighting scheme is required for apodizing the joint histogram required for the entropy-based cost functions. This is because the number of entries in each histogram bin becomes discontinuous as the intensity at a floating image location (calculated using interpolation) passes through the threshold value between intensity bins.

As it is the intensity passing through a threshold value that causes the discontinuities for the joint histogram, we propose a weighting function that is determined by the intensities and applied to every location. We choose, once again, a linear weighting function (as shown in Fig. 4) where w_k is the weight for bin k, I is

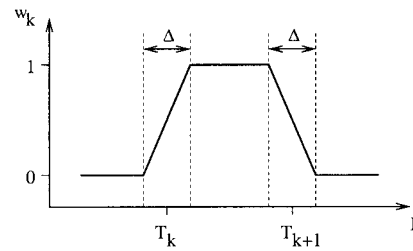

FIG. 4. Weighting function.

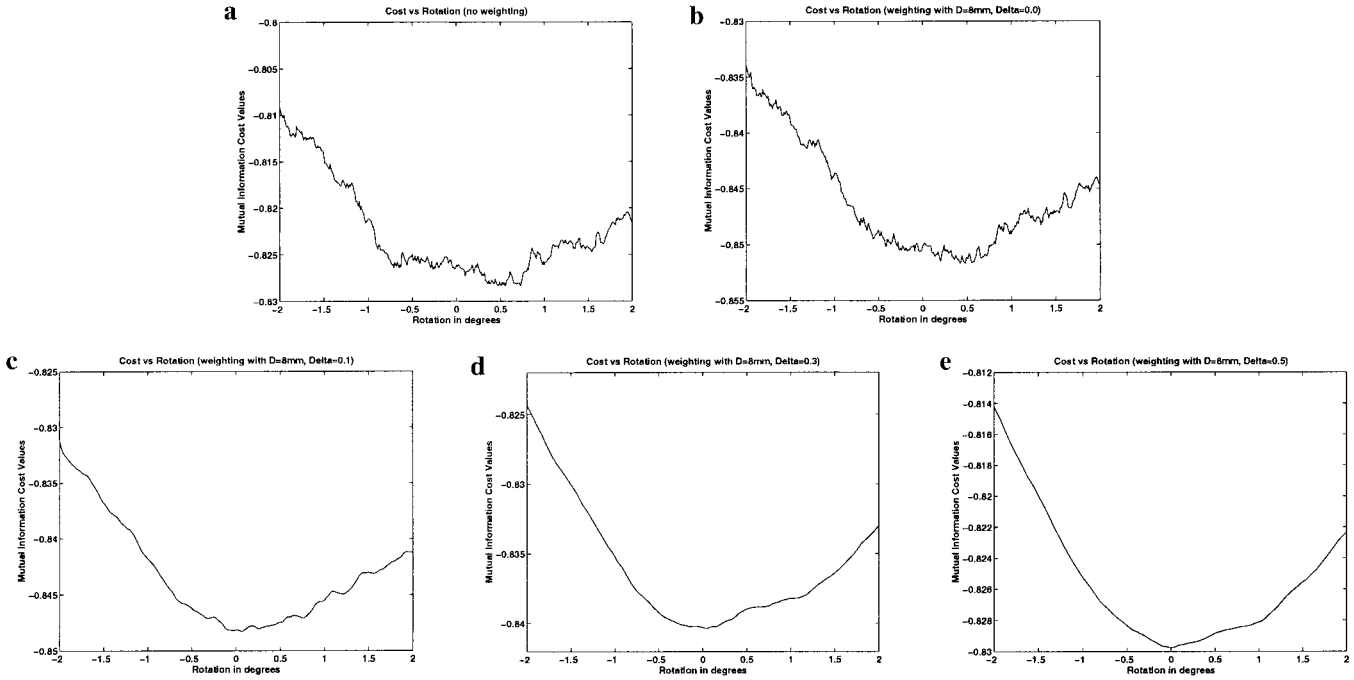

FIG. 5. Illustration of cost function apodization.

the intensity at the location under consideration, T_k is the intensity threshold between bins k-1 and k, and Δ is the smoothing threshold—the equivalent of D in the preceding section. This weight is then applied to the accumulation of intensity within the joint histogram bin, as well as the number of entries in the bin, which is no longer an integer. That is, for a point of intensity I, the updating equations for bin k are

$$N_k \rightarrow N_k + w_k(I)$$

$$S_k \rightarrow S_k + I \cdot w_k(I)\,,$$

where $w_k(\cdot)$ is the weighting function for the kth bin (see Fig. 4), N_k is the occupancy of bin k (a noninteger version of the number of elements), and S_k is the sum of intensities in the bin.

This approach is effectively fuzzy-binning, where each intensity bin no longer has sharp thresholds, but fuzzy membership functions. It also means that a given location can influence more than one bin entry. Because of the way the weighting function is calculated though, each location contributes equally, since the sum of weights for all bin entries is equal to 1.

As changing overlap will still create discontinuities, both the geometrical weighting and the fuzzy-binning must be applied to have a continuous joint histogram. Moreover, the parameter Δ will give a continuous joint histogram for any value greater than zero, although the value should not exceed the intensity bin width. The smoothing capacity of Δ is shown in Fig. 5 for the mutual information cost function. Results for the normalized mutual information cost function are very similar. Note that, in general, a value of $\Delta = 0.5$ together with D equal to the resolution scale (e.g., 8 mm) gives a desirably smooth cost function.

Note that the partial volume interpolation introduced by Maes (Maes *et al.*, 1997) also creates continuous joint histogram estimates if used in conjunction with the geometrical weighting (applied to the reference locations instead). However, as the name suggests, PVI is more than just an apodization scheme—in fact, it functions as an interpolation method too. Therefore, different interpolation methods cannot be used in conjunction with PVI, whereas for fuzzy-binning the interpolation method used can be freely chosen. Furthermore, the fuzzy-binning scheme provides an adjustable parameter, Δ, which controls the amount of smoothing of the cost function, allowing for different degrees of smoothing as desired.

Finally, it can be seen that fuzzy-binning can be made fully symmetric with respect to the two images (see (Cachier and Rey, 2000) for a discussion of symmetry in general registration cost functions). That is, both floating and reference intensities can have fuzzy-binning applied to them. However, there is an inherent asymmetry in the way that interpolation is applied only to the floating image. Therefore, although such a symmetric approach appears initially attractive, the

simpler and faster approach of only using fuzzy bins for the floating image was adopted in practice.

A Global–Local Hybrid Optimization Method

Of the many different approaches to global optimization, we have investigated two strategies and combined them with a simple but fast local optimization method to produce a hybrid optimization method. The two strategies are searching and multistart optimization.

Our hybrid optimization method (also described in (Jenkinson and Smith, 2001)) is specifically designed for the problem at hand, using prior knowledge about the transformation parameters and typical data size (FOV, voxel size, etc.) to help make the method efficient. The method cannot guarantee that the global solution is found, but then neither can any other global optimization method given a finite amount of time. Generally, only statistical "guarantees" are given, and these often require excessive run-times in order to be met. In contrast, our method is designed to give a reliable estimate of the global minimum given some time restriction (in our case, less than 1 h on a moderately powered standard workstation; e.g., registering two $1 \times 1 \times 1$ mm images typically takes 15 min on a 500-MHz Pentium III).

The method still uses a local optimization method with a multiresolution framework, and these are described in the next two sections, followed by descriptions of the global search and multistart optimization strategies employed.

Multiresolution. Currently, four different scales are used in our method: 8, 4, 2, and 1 mm. At each scale, the two images are resampled, after initial preblurring (using a Gaussian with FWHM equal to the ratio of the final and initial voxel sizes), so that they have isotropic voxels of size equal to the scale size. Note that an exception to this occurs if the scale is smaller than the data resolution, in which case the data are resampled to isotropic voxels of scale closest to the data resolution.

Furthermore, skew and anisotropic scaling changes are much less prominent than rotational, translational, and global scaling changes and so their effects are difficult to estimate reliably at low resolutions. Consequently, only similarity transformations (rigid-body + global scaling) are estimated at the 8- and 4-mm scales.

Local optimization. The choice of local optimization method used here is not critical, except that it must be efficient. Furthermore, since it will be used in a multiresolution framework, the low-resolution stages do not need to find highly accurate transformations. Therefore, the initial parameter bracketing and the parameter tolerances (the size of uncertainty on the optimized parameter values) are both made proportional to the scale size. This avoids many unnecessary cost function evaluations at low resolutions.

We initially chose Powell's method (Press *et al.*, 1995) as our local optimization method as it was efficient and did not require gradients to be calculated which are especially difficult given the apodizations applied to the cost functions. However, we discovered that a set of N 1D golden searches (Press *et al.*, 1995) gave equally good results, which can be reasonably expected if the parameterization is close to decoupled.

Global search. To estimate the final transformation sufficiently accurately, a brute-force search of the transformation space is infeasible, even for rigid-body transformations. However, at the lowest resolution (8-mm scale) only the gross image features still exist and so a coarse search of the cost function at this resolution should reflect the major changes in rotation, translation, and global scaling, allowing large misregistrations to be avoided.

Speed remains an issue, even for coarse searches at low resolution. Therefore, the search is restricted to the rotation parameters, as these are the most difficult to find and are the cause of many large misregistrations. Furthermore, the search is divided into three stages:

1. a coarse search over the rotation parameters with a full local optimization of translation and global scale forn each rotation tried;
2. a finer search over rotation parameters, but with only a single cost function evaluation at each rotation (for efficiency);
3. a full local optimization (rotation, translation, and global scale) for each local minimum detected from the previous stage.

The first of these stages is straightforward. Given a set of rotations to try (by default we use 60° increments in each of the Euler angles, leading to $6^3 = 216$ different rotations), the local optimization routine is called for the translation and global scale only. That is, the rotation is left fixed, and the best translation and global scale for this particular rotation is found. These results are then stored for use in the later stages.

The second stage takes a larger set of rotations (by default we use 18° degree increments, leading to $20^3 = 8000$ different rotations) but only evaluates the cost function once for each rotation. This contrasts with the previous stage where the cost function is typically evaluated between 10 and 30 times during the local optimization. However, in order for the evaluation to be a reasonable estimate of the best cost function with this rotation, the translation and global scale parameters must be close to the optimal values. These parameter values are supplied from the results of the previous stage, with the translation parameters determined by interpolating between the stored translation values. Global scale is fixed at the median global scale value over all the stored values. This is done differently from

the translation as the scale should not vary greatly with rotation, whereas the translation is highly coupled with the rotation values.

Finally, the third stage applies a full local optimization, allowing rotation to vary, at each local minimum detected from the results of the previous stage. These local minima are defined as rotations where the cost value found is less than for any of the "neighboring" rotation values. There are often several such local minima, and rather than force the selection of the best one at this stage, they are all optimized and passed onto the higher resolution stages.

Although it is unlikely that the first stage in this process will get very close to the correct rotation, the second stage should get close enough for the local optimization in the last stage to give a good estimate. Note that in most registration methods there is no equivalent of this search and a single local optimization is performed with the starting point being no rotation, no translation, and unity scaling (the identity transformation). As this method examines many more possible starting transformations, one cannot do any worse than these simple methods.

Multistart optimization with perturbations. Following the previous search stage (at 8-mm scale) there are usually several local minima selected as candidates for initializing more detailed searches for the global minimum. This stage (at 4-mm scale) performs a local optimization for the best of these candidate transformations. In addition, it takes several perturbations of the candidate transformations and performs local optimization of these perturbations. Finally, the single best (minimum cost) solution is selected from among these optimization results.

In practice, the three best candidates are taken from the previous stage, together with 10 perturbations of each candidate. Two perturbations are applied to each rotation parameter, each being half the magnitude of the fine search step size from the previous stage. As well as these 6 rotational perturbations, 4 perturbations in scale (± 0.1, ± 0.2) are also applied. The number and size of the perturbations used are arbitrary, with these values chosen here being selected largely from experience with the magnitude and type of typical misregistrations.

This approach of trying several candidate solutions is effectively a multistart strategy similar to that used in genetic algorithms and other global optimization methods. Furthermore, the use of "local" perturbations is similar to the way in which alternatives are generated in simulated annealing. It has been found empirically that the combination of these strategies, together with the initial search, avoids getting trapped in local minima to a much greater extent than for a local optimization method alone within a multiresolution framework.

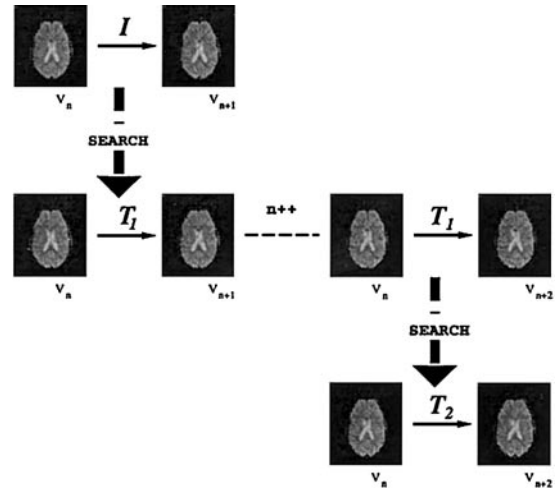

FIG. 6. MCFLIRT schedule.

Higher resolution stages. Following the 4-mm scale the single best candidate transformation is chosen, and it is only this transformation which is worked with from here on. At the 2-mm scale the skews and anisotropic scalings start to become significant. Consequently, these extra degrees of freedom (DOF) are progressively introduced by calling the local optimization method three times: first using only 7 DOF (rigid-body + global scale), then with 9 DOF (rigid-body + independent scalings), then with the full 12 DOF (rigid-body + scales + skews).

Since the cost function evaluations take 8 times longer at the 1-mm scale than at the 2-mm scale and 512 times longer than at the 8-mm scale, only a single pass of the local optimization is done at the 1-mm scale. The result of this single pass is returned as the registration solution, T^*.

Motion Correction

In broad terms, a motion correction algorithm must take a time series of fMRI images and register each image in the series to a reference image. This reference image may be of a different modality (Biswal and Hyde, 1997) but a more common approach is to select one image from the time-series itself (usually the first—c.f., SPM (Friston *et al.*, 1996)) and register the remaining images to this template image.

If we make the reasonable assumption that there is unlikely to be large motion from one image to the next (usually 3 s or less between images), we can use the result of one image's registration as an initial guess for the next image in the series. This is accomplished by assuming an initial identity transformation between the middle image V_n in a time-series and the next adjacent image V_{n+1} and then finding the optimal transformation T_1 by optimizing the cost function. The resulting solution is then used as a starting point for

the next optimization with the next image pair V_n, V_{n+2} (see Fig. 6). This is only done at the lowest resolution, as all higher resolutions use the transformations found at the next lower resolution for the initial estimates.

The final schedule carries out the following steps on the uncorrected data (optional stages are shown in italics):

- 8-mm optimization using the middle image as initial reference and then using each result to initialize the next pairwise optimization;
- 4-mm optimization using the middle image as reference and 8-mm stage results as initialization parameters;
- 4-mm optimization (lower tolerance) using the middle image as reference and 4-mm stage results as initialization parameters;
- *mean registration option*:

  Apply transformation parameters from high-tolerance 4-mm stage;

  Average corrected images to generate mean template image;

  Carry out 8-, 4-, and 4-mm (high-tolerance) optimizations as before but against mean image as reference;
- *since registration option*:

  Carry out additional 4-mm (high-tolerance) optimization using sinc interpolation (instead of trilinear as used in previous stages);
- apply current transformation parameters to uncorrected data and save.

End Slices

As the intensity values are of great interest after motion correction, attention must be paid to not only the estimation but also the application of the transformation. Interpolation probably has the largest impact on the quality of the transformed data, with sinc interpolation methods often being used, although no absolute consensus on the best method exists. However, the loss of information outside the FOV, usually seen in the end slices, can also be very detrimental to the final statistical maps in these areas.

Our motion correction implementation has also been designed to handle the potentially problematic issue of end-slice interpolation. It is frequently the case that under even small affine motion of the head, voxels in the top and bottom slices can move either in or out of the field of view (see Fig. 7). Other schemes approach this by assuming that all affected voxels are either zero (AIR) or can be completely excluded from further calculations (SPM). This clearly impacts later analysis as valuable spatial information may be lost.

We counter this situation by padding the end-slices when applying the estimated transformation (i.e, increasing the extent of each volume by two slices). This means that if data are to be interpolated from outside the FOV, it will take on "sensible" values (personal communication, Roger Woods, 1999).

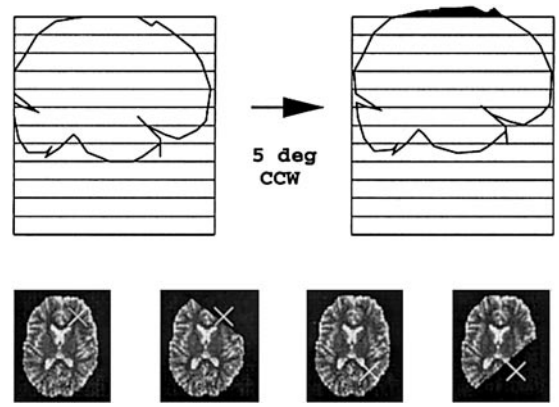

FIG. 7. End-slice correction.

RESULTS

This section presents several experiments that demonstrate the robustness and accuracy of the proposed registration and motion correction method. We begin by first stating the implementational choices made as these are often critical in creating a stable method that performs well. Following this, we present the experiments for registration which clearly demonstrate the improved robustness, and the following sections discuss motion correction and atrophy estimation, demonstrating the improved accuracy.

Implementation: FLIRT and MCFLIRT

The registration and motion correction methods described in the previous sections have been implemented in C++ and are called FLIRT (FMRIB's[2] linear image registration tool) and MCFLIRT (motion correction FLIRT). In each case several implementation choices needed to be made to obtain a robust, working method. The more important choices are: (1) the use of center of mass as the center of transformation (also used for initial alignment); (2) the parameterization of the transformations as three Euler angles, three translations, three scales, and three skews; and (3) the number of intensity histogram bins set to 256 *divided by the scale size* (i.e., 256 for 1-mm scaling but only 32 for 8-mm scaling) since the number of voxels (samples) is small for large scalings and so fewer bins must be used in order to get reliable statistics (Izenman, 1991). Each of these choices is detailed more fully in Jenkinson and Smith (2001).

Robustness Assessment: Registration

Consistency test. For many registration problems in practice, there is no ground truth available with

[2] The Oxford Centre for Functional MRI of the Brain.

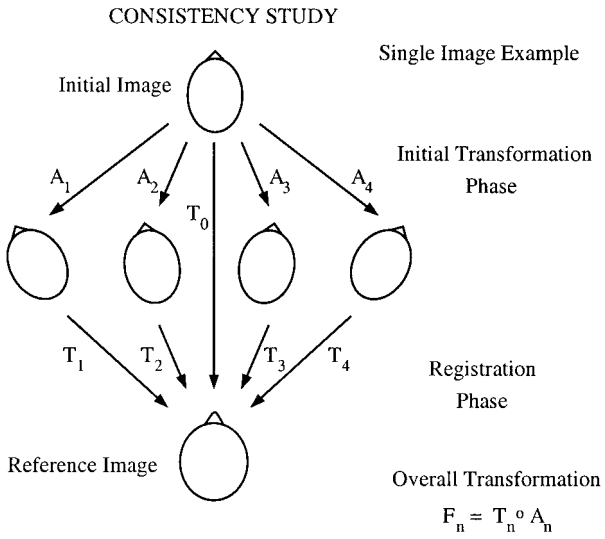

FIG. 8. Illustration of the consistency test.

which to evaluate the registration. This makes the quantitative assessment of methods quite difficult. Therefore, to test the method quantitatively, a comparative consistency test was performed that does not require knowledge of the actual ground truth.

The consistency test is based on comparing registrations obtained using various different, but known, initial starting positions of a given image. If the registrations are consistent then the final registered image will be the same, regardless of the starting position. Consistency is a *necessary, but not sufficient* condition that all correctly functioning registration methods must possess. This is essentially a measure of the robustness rather than the accuracy (West *et al.*, 1997) of the registration method. Robustness is defined here as the ability to get close to the global minimum on all trials, whereas accuracy is the ability to precisely locate a (possibly local) minimum of the cost function. Ideally, a registration method should be both robust and accurate.

More specifically, the consistency test for an individual image I involved taking the image and applying several predetermined affine transformations, A_j to it (with appropriate cropping so that no "padding" of the images was required). All these images (both transformed and untransformed) were registered to a given reference image, I^r, giving transformations T_j. If the method was consistent the composite transformations $T_j \circ A_j$ should all be the same, as illustrated in Fig. 8.

The transformations are compared quantitatively using the RMS deviation between the composite registration $T_j \circ A_j$ and the registration from the untransformed case T_0. This RMS deviation is calculated directly from the affine matrices (Jenkinson, 1999). That is,

$$d_{\mathrm{RMS}} = \sqrt{\frac{1}{5}R^2 \mathrm{Tr}(M^\top M) + t^\top t}, \qquad (6)$$

where d_{RMS} is the RMS deviation in mm, R is a radius specifying the volume of interest, and $\begin{pmatrix} M & t \\ 0 & 0 \end{pmatrix} = T_j \cdot A_j \cdot T_0^{-1}$ is used to calculate the 3×3 matrix M and the 3×1 vector t.

Comparison with existing methods. A comparison of FLIRT with several other registration packages was initially performed using the consistency test explained above. The other registration packages used were AIR (Woods *et al.*, 1993), SPM (Friston *et al.*, 1995), UMDS (Studholme *et al.*, 1996), and MRI-TOTAL (Collins *et al.*, 1994). These methods were chosen because the authors' implementations were avail-

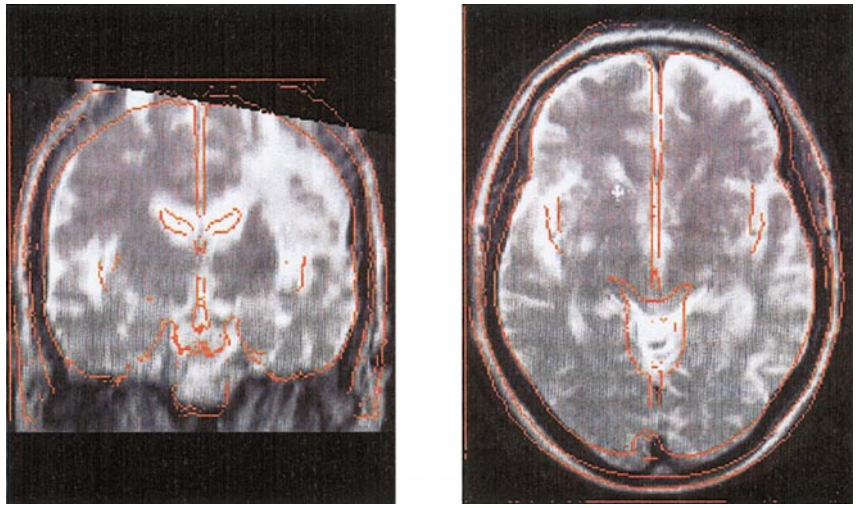

FIG. 9. Example slices from one of the images used in the consistency study (after registration). The red lines represent edges from the standard image (the reference image) overlayed on the transformed initial image.

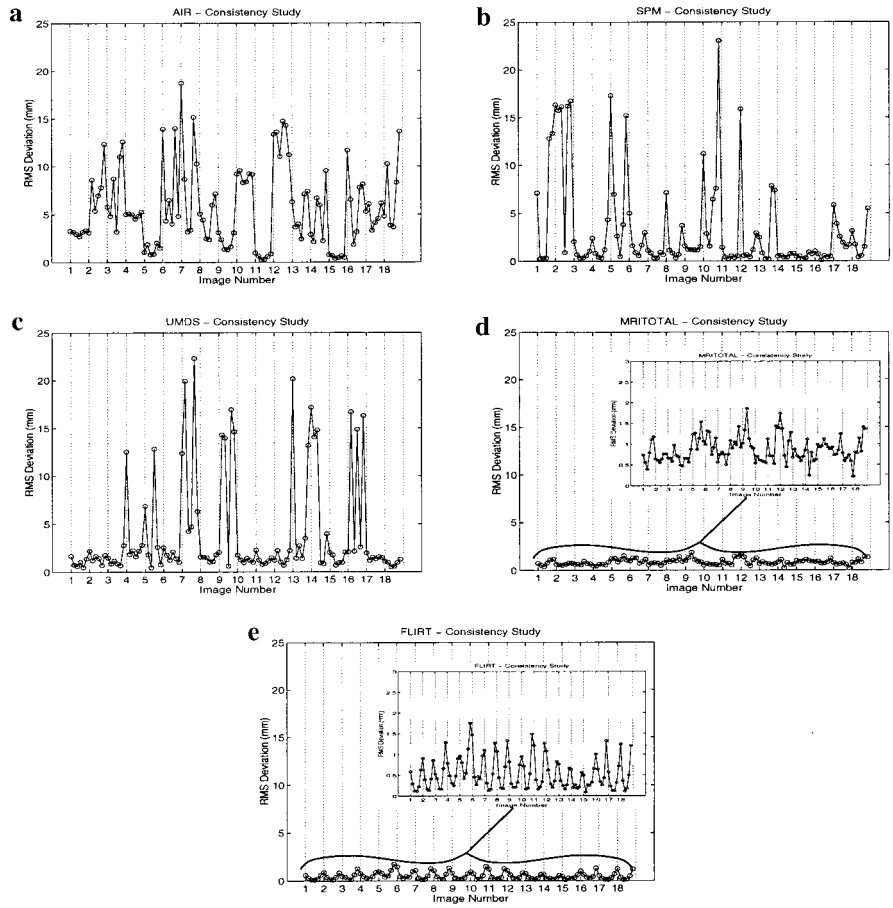

FIG. 10. Results of the consistency study.

able, and so this constituted a fair test as opposed to a reimplementation of a method described in a paper, where often the lack of precise implementation details makes it difficult to produce a good working method.

The particular experiment that was performed was intersubject and intermodal using 18 different images as the floating images (like the one shown in Fig. 9), all with the MNI 305 brain (Collins *et al.*, 1994) as the reference image. The 18 images were all 256 × 256 × 30, *T*2-weighted MR images with voxel dimensions of 0.93 by 0.93 by 5 mm, while the MNI 305 template is a 172 × 220 × 156, *T*1-weighted MR image with voxel dimensions of 1 by 1 by 1 mm.

The results of one such test, using six different rotations about the anterior–posterior axis, are shown in Fig. 10. It can be seen that only FLIRT and MRITOTAL performed consistently. This indicates that the other methods (AIR, SPM, and UMDS) frequently get trapped in local minima, i.e., are not as robust.

A further consistency test was then performed comparing only MRITOTAL and FLIRT. This test used initial scalings rather than rotations. The reason that this is important is that MRITOTAL uses a multiresolution local optimization method (gradient descent) but relies on initial preprocessing to provide a good starting position. This preprocessing is done by finding the principle axes of both images and initially aligning them. Consequently the initial alignment compensates for rotations but does not give any information, and hence correction, for scalings.

The results of the scaling consistency test are shown in Fig. 11. It can be seen that, although generally consistent, in three cases MRITOTAL produces registrations that deviate by more than 20 mm (RMS) from each other. In contrast, FLIRT was consistent (less than 2 mm RMS) in all cases.

Accuracy Assessment: Motion Correction

This section details the comparative accuracy of the motion correction scheme (MCFLIRT) when tested against two of the most widely used schemes, SPM and AIR.

In order to attempt to establish a "gold standard" for registration accuracy, initial tests use the RMS measure (Eq. 6) in combination with synthetic data, where the exact value of the motion is known, in order to quantify the scheme's ability to correct for subject mo-

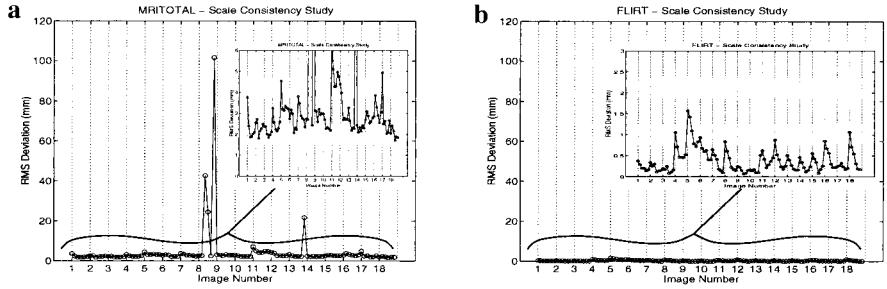

FIG. 11. Results of the scale consistency study.

tion. Later tests characterize the degree of correction by examining the residual variation (collapsed across time) in the corrected data. This measure is also used to test the scheme's effectiveness on real data where we have no absolute measure of the subject's movement.

Simulated data. The artificial data enabling gold standard comparisons were generated as follows: a high-resolution EPI volume ($2 \times 2 \times 2$ mm) was duplicated 180 times and each volume was transformed by an affine matrix corresponding to real motion estimates taken from one of two studies where the subject had been asked to move his or her head appreciably during the scan. Three further groups of images were generated using motion estimates from experiments where the subject had been asked to remain as still as possible. Within these five motion designs, three further groups of data were created corresponding to audiovisual activation at 0, 2.5, and 5% of the overall voxel intensities by modulating the intensity values according to a mask derived from real fMRI data. Once the activation (if any) had been applied and the volumes transformed by the corresponding parameters, the data were subsampled to $4 \times 4 \times 6$ mm voxels and appropriately cropped to avoid introducing any padding voxels. The use of a high-resolution template image which is then subsampled should minimize the effect of interpolation when applying such transformations to the data.

Within our correction scheme, there are a number of stages which can be tuned to optimize the accuracy of the correction. The remainder of this section aims to find a robust set of parameters which give consistently accurate results on all data presented. We begin by examining the comparative accuracy of several cost functions which can be used with our optimization scheme. Later we proceed to examine the impact made by the choice of interpolation scheme and registration schedule.

Cost functions. The test results shown in Fig. 12 show the relative accuracy of the available cost functions within the MCFLIRT optimization framework when applied to the problem of motion correction on our synthetic data.

Although there is no clear leader over all cost functions in terms of accuracy, we note that the most accurate results are predominantly yielded by the normalized correlation and correlation ratio cost functions. This observation is reinforced when we examine the number of data sets where a particular cost function is most accurate. This is summarized in Table 2.

Note that previous work (Freire and Mangin, 2001) which had demonstrated the superiority of entropy-based cost measures over alternatives in terms of motion correction without introducing further spurious activations in the data has only compared mutual information metrics against least squares (SPM) and Woods (AIR) measures.

The next stage of testing was to verify that these cost functions were in fact more accurate when smoothed (apodized) than unsmoothed (unapodized). The same

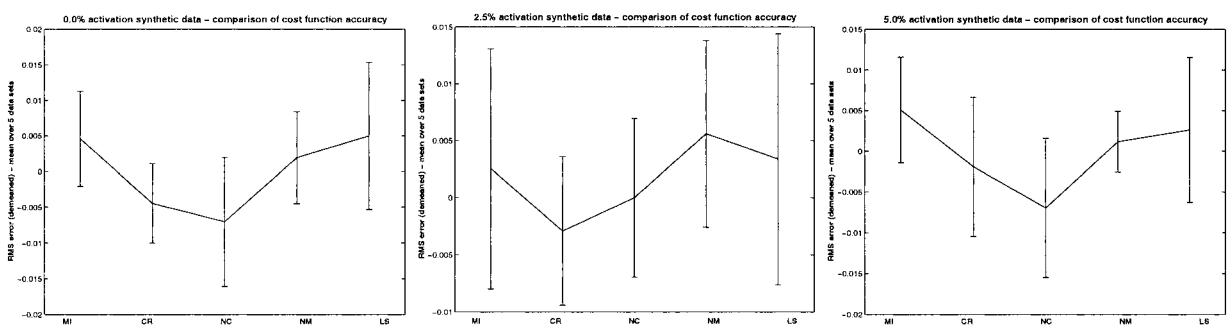

FIG. 12. Median (over time) RMS (over space) error results for the MCFLIRT scheme applied to synthetic data exhibiting known motion of one of five designs and audiovisual activation at increasing intensities. Cost function notation corresponds to Table 1.

TABLE 2

Accuracy Counts for the Five Cost Functions Offered by MCFLIRT

Cost	No. of sets most accurate	No. of sets second most accurate
Normalized correlation	8	5
Correlation ratio	2	7
Mutual information	2	1
Normalized mutual information	0	2
Least squares	3	0

RMS test measure and data sets were used as in the previous test and results are given in Fig. 13.

Overall, the smoothed cost functions outperform their unsmoothed versions.

Interpolation scheme. To further improve the accuracy of the motion estimates, the next parameter we experimented with was the choice of interpolation scheme for the motion estimation. In addition to the standard trilinear scheme, a windowed-sinc interpolation (using a Hanning window of size $7 \times 7 \times 7$) was tried. While considerably slower than trilinear interpolation, the sinc approach is able to further refine motion estimates after the initial trilinear stage has converged on a solution, thus providing greater accuracy. The results in Fig. 14 show the greater degree of accuracy achieved over using trilinear interpolation alone. Note that on the third data set (cropped to allow

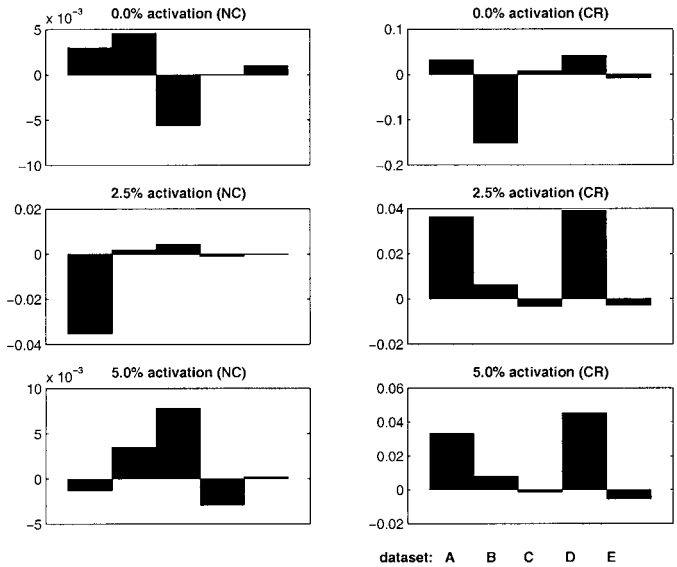

FIG. 13. Median (over time) RMS (over space) error results (unsmoothed minus smoothed) for the MCFLIRT scheme applied to five synthetic data sets (A–E) exhibiting known motion at increasing intensities. A positive value indicates improved accuracy as a result of smoothing the cost function. Cost function notation corresponds to Table 1 and the results demonstrate the improvement in accuracy achieved by using the smoothed cost functions.

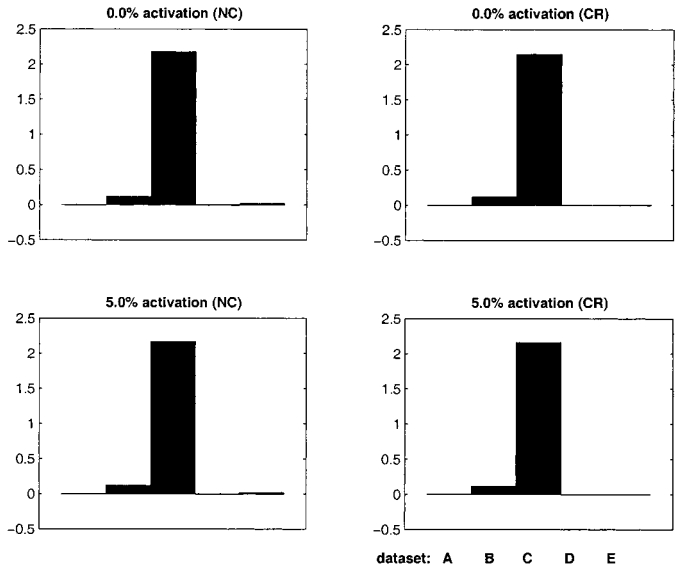

FIG. 14. Median (over time) RMS (over space) error results for the MCFLIRT scheme applied to synthetic data exhibiting known motion of one of five designs and audiovisual activation at increasing intensities. A positive value indicates improved accuracy as a result of incorporating the final sinc interpolation stage. Cost function notation corresponds to Table 1 and demonstrates the improvement in accuracy achieved by using smoothed cost functions and additional sinc interpolation when compared to the basic trilinear scheme reported in Fig. 13.

distinction between the other four sets), the improvement was consistently over a value of 2.0.

Choice of template image. In an attempt to increase the accuracy of the scheme and as a final parameter investigation, a method using a mean image template was implemented. This scheme generates a mean image for the series by averaging all the volumes over time after the first three stages of trilinear interpolation-based motion correction have been carried out. In doing so we hope to be registering all volumes to a more generalized target which exhibits less overall variation from each volume in the series than the original target (middle) volume previously used. This new mean image is a robust target to which the original time series is then registered, again using three trilinear interpolation stages and an optional final sinc interpolation stage.

Because we are registering to a mean image, we no longer have gold standard values for the transformations found by the correction scheme. Therefore, to quantify the accuracy of the correction, a median absolute residual variation (MARV) score was created by initially demeaning each voxel time-series and then measuring the median value of the residual absolute values in this time-series. That is,

$$\text{M ARV}(x, y, z) = \sum_{t=1}^{N} |I_t(x, y, z) - I_{\text{mean}}(x, y, z)|/N .$$

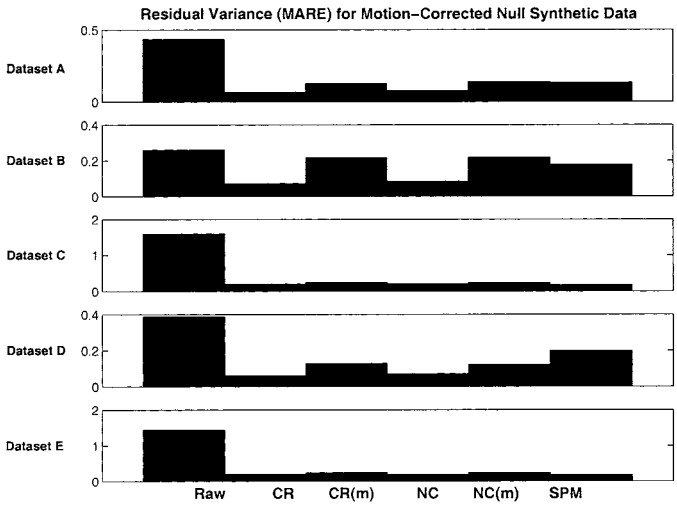

FIG. 15. Median absolute residual variation (MARV) values for corrected data processed by different motion correction schemes: Uncorrected, MCFLIRT w. CR, MCFLIRT w. NC, MCFLIRT w. CR and mean, MCFLIRT w. NC and mean, SPM99 run at full quality, with sinc interpolation and interpolation error adjustment.

This produces a volume of MARV scores for each voxel and the median of these values (over the volume) is then taken as a summary measure. This is effectively a measure characterizing the level of intervolume intensity variation (presumed to be due to subject motion) after retrospective motion correction has been applied. While this can only work for activation-free data (so that in perfect alignment the variance should be at minimum), it can give us a clear impression of the accuracy of the motion correction scheme. Because SPM rejects information outside a mask obtained from the data (end-slice effects), the corrected median images were masked according to the corrected SPM data so that the measure reflected a consistent comparison across the schemes. The results shown in Fig. 15 correspond to the MARV values generated after running MCFLIRT and SPM on the null-activation data set for both the low and the severe motion designs.

Results using the RMS measure (Table 3) revealed that, although all three schemes provide subvoxel accuracy, AIR 3.08 using least squares (which we found to give better results than the standard AIR measure) and windowed sinc interpolation was almost an order of magnitude worse than basic three-stage trilinear MCFLIRT. Accordingly, we decided not to compare it further. However, we note that AIR was primarily designed to solve several different registration problems that arise in tomographic data sets (Woods, 1998) rather than optimized for FMRI motion correction.

TABLE 3

RMS Deviation Values for Synthetic Null Data

	Uncorrected	AIR	SPM	MCFLIRT
Sum of squared intensity errors	936.5866	406.8876	1.6405	1.5171
RMS error (mm)	2.3360	1.7570	0.1064	0.1102

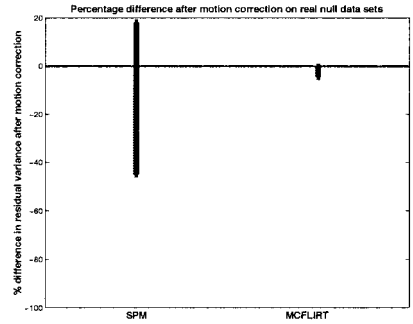

FIG. 16. Summary MARV statistics for three sets of real data, comparing MCFLIRT w. CR, and SPM99 correction using full quality, sinc interpolation, and interpolation error adjustment. The results show the range of percentage changes in residual variance as a result of the motion correction (i.e., comparing it to the uncorrected variance).

From these results we conclude that for some cases (generally the low motion data), MCFLIRT with the correlation ratio cost function produces significantly smaller errors than SPM99, while in other cases (some of the high motion data) both methods give similar results. This can be seen be comparing the heights (MARV values) of the SPM bars with the CR bars (typically the best MCFLIRT cost function) where a 30 to 40% reduction can be seen in the first, second, and fourth cases.

Slightly surprisingly, we found that the use of a mean image template gave no discernable improvement in accuracy. We conclude that for artificial data where the motion is purely rigid, there is no advantage to using an (possibly blurred) average image over an image from the original data. We would expect that the mean template scheme could yield greater accuracy where the data includes some physical motion-induced artifacts and the choice of a reference image from the original data set is not so obvious.

Null data study. Having established the accuracy of MCFLIRT on artificial data, we ran both our scheme and SPM99 on a number of real fMRI studies. In all instances, the subjects had been exposed to no stimulus (null data). The underlying assumption was that after motion correction on a null data study, we would expect the overall variation of the data to be lower than before correction as subject motion-induced variablity had been minimized. Again, results were masked according to the SPM data to give a fair comparison.

Results, given in Fig. 16, show considerable changes induced by SPM (both beneficial and detrimental) but only minimal changes induced by MCFLIRT. This is due to the fact that the amount of actual motion that occurred in these studies is very low, so that the

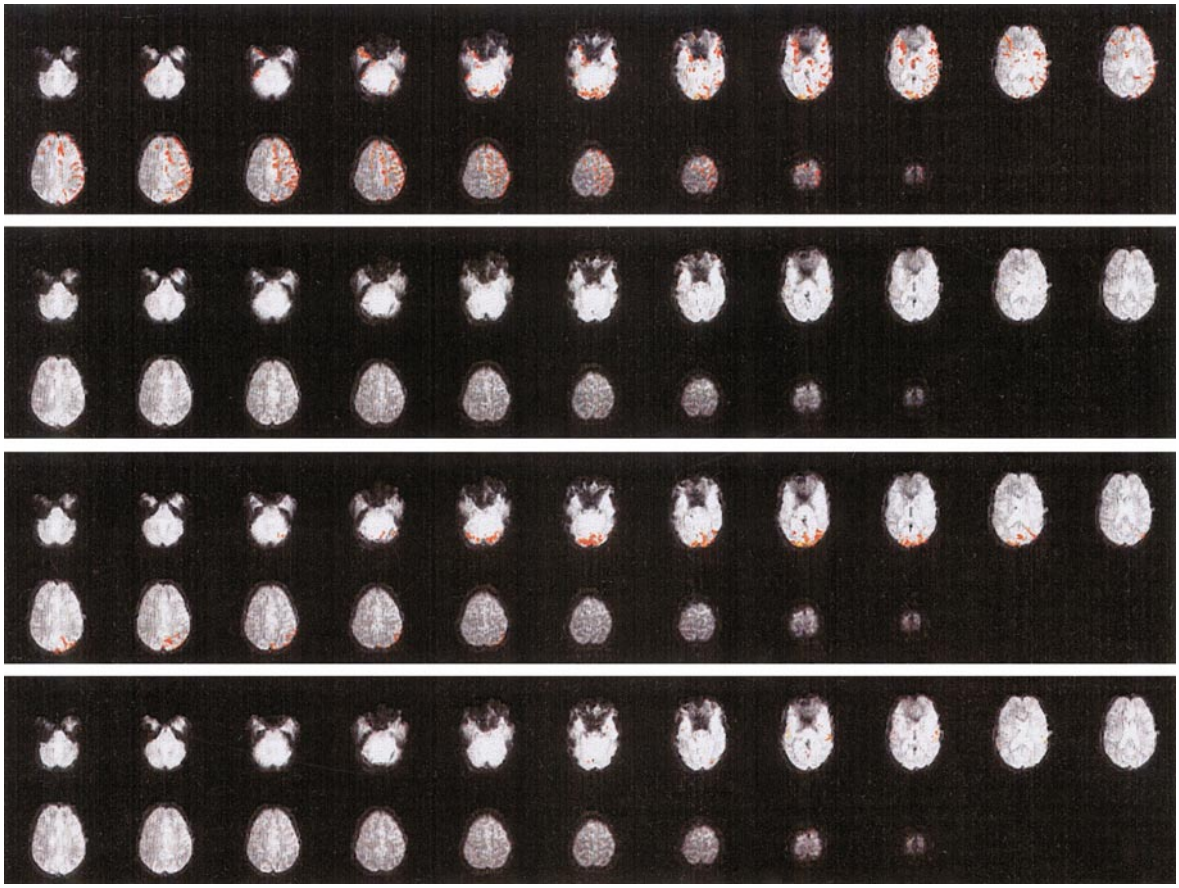

FIG. 17. Results of a FEAT analysis on motion-corrected audiovisual data.

changes in intensity in each voxel (as a result of motion correction) are also low—in fact, lower than the expected changes induced by physiological processes. Consequently, these test results are inconclusive since the test measure does not purely measure motion-induced changes but also physiological changes which dominate in this case. The only way one could expect to obtain quantitative analysis for real data would be to incorporate some form of position measurement into the scanner—a facility not available to us.

Real activation study. As we have described above, it is difficult to make accurate measurements pertaining to the accuracy of a motion correction scheme when presented with real data. With data exhibiting activation, examination of the time-series after correction using an animation tool reveals no visible extreme affine movement although some motion artifacts remain. In particular, we are able to show good localization of activations which would not be possible without motion correction first being carried out. The thresholded statistical maps shown in Fig. 17 correspond to a 180 volume audiovisual experiment. Analysis was carried out using FEAT, FMRIBs easy analysis tool using an improved linear model (Woolrich *et al.*, 2001). In order to test the effectiveness of MCFLIRT on real data, the subject was asked to move his or her head during the experiment.

It can be seen, by comparing the activations from the uncorrected and corrected data sets, that in the uncorrected set a large number of false positives exist in both visual and auditory results. While it can be argued that both data sets are highly corrupted by this large motion (indeed, even the corrected data set still exhibits some visible movement, albeit at a significantly smaller scale than the uncorrected data), the MCFLIRT-corrected visual stimulus is well localized and allows an otherwise corrupted set of experimental data to yield potentially useful results.

DISCUSSION

This paper has examined the problem of optimization for fully automatic registration of brain images. In particular, the problem of avoiding local minima is addressed in two ways. First, a general apodization (smoothing) method was formulated for cost functions in order to eliminate small discontinuities formed by discontinuous changes of the number of voxels in the overlapping field of view with changing transformation parameters. Second, a novel hybrid global–local opti-

mization method was proposed which uses prior knowledge about the problem (size of the brain, expected changes in scalings, etc.) while combining the speed of local optimization with an initial search phase and multistart optimization component. The latter two components are similar to those used in simulated annealing and genetic algorithms—two popular, but slow, global optimization techniques.

Only affine (linear) registration was examined in this paper as, although this is a much easier problem than general nonlinear registration, finding the global minimum is still difficult. Furthermore, many nonlinear methods rely on an initial affine registration to find a good starting position, and so having a good method of affine registration is important.

The global optimization method proposed here does not, however, guarantee finding the global minimum. This is typical though, as even methods such as simulated annealing and genetic algorithms only provide a statistical guarantee which cannot be met in practice. The results, though, are encouraging, and by using finer search grids, the likelihood of finding the global minimum can be increased. This requires that there be sufficient time at hand or a sufficiently fast computer. However, even with modest resources this method can find the global minimum and solve the registration problem within 1 h (and often much less) more reliably than the other methods tested.

Optimization is only one aspect of the registration problem, although it is practically a very important one. Other aspects such as interpolation, alternative cost functions, and understanding the properties of existing cost functions remain important areas for further work. In addition, a theoretical justification for the current method and finding a method suitable for higher dimensional transformations are important areas for future research.

The implementations of the registration and motion correction methods (FLIRT and MCFLIRT) were tested using experiments designed to demonstrate the improved robustness and accuracy. These issues are important and each is examined separately.

Robustness Study

Quantitative results for the robustness of FLIRT were ascertained using a consistency test. This test is designed to examine the robustness of a registration method by comparing registrations obtained using various different, but known, initial starting positions of a given image. Results showed that the method was highly consistent on a set of difficult images. Furthermore, several other available packages were tested on the same set of images and did not achieve the same level of consistency, sometimes demonstrating substantial inconsistencies. These tests (together with those presented in Jenkinson and Smith (2001)) showed that the robustness was due to the optimization method, not just the choice of cost function, and that in order to achieve the robust registrations, multiresolution local optimization alone is insufficient. Moreover, the newly proposed hybrid local–global optimization method achieves a much greater degree of robustness, which is necessary for fully automatic use, within a prescribed time limit (less than 1 h on a PC; e.g., registering two $1 \times 1 \times 1$ mm images typically takes 15 min on a 500-MHz Pentium III).

Accuracy Study

In the case of accuracy we have shown that MCFLIRT optimization routines, cost functions, and sinc interpolation consistently achieve high levels of accuracy. In particular, the RMS test measure shows that the error is typically around 0.1 mm, which is more than an order of magnitude less than the voxel size of 4 mm, but necessary to ensure that subsequent statistical analysis is valid. Furthermore, tests demonstrated that the average performance of the MCFLIRT scheme was superior to both SPM99 and AIR v3.08. Results of the tests using real data that contained activation, where the underlying ground truth was not known and the RMS measure could not be used, were inconclusive due to the presence of (an unknown amount of) physiologically induced intensity variations. Note that in all cases (for the robustness and accuracy studies) the data sets used for comparative testing were independent of those used to tune the empirical parameters of the methods used.

Early tests using synthetic data have revealed that in cases where the motion is moderate (up to 2 mm translation and 2° rotation), the sequential initialization (see Fig. 6) scheme yields an improvement in the accuracy of motion estimates compared to one where no sequential initialization is performed. Conversely, in cases where the amplitude of motion parameters was known to be high, there was no inherent disadvantage in making the assumption of an underlying smooth motion trend across time points.

We would be interested to see how robustly the schemes perform over time-series of varying length. If at all significant, we might expect to see some impact on the MCFLIRT mean image registration scheme where a longer time-series might provide a more general and robust template image. At present there is no guaranteed advantage in using the mean template in addition to the standard correction schedule but one would expect it to play a more beneficial role in correcting extended time-series exhibiting moderate to low motion artifacts.

The methods and results shown here are all for whole brain data sets, although the algorithms have also been successfully adapted to work with data sets containing very few slices (or just single slices) by

restricting the transformations to be two dimensional, as the optimization and apodizing methods also apply for these two-dimensional registrations. However, when general three-dimensional registration is required for images containing few slices, other approaches need to be used, such as those employed by some of the other packages tested here (where no global search is involved).

Practical registration packages usually require the setting of certain parameter values. The methods introduced here also contain several configurable parameters such as image resolutions for the various optimization stages. These values have been selected empirically, over a wide range of data sets, to be as robust as possible for general brain images. However, when dealing with particular data sets these general settings may not be optimal and so most methods allow these values to be changed, via configurable options. When using other packages for comparitive studies we have tried to select the best general parameter values (by consulting the appropriate documentation) but recognize that further improvement upon these results could be possible by careful selection of parameters. The tuning of registration methods is, at present, an undesirable necessity in many situations, which prevents easy automation. By using more robust algorithms this tuning of parameter settings can hopefully be minimized or avoided entirely.

In summary, the FLIRT and MCFLIRT packages are highly robust and accurate, as has been demonstrated by the quantitative experiments and by qualitative feedback. These methods have now been used to satisfactorily solve thousands of registration problems, some using extremely different imaging modalities. The binary and source code distributions for the MCFLIRT and FLIRT packages are available for downloading from www.fmrib.ox.ac.uk/fsl.

ACKNOWLEDGMENTS

The authors thank the Medical Research Council, the European MICRODAB project, and the Engineering and Physical Sciences Research Council for supporting this work.

REFERENCES

Biswal, B. B., and Hyde, J. S. (1997). Contour-based registration techique to differentiate between task-activated and head motion-induced signal variations in fMRI. *Magn. Reson. Med.* **38:** 470–476.

Cachier, P., and Rey, D. (2000). Symmetrization of the non-rigid registration problem using inversion-invariant energies: Application to multiple sclerosis. In *3rd International Conference on Medical Image Computing and Computer-Assisted Intervention,* pp. 472–481. Springer, Berlin.

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

Freire, L., and Mangin, J.-F. (2001). Motion correction algorithms of the brain mapping community create spurious functional activations. In *17th International Conference on Information Processing in Medical Imaging* (M. Insana and R. Leahy, Eds.), pp. 246–258. Springer-Verlag, Berlin.

Friston, K., Ashburner, J., Frith, C., Poline, J.-B., Heather, J., and Frackowiak, R. (1995). Spatial registration and normalization of images. *Hum. Brain Mapping* **2:** 165–189.

Friston, K., Williams, S., Howard, R., Frackowiak, R., and Turner, R. (1996). Movement-related effects in fMRI time-series. *Magn. Reson. Med.* **35:** 346–355.

Hajnal, J., Saeed, N., Oatridge, A., Williams, E., Young, I., and Bydder, G. (1995a). Detection of subtle brain changes using subvoxel registration and subtraction of serial MR images. *J. Comput. Assist. Tomogr.* **19**(5): 677–691.

Hajnal, J., Saeed, N., Soar, E., Oatridge, A., Young, I., and Bydder, G. (1995b). A registration and interpolation procedure for subvoxel matching of serially acquired MR images. *J. Comput. Assist. Tomogr.* **19**(2): 289–296.

Izenman, A. J. (1991). Recent developments in nonparametric density estimation. *J. Am. Statis. Assoc.* **86**(413): 205–224.

Jenkinson, M. (1999). Measuring transformation error by RMS deviation. Internal Technical Report TR99MJ1, FMRIB Centre, University of Oxford. Available at www.fmrib.ox.ac.uk/analysis/techrep for downloading.

Jenkinson, M., and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* **5**(2): 143–156.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. (1997). Multimodality image registration by maximisation of mutual information. *IEEE Trans. Med. Imag.* **16**(2): 187–198.

Maes, F., Vandermeulen, D., and Suetens, P. (1999). Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. *Med. Image Anal.* **3**(4): 373–386.

Maintz, J., and Viergever, M. (1998). A survey of medical image registration. *Med. Image Anal.* **2**(1): 1–36.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (1995). *Numerical Recipies in C*, second ed., Cambridge Univ. Press, Cambridge, UK.

Studholme, C., Hill, D., and Hawkes, D. (1996). Automated 3D registration of MR and CT images of the head. *Med. Image Anal.* **1**(2): 163–175.

West, J., *et al.* (1997). Comparison and evaluation of retrospective intermodality brain image registration techniques. *J. Comput. Assist. Tomogr.* **21**(4): 554–566.

Woods, R., Mazziotta, J., and Cherry, S. (1993). MRI-PET registration with automated algorithm. *J. Comput. Assist. Tomogr.* **17**(4): 536–546.

Woods, R. P. (1998). Automated Image Registration. Web page. http://www.bishopw.loni.ucla.edu/AIR3/.

Woolrich, M., Ripley, B., Brady, J., and Smith, S. (2001). Temporal autocorrelation in univariate linear modelling of FMRI data. *NeuroImage* **14**(6): 1370–1386.