

High-Dimensional Image Registration Using Symmetric Priors

John Ashburner, Jesper L. R. Andersson, and Karl J. Friston

The Wellcome Department of Cognitive Neurology, Institute of Neurology, Queen Square, London WC1N 3BG, United Kingdom

Received July 23, 1998

This paper is about warping a brain image from one subject (the object image) so that it matches another (the template image). A high-dimensional model is used, whereby a finite element approach is employed to estimate translations at the location of each voxel in the template image. Bayesian statistics are used to obtain a *maximum a posteriori* (MAP) estimate of the deformation field. The validity of any registration method is largely based upon the constraints or, in this instance, priors incorporated into the model describing the transformations. In this approach we assume that the priors should have some form of symmetry, in that priors describing the probability distribution of the deformations should be identical to those for the inverses (i.e., warping brain A to brain B should not be different probabilistically from warping B to A). The fundamental assumption is that the probability of stretching a voxel by a factor of n is considered to be the same as the probability of shrinking n voxels by a factor of n^{-1} . In the Bayesian framework adopted here, the priors are assumed to have a Gibbs form, where the Gibbs potential is a penalty function that embodies this symmetry. The penalty function of choice is based upon the singular values of the Jacobian having a lognormal distribution. This enforces a continuous one-to-one mapping. A gradient descent algorithm is presented that incorporates the above priors in order to obtain a MAP estimate of the deformations. We demonstrate this approach for the two-dimensional case, but the principles can be extended to three dimensions. A number of examples are given to demonstrate how the method works. © 1999 Academic Press

Key Words: registration; anatomy; imaging; stereo-taxy; spatial normalization; PET; MRI; functional mapping.

1. INTRODUCTION

Image registration is an important component in many neuroimaging applications. One of the objectives of image registration is to allow the characterization of the morphology of different subjects' brains. This allows anatomical comparisons to be made among different populations. Studies of brain morphometry have al-

ready revealed structural differences between a number of patient populations, and much of the focus of schizophrenia research in particular is based upon brain morphometry. We have previously presented a multivariate approach for identifying global structural differences based on very smooth (low spatial frequency) deformation fields (Ashburner *et al.*, 1998). The registration approach developed here should allow much more detailed deformation fields to be obtained and therefore allow one to identify more subtle structural differences with greater precision.

Another objective of this work is to provide a more precise form of spatial normalization. In functional imaging, the operation of spatial normalization facilitates intersubject averaging of data. Modalities such as positron emission tomography (PET) are limited by the radioactive dose that can be administered to any subject. This means that it is often necessary to average signals over a number of subjects in order to obtain a meaningful result. A template image is used to define the standard space into which the different subjects are warped. By using a template that conforms to the space of a standard coordinate system, such as that defined by Talairach and Tournoux (1988), it is possible to report anatomical positions in terms of meaningful Cartesian coordinates, relative to some reference space (i.e., the template).

Spatial transformations can be broadly divided into *label based* and *non-label based*. Label-based techniques identify homologous features (labels) in the image and template and find the transformations that best superpose them. The labels can be points, lines, or surfaces. If the labels are points, then the required transformations at each of those points is known. Between the points, the deforming behavior is not known, so it is forced to be as "smooth" as possible. There are a number of methods for modeling this smoothness. The simplest models include fitting splines through the points in order to minimize *bending energy* (Bookstein, 1997, 1989). More complex forms of interpolation are often used when the labels are surfaces. For example, Thompson *et al.* (1996) map surfaces together using a fluid model. Non-label-based approaches identify a spatial transformation that minimizes some index of the difference between an object and a tem-

plate image, where both are treated as unlabeled continuous processes. The matching criterion is usually based upon minimizing the sum of squared differences (equivalent to maximizing the correlation) between the images. For this criterion to be successful, it requires the template to appear like a warped version of the image. In other words, there must be correspondence in the gray levels of the different tissue types between the image and template.

There are a number of approaches to non-label-based registration. A potentially enormous number of parameters are required to describe the transformations that warp two images together (i.e., the problem is very high dimensional). The forms of registration tend to differ in how they cope with the large number of parameters. One approach is to reduce the number of parameters that model the deformations. Some groups simply use only a 9- or 12-parameter affine transformation to spatially normalize their images, accounting for differences in position, orientation, and overall brain size. Low spatial frequency global variability in head shape can be accommodated by describing deformations by a linear combination of low frequency basis functions (Amit *et al.*, 1991; Friston *et al.*, 1995; Ashburner and Friston, submitted for publication). The small number of parameters will not allow every feature to be matched exactly, but it will permit the global head shape to be modeled.

The deformations required to transform images to the same space are not clearly defined. Unlike rigid body transformations, where the constraints are explicit, those for warping are more arbitrary. Regularization schemes are therefore necessary when attempting high-dimensional image registration, thus ensuring that voxels remain close to their neighbors. Without any constraints it is of course possible to transform any image such that it matches another exactly. The issue is therefore less about the nature of the transformation and more about defining constraints or priors under which a transformation is effected. Priors are normally incorporated using some form of Bayesian scheme, using estimators such as the *maximum a posteriori* (MAP) estimate or the *minimum variance estimate* (MVE). The MAP estimate is the single solution that has the highest *a posteriori* probability of being correct, and it is this estimate that we use in this paper. The MVE is used by Miller *et al.* (1993, 1994) and is the solution that is the conditional mean of the posterior. In addition to minimizing the differences between the images, Bayesian schemes involve minimizing additional functions relating to the position of each voxel relative to its neighbors. Often, the *a priori* distributions used by these schemes are linear and include minimizing the *membrane energy* (or *Laplacians*) of the deformation field (Amit *et al.*, 1991; Gee *et al.*, 1997), the *bending energy* (Bookstein, 1997), and the *linear-elastic energy* (Miller *et al.*, 1993). However, none of

these linear regularization schemes necessarily preserve the topology of the warped images.

An alternative to using a Bayesian scheme incorporating some form of elastic prior would be to use a viscous fluid model (Christensen *et al.*, 1994, 1996) to describe the warps. In these models, finite difference methods are normally used to solve the partial differential equations that model one image as it “flows” to the same shape as the other. The major advantage of these methods is that they are able to account for large displacements and also ensure that the topology of the warped image is preserved, but they do have the disadvantage that they are computationally expensive. Viscous fluid models are almost able to warp any image so that it looks like any other image, while still preserving the original topology. In some respects these models may have too much freedom, in that extremely unlikely deformations are not penalized.

Viscous fluid models are one of many approaches that describe the spatial transformations in terms of a physical process. However, rather than obeying physical laws, the intensity-based registration model presented in this paper utilizes statistical rules. Unlikely deformations are penalized by incorporating prior information about the smoothness of the expected deformations using a MAP scheme. In addition, the topology of the deformed images is preserved by ensuring that the deformations are globally one to one. The remainder of this paper is divided into two sections. We first introduce the theoretical background and then some illustrative applications.

2. THEORY

A high-dimensional image registration can be considered as an optimization problem involving many thousands of parameters. In order to achieve a satisfactory result with this many parameters, it is essential that priors or constraints are imposed. A Bayesian framework has been used by a number of researchers (Amit *et al.*, 1991; Miller *et al.*, 1993; Gee *et al.*, 1995a,b) to incorporate prior probability distributions into a warping model, and we do the same here. Bayes rule can be expressed as

$$p(\mathbf{Y}|\mathbf{b}) \propto p(\mathbf{b}|\mathbf{Y}) p(\mathbf{Y}),$$

where $p(\mathbf{Y})$ is the *a priori* probability of parameters \mathbf{Y} being the true parameters, $p(\mathbf{b}|\mathbf{Y})$ is the likelihood of observing data \mathbf{b} given that parameters \mathbf{Y} are correct, and $p(\mathbf{Y}|\mathbf{b})$ is the *a posteriori* probability that \mathbf{Y} are correct given the data \mathbf{b} . Here, \mathbf{Y} are the parameters describing the deformation, and \mathbf{b} are the images to be matched. The estimate that we determine here is the MAP estimate, which is the value of \mathbf{Y} that maximizes $p(\mathbf{Y}|\mathbf{b})$. A probability is related to its Gibbs form by

$$p(\mathbf{Y}) \propto e^{-H(\mathbf{Y})}.$$

Therefore, the MAP estimate is identical to the parameter estimate that minimizes the Gibbs potential of the posterior distribution ($H(\mathbf{Y}|\mathbf{b})$), where

$$H(\mathbf{Y}|\mathbf{b}) = H(\mathbf{b}|\mathbf{Y}) + H(\mathbf{Y}) + c,$$

where c is a constant.

From an information theoretic perspective, it can be seen that the Gibbs potential is simply the information. Therefore minimizing $H(\mathbf{Y}|\mathbf{b})$ is equivalent to minimizing the information or surprise about a deformation given the data. In order to minimize $H(\mathbf{Y}|\mathbf{b})$, we must minimize the sum of $H(\mathbf{b}|\mathbf{Y})$ (the likelihood potential) and $H(\mathbf{Y})$ (the prior potentials). We now discuss these potentials in detail.

2.1. Likelihood Potentials ($H(\mathbf{b}|\mathbf{Y})$)

The registration matches an object image (\mathbf{f}) to a template image (\mathbf{g}). Our current model assumes that one is simply a spatially transformed version of the other (i.e., there are no intensity variations between them), where the only intensity differences are due to uniform additive Gaussian noise. The Gibbs potential for this situation is given by

$$H(\mathbf{b}|\mathbf{Y}) = \frac{1}{2\sigma^2} \sum_{i=1}^I (f(\mathbf{y}(\mathbf{x}_i)) - g(\mathbf{x}_i))^2,$$

where $g(\mathbf{x}_i)$ is the i th pixel value of \mathbf{g} and $f(\mathbf{y}(\mathbf{x}_i))$ is the corresponding pixel value of \mathbf{f} . In the model we use here, the variance (σ^2) is assumed to be the same for all voxels. A suitable value to use for each iteration is estimated by computing the residual sum of squared differences. For the early iterations, σ^2 has a higher value. This places more weight on the priors in the MAP optimization scheme, so that the deformations are smoother. When close to the final solution, σ^2 has decreased, and the algorithm is able to compute more detailed deformations.

2.2. Prior Potentials ($H(\mathbf{Y})$)

Consider the deformation fields that register two images \mathbf{f} and \mathbf{g} . The two fields that map from \mathbf{f} to \mathbf{g} and from \mathbf{g} to \mathbf{f} can be combined in order to map from \mathbf{f} to \mathbf{g} and then back to \mathbf{f} . If the registrations are perfect, then the resulting deformation should be uniformly zero. Any deviations must be due to registration errors. In order to minimize these errors, we suggest that there should be symmetry in the priors. In addition to considering the deforming forces that warp image \mathbf{f} to match image \mathbf{g} , we also need to consider what is happening to the forces mediating the inverse of the deformation field. In order to achieve this symmetry, we make the fundamental assumption that the probability of stretching a voxel by a factor of n is the same as the

probability shrinking n voxels by a factor of n^{-1} . For example, a deformation that stretches one voxel in the object image to fit two voxels in the template should incur the same penalty as the contraction of two voxels to fit one template voxel.

In order to compute these potentials, we consider the pixels of the template image (\mathbf{g}) as being on a regular grid, with unit spacing between them. A triangular mesh connects the centers of each pixel (as shown in Fig. 1). Within each triangle, there is assumed to be a uniform affine mapping between the images. If the coordinates of the vertices of an undeformed triangle are (x_{11}, x_{21}) , (x_{12}, x_{22}) , and (x_{13}, x_{23}) , and if they map to coordinates (y_{11}, y_{21}) , (y_{12}, y_{22}) , and (y_{13}, y_{23}) , respectively, then the 3×3 affine mapping (\mathbf{M}) can be obtained by

$$\mathbf{M} = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ 1 & 1 & 1 \end{pmatrix}^{-1}.$$

The Jacobian (\mathbf{J}) of this affine mapping is simply obtained from matrix \mathbf{M} by

$$\mathbf{J} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}.$$

The penalty for distorting each of these triangles is derived from its Jacobian. By using singular value decomposition, \mathbf{J} can be decomposed into two unitary matrixes (\mathbf{U} and \mathbf{V}) and a diagonal matrix (\mathbf{S}), such that $\mathbf{J} = \mathbf{U}\mathbf{S}\mathbf{V}^T$. The unitary matrixes simply represent

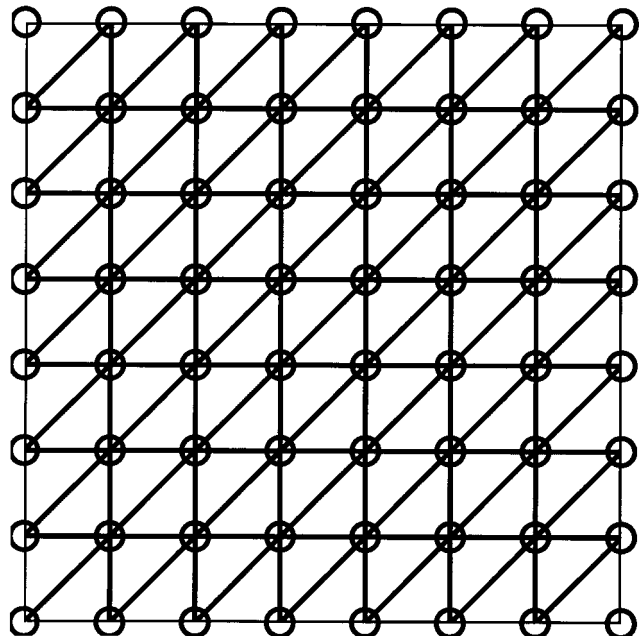


FIG. 1. The area of the template image (\mathbf{g}) is divided into a triangular mesh where the nodes are centered on the pixels.

rotations¹ and are therefore not important to the penalty function. Diagonal matrix \mathbf{S} contains the singular values, and these represent relative stretching in orthogonal directions. The determinant of \mathbf{J} ($\det(\mathbf{J})$) represents relative volume changes and is simply the product of the singular values.

A suitable prior potential function should preserve a one to one mapping between \mathbf{f} and \mathbf{g} , by constraining the determinants of the Jacobians to be positive. The inverse of the mapping also needs to be considered, in that the potential (per unit area) for \mathbf{J} should be identical to that which would be obtained for \mathbf{J}^{-1} . A penalty such as $\log(\det(\mathbf{J}))^2$ (or even $\det(\mathbf{J}) + \det(\mathbf{J})^{-1} - 2$) would realize both these criteria. However, relative lengths also need to be considered, and the length and volume changes should have similar distributions. A suitable form for this function is based upon the diagonal elements of \mathbf{S} being drawn from a lognormal distribution (i.e., the logs of the elements are normally distributed). The penalty per unit area is therefore $\lambda \log(s_{11})^2 + \lambda \log(s_{22})^2$, where λ is a “regularization parameter.”² A lognormal distribution for each diagonal element of \mathbf{S} also implies a lognormal distribution for $\det(\mathbf{J})$, since $\log(\det(\mathbf{J})) \equiv \log(s_{11}) + \log(s_{22})$ and both $\log(s_{11})$ and $\log(s_{22})$ are normally distributed. Each triangular patch has an area of $1/2$ pixel, and it will have an area of $\det(\mathbf{J})/2$ pixels when mapped to the space of image \mathbf{f} . The total area affected by the penalty in both the template and the object images is therefore $(1 + \det(\mathbf{J}))/2$, so the penalty for each triangle becomes $h = \lambda(1 + \det(\mathbf{J})) (\log(s_{11})^2 + \log(s_{22})^2)/2$. Examples of these penalties in terms of two-dimensional probability functions are illustrated in Fig. 2. The prior potential over the whole image is based on the sum of the potentials for each of the I triangles:

$$H(\mathbf{Y}) = \sum_{i=1}^I h_i$$

For simplicity, in the current description, the fact that the images have boundaries is ignored. The boundaries are fixed so that the deformation at the edges is always zero. This ensures that the topology is preserved, because deformations are globally one to one when they are locally one to one and have fixed boundaries.

¹ Complications arise when the determinant of \mathbf{J} is negative. In this case either \mathbf{U} or \mathbf{V} will also incorporate a reflection by having a negative determinant. However, this should not cause problems since the registration prevents the determinant of \mathbf{J} from becoming negative.

² Short of determining λ using a large number of “true” deformations, it is assigned some suitable value that facilitates rapid convergence to reasonable solutions.

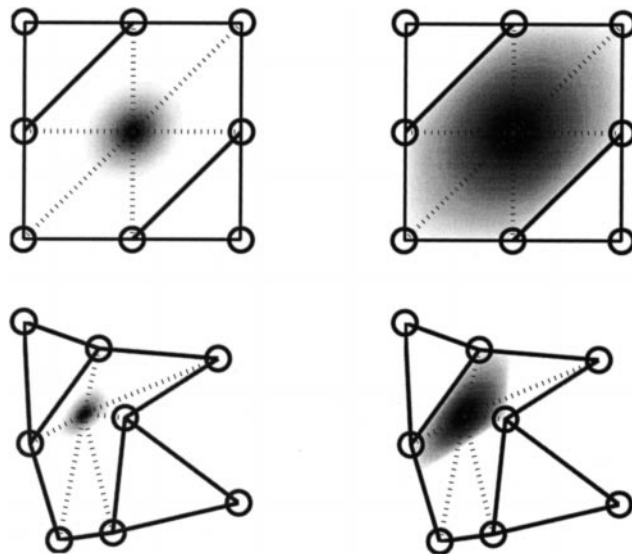


FIG. 2. Probability density functions relating to the position of the center voxel, assuming that all other voxels remain at fixed locations on a regular grid. Left column, using heavy regularization $\lambda = 10$. Right column, using light regularization $\lambda = 1$. Top, on a regular grid; bottom, on an irregular grid.

2.3. Steepest Descent Optimization

The images are matched by estimating the set of parameters (\mathbf{Y}) that maximizes their *a posteriori* probability. This involves beginning with a set of starting estimates and repeatedly making a tiny adjustment to the parameters such that the *a posteriori* potential is decreased. In each iteration, the voxel coordinates are updated *in situ*, by sequentially scanning from top to bottom and left to right. In the next iteration, the order of the updating is reversed (bottom to top and right to left), and this alternating sequence is continued until there is no improvement.

Each iteration of the optimization involves determining the rate of change of the *a posteriori* potential with respect to tiny changes in each element of \mathbf{Y} . For the n th iteration, the estimates for the i th element of \mathbf{Y} are modified according to

$$\mathbf{y}_i^{(n+1)} = \mathbf{y}_i^{(n)} - \epsilon \frac{\partial H(\mathbf{Y} | \mathbf{b})}{\partial \mathbf{y}_i} = \mathbf{y}_i^{(n)} - \epsilon \left(\frac{\partial H(\mathbf{b} | \mathbf{Y})}{\partial \mathbf{y}_i} + \frac{\partial H(\mathbf{Y})}{\partial \mathbf{y}_i} \right),$$

where the value of ϵ is chosen to be suitably small (see below).

$\partial H(\mathbf{b} | \mathbf{Y}) / \partial \mathbf{y}_i$ is the rate of change of likelihood potential with respect to changes in \mathbf{y}_i :

$$\frac{\partial H(\mathbf{b} | \mathbf{Y})}{\partial \mathbf{y}_i} = \frac{\partial (f(\mathbf{y}_i) - g(\mathbf{x}_i))^2 / (2\sigma^2)}{\partial \mathbf{y}_i} = \frac{(f(\mathbf{y}_i) - g(\mathbf{x}_i)) \partial f(\mathbf{y}_i)}{\sigma^2 \partial \mathbf{y}_i},$$

where σ^2 is estimated as described in Section 2.1.

$\partial H(\mathbf{Y})/\partial \mathbf{y}_i$ is dependent upon changes to the Jacobians of the six adjacent triangles shown in Fig. 3. Because the mathematics of computing these partial derivatives is algebraically dense, a C subroutine is provided in Fig. 4 that will compute the derivatives for a single triangular patch.

It is essential that none of the determinants become less than or equal to zero, because the registration algorithm would be unable to recover if this were to happen. However, preventing this is relatively straightforward. Before any \mathbf{y}_i is updated, the areas of all affected triangles are computed. If any of the areas would become less than or equal to zero, then the value of ϵ is decreased, for that point, until the change in \mathbf{y}_i results in a nonnegative area. By incorporating this test of stability, it is possible to use larger values of ϵ . This is especially useful for the early iterations because it allows faster convergence by allowing large deformations to be estimated more rapidly. For later iterations, smaller values of ϵ are more appropriate as they slow down convergence, thus reducing the amount of noise in the deformation fields.

3. ILLUSTRATIONS

3.1. Examples Using Simulated Data

Simulated data were used to demonstrate the reversibility of the deformation fields. Two images (of 100×100 pixels) were constructed, one of them a circle and the other a square. Pixel intensities ranged from 0 to 1. The circle was warped to match the square, and the square was warped to match the circle using a value of 1 for λ . No noise was added to the images, so a constant variance (σ^2) of 0.01 was assumed for all iterations. The final results of the registration are shown in Fig. 5.

In order to demonstrate the symmetry of the deforma-

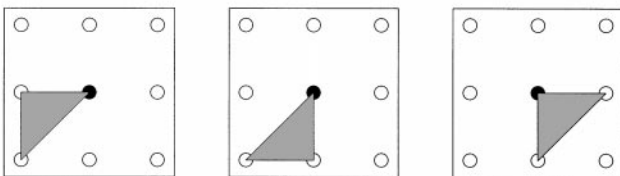
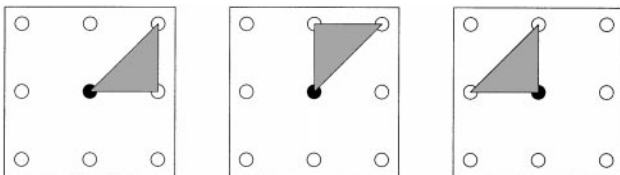


FIG. 3. The six neighboring triangles whose Jacobians are influenced by translating the central point.

tions, the two deformation fields were combined. These are shown in Fig. 6. If the deformations were perfectly symmetric, the combined deformations would be completely uniform. However, wrinkles can be seen which may be due to using finite approximations of continuous functions. Another contributing factor to the wrinkles may be because the likelihood potentials driving the registration are not symmetric. Only the gradients of one of the images is used. Future work may involve using the gradients of both images to drive the registration.

With some modifications, the method should allow large deformations to be estimated by reducing the value of λ . The limitations of using a finite element approach on a fixed lattice mean that a regriding scheme (Christensen *et al.*, 1996) may need to be incorporated to allow larger deformations to be modeled correctly. However, because brains are all of similar shape, the ability to model very large deformations in this way may not be necessary in neuroimaging applications.

3.2. Registering a Pair of Images

Approximately corresponding slices through two MR images of different subjects were registered together using the current approach. Each image contained 200×256 pixels (where each pixel was 1×1 mm), so the registration involved optimizing a total of 100,584 parameters.

In order to reduce the chance of the algorithm being caught in a local minimum, the first few iterations of the registration were carried out with the images smoothed using an 8-mm full width at half maximum Gaussian convolution kernel. Larger values for λ were also used for the early iterations in order to estimate the global head shape prior to estimating the more detailed deformations. The final results of the registration are shown in Fig. 7.

Two-hundred iterations were used for this particular example, each iteration taking approximately 2 s on a Sun SPARC Ultra 2. σ^2 was estimated from the residual squared difference between the images at each iteration, and the regularization parameter (λ) was set to 0.5 in the final iterations.

3.3. Toward a Canonical Brain

The main objective of spatial normalization is to transform images from a number of subjects to the same stereotactic space. In order to achieve this, the images of all the subjects must be matched to a template image. Obviously, some template images are "better" than others. For example, if the image of a subject with a particularly unusual pattern of cortical folding was used, then the validity of the resulting registrations would be compromised. In order to regis-

```

void dh_dy(double *h, double *dh1, double *dh2, double lambda,
           double x11, double x21, double y11, double y21,
           double x12, double x22, double y12, double y22,
           double x13, double x23, double y13, double y23)
{
    double j11, j12, j21, j22, dj1, dj2;
    double w, w1, w2, dt, dt1, dt2, tm, tm1, tm2;
    double s1, s2, d1s1, d1s2, d2s1, d2s2;
    double dtx, t1, t2, t3, t4;
    dtx = x11*(x22-x23)+x12*(x23-x21)+x13*(x21-x22);
    j11 = (y11*(x22-x23)+y12*(x23-x21)+y13*(x21-x22))/dtx;
    j12 = (y11*(x13-x12)+y12*(x11-x13)+y13*(x12-x11))/dtx;
    j21 = (y21*(x22-x23)+y22*(x23-x21)+y23*(x21-x22))/dtx;
    j22 = (y21*(x13-x12)+y22*(x11-x13)+y23*(x12-x11))/dtx;
    dj1 = (x22-x23)/dtx; dj2 = (x13-x12)/dtx;
    w = j11*j11+j12*j12+j21*j21+j22*j22;
    w1 = 2.0*(dj1*j11+dj2*j12); w2 = 2.0*(dj1*j21+dj2*j22);
    dt = j22*j11-j12*j21;
    dt1 = j22*dj1-dj2*j21; dt2 = dj2*j11-j12*dj1;
    t1 = w+2.0*dt; t2 = w-2.0*dt; t3 = t1*t2;
    if (t3>1e-6){
        t3 = sqrt(t3); tm = 2.0*t3;
        tm1 = (t2*(w1+2*dt1)+t1*(w1-2*dt1))/t3;
        tm2 = (t2*(w2+2*dt2)+t1*(w2-2*dt2))/t3;
    }
    else { tm = 0.0; tm1 = 1.0; tm2 = 1.0; }
    s1 = w *0.50 + tm *0.25; s2 = w *0.50 - tm *0.25;
    d1s1 = w1*0.50 + tm1*0.25; d1s2 = w1*0.50 - tm1*0.25;
    d2s1 = w2*0.50 + tm2*0.25; d2s2 = w2*0.50 - tm2*0.25;
    t1 = log(s1); t2 = log(s2);
    t3 = t1/s1; t4 = t2/s2;
    dtx = lambda*fabs(dtx)*0.5;
    t1 = 0.25*(t1*t1 + t2*t2 );
    t2 = 0.50*(t3*d1s1 + t4*d1s2);
    t3 = 0.50*(t3*d2s1 + t4*d2s2);
    *h = dtx*t1*(dt+1);
    *dh1 = dtx*(t1*dt1 + t2*(dt+1));
    *dh2 = dtx*(t1*dt2 + t3*(dt+1));
}

```

FIG. 4. C code for computing the rate of change of the prior potential (h) with respect to changes in y_{11} and y_{21} . The arguments passed to the routine are the original coordinates at the vertexes of the triangle. These are (x_{11}, x_{21}) , (x_{12}, x_{22}) , and (x_{13}, x_{23}) , and they map to (y_{11}, y_{21}) , (y_{12}, y_{22}) , and (y_{13}, y_{23}) respectively. The values returned are h , dh_1 , and dh_2 , and these correspond to the potential of the deformation of the triangular patch and the rate of change of the potential with respect to changes in y_{11} and y_{21} . Note that the singular values of a 2×2 matrix \mathbf{J} are $((w + ((w + 2d)(w - 2d))^{1/2})/2)^{1/2}$ and $((w - ((w + 2d)(w - 2d))^{1/2})/2)^{1/2}$, where $w = j_{11}^2 + j_{12}^2 + j_{21}^2 + j_{22}^2$ and $d = j_{22}j_{11} - j_{12}j_{21}$.

ter the images to an unusually shaped brain, the deformations would need to be greater and therefore more susceptible to error. The ideal solution would be to use a “canonical” or average shaped brain as a template. The concept of a canonical brain we shall use

here is one that other images can be matched to with the least amount of “energy” (Gibbs potential).

An iterative procedure has been used to generate such an average from a slice from the MR images of 29 subjects. The images were first registered to the same

stereotactic space using a 12-parameter affine registration (Ashburner *et al.*, 1997), and the same slice was extracted from all the registered images. The first step is to average the intensities of the unwarped images to create an initial estimate for the canonical template. This model also incorporates a spatially varying σ^2 , and this is computed from the residual variance at each pixel. Then an iteration of the registration procedure is used to bring each of the images slightly closer to the shape of the template. The warped brains are averaged again, and this average is used as the template for the next round of the procedure. The residual variance is also recomputed and used as the new estimate for the nonstationary σ^2 . This is continued until the algorithm converges and the Gibbs potential of the system is minimized. After many iterations, the end result is a template that satisfies our criterion. The results of this procedure are shown in Fig. 8. We suggest that this may be a very useful and principled technique to generate templates and “canonical” references.

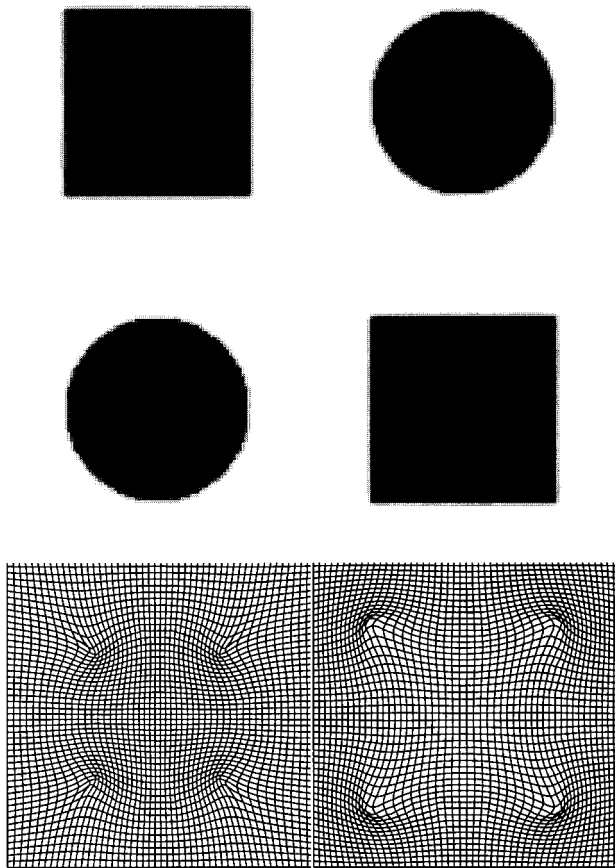


FIG. 5. Demonstration using simulated data. Top left, original square; top right, original circle; middle left, square deformed to match the circle; middle right, circle deformed to match the square; bottom left, deformation field applied to the circle in order to warp it to match the square. The deformation field shows where data should be resampled from in the original image in order to generate the warped version. Bottom right, deformation field required to deform the square to the circle.

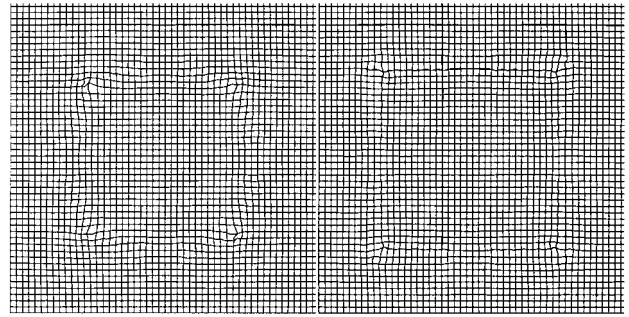


FIG. 6. Demonstration of the reversibility of the deformations obtained by combining forward and reverse deformations. Left, deformation field that warps from the circle to the square and back to the circle; right, deformation field that warps from the square to the circle and back to the square.

4. DISCUSSION

In summary, we have presented a novel high-dimensional image registration method that uses Bayesian statistics to achieve a *maximum a posteriori* estimate of the required deformations. The priors used by the model are symmetric, in that the probability distribution of the deformations is considered identical to those for the inverses. Although the method has only been demonstrated in two dimensions, the principles can be extended to three-dimensional registration. The two-dimensional prototype was developed, mostly because it makes the evaluation of the basic principles much more straightforward. Two-dimensional deformation fields can be visualized much more easily than three-dimensional deformations and also require less execution time than the equivalent implementation in three dimensions. Many hundreds of iterations are normally required for the algorithm to converge to the final solution, so a three-dimensional version would take several hours to run.

Like all image registration methods that rely on conventional optimization techniques, this method can still get caught in a local minimum. The chance of this occurring is reduced by assigning more regularization to the early stages of the registration (when the matching may also be performed using smoother images). This results in the global brain shape being estimated first. By gradually decreasing the amount of regularization (and smoothing), the algorithm can estimate higher frequency deformations. However, this “coarse to fine” approach will still not guarantee the globally optimum solution, especially for features on the cortical surface. For example, it would be very easy for a sulcus in one brain to be matched to an incorrect sulcus in another. In order to achieve a globally optimal solution, some form of robust stochastic method would be required. Unfortunately, these methods are very computationally intensive and are therefore not really practical for extremely high-dimensional image registration methods. However, if sulci and gyri can be easily labeled from the

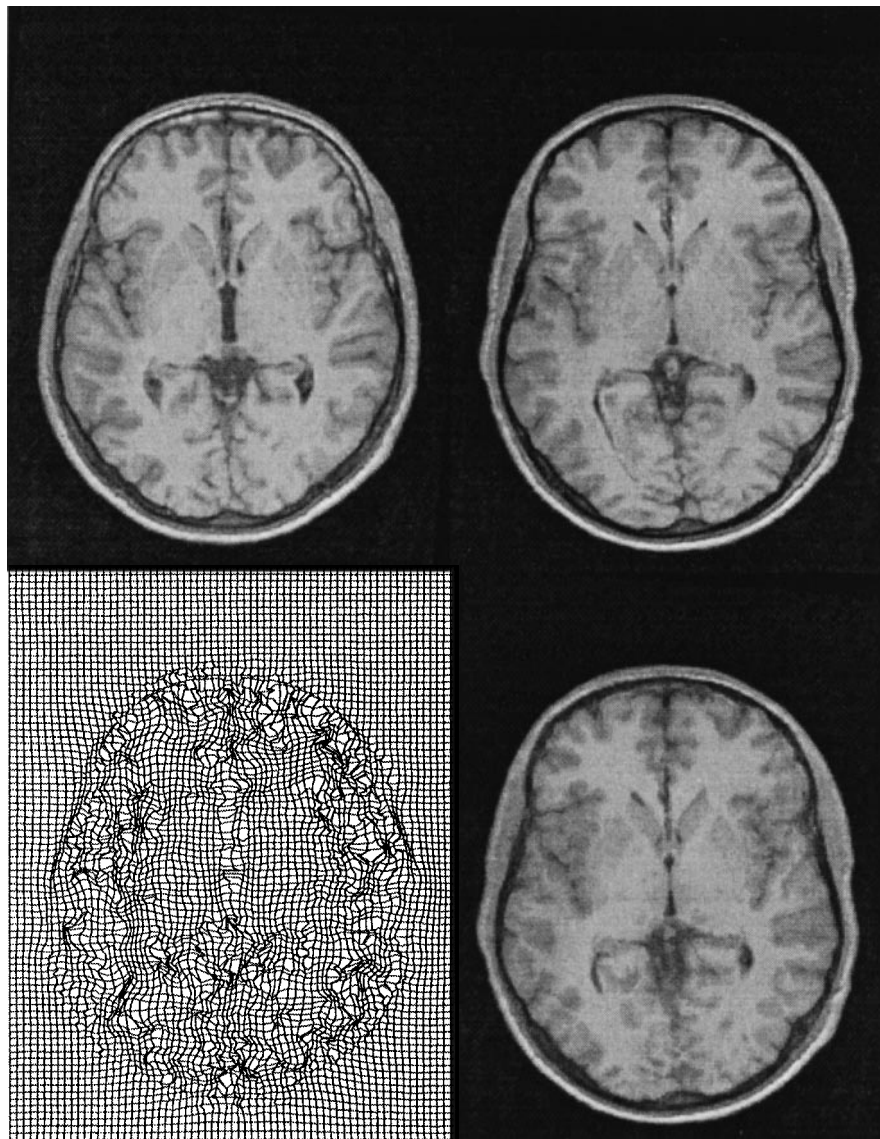


FIG. 7. Top left, the unwarped object image; top right, the template image; bottom left, the deformation field applied to the object image in order to warp it to match the template image; bottom right, the object image after warping to match the template.

brain images, then robust methods can be applied in order to match the labeled features. Robust methods become more practical when the amount of information is reduced to a few key features. The robust match can then be used to bias the high-dimensional registration (Joshi *et al.*, 1995; Thompson and Toga, 1996; Davatzikos, 1996), therefore increasing the likelihood of obtaining the globally optimum solution.

The emphasis of this paper is about how priors used for image registration should be symmetric. Unlike Bayesian models that use linear priors such as membrane energy (Amit *et al.*, 1991; Gee *et al.*, 1997), linear elasticity (Miller *et al.*, 1993), or bending energy (Bookstein, 1989, 1997), the Jacobians of the deformation field are increasingly penalized as they approach singularity. In theory, by decreasing the weight of the priors,

large distance deformations can be achieved that still preserve the image topology. Unfortunately, one of the disadvantages of using weaker priors (smaller λ) is that the search space is increased because there are effectively a greater number of independent parameters in the model. A larger search space means that there are more potential local minima, so it becomes more important to provide better starting estimates if large distance deformations are to be estimated.

The priors are currently spatially invariant. Future work may involve defining nonstationary priors. Because structural variability is often greater in certain directions (Thompson *et al.*, 1996), some form of a tensor field describing normal variability in each direction may be appropriate. A data representation of this form, together with a canonical brain template and

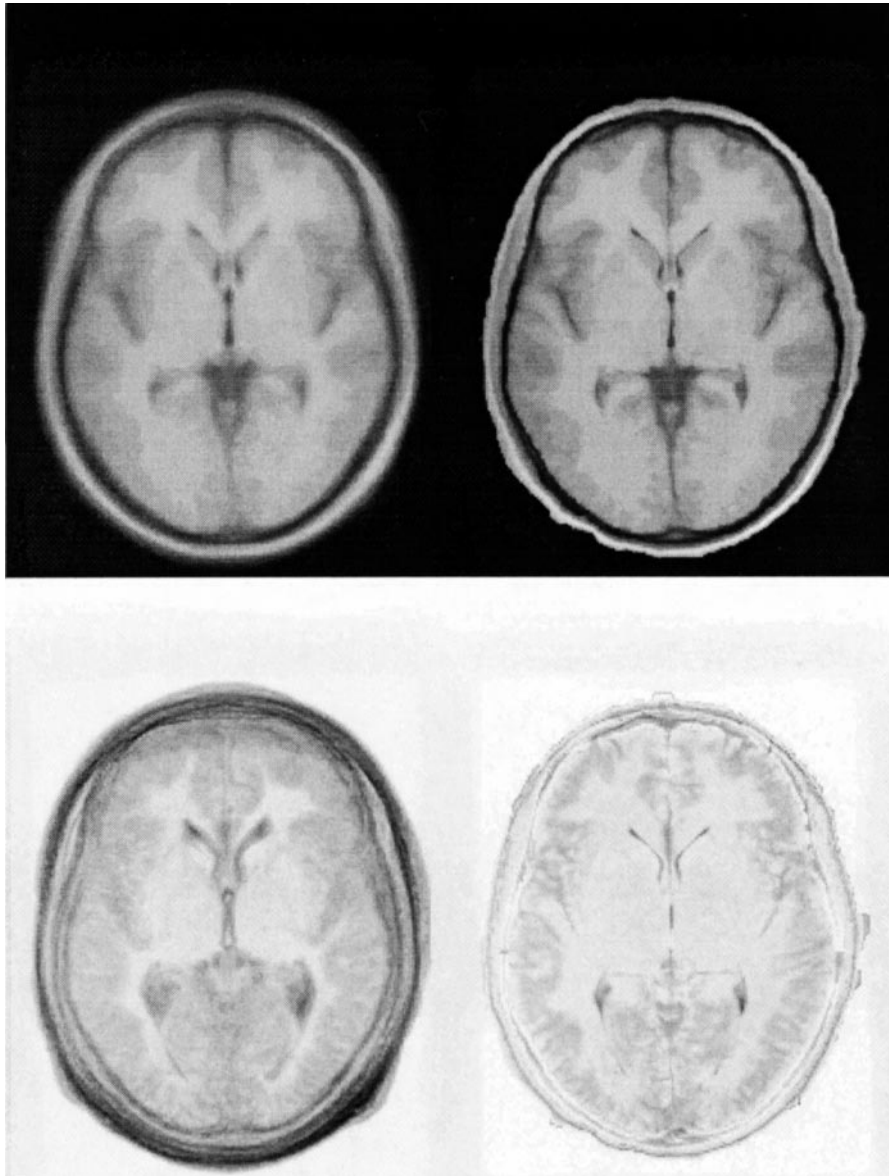


FIG. 8. Top left, the average of the MR images of 29 subjects registered together using a 12-parameter affine registration; top right, the average (both shape and intensity) of the same 29 MR images after registering together using the method described in Section 3.3; bottom left, the standard deviations of the affine registered MR images; bottom right, the standard deviations of the MR images after the registration using the current method. The images of standard deviation are shown using the same intensity scaling.

associated error variance image, would allow anatomical comparisons to be made against the normal population. This would facilitate the identification of abnormal morphological features by locating higher than normal values in the spatial distribution of the Gibbs potentials. If successful, this should be a significant contribution to the field of radiology and computational neuroanatomy in general.

ACKNOWLEDGMENT

This work was supported by the Wellcome Trust.

REFERENCES

- Amit, Y., Grenander, U., and Piccioni, M. 1991. Structural image restoration through deformable templates. *J. Am. Stat. Assoc.* **86**:376–387.
- Ashburner, J., and Friston, K. J. 1999. Nonlinear spatial normalization using basis functions. *Hum. Brain Map.* **7**(4): in press.
- Ashburner, J., Neelin, P., Collins, D. L., Evans, A. C., and Friston, K. J. 1997. Incorporating prior knowledge into image registration. *NeuroImage* **6**:344–352.
- Ashburner, J., Hutton, C., Frackowiak, R. S. J., Johnsrude, I., Price, C., and Friston, K. J. 1998. Identifying global anatomical differences: Deformation-based morphometry. *Hum. Brain Map.* **6**(5): 348–357.

- Bookstein, F. L. 1989. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(6):567–585.
- Bookstein, F. L. 1997. Landmark methods for forms without landmarks: Morphometrics of group differences in outline shape. *Med. Image Anal.* **1**(3):225–243.
- Christensen, G. E. 1994. *Deformable Shape Models for Anatomy*, doctoral thesis. Sever Institute of Washington University.
- Christensen, G. E., Rabbitt, R. D., and Miller, M. I. 1994. 3D brain mapping using a deformable neuroanatomy. *Phys. Med. Biol.* **39**:609–618.
- Christensen, G. E., Rabbitt, R. D., and Miller, M. I. 1996. Deformable templates using large deformation kinematics. *IEEE Trans. Image Process.* **5**:1435–1447.
- Davatzikos, C. 1996. Spatial normalization of 3D images using deformable models. *J. Comput. Assist. Tomogr.* **20**(4):656–665.
- Friston, K. J., Ashburner, J., Frith, C. D., Poline, J.-B., Heather, J. D., and Frackowiak, R. S. J. 1995. Spatial registration and normalization of images. *Hum. Brain Map.* **2**:165–189.
- Gee, J. C., Le Briquer, L., Barillot, C., Haynor, D. R., and Bajcsy, R. 1995a. Bayesian approach to the brain image matching problem. In *Proceedings, SPIE Medical Imaging 1995: Image Processing* (M. H. Loew, Ed.), Vol. 2434, pp. 145–156. SPIE, Bellingham, WA.
- Gee, J. C., Le Briquer, L., and Barillot, C. 1995b. Probabilistic matching of brain images. In *Information Processing in Medical Imaging* (Y. Bizais, C. Barillot, and R. Di Paola, Eds.), pp. 113–125. Kluwer, Dordrecht.
- Gee, J. C., Haynor, D. R., Le Briquer, L., and Bajcsy, R. K. 1997. Advances in elastic matching theory and its implementation. In *CVRMed-MRCAS'97* (P. Cinquin, R. Kikinis, and S. Lavalée, Eds.). Springer-Verlag, Heidelberg.
- Joshi, S. C., Miller, M. I., Christensen, G. E., Banerjee, A., Coogan, T. A., and Grenander, U. 1995. Hierarchical brain mapping via a generalized dirichlet solutions for mapping brain manifolds. In *Proceedings, SPIE International Symposium on Optical Science, Engineering and Instrumentation*.
- Miller, M. I., Christensen, G. E., Amit, Y., and Grenander, U. 1993. Mathematical textbook of deformable neuroanatomies. *Proc. Natl. Acad. Sci. USA* **90**:11944–11948.
- Talairach, J., and Tournoux, P. 1988. *Coplanar Stereotaxic Atlas of the Human Brain*. Thieme, New York.
- Thompson, P. M., and Toga, A. W. 1996. Visualization and mapping of anatomic abnormalities using a probabilistic brain atlas based on random fluid transformations. In *Proceedings of the International Conference on Visualization in Biomedical Computing*, pp. 383–392.
- Thompson, P. M., Schwartz, C., Lin, R. T., Khan, A. A., and Toga, A. W. 1996. 3D statistical analysis of sulcal variability in the human brain. *J. Neurosci.* **16**(13):4261–4274.