Chapter 4

High-Dimensional Image Warping

John Ashburner & Karl J. Friston
The Wellcome Dept. of Imaging Neuroscience,
12 Queen Square, London WC1N 3BG, UK.

Contents

Abstract

This chapter is about warping brain images of different subjects to the same stereotaxic space. However, unlike Chapter 3, this method uses thousands or millions of parameters, so is potentially able to obtain much more precision. 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 prior knowledge about the variability of the estimated parameters. In this approach it is assumed 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}. The penalty function of choice is based upon the logs of the singular values of the Jacobian matrices having normal distributions, which 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. Further consistency is achieved by registering images to their "averages", where this average is one of both intensity and shape.

4.1 Introduction

Two brain images from the same subject can be co-registered using a six-parameter rigid body transformation, which simply describes the relative position and orientation of the images. However, for matching brain images of different subjects (or the brain of the same subject that may have changed shape over time [9]), it is necessary to estimate a deformation field that also describes the relative shapes of the images. The previous chapter described a method of registering images of different subjects in order to match the overall shapes. However, many more parameters are required to describe the shape of a brain precisely, and estimating these can be very prone to error. This error can be reduced by ensuring that the deformation fields are internally consistent [4]. For example, suppose a deformation that matches brain **f** to brain **g** is estimated, and also a deformation that matches brain **g** to brain **f**. If one deformation field is not the inverse of the other, then at least one of them has to be wrong.

Often, the prior probability distributions used by Bayesian registration schemes are linear (see Section **??**), and include minimizing the *membrane energy* of the deformation field [1, 10], the *bending energy* [3] or the *linear-elastic energy* [17]. None of these linear penalties are symmetric, and they do not explicitly preserve the topology[1] of the warped images.

An alternative, to using a Bayesian scheme incorporating some form of elastic prior, could be to use a viscous fluid model [5, 6] to estimate 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 chapter utilizes statistical rules. Unlikely deformations are

[1] The word "topology" is used in the same sense as in "Topological Properties of Smooth Anatomical Maps" [7]. If spatial transformations are not one-to-one and continuous, then the topological properties of different structures can change.

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 the chapter is divided into three main sections. The methods section describes the Bayesian principles behind the registration, which is essentially an optimization procedure that simultaneously minimizes a likelihood function (i.e., the sum of squared differences between the images), and a penalty function that relates to the prior probability of obtaining the deformations. A number of examples of registered images are provided in the next section. The final section discusses the validity of the method, and includes a number of suggestions for future work.

4.2 Methods

Registering one image volume to another involves estimating a vector field (deformation field) that maps from co-ordinates of one image to those of the other. In this work, one image (the template image) is considered as fixed, and a mapping from this image to the second image (the source image) is estimated. The intensity of the ith voxel of the template is denoted by $g(\mathbf{x}_i)$, where $\mathbf{x}_i$ is a vector describing the co-ordinates of the voxel. The deformation field spanning the domain of the template is denoted by $\mathbf{y}_i$ (or $\mathbf{y}(\mathbf{x}_i)$) at each point, and the intensity of the source image at this mapped point by $f(\mathbf{y}_i)$. The source image is transformed to match the template by re-sampling it at the mapped co-ordinates.

This section begins by describing how the deformation fields are parameterized as piecewise affine transformations within a finite element mesh. The registration is achieved by matching the images while simultaneously trying to maximize the smoothness of the deformations. Bayesian statistics are used to incorporate this smoothness into the registration, and a method of optimization is presented for finding the *maximum a posteriori* (MAP) estimate of the parameters. Suitable forms for the smoothness priors are presented. The first of these is the ideal form, which for practical reasons has only been developed for registration in two dimensions. The second form is an approximation to the ideal, and it has been developed for both two and three dimensional image registration.

4.2.1 Bayesian Framework

This approach to image registration estimates the required spatial transformation at every voxel, and therefore requires many parameters. For example, to register two volumes of size $256 \times 256 \times 108$ voxels, needs 21,233,664 parameters. The number of parameters describing the transformations exceeds the number of voxels in the data. Because of this, it is essential that the effective degrees of freedom are reduced by imposing priors or constraints on the registration. As in the previous chapter, Bayesian statistics are used to incorporate a prior probability distribution into the warping model.

Bayes' theorem can be expressed as (see Eqn. **??**):

$$p(\mathbf{Y}|\mathbf{b}) \propto p(\mathbf{b}|\mathbf{Y})p(\mathbf{Y}) \tag{4.1}$$

where $p(\mathbf{Y})$ is the *a priori* probability of parameters $\mathbf{Y}$, $p(\mathbf{b}|\mathbf{Y})$ is the likelihood of observing data $\mathbf{b}$ given parameters $\mathbf{Y}$, and $p(\mathbf{Y}|\mathbf{b})$ is the *a posteriori* probability of $\mathbf{Y}$ 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 determined here is the *maximum a posteriori* (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:

$$p(\mathbf{Y}) \propto e^{-H(\mathbf{Y})} \tag{4.2}$$

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:

$$H(\mathbf{Y}|\mathbf{b}) = H(\mathbf{b}|\mathbf{Y}) + H(\mathbf{Y}) + c \tag{4.3}$$

where c is a constant. The registration is therefore a nonlinear optimization problem, whereby the cost function to be minimized is the sum of the likelihood potential ($H(\mathbf{b}|\mathbf{Y})$) and the prior potential ($H(\mathbf{Y})$). These potentials are now discussed in detail.

4.2.2 Likelihood Potentials

The registration matches a source image ($\mathbf{f}$) to a template image ($\mathbf{g}$). The 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:

$$H(\mathbf{b}|\mathbf{Y}) = \frac{1}{2\sigma^2} \sum_{i=1}^{I} \left(f(\mathbf{y}_i) - g(\mathbf{x}_i)\right)^2 \tag{4.4}$$

where $g(\mathbf{x}_i)$ is the ith pixel value of $\mathbf{g}$ and $f(\mathbf{y}_i)$ is the corresponding pixel value of $\mathbf{f}$.

In this model, 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[2]. For the early iterations, σ^2 has a higher value, placing more weight on the priors, so the deformations are smoother. When close to the final solution, σ^2 has decreased, and the algorithm is able to compute more detailed deformations.

The potential is computed by sampling I discrete points within the template image, and equivalent points within the source image are sampled using tri-linear interpolation. Gradients of the tri-linearly interpolated source image are required for the registration, and these are computed using a finite difference method.

4.2.3 Prior Potentials - 2D

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. There should be symmetry in the priors, in order to minimize these errors. In addition to considering the deforming forces that warp image $\mathbf{f}$ to match image $\mathbf{g}$, the forces mediating the inverse of the deformation also need to be considered. In order to achieve this symmetry, the fundamental assumption is made 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 source 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 in 2D, the pixels of the template image ($\mathbf{g}$) are considered as being on a regular grid, with unit spacing between them. A triangular mesh connects the centers

[2]Note that this is an Empirical Bayesian approach as the variance component is derived from the data itself, and that the variance estimate is just an approximation because the degrees of freedom are not properly computed.

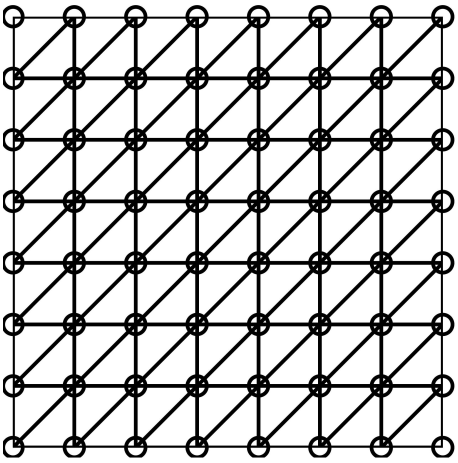

Figure 4.1: For the two dimensional registration problem, the area of the template image ($\mathbf{g}$) is divided into a triangular mesh where the nodes are centered on the pixels.

of each pixel (as shown in Figure 4.1). Within each triangle, there is assumed to be a uniform affine mapping between the images. If the co-ordinates 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 co-ordinates (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:

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ 1 & 1 & 1 \end{bmatrix}^{-1} \tag{4.5}$$

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

$$\mathbf{J} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \tag{4.6}$$

The penalty for distorting each of these triangles is derived from its Jacobian matrix. By using singular value decomposition, $\mathbf{J}$ can be decomposed into two unitary matrices ($\mathbf{U}$ and $\mathbf{V}$) and a diagonal matrix ($\mathbf{S}$), such that $\mathbf{J} = \mathbf{USV}^T$. The unitary matrices simply represent rotations [3], 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 $\mathbf{J}$ ($|\mathbf{J}|$) 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 Jacobian matrices 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 $log(|\mathbf{J}|)^2$ (or even $|\mathbf{J}| + |\mathbf{J}|^{-1} - 2$) 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 logs of the diagonal elements of $\mathbf{S}$ being drawn from a normal distribution. The penalty per unit area is therefore $\lambda log(s_{11})^2 + \lambda log(s_{22})^2$, where λ is a "regularization parameter" [4]. If the

[3] 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.

[4] Short of determining λ using a large number of "true" deformations, it is assigned some suitable value that facilitates rapid convergence to reasonable solutions.

logs of each diagonal element of $\mathbf{S}$ are normally distributed, then $\log |\mathbf{J}|$ is also normally distributed as $log(|\mathbf{J}|) \equiv log(s_{11}) + log(s_{22})$ 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 $|\mathbf{J}|/2$ pixels when mapped to the space of image $\mathbf{f}$. The total area affected by the penalty in both the template and source images is therefore $(1 + |\mathbf{J}|)/2$, so the penalty for each triangle becomes:

$$h = \lambda(1 + |\mathbf{J}|) \left(log(s_{11})^2 + log(s_{22})^2\right) /2 \tag{4.7}$$

Examples of these penalties in terms of two dimensional probability functions are illustrated in Figure 4.2. The prior potential over the whole image is based on the sum of the potentials for each of the I triangles:

$$H(\mathbf{Y}) = \sum_{i=1}^{I} h_i \tag{4.8}$$

For simplicity, in the current description, the fact that the images have boundaries is ignored. In practice, the boundaries are fixed so that the deformation at the edges is always zero.

4.2.4 Prior Potentials - 3D

Section 4.2.3 described deformation fields consisting of a patchwork of triangles. The situation is more complex when working with three dimensional deformations. For this case, the volume of the template image is divided into a mesh of tetrahedra, where the vertices of the tetrahedra are centered on the voxels. This is achieved by considering groups of eight voxels as little cubes. Each of these cubes is divided into five tetrahedra: one central one having a 1/3 of the cube's volume, and four outer ones, each having 1/6 of the cube's volume [12]. There are two possible ways of dividing a cube into five tetrahedra. Alternating between the two conformations in a three dimensional checkerboard pattern ensures that the whole template volume is uniformly covered (see Figure 4.3). A deformation field is generated by treating the vertices of the tetrahedra as control points. These points are moved iteratively until the best match is achieved. The deformations are constrained to be locally one-to-one by ensuring that a tetrahedron never occupies any of the same volume as its neighbors. When a deformation is one-to-one, it is possible to compute its inverse (Section 4.2.6).

The algorithm can use one of two possible boundary conditions. The simplest is when the vertices of tetrahedra that lie on the boundary remain fixed in their original positions (Dirichlet boundary condition). Providing that the initial starting estimate for the deformations is globally one-to-one, then the final deformation field will also satisfy this constraint [7]. The other boundary condition involves allowing the vertices on the surface to move freely (analogous to the Neumann boundary condition). It is possible for the global one-to-one constraints to be broken in this case, as the volumes of non-neighboring tetrahedra can now overlap. The examples shown later use the free boundary condition.

Within each tetrahedron, the deformation is considered as a uniform affine transformation from which the Jacobian matrix is extracted in a similar way to that described in Section 4.2.3. A penalty is applied to each of the tetrahedra that constitute the volume covered. For each tetrahedron, it is the product of a penalty per unit volume, and the total volume affected by the tetrahedron. The affected volume is the volume of the undeformed tetrahedron in the template image, plus the volume that the deformed tetrahedron occupies within the source image ($v\,(1 + |\mathbf{J}|)$, where v is the volume of the undeformed tetrahedron).

A good penalty against deforming each tetrahedron is based on the logs of the singular values of the Jacobian matrices at every point in the deformation being drawn from a normal distribution.

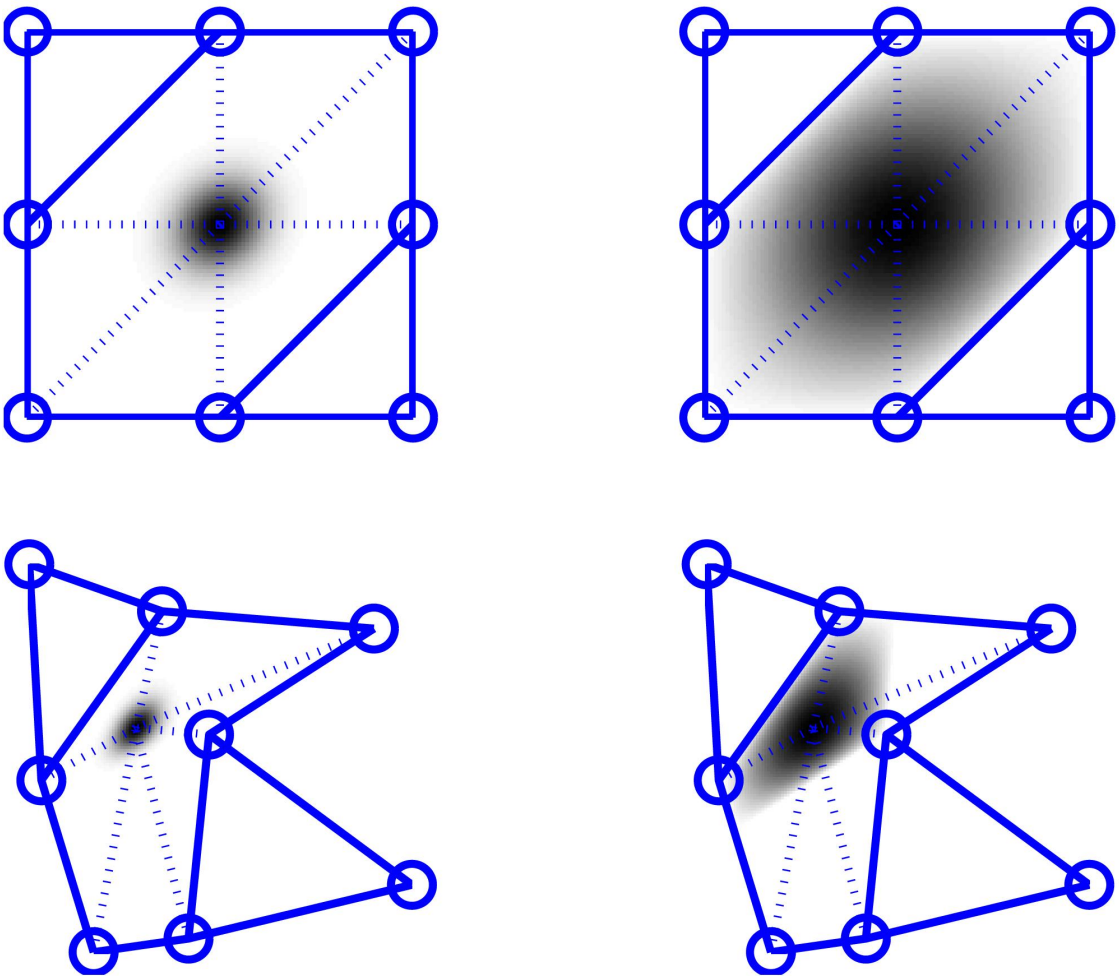

Figure 4.2: Probability density functions relating to the position of the center node, assuming that all other voxels remain at fixed locations on a regular grid. *Left column*: using heavy regularization $\lambda = 10$. *Right column*: using light regularization $\lambda = 1$. *Above*: on a regular grid. *Below*: on an irregular grid. The probabilities are defined as the exponent of the negative cost function (as in Eqn. 4.2), where the cost function is derived from the Jacobian matrices of the triangles that have the node as a vertex. The cost function is based on the sum of squares of the logs of the singular values. The dotted lines show the position of the node in the conformation with the lowest cost function.

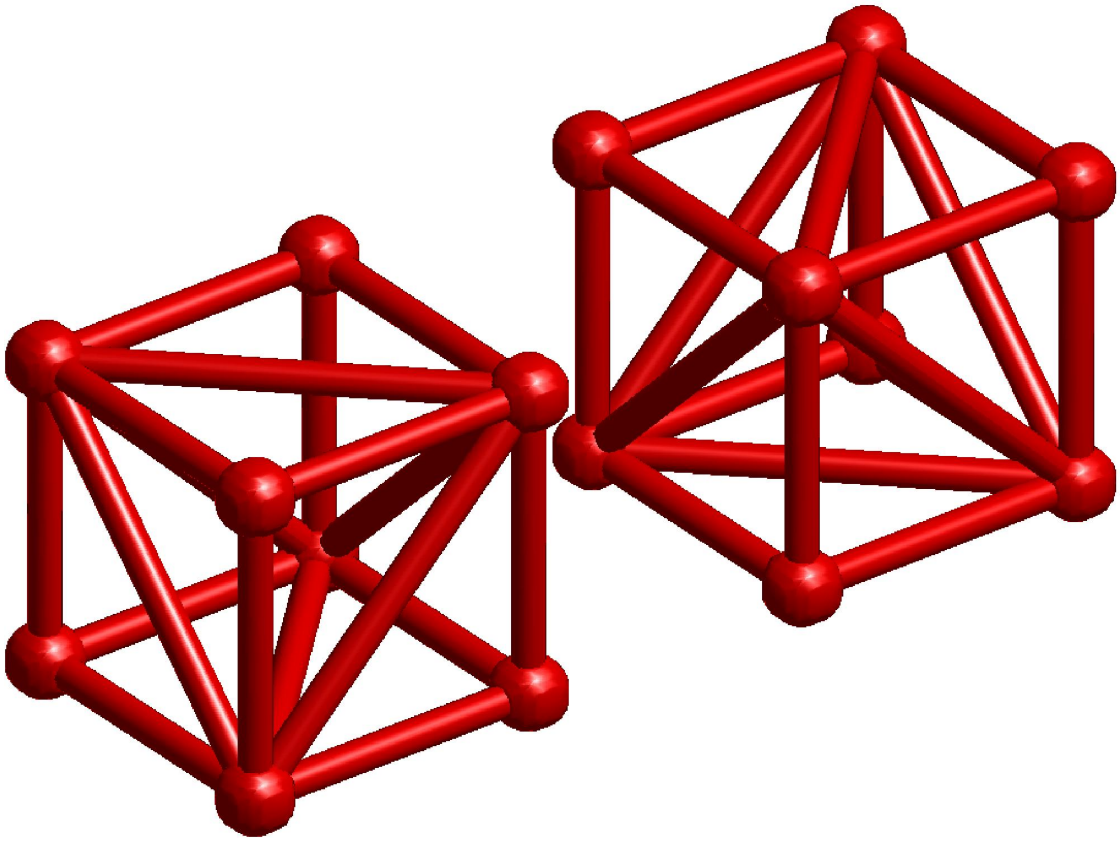

Figure 4.3: The volume of the template image is divided into a mesh of irregular tetrahedra, where the vertices of the tetrahedra are centered on the voxels. Groups of eight voxels are considered as little cubes. The volume of each cube is divided into five tetrahedra, in one of the two possible arrangements shown here. A face of a cube that is divided according to one arrangement, opposes with the face of a cube that has been divided the other way. Because of this, it is necessary to arrange the two conformations in a three dimensional checkerboard pattern.

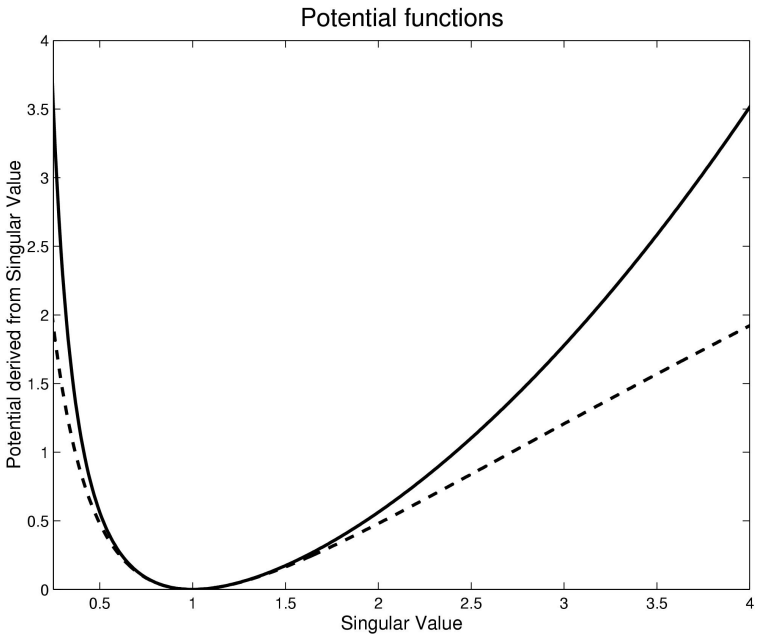

Figure 4.4: A comparison of the different cost functions. The dotted line shows the potential estimated from $log(s_{ii})^2$, where s_{ii} is the ith singular value of a Jacobian matrix. The solid line shows the new potential, which is based on $(s_{ii}^2 + s_{ii}^{-2} - 2)/4$. For singular values very close to one, the potentials are almost identical.

The use of conventional methods for computing the SVD of a 3×3 matrix is currently too slow to be used within an image registration procedure.

Using the SVD regularization, the penalty per unit volume is $\sum_{i=1}^{3} log(s_{ii})^2$, where s_{ii} is the ith singular value of the Jacobian matrix. This function is equivalent to $\sum_{i=1}^{3} log(s_{ii}^2)^2/4$. By using an approximation that $log(x)^2 \simeq x + 1/x - 2$ for values of x very close to one, we now have the function $\sum_{i=1}^{3}(s_{ii}^2 + 1/s_{ii}^2 - 2)/4$. This function is relatively simple to evaluate, because the sum of squares of the singular values of a matrix is equivalent to the sum of squares of the individual matrix elements. This derives from the facts that the trace of a matrix is equal to the sum of its eigenvalues, and the eigenvalues of $\mathbf{J}^T\mathbf{J}$ are the squares of the singular values of $\mathbf{J}$. The trace of $\mathbf{J}^T\mathbf{J}$ is equivalent to the sum of squares of the individual elements of $\mathbf{J}$. Similarly, the sum of squares of the reciprocals of the singular values is identical to the sum of squares of the elements of the inverse matrix. The singular values of the matrix need not be calculated, and there is no longer a need to call the log function (which is relatively slow to compute). The penalty for each of the tetrahedra is now:

$$h = \lambda v \left(1 + |\mathbf{J}|\right) tr\left(\mathbf{J}^T\mathbf{J} + (\mathbf{J}^{-1})^T\mathbf{J}^{-1} - 2\mathbf{I}\right)/4 \tag{4.9}$$

where tr is the trace operation, $\mathbf{I}$ is a 3×3 identity matrix, v is the volume of the undeformed tetrahedron (either $1/6$ or $1/3$), and λ is a regularization constant. The prior potential for the whole image is the sum of these penalty functions over all tetrahedra. Figure 4.4 shows a comparison of the potential based on the original $(log(s_{ii}))^2$ cost function, and the potential based on $(s_{ii}^2 + s_{ii}^{-2} - 2)/4$.

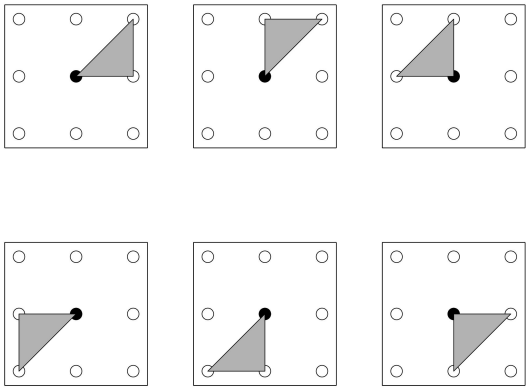

Figure 4.5: The six neighboring triangles whose Jacobian matrices are influenced by translating the central point.

4.2.5 The Optimization Algorithm

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 tiny adjustments such that the posterior potential is decreased. In each iteration, the positions of the control points (nodes) are updated *in situ*, by sequentially scanning through the template volume. During one iteration of the three dimensional registration, the looping may work from inferior to superior (most slowly), posterior to anterior, and left to right (fastest). In the next iteration, the order of the updating is reversed (superior to inferior, anterior to posterior, and right to left). This alternating sequence is continued until there is no longer a significant reduction to the posterior potential, or for a fixed number of iterations.

Each iteration of the optimization involves determining the rate of change of the posterior potential with respect to tiny changes in each element of $\mathbf{Y}$. For the nth iteration, the estimates for the ith element of $\mathbf{Y}$ are modified according to:

$$\mathbf{y}_i^{(n+1)} = \mathbf{y}_i^{(n)} - \epsilon \frac{\partial H(\mathbf{Y}|\mathbf{b})}{\partial \mathbf{y}_i} = \mathbf{y}_i^{(n)} - \epsilon \left(\frac{\partial H(\mathbf{b}|\mathbf{Y})}{\partial \mathbf{y}_i} + \frac{\partial H(\mathbf{Y})}{\partial \mathbf{y}_i} \right) \tag{4.10}$$

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 $\mathbf{y}_i$:

$$\frac{\partial H(\mathbf{b}|\mathbf{Y})}{\partial \mathbf{y}_i} = \frac{\partial \left(f(\mathbf{y}_i) - g(\mathbf{x}_i) \right)^2 / (2\sigma^2)}{\partial \mathbf{y}_i} = \frac{\left(f(\mathbf{y}_i) - g(\mathbf{x}_i) \right)}{\sigma^2} \frac{\partial f(\mathbf{y}_i)}{\partial \mathbf{y}_i} \tag{4.11}$$

where σ^2 is estimated as described in Section 4.2.2. In the updating, each node is moved along the direction that most rapidly decreases the *a posteriori* potential (a gradient descent method).

For the two dimensional registration, $\partial H(\mathbf{Y})/\partial \mathbf{y}_i$ is dependent upon changes to the Jacobian matrices of the six adjacent triangles shown in Figure 4.5. Because the mathematics of computing these partial derivatives is algebraicly dense, a C subroutine is provided in Figure 4.6 that will compute the derivatives for a single triangular patch.

For the three dimensional case, moving a node in the mesh influences the Jacobian matrices of the tetrahedra that have a vertex at that node, so the rate of change of the posterior potential is equal to the rate of change of the likelihood plus the rate of change of the prior potentials

```
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));
}
```

Figure 4.6: C code for computing the rate of change of the prior potential (h) with respect to changes in y11 and y21. The arguments passed to the routine are the original co-ordinates at the vertices of the triangle. These are (x11,x21), (x12,x22) and (x13,x23), and they map to (y11,y21), (y12,y22) and (y13,y23) respectively. The values returned are h, dh1 and dh2, 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 y11 and y21. 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 $d = j_{22}j_{11} - j_{12}j_{21}$.

from these local tetrahedra. Approximately half of the nodes form a vertex in 8 neighboring tetrahedra, whereas the other half are vertices of 24 tetrahedra. The rate of change of the penalty function for each tetrahedron with respect to changes in position of one of the vertices is required. The Matlab 5.3 Symbolic Toolbox (The MathWorks, Natick, Mass., USA) was used to derive expressions for analytically computing these derivatives, but these formulae are not given here. The ideas presented above assume that the voxel dimensions are isotropic, and the same for both images. Modifications to the method that are required in order to account for the more general cases are trivial, and are also not shown here.

If a node is moved too far, then the determinant of one or more of the Jacobian matrices associated with a neighboring triangle or tetrahedron may become negative. This would mean a violation of the one-to-one constraint in the mapping (since neighboring tetrahedra would occupy the same volume), so it is prevented by a bracketing procedure. The initial attempt moves the node by a small distance ϵ. If any of the determinants become negative, then the value of ϵ is halved and another attempt made to move the node the smaller distance from its original location. This continues for the node until the constraints are satisfied. A similar procedure is then repeated whereby the value of ϵ continues to be halved until the new potential is less than or equal to the previous value. By incorporating this procedure, the potential will never increase as a node is moved, therefore ensuring that the potential over the whole image will decrease with every iteration.

4.2.6 Inverting a Deformation Field

Occasionally it is desirable to compute the inverse of a deformation field. For example, if a deformation has been defined that warps brain A to match brain B, it may be useful to have the inverse of this deformation in order to warp brain B to match brain A. This section describes how to do this for three dimensional transformations. The registration method estimates a deformation field that describes a mapping from points in the template volume to those in the source volume. Each point within the template maps to exactly one point within the source image, and every point within the source maps to a point in the template. For this reason, a unique inverse of the spatial transformation exists. To invert the deformation field, it is necessary to find the mapping from the voxels in the source image to their equivalent locations in the template.

The template volume is covered by a large number of contiguous tetrahedra. Within each tetrahedron, the mapping between the images is described by an affine transformation. Inverting the transformation involves sequentially scanning through all the deformed tetrahedra to find any voxels of the source image that lie inside. The vertices of each tetrahedron are projected onto the space of the source volume, and so form an irregular tetrahedron within that volume. All voxels within the source image (over which the deformation field is defined) should fall into one of these tetrahedra. Once the voxels within a tetrahedron are identified, the mapping to the template image is achieved simply by multiplying the co-ordinates of the voxels in the source image by the inverse of the affine matrix $\mathbf{M}$ for the tetrahedron (from Section 4.2.4).

The first problem is to locate the voxels of the source image that lie within a tetrahedron, given the locations of the four vertices. This involves finding locations where the x, y and z co-ordinates assume integer values within the tetrahedral volume. First of all, the vertices of the tetrahedron are sorted into increasing z co-ordinates. Planes where z takes an integer value are identified between the first and second, the second and third, and the third and fourth vertices. Between the first and second vertices, the cross-sectional shape of a tetrahedron (where it intersects a plane where z is an integer) is triangular. The corners of the triangle are at the locations where lines connecting the first vertex to each of the other three vertices intersect the plane. Similarly, between the third and fourth vertices, the cross-section is again triangular, but this time the corners are at the intersects of the lines connecting the first, second and third vertices to the

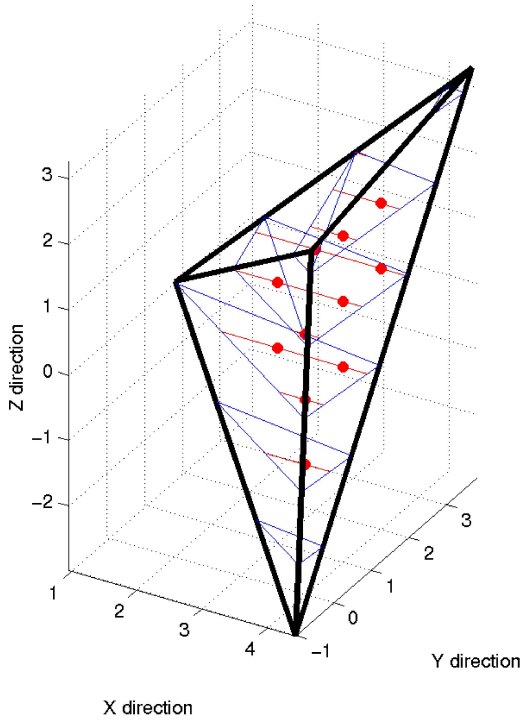

Figure 4.7: An illustration of how voxels are located within a tetrahedron.

fourth. Between the second and third vertex, the cross-section is a quadrilateral, and this can be described by two triangles. The first can be constructed from the intersects of the lines connecting vertices one to four, two to four, and two to three. The other is from the intersects of the lines connecting vertices one to four, one to three, and two to three. The problem has now been reduced to the more trivial one of finding co-ordinates within the area of each triangle for which x and y are both integer values (see Figure 4.7).

The procedure for finding points within a triangle is broken down into finding the ends of line segments in the triangle where y takes an integer value. In order to find the line segments, the corners of the triangle are sorted into increasing y co-ordinates. The triangle is divided into two smaller areas, separated by a line at the level of the second vertex. The method for identifying the ends of the line segments is similar to the one used for identifying the corners of the triangles. The voxels are then simply located by finding points on each of the lines where x is integer.

4.3 Examples

A number of sets of examples are provided in this section. The first is based on simulated data, and is designed to show the symmetric nature of the estimated warps. This is followed by examples of registering one brain to another, both in two and three dimensions. Finally, examples of several brains registered simultaneously are given, again in two and three dimensions.

4.3.1 Two Dimensional Warping Using Simulated Data

Simulated data are used to demonstrate the reversibility of the deformation fields. Two images were constructed, one of them a circle, and the other a square. The circle was warped to match the square, and the square to match the circle. No noise was added to the images, so a constant variance was assumed for all iterations. The final results of the registration are shown in Figure 4.8. In order to demonstrate the symmetry of the deformations, the two deformation fields were combined. These are shown in Figure 4.9. 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.

4.3.2 Registering Pairs of Images

Here are examples of registering pairs of images together, first in two dimensions, and then in three dimensions.

Two Dimensional Example

Approximately corresponding slices through two MR images of different subjects were registered together using the current approach. 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 8mm 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 Figure 4.10.

Three Dimensional Example

A pair of three dimensional brain images were first registered using the global registration methods described in Chapter 3, which provide a good starting point for estimating the optimum high-dimensional deformation field. It took about $15\frac{1}{2}$ hours to estimate the 21,233,664 parameters on one of the processors of a SPARC Ultra 2 (Sun Microsystems, USA). Figure 4.11 shows the two registered brain images, and the corresponding deformation fields are shown in Figure 4.12.

4.3.3 Registering to an Average

One of the themes of this chapter is about achieving internal consistency in the estimated warps. So far, only mappings between pairs of images have been discussed. When a number of images are to be registered to the same stereotaxic space, then there are more possible ways in which this can be achieved. The different routes that can be taken to achieve the same goal may not always produce consistent results [14, 20, 4]. In order to achieve more consistency in the warps, the images should all be registered to a template that is some form of average of all the individual images. A mapping between any pair of brains can then be obtained by combining the transformation from one of the brains to the template, with the inverse of the transformation that maps from the other brain to the template.

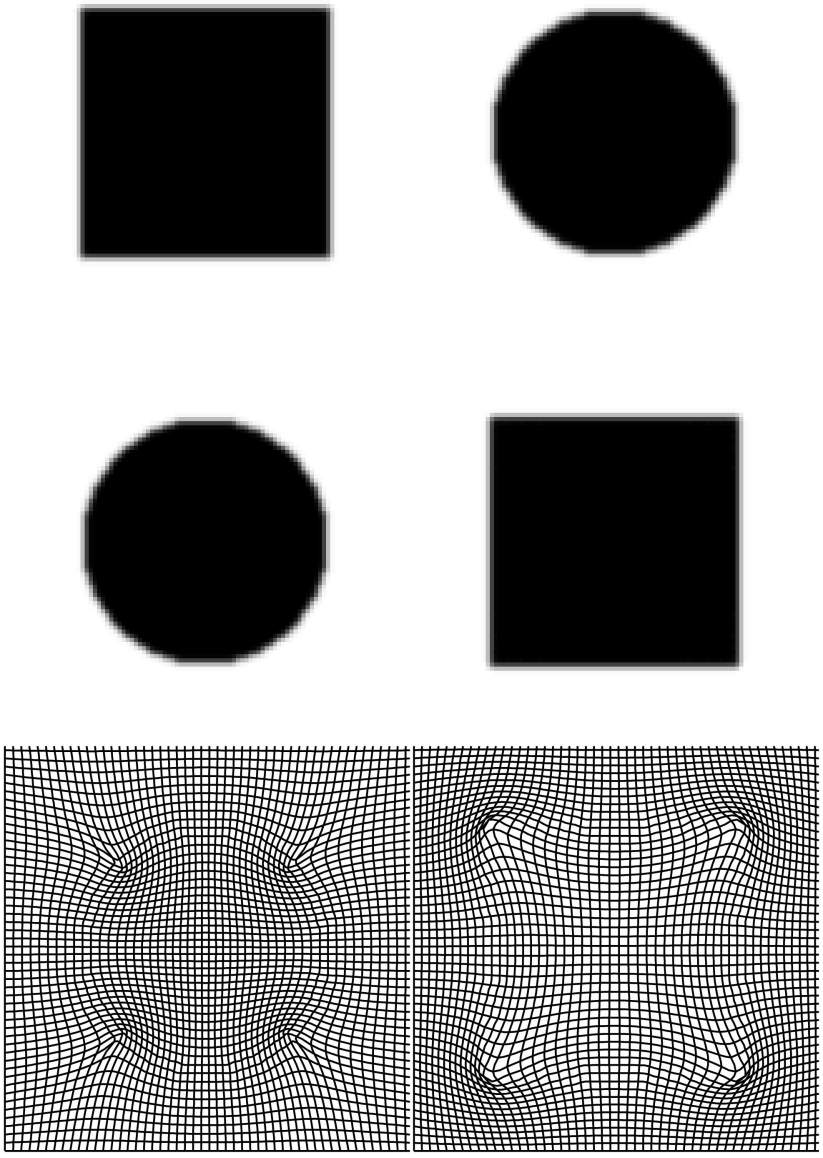

Figure 4.8: Demonstration using simulated data. *Above left*: original square. *Above right*: original circle. *Center left*: square deformed to match the circle. *Center right*: circle deformed to match the square. *Below left*: deformation field applied to the circle in order to warp it to match the square. The deformation field shows where data should be re-sampled from in the original image in order to generate the warped version. *Below right*: deformation field required to deform the square to the circle.

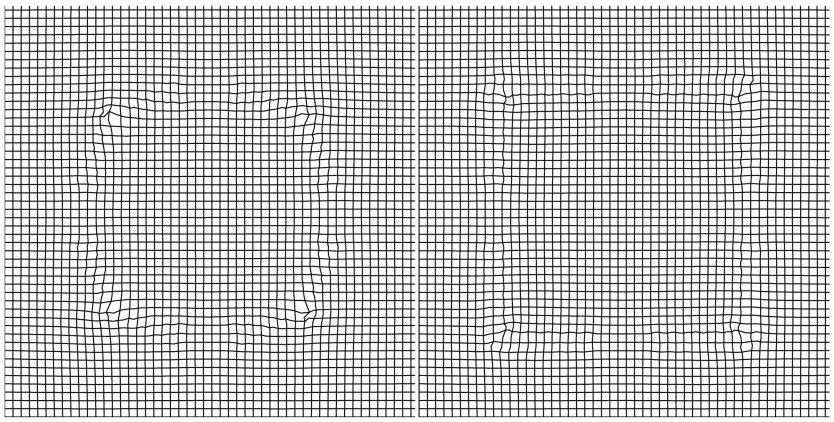

Figure 4.9: Demonstration of the reversibility of the deformations obtained by combining forward and reverse deformations. *left*: deformation field that warps from the circle to the square and back to the circle. *right*: deformation field that warps from the square to the circle and back to the square.

Two Dimensional Example

An iterative procedure was used to generate such an average from a slice from the MR images of 29 subjects. The images were first registered to the same stereotaxic space using a 12 parameter affine registration, and the same slice extracted from all the registered images. The first step involved averaging the intensities of the unwarped images to create an initial estimate for the template. Then an iteration of the registration procedure was used to bring each of the images slightly closer to the shape of the template. The warped brains were then averaged again, and this average was used as the template for the next round of the procedure. This was continued until the algorithm converged, and the Gibbs potential of the system was minimized. The template resulting from this procedure is shown in Figure 4.13. A procedure similar to this may be a very useful and principled technique to generate templates and "canonical" references.

Three Dimensional Example

A similar procedure was performed using the three dimensional registration method, where an image that is the average of six normal subjects brains was created. The image was an average not only of intensity, but also of shape. Again, the global registration methods described in Chapter 3 were first used. The procedure began by estimating the approximate deformations that map each of the images to a reference template, using the 12-parameter affine registration followed by the basis function approach (see Figures **??** and **??**). After the registration, each of the images were transformed according to the estimated parameters. The transformed images contained $121 \times 145 \times 121$ voxels, with a resolution of approximately $1.5 \times 1.5 \times 1.5$mm. The first estimate of the new template was computed as the average of these images. The estimated 4×4 affine transformation matrices and basis function coefficients were used to generate starting parameters for estimating the high dimensional deformation fields.

For each of the six images, ten iterations of the current algorithm were used to bring the images slightly closer in to register with the template. The spatially transformed images were averaged again to obtain a new estimate for the template, after which the images were again registered to the template using a further ten iterations. This process continued for a total of

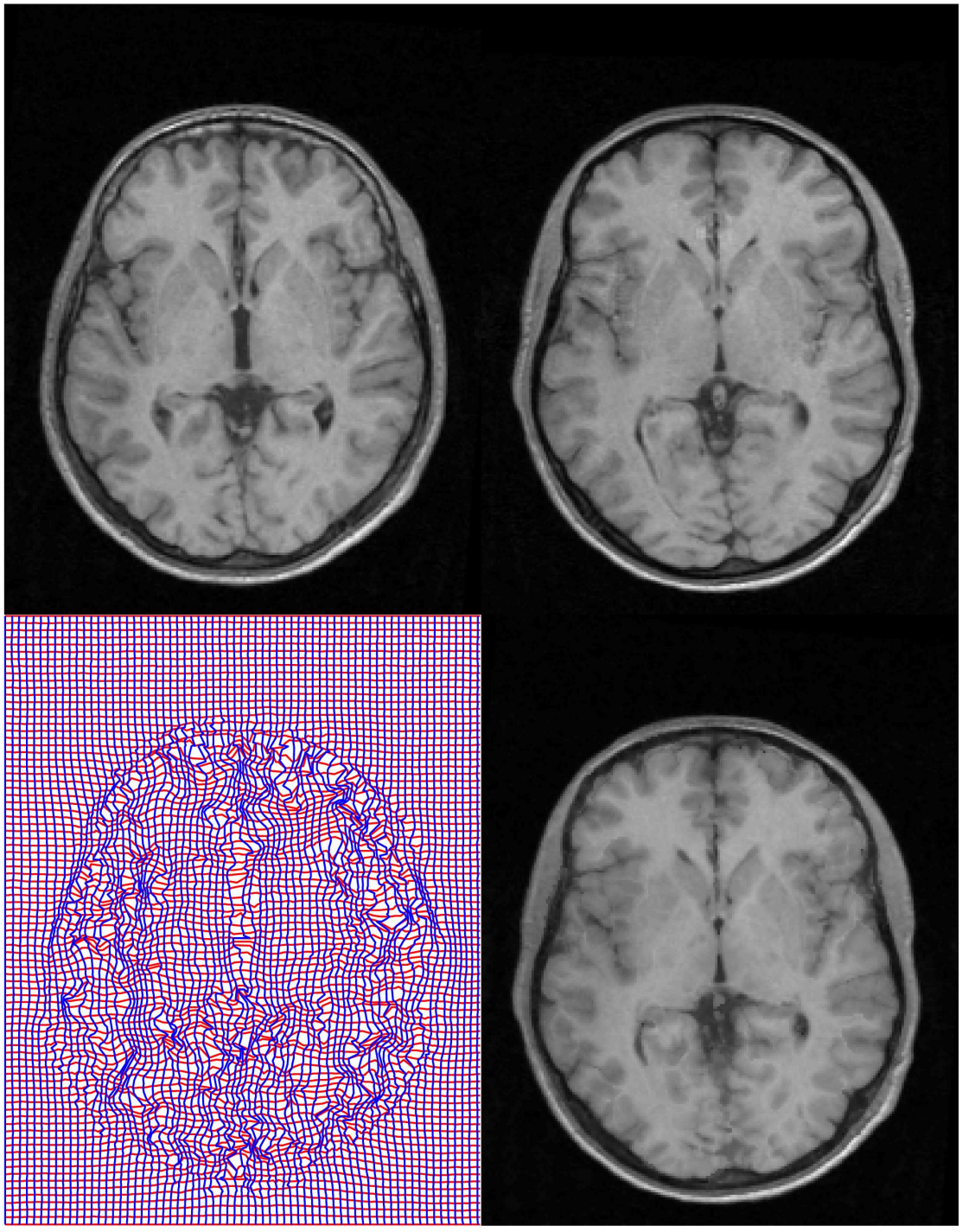

Figure 4.10: *Above left*: the unwarped source image. *Above right*: the template image. *Below left*: the deformation field applied to the source image in order to warp it to match the template image. *Below right*: the source image after warping to match the template.

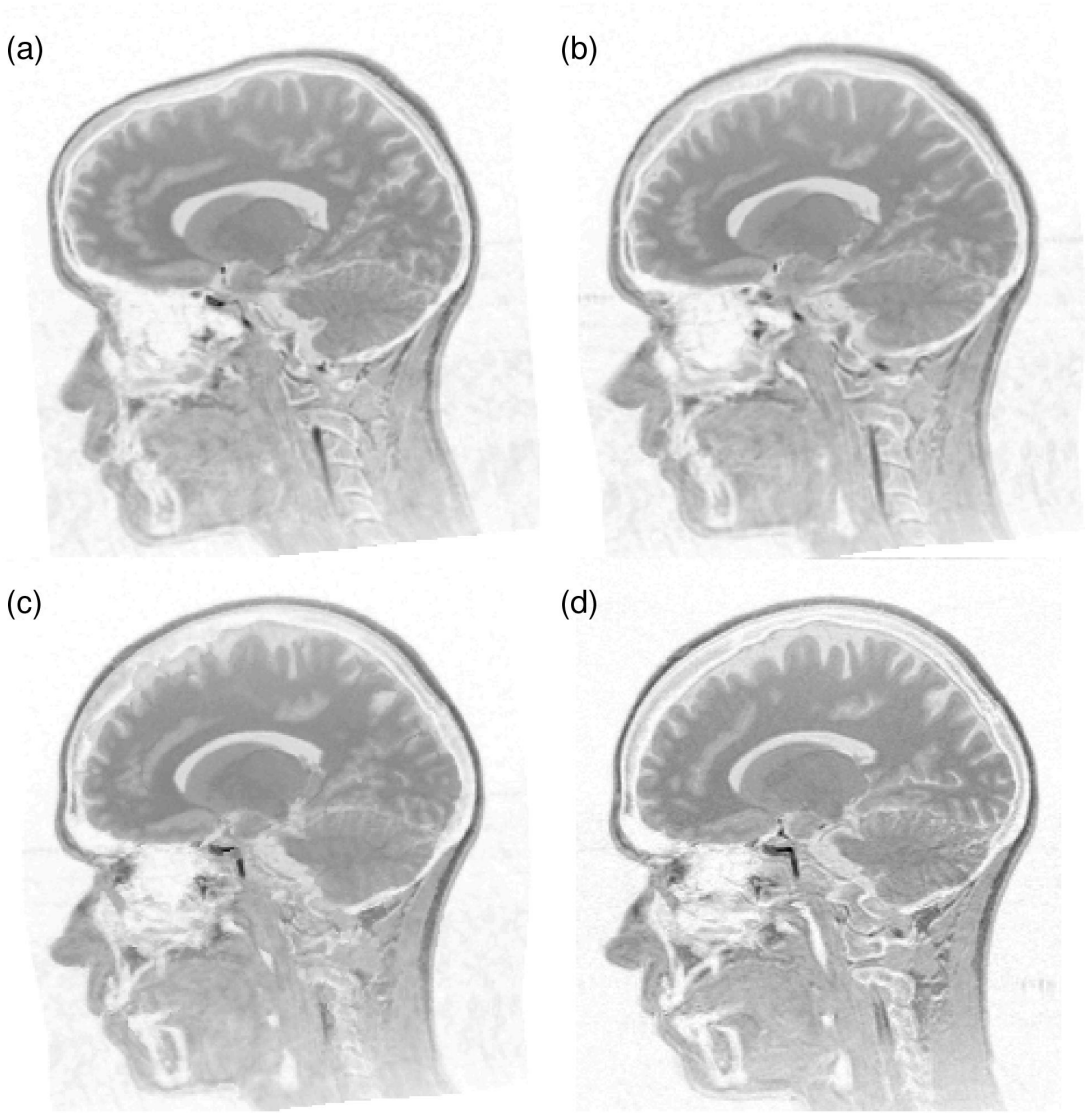

Figure 4.11: A sagittal plane from two images registered together. The template (reference) image is shown in (d). (a) shows the source image after affine registration to the template image. The source image after the basis function registration is shown in (b), and the final registration result is in (c). The deformation fields are shown in Figure 4.12.

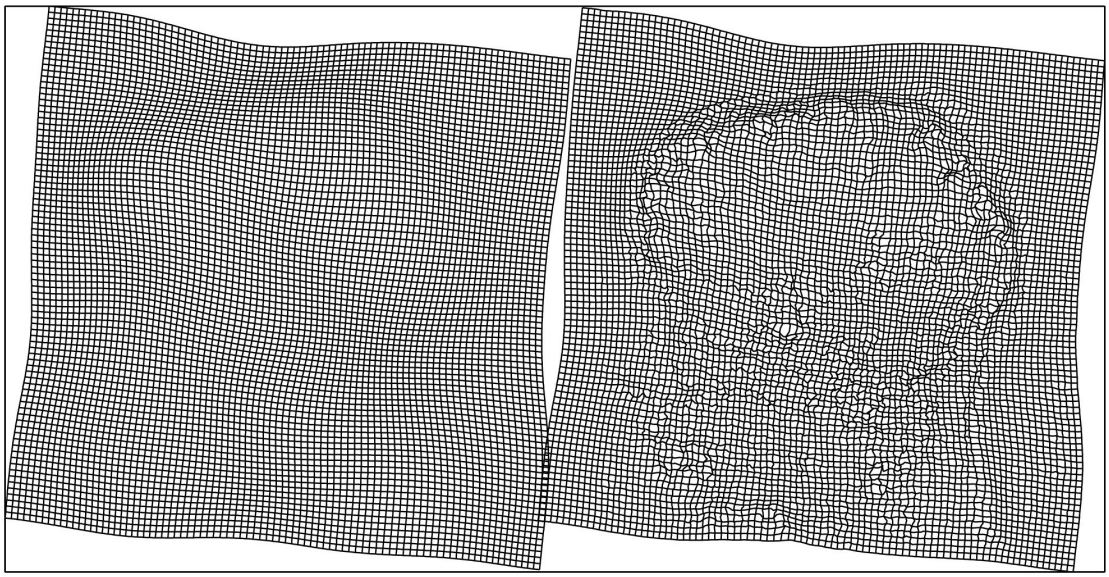

Figure 4.12: The deformation fields corresponding to the images in Figure 4.11. Two components (vertical and horizontal translations) of the field after affine and basis function registration are shown on the left, whereas the final deformation field is shown on the right.

four times. A plane from each of the spatially transformed images is shown in Figure 4.14.

Visually, the registered images appear very similar. This is not always a good indication of the quality of the registration (but it does confirm that the optimization algorithm has reduced the likelihood potentials). In theory, the wrong structures could be registered together, but distorted so that they look identical.

The brain surfaces of the original images were extracted using the procedures described in Chapter 5. This involved a segmentation of grey and white matter on which morphological operations were performed to remove the small amounts of remaining non-brain tissue. The surfaces were then rendered. A number of points were selected that were near the surface of the template brain (the average of the spatially transformed images). By using the computed spatial transformations, these points were projected on to the closest corresponding location on the rendered brain surfaces. Figure 4.15 shows the rendered surfaces. Note that the large amount of cortical variability means that it is very difficult to objectively identify homologous locations on the surfaces of the different brains. The shapes of the internal brain structures are less variable, so the method is able to estimate transformations for these regions much more precisely.

Although the deformation fields contain a mapping from the average shaped template image to each of the individual images, it is still possible to compute the mapping between any image pair. This is done by combining a forward deformation that maps from the template image to the first image, with an inverse deformation (computed as described in Section 4.2.6) that maps from the second image to the template. Figure 4.16 shows five images that have been transformed in this way to match the same image.

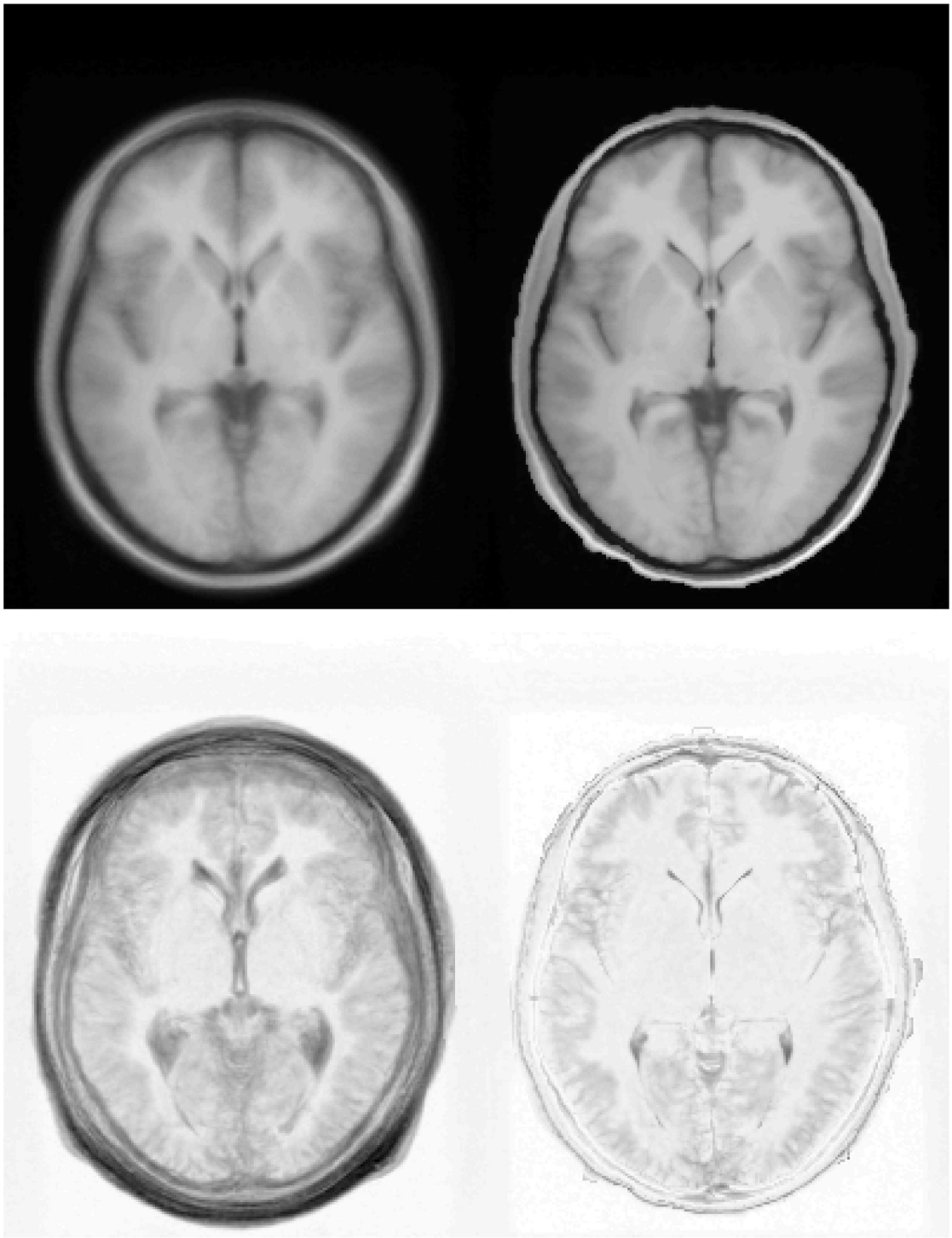

Figure 4.13: *Above left*: the average of the MR images of 29 subjects registered together using a 12-parameter affine registration. *Above right*: the average (both shape and intensity) of the same 29 MR images after registering together using the method described in Section 4.3.3. *Below left*: the standard deviations of the affine registered MR images. *Below 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.

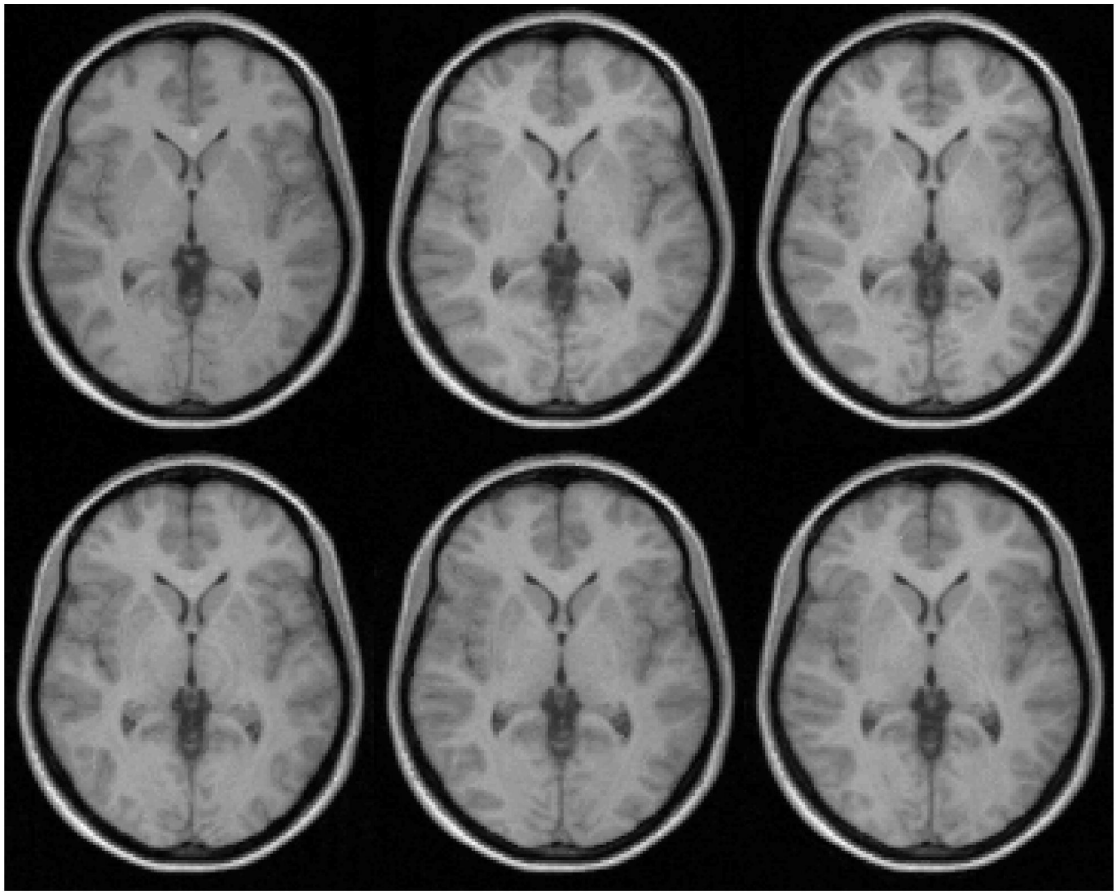

Figure 4.14: The images of the six brains after affine and basis function registration, followed by high-dimensional image registration using the methods described in this chapter (see also Figures **??** and **??**). The high-dimensional transformations are able to model high frequency deformations that cannot be achieved using the basis function approach alone.

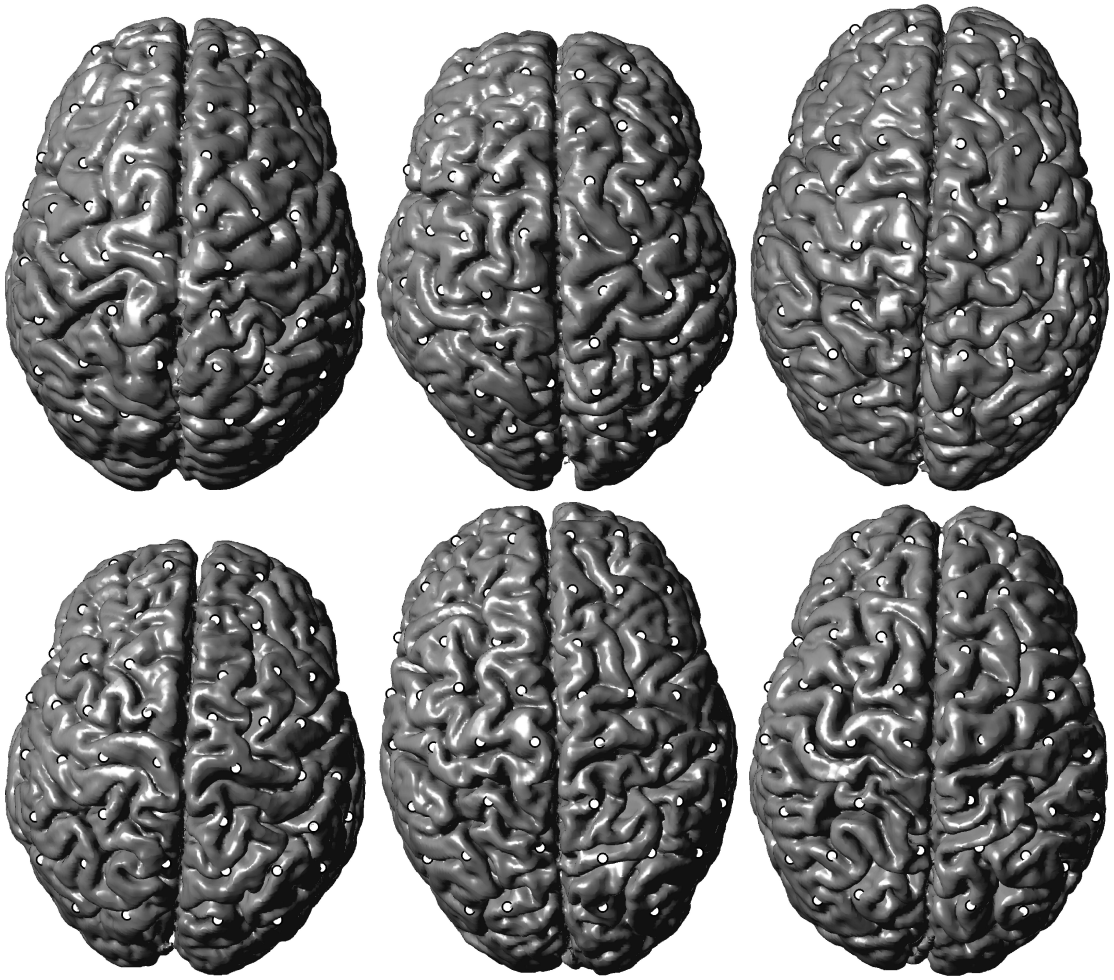

Figure 4.15: Rendered surfaces of the original six brains. The white markers correspond to equivalent locations on the brain surfaces as estimated by the registration algorithm.

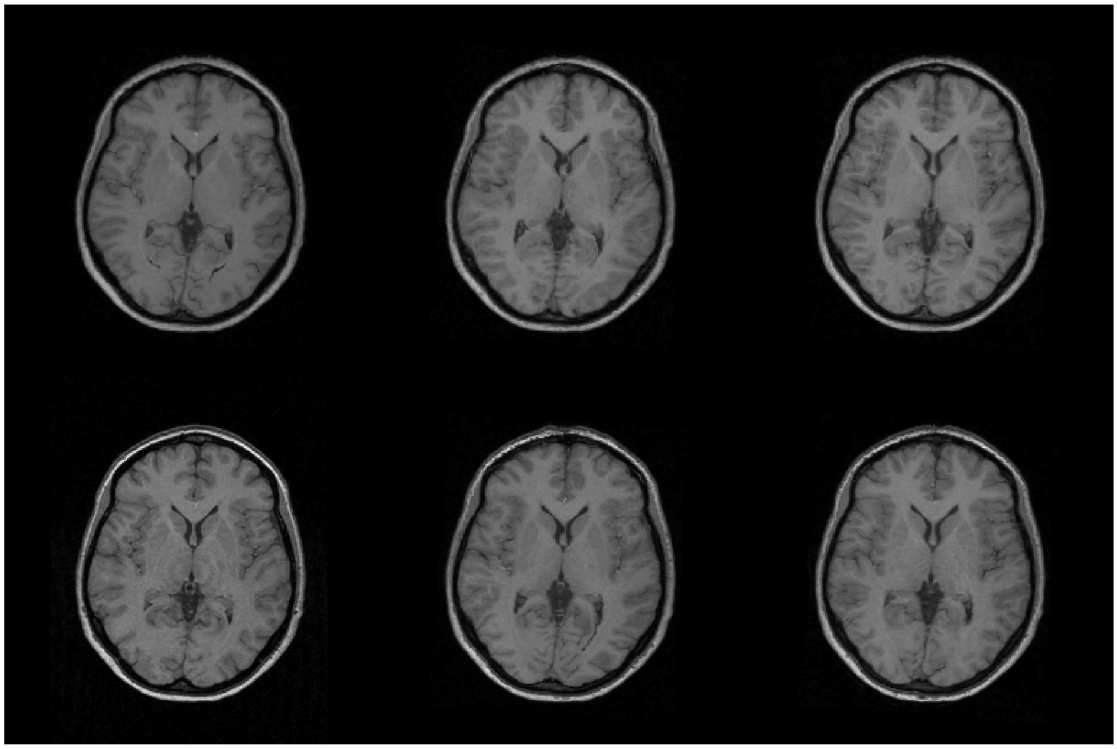

Figure 4.16: By combining the warps, it is possible to compute a mapping between any pair of images. In this example, the remaining images were all transformed to match the one shown at the lower left.

4.4 Discussion

Validation of warping methods is a complex area. The appropriateness of an evaluation depends on the particular application that the deformations are to be used for. For example, if the application was spatial normalization of functional images of different subjects, then the most appropriate evaluation may be based on assessing the sensitivity of voxel-wise statistical tests. Because the warping procedure is based only on structural information, it is blind to the locations of functional activation. If the locations of activations can be brought into close correspondence in different subjects, then it is safe to say that the spatial normalization procedure is working well. The best measure of correspondence depends on how much the images are smoothed prior to performing the statistical tests. Different registration methods will perform differently depending on the amount of smoothing used. For example, the difference in performance of high- versus low-dimensional methods will be less when lots of smoothing is used. More discussion of this can be found in Section **??**.

Another application may involve identifying shape differences among populations of subjects. In this case, the usefulness of the warping algorithm would be assessed by how well the deformation fields can be used to distinguish between the populations.

In order to know how well a new warping method works, it needs to be considered in relation to other available methods. There are also likely to be many hyper-parameters to be tweaked for each of the models in order to obtain the best results. Optimum values for these hyper-parameters are likely to depend on the types of images being registered. Rather than focusing on one of many possible evaluation strategies, this section considers the validity of the different components of the warping method described in the current chapter. The validity of the registration method is dependent on four main elements: the parameterization of the deformations, the matching criteria, the priors describing the nature of the warps, and the algorithm for estimating the spatial transformations.

4.4.1 Parameterizing the Deformations

The deformations are parameterized using regularly arranged piecewise affine transformations. The same principles described in this chapter can also be applied to more irregular arrangements of tetrahedra. Because much of an estimated deformation field is very smooth, whereas other regions are more complex, it may be advantageous in terms of speed to arrange the tetrahedra more efficiently. The layout of the triangles or tetrahedra described in this chapter is relatively simple, and it does have the advantage that no extra memory is required to store the original co-ordinates of vertices. It also means that some of the calculations required to determine the Jacobian matrices (part of a matrix inversion) can be pre-computed and stored efficiently.

An alternative to using the linear mappings could be to use piecewise nonlinear mappings such as those described by [11]. However, such mappings would not fit easily into the current framework, as there would be no simple expression for the Gibbs potential for each of the patches of deformation field. This is because the Jacobian matrix is not constant within a patch, so computing the Gibbs potential would require a complicated integration procedure.

In terms of speed, this method does not compare favorably with some other high-dimensional intensity based registration algorithms [18], and this chapter has not concentrated on describing ways of making the algorithm more efficient. One way of achieving this would be to use an increasing density of nodes. For the early iterations, when estimating smoother deformations, less nodes are required to adequately define the deformations. The number of parameters describing the deformations is equal to three times the number of nodes, and a faster convergence should be achieved using fewer parameters. It is worth noting that a coarse to fine scheme for the

arrangement of the nodes is not necessary in terms of the validity of the method. A coarse to fine approach (in terms of using smoother images and deformations for the early iterations) can still be achieved even when the deformation field is described by an equally large number of nodes from start to finish.

4.4.2 The Matching Criterion

The matching criterion described here is fully automatic, and produces reproducible and objective estimates of deformations that are not susceptible to bias from different investigators. It also means that relatively little user time is required to perform the registrations. However, this does have the disadvantage that human expertise and understanding (that is extremely difficult to encode into an algorithm) is not used by the registration. More accurate results may be possible if the method was semi-automatic, by also allowing user identified features to be matched.

The current matching criterion involves minimizing the sum of squared differences between source and template images. This same criterion is also used by many other intensity based nonlinear registration methods and assumes that one image is just a spatially transformed version of the other, but with white Gaussian noise added. It should be noted that this is not normally the case. After matching a pair of brain images, the residual difference is never purely uniform white noise, but tends to have a spatially varying magnitude. For example, the residual variance in background voxels is normally much lower than that in grey matter. An improved model would use a non-stationary variance map and possibly model covariance between neighboring voxels, or even covariance between intensities in different regions (e.g., see Section **??**).

The validity of the matching criterion depends partly upon the validity of the template image. If the intensity values of the different tissues of the template image differ systematically from those of the corresponding tissues in the source image, then the validity of the matching will be impaired because the correlations introduced into the residuals are not accounted for by the model. Pathology is another case where the validity of the registration is compromised. This is because there is no longer a one-to-one correspondence between the features of the two images. An ideal template image should contain a "canonical" or average shaped brain. On average, registering a brain image to a canonical template requires smaller (and therefore less error prone) deformations than would be necessary for registering to an unusually shaped template.

Although the priors for the registration model are symmetric, the matching criterion is not. One effect of this is that the gradients of only one image are used to drive the registration, rather than the gradients of both. This is illustrated in Figure 4.17. With a fully symmetric matching criterion, the evaluations in Sections 4.3.1 and 4.3.2 would be expected to produce more consistent results. Note however, that the matching used in Section 4.3.3 can be considered as symmetric. When registering a pair of images together by matching them both to their average, the gradients of both images are considered equally. The result of this procedure would be two deformation fields that map "half way". By combining the "half way" deformations in the appropriate way (as shown in Figure 4.16), a pair of deformation fields can be obtained that map between the images, and are both inverses of each other.

4.4.3 The Priors

Consider the transformations mapping between images $\mathbf{f}$ and $\mathbf{g}$. By combining the transformation mapping from image $\mathbf{f}$ to image $\mathbf{g}$, with the one that maps image $\mathbf{g}$ to image $\mathbf{f}$, a third transformation can be obtained that maps from $\mathbf{f}$ to $\mathbf{g}$ and then back to $\mathbf{f}$. Any non-uniformities in this resulting transformation represent errors in the registration process. The priors adopted in this chapter attempt to reduce any such inconsistencies in the deformation fields. The extreme

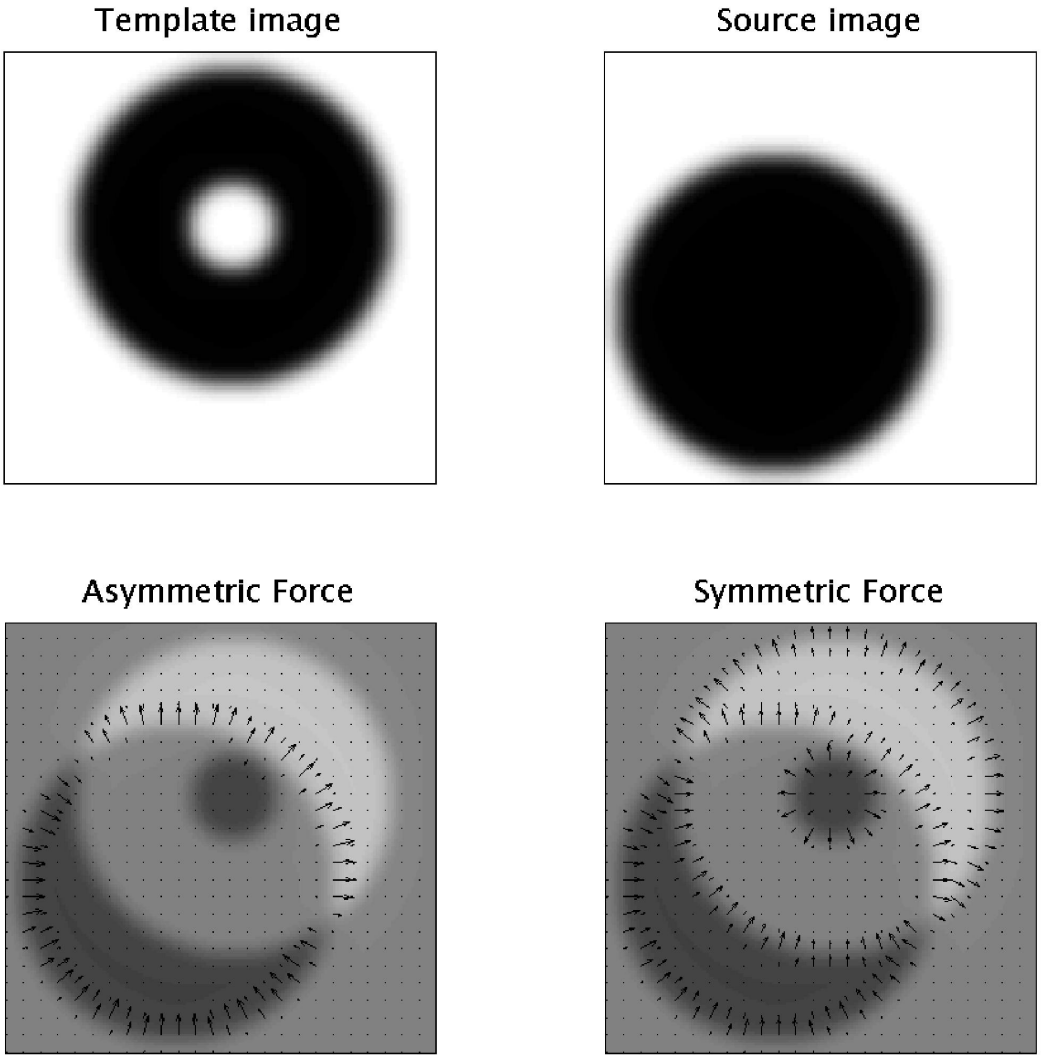

Figure 4.17: A comparison of a symmetric with an asymmetric likelihood potential. The arrows on the lower images show the directions in which the source image would be deformed. The template image contains a feature that is not found in the source image. If a registration is based only on gradients of the source image, then this feature is likely to have no effect on the final estimated spatial transformation. However, if the likelihood potential is symmetric, then this feature would drive a local expansion of the source image, until the likelihood potential is balanced by the prior potential.

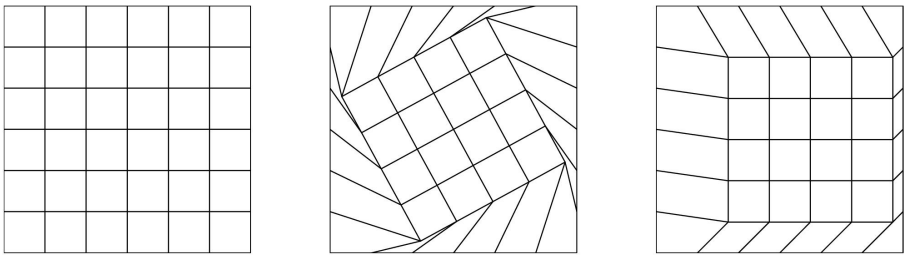

Figure 4.18: The objective of including the prior probability distribution in the registration model is to penalize shape changes of the source image. A rigid body rotation of a region of brain does nothing to the shape of that region. However, in order to rotate relative to the neighboring regions then shears and zooms are necessary and these do change the shape of the image.

case of an inconsistency between a forward and inverse transformation is when the one-to-one mapping between the images breaks down. Unlike many Bayesian registration methods that use linear priors [1, 10, 17, 3, 2], the Bayesian scheme here uses a penalty function that approaches infinity if a singularity begins to appear in the deformation field. This is achieved by considering both the forward and inverse spatial transformations at the same time. For example, when the length of a structure is doubled in the forward transformation, it means that the length should be halved in the inverse transformation. Because of this, the penalty function used here is the same for both the forward and inverse of a given spatial transformation. The ideal form for this function should be based on the logs of the singular values of the Jacobian matrices having normal distributions, but the more rapidly computed function described in Section 4.2.4 is a close enough approximation.

The penalty function is invariant to the relative orientations of the images. It does not penalize rotations or translations in isolation, only those relative to the position of neighboring voxels. In order to reposition a region relative to its neighbors, it is necessary to introduce scaling and shearing into the affine transformations. It is this scaling and shearing that the model penalizes, rather than the position and orientation itself (see Figure 4.18).

Only the form of the prior potential has been stated, and little has been said about its magnitude relative to the likelihood potential. This is because it is not clear what the relative magnitudes of the two sets of potentials should be. λ relates to our belief in the amount of brain structural variability that is likely to be observed in the population. A relatively large value for λ results in the deformations being more smooth, at the expense of a higher residual squared difference between the images, whereas a small value for λ will result in a lower residual squared difference, but less smooth deformations. The prior distributions described in this chapter are stationary (since λ is constant throughout). In reality, the true amount of brain structural variability is very likely to be different from region to region [16], so a set of non-stationary priors should, in theory, produce more valid MAP estimates.

Much of the non-stationary variability will be higher in some directions than others. An alternative way of understanding the penalty function based on normally distributed logs of singular values, is to consider the Hencky strain tensors of the deformations (see Chapter 6). The prior potential model described in this chapter is essentially minimizing the sum of squares of the Hencky tensor elements. Anisotropic variability could be modeled by assuming different variances for each Hencky tensor element, thus allowing more stretching or contraction in some directions than others. Further material properties could be introduced by also modeling the covariance between the tensor elements. For example, by making all the diagonal elements of the tensor correlated, then the deformations could be forced to be the same in all directions. Another

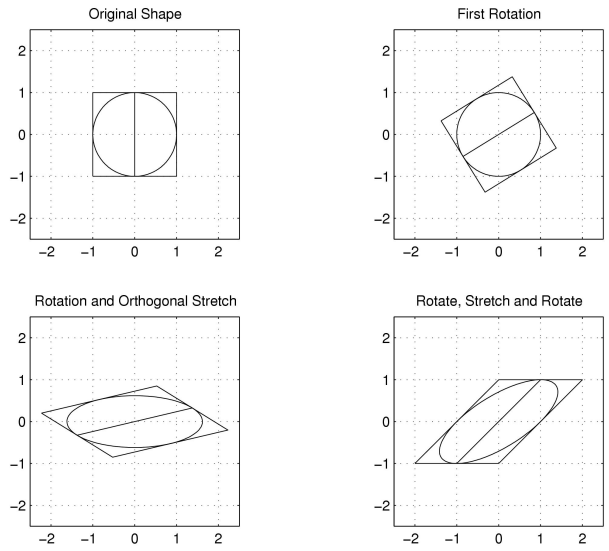

Figure 4.19: An affine transformation matrix that performs a shear must have singular values that are not equal to one. This figure shows a shear applied to a square and circle decomposed into the steps defined by singular value decomposition. The stretching that occurs between the upper right sub-figure, and that in the lower left, is what would be penalized. Note that shearing does not change the area of the objects, so a penalty based only on the Jacobian determinants would have no effect on this type of distortion.

covariance model could be used to force the deformations to be volume preserving (isochoric) by forcing the trace of the Hencky tensor to equal zero. Even more complex prior probability distributions could be devised that involve modeling covariance between the strain tensors of neighboring (or even remote) triangles or tetrahedra. In theory, such models could be used to make whole regions stretch or contract uniformly.

Estimating the normal amount of structural variability is not straightforward. Registration methods could be used to to do this by registering a large number of brain images to a canonical template. However, the estimates of structural variability will be heavily dependent upon the priors used by the algorithm. A "chicken and egg" situation arises, whereby the priors are needed to estimate the optimum deformation fields, and the deformation fields are needed to estimate the correct priors. It may be possible to overcome this problem using an Empirical Bayes method (see Chapter 13) to estimate the unknown hyper-parameters (i.e., σ^2 and λ), which describe the relative importance of the different components of the objective function.

4.4.4 The Optimization Algorithm

The method searches for the MAP solution, which is the single most probable realization of all possible deformation fields. The steepest descent algorithm that is used does not guarantee that the globally optimum MAP solution will be achieved, but it does mean that a local optimum solution can be reached – eventually. Robust optimization methods that almost always find the global optimum would take an extremely long time to run with a model that uses millions of parameters. These methods are simply not feasible for routine use on problems of this scale. However, if sulci and gyri can be easily labeled from the 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 [13, 19, 8], therefore increasing the likelihood of obtaining the global optimum.

If the starting estimates are sufficiently close to the global optimum, then the algorithm is more likely to find the true MAP solution. Therefore, the choice of starting parameters can influence the validity of the final registration result. An error surface based only on the prior potential does not contain any local minima. However there may be many local minima when the likelihood potential is added to this. Therefore, if the posterior potential is dominated by the likelihood potential, then it is much less likely that the algorithm will achieve the true MAP solution. If very high frequency deformations are to be estimated, then the starting parameters must be very close to the optimum solution.

One method of increasing the likelihood of achieving a good solution is to gradually reduce the value of λ relative to $1/\sigma^2$ over time. This has the effect of making the registration estimate the more global deformations before estimating more detailed warps. Most of the spatial variability is low frequency, so the algorithm can get reasonably close to a good solution using a relatively high value for lambda. This also reduces the number of local minima for the early iterations. The images should also be smoother for the earlier iterations in order to reduce the amount of confounding information and the number of local minima. A review of such approaches can be found in [15].

A value for σ^2 is used that is based on the residual squared difference between the images after the previous iteration. σ^2 is larger for the early iterations, so the posterior potential is based more on the priors. It decreases over time, thus decreasing the influence of the priors and allowing higher frequency deformations to be estimated. Similarly, for the example where images were registered to their average, the template image was smoothest at the beginning. Each time the template was recreated, it was slightly crisper than the previous version. High frequency information that would confound the registration in the early iterations is gradually re-introduced to the template image as it is needed.

At first sight, it would appear that optimizing the millions of parameters that describe a deformation field would be an impossible task. It should be noted that these parameters are all related to each other since the regularization tends to preserve the shape of the image, and so reduces the effective number of degrees of freedom that the model has to fit. The limiting case would be to set the regularization parameter λ to infinity. Providing that the boundary conditions allowed it, this would theoretically reduce the dimensionality of the problem to a six parameter rigid body transformation (although the current implementation would be unable to cope with a λ of infinity).

Bibliography

[1] Y. Amit, U. Grenander, and M. Piccioni. Structural image restoration through deformable templates. *Journal of the American Statistical Association*, 86:376–387, 1991.

[2] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989.

[3] F. L. Bookstein. Landmark methods for forms without landmarks: Morphometrics of group differences in outline shape. *Medical Image Analysis*, 1(3):225–243, 1997.

[4] G. E. Christensen. Consistent linear elastic transformations for image matching. In A. Kuba et al., editor, *Proc. Information Processing in Medical Imaging*, pages 224–237, Berlin, Heidelberg, 1999. Springer-Verlag.

[5] G. E. Christensen, R. D. Rabbitt, and M. I. Miller. 3D brain mapping using using a deformable neuroanatomy. *Physics in Medicine and Biology*, 39:609–618, 1994.

[6] G. E. Christensen, R. D. Rabbitt, and M. I. Miller. Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing*, 5:1435–1447, 1996.

[7] G. E. Christensen, R. D. Rabbitt, M. I. Miller, S. C. Joshi, U. Grenander, T. A. Coogan, and D. C. Van Essen. Topological properties of smooth anatomic maps. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Proc. Information Processing in Medical Imaging*, pages 101–112, Dordrecht, The Netherlands, 1995. Kluwer Academic Publishers.

[8] C. Davatzikos. Spatial normalization of 3D images using deformable models. *Journal of Computer Assisted Tomography*, 20(4):656–665, 1996.

[9] P. A. Freeborough and N. C. Fox. Modelling brain deformations in alzheimer disease by fluid registration of serial MR images. *Journal of Computer Assisted Tomography*, 22(5):838–843, 1998.

[10] J. C. Gee, D. R. Haynor, L. Le Briquer, and R. K. Bajcsy. Advances in elastic matching theory and its implementation. In P. Cinquin, R. Kikinis, and S. Lavallee, editors, *Proc. CVRMed-MRCAS'97*, Heidelberg, 1997. Springer-Verlag.

[11] A. Goshtasby. Piecewise cubic mapping functions for image registration. *Pattern Recognition*, 20(5):525–533, 1987.

[12] A. Guéziec and R. Hummel. Exploiting triangulated surface extraction using tetrahedral decomposition. *IEEE Transactions on Visualization and Computer Graphics*, 1:328–342, 1995.

[13] S. C. Joshi, M. I. Miller, G. E. Christensen, A. Banerjee, T. A. Coogan, and U. Grenander. Hierarchical brain mapping via a generalized dirichlet solutions for mapping brain manifolds. In *Proc. SPIE International Symposium on Optical Science, Engineering and Instrumentation*, 1995.

[14] L. Le Briquer and J. C. Gee. Design of a statistical model of brain shape. In J. Duncan and G. Gindi, editors, *Proc. Information Processing in Medical Imaging*, pages 477–482, Berlin, Heidelberg, New York, 1997. Springer-Verlag.

[15] H. Lester and S. R. Arridge. A survey of hierarchical non-linear medical image registration. *Pattern Recognition*, 32:129–149, 1999.

[16] H. Lester, S. R. Arridge, K. M. Jansons, L. Lemieux, J. V. Hajnal, and A. Oatridge. Non-linear registration with the variable viscosity fluid algorithm. In A. Kuba et al, editor, *Proc. Information Processing in Medical Imaging*, pages 238–251, Berlin, Heidelberg, 1999. Springer-Verlag.

[17] M. I. Miller, G. E. Christensen, Y. Amit, and U. Grenander. Mathematical textbook of deformable neuroanatomies. *Proc. National Academy of Sciences*, 90:11944–11948, 1993.

[18] J.-P. Thirion. Fast non-rigid matching of 3D medical images. Technical Report 2547, Institut National de Recherche en Informatique et en Automatique, May 1995. Available from http://www.inria.fr/RRRT/RR-2547.html.

[19] P. M. Thompson and A. W. Toga. Visualization and mapping of anatomic abnormalities using a probablistic brain atlas based on random fluid transformations. In *Proc. Visualization in Biomedical Computing*, pages 383–392, 1996.

[20] R. P. Woods, S. T. Grafton, C. J. Holmes, S. R. Cherry, and J. C. Mazziotta. Automated image registration: I. General methods and intrasubject, intramodality validation. *Journal of Computer Assisted Tomography*, 22(1):139–152, 1998.