Computational Neuroanatomy

by

John Towers Ashburner

Prepared under the supervision of Prof. Karl J. Friston.

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
of the
University of London.

The Wellcome Department of Cognitive Neurology
Institute of Neurology
University College London
London
UK

July 2000

Abstract

A number of procedures have been developed for brain morphometry, many of which are essential in functional imaging applications. The theme developed in the thesis is computational neuroanatomy, which relies upon a series of image registration methods, image segmentation and statistical methods for characterising brain structure.

The simplest registration is for rigid bodies, and is normally applied within subject. Methods are described for rigid registration of both inter- and intra-modality images. More complex models are required for registering images of different subjects into the same stereotactic space. A coarse, but fast method is described, which begins with a 12-parameter affine registration, followed by nonlinear warps modelled by a linear combination of spatial basis functions. The registration proceeds within a Bayesian framework, which is used for penalising unlikely shape changes. A high-dimensional warping method follows for refining the initial registration. In order to estimate more accurate warps, this method emphasises consistency of the deformations, by considering that warping brain A to brain B should not be different, probablistically, from warping B to A. A Bayesian framework is used again, whereby a prior probability distribution for the warps is assumed that embodies this symmetry.

A new scheme for segmenting grey and white matter from MR images has been developed, which is based on Mixture Model cluster analysis. Registered prior probability maps of different tissue classes are used to make the classification more robust, and MRI intensity nonuniformity correction is also incorporated into this model.

Finally, a taxonomy of morphometric methods is described, which characterise the regional distribution of different tissue types, or shape differences inherent in the deformations computed by the warping methods.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Contents

1.1 Motivation and Aims

The initial motivation for this work was to develop improved methods of image registration for functional imaging. Much of it has now been incorporated into the SPM99 package, and is used by several hundred researchers around the world for analysing functional imaging data. The second motivation was to facilitate the development of methods for studying brain shape among different populations. Recently, the term *computational neuroanatomy* has been coined for this area of research. These areas are now reviewed in more detail.

1.1.1 Functional Imaging

The principle behind detecting activations using functional imaging methods such as *Positron Emission Tomography* (PET) or *functional Magnetic Resonance Imaging* (fMRI) is essentially a voxel by voxel t-test on a series of images acquired under different conditions (Friston *et al.*, 1995d; Worsley & Friston, 1995). This analysis results in a statistical parametric map (SPM) showing significant differences in cerebral blood flow that are explained according to the different conditions experienced by the subject (or subjects) in the scanner.

The first modalities used for this type of study were PET and SPECT (*Single Photon Emission Computed Tomography*). Recent advances in MR methods, the large number of existing scanners, coupled with the relatively high cost of producing radioactive tracers, and the invasive nature of PET and SPECT have meant that most studies currently use fMRI.

PET and SPECT methods involve generating images from the photons of radiation emitted by tracers injected into, or inhaled by, the subjects. Tracers used in PET studies emit positrons when they decay, and include ^{11}C, ^{15}O and ^{18}F, although most studies are based on identifying regional differences in cerebral blood flow, and involve injecting the subject with ^{15}O labelled water, or the subject breathing ^{15}O labelled carbon dioxide. When a positron meets an electron, the two annihilate each other resulting in the emission of two gamma ray photons in opposite directions. By recording the paths of the gamma ray pairs, it is possible to reconstruct a three dimensional image of the tracer concentration. For PET studies, a typical protocol may involve about 12 scans, with an injection of ^{15}O labelled water prior to each of them. The subject may have one task to do for six of the scans, and a different task for the other six. More active brain regions have a higher rate of blood flow, and so receive the tracer earlier than the other areas. By imaging the brain as the radioactivity is entering, and comparing the images resulting from the different tasks, it is possible to create a picture of where the tasks have most influence over the cerebral blood flow.

The mechanisms of MRI are very different from those of PET, and rely on the nuclei of certain atoms (normally ^{1}H) absorbing and then re-emitting radio waves when in a magnetic field. The frequency of the absorbed and emitted waves depends on the strength of the magnetic field, so by varying the field over the head it is possible to record waves of different frequencies from different regions. Fourier transform methods are used to reconstruct images of where the signals emanate from. Depending on the properties of the surrounding tissue, the amplitudes of the signals will decay at different rates, so not only does MRI produce a map of the density of ^{1}H atoms, but it also says something about the environment in which the atoms are found.

Currently, the index of neuronal activity most commonly used for fMRI is the *Blood Oxygenation Level Dependent* (BOLD) contrast (Ogawa *et al.*, 1990). The assumption is that an increase in neuronal activity within a brain region results in an increase in local blood flow, leading to reduced concentrations of deoxyhæmoglobin in the blood vessels. Unlike oxyhæmoglobin, deoxyhæmoglobin has a differential magnetic susceptibility in relation to the surrounding tissue. Therefore, relative decreases in deoxyhæmoglobin concentration lead to a reduction in local field inhomogeneity and a slower decay of the MR signal, resulting in higher intensities in the images. fMRI allows whole brain images to be collected in about six seconds, giving it a much better temporal resolution than PET (as there is typically a wait of about 10 minutes between scans in order for the radiation from the previous scan to decay). This means that hundreds of fMRI volumes are often collected for each subject.

Image registration is important in many aspects of functional image analysis. In imaging neuroscience, particularly for fMRI, the signal changes due to any hæmodynamic response can be small compared to apparent signal differences that can result from subject motion, so it is important that the images are as closely aligned as possible prior to performing the statistical tests. Subject head movement in the scanner can not be completely eliminated, so motion correction needs to be performed as a preprocessing step. Most current algorithms for movement correction consider the head as a rigid object. In three dimensions, six parameters are needed to define a rigid body transformation (three translations and three rotations). The first step in the correction is image registration, which involves determining the parameter values for a rigid body transformation that optimise some criteria for matching each image with a reference image. In order to be effective, the accuracy of the registration needs to be within a fraction of a voxel. Following the registration, the images are transformed by resampling according to the determined parameters.

Motion correction is especially important for experiments where subjects may move in the scanner in a way that is correlated with the different conditions (Hajnal *et al.*, 1994). Even tiny systematic differences can result in a significant signal accumulating over numerous scans. Without suitable corrections, artifacts arising from subject movement correlated with the experimental paradigm may appear as activations. A second reason why motion correction is important is that it increases sensitivity. The t-test is based on the signal change relative to the residual variance. The residual variance is computed from the sum of squared differences between the data and the linear model to which it is fitted. Movement artifacts add to this residual variance, and so reduce the sensitivity of the test to true activations.

For studies of a single subject, sites of activation can be accurately localised by superimposing them on a high resolution structural image of the subject (typically a T1 weighted MRI). This requires registration of the functional images with the structural image. As in the case of movement correction, this is normally performed by optimising a set of parameters describing a rigid body transformation, but the matching criterion needs to be more complex because the structural and functional images normally look very different. A further use for this registration is that a more precise spatial normalisation can be achieved by computing it from a more detailed structural image. If the functional and structural images are in register, then a warp computed from the structural image can be applied to the functional images.

Sometimes it is desirable to warp images from a number of individuals into roughly the same standard space to allow signal averaging across subjects. This procedure is known as spatial normalisation. Because different people may have different strategies for performing tasks in the scanner, spatial normalisation of the images is useful for determining what happens generically over individuals. A further advantage of using spatially normalised images is that activation sites can be reported according to their Euclidian co-ordinates within a standard space (Fox, 1995). The most commonly adopted co-ordinate system within the brain imaging community is that described by Talairach & Tournoux (1988), although new standards are now emerging that are based on digital atlases (Evans *et al.*, 1993; Evans *et al.*, 1994; Mazziotta *et al.*, 1995).

1.1.2 Computational Neuroanatomy

A large number of approaches for characterising differences in the shape and neuroanatomical configuration of different brains have recently emerged due to improved resolution of anatomical human brain scans and the development of new sophisticated image processing techniques.

One of the simplest morphometric approaches involves identifying shape changes within single subjects by subtracting coregistered images acquired at different times. The changes could be because of a number of different reasons, but most are related to pathology. Because the scans are of the same subject, the first step for this kind of analysis involves registering the images together by a rigid body transformation.

Other approaches require the images of multiple subjects to be registered together by some form of spatial normalisation. The primary result of spatially normalising a series of images is that they all conform to the same stereotactic space, enabling region-by-region comparisons to be performed. A second result is a series of deformation fields that describe the spatial transformations required to match the different shaped brains to the same template. Encoded within each deformation field is information about the individual image shapes, which can be

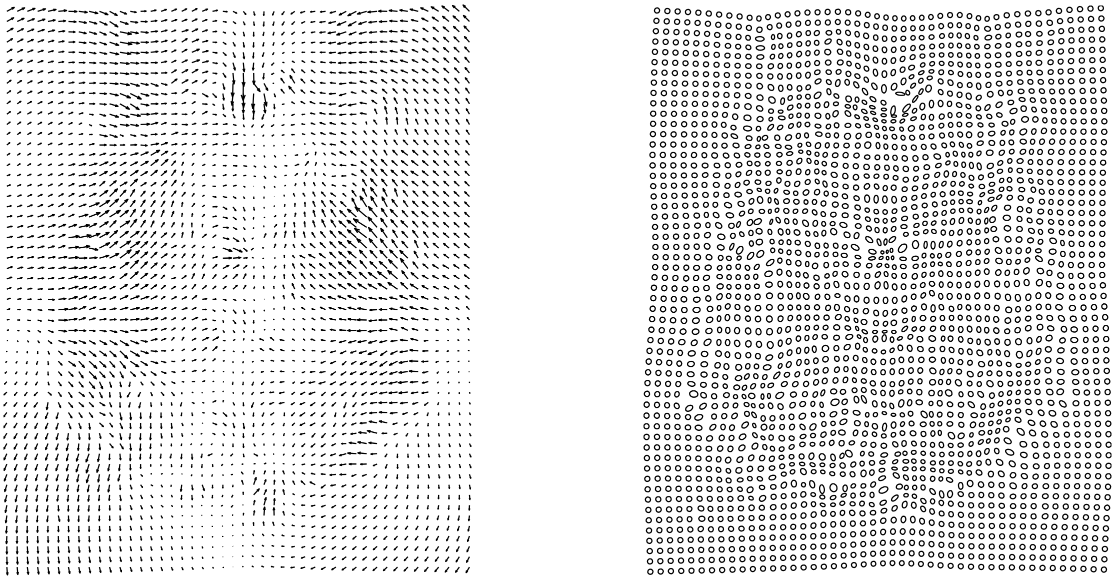

Figure 1.1: The term "deformation-based morphometry" will be used to describe methods of studying the positions of structures within the brain (left), whereas the term "tensor-based morphometry" will be used for methods that look at local shapes (right). Currently, the main application of tensor-based morphometry involves using the Jacobian determinants to examine the relative volumes of different structures. However, there are other features of the Jacobian matrices that could be used, such as those representing elongation and contraction in different directions. The arrows in the panel on the left show absolute displacements after making a global correction for rotations and translations, whereas the ellipses on the right show how the same circles would be distorted in different parts of the brain.

further characterised using a number of statistical procedures.

The terms deformation-based and tensor-based morphometry will be used to denote methods of studying brain shape that are based on deformation fields. When comparing groups, deformation-based morphometry (DBM) uses deformation fields to identify differences in the relative positions of structures within subjects' brains. Tensor-based morphometry (TBM) refers to those methods that identify differences in the local shape of brain structures (see Figure 1.1).

Characterisation using DBM can be global, pertaining to the entire field as a single observation, or can proceed on a voxel by voxel basis to make inferences about regionally specific positional differences. This simple approach to the analysis of deformation fields involves treating them as vector fields representing absolute displacements. However in this form, in addition to shape information, the vector fields also contain information on position and size that is likely to confound an analysis. Much of the confounding information is first removed by global rotations, translations and a zoom of the fields (Bookstein, 1997a).

DBM can be applied on a global scale to simply identify whether there are significant differences in overall shapes (based on a small number of parameters) among the brains of different populations. Generally, a single multi-variate test is performed using parameters describing the deformations - often after parameter reduction with singular value decomposition. The Hotelling's

T^2 statistic can be used for such simple comparisons between two groups of subjects (Bookstein, 1997a; Bookstein, 1999), but for more complex experimental designs, a multi-variate analysis of covariance can be used to identify differences via the Wilk's λ statistic.

An alternative approach to DBM involves producing a statistical parametric map that locates any regions of significant positional differences among the groups of subjects. An example of this approach involves using a voxel-wise Hotelling's T^2 test on the vector field describing the displacements at each and every voxel (Thompson & Toga, 1999; Gaser *et al.*, 1999). The significance of any observed differences can be assessed by modelling the statistic field as a T^2 random field (Cao & Worsley, 1999). Note that this approach does not directly localise brain regions with different shapes, but rather identifies those brain structures that are in relatively different positions.

If the objective is to localise structures whos shapes differ among groups, then some form of tensor-based morphometry is required to produce statistical parametric maps of regional shape differences. A deformation field that maps one image to another can be considered as a discrete vector field. By taking the gradients at each element of the field, a Jacobian matrix field is obtained, in which each element is a tensor describing the relative positions of neighbouring elements. Morphometric measures derived from such a tensor field can be used to locate regions with different shapes. The field obtained by taking the determinants at each point gives a map of structural volumes relative to those of a reference image (Freeborough & Fox, 1998; Gee & Bajcsy, 1999). Statistical parametric maps of these determinant fields can then be used to compare the anatomy of groups of subjects. A number of other measures derived from tensor fields have been used by other researchers, and these are described by Thompson and Toga (1999).

Another form of morphometry involves examining the local composition of brain images. Grey and white matter voxels can be identified by image segmentation, before applying morphometric methods to study the spatial distribution of the tissue classes. These techniques will be referred to as voxel-based morphometry (VBM). Currently, the difficulty of computing very high resolution deformation fields (required for TBM at small scales) makes voxel-based morphometry a simple and pragmatic approach to addressing small scale differences that is within the capabilities of most research units.

To summarise, computational neuroanatomic techniques can either use the deformation fields themselves or use these fields to normalise images that are then entered into an analysis of regionally specific differences. In this way, information about overall shape (deformations fields) and residual anatomic differences inherent in the data (registered images) can be partitioned.

1.2 Overview of Chapters

The remaining chapters of this thesis are organised as follows.

Rigid Body Registration

Rigid body registration is one of the simplest forms of image registration, so this chapter provides an ideal framework for introducing some of the concepts that will be used by the more complex registration methods described in later chapters. The shape of a human brain changes very little with head movement, so rigid body transformations can be used to model different head positions of the same subject. Registration methods described in this chapter include within modality, or between different modalities such as PET and MRI. Matching of two images is performed by finding the rotations and translations that optimise some mutual function of the images. Within modality registration generally involves matching the images by minimising the sum of squared difference between them. For between modality registration, the matching criterion needs to be more complex. A method for co-registering brain images of the same subject that have been acquired in different modalities is presented. The basic idea is that instead of matching two images directly, one performs intermediate within modality registrations to two template images that are already in register. One can use a least squares minimisation to determine the affine transformations that map between the templates and the images. By incorporating suitable constraints, a rigid body transformation that directly maps between the images can be extracted from these more general affine transformations. A further refinement capitalises on the implicit normalisation of both images into a standard space. This facilitates partitioning both original images into homologous tissue classes. Once extracted, the partitions are jointly matched further increasing the accuracy of the co-registration.

Image Warping with Basis Functions

This chapter describes the steps involved in registering images of different subjects into roughly the same co-ordinate system, where the co-ordinate system is defined by a template image (or series of images). The method only uses up to a few hundred parameters, so can only model global brain shape. It works by estimating the optimum coefficients for a set of bases, by minimising the sum of squared differences between the template and source image, while simultaneously maximising the smoothness of the transformation using a *maximum a posteriori* (MAP) approach. In order to adopt the MAP approach, it is necessary to have estimates of the likelihood of obtaining the fit given the data, which requires prior knowledge of spatial variability, and also knowledge of the variance associated with each observation. True Bayesian approaches assume that the variance associated with each voxel is already known, whereas the approach developed here is a type of Empirical Bayesian method, which attempts to estimate this variance from the residual errors. Because the registration is based on smooth images, correlations between neighbouring voxels are considered when estimating the variance. This makes the same approach suitable for the spatial normalisation of both high quality MR images, and low resolution noisy PET images. A fast algorithm has been developed that utilises Taylor's Theorem and the separable nature of the basis functions, meaning that most of the nonlinear spatial variability between images can be automatically corrected within a few minutes.

The approach begins by matching the images using an affine transformation. Unlike Chapter 2 – where the images to be matched together are from the same subject – zooms and shears are needed to register heads of different shapes and sizes. Knowledge of the variability of head sizes is included within a Bayesian framework in order to increase the robustness and accuracy of the method. Following this step, gross differences in head shapes, that can not be accounted for by affine normalisation alone, are corrected by a nonlinear spatial normalisation procedure. In order to reduce the number of parameters to be estimated, the nonlinear warps are described by a linear combination of low spatial frequency discrete cosine transform basis functions. Regularisation of the problem involves biasing the warps to be smooth by simultaneously minimising their membrane energy.

High Dimensional Image Warping

This chapter is also about warping brain images of different subjects to the same stereotactic 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 probablistically 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 singular values of the Jacobian matrices having log-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.

Segmentation

A tissue classification method was originally developed to be part of the between modality registration procedure described in Chapter 2, but the classification results are also useful for various types of morphometry, as well as having potential applications in other registration techniques. This chapter describes a method of segmenting MR images into different tissue classes, using a modified Gaussian Mixture Model. By knowing the prior spatial probability of each voxel being grey matter, white matter or cerebro-spinal fluid, it is possible to obtain a more robust classification. In addition, a step for correcting intensity non-uniformity is also included, which makes the method more applicable to images corrupted by smooth intensity variations. Evaluations of the method show that the non-uniformity correction improves the segmentation of images containing this artifact.

Morphometry

The chapter on morphometry covers three principle morphometric methods, that will be called *voxel-based*, *deformation-based* and *tensor-based* morphometry.

At its simplest, voxel-based morphometry (VBM) involves a voxel-wise comparison of the local concentration of grey matter between two groups of subjects. The procedure is relatively straightforward, and involves spatially normalising high resolution MR images from all the subjects in the study into the same stereotactic space. This is followed by segmenting the grey matter from the spatially normalised images, and smoothing these grey-matter segments. Voxel-wise parametric statistical tests are performed which compare the smoothed grey-matter images from the groups. Corrections for multiple comparisons are made using the theory of Gaussian random fields. This chapter describes the steps involved in VBM, and provides evaluations of the assumptions made about the statistical distribution of the data.

Deformation-based morphometry (DBM) is a method for identifying macroscopic anatomical differences among the brains of different populations of subjects. The method involves spatially normalising the structural MR images of a number of subjects so that they all conform to the same stereotactic space. Multivariate statistics are then applied to the parameters describing the estimated nonlinear deformations that ensue. To illustrate the method, the gross morphometry of male and female subjects are compared. Brain asymmetry, the effect of handedness, and the interactions among these effects are also assessed.

Tensor-based morphometry (TBM) is introduced as a method of identifying regional structural differences from the gradients of deformations fields. Deformation fields encode the relative positions of different brain structures, but local shapes (such as volumes, lengths and areas) are encoded in their gradients (Jacobian matrix field). Various functions of these tensor-fields can be used to characterise shape differences.

Chapter 2

Rigid Body Registration

Contents

2.1 Introduction

Rigid body registration is normally used for registering images of the same subject. This chapter describes methods of within subject registration for images of the same or different modalities.

For every image registration, the spatial transformation should be described by a set of parameters. In three dimensions, rigid registration requires six parameters: three translations and three rotations. There are two steps involved in registering a pair of images together. There is the *registration* itself, whereby the parameters describing a transformation are estimated. Then there is the *transformation*, where one of the images is transformed according to the set of parameters.

At its simplest, image registration involves estimating a mapping between a pair of images. One image is assumed to remain stationary (the target or template image), whereas the other (the source image) is spatially transformed to match it. In order to transform the source to match the target, it is necessary to determine a mapping from each voxel position ($\mathbf{x}$) in the target to a corresponding position ($\mathbf{y}$) in the source. The source is then resampled at the new positions. The vector function $\mathbf{y}$ can be thought of as a function of $\mathbf{x}$, and a set of transformation parameters $\mathbf{q}$ that are estimated in order to register the images.

This chapter will touch first on how rigid transformations are parameterised in terms of affine transformations. The next section explains how the images are transformed via the process of resampling, before the optimisation section explains how the best values for the parameters ($\mathbf{q}$) are estimated. The simplest form of within subject registration involves registering together two images of the same modality. A method for doing this is briefly described, before the final section describes a more complex technique for performing between modality registration.

2.2 Affine Transformations

Rigid body transformations are a subset of the more general affine transformations. For each point (x_1, x_2, x_3) in an image, an affine mapping can be defined into the co-ordinates of another space (y_1, y_2, y_3). This is expressed as:

$$\begin{array}{rclllll} y_1 & = & m_{11}x_1 & + m_{12}x_2 & + m_{13}x_3 & + m_{14} \\ y_2 & = & m_{21}x_1 & + m_{22}x_2 & + m_{23}x_3 & + m_{24} \\ y_3 & = & m_{31}x_1 & + m_{32}x_2 & + m_{33}x_3 & + m_{34} \end{array} \tag{2.1}$$

This mapping is often expressed as a simple matrix multiplication ($\mathbf{y} = \mathbf{Mx}$):

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix} \tag{2.2}$$

The elegance of formulating these transformations in terms of matrices is that several transformations can be combined simply by multiplying the matrices together to form a single matrix. This means that repeated resampling of data can be avoided when reorienting an image.

Translations

Translations are simple to implement. If a point $\mathbf{x}$ is to be translated by $\mathbf{q}$ units, then the transformation is simply:

$$\mathbf{y} = \mathbf{x} + \mathbf{q} \tag{2.3}$$

In matrix terms, this transformation can be considered as:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & q_1 \\ 0 & 1 & 0 & q_2 \\ 0 & 0 & 1 & q_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix} \tag{2.4}$$

Rotations

In two dimensions, a rotation is described by a single angle. Consider a point at co-ordinate (x_1, x_2) on a two dimensional plane. A rotation of this point to new co-ordinates (y_1, y_2), by θ

radians around the origin, can be generated by the transformation:

$$\begin{array}{rclcl} y_1 & = & cos(\theta)x_1 & + & sin(\theta)x_2 \\ y_2 & = & -sin(\theta)x_1 & + & cos(\theta)x_2 \end{array} \tag{2.5}$$

This is another example of an affine transformation. For the three dimensional case, there are three orthogonal planes that an object can be rotated in. For simplicity, the planes of rotation are normally expressed as being around the axes. A rotation of q_1 radians about the first (x) axis is normally called pitch, and is performed by:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos(q_1) & sin(q_1) & 0 \\ 0 & -sin(q_1) & cos(q_1) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix} \tag{2.6}$$

Similarly, rotations about the second (y) and third (z) axes (called roll and yaw respectively) are carried out by the following matrices:

$$\begin{bmatrix} cos(q_2) & 0 & sin(q_2) & 0 \\ 0 & 1 & 0 & 0 \\ -sin(q_2) & 0 & cos(q_2) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} cos(q_3) & sin(q_3) & 0 & 0 \\ -sin(q_3) & cos(q_3) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Rotations are combined by multiplying these matrices together in the appropriate order. The order in which the operations are performed is important. For example, a rotation about the first axis of $\pi/2$ radians followed by an equivalent rotation about the second axis would produce a very different result to that obtained if the order of the operations was reversed.

Zooms

The affine transformations described so far will perform purely rigid mappings. Zooms are needed to change the size of an image, or to work with images whos voxel sizes are not isotropic, or differ between images. These merely represent scalings along the orthogonal axes, and can be represented via:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ 1 \end{bmatrix} = \begin{bmatrix} q_1 & 0 & 0 & 0 \\ 0 & q_2 & 0 & 0 \\ 0 & 0 & q_3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix} \tag{2.7}$$

A single zoom by a factor of -1 will flip the object (see Section 2.2.3). Two flips in different directions will merely rotate the object by π radians (a rigid body transformation). In fact, any affine transformation that has a negative determinant will render the object flipped.

Shears

Shearing by parameters q_1, q_2 and q_3 can be performed by the following matrix:

$$\begin{bmatrix} 1 & q_1 & q_2 & 0 \\ 0 & 1 & q_3 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

A shear by itself is not a rigid body transformation, but it is possible to combine shears in order to generate rigid rotations. For a simple two dimensional case, a matrix encoding a rotation of θ radians about the origin (see Section 2.5) can be constructed by multiplying together three matrices that effect shears:

$$\begin{bmatrix} cos(\theta) & sin(\theta) & 0 \\ -sin(\theta) & cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \equiv \begin{bmatrix} 1 & tan(\theta/2) & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ sin(\theta) & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & tan(\theta/2) & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.8}$$

This approach has been useful for rigid registration of MR images (Eddy *et al.*, 1996), and subsequently improved by a more efficient reformulation for three dimensional transformations (Cox & Jesmanowicz, 1999).

2.2.1 Parameterising a Rigid Body Transformation

When registering a pair of images together via a rigid body transformation, it is necessary to estimate six parameters that describe the rigid-body transformation matrix. There are many ways of parameterising a rigid body transformation in terms of six parameters ($\mathbf{q}$), but the parameterisation chosen here is:

$$\mathbf{M} = \mathbf{TR} \tag{2.9}$$

where:

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & q_1 \\ 0 & 1 & 0 & q_2 \\ 0 & 0 & 1 & q_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.10}$$

and:

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos(q_4) & sin(q_4) & 0 \\ 0 & -sin(q_4) & cos(q_4) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} cos(q_5) & 0 & sin(q_5) & 0 \\ 0 & 1 & 0 & 0 \\ -sin(q_5) & 0 & cos(q_5) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} cos(q_6) & sin(q_6) & 0 & 0 \\ -sin(q_6) & cos(q_6) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.11}$$

Extracting the parameters $\mathbf{q}$ from $\mathbf{M}$ is relatively straightforward. Determining the translations is trivial, as they are simply contained in the fourth column of $\mathbf{M}$. This just leaves the rotations:

$$\mathbf{R} = \begin{bmatrix} c_5c_6 & c_5s_6 & s_5 & 0 \\ -s_4s_5c_6 - c_4s_6 & -s_4s_5s_6 + c_4c_6 & s_4c_5 & 0 \\ -c_4s_5c_6 + s_4s_6 & -c_4s_5s_6 - s_4c_6 & c_4c_5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.12}$$

where s_4, s_5 and s_6 are the sines, and c_4, c_5 and c_6 are the cosines of parameters q_4, q_5 and q_6 respectively. Therefore, providing that c_5 is not zero, then:

$$\begin{aligned} q_5 &= \quad sin^{-1}(r_{13}) \\ q_4 &= \quad atan2(r_{23}/cos(q_5), r_{33}/cos(q_5)) \\ q_6 &= \quad atan2(r_{12}/cos(q_5), r_{11}/cos(q_5)) \end{aligned} \tag{2.13}$$

where $atan2$ is the four quadrant inverse tangent. See Section 6.4 for more on decomposing affine transformations containing zooms and shears.

2.2.2 Working with Volumes of Differing or Anisotropic Voxel Sizes

Image voxel sizes need be considered when performing rigid body registration. Often, the images (say $\mathbf{f}$ and $\mathbf{g}$) will have voxels that are anisotropic. The dimensions of the voxels are also likely to differ between images of different modalities. For simplicity, a Euclidean space is used, where measures of distance are expressed in millimetres. Rather than interpolating the images such that the voxels are cubic and have the same dimensions in all images, one can simply define affine transformation matrices that map from voxel co-ordinates into this Euclidean space. For example, if image $\mathbf{f}$ is of size $128 \times 128 \times 43$ and has voxels that are $2.1mm \times 2.1mm \times 2.45mm$, the following matrix can be defined:

$$\mathbf{M_f} = \begin{bmatrix} 2.1 & 0 & 0 & -134.4 \\ 0 & 2.1 & 0 & -134.4 \\ 0 & 0 & 2.45 & -52.675 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.14}$$

This transformation matrix maps voxel co-ordinates to a Euclidean space who's axes are parallel to those of the image and distances are measured in millimetres, with the origin at the centre of the image. A similar matrix can be defined for $\mathbf{g}$ ($\mathbf{M_g}$). Because modern MR image formats such as SPI (Standard Product Interconnect) generally contain information about image orientations in their headers, it is possible to extract this information to automatically compute values for $\mathbf{M_f}$ or $\mathbf{M_g}$. This makes it possible to easily register images together that were originally acquired in completely different orientations.

The objective of any co-registration is to determine the rigid body transformation that maps the co-ordinates of image $\mathbf{g}$, to that of $\mathbf{f}$. To accomplish this, a rigid body transformation matrix $\mathbf{M_r}$ is determined, such that $\mathbf{M_f}^{-1}\mathbf{M_r}^{-1}\mathbf{M_g}$ will map from voxels in $\mathbf{g}$ to those in $\mathbf{f}$. The inverse of this matrix maps from $\mathbf{f}$ to $\mathbf{g}$. Once $\mathbf{M_r}$ has been determined, $\mathbf{M_f}$ can be set to $\mathbf{M_r M_f}$. From there onwards the mapping between the voxels of the two images can be achieved by $\mathbf{M_f}^{-1}\mathbf{M_g}$. Similarly, if another image ($\mathbf{h}$) is also co-registered to image $\mathbf{g}$ in the same manner, then not only is there a mapping from $\mathbf{h}$ and $\mathbf{g}$ (via $\mathbf{M_g}^{-1}\mathbf{M_h}$), but there is also one from $\mathbf{h}$ to $\mathbf{f}$ which is simply $\mathbf{M_f}^{-1}\mathbf{M_h}$ (derived from $\mathbf{M_f}^{-1}\mathbf{M_g}\mathbf{M_g}^{-1}\mathbf{M_h}$).

2.2.3 Left- and Right-handed Co-ordinate Systems

Positions in space can be represented in either a left- or right-handed co-ordinate system (see Figure 2.1), where one system is a mirror image of the other. For example, the system used by the Talairach Atlas (Talairach & Tournoux, 1988) is right-handed, because the first dimension

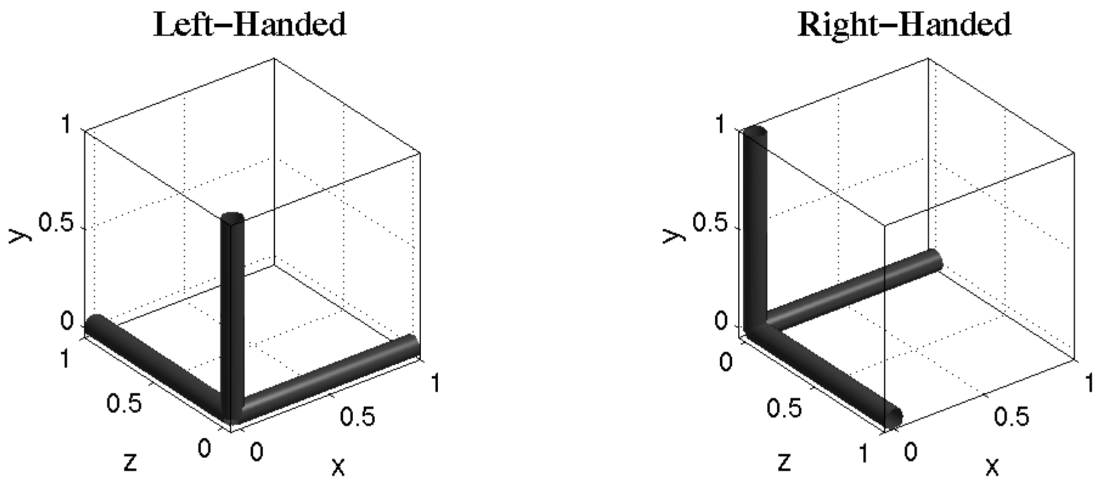

Figure 2.1: Left- and right-handed co-ordinate systems.

(often referred to as the x direction) increases from left to right, the second dimension goes from posterior to anterior (back to front) and the third dimension increases from inferior to superior (bottom to top). The axes can be rotated by any angle, and they still retain their handedness. An affine transformation that maps between left and right-handed co-ordinate systems has a negative determinant, whereas one that maps between co-ordinate systems of the same kind will have a positive determinant. Because the left and right sides of a brain have similar appearances, care must be taken when reorienting brain image volumes. Consistency of the co-ordinate systems can be achieved by performing any reorientations using affine transformations, and checking the determinants of the transformation matrices.

2.3 Resampling Images

Once there is a mapping between the original and transformed co-ordinates of an image, it is necessary to resample the image in order to apply the spatial transform. Spatially transforming images is usually implemented as a "pulling" operation (where pixel values are pulled from the original image into their new location) rather than a "pushing" one (where the pixels in the original image are pushed into their new location). This involves determining for each voxel in the transformed image, the corresponding intensity in the original image. Usually, this requires sampling between the centres of voxels, so some form of interpolation is needed.

The simplest approach is to take the value of the closest neighbouring voxel. This is referred to as *nearest neighbour* or *zero-order hold* resampling. This has the advantage that the original voxel intensities are preserved, but the resulting image is degraded quite considerably.

Another approach is to use *tri-linear interpolation* (*first-order hold*) to resample the data. This is slightly slower than nearest neighbour, but the resulting images have a less "blocky" appearance. However, tri-linear interpolation has the effect of losing some high frequency information from the image.

Figure 2.2 will now be used to illustrate bilinear interpolation in two dimensions. Assuming

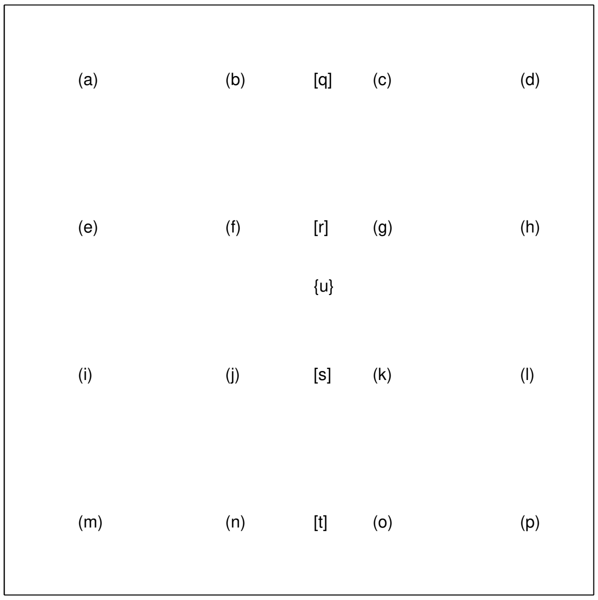

Figure 2.2: Illustration of image interpolation in two dimensions. Points a through to p represent the original regular grid of pixels. Point u is the point who's value is to be determined. Points q to t are used as intermediates in the computation.

that there is a regular grid of pixels at co-ordinates x_a, y_a to x_p, y_p, having intensities v_a to v_p, and that the point to resample is at u. The value at points r and s are first determined (using linear interpolation) as follows:

$$\begin{aligned} v_r &= \frac{(x_g - x_r)v_f + (x_r - x_f)v_g}{x_g - x_f} \\ v_s &= \frac{(x_k - x_s)v_j + (x_s - x_j)v_k}{x_k - x_j} \end{aligned} \tag{2.15}$$

Then v_u is determined by interpolating between v_r and v_s:

$$v_u = \frac{(y_u - y_s)v_r + (y_r - y_u)v_s}{y_r - y_s} \tag{2.16}$$

The extension of the approach to three dimensions is trivial.

Rather than using only the 8 nearest neighbours (in 3D) to estimate the value at a point, more neighbours can be used in order to fit a smooth function through the points, and then read off the value of the function at the desired location. *Polynomial interpolation* is one such approach (zero- and first-order hold interpolations are simply low order polynomial interpolations). It is now illustrated how v_q can be determined from pixels a to d. The coefficients ($\mathbf{q}$) of a polynomial that runs through these points can be obtained by computing:

$$\mathbf{q} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & (x_b - x_a) & (x_b - x_a)^2 & (x_b - x_a)^3 \\ 1 & (x_c - x_a) & (x_c - x_a)^2 & (x_c - x_a)^3 \\ 1 & (x_d - x_a) & (x_d - x_a)^2 & (x_d - x_a)^3 \end{bmatrix}^{-1} \begin{bmatrix} v_a \\ v_b \\ v_c \\ v_d \end{bmatrix} \tag{2.17}$$

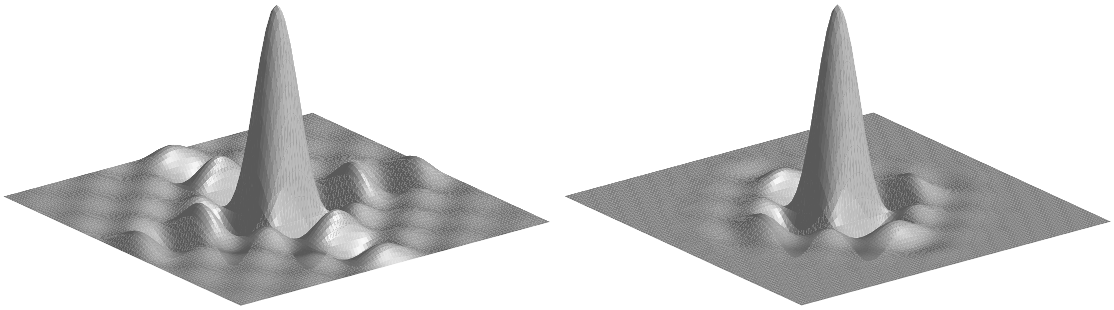

Figure 2.3: Sinc function in two dimensions, both with (right) and without (left) a Hanning window.

Then v_q can be determined from these coefficients by:

$$v_q = \begin{bmatrix} 1 & (x_q - x_a) & (x_q - x_a)^2 & (x_q - x_a)^3 \end{bmatrix} \mathbf{q} \tag{2.18}$$

To determine v_u, a similar polynomial would be fitted through points q, r, s and t. The Vandermonde matrices required for polynomial interpolation are very ill conditioned, especially for higher orders. A better way of doing polynomial interpolation involves using *Lagrange polynomials* (see Press *et al.*(1992) or Jain (1989)).

The optimum method of applying rigid body transformations to images with minimal interpolation artifact is to do it in Fourier space. In real space, the interpolation method that gives results closest to a Fourier interpolation is *sinc* interpolation. This involves convolving the image with a sinc function centred on the point to be resampled. To perform a pure sinc interpolation, every voxel in the image should be used to sample a single point. This is not feasible due to speed considerations, so an approximation using a limited number of nearest neighbours is used. Because the sinc function extends to infinity, it is often truncated by modulating with a Hanning window (see Figure 2.3). Because the function is separable, the implementation of sinc interpolation is similar to that for polynomial interpolation, in that it is performed sequentially in the three dimensions of the volume. For one dimension the windowed sinc function using the I nearest neighbours would be:

$$\sum_{i=1}^{I} v_i \frac{\frac{sin(\pi d_i)}{\pi d_i} \frac{1}{2} \left(1 + cos\left(2\pi d_i / I\right)\right)}{\sum_{j=1}^{I} \frac{sin(\pi d_j)}{\pi d_j} \frac{1}{2} \left(1 + cos\left(2\pi d_j / I\right)\right)} \tag{2.19}$$

where d_i is the distance from the centre of the ith voxel to the point to be sampled, and v_i is the value of the ith voxel.

Sinc interpolation is slow when many neighbouring voxels are used. A slightly better alternative may be to use a Fourier interpolation method for effecting a rigid body transformation. In one dimension, a translation is simply a convolution with a translated delta function. For translations that are not whole numbers of pixels, the delta function is replaced by a sinc function centred at the translation distance. The use of fast Fourier transforms means that the convolution can be performed most rapidly as a multiplication in Fourier space. It is clear how translations can be performed in this way, but rotations are less obvious. One way that rotations can be effected involves replacing the rotations by a series of shears as described previously (Section 2.2). A shear simply involves translating different rows or columns of an image by different amounts, so each shear can be performed by a series of one dimensional convolutions in Fourier space. Alternatively, the method of rotating and translating using shears can also be done using a windowed sinc or polynomial interpolation. Each interpolation is in just one dimension, requiring much less computation than it would in three dimensions.

In addition to resampling images, many image registration methods also require the image gradients to be computed. This procedure is similar to the straightforward interpolation methods described above.

2.4 Optimisation

The objective of optimisation is to determine the values for a set of parameters for which some function of the parameters is minimised (or maximised). One of the simplest cases involves determining the optimum parameters for a model in order to minimise the sum of squared differences between a model and a set of real world data (χ^2). Normally there are many parameters, and it is not possible to exhaustively search through the whole parameter space. The usual approach is to make an initial parameter estimate, and begin iteratively searching from there. At each iteration, the model is evaluated using the current parameter estimates, and χ^2 computed. A judgement is then made about how the parameter estimates should be modified, before continuing on to the next iteration. The optimisation is terminated when some convergence criterion is achieved (usually when χ^2 stops decreasing).

The image registration approach described here is essentially an optimisation. One image (the source image) is spatially transformed so that it matches another (the target image), by minimising χ^2. The parameters that are optimised are those that describe the spatial transformation (although there are often other nuisance parameters required by the model, such as intensity scaling parameters). For rigid registration, the algorithm chosen (Friston *et al.*, 1995c) is *Gauss-Newton* optimisation, and it is illustrated here:

Suppose that $b_i(\mathbf{q})$ is the function describing the difference between the source and target images at voxel i, when the vector of model parameters have values $\mathbf{q}$. For each voxel, a first approximation of Taylor's Theorem can be used to estimate the value that this difference will take if the parameters $\mathbf{q}$ are decreased by $\mathbf{t}$:

$$b_i(\mathbf{q}-\mathbf{t}) \simeq b_i(\mathbf{q}) - t_1\frac{\partial b_i(\mathbf{q})}{\partial q_1} - t_2\frac{\partial b_i(\mathbf{q})}{\partial q_2}\dots \tag{2.20}$$

This allows the construction of a set of simultaneous equations (of the form $\mathbf{At} \simeq \mathbf{b}$) for estimating the values that $\mathbf{t}$ should assume to in order to minimise $\sum_i b_i(\mathbf{q}-\mathbf{t})^2$:

$$\begin{bmatrix} \frac{\partial b_1(\mathbf{q})}{\partial q_1} & \frac{\partial b_1(\mathbf{q})}{\partial q_2} & \dots \\ \frac{\partial b_2(\mathbf{q})}{\partial q_1} & \frac{\partial b_2(\mathbf{q})}{\partial q_2} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ \vdots \end{bmatrix} \simeq \begin{bmatrix} b_1(\mathbf{q}) \\ b_2(\mathbf{q}) \\ \vdots \end{bmatrix} \tag{2.21}$$

From this, an iterative scheme can be derived for improving the parameter estimates. For iteration n, the parameters $\mathbf{q}$ are updated as:

$$\mathbf{q}^{(n+1)} = \mathbf{q}^{(n)} - \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b} \tag{2.22}$$

where $\mathbf{A} = \begin{bmatrix} \frac{\partial b_1(\mathbf{q})}{\partial q_1} & \frac{\partial b_1(\mathbf{q})}{\partial q_2} & \cdots \\ \frac{\partial b_2(\mathbf{q})}{\partial q_1} & \frac{\partial b_2(\mathbf{q})}{\partial q_2} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1(\mathbf{q}) \\ b_2(\mathbf{q}) \\ \vdots \end{bmatrix}$.

This process is repeated until χ^2 can no longer be decreased - or for a fixed number of iterations. There is no guarantee that the best global solution will be reached, because the algorithm can get caught in a local minimum. To reduce this problem, the starting estimates for $\mathbf{q}$ should be set as close as possible to the optimum solution. The number of potential local minima can also be decreased by working with smooth images. This also has the effect of making the first order Taylor approximation more accurate for larger displacements. Once the registration is close to the final solution, it can continue with less smooth images.

In practice, $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}^T\mathbf{b}$ from Eqn. 2.22 are often computed 'on the fly' for each iteration. By computing these matrices using only a few rows of $\mathbf{A}$ and $\mathbf{b}$ at a time, much less computer memory is required than is necessary for storing the whole of matrix $\mathbf{A}$. Also, the partial derivatives $\partial b_i(\mathbf{q})/\partial q_j$ can be rapidly computed from the gradients of the images using the chain rule.

It should be noted that element i of $\mathbf{A}^T\mathbf{b}$ is equal to $\frac{1}{2}\frac{\partial\chi^2}{\partial q_i}$, and that element i,j of $\mathbf{A}^T\mathbf{A}$ is approximately equal to $\frac{1}{2}\frac{\partial^2\chi^2}{\partial q_i \partial q_j}$ (one half of the Hessian matrix, often referred to as the curvature matrix - see Press *et al.*(1992), Section 15.5). Another way of thinking about the optimisation is that it fits a quadratic function to the error surface at each iteration. Successive parameter estimates are chosen such that they are at the minimum point of this quadratic (illustrated for a single parameter in Figure 2.4).

2.5 Within Modality Image Registration

Within modality image registration has a number of uses, both within morphometry and for processing functional images. Morphometric studies sometimes involve looking at changes in brain shape over time, often to study the progression of a disease such as Altzheimers, or to monitor tumour growth or shrinkage. Differences between structural MR scans acquired at different times are identified, by first co-registering the images and then looking at the difference between the registered images. Rigid registration can also be used as a pre-processing step before using nonlinear registration methods for identifying shape changes (Freeborough & Fox, 1998).

The most common application of within modality registration in functional imaging is to reduce motion artifacts by realigning the volumes in image time-series. The objective of realignment is to determine the rigid body transformations that best map the series of functional images to the same space. This can be achieved by minimising the sum of squared differences between each of the images and a reference image, where the reference image could be one of the images in the series. For slightly better results, this procedure could be repeated, but instead of matching to one of the images from the series, the images would be registered to the mean of all the realigned images. Because of the nonstationary variance in the images, a variance image could be computed

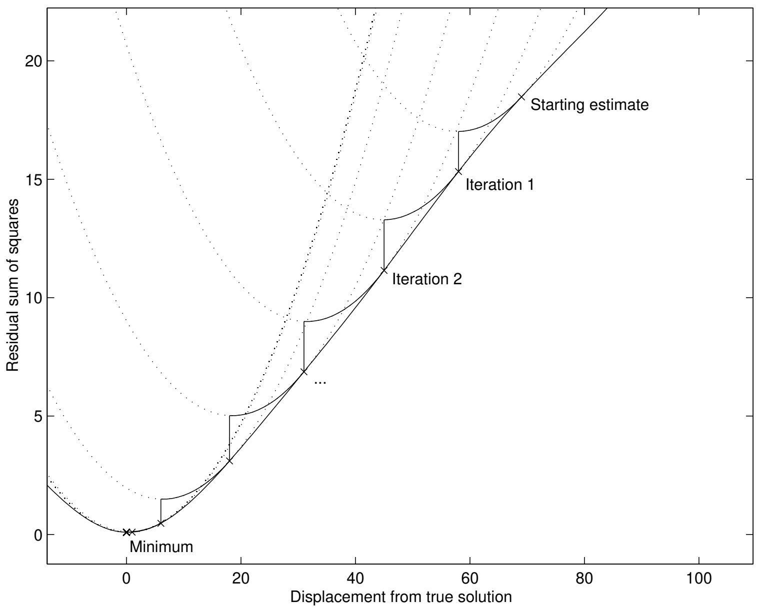

Figure 2.4: The optimisation can be thought of as fitting a series of quadratics to the error surface. Each parameter update is such that it falls at the minimum of the quadratic.

at the same time as the mean, in order to provide better weighting for the registration. Voxels with a lot of variance should be given lower weighting, whereas those with less variance should be weighted more highly.

2.5.1 Methods

To register a source image $\mathbf{f}$ to a reference image $\mathbf{g}$, a six parameter rigid body transformation (parameterised by q_1 to q_6) would be used. To perform the registration, a number of points in the reference image (each denoted by $\mathbf{x}_i$) are compared with points in the source image (denoted by $\mathbf{Mx}_i$, where $\mathbf{M}$ is the rigid body transformation matrix constructed from the six parameters). The images may be scaled differently, so an additional intensity scaling parameter (q_7) may be included in the model. The parameters ($\mathbf{q}$) are optimised by minimising the sum of squared differences[1] between the images according to the algorithm described in Sections 2.2.1 and 2.4 (Eqn. 2.22). The function that is minimised is:

$$\sum_i (f(\mathbf{Mx}_i) - q_7 g(\mathbf{x}_i))^2 \tag{2.23}$$

[1]Strictly speaking, it is the mean squared difference that is minimised, rather than the sum of squared differences. Inevitably, some values of $\mathbf{Mx}_i$ will lie outside the domain of $\mathbf{f}$, so nothing is known about what the image intensity should be at these points. The computations are only performed for points where both $\mathbf{x}_i$ and $\mathbf{Mx}_i$ lie within the field of view of the images.

where $\mathbf{M} = \mathbf{M_f}^{-1}\mathbf{M_r}^{-1}\mathbf{M_g}$, and $\mathbf{M_r}$ is constructed from parameters $\mathbf{q}$ (refer to Section 2.2.2). Vector $\mathbf{b}$ is generated for each iteration as:

$$\mathbf{b} = \begin{bmatrix} f(\mathbf{Mx}_1) - q_7 g(\mathbf{x}_1) \\ f(\mathbf{Mx}_2) - q_7 g(\mathbf{x}_2) \\ \vdots \end{bmatrix} \tag{2.24}$$

Each column of matrix $\mathbf{A}$ is constructed by differentiating $\mathbf{b}$ with respect to parameters q_1 to q_7:

$$\mathbf{A} = \begin{bmatrix} \frac{\partial f(\mathbf{Mx}_1)}{\partial q_1} & \frac{\partial f(\mathbf{Mx}_1)}{\partial q_2} & \cdots & \frac{\partial f(\mathbf{Mx}_1)}{\partial q_6} & -g(\mathbf{x}_1) \\ \frac{\partial f(\mathbf{Mx}_2)}{\partial q_1} & \frac{\partial f(\mathbf{Mx}_2)}{\partial q_2} & \cdots & \frac{\partial f(\mathbf{Mx}_2)}{\partial q_6} & -g(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \end{bmatrix} \tag{2.25}$$

Because non-singular affine transformations are easily invertible, it is possible to make the registration more robust by also considering what happens with the inverse transformation. By swapping around the source and reference image, the registration problem also becomes one of minimising:

$$\sum_j (g(\mathbf{M}^{-1}\mathbf{y}_j) - q_7^{-1} f(\mathbf{y}_j))^2 \tag{2.26}$$

In theory, a more robust solution could be achieved by simultaneously including the inverse transformation to make the registration problem symmetric (Woods *et al.*, 1998a). The cost function would then be:

$$\lambda_1 \sum_i (f(\mathbf{Mx}_i) - q_7 g(\mathbf{x}_i))^2 + \lambda_2 \sum_j (g(\mathbf{M}^{-1}\mathbf{y}_j) - q_7^{-1} f(\mathbf{y}_j))^2 \tag{2.27}$$

Normally, the intensity scaling of the image pair will be similar, so equal values for the weighting factors (λ_1 and λ_2) can be used. Matrix $\mathbf{A}$ and vector $\mathbf{b}$ would then be formulated as:

$$\mathbf{b} = \begin{bmatrix} \lambda_1^{\frac{1}{2}}(f(\mathbf{Mx}_1) - q_7 g(\mathbf{x}_1)) \\ \lambda_1^{\frac{1}{2}}(f(\mathbf{Mx}_2) - q_7 g(\mathbf{x}_2)) \\ \vdots \\ \lambda_2^{\frac{1}{2}}(g(\mathbf{M}^{-1}\mathbf{y}_1) - q_7^{-1} f(\mathbf{y}_1)) \\ \lambda_2^{\frac{1}{2}}(g(\mathbf{M}^{-1}\mathbf{y}_2) - q_7^{-1} f(\mathbf{y}_2)) \\ \vdots \end{bmatrix} \tag{2.28}$$

and

$$\mathbf{A} = \begin{bmatrix} \lambda_1^{\frac{1}{2}} \frac{\partial f(\mathbf{Mx}_1)}{\partial q_1} & \cdots & \lambda_1^{\frac{1}{2}} \frac{\partial f(\mathbf{Mx}_1)}{\partial q_6} & -\lambda_1^{\frac{1}{2}} g(\mathbf{x}_1) \\ \lambda_1^{\frac{1}{2}} \frac{\partial f(\mathbf{Mx}_2)}{\partial q_1} & \cdots & \lambda_1^{\frac{1}{2}} \frac{\partial f(\mathbf{Mx}_2)}{\partial q_6} & -\lambda_1^{\frac{1}{2}} g(\mathbf{x}_2) \\ \vdots & \ddots & \vdots & \vdots \\ \lambda_2^{\frac{1}{2}} \frac{\partial g(\mathbf{M}^{-1}\mathbf{y}_1)}{\partial q_1} & \cdots & \lambda_2^{\frac{1}{2}} \frac{\partial g(\mathbf{M}^{-1}\mathbf{y}_1)}{\partial q_6} & \lambda_2^{\frac{1}{2}} q_7^{-2} f(\mathbf{y}_1) \\ \lambda_2^{\frac{1}{2}} \frac{\partial g(\mathbf{M}^{-1}\mathbf{y}_2)}{\partial q_1} & \cdots & \lambda_2^{\frac{1}{2}} \frac{\partial g(\mathbf{M}^{-1}\mathbf{y}_2)}{\partial q_6} & \lambda_2^{\frac{1}{2}} q_7^{-2} f(\mathbf{y}_2) \\ \vdots & \ddots & \vdots & \vdots \end{bmatrix} \tag{2.29}$$

Symmetric formulation of registration problems is a theme that will be returned to in Chapter 4.

2.5.2 Residual Artifacts from PET and fMRI

Even after realignment, there may still be some motion related artifacts remaining in functional data. After retrospective realignment of PET images with large movements, the primary source of error is due to incorrect attenuation correction. In emission tomography methods, many photons are not detected because they are attenuated by the subject's head. Normally, a transmission scan (using a moving radioactive source external to the subject) is acquired before collecting the emission scans. The ratio of the number of detected photon pairs from the source, with and without a head in the field of view, produces a map of the proportion of photons that are absorbed along any line-of-response. If a subject moves between the transmission and emission scans, then the applied attenuation correction is incorrect because the emission scan is no longer aligned with the transmission scan. There are methods for correcting these errors (Andersson *et al.*, 1995), but they are beyond the scope of this thesis.

In fMRI, there are many sources of motion related artifacts. The most obvious ones are:

- Interpolation error from the resampling algorithm used to transform the images can be one of the main sources of motion related artifacts. When the image series is resampled, it is important to use a very accurate interpolation method such as sinc or Fourier interpolation.
- When MR images are reconstructed, the final images are usually the modulus of the initially complex data, resulting in any voxels that should be negative being rendered positive. This has implications when the images are resampled, because it leads to errors at the edge of the brain that can not be corrected however good the interpolation method is. Possible ways to circumvent this problem are to work with complex data, or possibly to apply a low pass filter to the complex data before taking the modulus.
- The sensitivity (slice selection) profile of each slice also plays a role in introducing artifacts (Noll *et al.*, 1997).
- fMRI images are spatially distorted, and the amount of distortion depends partly upon the position of the subject's head within the magnetic field. Relatively large subject movements result in the brain images changing shape, and these shape changes can not be corrected by a rigid body transformation (Jezzard & Clare, 1999).
- Each fMRI volume of a series is currently acquired a plane at a time over a period of a few seconds. Subject movement between acquiring the first and last plane of any volume leads to another reason why the images may not strictly obey the rules of rigid body motion.
- After a slice is magnetised, the excited tissue takes time to recover to its original state, and the amount of recovery that has taken place will influence the intensity of the tissue in the image. Out of plane movement will result in a slightly different part of the brain being excited during each repeat. This means that the spin excitation will vary in a way that is related to head motion, and so leads to more movement related artifacts.
- Ghost artifacts in the images do not obey the same rigid body rules as the head, so a rigid rotation to align the head will not mean that the ghosts are aligned.
- The accuracy of the estimated registration parameters is normally in the region of tens of μm. This is dependent upon many factors, including the effects just mentioned. Even

the signal changes elicited by the experiment can have a slight effect (a few μm) on the estimated parameters.

These problems can not be corrected by simple image realignment, and so may be sources of possible stimulus correlated motion artifacts. Systematic movement artifacts resulting in a signal change of only one or two percent can lead to highly significant false positives over an experiment with many scans. This is especially important for experiments where some conditions may cause slight head movements (such as motor tasks, or speech), because these movements are likely to be highly correlated with the experimental design. In cases like this, it is difficult to separate true activations from stimulus correlated motion artifacts. Providing there are enough images in the series and the movements are small, some of these artifacts can be removed by using an ANCOVA model to remove any signal that is correlated with functions of the movement parameters (Friston *et al.*, 1996). However, when the estimates of the movement parameters are related to the the experimental design, it is likely that much of the true fMRI signal will also be lost. These are still unresolved problems.

2.6 Between Modality Image Registration

Co-registration of brain images of the same subject acquired in different modalities has proved itself to be useful in many areas, both in research and clinically. Two images from the same subject acquired using the same modality or scanning sequences generally look similar, so it suffices to find the rigid-body transformation parameters that minimise the sum of squared differences between them. However, for co-registration between modalities there is nothing quite as obvious to minimise.

Older methods of registration involved the manual identification of homologous landmarks in the images. These landmarks are aligned together, thus bringing the images into registration. This is time-consuming, requires a degree of experience, and can be rather subjective. One of the first widely used semi-automatic co-registration methods was that known as the "head-hat" approach (Pelizzari *et al.*, 1988). This method involved extracting brain surfaces of the two images, and then matching the surfaces together. Another method that has been widely used for a number of years for registering PET to MR images is AIR (Woods *et al.*, 1992). This method divides the MR images into a number of partitions based on intensity. The registration is based on minimising the standard deviation of the corresponding PET voxel intensities for each partition. It makes a number of assumptions about how the PET intensity varies with the MRI intensity, which are generally valid within the brain, but do not work when non-brain tissue is included. Because of this, the method has the disadvantage of requiring the MR images to be pre-processed, which normally involves laborious manual editing in order to remove any non-brain tissue.

More recently, the idea of matching images by maximising the *mutual information* in their histograms is becoming more widespread (Collignon *et al.*, 1995). For this elegant approach, the 2D histogram is normalised so that the bins integrate to unity. This is considered as an I by J matrix ($\mathbf{H}$, see Figure 3.7 for examples based on smooth images), and the registration involves maximising the following objective function (where element i, j of $\mathbf{H}$ is denoted by h_{ij}):

$$\sum_{j=1}^{J}\sum_{i=1}^{I} h_{ij} log \left(\frac{h_{ij}}{(\sum_{k=1}^{J} h_{ik})(\sum_{l=1}^{I} h_{lj})} \right) \tag{2.30}$$

Maximising mutual information is a very general approach, which has been successfully applied to the registration of a wide variety of imaging modalities.

The work developed here concentrates on a different fully automatic method of registering magnetic resonance (MR) images with positron emission tomography (PET) images, and on registering MR images from different scanning sequences. The assumption that it makes is that brains consist of two tissue types (grey and white matter) that can clearly be identified from the images.

2.6.1 Methods

This image co-registration method relies on images other than those that are to be registered ($\mathbf{f}$ and $\mathbf{g}$). These are template images of the same modalities ($\mathbf{t_f}$ and $\mathbf{t_g}$ respectively), and prior probability images of grey matter (GM), white matter (WM) and cerebro-spinal fluid (CSF). The template images and probability images conform to the same anatomical space, and examples of these are shown in Figure 3.6 on page 55.

The between modality co-registration is a three step approach that essentially reduces the problem to a series of within-modality approaches:

1. *Determine the affine transformations that map between the images and templates* by minimising the sum of squared differences between $\mathbf{f}$ and $\mathbf{t_f}$, and $\mathbf{g}$ and $\mathbf{t_g}$. These transformations are constrained such that only parameters that describe rigid body component are allowed to differ between the two registrations.

2. *Segment or partition the images using the probability images* and a modified *mixture model* algorithm (described in Chapter 5). The mapping between the probability images to images $\mathbf{f}$ and $\mathbf{g}$ having been determined in step 1.

3. *Co-register the image partitions* generated by the previous step using the rigid body transformations computed from step 1 as a starting estimate an using a within modality approach.

Determining the mappings from images to templates

It is possible to obtain a reasonable match of images of most normal brains to a template image using just a twelve parameter affine transformation. One can register image $\mathbf{g}$ to template $\mathbf{t_g}$, and similarly register $\mathbf{f}$ to $\mathbf{t_f}$ using this approach. These transformation matrices will be called $\mathbf{M_{gt}}$ and $\mathbf{M_{ft}}$ respectively. Thus a mapping from voxels in $\mathbf{g}$ to those in $\mathbf{f}$ is $\mathbf{M_f}^{-1}\mathbf{M_{ft}}\mathbf{M_{gt}}^{-1}\mathbf{M_g}$ (see Section 2.2.2 for a description of matrices $\mathbf{M_g}$ and $\mathbf{M_f}$). However, this affine transformation between $\mathbf{f}$ and $\mathbf{g}$ has not been constrained to be rigid body. This simple approach is modified in order to incorporate this constraint, by decomposing matrix $\mathbf{M_{gt}}$ into matrices that perform a rigid body transformation ($\mathbf{M_{gr}}$), and one that performs the scaling and shearing ($\mathbf{M_{ta}}$). i.e., $\mathbf{M_{gt}} = \mathbf{M_{gr}}\mathbf{M_{ta}}$. Similarly $\mathbf{M_{ft}} = \mathbf{M_{fr}}\mathbf{M_{ta}}$, where $\mathbf{M_{ta}}$ is common to both $\mathbf{M_{gt}}$ and $\mathbf{M_{ft}}$ so that the same zooms and shears are used for registering both images to their respective templates. Now the mapping becomes $\mathbf{M_f}^{-1}\mathbf{M_{fr}}(\mathbf{M_{ta}}\mathbf{M_{ta}}^{-1})\mathbf{M_{gr}}^{-1}\mathbf{M_g}$, and is a rigid body transformation. These matrices are parameterised by 18 elements of a vector $\mathbf{q}$.

$$\mathbf{M_{gr}} = \begin{bmatrix} 1 & 0 & 0 & q_1 \\ 0 & 1 & 0 & q_2 \\ 0 & 0 & 1 & q_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos(q_4) & sin(q_4) & 0 \\ 0 & -sin(q_4) & cos(q_4) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times$$
$$\begin{bmatrix} cos(q_5) & 0 & sin(q_5) & 0 \\ 0 & 0 & & 0 \\ -sin(q_5) & 0 & cos(q_5) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} cos(q_6) & sin(q_6) & 0 & 0 \\ -sin(q_6) & cos(q_6) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
$$\mathbf{M_{fr}} = \begin{bmatrix} 1 & 0 & 0 & q_7 \\ 0 & 1 & 0 & q_8 \\ 0 & 0 & 1 & q_9 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos(q_{10}) & sin(q_{10}) & 0 \\ 0 & -sin(q_{10}) & cos(q_{10}) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times$$
$$\begin{bmatrix} cos(q_{11}) & 0 & sin(q_{11}) & 0 \\ 0 & 0 & & 0 \\ -sin(q_{11}) & 0 & cos(q_{11}) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} cos(q_{12}) & sin(q_{12}) & 0 & 0 \\ -sin(q_{12}) & cos(q_{12}) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
$$\mathbf{M_{ta}} = \begin{bmatrix} q_{13} & 0 & 0 & 0 \\ 0 & q_{14} & 0 & 0 \\ 0 & 0 & q_{15} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & q_{16} & q_{17} & 0 \\ 0 & 1 & q_{18} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.31}$$

The parameter set $\mathbf{q}$ can now be optimised in order to determine the transformations that minimise the sum of squares difference between the images and templates. The iterative optimisation method described in Section 2.4 is used, which generally converges within a few iterations. The chance of finding a local minimum is reduced by using smoothed data (typically by convolving with an 8mm full width at half maximum [FWHM] Gaussian kernel). Each iteration involves generating a linear approximation to the problem using Taylor's Theorem. This can be expressed as computing $\mathbf{q}^{(n+1)} = \mathbf{q}^{(n)} - \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}$. The vector $\mathbf{q}^{(n)}$ contains the parameter estimates at iteration n, and vector $\mathbf{b}$ contains the differences between the templates and the images that have been spatially transformed according to the latest parameter estimates. $\mathbf{A}$ is a matrix of derivatives of $\mathbf{b}$, with respect to changes to each element of $\mathbf{q}$. For the purpose of this optimisation, two matrices, $\mathbf{M_1} = \mathbf{M_f}^{-1}\mathbf{M_{ft}}^{-1}\mathbf{M_t}$ and $\mathbf{M_2} = \mathbf{M_g}^{-1}\mathbf{M_{gt}}^{-1}\mathbf{M_t}$, are defined:

$$\mathbf{b} = \begin{bmatrix} \lambda_1^{\frac{1}{2}}\left(f(\mathbf{M_1}\mathbf{x}_1) - q_{19}t_f(\mathbf{x}_1)\right) \\ \lambda_1^{\frac{1}{2}}\left(f(\mathbf{M_1}\mathbf{x}_2) - q_{19}t_f(\mathbf{x}_2)\right) \\ \vdots \\ \lambda_2^{\frac{1}{2}}\left(g(\mathbf{M_2}\mathbf{x}_1) - q_{20}t_g(\mathbf{x}_1)\right) \\ \lambda_2^{\frac{1}{2}}\left(g(\mathbf{M_2}\mathbf{x}_2) - q_{20}t_g(\mathbf{x}_2)\right) \\ \vdots \end{bmatrix} \tag{2.32}$$

$$\mathbf{A} = \begin{bmatrix} \lambda_1^{\frac{1}{2}} \frac{\partial f(\mathbf{M_1 x_1})}{\partial q_1} & \dots & 0 & \dots & \lambda_1^{\frac{1}{2}} \frac{\partial f(\mathbf{M_1 x_1})}{\partial q_{13}} & \dots & -\lambda_1^{\frac{1}{2}} t_f(\mathbf{x}_1) & 0 \\ \lambda_1^{\frac{1}{2}} \frac{\partial f(\mathbf{M_1 x_2})}{\partial q_1} & \dots & 0 & \dots & \lambda_1^{\frac{1}{2}} \frac{\partial f(\mathbf{M_1 x_2})}{\partial q_{13}} & \dots & -\lambda_1^{\frac{1}{2}} t_f(\mathbf{x}_2) & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots \vdots & \vdots & \\ 0 & \dots & \lambda_2^{\frac{1}{2}} \frac{\partial g(\mathbf{M_2 x_1})}{\partial q_7} & \dots & \lambda_2^{\frac{1}{2}} \frac{\partial g(\mathbf{M_2 x_1})}{\partial q_{13}} & \dots & 0 & -\lambda_2^{\frac{1}{2}} t_g(\mathbf{x}_1) \\ 0 & \dots & \lambda_2^{\frac{1}{2}} \frac{\partial g(\mathbf{M_2 x_2})}{\partial q_7} & \dots & \lambda_2^{\frac{1}{2}} \frac{\partial g(\mathbf{M_2 x_2})}{\partial q_{13}} & \dots & 0 & -\lambda_2^{\frac{1}{2}} t_g(\mathbf{x}_2) \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots \vdots & \vdots & \end{bmatrix} \tag{2.33}$$

The parameters describing the non-rigid transformations (q_{13} to q_{18}) could in theory be derived from either $\mathbf{f}$ or $\mathbf{g}$. In practice, a better solution is obtained by estimating these parameters using both images, and by biasing the result so that the image that fits the template better has a greater influence over the parameter estimates. This is achieved by weighting the rows of $\mathbf{A}$ and $\mathbf{b}$ that correspond to the different images. The weights (λ_1 and λ_2) are derived from the residual variance between the template and source images, obtained from the previous solution for $\mathbf{q}$. These are:

$$\lambda_1 = \frac{\nu_1}{\sum_{i=1}^{I}(f(\mathbf{M_1 x}_i) - q_{19} t_f(\mathbf{x}_i))^2} \text{ and } \lambda_2 = \frac{\nu_2}{\sum_{i=1}^{I}(g(\mathbf{M_2 x}_i) - q_{20} t_g(\mathbf{x}_i))^2} \tag{2.34}$$

where ν_1 and ν_2 are the degrees of freedom for the two parts of the problem. Also note that this first step is made more robust by including the regularisation that will be described in the next chapter.

Once the optimisation has converged to the final solution, the rigid body transformation that approximately maps between $\mathbf{g}$ and $\mathbf{f}$ can be obtained, and also the affine transformation matrices that map between the source images and templates. These are used by the next step.

Partitioning the Images

The result of the previous step includes affine mappings between source and template images, which are used to assist image segmentation by allowing prior probability images GM, WM and CSF to be automatically overlaid on to the source images. The extraction of GM and WM proceeds as described in Chapter 5 (although no correction for image intensity nonuniformity is included for the evaluations). The result of the partitioning are images representing the probability of the voxels belonging to each tissue class. All voxel intensities are between zero and one, and most lie close to one or the other extreme.

Co-registering the Image Partitions

The previous step produces images of GM and WM from the original images $\mathbf{f}$ and $\mathbf{g}$. These image partitions can then be simultaneously co-registered together to produce the final solution.

This optimisation stage only needs to search for the six parameters that describe a rigid body transformation. A voxel-to-voxel affine transform matrix $\mathbf{M}$ is defined by $(\mathbf{M_f}^{-1}\mathbf{M_{fg}}^{-1}\mathbf{M_g})$, where the rigid body transformation matrix $\mathbf{M_{fg}}$ is parameterised by q_1 to q_6. Starting estimates for $\mathbf{q}$ are obtained by extracting them from the rigid transformation matrix $\mathbf{M_{gr}}\mathbf{M_{fr}}^{-1}$ as described in Section 2.2. Convergence should be achieved within a few iterations because of the good starting estimates obtained from the first step. No scaling parameters are needed, because the probability images derived from $\mathbf{f}$ have similar intensities to those derived from $\mathbf{g}$. Again, the method described in Section 2.4 is used to optimise the parameters, where (using notation where $p_{g1}(\mathbf{x}_2)$ means 'probability of voxel at $\mathbf{x}_2$ from image $\mathbf{g}$ belonging to cluster 1') $\mathbf{b}$ and $\mathbf{A}$ are defined by:

$$\mathbf{b} = \begin{bmatrix} p_{f1}(\mathbf{Mx}_1) - p_{g1}(\mathbf{x}_1) \\ p_{f1}(\mathbf{Mx}_2) - p_{g1}(\mathbf{x}_2) \\ \vdots \\ p_{f2}(\mathbf{Mx}_1) - p_{g2}(\mathbf{x}_1) \\ p_{f2}(\mathbf{Mx}_2) - p_{g2}(\mathbf{x}_2) \\ \vdots \end{bmatrix} \tag{2.35}$$

$$\mathbf{A} = \begin{bmatrix} \frac{\partial p_{f1}(\mathbf{Mx}_1)}{\partial q_1} & \frac{\partial p_{f1}(\mathbf{Mx}_1)}{\partial q_2} & \cdots & \frac{\partial p_{f1}(\mathbf{Mx}_1)}{\partial q_6} \\ \frac{\partial p_{f1}(\mathbf{Mx}_2)}{\partial q_1} & \frac{\partial p_{f1}(\mathbf{Mx}_2)}{\partial q_2} & \cdots & \frac{\partial p_{f1}(\mathbf{Mx}_2)}{\partial q_6} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial p_{f2}(\mathbf{Mx}_1)}{\partial q_1} & \frac{\partial p_{f2}(\mathbf{Mx}_1)}{\partial q_2} & \cdots & \frac{\partial p_{f2}(\mathbf{Mx}_1)}{\partial q_6} \\ \frac{\partial p_{f2}(\mathbf{Mx}_2)}{\partial q_1} & \frac{\partial p_{f2}(\mathbf{Mx}_2)}{\partial q_2} & \cdots & \frac{\partial p_{f2}(\mathbf{Mx}_2)}{\partial q_6} \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix} \tag{2.36}$$

The final solution is obtained after this co-registration. It is now possible to map voxel $\mathbf{x}$ of image $\mathbf{g}$, to the corresponding voxel $\mathbf{Mx}$ of image $\mathbf{f}$. Examples of PET/T1-MRI and T1/T2-MRI co-registration using this approach are illustrated in figures 2.5 and 2.6.

2.6.2 Evaluation

The co-registration methods were evaluated for PET to T1 weighted MRI, using data from the "Evaluation of Retrospective Image Registration" project (National Institutes of Health, Project Number 1 R01 NS33926-01, Principal Investigator, J. Michael Fitzpatrick, Vanderbilt University, Nashville, TN.) (West *et al.*, 1996; West *et al.*, 1997). This involved obtaining both PET and MRI data from Vanderbilt University, and performing inter-modality registrations on the volumes. Fiducial markers during the acquisition of these datasets enabled investigators at Vanderbilt to know the true registration parameters, but any visible traces of these markers had been removed from the images prior to their distribution to other investigators.

Registrations were performed on 11 volume images. Four of the images had been geometry distortion corrected at Vanderbilt using their own software that uses a pair of distorted images acquired with reversed readout gradients. Seven of the images had not been corrected. The registrations were done using only the first step of the registration process (constrained simultaneous affine registration) and also using all the steps. The evaluations were performed on a Sun SPARC Ultra 2, using an implementation of the method written in C and Matlab (from The Mathworks, Natick, Mass., USA). The starting estimates for the registration parameters matched the centres of each volume together, and assumed that the images were in the same orientation. No manual

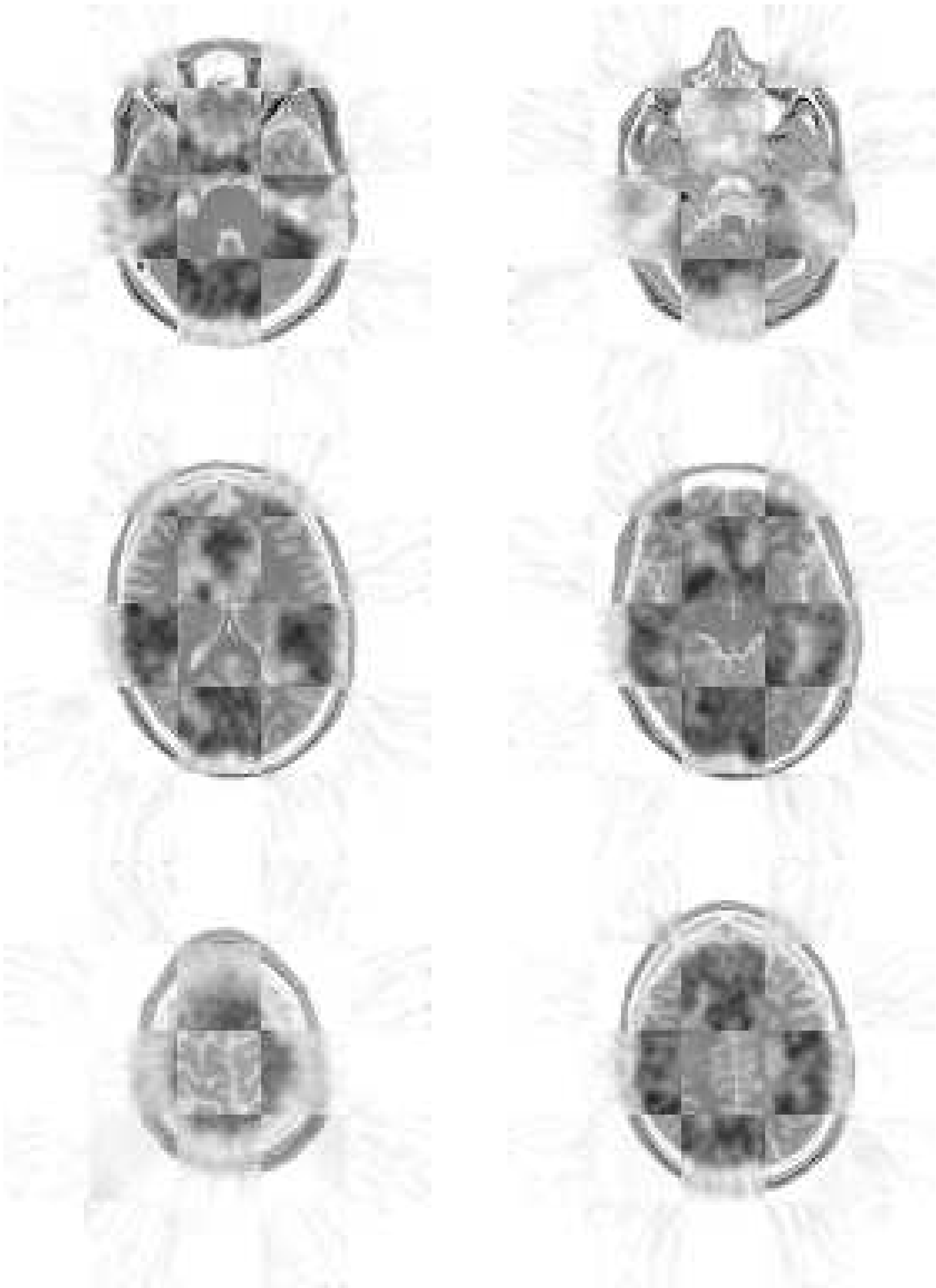

Figure 2.5: An example of PET-MRI co-registration, achieved using the techniques described here.

Figure 2.6: An example of T1-weighted and T2-weighted MR images of the same subject registered using the techniques described here.

	First step only		All three steps	
	Uncorrected	Corrected	Uncorrected	Corrected
mean error (mm)	5.57	3.77	4.14	3.20
median error (mm)	5.11	3.17	4.20	3.36
maximum error (mm)	11.62	8.54	7.46	5.76
N	7	4	7	4

Table 2.1: Errors for PET-MRI registration. Errors are presented, for both the uncorrected and the distortion corrected MR images. The results in the left hand column were derived after using only the first step of the registration process. The right hand column shows the results of using all three steps of the registration process.

	BA	CO	HA	HI	MAI	MAL
median error uncorrected	4.6	3.6*	2.8*	3.2*	3.5*	4.2*
median error corrected	3.2*	2.8*	3.6	2.5*	3.9	3.6
maximum error uncorrected	11.5	12.7	12.1	9.3	10.6	8.5
maximum error corrected	6.0	3.7*	17.7	6.0	7.7	8.4

	NO	PE	RO3	RO4	WO1	WO2
median error uncorrected	3.6*	2.9*	4.0*	3.4*	2.3*	3.1*
median error corrected	3.9	2.8*	3.8	3.6	2.0*	2.0*
maximum error uncorrected	11.4	10.0	9.4	5.9*	5.8*	6.0*
maximum error corrected	14.2	7.9	7.3	8.9	4.3*	5.0*

Table 2.2: Errors for PET-MRI registration from other methods. Asterisked values indicate results from methods that performed better than or the same as the one presented here. Method NO involved matching manually specified landmark pairs. Methods BA, HA, MAI, MAL, PE, RO3 and RO4 involved matching surfaces, contours or edges. WO1 and WO2 both used the AIR software (Woods *et al.*, 1992), but with different amounts of manual editing. CO and HI were based on maximising mutual information.

intervention was involved. The first step of the registration process (including the initial smoothing) took an average of 66 seconds, whereas the complete three step registration required about 350 seconds.

The resulting parameter estimates were communicated back to Vanderbilt, where their accuracy was evaluated. The results presented in Table 2.1 are the mean, median and maximum errors for the registration, and can be compared directly with those in Table 2.2 (taken from West *et al.*(1997)). They show that the first step of the process quickly registered the images to within about 6 mm, before the remaining steps further refined the parameters. The median registration errors using the current approach tended to be slightly larger than those from most other methods, although the maximum errors were generally smaller. Registration using AIR was found to produce the most consistently accurate solutions, but this requires initial manual pre-processing of the images.

Although the accuracy of the registrations was found to be comparable with the other techniques evaluated by West *et al.*, the data used in the evaluation did violate a number of the assumptions made by the current approach. Image intensity nonuniformity and low grey/white matter contrast resulted in a considerable amount of tissue misclassification of the MR images. Also, the assumption that brain tissue can be broadly classified as grey or white matter was complicated by the presence of tumours, which were classified as grey matter in the MR images, and white matter in the PET images. This would be expected to introduce additional registration errors, because the final step is based upon matching corresponding image partitions together. The registration should be much more accurate for images where the assumptions hold, as has been shown by more recent evaluations of SPECT and MRI co-registration methods (Barnden *et al.*, 2000).

In summary, the current technique is valid in relation to existing techniques. Unlike some of the existing approaches, the present method does not require manual intervention. The only occasional intervention that may be needed is to provide starting estimates to the first step. The procedure has so far been successfully applied to registering T1 weighted MRI to PET (blood flow), T1 to T2 weighted MRI, and T2 weighted MRI to PET.

Chapter 3

Image Warping using Basis Functions

Contents

3.1 Introduction

Methods of registering images 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. Homologous features are often identified manually, but this process is time consuming and subjective. Another disadvantage of using points or lines as landmarks is that there are very few readily identifiable discrete points or lines in the brain. However, surfaces are more readily identified, and in many instances they can be extracted automatically (or at least semi-automatically). Once they are identified, the spatial transformation is effected by bringing the homologies together. If the labels are points, then the required transformations at each of those points is known. Between the points, the deforming behaviour is not known, so it is forced to be as 'smooth' as possible. There are a number of methods for modelling this smoothness. The simplest models include fitting splines through the points in order to minimise *bending energy* (Bookstein, 1997a; Bookstein, 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 minimises some index of the difference between a source and a template image, where both are treated as unlabeled continuous processes. The matching criterion is usually based upon minimising the sum of squared differences or maximising 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 grey levels of the different tissue types between the source image and template.

A potentially enormous number of parameters are required to describe the nonlinear transformations that warp two images together (i.e., the problem is very high dimensional). However, much of the spatial variability can be captured using just a few parameters. Low spatial frequency global variability of head shape can be accommodated by describing deformations by a linear combination of low frequency basis functions. One very widely used basis function registration method is part of the AIR package (Woods *et al.*, 1998a; Woods *et al.*, 1998b), which uses polynomial basis functions to model shape variability. For example, a two dimensional third order polynomial basis function mapping can be defined something like:

$$
\begin{aligned}
y_1 = {} & q_1 + q_2 x_1 + q_3 x_1^2 + q_4 x_1^3 + \\
& q_5 x_2 + q_6 x_1 x_2 + q_7 x_1^2 x_2 + \\
& q_8 x_2^2 + q_9 x_1 x_2^2 + \\
& q_{10} x_2^3 \\
y_2 = {} & q_{11} + q_{12} x_1 + q_{13} x_1^2 + q_{14} x_1^3 + \\
& q_{15} x_2 + q_{16} x_1 x_2 + q_{17} x_1^2 x_2 + \\
& q_{18} x_2^2 + q_{19} x_1 x_2^2 + \\
& q_{20} x_2^3
\end{aligned}
\tag{3.1}
$$

The small number of parameters will not allow every feature to be matched exactly, but it will permit the global head shape to be modelled. The method of nonlinear registration described in this chapter is a similar approach, but uses discrete cosine transform basis functions instead of polynomials. The rational for adopting the low dimensional approach is that it allows rapid modelling of the global brain shape.

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. Regularisation schemes are therefore necessary when attempting image registration with many parameters, thus ensuring that voxels remain close to their neighbours. Regularisation is normally incorporated by 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 posterior probability of being correct, and is the estimate used for the fully automatic non-label based spatial normalisation method described in this chapter.

3.2 Methods

This section begins by introducing a modification to the optimisation method described in Section 2.4, such that more robust *maximum a posteriori* (MAP) parameter estimates can be obtained. The first step in registering images from different subjects involves determining the optimum 12 parameter affine transformation. A procedure for doing this using the MAP optimisation scheme is described. Because the variability of head sizes is known *a priori*, the registration can be made more robust by incorporating this knowledge. The next part describes nonlinear registration for correcting gross differences in head shapes that can not be accounted for by the affine normalisation alone. The nonlinear warps are modelled by linear combinations of smooth basis functions, and a fast algorithm for determining the optimum combination of basis functions is described. For speed and simplicity, a relatively small number of parameters (approximately 1000) are used to describe the nonlinear components of the registration. The MAP scheme requires some form of prior distribution for the basis function coefficients, so a number of different forms for this distribution are then presented. The last part of this section describes a variety of possible models for intensity transforms. In addition to spatial transformations, it is sometimes desirable to also include intensity transforms in the registration model, as one image may not look exactly like a spatially transformed version of the other.

3.2.1 A *Maximum A Posteriori* Solution

A Bayesian registration scheme is used in order to obtain a *maximum a posteriori* estimate of the registration parameters. Given some prior knowledge of the variability of brain shapes and sizes that may be encountered, a MAP registration scheme is able to give a more accurate (although biased) estimate of the true shapes of the brains. This is illustrated by a very simple one dimensional example in Figure 3.1. The use of a MAP parameter estimate reduces any potential over-fitting of the data, which may lead to unnecessary deformations that only reduce the residual variance by a tiny amount. It also makes the registration scheme more robust by reducing the search space of the algorithm, and therefore the number of potential local minima.

Bayes' rule can be expressed as:

$$p(\mathbf{q}|\mathbf{b}) \propto p(\mathbf{b}|\mathbf{q})p(\mathbf{q}) \tag{3.2}$$

where $p(\mathbf{q})$ is the prior probability of parameters $\mathbf{q}$, $p(\mathbf{b}|\mathbf{q})$ is the conditional probability that $\mathbf{b}$ is observed given $\mathbf{q}$ and $p(\mathbf{q}|\mathbf{b})$ is the posterior probability of $\mathbf{q}$, given that measurement $\mathbf{b}$ has been made. The *maximum a posteriori* (MAP) estimate for parameters $\mathbf{q}$ is the mode of $p(\mathbf{q}|\mathbf{b})$. The *maximum likelihood* (ML) estimate is a special case of the MAP estimate, in which $p(\mathbf{q})$ is uniform over all values of $\mathbf{q}$. For our purposes, $p(\mathbf{q})$ represents a known prior probability distribution from which the parameters are drawn, $p(\mathbf{b}|\mathbf{q})$ is the likelihood of obtaining the data $\mathbf{b}$ given the parameters, and $p(\mathbf{q}|\mathbf{b})$ is the function to be maximised. The optimisation can be simplified by assuming that all probability distributions can be approximated by multi-normal (multidimensional and normal) distributions, and can therefore be described by a mean vector and a covariance matrix.

A probability is related to its Gibbs form by $p(a) \propto e^{-H(a)}$. Therefore the posterior probability is maximised when its Gibbs form is minimised. This is equivalent to minimising $H(\mathbf{b}|\mathbf{q}) +$

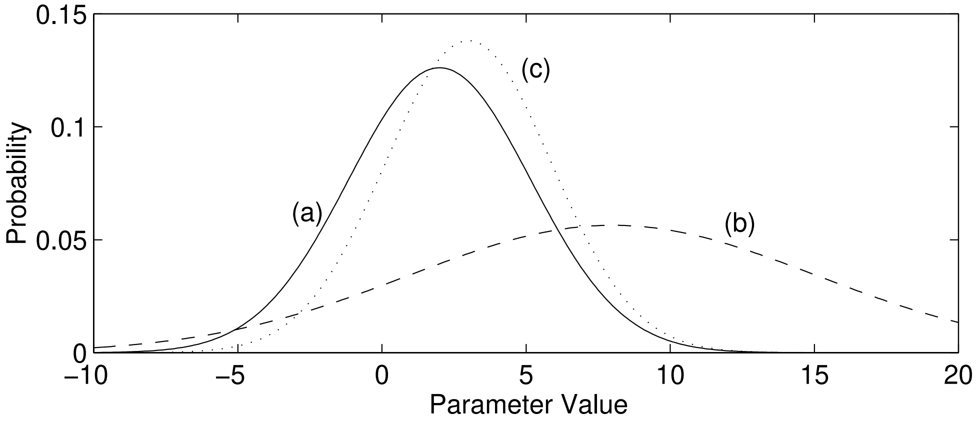

Figure 3.1: This figure illustrates a hypothetical example with one parameter, where the prior probability distribution is better described than the likelihood. The solid Gaussian curve (a) represents the prior probability distribution (p.d.f), and the dashed curve (b) represents a maximum likelihood parameter estimate (from fitting to observed data) with its associated certainty. The true parameter is known to be drawn from distribution (a), but it can be estimated with the certainty described by distribution (b). Without the MAP scheme, a more precise estimate would probably be obtained for the true parameter by taking the most likely *a priori* value, rather than the value obtained from a maximum likelihood fit to the data. This would be analogous to cases where the number of parameters is reduced in a maximum likelihood registration model in order to achieve a better solution (e.g., see page 45). The dotted line (c) shows the posterior p.d.f obtained using Bayesian statistics. The maximum value of (c) falls at the MAP estimate. It combines previously known information with that from the data to give a more accurate estimate.

$H(\mathbf{q})$ (the posterior potential). In this expression, $H(\mathbf{b}|\mathbf{q})$ (the likelihood potential) is related to the residual sum of squares. If the parameters are assumed to be drawn from a multi-normal distribution described by a mean vector $\mathbf{q_0}$ and covariance matrix $\mathbf{C_0}$, then $H(\mathbf{q})$ (the prior potential) is simply given by:

$$H(\mathbf{q}) = (\mathbf{q} - \mathbf{q_0})^T \, \mathbf{C_0}^{-1} \, (\mathbf{q} - \mathbf{q_0}) \tag{3.3}$$

Eqn. 2.22 gives the following maximum likelihood updating rule for the parameter estimation:

$$\mathbf{q_{ML}}^{(n+1)} = \mathbf{q}^{(n)} - \left(\mathbf{A}^T\mathbf{A}\right)^{-1} \mathbf{A}^T\mathbf{b} \tag{3.4}$$

Assuming equal variance for each observation (σ^2) and ignoring covariances among them, the formal covariance matrix of the fit on the assumption of normally distributed errors is given by $\sigma^2 \left(\mathbf{A}^T\mathbf{A}\right)^{-1}$. When the distributions are normal, the MAP estimate is simply the average of the prior and likelihood estimates, weighted by the inverses of their respective covariance matrices:

$$\mathbf{q}^{(n+1)} = \left(\mathbf{C_0}^{-1} + \mathbf{A}^T\mathbf{A}/\sigma^2\right)^{-1} \left(\mathbf{C_0}^{-1}\mathbf{q_0} + \mathbf{A}^T\mathbf{A}/\sigma^2\mathbf{q_{ML}}^{(n+1)}\right) \tag{3.5}$$

The MAP optimisation scheme is obtained by combining Eqns. 3.4 and 3.5.

$$\mathbf{q}^{(n+1)} = \left(\mathbf{C_0}^{-1} + \mathbf{A}^T\mathbf{A}/\sigma^2\right)^{-1} \left(\mathbf{C_0}^{-1}\mathbf{q_0} + \mathbf{A}^T\mathbf{A}\mathbf{q}^{(n)}/\sigma^2 - \mathbf{A}^T\mathbf{b}/\sigma^2\right) \tag{3.6}$$

For the sake of the registration, it is assumed that the exact form for the *a priori* probability distribution ($\mathbf{q_0}$ and $\mathbf{C_0}$) is known. However, because the registration may need to be done on a wide range of different image modalities, with differing contrasts and signal to noise ratios, it is not possible to easily and automatically know what value to use for σ^2. In practice, σ^2 is assumed to be the same for all observations, and is estimated from the sum of squared differences from the current iteration:

$$\sigma^2 = \sum_{i=1}^{I} b_i(\mathbf{q})^2/\nu \tag{3.7}$$

where ν refers to the degrees of freedom. If the sampling is sparse relative to the smoothness, then $\nu \simeq I - J$, where I is the number of sampled locations in the images and J is the number of estimated parameters [1].

However, complications arise because the images are smooth, resulting in the observations not being independent, and a reduction in the effective number of degrees of freedom. The degrees of freedom are corrected using the principles described by Friston (1995a) [although this approach is not strictly correct (Worsley & Friston, 1995), it gives an estimate that is close enough for these purposes]. The effective degrees of freedom are estimated by assuming that the difference between $\mathbf{f}$ and $\mathbf{g}$ approximates a continuous, zero-mean, homogeneous, smoothed *Gaussian random field*. The approximate parameter of a Gaussian point spread function describing the smoothness in direction k (assuming that the axes of the Gaussian are aligned with the axes of the image co-ordinate system) can be obtained by (Poline *et al.*, 1995):

$$w_k = \sqrt{\frac{\sum_{i=1}^{I} b_i(\mathbf{q})^2}{2\sum_{i=1}^{I}(\nabla_k b_i(\mathbf{q}))^2}} \tag{3.8}$$

[1] Strictly speaking, the computation of the degrees of freedom should be more complicated than this, as this simple model does not account for the regularisation (See Section 7.3).

Multiplying w_k by $\sqrt{8\log_e(2)}$ produces an estimate of the full width at half maximum of the Gaussian. If the images are sampled on a regular grid where the spacing in each direction is s_k, the number of effective degrees of freedom[2] becomes approximately:

$$\nu = (I - J)\prod_k \frac{s_k}{w_k(2\pi)^{1/2}} \tag{3.9}$$

This is essentially a scaling of $I - J$ by the number of resolution elements per voxel.

This approach has the advantage that when the parameter estimates are far from the solution, σ^2 is large, so the problem becomes more heavily regularised with more emphasis being placed on the prior information. For nonlinear warping, this is analogous to a coarse to fine registration scheme. The penalty against higher frequency warps is greater than that for those of low frequency (see Section 3.2.4). In the early iterations, the estimated σ^2 is higher leading to a heavy penalty against all warps, but with more against those of higher frequency. The algorithm does not fit much of the high frequency information until σ^2 has been reduced. In addition to a gradual reduction in σ^2 due to the decreasing residual squared difference, σ^2 is also reduced because the estimated smoothness is decreased, leading to more effective degrees of freedom. Both these factors are influential in making the registration scheme more robust to local minima.

3.2.2 Affine Registration

Almost all between subject co-registration or spatial normalisation methods for brain images begin by determining the optimal nine or twelve parameter affine transformation that registers the images together. This step is normally performed automatically by minimising (or maximising) some mutual function of the images. The objective of affine registration is to fit the source image $\mathbf{f}$ to a template image $\mathbf{g}$, using a twelve parameter affine transformation (via a matrix $\mathbf{M}$ generated from parameters q_1 to q_{12}). The images may be scaled quite differently, so an additional intensity scaling parameter (q_{13}) is included in the model. The objective function that is minimised is therefore:

$$\sum_i \left(f(\mathbf{M}\mathbf{x}_i) - q_{13}g(\mathbf{x}_i)\right)^2 \tag{3.10}$$

Without constraints and with poor data, simple ML parameter optimisation (similar to that described in Section 2.5) can produce some extremely unlikely transformations. For example, when there are only a few slices in the image, it is not possible for the algorithms to determine an accurate zoom in the out of plane direction. Any estimate of this value is likely to have very large errors. When a regularised approach is not used, it may be better to assign a fixed value for this difficult-to-determine parameter, and simply fit for the remaining ones.

By incorporating prior information into the optimisation procedure, a smooth transition between fixed and fitted parameters can be achieved. When the error for a particular fitted parameter is known to be large, then that parameter will be based more upon the prior information. In order to adopt this approach, the prior distribution of the parameters should be known. A suitable *a priori* distribution of the parameters ($\mathbf{q_0}$ and $\mathbf{C_0}$ from Eqn. 3.6) was determined from

[2]Note that this only applies when $s_k < w_k(2\pi)^{1/2}$, otherwise $\nu = I - J$. Alternatively, to circumvent this problem the degrees of freedom can be better estimated by $(I-J)\prod_k \mathrm{erf}(2^{-3/2}s_k/w_k)$. This gives a similar result to the approximation by Friston (1995a) for smooth images, but never allows the computed value to exceed $I - J$.

affine transformations estimated from 51 high resolution T_1 weighted brain MR images using the basic least squares optimisation algorithm. The transformation matrices were defined by $\mathbf{M} = \mathbf{M_f}^{-1}\mathbf{M_a}^{-1}\mathbf{M_g}$, where $\mathbf{M_a}$ (refer back to Sections 2.2.1 and 2.2.2) is constructed from parameters $\mathbf{q}$:

$$\mathbf{M_a} = \begin{bmatrix} 1 & 0 & 0 & q_1 \\ 0 & 1 & 0 & q_2 \\ 0 & 0 & 1 & q_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos(q_4) & sin(q_4) & 0 \\ 0 & -sin(q_4) & cos(q_4) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} cos(q_5) & 0 & sin(q_5) & 0 \\ 0 & 0 & & 0 \\ -sin(q_5) & 0 & cos(q_5) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \ldots$$

$$\begin{bmatrix} cos(q_6) & sin(q_6) & 0 & 0 \\ -sin(q_6) & cos(q_6) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} q_7 & 0 & 0 & 0 \\ 0 & q_8 & 0 & 0 \\ 0 & 0 & q_9 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & q_{10} & q_{11} & 0 \\ 0 & 1 & q_{12} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.11)$$

The results for the translation and rotation parameters (q_1 to q_6) are ignored, since these depend only on the positioning of the subjects in the scanner, and do not reflect variability of head shape and size.

The mean zooms required to fit the individual brains to the space of the template (parameters q_7 to q_9) were 1.10, 1.05 and 1.17 in the left-right, posterior-anterior and inferior-superior directions respectively, reflecting the fact that the template was larger than a typical head. The variance-covariance matrix for these parameters was:

$$\begin{bmatrix} 0.00210 & 0.00094 & 0.00134 \\ 0.00094 & 0.00307 & 0.00143 \\ 0.00134 & 0.00143 & 0.00242 \end{bmatrix}$$

giving a correlation coefficient matrix of:

$$\begin{bmatrix} 1.00 & 0.37 & 0.59 \\ 0.37 & 1.00 & 0.52 \\ 0.59 & 0.52 & 1.00 \end{bmatrix}$$

As expected, these parameters are correlated, since larger brains are generally larger in all dimensions. This allows partial prediction of the optimal zoom in one direction given the zooms in the others, a fact that is useful for spatially normalising images containing a limited number of transverse slices. The means of the parameters defining shear were close to zero (-0.0024, 0.0006 and -0.0107 for q_{10}, q_{11} and q_{12} respectively). The variances of the parameters are 0.000184, 0.000112 and 0.001786, with very little covariance.

A number of affine registrations were evaluated both with and without incorporating the MAP scheme. This evaluation is described in Section 3.3.1.

3.2.3 Nonlinear Registration

The nonlinear spatial normalisation approach described here assumes that the image has already been approximately registered with the template according to a twelve-parameter affine registration. This section illustrates how the parameters describing global shape differences (not accounted for by affine registration) between an image and template can be determined.

The model for defining nonlinear warps uses deformations consisting of a linear combination of low-frequency periodic basis functions. The spatial transformation from co-ordinates $\mathbf{x}_i$, to co-ordinates $\mathbf{y}_i$ is:

$$\begin{aligned} y_{1i} &= x_{1i} + u_{1i} = x_{1i} + \sum_j q_{j1} d_j(\mathbf{x}_i) \\ y_{2i} &= x_{2i} + u_{2i} = x_{2i} + \sum_j q_{j2} d_j(\mathbf{x}_i) \\ y_{3i} &= x_{3i} + u_{3i} = x_{3i} + \sum_j q_{j3} d_j(\mathbf{x}_i) \end{aligned} \tag{3.12}$$

where q_{jk} is the jth coefficient for dimension k, and $d_j(\mathbf{x})$ is the jth basis function at position $\mathbf{x}$.

The choice of basis functions depend upon the distribution of warps likely to be required, and also upon how translations at borders should behave. If points at the borders over which the transform is computed are not required to move in any direction, then the basis functions should consist of the lowest frequencies of the three dimensional discrete sine transform (DST). If there are to be no constraints at the borders, then a three dimensional discrete cosine transform (DCT) is more appropriate. Both of these transforms use the same set of basis functions to represent warps in each of the directions. Alternatively, a mixture of DCT and DST basis functions can be used to constrain translations at the surfaces of the volume to be parallel to the surface only (*sliding* boundary conditions). By using a different combination of DCT and DST basis functions, the corners of the volume can be fixed and the remaining points on the surface can be free to move in all directions (*bending* boundary conditions) (Christensen, 1994). These various boundary conditions are illustrated in Figure 3.2.

The basis functions used here are the lowest frequency components of the three (or two) dimensional DCT. In one dimension, the DCT of a function is generated by pre-multiplication with the matrix $\mathbf{D}^T$, where the elements of the $I \times M$ matrix $\mathbf{D}$ are defined by:

$$\begin{aligned} d_{i1} &= \frac{1}{\sqrt{I}} \quad i = 1..I \\ d_{im} &= \sqrt{\frac{2}{I}} \cos\left(\frac{\pi(2i-1)(m-1)}{2I}\right) \quad i = 1..I, m = 2..M \end{aligned} \tag{3.13}$$

A set of low frequency two dimensional DCT basis functions are shown in Figure 3.3, and a schematic example of a two dimensional deformation based upon the DCT is shown in Figure 3.4.

As for affine registration, the optimisation involves minimising the sum of squared differences between a source ($\mathbf{f}$) and template image ($\mathbf{g}$). The images may be scaled differently, so an additional parameter (w) is needed to accommodate this difference. The minimised function is then:

$$\sum_i (f(\mathbf{y}_i) - w g(\mathbf{x}_i))^2 \tag{3.14}$$

The approach described in Section 2.4 is used to optimise the parameters $\mathbf{q_1}$, $\mathbf{q_2}$, $\mathbf{q_3}$ and w, and requires derivatives of the function $f(\mathbf{y}_i) - w g(\mathbf{x}_i)$ with respect to each parameter. These can be obtained using the chain rule:

$$\frac{\partial f(\mathbf{y}_i)}{\partial q_{j1}} = \frac{\partial f(\mathbf{y}_i)}{\partial y_{1i}} \frac{\partial y_{1i}}{\partial q_{j1}} = \frac{\partial f(\mathbf{y}_i)}{\partial y_{1i}} d_j(\mathbf{x}_i)$$

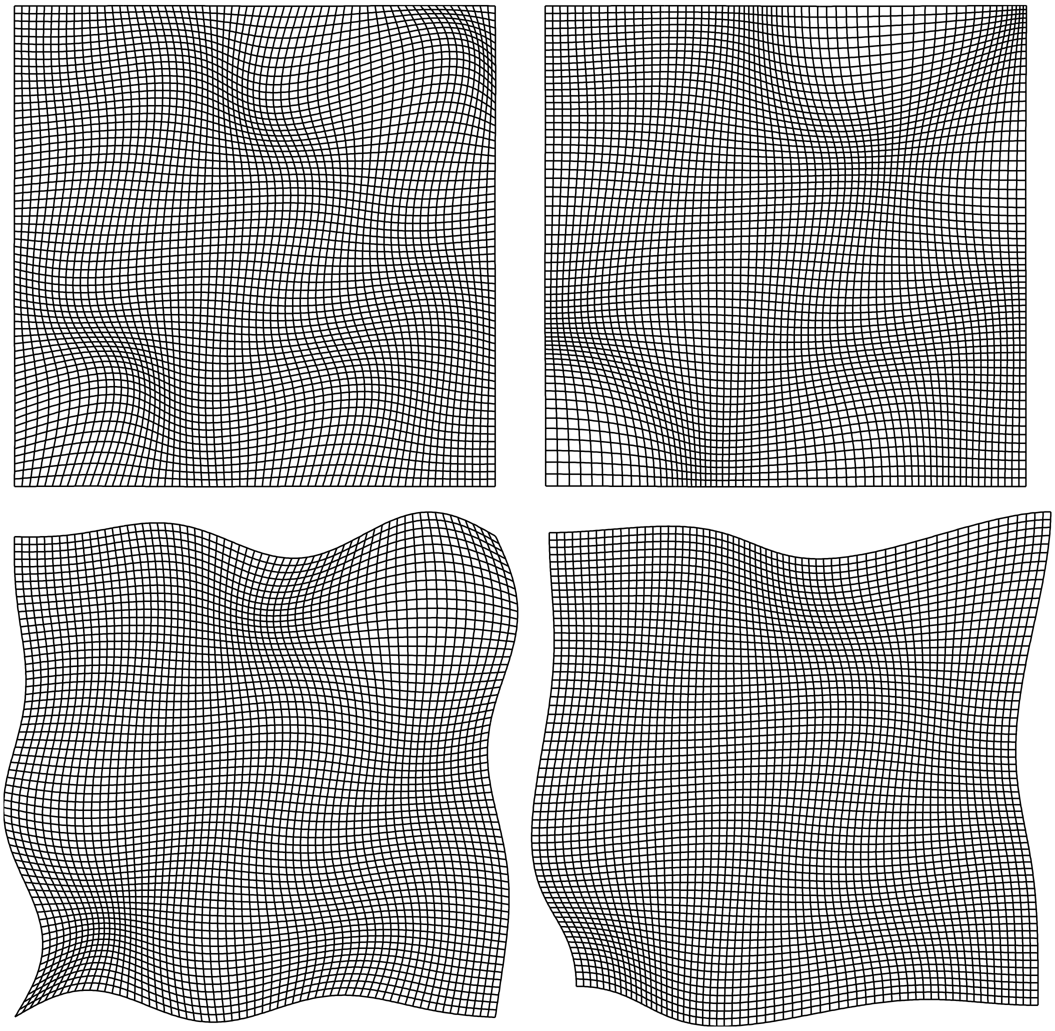

Figure 3.2: Different boundary conditions. Above left: fixed boundaries (generated purely from DST basis functions). Above right: sliding boundaries (from a mixture of DCT and DST basis functions). Below left: bending boundaries (from a different mixture of DCT and DST basis functions). Below right: free boundary conditions (purely from DCT basis functions). These deformation fields were all computed using the same 4×4 randomly generated coefficients (normal distribution of unit variance), and are assumed to cover a unit square.

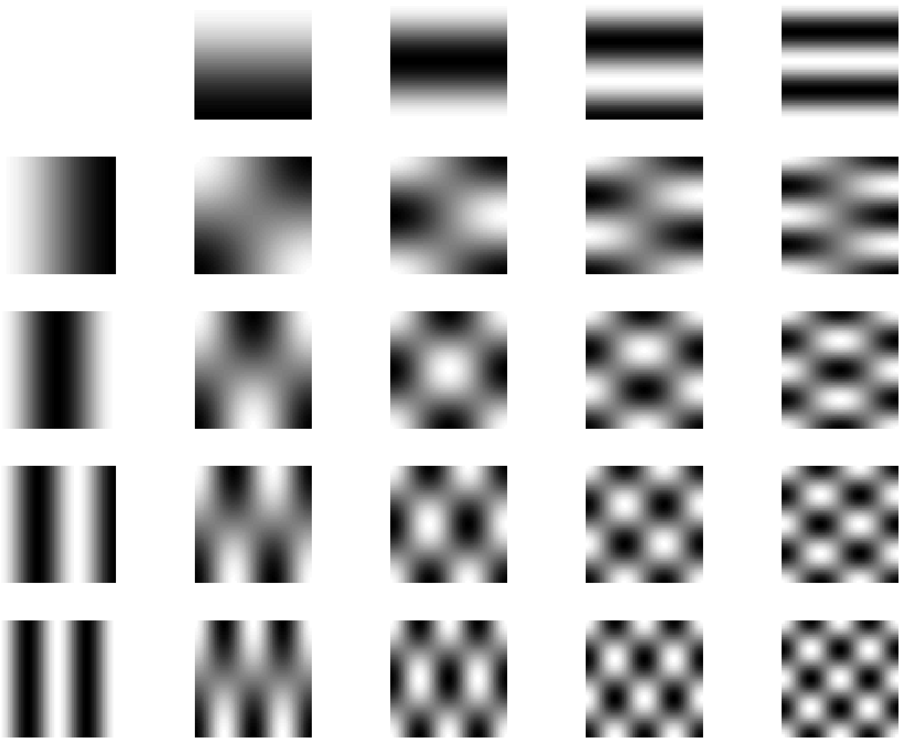

Figure 3.3: The lowest frequency basis functions of a two dimensional Discrete Cosine Transform.

$$\begin{aligned}\frac{\partial f(\mathbf{y}_i)}{\partial q_{j2}} &= \frac{\partial f(\mathbf{y}_i)}{\partial y_{2i}}\frac{\partial y_{2i}}{\partial q_{j2}} = \frac{\partial f(\mathbf{y}_i)}{\partial y_{2i}}d_j(\mathbf{x}_i)\\ \frac{\partial f(\mathbf{y}_i)}{\partial q_{j3}} &= \frac{\partial f(\mathbf{y}_i)}{\partial y_{3i}}\frac{\partial y_{3i}}{\partial q_{j3}} = \frac{\partial f(\mathbf{y}_i)}{\partial y_{3i}}d_j(\mathbf{x}_i)\end{aligned} \tag{3.15}$$

The approach involves iteratively computing $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}^T\mathbf{b}$. However, because there are many parameters to optimise, these computations can be very time consuming. There now follows a description of a very efficient way of computing these matrices.

A Fast Algorithm

A fast algorithm for computing $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}^T\mathbf{b}$ is shown in Figure 3.5. The remainder of this section explains the matrix terminology used, and why it is so efficient.

For simplicity, the algorithm is only illustrated in two dimensions, although it has been implemented to estimate warps in three dimensions. Images $\mathbf{f}$ and $\mathbf{g}$ are considered as $I \times J$ matrices $\mathbf{F}$ and $\mathbf{G}$ respectively. Row i of $\mathbf{F}$ is denoted by $\mathbf{f}_{i,:}$, and column j by $\mathbf{f}_{:,j}$. The basis functions used by the algorithm are generated from a separable form from matrices $\mathbf{D_1}$ and $\mathbf{D_2}$, with dimensions $I \times M$ and $J \times N$ respectively. By treating the transform coefficients as $M \times N$ matrices $\mathbf{Q_1}$ and $\mathbf{Q_2}$, the deformation fields can be rapidly constructed by computing $\mathbf{D_1 Q_1 D_2}^T$ and $\mathbf{D_1 Q_2 D_2}^T$.

Between each iteration, image $\mathbf{F}$ is resampled according to the latest parameter estimates. The derivatives of $\mathbf{F}$ are also resampled to give $\nabla_\mathbf{1}\mathbf{F}$ and $\nabla_\mathbf{2}\mathbf{F}$. The ith element of each of these matrices contain $f(\mathbf{y}_i)$, $\partial f(\mathbf{y}_i)/\partial y_{1i}$ and $\partial f(\mathbf{y}_i)/\partial y_{2i}$ respectively.

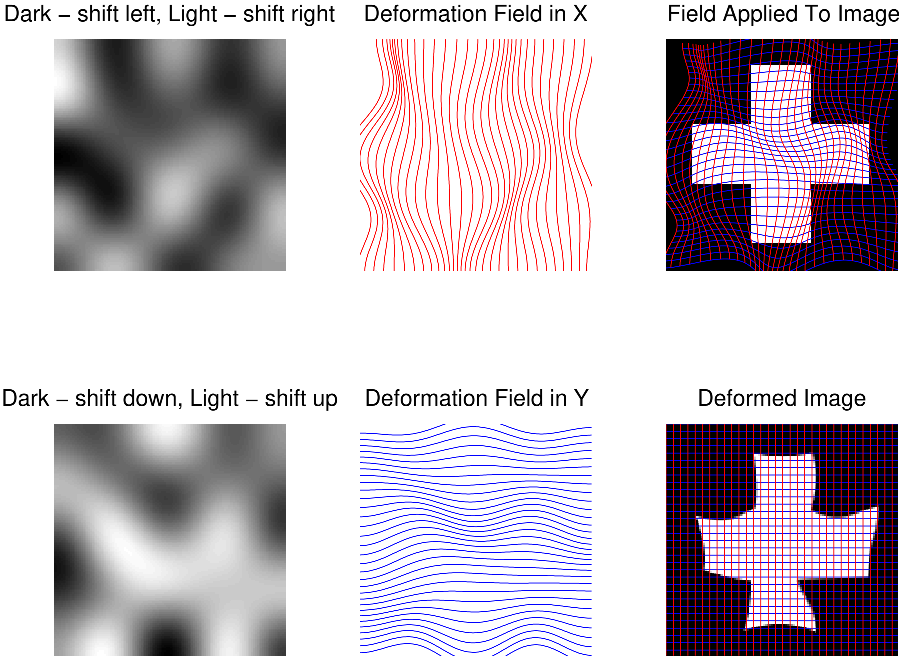

Figure 3.4: In two dimensions, a deformation field consists of two scalar fields. One for horizontal deformations, and the other for vertical deformations. The images on the left show deformations as a linear combination of basis images (see Figure 3.3). The centre column shows the same deformations in a more intuitive sense. The deformation is applied by overlaying it on a source image, and re-sampling (right).

$$\alpha = \begin{bmatrix}\mathbf{0}\end{bmatrix}$$
$$\beta = \begin{bmatrix}\mathbf{0}\end{bmatrix}$$

$for \quad j = 1 \ldots J$

$$\mathbf{C} = \mathbf{d_2}_{j,:}{}^T \mathbf{d_2}_{j,:}$$
$$\mathbf{E_1} = diag(\nabla_\mathbf{1} \mathbf{f}_{:,j})\mathbf{D_1}$$
$$\mathbf{E_2} = diag(\nabla_\mathbf{2} \mathbf{f}_{:,j})\mathbf{D_1}$$

$$\alpha = \alpha + \begin{bmatrix} \mathbf{C} \otimes (\mathbf{E_1}^T \mathbf{E_1}) & \mathbf{C} \otimes (\mathbf{E_1}^T \mathbf{E_2}) & -\mathbf{d_2}_{j,:}{}^T \otimes (\mathbf{E_1}^T \mathbf{g}_{:,j}) \\ (\mathbf{C} \otimes (\mathbf{E_1}^T \mathbf{E_2}))^T & \mathbf{C} \otimes (\mathbf{E_2}^T \mathbf{E_2}) & -\mathbf{d_2}_{j,:}{}^T \otimes (\mathbf{E_2}^T \mathbf{g}_{:,j}) \\ (-\mathbf{d_2}_{j,:}{}^T \otimes (\mathbf{E_1}^T \mathbf{g}_{:,j}))^T & (-\mathbf{d_2}_{j,:}{}^T \otimes (\mathbf{E_1}^T \mathbf{g}_{:,j}))^T & \mathbf{g}_{:,j}{}^T \mathbf{g}_{:,j} \end{bmatrix}$$

$$\beta = \beta + \begin{bmatrix} \mathbf{d_2}_{j,:}{}^T \otimes (\mathbf{E_1}^T (\mathbf{f}_{:,j} - w\mathbf{g}_{:,j})) \\ \mathbf{d_2}_{j,:}{}^T \otimes (\mathbf{E_2}^T (\mathbf{f}_{:,j} - w\mathbf{g}_{:,j})) \\ \mathbf{g}_{:,j}{}^T (\mathbf{f}_{:,j} - w\mathbf{g}_{:,j}) \end{bmatrix}$$

end

Figure 3.5: A two dimensional illustration of the fast algorithm for computing $\mathbf{A}^T\mathbf{A}$ (α) and $\mathbf{A}^T\mathbf{b}$ (β).

The notation $diag(\nabla_\mathbf{1}\mathbf{f}_{:,j})\mathbf{D_1}$ simply means multiplying each element of row i of $\mathbf{D_1}$ by $\nabla_\mathbf{1}\mathbf{f}_{i,j}$, and the symbol '$\otimes$' refers to the *Kronecker tensor product*. If $\mathbf{D_2}$ is a matrix of order $J \times N$, and $\mathbf{D_1}$ is a second matrix, then:

$$\mathbf{D_2} \otimes \mathbf{D_1} = \begin{bmatrix} d_{\mathbf{2}11}\mathbf{D_1} & \dots & d_{\mathbf{2}1N}\mathbf{D_1} \\ \vdots & \ddots & \vdots \\ d_{\mathbf{2}J1}\mathbf{D_1} & \dots & d_{\mathbf{2}JN}\mathbf{D_1} \end{bmatrix} \tag{3.16}$$

The advantage of the algorithm shown in Figure 3.5 is that it utilises some of the useful properties of Kronecker tensor products. This is especially important when the algorithm is implemented in three dimensions. The performance enhancement results from a reordering of a set of operations like $(\mathbf{D_3} \otimes \mathbf{D_2} \otimes \mathbf{D_1})^T (\mathbf{D_3} \otimes \mathbf{D_2} \otimes \mathbf{D_1})$, to the equivalent $(\mathbf{D_3}^T\mathbf{D_3}) \otimes (\mathbf{D_2}^T\mathbf{D_2}) \otimes (\mathbf{D_1}^T\mathbf{D_1})$. Assuming that the matrices $\mathbf{D_3}$, $\mathbf{D_2}$ and $\mathbf{D_1}$ all have order $I \times M$, then the number of floating point operations is reduced from $I^3M^3(M^3+2)$ to approximately $3I(M^2+M)+M^6$. If I equals 32, and M equals 4, then a performance increase of about a factor of 20,000 would be expected. The limiting factor to the algorithm is no longer the time taken to create the curvature matrix $(\mathbf{A}^T\mathbf{A})$, but is now the amount of memory required to store it and the time taken to invert it.

3.2.4 Linear Regularisation for Nonlinear Registration

Without regularisation in the nonlinear registration, it is possible to introduce unnecessary deformations that only reduce the residual sum of squares by a tiny amount. This could potentially make the algorithm very unstable. Regularisation is achieved by minimising the sum of squared difference between the template and the warped image, while simultaneously minimising some function of the deformation field. The principles are Bayesian and make use of the MAP scheme described in Section 3.2.1.

The first requirement for a MAP approach is to define some form of prior distribution for the parameters. For a simple linear[3] approach, the priors consist of an *a priori* estimate of the mean of the parameters (assumed to be zero), and also a covariance matrix describing the distribution of the parameters about this mean. There are many possible forms for these priors, each of which describes some form of 'energy' term. If the true prior distribution of the parameters is known (somehow derived from a large number of subjects), then $\mathbf{C_0}$ could be an empirically determined covariance matrix describing this distribution. This approach would have the advantage that the resulting deformations are more typically "brain like", and so increase the face validity of the approach.

The three distinct forms of linear regularisation that will now be described are based upon *membrane energy*, *bending energy* and *linear-elastic energy*. None of these schemes enforce a strict one to one mapping between the source and template images, but this makes little difference for the small deformations required here. Each of these models needs some form of elasticity constants (λ and sometimes μ). Values of these constants that are too large will provide too much regularisation and result in greatly underestimated deformations. If the values are too

[3] Although the cost function associated with these priors is quadratic, the priors are linear in the sense that they minimise the sum of squares of a linear combination of the model parameters. This is analogous to solving a set of linear equations by minimising a quadratic cost function.

small, there will not be enough regularisation and the resulting deformations will overfit the data. Section 7.3 will introduce one possible method of estimating what the best values for these constants should be.

Membrane Energy

The simplest model used for linear regularisation is based upon minimising the *membrane energy* of the deformation field $\mathbf{u}$ (Amit *et al.*, 1991; Gee *et al.*, 1997). By summing over i points in three dimensions, the membrane energy of $\mathbf{u}$ is given by:

$$\sum_i \sum_{j=1}^{3} \sum_{k=1}^{3} \lambda \left(\frac{\partial u_{ji}}{\partial x_{ki}} \right)^2 \tag{3.17}$$

where λ is simply a scaling constant. The membrane energy can be computed from the coefficients of the basis functions by $\mathbf{q_1}^T \mathbf{H} \mathbf{q_1} + \mathbf{q_2}^T \mathbf{H} \mathbf{q_2} + \mathbf{q_3}^T \mathbf{H} \mathbf{q_3}$, where $\mathbf{q_1}$, $\mathbf{q_2}$ and $\mathbf{q_3}$ refer to vectors containing the parameters describing translations in the three dimensions. The matrix $\mathbf{H}$ is defined by:

$$\begin{aligned}\mathbf{H} = \;& \lambda\left(\dot{\mathbf{D}}_\mathbf{3}^T \dot{\mathbf{D}}_\mathbf{3}\right) \otimes \left(\mathbf{D_2}^T \mathbf{D_2}\right) \otimes \left(\mathbf{D_1}^T \mathbf{D_1}\right) \\ +\;& \lambda\left(\mathbf{D_3}^T \mathbf{D_3}\right) \otimes \left(\dot{\mathbf{D}}_\mathbf{2}^T \dot{\mathbf{D}}_\mathbf{2}\right) \otimes \left(\mathbf{D_1}^T \mathbf{D_1}\right) \\ +\;& \lambda\left(\mathbf{D_3}^T \mathbf{D_3}\right) \otimes \left(\mathbf{D_2}^T \mathbf{D_2}\right) \otimes \left(\dot{\mathbf{D}}_\mathbf{1}^T \dot{\mathbf{D}}_\mathbf{1}\right)\end{aligned} \tag{3.18}$$

where the notation $\dot{\mathbf{D}}_\mathbf{1}$ refers to the first derivatives of $\mathbf{D_1}$.

Assuming that the parameters consist of $\left[\mathbf{q_1}^T \mathbf{q_2}^T \mathbf{q_3}^T w\right]^T$, matrix $\mathbf{C_0}^{-1}$ from Eqn. 3.6 can be constructed from $\mathbf{H}$ by:

$$\mathbf{C_0}^{-1} = \begin{bmatrix} \mathbf{H} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{H} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 \end{bmatrix} \tag{3.19}$$

$\mathbf{H}$ is all zeros, except for the diagonal. Elements on the diagonal represent the reciprocal of the *a priori* variance of each parameter. If all the DCT matrices are $I \times M$, then each diagonal element is given by:

$$\begin{aligned} h_{j+M(k-1+M(l-1))} &= \lambda\pi^2 I^{-2}\left((j-1)^2 + (k-1)^2 + (l-1)^2\right) \\ &\text{over } j = 1 \ldots M,\ k = 1 \ldots M \text{ and } l = 1 \ldots M. \end{aligned} \tag{3.20}$$

The nonlinear registration algorithm described here is implemented in three dimensions using membrane energy as the cost functions. For completeness, the other two commonly used cost functions will now be described, but only for the two dimensional case.

Bending Energy

Bookstein's thin plate splines (1997b; 1997a) minimise the *bending energy* of deformations. For a two dimensional deformation, the bending energy is defined by:

$$\lambda \sum_i \left(\left(\frac{\partial^2 u_{1i}}{\partial x_{1i}^2} \right)^2 + \left(\frac{\partial^2 u_{1i}}{\partial x_{2i}^2} \right)^2 + 2\left(\frac{\partial^2 u_{1i}}{\partial x_{1i} \partial x_{2i}} \right)^2 \right) \quad +$$

$$\lambda \sum_i \left(\left(\frac{\partial^2 u_{2i}}{\partial x_{1i}^2} \right)^2 + \left(\frac{\partial^2 u_{2i}}{\partial x_{2i}^2} \right)^2 + 2 \left(\frac{\partial^2 u_{2i}}{\partial x_{1i} \partial x_{2i}} \right)^2 \right) \tag{3.21}$$

This can be computed by:

$$\begin{aligned} \lambda \mathbf{q_1}^T (\ddot{\mathbf{D}}_\mathbf{2} \otimes \mathbf{D_1})^T (\ddot{\mathbf{D}}_\mathbf{2} \otimes \mathbf{D_1}) \mathbf{q_1} &+ \lambda \mathbf{q_1}^T (\mathbf{D_2} \otimes \ddot{\mathbf{D}}_\mathbf{1})^T (\mathbf{D_2} \otimes \ddot{\mathbf{D}}_\mathbf{1}) \mathbf{q_1} + \\ 2\lambda \mathbf{q_1}^T (\dot{\mathbf{D}}_\mathbf{2} \otimes \dot{\mathbf{D}}_\mathbf{1})^T (\dot{\mathbf{D}}_\mathbf{2} \otimes \dot{\mathbf{D}}_\mathbf{1}) \mathbf{q_1} &+ \lambda \mathbf{q_2}^T (\ddot{\mathbf{D}}_\mathbf{2} \otimes \mathbf{D_1})^T (\ddot{\mathbf{D}}_\mathbf{2} \otimes \mathbf{D_1}) \mathbf{q_2} + \\ \lambda \mathbf{q_2}^T (\mathbf{D_2} \otimes \ddot{\mathbf{D}}_\mathbf{1})^T (\mathbf{D_2} \otimes \ddot{\mathbf{D}}_\mathbf{1}) \mathbf{q_2} &+ 2\lambda \mathbf{q_2}^T (\dot{\mathbf{D}}_\mathbf{2} \otimes \dot{\mathbf{D}}_\mathbf{1})^T (\dot{\mathbf{D}}_\mathbf{2} \otimes \dot{\mathbf{D}}_\mathbf{1}) \mathbf{q_2} \end{aligned} \tag{3.22}$$

where the notation $\dot{\mathbf{D}}_\mathbf{1}$ and $\ddot{\mathbf{D}}_\mathbf{1}$ refer to the column-wise first and second derivatives of $\mathbf{D_1}$. This is simplified to $\mathbf{q_1}^T \mathbf{H} \mathbf{q_1} + \mathbf{q_2}^T \mathbf{H} \mathbf{q_2}$ where:

$$\mathbf{H} = \lambda \left(\left(\ddot{\mathbf{D}}_\mathbf{2}{}^T \ddot{\mathbf{D}}_\mathbf{2} \right) \otimes \left(\mathbf{D_1}^T \mathbf{D_1} \right) + \left(\mathbf{D_2}^T \mathbf{D_2} \right) \otimes \left(\ddot{\mathbf{D}}_\mathbf{1}{}^T \ddot{\mathbf{D}}_\mathbf{1} \right) + 2 \left(\dot{\mathbf{D}}_\mathbf{2}{}^T \dot{\mathbf{D}}_\mathbf{2} \right) \otimes \left(\dot{\mathbf{D}}_\mathbf{1}{}^T \dot{\mathbf{D}}_\mathbf{1} \right) \right) \tag{3.23}$$

Matrix $\mathbf{C_0}^{-1}$ from Eqn. 3.6 can be constructed from $\mathbf{H}$ as:

$$\mathbf{C_0}^{-1} = \begin{bmatrix} \mathbf{H} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 0 \end{bmatrix} \tag{3.24}$$

with values on the diagonals of $\mathbf{H}$ given by:

$$h_{j+(k-1)\times M} = \lambda \left(\left(\frac{\pi(j-1)}{I} \right)^4 + \left(\frac{\pi(k-1)}{I} \right)^4 + 2 \left(\frac{\pi(j-1)}{I} \right)^2 \left(\frac{\pi(k-1)}{I} \right)^2 \right) \quad \text{over } j = 1 \ldots M \text{ and } k = 1 \ldots M \tag{3.25}$$

Linear-Elastic Energy

The *linear-elastic* energy (Miller *et al.*, 1993) of a two dimensional deformation field is:

$$\sum_{j=1}^{2} \sum_{k=1}^{2} \sum_i \frac{\lambda}{2} \left(\frac{\partial u_{ji}}{\partial x_{ji}} \right) \left(\frac{\partial u_{ki}}{\partial x_{ki}} \right) + \frac{\mu}{4} \left(\frac{\partial u_{ji}}{\partial x_{ki}} + \frac{\partial u_{ki}}{\partial x_{ji}} \right)^2 \tag{3.26}$$

where λ and μ are the *Lamé* elasticity constants. The elastic energy of the deformations can be computed by:

$$\begin{aligned} (\mu + \lambda/2) \mathbf{q_1}^T (\mathbf{D_2} \otimes \dot{\mathbf{D}}_\mathbf{1})^T (\mathbf{D_2} \otimes \dot{\mathbf{D}}_\mathbf{1}) \mathbf{q_1} &+ (\mu + \lambda/2) \mathbf{q_2}^T (\dot{\mathbf{D}}_\mathbf{2} \otimes \mathbf{D_1})^T (\dot{\mathbf{D}}_\mathbf{2} \otimes \mathbf{D_1}) \mathbf{q_2} \\ +\mu/2 \mathbf{q_1}^T (\dot{\mathbf{D}}_\mathbf{2} \otimes \mathbf{D_1})^T (\dot{\mathbf{D}}_\mathbf{2} \otimes \mathbf{D_1}) \mathbf{q_1} &+ \mu/2 \mathbf{q_2}^T (\mathbf{D_2} \otimes \dot{\mathbf{D}}_\mathbf{1})^T (\mathbf{D_2} \otimes \dot{\mathbf{D}}_\mathbf{1}) \mathbf{q_2} \\ +\mu/2 \mathbf{q_1}^T (\dot{\mathbf{D}}_\mathbf{2} \otimes \mathbf{D_1})^T (\mathbf{D_2} \otimes \dot{\mathbf{D}}_\mathbf{1}) \mathbf{q_2} &+ \mu/2 \mathbf{q_2}^T (\mathbf{D_2} \otimes \dot{\mathbf{D}}_\mathbf{1})^T (\dot{\mathbf{D}}_\mathbf{2} \otimes \mathbf{D_1}) \mathbf{q_1} \\ +\lambda/2 \mathbf{q_1}^T (\mathbf{D_2} \otimes \dot{\mathbf{D}}_\mathbf{1})^T (\dot{\mathbf{D}}_\mathbf{2} \otimes \mathbf{D_1}) \mathbf{q_2} &+ \lambda/2 \mathbf{q_2}^T (\dot{\mathbf{D}}_\mathbf{2} \otimes \mathbf{D_1})^T (\mathbf{D_2} \otimes \dot{\mathbf{D}}_\mathbf{1}) \mathbf{q_1} \end{aligned} \tag{3.27}$$

A regularisation based upon this model requires an inverse covariance matrix ($\mathbf{C_0}^{-1}$) that is not a simple diagonal matrix. This matrix is constructed as follows:

$$\mathbf{C_0}^{-1} = \begin{bmatrix} \mathbf{H_1} & \mathbf{H_3} & \mathbf{0} \\ \mathbf{H_3}^T & \mathbf{H_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 0 \end{bmatrix} \tag{3.28}$$

where:

$$
\begin{aligned}
\mathbf{H_1} = &\quad (\mu + \lambda/2)(\mathbf{D_2}^T\mathbf{D_2}) \otimes (\dot{\mathbf{D}}_\mathbf{1}^T\dot{\mathbf{D}}_\mathbf{1}) + \mu/2(\dot{\mathbf{D}}_\mathbf{2}^T\dot{\mathbf{D}}_\mathbf{2}) \otimes (\mathbf{D_1}^T\mathbf{D_1}) \\
\mathbf{H_2} = &\quad (\mu + \lambda/2)(\dot{\mathbf{D}}_\mathbf{2}^T\dot{\mathbf{D}}_\mathbf{2}) \otimes (\mathbf{D_1}^T\mathbf{D_1}) + \mu/2(\mathbf{D_2}^T\mathbf{D_2}) \otimes (\dot{\mathbf{D}}_\mathbf{1}^T\dot{\mathbf{D}}_\mathbf{1}) \\
\mathbf{H_3} = &\quad \lambda/2(\mathbf{D_2}^T\dot{\mathbf{D}}_\mathbf{2}) \otimes (\dot{\mathbf{D}}_\mathbf{1}^T\mathbf{D_1}) + \mu/2(\dot{\mathbf{D}}_\mathbf{2}^T\mathbf{D_2}) \otimes (\mathbf{D_1}^T\dot{\mathbf{D}}_\mathbf{1})
\end{aligned}
\tag{3.29}
$$

3.2.5 Templates and Intensity Transformations

Sections 3.2.2 and 3.2.3 have modelled a single intensity scaling parameter (q_{13} and w respectively), but more generally, the optimisation can be assumed to minimise two sets of parameters: those that describe spatial transformations ($\mathbf{q_s}$), and those for describing intensity transformations ($\mathbf{q_t}$). This means that the difference function can be expressed in the generic form:

$$b_i(\mathbf{q}) = f(\mathbf{s}(\mathbf{x}_i, \mathbf{q_s})) - t(\mathbf{x}_i, \mathbf{q_t}) \tag{3.30}$$

where $\mathbf{f}$ is the source image, $\mathbf{s}()$ is a vector function describing the spatial transformations based upon parameters $\mathbf{q_s}$ and $t()$ is a scalar function describing intensity transformations based on parameters $\mathbf{q_t}$. $\mathbf{x}_i$ represents the co-ordinates of the ith sampled point.

The previous subsections simply considered matching one image to a scaled version of another, in order to minimise the sum of squared differences between them. For this case, $t(\mathbf{x}_i, \mathbf{q_t})$ is simply equal to $q_{t1}g(\mathbf{x}_i)$, where q_{t1} is a simple scaling parameter and $\mathbf{g}$ is a template image. This is most effective when there is a linear relation between the image intensities. Typically, the template images used for spatial normalisation will be similar to those shown in the top row of Figure 3.6. The simplest least squares fitting method is not optimal when there is not a linear relationship between the images. Examples of nonlinear relationships are illustrated in Figure 3.7, which shows histograms (scatter-plots) of image intensities plotted against each other.

An important idea is that a given image can be matched not to one reference image, but to a series of images that all conform to the same space. The idea here is that (ignoring the spatial differences) any given image can be expressed as a linear combination of a set of reference images. For example these reference images might include different modalities (e.g., PET, SPECT, ^{18}F-DOPA, ^{18}F-deoxy-glucose, T_1-weighted MRI T_2^*-weighted MRI .. etc.) or different anatomical tissues (e.g., grey matter, white matter, and CSF segmented from the same T_1-weighted MRI) or different anatomical regions (e.g., cortical grey matter, sub-cortical grey mater, cerebellum ... etc.) or finally any combination of the above. Any given image, irrespective of its modality could be approximated with a function of these images. A simple example using two images would be:

$$\sum_i \left(f(\mathbf{M}\mathbf{x}_i) - (q_{t1}g_1(\mathbf{x}_i) + q_{t2}g_2(\mathbf{x}_i))\right)^2 \tag{3.31}$$

In Figure 3.8, a plane of a T1 weighted MRI is modelled by a linear combination of the five other template images shown in Figure 3.6. Similar models were used to simulate T2 and PD weighted MR images. The linearity of the scatter-plots (compared to those in Figure 3.7) shows that MR images of a wide range of different contrasts can be modelled by a linear combination of a limited number of template images. Visual inspection shows that the simulated images are very similar to those shown in Figure 3.6.

Alternatively, the intensities could vary spatially (for example due to inhomogeneities in the MRI scanner). Linear variations in intensity over the field of view can be accounted for by

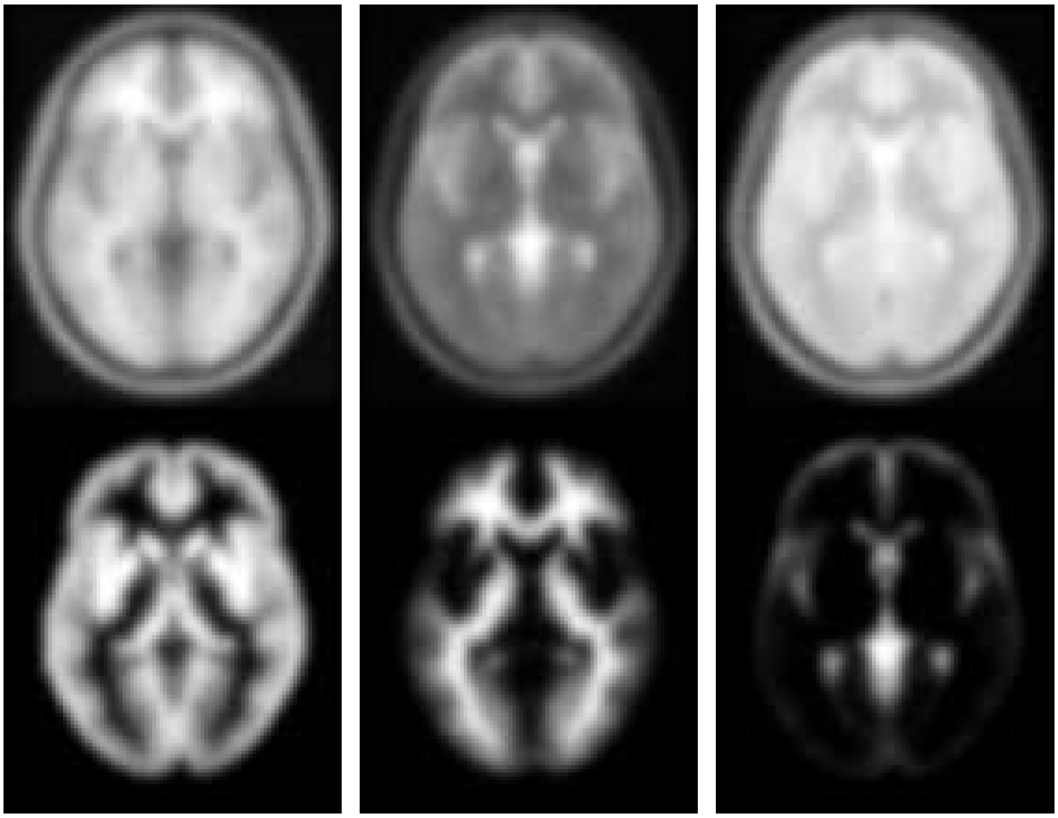

Figure 3.6: Example template images. Above: T1 weighted MRI, T2 weighted MRI and PD weighted MRI. Below: Grey matter probability distribution, White matter probability distribution and CSF probability distribution. All the data were generated at the McConnel Brain Imaging Centre, Montréal Neurological Institute at McGill University, and are based on the averages of about 150 normal brains. The original images were reduced to 2mm resolution and convolved with an 8mm FWHM Gaussian kernel to be used as templates for spatial normalisation.

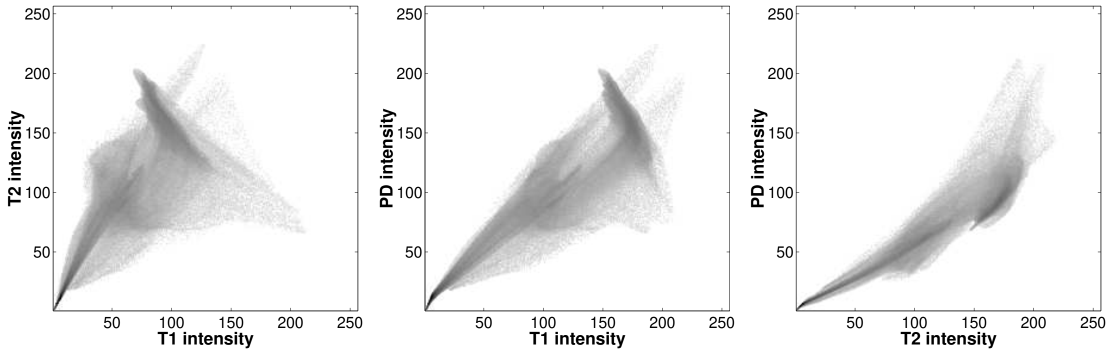

Figure 3.7: Two dimensional histograms of template images (intensities shown as $log(1+n)$, where n is the value in each bin). The histograms were based on the whole volumes of the template images shown in the top row of Figure 3.6.

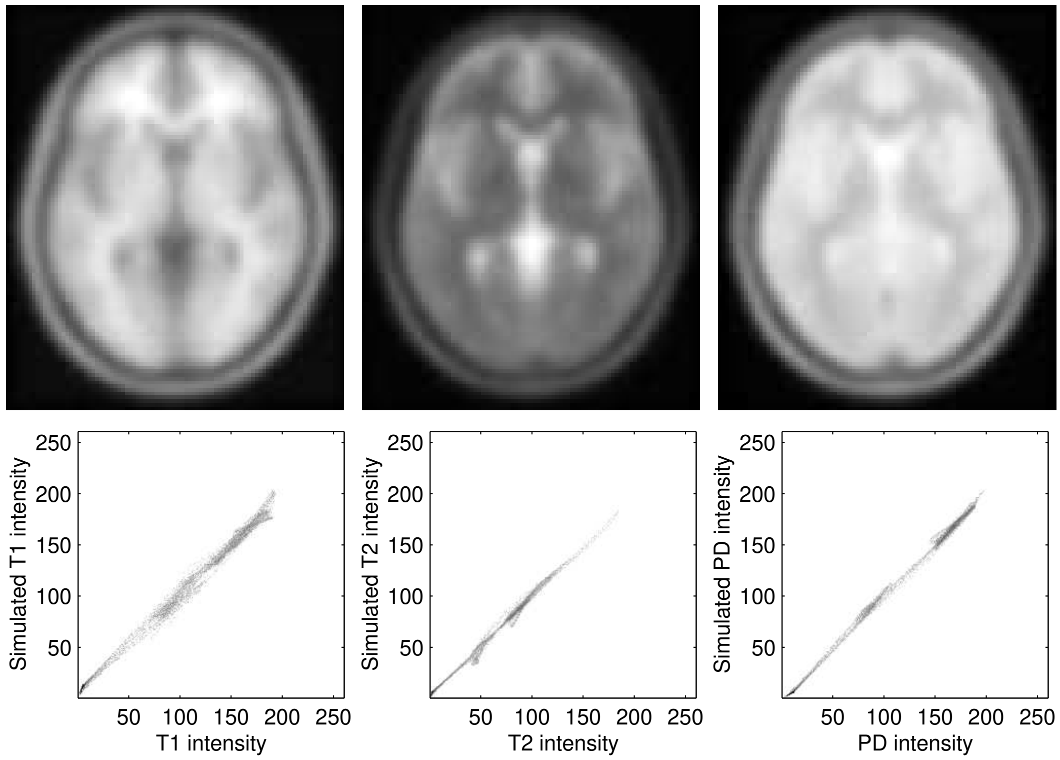

Figure 3.8: Simulated images of T1, T2 and PD weighted images, and histograms of the real images versus the simulated images.

optimising a function of the form:

$$\sum_i (f(\mathbf{x}_i, \mathbf{q_s}) - (q_{t1}g(\mathbf{x}_i) + q_{t2}x_{1i}g(\mathbf{x}_i) + q_{t3}x_{2i}g(\mathbf{x}_i) + q_{t4}x_{3i}g(\mathbf{x}_i)))^2 \tag{3.32}$$

More complex variations could be included by modulating with other basis functions (such as the DCT basis function set described in Section 3.2.3) (Friston *et al.*, 1995c). The examples shown so far have been linear in their parameters describing intensity transformations. A simple example of an intensity transformation that is nonlinear would be:

$$\sum_i (f(\mathbf{x}_i, \mathbf{q_s}) - q_{t1}g(\mathbf{x}_i)^{q_{t2}})^2 \tag{3.33}$$

Collins *et al.*(1994b) suggested that – rather than matching the image itself to the template – some function of the image should be matched to a template image transformed in the same way. He found that the use of gradient magnitude transformations lead to more robust solutions, especially in cases of partial volume coverage or intensity inhomogeneity artifacts (in MR images). Other spatially invariant moments may also contain other useful matching information. The algorithms described here perform most efficiently with smooth images. Much of the high frequency information in the images is lost in the smoothing step, but information about important image features may be retained in separate (smoothed) moment images. Simultaneous registrations (comparable to those in the previous chapter that matched grey matter with grey matter, and white matter with white matter) using these extracted features may be a useful technique for preserving information, while still retaining the advantages of using smooth images in the registration.

Another idea for introducing more accuracy would be to simultaneously spatially normalise co-registered images to corresponding templates. For example, by simultaneously matching a PET image to a PET template, at the same time as matching a structural MR image to a corresponding MR template, more accuracy could be obtained than by matching the images individually. Section 2.6 described a method of between modality registration where the first step involves simultaneous affine registration of a pair of images to a pair of templates in order to extract the rigid body transformation that maps the images together. There is no reason why nonlinear warping can not also be included in this model to further increase the accuracy of the rigid registration, while also improving the spatial normalisation.

3.3 Evaluation

The criteria for 'good' spatial transformations can be framed in terms of validity, reliability and computational efficiency. The validity of a particular transformation device is not easy to define or measure and indeed varies with the application. For example a rigid body transformation may be perfectly valid for realignment but not for spatial normalisation of an arbitrary brain into a standard stereotactic space. Generally the sorts of validity that are important in spatial transformations can be divided into (i) *Face validity*, established by demonstrating the transformation does what it is supposed to and (ii) *Construct validity*, assessed by comparison with other techniques or constructs. Face validity is a complex issue in functional mapping. At first glance, face validity might be equated with the co-registration of anatomical homologues in two images. This would be complete and appropriate if the biological question referred to structural differences or modes of variation. In other circumstances however this definition of face validity is not appropriate. For example, the purpose of spatial normalisation (either within or between subjects) in functional mapping studies is to maximise the sensitivity to neurophysiological change elicited by experimental manipulation of sensorimotor or cognitive state. In this case a better definition of a valid normalisation is that which maximises condition-dependent effects with respect to error (and if relevant inter-subject) effects. This will probably be effected when functional anatomy is congruent. This may or may not be the same as registering structural anatomy.

One approach is illustrated for assessing validity by comparing affine registrations both with and without the incorporation of the MAP scheme. It was found that the affine transformations derived using the Bayesian scheme are much more robust, and that the rate of convergence is greater. The final part of the evaluations illustrate that the nonlinear registration reduces structural variability on a global scale.

3.3.1 Evaluation of the MAP Scheme for Affine Registration

The affine registration scheme relies on optimising a set of 12 parameters that define the overall size and position of the head. The optimisation searches for the closest local minimum to the initial estimates, so it relies on these starting estimates being reasonably close to the optimum solution. In practice, this should not be problem. For PET images of the brain, the position of the subject within the scanner should not vary greatly. Also, the images are almost exclusively reconstructed in the same transverse orientation. Once a suitable set of starting estimates for one subject has been determined, it should be possible to use the same one for all subsequent subjects. The situation is slightly more complicated for MR images, where the images can be acquired in any orientation. However, in most medical image format standards, the orientation and position of the images is stored within the headers. This information can be automatically extracted and used as starting estimates for the registration.

The MAP optimisation scheme from Section 3.2.1 was evaluated for affine registrations with respect to the same optimisations performed without using the MAP scheme. It was found to converge more rapidly to a good solution, and also give much more robust and reliable results for limited data. These evaluations are detailed below.

Plots of convergence – with and without Bayesian extension

The affine registration algorithm was applied to 100 T1 weighted images, in order to match the images to a T1 template image. All images were smoothed with a Gaussian kernel of 8mm full width at half maximum. The voxels were reduced to $2 \times 2 \times 4$mm with a field of view of $256 \times 256 \times 128$mm in x, y and z respectively, in order to facilitate faster computations.

The optimisations were performed three times: (A) Without the Bayesian scheme, for a 12 parameter affine transformation. (B) With the Bayesian scheme, for a 12 parameter affine transformation. (C) Without the Bayesian scheme, for a six parameter rigid body transformation (to demonstrate that the Bayesian scheme is not simply optimising a rigid body transformation).

During the optimisation procedure, the images were sampled approximately every 8mm. 32 iterations were used, and the residual variance (χ^2) recorded for each iteration. 50 of the subjects were given good starting estimates (i), and 50 were given starting estimates that deviated from the optimal solution by about 10cm (ii).

There were 2 cases from (ii) in which the starting estimates were insufficiently close to the solution, for either (A) or (B) to converge satisfactorily. These cases have been excluded from the results.

Figure 3.9 shows the average χ^2 for all images plotted against iteration number. As can be seen from these plots, (B) leads to a more rapid estimation of the optimal parameters, even though convergence appears faster at the start of (A). The plot of convergence for (C) illustrates the point that the Bayesian method is not over-constrained and simply optimising a set of rigid body parameters.

Figure 3.10 compares the number of iterations required by (A) and (B) in order to reduce the χ^2 to within 1% of the minimum of both schemes. In several cases of (A), the optimisation had not converged within the 32 iterations. There were only 5 cases where (B) does not obtain a value for χ^2 that is as low as that from (A). In two of the cases, the results from (A) were very close to those from (B). However, in the other three cases, examination of the parameter estimates from scheme (A) showed that it had found a minimum that was clearly not a proper solution. The zooms determined, after 32 iterations, were (0.96,0.98,0.11), (2.10,0.72,0.0003) and (1.09,0.24,0.02). These are clearly not correct!

The algorithm requires relatively few iterations to reach convergence. The speed of each iteration for the affine normalisation depends upon the number of sampled voxels. On a SPARC Ultra 2, an iteration takes one second when about 26000 points (and their gradients) are sampled using tri-linear interpolation.

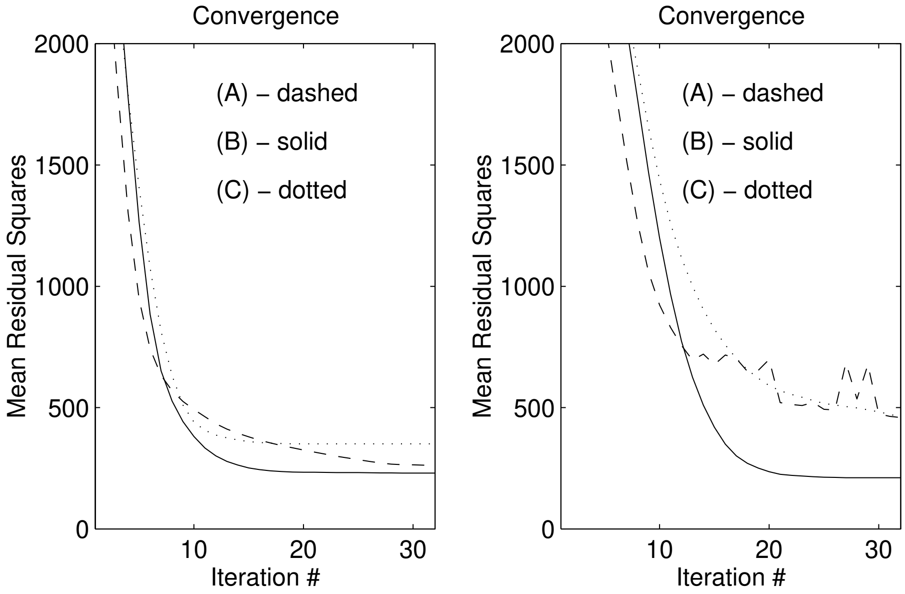

Figure 3.9: The average χ^2 for the images plotted against iteration number. Left: given good starting estimates (i). Right: given poor starting estimates (ii). The dashed lines (A) show convergence for a 12 parameter affine transformation without using the Bayesian scheme. The solid lines (B) show the same, but with the Bayesian scheme. Convergence for a six parameter rigid body transformation (C) is shown in the dotted lines.

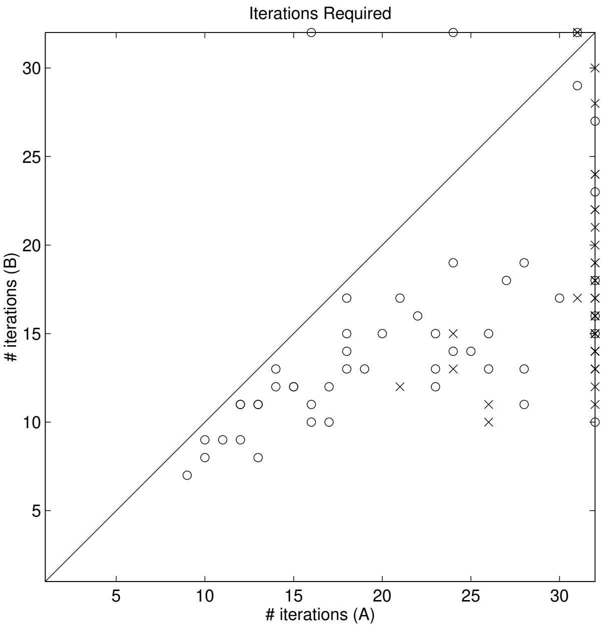

Figure 3.10: The number of iterations in which convergence to within 1% of the minimum mean residual sum of squares had not been reached. The non-Bayesian scheme (A) is on the X axis, with the Bayesian scheme (B) on the Y axis. Results from optimisations given good starting estimates are shown as circles, whereas those with bad starting estimates are shown as crosses.

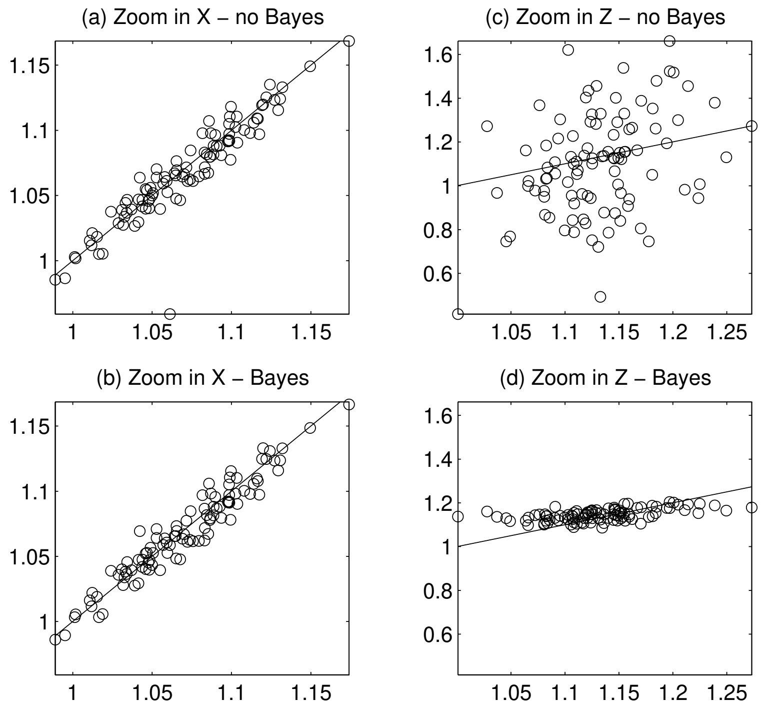

Figure 3.11: Plots of the parameter estimates from reduced data, against estimates using the complete data. As expected, the Bayesian scheme makes little difference for the estimates of the zoom in the X direction [(a) and (b)], whereas the Bayesian scheme heavily biases the zoom in Z towards the mean of the prior distribution [(c) and (d)].

Comparisons of Affine Normalisation with Limited Data

Occasionally the image that is to be spatially normalised is of poor quality. It may have a low signal to noise ratio, or it may contain only a limited number of slices. When this is the case, the parameter estimates for the spatial normalisation are likely to be unreliable. Here, a further comparison of affine registrations, with and without the incorporation of prior information [(E) and (D) respectively], is presented. This time, only four planes from the images were sampled, to simulate an effective field of view of 16 mm. The optimisations were given good initial parameter estimates, and the results compared with those obtained using the complete data.

The resulting parameter estimates from (D) and (E) are plotted against those from (B) in Figure 3.11. As can be seen from the plots, where the parameters can be estimated accurately, the results from (D) and (E) are similar. However, where there is not enough information in the images to determine an accurate parameter estimate, the results of (E) are properly biased towards the prior estimate.

3.3.2 Comparing Spatial Normalisation both With and Without Nonlinear Deformations

This section provides an anecdotal evaluation of the nonlinear warping techniques. Spatial normalisation is compared both with and without nonlinear deformations, and nonlinear deforma-

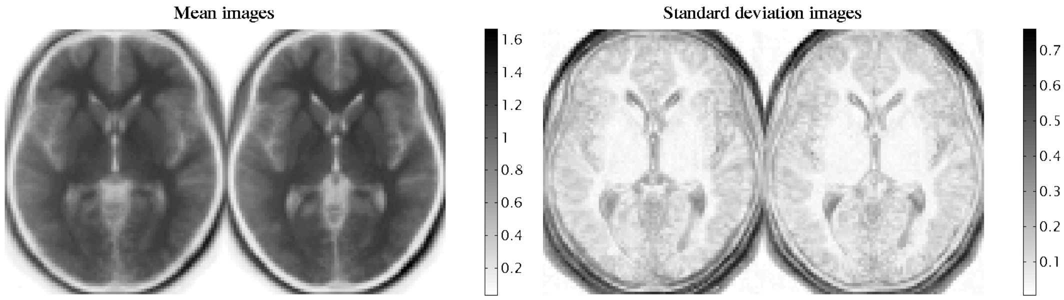

Figure 3.12: Means and standard deviations of spatially normalised T1 weighted images from 12 subjects. The images on the left of each pair were derived using only affine registration. Those on the right used nonlinear registration in addition to the affine registration.

tions compared with and without using Bayesian priors.

T1 weighted MR images of 12 subjects were spatially normalised to the same anatomical space. The normalisations were performed twice, first using only 12 parameter affine transformations and then using affine transformations followed by nonlinear warps. In both cases, the registration was weighted using the image shown in Figure 6.5 (page 129), so that any confounding effects of skull and scalp differences could be discounted. The nonlinear transformation used 392 ($7\times8\times7$) parameters to describe deformations in each of the directions, and four parameters to model a linear scaling and simple linear image intensity inhomogeneities (making a total of 1180 parameters in all). The basis functions were those of a three dimensional DCT, and the regularisation minimised the membrane energy of the deformation fields (using a value of 0.01 for λ). Twelve iterations of the nonlinear registration algorithm were performed.

Figure 3.12 shows pixel by pixel means and standard deviations of the normalised images. In order to create these mean and standard deviation images, the spatially normalised images were first scaled such that each of their weighted mean intensities was unity, where the weighting was done using the image in Figure 6.5. The mean image from the nonlinear normalisation shows more contrast and has edges that are slightly sharper. The standard deviation image for the nonlinear normalisation shows decreased intensities, demonstrating that the intensity differences between the images have been reduced. However, the differences tend to reflect changes in the global shape of the heads, rather than differences between the cortical anatomy. More examples of affine versus nonlinearly warped images are shown in Figures 4.15 and 4.16 of the next chapter. The average weighted residual squared difference between the normalised images and the mean image of the group was computed. Again, the weighting was done so that the residual squared differences were derived primarily from voxels in the brain. The average squared difference was 0.0237 for the affine only normalised images and 0.0187 for those that had also been nonlinearly warped. This shows that a 20% reduction of residual variance can be achieved by following an affine registration by the nonlinear warping described here.

This evaluation should illustrate the fact that nonlinear normalisation clearly reduces the sum of squared intensity differences between the images. The amount of residual variance could have been reduced further by decreasing the amount of regularisation. This however, may lead to some very un-natural looking distortions being introduced, due to an over-estimation of the *a priori* variability.

Evaluations like this tend to show more favourable results for less heavily regularised algorithms. With less regularisation, the optimum solution is based more upon minimising the difference between the images, and less upon knowledge of the *a priori* distribution of the parameters. This is illustrated for a single subject in Figure 3.13, where the distortions of gyral anatomy clearly have a very low face validity (lower right panel).

3.4 Discussion

Because the deformations are only defined by a few hundred parameters, the nonlinear registration method described here does not have the potential precision of some other methods. High frequency deformations can not be modelled because the deformations are restricted to the lowest spatial frequencies of the basis functions. This means that the current approach is unsuitable for attempting exact matches between fine cortical structures.

The current method is relatively fast, (taking in the order of 30 seconds per iteration – depending upon the number of basis functions used). The speed is partly a result of the small number of parameters involved, and the simple optimisation algorithm that assumes an almost quadratic error surface. Because the images are first matched using a simple affine transformation, there is less 'work' for the algorithm to do, and a good registration can be achieved with only a few iterations (less than 20). The method does not rigorously enforce a one-to-one match between the brains being registered. However, by estimating only the lowest frequency deformations and by using appropriate regularisation, this constraint is rarely broken.

The approach in this chapter searches for a MAP estimate of the parameters defining the warps. However, optimisation problems for complex nonlinear models such as those used for image registration can easily get caught in local minima, so there is no guarantee that the estimate determined by the algorithm is globally optimum. Even if the best MAP estimate is achieved, there will be many other potential solutions that have similar probabilities of being correct. A further complication arises from the fact that there is no one-to-one match between the small structures (especially gyral and sulcal patterns) of any two brains. This means that it is not possible to obtain a single objective high frequency match however good an algorithm is for determining the best MAP estimate. Because of these issues, registration using the minimum variance estimate (MVE) may be more appropriate. Rather than searching for the single most probable solution, the MVE is the average of all possible solutions, weighted by their individual probabilities of being correct. Although useful approximations have been devised (Miller *et al.*, 1993; Christensen, 1994), this estimate is still difficult to achieve in practice because of the enormous amount of computing power required. The MVE is probably more appropriate than the MAP estimate for spatial normalisation, as it is (on average) closer to the "true" solution. However, if the errors associated with the parameter estimates and also the priors are normally distributed, then the MVE and the MAP estimate are identical. This is partially satisfied by smoothing the images before registering them.

When higher spatial frequency warps are to be fitted, more DCT coefficients are required to describe the deformations. There are practical problems that occur when more than about the $8 \times 8 \times 8$ lowest frequency DCT components are used. One of these is the problem of storing and inverting the curvature matrix ($\mathbf{A}^T\mathbf{A}$). Even with deformations limited to $8 \times 8 \times 8$ coefficients,

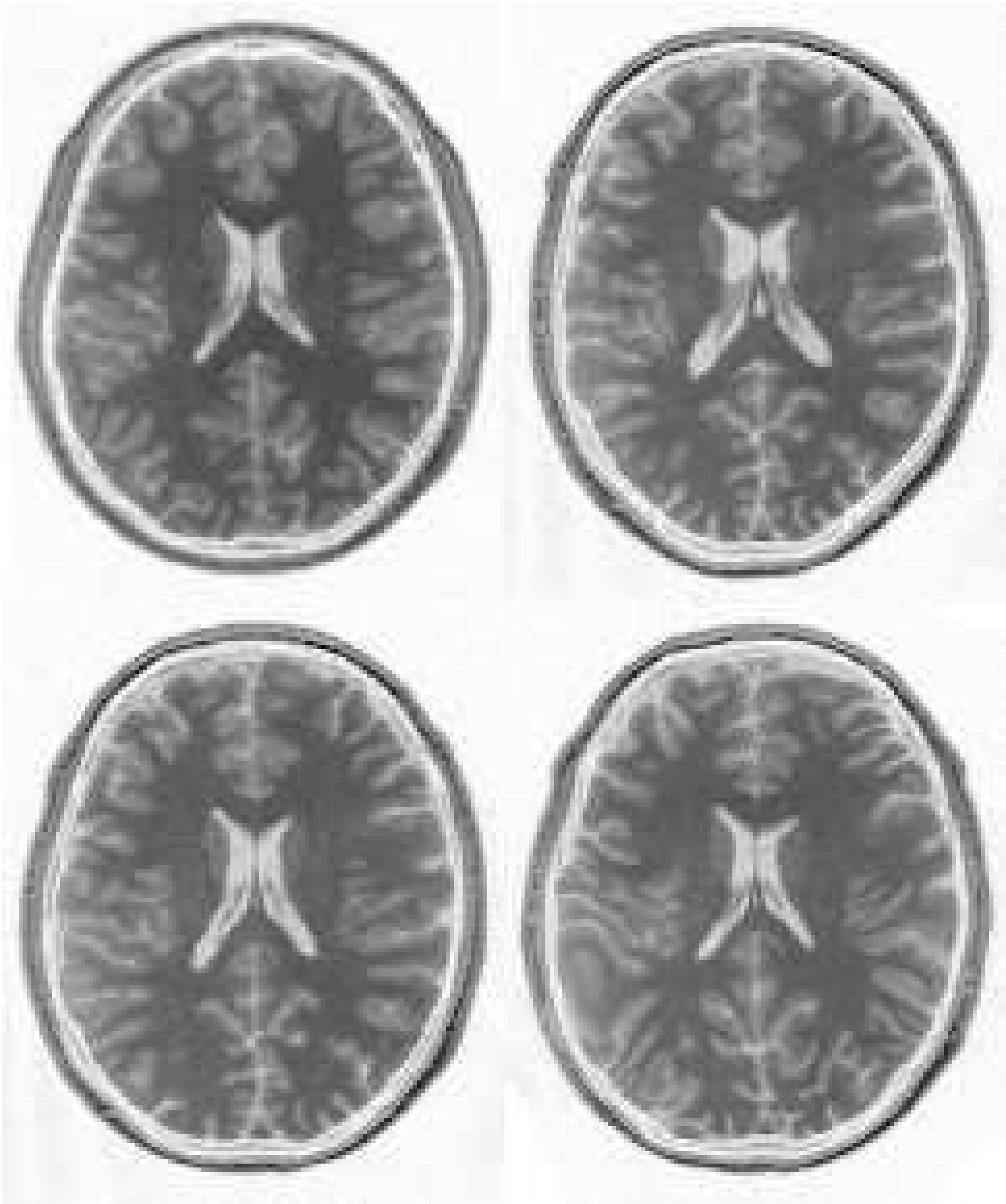

Figure 3.13: The image shown at the top-left is the template image. At the top-right is an image that has been registered with it using a 12-parameter affine registration. The image at the bottom-left is the same image registered using the 12-parameter affine registration, followed by a regularised global nonlinear registration (using 1180 parameters, 12 iterations and a λ of 0.01). It should be clear that the shape of the image approaches that of the template much better after nonlinear registration. At the bottom right is the image after the same affine transformation and nonlinear registration, but this time without using any regularisation. The mean squared difference between the image and template after the affine registration was 472.1. After the regularised nonlinear registration this was reduced to 302.7. Without regularisation, a mean squared difference of 287.3 is achieved, but this is at the expense of introducing a lot of unnecessary warping.

there are at least 1537 unknown parameters, requiring a curvature matrix of about 18Mbytes (using double precision floating point arithmetic). High-dimensional registration methods that search for more parameters should be used when more precision is required in the deformations. These include the method of Collins *et al.*(1994a), high dimensional linear-elasticity model (Miller *et al.*, 1993) and the viscous fluid models (Christensen *et al.*, 1996; Thompson & Toga, 1996). The next chapter also describes one such method.

In practice however, it may be meaningless to even attempt an exact match between brains beyond a certain resolution. There is not a one-to-one relationship between the cortical structures of one brain and those of another, so any method that attempts to match brains exactly must be folding the brain to create sulci and gyri that do not really exist. Even if an exact match is possible, because the registration problem is not convex, the solutions obtained by high dimensional warping techniques may not be truly optimum. High-dimensional registrations methods are often very good at registering grey matter with grey matter (for example), but there is no guarantee that the registered grey matter arises from homologous cortical features.

Also, structure and function are not always tightly linked. Even if structurally equivalent regions can be brought into exact register, it does not mean that the same is true for regions that perform the same or similar functions. For inter-subject averaging, an assumption is made that functionally equivalent regions lie in approximately the same parts of the brain. This leads to the current rationale for smoothing images from multi-subject functional imaging studies prior to performing statistical analyses. Constructive interference of the smeared activation signals then has the effect of producing a signal that is roughly in an average location. In order to account for substantial fine scale warps in a spatial normalisation, it is necessary for some voxels to increase their volumes considerably, and for others to shrink to an almost negligible size. The contribution of the shrunken regions to the smoothed images is tiny, and the sensitivity of the tests for detecting activations in these regions is reduced. This is another argument in favour of registering only on a global scale.

The constrained normalisation described here assumes that the template resembles a warped version of the image. Modifications are required in order to apply the method to diseased or lesioned brains. One possible approach is to assume different weights for different brain regions. Lesioned areas can be assigned lower weights, so that they have much less influence on the final solution.

The registration scheme described in this chapter is constrained to describe warps with a few hundred parameters. More powerful and less expensive computers are rapidly evolving, so algorithms that are currently applicable will become increasingly redundant as it becomes feasible to attempt more precise registrations. Scanning hardware is also improving, leading to improvements in the quality and diversity of images that can be obtained. Currently, most registration algorithms only use the information from a single image from each subject. This is typically a T1 MR image, which provides limited information that simply delineates grey and white matter. For example, further information that is not available in the more conventional sequences could be obtained from diffusion weighted imaging. Knowledge of major white matter tracts should provide structural information more directly related to connectivity and implicitly function, possibly leading to to improved registration of functionally specialised areas.

Chapter 4

High-Dimensional Image Warping

Contents

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 (Freeborough & Fox, 1998)), 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. 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.

The current approach assumes that the probability distribution from which the warps are assumed to be drawn (the prior probability distribution) should have some form of symmetry, in that the probability of obtaining a particular deformation should be identical to that for the inverse. 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 matrix at each point of the deformation having a log-normal distribution. The priors penalise unlikely deformations and enforce a continuous one-to-one mapping.

Often, the *a priori* probability distributions used by Bayesian registration schemes are linear (see Section 3.2.4), and include minimising the *membrane energy* of the deformation field (Amit *et al.*, 1991; Gee *et al.*, 1997), the *bending energy* (Bookstein, 1997a) or the *linear-elastic energy* (Miller *et al.*, 1993). 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 (Christensen *et al.*, 1994; Christensen *et al.*, 1996) 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 penalised.

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 utilises statistical rules. Unlikely deformations are penalised 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 optimisation procedure that simultaneously minimises 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.

[1] The word "topology" is used in the same sense as in "Topological Properties of Smooth Anatomical Maps" (Christensen *et al.*, 1995). If spatial transformations are not one-to-one and continuous, then the topological properties of different structures can change.

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 donated 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 resampling it at the mapped co-ordinates.

This section begins by describing how the deformation fields are parameterised as piecewise affine transformations within a finite element mesh. The registration is achieved by matching the images while simultaneously trying to maximise the smoothness of the deformations. Bayesian statistics are used to incorporate this smoothness into the registration, and a method of optimisation 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. 3.2):

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

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 maximises $p(\mathbf{Y}|\mathbf{b})$. 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 minimises 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 optimisation problem, whereby the cost function to be minimised 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 minimise 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 centres 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

[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.

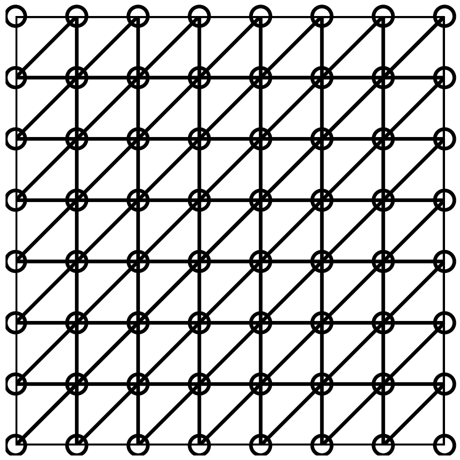

Figure 4.1: For the two dimensional registration problem, the area of the template image (**g**) is divided into a triangular mesh where the nodes are centered on the pixels.

(y_{13}, y_{23}) respectively, then the 3×3 affine mapping (**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 (**J**) of this affine mapping is simply obtained from matrix **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, **J** can be decomposed into two unitary matrices (**U** and **V**) and a diagonal matrix (**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 **S** contains the singular values, and these represent relative stretching in orthogonal directions. The determinant of **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 **f** and **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 **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 realise 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 **S** being drawn from a log-normal 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 "regularisation parameter" [4] . A log-normal distribution for each diagonal element of **S**, also implies a log-normal distribution for $|\mathbf{J}|$, since $log(|\mathbf{J}|) \equiv log(s_{11}) + log(s_{22})$ and both

[3] Complications arise when the determinant of **J** is negative. In this case either **U** or **V** will also incorporate a reflection by having a negative determinant. However, this should not cause problems since the registration prevents the determinant of **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.

$log(s_{11})$ 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. 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 (Guéziec & Hummel, 1995). 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 neighbours. When the deformations are one-to-one, it is possible to compute their inverses (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 (Christensen *et al.*, 1995). 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, since the volumes of non-neighbouring 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 tem-

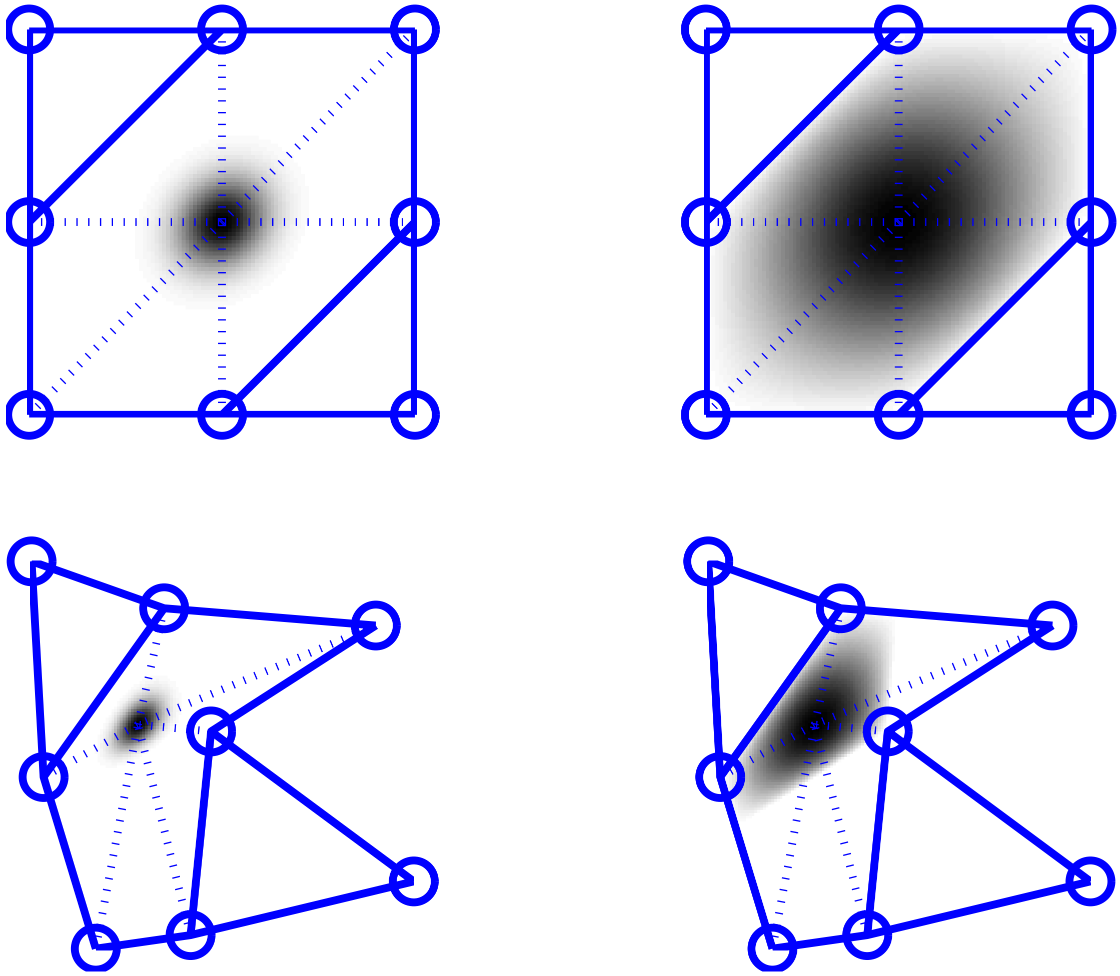

Figure 4.2: Probability density functions relating to the position of the centre node, assuming that all other voxels remain at fixed locations on a regular grid. *Left column*: using heavy regularisation $\lambda = 10$. *Right column*: using light regularisation $\lambda = 1$. *Above*: on a regular 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.

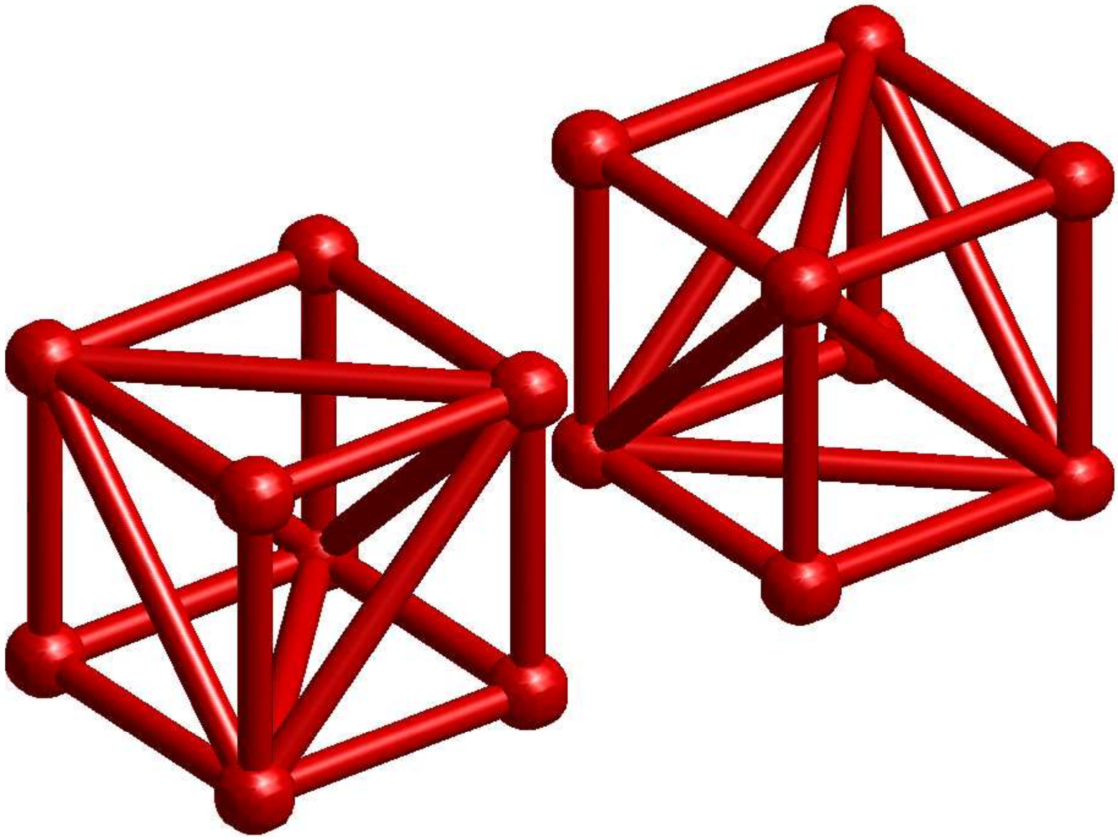

Figure 4.3: The volume of the template image is divided into a mesh of irregular tetrahedra, where the vertices of the tetrahedra are centered on the voxels. Groups of eight voxels are considered as little cubes. The volume of each cube is divided into five tetrahedra, in one of the two possible arrangements 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.

plate 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 singular values of the Jacobian matrices at every point in the deformation being drawn from a log-normal distribution. 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. C code for an analytic form for computing the singular values of a 3×3 matrix was developed using the Symbolic Toolbox of Matlab 5.3 (The MathWorks, Natick, Mass., USA), and also a function describing the rates of change of the singular values with respect to changes to the Jacobian matrix. Unfortunately, the functions were very susceptible to severe rounding errors – particularly for matrices close to the identity matrix. This instability, and the complexity of computing a SVD meant that an alternative method was required.

Using the SVD regularisation, 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:

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

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 regularisation 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 $(s_{ii}^2 + s_{ii}^{-2} - 2)/4$.

4.2.5 The Optimisation Algorithm

The images are matched by estimating the set of parameters ($\mathbf{Y}$) that maximises 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 optimisation involves determining the rate of change of the posterior

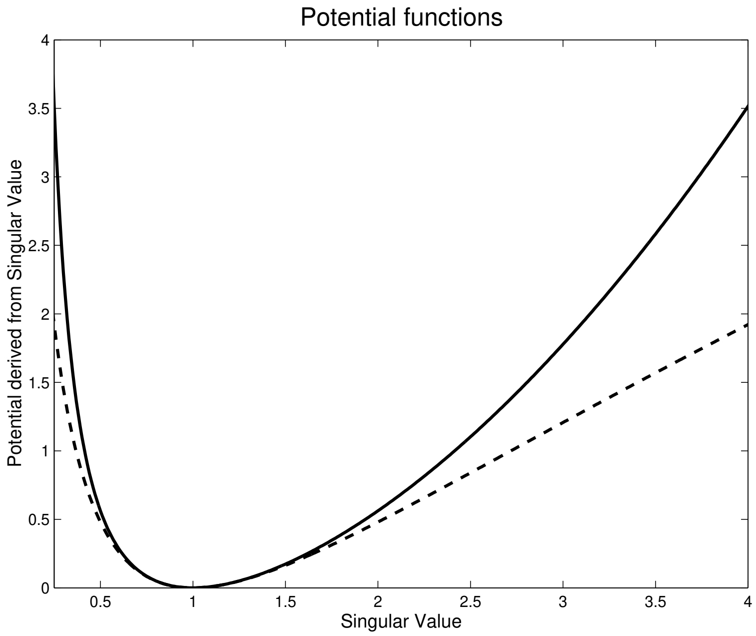

Figure 4.4: A comparison of the different cost functions. The dotted line shows the potential estimated from $(log(s_{ii}))^2$, where s_{ii} is the ith singular value of a Jacobian matrix. The solid line shows the new potential, which is based on $(s_{ii}^2 + s_{ii}^{-2} - 2)/4$. For singular values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

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 $\mathbf{y}_i$:

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

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 from these local tetrahedra. Approximately half of the nodes form a vertex in 8 neighbouring 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

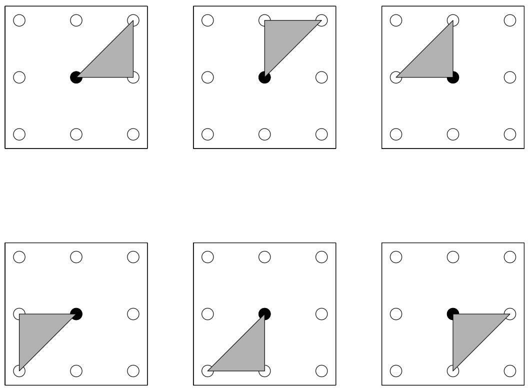

Figure 4.5: The six neighbouring triangles whos Jacobian matrices are influenced by translating the central point.

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 neighbouring triangle or tetrahedron may become negative. This would mean a violation of the one-to-one constraint in the mapping (since neighbouring 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. 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

```
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 $d = j_{22}j_{11} - j_{12}j_{21}$.

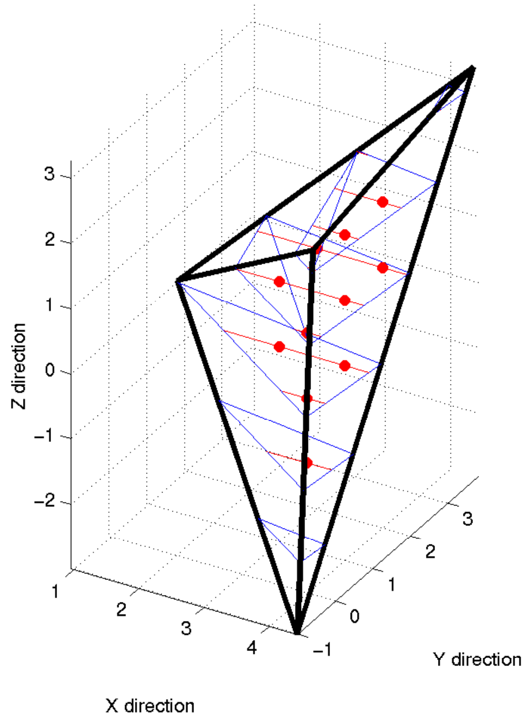

Figure 4.7: An illustration of how voxels are located within a tetrahedron.

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 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 example 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 (of 100×100 pixels) were constructed, one of them a circle, and the other a square, with pixel intensities ranging from zero to one. The circle was warped to match the square, and the square to match the circle using a value of one 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 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. 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 some form of regridding scheme (Christensen *et al.*, 1996) may need to be incorporated to allow larger deformations to be modelled correctly. However, because brains are all of similar shape, the ability to model very large deformations in this way may not be necessary in neuro-imaging applications.

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. Each image contained 200×256 pixels (where each pixel was 1×1mm), so the registration involved optimising a total of 100584 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 8mm full

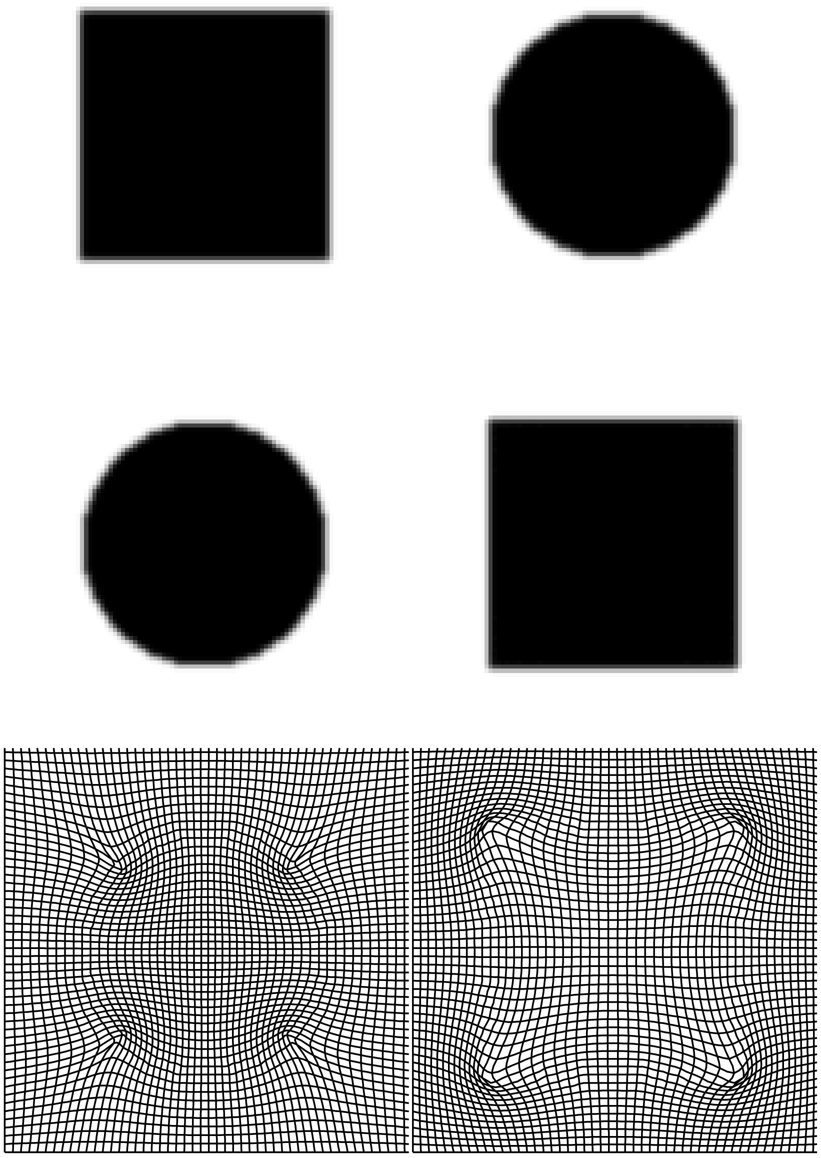

Figure 4.8: Demonstration using simulated data. *Above left*: original square. *Above right*: original circle. *Centre left*: square deformed to match the circle. *Centre right*: circle deformed to match the square. *Below left*: deformation field applied to the circle in order to warp it to match the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

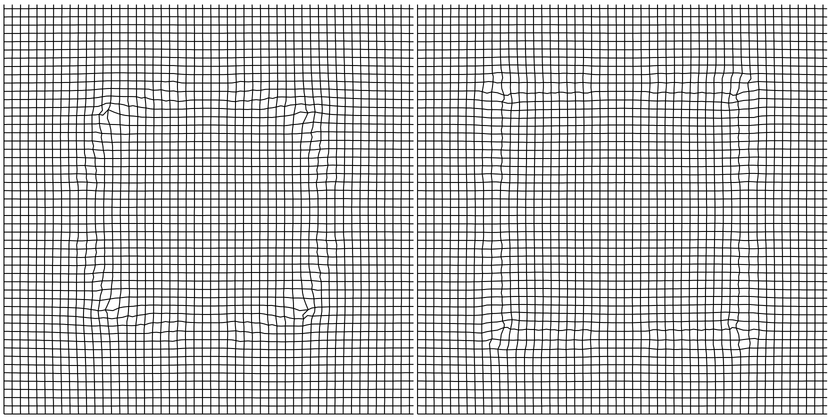

Figure 4.9: Demonstration of the reversibility of the deformations obtained by combining forward and reverse deformations. *left*: deformation field that warps from the circle to the square and back to the circle. *right*: deformation field that warps from the square to the circle and back to the square.

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

Three Dimensional Example

A pair of three dimensional brain images were first registered using the global registration methods described in Chapter 3. Differences in size and orientation between the images were first removed by performing a 12-parameter affine registration. Most of the remaining measurable shape differences were low frequency, and estimated using the DCT basis function approach (with 1176 parameters). The combination of these two methods provided a good starting point for estimating the optimum high-dimensional deformation field. A value of one was used for λ, with 40 iterations of the algorithm. 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). 94.5 Mbytes of memory were required by the programme (6.75 Mbytes for each of the 8-bit images, and 81 Mbytes for a single precision floating point representation of the deformation field). This meant that all the data could be stored in random access memory. Figure 4.11 shows the two registered brain images, and the corresponding deformation fields are shown in Figure 4.12.

The symmetry of the registration process was examined by repeating the registration, but swapping the source and template images. This gave a second deformation field, which, if the registration procedure is symmetric, should be the inverse of the first. In order to do the comparisons, the inverses of the two deformation fields were computed as described in Section 4.2.6. An example of part of a deformation field computed both ways is shown in Figure 4.13. The average absolute discrepancy between the first and the inverse of the second deformation, and

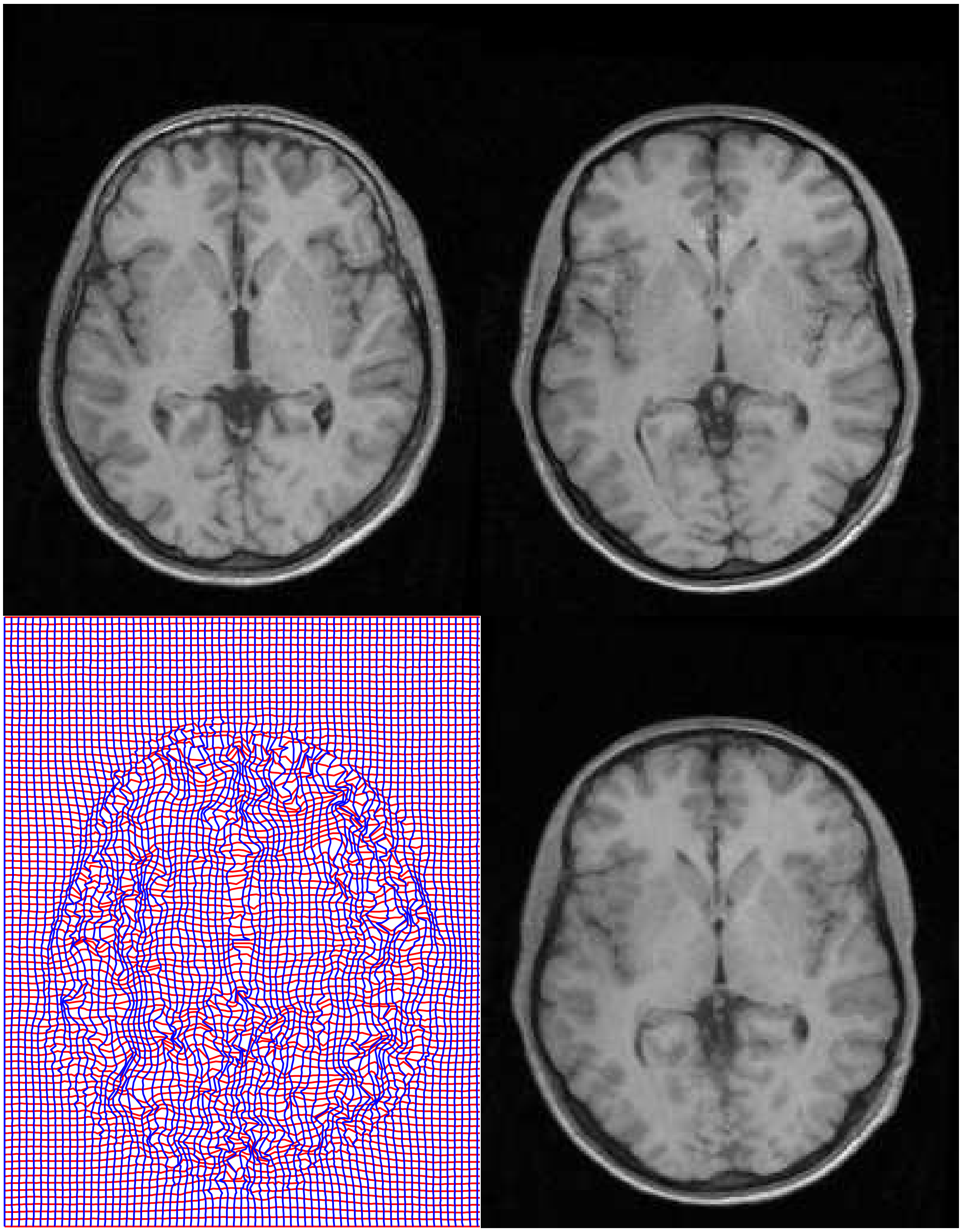

Figure 4.10: *Above left*: the unwarped source image. *Above right*: the template image. *Below left*: the deformation field applied to the source image in order to warp it to match the template image. *Below right*: the source image after warping to match the template.

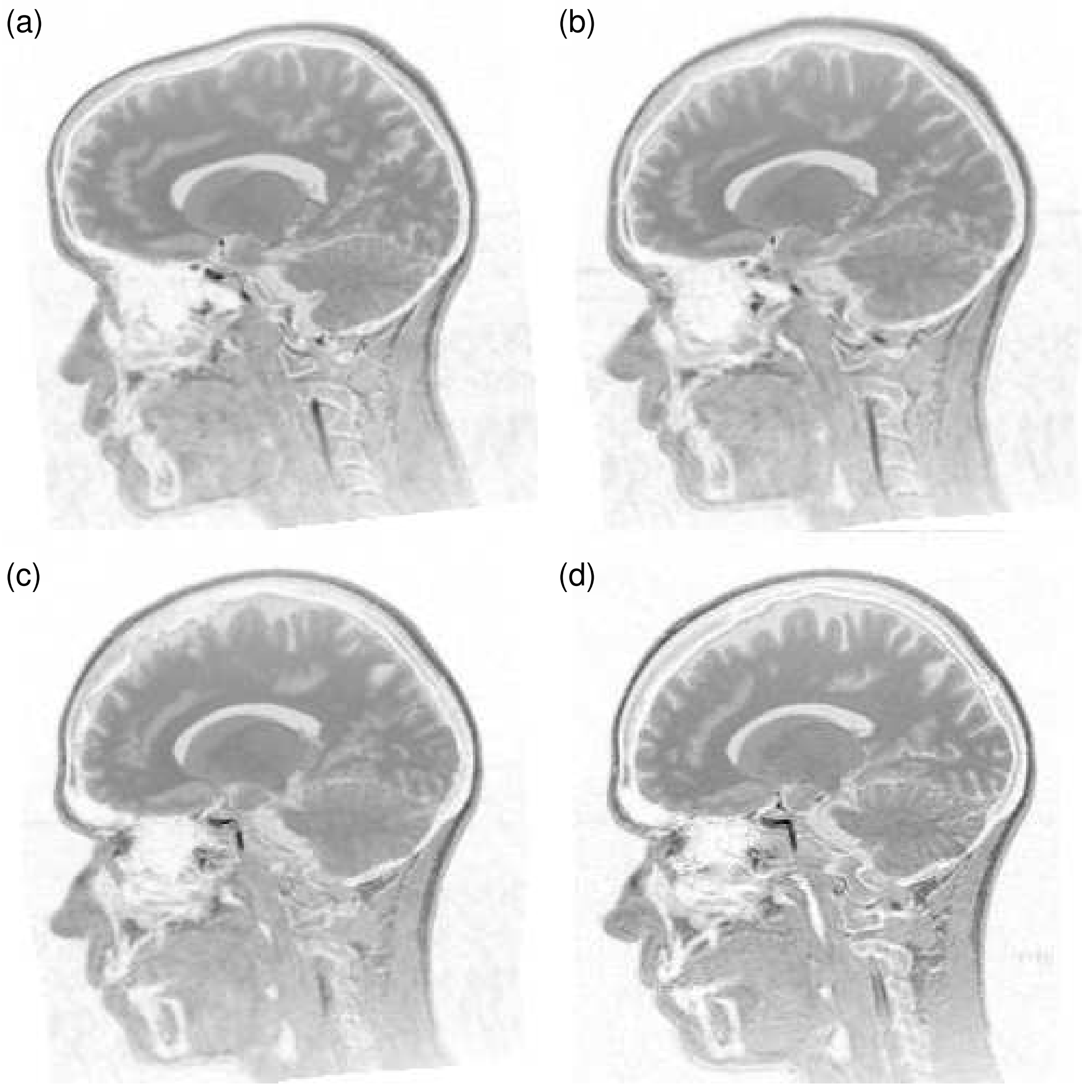

Figure 4.11: A sagittal plane from two images registered together. The template (reference) image is shown in (d). (a) shows the source image after affine registration to the template image. The source image after the basis function registration is shown in (b), and the final registration result is in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

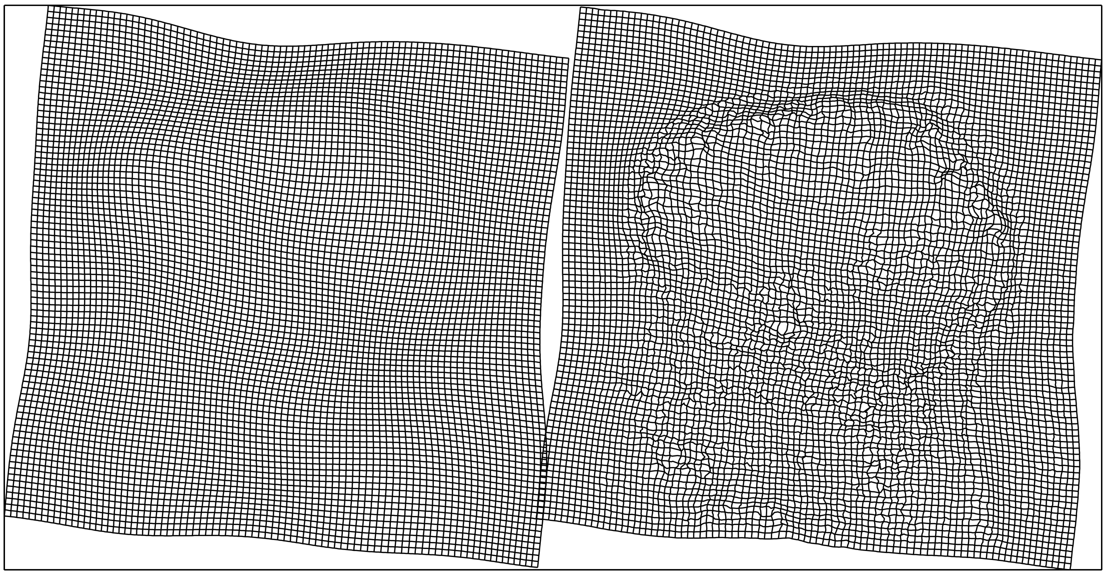

Figure 4.12: The deformation fields corresponding to the images in Figure 4.11. Two components (vertical and horizontal translations) of the field after affine and basis function registration are shown on the left, whereas the final deformation field is shown on the right.

between the second and the inverse of the first was found to be 1.36 and 1.37mm. Part of the error is because the likelihood function is not symmetric, as it uses only the gradient of one of the images. A fully symmetric likelihood function (Christensen, 1999) would be required to make the transformations more consistent.

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 stereotactic 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 (Le Briquer & Gee, 1997; Woods *et al.*, 1998a). In order to achieve more consistency in the warps, the images should probably 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.

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 stereotactic 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. This model also incorporated a spatially varying σ^2, which was computed from the

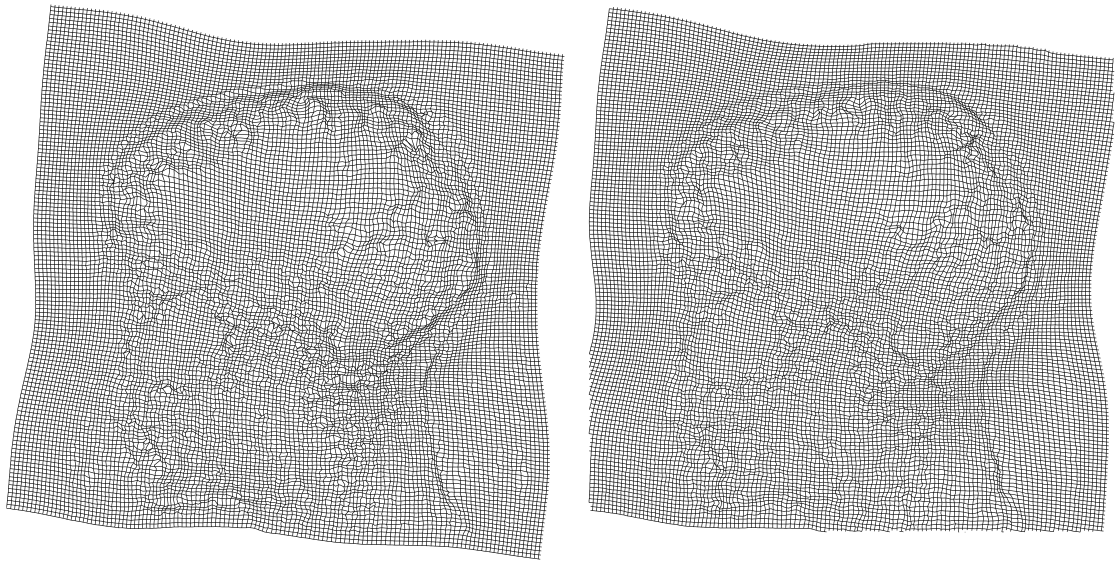

Figure 4.13: A deformation computed by warping the first image to the second (left), and by taking the inverse of the deformation computed by warping the second to the first (right).

residual variance at each pixel. 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. The residual variance was also re-computed and used as the new estimate for the nonstationary σ^2. This was continued until the algorithm converged, and the Gibbs potential of the system was minimised. The template resulting from this procedure is shown in Figure 4.14. 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 begin 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 4.15 and 4.16). 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. A value of four was used for λ. 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

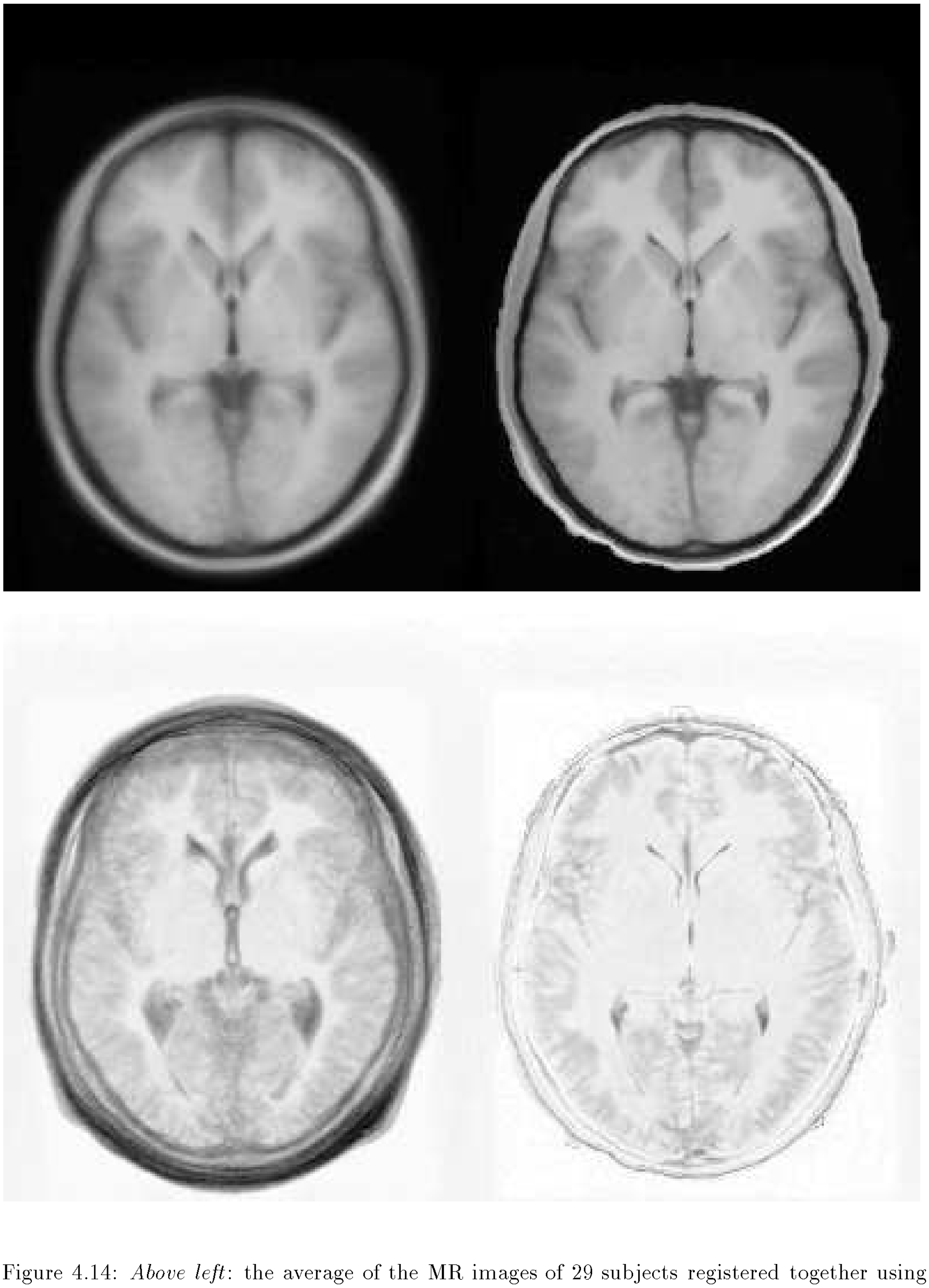

Figure 4.14: *Above left*: the average of the MR images of 29 subjects registered together using a 12-parameter affine registration. *Above right*: the average (both shape and intensity) of the same 29 MR images after registering together using the method described in Section 4.3.3. *Below left*: the 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.

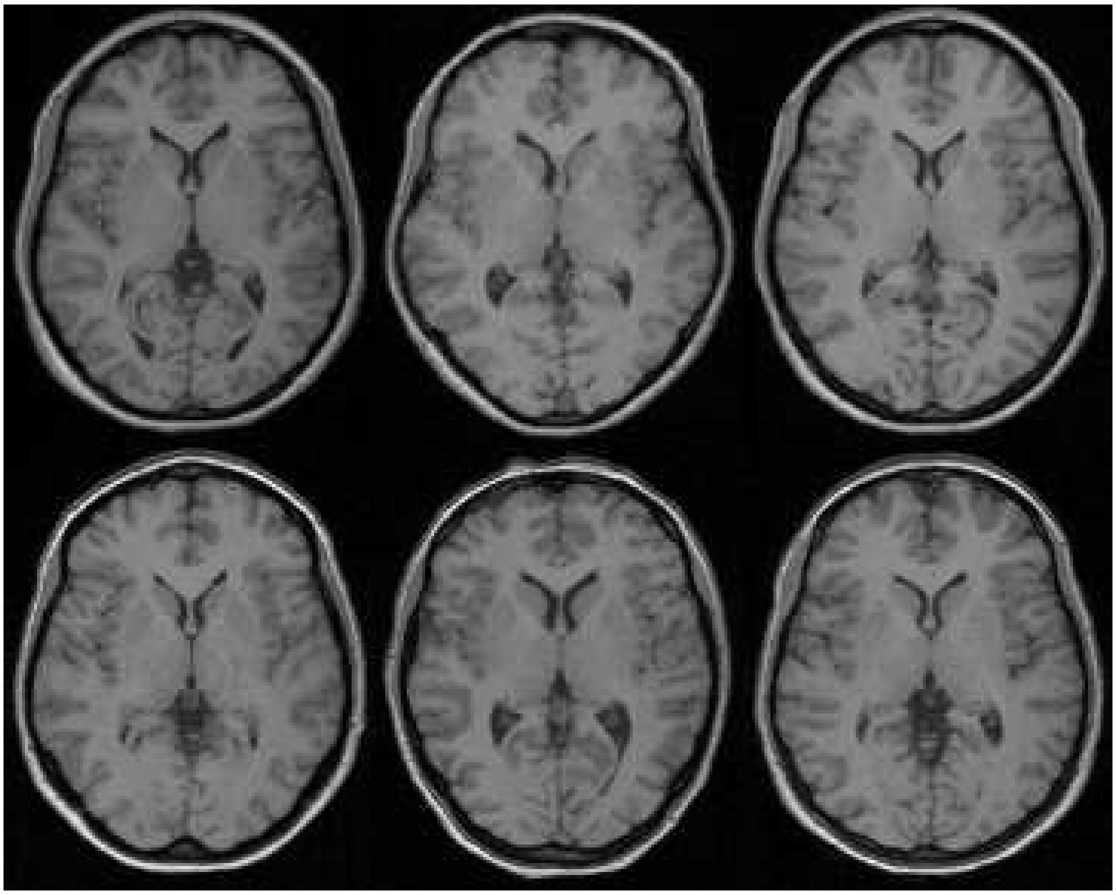

Figure 4.15: Images of six subjects registered using a 12-parameter affine registration (see also Figures 4.16 and 4.17). The affine registration matches the positions and sizes of the images.

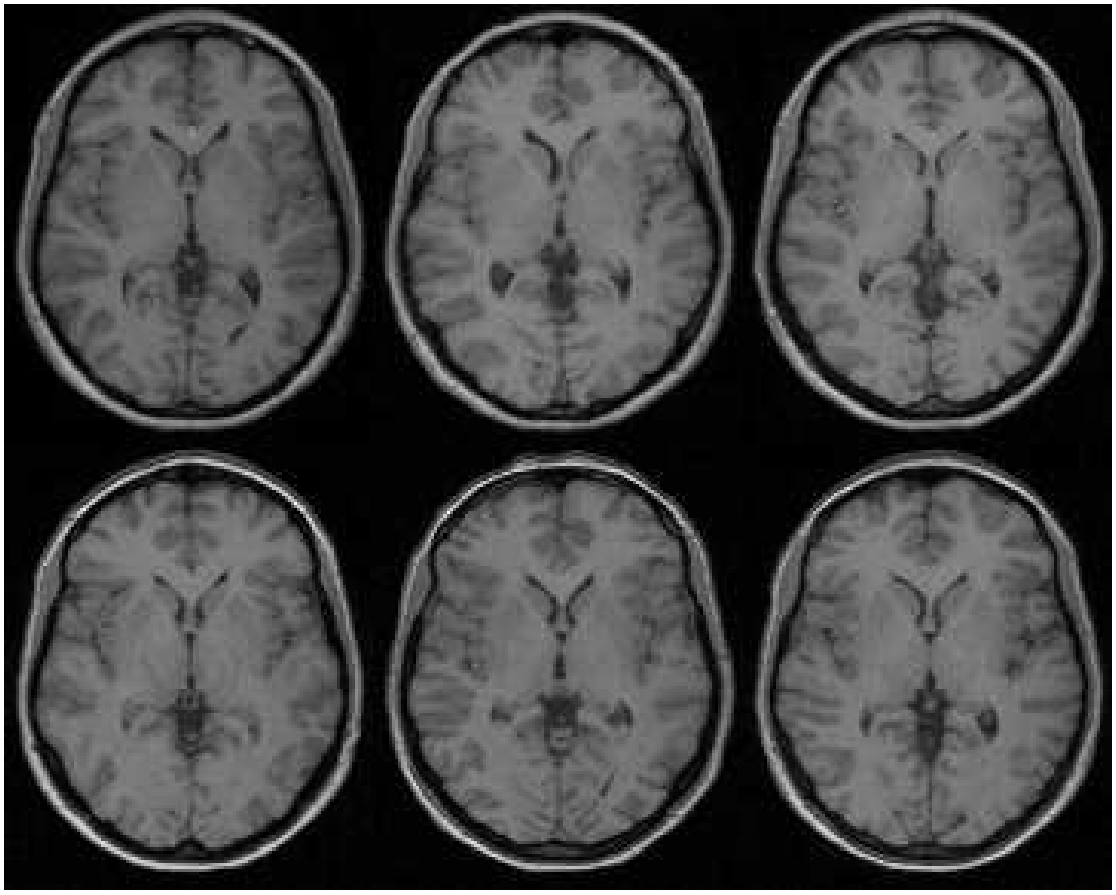

Figure 4.16: Six subjects brains registered with both affine and basis function registration (see also Figures 4.15 and 4.17). The basis function registration estimates the global shapes of the brains, but is not able to account for high spatial frequency warps.

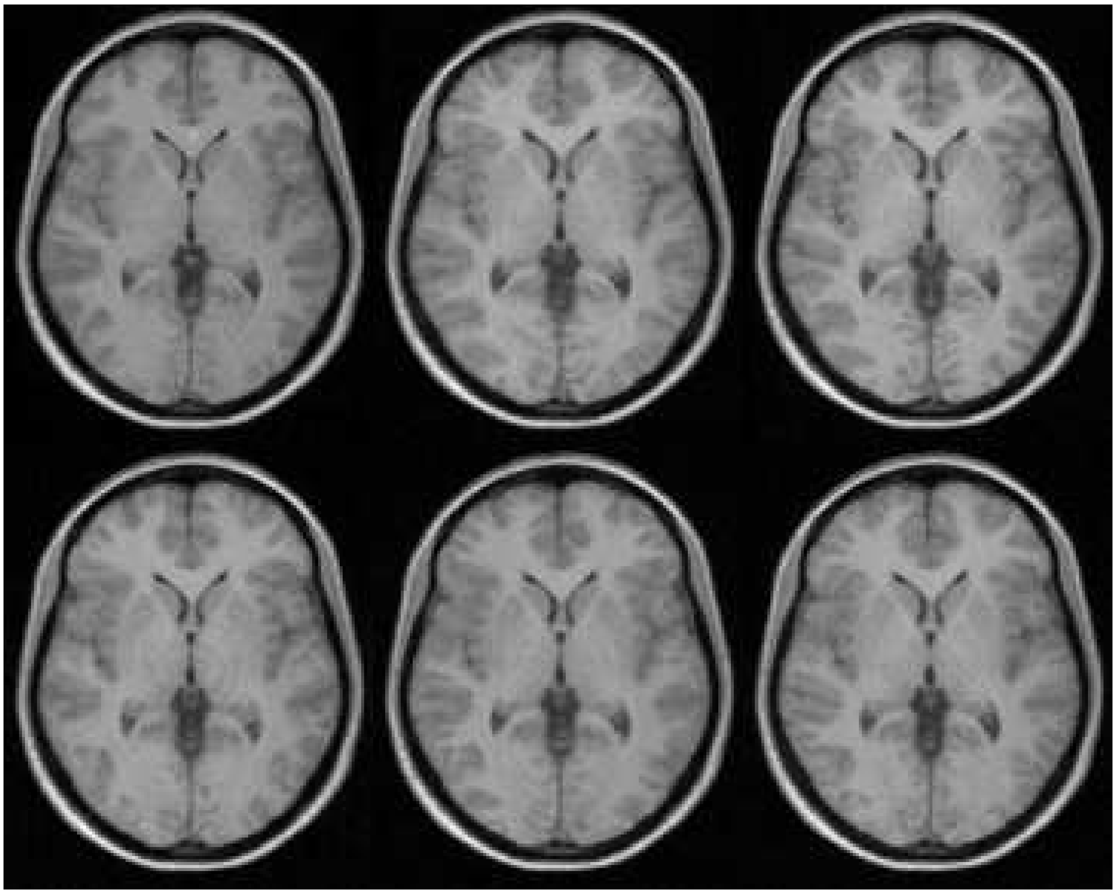

Figure 4.17: The images of the six brains after affine and basis function registration, followed by high-dimensional image registration using the methods described in this chapter (see also Figures 4.15 and 4.16). The high-dimensional transformations are able to model high frequency deformations that can not be achieved using the basis function approach alone.

process continued for a total of four times. A plane from each of the spatially transformed images is shown in Figure 4.17.

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 optimisation algorithm has reduced the likelihood potentials). In theory, the wrong structures could be registered together, but distorted so that they look identical. For the examples shown here, the mapping between the images appeared satisfactory for brain structures that are consistently present and identifiable.

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 using tools within Matlab 5.3 (The MathWorks, Natick, Mass., USA). A number of points were selected that were near the surface of the template brain (the average of the spatially transformed images). The points did not refer to any particular distinctive features, but were randomly selected from coronal slices spaced approximately 20mm apart. By using the computed spatial transformations, these points were projected on to the closest corresponding

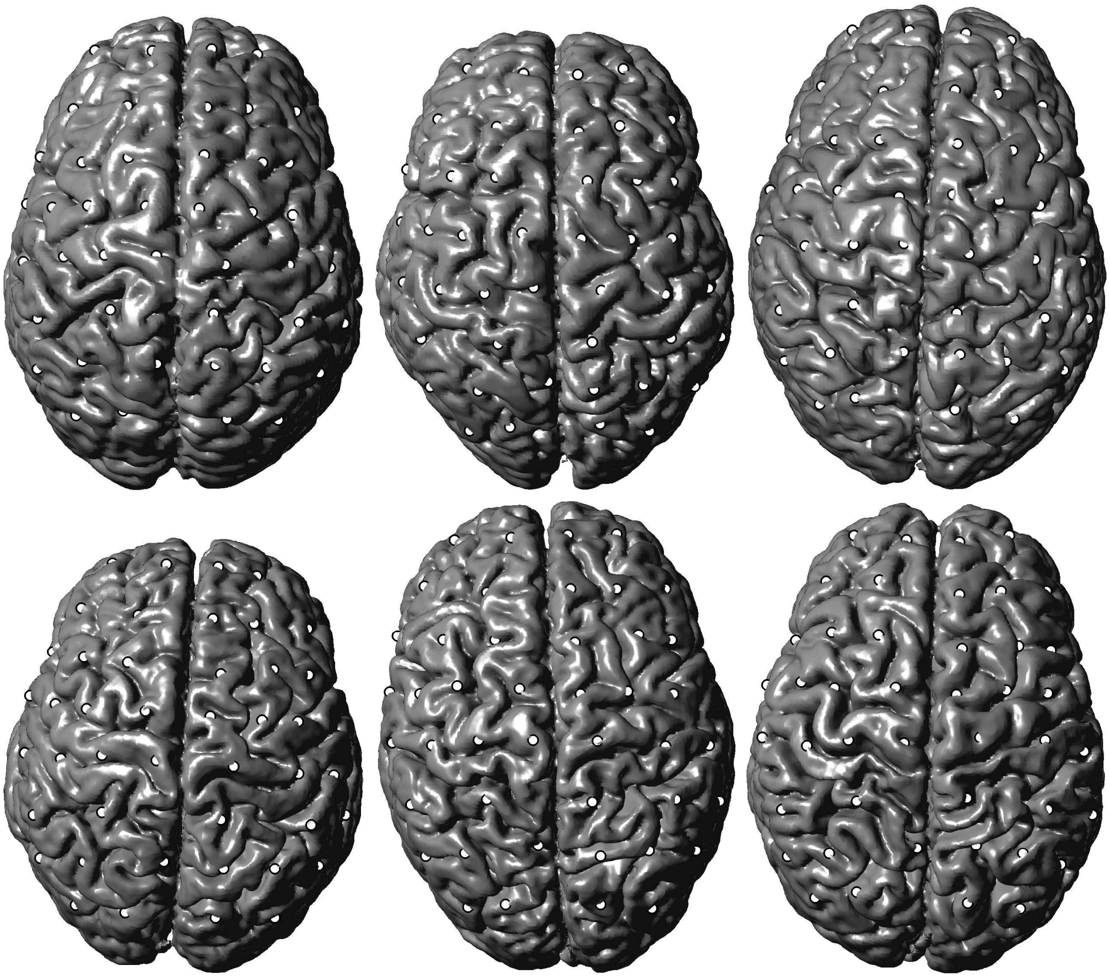

Figure 4.18: Rendered surfaces of the original six brains. The white markers correspond to equivalent locations on the brain surfaces as estimated by the registration algorithm.

location on the rendered brain surfaces. Figure 4.18 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 4.19 shows five images that have been transformed in this way to match the same image.

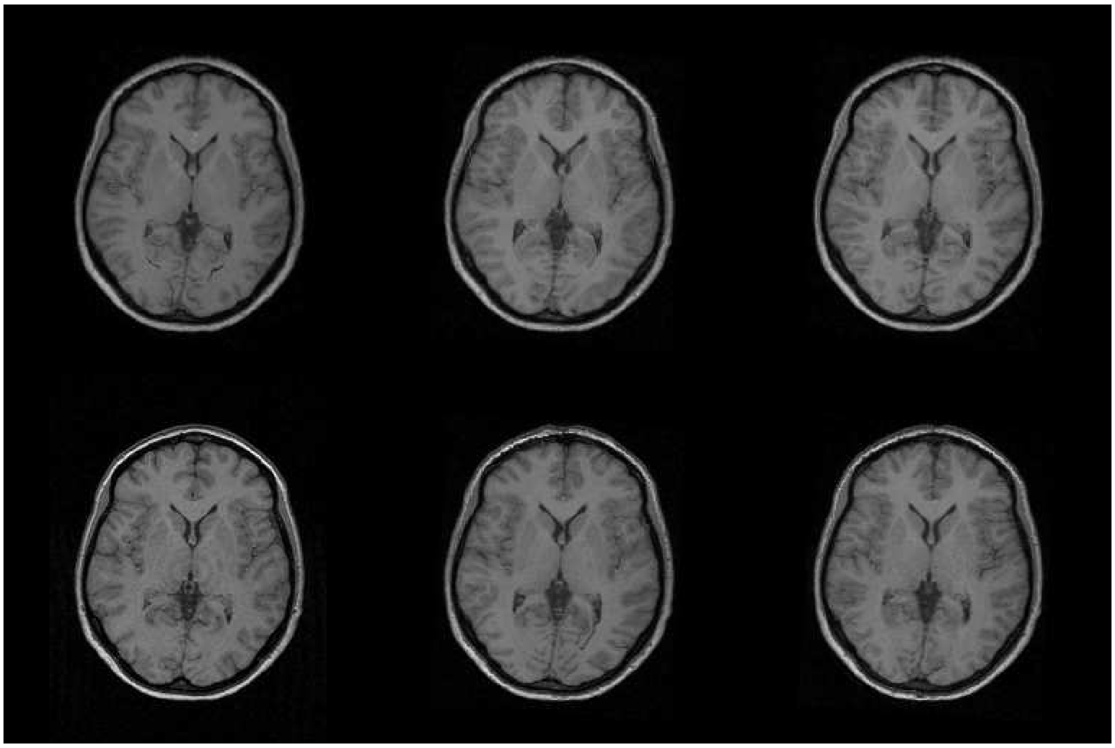

Figure 4.19: By combining the warps, it is possible to compute a mapping between any pair of images. In this example, the remaining images were all transcribed to match the one shown at the lower left.

4.4 Discussion

Validation of warping methods is a complex area, and no empirical evaluation is presented here. 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 normalisation 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 normalisation 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 3.4.

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 the registration method is dependent on four main elements: the parameterisation 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 Parameterising the Deformations

The deformations are parameterised 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 Goshtasby (1987). 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 probably require some complicated integration method.

In terms of speed, this method does not compare favourably with some other high-dimensional intensity based registration algorithms (Thirion, 1995), 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 minimising 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 nonstationary variance map and possibly model covariance between neighbouring voxels, or even covariance between intensities in different regions (e.g., see Section 3.2.5).

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 4.20. 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 4.19), 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 nonuniformities 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 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 (Amit *et al.*, 1991; Gee *et al.*, 1997; Miller *et al.*, 1993; Bookstein, 1997a; Bookstein, 1989), 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 singular values of the Jacobian matrices having log-normal distributions, but the more rapidly computed function described in Section 4.2.4 is probably a close enough approximation.

The penalty function is invariant to the relative orientations of the images. It does not penalise rotations or translations in isolation, only those relative to the position of neighbouring voxels. In order to reposition a region relative to its neighbours, it is necessary to introduce scaling and shearing into the affine transformations. It is this scaling and shearing that the model penalises, rather than the position and orientation itself (see Figure 4.21).

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 (Lester *et al.*, 1999), so a set of nonstationary priors should, in theory, produce more valid MAP estimates.

Much of the nonstationary variability will be higher in some directions than others. An alternative way of understanding the penalty function based on log-normally distributed 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 minimising the sum of squares of the Hencky tensor elements. Anisotropic variability could be modelled by assuming different variances for each Hencky tensor element, thus allowing more stretching or contraction in some directions

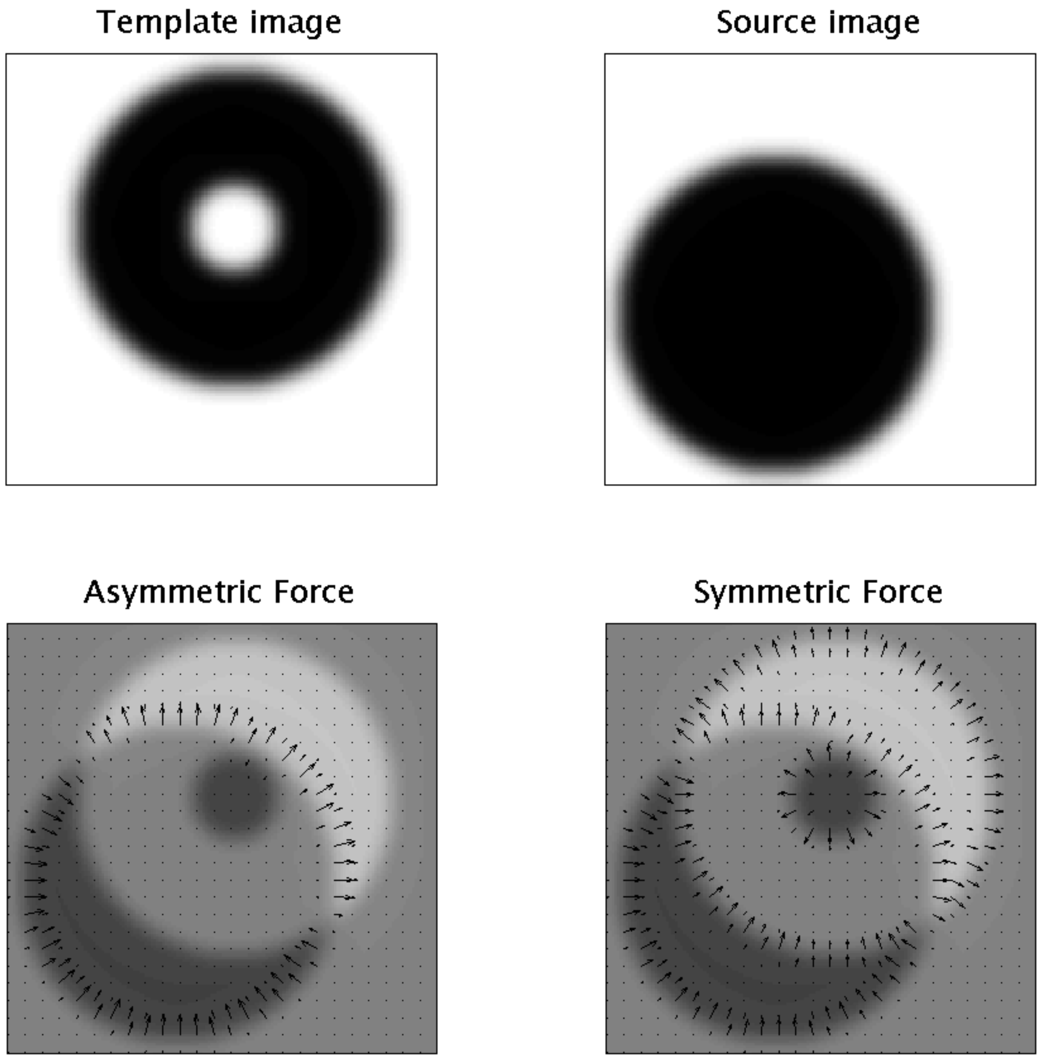

Figure 4.20: A comparison of a symmetric with an asymmetric likelihood potential. The arrows on the lower images show the directions in which the source image would be deformed. The template image contains a feature that is not found in the source image. If a registration is based only on gradients of 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.

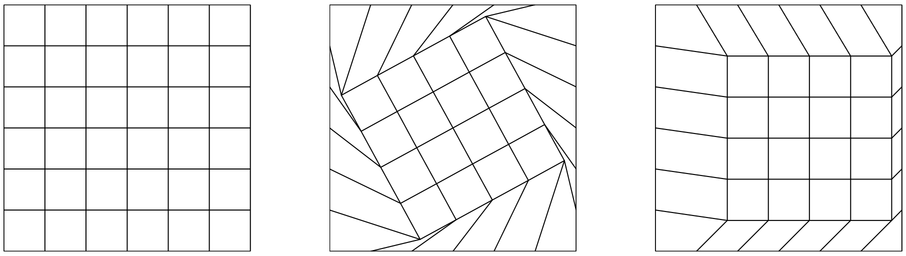

Figure 4.21: The objective of including the prior probability distribution in the registration model is to penalise shape changes of the source image. A rigid body rotation of a region of brain does nothing to the shape of that region. However, in order to rotate relative to the neighbouring regions then shears and zooms are necessary and these do change the shape of the image.

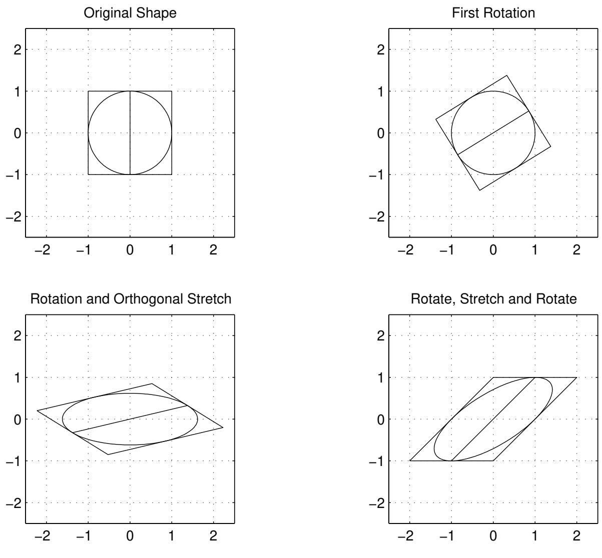

Figure 4.22: An affine transformation matrix that performs a shear must have singular values that are not equal to one. This figure shows a shear applied to a square and circle decomposed into the steps defined by singular value decomposition. The stretching that occurs between the upper right sub-figure, and that in the lower left, is what would be penalised. 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.

than others. Further material properties could be introduced by also modelling 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 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 modelling covariance between the strain tensors of neighbouring (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 some form of *restricted maximum likelihood estimation* (REML) (Harville, 1974) approach. REML algorithms are normally used for fitting weighted linear least squares models where the weights are also treated as unknown hyper-parameters. This is a similar situation to that described here, since there are unknown hyper-parameters (i.e., σ^2 and λ) describing the relative importance of the likelihoods versus the priors. Section 7.3 says more on this subject.

4.4.4 The Optimisation Algorithm

The method searches for the MAP solution, which is the single most probable realisation 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. Robust optimisation 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 (Joshi *et al.*, 1995; Thompson & Toga, 1996; Davatzikos, 1996), 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 Lester (1999).

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. The current method of computing σ^2 at each iteration is relatively ad hoc. A more rigorous approach would properly compute the degrees of freedom using a procedures similar to those described in Section 7.3.

At first sight, it would appear that optimising 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 regularisation 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 regularisation 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).

Chapter 5

Image Segmentation

Contents

5.1 Introduction

Healthy brain tissue can generally be classified into three broad tissue types on the basis of an MR image. These are grey matter (GM), white matter (WM) and cerebro-spinal fluid (CSF). This classification can be performed manually on a good quality T1 image, by simply selecting suitable image intensity ranges which encompass most of the voxel intensities of a particular tissue type. However, this manual selection of thresholds is highly subjective.

Some groups have used clustering algorithms to partition MR images into different tissue types, either using images acquired from a single MR sequence, or by combining information from two or more registered images acquired using different scanning sequences or echo times (eg. proton-density and T2-weighted). The approach adopted here is a modified version of one of these clustering algorithms. The clustering algorithm of choice is the maximum likelihood 'mixture model' algorithm (Hartigan, 1975), which has been extended to include spatial maps of prior belonging probabilities, and also a correction for image intensity non-uniformity that arises for many reasons in MR imaging. Because the tissue classification is based on voxel intensities, partitions derived without the correction can be confounded by these smooth intensity variations.

The model assumes that the MR image (or images) consists of a number of distinct tissue types (clusters) from which every voxel has been drawn. The intensities of voxels belonging to

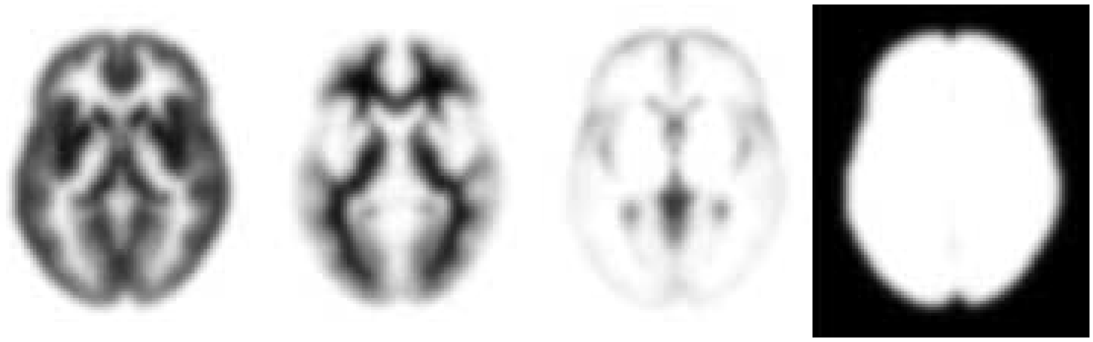

Figure 5.1: The *a priori* probability images of GM, WM, CSF and non-brain tissue. Values range between zero (white) and one (black).

each of these clusters conform to a multivariate normal distribution, which can be described by a mean vector, a covariance matrix and the number of voxels belonging to the distribution. In addition, the model has approximate knowledge of the spatial distributions of these clusters, in the form of prior probability images.

Prior to classifying an image, it is necessary to determine the spatial transformation that maps from each voxel in the image to its equivalent location in the *a priori* probability images, thus allowing simple 'on-the-fly' sampling of the probability images. The mapping is normally achieved by least squares matching with template images in the same stereotactic space as the prior probability images. This can be done using nonlinear warping, but the examples provided in this chapter were done using affine registration.

One of the greatest problems faced by tissue classification techniques is non-uniformity of the images intensity. Many groups have developed methods for correcting intensity non-uniformities, and the scheme developed here shares common features. There are two basic models describing image noise properties: multiplicative noise and additive noise. The multiplicative model describes images that have noise added before being modulated by the non-uniformity field (i.e., the standard deviation of the noise is multiplied by the modulating field), whereas the additive version models noise that is added after the modulation (standard deviation is constant). The current method uses a multiplicative noise model, which assumes that the errors originate from tissue variability rather than additive Gaussian noise from the scanner. Figure 5.2 illustrates the model used by the classification.

Non-uniformity correction methods all involve estimating a smooth function that modulates the image intensities. If the function is is not forced to be smooth, then it will begin to fit the higher frequency intensity variations due to different tissue types, rather than the low frequency intensity non-uniformity artifact. Spline (Yan & Karp, 1995; Sled *et al.*, 1998) and polynomial (Van Leemput *et al.*, 1999a; Van Leemput *et al.*, 1999b) basis functions are widely used for modelling the intensity variation. In these models, the higher frequency intensity variations are restricted by limiting the number of basis functions. In the current method, a Bayesian model is used, where it is assumed that the modulation field ($\mathbf{U}$) has been drawn from a population for which the *a priori* distribution is known, thus allowing high frequency variations of the modulation field to be penalised.

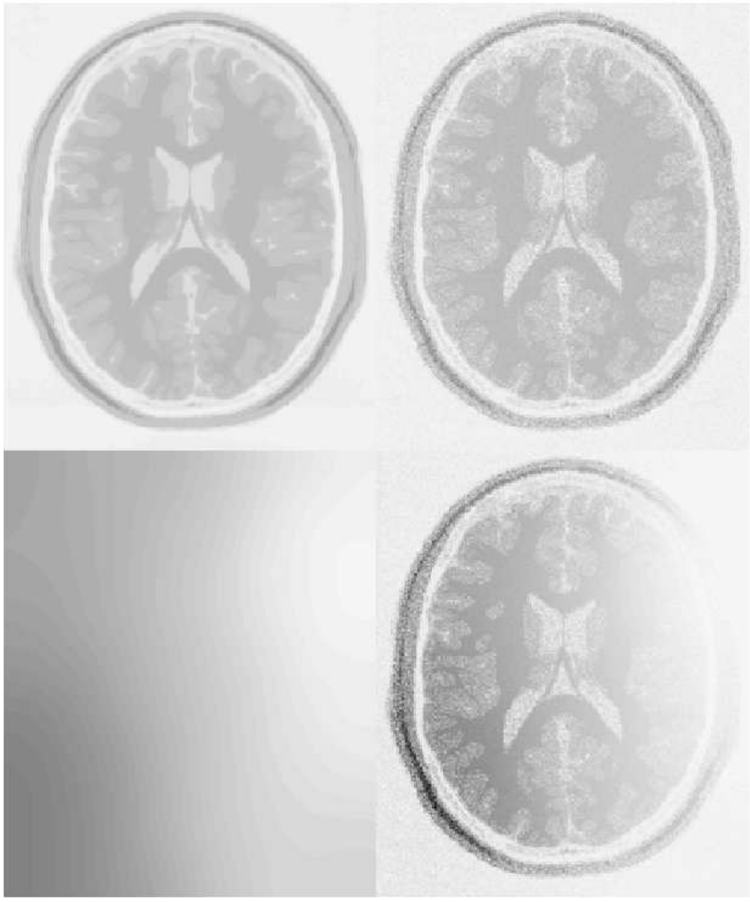

Figure 5.2: The MR images are modelled as a number of distinct clusters (top left), with different levels of Gaussian random noise added to each cluster (top right). The intensity modulation is assumed to be smoothly varying (bottom left), and is applied as a straightforward multiplication of the modulation field with the image (bottom right).

5.2 Methods

Although a three dimensional implementation of the tissue classification method has been developed, that can also be applied to multi-spectral images, the explanation of the algorithm will be simplified by describing its application to a single two dimensional image.

The tissue classification model makes a number of assumptions. The first is that each of the $I \times J$ voxels of the image ($\mathbf{F}$) has been drawn from a known number (K) of distinct tissue classes (clusters). The distribution of the voxel intensities within each class is normal (or multi-normal for multi-spectral images) and initially unknown. The distribution of voxel intensities within cluster k is described by the number of voxels within the cluster (h_k), the mean for that cluster (v_k), and the variance around that mean (c_k).

Because the images are matched to a particular stereotactic space, prior probabilities of the voxels belonging to the grey matter (GM), white matter (WM) and cerebro-spinal fluid (CSF) classes are known. This information is in the form of probability images – provided by the Montréal Neurological Institute (Evans *et al.*, 1992; Evans *et al.*, 1993; Evans *et al.*, 1994) as part of the ICBM, NIH P-20 project (Principal Investigator John Mazziotta). They were derived from scans of 152 young healthy subjects (66 female and 86 male, 129 right handed, 14 left handed and 9 unknown handedness, aged between 18 and 44, with a mean age of 25 and median age of 24) that were segmented into binary images of GM, WM and CSF, and all normalised into the same space using a 9 parameter (3 translations, 3 rotations and 3 orthogonal zooms) affine transformation. The images were originally classified using a neural network approach, and misclassified non-brain tissue was removed by a masking procedure. The probability images are the means of these binary images, and contain values in the range of zero to one. They represent the *a priori* probability of a voxel being either GM, WM or CSF after an image has been normalised to the same space using a 9 parameter affine transformation (see Figure 5.1). To increase the stability of the classification with respect to small registration errors, the prior probability images have been convolved with an 8mm full width at half maximum Gaussian smoothing kernel. The prior probability of a voxel at co-ordinate i, j belonging to cluster k is denoted by b_{ijk}[1].

The final assumption is that the intensity and noise associated with each voxel in the image has been modulated by multiplication with an unknown smooth scalar field.

There are many unknown parameters to be determined by the classification algorithm, and estimating any of these requires knowledge of the others. Estimating the parameters that describe a cluster (h_k, v_k and c_k) relies on knowing which voxels belong to the cluster, and also the form of the intensity modulating function. Estimating which voxels should be assigned to each cluster requires the cluster parameters to be defined, and also the modulation field. In turn, estimating the modulation field needs the cluster parameters and the belonging probabilities.

The problem requires an iterative algorithm (see Figure 5.3). It begins by assigning starting estimates for the various parameters. The starting estimate for the modulation field is typically uniformly one. Starting estimates for the belonging probabilities of the GM, WM and CSF partitions are based on the prior probability images. Since there are no prior probability maps for background and non-brain tissue clusters, they are estimated by subtracting the prior probabilities

[1]Note that ij subscripts are used for voxels rather than the single subscripts used in the previous chapters. This is to facilitate the explanation of how the modulation field is estimated for 2D images as described in Section 5.2.3.

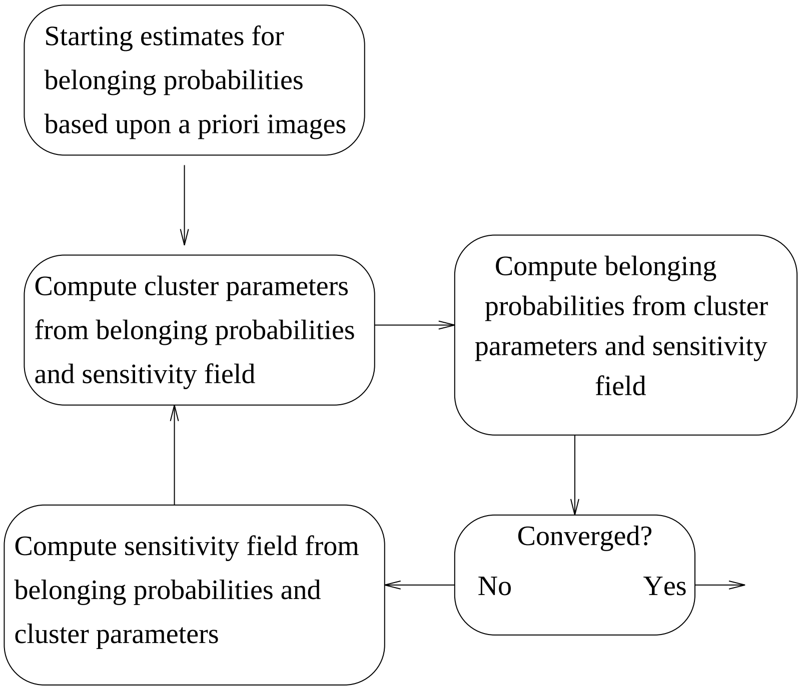

Figure 5.3: A flow diagram for the tissue classification.

for GM, WM and CSF from a map of all ones, and dividing the result equally between the remaining clusters [2].

Each iteration of the algorithm involves estimating the cluster parameters from the non-uniformity corrected image, assigning belonging probabilities based on the cluster parameters, checking for convergence, and re-estimating and applying the modulation function. With each iteration, the parameters describing the distributions move towards a better fit and the belonging probabilities ($\mathbf{P}$) change slightly to reflect the new distributions. This continues until a convergence criterion is satisfied. The parameters describing clusters with corresponding *a priori* probability images tend to converge more rapidly than the others. This may be partly due to the better starting estimates. The final values for the belonging probabilities are in the range of 0 to 1, although most values tend to stabilise very close to one of the two extremes. The algorithm is in fact an *expectation maximisation* (EM) approach, where the *E-step* is the computation of the belonging probabilities, and the *M-step* is the computation of the cluster and non-uniformity correction parameters. The individual steps involved in each iteration are now described in more detail.

5.2.1 Estimating the Cluster Parameters

This stage requires the original image to be intensity corrected according to the most recent estimate of the modulation function. Each voxel of the intensity corrected image is denoted by g_{ij}, and the current estimate of the probability of voxel i, j belonging to class k is denoted by p_{ijk}. The first step is to compute the number of voxels ($\mathbf{h}$) belonging to each of the K clusters

[2] Where identical prior probability maps are used for more than one cluster, the affected cluster parameters need to be modified slightly. This is typically done after the first iteration, by assigning different values for the means uniformly spaced between zero and the intensity of the white matter cluster.

as:

$$h_k = \sum_{i=1}^{I}\sum_{j=1}^{J} p_{ijk} \text{ over } k = 1..K. \tag{5.1}$$

Mean voxel intensities for each cluster ($\mathbf{v}$) are computed. This step effectively produces a weighted mean of the image voxels, where the weights are the current belonging probability estimates:

$$v_k = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} p_{ijk} g_{ij}}{h_k} \text{ over } k = 1..K. \tag{5.2}$$

Then the variance of each cluster ($\mathbf{c}$) is computed in a similar way to the mean:

$$c_k = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} p_{ijk} (g_{ij} - v_k)^2}{h_k} \text{ over } k = 1..K. \tag{5.3}$$

5.2.2 Assigning Belonging Probabilities

The next step is to to re-calculate the belonging probabilities. It uses the cluster parameters computed in the previous step, along with the prior probability images and the intensity modulated input image. Bayes rule is used to assign the probability of each voxel belonging to each cluster:

$$p_{ijk} = \frac{r_{ijk} s_{ijk}}{\sum_{l=1}^{K} r_{ijl} s_{ijl}} \text{ over } i = 1..I,\ j = 1..J \text{ and } k = 1..K. \tag{5.4}$$

where p_{ijk} is the *a posteriori* probability that voxel i, j belongs to cluster k given its intensity of g_{ij}, r_{ijk} is the likelihood of a voxel in cluster k having an intensity of g_{ik}, and s_{ijk} is the *a priori* probability of voxel i, j belonging in cluster k.

The likelihood function is obtained by evaluating the probability density functions for the clusters at each of the voxels:

$$r_{ijk} = (2\pi c_k)^{-1/2} \exp\left(\frac{-(g_{ij} - v_k)^2}{2c_k}\right) \text{ over } i = 1..I,\ j = 1..J \text{ and } k = 1..K. \tag{5.5}$$

The prior (s_{ijk}) is based on two factors: the number of voxels currently belonging to each cluster (h_k), and the prior probability images derived from a number of images (b_{ijk}). With no knowledge of the *a priori* spatial distribution of the clusters or the intensity of a voxel, then the *a priori* probability of any voxel belonging to a particular cluster is proportional to the number of voxels currently included in that cluster. However, with the additional data from the prior probability images, a better estimate for the priors can be obtained:

$$s_{ijk} = \frac{h_k b_{ijk}}{\sum_{l=1}^{I}\sum_{m=1}^{J} b_{lmk}} \text{ over } i = 1..I,\ j = 1..J \text{ and } k = 1..K. \tag{5.6}$$

Convergence is ascertained by following the log-likelihood function:

$$\sum_{i=1}^{I}\sum_{j=1}^{J} \log\left(\sum_{k=1}^{K} r_{ijk} s_{ijk}\right) \tag{5.7}$$

The algorithm is terminated when the change in log-likelihood from the previous iteration becomes negligible.

5.2.3 Estimating and Applying the Modulation Function

To reduce the number of parameters describing an intensity modulation field, it is modelled by a linear combination of low frequency discrete cosine transform (DCT) basis functions (see Section 3.2.3), which were chosen because there are no constraints at the boundary. The sensitivity correction field is computed by estimating the coefficients ($\mathbf{q}$) of the DCT basis functions that minimise the weighted sum of squared differences between the data and the model, and also a penalty function based on the smoothness of the modulation field. This can be expressed using matrix terminology as a regularised weighted least squares fitting:

$$\mathbf{q} = \left(\mathbf{A_1}^T\mathbf{A_1} + \mathbf{A_2}^T\mathbf{A_2}\cdots + \mathbf{C_0}^{-1}\right)^{-1}\left(\mathbf{A_1}^T\mathbf{b_1} + \mathbf{A_2}^T\mathbf{b_2}\cdots + \mathbf{C_0}^{-1}\mathbf{q_0}\right) \quad (5.8)$$

where $\mathbf{q_0}$ and $\mathbf{C_0}$ are the means and covariance matrices describing the *a priori* distribution of the coefficients. Matrix $\mathbf{A_k}$ and column vector $\mathbf{b_k}$ are constructed for cluster k from:

$$\mathbf{A_k} = \mathrm{diag}\left(\mathbf{p_k}c_k^{-1/2}\right)\mathrm{diag}\left(\mathbf{f}\right)\mathbf{D} \text{ and } \mathbf{b_k} = \mathbf{p_k}c_k^{-1/2}v_k \quad (5.9)$$

where matrix $\mathbf{D}$ contains the two or three dimensional DCT basis functions, and $\mathbf{p_k}$ refers to the belonging probabilities for the kth cluster considered as a column vector. Image $\mathbf{F}$ is similarly considered as a column vector $\mathbf{f}$. This model aims to find the smooth modulating function (described by its DCT coefficients), that will bring the voxel intensities of each cluster as close as possible (in the least squares sense) to the cluster means, where the vectors $\mathbf{p_k}c_k^{-1/2}$ are voxel by voxel weighting functions.

A two (or three) dimensional discrete cosine transform (DCT) is performed as a series of one dimensional transforms, which are simply multiplications with the DCT matrix. The elements of a matrix ($\mathbf{D}$) for computing the first M coefficients of the one dimensional DCT of a vector of length I is given by:

$$\begin{aligned} d_{i1} &= \frac{1}{\sqrt{I}} \quad i = 1..I \\ d_{im} &= \sqrt{\frac{2}{I}}\cos\left(\frac{\pi(2i-1)(m-1)}{2I}\right) \quad i = 1..I, m = 2..M \end{aligned} \quad (5.10)$$

The matrix notation for computing the first $M \times N$ coefficients of the two dimensional DCT of a modulation field $\mathbf{U}$ is $\mathbf{Q} = \mathbf{D_1}^T\mathbf{U}\mathbf{D_2}$, where the dimensions of the DCT matrices $\mathbf{D_1}$ and $\mathbf{D_2}$ are $I \times M$ and $J \times N$ respectively, and $\mathbf{U}$ is an $I \times J$ matrix. The approximate inverse DCT is computed by $\mathbf{U} \simeq \mathbf{D_1}\mathbf{Q}\mathbf{D_2}^T$. An alternative representation of the two dimensional DCT is obtained by reshaping the $I \times J$ matrix $\mathbf{U}$ so that it is a vector ($\mathbf{u}$). Element $i + (j-1) \times I$ of the vector is then equal to element i, j of the matrix. The two dimensional DCT can then be represented by $\mathbf{q} = \mathbf{D}^T\mathbf{u}$, where $\mathbf{D} = \mathbf{D_2} \otimes \mathbf{D_1}$ (the Kronecker tensor product of $\mathbf{D_2}$ and $\mathbf{D_1}$), and $\mathbf{u} \simeq \mathbf{D}\mathbf{q}$.

Computing $\mathbf{A_k}^T\mathbf{A_k}$ and $\mathbf{A_k}^T\mathbf{b_k}$ could be potentially very time consuming, especially when applied in three dimensions. However, this operation can be greatly speeded up using the properties of Kronecker tensor products (see Chapter 3). Figure 5.4 shows how this can be done in two dimensions using Matlab as a form of pseudo-code.

The Prior Probability Distribution

In Eqn. 5.8, $\mathbf{q_0}$ and $\mathbf{C_0}$ represent a multi-normal *a priori* probability distribution for the basis function coefficients. The mean of the *a priori* distribution is such that it would generate a field

```
alpha_k = zeros(M*N,M*N);
beta_k  = zeros(M*N,1);
weight  = P_k*(c_k^(-0.5));
img1    = weight.*F;
img2    = weight*v_k;
for j = 1:J,
   tmp     = (img1(:,j)*ones(1,M)).*D1;
   alpha_k = alpha_k + kron(D2(j,:)'*D2(j,:), tmp'*tmp);
   beta_k  = beta_k  + kron(D2(j,:)', tmp'*img2(:,j));
end;
```

Figure 5.4: The algorithm for computing $\mathbf{A_k}^T\mathbf{A_k}$ (*alpha_k*) and $\mathbf{A_k}^T\mathbf{b_k}$ (*beta_k*) in two dimensions using Matlab as a pseudocode (c.f., Figure 3.5). The symbol "$*$" refers to matrix multiplication, whereas "$.*$" refers to element by element multiplication. " $'$ " refers to a matrix transpose and "$\hat{}$" to a power. The jth row of matrix "$D2$" is denoted by "$D2(j,:)$", and the jth column of matrix "$img2$" is denoted by "$img2(:,j)$". The functions "$zeros(a,b)$" and "$ones(a,b)$" would produce matrices of size a$\times$b of either all zero or all one. A Kronecker tensor product of two matrices is represented by the "$kron$" function. Matrix "F" is the $I\times J$ non-uniformity corrected image. Matrix "*P_k*" is the $I\times J$ current estimate of the probabilities of the voxels belonging to cluster k. Matrices "$D1$" and "$D2$" contain the DCT basis functions, and have dimensions $I\times M$ and $J\times N$. "*v_k*" and "*c_k*" are scalers, and refer to the mean and variance of the kth cluster.

that is uniformly one. For this, all the elements of the mean vector are set to zero, apart from the first element that is set to $\sqrt{IJ}$.

The covariance matrix $\mathbf{C_0}$ is such that $(\mathbf{q}-\mathbf{q_0})^T\mathbf{C_0}^{-1}(\mathbf{q}-\mathbf{q_0})$ produces an "energy" term that penalises modulation fields that would be unlikely *a priori*. There are many possible forms for this penalty function (see Section 3.2.4). Some widely used simple penalty functions include the "membrane energy" and the "bending energy", which (in three dimensions) have the forms $h=\sum_i\sum_{j=1}^3\lambda\left(\frac{\partial u(\mathbf{x}_i)}{\partial x_{ji}}\right)^2$ and $h=\sum_i\sum_{j=1}^3\sum_{k=1}^3\lambda\left(\frac{\partial^2 u(\mathbf{x}_i)}{\partial x_{ji}\partial x_{ki}}\right)^2$ respectively. In these formulae, $\frac{\partial u(\mathbf{x}_i)}{\partial x_{ji}}$ is the gradient of the modulating function at the ith voxel in the jth orthogonal direction, and λ is a user assigned constant. However, for the purpose of modulating the images, a smoother cost function is used that is based on the squares of the third derivatives (third order regularisation):

$$h=\sum_i\sum_{j=1}^3\sum_{k=1}^3\sum_{l=1}^3\lambda\left(\frac{\partial^3 u(\mathbf{x}_i)}{\partial x_{ji}\partial x_{ki}\partial x_{li}}\right)^2 \tag{5.11}$$

This model was chosen because it produces slowly varying modulation fields that can represent the variety of non-uniformity effects that are likely to be encountered in MR images (see Figure 5.5). In two dimensions it can be computed from:

$$\begin{aligned}\mathbf{C_0}^{-1} = \;& \lambda\left(\dddot{\mathbf{D}}_\mathbf{2}^T\dddot{\mathbf{D}}_\mathbf{2}\right)\otimes\left(\mathbf{D_1}^T\mathbf{D_1}\right)+3\lambda\left(\ddot{\mathbf{D}}_\mathbf{2}^T\ddot{\mathbf{D}}_\mathbf{2}\right)\otimes\left(\dot{\mathbf{D}}_\mathbf{1}^T\dot{\mathbf{D}}_\mathbf{1}\right)+\\ & 3\lambda\left(\dot{\mathbf{D}}_\mathbf{2}^T\dot{\mathbf{D}}_\mathbf{2}\right)\otimes\left(\ddot{\mathbf{D}}_\mathbf{1}^T\ddot{\mathbf{D}}_\mathbf{1}\right)+\lambda\left(\mathbf{D_2}^T\mathbf{D_2}\right)\otimes\left(\dddot{\mathbf{D}}_\mathbf{1}^T\dddot{\mathbf{D}}_\mathbf{1}\right)\end{aligned} \tag{5.12}$$

where the notation $\dot{\mathbf{D}}_\mathbf{1}$, $\ddot{\mathbf{D}}_\mathbf{1}$ and $\dddot{\mathbf{D}}_\mathbf{1}$ refer to the first, second and third derivatives (by differentiating Eqn. 5.10 with respect to i) of $\mathbf{D_1}$, and λ is a user specified hyper-parameter.

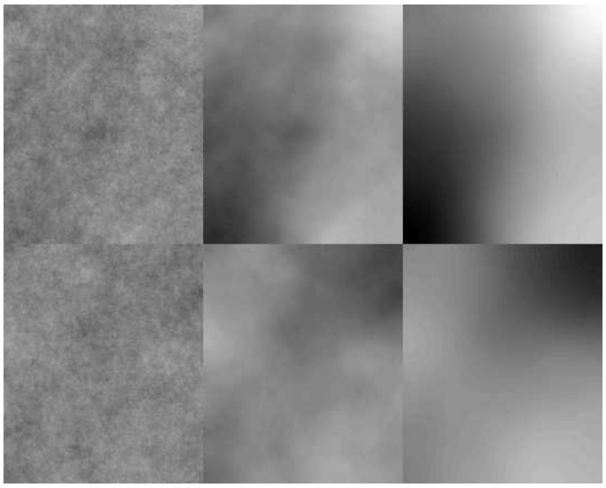

Figure 5.5: Randomly generated modulation fields generated using the membrane energy cost function (left), the bending energy cost function (centre) and the squares of the third derivatives (right). These can be referred to as 1st, 2nd and 3rd order regularisation.

Applying the Correction

Finally, once the coefficients have been estimated, then the modulation field $\mathbf{U}$ can be computed from the estimated coefficients ($\mathbf{Q}$) and the basis functions ($\mathbf{D_1}$ and $\mathbf{D_2}$).

$$u_{ij} = \sum_{n=1}^{N} \sum_{m=1}^{M} d_{2jn} q_{mn} d_{1im} \text{ over } i = 1..I \text{ and } j = 1..J. \tag{5.13}$$

The new estimate for the sensitivity corrected images are then obtained by a simple element by element multiplication with the modulation field.

$$g_{ij} = f_{ij} u_{ij} \text{ over } i = 1..I \text{ and } j = 1..J. \tag{5.14}$$

5.3 Evaluation

In order to provide a qualitative example of the classification, Figure 5.6 shows a single sagittal slice through six randomly chosen T1-weighted images. The initial registration to the prior probability images was via the 12-parameter affine transformation described in Section 3.2.2. The images were automatically classified using the method described here, and contours of extracted grey and white matter are shown superimposed on the images.

Tissue classification was also evaluated using a number of simulated images ($181 \times 217 \times 181$ voxels of $1 \times 1 \times 1$ mm) of the same brain generated by the BrainWeb simulator (Cocosco *et al.*, 1997; Kwan *et al.*, 1996; Collins *et al.*, 1998) with 3% noise (relative to the brightest tissue in the images). The contrasts of the images simulated T1-weighted, T2-weighted and proton density (PD) images (all with 1.5 Tesla field strength), and they were classified individually and in a multi-spectral manner [3]. The T1-weighted image was simulated as a spoiled FLASH sequence,

[3]Note that different modulation fields that account for non-uniformity were assumed for each image of the multi-spectral data-sets.

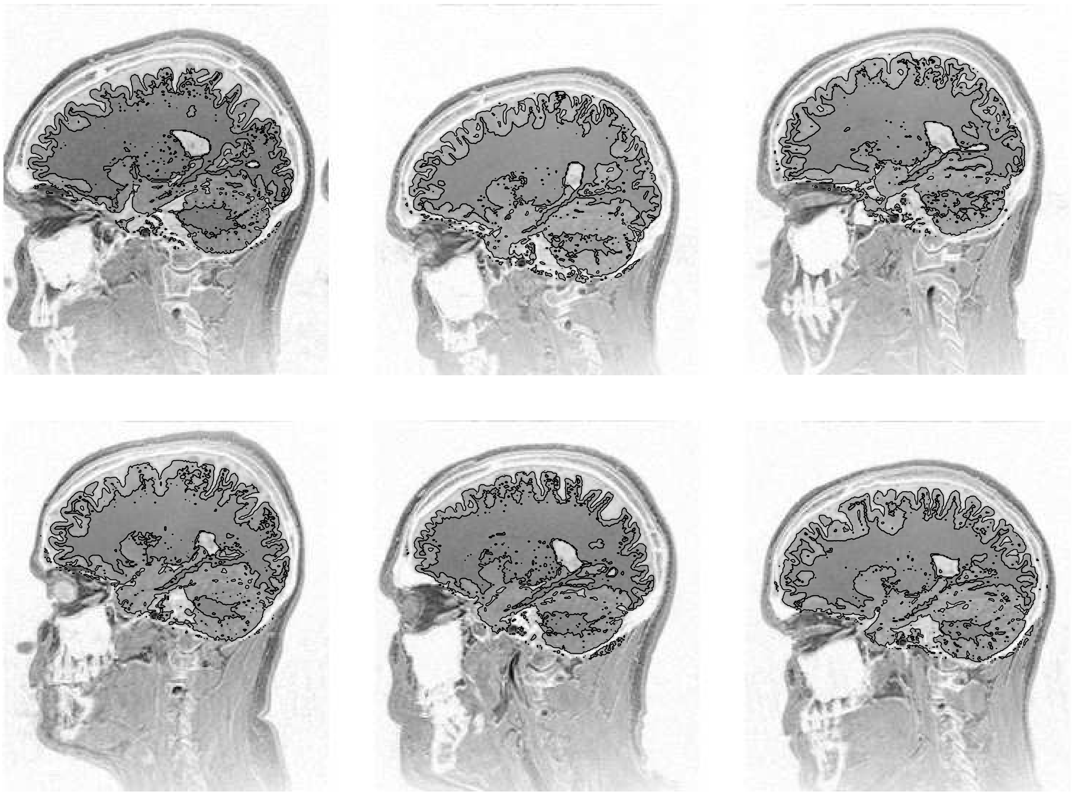

Figure 5.6: A single sagittal slice through six T1-weighted images (2 Tesla scanner, with an MPRAGE sequence, 12^o tip angle, 9.7ms repeat time, 4ms echo time and 0.6ms inversion time). Contours of extracted grey and white matter are shown superimposed on the images.

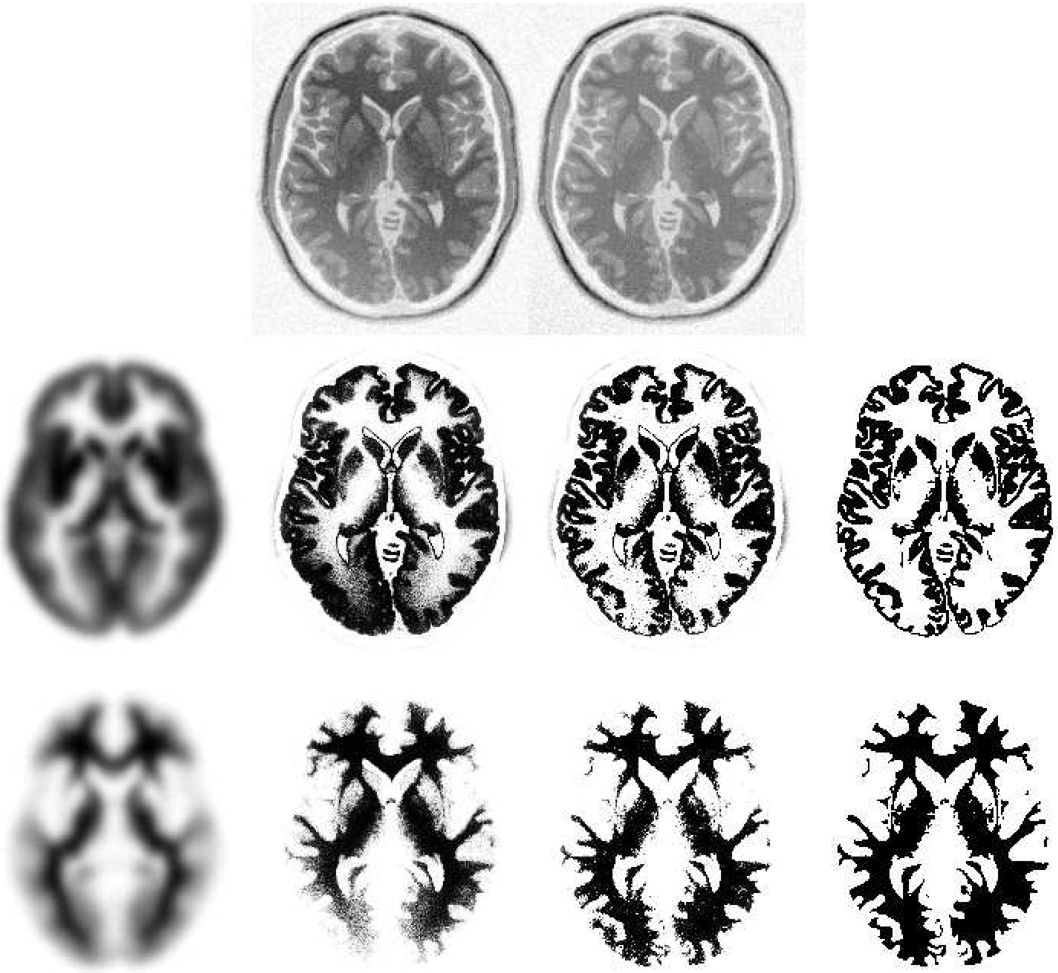

Figure 5.7: The classification of the simulated BrainWeb image. The top row shows the original simulated T1-weighted MR image with 100% non-uniformity, and the non-uniformity corrected version. From left to right, the middle row shows the *a priori* spatial distribution of grey matter used for the classification, grey matter extracted without non-uniformity correction, grey matter extracted with non-uniformity correction and the "true" distribution of grey matter (from which the simulated images were derived). The bottom row is the same as the middle, except that it shows white matter rather than grey. Without non-uniformity correction, the intensity variation causes some of the white matter in posterior areas to be classified as grey. This was also very apparent in the cerebellum because of the intensity variation in the inferior-superior direction.

with a 30^o flip angle, 18ms repeat time, 10ms echo time. The T2 and PD images were simulated by a dual echo, spin echo technique, with 90^o flip angle, 3300ms repeat time and echo times of 35 and 120ms. Three different levels of image non-uniformity were used: 0%RF - which assumes that there is no intensity variation artifact, 40%RF - that assumes a fairly typical amount of non-uniformity, and 100%RF which is more non-uniformity than would normally be expected. The simulated images were classified, both with and without sensitivity correction. Three partitions were considered in the evaluation: grey matter, white matter and other (not grey or white), and each voxel was assigned to the most likely partition. Because the data from which the simulated images were derived was available, it was possible to compare the classified images with ground truth images of grey and white matter using the κ statistic (a measure of inter-rater agreement):

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{5.15}$$

where p_o is the observed proportion of agreement, and p_e is the expected proportion of agreements by chance. If there are N observations in K categories, the observed proportional agreement is:

$$p_o = \sum_{k=1}^{K} f_{kk}/N \tag{5.16}$$

where f_{kk} is the number of agreements for the kth category. The expected proportion of agreements is given by:

$$p_e = \sum_{k=1}^{K} r_k c_k / N^2 \tag{5.17}$$

where r_k and c_k are the total number of voxels in the kth class for both the "true" and estimated partitions.

The classification of a single plane of the simulated T1 weighted BrainWeb image with the 100% non-uniformity is illustrated in Figure 5.7. It should be noted that no pre-processing to remove scalp or other non-brain tissue was performed on the image. In theory, the tissue classification method should produce slightly better results if this non-brain tissue is excluded from the computations. As the algorithm stands, a small amount of non-brain tissue remains in the grey matter partition, which has arisen from voxels that lie close to grey matter and have similar intensities.

The resulting κ statistics from classifying the different simulated images are shown in table 5.1. These results show that the non-uniformity correction made little difference to the tissue classification of images without any non-uniformity artifact. For images containing non-uniformity artifact, classifications using the correction were of about the same quality as the classifications without the artifact, and very much better than classifications without the correction. These results show the same general trends as the results presented by Van Leemput *et. al.* (1999b). The κ statistics by themselves do not really provide much intuition about the quality of the segmentation. As a guide, the segmentations shown in Figure 5.7 produced κ values of 0.85 and 0.94.

A by-product of the classification is the estimation of an intensity non-uniformity field. Figure 5.8 shows a comparison of the intensity non-uniformity present in a simulated T1 image with 100% non-uniformity (created by dividing noisless simulated images with 100% non-uniformity and no non-uniformity) with that recovered by the classification method. A scatter-plot of "true" versus recovered non-uniformity shows a straight line, suggesting that the accuracy of the estimated non-uniformity is very good.

5.3.1 Stability With Respect to Misregistration with the Prior Probability Images

In order for the Bayesian classification to work properly, an image volume must be in register with a set of *a priori* probability images used to instate the priors. Here the effects of misregistration on the accuracy of segmentation are examined, by artificially translating (in the left-right direction) the prior probability images by different distances prior to segmenting the image. The 1mm slice thickness, 40% non-uniformity, and 3% noise simulated T1-weighted image (described above) was used for the classification, which included the non-uniformity correction. The κ statistic was computed with respect to the "true" grey and white matter for the different translations, and the results are plotted in Figure 5.9.

In addition to illustrating the effect of misregistration, the figure also gives an indication of how far a brain can deviate from the normal population of brains (that constitute the prior probability images) in order for it to be segmented adequately. Clearly, if the brain cannot be well registered with the probability images, then the segmentation will not be as accurate. This fact also has implications for severely abnormal brains, as they are more difficult to register with images that represent the prior probabilities of voxels belonging to different classes. Segmenting such abnormal brains can be a problem for the algorithm, as the prior probability images are based on normal healthy brains. The profile in Figure 5.9 depends on the smoothness or resolution of the prior probability images. By not smoothing the prior probability images, the segmentation would be optimal for normal, young and healthy brains. However, these images may need to be smoother in order to encompass more variability when patient data are to be analysed.

5.4 Discussion

The segmentation method has been found to be robust and accurate for high quality T1 weighted images, but is not beyond improvement. Currently, each voxel is assigned a probability of belonging to a particular tissue class based only on its intensity and information from the prior probability images. There is a great deal of other knowledge that could be incorporated into the classification. For example, if all a voxel's neighbours are grey matter, then there is a high probability that it should also be grey matter. Other researchers have successfully used Markov random field models to include this information in a tissue classification model (Yan & Karp, 1995; Vandermeulen *et al.*, 1996; Van Leemput *et al.*, 1999b). Another very simple prior, that can be incorporated, is the relative intensity of the different tissue types. For example, when segmenting a T1 weighted image, it is known that the white matter should have a higher intensity than the grey matter, which in turn should be more intense than the CSF. When computing the means for each cluster, this prior information could sensibly be used to bias the estimates.

In order to function properly, the classification method requires good contrast between the different tissue types. However, many central grey matter structures have image intensities that

	Single image			Multi-spectral			
	T1	T2	PD	T2/PD	T1/T2	T1/PD	T1/T2/PD
0%RF - uncorreced	0.95	0.90	0.90	0.93	0.94	0.96	0.94
0%RF - corrected	0.95	0.90	0.90	0.93	0.94	0.96	0.95
40%RF - uncorreced	0.92	0.88	0.79	0.90	0.93	0.95	0.94
40%RF - corrected	0.95	0.90	0.90	0.93	0.94	0.96	0.94
100%RF - uncorrected	0.85	0.85	0.67	0.87	0.92	0.94	0.93
100%RF - corrected	0.94	0.90	0.88	0.92	0.93	0.95	0.94

Table 5.1: This table shows the different κ statistics that were computed after classifying the simulated images.

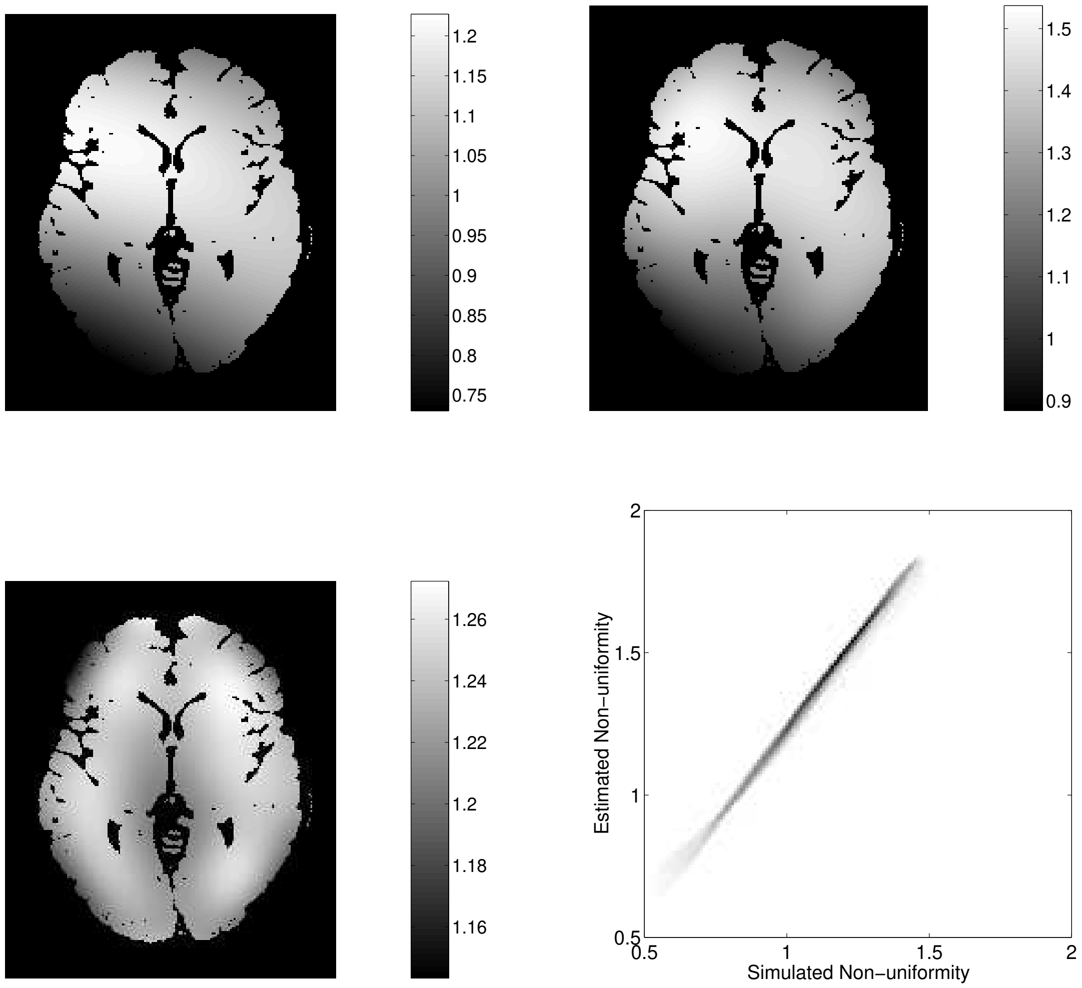

Figure 5.8: Top Left: The true intensity non-uniformity field of the simulated T1 image. Top Right: The non-uniformity recovered by the classification algorithm. Below Left: The recovered divided by the true non-uniformity. Below Right: A scatter-plot of true intensity non-uniformity versus recovered non-uniformity, derived from voxels throughout the whole volume classified as either white or grey matter. Note that the plot is a straight line, but that its gradient is not one because it is not possible to recover the absolute scaling of the field.

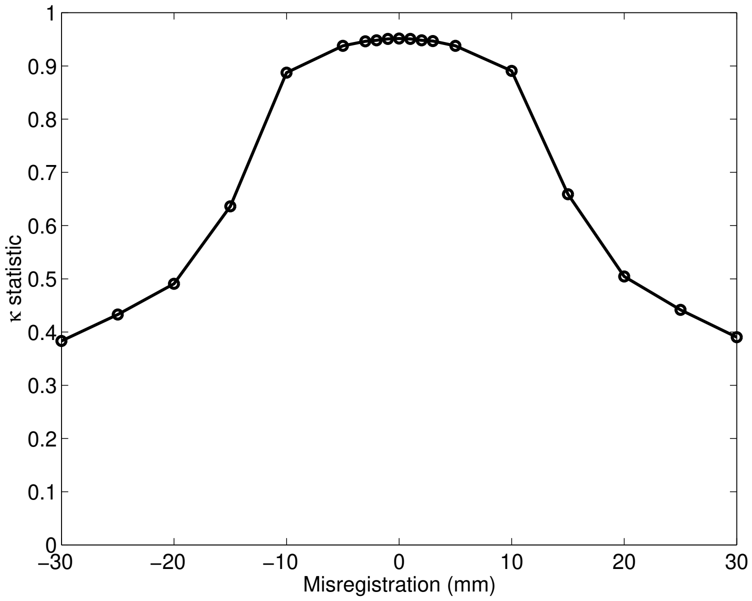

Figure 5.9: Segmentation accuracy with respect to misregistration with the *a priori* images.

are almost indistinguishable from that of white matter, so the tissue classification is not always very accurate in these regions. Another related problem is that of partial volume. Because the model assumes that all voxels contain only one tissue type, the voxels that contain a mixture of tissues may not be modelled correctly. In particular, those voxels at the interface between white matter and ventricles will often appear as grey matter. This can be seen to a small extent in Figures 5.6 and 5.7. Each voxel is assumed to be of only one tissue type, and not a combination of different tissues, so the model's assumptions are violated when voxels contain signal from more than one tissue type. This problem is greatest when the voxel dimensions are large, or if the images have been smoothed, and is illustrated using simulated data in Figure 5.10. The effect of partial volume is that it causes the distributions of the intensities to deviate from normal. Some authors have developed more complex models to describe the intensity distributions of the classes. For example, each non-Gaussian distribution could be modelled by a mixture of Gaussians. An alternative approach is to use logistic discriminant analysis to parameterise a function that best distinguishes between the classes (Bullmore *et al.*, 1995). In this approach, a training set consisting of voxels belonging to each class is first identified. From this training set, the discriminant functions are optimised so that they best distinguish between the classes, before being applied to the whole image. In theory, it may be possible to extend the method so that it uses an EM approach similar to that described in this chapter.

The incorporation of prior probability images into the clustering algorithm produces a much more robust solution when segmenting brains from a similar population to those from which the spatial priors were derived (refer to Section 5.3.1). However, for severely abnormal brains, the priors may not be representative, and this can lead to problems. For example, if a subject has very large ventricles, then CSF may appear where the priors suggest that tissue should always be WM. These CSF voxels are forced to be misclassified as WM, and the intensities of these voxels are incorporated into the computation of the WM means and variances. This results in the WM being characterised by a very broad distribution, so the algorithm is unable to distinguish it

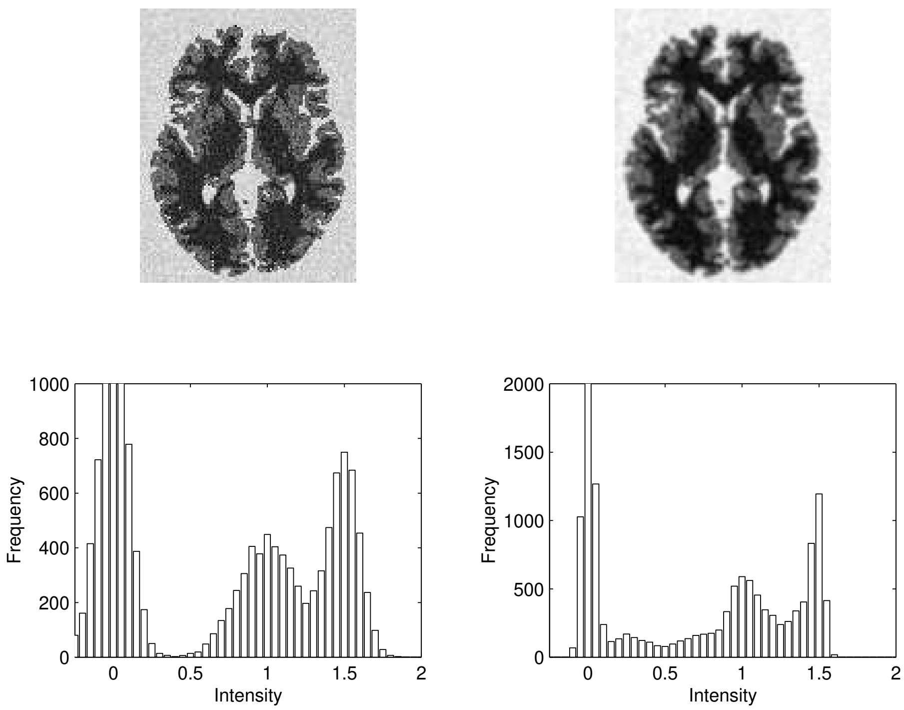

Figure 5.10: Simulated data showing the effects of partial volume on the intensity histograms. On the upper left is a simulated image consisting of three distinct clusters. The intensity histogram of this image is shown on the lower left and consists of three Gaussian distributions. The image at the top right is the simulated image after a small amount of smoothing. The corresponding intensity histogram no longer shows three distinct Gaussian distributions.

from any other tissue. For young healthy subjects, the classification is very good, but caution is required when the method is used for severely pathological brains.

MR images are normally reconstructed by taking the modulus of complex images. Normally distributed complex values are not normally distributed when the magnitude is taken. Instead, they obey a Rician distribution. This means that any clusters representing the background are not well modelled by a single Gaussian, but it makes very little difference for most of the other clusters.

The segmentation is normally run on unprocessed brain images, where non-brain tissue is not first removed. This results in a small amount of non-brain tissue being classified as brain. However, by using morphological operations on the extracted GM and WM segments, it is possible to remove most of this extra tissue. The procedure begins by eroding the extracted WM image, so that any small specs of misclassified WM are removed. This is followed by conditionally dilating the eroded WM, such that dilation can only occur where GM and WM were present in the original extracted segments. Although some non-brain structures (such as part of the sagittal sinus) may remain after this processing, most non-brain tissue is removed. Figure 5.11 shows how the GM and WM partitions can be cleaned up using this procedure, and surface rendered images of brains automatically extracted this way are shown in Figure 4.18. When applied to the segmented simulated T1 weighted image with 3% noise and no nonuniformity artifact, this cleanup procedure increased κ by 0.001.

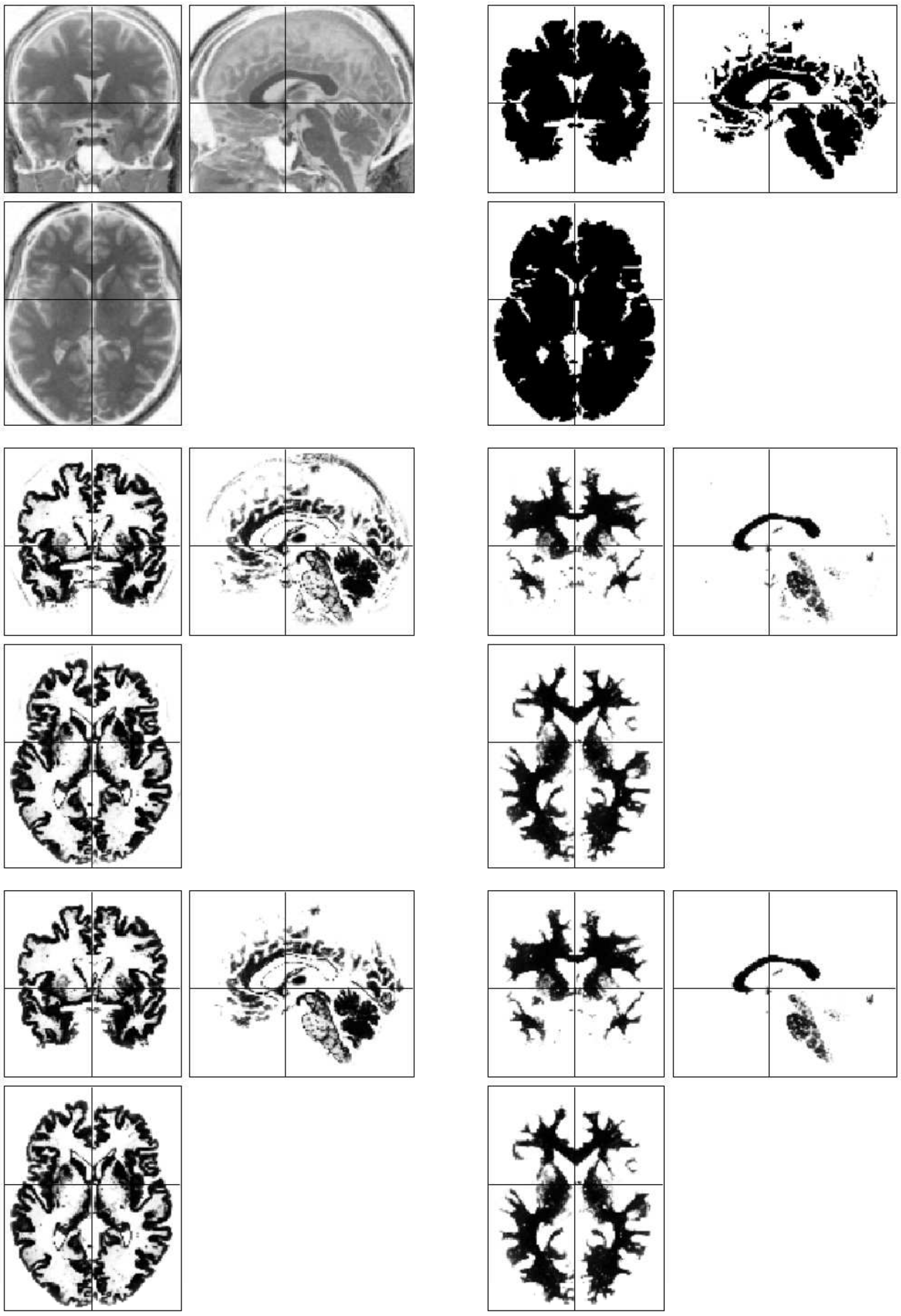

Figure 5.11: Example of automatically cleaned up segmented images. The top row shows the original T1 weighted MR image, next to an automatically generated mask of brain derived from the initial grey and white matter partitions. The second row shows the initial extracted grey and white matter. The bottom row shows the grey and white matter partitions after cleaning up by multiplying with the brain mask.

Chapter 6

Morphometry

Contents

6.1 Introduction

Studies of brain morphometry have been carried out by many researchers on a number of different populations, including patients with schizophrenia, autism, dyslexia and Turner's syndrome. Often, the morphometric measurements used in these studies have been obtained from brain regions that can be clearly defined, resulting in a wealth of findings pertaining to these particular measurements. The measures are typically volumes of unambiguous structures such as the hippocampi or the ventricles. However, there are a number of morphometric features that may be more difficult to quantify by inspection, meaning that many observable structural differences may be overlooked. This chapter describes morphometric approaches that are not biased to one particular structure or tissue and give even-handed and comprehensive assessments of anatomical differences throughout the brain.

Morphometric methods have a number of different aims. They can be used for localising significant structural differences among populations, or for showing that overall brain structure is related to some effect of interest. When testing the overall brain structure, multivariate statistical methods are used to analyse groups of parameters for the whole brain (e.g., the deformation-based morphometry described in this chapter). The result of the forms of morphometry that localise structural differences would typically be a statistical parametric map (Friston *et al.*, 1995d) of regional differences. Statistical parametric maps (SPMs) can be derived from uni-variate data where there is a single variable at each voxel (e.g., the voxel-based morphometry method described here), or from multi-variate data, where there are several different variables at each voxel (e.g., tensor-based morphometry).

Another use for morphometric methods is for characterising essential differences, or for producing some form of classification. Linear methods such as canonical correlation analysis, or nonlinear classification methods can be used for these purposes. This chapter will be restricted to simple linear classification methods, which model data as multivariate normal distributions. Nonlinear classification methods can assume more complex distributions for the data, but they tend to be much more computationally expensive.

Statistical tests generally involve disproving a null hypothesis with a particular level of confidence. In morphometry, the null hypothesis is usually that there are no significant structural differences among a number of populations, or due to particular covariates, such as age. The objective of the tests is to demonstrate improbability of the null hypothesis. For example, if p values of less than 0.01 are deemed to be significant, then false positive results would be expected only about once out of a hundred tests.

SPMs of univariate statistical measures often allow relatively simple questions to be addressed, such as where is there significantly more of a particular measure, that happens to correlate with a particular effect of interest. Standard parametric statistical procedures (t-tests and F-tests) can be used to test the hypotheses within the framework of the general linear model (LM), whereby a vector of observations is modelled by a linear combination of user specified regressors (Friston *et al.*, 1995d). The LM is a flexible framework that allows many different tests to be applied, ranging from group comparisons and identifying differences that are related to specified covariates such as disease severity or age, to complex interactions between different effects of interest.

Performing comparisons at each voxel results in many statistical tests being performed. Without any correction, the number of false-positive results would be proportional to the number of independent tests. A Bonferroni correction would be applied if the tests were independent, but this is not normally the case because of the inherent spatial smoothness of the data. In practice, the effective number of independent statistical tests is determined using Gaussian Random Field (GRF) theory (Friston *et al.*, 1995b; Worsley *et al.*, 1996). By using GRF theory, a correction for multiple dependent comparisons can be made to produce the appropriate rate of false-positive results.

SPMs can also be obtained from the results of voxel-wise multi-variate tests. Instead of one image per subject, multivariate tests effectively involve two or more images, and use a statistic such as Wilk's Λ. Following the tests, similar corrections based on GRF theory can be applied as in the univariate case. GRF theory has not yet been worked out for Wilk's Λ fields, so approximations are made that involve transforming the resulting Wilk's Λ fields to random fields of other statistics, such as χ^2 or F fields. Subsequent processing assumes that the transformed fields have the same properties as true χ^2 or F fields.

There now follows a description of how multivariate statistical tests can be performed using Multivariate Analysis of Covariance. These can be used to produce SPMs that localise structural differences, or can be applied to sets of parameters describing whole brains. This is followed by a brief description of how differences between groups can be characterised by multivariate methods. The remainder of the chapter will describe three different morphometric methods: voxel-based, deformation-based and tensor-based morphometry.

6.1.1 Multivariate Analysis of Covariance

A multivariate analysis of covariance (MANCOVA) assumes that there are several dependent variables for each observation, where an observation refers to a collection of data for a subject. This data can be represented by an $M \times I$ matrix $\mathbf{X}$, where M is the number of subjects in the analysis, and I is the number of variables for each subject.

The columns of $\mathbf{X}$ are modelled by linear combination of basis functions. Some of these basis functions represent effects that are not considered interesting in the study, but may still be significant. For example, linear age effects may confound a study of handedness. If a left-handed group is not perfectly age matched with a right-handed group, then differences due to age could be attributed to handedness differences. In addition, the inclusion of confounding effects (such as age) in a model can result in a better fit to the data, and possibly make the test more sensitive to the effects of interest.

Confounding effects are modelled by an $M \times K$ design matrix $\mathbf{G}$. Each column of $\mathbf{G}$ can be a vector of covariates (e.g., the age of each subject), or alternatively can be arranged in blocks (e.g., there may be a column containing ones for left-handed subjects, and zeros for right-handed). In almost all cases, a column of ones is included in order to model a mean effect. First of all, any variance in the data that could be attributed to the confounds is removed by:

$$\mathbf{X_a} = \mathbf{X} - \mathbf{G}\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{X} \tag{6.1}$$

Similarly, the effects of interest are modelled by an $M \times J$ design matrix $\mathbf{Y}$. The columns of $\mathbf{Y}$ can represent group memberships, where elements contain a one for subjects who are in a particular group, or a zero if they do not. Alternatively, the columns may contain covariates of interest such as disease severity for each subject. The columns in this design matrix are orthogonalised with respect to matrix $\mathbf{G}$:

$$\mathbf{Y_a} = \mathbf{Y} - \mathbf{G}\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{Y} \tag{6.2}$$

A MANCOVA involves assessing how the predictability of the observations change when the effects of interest are discounted. This is based on the distributions of the residuals, which are assumed to be multinormal. The statistic is related to the determinants of the covariance matrices describing these distributions. In practice, the residual sum of squares and products (SSP) matrix ($\mathbf{W}$), is compared to the SSP matrix of the fitted effects ($\mathbf{B}$). These matrices are obtained by:

$$\mathbf{T} = \quad \mathbf{Y_a}(\mathbf{Y_a}^T\mathbf{Y_a})^{-1}\mathbf{Y_a}^T\mathbf{X_a}$$

$$\begin{aligned} \mathbf{B} = &\quad \mathbf{T}^T\mathbf{T} \\ \mathbf{W} = &\quad (\mathbf{X_a} - \mathbf{T})^T (\mathbf{X_a} - \mathbf{T}) \end{aligned} \tag{6.3}$$

The statistic is called *Wilk's Lambda* (Λ), and is based upon the ratios of the determinants:

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|} \tag{6.4}$$

The Wilk's Lambda statistic can range between zero and one, where a value of one suggests no relationship between the effects of interest and the data, and a value of zero indicates a perfect relationship. This statistic is transformed to a χ^2 statistic (with IJ degrees of freedom under the null hypothesis) using the approximation of Bartlett [1]:

$$\chi^2 \approx -(M - K - (I + J + 1)/2) log_e(\Lambda) \tag{6.5}$$

Finally, the cumulative χ^2 distribution function is used to make inferences about whether the null hypothesis (that there is no difference between the distributions) can be rejected.

This multivariate approach fails when the number of variables approaches the number of subjects ($M - K - (I + J + 1)/2$ approaches zero, or becomes negative). In many situations, it is necessary to regularise the problem by reducing the number of variables with respect to the number of subjects. One way of doing this involves using singular value decomposition to decompose the original data matrix $\mathbf{X}$ into unitary matrices $\mathbf{U}$ and $\mathbf{V}$, and diagonal matrix $\mathbf{S}$, such that $\mathbf{X} = \mathbf{USV}^T$. The diagonal elements of $\mathbf{S}$ are arranged in decreasing importance, so it is possible to reconstruct an approximation of $\mathbf{X}$ using only the first L diagonal elements of $\mathbf{S}$ and the first L columns of $\mathbf{U}$ and $\mathbf{V}$, such that $\mathbf{X} \simeq \mathbf{U}^*\mathbf{S}^*\mathbf{V}^{*T}$. The MANCOVA would be performed using $\mathbf{U}^*$ (the first L columns of $\mathbf{U}$).

6.1.2 Canonical Correlation Analysis

Canonical correlation analysis (CCA) is used to measure the strength of association between two sets of variables. In this case, the variables are $\mathbf{X_a}$ and $\mathbf{Y_a}$, which are the data and design matrix from Section 6.1.1 after having been orthogonalised with respect to a set of confounding effects. The first canonical variate pair is the linear combination of columns of $\mathbf{X_a}$ and the linear combination of columns of $\mathbf{Y_a}$, that has the maximum correlation. The second canonical variate pair consists of linear combinations that maximise the correlation subject to the constraint that they are orthogonal to the first pair of canonical variables. Similarly, all subsequent pairs maximise the correlations and are orthogonal to all the previous pairs.

The weights used to determine the linear combinations are derived from the unitary matrices ($\mathbf{U}$ and $\mathbf{V}$) obtained by singular value decomposition:

$$\mathbf{USV}^T = \left(\mathbf{X_a}^T\mathbf{X_a}\right)^{-\frac{1}{2}} \left(\mathbf{X_a}^T\mathbf{Y_a}\right) \left(\mathbf{Y_a}^T\mathbf{Y_a}\right)^{-\frac{1}{2}} \tag{6.6}$$

Then the weights ($\mathbf{A}$ and $\mathbf{B}$) are derived by:

$$\mathbf{A} = \left(\mathbf{X_a}^T\mathbf{X_a}\right)^{-\frac{1}{2}} \mathbf{U} \text{ and } \mathbf{B} = \left(\mathbf{Y_a}^T\mathbf{Y_a}\right)^{-\frac{1}{2}} \mathbf{V} \tag{6.7}$$

The canonical variate pairs are obtained from the columns of $\mathbf{X_a A}$ and $\mathbf{Y_a B}$.

[1] This is true providing the matrices do not contain any redundant columns, otherwise matrix pseudo-inverses are required in the computations, and the numbers of columns replaced by the matrix ranks.

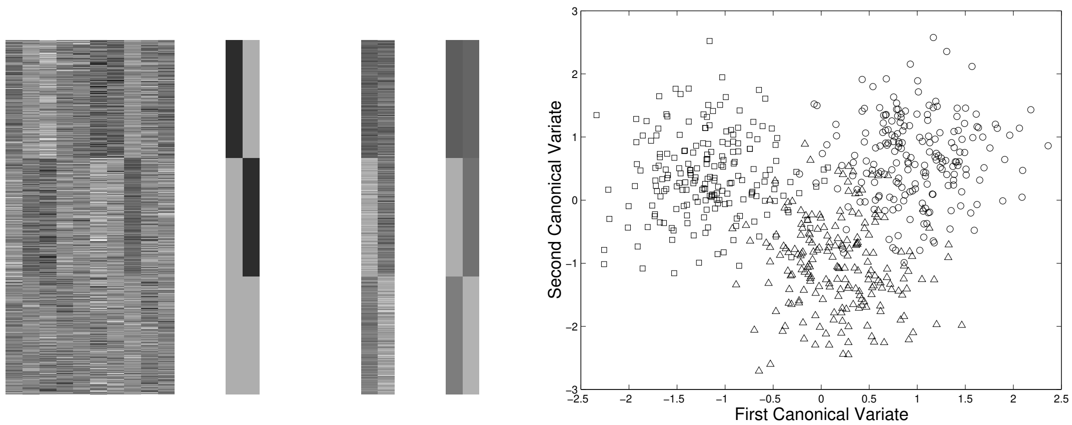

Figure 6.1: This figure illustrates canonical correlation analysis using simulated data. The two matrices on the left are a 600×10 data matrix ($\mathbf{X_a}$) and a 600×2 design matrix ($\mathbf{Y_a}$) after centering by orthogonalising with respect to a column of ones as described in Equations 6.1 and 6.2. The design matrix represents a partitioning of the data into three groups. The canonical variates for $\mathbf{X_a}$ and $\mathbf{Y_a}$ are shown in the third and fourth matrices ($\mathbf{X_a A}$ and $\mathbf{Y_a B}$). The left-hand columns of these two matrices contain the first canonical variate pair, whereas the second pair are in the right-hand columns. The graph on the right shows the two columns of $\mathbf{X_a A}$ plotted against each other, where the different symbols used represent memberships of the three different groups.

When the number of variables approaches the number of observations, the problem needs to be regularised. This can be done by computing the canonical variates from data that has been compacted using singular value decomposition as in Section 6.1.1. If $\mathbf{X}$ has been decomposed such that it can be approximated by $\mathbf{X} \simeq \mathbf{U}^*\mathbf{S}^*\mathbf{V}^{*T}$, and canonical correlation analysis performed on $\mathbf{U}^*$ and $\mathbf{Y}$ to give weight matrices $\mathbf{A}^*$ and $\mathbf{B}^*$, then the weights to be applied to the original data ($\mathbf{X}$) can be reconstructed by $\mathbf{V}^*\mathbf{S}^{*-1}\mathbf{A}^*$.

Figure 6.1 shows CCA as it would be used to graphically describe the differential features of three groups. However, it can also be used to aid classification. Once derived, the same weighting matrix ($\mathbf{A}$) can be applied to new data-sets that were not involved in its derivation. If there are only two groups involved in a study, then CCA can be used directly to assign group memberships to new observations[2]. With more groups, CCA serves as a graphical aid for assigning new observations to the groups.

6.2 Voxel-Based Morphometry

A number of studies have already demonstrated structural brain differences among different patient populations using the technique of *voxel-based morphometry* (VBM) (Wright *et al.*, 1995), or a related method that involves parcelling the images into different regions (Goldszal *et al.*, 1998). This section summarises the procedures involved in VBM, and provides evaluations of its statistical component.

[2]Providing that the signs of the canonical variates are adjusted accordingly.

6.2.1 Methods

Voxel-based morphometry of MRI data involves spatially normalising all the the subjects' images to the same stereotactic space, extracting the grey matter from the normalised images, smoothing, and finally performing a statistical analysis to localise, and make inferences about, group differences. The output from the method is an SPM showing regions where grey matter concentration differs significantly among the groups.

Spatial Normalisation

Spatial normalisation involves transforming all subjects' data to the same stereotactic space. This is achieved by registering each image to the same template image by minimising the residual sum of squared differences between them according to the methods described in Chapter 3. It should be noted that this method of spatial normalisation does not attempt to match every cortical feature exactly, but merely accommodates differences in global brain shape. If the spatial normalisation was perfectly exact, then all the segmented images would appear identical and no significant differences could be detected: VBM tries to detect differences in the regional concentration of grey matter at a local scale, having discounted global shape differences.

It is important that the quality of the registration is as high as possible, and that the choice of template image does not bias the final solution. An ideal template consists of the average of a large number of MR images that have been registered to within the accuracy of the spatial normalisation technique. The spatially normalised images should have a relatively high resolution (1mm or 1.5mm isotropic voxels), so that the grey matter extraction method is not excessively confounded by partial volume effects, where voxels contain a mixture of different tissue types.

Image Partitioning

The spatially normalised images are then partitioned into grey matter (GM), white matter (WM), cerebro-spinal fluid (CSF), and three other background classes, using the tissue classification technique described in Chapter 5.

Smoothing the Grey Matter Segments

The grey matter images are now smoothed by convolving with an isotropic Gaussian kernel. This makes the subsequent voxel by voxel analysis comparable to a region of interest approach, because each voxel in the smoothed images contains the average concentration of grey matter from around the voxel (where the region around the voxel is defined by the form of the smoothing kernel). This is often referred to as "grey matter density", but should not be confused with cell packing density measured cytoarchitectonically. It will be referred to as "concentration" to avoid confusion. By the central limit theorem, smoothing also has the effect of rendering the data more normally distributed, thus increasing the validity of parametric statistical tests. Whenever possible, the size of the smoothing kernel should be comparable to the size of the expected regional differences between the groups of brains. The smoothing step also helps to compensate for the inexact nature of the spatial normalisation.

Logit Transform (optional)

In effect, each voxel in the smoothed image segments represents the local GM concentration (between zero and one). Often, prior to performing statistical tests on measures of concentration, the data are transformed using the *logit* transformation in order to render them more normally distributed. The logit transformation of a concentration p is given by:

$$logit(p) = \log_e\left(\frac{p}{1-p}\right) \tag{6.8}$$

For concentrations very close to either one or zero, it can be seen that the logit transform rapidly approaches infinite values. Because of this instability, it is advisable to exclude voxels from subsequent analyses that are too close to one or other extreme. An alternative method of analysis involves using logistic regression within the framework of the general*ised* linear model (GLM) (Taylor *et al.*, 1998), but this is beyond the scope of this chapter as it requires iterative re-weighted least-squares methods. Whether or not the logit transform is a necessary processing step for VBM will be addressed later.

Statistical Analysis

Statistical analysis using the general linear model (LM) is used to identify regions of grey matter concentration that are significantly related to the particular effects under study (Friston *et al.*, 1995d). The LM is a flexible framework that allows many different tests to be applied, ranging from group comparisons and identification of regions of grey matter concentration that are related to specified covariates such as disease severity or age, to complex interactions between different effects of interest. Standard parametric statistical procedures (t-tests and F-tests) are used to test the hypotheses, so they are valid providing the residuals, after fitting the model, are independent and normally distributed. If the statistical model is appropriate there is no reason why the residuals should not be independent, but there are reasons why they may not be normally distributed. The original segmented images contain values between zero and one, where most of the values are very close to either of the extremes. Only by smoothing the segmented images does the behaviour of the residuals become more normally distributed.

Following application of the LM, the significance of any differences is ascertained using GRF theory (Friston *et al.*, 1995b; Worsley *et al.*, 1996). A voxel-wise statistical parametric map (SPM) comprises the result of many statistical tests, so it is necessary to correct for multiple dependent comparisons.

6.2.2 Evaluations

A number of assumptions need to hold in order for VBM to be useful. Confounding effects must be eliminated or modelled as far as possible. For example, it is not appropriate to compare two different groups if the images are acquired on different scanners, or with different MR sequences. In cases such as this, any group differences may be attributable to scanner differences rather than to the subjects themselves. Subtle but systematic differences in image contrast or noise can easily become statistically significant when a large number of subjects are entered in a study. An issue of validity concerns the assumptions required by the statistical tests. For parametric tests, it is often important that the data are normally distributed. If the data are not well behaved, then it is useful to know what effects there may be on the statistical tests. If there is doubt about the validity of the assumptions, it is better to use a non-parametric statistical analysis (Holmes *et al.*, 1996; Bullmore *et al.*, 1999).

Evaluation of the Assumptions About Normally Distributed Data

The statistics used to identify structural differences make the assumption that the residuals after fitting the model are normally distributed. Statistics cannot prove that data are normally distributed – they can only be used to disprove the hypothesis that they are normal. For normally distributed data, a *Q-Q plot* of the data should be a straight line. A significant deviation from a straight line can be identified by computing the correlation coefficient of the plot as described in Johnson and Wichern (1998).

A Q-Q plot is a plot of the sample quantile versus the sample quantile that would be expected if the residuals were normally distributed. Computing the sample quantile involves first sorting the J residuals (after applying the appropriate corrections derived from residual forming matrix) into increasing order $(x_1, x_2, \dots x_J)$. The inverse cumulative distribution of each of the J elements is then computed as:

$$q_j = \sqrt{2} erf^{-1}\left(2\frac{j-\frac{3}{8}}{J+\frac{1}{4}} - 1\right) \tag{6.9}$$

where erf^{-1} is the inverse error function. A Q-Q plot is simply a plot of $\mathbf{q}$ versus $\mathbf{x}$, and should be a straight line if the data in $\mathbf{x}$ are normally distributed. To test normality, the correlation coefficient for the Q-Q plot is used to identify any significant deviation from a straight line. A lookup table is used to reject the null hypothesis if the correlation coefficient falls below a particular value, given a certain sample size. However in this work, the correlation coefficient is simply used as a "normality statistic", and its distribution is examined over voxels.

The data used to test the assumptions were T1 weighted MRI scans of 50 normal male right handed subjects aged between 17 and 62 (median 26, mean 29). The structural scans had been acquired as part of an ongoing programme of functional imaging research. The scans were performed on a Siemens MAGNETOM Vision scanner operating at 2 Tesla. An MPRAGE sequence was used with a 12^o tip angle, 9.7ms repeat time, 4ms echo time and 0.6ms inversion time, to generate sagittal images of the whole brain with voxel sizes of $1\times1\times1.5$mm. The images were spatially normalised, segmented and smoothed using a Gaussian kernel of 12mm full width at half maximum (FWHM).

Voxel-by-voxel correlation coefficients of the Q-Q plots were computed over all voxels where the mean intensity over all images was greater than 0.05. Voxels of low mean intensity were excluded from the computations, because they would not be included in a VBM analysis. This is because it is known that these low intensity voxels are most likely to deviate strongly from the assumptions about normality. Q-Q plots were computed using two different linear models. The first model involved looking at the residuals after fitting the mean, whereas the second was more complex, in that it also modelled the confounding effect of the total amount of grey matter in each volume. Q-Q plots were computed both with and without the logit transform. Histograms of the correlation coefficients were computed over the whole image (717191 voxels), along with histograms generated from simulated Gaussian noise. These are plotted in Figure 6.2, and show

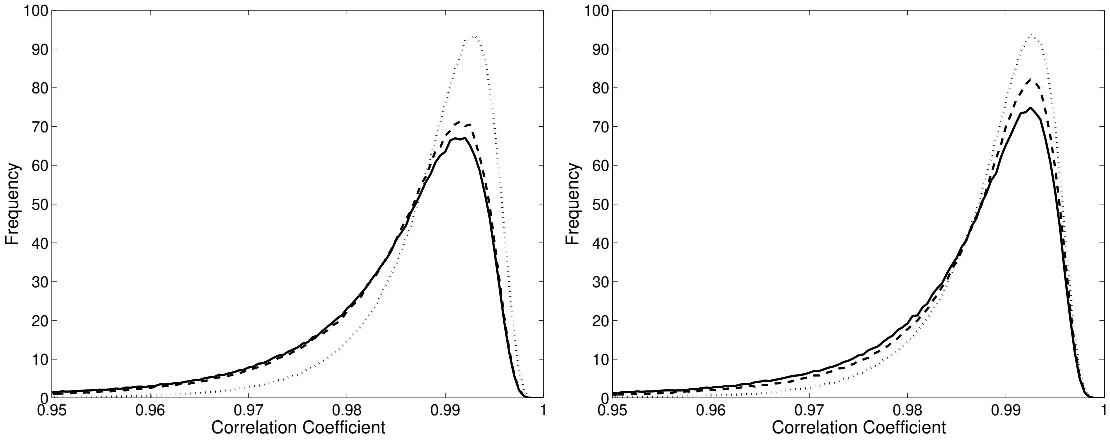

Figure 6.2: Histogram of correlation coefficients taken over the whole image volumes (using a total of 717191 voxels where the mean intensity over all images was greater than 0.05). The dotted line is the histogram that would be expected if the data were perfectly normally distributed. The solid line shows the histogram of the data without the logit transform, and the dashed line shows the histogram obtained using the logit transformed data. The plot on the left is based on the model that does not include global grey matter as a confound, whereas that on the right does model this confounding effect.

that the data does deviate slightly from a normal distribution. The logit transform appeared to make the residuals slightly more normally distributed. The normality of the residuals was also improved by modelling the total amount of grey matter as a confounding effect.

Testing the Rate of False Positives using Randomisation

The previous section showed that the data are not quite normally distributed, but it does not show how the non-normality influences any subsequent statistics. Ultimately, we wish to protect against false-positive results, and in this section, the frequency with which they arise is tested (Bullmore *et al.*, 1999). The statistics were evaluated using the same pre-processed structural brain images of 50 subjects as were used in the previous section. The subjects were randomly assigned, with replacement, to two groups of 12 and 38, and statistical tests performed using SPM99 (Wellcome Department of Cognitive Neurology, London, UK) to compare the groups. The particular numbers in the groups were chosen as many studies typically involve comparing about a dozen patients with a larger group of control subjects. This was repeated a total of 50 times, looking for both significant increases and decreases in the grey matter concentration of the smaller group. The end result is a series of 100 statistical parametric parametric maps of the t-statistic.

Within each of these SPMs, the local maxima of the t-statistic fields were corrected for the number of dependent tests performed, and a p value assigned to each (Friston *et al.*, 1995b; Worsley *et al.*, 1996). Using a corrected threshold[3] of p=0.05, one would expect about five

[3] A sufficiently smooth random field can be thresholded to reveal a number of distinct blobs, (more technically known as connected components of the excursion set, and often referred to as clusters). The number of clusters in the field minus the number of holes is the Euler characteristic of the excursion set. For high thresholds, the

clusters containing p values below this threshold by chance alone. Over the 100 SPMs, there were six clusters containing corrected p values below 0.05. The same 50 subjects were randomly assigned to either of the two groups and the statistics performed a further 50 times, but this time modelling the total amount of grey matter as a confounding effect. The results of this analysis produced four significant clusters with corrected p values below 0.05. These results suggest that the inference procedures employed are robust to the mild deviations from normality incurred by using smooth image partitions.

Another test available within SPM99 is based on the number of connected voxels in a cluster defined by a pre-specified threshold (extent statistic) (Friston *et al.*, 1995b). In order to be valid, this test requires the smoothness of the residuals to be spatially invariant, but this is known not to be the case by virtue of the highly nonstationary nature of the underlying neuro-anatomy. As noted by Worsley (1999), this nonstationary smoothness leads to inexact p values:

> "The reason is simple: by chance alone, large size clusters will occur in regions where the images are very smooth, and small size clusters will occur in regions where the image is very rough. The distribution of cluster sizes will, therefore, be considerably biased towards more extreme cluster sizes, resulting in more false positive clusters in smooth regions. Moreover, true positive clusters in rough regions could be overlooked because their sizes are not large enough to exceed the critical size for the whole region."

Corrected probability values were assigned to each cluster based on the number of connected voxels exceeding a t value of 3.27 (spatial extent test). Approximately five significant clusters would be expected from the 100 SPMs if the smoothness was stationary. Eighteen significant clusters were found when the total amount of grey matter was not modelled as a confound, and fourteen significant clusters were obtained when it was. Assuming Poisson statistics, the probabilities of obtaining this many significant clusters, under the null hypothesis that the extent statistic is valid, are 0.000001 and 0.000226 respectively. These tests confirmed that the voxel-based extent statistic should not be used in VBM.

Under the null hypothesis, repeatedly computed t-statistics should assume the probability density function of the Student's t distribution. This was verified using the computed t-fields, where each t-field contains 717191 voxels. Plots of the resulting histograms are shown in Figure 6.3. The top row presents distributions when global differences in grey matter are not removed as a confound. Note that global variance biases the distributions of t values from any particular comparison.

Further experiments were performed to test whether false positives occurred evenly throughout the brain, or were more specific to particular regions. The tests were done on a single slice through the same 50 subjects pre-processed brain images, but used the total count of grey matter in the brains as a confound. Each subject was randomly assigning to two groups of 12 and 38, pixel by pixel two-tailed t-tests were done, and locations of t-scores higher than 3.2729, or lower than -3.2729 were recorded (corresponding to an uncorrected probability of 0.002). This procedure was repeated 10000 times, and Figure 6.4 shows an image of the number of false positives occurring

number of holes becomes negligible, so the Euler characteristic is approximately equal to the number of clusters. Therefore, it is possible to control the expected number of false-positive clusters arising in an SPM by choosing an appropriate threshold. For example, a particular threshold on the random field could predict five false positive clusters in every 100 fields. The threshold can therefore be equated with a particular "corrected probability value" of 0.05. Similarly, any reasonably high value in the field can have a corrected p value associated with it.

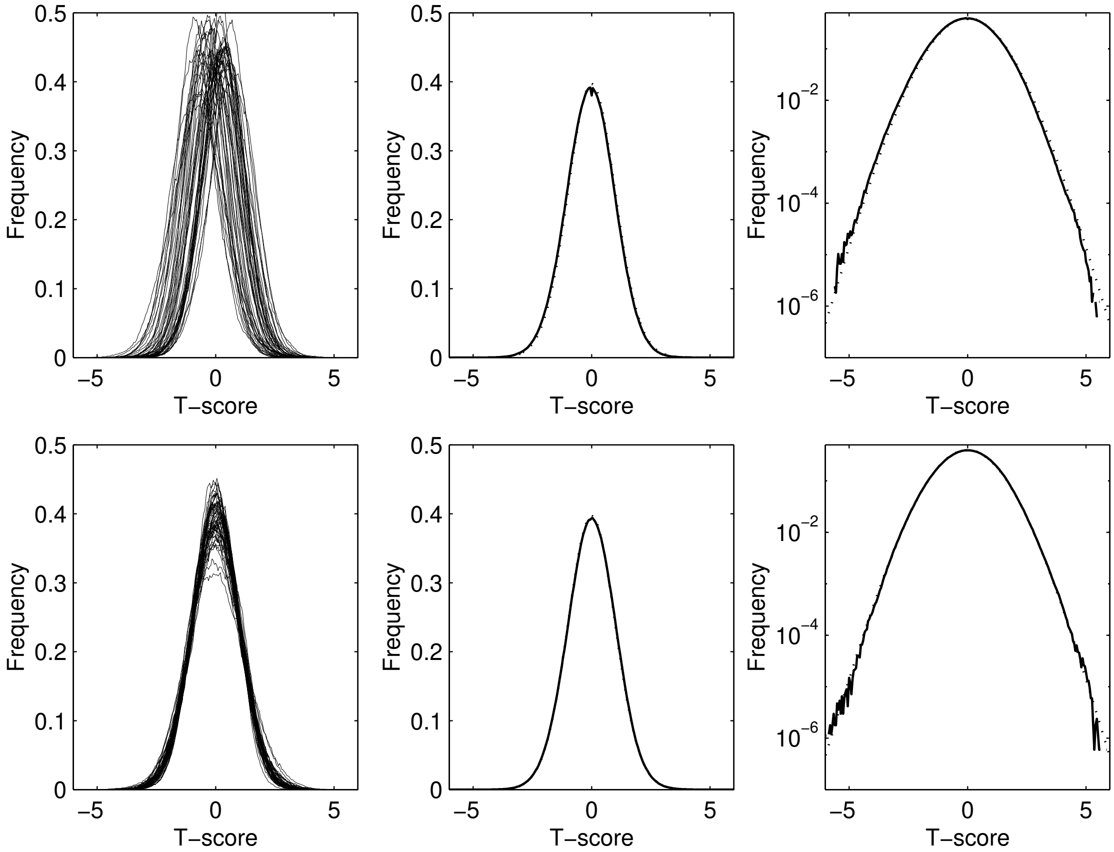

Figure 6.3: Histograms of t-scores from randomly generated tests. Above: Not modelling mean effect (48 degrees of freedom). Below: Modeling a mean effect as a confound (47 degrees of freedom). Left: 50 histograms of t-scores testing randomly generated effects of interest. Centre: the mean (i.e., cumulative distribution over all voxels and volumes) of the 50 histograms is plotted as a solid line, and the probability density function of the Students t distribution for 47/48 degrees of freedom is shown by the dotted line. Right: The same as centre, except plotted on a logarithmic scale.

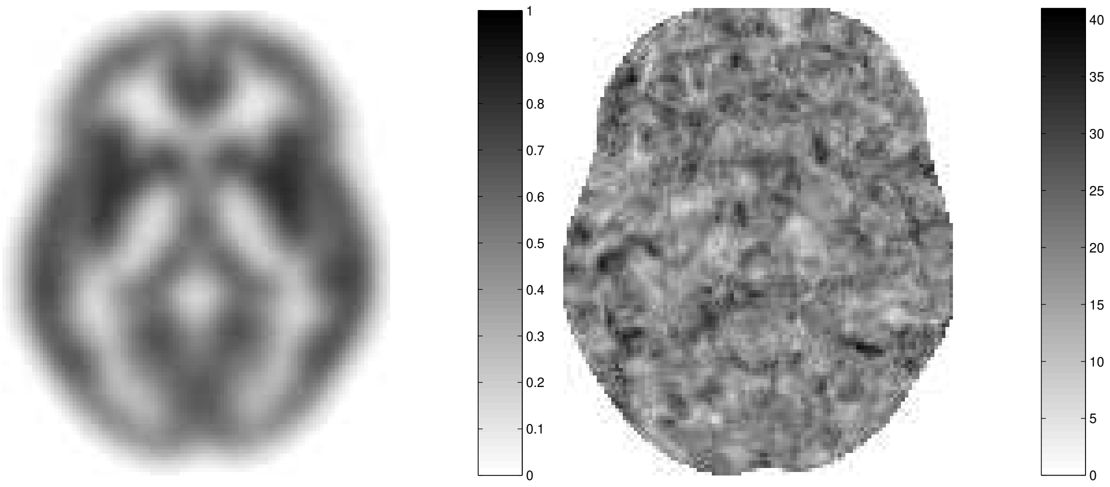

Figure 6.4: Left: Mean of 50 subjects pre-processed brain images. Right: Number of false positives occurring at each voxel at the uncorrected 0.002 level, after 10000 randomisations.

at each of the 10693 pixels. Visually, the false positives appear to be uniformly distributed. According to the theory, the number of false positives occurring at each pixel should be 20 (10000×0.002). An average of 20.171 false positives were found, showing that the validity of statistical tests based on uncorrected t statistics are not severely compromised.

The results from randomisation have involved comparisons between two groups of subjects. Further evaluations are required in order to assess the rate of false positive results arising from comparisons between single subjects and whole groups. It is expected that non-normally distributed errors may have a much more detrimentel effect on these types of parametric tests (see Section 6.4).

6.3 Deformation Based Morphometry

This section introduces a different technique for characterising differences among structural or anatomical brain images. The anatomical differences between any two brains can be expressed at a microscopic scale (e.g., differences in cytoarchitectonics or myeloarchitectonics), at a mesoscopic scale (e.g., cortical dysplasia) or at a macroscopic level (e.g., ventricular enlargement or abnormal temporal lobe asymmetry). From the perspective of neuroimaging, differences at a mesoscopic and macroscopic level are amenable to measurement. The technique described here characterises global differences in macroscopic anatomy that complements voxel-based morphometry, allowing one to examine differences at both mesoscopic and macroscopic scales.

Deformation-based morphometry (DBM) is a characterisation of the differences in the vector fields that describe global or gross differences in brain shape. These vector fields are the deformation fields used to effect nonlinear variants of spatial normalisation, when one of the images is a template that conforms to some standard anatomical space. In what follows, the deformation fields that map a series of images onto the same template are compared to see if there are any systematic differences. Because the deformation fields are multivariate, standard multivariate statistical techniques are employed to estimate the nature of the differences and to make inferences about them. The endpoint of DBM is a p value pertaining to the significance of the effect and one or more canonical vectors, or deformations, that characterise their nature. These results are obtained using multivariate analysis of covariance (MANCOVA) and canonical correlation analysis (CCA) respectively.

In order to demonstrate the technique, a study of dimorphism in relation to handedness and sex has been chosen. This should be seen as a vehicle to explain and illustrate how to do these analyses.

6.3.1 Methods

High resolution structural T1 MR images were collected from 61 normal healthy volunteers. These were all acquired on the same 2 Tesla Siemens Magnetom Vision scanner, using an MPRAGE sequence. The resolution of the images was $1mm \times 1mm \times 1.5mm$. The data consisted of 15 female right-handed subjects, 5 female left-handed subjects, 30 male right-handed subjects and 11 male left-handed subjects, all between the ages of 20 and 37. The scans were all acquired as part of ongoing functional imaging projects within the department, and all subjects had no neurological or psychiatric history.

The data were spatially normalised to the same template, and their shapes extracted from the estimated deformations. The shapes were compacted into their most significant "eigen-warps", before multi-variate statistical analysis was performed.

Spatial Normalisation

The images were spatially normalised using the methods described in Chapter 3. The template image consisted of an average of twelve 12-parameter affine registered T1 MR images of the head, and was rendered symmetric (so it could be used to examine brain asymmetry) by averaging with itself reflected across the sagittal midplane. The MRI sequence used to generate the images that constituted the template, was identical to that used for all the other images, thus ensuring that more accurate registrations could be achieved.

The first step of the normalisation was to determine the optimum 12-parameter affine transformation. Initially, the registration was performed by matching the whole of the head (including the scalp) to the template. Following this, the registration proceeded by only matching the brains together, by appropriate weighting of the template voxels (see Figure 6.5). This is a completely automated procedure that largely discounts the confounding effects of skull and scalp differences.

The affine registration was followed by estimating nonlinear deformations. Each of the deformation fields was described by 1176 parameters, where these represent the coefficients of the deformations in three orthogonal directions. The mean of the spatially normalised images is shown for each group in Figure 6.6. It can be seen that in terms of gross anatomy, following normalisation, they are virtually indistinguishable.

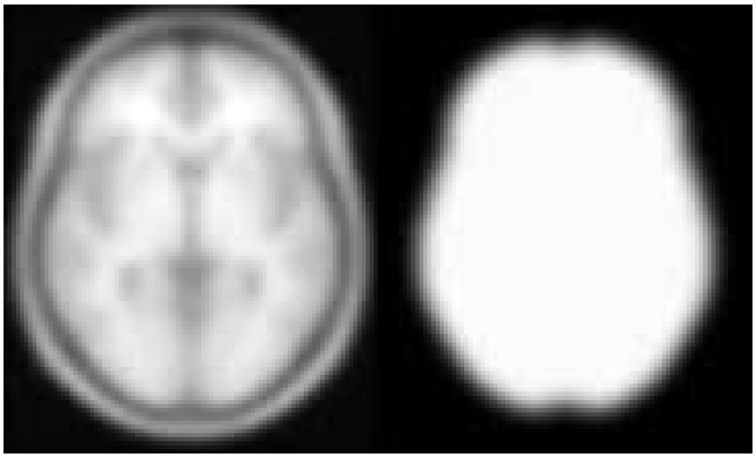

Figure 6.5: The template (left) and weighting image (right) used by the registration. Note that the images have been smoothed using an 8mm FWHM isotropic Gaussian kernel in order to facilitate the registration.

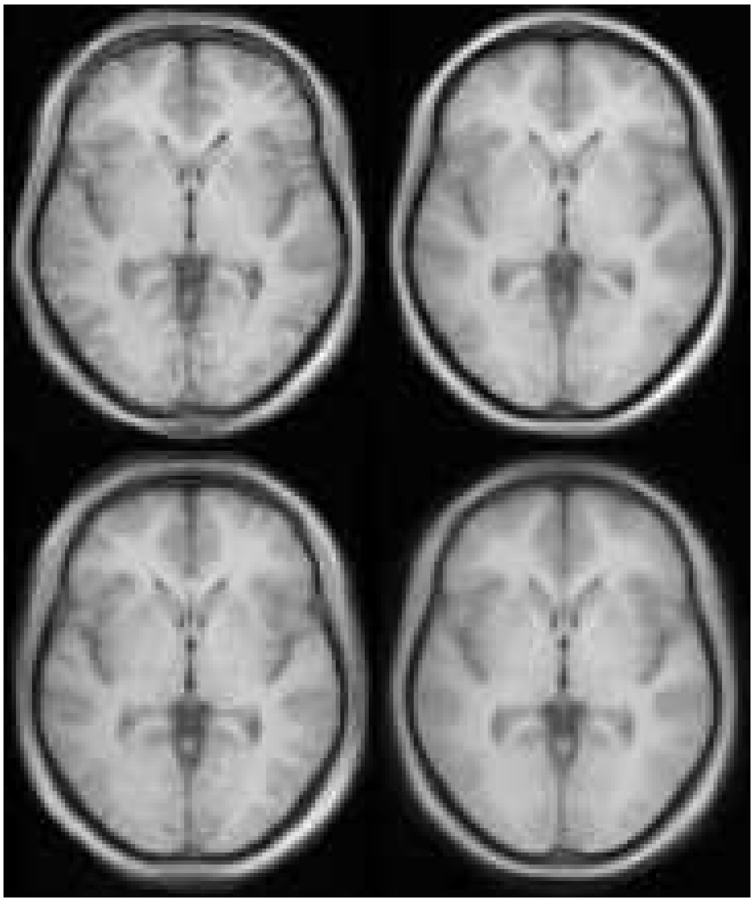

Figure 6.6: The mean of the spatially normalised images for each group: left handed females (above left), right handed females (above right), left handed males (below left) and right handed males (below right).

Extracting Shape Information

Each set of spatial normalisation parameters (affine and nonlinear) encodes a deformation field relating to the position, size and shape of the brain. For the analysis presented here, only the information relating to the shapes of the brains was used, by removing the effects of size and position.

The deformation fields are defined by both nonlinear and affine components. In order to proceed, it is necessary to decompose the transformation into components relating to position and size (uninteresting components), and shape (the components of interest). In order to effect this decomposition, each deformation field was reconstructed from its parameters. Each field provides a mapping from points in the template to points in the source image, allowing the landmark methods described by Bookstein (1997b) to be used to extract the size and positional information. The extracted measures were such that the remaining shapes minimised the squared Procrustes distance between the template and images. Rather than basing the procedure on a few landmarks, all the elements of the deformation field corresponding to voxels within the brain were considered (by weighting with the image shown in Figure 6.5). This involved first determining translations by computing centres of mass:

$$\begin{aligned}\bar{\mathbf{x}} &= \frac{\sum_{i=1}^{I} \mathbf{x}_i w_i}{\sum_{i=1}^{I} w_i} \\ \bar{\mathbf{y}} &= \frac{\sum_{i=1}^{I} \mathbf{y}_i w_i}{\sum_{i=1}^{I} w_i}\end{aligned} \tag{6.10}$$

where $\mathbf{x}_i$ is the co-ordinate of the ith voxel of the template, $\mathbf{y}_i$ is the location that it maps to, and w_i is the weighting for that element. The rotations were computed from the cross-covariance matrix ($\mathbf{C}$) between the elements and deformed elements (after removing the effects of position):

$$c_{jk} \propto \sum_{i=1}^{I} w_i (x_{ij} - \bar{x}_j)(y_{ik} - \bar{y}_k) \tag{6.11}$$

The 3×3 matrix $\mathbf{C}$ was decomposed using singular value decomposition to give three matrices, $\mathbf{U}$, $\mathbf{S}$ and $\mathbf{V}$ (such that $\mathbf{C} = \mathbf{USV}^T$, where $\mathbf{U}$ and $\mathbf{V}$ are unitary, and $\mathbf{S}$ is a diagonal matrix). The rotation matrix ($\mathbf{R}$) could then be reconstituted from these matrices by $\mathbf{R} = \mathbf{UV}^T$. Finally, moments around the centres were used to correct for relative size differences (z):

$$z = \sqrt{\frac{\sum_{j=1}^{3} \sum_{i=1}^{I} (x_{ij} - \bar{x}_j)^2 w_i}{\sum_{j=1}^{3} \sum_{i=1}^{I} (y_{ij} - \bar{y}_j)^2 w_i}} \tag{6.12}$$

After removing the effects of translation, rotation and scaling from the deformation fields, they were then reparameterised by the lowest frequency coefficients of their three dimensional DCTs.

Statistical Analysis

Following this, a data matrix was generated, where each row contained the coefficients of the nonlinear basis functions describing the difference in shape between the template and each image. For the multivariate analysis that followed, it was necessary to reduce the number of these coefficients relative to the number of images. Singular value decomposition was used to compact this information, such that about 96% of the variance of the nonlinear deformations was represented by 20 parameters for each image.

MANCOVA was used to make inferences about the effects of interest (i.e. provide p values). In the simplest case of comparing two groups, the MANCOVA becomes the special case of Hotelling's T^2 test. MANCOVA does not simply tell one what the difference is. To characterise these differences, one usually uses CCA based upon the parameters estimated by the MANCOVA. CCA is a device that finds the linear combination of the dependent variables (in this case the deformations) that is maximally correlated with the explanatory variables (e.g., male vs. female). In the simple case of one categorical explanatory variable (e.g., sex) this will be the deformation field that best discriminates between males and females. Note that this is not the same as simply subtracting the deformation fields of two groups. This is because (i) the MANCOVA includes the effects of confounds that are removed and (ii) some aspects of the deformations may be less reliable than others (CCA gives deformations that explicitly discount error in relation to predicted differences). The canonical deformations can either be displayed directly as deformation fields, or can be applied to some image to "caricature" the effect detected. In this section, both are combined in order to illustrate the deformations more clearly.

6.3.2 Results

Tests for significant differences between groups of subjects were performed using a MANCOVA on the deformation parameters. A number of tests were performed, including tests relating to the handedness of the subjects, of sexual dimorphism, looking at brain asymmetry, and interactions among these factors.

Handedness, Sex and the Interaction Between Them

A MANCOVA testing for the effects of both handedness and sex simultaneously, suggested extremely significant effects ($p = 2.1 \times 10^{-7}$). Because there were two effects of interest, CCA could be used to generate a scatter-plot representing the optimum separation of the groups (see Figure 6.7) in terms of the two corresponding canonical variates. It can be seen that the first canonical variate is mainly sensitive to the differences between men and women, whereas the second discriminates between handedness.

The effects of sex and handedness were then tested individually, both showing significant differences ($p = 0.00014$ and $p = 0.00020$ respectively). The test for sex differences used handedness as a confound and that for handedness used sex as a confound. A further test failed to show any interaction between handedness and sex ($p = 0.90$). Differences between the groups were characterised using CCA, the results of which are illustrated in Figure 6.8. These can be compared to the difference between the brain shapes (after removing confounding effects) as shown in Figure 6.9.

The effect of sex can be most clearly seen in the sagittal view and suggests that men have a more protruding occipital pole, whereas women have more prominent frontal poles. The effect of handedness involves more asymmetric differences affecting predominantly the right frontal lobe (transverse section, middle row of Figure 6.8).

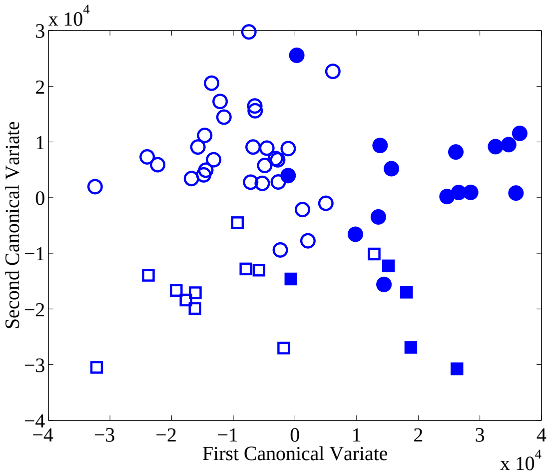

Figure 6.7: Separation of subjects using canonical correlation analysis. Right handed females (filled circles), left handed females (filled squares), right handed males (empty circles) and left handed males (empty squares).

Brain Asymmetry

Because the template used by the spatial normalisation was symmetric, it was possible to look at left/right brain asymmetry. The coefficients of the DCT can be divided into those that account for nonlinear deformations that are symmetric, and those that relate to anti-symmetric deformations. Very significant brain asymmetries were detected ($p < 0.001$) by testing that the coefficients of the anti-symmetric warps differed from zero. Geschwind & Galaburda (1987) discuss many of the asymmetries found by a number of researchers. These include the fact that the left occipital lobe is broader and longer than that on the right, which is confirmed in Figure 6.9. However, because of the large amount of variability in the occipital lobes, this is not a feature of brain asymmetry that is strongly characterised by CCA (see Figure 6.8). Another asymmetry (that was not really confirmed in Geschwind & Galaburda) is that the right frontal lobe is usually larger than that on the left. However, the results obtained here contradict this finding, in that the left frontal lobe appears to be the larger of the two. From Figure 6.9 one can see that the magnitude of the difference is relatively small, but it is still a feature that is strongly characterised by CCA. Differences in asymmetry between males and females and between left and right handed subjects were both found to be significant ($p = 0.026$ and $p = 0.0076$ respectively).

In short, reliable features of asymmetry and dimorphism may not necessarily be the biggest or most obvious. Furthermore, the approach presented here gives estimates of dimorphism that explicitly discount differences due to other factors.

6.4 Tensor-Based Morphometry

The objective of *tensor-based morphometry* (TBM) is to localise regions of shape differences among groups of brains, based on deformation fields that map points in a template (x_1, x_2, x_3) to equivalent points in individual source images (y_1, y_2, y_3). In principle, the Jacobian matrices of the deformations (a 2nd order tensor field relating to the spatial derivatives of the transformation)

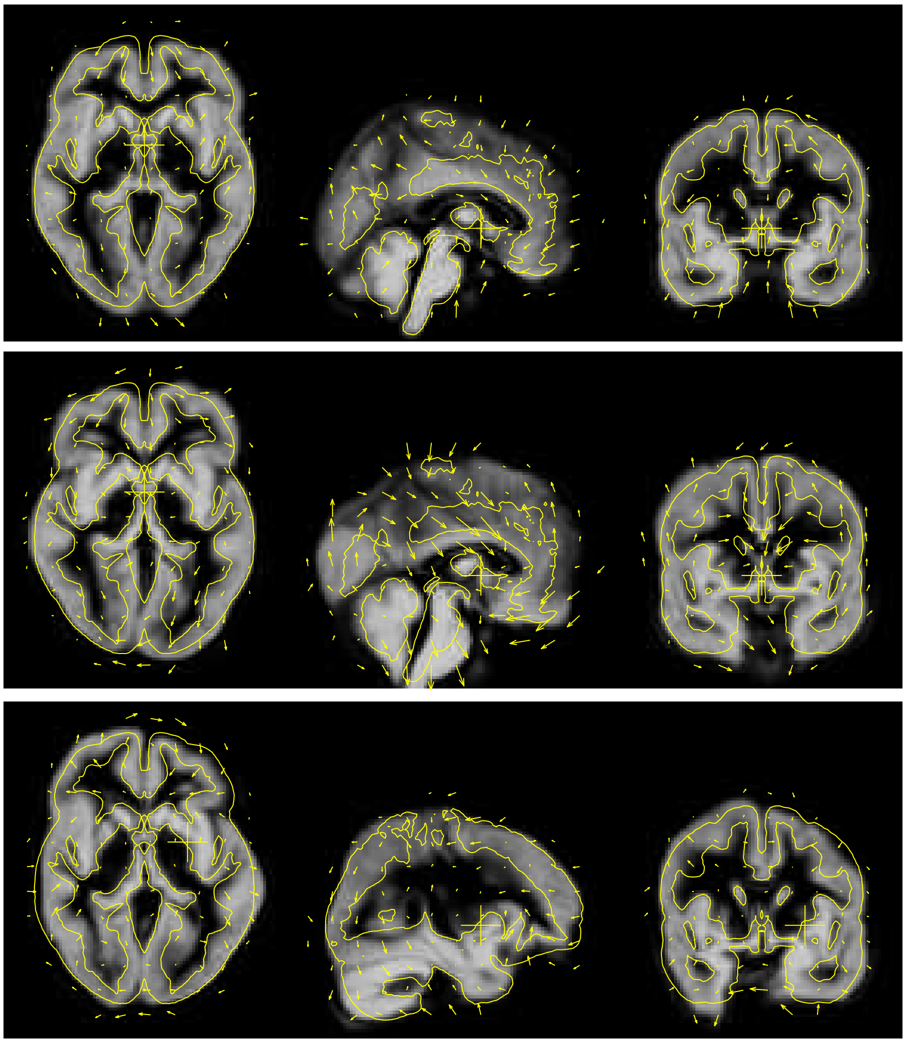

Figure 6.8: Nonlinear warps that caricature a male brain (above), a right-handed brain (centre), and natural brain asymmetry (below). These have been characterised by canonical correlation analysis. The images of grey matter show the caricature of the deformations. Superimposed on this is a contour from the undeformed image. The arrows show the direction of the nonlinear warps characterised by CCA (from undeformed to deformed). In this figure, the deformations have been arbitrarily scaled for better visualisation. These are not the mean differences between the brain shapes, but rather the differences that most clearly distinguish them. In it's most general form, CCA produces a set of vectors that best partition the data according to the design matrix. If there are multiple effects of interest, then there is no simple relationship between these effects and the canonical variates, but with only one factor of interest (as in these examples), the canonical variates can be directly related to the factor. In the transverse and coronal sections, the left side of the brain is on the left side of the figure.

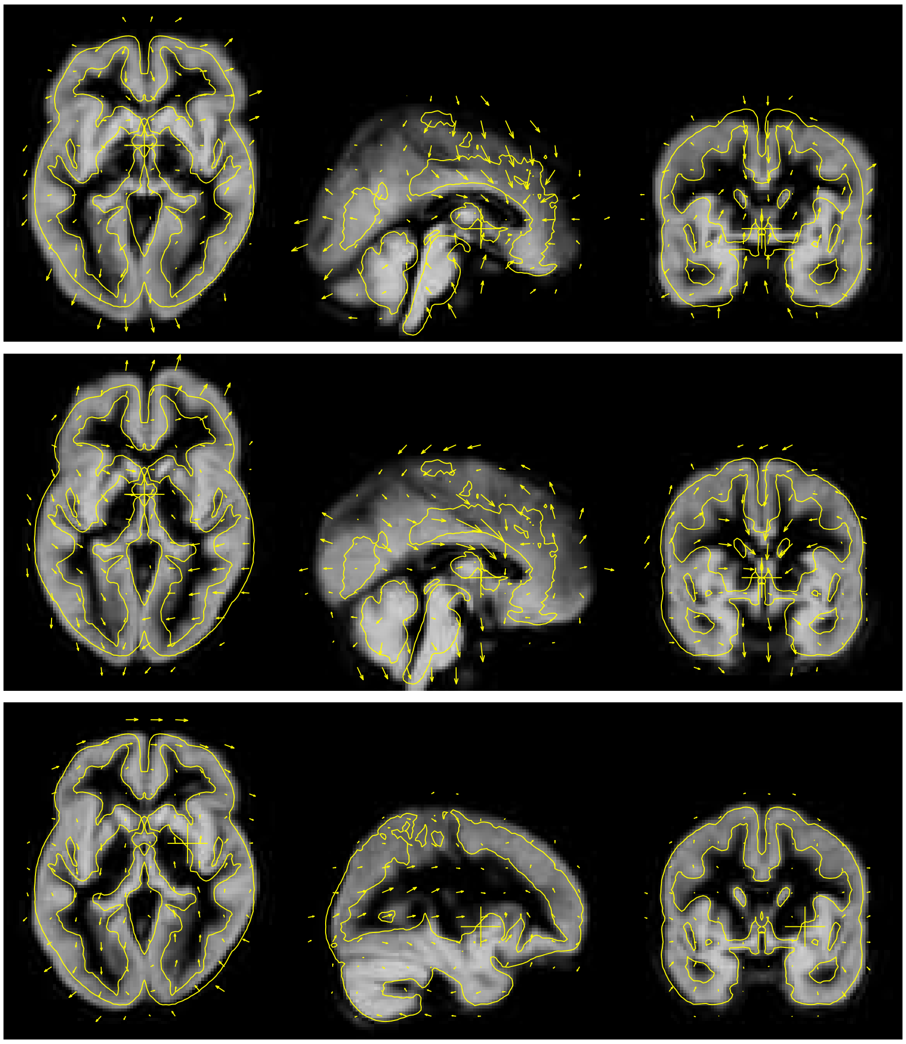

Figure 6.9: The deformation required to map from a female to male brain (above), left-handed to right-handed brain (centre), and antisymmetric deformations from a symmetric template to an asymmetric brain (below), all multiplied by a factor of 5. The deformations were computed after first removing the effects of confounds, and are a direct characterisation of the parameter estimates without referring to the errors or reliability of the differences (c.f., CCA).

should be more reliable indicators of local brain shape than absolute deformations. Absolute deformations represent positions of brain structures, rather than local shape, and need to be quantified relative to some arbitrary reference position.

6.4.1 Theory

A Jacobian matrix contains information about the local stretching, shearing and rotation involved in the deformation, and is defined at each point by:

$$\mathbf{J} = \begin{bmatrix} \partial y_1/\partial x_1 & \partial y_1/\partial x_2 & \partial y_1/\partial x_3 \\ \partial y_2/\partial x_1 & \partial y_2/\partial x_2 & \partial y_2/\partial x_3 \\ \partial y_3/\partial x_1 & \partial y_3/\partial x_2 & \partial y_3/\partial x_3 \end{bmatrix} \tag{6.13}$$

A simple form of TBM involves comparing relative volumes of different brain structures, where the volumes are derived by taking Jacobian determinants at each point (see Figures 6.10 and 6.11). Simple uni-variate statistics (t- or F- tests) can then be used to make inferences about regional volume differences among populations. This type of morphometry is useful for studies that have specific questions about whether growth or volume loss has occurred.

When many subjects are included in a study, a potentially more powerful form of TBM can be attained using multi-variate statistics on other shape measures derived from the Jacobian matrices. This use of multivariate statistics does not explicitly test for growth or shrinkage, but indicates whether there is any shape difference. It is therefore useful when there is no clear hypothesis about the nature of the differences, as may be the case when studying the effects of maturation on the human brain. This form of morphometry should be able to identify shape differences even when volumes are the same.

Because a Jacobian matrix encodes both local shape (zooms and shears) and orientation, it is necessary to remove the latter before making inferences about shape. According to the polar decomposition theorem (Ogden, 1984), a non-singular Jacobian matrix can be decomposed into a rotation matrix ($\mathbf{R}$) and a symmetric positive definite matrix ($\mathbf{U}$ or $\mathbf{V}$), such that $\mathbf{J} = \mathbf{RU} = \mathbf{VR}$. Matrices $\mathbf{U}$ and $\mathbf{V}$ (called the *right* and *left stretch* tensors respectively) are derived by $\mathbf{U} = \left(\mathbf{J}^T\mathbf{J}\right)^{1/2}$ and $\mathbf{V} = \left(\mathbf{J}\mathbf{J}^T\right)^{1/2}$. Matrix $\mathbf{R}$ is then given by $\mathbf{R} = \mathbf{J}\mathbf{U}^{-1}$ or $\mathbf{R} = \mathbf{V}^{-1}\mathbf{J}$.

For a purely rigid body transformation, $\mathbf{U} = \mathbf{V} = \mathbf{I}$ (the identity matrix). Similarly, if $\mathbf{R} = \mathbf{I}$, then the deformation can be considered as *pure strain*. Deviations of $\mathbf{U}$ or $\mathbf{V}$ away from $\mathbf{I}$ indicate a shape change, which can be represented by a strain tensor $\mathbf{E}$. Depending upon the reference co-ordinate system, the strain tensor is referred to as either a *Lagrangean* or an *Eulerian* strain tensor. When the strain tensor is derived from $\mathbf{U}$, it is referred to as a Lagrangean strain tensor, whereas when it is derived from $\mathbf{V}$, it is Eulerian. Spatial normalisation of a series of source images involves determining a mapping from each point in the template image, to corresponding points in the source images. In order to compare image shapes, it is necessary to derive measures of shape within the co-ordinate system of the template image, rather than within the different co-ordinate systems of the individual source images. Therefore, the Lagrangean framework should be used.

For any deformation, there is a whole continuum of ways of defining strain tensors, based on a parameter m. When m is non-zero, the family of Lagrangean strain tensors $\mathbf{E}$ are given by $\mathbf{E}^{(m)} = m^{-1}\left(\mathbf{U}^m - \mathbf{I}\right)$. For the special case when m is zero, the strain tensor is given by

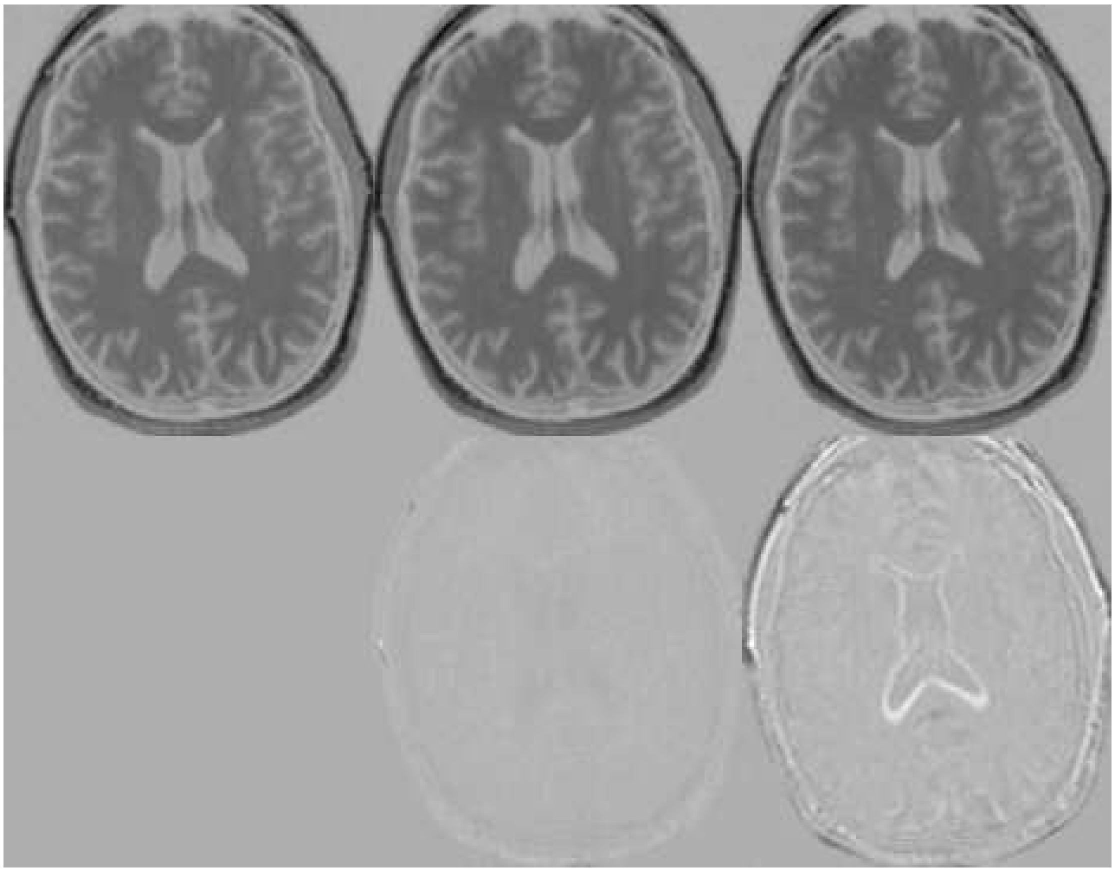

Figure 6.10: This figure illustrates warping together a pair of serial scans of the same Alzheimers subject, using the three dimensional method presented in Chapter 4. The data were provided by the Dementia Research Group, National Hospital for Neurology and Neurosurgery, London, UK. From left to right, the top row shows a slice from the late scan (used as the template), the early scan after rigid body registration followed by warping, and the early scan after rigid registration only. The bottom row shows the difference between the late and early scan, with and without warping. All images are displayed with the same intensity scaling. With only rigid registration, the difference between the early and late scan exhibits a difference at the edges of the ventricles (among other regions). This difference is greatly reduced by including warping in the spatial transformation. The information contained in the difference image is transferred to the deformation fields. One informative feature of the deformation fields are the Jacobian determinants, which reflect volume changes. These are shown in Figure 6.11.

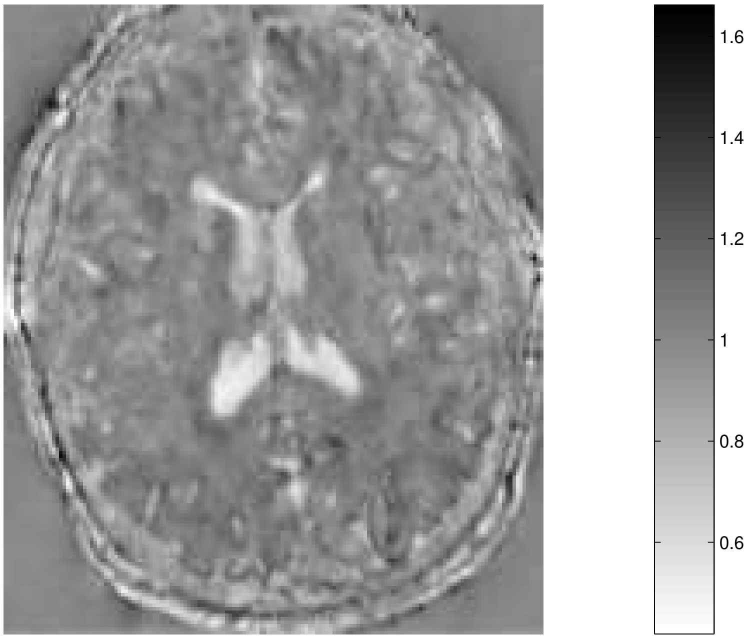

Figure 6.11: This figure illustrates the volume changes estimated by warping together the images shown in Figure 6.10. The relative volumes are the Jacobian determinants of the deformation field. Smaller determinants are obtained when a region of the template maps to a smaller region in the source image. In this example, they represent regions that have expanded between the early and late scans. Regions where there are no measurable volume changes have Jacobian determinants with a value of one.

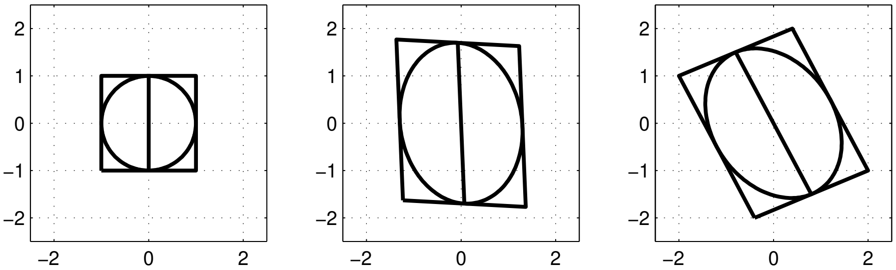

Figure 6.12: This figure illustrates polar decomposition, whereby the regular shape in the left-hand panel is transformed via shears and zooms to give the shape in the centre, before being rotated to give the right-hand shape.

$\mathbf{E}^{(0)} = ln(\mathbf{U})$, where the ln refers to a matrix logarithm. When m assumes values of -2, 0, 1 or 2, then the tensors are called the *Almansi*, *Hencky*, *Biot* or *Green* strain tensors respectively.

For deformations derived using the methods in Chapter 4[4], the Hencky tensor may be the most appropriate measure of local shape, providing the priors exert enough influence on the resulting deformation fields. One advantage of basing measures on the Hencky tensor is that it allows relative changes in length, volume and area to be modelled using similar log-normal distributions. This would not be possible if the measures were modelled by normal distributions, since if a variable l is normally distributed, then l^3 is not. Also, it would not make sense to assume distributions that allow negative lengths or volumes, if these are explicitly prohibited by the warping algorithm.

In three dimensions, a strain tensor has six unique elements (as it is a symmetric 3×3 matrix). Multivariate statistics on these elements can model their means, and also their distribution about the mean via a variance-covariance matrix. Such a covariance matrix is able to describe many properties of the warps. The simplest of these are modelled by the diagonal elements of the variance-covariance matrix, which are able to characterise different amounts of warping in different directions. The off diagonal elements show correlations between the strains in different directions, allowing more complex material properties to be modelled. For example, a covariance matrix that enforces the trace of the Hencky tensor to equal zero would model transformations where volumes are preserved (isochoric transformations). In two dimensions, such a covariance matrix could take the following form:

$$\mathbf{C} = \begin{bmatrix} t_1 & -t_1 & -t_3 \\ -t_1 & t_1 & t_3 \\ -t_3 & t_3 & t_2 \end{bmatrix} \tag{6.14}$$

where t_1, t_2 and t_3 are arbitrary values, and the Hencky tensor has been parameterised by $\mathbf{p}$ which has been drawn from a multi-normal distribution of zero mean and variance-covariance $\mathbf{C}$:

$$\mathbf{E}^{(0)} = \begin{bmatrix} p_1 & p_3 \\ p_3 & p_2 \end{bmatrix} \tag{6.15}$$

6.4.2 Data for Evaluations

The data used for the evaluation were MR scans of 58 male right-handed subjects. Sagittal images of the whole brain with voxel sizes of $1 \times 1 \times 1.5$mm were acquired on a Siemens MAGNETOM Vision scanner operating at 2 Tesla using an MPRAGE sequence with a 12^o tip angle, 9.7ms repeat time, 4ms echo time and 0.6ms inversion time.

The images were all approximately warped to the same stereotactic space using the methods described in Chapter 3. Following this, a $41 \times 67 \times 59$mm region around the right hippocampus was identified, and the deformation fields refined for this region using a similar procedure to that described in Section 4.3.3. First of all, the approximately spatially normalised images were used to create an average image for this region. This average was used as a template for subsequent high-dimensional registrations, but was modified every four iterations to reflect the new warped

[4]It should be noted that the sum of squares of the Hencky tensor elements is equivalent to the sum of squares of the logs of the singular values of $\mathbf{J}$, meaning that the priors described in Chapter 4 are effectively minimising the squares of the Hencky tensor.

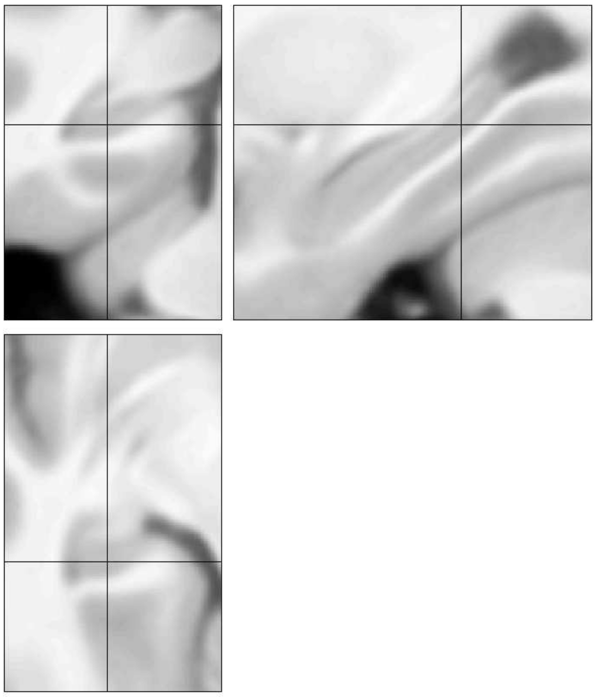

Figure 6.13: Orthogonal views through the mean of 58 warped hippocampus images.

average. A value of 4 was used for λ (the regularisation parameter), and a total of approximately 220 iterations were performed for each image. The resulting deformation fields had an isotropic resolution of 1mm, meaning that each of the 58 deformation fields was defined by 486219 parameters. Figure 6.13 shows orthogonal views through the mean of the normalised images. Six randomly chosen warped images from the dataset are shown in Figure 6.14.

A number of evaluations were performed using this dataset, both in the context of comparing Jacobian determinants between the groups, and also in the context of using multivariate methods to compare strain tensors. No evaluation of the validity of using GRF theory for correcting for multiple dependent comparisons was performed. This was partly because a solid theory for Wilk's Lambda fields has not yet been formulated. Also, some of the transforms applied to the fields involved smoothing, followed by taking logs, which would lead to violations of the GRF theories as they currently stand.

6.4.3 Morphometry on Jacobian Determinants

One form of TBM involves identifying structure volume differences among groups, based on the Jacobian determinants of the deformation fields. Generally, simple voxel-by-voxel t or F tests would be used to produce an SPM of significant volume differences among the groups.

The Jacobian determinant fields from the deformations described in Section 6.4.2 were written. Each element in the field was considered a node in a tetrahedral mesh. A Jacobian determinant was computed at each node from the average volume of the tetrahedra with which it formed a vertex. For half the elements, the value was the average of 8 tetrahedral volumes, whereas it was the average of 24 volumes for the other half (see Section 4.2.4). Images of six determinant fields

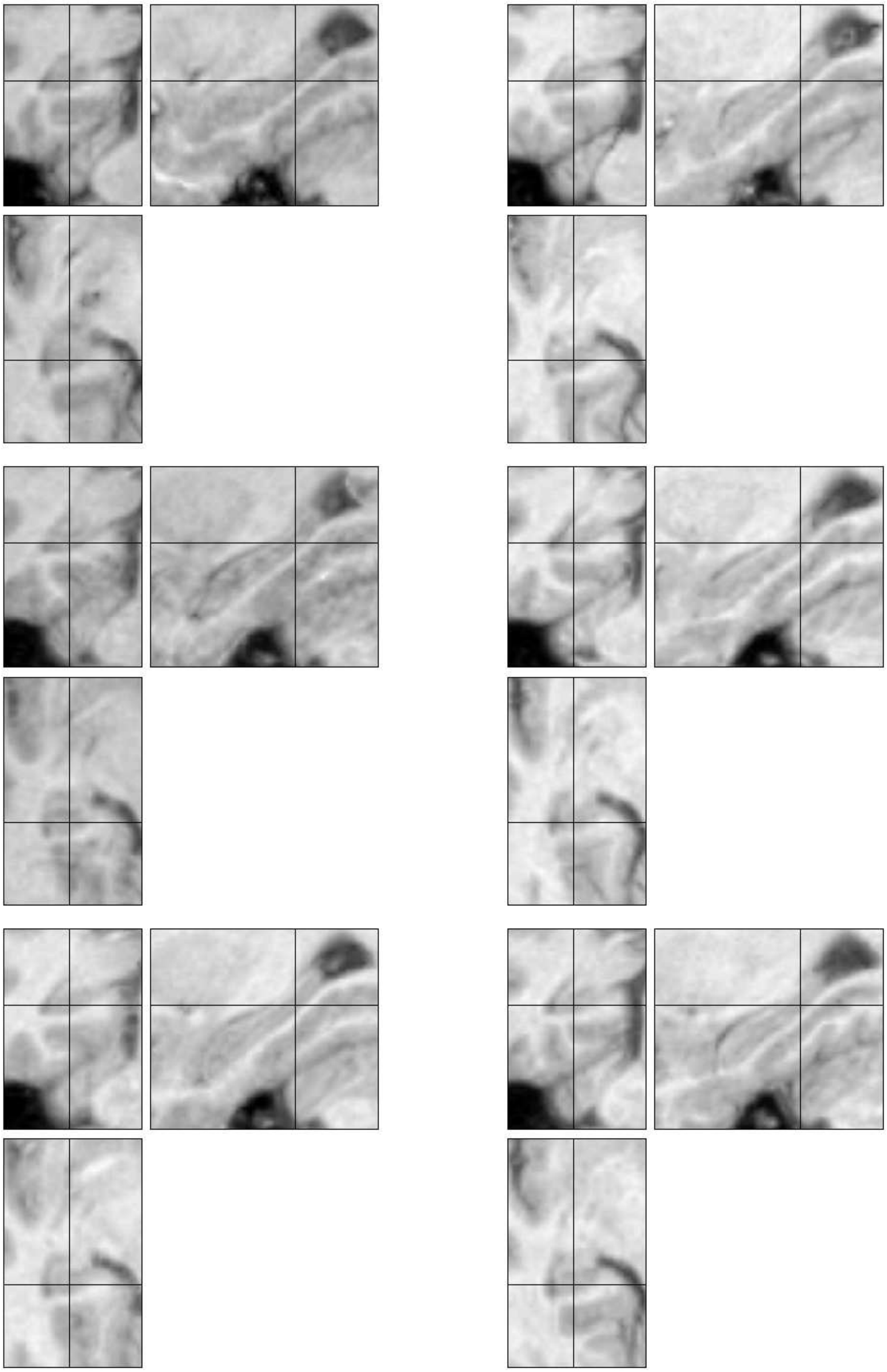

Figure 6.14: A selection of orthogonal views through randomly chosen warped images of hippocampi (see also Figure 6.15).

are shown in Figure 6.15.

Distributions of Jacobian Determinants

A number of assumptions are required in order to perform statistical tests, and when parametric tests are used within the framework of the general linear model, one of the requirement is that the residuals from the model are normally distributed. Section 6.2.2 described a method of testing normality based on correlation coefficients from Q-Q plots. A similar procedure is used here in order to examine the distributions of the residuals. A number of different transformations were also applied to the Jacobian determinants in order to assess their effects on the residuals. The transformations include spatial smoothing by an isotropic 8mm FWHM Gaussian kernel, and taking logs of the Jacobians both before and after smoothing. Histograms of the correlation coefficients over all voxels were computed as described in Section 6.2.2.

The data used in the analysis were derived from 42 of the subjects, the results of which are plotted in Figure 6.16. The first conclusion to be drawn from this figure is that when the warping scheme of Chapter 4 is used, the logs of the Jacobian determinants show more normally distributed residuals than the Jacobian determinants themselves. Smoothing also rendered the residuals more normally distributed, both with and without taking logs of the determinants. Whether logs were taken before or after smoothing appeared to make little difference to the behaviour of the residuals, but this still has implications for the analysis. When logs are taken before smoothing, any analysis is based on logs of the geometric mean of the relative volumes (integrated under the smoothing kernel), and therefore does not directly relate to arithmetic mean volumes. If logs are taken after smoothing, then any corrections for multiple comparisons made using GRF theory are likely to be compromised.

False Positive Rates from T-Tests

The previous section showed that the Jacobian determinants were not normally distributed, but that the normality can be improved by taking logs of the determinants, and also by smoothing. This section evaluates the rate of false positives that are likely to be encountered when performing voxelwise t tests on the logs of the smoothed Jacobian determinant field.

Voxelwise t tests were performed 1000 times over all 58 logged, 8mm smoothed determinant fields. The tests involved randomly assigning the fields to two groups, where one group contained 12 subjects and the other contained 46. The results are histograms of 162073000 t values, which are plotted in Figure 6.17. As hoped for, the resulting t values appear to be distributed reasonably well according to a t distribution with 56 degrees of freedom, implying that voxelwise t tests should produce the appropriate rate of false positive results when comparing groups containing several subjects.

Further tests were performed that involved comparing Jacobian determinants from each single subject with those of the remaining 57. The evaluations were done on the original determinant fields, on the fields after 8mm smoothing and after taking logs of the smoothed and unsmoothed data. Higher t-scores indicate larger Jacobian determinants in the deformation field of the single subject. The results (shown in Figure 6.18) suggest that parametric tests may be less valid for comparing single subjects with groups. Many of the effects of non-normally distributed data are

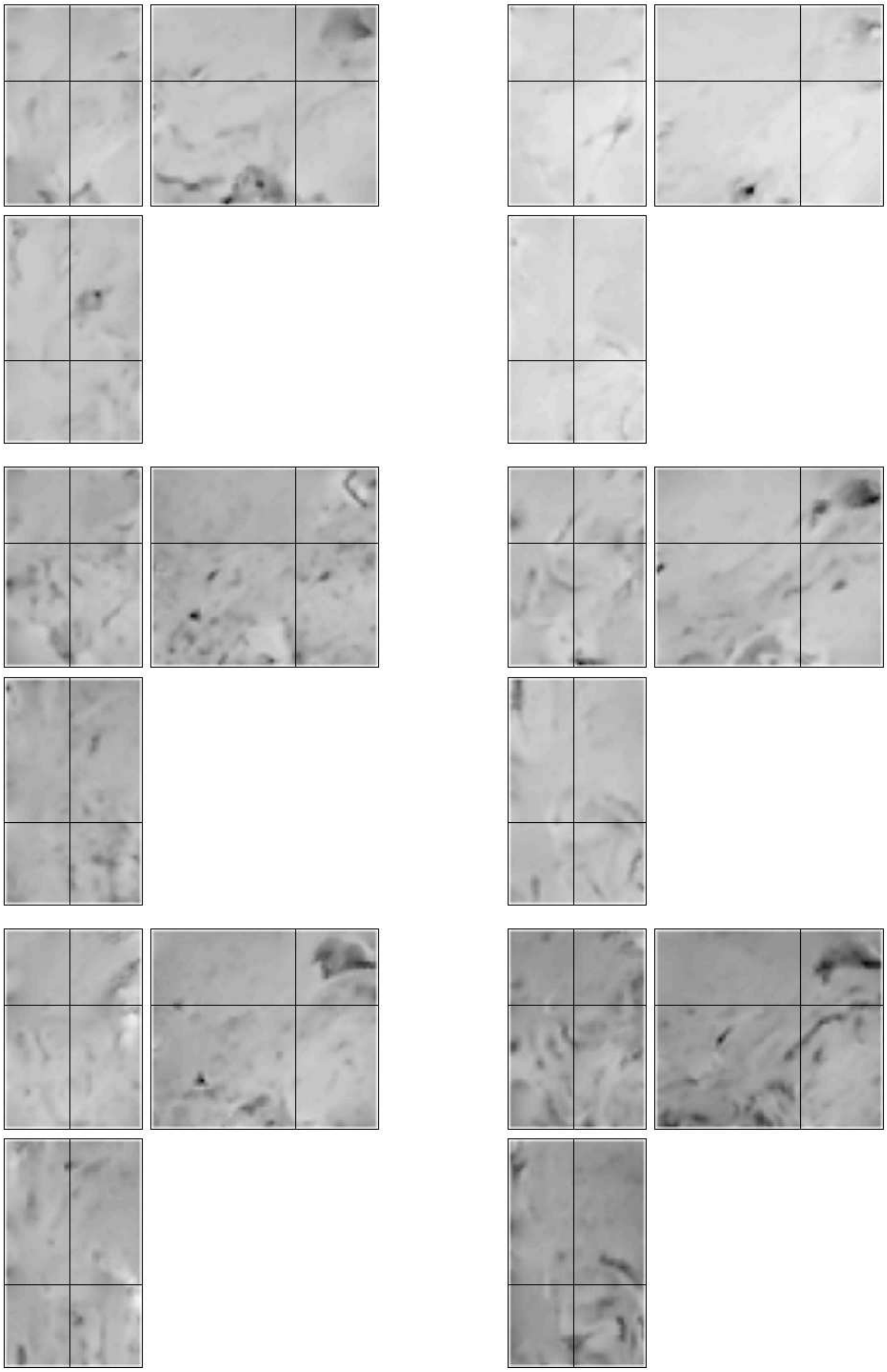

Figure 6.15: A selection of orthogonal views through the Jacobian determinant fields from the same images shown in Figure 6.14.

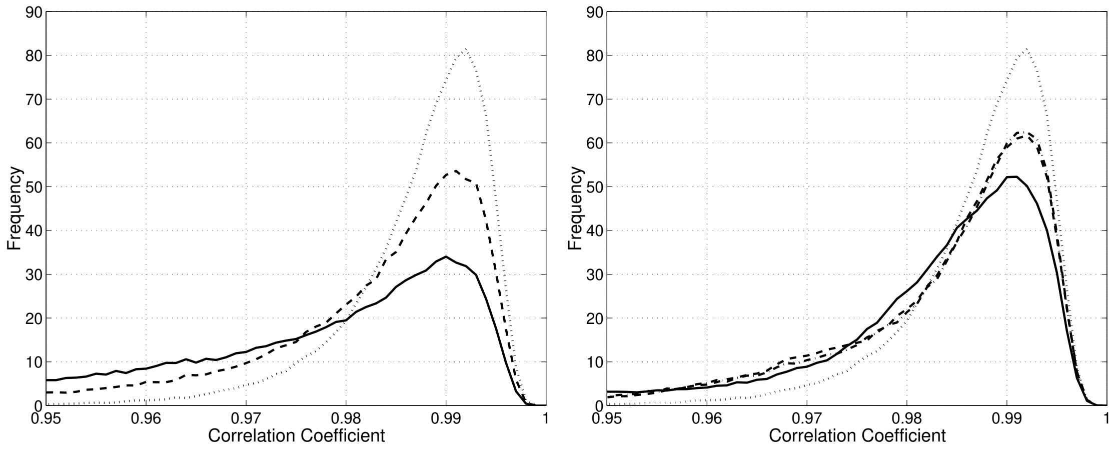

Figure 6.16: Histogram of correlation coefficients taken over the whole volume volumes (144495 voxels). The dotted line in both panels is a histogram obtained from simulated random data, and represents the histogram that would be expected if the Jacobian determinants were perfectly normally distributed. The left panel shows histograms of Q-Q plot correlation coefficients obtained from the unsmoothed Jacobian determinants, both without (solid line) and with taking logs. The panel on the right shows histograms of Q-Q plot correlation coefficients after taking 8mm FWHM Gaussian smoothing the Jacobian determinants. The solid line is for the smoothed determinants, the dot-dash line represents determinants that were smoothed before taking logs, and the dash-dash line was generated by taking logs and then smoothing.

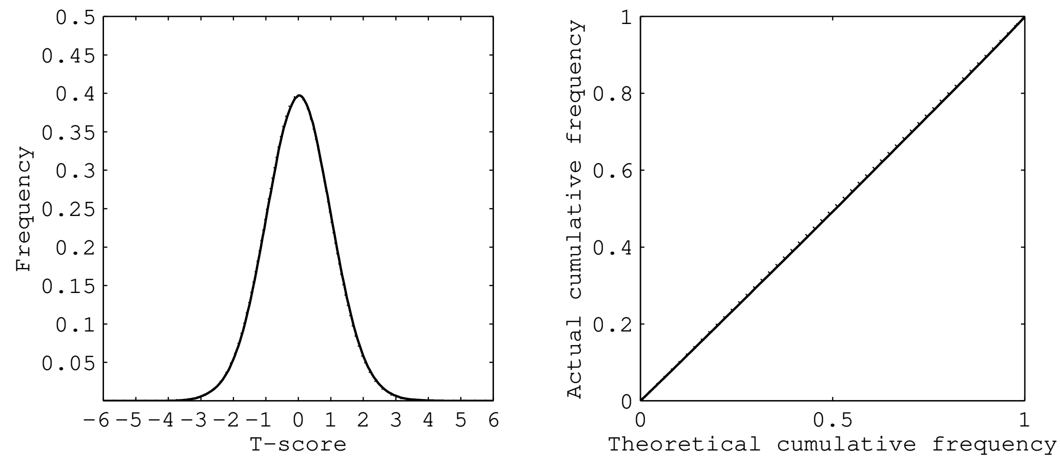

Figure 6.17: Histograms of t-scores from randomly generated tests involving comparing a group of 12 subjects with a group of 46. Left: the histogram resulting from 1000 random voxelwise t tests is plotted as a solid line, and the probability density function of the Students t distribution for 56 degrees of freedom is shown by the dotted line. For each test, a group of 12 subjects was compared with the remaining 46. Right: The same data plotted as a cumulative histogram against the theoretical cumulative histogram.

hidden when making group comparisons, because the group averages tend to be more normally distributed because of the central-limit theorem. This does not apply when comparing single subjects, so the validity of the results is strongly influenced by any non-normality of the data.

The least valid results were obtained from t-tests on the original determinant fields. Slight improvements in the validity of the tests were obtained after smoothing the fields or taking logs. However, closer examination of the cumulative histograms showed that much higher rates of false positive results than expected would occur, even after these transformations.

6.4.4 Morphometry on Strain Tensors

TBM based on Jacobian determinants has just been illustrated. Although this method identifies relative volume differences, there are some shape differences that it can not identify. A potentially more powerful approach may involve using multivariate statistics on strain tensors derived from the deformation fields.

Again, the deformation fields from Section 6.4.2 were used in the evaluations. Hencky and Biot strain tensors fields were derived from the deformations and written to disk. These tensor fields were computed from the deformations in much the same way as the Jacobian determinants used previously. Each element in a tensor field was computed from the average of the strain tensors of the neighbouring tetrahedra.

Distribution of Tensor Field Elements

Ideally, the tensor elements on which subsequent multivariate analysis is performed should be multi-normally distributed. The parameterisation of the distribution should be related to the registration scheme used to estimate the deformations. Because the schemes in Chapter 4 attempt to enforce log-normal distributions on the relative lengths, areas and volumes, then it should bias the elements of the Hencky tensor fields derived from the deformations to be multi-normally distributed.

This section examines the distributions of various tensor fields derived from the deformations, by looking at histograms of the correlation coefficients obtained from Q-Q plots. For the analysis, the tensor fields can be considered as consisting of six scalar fields. Univariate normality of each scalar field was assessed by the correlation coefficients of Q-Q plots, and the results combined into single histograms. First of all the the distributions of unsmoothed Biot and Hencky strain tensors were examined. This was followed by convolving the the Biot strain tensor fields with an 8mm FWHM Gaussian kernel, to identify any increases in normality that would be expected because of the central limit theorem. Finally, the smoothed Biot strain tensor fields were converted to Hencky tensors, and their distributions examined. All the resulting histograms are plotted in Figure 6.19.

As expected, smoothing made the data more normally distributed. Similarly, taking logs of the unsmoothed data also increased normality. However, one unexpected result is that the smoothed Hencky tensor data and the smoothed Biot tensor data transformed to Hencky tensors appears to be slightly less normally distributed than the smoothed Biot tensor data.

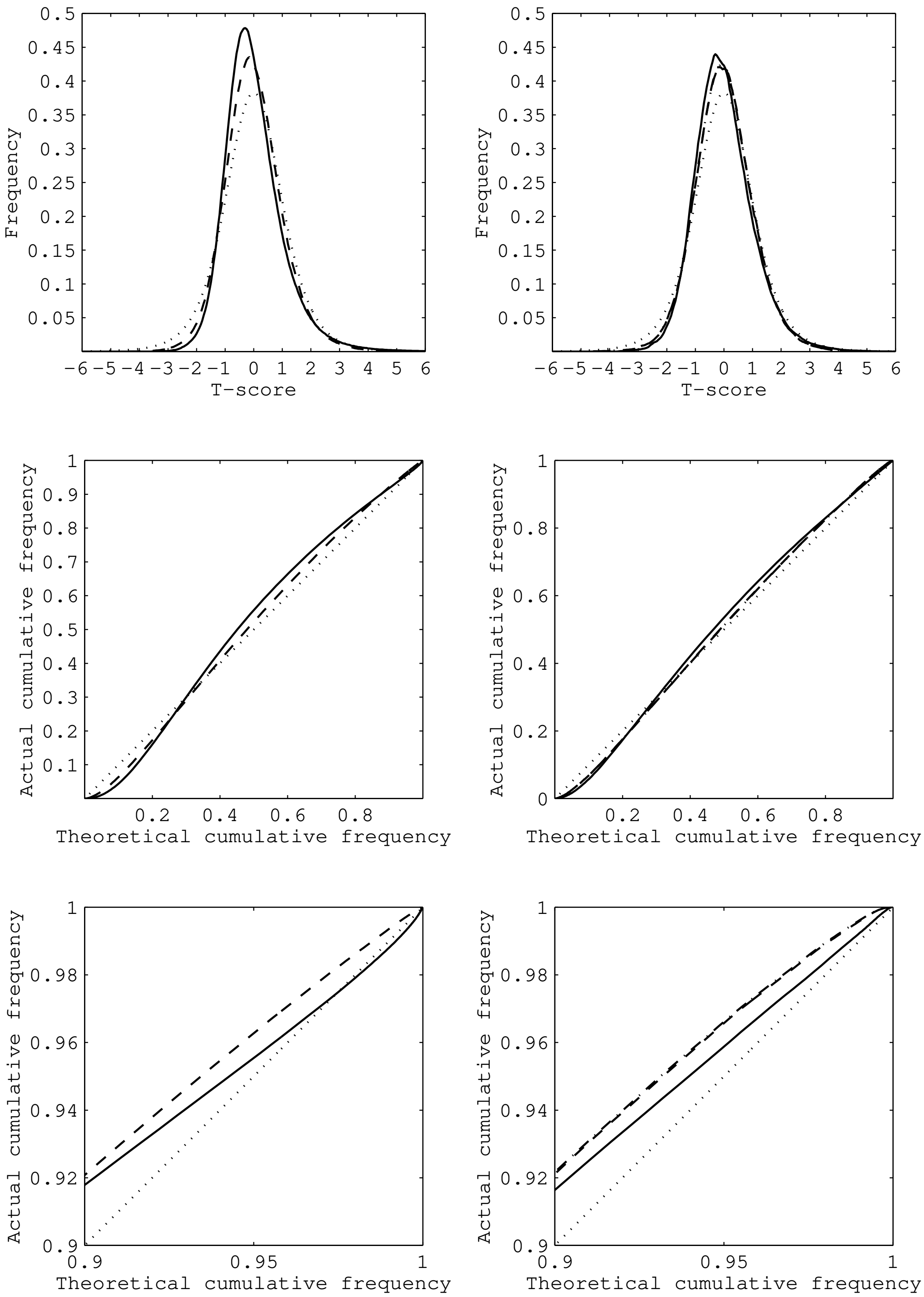

Figure 6.18: Histograms of t-scores from comparing single subjects with a group of 57. The top row shows histograms of the computed t-statistics, whereas the centre and bottom rows show plots of theoretical versus measured cumulative histograms. In all plots, the dotted line shows the theoretical t-distribution for 56 degrees of freedom. Left: Statistics derived from Jacobian determinants (solid line), and logged determinants (dashed line). Right: Statistics derived from 8mm smoothed Jacobian determinants (solid line), logged smoothed determinants (dash-dash line) and smoothed logged determinants (dot-dash line).

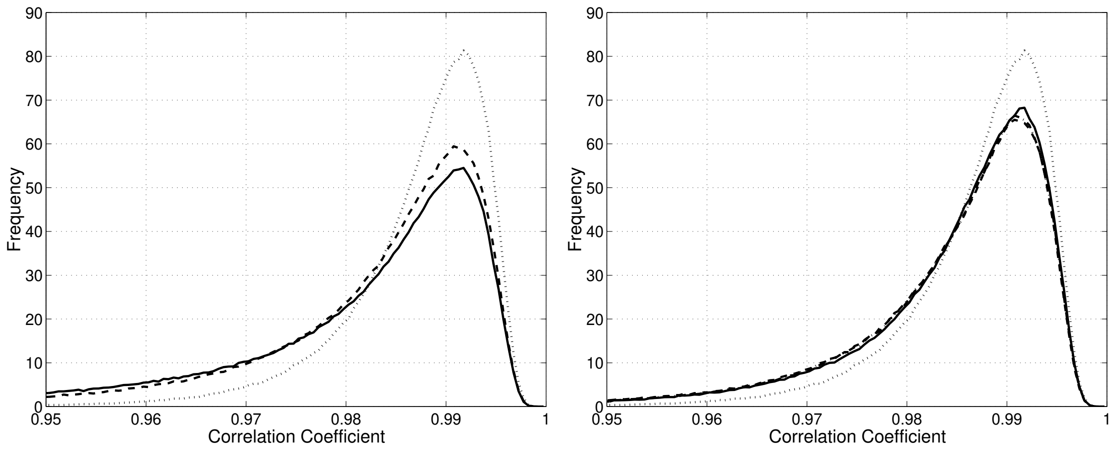

Figure 6.19: Histogram of Q-Q plot correlation coefficients taken over tensor fields. The dotted line in both panels is a histogram obtained from simulated random data, and represents the histogram that would be expected if the tensor elements were perfectly normally distributed. The left panel shows histograms of Q-Q plot correlation coefficients obtained from the all six unsmoothed sets of strain tensor elements. Histograms from Biot (solid line) and Hencky tensors (dashed line) are shown. The panel on the right shows histograms of Q-Q plot correlation coefficients of 8mm FWHM smoothed tensor fields. The solid line is derived from the Biot tensor field, the dot-dash line is from Hencky tensors that were derived from smoothed Biot tensors, and the dash-dash line was generated from 8mm smoothed Hencky tensor fields.

False Positive Rates from MANCOVA

This section assesses the rate of false positives that are likely to occur when multivariate methods are used to compare strain tensors among groups. The tests were for the worst case scenario, and involved comparing each of the 58 tensor fields with the remaining 57. A voxelwise Wilk's Λ statistic was computed at each point in the tensor fields. The same transformations of the tensor fields were used for the analysis as were used in Section 6.4.4, and the results are presented in Figure 6.20.

The plots of cumulative histograms show that a much higher rate of false positives than theoretically predicted would be obtained when comparing single subjects with groups. Smoothing the fields slightly reduces this rate of false positives, but smoothing by more than 8mm FWHM would be needed to bring the rate of false positive results to an acceptable rate.

6.5 Discussion

The morphometric methods developed here all attempt to automatically identify neuroanatomical differences, either from features of spatially normalised images, or from deformation fields that encode information about image shapes. Rather than focusing on particular structures, one important aspect of these methods is that the entire brain can be examined in a balanced way.

Two different types of results are obtained from the methods. One form localises structural differences among groups of brains (VBM and TBM) by producing SPMs of local shape differ-

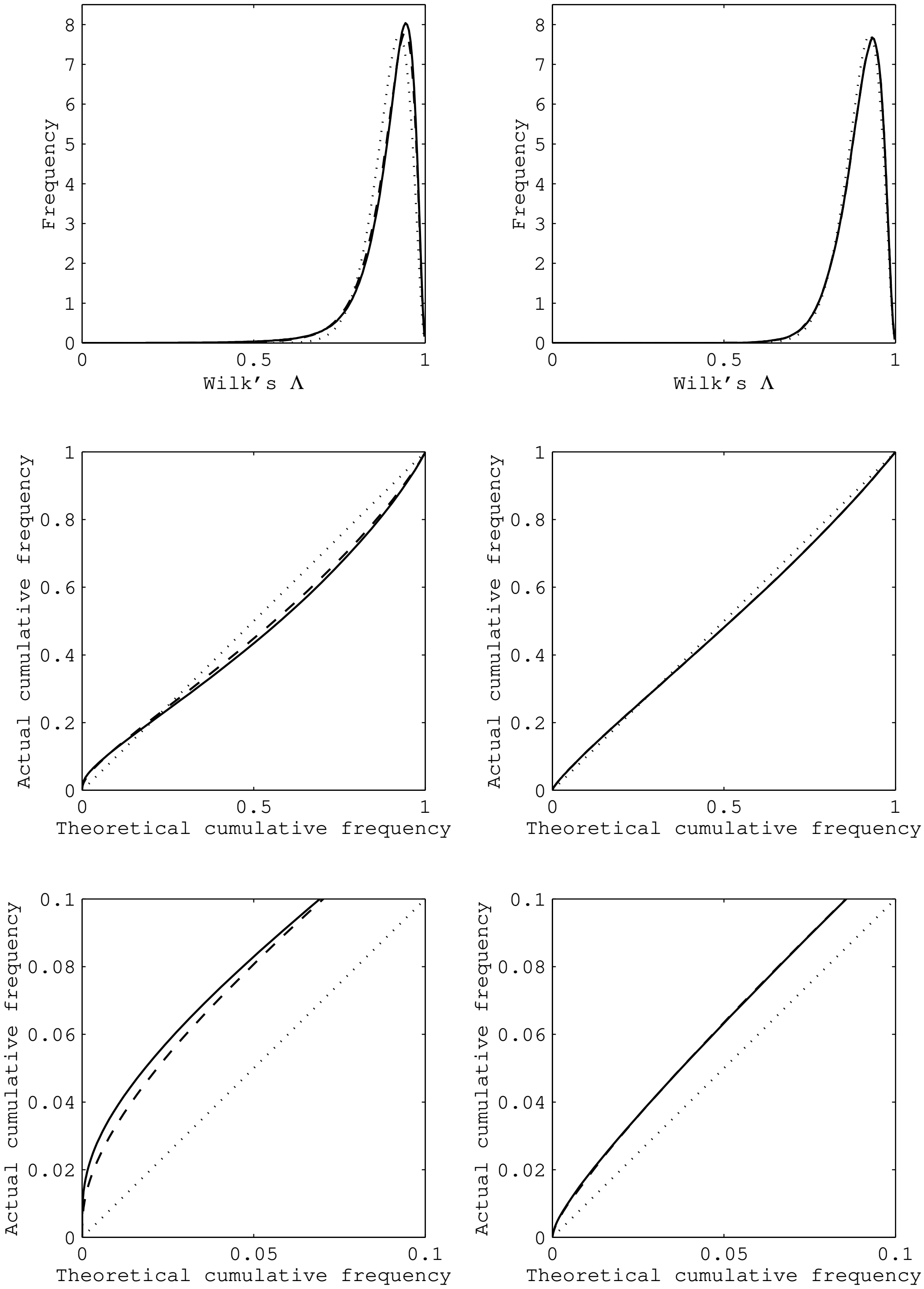

Figure 6.20: Histograms of Wilk's Λ statistics from comparing single subjects with a group of 57. The top row shows histograms of the computed Wilk's Λ statistics, whereas the centre and bottom rows show different parts of the cumulative histograms. In all plots, the dotted line shows the theoretical Wilk's Λ distributions. Left: Wilk's Λ statistics derived from Biot strain tensors (solid line), and Hencky strain tensors (dashed line). Right: Wilk's Λ statistics derived from 8mm smoothed Biot strain tensor field (solid line) and smoothed Biot strain tensor field converted to Hencky tensor field (dashed line).

ences. The other form performs a single, global multivariate test to compare the images (DBM). In principle, the difference between these approaches need not be quite so distinct. Rather than basing the tests either on information at each voxel, or on information from the whole brain, it is easy to see that forms of morphometry could be based on regional analyses. For example, DBM could be done region by region (after factoring out positional information), or TBM could be applied such that the tests include information from surrounding voxels. Both of these approaches should produce very similar results. The regions could be based on pre-defined structures on the template image. Alternatively, they could involve some weighted combination of information from the neighbourhood of each voxel, with weights based on a function such as a Gaussian kernel, or some other windowing function.

The different forms of morphometry have their own advantages and disadvantages, and the optimum approach may depend on the types of structural difference expected among the data. Where there are global patterns of difference, then the global approaches (that do not produce a SPM) may be more powerful, as they can model covariances between shapes of different structures. In contrast, the SPM approaches may be preferable where discrete focal differences are expected. The main disadvantages of global testing methods (as opposed to SPM methods), is that there are generally far more variables for each brain than there are brains in a study. This implies that covariances between all variables can not be computed, so some form of dimensionality reduction (or regularisation) is required, leading to inevitable data loss.

The most powerful forms of morphometry may ultimately depend on how well covariances can be modelled using as few parameters as possible. The current implementation of DBM effectively models covariance matrices by their most prominent eigenvectors, whereas the VBM implementation (and the univariate TBM) models a diagonal covariance matrix[5]. Voxelwise and regionwise multivariate approaches model banded covariance matrices.

A related issue is about how to normalise morphometric measures in order to produce the most useful results. This is particularly true for voxelwise approaches, because relationships between other, more global, factors are not modelled. For example, consider a group of subjects with smaller hippocampi, but also smaller brains. Should the reduced hippocampi be considered significantly different if they correlate with a smaller brain? This size difference could also relate to lengths or volumes of temporal lobe, or to a whole number of other measurements.

Another problem concerns visualising and communicating the results of morphometric tests. Three dimensional volumes are quite difficult to visualise, especially within the limited space of most journals. However, the results of morphometric tests are often vector or tensor fields. These are quite difficult to visualise in two dimensions, but in three dimensions the problem becomes much worse. The more useful results of global morphometric methods are the canonical variates that characterise the differences, and these are often some form of three dimensional vector or tensor field. In comparison, differences localised by voxelwise methods can be relatively easily presented as a SPM. Although the reasons may appear trivial, voxelwise approaches will probably come to dominate because their results can be explained and presented much more easily.

A potential future application for these approaches may be to automatically identify and locate unusual or abnormal brain regions in single subjects. This may involve comparing single subject images with images derived from a large group of controls. If the objective is to find anything

[5] With a few assumptions about some of the off-diagonal elements, which are modelled by Gaussian random field theory

unusual, then the most powerful methods may involve multivariate methods that combine VBM and TBM. Provided that the database of controls is sufficiently large, then the more variables that are included in a multivariate voxelwise test, then the more powerful the tests should be at localising differences.

When parametric statistical methods are used, comparisons between single subjects and whole groups may not be as reliable as comparisons between groups containing several subjects. It may be necessary to resort to non-parametric methods for such cases. The alternative may involve developing suitable transforms for the data that render them more normally distributed. The distributions of estimated deformation fields can be partially controlled by the amount of regularisation imposed on the warping model used. For example, approaches similar to those described in Section 7.3 would attempt to find the optimal amount of regularisation such that the distribution of the residuals is as close to normal as possible. This not only includes the residual difference between image and template, but also the distributions of the parameter estimates around their expected mean.

Other factors also influence the validity and sensitivity of the different morphometric approaches. In particular, the warping method, used to register the images to the same stereotactic space, has a significant influence on the results. For DBM and TBM, the statistical tests are based entirely on the deformation fields produced by the warping methods. The warping methods developed in this thesis are all nonlinear optimisation procedures, and therefore can be susceptible to reaching local minima, and hence non-optimal solutions. These local minima have negative consequences for subsequent statistical tests, as the estimated shapes do not reflect the true shapes of anatomical structures. Other problems occur when warping brains containing severe pathologies. For example, if a brain contains features that are not present in the template image, then an accurate match can not be achieved. The effects of this mismatch may also propagate to other brain regions because of the inherent smoothness of the deformation fields. Much more work is necessary in order to develop warping methods that can model the various forms of severe pathology that may be encountered.

In VBM, nonlinear spatial normalisation results in the volumes of certain brain regions increasing, whereas others decrease (Goldszal *et al.*, 1998). This has implications for the interpretation of what VBM actually tests. The objective of VBM is to identify regional differences in the concentration of a particular tissue (grey or white matter). In order to preserve the actual amounts of grey matter within each structure, a further processing step can be incorporated that multiplies the partitioned images by the relative voxel volumes. These relative volumes are simply the Jacobian determinants of the deformation field. VBM can be thought of as comparing the *relative concentration* of grey matter (i.e., the proportion of grey matter to other tissue types within a region). With the adjustment for volume change, VBM compares the *absolute amount* of grey matter in different regions. As mentioned in Section 6.2.1, if spatial normalisation was perfect, then no grey matter differences would be observed if a volume change adjustment was not applied. In this instance, all the information would be in the deformation fields and would be tested using TBM. However, if the spatial normalisation is only removing global differences in brain shape, the results of VBM show relative grey matter concentration differences. As faster and more precise registration methods emerge, then a TBM volume change adjustment may become more important. It is envisaged that, by incorporating such a correction, a continuum will arise with simple VBM (with low resolution spatial normalisation) at one end of the methodological spectrum, and statistical tests based on Jacobian determinants (with high resolution spatial normalisation) at the other.

Another perspective on what VBM compares can be obtained by considering how a similar analysis would be done using volumes of interest (VOIs). To simplify the analogy, consider that the smoothing kernel is the shape of a sphere (values of one inside, and zero outside) rather than a 3D Gaussian point spread function. After convolving an image with this kernel, each voxel in the smoothed image will contain a count of the grey matter voxels from the surrounding spherical VOI. Now consider the effects of spatial normalisation, and where the voxels within each VOI come from in the original grey matter images. The spheres can be thought of as projected on to the original anatomy, but in doing so, their shapes and sizes will be distorted. Without multiplying by the relative voxel sizes, what would be measured would be the proportion of grey matter within each projected VOI (relative to other tissue types). With the multiplication, the total amount of grey matter within the VOI is measured.

Chapter 7

Discussion

Contents

7.1 Original Contributions

This thesis has described a number of computational tools that can be used for both functional image analysis, and computational neuro-anatomy. These include various registration approaches, and image segmentation. Many of the methods have been incorporated into the SPM99 package, are already used widely by a large number of researchers in the field of functional imaging, and are also increasingly used in studies of morphometry.

A number of original contributions have been made within this thesis. Some of these may be completely original (or independently re-invented), but others simply involve combining different parts of pre-existing methods in a new way. For example, most of the techniques developed in Chapters 2 and 3 are significant improvements to methods previously presented by Friston *et. al.* (1995c). Within Chapter 2, the original contributions were predominantly in Section 2.6, which described a new approach to between modality registration. Although the approach is not as flexible as mutual information registration methods, it generally performs well when the various assumptions are satisfied.

The main novel component of Chapter 3 is the fast algorithm for computing the Hessian matrix (ignoring second derivatives) for the Gauss-Newton optimisation scheme. A simplified version of this same algorithm is also used for estimating intensity modulation fields in Chapter 5. In addition, Friston's original registration scheme (Friston *et al.*, 1995c) was placed within an Empirical Bayesian framework.

Chapter 4 introduced two main original ideas. The first is the use of prior probability distributions that are the same for both the forward and inverse deformations. The more pure version assumes log-normal distributions for length, area and volume changes. Because the computations required for this version would be prohibitively slow, a much faster approximation was also devised. The second main idea involves methods for making deformations internally consistent by simultaneously matching several images at the same time. A method for inverting deformation fields was also included in order to facilitate combining deformations with the inverses of other deformations.

A number of novel ideas can be found in Chapter 5. These include the use of prior probability maps to assist classification, and some aspects of the intensity non-uniformity correction. These include the use of a third-order regularisation scheme, and a rapid algorithm similar to that of Chapter 3.

The first original contribution in Chapter 6 are the evaluations of the applicability of the statistical parametric modelling component of SPM99 to voxel-based morphometry. The deformation-based morphometry section describes the use of standard multivariate statistical methods for comparing deformation fields among groups. Finally, a method of identifying regional shape differences is presented, which is based on applying multi-variate statistical tests to strain tensors.

A few recurring themes are encountered within this thesis. The first of these is about achieving consistency, not only in terms of between different data-sets, but also between different processing steps. A certain amount of modularity may need to be sacrificed in order to achieve optimal consistency between procedures. A second theme is the problem of assigning optimal hyper-parameters (weights) to different components of the various procedures. Although none of the chapters so far have presented any good methods for doing this, a section is included here that may provide a useful framework in which future methods may achieve this end.

7.2 Modularity

Currently, different image processing components are often thought of as distinct modules, whereby one set of processing is completed before another one begins. However, this may not be the best approach. A few examples will now be given that suggest that the whole should be considered greater than the sum of the parts.

Chapter 2, includes a section on between modality image registration. Part of the method includes simultaneously spatially normalising a pair of source images to corresponding templates, while maintaining a rigid body relationship between the source images. This component could be extended to include nonlinear warping in addition to the current zooms and shears. In theory, this should provide a framework for improved rigid registration, as well as better estimates of the deformations that spatially normalise the images. One of the themes of the thesis is about improving internal consistency, of which this is another example, as all the information would be included within the same internally consistent registration model. Another matching criterion such as mutual information could also be recruited at the same time, which should effectively bring additional information to the problem, further enhancing the final result. The more useful information that is included, then the better the parameter estimates should be – providing that the optimal weighting for the different components can be found.

Section 2.6 also suggests another move away from a modular image processing stream. Simultaneous rigid registration and spatial normalisation has just been mentioned, but the section also includes a segmentation component. Segmentation can be performed more efficiently when the prior probability of each voxel belonging to particular tissue classes is known, and in order to achieve this, the images need to be in register with prior probability images. Although the chapter only describes using affine registration to effect this mapping, more accurate results could be expected after nonlinear matching. Conversely, segmented images can be useful for image registration. For example, a number of nonlinear image registration methods are based on matching brain surfaces together, either in 3D (Thompson & Toga, 1996), or in 2D on a flattened cortical surface (see Drury *et. al.* (1999) or Thompson and Toga (1999)). It seems that segmentation and spatial normalisation could be performed most accurately when both done simultaneously.

Many issues relating to morphometry are also applicable to image warping. Image registration methods generally involve some model of how brains should deform. These can generally be thought of as some form of multi-dimensional probability distribution. Similarly, morphometric methods also require models of brain shape variability. It is envisaged that future work on morphometry should develop in concurrence with the methods used for estimating deformations. The parameter distributions imposed upon the deformations by the registration method could be used in the morphometry studies. Similarly, knowledge of the variability of brain shapes obtained from morphometry could be used as *a priori* information for Bayesian image registration methods. Both fields would clearly benefit by having a compact and concise representation of the anatomical variability of brains.

The link between warping methods and morphometric methods may be even tighter. Not only do they both need similar models of shape variability, but they also need the hyper-parameters that describe the amount of variability. These can not be estimated properly by warping a number of images to the same template, followed by simply computing the variance of some features of the resulting deformation fields. This is because the amount of regularisation used by the warping model would greatly influence the estimated result. In order to obtain unbiased estimates of shape variability, this estimation would need to be a component of the warping model. The next section elaborates on this theme.

7.3 Hyper-parameter estimation

An ongoing argument in the field of image registration is something to the effect of "if you use too many parameters in a registration model, then the model will be overfitted", versus "if not enough parameters are used, then the images can not be registered properly". Increasing the relative weighting of the prior potential component effectively reduces the search space of the optimisation problem, so this argument can be rephrased in terms of the relative weights used for the likelihood and prior potential components. This section elaborates on a possible approach that may help to resolve these issues.

This thesis contains many examples of problems where a model is used, but optimal weights for determining the best parameter estimates are not known. These problems occur in situations where the optimised functions consist of several terms, where each term is weighted differently. For example, in Section 2.6, more than one pair of images is registered simultaneously. Optimal results are obtained only when the contribution from each image pair to the cost function, is properly balanced. Another area is in image warping, in which it is necessary to know the optimal balance between prior and likelihood potential (see Sections 3.2.1 and 4.2.2). Also, some method of estimating the *a priori* variability of the non-uniformity is needed to improve the segmentation (see Section 5.2.3). In order to obtain the most accurate results, optimal weighting should be assigned to the different components of the problems.

One method of choosing weights empirically is by using L-curves, whereby a model is fitted using a range of different weights. After all the fits are complete, the logs of one cost function term are plotted against the logs of the other, over the range of different weights. The weighting is deemed optimal for values that fall close to the inflection of the plot. This method is time consuming, and is not practical for complex models where there may be many different terms. Fortunately, there may be more practical solutions to the problem of hyper-parameter estimation, which would allow the unknown weights, in addition to the parameters of interest, to be estimated.

So far, any hyper-parameter estimation in this thesis has been performed in a pragmatic, but rather ad hoc manner. In particular, the estimation of the degrees of freedom in the models has involved slightly questionable methods. There now follows a more rigorous framework for hyper-parameter estimation that could be used for future work. As an illustration, consider the iterative parameter updating scheme from Chapter 3. After rearranging and a slight modification to the notation[1], Eqn. 3.6 can be written as:

$$\mathbf{q}^{(n+1)} = \mathbf{q}^{(n)} - \left(\lambda_1 \mathbf{A_1}^T \mathbf{A_1} + \lambda_2 \mathbf{C_0}^{-1}\right)^{-1} \left(\lambda_1 \mathbf{A_1}^T \mathbf{b_1} + \lambda_2 \mathbf{C_0}^{-1} \left(\mathbf{q}^{(n)} - \mathbf{q_0}\right)\right) \tag{7.1}$$

Obtaining the best balance between minimising the residual squared difference between the images, and maximising some form of deformation smoothness, can be considered as finding some optimum estimates for λ_1 and λ_2. It is possible to do this from the data using an empirical Bayes approach, providing that a few assumptions about the forms of the errors are made. In this example, both the parameters themselves and the residual difference between the image pair are assumed to be independent and identically distributed. An algorithm similar to expectation maximisation (EM) can be used for estimating the hyper-parameters.

EM is an iterative approach that alternates between two steps: the expectation (E) step and the maximisation (M) step (see Chapter 5). The algorithm is initialised by assigning some (non-zero) starting estimates to the hyper-parameters (λ_1 and λ_2). The first E step involves computing the expectation of a set of unobserved parameters while holding the hyper-parameters constant. Within a Bayesian registration scheme, this would involve computing a minimum variance estimate of the parameters ($\mathbf{q}$). This is where this approach deviates from a pure EM algorithm, as a MAP estimate of $\mathbf{q}$ is used instead, as it is much faster to compute. This can be achieved by repeated application of Eqn. 7.1 until convergence is attained.

The M step involves re-estimating the hyper-parameters such that the likelihood of observing the data is maximised, while holding the parameters fixed. The hyper-parameters are computed from the reciprocals of variance estimates derived from the observations and parameters. This step involves estimating the residual variance attributed to the different parts of the model, divided by the appropriate degrees of freedom, computed by:

$$\begin{aligned} p_1 &= m_1 - tr\left(\left(\lambda_1 \mathbf{A_1}^T \mathbf{A_1} + \lambda_2 \mathbf{C_0}^{-1}\right)^{-1} \lambda_1 \mathbf{A_1}^T \mathbf{A_1}\right) \\ p_2 &= m_2 - tr\left(\left(\lambda_1 \mathbf{A_1}^T \mathbf{A_1} + \lambda_2 \mathbf{C_0}^{-1}\right)^{-1} \lambda_2 \mathbf{C_0}^{-1}\right) \end{aligned} \tag{7.2}$$

where m_1 is the number of rows an $\mathbf{A_1}$, and m_2 is the number of columns (also the same as the

[1] $\mathbf{C_0}$, $\mathbf{A}$, $\mathbf{b}$ and σ^2 become $\lambda_2\mathbf{C_0}$, $\mathbf{A_1}$, $\mathbf{b_1}$ and $1/\lambda_1$ respectively.

number of rows or columns of ${\mathbf{C_0}}^{-1}$. The new estimates of λ_1 and λ_2 are then:

$$\begin{aligned} \lambda_1 &= \frac{p_1}{\left(\mathbf{b_1} - \mathbf{A_1 q}\right)^T \left(\mathbf{b_1} - \mathbf{A_1 q}\right)} \\ \lambda_2 &= \frac{p_2}{\left(\mathbf{q} - \mathbf{q_0}\right)^T {\mathbf{C_0}}^{-1} \left(\mathbf{q} - \mathbf{q_0}\right)} \end{aligned} \tag{7.3}$$

This alternating EM procedure is continued until a stopping criterion is reached.

More generally, Eqn. 7.1 can be expressed as:

$$\mathbf{q}^{(n+1)} = \mathbf{q}^{(n)} - \left(\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A}\right)\mathbf{A}^T\mathbf{C}^{-1}\mathbf{b} \tag{7.4}$$

where $\mathbf{A} \equiv \begin{bmatrix}\mathbf{A}\\ \mathbf{I}\end{bmatrix}$, $\mathbf{b} \equiv \begin{bmatrix}\mathbf{b_1}\\ \mathbf{q}^{(n)} - \mathbf{q_0}\end{bmatrix}$ and $\mathbf{C} \equiv \begin{bmatrix}\lambda_1^{-1}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \lambda_2^{-1}\mathbf{C_0}\end{bmatrix}$.

This EM approach maximises the likelihood of the data given the hyper-parameters:

$$p(\mathbf{b}|\lambda) = \sqrt{\frac{|\mathbf{A}^T\mathbf{A}|}{(2\pi)^{m-n}|\mathbf{C}||\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A}|}}\, e^{-\frac{1}{2}(\mathbf{b}-\mathbf{Aq})^T\mathbf{C}^{-1}(\mathbf{b}-\mathbf{Aq})} \tag{7.5}$$

This is equivalent to minimising the *restricted maximum likelihood* (REML) objective function (Patterson & Thompson, 1971; Harville, 1974):

$$-\log(p(\mathbf{b}|\lambda)) = \frac{1}{2}\log|\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A}| + \frac{1}{2}\log|\mathbf{C}| + \frac{1}{2}\left(\mathbf{b} - \mathbf{Aq}\right)\mathbf{C}^{-1}\left(\mathbf{b} - \mathbf{Aq}\right) + const \tag{7.6}$$

There are a number of more general EM approaches for solving these problems (Harville, 1977), but their complete description is beyond the scope of this thesis. The idea would be to estimate coefficients for a linear combination of basis functions that parameterise $\mathbf{C}$ such that the REML objective function is optimised[2]. For image registration methods with nonlinear priors, such as those described in Chapter 4, the covariance matrices would be replaced by inverse Hessian matrices (second partial derivatives of the prior potential with respect to changes in the parameters - see Section 2.4). An extension of the scheme described in Section 4.3.3 could allow spatially varying estimates of λ_1 and λ_2 to be obtained. A more useful model of spatial variability would allow anisotropic variability, with different amounts of distortion in different directions, and possibly also model some covariance among distortions in different directions, and also among distortions of neighbouring tetrahedra. This could be based on a multi-normal model of Hencky tensor elements (see Section 6.4). With more complex models of structural variability, and lots of example datasets, it should be possible for algorithms to empirically learn to represent neuroanatomical variability. Given improved variability models, the estimates of the deformations should become more accurate, which, in turn should lead to better representations of neuroanatomical variability.

[2] In the current example, the coefficients are λ_1^{-1} and λ_2^{-1}, with basis functions $\begin{bmatrix}\mathbf{I} & 0\\ 0 & 0\end{bmatrix}$ and $\begin{bmatrix}0 & 0\\ 0 & \mathbf{C_0}\end{bmatrix}$.

Acknowledgments

First of all, I am especially grateful to my supervisor, Karl Friston, for his constant support, guidance and encouragement. Karl's work provided a good foundation from which many of the ideas included in this thesis were derived from. I think that he and Richard Frackowiak should also be thanked for twisting my arm and pursuading me to undertake a part-time PhD in the first place.

I also owe a great deal of gratitude to a number of other friends and colleagues at the Wellcome Department of Cognitive Neurology – both past and present. Those especially deserving of a mention include Andrew Holmes, Chloe Hutton, Christophe Phillips, Jean-Baptiste Poline, Jesper Andersson, Oliver Josephs, Ralf Deichman, and Stefan Kiebel. I have learned a great deal from all of them – and from their books that I was constantly borrowing.

Other WDCN members have also been immensely helpful by allowing me to use some of their data – either knowingly or unknowingly. Catriona Good, Eleanor Maguire and Ingrid Johnsrude have donated particularly large numbers of images.

Thanks are also in order for a number of people outside the WDCN. Our neighbours, Nick Fox and Bill Crumb, of the Dementia Research Group have provided some especially useful serial images. I also owe thanks to Hava Lester who tought me a great deal, both from conversation and from the large number of important references for warping methods that she sent me.

Further afield, Keith Worsley has provided much more of an input to this thesis than he probably realises. I would also like to thank all the other people at McGill who made me feel so welcome during my three month period at Montréal Neurological Institute a few years ago.

This research was supported by the Wellcome Trust. Finally, I have to thank Steve, Suren, Isabelle and Ann Marie – mostly for their support and encouragement – but also because they want a mention in this thesis.

References

Amit, Y., Grenander, U., & Piccioni, M. 1991. Structural Image Restoration through Deformable Templates. *Journal of the American Statistical Association*, **86**, 376–387.

Andersson, J. L. R., Vagnhammar, B. E., & Schneider, H. 1995. Accurate Attenuation Correction Despite Movement during PET Imaging. *Journal of Nuclear Medicine*, **36**(4), 670–678.

Barnden, L., Kwiatek, R., Lau, Y., Hutton, B., Thurfjell, L., Pile, K., & Rowe, C. 2000. Validation of Fully Automatic Brain SPECT to MR Co-registration. *European Journal of Nuclear Medicine*, **27**, 147–154.

Bookstein, F. L. 1989. Principal Warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans Pattern Anal Machine Intelligence*, **11**(6), 567–585.

Bookstein, F. L. 1997a. Landmark Methods for Forms Without Landmarks: Morphometrics of Group Differences in Outline Shape. *Medical Image Analysis*, **1**(3), 225–243.

Bookstein, F. L. 1997b. Quadratic Variation of Deformations. *Pages 15–28 of:* Duncan, J., & Gindi, G. (eds), *Information Processing in Medical Imaging*. Berlin, Heidelberg, New York: Springer-Verlag.

Bookstein, F. L. 1999. *Brain Warping*. Academic Press. Chap. 10, pages 157–182.

Bullmore, E., Brammer, M., Rouleau, G., Everitt, B., Simmons, A., Sharma, T., Frangou, S., Murray, R., & Dunn, G. 1995. Computerized Brain Tissue Classification of Magnetic Resonance Images: A New Approach to the Problem of Partial Volume Artifact. *NeuroImage*, **2**, 133–147.

Bullmore, E., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., & Brammer, M. 1999. Global, Voxel, and Cluster Tests, by Theory and Permutation, for a Difference Between Two Groups of Structural MR Images of the Brain. *IEEE Transactions on Medical Imaging*, **18**(1), 32–42.

Cao, J., & Worsley, K. J. 1999. The Geometry of the Hotelling's T^2 Random Field with Applications to the Detection of Shape Changes. *Annals of Statistics*, **27**(3), 925–942.

Christensen, G. E. 1994. Deformable Shape Models for Anatomy. *Doctoral Thesis*.

Christensen, G. E. 1999. Consistent Linear Elastic Transformations for Image Matching. *Pages 224–237 of:* et al., A. Kuba (ed), *Information Processing in Medical Imaging*. Berlin, Heidelberg: Springer-Verlag.

Christensen, G. E., Rabbitt, R. D., & Miller, M. I. 1994. 3D Brain Mapping Using Using a Deformable Neuroanatomy. *Physics in Medicine and Biology*, **39**, 609–618.

Christensen, G. E., Rabbitt, R. D., Miller, M. I., Joshi, S. C., Grenander, U., Coogan, T. A., & Van Essen, D. C. 1995. Topological Properties of Smooth Anatomic Maps. *Pages 101–112 of:* Bizais, Y., Barillot, C., & Di Paola, R. (eds), *Information Processing in Medical Imaging*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Christensen, G. E., Rabbitt, R. D., & Miller, M. I. 1996. Deformable Templates Using Large Deformation Kinematics. *IEEE Transactions on Image Processing*, **5**, 1435–1447.

Cocosco, C.A., Kollokian, V., Kwan, R.K.-S., & Evans, A.C. 1997. BrainWeb: Online Interface to a 3D MRI Simulated Brain Database. *NeuroImage*, **5**(4), S425.

Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., & Marchal, G. 1995. Automated Multi-Modality Image Registration Based on Information Theory. *Pages 263–274 of:* Bizais, Y., Barillot, C., & Di Paola, R. (eds), *Information Processing in Medical Imaging*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Collins, D. L., Peters, T. M., & Evans, A. C. 1994a. An automated 3D Non-linear image deformation procedure for determination of gross morphometric variability in human brain. *Proc. Conference on Visualisation in Biomedical Computing*, 180–190.

Collins, D. L., Neelin, P., Peters, T. M., & Evans, A. C. 1994b. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J. Comput. Assist. Tomogr.*, **18**, 192–205.

Collins, D.L., Zijdenbos, A.P., Kollokian, V., Sled, J.G., Kabani, N.J., Holmes, C.J., & Evans, A.C. 1998. Design and Construction of a Realistic Digital Brain Phantom. *IEEE Transactions on Medical Imaging*, **17**(3), 463–468.

Cox, R. W., & Jesmanowicz, A. 1999. Real-Time 3D Image Registration for Functional MRI. *Magnetic Resonance In Medicine*, **42**, 1014–1018.

Davatzikos, C. 1996. Spatial Normalization of 3D Images Using Deformable Models. *J. Comput. Assist. Tomogr.*, **20**(4), 656–665.

Drury, H. A., Van Essen, D. C., Corbetta, M., & Snyder, A. Z. 1999. *Brain Warping*. Academic Press. Chap. 19, pages 337–362.

Eddy, W. F., Fitzgerald, M., & Noll, D. C. 1996. Improved Image Registration by Using Fourier Interpolation. *Magnetic Resonance in Medicine*, **36**, 923–931.

Evans, A. C., Collins, D. L., & Milner, B. 1992. An MRI-based stereotactic atlas from 250 young normal subjects. *Soc. Neurosci. Abstr.*, **18**, 408.

Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., & Peters, T. M. 1993. 3D statistical neuroanatomical models from 305 MRI volumes. *Pages 1813–1817 of: Proc. IEEE-Nuclear Science Symposium and Medical Imaging Conference.*

Evans, A. C., Kamber, M., Collins, D. L., & Macdonald, D. 1994. An MRI-based probabilistic atlas of neuroanatomy. *Pages 263–274 of:* Shorvon, S., Fish, D., Andermann, F., Bydder,

G. M., & H, Stefan (eds), *Magnetic Resonance Scanning and Epilepsy.* NATO ASI Series A, Life Sciences, vol. 264. Plenum Press.

Fox, P. T. 1995. Spatial Normalization Origins: Objectives, Applications, and Alternatives. *Human Brain Mapping*, **3**, 161–164.

Freeborough, P. A., & Fox, N. C. 1998. Modelling Brain Deformations in Alzheimer Disease by Fluid Registration of Serial MR Images. *J. Comput. Assist. Tomogr.*, **22**(5), 838–843.

Friston, K. J., Holmes, A. P., Poline, J.-B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S. J., & Turner, R. 1995a. Analysis of fMRI Time Series Revisited. *NeuroImage*, **2**, 45–53.

Friston, K. J., Holmes, A. P., Poline, J.-B., Price, C. J., & Frith, C. D. 1995b. Detecting Activations in PET and fMRI: Levels of Inference and Power. *NeuroImage*, **4**, 223–235.

Friston, K. J., Ashburner, J., Frith, C. D., Poline, J.-B., Heather, J. D., & Frackowiak, R. S. J. 1995c. Spatial Registration and Normalization of Images. *Human Brain Mapping*, **2**, 165–189.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-B., Frith, C. D., & Frackowiak, R. S. J. 1995d. Statistical Parametric Maps in Functional Imaging: A General Linear Approach. *Human Brain Mapping*, **2**, 189–210.

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J., & Turner, R. 1996. Movement-Related Effects in fMRI Time-Series. *Magnetic Resonance in Medicine*, **35**, 346–355.

Gaser, C., Volz, H.-P., Kiebel, S., Riehemann, S., & Sauer, H. 1999. Detecting Structural Changes in Whole Brain Based on Nonlinear Deformations - Application to Schizophrenia Research. *NeuroImage*, **10**, 107–113.

Gee, J. C., & Bajcsy, R. K. 1999. *Brain Warping.* Academic Press. Chap. 11, pages 183–198.

Gee, J. C., Haynor, D. R., Le Briquer, L., & Bajcsy, R. K. 1997. Advances in Elastic Matching Theory and its Implementation. *In:* Cinquin, P., Kikinis, R., & Lavallee, S. (eds), *CVRMed-MRCAS'97.* Heidelberg: Springer-Verlag.

Geschwind, N., & Galaburda, A. M. 1987. *Cerebral Lateralization, Biological Mechanisms, Associations and Pathology.* Cambridge Mass.: MIT Press. Pages 21–34.

Goldszal, A. F., Davatzikos, C., Pham, D. L., Yan, M. X. H., Bryan, R. N., & Resnick, S. M. 1998. An Image-Processing System for Qualitative and Quantitative Volumetric Analysis of Brain Images. *Journal of Computer Assisted Tomography*, **22**(5), 827–837.

Goshtasby, A. 1987. Piecewise Cubic Mapping Functions for Image Registration. *Pattern Recognition*, **20**(5), 525–533.

Guéziec, A., & Hummel, R. 1995. Exploiting Triangulated Surface Extraction Using Tetrahedral Decomposition. *IEEE Transactions on Visualization and Computer Graphics*, **1**, 328–342.

Hajnal, J. V., Mayers, R., Oatridge, A., Schwieso, J. E., Young, J. R., & Bydder, G. M. 1994. Artifacts due to stimulus correlated motion in functional imaging of the brain. *Magnetic Resonance in Medicine*, **31**, 289–291.

Hartigan, J. A. 1975. *Clustering Algorithms.* New York: John Wiley & Sons, Inc. Pages 113–129.

Harville, D. A. 1974. Bayesian inference for variance components using only error contrasts. *Biometrika*, **61**, 383–385.

Harville, D. A. 1977. Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association*, **72**(358), 320–338.

Holmes, A. P., Blair, R. C., J, D. G. Watson, & Ford, I. 1996. Non-Parametric Analysis of Statistic Images From Functional Mapping Experiments. *Journal of Cerebral Blood Flow and Metabolism*, **16**, 7–22.

Jain, A. K. 1989. *Fundamentals of Digital Image Processing*. Pretence-Hall.

Jezzard, P., & Clare, S. 1999. Source Distortion in Functional MRI Data. *Human Brain Mapping*, **8**(2), 80–85.

Johnson, R. A., & Wichern, D. W. 1998. *Applied Multivariate Statistical Analysis. Fourth Edition.* Upper Saddle River, NJ: Prentice-Hall.

Joshi, S. C., Miller, M. I., Christensen, G. E., Banerjee, A., Coogan, T. A., & Grenander, U. 1995. Hierarchical Brain Mapping via a Generalized Dirichlet Solutions for Mapping Brain Manifolds. *Proc. SPIE International Symposium on Optical Science, Engineering and Instrumentation.*

Kwan, R. K.-S., Evans, A. C., & Pike, G. B. 1996. An Extensible MRI Simulator for Post-Processing Evaluation. *Pages 135–140 of: Conference on Visualisation in Biomedical Computing.*

Le Briquer, L., & Gee, J. C. 1997. Design of a Statistical Model of Brain Shape. *Pages 477–482 of:* Duncan, J., & Gindi, G. (eds), *Information Processing in Medical Imaging*. Berlin, Heidelberg, New York: Springer-Verlag.

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

Lester, H., Arridge, S. R., Jansons, K. M., Lemieux, L., Hajnal, J. V., & Oatridge, A. 1999. Non-linear registration with the Variable Viscosity Fluid Algorithm. *Pages 238–251 of:* et al, A. Kuba (ed), *Information Processing in Medical Imaging*. Berlin, Heidelberg: Springer-Verlag.

Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P., & Lancaster, J. 1995. A Probablistic Atlas of the Human Brain: Theory and Rationale for Its Development. *NeuroImage*, **2**, 89–101.

Miller, M. I., Christensen, G. E., Amit, Y., & Grenander, U. 1993. Mathematical Textbook of Deformable Neuroanatomies. *Proc. Natl. Acad. Sci.*, **90**, 11944–11948.

Noll, D. C., Boada, F. E., & Eddy, W. F. 1997. A Spectral Approach to Analyzing Slice Selection in Planar Imaging: Optimization for Through-Plane Interpolation. *Magnetic Resonance in Medicine*, **38**, 151–160.

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci.*, **89**, 9868–9872.

Ogden, R. W. 1984. *Non-linear Elastic Deformations*. Dover.

Patterson, H. D., & Thompson, R. 1971. Recovery of Inter-Block Information when Block Sizes are Unequal. *Biometrika*, **58**(3), 545–554.

Pelizzari, C. A., Chen, G. T. Y., Spelbring, D. R., Weichselbaum, R. R., & Chen, C. T. 1988. Accurate three-dimensional registration of CT, PET and MR images of the brain. *J. Comput. Assist. Tomogr.*, **13**, 20–26.

Poline, J.-B., Friston, K. J., Worsley, K. J., & Frackowiak, R. S. J. 1995. Estimating Smoothness in Statistical Parametric Maps: Confidence Intervals on p-Values. *J. Comput. Assist. Tomogr.*, **19**(5), 788–796.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992. *Numerical Recipes in C (Second Edition).* Cambridge: Cambridge.

Sled, J. G., Zijdenbos, A. P., & Evans, A. C. 1998. A non-parametric method for automatic correction of intensity non-uniformity in MRI data. *IEEE Transactions on Medical Imaging*, **17**(1), 87–97.

Talairach, J., & Tournoux, P. 1988. *Coplanar stereotaxic atlas of the human brain.* New York: Thieme Medical.

Taylor, J., Worsley, K. J., Zijdenbos, A. P., Paus, T., & Evans, A. C. 1998. Detecting anatomical changes using logistic regression of structure masks. *NeuroImage*, **7**, S753.

Thirion, J.-P. 1995 (May). *Fast Non-Rigid Matching of 3D Medical Images.* Tech. rept. 2547. Institut National de Recherche en Informatique et en Automatique. Available from http://www.inria.fr/RRRT/RR-2547.html.

Thompson, P. M., & Toga, A. W. 1996. Visualization and Mapping of Anatomic Abnormalities using a Probablistic Brain Atlas Based on Random Fluid Transformations. *Pages 383–392 of: Proceedings of the International Conference on Visualization in Biomedical Computing.*

Thompson, P. M., & Toga, A. W. 1999. *Brain Warping.* Academic Press. Chap. 18, pages 311–336.

Van Leemput, K., Maes, F., Vandermeulen, D., & Suetens, P. 1999a. Automated model-based bias field correction of MR images of the brain. *IEEE Transactions on Medical Imaging*, **18**(10), 885–896.

Van Leemput, K., Maes, F., Vandermeulen, D., & Suetens, P. 1999b. Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging*, **18**(10), 897–908.

Vandermeulen, D., Descombes, X., Suetens, P., & Marchal, G. 1996. Unsupervised Regularized Classification of Multi-Spectral MRI. *Pages 229–234 of: Proceedings of the International Conference on Visualization in Biomedical Computing.*

West, J., Fitzpatrick, J. M., Wang, M. Y., Dawant, B. M., Maurer, C. R., Kessler, R. M., Maciunas, R. J., Barillot, C., Lemoine, D., Collignon, A., Maes, F., Suetens, P., Vandermeulen, D., van den Elsen, P. A., Hemler, P. F., Napel, S., Sumanaweera, T. S., Harkness, B., Hill, D. L. G., Studholme, C., Malandain, G., Pennec, X., Noz, M. E., Maguire, G. Q., Pollack, M., Pelizzari, C. A., Robb, R. A., Hanson, D., & Woods, R. P. 1996. Comparison and Evaluation of Retrospective Intermodality Image Registration Techniques. *Pages 332–347 of: Medical Imaging 1996: Image Processing, Proc. SPIE 2710.*

West, J., Fitzpatrick, J. M., Wang, M. Y., Dawant, B. M., Maurer, C. R., Kessler, R. M., Maciunas, R. J., Barillot, C., Lemoine, D., Collignon, A., Maes, F., Suetens, P., Vandermeulen, D., van den Elsen, P. A., Napel, S., Sumanaweera, T. S., Harkness, B., Hemler, P. F., Hill, D. L. G., Hawkes, D. J., Studholme, C., Maintz, J. B. A., Viergever, M. A., Malandain, G., Pennec, X., Noz, M. E., Maguire, G. Q., Pollack, M., Pelizzari, C. A., Robb, R. A., Hanson, D., & Woods, R. P. 1997. Comparison and Evaluation of Retrospective Intermodality Brain Image Registration Techniques. *J. Comput. Assist. Tomogr.*, **21**, 554–566.

Woods, R. P., Cherry, S. R., & Mazziotta, J. C. 1992. Rapid automated algorithm for aligning and reslicing PET images. *J. Comput. Assist. Tomogr.*, **16**, 620–633.

Woods, R. P., Grafton, S. T., Holmes, C. J., Cherry, S. R., & Mazziotta, J. C. 1998a. Automated Image Registration: I. General Methods and Intrasubject, Intramodality Validation. *J. Comput. Assist. Tomogr.*, **22**(1), 139–152.

Woods, R. P., Grafton, S. T., Watson, J. D. G., Sicotte, N. L., & Mazziotta, J. C. 1998b. Automated Image Registration: II. Intersubject Validation of Linear and Nonlinear Models. *J. Comput. Assist. Tomogr.*, **22**(1), 153–165.

Worsley, K. J., & Friston, K. J. 1995. Analysis of fMRI Time-Series Revisited - Again. *NeuroImage*, **2**, 173–181.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, J. J., & Evans, A. C. 1996. A unified statistical approach for determining significant voxels in images of cerebral activation. *Human Brain Mapping*, **4**, 58–73.

Worsley, K. J., Andermann, M., Koulis, T., MacDonald, D., & Evans, A. C. 1999. Detecting Changes in Non-Isotropic Images. *Human Brain Mapping*, **8**(2), 98–101.

Wright, I. C., McGuire, P. K., Poline, J.-B., Travere, J. M., Murray, R. M., Frith, C. D., Frackowiak, R. S. J., & Friston, K. J. 1995. A Voxel-Based Method for the Statistical Analysis of Gray and White Matter Density Applied to Schizophrenia. *NeuroImage*, **2**, 244–252.

Yan, M. X. H., & Karp, J. S. 1995. An Adaptive Baysian Approach to Three-Dimensional MR Brain Segmentation. *Pages 201–213 of:* Bizais, Y., Barillot, C., & Di Paola, R. (eds), *Information Processing in Medical Imaging*. Dordrecht, The Netherlands: Kluwer Academic Publishers.