Analysis of the developing brain using image registration

Kanwal K. Bhatia

A dissertation submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

of the

University of London

March 2007

Department of Computing

Imperial College London

In loving memory of my grandfather

Ooi Taw Hock (T.F. Hwang), 1926-2006.

Abstract

Preterm birth affects around 5% of births in industrialised countries and its consequences contribute to significant individual, medical and social problems. The principle morbidity among survivors is neurological, resulting from the profound effect of preterm birth on the developing brain: half of all infants born at less than 25 weeks have neurodevelopmental impairment at 30 months of age, and in less immature infants, neuropsychiatric problems are common in the teenage years. The structural correlates of functional disorders are, however, poorly characterised.

This motivates the study of the growth of the preterm brain from birth through infancy. However, difficulties in analysis arise due to the absence of a standard anatomical template for either the neonatal or the infant brain. This thesis focuses on the unbiased, average atlas construction for populations. Methods for groupwise registration have been developed in order to create atlases representing the average shape of a population. In addition, groupwise segmentation techniques have been developed to segment a population of aligned subjects, in order to obtain probabilistic atlases of a population. Finally, groupwise registration and segmentation have been combined in order to obtain more accurate representations and segmentations of the average shape.

Groupwise registration has been used to create average atlases of the preterm and term-born neonate at term-equivalent age, and these atlases have been compared using deformation-based morphometry to determine quantitative differences between the populations. Groupwise registration and segmentation have furthermore been used to create average intensity and probabilistic segmentations of populations of one- and two-year-old subjects. The growth occurring between these two time points has also been quantified.

Acknowledgements

I would first of all like to express my sincere thanks to my supervisor Daniel Rueckert for all his help, advice and enthusiasm over the past years. I would also like to thank my co-supervisors, Jo Hajnal and David Edwards, particularly for their assistance on the the practical aspects of MR and the clinical aspects of the work. It has really been a pleasure to work on this project.

Many thanks also to my colleagues on associated neonatal projects, in particular to Paul Aljabar, James Boardman, Mary Rutherford, Serena Counsell, Latha Srinivasan, Maria Murgasova and Lara Leijser, for their assistance.

The Visual Information Processing Group has been a stimulating and entertaining place to work and I thank everyone there, particularly those who started with me in Room 343, for making the working environment so interesting.

Finally, I would like to thank my family whose unending support has made this thesis possible.

Contents

List of Figures

List of Tables

Nomenclature

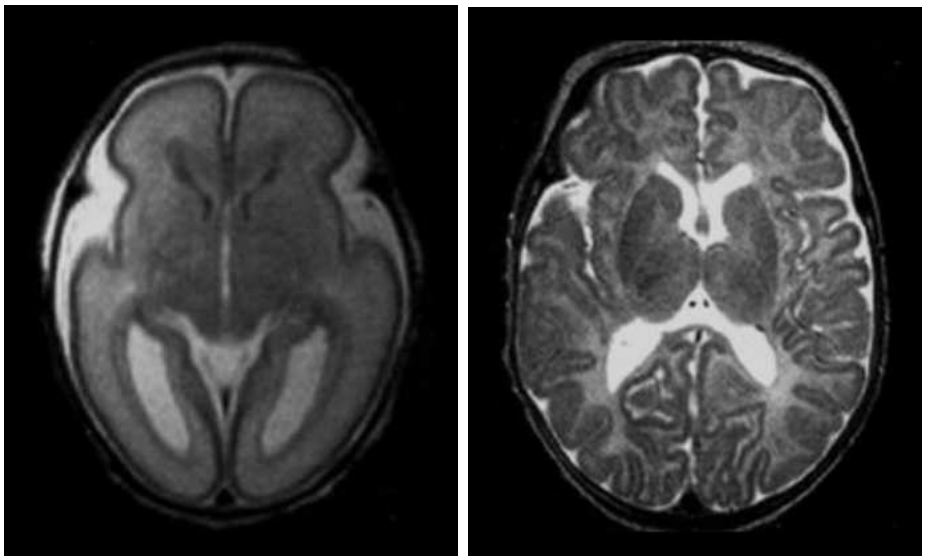
I_i	Image i
Ω	Image domain
$\mathbf{x}$	Location in image domain $\mathbf{x} = (x, y, z)$
$\mathbf{x}_{I_i}$	Location in domain of image i
$\mathbf{T}$	Transformation
$\mathbf{T}^{-1}$	Inverse of transformation
$\mathbf{T}(\mathbf{x})$	Transformed position of location $\mathbf{x}$
$\mathbf{u}(\mathbf{x})$	Deformation of location $\mathbf{x}$
$I_i(\mathbf{x})$	Intensity of location $\mathbf{x}$ in image i
$I(\mathbf{T}(\mathbf{x}))$	Intensity of transformed location of $\mathbf{x}$
$\phi_{i,j,k}$	Location of control point (i, j, k)
Φ	Control point domain
$\mathbf{v}$	Velocity field
$p(i)$	Probability of a voxel with an intensity of i
S	Similarity metric
C	Cost function
n_I	Number of images in a group
n_Ω	Number of voxels in an image domain
n_K	Number of tissue classes
$H(I)$	Entropy of image I
$\mathbf{J}$	Jacobian of deformation field
μ	Mean
σ^2	Variance
θ	Segmentation parameters
$\mathbf{Z}$	Hidden data / unknown segmentation labels
$\mathbf{Y}$	Set of image intensities
j	Index to voxel
k	Index to segmentation class
l	Segmentation label
$G_k(y_j)$	Gaussian distribution at jth voxel for class k
π_{jk}	Prior probability of jth voxel being labelled as class k

Chapter 1

Introduction

Preterm birth is defined as the delivery of a baby before 37 completed weeks of gestation (compared to 40 weeks for a full-term infant). In most industrialised countries, preterm birth typically occurs with an incidence of around 5-7% [250], and its consequences contribute to significant individual, medical, social and economic problems globally. Preterm birth is associated with long-term neurodevelopmental impairment including cognitive and behavioral problems [157, 162].

Around 75% of all perinatal deaths are of infants born preterm [218], with most morbidity and mortality occurring in *very preterm* (delivered before 32 weeks gestation) and *extremely preterm* (delivered before 28 weeks gestation) infants. Over the last 20-30 years, developments in neonatal medicine have improved the outcomes of infants born preterm. However, the effects of preterm birth still extend into later life, and it is a major cause of neurocognitive impairment in childhood [1, 157]. This is likely to be due to the profound effect of preterm birth on the developing brain. Figure 1.1 shows a magnetic resonance (MR) scan of an infant born and imaged at 24 weeks, and imaged again at 40 weeks (term-equivalent age). It can be seen that a huge amount of growth and increased complexity of structures occurs between these two time-points. A comparison of a preterm and term-born infant at term-equivalent age is shown in Figure 1.2. There is increased fluid-filled



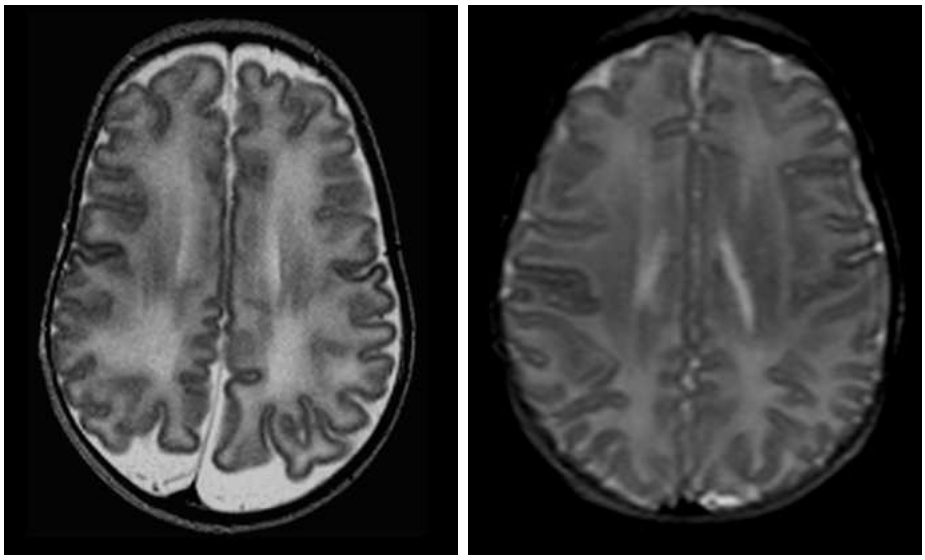
(a) (b)

Figure 1.1: T2 images of a preterm-born infant (a) at 24 weeks (b) at 40 weeks.

space surrounding the brain structures and reduced cortical folding [4, 209] in the preterm. This shows that the exposure to an extra-uterine environment has affected the growth of brain in the preterm infant.

The objective of this thesis is to develop computational techniques that can be used to analyse the growth of the brain in infants born preterm. To do this, we aim to create average representations of anatomy at various time-points through infancy. These can be compared to the anatomy of term-born control subjects at an equivalent age to determine differences between the groups, or to each other, to analyse growth between time-points.

Advances in MR imaging techniques have made the non-invasive acquisition of high-resolution three-dimensional (3D) images of the brain increasingly feasible. By acquiring and analysing images from large populations of subjects, structural and functional trends in the population can be determined. Central to the task of structural analysis of a population is the concept of an anatomical atlas. This specifies a standard coordinate system for analysis and defines typical or "normal"



(a) (b)

Figure 1.2: T2 images of (a) a preterm infant born at 26 weeks and imaged at 40 and (b) a term-born infant, born and imaged at 40 weeks.

anatomy for the group. Atlases can be constructed either from a carefully-chosen individual (e.g. the Talairach atlas [234]), or from combining the scans of many subjects (e.g. the MNI (Montreal Neurological Institute) atlas [80]). The formation of a representative altas enables the comparison of individuals, the comparison of groups of subjects, or the tracking of changes over time. This type of morphometric analyis has been extensively used in the study of adult neurodegenerative disorders, such as Alzheimer's, Schizophrenia and autism [87, 86, 244, 8, 61, 258, 174].

However, unlike the case for adults, no standard anatomical atlas exists for the neonatal or infant brain, either using an individual subject or the combination of multiple subjects. This thesis focuses on the building of unbiased atlases of the neonatal and infant brain and the subsequent analysis of these atlases (although the techniques developed are applicable to any population). The aim is to use MR scans of multiple individuals to gain information about the development of the brain of infants born preterm. To do this, two methods of MR image analysis are developed:

1. *Image registration*: the geometric alignment of images such that equivalent

features are brought into spatial correspondence.

2. *Image segmentation*: the automatic delineation of structures within an MR scan.

The ability to achieve the above allows individual structures to be easily compared across individuals, groups or time.

1.1 Neurodevelopmental outcome following preterm birth

The EPICure study [261] of 283 infants provided a large-scale survey of infants born extremely preterm, assessed at 30 weeks gestational age, and followed up at six years of age when disabilities are better able to predict long-term impairment. At 30 weeks, it was found that 49% of survivors had impairment of one or more of neuromotor, mental, psychomotor, sensory or communication developmental domains, with 23% meeting the criteria for severe disability. These infants were reassessed at six years, with results compared to age-matched classmate controls. It was found that 34% suffered from mild disabilities, indicated by neurological signs and minimal functional impairment such as squints. 22% were classified as severely disabled and dependent on care-givers. These children had IQs of more than three standard deviations below the mean, sensorineural hearing loss and impaired visual function. A further 12% of the population had disabling cerebral palsy [157]. Overall, 41% of the children studied displayed cognitive deficits, compared to 1% of the classmate controls.

Importantly, most impairment seen is neurological, with the cognitive domain more frequently affected than neuromotor function, hearing or vision. Other studies globally have found similar results [73, 79, 121, 187]. Studies of very preterm infants have also shown agreeing results on neurologically-based problems faced in adolescence

and later life:

- Poor educational achievement and even school failure [101, 162, 210].
- Behavioural and social difficulties [90, 109].
- Attention problems including Attention Deficit Hyperactivity Disorder (ADHD) [27, 33].
- Fine motor skill impairment [156].

1.2 Neonatal and infant brain imaging methods

To evaluate the effect of preterm birth on the developing brain, *in-vivo* images of the brain of both term-born and preterm-born infants need to be obtained at various time points. These are needed for clinical research purposes to more fully understand how structures in both groups grow over time and the differences between the groups. Additionally, imaging enables the evaluation of the efficacy of potential treatments for damage caused by preterm delivery. Only the techniques applicable to neonatal and child brain imaging are described in this section. A full description of general medical imaging techniques is given in [229].

1.2.1 Cranial Ultrasonography (US)

Cranial ultrasonography (or ultrasound or sonography) for imaging the preterm brain is routinely performed to detect complications associated with preterm birth. It is one of the frequently-used imaging techniques due to the portability of the apparatus (imaging can be performed at the patient's bedside), the lack of use of any ionising radiation, and because it is relatively inexpensive. Sonography is additionally commonly used for foetal monitoring.

Sonography involves the creation of ultrasound waves which are transmitted in pulses

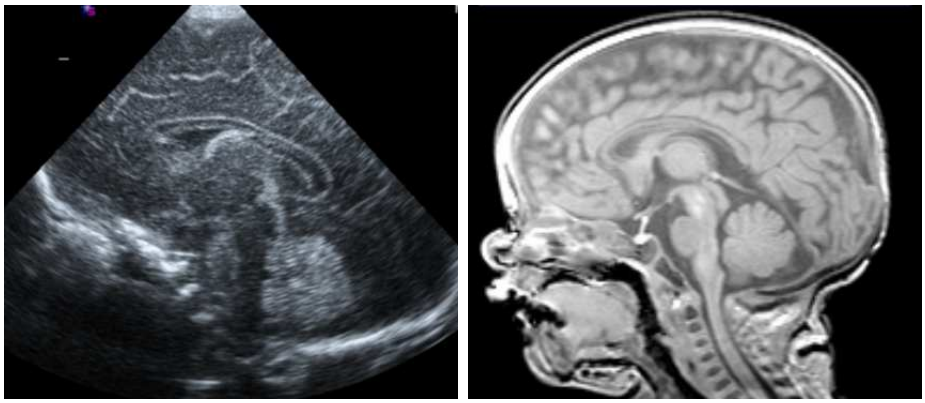
through the body. The reflection of these waves off tissue boundaries produces an echo which can be received and processed to determine the distance of the tissue from the transducer. By using a curvilinear array of transducers a 2D image of the brain can be reconstructed. As sound waves cannot pass through bone, sonography is only used for brain imaging in infants when the cranium is not fully formed. This eliminates its use for tracking structural growth in later childhood.

US has been extensively used to assess preterm infants at birth and up to term-equivalent age. It has proven a useful tool in the detection of haemorrhages and ventriculomegaly (enlargement of ventricles) [67]. It has also been used to effectively diagnose cystic periventricular leucomalacia (cPVL), which is manifested by focal lesions or cysts in the periventricular white matter [248]. Cystic PVL is prevalent in around 3-10% of very low birth weight infants. Its principal clinical correlate is spastic diplegia (the most common form of cerebral palsy following preterm birth) [222, 259]. However, its relatively low prevalence indicates that cPVL cannot be solely responsible for the neurocognitive and behavioural impairments often seen in survivors of preterm birth, and it has been shown to be a poor detector of such impairments [189]. To detect the more subtle structural correlates of these therefore requires a more sensitive image technology.

Figure 1.3(a) shows an example of an ultrasonography image, alongside comparable anatomy obtained using magnetic resonance imaging (b).

1.2.2 Magnetic Resonance (MR)

Magnetic resonance imaging uses the quantum mechanical properties of hydrogen protons, present in different tissues in different quantities (as water and fat). Nuclei which have at least one unpaired proton, like the hydrogen proton, possess inherent spins. Ordinarily, the random alignment of these spins means there is no net mag-



(a) (b)

Figure 1.3: US image (a) and MR image (b) of comparable anatomy in a neonate.

netisation. When a radiofrequency (RF) pulse is applied to the protons, the spins align and begin precessing in phase with each other, creating their own magnetic field. The strength of the magnetic field produced is dependent on the frequency and phase coherence of the spins, the greater the phase coherence, the stronger the field. When the RF pulse is removed, the spins lose energy and return to their equilibrium position. This loss of magnetisation is used to create an MR image. The energy loss occurs through two main ways:

1. Spin-lattice interactions.

2. Spin-spin iterations.

Spin-lattice interactions involve an exchange of energy between the spins and their surroundings. The results in the recovery of the longitudinal component of the magnetisation after a time T1. Spins also interact with themselves in a more rapid process than spin-lattice interactions, and this leads to a loss of phase coherence amongst the spins. The time for the resulting loss in the transverse component of the magnetisation is the T2 time. In general, T2 << T1. Different tissues have different T1 and T2 time constants, for example, myelinated white matter has a shorter T1 than grey matter. This means that in adult MR images, white matter recovers faster, and therefore appears brighter than grey matter, in T1-weighted images. In T2-weighted images of adult brains, grey matter is brighter than white

matter.

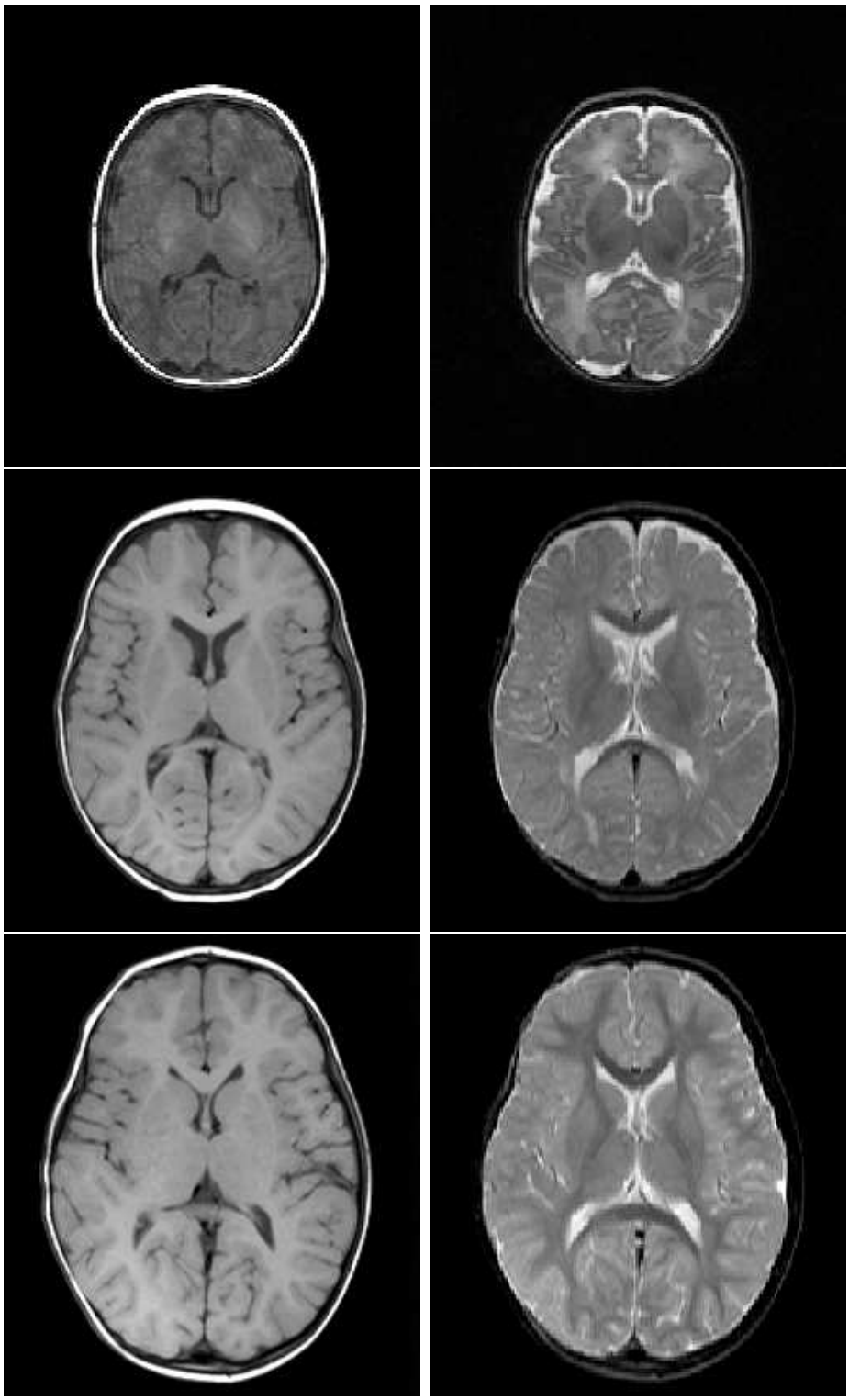
In newborn infants, white matter is largely unmyelinated. Myelination is a process whereby white matter fibres are covered in an insulating lipid sheath in order to aid the transmission of neural impulses. Much of this develops after birth, completing by around the end of the second year [77]. This layer of lipids alters the MR signal of white matter. At birth, the white matter to grey matter contrast is inverted as compared to images of adult brains. Figure 1.4 shows T1- and T2-weighted MR images of a neonate and a two-year-old infant showing inverted contrasts between grey and white matter.

MR has been shown to have greater sensitivity than US at detecting subtle cerebral abnormalities such as diffuse white matter abnormalities and small focal lesions. Additionally, these studies have shown increased correspondence between IQ in later life and predictions based on MR than with those made using US [249, 117, 148].

Previous serial MR studies have shown that early focal lesions frequently regress but are superceded by local or global growth failure and diffuse white matter changes (diffuse excessive high signal intensity - DEHSI - on T2-weighted images). According to [118, 149], abnormal white matter signal intensity is present in half to two-thirds of preterms at term-equivalent age. MR has also detected enlargement of the ventricular system and extra-cerebral space and immature gyral development [118, 149].

1.2.3 Other imaging modalities

Other imaging techniques have been used to investigate the development of the neonatal brain. However, the goal of these methods is to image features other than tissue structure.



(a) (b)

Figure 1.4: T1 (a) and T2-weighted (b) MR images of a neonate (top row), one-year-old (middle row) and two-year-old (bottom row) infant, showing changing contrast between white and grey matter as myelination develops.

1.2.3.1 Diffusion Tensor Imaging (DTI)

Diffusion tensor imaging [135, 167, 235] uses MR to image the movement of water through white matter in the brain, in order to image white matter tracts. Bipolar magnetic field gradient pulses are applied which cause water molecules to diffuse radomly through the tissue. The resulting image represents the probabilistic displacement distribution of the water in each voxel. As white matter consists largely of parallel axonal fibres, diffusion in the direction of the fibres is easier than in the perpendicular direction. This anisotropy allows DTI to be used to image white matter tracts in the brain. Although diffusion anisotropy is present even in unmyelinated white matter, it has been shown that the degree of anisotropy increases with increasing myelination. This has been used to assess brain maturation in children and neonates [173, 268, 115, 208, 247, 232]. More recently, the use of DTI to investigate how white matter connectivity damage in preterm neonates correlates with neurological impairment in later life, has been assessed in [56].

1.2.3.2 Magnetic Resonance Angiography (MRA)

Magnetic Resonance Angiography uses MR techniques to image blood flow through arteries. Time-of-flight MRA [175] allows blood flow, and therefore arteries, to be visualised without the use of any contrast agent. This has been used to show reduced toruosity in the cerebral arteries preterm infants [154].

1.2.3.3 Near-infrared methods

Near-infrared spectroscopy (NIRS) is based on the varying ability of oxygenated and deoxygenated blood (in particular haemoglobin) to absorb radiation at near-infrared wavelengths. This has been used to measure cerebral oxygenation, haemoglobin concentration, cerebral blood volume and flow in newborn infants [35, 199, 78, 265]. By acquiring reflectance measurements at multiple sites over the head, near-infrared can be used as an imaging tool. Two main methods exist for this: optical topography

and optical tomography.

In optical topography, the separation of the near-infrared source and detectors are kept low in order to enable high signals to be acquired quickly. This allows for fast (around 100ms) haemodynamic changes to be detected. Optical topography has been used to study functional activation at sites within the infant cortex in response to a variety of external stimuli [37, 110, 231, 134]. However, the small distance between source and detector results in extreme sensitivity to changes occurring near the surface of the head (such as in the cortex), and so limits its application as an imaging tool for deeper brain regions.

Optical tomography [9] can instead be used to obtain 3D volume images of cerebral oxygenation. By measuring the light transmitted between pairs of points on the surface of the head, a 2D slice or 3D volume, representing the internal distribution of light scatterers and absorbers, can be reconstructed. The use of large source-detector distances, allows greater sensitivity to deep tissue responses. However, the resulting increase in time required to obtain signals of adequate strength constrains its use to the assessment of long-term oxygentation changes (occurring over hours or days). Optical tomography has been used to obtain 2D scans used to identify intercranial haemorrhage [201, 111, 112]. Methods for 3D reconstruction have been developed in [10, 9]. These have been used to image blood volume and oxygenation, and to detect interventricular haemorrhage, in the preterm brain [104, 16].

1.3 Neonatal and infant brain image analysis

The focus of this thesis is on the structural changes that occur to the preterm brain and differences between term and preterm populations. In order to be able to detect subtle abnormalities, and also to be able to relate scans of subject at varying time-points at later ages, MR imaging is used. Methods for computational morphometry

[64, 14] exist for the analysis of adult brain images. However, the analysis of neonatal and child brains poses additional difficulties:

1. There is a much larger variation of brain and skull shapes in neonates as shown in Figure 4.6.

2. Cortical growth (as measured by surface area) increases logarithmically wih respect to unit cerebral volume. Additionally, new structures arise [77], as shown in Figures 1.1 and 1.4.

3. Myelination - the formation of an insulating layer on neurons forming white matter - develops after birth at varying rates until around two years. This changes the signal intensity of the white matter, despite the fact that the underlying structure does not change [77].

4. Infant, and in particular neonatal, brain images have a lower contrast-to-noise ratio than adult brain images. This is due to the absence of fully-developed myelin on white matter, resulting in lower contrast between white and grey matter. Additionally, the small size of infant brains and the short scanning time contribute to overall low contrast-to-noise [194, 95].

5. Obtaining scans of control subjects is difficult. Parental consent for scanning healthy, term-born infants is not always easy to obtain. Even when this is received, the inability to sedate infants often results in increased motion artifacts in images (see Figure 1.6).

6. To our knowledge, no standard anatomical or statistical atlases exist for these populations.

This thesis is concerned solely with the processing of images once acquired, in order to obtain information on neonatal and child brain development. Only images from healthy individuals (with no cPVL) and without artifacts were used in the analysis. The focus is on the construction of atlases representing the *average* image of a

population, in order to reduce the impact of the high variance associated with this particular group.

1.4 Contributions

The aim of the work in this thesis is to analyse the development of structures in the brain of preterm-born infants. To do this, atlases representing average structure of the brain of preterm infants are constructed at term-equivalent age, one year and two years. This enables the comparison of average neuroanatomy of preterms to that of term-born controls at equivalent age. Furthermore, the growth of structures over the first two years can be analysed. To do this requires the construction of representative atlases of each population. In particular, the construction of *unbiased* atlases is desirable. In this thesis, it is proposed that the least biased atlas is one which requires the least deformation from itself to all other subjects in the population. The contributions form the work presented in Chapter 4-6:

- A novel, groupwise, non-rigid registration algorithm for average atlas construction is developed. This defines a common, average coordinate system for atlas construction, such that the sum of deformations from this space to all subjects is zero. To do this, a method of constrained optimization for non-rigid registration is developed. Additionally, similarity measures to assess the similarity of a group of images are developed and compared. The algorithm is tested on simulated 2D MR data and real 3D MR data. The algorithm does not require the choice of any arbitrary reference subject.

- This groupwise non-rigid registration technique is then used to construct average neuroanatomical atlases of term-born and preterm infants at term-equivalent age. These atlases are compared to determine quanititative differences between the two groups. Additionally, average atlases of the brain of preterm infants at one- and two-years-old are created (scans of one- and two-year old controls

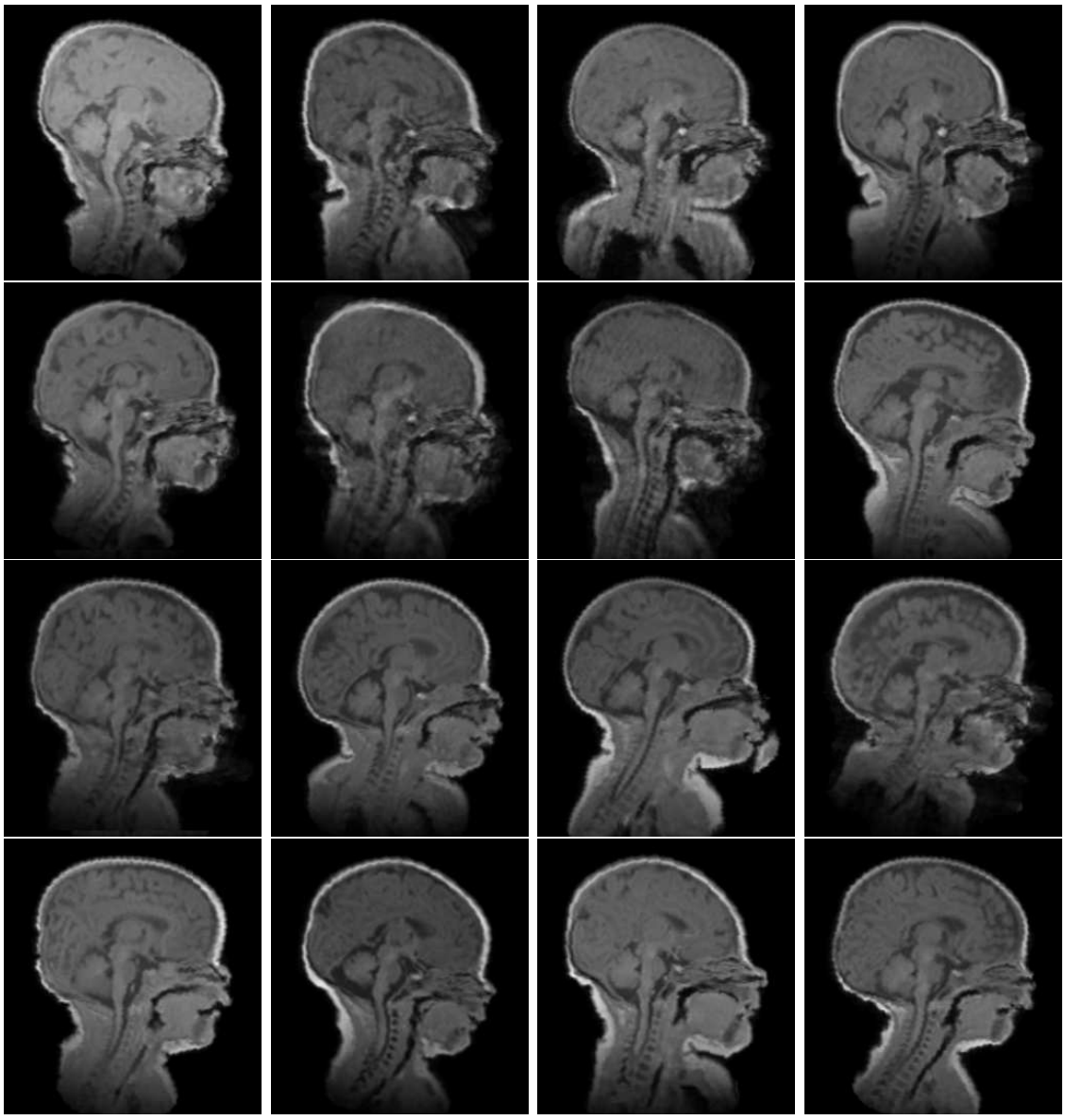
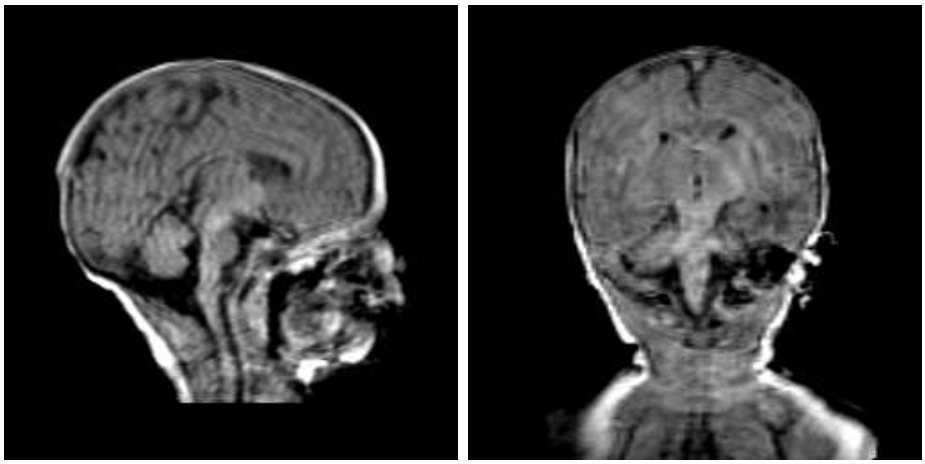


Figure 1.5: Sagittal slices of 16 neonatal subjects, showing variation is shape and contrast.



(a) (b)

Figure 1.6: Sagittal and axial slices of an MR image of a neonate corrupted by motion artifacts.

are not available due to difficulty in recruiting volunteers), and the growth of structures between these two timepoints is determined.

- A novel, groupwise segmentation algorithm is developed. This uses the alignment of multiple images in a common space to aid in the segmentation of each subject in the group, as well as the segmentation of the average shape. Furthermore, algorithms to combine the groupwise segmentation and registration are proposed, with the premise that the improvement of one leads to the improvement of the other. Two methods are developed and evaluated: an interleaved method of segmentation and an integrated method of combining the registration into a Bayesian framework of segmentation. These methods are evaluated on a simulated population of 2D MR data. Groupwise segmentation is used to segment a population of one-year-old preterm subjects, and a population of two-year-old preterm subjects and to create representations of the average shape of each population, as well as the average intensity, hard and soft segmentations in this coordinate system.

1.4.1 Overview of thesis

Chapters 2 and 3 contain the introductory and background material on this topic. Chapter 2 reviews image registration techniques, with particular emphasis on brain image analysis applications. In Chapter 3, methods of average atlas construction are discussed. The subsequent chapters contain the methods developed in this thesis. Chapter 4 develops an algorithm for the groupwise non-rigid registration of a population of subjects to their average shape, in order to construct an unbiased, average anatomical atlas of the population. In Chapter 5, this algorithm and deformation-based morphometry are used to determine the differences between preterm and term-born infants at term-equivalent age. Average atlases of infants born preterm at one and two years are also constructed, and the growth of structures between these ages found. Chapter 6 develops a groupwise segmentation algorithm for aligned images. Additionally, algorithms to combine groupwise segmentation and groupwise registration for the simultaneous segmentation and registration of a population of subjects, are presented. A summary of the work presented in this thesis and potential future work is given in Chapter 7.

Chapter 2

Image Registration

Image registration involves the deformation or transformation of images such that corresponding features are brought into spatial alignment. In brain image analysis, registration allows information from multiple sources to be combined in a common frame of reference, in order to aid clinical interpretation. For example, by comparing the scan of a subject aligned with a model or an atlas of anatomy, or with the scan of another individual (*inter-subject registration*), shape and size differences between subjects or populations can be found [64, 61, 8, 174]. Alternatively, images from the same subject could be registered and compared (*intra-subject registration*). This could be used to determine changes in anatomy occurring over time [87, 244, 242, 6]. Another use of intra-subject registration is in the fusion of scans obtained using different imaging modalities. For example, combining MR with Computed Tomography (CT) scans, has been used to assist in surgical planning [107, 108, 89], while the fusion of MR with Positron Emission Tomography (PET) has been used in the detection of tumours [34]. The clinical applications of brain image registration will be discussed in detail at the end of this chapter.

The goal of image registration is to find the mapping $\mathbf{T}$ between a *reference* (also referred to as *target* or *template*) image I_1 and a *source* image I_2 that maximises the similarity between the images. The following are therefore necessary:

1. *Deformation*, or *transformation models* which deform the source image to the reference space.

2. *Similarity measures* to evaluate the similarity of the images when the mapping is applied (these can be derived from features or intensities in the image).

3. *Optimisation methods* to select the transformation that gives the best similarity.

This chapter reviews image registration techniques applicable to brain image registration. General surveys of image registration can be found in Fitzpatrick et al. [84] and Maintz et al. [153].

2.1 Deformation models

The alignment of images requires finding the transformation (also known as deformation or spatial mapping), that relates the position of features in one image (or coordinate system), to the position of the corresponding features in the other image (or coordinate system). Transformations can be modelled either with respect to a *Lagrangian* frame of reference, which is fixed with respect to a given coordinate system (such as an image), or in a *Eulerian* frame of reference which moves with the deformation. While the latter is frequently used to model fluid deformations (which will be discussed later in this section), most other deformation models adopt the fixed, Lagrangian frame of reference. In this formulation, the transformation which maps a position $\mathbf{x}_{I_2}$ in image I_2 to $\mathbf{x}_{I_1}$ in image I_1 is given by:

$$\mathbf{T} : \mathbf{x}_{I_2} \mapsto \mathbf{x}_{I_1} = \mathbf{T}(\mathbf{x}_{I_2}) = \mathbf{x}_{I_1} \tag{2.1}$$

where $\mathbf{x} = (x, y, z)$ represents a voxel location in the image. The intensity of position of image I_1 at $\mathbf{x}$ is denoted by $I_1(\mathbf{x})$.

Transformation models can be broadly classified into two types: those that preserve the straightness of lines (rigid and affine) and those that do not (non-linear or non-rigid). Non-linear transformations allow for more detailed and localised deformations, which is useful when the images show high levels of anatomical variability between subjects.

2.1.1 Rigid transformations

Rigid transformations preserve distances and angles in the object to which they are applied. They therefore only allow for rotations and translations. In 3D, this gives six degrees of freedom: translations in x, y and z directions, and rotations about the same three axes. The rigid body transformation is given by a rotation $\mathbf{R}$ followed by a translation $\mathbf{t}$, and maps point $\mathbf{x} = (x, y, z)$ to point $\mathbf{x}' = (x', y', z')$.

$$\mathbf{x}' = \mathbf{R}\mathbf{x} + \mathbf{t} \tag{2.2}$$

where $\mathbf{R} = \{r_{ij}\}\ \ i, j \in \{0, 1, 2\}$ is the matrix describing the rotational component of the transformation and $\mathbf{t} = (t_x, t_y, t_z)$ is the vector describing the translational component:

$$\mathbf{T}_{rigid}(\mathbf{x}) = \begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{pmatrix} r_{01} & r_{02} & r_{03} & t_x \\ r_{11} & r_{12} & r_{13} & t_y \\ r_{21} & r_{22} & r_{23} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{2.3}$$

For translation only, this simplifies to:

$$\mathbf{T}_{rigid}(\mathbf{x}) = \begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{2.4}$$

For rotation of θ degrees around the x-axis only, the rigid transformation matrix will be given by:

$$\mathbf{T}_{rigid}(\mathbf{x}) = \begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & t_x \\ 0 & \cos\theta & \sin\theta & t_y \\ 0 & -\sin\theta & \cos\theta & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{2.5}$$

2.1.2 Affine transformations

Affine transformations add scaling and shearing to rigid transformations. Parallel lines are maintained during the transformation:

$$\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{t} \tag{2.6}$$

$$\mathbf{T}_{affine}(\mathbf{x}) = \begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{pmatrix} a_{01} & a_{02} & a_{03} & t_x \\ a_{11} & a_{12} & a_{13} & t_y \\ a_{21} & a_{22} & a_{23} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{2.7}$$

where $\mathbf{A} = \{a_{ij}\}, i, j \in \{0, 1, 2\}$ gives the matrix describing the rotational, shear and scale parameters of the transformation and $\mathbf{t}$ denotes the translational component of the transformation.

Figure 2.1 shows examples of the deformation that can be obtained using rigid and affine transformation models. Limited deformation can be obtained using these models. Although they can be used to account for global differences in size and shape, they cannot accurately describe local variation between subjects. Structures in human brains vary greatly between subjects, which makes rigid and affine registration models insufficient for inter-subject registration. For more localised deformation, non-linear, or *non-rigid* transformation models need to be used. These

are described in the following sections. An overview of these methods can be found in [102].

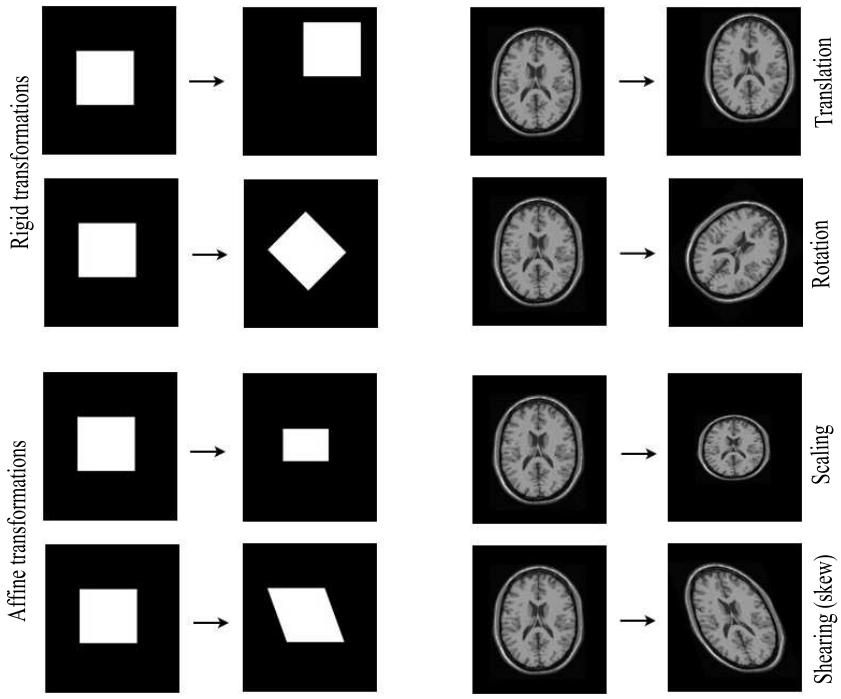


Figure 2.1: Examples of obtainable deformation using rigid and affine transformation models.

2.1.3 Spline-based deformations

Spline-based registration techniques typically require a set of corresponding control points or landmarks to be identified in both source and target images. These points could represent corresponding anatomical features [31] and can be updated as the registration proceeds [168]. Alternatively, the control points could be used only to parameterise the transformation and can be equally spaced across the image to form a regular grid, as in Davis [66], without corresponding to any anatomical landmark. These are referred to as *pseudo-* or *quasi*-landmarks.

The location of the control point in the target image is mapped to the corresponding point in the source image. Between these control points, splines are used to either interpolate or approximate, giving a smoothly-varying displacement field. This in-

terpolation condition can be written as:

$$\mathbf{T}(\phi_a) = \phi'_a \qquad a = 1, ..., n \tag{2.8}$$

where ϕ_a denotes the location of the control point a in the target image, and ϕ'_a represents the location of the corresponding point in the source image.

The term spline originates from engineering where thin strips of metal or wood are used to model ships and planes. In order to bend these strips into shape, weights are applied at certain points along the metal. Image registration uses geometric splines where the displacement of the control points corresponds to the applied weights in engineering splines.

2.1.3.1 Thin-plate splines

Thin-plate splines were originally developed for the interpolation of scattered data [74, 164]. They are based on radial-basis functions, that is, functions whose values depend solely on the distance from the origin. In terms of image registration, the aim is to find a smooth function which interpolates between fixed control point displacements, such that the bending energy of the spline is minimised. The deformation is defined as a linear combination of n radial basis functions, $U(r)$. For 3D deformation, this is given by:

$$t(x, y, z) = \alpha_1 + \alpha_2 x + \alpha_3 y + \alpha_4 z + \sum_{a=1}^{n} w_a U(|\phi_a - (x, y, z)|) \tag{2.9}$$

The transformation can then be regarded as a combination of three separate thin-plate spline functions, one for each coordinate:

$$\mathbf{T}(\mathbf{x}) = (t_1, t_2, t_3)^T \tag{2.10}$$

Given this formulation, the coefficients α represent the affine component of the transformation, and w represent the non-rigid component, which weights each radial basis function. This also gives $3n$ interpolation equations (Equation 2.8) and $3(n+4)$ unknown coefficients. An additional twelve equations are therefore required to determine each of the coefficients uniquely. These equations are chosen to guarantee that the sum of all the non-rigid weighting coefficients is zero and that the sum of their cross-products with the x, y and z locations of the control points is also zero (this guarantees that the final term in Equation 2.9 contains only non-affine terms) [30].

The radial basis function of the thin-plate spline is given by:

$$U(s) = \begin{cases} |s|^2 \, log(|s|) & \text{in 2D} \\ |s| & \text{in 3D} \end{cases} \tag{2.11}$$

Thin-plate splines have been used for image registration in [96, 30, 31]. They also allow additional constraints to be incorporated in the transformation model. For example, rigid body constraints have been used in [142] and directional constraints in [32].

2.1.3.2 B-splines (Free-form deformations)

Radial basis functions generally have infinite support. This means that every basis function, and therefore every control point, contributes to the whole of the transformation. This makes modelling local deformations difficult, and furthermore, prohibits the use of very large numbers of control points due to the increased computational complexity. *Free-form deformations* (FFDs), developed by Sederberg and Parry [214] for computer graphics applications, provide an alternative. FFDs deform an object by manipulating an underlying mesh of control points, producing a smooth transformation. This requires a regular mesh of control points with uniform

spacing. By using locally-controlled blending functions to smoothly approximate the control point displacements, this produces an efficient, yet powerful tool for modelling 3D deformable objects.

Let the spatial domain of an image volume be denoted by $\Omega_{I_1} = \{(x, y, z) \mid 0 \leq x < X, 0 \leq y < Y, 0 \leq z < Z\}$ and Φ denote a mesh of $n_x \times n_y \times n_z$ control points $\phi_{a,b,c}$, with uniform spacing δ. A point $\mathbf{x} = (x, y, z)$ is transformed to its new location, $\mathbf{x}' = (x', y', z')$, using the following equation:

$$\mathbf{T}(\mathbf{x}) = \mathbf{x}' = \mathbf{x} + \sum_{l} \sum_{m} \sum_{n} b_{l,m,n} \phi_{a+l,b+m,c+n} \tag{2.12}$$

where $b_{l,m,n}$ depends on the choice of the blending function. One such function is the Bernstein polynomial. The deformation $\mathbf{u}(x, y, z)$ is given by the trivariate tensor product of Bernstein polynomials:

$$\mathbf{u}(x, y, z) = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} B_{l,n_B}(x) B_{m,n_B}(y) B_{n,n_B}(z) \phi_{a+l,b+m,c+n} \tag{2.13}$$

where

$$B_{i,n_B}(s) = \binom{n_B}{i} s^i (1-s)^{n_B - i} \tag{2.14}$$

where n_B is the order of the polynomial. Bernstein polynomials have previously been used to deform geometric models [214].

Another choice of function is the symmetric cubic B-spline [136, 137]. FFDs based on B-splines have been used in a number of image registration problems [68, 81, 207]. The displacement field $\mathbf{u}$ given by the FFD can be expressed as the 3D tensor product of 1D cubic B-splines [137]:

$$\mathbf{u}(x, y, z) = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} B_l(u) B_m(v) B_n(w) \phi_{a+l,b+m,c+n} \tag{2.15}$$

where $a = \lfloor \frac{x}{n_x} \rfloor - 1, b = \lfloor \frac{y}{n_y} \rfloor - 1, c = \lfloor \frac{z}{n_z} \rfloor - 1, u = \frac{x}{n_x} - \lfloor \frac{x}{n_x} \rfloor, v = \frac{y}{n_y} - \lfloor \frac{y}{n_y} \rfloor, w = \frac{z}{n_z} - \lfloor \frac{z}{n_z} \rfloor$ and where B_l represents the l-th basis function of the B-spline:

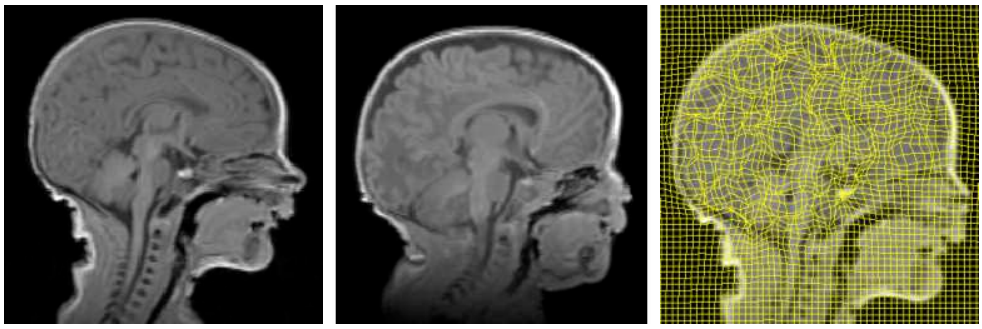
$$\begin{aligned} B_0(u) &= (1-u)^3/6 \\ B_1(u) &= (3u^3 - 6u^2 + 4)/6 \\ B_2(u) &= (-3u^3 + 3u^2 + 3u + 1)/6 \\ B_3(u) &= u^3/6 \end{aligned}$$

B-splines are locally-controlled and have limited support. This means that changing a control point $\phi_{a,b,c}$ only changes the transformation in the neighbourhood of that point, making their use computationally efficient, even for large numbers of control points. While B-splines provide a smooth interpolation between control points, the movement of the control points allows for folding and tearing of the overall deformation, which may be unrealistic in medical image registration. Smoothness regularisation terms may therefore have to be incorporated into the registration process as in [207].

Rueckert et al. [207] introduced an overall transformation model which accounts for both global and local deformations:

$$\mathbf{T}(x', y', z') = \mathbf{T}_{global}(x, y, z) + \mathbf{T}_{local}(x, y, z) \tag{2.16}$$

where $\mathbf{T}_{local}(x, y, z) = \mathbf{u}(x, y, z)$ represents the deformation obtained using an FFD model based on B-splines (Equation 2.15). The global deformation is handled using an affine transformation (as described in 2.1.2). This was initially developed to model the non-rigid deformation of breast tissue [207] and has also been applied to liver and brain image registration [213] as well as to cardiac image registration [182]. Figure 2.2 shows an example of a source brain image being warped into target space by deforming a grid of control points and using an FFD based on B-splines.



(a) (b) (c)

Figure 2.2: Non-rigid registration of two images using FFD transformation based on B-splines. (a): target image I_1; (b): source image I_2; (c): source image mapped into target space using FFD based on B-splines.

In this example, the brain of the target image (a) was manually segmented so the deformation only occurs within the brain area.

2.1.4 Physical models of deformation

2.1.4.1 Elastic deformation

Bajcsy et al. [18] proposed elastic registration techniques to match an atlas of the brain with a CT image of a new subject. This method models the deformation required to match two images as a physical process akin to the deformation of an elastic material (such as rubber). In elastic materials, any applied external force is counteracted by an internal force (a property of the material itself) which resists change from the equilibrium state. When these two forces are equal, the deformation stops. At equilibrium:

$$\mu\nabla^2\mathbf{u} + (\lambda + \mu)\nabla(\nabla \cdot \mathbf{u}) + \mathbf{f} = 0 \tag{2.17}$$

where $\mathbf{u}$ represents the displacement field and $\mathbf{f}$ represents the applied force used to drive the registration process. μ and λ are constants of elasticity (which can be combined to give the Young's modulus and Poisson's ratio of a material).

The applied force is commonly chosen to be the gradient of a similarity measure between the two images. These similarity measures could be based on intensities [18], intensity differences [39] or intensity features (for example curvature) [91]. Alternatively, the difference in the anatomical structures themselves could be used to drive the registration: [65] use the distance between the curves, and [241] use the distance between the surfaces, of corresponding anatomies.

The partial differential equation in Equation 2.17 can be solved using finite differences, giving a displacement field for each voxel. Alternatively, it might be more efficient only to consider the voxels corresponding to the nodes of a finite element model and interpolating between these nodes [91].

Davatzikos [63] has proposed an extension to the original method to allow for spatially-varying elasticity parameters. This enables different anatomical structures to deform by different amounts; more variable structures, such as brain ventricles, are allowed to deform more freely than those which typically show less variation.

2.1.4.2 Fluid deformation

The amount of deformation obtained using elastic registration is proportional to the force. For this reason elastic deformations cannot easily model highly localised deformations. This has led to interest in fluid registration techniques which enable large as well as local deformations (including corners) to be smoothly recovered. Fluid motion is commonly described in an *Eulerian* frame of reference, that is, one that moves with the motion of the fluid. Instead of using the displacement of the deformation, the velocity of the fluid is therefore considered in the Navier-Stokes equation:

$$\mu\nabla^2\mathbf{v} + (\lambda + \mu)\nabla(\nabla \cdot \mathbf{v}) + \mathbf{f} = \mathbf{0} \tag{2.18}$$

This partial differential equation is similar to that for elastic deformation (Equation 2.17). However, here $\mathbf{v} = \mathbf{v}(x, y, z)$ represents the velocity field and is solved at each time step, and λ and μ represent coefficients of viscosity.

To solve equation 2.18, Christensen et al. [43] use successive over-relaxation (SOR) [195]. However, this can be slow and computationally expensive. Bro-Nielsen et al. [36] propose a faster alternative by using a convolution filter. However, this requires that the viscosity is constant throughout, which is not always the case. As in the case for elastic registration, spatially-varying models of viscosity have been proposed [140], which allow for varying degrees of deformation among structures. These require the use of the conventional numerical schemes, such as SOR, to solve 2.18.

2.1.4.3 Large Deformation Diffeomorphic Metric Mappings (LDDMMs)

A related registration technique proposed by Beg et al. [24] is the Large Deformation Diffeomorphic Metric Mapping (LDDMM) method. The two images to be registered are assumed to be connected via a geodesic flow, which is estimated using a variational framework:

$$\hat{\mathbf{v}} = \underset{\mathbf{v}}{\operatorname{argmin}} \int_0^1 \|L\mathbf{v}_t\|^2 \, dt + \frac{1}{\sigma^2} \left\| I_1(\mathbf{T}^{-1}) - I_2 \right\|_{L^2}^2 \tag{2.19}$$

where $\|L\mathbf{v}_t\|$ is an appropriate Sobolev norm on the velocity field $\mathbf{v}_t$ and $\|\cdot\|_{L^2}$ denotes the squared-error norm for integrable functions and represents the difference in similarity of the images. This method ensures that all mappings are smooth and diffeomorphic. Additionally, in contrast to the fluid registration technique described in the previous section, this mapping gives a metric on the length of the shortest path connecting the two images:

$$\inf \int_0^1 \|\mathbf{v}_t\|_V \, dt \tag{2.20}$$

2.1.4.4 Optical flow and the demon's algorithm

The demon's algorithm, as described by Thirion [238], takes its name from the analogy to Maxwell's demons from thermodynamics, which (contrary to the second law of thermodynamics), aims to reduce the entropy of two substances by separating them at a boundary of "demons". In relation to image registration, it is assumed that for each point in a given object in image I_2, it is possible to determine whether it is inside or outside the boundary of the same object in image I_1. The demons, situated on the object boundary, then act to push points outside the boundary inside, and vice-versa.

Optical flow (Horn et al. [114]) can be thought of as a variant on the demon's algorithm. It was however, originally developed as a computer vision tool to recover relative motion of objects between two frames of a temporal image sequence. Optical flow represents the distribution of velocities of movement of brightness patterns in an image, and is comparable to the equation of motion for an ideal, incompressible fluid. The basic premise is that the intensity of a moving object is constant with time. When using this for image registration, it means that a structure in one image "moves" to form the structure in the next image, and should thus have the same intensities. This means:

$$I_1(x, y, z, t) = I_1(x + \delta x, y + \delta y, z + \delta z, t + \delta t) \tag{2.21}$$

Taylor's expansion of the right hand side gives the equation for optical flow:

$$\frac{\partial I_1}{\partial x}\frac{dx}{dt} + \frac{\partial I_1}{\partial y}\frac{dy}{dt} + \frac{\partial I_1}{\partial z}\frac{dz}{dt} + \frac{\partial I_1}{\partial t} = 0 \tag{2.22}$$

ignoring higher order terms. Rearranging (see [114] for full details), gives:

$$\left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right) = -\frac{I_{1,t}}{\sqrt{I_{1,x}^2 + I_{1,y}^2 + I_{1,z}^2}} \tag{2.23}$$

which represents the component of the movement in the direction of the brightness gradient. However, if every point is allowed to move freely, determining these velocities becomes infeasible. It is therefore necessary to incorporate additional smoothness constraints (such as applying a Gaussian filter to each component), to the formulation, as described in [238, 20].

This section has described a number of non-rigid registration techniques. The efficacy of each technique depends on the underlying objects to be registered as well as any computational constraints. Much research has been presented using the above techniques with little consensus as to the "best" method. However, for the non-rigid alignment of intersubject infant and adult brain images, a model which can efficiently model highly localised deformations is necessary. Additionally, the performance of the registration method is dependent on the similarity metric and optimisation procedure chosen. These will be discussed in the following sections.

2.2 Similarity metrics

The goal of image registration is to match one image (source) to a reference image (target). The deformation models described previously warp the images concerned until the alignment between the two images is maximised. To do this requires some measure of the similarity between the two images. A full review of similarity measure for image registration can be found in [102].

2.2.1 Point-based methods

Point-based similarity metrics rely upon having corresponding sets of points identified in target and source images. These points may be obtained by external objects introduced to the image by rigid structures such as stereotactic frames placed around the head, or by markers placed in the skin. Alternatively, internal markers based on anatomical features visible in the images may be used. The internal points may

be manually or, if they are clearly discernable, automatically located. For a set of points $\{\mathbf{x}_p : p \in \{0, 1...P\}\}$ in the target image and $\{\mathbf{y}_p : p \in \{0, 1...P\}\}$ in the source image, where P is the total number of points, the alignment of the images is found by minimising the distance between the target points and transformed source points:

$$S = \sum_p w_p^2 \left\| \mathbf{x}_p - \mathbf{T}(\mathbf{y}_p) \right\|^2 \tag{2.24}$$

w_p denotes a weighting term representing the degree of confidence with which the point p has been located. For rigid transformations, a least-squares fitting approach can be used to solve Equation 2.24 [12]. Additionally, methods for spline-based transformations also exist [84].

2.2.2 Voxel-based metrics

These metrics look at differences between the voxel intensities at corresponding locations in the two images. Given two images, a target I_1 and a source I_2, and a transformation $\mathbf{T}$, the overall similarity of these images is given by the sum of the distances at each corresponding voxel location $\mathbf{x}$ over the image domain Ω. In the following, $I(\mathbf{x})$ denotes the image intensity of voxel location $\mathbf{x}$ in the target image, and $I_2(\mathbf{T}(\mathbf{x}))$ represents the intensity of voxel $\mathbf{x}$ in I_2, transformed by transformation $\mathbf{T}$.

2.2.2.1 Cross-correlation (CC)

Cross-correlation was one of the earliest intensity-based measures used for image registration [200]. The cross-correlation of two images is given by:

$$S_{CC} = \sum_{\mathbf{x} \in \Omega} I_1(\mathbf{x}) \cdot I_2(\mathbf{T}(\mathbf{x})) \tag{2.25}$$

This measure assumes a linear relationship between corresponding intensities in the images, and so is sensitive to differences in brightness and constrast. To reduce this

dependency, the *normalised cross-correlation* can be used:

$$\sum_{\mathbf{x}\in\Omega} \frac{(I_1(\mathbf{x}) - \langle I_1(\mathbf{x})\rangle) \cdot (I_2(\mathbf{T}(\mathbf{x})) - \langle I_2(\mathbf{T}(\mathbf{x}))\rangle)}{\sqrt{(I_1(\mathbf{x}) - \langle I_1(\mathbf{x})\rangle)^2 \cdot (I_2(\mathbf{T}(\mathbf{x})) - \langle I_2(\mathbf{T}(\mathbf{x}))\rangle)^2}} \tag{2.26}$$

where $\langle\cdot\rangle$ represents the mean intensity.

2.2.2.2 Sum-of-squared/absolute difference (SSD)

The sum-of-squared, or Euclidean, distance a between target and source image is given by:

$$\sum_{\mathbf{x}\in\Omega} (I_1(\mathbf{x}) - I_2(\mathbf{T}(\mathbf{x})))^2 \tag{2.27}$$

when the voxels in both images have exactly the same intensities after the transformation has been applied, this value is at a minimum of zero. This similarity metric assumes that the images will be identical when registered, except for Gaussian noise.

As with the cross-correlation metric, SSD can be strongly affected by a small number of voxels having large intensity differences. A similar metric which is less sensitive is the sum-of-absolute differences (SAD).

$$\sum_{\mathbf{x}\in\Omega} |I_1(\mathbf{x}) - I_2(\mathbf{T}(\mathbf{x}))| \tag{2.28}$$

2.2.3 Entropy-based metrics

The variability of intensities in MR images means that corresponding structures in different images, need not have the same voxel intensities. Taking this into account, entropy-based similarity metrics which use information from the whole image are attractive. A survey of entropy-based registration can be found in [190].

2.2.3.1 Joint Entropy (JE)

When considering how well two images are aligned, their joint entropy [226] can be considered. For two images I and I_2 with intensities $i_1(\mathbf{x}) \in I_1$ and $i_2(\mathbf{x}) \in I_2$ respectively, their joint entropy, $H(I_1, I_2)$, is given by:

$$S_{JE}(I_1, I_2) = H(I_1, I_2) = -\sum_{i_1 \in I_1, i_2 \in I_2} p(i_1, i_2) \log p(i_1, i_2) \tag{2.29}$$

where $p(i_1, i_2)$ represents the joint probability density funtion of the images I_1 and I_2. It has been shown heuristically that as the images get better aligned, their JE often decreases, indicating less disorder in the overlap. However, as shown in [49], low values of JE can be found with very poor alignment. For example, if the images are transformed in such a way that only background (and not anatomical structure) is aligned, this will still result in a good JE.

2.2.3.2 Mutual Information (MI)

An alternative measure is mutual information (MI) [256, 49], which additionally takes into account the individual entropies of the images. MI gives a measure of how much information one variable gives about another (in this case, the variables being the intensities in each image), instead of comparing intensities directly.

The marginal entropies are defined to be:

$$H(I_1) = -\sum_{i_1 \in I_1} p(i_1) \log p(i_1) \tag{2.30}$$

$$H(I_2) = -\sum_{i_2 \in I_2} p(i_2) \log p(i_2) \tag{2.31}$$

where $p(i_1)$ and $p(i_2)$ are the probabilities of voxels with intensities i_1 and i_2 occurring in the corresponding image. Methods to estimate these probabilities will be described in Section 2.2.3.4. The mutual information, which needs to be maximised,

is then given by:

$$S_{MI}(I_1, I_2) = H(I_1) + H(I_2) - H(I_1, I_2) = \sum_{i_1 \in I_1, i_2 \in I_2} p(i_1, i_2) \log \frac{p(i_1, i_2)}{p(i_1) \cdot p(i_2)} \quad (2.32)$$

As can be seen from Equation 2.32, minimising the joint entropy still increases the mutual information. However, the addition of the marginal entropies for each individual image penalises a reduction in the amount of information in each image. It is therefore less sensitive to overlap than the joint entropy.

Equation 2.32 is equivalent to the Kullback-Leibler distance [132] between the probability distributions $p(i_1, i_2)$ and $p(i_1) \cdot p(i_2)$. If I_1 and I_2 are completely independent, then $p(i_1, i_2) = p(i_1) \cdot p(i_2)$ and $S_{MI}(I_1, I_2) = 0$. Mutual information can therefore be viewed as a measure of the dependence of two images: the more dependent, the higher the value.

Additionally, MI can be written as:

$$\begin{aligned} S_{MI}(I_1, I_2) &= H(I_1) + H(I_2) - H(I_1, I_2) \\ &= H(I_1) - H(I_1|I_2) \\ &= H(I_2) - H(I_2|I_1) \end{aligned} \quad (2.33)$$

where $H(I_1|I_2)$ is the conditional entropy defined as:

$$H(I_1|I_2) = -\sum_{i_1 \in I_1, i_2 \in I_2} p(i_1, i_2) \log p(i_1|i_2) \quad (2.34)$$

This formulation interprets mutual information as the reduction in the uncertainty of I_1 (or I_2), due to knowledge of I_2 (or I_1).

2.2.3.3 Normalized Mutual Information (NMI)

Although MI has been shown to be less sensitive to overlap than JE, it is still not invariant. In certain cases, a reduction in overlap still leads to an increase in mutual information, yet also casuses an increase in misalignment. To tackle this overlap problem, Studholme et al. [227] proposed an alternate measure, Normalised Mutual Information (NMI):

$$S_{NMI}(I_1, I_2) = \frac{H(I_1) + H(I_2)}{H(I_1, I_2)} \tag{2.35}$$

which was shown to be invariant to image overlap. This has been used as a similarity measure in [207].

2.2.3.4 Density estimation

Entropy-based similarity metrics require the ability to estimate probability density functions of image intensities in the target and source images. Two methods to do this are based on histograms and kernel-density estimators.

Histogram-based estimation. This provides a frequentist approach to density estimation and its simplicity has made it a popular choice for image registration [49, 228, 227]. Histograms partition the range of image intensities of the target and source images into distinct intervals, known as bins, of fixed width. A 2D histogram of target and source bins is created. Each entry in the histogram corresponds to the number of times an intensity in the target image coincides with an intensity in the source image. The probability that a voxel lies within a particular range of intensities is then simply the number of samples in the corresponding bin divided by the total number of samples in the histogram.

Issues arise however, as to what width and number of bins to use. Too many bins gives a spiky histogram, displaying structure that may not be present in the original dataset, while also increasing computational requirements. Conversely, too

few bins gives too smooth a distribution where the information lost may adversely affect accuracy. Another problem is that the use of histograms gives discontinuities at the edges of bins, which may not reflect the actual underlying distribution.

Kernel density estimation. The use of binning and sampling in the histogram approach can lead to errors in the estimation of the probability density function. An alternative approach is through the use of kernel density estimators or Parzen windows [237, 75]. The parzen window gives the PDF as:

$$p(\mathbf{x}) = \frac{1}{n}\sum_{n=1}^{N}\frac{1}{Nh}K\frac{(\mathbf{x}-\mathbf{x}_n)}{h} \tag{2.36}$$

where h is the bandwidth (a smoothing parameter) and K is the kernel. The selection of the bandwidth is non-trivial and depends on factors such as the sample size and variance. A review of methods used to choose the bandwidth is given in [257]. The kernel is often chosen to be Gaussian [260, 256, 19, 212]:

$$K(\mathbf{x}) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2} \tag{2.37}$$

although exponentials [129] and splines [237] have also been used.

2.3 Optimisation

In order to find the transformation which maximises the similarity between images, methods of optimisation are needed. A full description of optimisation methods can be found in [147, 21]. The most appropriate methods are generally iterative methods which improve the correspondence between the images at each iteration, until a maximum is found. However, these methods generally can only find local optima and require that the gradient of the function can be computed (that is, the function should be smooth and differentiable). The type of optimisation method depends to some extent on the shape of the search space of function; this is turn

depends on the similarity metric chosen. For image registration based on mutual information, a review of optimisation techniques can be found in [151].

2.3.1 Descent-based methods

The gradient of a function represents the direction of steepest descent (or ascent). Steepest descent optimisation traverses the function in the direction of the gradient of the function, until a local optimum is reached. At each iteration, the gradient is found, and the search is moved in that direction. However, the steepest descent direction is a local direction, and moving in this direction, while decreasing the value of the function in this region, may overall not be moving in the best direction to reach a minimum. For search spaces with two or more dimensions, this can lead to a *zig-zag* path which can be slow to reach the optimum.

2.3.2 Conjugate direction methods

This method guarantees a quadratic function will converge in a finite number of steps. In general, any function can be well-approximated by a quadratic function in the region of an optimum point. The conjugate gradient method of optimisation is similar to the method of steepest descent, but instead of moving in the direction of the gradient, the search proceeds in the direction of the *conjugate* direction. This direction is a linear combination of the previous search directions, together with the new gradient at that point. This method is preferable for long, narrow-shaped functions, where steepest descent methods take many iterations to converge. In general it converges faster, although computing the conjugate direction is slightly more complicated.

2.3.3 Hierarchical registration

In the particular case of non-linear registration, the number of degrees of the freedom in the optimisation is usually very large, leading to a complex search space with

many local optima. To account for this, hierarchical registrations are often used [140]. These aim to reduce the number of degrees of freedom initially, and gradually build up to the global optimum. This can be done by blurring or resampling the images initially, or by using deformations which allow fewer degrees of freedom (for example, by using a coarsley-spaced control point grid in [207]).

2.4 Applications of image registration to brain image analysis

Image registration allows information from different sources to be combined and compared, by bringing them into a common alignment. Its application to clinical analysis falls into two broad categories: registration of images from the same subject (intra-subject registration) and registration between images of different subjects (inter-subject).

2.4.1 Intra-subject registration

2.4.1.1 Longitudinal studies: tracking growth and atrophy

The same subject can be imaged over time to track the growth or atrophy of anatomical structures. For example, by non-rigidly registering MR scans of infants at 1- and 2-years-old, growth of anatomical structures in the developing brain over this time period have been analysed [6]. [242] map the development of the brain in older children from 3-15 years. There have also been studies into the progression of neurodegenerative diseases such as Alzheimer's and Multiple Sclerosis, and their associated responses to treatment. In [87, 244], changes in the size and structure of brain anatomy in Alzheimer's patients have been compared with normal aging subjects over time. Registration-based methods to quantify Multiple Scelrosis lesion volumes are developed in [267].

2.4.1.2 Image fusion

Another use of intra-subject registration is to combine information from different modality scans of the same subject [181, 188], taken at the same time (*image fusion*). This could be used to aid in surgical planning. For example, scans from CT and MR could be combined to give clear visualisation of the relative position of bone and soft tissue, which is necessary in planning surgery of the skull base [107, 108, 89]. If the scans have been taken at the same time and without any interventions, then rigid registration may be sufficient to align the images (registration of bony structures should be attainable with only rigid deformations, due to the physical constraints of bone motion).

The registration of PET to MR/CT images can be used to combine information about structure and function. PET and MR/CT scans can be aligned to determine if structural abnormalities are likely to be caused by tumours or infarction.

Another use of image fusion is in radiotheraphy planning [196, 127, 126], where radiation doses need to be calculated and beams need to be accurately located. CT images enable geometrically accurate localisation of bony structures while also providing electron density information needed for accurate calculation of radiation doses. This information can be combined with MR images, which provide better constrast between soft tissues, and therefore better visualisation of the tumour itself.

2.4.2 Inter-subject registration

2.4.2.1 Atlas construction and population comparison

The human brain is extremely variable in structure and this makes analysing individual subjects difficult. The development of brain atlases representing typical anatomy are therefore a critical tool for brain image analysis. Subjects from a population can be registered together to create an atlas for that population. If low-dimensional

transformation models, such as rigid or affine models, are used, the variation of the population is visible in the resulting atlas. If non-rigid transformations are used, the population variation is instead encoded in the deformation fields, and the atlas produced displays crisp anatomy. Individual subjects can then be registered to an atlas to detect differences between structures in the subject and "normal" anatomy. Furthermore, the creation of atlases of different populations of subjects allows the comparison of typical anatomies for each group. Often a test group of subjects is compared to a group of controls or normals. For example, this has been used to find volumetric differences in the brains of term-born and preterm-born neonates [29] and to determine structural differences between males and females in an elderly population [64]. Other work has used the comparison of atlases to analyse the neuroanatomical correlates of autism [174] and to ascertain structural abnormalities in Alzheimer's [8, 244] and Schizophrenia [61] patients.

2.4.2.2 Image segmentation

A further use of inter-subject registration is in the segmentation of structures in brain images. Two commonly-used methods of segmentation are probabilistic segmentation, using the Expectation-Maximisation (EM) algorithm [116, 138, 139], and label propogation [105, 202, 203]. In EM-segmentation, an image can be non-rigidly or affinely aligned to an atlas containing prior information on the segmentation of tissue classes or structures to aid in its segmentation. Label propogation requires an initial segmented image or atlas. The corresponding intensity image is then non-rigidly registered to a new image to be segmented, and the labels are transferred. A more detailed discussion of image segmentation is given in Chapter 6.

2.5 Summary

Image registration is an essential tool for brain image analysis, enabling the integration of information from multiple sources. It has been extensively used for medical

research as well as for clinical diagnosis and surgical planning.

Two important aspects that need to be considered when registering images are the type of transformation and the similarity measure used. The choice of these depends on the data to be registered and the aim of the registration. For example, the increased variation between subjects means that inter-subject registration requires non-rigid transformation models, whereas rigid deformations may be sufficient for intra-subject registration. If very large deformations are required, fluid registration techniques may be considered. The similarity measure used depends on the images being registered. If a simple relationship between images exists, then voxel-based similarity metrics can be used. However, many brain images are highly variable both in terms of anatomy and intensities and so entropy-based metrics may be favoured. These metrics can be particularly useful for multi-modal registration.

Registration is key for the construction of anatomical or statistical atlases, which can be used to measure, visualise and compare anatomy of subjects. Atlases contain information about typical structures in populations of subjects and facilitate the analysis and interpretation of images of individuals or populations. How to use image registration to construct atlases is the subject of the next chapter.

Chapter 3

Average Atlas Construction

Anatomical atlases, such as those of the brain, enable spatial characteristics such as size and location of structures, or regions of functional activation, to be determined. They are generally built from one or more representations of that anatomy and should represent the typical structure or function of a given population. The complex and highly variable nature of human brains means that atlases have an important role to play in the analysis and interpretation of brain images. For example, atlases of different populations can be compared to determine differences between these populations [29]. Alternatively, comparison of a subject to an atlas representing the normal anatomy for that subject's population, allows the detection of abnormalities and of potential disease. Furthermore, information available from an atlas, such as segmentations or labels, can be transferred to new subjects [171].

3.1 Single subject atlases

The effective use of atlases requires a common coordinate system in which subjects can be compared. Finding such a space has been a growing topic of research. The earliest atlases, such as the Talairach atlas [233, 234], were constructed from single-subject anatomical specimens and were aimed at the brain as a structure of interest. The Talairach atlas is a highly detailed and richly labelled 3D dataset. However, the

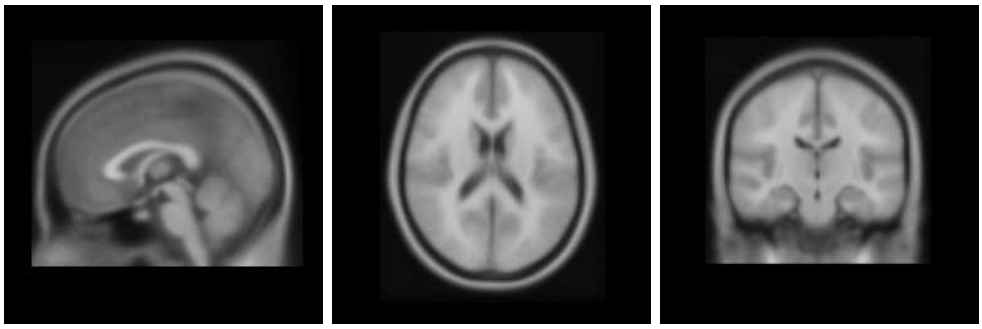
atlas was constructed by imaging the post-mortem brain of one 60-year old woman. Additionally, only one hemisphere was imaged and lateral symmetry was assumed (which is typically not the case in real brains). Although the Talairach atlas space is often used as a standard coordinate system, the atlas is not representative of normal brain structure.

3.2 Population atlases

Given that the anatomy of a single subject can never represent all the variation in a population, there has been interest in developing population atlases, using information from multiple subjects within a group. This form of atlas can be formed using three methods: intensity-based, segmentation-based and deformation-based approaches. It should be noted that there is a philosophical debate as to how atlases should be constructed based on the homology [159] of human brains. The "structure/function problem" described in [83] occurs as the same structure can have different functions in different people, and, likewise, different structures can have the same function. In this section we discuss methods of atlas construction which only consider correspondences in anatomical structure. A detailed discussion on the philosophical arguments of image registration, correspondence and homology can be found in [60].

3.2.1 Intensity-based

Intensity-based atlases involve generating an "average" representation of anatomy by taking the voxel-wise average of scans of multiple subjects. An example of such an atlas is the MNI305 atlas from the Montreal Neurological Institute [80]. This is an atlas created by averaging 305 MR images, linearly aligned to the Talairach space. The subjects used in this atlas were all right-handed and consisted of 239 male subjects and 66 female subjects, aged 19-28 years. Since then, the *International Consortium for Brain Mapping* (ICBM) [3, 160, 161] has affinely registered a further



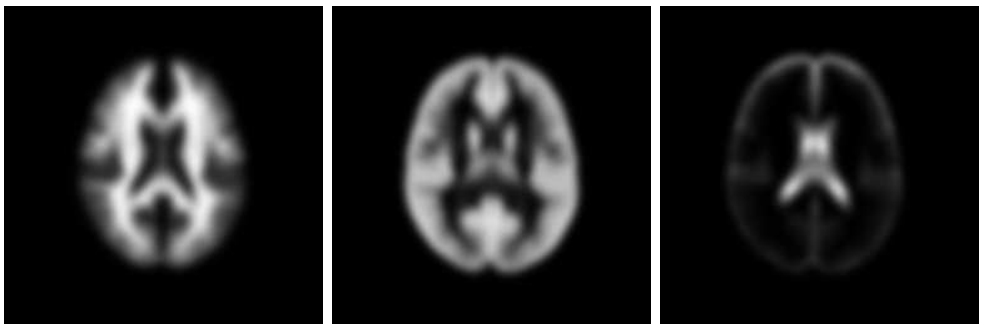
(a) (b) (c)

Figure 3.1: Sagittal (a), axial (b) and coronal (c) slices of the ICBM152 atlas.

152 scans to the MNI305 atlas to produce another intensity-based probabilistic atlas known as the ICBM152, shown in Figure 3.1. A further aim of the ICBM is to use a much larger population (around 7000 subjects) from wide-ranging age, gender and ethnic backgrounds to create probabilistic atlases of the human brain.

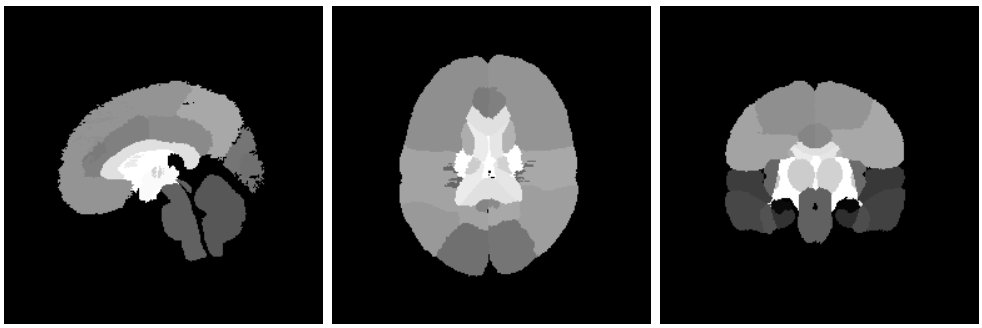
3.2.2 Segmentation-based

The images of the 305 affinely-aligned subjects used have additionally been segmented into white matter (WM), grey matter (GM) and cerebro-spinal fluid (CSF), to give spatial probabilities of the presence of each structure (*priors*) at each voxel. An axial slice of these probabilistic atlases are shown in Figure 3.2. Other similar approaches [176, 180] have been used to create probabilistic atlases of smaller structures within the brain, by affinely aligning and segmenting images of multiple subjects into sub-volumes. An alternative method of probabilistic atlas construction is the Maximum Probability estimate developed by Hammers et al. [103]. 20 subjects have been individually segmented into 49 regions of interest and then aligned with the MNI atlas space. Each voxel in atlas space is assigned to be the most frequently encountered tissue class at that location (i.e.: the modal tissue class encountered), creating a maximum probability atlas for the 49 tissue classes. An example of such an atlas obtained using 20 labels is shown in Figure 3.3.



(a) (b) (c)

Figure 3.2: Axial slices of white matter (a), grey matter (b) and CSF (c) priors created by segmenting tissue classes in affinely-aligned subjects.



(a) (b) (c)

Figure 3.3: Sagittal (a), axial (b) and coronal (c) slices of a maximum probability atlas using 20 labels.

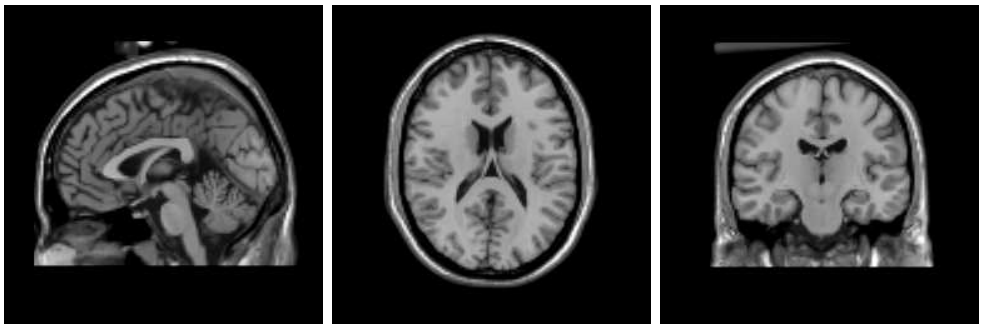


Figure 3.4: MNI Brainweb image of a single subject scanned 27 times. The images have been co-registered and averaged to give a sharp atlas.

In addition to the MNI305, a single subject has been scanned by the MNI lab 27 times. These images have been co-registered and the average intensity image found. The result is a very sharp, low noise atlas, shown in Figure 3.4, which has been used to build realistic brain MR phantoms in [51].

The atlases created in the above studies will always be representative of the subjects used to create them. In terms of medical image analysis, this is a problem if the subject being studied is from a different population. For example, subtle or diffuse changes due to disease, may be difficult to detect if the "normal" population is itself highly variable. An atlas which corresponds more closely with the (undiseased) population of the subject to be studied (for example, in terms of age), is therefore desirable. Hill et al. [106] propose the development of dynamic brain atlases for this. These are atlases which can be created on-demand, with particular attributes (such as age, gender or disease classification), configured to the research question of interest. The use of Grid technologies [205] to access distributed resources enable such atlases to be created 'on-the-fly' as required by clinical demands at the time. However, the need for fast results means that only affine alignment is currently feasible.

3.2.3 Deformation-based atlases

The linear alignment used in the constructions of probabilistic atlases is not sufficient to account for all the variation in a population. This results in a blurred atlas when a voxel-wise average is taken. For smaller structures such as gyri in the cortex, this may lead to information being lost. To account for this, non-linear alignment of the population is needed. In deformation-based methods of atlas construction [29, 245, 246], non-rigid registration is used to match structures locally to the same coordinate system. The resulting intensity atlas does not display much variation and so structures are clearly defined with low noise. The variation in the population is instead encoded in the deformation fields produced by the registration. These can be analysed to find volumetric differences between subjects. However, an important question of which coordinate system to construct a deformation-based atlas in, still remains.

3.3 Image registration

Central to the task of atlas creation is image registration. This is needed to bring subjects into common alignment in order to form the atlas. Typically, image registration is done pairwise: a single subject is chosen as a reference image, and this is registered to all other subjects in the population. While it is possible for the subjects to then be transformed to the average space (how to do this will be discussed in Section 3.5), this method of registration has a number of issues associated with it:

1. **Inconsistency**: registering image I_1 to image I_2 does not necessarily produce the inverse transform to registering image I_2 to image I_1. This could be due to errors in the registration process or to interpolation occurring in different images in each case. Also, it may not even be possible to represent the inverse of the transformation in the required form, for example in B-Spline registration

[207]. While pairwise methods for consistent registration have been developed (and will be described in the following section), these require the transformation model used to be invertible. Additionally, when constructing an atlas of the population, a reference subject still needs to be chosen. Any inferences made from the registration (such as volume measurements), are therefore dependent on this choice. Groupwise registration removes the need to choose a reference at all, and the simultaneous nature of the registration means that the order in which images are considered is irrelevant.

2. **Bias**: The atlas produced represents the anatomy of the chosen reference, and this may not adequately reflect the population.

3. **Distance**: If the reference subject is at an extreme of the population, larger deformations may be required. This may degrade the registration performance. The choice of a suitable reference subject is therefore very important.

This motivates the development of template-free registration - where the reference image is automatically selected or created. Consistent methods of registration also need to be found to reduce the influence of the choice of reference image. The remainder of this chapter discusses previous efforts to develop methods of atlas construction which are less dependent on the choice of reference subject, and which aim to construct the atlas in a coordinate system representative of its population.

3.4 Consistent pairwise registration

Increasing the consistency of pairwise image registration is an important step in unbiased atlas construction. The aim here is to ensure that the order of registration is not important and registering image I_1 to image I_2 will produce the inverse of the transformation obtained when registering image I_2 to image I_1. The importance of choosing one subject to be a reference target is therefore reduced.

3.4.1 Inverse-consistent registration

An inverse consistent linear elastic image registration (ICLEIR) algorithm has been developed by Christensen et al. [40, 133], which jointly estimates the forward and reverse transformations between two images, while constraining these transformations to be inverses of each other. The forward transformation, $\mathbf{T}_{1,2}$ from template image I_1 to target image I_2 and the reverse transformation, $\mathbf{T}_{2,1}$ from I_2 to I_1 are estimated concurrently, subject to the constraint that $\mathbf{T}_{1,2} = \mathbf{T}_{2,1}^{-1}$. This gives the following symmetric cost function to be minimised:

$$S = \sigma \int_{\Omega} |I_1(\mathbf{T}_{1,2}(\mathbf{x})) - I_2(\mathbf{x})|^2 + |I_2(\mathbf{T}_{2,1}(\mathbf{x})) - I_1(\mathbf{x})|^2 d\mathbf{x}$$
$$+\rho \int_{\Omega} \|\ell \mathbf{u}_{1,2}(\mathbf{x})\|^2 + \|\ell \mathbf{u}_{2,1}(\mathbf{x})\|^2 d\mathbf{x} + \chi \int_{\Omega} \|\mathbf{T}_{1,2}(\mathbf{x}) - \mathbf{T}_{2,1}^{-1}(\mathbf{x})\|^2 d\mathbf{x} \qquad (3.1)$$

where $\mathbf{u}_{1,2} = \mathbf{x} - \mathbf{T}_{1,2}(\mathbf{x})$ represents the displacement field from image $I_1(\mathbf{x})$ to $I_2(\mathbf{x})$. The first integral defines the cost of the cumulative squared error similarity between the template image $I_1(\mathbf{T}_{1,2}(\mathbf{x}))$ and the target image $I_2(\mathbf{x})$ and the template image $I_2(\mathbf{T}_{2,1}(\mathbf{x}))$ and the target image $I_1(\mathbf{x})$: minimising this minimises the difference in intensities between the images. The second integral in Equation 3.1 is a regularisation term, constraining the transformations with a linear elasticity constraint given by: $\ell \mathbf{u}(\mathbf{x}) = \alpha \nabla^2 \mathbf{u}(\mathbf{x}) + \beta \nabla(\nabla \cdot \mathbf{u}(\mathbf{x})) + \gamma \mathbf{u}(\mathbf{x})$; α, β, γ are constants. The final integral in the above equation represents the inverse consistency constraint (or inverse consistency error). The influence of each of these three terms is determined by the weighting parameters σ, ρ and χ.

Johnson et al. [122] have additionally developed a similar inverse-consistent algorithm to match landmarks. This has been combined with the intensity matching to give an inverse-consistent algorithm based on the simultaneous matching of both landmarks and intensities, with improved results. The work has been augmented by Magnotta et al. [152] with the additional usage of segmentations to aid the

performance of the registration.

3.4.2 Transitive inverse-consistent registration

An extension to the minimisation of the inverse-consistent registration error for computing pairwise registrations between three subjects has been developed in [93]. This work additionally aims to ensure registrations are *transitive* [41]. This means that for pairwise registrations between images I_1, I_2 and I_3, $\mathbf{T}_{1,2}(\mathbf{x}) = \mathbf{T}_{13}(\mathbf{T}_{32}(\mathbf{x}))$ (see Figure 3.5). A cost function to achieve this is therefore included in the registration:

$$C_{trans} = \sum_{i=1}^{3} \sum_{j=1, j\neq i}^{3} \sum_{k=1, k\neq j\neq i}^{3} \int_{\Omega} \left\| \mathbf{T}_{i,k} \left(\mathbf{T}_{kj}(\mathbf{x}) \right) \right\|^2 d\mathbf{x} \tag{3.2}$$

However, extending this idea beyond three images is not straightforward as with increasing numbers of images, the number of pairwise registrations and the number of paths between images increase significantly.

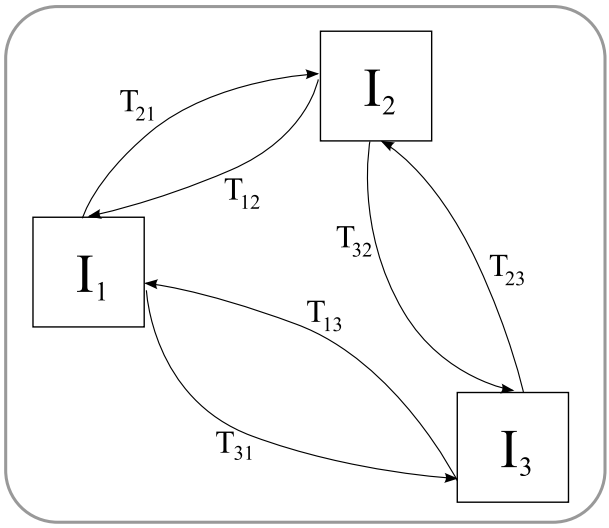


Figure 3.5: Transformations satisfy the transitivity property if $\mathbf{T}_{i,j}(\mathbf{x}) = \mathbf{T}_{i,k}(\mathbf{T}_{k,j}(\mathbf{x}))\forall \mathbf{x} \in \Omega$ and $i \neq j \neq k$.

3.4.3 Partitioning the transformation domain

Skrinjar et al. [217] have also developed a pairwise inverse-consistent algorithm where the transfomation domain is partitioned into triangles. Affine transformations

- which are fully invertable - are then applied over each of these triangles. Given two images I_1 and I_2, the aim is to find the optimum transformation $\mathbf{T}_{opt}$ and its inverse, which maximise the normalised mutual information (NMI) of the first image and the transformed second image, added to the NMI of the second image and the transformed first image:

$$S = \{S_{NMI}\left(I_1, I_2(\mathbf{T}_{opt}(\mathbf{x}))\right) + S_{NMI}\left(I_2, I_1(\mathbf{T}_{opt}^{-1}(\mathbf{x}))\right)\} \tag{3.3}$$

While the methods previously described improve the consistency of registration between two images, they do not directly aid the construction of representative atlases, as the choice of which coordinate system to produce this atlas is still undefined. The methods described in the following sections aim to address this issue through either the sequential pairwise or simultaneous registration of multiple subjects, in order to create an atlas space representing the average shape of the population.

3.5 Average affine atlas construction

Methods have been developed to find the average affine shape of a population. Although many studies [64, 243, 225] discount the global shape and size of brains when determining volumetric changes occurring to specific structures, it can still be important to find the average affine shape and size. For example, when investigating growth of the developing brain in neonates, the global volumetric changes of structures are just as important as local changes.

3.5.1 Geometric averaging of affine matrices

Aljabar et al. [6] have developed a method to average a group of affine transformations. An affine transformation consists of translations, rotations (rigid parameters),

scalings and skews. As the translational and rotational differences between scans are purely due to patient positioning in the scanner, and not due to anatomical differences, it is not necessary to find the average of these.

A simple arithmetic mean is inappropriate for affine transformations. For example, in an extreme case of one subject being scaled by a factor of 0.1, and another subject being scaled by a factor of 10, over a common space, the arithmetic mean would give an incorrect average scale factor of 5.05. Instead, the geometric mean can be taken. The geometric mean of a set of n_I numbers $a_i, i = 1...n_I$ is the n_Ith root of the product of the numbers:

$$\left(\prod_{i=1}^{n_I}\right)^{\frac{1}{n_I}} a_i \tag{3.4}$$

which in the example above, gives an average scale factor of 1. The geometric mean has previously been used to average more general transformations [5] and for tensor data [11]. Equation 3.4 can be rewritten using logarithmic identities to give the geometric mean of a set of n_I affine matrices $\mathbf{A}$:

$$\mathbf{A}_{GM} = \exp\left(\frac{1}{n_I}\sum_{i=1}^{n_I}\log(\mathbf{A}_i)\right) \tag{3.5}$$

In order to construct an atlas representing the average affine shape of the population, pairwise affine registrations are performed to a chosen reference subject. The average affine matrix is calculated, and the inverse of this is concatenated with each of the individual transformations. This new transformation is applied to the corresponding image. The affine transformation that warps each image, i, to the average affine space is therefore given by:

$$\mathbf{A}_{i(ave)} = \mathbf{A}_i \circ \mathbf{A}_{GM}^{-1} \tag{3.6}$$

where $\mathbf{A}_i$ is the affine matrix obtained when registering subject i to a chosen reference subject, and $\mathbf{A}_{GM}^{-1}$ is the inverse of the geometric mean affine transformation.

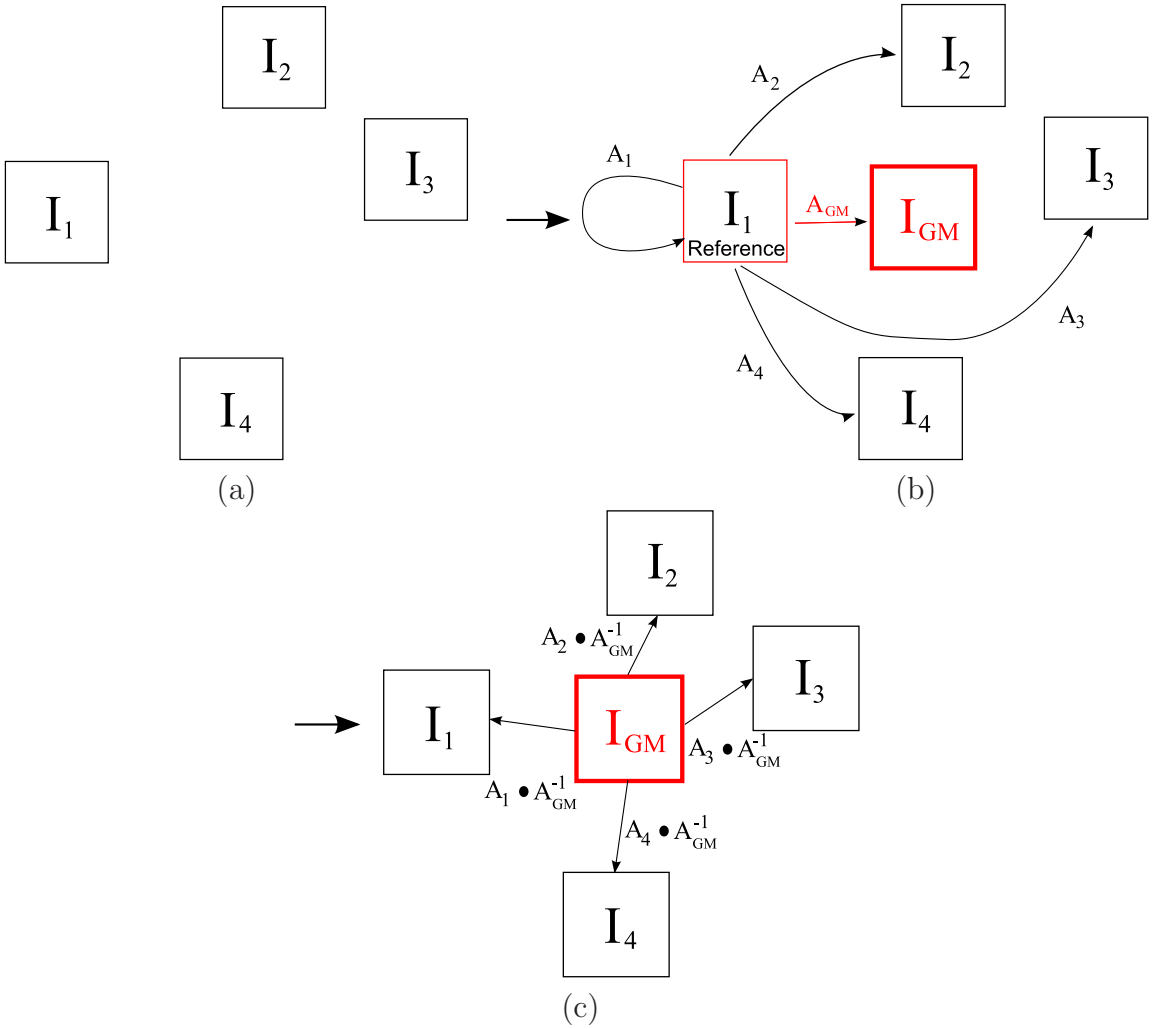


Figure 3.6: Finding the average affine space using geometric averaging of pairwise affine registrations. (a): initial population; (b): calculating the mean transformation from individual registrations to an arbitrary reference; (c): mappings from the population to mean shape.

A schematic for this process is shown in Figure 3.6. This method of averaging affine transformations was used in [6] to construct average atlases of the brain of one- and two-year old infants, and to quantify the development of brain structures between these ages.

3.5.2 Woods's Matrix Averaging

In [262, 263], Woods et al. describe another method to find the affine average brain image of a population, which has shape, size and orientation that are intermediate to those of the original population. The method is based on the fact that any linear mapping between images can be broken down into an arbitrary number of identical smaller linear transformations, by computing the required positive root of the original transformation. These smaller transformations preserve the geometric properties of the original transformation. For example, given two images, I_1 and I_2, linearly registered to a common reference by transformations $\mathbf{T}_1$ and $\mathbf{T}_2$, the direct linear transformation from I_1 to I_2 is given by $\mathbf{T}_1 \circ \mathbf{T}_2^{-1}$. The square root of this transformation (obtained by taking the square root of the elements of the transformation matrix) defines the mapping to a position halfway between the two images: applying the same transformation to the intermediate image will complete the mapping to I_2. A third image, I_3 (also registered to the same reference) can be incorporated into the average by finding the direct transformation from the current average atlas to I_3. Since the original atlas should be weighted twice as heavily as the new image, the cube root of the direct transformation is taken to obtain a new average. This can be repeated for any number of initial transformations to a common reference. This is shown diagrammatically in Figure 3.7.

Furthermore, this method of averaging can be extended to be used for reconciling differences among pairs of pairwise registrations in a set of images. All possible pairwise registrations from each image to all other images are conducted. A matrix

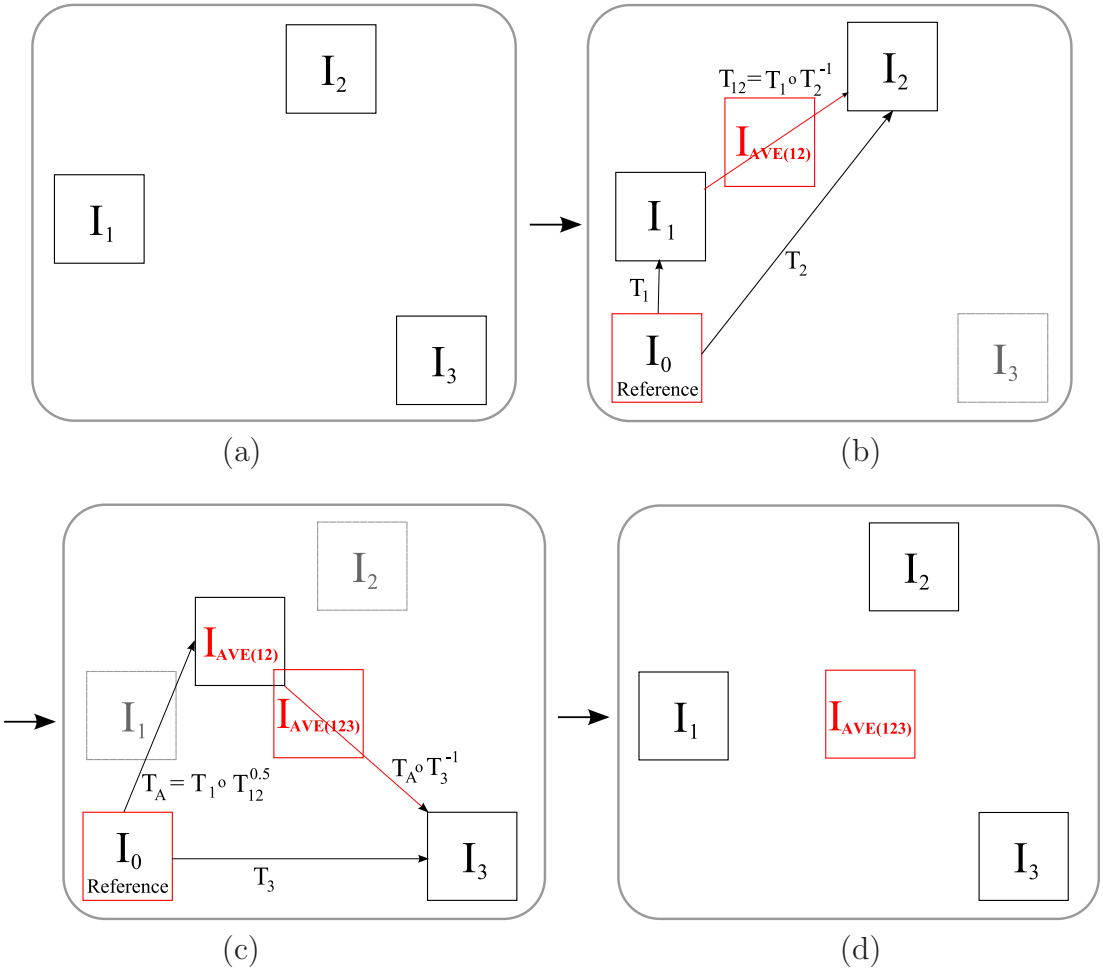


Figure 3.7: Finding the average affine space using Woods's matrix averaging of pairwise affine registrations. (a): initial population; (b): Calculating the average of two images registered to common reference; (c): Calculating the average of the third average and previous average space; (d) final average atlas.

averaging and reconcilliation scheme is then used to find the average space. The transformation that maps image I_1 to I_2 directly, is averaged with all the indirect transformations that map image I_1 to I_X and image I_X to I_2. The original transformation is then replaced with this average, and the process is repeated for all subjects until a convergent, average transformation is found. This reconcilliation scheme does, however, increase the computational complexity non-linearly. For example, for a population of 10 subjects, 45 registrations are required, whereas for a population of 20 subjects, 190 registrations are required.

The similarity metric used to evaluate the quality of the affine alignment is the ratio image uniformity (RIU): for each voxel, the intensity in the transformed image is divided by the intensity in the reference, forming a ratio image. The standard deviation of this ratio image, normalised by its mean, is used as the cost function which is optimised for each image, separately:

$$S_{RIU} = \frac{\sqrt{\frac{1}{n_\Omega}\sum_{\mathbf{x}\in\Omega}(r_{\mathbf{x}} - \bar{r})}}{\bar{r}} \tag{3.7}$$

where $r_{\mathbf{x}}$ is the ratio of the intensity of the transformed image and the intensity of the reference image at location $\mathbf{x}$, and $\bar{r}$ is the mean ratio for the image under consideration.

This only produces an atlas in the average affine coordinate system. It is, however, computationally expensive, requiring $n_I(n_I - 1)/2$ registrations for n_I subjects in the group.

3.5.3 Average affine construction using congealing

Zollei et al. [269] have developed an algorithm to affinely align a population of 3D images to the *central tendency* of the population. To do this, they use a technique formerly applied to hand-written digit recognition known as *congealing*. The aim

is to find the transformation for each subject which minimises the total voxel-wise entropy of the input image volumes when applied, and reduces the overall entropy of the atlas image. The affine registration uses 12 parameters representing rotation, scaling and skews, followed by translations.

The objective function to be minimised is given by:

$$S = \sum_{i=1}^{n_I} H(I_i(\mathbf{T}_i(\mathbf{x}))) \tag{3.8}$$

where $I_i(\mathbf{T}(\mathbf{x}))$ represents the intensity of transformed location $\mathbf{x}$ in the space of image i, and $H(I_i(\mathbf{T}(\mathbf{x})))$ represents the entropy of this voxel given the set of images I.

Given that groupwise registration often requires many images, it is time-consuming to calculate this metric for every voxel location. Instead, stochastic sampling is used to consider only a random sample of voxel locations. Approximating the expectation to be the sample mean of the population, this gives an objective function of:

$$S = -\frac{1}{n_I} \sum_{j=1}^{n_S} \sum_{i=1}^{n_I} \log p(I_i(\mathbf{T}_i(\mathbf{x}))) \tag{3.9}$$

where n_S denotes the number of random sample locations. These are not held fixed during the process, but are re-generated at every iteration of the algorithm.

To ensure that the atlas image is at the centre of the population, after each iteration, each current transformation is composed with the inverse of the mean of the transformation matrices, in the same way as in Section 3.5.1.

3.6 Pairwise nonrigid registration for average atlas construction

As discussed in the previous chapter, affine registration is not enough to account for the variation seen in brain images. Non-rigid registration is needed to align structures more locally. Methods for average non-rigid atlas construction based on existing methods of pairwise registration can be classified into two groups:

1. Post-processing of transformations, obtained using pairwise registration from subjects in a population to a chosen reference image, to give an average transformation. The inverse of the average transformation can be composed with the original transformations to give transformations from the average shape to each subject of the population. This is shown diagrammatically in Figure 3.6, where the affine transformations can be replaced by non-rigid transformations.

2. Repeated registration to the current approximation of the average shape, each time improving the consistency of the registration and the moving closer to the actual average shape. These methods can additionally involve averaging the transformations at each iteration to ensure that the atlas shape is indeed the average of the population.

As these methods are still essentially pairwise, no new algorithms for simultaneous registration or for multi-subject similarity assessment need to be developed.

3.6.1 Averaging transformations

Christensen et al. [42] extend the work by Miller et al. [169, 98] which defines the average shape as that which has the minimum mean squared error to the rest of the population. A subject from the population is chosen to be a reference image, and this is registered to all other subjects in the population using the ICLEIR method described in Section 3.4.1. These registrations are performed in an Eulerian frame of reference which moves with the transformation. The registration of a target

reference I_1 to a source I_2 gives a transformation $\mathbf{T}_{1,2}$ in the coordinate system of I_2. The coordinate system is therefore different for different source images. However, to compute their average, all transformations have to be in a common (Lagrangian) coordinate system. This can be achieved by using the inverse transformation $T_{1,2}^{-1} = T_{2,1}$, since the same reference subject always used, and the registration is inverse consistent. The average Lagrangian transformation from the reference coordinate to the average coordinate system is then given by:

$$\mathbf{T}_{ref,ave}^{-1}(\mathbf{x}) = \frac{1}{n_I} \sum_{i=1}^{n_I} \mathbf{T}_{ref,i}^{-1}(\mathbf{x}) \ \forall \mathbf{x} \in \Omega \tag{3.10}$$

An arithmetic mean is used since the parameterisation of the transformation is based on displacements and not on explicit scaling.

Avants [17] and Beg [23] have separately proposed methods for averaging geodesic flows. The approach taken in [23] relies on the conservation of the momentum of flow. Given a set of transformations to a chosen subject, an average initial velocity vector can be computed:

$$\bar{\mathbf{v}}_0 = \frac{1}{n_I} \sum_{i}^{n_I} \mathbf{v}_{i,0} \tag{3.11}$$

This average velocity vector is propogated forward using the principle of the conservation of momentum to reach an average shape. The geodesic evolution guarantees that the transformation is diffeomorphic. This process can be iterated by re-registering the subjects in the population to the new atlas and repeating the procedure. The mean velocity calculated in Equation 3.11 is also used in [17]. Additionally, the mapping of each subject to the average image is calculated such that:

$$\forall i \quad I_i(\mathbf{T}_i^{-1}) = \bar{I} \tag{3.12}$$

and the energy of the transformation, $E(\mathbf{T}_i)$, is minimal. The energy is given by the sum of squared distances as defined by Equation 2.20, in Section 2.1.4.3. In the

pairwise case, this ensures that the registrations are additionally inverse consistent (proof of this is given in [17]).

A method for averaging FFDs, which has been used to construct an unbiased atlas from 25 MR images of a population with schizophrenia, was developed in work by Rueckert et al. [206]. To do this, conventional pairwise registration between a chosen reference subject and each of the other subjects is used to create an initial atlas, using a registration model based on global and local components. The global transformation is represented by an affine transformation allowing for rotations, translations, scaling and shearing. The local component allows more detailed deformation to be obtained and uses a free-form deformation model based on B-splines. A uniformly-spaced grid of control points is overlaid onto the reference subject. Displacing the control points deforms the underlying image, which is interpolated using B-splines:

$$\mathbf{T}(x,y,z) = \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} B_l(u)B_m(v)B_n(w)\phi_{a+l,b+m,c+n} \tag{3.13}$$

where $a = \lfloor \frac{x}{n_x} \rfloor - 1, b = \lfloor \frac{y}{n_y} \rfloor - 1, c = \lfloor \frac{z}{n_z} \rfloor - 1, u = \frac{x}{n_x} - \lfloor \frac{x}{n_x} \rfloor, v = \frac{y}{n_y} - \lfloor \frac{y}{n_y} \rfloor, w = \frac{z}{n_z} - \lfloor \frac{z}{n_z} \rfloor$ and where B_l represents the l-th basis function of the B-spline. To guide the transformation, the model is optimised such that the similarity between the two images is maximised. This similarity is determined by evaluating the normalised mutual information of the two images:

$$S_{NMI} = \frac{H(I_1) + H(I_2)}{H(I_1, I_2)} \tag{3.14}$$

where $H(I_X)$ represents the marginal entropy of image I_X and $H(I_1, I_2)$ represents the joint entropy of the two images. The transformation parameters are varied until the NMI is maximised, using the steepest descent algorithm for unconstrained, non-linear optimisation.

In this way, deformation fields are obtained which map each subject to the reference

subject individually. An initial atlas, in the coordinate system of this reference, is created by transforming each subject to this space, using these mappings. The inverse of the mean of all the deformation fields is then used to transform the initial atlas into an atlas at the natural coordinate system of the population. If the registration was perfect, this atlas would represent the average of the population regardless of the initial reference chosen. In practice, however, although the bias is greatly reduced, residual errors in the registration process can influence the average deformation field, and thus also the final atlas produced.

As the transformations are parameterised on a grid of control points, where the movement of the control points is small and constrained to be linear displacements, the arithmetic mean of these displacements can be used. As the deformations are linearly-dependent on the control point displacements (see Equation 3.13), taking the mean of the displacements at each control point will give the mean deformation field for the population.

However, it should be noted, that it is not trivial to invert a deformation field represented by B-splines. One method to do this is to approximate the inverse transformation using a numerical method such as the Newton-Raphson process [198].

3.6.2 Iterative pairwise registration to average shape

Guimond et al. [100] have worked on creating an atlas at the average of a population. The registration method uses the demons algorithm [239], a variant of optical flow, described in Section 2.1.4.4. To find the average atlas, the subjects in the population are first registered to a chosen reference using a low-dimensional transformation to account for global shape and intensity differences only. An atlas (model) is created in the coordinate system of the chosen reference. Elastic registration is then used to correct for residual differences, by registering each subject in turn to the atlas

created. The inverse of the average elastic transformation is then used to create a second atlas in the average space of the population. This new atlas is then used as a reference for the next iteration when it is registered to all subjects of the population. The process of average atlas construction, followed by registration of this atlas to the population of subjects, is repeated until the atlas converges. The atlas created at each iteration therefore acts as the reference subject for the next iteration.

One potential source of errors with this is that elastic registration first takes place using an affine atlas where the structures are not clearly defined. It is questionable as to whether structural details could ever be recovered during the registration. Also, this method still requires the initial choice of a reference subject, which may influence the final atlas.

Kovacevic et al. [130] have also developed an iterative pairwise registration method to align images in the average space. They first create an average affine atlas using Woods's matrix averaging and reconcilliation scheme described in Section 3.5.2. The similarity metric used is also the ratio image uniformity. A non-rigid registration is then used to individually register each subject in the population to the average atlas produced after an affine alignment. This initially uses subsampled images with lower resolution. The process is then repeated, with increasing image resolutions, until convergence. To evaluate the similarity of the images at the non-rigid stage, a similarity function based on the cross-correlation of image intensities is used. Each transformation is then composed with the average of the inverse of all the transformations to centre the reference space at the average shape.

Unlike the work by Guimond, this method does not require any initial choice of reference image. However, a similar potential problem is that the first stage of the non-rigid registration involves registering a reference formed from an affine atlas of the population. This could lack structural detail and might lead to information

being lost at the start of the registration.

3.7 Groupwise non-rigid registration

The previously described work on non-rigid atlas construction involves pairwise registrations. In the methods which require an initial choice of reference, the final atlas may still be biased by this choice. Although this is not a feature of the work by Kovacevic et al., their method involves pairwise registration to the current best estimate of the atlas. The initial estimate of the average is an affinely-aligned atlas. Although this estimate gets progressively sharper, it may still lack the structural detail obtainable when registering two original intensity images. An additional issue with some of the above methods is the need to accurately invert transformations; this is not always possible. These problems motivate the development of pure *groupwise* registration techniques, where all subjects in the population are considered simultaneously. The following methods are not biased by the choice of any particular reference image.

3.7.1 Penalising displacement from average shape

The work by Studholme et al. [223, 224] simultaneously aligns the group of images to a common space using high-dimensional non-rigid registration. A cost function is optimized with the aim of maximising the similarity between the images, while penalizing displacement of the reference space from the average shape.

From the reference space, $\mathbf{x}_R$, the displacement of each location to each subject is given by $\mathbf{u}_i(\mathbf{x}_R)$. The aim is to make the total displacement:

$$U(\mathbf{x}) = \left\| \sum_{i \in 1,\ldots,N} \mathbf{u}_i(\mathbf{x}) \right\|^2 \tag{3.15}$$

have zero length. This is used as a penalty term in the optimization.

The groupwise similarity is evaluated using a local measure of self-information. For each location $\mathbf{x}$, the similarity is given by:

$$S(\mathbf{x}) = -\log p(i'_1, i'_2, ..., i'_{n_I}) \tag{3.16}$$

where $i'_. = I(\mathbf{T}_i(\mathbf{x}))$ is the intensity of the transformed voxel. This is averaged over the reference volume to give a measure of the overall average information. This results in a very sparse, high dimensional distribution, making the probabilities difficult to estimate: using a traditional histogram is computationally impractical. Instead a two-step approach is used: the data is pre-clustered (binned) into a sparse matrix type data structure and then the cluster count is evaluated. The clustering stage is done in parallel, taking regular subregions of the image. The cluster count, $c(\mathbf{B})$, is evaluated using:

$$c(\mathbf{B}) = \sum_{\mathbf{x}\in\Omega} \tau(\mathbf{B}, I(\mathbf{x})) \tag{3.17}$$

where $\tau()$ is an intensity kernel determining the contribution of a set of intensities $\mathbf{I} = \{i_1, i_2, ..., i_{n_I}\}$ to the cluster centred at $\mathbf{B} = \{b_1, b_2, ..., b_{n_I}\}$.

Overall, this produces a cost function to be minimised:

$$C = \int_{\mathbf{x}} S(\mathbf{x}) + \lambda_1 \int_x U(\mathbf{x}) + \lambda_2 \int_x R(\mathbf{x}) \tag{3.18}$$

where $R(\mathbf{x})$ represents an regularising penalty term to ensure smoothness of the transformations. λ_1 and λ_2 are constants which determine the influence of each of the terms representing each geometric constraint.

The algorithm has been tested on a population of synthetically-generated deformed spheres and on 32 adult brain MR images. However, the method requires explicitly

choosing weighting parameters to specify the influence of the penalty terms and thus how well the average shape constraint is satisfied: the final atlas therefore need not necessarily be the average shape of the population. An additional issue is that the similarity metric proposed does not scale linearly with increasing numbers of subjects, making its use for large numbers of subjects computationally expensive.

3.7.2 Minimum message length

There has also been much work on groupwise diffeomorphic non-rigid registration using bounded diffeomorphic deformations (warps) [158, 54], involving Minimum Description Length (MDL) and Minimum Message Length (MML) similarity measures. These principles state that given data and a choice of models, the model which gives the shortest description of the model plus the conditional description of the data should be chosen. This description is given by the algorithmic, or Kolmogorov, complexity, which can be shown to approximate entropy [58]. The difference between MDL and MML is that in MDL, data which are not of interest are contained in the second part of the code, whereas in MML, these data are contained in the first part of the code (the model). In its application to groupwise registration in [158, 54], the MML is used to picked the "best" reference subject from a population. The MML is calculated using all but one of the images in the group in turn. The length of the message required to transmit the left-out subject is then found (Equation 3.20). The reference subject is chosen to be the subject from the population that minimises the MML. Pairwise registration to a selected reference is initially carried out to obtain an initial estimate of deformations.

Leaving out each subject in turn, the similarity of the group is then calculated using a sum-of-absolute-differences metric:

$$S_{SAD}(i) = \sum_{\mathbf{x}\in\Omega} \frac{|I_i(\mathbf{x}) - \bar{I}(\mathbf{x})|}{w_{\mathbf{x}}} \tag{3.19}$$

where $\bar{I}(\mathbf{x})$ is the mean intensity at location $\mathbf{x}$, excluding image i, and $w(\mathbf{x})$ is the mean absolute difference from the mean intensity at $\mathbf{x}$.

The individual deformation of the omitted subject is then optimised to minimise the message length of the group, defined by:

$$C_i(\mathbf{T}) = -\log P_i^{(I)}(I_i(\mathbf{T}_{ref,i})) - \lambda \log P_i^{(S)}(I_i(\mathbf{T}_{ref,i})) \tag{3.20}$$

where $I_i(\mathbf{T}_{ref,i})$ gives the image to be transmitted given the reference subject I_{ref} and the associated transformation $\mathbf{T}_{ref,i}$ and $P_i^{(I)}$ and $P_i^{(S)}$ represent probability density functions associated with shape and texture of the group, excluding example i, and λ represents a weighting term to determine their relative influence.

The process is repeated for all subjects until all the deformations are updated, and the results converge to select a reference subject from the population. The method has been applied to 16 2D MR brain slices and 51 face images. Although this method aims to optimize the correspondences between the group of images, it still requires choosing one of the subjects to be a reference. The algorithm is also computationally expensive for large numbers of 3D subjects, requiring multiple registrations between the population.

A further extension to this approach for groupwise registration is given in [252, 251]. These aim to minimise the total description length, $\mathcal{L}$, of the problem given by:

$$\mathcal{L}_{total} = \mathcal{L}_{ref}(R, I_{ref}) + \mathcal{L}_{params} + \mathcal{L}_{group} + \mathcal{L}_{residuals} \tag{3.21}$$

where $\mathcal{L}_{ref}(R, I_{ref})$ gives the length of the reference frame and reference image, $\mathcal{L}_{params}$ is the length of the parameters of the groupwise model, $\mathcal{L}_{group}$ is the encoding of the groupwise model and $\mathcal{L}_{residuals}$ is the encoding of the residuals. When

transmitting a histogram of an image with n_m voxels in the image having an intensity of m, and occupied bins situated at m_α, the associated probability is given by: $p_m = \frac{n_m}{n_\Omega}$. The description length for transmitting this image histogram is then given by:

$$\mathcal{L}_{hist} = -\sum_{\alpha} \ln \frac{m_\alpha}{n_h} + \sum_{\alpha} \mathcal{L}(n_{m_\alpha}) - \sum_{\mathbf{x}} \ln p(I(\mathbf{x})) \tag{3.22}$$

where n_h is the number of bins in the histogram.

3.7.3 Large deformation diffeomorphisms

Joshi et al. [123] have developed an algorithm for the simultaneous registration of subjects using large deformation diffeomorphisms. This generates deformations, $\mathbf{T}_i$, which solve the Lagrangian ordinary differential equation:

$$\frac{d}{dt}\mathbf{T}_i(\mathbf{x}, t) = \mathbf{v}_i(\mathbf{T}_i(\mathbf{x}, t), t); t \in [0, 1] \tag{3.23}$$

This registration is inverse-consistent as the deformations are obtained by integrating velocity fields forward in time and the negative velocity fields backward in time.

In the case of large deformations using diffeomorphisms, a straightforward linear averaging is inappropriate, as the addition of two diffeomorphisms is not necessarily a diffeomorphism. Instead, the reference space is defined to be the space which minimises the sum-of-squared distances to each of the data points. For the non-rigid case, the problem is to estimate the reference space, $\hat{I}$, that requires the minimum amount of deformation, represented by the diffeomorphism $\hat{\mathbf{T}}_i(\mathbf{x})$, to transform itself to the every subject, I_i in the population. If the deformations in the group S are defined by the metric D, and an image dissimilarity metric given by E, then the reference space is found by:

$$\{\hat{\mathbf{T}}_i, \hat{I}\} = \arg \min_{\mathbf{T}_i \in S, I} \sum_{i=1}^{n_I} E(I_i(\mathbf{T}_i), I)^2 + D(\mathbf{T}_e, \mathbf{T}_i)^2 \tag{3.24}$$

where $\mathbf{T}_e$ is the identity transformation, and:

$$D^2(\mathbf{T}_e, \mathbf{T}) = \min_{\nu} \int_0^1 \int_\Sigma \|L\mathbf{v}(\mathbf{x}, t)\|^2 dx dt \tag{3.25}$$

subject to:

$$\mathbf{T}(\mathbf{x}) = \mathbf{x}t \int_0^1 \mathbf{v}(\mathbf{T}(\mathbf{x}, t), t) dt \tag{3.26}$$

L represents a partial differential operator used to introduce a Sobolev norm. The measure of dissimilarity used is the squared error dissimilarity:

$$\sum_{i=1}^{n_I} E(I_i(\mathbf{T}_i), I)^2 = \sum_{i=1}^{n_I} \int_\Omega \left(I_i(T_i(\mathbf{x})) - \frac{1}{n_I} \sum_{j=1}^{n_I} I_j(\mathbf{T}_i(\mathbf{x})) \right)^2 d\mathbf{x} \tag{3.27}$$

This effectively aims to minimise the dissimilarity between each image and the reference space, as well as to minimise the deformation required. It is symmetric because $\mathbf{T}(\mathbf{x})^{-1}$ is calculated by integrating backward in time the negative of the velocity field used to generate $\mathbf{T}(\mathbf{x})$. The minimiser is thus the same for both $\mathbf{Tx}$) and $\mathbf{T}(\mathbf{x})^{-1}$.

Overall, the deformations that optimise the problem solve:

$$\hat{\mathbf{T}}_i = \text{argmin} \sum_{i=1}^{n_I} \int_\Omega [I_i(\mathbf{T}_i(\mathbf{x})) - \frac{1}{n_I} \sum_{j=1}^{n_I} I_j(\mathbf{T}_i(\mathbf{x}))]^2 d\mathbf{x} + \int_0^1 \int_\Omega \|L\mathbf{v}_i(\mathbf{x}, t)\|^2 d\mathbf{x} dt \tag{3.28}$$

The algorithm has been applied to eight, intensity-adjusted, 3D MR brain images of different subjects.

Lorenzen et al. [146] also use the large deformation diffeomorphism framework for groupwise registration, but this time on probabilistic segmentations, k, of the images instead of on the intensity images themselves. It can therefore also be used for the registration of multi-modal images. This aims to find the representative atlas class posterior $\hat{p}$ that requires the minimum amount of energy to deform into all the

other population posteriors p_i. The distance between two probability mass functions P and Q can be represented by the Kullback-Leibler divergence metric [132]:

$$D_{KL}(P||Q) = \sum_k P_k \log \frac{P_k}{Q_k} \tag{3.29}$$

The energy minimisation problem, corresponding to Equation 3.24, but now using probability density functions, becomes:

$$\{\hat{\mathbf{T}}_i,\hat{p}\} = \text{argmin} \sum_{i=1}^{n_I} E(p_i \circ \mathbf{T}_i, p) + D(\mathbf{T}_e, \mathbf{T}_i)^2$$

$$\{\hat{\mathbf{T}}_i,\hat{p}\} = \text{argmin} \sum_{i=1}^{n_I} \int_\Omega D_{KL}(p_i(\mathbf{x})||p_i(\mathbf{T}_i(\mathbf{x})))d\mathbf{x} + \int_0^1 \int_\Omega \|L\mathbf{v}_i(\mathbf{x},t)\|^2 d\mathbf{x}dt \tag{3.30}$$

The PDF, $\hat{p}$, which minimises the above function, is given by the normalised geometric mean of the PDFs of the population:

$$\hat{p}(k_l(\mathbf{x})) = \frac{(\prod_{i=1}^{n_I} p_i(\mathbf{T}_i(k_l(\mathbf{x}))))^{\frac{1}{n_I}}}{\sum_k (\prod_{i=1}^{n_I} p_i(\mathbf{T}_i(k_l(\mathbf{x}))))^{\frac{1}{n_I}}} \tag{3.31}$$

giving the final minimsation problem of:

$$\hat{\mathbf{T}}_i, = \text{argmin} \sum_{i=1}^{n_I} \int_\Omega D_{KL}(\hat{p}(\mathbf{x})||p_i(\mathbf{T}_i(\mathbf{x})))d\mathbf{x} + \int_0^1 \int_\Omega \|L\mathbf{v}_i(\mathbf{x},t)\|^2 d\mathbf{x}dt \tag{3.32}$$

This algorithm has been used to create an unbiased atlas of the brain of normal two-year-olds, using MR images of five subjects. The posterior probabilities of these subjects were derived from multi-modal (T1-weighted, T2-weighted and proton density) scans, using the Expectation-Maximisation algorithm.

3.8 Summary

The ability to construct a representative anatomical atlas of a population is an important tool in medical image analysis, and is particularly useful for brain image

analyis. There are many issues associated with ensuring this atlas is not biased to an arbitrary member of the population. This has motivated the recent development of groupwise registration methods and the construction of atlases at the average coordinate system of the population. These techniques have been used to create atlases representing the average global geometry, the average local geometry and the average intensity of the population. The first two steps ensure crisp detail in the atlas as corresponding structures are locally aligned, while the final step reduces the noise in the atlas compared to the individual intensity images. Furthermore, non-rigid registration allows the structural variation of the population to be encoded within the deformation fields obtained.

However, groupwise registration presents new challenges, in addition to those existing in standard image registration. The aim of the work presented in this thesis is to develop and use groupwise registration to construct unbiased atlases of given populations of subjects. The methods should not require the choice of any arbitrary subject from the population. To do this, it is necessary to develop non-rigid registration techniques that allow for the simultaneous warping of all subjects in the population. Another area which needs to be considered is how to measure the similarity of a group of images. Both these issues must be effectively addressed using methods that are scalable with increasing numbers of subjects.

The work is closest to those developed in Section 3.7. However, in this work, it is asserted that the most representative atlas is one which represents the geometrical average shape of the population. Due to its ability to model complex variation in subjects, a free-form deformation method of registration [207] is extended to allow for the simultaneous groupwise registration of a population of subjects to the average shape of this population.

Chapter 4

Groupwise Registration

The last chapter outlined the advantages of constructing an *average* atlas space for a given population. Most previous work which uses non-rigid registration to do this, has primarily involved the postprocessing of deformation fields produced from *pairwise* registrations between subjects in the population. This often means that a single reference subject still needs to be chosen initially, and this may still bias the final model. Work on pure groupwise registration has mainly used only affine deformations. The methods discussed in this chapter differ in that non-rigid registration is used, acting on all subjects in the population *simultaneously*. The registration itself is used to find the average shape: it is truly *groupwise*. In this respect, it is most similar to the recent work by Lorenzen et al. [146]. They have created a groupwise algorithm for the registration of soft (probabilistic) segmentations to an average space. However, the algorithms which will be described in this chapter can be used to register intensity as well as segmented (both hard and soft) images. Additionally, the deformation model used is different. Whereas Lorenzen uses a fluid deformation model, the work here extends the B-spline algorithm, developed by Rueckert et al. [207] for pairwise registration (as described in Chapter 2).

The aim of this chapter is to develop an *unbiased, groupwise, non-rigid registration* algorithm which:

1. Eliminates the need to choose any reference subject, thus removing any bias from the construction.

2. Creates an atlas space representing the mean shape of the population, via the deformation of all subjects in the population simultaneously.

The first problem encountered when doing this, is that it is not possible to know what the average shape is beforehand. The average space can be defined as the coordinate system which requires the least total deformation to each subject. It is therefore the shape produced when the sum of all deformations is equal to zero. By constraining the deformation to sum to zero, the average shape can be calculated implicitly, (see Figure 4.1). To achieve the above aims, it is necessary to develop

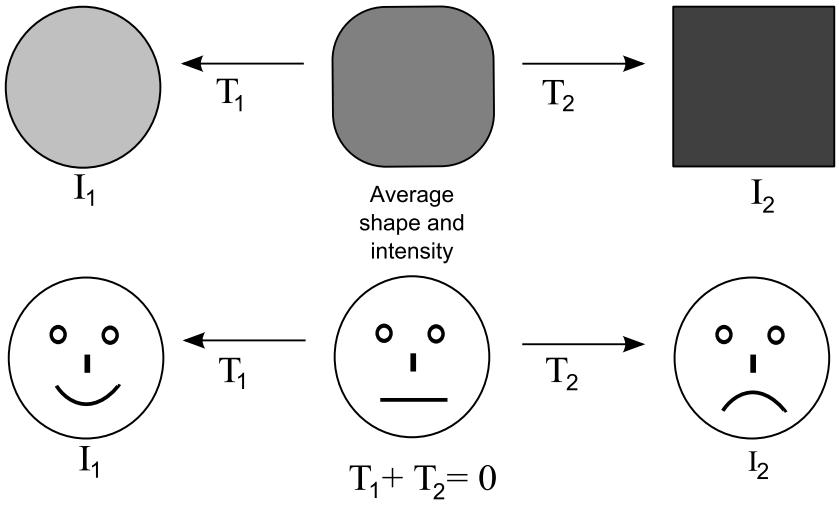


Figure 4.1: Average atlas construction requires the sum of the deformations to equal zero.

the following methods for multi-subject registration:

- **Deformation** (or transformation) **models**, **d**, describing how the reference space maps to each subject in the population.

- A **similarity metric**, S, to determine how well-aligned the whole group of transformed subjects is.

- A **constrained optimisation** strategy, aiming to find the set of deformations

which maximise the similarity metric, while enforcing the constraint that the deformations sum to zero.

These will be discussed in the sections following. The remainder of the chapter is devoted to experiments, using these new methods, on synthetic and real MR data.

4.1 Transformation models

For n images, given a set of points in the spatial domain of a common reference volume, Ω_r, and a set of points in the spatial domain of each individual image $\Omega_i, i = 1...n$, the goal of the registration is to find a set of transformations τ, each of which maps any point $\mathbf{x}_r$ in the reference space to a corresponding point $\mathbf{x}_i$ in image i: $\tau = \{\mathbf{T}_i : \mathbf{x}_r \mapsto \mathbf{x}_i, i = 1...n\}$.

To do this, a registration algorithm based on global and local components is used. For each image, i:

$$\mathbf{T}_i(x, y, z) = \mathbf{T}_{i,global}(x, y, z) + \mathbf{T}_{i,local}(x, y, z) \tag{4.1}$$

The global component of the transformation consists of differences in translations, rotations, scalings and skews. The position and orientation of an image in the field of view can vary even when the same patient is scanned, due to their position in the scanner at the time. Differences in these do not therefore provide any information on differences between subjects in terms of anatomy. To account for global skew and scale differences, the log-averaging method of Aljabar et al. [6] as described in Chapter 3, can be used. The local, non-rigid, component of the transformation describes the detailed differences between brain shapes after affine transformation. It is this part of the transformation which will be developed in this chapter. When using multiple images, the number of degrees of freedom of the problem increases, adding to the importance of a deformation model with local control. Additionally,

because of the large complexity and variation between subjects of the human brain, a free-form deformation (FFD) model based on B-splines is used. The FFD model is a powerful tool for modelling 3D deformable objects, and has previously been applied successfully to numerous medical image registration problems such as [207, 213].

In 3D, the FFD model embeds an $n_x \times n_y \times n_z$ mesh of uniformly-spaced control points $\phi_{a,b,c}$ onto each image. Manipulating the positions of these control points deforms the underlying object, using the B-spline model described in Chapter 2. Increasing the resolution of the control point lattice increases the amount of localised deformation that can be achieved. For the application to groupwise registration, $\mathbf{T}_{i,local}$ is represented by a collection of FFDs, $\mathbf{d}_i$, which are used to deform each subject i to the common reference coordinate system. Each deformation field has the same number and spacing of control points and can be written as the 3D tensor product of 1D cubic B-splines:

$$\mathbf{d}_i(x,y,z) = \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} B_l(u)B_m(v)B_n(w)\phi^i_{a+l,b+m,c+n} \tag{4.2}$$

Moving a control point only deforms an area in a $4 \times 4 \times 4$ vicinity of that control point. The use of B-splines is therefore computationally efficient even with large numbers of control points. Additionally, Equation 4.2 shows that the deformation produced is linearly dependent on the displacement of the control points.

4.2 Multi-subject similarity

In order to determine how well a group of subjects is aligned, it is necessary to define a measure of similarity between the n images. For groupwise registration without a known reference image, this is not so straightforward, as pairwise metrics discussed in Chapter 2 are not easily extendable: even if the formulation itself can be extended, computational difficulties arise, as the dimensionality of the problem

increases. This is sometimes known as the "curse of dimensionality" [25].

4.2.1 Voxel intensity-based metrics

The simplest measure would be a measure of intensity differences from the mean at each voxel location. The sample variance is a potential choice where differences from the mean voxel intensity are summed over the image domain Ω. This has to be normalised by the number of voxels n_V in the image domain, giving:

$$S_{SV} = \sum_{i=1}^{n_I} \sum_{\mathbf{x} \in \Omega} \frac{\left(I_i(\mathbf{x}) - \bar{I}(\mathbf{x})\right)^2}{n_\Omega} \tag{4.3}$$

However, as with all intensity-based metrics, largely varying differences in intensity through the group may adversely affect the above measure. Using this metric on real MR data would therefore require normalisation of the intensities of the images. This can be done using linear intensity correction where linear regression and outlier detection can be used to find the line of best fit on a joint histogram of the intensities of two images [100].

4.2.2 Entropy-based metrics

The variability of MR images means that corresponding structures in different images do not necessarily have the same intensities. Given this, entropy-based metrics are an attractive measure of image similarity, as discussed in Chapter 2. These require methods of estimating probability density functions (PDFs) using kernel density estimators or histograms.

Kernel density estimators scale better with dimensionality [75]. However, in order to achieve good scaling while maintaining accuracy, evaluating the density at each stage is computationally expensive [97, 144]. Although such algorithms are effective for estimating one-off densities of orders even up to 10^6, the need to repeatedly re-calculate densities in image registration, makes their use inefficient.

When using histograms for two images, evaluating the PDFs requires the construction of a 2D histogram. The histogram bins represent ranges of intensities in the two images. At each voxel location, the pair of corresponding intensities are added to the appropriate bin. When the correspondence between the images is high, this is represented by sharper peaks in the histogram.

Generalising Normalised Mutual Information (NMI) to n_I images gives:

$$S = \frac{1}{H(I_1, I_2 ... I_{n_I})} \sum_{i=1}^{n_I} H(I_i) \tag{4.4}$$

where $H(I_i)$ represents the marginal entropy of image i and $H(I_1, I_2, ...I_{n_I})$ represents the joint entropy of all the images. Evaluating n_I-dimensional NMI, would therefore require an n_I-dimensional histogram. Apart from the exponentially increasing memory requirements, this also leads to increasing sparsity as the histogram size becomes very much larger than the number of samples in the histogram. For example given 10 images of size $256 \times 256 \times 256$ and with 64 intensity bins, the histogram size will be $64^{10} = 2^{60}$, but the number of samples will only be $256^3 = 2^{24}$. The evaluation of such sparse histograms is computationally infeasible and so this method cannot be used even for moderate numbers of subjects.

An alternative strategy would be to select one arbitrary image to act as an intensity - but not as an anatomical - reference. The NMI of this reference image and the other images can then be evaluated, with all images (including the reference) warped into the atlas space, using:

$$S_{NMI}(I_{ref}, I) = \frac{H(I_{ref}) + H(I)}{H(I_{ref}, I)} \tag{4.5}$$

where $H(I_{ref})$ denotes the marginal entropy of the intensity reference and $H(I_{ref}, I)$ denotes the joint entropy of the reference and the group. All pairs of intensities, comprising the voxel intensity in the reference and the corresponding intensity in each other subject, could be added to the same joint histogram. A problem with this method however, is that as the number of subjects increases, the effect of each image on the overall similarity measure decreases (the gradient of the similarity for a single image will be reduced as the overall change in the PDF of the histogram will be reduced). This similarity measure may therefore not be strong enough to accurately model small deformations.

Instead the sum of the values of NMI between the reference and each subject could be used, i.e.:

$$S_{NMI}(I_{ref}, I) = \sum_{i=1}^{n} \left(\frac{H(I_{ref}) + H(I_i)}{H(I_{ref}, I_i)} \right) \tag{4.6}$$

where $H(I_i)$ represents the marginal intensity of image i and $H(I_{ref}, I_i)$ denotes the joint entropy between the chosen arbitrary image I_{ref} and image I_i.

The above method still has the undesirable property that an arbitrary image needs to be selected from the population. An alternative is to create an intensity reference image using the average intensities of the population, given the current transformation. Once again, the similarity measure used is the sum of the NMI values of individual histograms, constructed between reference and single subject:

$$S_{ANMI} = \sum_{i=1}^{n_I} \left(\frac{H(\bar{I}) + H(I_i)}{H(\bar{I}, I_i)} \right) \tag{4.7}$$

The reference image $\bar{I}$, is the voxel-wise mean intensity of the group of images and is updated at every iteration.

4.2.3 Segmentation-based metrics

Mutual information is not a strictly additive measure and two equally-well-registered pair of images may not have the same numerical value of MI or NMI, if their intensity or noise profiles differ. This poses a particular problem for groupwise registration where there will be variation in intensities among subjects. This motivates the use of segmentation-based similarity metrics. A segmentation of an image divides the voxels into labelled tissue classes or anatomical structures. Hard segmentations label every voxel in a given structure with the same value. Soft, or probabilistic segmentations assign, to each voxel, the probability of being each structure. Methods for segmentation include clustering algorithms such as the K-nearest neighbour algorithm [85] or likelihood measures such as the Expectation-Maximisation (EM) algorithm [70]. Examples of segmentation algorithms for brain image analysis can be found in [116, 138, 139, 266] and are described in more detail in Chapter 6.

When working with hard segmentations, a label-consistency metric can be used, which measures the overlap of segmentations:

$$S_{LC} = \sum_{i=1}^{n_I} \frac{N(I_i \cap I_{ref})}{N(I_i \cup I_{ref})} \tag{4.8}$$

where $N(I_i \cap I_{ref})$ represents the number of voxels in both image I_i and the reference image having the same label and $(I_i \cup I_{ref})$ is the total of the number of voxels labelled in I_i and the reference. The reference model is maximum probability estimate [103, 105], created by assigning to each voxel the class representing the mode of the group (i.e. the most commonly occuring class for that voxel).

A problem with registering hard segmentations is that it lacks sensitivity as large regions are labelled homogeneously. If labelled regions span too many structures, there is no way of determining how well individual structures are aligned within the labelled area. For registration with hard segmentations to work well, many individ-

ual structures would have to be labelled in each image, or the optimisation would have to start at some initial solution close to the optimum. An alternative is to use soft (probabilistic) segmentations. The *relative entropy* or *Kullback-Leibler divergence* (D_{KL}), [132, 58] is an information theory metric that represents the distance between a model probability distribution, Q, and an observed data distribution P. It can be regarded as the inefficiency of assuming a distribution of Q, when the true distribution is P. It has recently been used as a similarity measure for image registration in [44, 88] and in particular for groupwise registration by Lorenzen et al. in [145, 146].

$$D_{KL}(P||Q) = \sum_k P(k) \log \frac{P(k)}{Q(k)} \tag{4.9}$$

Given a set of n_I images, probabilistic segmentations for each of n_k tissue classes can be obtained via a segmentation algorithm such as the EM algorithm [138, 139]. The model PDF for a single class is then the mean PDF of the group. The Kullback-Leibler divergence between this reference model and each of the other images (observed data) can then be calculated. As it is an additive measure, the overall similarity is simply the sum of the distances between the model and each individual PDF:

$$S = \sum_{i=1}^{n_I} \sum_{\mathbf{x}\in\Omega} \sum_{k}^{n_k} p_{i,\mathbf{x},k} \log \frac{p_{i,\mathbf{x},k}}{p_{ref,\mathbf{x},k}} \tag{4.10}$$

where $p_{i,\mathbf{x},k}$ is the probability of voxel location $\mathbf{x}$ in image i being classified as tissue class k, and:

$$p_{ref,\mathbf{x},k} = \frac{\sum_i p_{i,\mathbf{x},k}}{n_I} \tag{4.11}$$

The Kullback-Leibler divergence can be related to entropy measures as follows:

$$D_{KL}(P||Q) = \sum_k P(k) \log \frac{P(k)}{Q(k)} = -\sum_x P(k) \log Q(k) + \sum_k P(k) \log P(k) \tag{4.12}$$

which is the difference between the *cross-entropy* of P and Q and the entropy of P. D_{KL} is always non-negative [58] and is equal to zero if and only if P=Q, i.e. the assumed and actual probability density functions are exactly equal. The measure is

not symmetric and so, in general, $D_{KL}(P||Q) \neq D_{KL}(Q||P)$. However, since we are trying to minimise the distance between the model and the individual subjects, and not the distances between individual subjects, this asymmetry is acceptable as the same model is used in each case.

4.3 Constrained optimisation: primal methods

Regardless of the metric used, a set of transformations needs to be found that will maximise the similarity. However, in addition to this, it is not possible to know what the average shape of the population is before the registration. The average coordinate system represents the coordinate system that requires least total deformation from itself to all other members of the population. The average space is therefore calculated implicitly, by constraining the sum of all deformations, from this space to each subject, to be equal to zero: *Maximize:* $S(\mathbf{d})$
Subject to:

$$\sum_{i=1}^{n_I} \mathbf{d}_i(\mathbf{x}) = 0 \quad \forall \mathbf{x} \in \Omega \tag{4.13}$$

where the objective function $S(\mathbf{d})$ denotes a measure of the similarity, such as those given in the previous section. Since the deformation fields are linearly-dependent on the control points displacements (Equation 4.2), this is equivalent to constraining the sum of the displacements, d, of each control point ϕ, to be zero:

$$\sum_{i=1}^{n_I} d_{\phi i} = 0 \quad \forall \phi \in \Phi \tag{4.14}$$

A control point lattice is created for each subject. These control points are displaced, deforming the underlying image, until the similarity between the group of images is maximised, subject to the constraint being satisfied. This could be solved using a number of optimisation schemes for constrained optimisation. However, as the objective function is non-linear, but the constraints are linear equality constraints (and are therefore all active throughout the process), *primal methods* [147, 22] are

attractive methods to use. These methods are search methods where the original objective function is optimised by searching through regions in which the constraints are always valid. Each solution generated by the process therefore increases the objective function while always being feasible. Additionally, primal methods do not require any particular structure of the initial function (e.g. convexity). In this work, the *Gradient Projection Method (Rosen)* method is used to optimise the deformations as this integrates well with the steepest descent method of unconstrained optimisation used in [207] for registration. This enables registration to be easily performed with or without the constraint according to the user requirements. Full details of constrained optimisation techniques can be found in [147] and [22]. For linear constraints, as in this formulation, convergence rates of these primal methods are often very efficient [147].

Rosen's Gradient Projection Method [204], is comparable to the method of steepest descent for unconstrained optimization. For the simple case of a single, linear equality constraint, Gradient Projection is a suitable strategy. This projects the negative of the objective function gradient onto the active constraints, by multiplying the gradient by a projection matrix, $\mathbf{P}$. The projection matrix is calculated by:

$$\mathbf{P} = \mathbf{I} - \mathbf{A}(\mathbf{A^T A})^{-1}\mathbf{A^T} \tag{4.15}$$

where $\mathbf{A}$ represents a vector containing the coefficients of each constraint and $\mathbf{I}$ is the identity matrix. For each degree of freedom, a separate constraint such that the sum of movements for each image sum to zero is needed. This gives a square $\mathbf{P}$ matrix of side *number of degrees of freedom* $\times$ *number of subjects*. However, since the constraints are simple and repeated, the matrix reduces to a block diagonal matrix, and is thus computationally feasible. For example, in the case of two images and four control points each with one degree of freedom (Figure 4.2), the problem becomes:

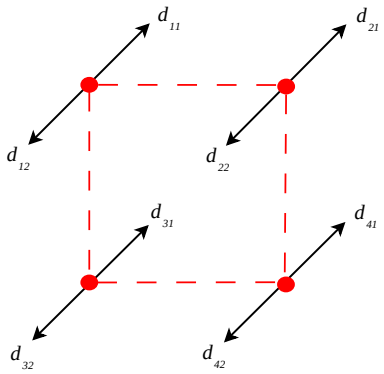


Figure 4.2: A grid of four control points, showing the displacements $d_{\phi i}$ for each image i, at each control point ϕ.

Maximise: $S(\mathbf{d})$

Subject to:

$$\begin{aligned} d_{11} + d_{12} &= 0 \\ d_{21} + d_{22} &= 0 \\ d_{31} + d_{32} &= 0 \\ d_{41} + d_{42} &= 0 \end{aligned}$$

The projection matrix then becomes:

$$\mathbf{P} = \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} \tag{4.16}$$

If there were three images, the projection matrix obtained from Equation 4.15 would be:

$$\mathbf{P} = \frac{1}{3}\begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 2 & 0 \\ 0 & 0 & 0 & \ddots \end{bmatrix} \tag{4.17}$$

At each iteration, the directions of movement of the control points δ are found by calculating the gradient $\mathbf{g}$ of the similarity function with respect to the control point displacements. To satisfy the constraints, the gradient vector is then multiplied by the projection matrix and then normalised. At this stage the method reduces to steepest descent along the constraint.

$$\delta = -\frac{\mathbf{Pg}}{\|\mathbf{Pg}\|} \tag{4.18}$$

4.3.1 Initial feasible solution

The Gradient Projection method of optimisation requires starting from a initial feasible position on the constraint surface, such as zero displacement for all control points and for all subjects. However, a more efficient method, would be to start at a solution closer to an optimal solution. To do this, a pairwise registration can be done first, as described in Chapter 3. The inverse of the mean of the pairwise deformations can be calculated using a numerical scheme [198]. This can be concatenated with each of the initial pairwise transformations to obtain an initial set of feasible deformations ($\mathbf{T}'_i, i \in \{1...n_I\}$) which sum to zero (see Figure: 4.3):

$$\mathbf{T}'_i = \mathbf{T}_i \circ \bar{\mathbf{T}}^{-1} \tag{4.19}$$

where

$$\sum_{i=1}^{n_I} \mathbf{T}_i \circ \bar{\mathbf{T}}^{-1} = 0 \tag{4.20}$$

where $\bar{\mathbf{T}}^{-1}$ is the inverse of the mean deformation field given by:

$$\bar{\mathbf{T}} = \frac{\sum_i^{n_I} \mathbf{T}_i}{n_I} \tag{4.21}$$

Instead of starting the groupwise registration algorithm with deformation fields with no initial deformation, the deformation fields calculated by Equation 4.19 can be used as initial solutions to the algorithm, and updated using the methods described in the previous section. This has the advantage that the initial solution can be calculated from pairwise registrations. Each registration can therefore be run simultaneously on different computers (for example, by using Condor [143, 236]), thereby saving overall processing time.

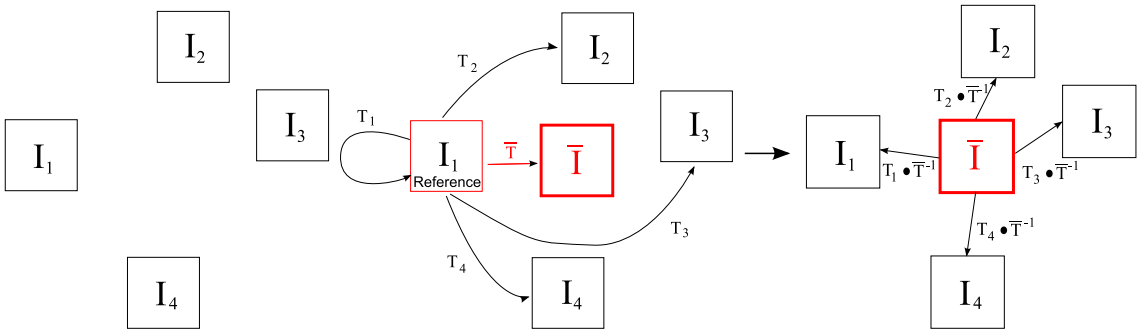


Figure 4.3: How an initial average can be calculated from pairwise registrations

4.4 Parallelisation of algorithm

Given the size and number of images, running the algorithm on a single processor can take a long time. The largest bottleneck will be in the evaluation of the derivative of the similarity measure for each control point [119]. A single 100×100×100 control point grid already has 1000000×3 degrees of freedom, making groupwise registration of multiple images very computationally expensive. Fortunately, the local nature of the transformation means that the derivative of the similarity measure at each control point is only affected by control points in a small neighbourhood around it (see Section 4.1). This makes this part of the algorithm very suitable for parallelisa-

tion: each region of control points can be handled separately by a different processor.

A Message-Passing Interface (MPI) [99] formulation has been used to distribute the calculation of the derivative of the similarity metric across multiple computers. For simplicity, all machines being used run the whole of the remainder of the code. Further extensions to this would be to store the images across multiple machines, increasing the number of images that could be simultaneously registered, as in [119], and to parallelise the evaluation of the similarity metric, which forms the next largest bottleneck.

4.5 Experiments

4.5.1 Synthetic Images

To test the algorithm, we have first created purely synthetic 3D images of a cube and a sphere (shown in the diagrams only in 2D). We know that the average shape of these shapes is in-between the pair. The results, obtained using the sample variance similarity measure (Equation 4.3), are shown in Figure 4.4. As can be seen, the final images are of the same expected shape, showing that the alignment has been successful. Additionally, the deformation fields are shown. When the deformations at each control point are added together for each control point spacing, deformation fields with zero deformation are obtained, showing that the resulting shape does represent the average of the group.

4.5.2 Artificially-deformed data

The groupwise registration algorithm and the similarity measures developed have been tested on a dataset of simulated brain data. To create a population of subjects, an initial image can be deformed by known deformation fields. If the total deformation for each degree of freedom over all the deformation fields is equal to

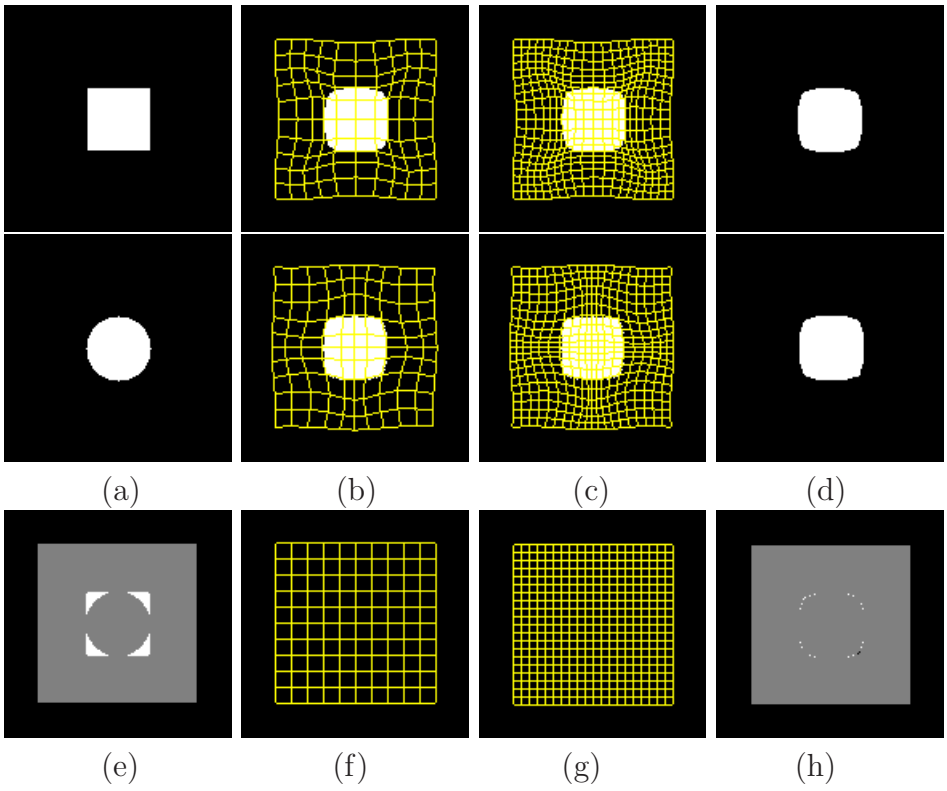


Figure 4.4: Groupwise registration of a cube (top row) and sphere (second row). (a) Initial images; (b),(c) images and deformation fields using grids of 10mm and 5mm control point spacings respectively; (d) final images; (e) initial difference image (f),(g): total deformation field at 10mm and 5mm spacings respectively; (h): final difference image.

zero, the original image will represent the mean shape of the population. An initial set of 50 deformation grids was generated, each with a uniform control point spacing of 10mm. At each corresponding control point across the 50 images, a random displacement is applied such that the displacements across the images form a Gaussian distribution with zero mean at that location. Within any given image, the displacements at different control points are generated independently. The average displacement at each control point can therefore be expected to be close to, but not necessarily exactly, zero. Given the large number of images used, the total displacement at each control point could therefore be quite high. To ensure that the sum of the displacements at each control point do sum to exactly zero, and therefore that the average of the population is the original image, a further 50 subjects were produced. For each existing deformation field, another is created to have the negative of the existing displacement at each control point. This results in a set of 100 deformation fields which have a total of zero displacement at each control point. The *inverse* of each deformation was applied to a single 2D slice of the MNI Brainweb image [51] (Figure 4.5) to produce a population of 100 subjects, each of size 216×180 voxels (Figure 4.6), having a mean shape which is the original image. The original MNI Brainweb image itself was not included in the population. The reason for the use of the inverse transformation is that B-spline deformation fields are themselves difficult to exactly invert. It is easier to transform an image by the inverse of the deformation field using an iterative numerical scheme such as Newton-Raphson [198], which gives accurate inversion given a smooth deformation. By doing this, the deformation field needed to recover the MNI Brainweb image for each subject is known. When assessing the performance of the registration, it should be noted that there is inherent bias in that the deformation fields are defined on the same grid as the registration, and the interpolation the same (B-spline) in both cases. Additionally, all images have the same intensities and same levels of noise. However, the aim of these experiments is to investigate the accuracy of each groupwise similarity metric, and comparison with pairwise approaches, and in all

cases, the same B-spline formulation is used for the registration.

The MNI Brainweb image also has ground truth segmentations for four tissue classes - white matter (WM), grey matter (GM), cerebro-spinal fluid (CSF) and background (BG) - as well as probabilistic segmentations for these same classes. These segmentations were transformed to align with the new population in the same way as with the intensity images. The resulting population was non-rigidly registered using the groupwise registration approach described, using control point grids of 20mm, 10mm and 5mm. The following similarity measures were tested:

- Sample variance (SV)
- Sum of NMI using the average intensity as a reference (ANMI)
- Kullback-Leibler divergence using known probabilistic segmentations of the slice (KL)
- Label consistency using ground truth hard segmentations (LC)

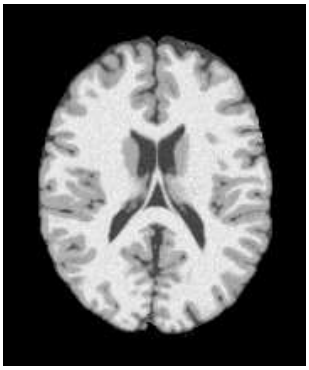


Figure 4.5: Original MNI Brainweb slice

These measures were also compared to registration using a pairwise scheme. In the pairwise case, one of the sample images was chosen to be a reference (in this case, the first image shown in Figure 4.6), and was registered to all the subjects in the population using 20mm and 10mm control point lattices. The same B-spline deformation model was used with a normalised mutual information similarity metric.

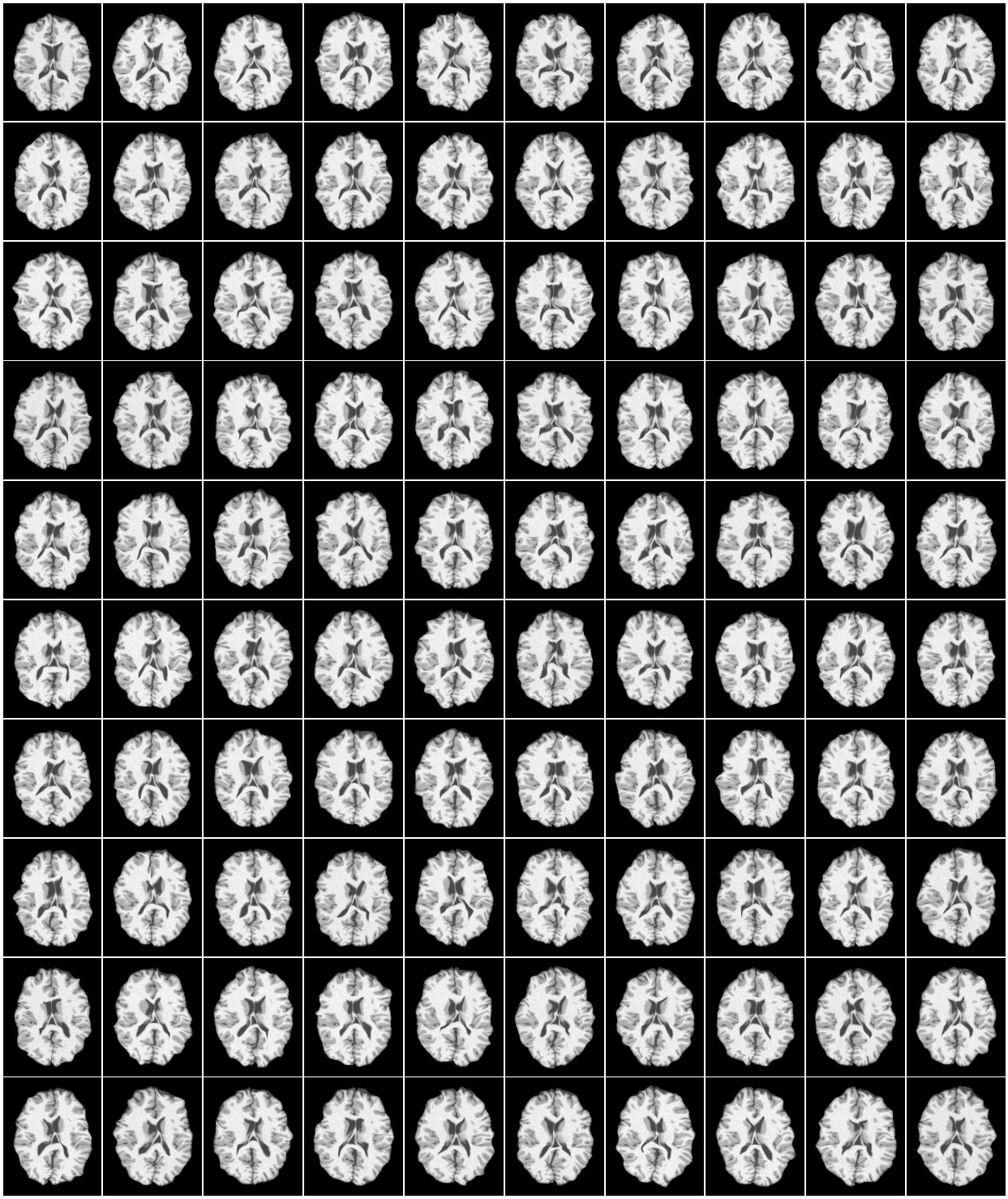


Figure 4.6: The population of 100 artificially deformed subjects. The deformations were produced by applying Gaussian random displacements to each control point of a 10mm grid, with a maximum displacement of 21mm and mean displacement of zero. At each control point, the displacements sum to zero over the population.

The resulting atlases (in the coordinate system of the chosen reference), were transformed by the inverse of the mean deformation field (using the numerical scheme described in [198]), into a coordinate system which should, if the registration was perfect, lie at the centre of the population (see Figure 4.3). This method of average atlas construction was proposed by Rueckert et al. in [206]. In practice, however, errors in the registration process are likely to occur. To account for these, the process is then repeated using the average atlas, produced as above, as the new reference subject [100]. This is re-registered to all subjects and a new average is then found using the mean of the new deformation fields. This new average image then acts as a reference for the next iteration. The results here show four iterations of this process of calculating an average atlas and re-registering to this atlas. These are denoted in the graphs by $P(i = x)$ where x is the iteration number.

A sample of the original population and the final transformed subjects after various registration techniques can be found in Figure 4.7. The final atlases created using each method are compared to the MNI Brainweb average in Figure 4.8. This shows how all the registration techniques transform the samples to shapes closer to the average shape. Although it is difficult to distinguish from these images alone which similarity measure is best, it can be seen from Figure 4.7 (where the individual images shown are less similar to the Brainweb image, and the final atlas is more blurry than using the other methods), that most groupwise techniques outperform one iteration of the pairwise method. The exception to this, is when registering using hard segmentations of the subjects, using the label consistency metric. This may be because using hard segmentations lacks the sensitivity required for accurate registration.

To analyse the results in more detail, two quantities are considered:

1. ***Accuracy***: how well the registration recovers the average MNI Brainweb shape for each image.

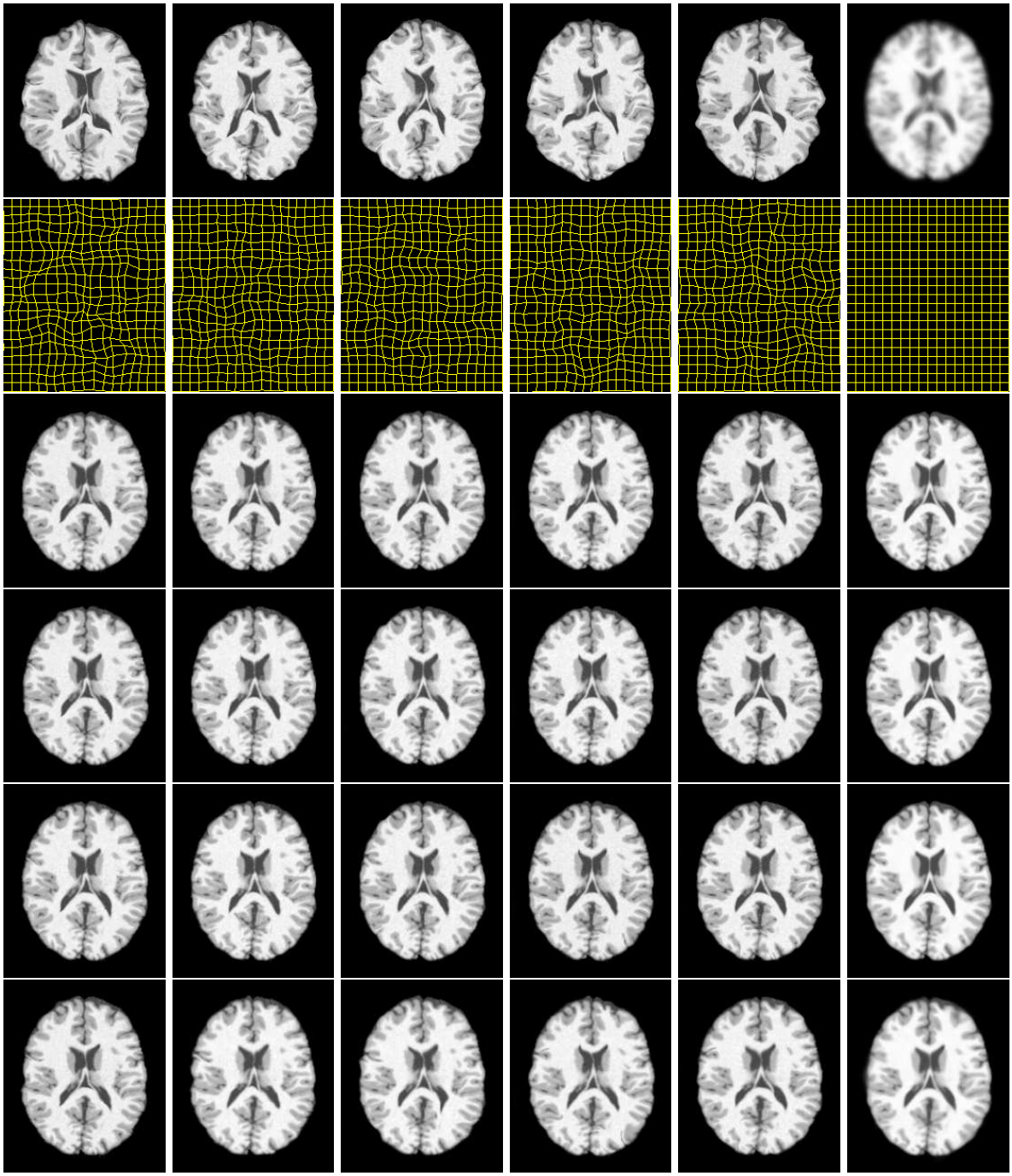
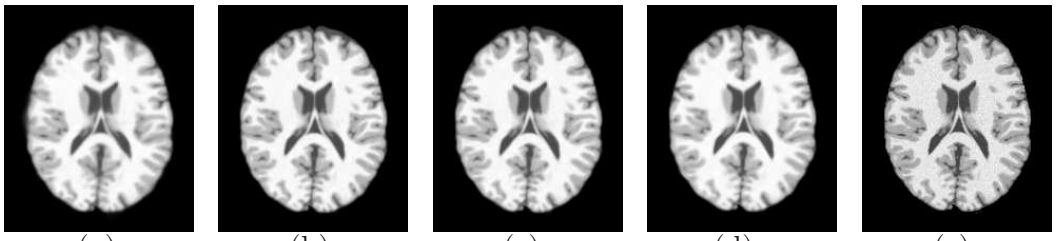


Figure 4.7: Top row: five samples of the unregistered population and the atlas of the whole population of 100 subjects (far right). Second row: the deformation fields produced using groupwise registration (KL) for each image and the total deformation field (far right). The same samples after groupwise registration with: Kullback-Leibler (row 3), ANMI (row 4) and sample variance (row 5). The samples after pairwise registration and transformation to average space (row 6). The images in the far right column show the atlases of the population after registration using the corresponding metric.



(a) (b) (c) (d) (e)

Figure 4.8: Atlases after pairwise (a) and groupwise (using KL (b), ANMI (c), SV (d) similarity metrics) registration. Far right: original Brainweb image representing actual mean shape of population.

2. ***Consistency***: how well-aligned the subjects in the group are with each other.

To determine (1), three measures have been computed. The sum-of-squared differences (SSD) between the voxel intensities in the MNI Brainweb image and in each resulting image have been found:

$$SSD = \sum_{\mathbf{x}\in\Omega} \frac{(I_{ref}(\mathbf{x}) - I(\mathbf{x}))^2}{n_\Omega} \tag{4.22}$$

The mean and standard deviations of the SSDs of the population, produced using each similarity measure, are shown in Figure 4.9, together with the mean and standard deviation of the SSD between the original population and the MNI Brainweb image.

The second measure looks at the overlap of the tissue classes of the MNI Brainweb image and each transformed image. This is determined using the Dice similarity metric [72]:

$$D = \frac{2 \times N(I \cap I_{ref})}{N(I \cup I_{ref})} \tag{4.23}$$

which is twice the ratio of the number of voxels correctly labelled to the total number of voxels with that label in both the reference and image under consideration. The average Dice overlaps of the population for BG, CSF, GM and WM are shown in Figure 4.10.

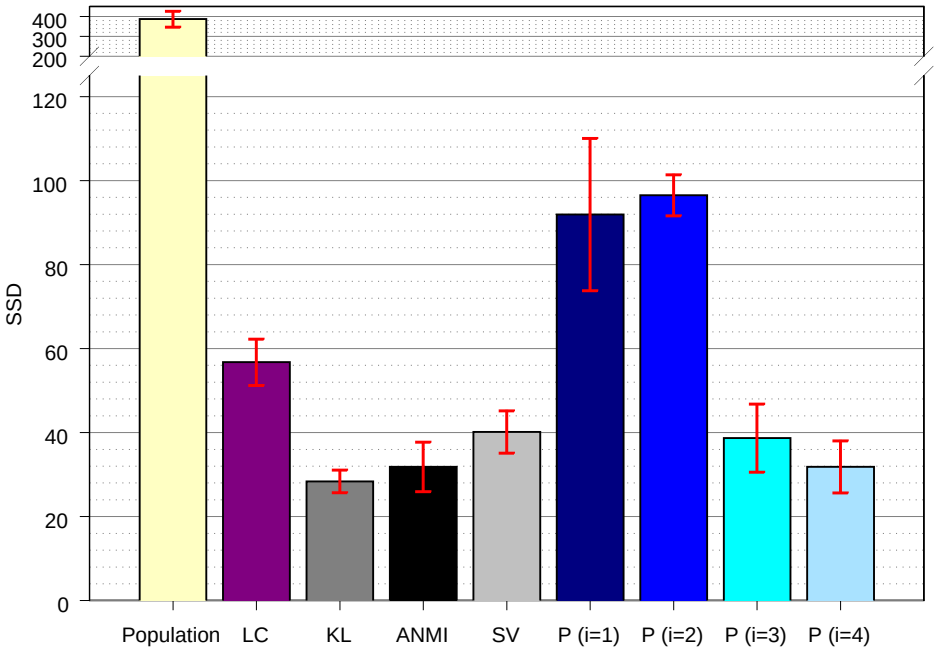


Figure 4.9: The mean (histogram bar) and standard deviations (red error bars) of the sum-of-squared differences between the transformed sources and the original MNI Brainweb image for each similarity metric show the lowest error obtained using Kullback-Leibler.

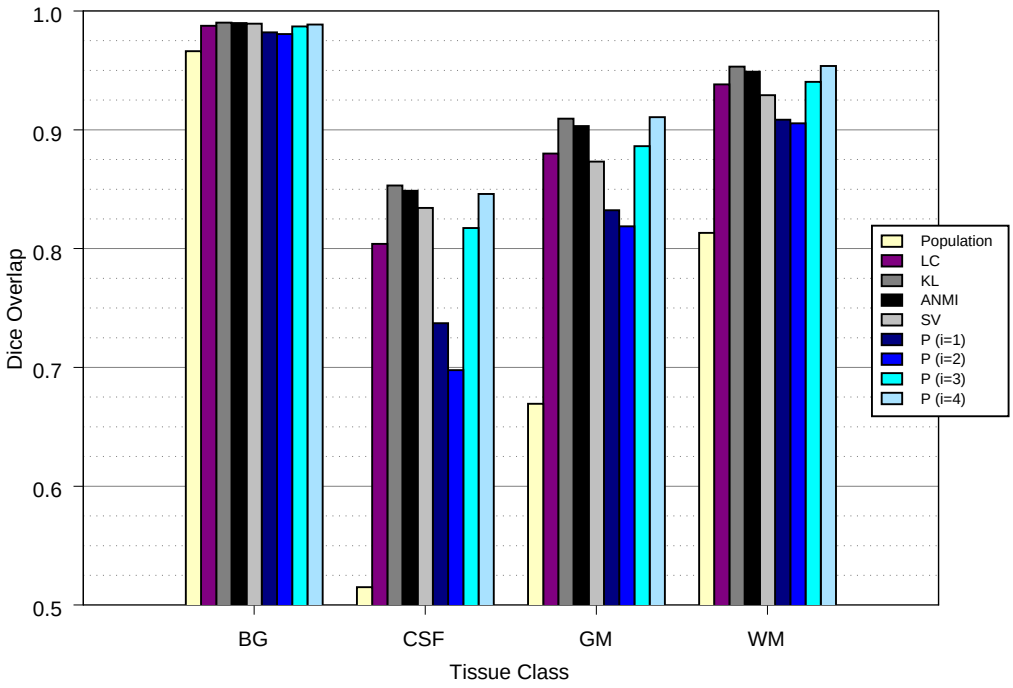


Figure 4.10: Average Dice overlap measures between known segmentation and segmentation in space found by groupwise registration using various similarity metrics.

Finally, as each subject in the population was created by transforming the MNI Brainweb image by the inverse of a known transformation, an exact registration should recover the original deformation. The average absolute displacement error for each voxel, $\mathbf{x}$, has been computed for each similarity measure:

$$Error = \sum_{\mathbf{x} \in \Omega} \frac{|d_{simulated}(\mathbf{x}) - d_{recovered}(\mathbf{x})|}{n_{\Omega}} \tag{4.24}$$

The results of these are shown in Figure 4.11. In the first iteration of the pairwise method, a chosen subject acts as the reference image for the registration. For this reason, the deformation fields produced by this iteration are not considered. In subsequent iterations, the current average image is used as a reference image.

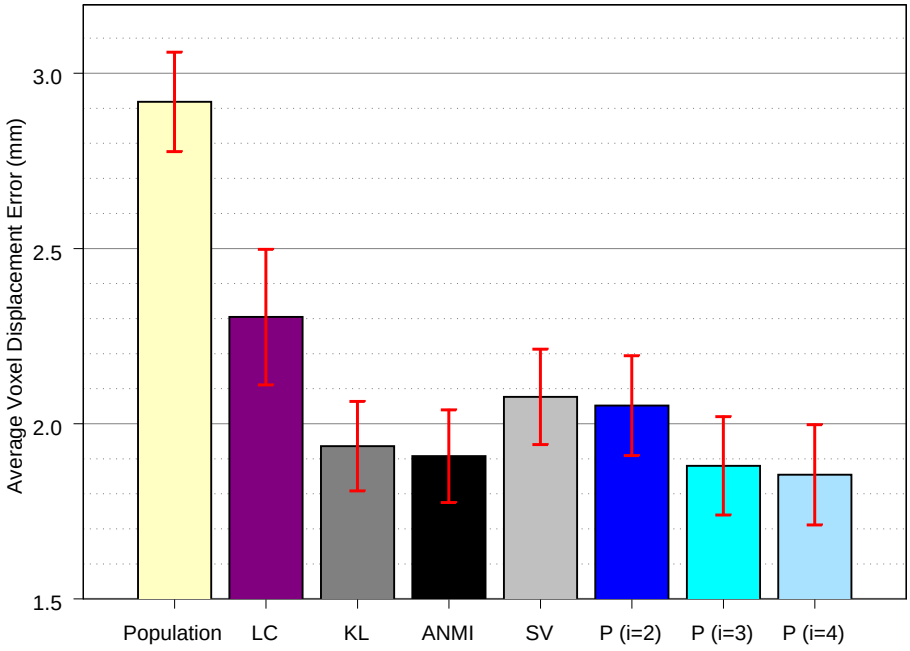


Figure 4.11: Mean and standard deviation of absolute displacement error of each voxel for each metric (mm).

These results show that the best-performing groupwise similarity metrics overall are Normalised Mutual Information with an average intensity reference (ANMI), and

the Kullback-Leibler (KL) divergence, using soft segmentations of the population. The results are comparable to those obtained after four iterations of the pairwise averaging process. With the exception of the label consistency, all of the groupwise similarity metrics outperform the results of using one iteration of the pairwise averaging technique. Additionally, the groupwise measures show lower standard deviations of the SSD error than the pairwise metrics, showing the alignment is more consistent.

To further assess the consistency of the registration, how well-registered the group, as a whole, is considered. The more well-aligned a population is, the sharper the resulting final atlas (a mean of the intensities of the individual transformed images) should be. However, it is not easy to distinguish between the atlases in Figure 4.8 by visual inspection alone. The entropy, H(A) of each atlas has therefore been computed:

$$H(A) = -\sum_{\mathbf{x}} p(A(\mathbf{x})) \log p(A(\mathbf{x})) \tag{4.25}$$

where $p(A(\mathbf{x}))$ is the probability of the intensity of voxel $\mathbf{x}$. As the atlas gets sharper, the entropy of the atlas should decrease. The results are presented in Figure 4.12.

Overlap coefficients of the whole group have also been computed, as shown in Figure 4.13. These are found using the metric proposed by Zollei in [269], for each tissue class:

$$Overlap = \frac{N(I_1 \cap I_2 \cap ... \cap I_{n_I})}{\min\left(N(I_1), N(I_2), ..., N(I_{n_I})\right)} \tag{4.26}$$

Here, the numerator represents the area of overlapping labels, and the denominator represents the minimum input area for that label. This metric is fairly sensitive. For example, for a group of 25 2D circles, each with radius 50 voxels, but displaced by 1 and 2 voxels in each x, y and xy direction from the origin, the overlap coefficient already drops to 0.9. The Dice coefficient for the two *most mis-registered* samples of the population is 0.98. For the relatively small structures of CSF and grey matter,

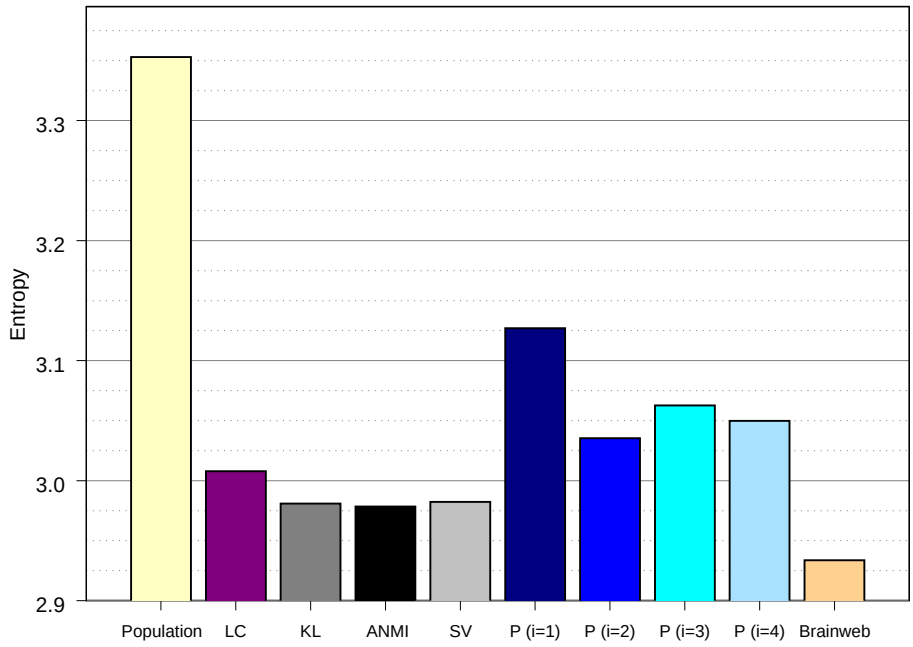


Figure 4.12: Entropy of the final atlas created using the various similarity metrics and the entropy of the original MNI Brainweb image and the atlas of the original population.

the overlaps of the original population, as computed by Equation 4.26, are 0.0013 and 0.0012 respectively. The obtained coefficients of around 0.2, after registration, therefore represent a significant improvement in alignment.

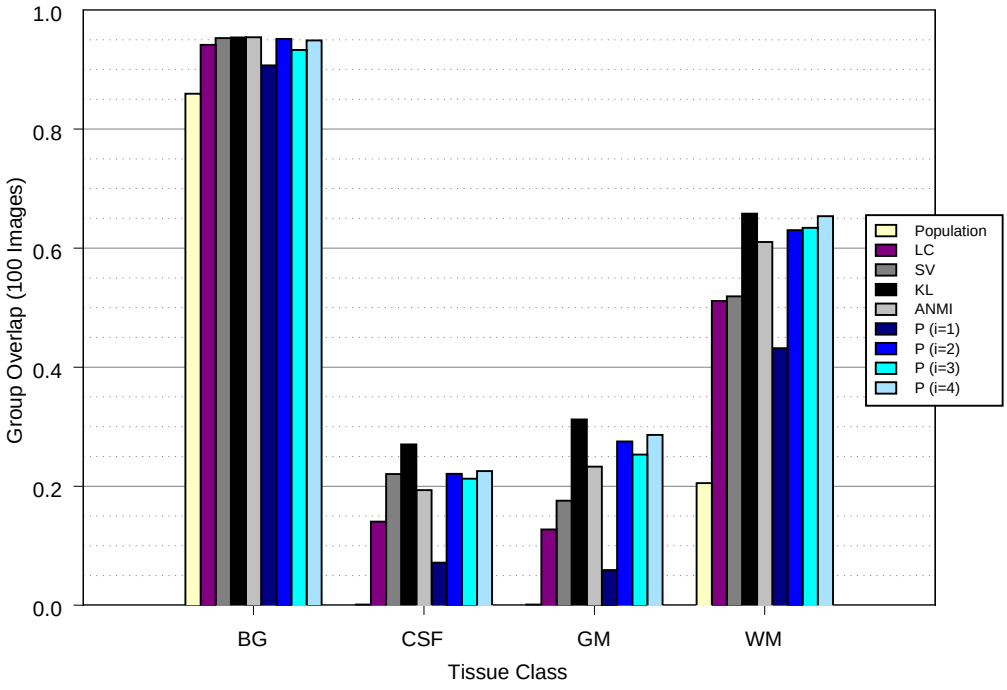


Figure 4.13: Groupwise overlaps of tissue classes

Groupwise accumulated overlaps based on fuzzy set theory, developed by Crum et al. [59], have also been used to assess the registration:

$$Overlap = \frac{\sum_{pairs} \sum_{labels} \sum_{voxels} \min(I_1, I_2)}{\sum_{pairs} \sum_{labels} \sum_{voxels} \max(I_1, I_2)} \tag{4.27}$$

where A and B are the segmentation values in a given pair of images for a given voxel and label. For these experiments, only binary segmentation values are considered. This measure implicitly weights the effect of structures according to their area or volume. In Figure 4.14, the results shown do not include the alignment of the very large background class.

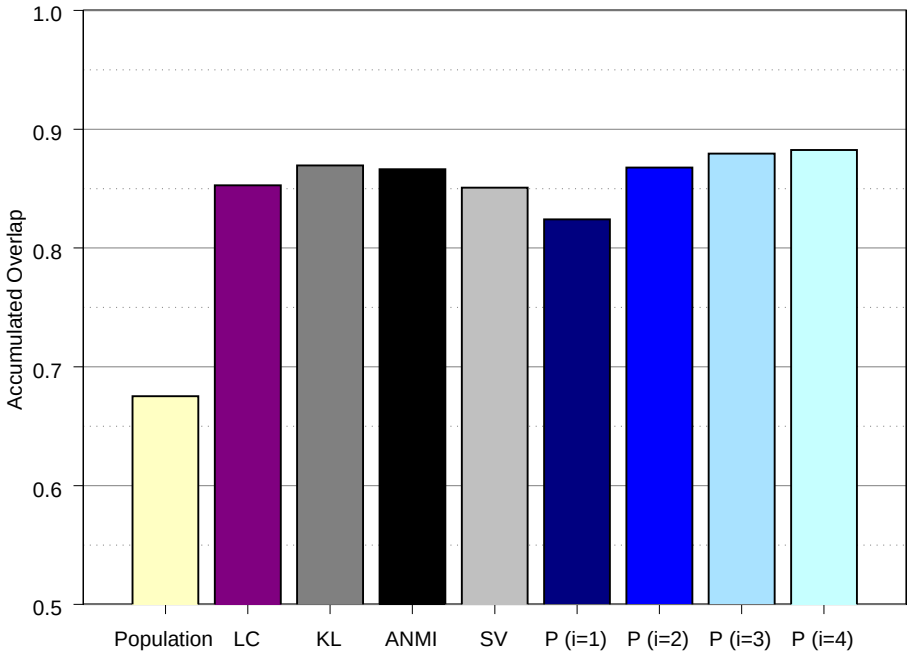


Figure 4.14: Accumulated overlaps

The results above show a significant benefit of using groupwise registration over a single iteration of the pairwise scheme for this data, regardless of the similarity metric used. Using the Kullback-Leibler metric generally produces more consistent registrations than even four iterations of the pairwise process. Additionally, there appears to be no guarantee that increasing the number of pairwise iterations increases the consistency of the registration.

4.5.3 Groupwise Computational Complexity

When registering many subjects simultaneously, it is important that the time taken does not increase more than linearly with increasing numbers of subjects. The areas of the registration process that could affect the optimisation are the number of degrees of freedom of the problem and the calculation of the similarity metric. The number of degrees of freedom increases in proportional to the total number

of control points, and this increases linearly with the number of images used. The similarity measures have also been designed to scale linearly with increasing numbers of subjects. Figure 4.15 shows how the time taken for the registration to converge, using 20mm and 10mm meshes, varies using increasing numbers of subjects, using the sample variance metric. The images used were from the population above and each image was of size 180×216 voxels. The registrations were run on a 2GHz Intel Pentium 4 processor.

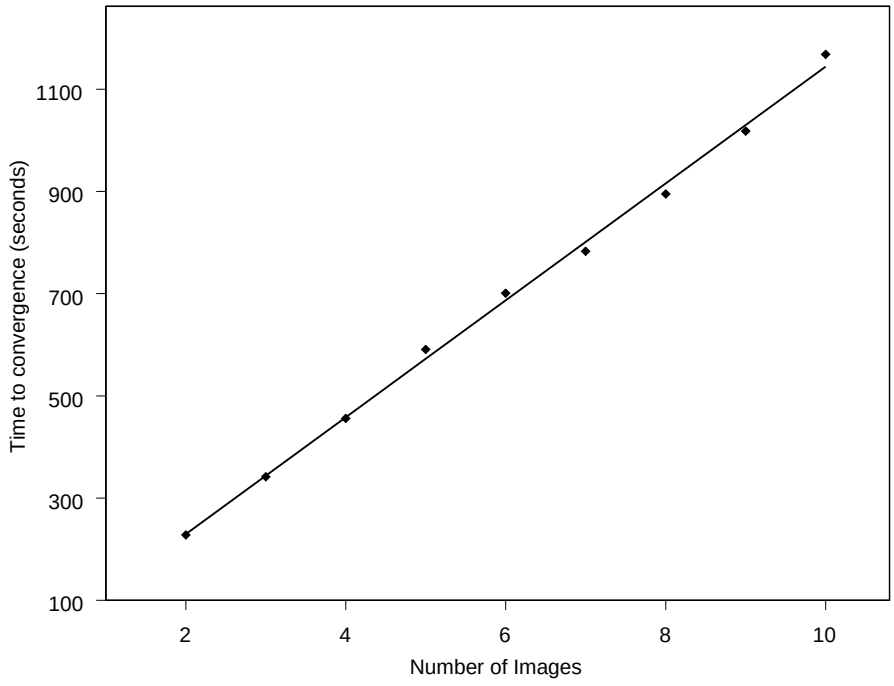


Figure 4.15: Time taken for convergence of groupwise algorithm, using SV similarity measure, with increasing number of subjects

4.5.4 Real 3D Adult Data

The algorithm has also been tested on 3D, real, MR datasets. The results of these do not necessarily follow easily from the 2D case above. First of all, the number of degrees of freedom that parameterise the transformations (proportional to the number of control points), increases significantly. This may make the search space

for the optimisation more complicated and increase the presence of local maxima. Additionally, the variable intensities of real data mean that the choice of similarity metric becomes more of an issue. In this section, we investigate the performance of the best-performing segmentation- and intensity-based similarity measures according to the results in the 2D case.

Twelve subjects were taken from a population of 30 volunteers (15 male, 15 female) of ages ranging from 20-54 years (median age 30.5 years). The images used are T1-weighted 3D volumes, acquired using a TE of 4.2ms and a TR of 15.5ms, with a flip angle of 20°. The scan data were resliced to create isotropic voxels of $0.9375 \times 0.9375 \times 0.9375$mm, using windowed sinc interpolation. This dataset also has hard segmentations of 83 tissue classes, obtained by manual segmentation by an expert, using an extension of an existing protocol [103].

The images are first aligned affinely to the average space, using the method described in [6] (Figure 4.16(a-c)). To speed up computation time, and to bypass any local maxima far from a good solution, an initial pairwise estimate of the average is found, using the method described in Section 4.3.1. The atlas of the population after this stage of the registration is shown in Figure 4.16(d-f). This is then used as starting solution to the various groupwise and pairwise measures. The following similarity measures have been considered:

1. KL using probabilistic segmentations of each subject, obtained using the Expectation-Maximisation algorithm.

2. ANMI.

3. Pairwise registration from average atlas (four iterations).

The final atlases are shown in Figure 4.16. Although it is hard to determine visual differences between the atlases, it can be seen that the pairwise average deformation field is not zero, indicating that the atlas is not exactly at the average of the

population. Additionally, the full groupwise registration is run on the same dataset, but without using any initial solution to the non-rigid registration. For this, the Kullback-Leibler similarity metric was used. The resulting atlas is shown in Figure 4.17.

The accumulated overlap as given by Crum (Equation 4.27), has been computed for each metric over the 83 tissue classes. The results are shown in Figure 4.18. In this, the background segmentation has been ignored due to its large size relative to the other structures, and its consequently large effect on the measure. Figure 4.16 also shows the total deformation fields (formed by summing the displacement at each control point over all images) produced using each method. The closer the deformation field is to a uniformly-spaced grid, the closer to the average shape the atlas is.

Once again, the best performing groupwise similarity metric is the Kullback-Leibler divergence, which produces segmentation results close to those produced using four iterations of the pairwise method of re-registering to the average atlas.

4.6 Summary

An unbiased, groupwise, non-rigid registration algorithm has been developed in this chapter, which simultaneously registers a population of subjects to find the average shape of the population. This algorithm does not require the *a-priori* selection of any reference subject. The method has been tested on, and validated using, a population of 100 2D images, created using known deformations, and on 12 3D, real MR images. The algorithm scales linearly with increasing numbers of subjects.

Groupwise similarity metrics have been developed to assess the similarity of a group

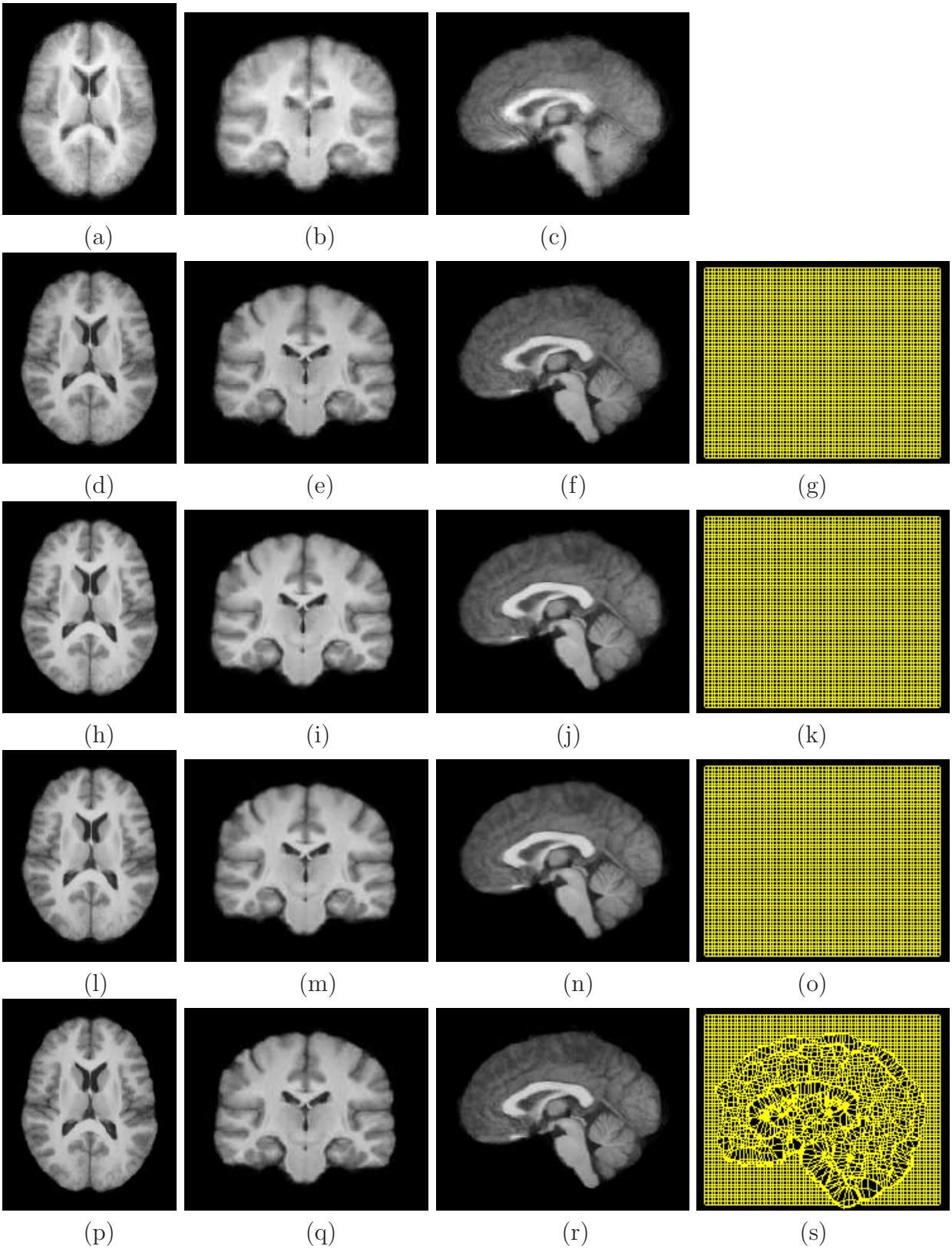
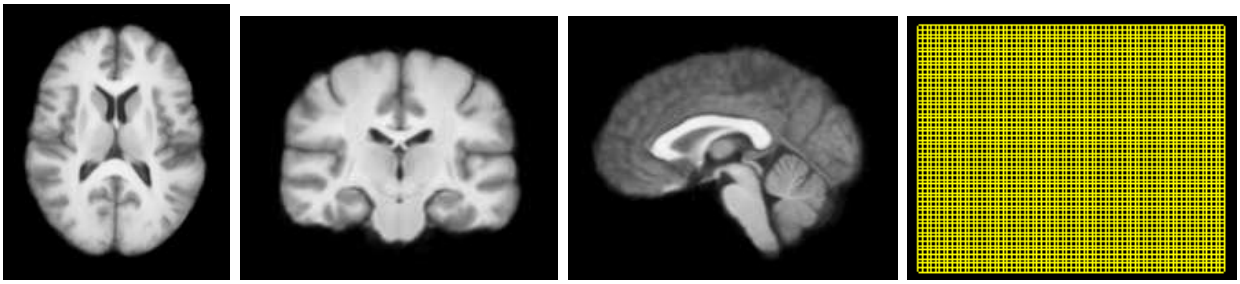


Figure 4.16: Top row: axial, coronal and sagittal sections of affinely-aligned 12 3D adult subjects; (d)-(f) initial pairwise estimate (h)-(j) groupwise registration using KL; (l)-(n) groupwise registration using ANMI (p)-(r) fourth iteration of pairwise re-registration to average shape. Far right column: total deformation field (sagittal section) using each similarity metric.



(a) (b) (c) (d)

Figure 4.17: Atlases obtained using groupwise registration of 12 3D adult subject, using the KL similarity metric without an initial estimate of the non-rigid transformation.

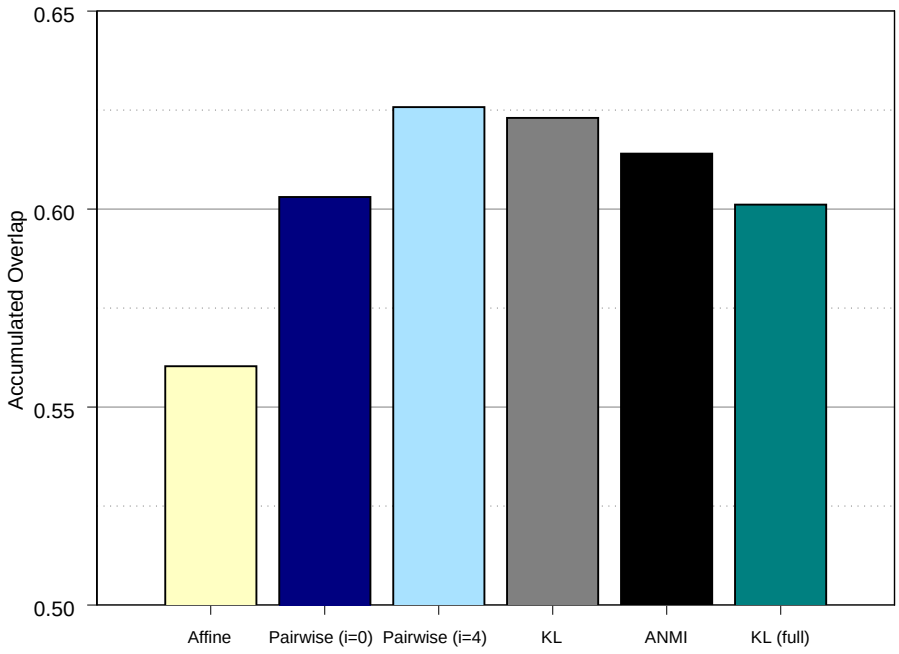


Figure 4.18: Accumulated overlaps of real 3D MR data aligned using groupwise registration. KL (full) represents the overlap obtained when running the groupwise registration using the Kullback-Leibler divergence, without an initial solution.

of images in the registration process. Methods based on image intensities (sample variance, (SV)), entropies (sum of NMI values using an average intensity image (ANMI)), hard (label-consistency (LC)) and soft (Kullback-Leibler (KL)) segmentations have been developed. It was found that using hard segmentations lacks the sensitivity required for accurate groupwise, non-rigid registration. However, the removal of intensity variations within tissue classes, and across subjects, did improve the performance of the registration algorithm. The best-performing groupwise similarity metric (in terms of both accuracy and consistency), was found to be the Kullback-Leibler divergence for registering probabilistic segmentations. The use of the ANMI metric was a viable alternative and can be used if soft segmentations cannot be made available. Apart from the LC metric, all other groupwise metrics outperformed a single iteration of the pairwise method of creating an average shape (using the inverse of the mean deformations from a chosen reference). Additionally, the low standard deviations of the results, low entropies of the final atlases and the results of the groupwise overlaps, show improved consistency when using the groupwise registration algorithm.

To conclude, the groupwise algorithm developed has been used to align populations to their average shape with promising results in terms of both accuracy and consistency. In the next chapter, this algorithm is applied to populations of neonatal and child brain images.

Chapter 5

Neonatal Image Analysis

As seen in Chapter 1, the development of brain structures in infants born prematurely proceeds differently to infants of the same age, developing in the uterus. Given the significant neuropsychiatric and neurological issues that preterm-born infants face in later life, it is important to be able to determine these differences, in order to potentially treat any problems. Until fetal MR imaging becomes more developed, the first point at which these two groups can be compared, is at term-equivalent age (around 40 weeks). Additionally, it is also important to consider how the brain continues to develop in the following years. The aims of this chapter are to investigate:

1. How extrauterine development of preterm infants affects the growth of brain structures up to term-equivalent age.

2. How the preterm brain continues to grow from from 1 to 2 years.

It is difficult to quantitatively map the growth of the brain between 40 weeks and 1 year because of the huge development of new structures in the brain and the increased myelination, which alters the MR signal (as shown in Figure 5.1), without a large number of intermediate scans. At present, there is not enough data available for such a study. Additionally, comparison with normal subjects is challenging due to the difficulty in acquiring parental approval for the scanning of healthy infants

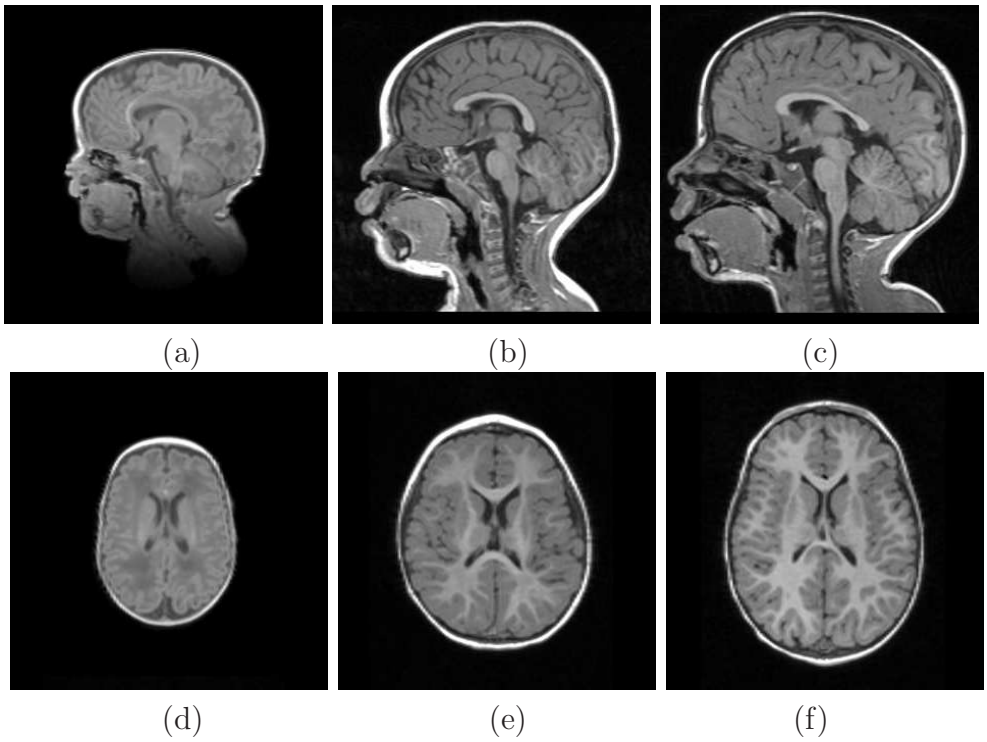


Figure 5.1: Examples of sagittal (top row) and axial (bottom row) sections through the brain of infants born preterm at term-equivalent age (a,d), one-year (b,e) and two-years (c,f).

and in obtaining the cooperation of young children without sedation. MR images of term-born neonates and children are therefore susceptible to motion artifacts.

The aim of this chapter is to determine how volumes of brain structures differ between populations. Only anatomical structures imaged through MR will be considered in this work, and not vasculature, tracts or functional development which require the use of other imaging techniques. Groupwise registration is used to create representative brain atlases of preterm and control infants at term-equivalent age. Deformation-based morphometry (DBM) is then used to analyse volumetric differences between the two groups. Average atlases of preterm populations at 1 year and 2 years are also created, and the growth of structures during this phase is calculated using DBM.

5.1 Volumetric analysis

Segmentation-based volumetric analyses [92, 197, 240] involve the labelling of corresponding structures of interest in the atlas of each group or individual being compared. The total volume of each structure for each atlas can then be calculated by counting the number of voxels in the segmentation. This generally requires the prior knowledge of which structures are particularly important, and the manual segmentation of these structures.

More recently, computational morphometry has been developed for automated, volumetric analysis, without the need for *a priori* segmentation of regions of interest. These methods fall broadly into two categories: those which compare the local composition of brain tissue at each voxel after global shape differences have been discounted (voxel-based morphometry), and those which assess differences in brain shape, by non-rigidly aligning images into the same coordinate system (deformation-based morphometry). Both methods examine the whole brain without the need for any a-priori hypothesis about which structures are likely to change.

5.1.1 Voxel-based morphometry (VBM)

Voxel-based morphometry [264, 14, 211] involves comparing the concentrations of tissue (e.g.: grey matter) on a voxel-wise basis. To do this, images are normalised into the same global coordinate system using a 12 parameter affine transformation, followed by a low-dimensional warping using basis functions [13]. This normalisation does not aim to align corresponding features exactly, but only to correct for global shape differences.

The normalised images are then partitioned into grey matter (GM), white matter (WM), cerebro-spinal fluid (CSF) and background classes using a clustering algorithm for mixture models [14]. These are then smoothed by convolving with

an isotropic Gaussian kernel. Each voxel in the smoothed image therefore contains the average concentration of tissue from the neighbouring region determined by the width of the kernel. Smoothing also renders the data more normally-distributed. This allows for the use of parametric statistical techniques, such as the T-test, for finding voxel-wise differences in tissue concentration, represented by the intensities of the smoothed images.

VBM provides a framework for analysis of differences between groups, and can be implemented using the standard SPM software [2]. However, the specific characteristics of neonatal MR images pose particular challenges for VBM. The high variation of size and shape of neonates means that low-dimensional normalisation may not be sufficient to capture enough of the variation in the population. Additionally, neonatal images are particularly susceptible to varying contrasts in the same tissue type, between subjects and within the same image, due to the variable development of myelination. This might confound methods, such as voxel-based morphometry, which are based on the intensity difference between corresponding voxel locations. Systematic differences in contrast can become statistically significant if large enough numbers of subjects are used. Finally, there are no existing protocols to reliably classify neonatal cerebral tissue, as required by VBM.

5.1.2 Deformation-based morphometry (DBM)

Instead of analysing intensity differences between corresponding voxels, deformation-based morphometry [64, 45] analyses the deformations required to warp one subject non-linearly to the coordinate system of another. All the differences between the two images are captured by the high-dimensional deformation field. DBM requires no segmentation, either of regions-of-interest or tissue classes.

The Jacobian of a deformation field, obtained when warping a subject to a spec-

ified coordinate system using image registration, is given by the gradient of the deformation field:

$$|\mathbf{J}(\mathbf{u})| = \begin{vmatrix} \frac{\partial u_x}{\partial x} & \frac{\partial u_x}{\partial y} & \frac{\partial u_x}{\partial z} \\ \frac{\partial u_y}{\partial x} & \frac{\partial u_y}{\partial y} & \frac{\partial u_y}{\partial z} \\ \frac{\partial u_z}{\partial x} & \frac{\partial u_z}{\partial y} & \frac{\partial u_z}{\partial z} \end{vmatrix} \tag{5.1}$$

where u_x, u_y and u_z represent the displacements in the x, y and z directions respectively. By definition, the Jacobian of a deformation is the volume change of the unit cube after the deformation is applied. By calculating the Jacobian of the deformation field at each voxel, the volume change at each voxel can be found. For example, in the case of an FFD model based on B-splines, the deformation is given by:

$$\mathbf{u}(x,y,z) = \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} B_l(u)B_m(v)B_n(w)\phi_{a+l,b+m,c+n} \tag{5.2}$$

where $a = \lfloor \frac{x}{n_x} \rfloor - 1, b = \lfloor \frac{y}{n_y} \rfloor - 1, c = \lfloor \frac{z}{n_z} \rfloor - 1, u = \frac{x}{n_x} - \lfloor \frac{x}{n_x} \rfloor, v = \frac{y}{n_y} - \lfloor \frac{y}{n_y} \rfloor, w = \frac{z}{n_z} - \lfloor \frac{z}{n_z} \rfloor$ and where B_l represents the l-th basis function of the B-spline:

$$\begin{aligned} B_0(s) &= (1-s)^3/6 \\ B_1(s) &= (3s^3 - 6s^2 + 4)/6 \\ B_2(s) &= (-3s^3 + 3s^2 + 3s + 1)/6 \\ B_3(s) &= s^3/6 \end{aligned}$$

For example, the component of the Jacobian $\frac{\partial u_x}{\partial x}$ is given by:

$$\begin{aligned} \frac{\partial u_x}{\partial x} &= \sum_{m=0}^{3}\sum_{n=0}^{3} -\frac{(1-u)^2}{2} B_m(v)B_n(w)\phi_{a,b+j,c+k} \\ &+ \sum_{m=0}^{3}\sum_{n=0}^{3} \frac{3u^2-4u}{2} B_m(v)B_n(w)\phi_{a+1,b+j,c+k} \\ &+ \sum_{m=0}^{3}\sum_{n=0}^{3} \frac{-3u^2+2u+1}{2} B_m(v)B_n(w)\phi_{a+2,b+j,c+k} \\ &+ \sum_{m=0}^{3}\sum_{n=0}^{3} \frac{u^2}{2} B_m(v)B_n(w)\phi_{a+3,b+j,c+k} \end{aligned} \tag{5.3}$$

Other derivatives are calculated in a similar fashion. For each subject, the determinant of the Jacobian at each voxel gives the volume change obtained when deforming that subject into the space of the reference coordinate system. The value of this determinant can be interpreted as in Table 5.1.

$\lvert\mathbf{J}(\mathbf{u})\rvert$	Local properties of deformation
$= 1$	No volume change
< 1	Local contraction
> 1	Local expansion
$= \infty$	Tearing
< 0	Folding

Table 5.1: Table showing how the value of the Jacobian indicates volume change from source to reference.

The use of DBM therefore allows for the analysis of the whole brain without the need for any prior hyphothesis or any tissue classification. As the analysis is done on the deformation fields themselves, it is not dependent on intensity variations, but requires an effective non-rigid registration from the reference coordinate system.

5.1.3 Data analysis

5.1.3.1 Effect size

The determinant of the Jacobian gives the absolute volume changes at each voxel, relative to a specified template. However, it does not account in any way for the variance of the group or give any statistically significant threshold. To assess which changes are most *consistent* across a group, Cohen's *effect size* [47] can be used, which has previously been used to assess volumetric differences between populations in [28, 64, 225]. This measures the standardised difference between the mean volume changes of two groups, C and P, therefore accounting for the variance of volume changes for each group:

$$\epsilon(\mathbf{x}) = \frac{\mu_C(\mathbf{x}) - \mu_P(\mathbf{x})}{\sigma_{C \cup P}(\mathbf{x})} \tag{5.4}$$

$\mu_C(\mathbf{x})$ and $\mu_P(\mathbf{x})$ represent the mean value of the determinant of the Jacobian across each group, C and P, respectively, at voxel location $\mathbf{x}$. $\sigma_{C\cup P}(\mathbf{x})$ represents the standard deviation of the Jacobian determinant of the pooled group at the same location. In this formulation, a positive effect size means tissue expansion from group C to group P; conversely, a negative effective size means tissue contraction from group C to group P.

Effect sizes can also be interpreted as the percentage overlap between the distributions of the two groups (see Cohen [47]). The various percentage overlaps for given effect size values can be found in Table 5.2. Cohen suggests that values of

$\epsilon(\mathbf{x})$	Overlap (%)
0.0	100.0
0.1	92.3
0.2	85.3
0.3	78.7
0.4	72.6
0.5	67.0
0.6	61.8
0.7	57.0
0.8	52.6
0.9	48.4
1.0	44.6
1.1	41.1
1.2	37.8
1.3	34.7
1.4	31.9
1.5	29.3

Table 5.2: Table showing how the percentage overlaps of two groups for different effect size values.

effect size ≈ 0.8 can be considered to be "large", values ≈ 0.5 "medium" and ≈ 0.2 "small". However, these values do not necessarily translate directly to a level of statistical significance.

5.1.3.2 Standardised two sample t-test

The two-sample t-test allows the determination of statistically significant differences across two groups. The *t-value* at each voxel is given by the formula:

$$t(\mathbf{x}) = \frac{\mu_P(\mathbf{x}) - \mu_C(\mathbf{x})}{\sqrt{\frac{\sigma_P^2(\mathbf{x})}{n_P} + \frac{\sigma_C^2(\mathbf{x})}{n_C}}} \tag{5.5}$$

which represents the difference in mean values (μ) divided by the *standard error* of the two populations, P and C. σ^2 represents the variance and N represents the number of samples in one of the groups. A t-value of over some statistically-significant threshold (*p-value*) indicates that the finding is more than that that would be expected by chance alone. Generally, this threshold is set at $p = 0.05$. However, since comparisons are made on a voxel-by-voxel basis, for an image of size $100 \times 100 \times 100$, this represents 50000 false positives - a very large number in absolute terms. One method to account for this is to use the *Bonferroni correction* [124]. Using this, when n tests are performed, the significance level is corrected to be p/n. In practice with neuroimaging data, however, this has the effect of elminating true as well as false positives [94]. An alternative method is to control the *false discovery rate* (FDR) [94, 26]. This is the proportion of false positives among only those tests which give a positive result.

5.2 Structural Differences at Term-Equivalent Age

Average atlases of a group of 16 preterm infants, scanned at term-equivalent age, and a group of 16 controls, born and scanned at term, were created. The aim is to analyse the differences between the two groups at this common time-point to assess how exposure of the preterm infant to an *extra-uterine* environment affects the development of brain structures.

5.2.1 Subjects and image acquisition

5.2.1.1 Image acquisition

A 1.5 Tesla Eclipse MR system (Philips Medical Systems) was used to acquire high resolution T1-weighted (TR=30ms, TE=4.5ms, flip angle=30°), volume datasets in contiguous sagittal slices (in-plane matrix size 256×256, field of view 25cm), and with a voxel size of $1.0 \times 1.0 \times 1.6$mm.

5.2.1.2 Population

Images of 16 control subjects and 16 preterm-born subjects were acquired using the above protocols, and used in the groupwise experiments following. Preterm infants were sedated during the acquisition using chloral hydrate. Control infants were examined in natural sleep. All subjects were imaged at the same, term-equivalent age for comparison. Infants with white matter brain injury were excluded from the study. Table 5.3 shows the age ranges of the infants at birth and at the time of the image acquisition. These images are shown in Figures 5.2 (controls) and 5.2 (preterms), after an initial affine alignment to the average space, for ease of documentation.

Gestational Age (weeks)	Controls	Preterms
At birth:		
Median	39.57	29.71
Range	36-41.86	24-34
At scan:		
Median	40.14	40.43
Range	36.57-43.14	37-44.57

Table 5.3: Ages of population at birth and scan

5.2.2 Groupwise registration parameters

The groupwise registration algorithm was separately run on three populations:

1. The 16 control subjects

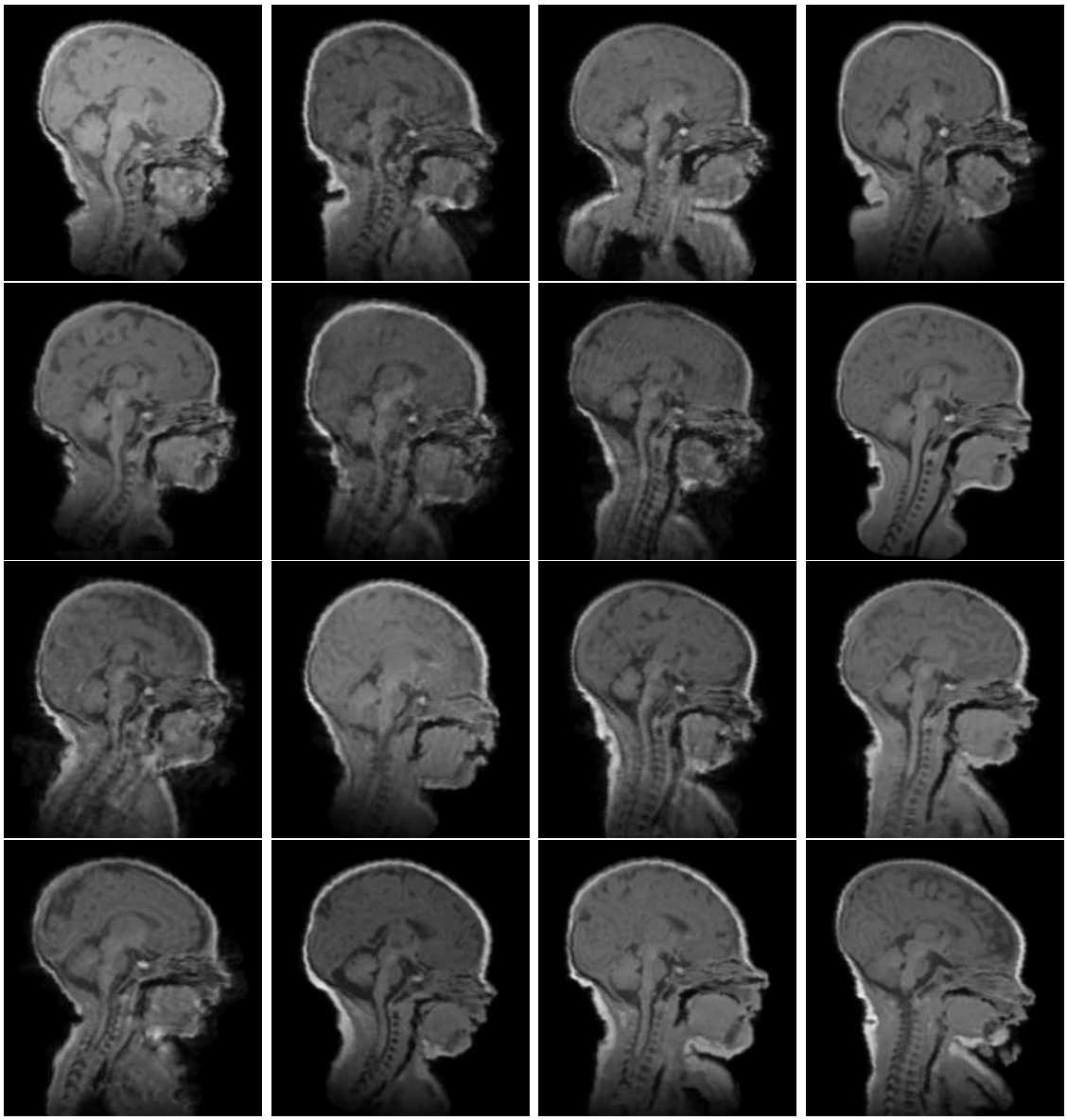


Figure 5.2: Sagittal slices of 16 control subjects, after affine alignment to an average space, showing variation is shape and contrast.

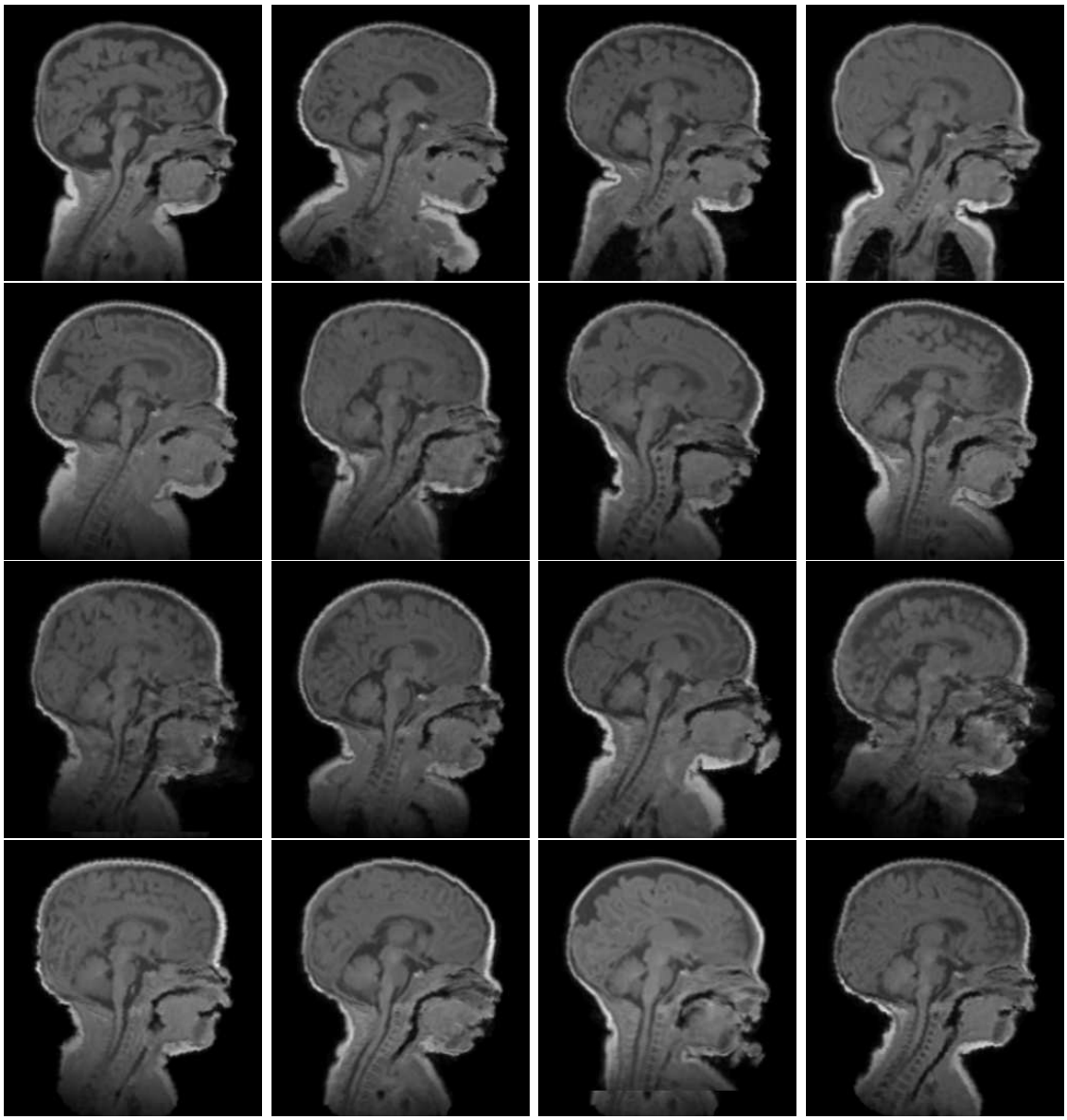


Figure 5.3: Sagittal slices of 16 preterm subjects, after affine alignment to an average space, showing variation is shape and contrast.

2. The 16 preterm subjects

3. The combined group of 32 subjects

Pairwise registrations to a chosen arbitrary subject are first used to find the mean deformation field. This is inverted and used to create an initial estimate of the average coordinate system as described in Chapter 3. A parallel (MPI) implementation of the groupwise registration algorithm was then run using the sum of normalised information scores between the average intensity image and each of the individual subjects, using cluster of eight 3.06GHz Intel Xeon machines.

5.2.3 Results

5.2.3.1 Neonatal atlases at term-equivalent age

Groupwise atlases of groups of 16 control subjects and 16 preterm subjects have been created using deformation grids of 2.5mm and are shown in Figure 5.4. Additionally, also shown is the atlas formed when registering all 32 subjects simultaneously, forming an atlas with a shape in between the two groups. The atlases show important differences between the two populations at term-equivalent age:

1. *Scaphocephalic brain shape in the preterm group.* The elongated shape of the preterm brain is probably due to the infant lying on a bed while the brain is developing, compared to the term-born infants growing while supported in all directions by amniotic fluid in the uterus.

2. *Enlargement of the lateral ventricular system in preterm infants.* Figures 5.4 (h) and (k) show that the fluid-filled ventricles are much larger in the preterm group than in the control group. It is possible that this also corresponds to lower tissue development in areas surrounding the ventricles. This is consistent with previously reported findings in clinical studies [177, 131, 29].

3. *Increased myelination in the control group.* The internal capsule in Figure 5.4(i) (controls) is much more pronounced than in Figure 5.4(l) (preterms),

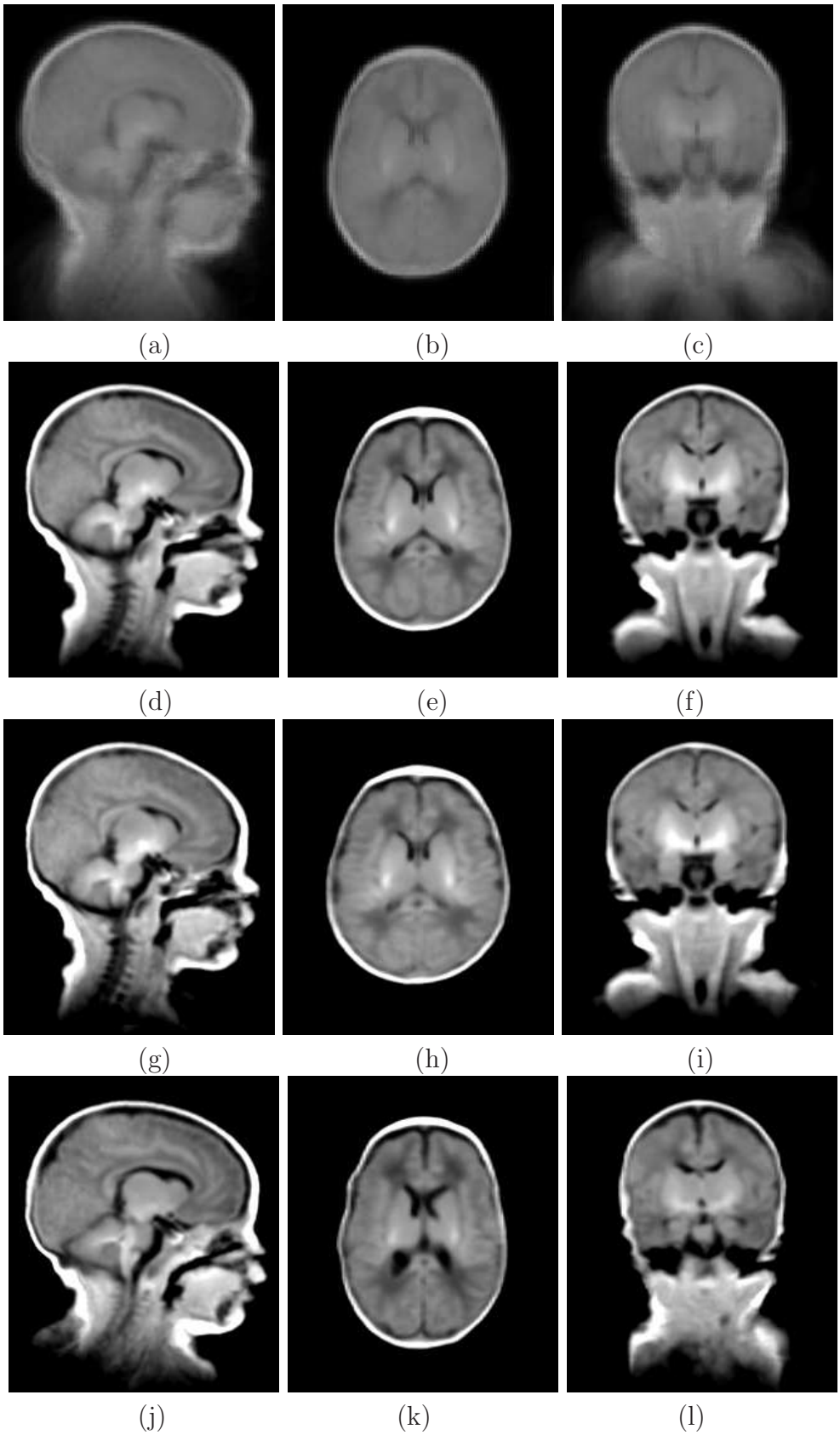


Figure 5.4: Atlases produced by groupwise registration showing sagittal, axial and coronal sections of the total group of 32 subjects affinely aligned (a)-(c), and after groupwise non-rigid registration (d)-(f); atlas produced using only 16 control subjects (g)-(i); atlas produced using only 16 preterm subjects (j)-(l). Major differences in ventricular size and myelination can be seen in (h),(k) and (i),(l), respectively.

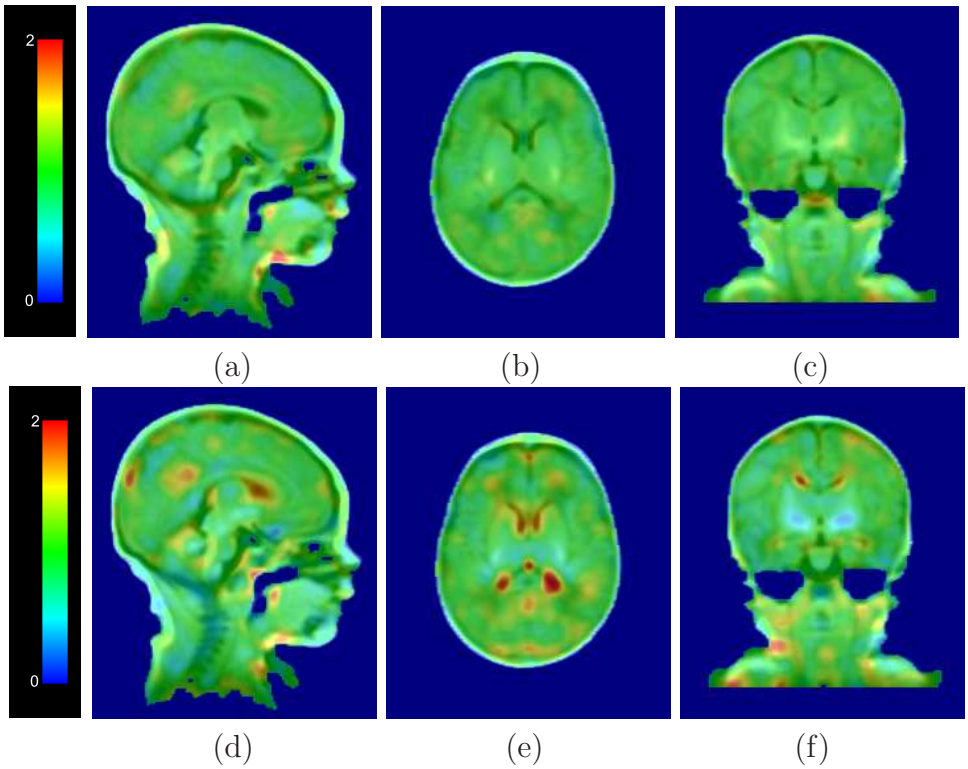


Figure 5.5: Determinant of Jacobians for controls (top row) and preterms (bottom row) shown in the common coordinate system. The scale on the far left shows how the colours vary for $0 \leq |\mathbf{J}| \leq 2$. Values < 1 indicate tissue contraction (shown by colours from green to blue) and values > 1 indicate tissue expansion (shown by colours from green to red).

showing that more myelination has occurred in the control group than in the preterm group at the same equivalent age. This result has also been previously reported in [253, 57].

One potential issue with the registration of the combined group of controls and preterms, is that the average shape may not form a realistic brain shape. However, in practice, this only happens if the two populations are very different from each other; here the two groups are sufficiently similar for this not to be a problem.

5.2.3.2 Volumetric changes using DBM

The atlases in Figure 5.4 show qualitative differences between the two populations and are particularly useful for visualising differences in brain shape and myelination. In order to quantify volumetric differences, it is possible to register the atlases of

the controls and preterms together, and analyse the resulting deformation field as described in Section 5.1.2. However, as this involves registering only the average atlases, it is not possible to perform any statistical significance testing on the volumetric changes obtained. Instead, the deformation fields obtained when registering the entire population to their average coordinate system (shown in Figure 5.4(d-f)) are analysed. One caveat with this approach is that the inconsistency in myelination across the population may adversely affect the alignment of the images, and so each registration was checked visually. Figure 5.5 shows the volume changes for each population, calculated by finding the average determinant of the Jacobian of these deformation fields. The distribution of the effect size metric across the brain, again in the overall average coordinate system, is shown in Figure 5.6. A large and positive effect size indicates a structure larger in the preterm group than in the control group. Figures 5.5(d-f) and 5.6(g-i) show a volume increase in the posterior horns of the lateral ventricles. This finding has also been described in previous studies [166, 183]. Large, negative effect sizes show a reduction of volume in the preterm group, compared to the controls. Figures 5.5(d-f) and 5.6(d-f) show noticable reduction in the volume of deep grey matter in an area corresponding to the basal ganglia. This has also been suggested previously in [141]. However their use of small selected patient groups with cystic white matter disease does not necessarily extend to infants without the tissue damage. The finding does correspond to that of Boardman et al. [29] who use deformation-based morphometry on a larger group of subjects - which include the ones used in this study - using pairwise registrations with a carefully-chosen reference subject. These same differences are also detected using a T-test to compare the volume changes in the two groups (Figure 5.7), using a significance level of 5% with correction for multiple comparisons to control the false discovery rate using SPM [2].

Volume changes can also be seen outside the brain area. This is due to the difficultly in accurately and consistently segmenting neonatal brain images. The skull

and part of the neck were therefore also included in the registration process, in addition to a layer of air surrounding the skull. The registration deforms the images to align the intensities of the surrounding air, thus producing the extraneous volume changes outside the brain.

5.2.4 Standard deviations of volumetric changes

By registering to the average image, it is also possible to determine the standard deviations of the volumetric change at each voxel, for each population:

$$\sigma = \sqrt{\frac{1}{n_I} \sum_{\Omega} \left(V(\mathbf{x}) - \bar{V(\mathbf{x})} \right)^2} \tag{5.6}$$

where $V(\mathbf{x})$ represents the volume change at voxel $\mathbf{x}$. Figure 5.8 shows the variation from the mean volume change in the average coordinate system of the combined population. It can be seen that the preterm group shows larger variation, particularly in the ventricles, while the basal ganglia area appears to have greater variation in the control group. While this may represent biological variation, it is also an area which is undergoing myelination in this age group. It is also possible that the apparent variation is due to erroneous alignment of varying myelination. Figure 5.9 shows the standard deviation from the mean volume change in the average coordinate system of each individual group. Once again, the preterm group (d-f) shows greater variation than the control group (a-c), particularly towards the back of the brain.

5.3 Growth Between 1 and 2 Years

As discussed in Chapter 1, preterm infants often suffer from neurological impairments in later life. An important factor to consider in the analysis of injury to the brain resulting from preterm birth, therefore, is how the brain continues to grow into childhood. In this section, the growth between one and two years in analysed

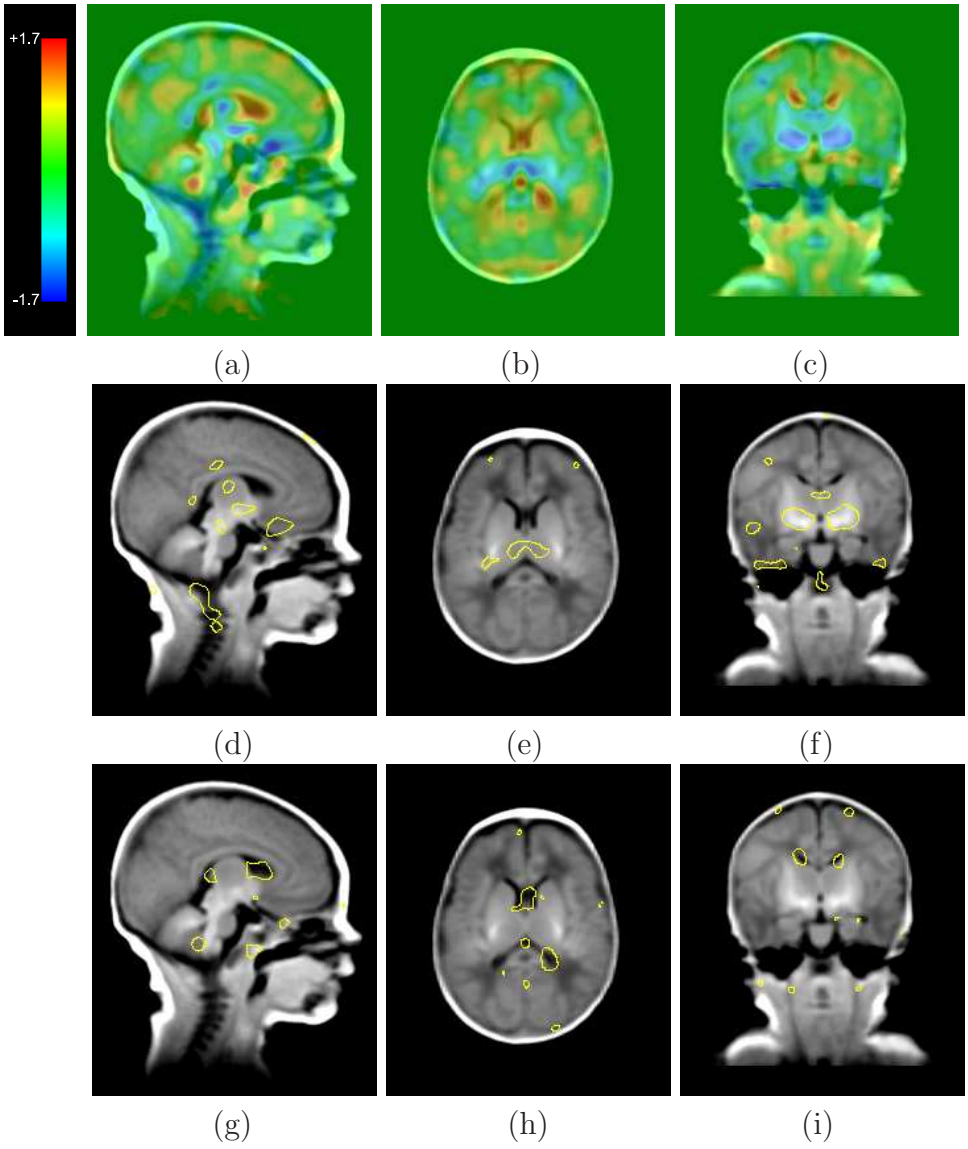


Figure 5.6: Effect sizes shown in the common coordinate system. Top row: overall distribution across the brain; middle row: areas where $\epsilon < -1$ (showing tissue contraction in the preterm group); bottom row: areas where $\epsilon > +1$ (showing tissue expansion in the preterm group).

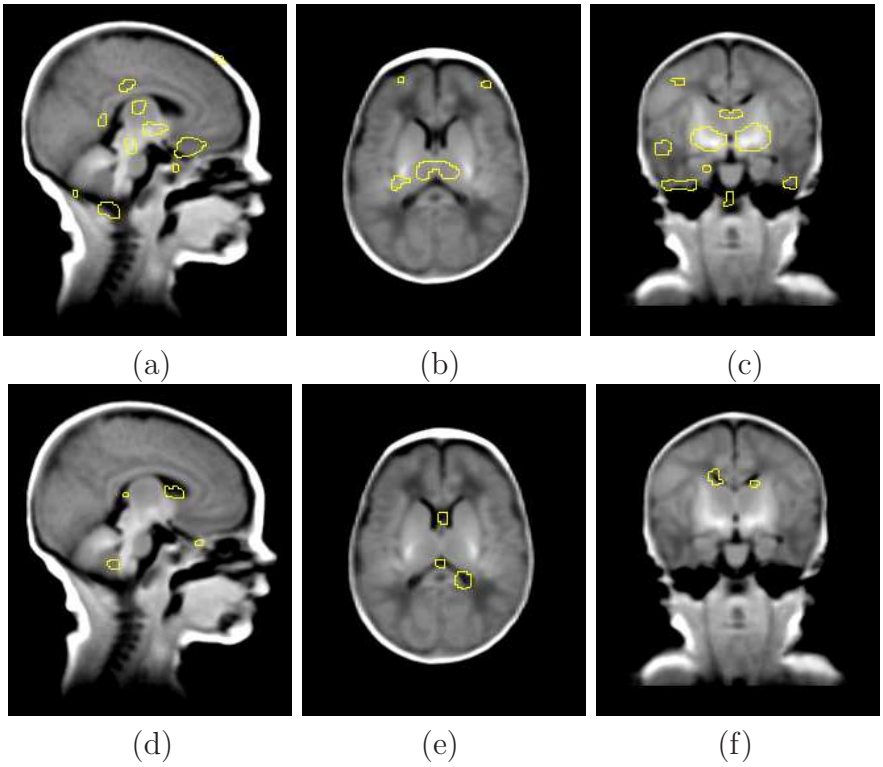


Figure 5.7: Statistically significant differences between controls and preterms at term-equivalent age. (a)-(c) areas smaller in the preterm group than the control group; (d)-(f) areas larger in the preterm group than the constrol group.

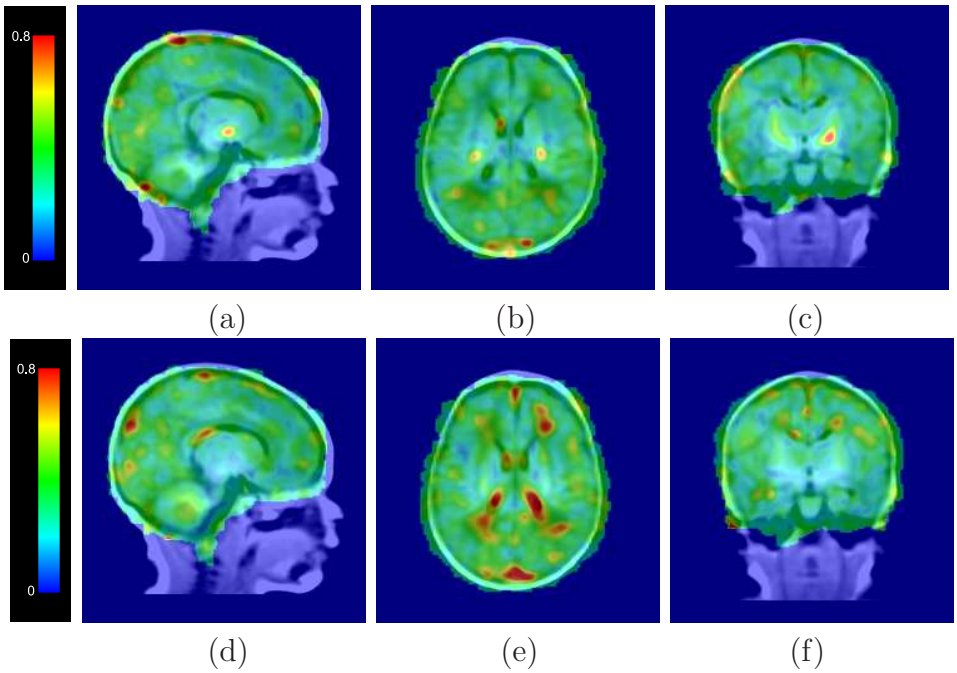


Figure 5.8: Standard deviations of volumetric changes of control (top row) and preterm (bottom row) populations in the average coordinate system of the combined population.

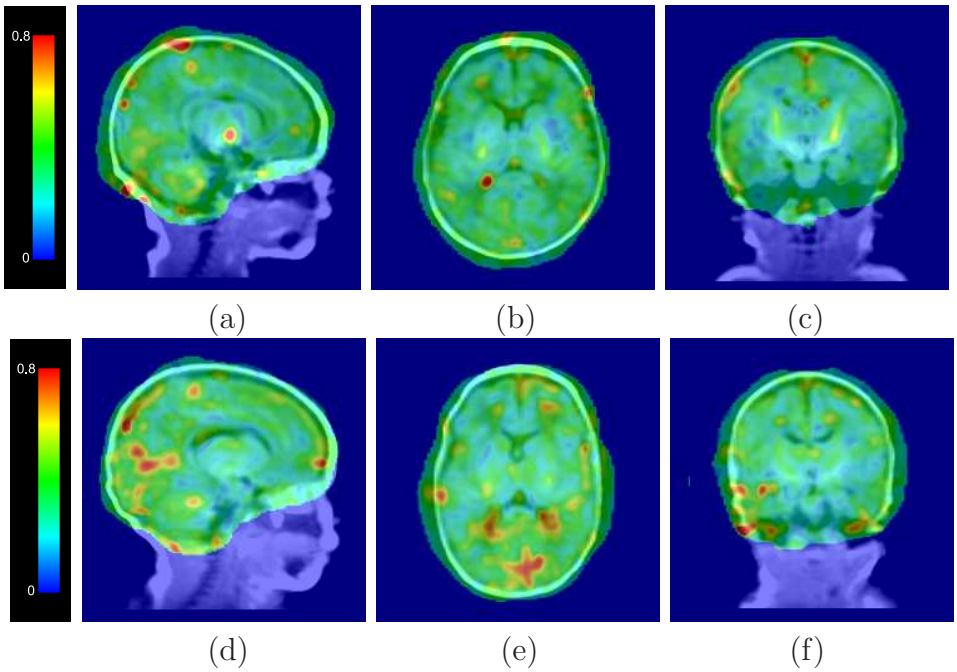


Figure 5.9: Standard deviations of volumetric changes in the average coordinate system of each individual population, showing increased variation in preterm population. Top row: controls. Bottom row: preterms.

by creating and comparing average atlases at these two time-points. The calculation of volumetric changes in infant growth additionally needs to account for overall growth in brain size, as well as local changes. The average affine shape and size of each group therefore also needs to be calculated. To do this, the log-averaging technique developed in [6] and discussed in Chapter 3 was used. A chosen arbitrary subject was affinely registered to the other subjects in the population. The inverse of the log (geometric) average of these affine transformations was then calculated using Equation 3.5, Chapter 3. For each subject, the transformation to the reference coordinate system was concatenated with the inverse of the average to give a transformation to the average affine space. Groupwise non-rigid deformation fields were then calculated in the same was as described in Section 5.2.2.

5.3.1 Subjects and image acquisition

18 preterm-born infants were scanned at one and two years of age. At birth, the mean gestational age of the group was 27.7 weeks, with a standard deviation of 2.2

weeks. The corrected gestational ages (the age corrected for gestational age at birth) at scan were 54.0 weeks and 106.4 weeks with standard deviations of 5.8 weeks and 4.4 weeks respectively.

The images used here are T1-weighted MR volumes. The scans of seven subjects were acquired using a Marconi 0.5T Apollo scanner (TR=23ms, TE=6ms, flip angle=30°), while the remaining subjects were scanned using a 1.0T HPQ system (TR=23ms, TE=6ms, flip angle=35°). The voxel dimensions were $1.0 \times 1.0 \times 1.6$mm in all cases.

5.3.2 Structural growth between 1 and 2 years

Average atlases of the one-year-old and two-year-old groups are shown in Figure 5.10. A visual comparison shows the increase in overall brain size.

The atlas of the two-year-olds has been used as a template and registered to the atlas of the one-year-olds using conventional pairwise registration. The process of finding the growth is shown schematically in Figure 5.11. This enables the volumetric changes that occur between the two ages to be determined, by calculating the determinant of the resulting deformation field. The volume changes from 1 to 2 years is shown in Figure 5.10(g-i). Although, as expected, most areas increase in volume, there are some areas (shown by the blue colouration) which actually shrink during this time. Such areas include the parts of the ventricles. This is not surprising as these are essentially fluid-filled spaces. This correlates with a previous study on seven preterm subjects scanned at one-year and two-years [6], with examples of volume changes obtained shown in Figure 5.12. To obtain these growth maps, the individual growth of each subject was found and averaged in a common space, resulting in a less noisy volume change map. However, corresponding areas appear to be changing in the same way.

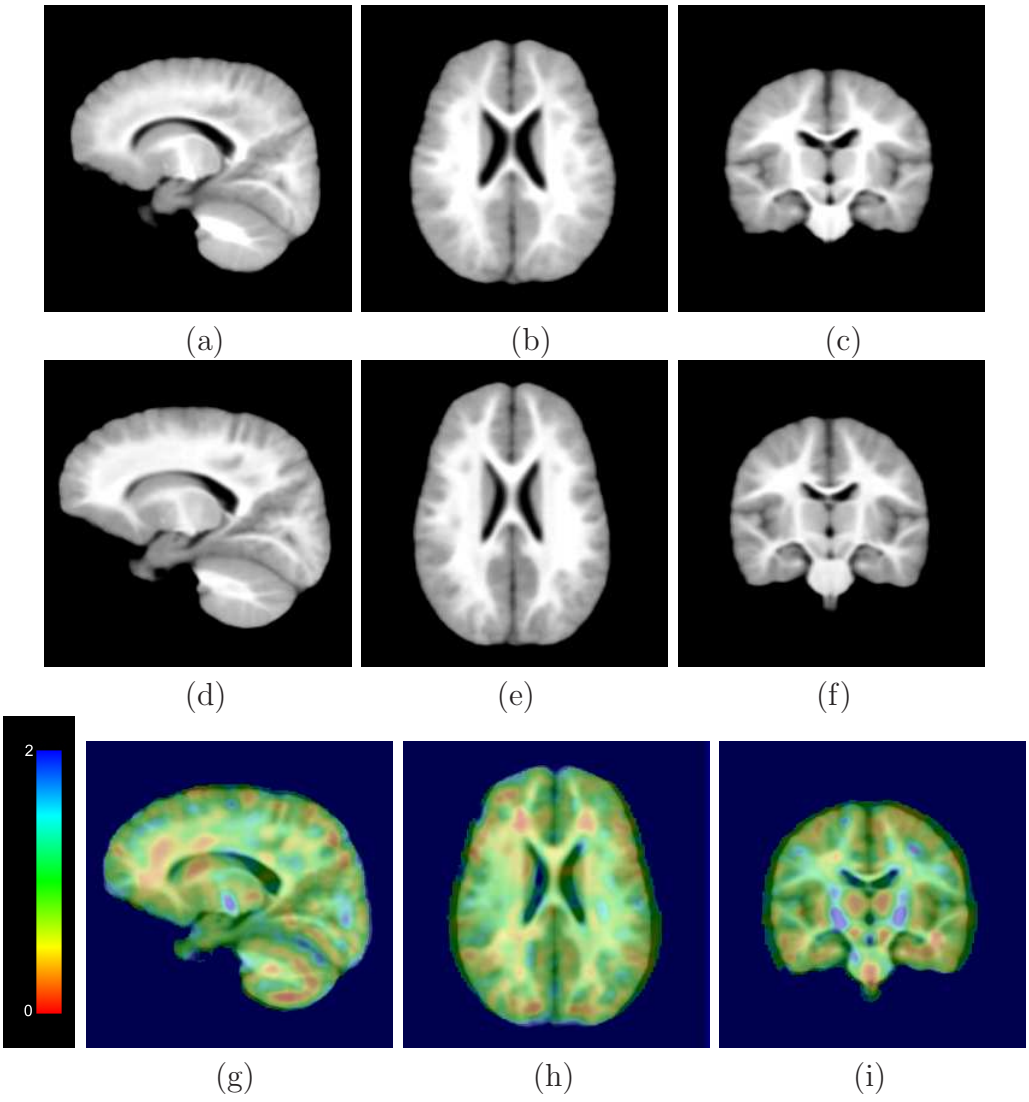


Figure 5.10: Comparison of average atlases of one- (a-c) and two-year-olds (d-f) and growth maps (g-i) showing average volumetric changes between the two ages (shown in the coordinate system of the two-year old atlas). Areas where $|\mathbf{J}| < 1$ indicate growth of tissue from one to two years. Areas where $|\mathbf{J}| > 1$ indicate shrinking of tissue from one to two years. The same population is used in each age group.

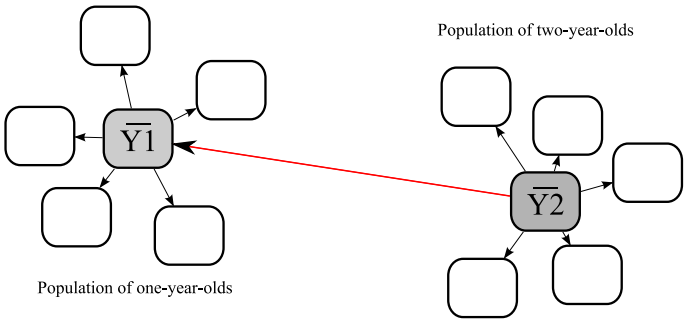


Figure 5.11: Mapping the growth between the one and two-year old populations.

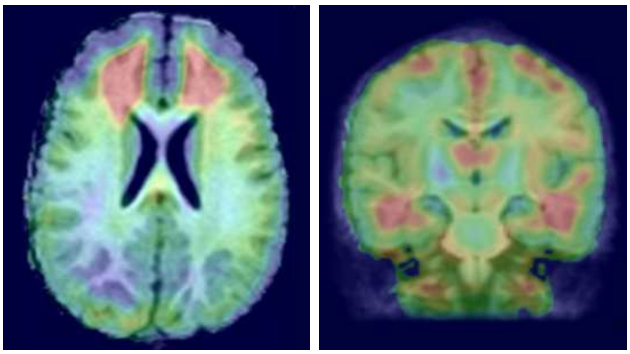


Figure 5.12: Growth between one and two years found by averaging growth of individual subjects in common space [6].

5.4 Summary

Groupwise non-rigid registration has been used to create unbiased average atlases of preterm and term-born infants at term-equivalent age. This has enabled the visual comparison of these groups to detect general differences in size and shape and myelination. Additionally, deformation-based morphometry has been used to determine volumetric differences between the groups. The results obtained are comparable to those found in previous studies [6, 7] and are consistent with clinical outcomes. The use of groupwise registration means that the results are not biased towards any particular reference subject and the differences found can therefore be described as typical for the populations being considered.

Average atlases of preterm infants scanned at one- and two-years-old have also been created. This has allowed the growth of structures over this time period to be determined. The vast changes that occur between term and one-year mean that registration of subjects at these two time-points is infeasible. In order to track growth of structures over this time period, it is necessary to obtain scans at many intermediate time-points.

This work could be further extended to quantify the growth and differences in individual structures. This requires the segmentation of structures or tissue classes in

each atlas, or in images of individual subjects. However, automatic segmentation for neonatal and child brain images is a difficult task due to the lack of prior atlases on these populations. The following chapter introduces methods to combine groupwise registration to the average coordinate system with automated methods of segmentation, in order to segment a population of subjects.

Chapter 6

Combined Groupwise Segmentation-Registration

6.1 Introduction

Image segmentation of MR images involves assigning, for each voxel in the image, a label indicating to which tissue or structure that voxel belongs. For example, in MR brain images, voxels can be separated into those representing white matter (WM), grey matter (GM) and cerebro-spinal fluid (CSF), or into more specific anatomical structures such as the ventricles, thalamus, or caudate. Segmentation is a crucial tool for medical image analysis. It allows for the quanitification of structural volumes, which can be used to analyse morphological differences over time or between subjects. For example, in longitudinal studies, the changes in size of the hippocampus of Alzheimer's or Schizophrenia patients could be tracked, as in [61] [8], or the growth of white and grey matter in neonatal infants could be measured [6]. Alternatively, in cross-sectional studies, differences between control subjects (such as term-born infants) and study subjects (such as pre-term born infants) can be detected and quantified [29]. The ability to segment an image into GM, WM and

CSF additionally assists in the 3D visualisation and morphometric analysis of the cortex [128].

In this chapter a novel *groupwise segmentation algorithm* is developed to segment images of a *population* of subjects. This utilises the information provided by the group when all images are aligned to the average space of the population. Additionally, the groupwise segmentation algorithm is incorporated with the groupwise registration algorithm proposed in Chapter 4 to produce methods for *combined groupwise segmentation and registration*.

6.1.1 Single image segmentation methods

MR image segmentation is still often done manually by a trained expert outlining the structures of interest. This can be time-consuming and laborious. Moreover, manual segmentations are difficult to reproduce and are prone to inter- and intra-operator variability, potentially suffering from operator bias and fatigue. There is therefore a need for automatic methods of segmentation.

Automatic segmentation, however, is rarely a trivial task. Methods such as thresholding, which create binary partitions of the image intensities in order to classify voxels, are often insufficient. This method labels all voxels with an intensity above a certain threshold as one class, and all those below the threshold as another. This may result in an estimate of the classes, but, the nature of MR images prevents this from being an effective tool. In particular, the segmentation of MR images is complicated by:

1. Low tissue contrast. The strength of MR magnets and the need to obtain scans in reasonable times often results in low contrast between different tissues [266].

2. Intensity inhomogeneity (bias field). This results in "shading" of the image giving different intensity profiles in different regions of the image. This is

primarily caused by inhomogeneities in the applied magnetic field during the scan [219]. As these inhomogeneities are due in part to the shape of the subject being scanned, it is not possible to use prior calibration to account for these [46]. Techniques do exist to estimate and correct for the bias field. These include the N3 (nonparametric, nonuniform intensity normalisation) algorithm [220] which models the intensity nonuniformity as a smooth multiplicative field. Such methods can additionally be built into the segmentation algorithm [116, 15, 191].

3. Partial volumes. The limited spatial resolution of MR scanners means that a single voxel sometimes represents tissues from more than one class [192].

4. Noise. The noise in MR images follows a Rician distribution [216], the shape of which is dependent on the signal-to-noise (SNR) ratio. For high SNRs (above three), the distribution approaches a Gaussian distribution [216] and, in practice, much research in image segmentation assumes a Gaussian distribution for MR noise.

Many different methods have been used for medical image segmentation. A brief description of some popular methods is given here. For a full review of such methods, see [186].

6.1.1.1 Supervised and unsupervised classifiers

Classification techniques aim to partition the data (feature space) into known categories. In the case of MR image segmentation, this involves allocating, for each voxel in an image, a label denoting the tissue class it represents. This is assigned depending of the distribution of the intensities of the original image. Classifiers fall into two categories: *unsupervised* (which do not require training data) and *supervised* (which require the model to be trained beforehand). Further details of classification techniques can be found in [75].

Clustering algorithms, such as the K-means and fuzzy c-means, algorithms are unsupervised methods as they require no prior training of the model. However, they do require an initial estimation of the segmentation or of the parameters of the model. Starting with an initial estimate of the segmentation, the K-means algorithm [48] iteratively computes the mean intensity for each tissue class. A voxel is labelled by assigning it with the tissue class label with the closest mean. This gives a hard segmentation. The fuzzy c-means clustering algorithm [185] applies the same technique to soft segmentations.

In contrast, supervised classifiers require prior training of the model on similar, manually-segmented, data. *Non-parametric* classifiers, such as Parzen windowing or k-Nearest Neighbour (kNN), make no assumptions about the distribution of the underlying data. The kNN algorithm [85] compares the intensity of the voxel to be segmented with the k closest intensities from the training data. It then assigns the most popular label of these neighbours to that voxel. Parzen windows [179] instead perform the classification according to the majority vote within a predefined window of voxels, centred on the voxel to be labelled.

An example of a supervised, *parametric* classifier is the maximum likelihood (ML), or Bayes, classifier [82]. This assumes that the distribution of the underlying data forms a finite mixture model (usually a mixture of Gaussians for MR brain images). Each tissue class represents one of the components of the mixture. The aim is to calculate the mixture and distribution parameters that lead to the highest probability of obtaining the image intensities. This method can be trained by estimating parameters of the distribution for each mixture (tissue class) on pre-segmented sample images. For example, for a Gaussian mixture model, the mean and variance of each Gaussian distribution, as well as the mixing parameters (which represent how much of each mixture there is), could be estimated from training samples. This method provides a soft, or probabilistic segmentation, and the hard labelling can be

obtained by assigning to the voxel the class with the highest probability.

However, the ML problem can also be solved without the use of training data. Instead, a good estimate of the initial probabilistic segmentation (known as a *prior* model) can be used. This method has been adopted in many brain MR image segmentation problems [138, 139, 116]. It is solved using an optimisation algorithm known as the Expectation-Maximisation (EM) algorithm [70], which iteratively improves the soft segmentation and the Gaussian parameters. This will be discussed in more detail in Section 6.2.

6.1.1.2 Deformable models

Deformable models are based on physical models of elasticity. These are used for segmentation by modelling image intensities (or their derivatives) as forces which act on the deformable model. Segmentation occurs when forces acting on the model are minimised or balanced. Deformable models fall into two categories: *parametric* and *geometric*. Parametric models were introduced by Kass et al. in [125]; a survey of these can be found in [163]. These use splines (or snakes) which are subjected to internal, external and image forces - these forces give rise to a corresponding energy of the spline. The aim is to guide the spline to the position of least energy, which should occur at image boundaries. However, splines cannot easily handle changes in topology (e.g.: region splitting and merging), and often need to be initialised close to the boundary requiring segmentation. Additionally, the internal energy constraints limit their geometric flexibility, thus reducing their effectiveness on more complicated structures.

Geometric models such as level-set methods, developed in [215], address the problems of changing topology. These are based on the geometric evolution of fronts with curvature-dependent speeds. The contour used for segmentation is embedded as the zero level-set of a higher-dimensional function. The propagation of the con-

tour is stopped in the region of image boundaries by an external force based on the intensities of the image (for example, on the gradient of the image intensities [155]).

6.1.1.3 Active shape/appearance models (ASMs/AAMs)

Active shape models (ASMs) [55] are statistical models of shapes of objects which iteratively deform to fit the shape of a new structure. A set of corresponding landmarks are positioned along the contours of known segmentations of the same structure in different training examples. These are used to build statistical shape models of the structure to be segmented. The statistical shape model is used to constrain the ASM to vary only in ways seen in the training data. A model of appearance around each landmark is also used to guide the segmentation. A simple example of such a model is to assume that landmarks should lie along strong edges. An extension to ASMs are active appearance models (AAMs) [53] which additionally incorporate statistical models of intensity variation across a whole of the region of interest, instead of just near modelled edges.

6.1.1.4 Atlas-based approaches

Atlas-based approaches [71, 50, 203, 105] essentially treat segmentation as a registration problem. A chosen reference, or atlas, image is manually segmented into the required tissue classes. This image is then registered to the image that needs to be segmented. When the images are aligned, the tissue labels can be transferred across corresponding voxels to give a segmentation of the new image. By using a non-rigid deformation field for the registration process, this additionally provides information on volumetric changes for different structures between subjects.

6.1.2 Segmentation and registration as complementary processes

Chapter 4 showed how probabilistic (soft) segmentations of subjects can be used to achieve groupwise registration of the population, with better results than when using similarity metrics based on image intensities. This is one example where the segmentation of an image helps in its registration. In general, segmentation helps to identify anatomical structures even when the intensities within a given structure may vary, and even when these intensities may be similar to those in a different structure. If every structure (in, say, a brain image) could be accurately labelled, this would greatly simplify the registration process.

However, it can also be argued that registration can contribute to solving the segmentation problem. When images are aligned, there is more information about the same structures available to guide the segmentation. This has been the motivation behind multi-modal segmentation where images from the same subject, acquired at the same time, but using different imaging modalities, are used to help the segmentation [254]. There has also been work produced showing how the alignment of prior information about segmentation with the image to be segmented can improve the segmentation [62, 191, 15].

If a perfect registration between a pre-existing labelled image and an image to be segmentated exists, the labelling could be transferred directly to new image. This method of atlas-based label propagation has been widely investigated in MR brain image segmentation: a reference image is labelled (perhaps by manual deliniation of the structures) and is registered to a new subject to be segmented. The labels can then be transferred to this new image [105, 202, 230]. This technique is often used to segment subjects from a similar population to the atlas. Rohlfing et al. [202] showed that the segmentation can be improved by using a pre-labelled atlas

which represents the average of the population to be segmented. In another recent work, Heckemann et al. [105] showed that the accuracy of the segmentations can be additionally improved by combining multiple segmentations using decision fusion.

It can be seen therefore that there is a link between segmentation and registration and the improvement of one is likely to lead to an improvement in the other. Only in recent years, however, have there been developments in methods to combine classification methods of segmentation with registration. Ashburner [15] and Pohl [191] aim to segment an intensity image by integrating the registration of a probabilistic prior atlas with the intensity image. The registration is considered to be a parameter to be optimised to get the best segmentation, using the EM algorithm. However, no attempt made to use segmentations to assist in the registration of images. This is done in [38], where two images (a segmented target and a floating image) are registered. The transformation between the two images is used to improve the segmentation in the target space. In this chapter, we aim to simultaneously segment and register a *population* of subjects, without the need for existing segmentations of any of the subjects. The work is most similar in its aim to Petrovic et al. [184], which provides a framework for registration, segmentation and modelling of a set of images.

The EM algorithm has been widely used for MR brain image segmentation [116] [138] [254]. However, its use generally requires known tissue class priors. Commonly-used priors are the MNI 305 priors [80] which have been formed by taking the voxel-wise mean of hard segmentations of WM, GM and CSF of 305 affinely-aligned subjects. Priors used should be representative of the population from which the subject to be segmented is taken. Using unrepresentative priors will bias the segmentation towards the priors, potentially resulting in errors in the final segmentation. However, representative priors may not be easily obtained as they themseleves are created from the segmentation of multiple subjects of the same population.

This chapter introduces a novel *groupwise combined segmentation and registration algorithm*. The aim of this is to concurrently align to, and segment a population of images in, the average coordinate system of the population. Information gained from the alignment of multiple subjects is used to aid the segmentation of each individual. The updated segmentations are then used to improve the alignment of the group. At each iteration, an updated atlas of priors is created, representing the population at that stage. By design, these prior atlases are already non-rigidly aligned to the average shape of the population. Two methods are proposed here: an *interleaved* method, which iterates between the segmentation and registration processes, and an *integrated* method, which uses a Bayesian framework to combine the registration parameters into the segmentation using a Maximum A-Posteriori approach.

6.2 The Expectation-Maximisation (EM) algorithm for brain MR image segmentation

As discussed previously, a commonly-used method for the segmentation of brain MR images is the EM algorithm [70, 172]. This has been used successfully in [138, 139, 116]. Additionally, recent work on joint segmentation and registration has used EM or Bayesian type methods for segmentation [38, 15, 191]. The EM algorithm is a parametric method which assumes that voxel intensities are independent samples taken from a mixture of tissue classes. For the purposes of brain MR image segmentation, each tissue class is represented by a Gaussian distribution (Figure 6.1), although the EM algorithm can be equally applied to other distributions.

For a given image, let $\mathbf{Y}$ be the collection of J voxels, each with intensity y_j, i.e. $\mathbf{Y} = \{y_j, j = 1, 2...J\}$. Assume we wish to segment this image into K tissue classes and let $l_j \in \{1, 2...K\}$ be the tissue class to which voxel j belongs (its

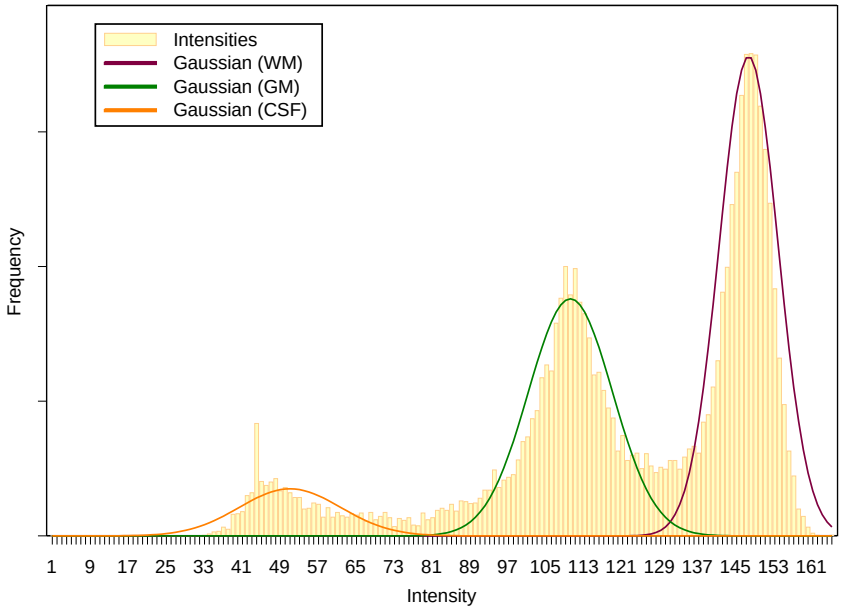


Figure 6.1: Intensity histograms of an axial slice of the MNI Brainweb image and Gaussian distribution approximations for white matter, grey matter and CSF.

label). Each tissue class, k, can be assumed to be approximately Gaussian distributed, with mean instensity $\mu_{j,k}$ and variance $\sigma^2_{j,k}$. These form the distribution parameters $\theta = \{\mu_k, \sigma^2_k, k = 1, 2...K\}$. The overall image can be considered to be a mixture of Gaussian distributions. The labelling of each voxel can be denoted by a vector $\mathbf{z}$ of length K, where $z_k \in [0, 1]$. When a voxel is labelled as belonging to class l, $z_l = 1$ and $z_{k \neq l} = 0$ everywhere else. The probability that a voxel belongs to class l can be represented by $p(z_l = 1) = \pi_l$.

The goal of image segmentation is to find, for each voxel, the tissue label l that best explains the voxel intensity y (*classification*). This is also dependent on the model parameters μ_k and σ_k chosen for each tissue class, k. However, the model parameters are, in turn, best estimated when the tissue classification is known. This suggests that an approach to iteratively update the model parameters and the tissue classification would result in both better estimation and better classification. This is achieved by the EM algorithm.

The EM algorithm aims to maximise the likelihood of the model parameters (here the mean and variance of the tissue classes) of a Gaussian mixture model, given a set of observations (the image intensities). It provides an iterative solution to maximum likelihood estimates for observations which form incomplete (or missing) data. In terms of medical image segmentation, the observations are the intensities of the voxels at each location, and the missing data are the tissue classes to which these voxels belong. The EM algorithm essentially takes the expectation over the missing data. The likelihood of the data set is given by:

$$P(\mathbf{Y}|\boldsymbol{\theta}) \tag{6.1}$$

This is maximised using the EM algorithm. First, the general theory of the EM algorithm will be considered. Its application to the specific problem of MR image segmentation follows.

6.2.1 General Theory of EM

The EM algorithm is an iterative approach to solving optimisation problems where some hidden (or *latent*) variables are unknown. It is used to estimate parameters, $\boldsymbol{\theta}$, of the model, given observations of the data, $\mathbf{Y}$. The derivation of the algorithm presented here is based on the lower-bound maximisation derivation by Dellaert [69] and Minka [170]. The maximum a-posteriori (MAP) estimate aims to maximise the posterior probability of the parameters, given the observations:

$$\hat{\theta} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, P(\boldsymbol{\theta}|\mathbf{Y}) \tag{6.2}$$

There is, however, no analytical solution to this equation. However, incorporating additional variables, $\mathbf{Z}$, of the model, which are not yet known and cannot be directly observed - hidden variables - makes the problem easier to solve. Marginalising over

$\mathbf{Z}$ gives:

$$\hat{\theta} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \sum_{\mathbf{Z}} P(\boldsymbol{\theta}, \mathbf{Z}|\mathbf{Y}) \tag{6.3}$$

From the laws of conditional probability, the posterior is proportional to the joint distribution, and this can be used instead:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \sum_{\mathbf{Z}} P(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}) \tag{6.4}$$

Maximising this is equivalent to maximising the natural logarithm of the same function:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \log P(\mathbf{Y}, \boldsymbol{\theta}) = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \log \sum_{\mathbf{Z}} P(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}) \tag{6.5}$$

However, this new term depends on taking the logarithm of a (potentially very large) sum term. To make the computation more feasible, a lower bound of the function is found. Finding and maximising this lower bound will maximise the original function also. To do this, requires the use of an arbitrary probability function over the hidden variables: $w^{(t)}(\mathbf{Z})$ where $\sum_Z w^{(t)}(\mathbf{Z}) = 1$.

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \log \sum_{\mathbf{Z}} \frac{P(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta})}{w^{(t)}(\mathbf{Z})} \cdot w^{(t)}(\mathbf{Z}) \\ &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \log \left(E_{w^{(t)}(\mathbf{Z})} \left(\frac{P(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta})}{w^{(t)}(\mathbf{Z})} \right) \right) \end{aligned} \tag{6.6}$$

where $E_{w^{(t)}(\mathbf{Z})}$ denotes the expectation. This is now in a form where Jensen's inequality [150] can be exploited:

$$\log E_{w^{(t)}(\mathbf{Z})} \left(\frac{P(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta})}{w^{(t)}(\mathbf{Z})} \right) \geq E_{w^{(t)}(\mathbf{Z})} \log \left(\frac{P(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta})}{w^{(t)}(\mathbf{Z})} \right) = Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}) \tag{6.7}$$

The right-hand-side of Equation 6.7 forms the lower bound of the objective function. By maximising this, the function itself will eventually be maximised. In this format, the log of sums has been transformed into the sum of logs which is much easier to

work with. The lower bound of the objective function is therefore given by:

$$Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(t)}) = E_{w^{(t)}(\mathbf{Z})} \log\left(\frac{P(\mathbf{Y},\mathbf{Z},\boldsymbol{\theta})}{w^{(t)}(\mathbf{Z})}\right) = \sum_{\mathbf{Z}} \log\frac{P(\mathbf{Y},\mathbf{Z},\boldsymbol{\theta})}{w^{(t)}(\mathbf{Z})} \cdot w^{(t)}(\mathbf{Z}) \tag{6.8}$$

6.2.1.1 Expectation

The optimal bound, given by $Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(t)})$, touches the objective function at the current estimate $\boldsymbol{\theta}^{(t)}$. That is $Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(t)}) = \log P(\boldsymbol{\theta}^{(t)}|\mathbf{Y})$ at $\boldsymbol{\theta}^{(t)}$. It is therefore necessary to find the $w^{(t)}(\mathbf{Z})$ which maximises the function:

$$Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(t)}) = \sum_{\mathbf{Z}} w^{(t)}(\mathbf{Z}) \log\left(\frac{P(\mathbf{Y},\mathbf{Z},\boldsymbol{\theta}^{(t)})}{w^{(t)}(\mathbf{Z})}\right) \tag{6.9}$$

Adding a Lagrange multiplier, λ, to enforce the constraint that $\sum_{\mathbf{Z}} w^{(t)}(\mathbf{Z}) = 1$, and rewriting gives an objective function of:

$$F = \lambda\left(1 - \sum_{\mathbf{Z}} w^{(t)}(\mathbf{Z})\right) + \sum_{\mathbf{Z}} w^{(t)}(\mathbf{Z}) \log P(\mathbf{Y},\mathbf{Z},\boldsymbol{\theta}^{(t)}) - \sum_{\mathbf{Z}} w^{(t)}(\mathbf{Z}) \log w^{(t)}(\mathbf{Z}) \tag{6.10}$$

Taking the derivative with respect to $w^{(t)}(\mathbf{Z})$ and solving gives:

$$w^{(t)}(\mathbf{Z}) = \frac{P(\mathbf{Y},\mathbf{Z},\boldsymbol{\theta}^{(t)})}{\sum_{\mathbf{Z}} P(\mathbf{Y},\mathbf{Z},\boldsymbol{\theta}^{(t)})} = P(\mathbf{Z}|\mathbf{Y},\boldsymbol{\theta}^{(t)}) \tag{6.11}$$

The optimal bound, occurring at $\boldsymbol{\theta}^{(t)}$ is therefore given by:

$$Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(t)}) = E_{\mathbf{Z}|\mathbf{Y},\boldsymbol{\theta}^{(t)}} \log\left(\frac{P(\mathbf{Y},\mathbf{Z},\boldsymbol{\theta})}{P(\mathbf{Z}|\mathbf{Y},\boldsymbol{\theta}^{(t)})}\right) = \log P(\mathbf{Y},\boldsymbol{\theta}^{(t)}) \tag{6.12}$$

The function:

$$w = P(\mathbf{Z}|\mathbf{Y},\boldsymbol{\theta}) \tag{6.13}$$

represents the posterior probabilities of $\mathbf{Z}$ given the observations $\mathbf{Y}$ and the parameters $\boldsymbol{\theta}$. Finding the value of this function constitutes the Expectation step of the EM algorithm.

6.2.1.2 Maximisation

Once the lower bound of the function is found, it needs to be maximised. The parameters at the next iteration are therefore given by:

$$\boldsymbol{\theta}^{(t+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}) = E_{w^{(t)}(\mathbf{Z})} \log \left(\frac{P(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta})}{w^{(t)}(\mathbf{Z})} \right) \tag{6.14}$$

Rearranging using rules of conditional probabilities and dropping terms that do not depend on $\boldsymbol{\theta}$ gives, for the expected log-likelihood:

$$\boldsymbol{\theta}^{(t+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}) = E_{w^{(t)}(\mathbf{Z})} \log \left(\frac{P(\mathbf{Y}, \mathbf{Z}|\boldsymbol{\theta}) P(\boldsymbol{\theta})}{w^{(t)}(\mathbf{Z})} \right) \tag{6.15}$$

$$\boldsymbol{\theta}^{(t+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, E_{w^{(t)}(\mathbf{Z})} \left(\log P(\mathbf{Y}, \mathbf{Z}|\boldsymbol{\theta}^{(t)}) + \log P(\boldsymbol{\theta}) \right) \tag{6.16}$$

Here $P(\boldsymbol{\theta})$ denotes the prior information of the probability of the parameters occurring. Solving this step represents the maximisation of the lower bound of the objective function - the Maximisation step. Alternatively, the bound can also be maximised using the expected log-posterior:

$$\boldsymbol{\theta}^{(t+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}) = E_{w^{(t)}(\mathbf{Z})} \log \left(\frac{P(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{Z})}{w^{(t)}(\mathbf{Z})} \right) \tag{6.17}$$

which leads to the same result.

6.2.1.3 Summary of EM algorithm

The aim of the EM algorithm is to find the optimal lower bound of the objective function and then maximise this bound to maximise the function. This involves two steps:

- *Expectation:* Calculate $w^{(t)} = P(\mathbf{Z}|\mathbf{Y}, \boldsymbol{\theta}^{(t)})$ using the current estimate of $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$.
- *Maximisation:* Solve for $\boldsymbol{\theta}^{(t+1)} = \operatorname{argmax}_{\boldsymbol{\theta}} \log \left(P(\mathbf{Y}|\mathbf{Z}, \boldsymbol{\theta}) + \log P(\boldsymbol{\theta}) \right)$.

This converges to a local maximum of $\log P(\mathbf{Y}|\boldsymbol{\theta})$ and also maximises the log-likelihood $\log P(\boldsymbol{\theta}|\mathbf{Y})$. A full proof of this is given in [70]. It should also be noted that simply improving the lower bound will also lead to an improvement in the solution. However, it will not maximise the function and so convergence will be slower.

6.2.2 Gaussian Mixture Model

For the application to MR image segmentation, each tissue class, k, is modelled as a Gaussian distribution, with μ_k and σ_k^2 representing the mean and variance, respectively, of the intensities of each tissue class, k: $\boldsymbol{\theta} = \{\mu_k, \sigma_k^2, k = 1, 2, ...K\}$. The Gaussian distribution for a voxel j is then given by:

$$G_k(y_j) = \mathcal{N}(y_j|\mu_k, \sigma_k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_j - \mu_k)^2}{2\sigma_k^2}\right] \tag{6.18}$$

for a single variable. For a multi-dimensional vector $\mathbf{y}$, the multivariate Gaussian distribution is given by:

$$G_k(\mathbf{y}_j) = \mathcal{N}(\mathbf{y}_j|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}_k|}} \exp\left(-\frac{1}{2}(\mathbf{y}_j - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{y}_j - \boldsymbol{\mu}_k)\right) \tag{6.19}$$

The overall image can then be regarded as a mixture of Gaussian distributions: a linear combination of many single Gaussian distributions. The Gaussian density of each tissue class forms one component of the mixture, with its own mean and (co)variance. The overall probability density function of the intensity is therefore:

$$p(\mathbf{y}_j) = \sum_k G_k(\mathbf{y}_j) \cdot \pi_{j,k} \tag{6.20}$$

where π_k denotes the mixing coefficients. These coefficients represent the prior probability of a voxel being generated by the component k of the mixture.

6.2.3 Incorporating tissue labelling

In addition to the tissue classes being modelled as Gaussian distributions with certain means and variances, each voxel in an image can be labelled as belonging to a single tissue class, l. These tissue labels represent the hidden variables of the problem. Let the hidden variables $\mathbf{Z}$ be represented by a vector $\mathbf{z}$ of length K. When a voxel is classified as being tissue class l, the value of the vector is $z_l = 1$. Everywhere else, $z_{k \neq l} = 0$. Therefore $z_k \in [0, 1]$ and $\sum_k z_k = 1$. The mixing coefficients represent the prior probability of one of these labels:

$$p(z_k = 1) = \pi_k \tag{6.21}$$

where $\sum_k \pi_k = 1$ and $0 \leq \pi_k \leq 1$. Assuming this is known, and that the tissue labellings of the voxels are statistically independent, this gives the overall probability density of the labelled image:

$$p(\mathbf{Z}) = \prod_k \pi_k^{z_k} \tag{6.22}$$

Therefore:

$$p(\mathbf{y}|z_k = 1) = G_k(\mathbf{y}) \tag{6.23}$$

and

$$p(\mathbf{y}|\mathbf{z}) = \prod_k \mathcal{N}(\mathbf{y}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_k} = \prod_k G_k(\mathbf{y})^{z_k} \tag{6.24}$$

Noting that the value of $z_k = 1$ if the label is k and $z_k = 0$ elsewhere, this gives the marginal distribution of $\mathbf{y}$ by summing over all possible labels:

$$p(\mathbf{y}) = \sum_{\mathbf{z}} p(\mathbf{y}|\mathbf{z}) p(\mathbf{z}) = \sum_k G_k(\mathbf{y}) \cdot \pi_k \tag{6.25}$$

Each of J voxel locations provides a different observation (the intensity of the voxel), y_j and has an associated, hidden (or unknown) vector $\mathbf{z}$, indicating the tissue class label of the voxel. Equation 6.25 is the mixture model. This models the intensities in the image as a mixture of Gaussian distributions, each weighted by a prior

probability π_k. How these probabilities are found for MR image segmentation are discussed in Section 6.2.6.

6.2.4 Maximum Likelihood

The maximum likelihood problem, as related to MR image segmentation aims to assign to each voxel j, a label, $z_{jl} = 1$, indicating to which tissue class that voxel belongs. The image is modelled as a mixture of Gaussians, given observations of the image intensity, y_j, at each voxel. To do this, the probability of the observations, given the image parameters is maximised. This is equivalent to maximising the log of the likelihood function, summed over all the voxels (assuming statistical independence of each voxels):

$$p(\mathbf{y}|\mu, \mathbf{\Sigma}) = \prod_j \left(\sum_k G_k(\mathbf{y}) \cdot \pi_k \right) \tag{6.26}$$

$$\log p(\mathbf{y}|\mu, \mathbf{\Sigma}) = \sum_j \log \left(\sum_k G_k(\mathbf{y}) \cdot \pi_k \right) \tag{6.27}$$

Expectation step for Gaussian Mixture Model segmentation

Find the function:

$$Q(\boldsymbol{\theta}_j|\boldsymbol{\theta}_j^{(t)}) = E\left[\log p(y_j, z_j|\boldsymbol{\theta}_j)|y_j, \boldsymbol{\theta}^{(t-1)}\right] \tag{6.28}$$

This results in a classification step equivalent to finding the posterior probabilities for the data. By Bayes's theorem, the posteriors are given by:

$$p(z_l = 1|\mathbf{y}) = \frac{p(\mathbf{y}|z_l = 1)p(z_l = 1)}{\sum_k p(\mathbf{y}|z_k = 1)p(z_k = 1)} = \frac{G_l(\mathbf{y}) \cdot \pi_l}{\sum_k G_k(\mathbf{y}) \cdot \pi_k} \tag{6.29}$$

Maximisation step for Gaussian Mixture Model segmentation

It is now required to maximise the log-likelihood with respect to the model param-

eters. Setting the derivative with respect to μ_k to zero gives:

$$\frac{\partial}{\partial \mu_k}\left(\sum_j \log\left(\sum_k p(y_j|z_{j,k}=1,\boldsymbol{\theta}_j)\right)\right)=0 \tag{6.30}$$

By differentiating and substituting in the Gaussian distribution, the value of μ_k that maximises the log-likelihood can be found to be:

$$\mu_k^{(t+1)}=\frac{\sum_j p(z_{j,k}=1|y_j,\boldsymbol{\theta}_j^{(t)})\cdot y_j}{\sum_j p(z_{j,k}=1|y_j,\boldsymbol{\theta}_j^{(t)})} \tag{6.31}$$

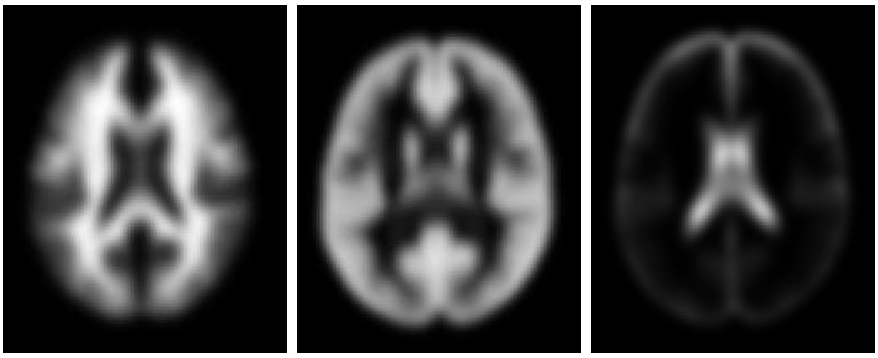
A similar approach can be taken to find the value of σ_k^2 which maximises the function:

$$(\sigma_k^2)^{(t+1)}=\frac{\sum_j p(z_{j,k}=1|y_j,\boldsymbol{\theta}_j^{(t)})\cdot (y_j-\mu_k^{(t)})^2}{\sum_j p(z_{j,k}=1|y_j,\boldsymbol{\theta}_j^{(t)})} \tag{6.32}$$

Since μ_k does not depend on σ_k^2, calculating the mean before the variance is necessary to maximise the objective function. This process improves both the classification and the parameter estimates at every iteration, leading to an increase in the log-likelihood.

6.2.5 Multi-channel EM

A set of aligned subjects can be viewed as a set of samples from the same underlying distribution. The more samples present, the easier it should be to determine the distribution. For example, it would be easier to manually segment an atlas formed from a number of aligned subjects, than it would be to manually segment each of the individual images. This has led to algorithms for multi-channel segmentation. For example, Van Leemput et al. [139, 254] use the EM algorithm with more than one input channel to segment adult brain images. These channels typically consist of two or three images of the same subject, acquired using different imaging modalities, such as T1-, T2- and proton-density (PD)-weighted MR. Here, the Gaussian distributions of classes k become multivariate normals with mean $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$.



(a) (b) (c)

Figure 6.2: The MNI 305 prior probability maps for white matter (a) grey matter (b) and CSF (c).

These are updated in the EM algorithm as follows:

$$G_k(\mathbf{y}_j) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}_k|}} exp\left(-\frac{1}{2}\left(\mathbf{y}_j - \boldsymbol{\mu}_k\right)^T \boldsymbol{\Sigma}_k^{-1}\left(\mathbf{y}_j - \boldsymbol{\mu}_k\right)\right) \tag{6.33}$$

where

$$\mu_{k,c} = \frac{\sum_j y_{j,c} p(z_{j,k} = 1|y_j, \boldsymbol{\theta})}{\sum_j p(z_{j,k} = 1|y_j, \boldsymbol{\theta})} \tag{6.34}$$

and

$$\Sigma_{k,c_1,c_2} = \frac{\sum_j p(z_{j,k} = 1|y_j, \boldsymbol{\theta}) \cdot (y_j - \mu_{kc_1})(y_j - \mu_{kc_2})}{\sum_j p(z_{j,k} = 1|y_j, \boldsymbol{\theta})} \tag{6.35}$$

for channels c_1 and c_2 and voxel location j.

6.2.6 Prior probability maps

The EM algorithm requires some knowledge of the probability of a voxel location being a certain class - a *prior* probability for each class for each voxel. For adult brain images, the MNI priors [80] for WM, GM, CSF and background (BG) are often used (shown in Figure 6.2). These have been formed by taking the average of segmentations of 305 affinely-aligned adult brain images segmented into the above tissue classes. The use of so many subjects creates smooth priors for each tissue class and aims to encapsulate the variation of most normal subjects.

6.3 Atlas-to-image registration

In many implementations of the EM algorithm, priors, such as the MNI priors, are first affinely-aligned to the subject to be segmented, in order to initialise the process [138, 139, 2]. However, there is a large anatomical variation in the brains of different subjects - the cortex in particular varies largely in shape from subject to subject, or the subject may not represent normal pathology (e.g. larger ventricles are commonly present in Alzheimer's sufferers). For this reason, it would be preferable to use a prior atlas created from a population of subjects which is representative of the subject to be segmented. More representative priors should then lead to a better segmentation. Recent methods have aimed to compensate for unrepresentative atlases by non-rigidly aligning the prior atlas to the image to be segmented (instead of the more commonly-used affine alignment). D'Agostino et al. [62] have proposed a method to register a probabilistic atlas to an intensity image directly such that the likelihood of the intensities, given the spatially-deformed prior model, is maximised. For this, they develop an information-theoretic similarity measure. The similarity measure between image intensities, Y, and class label probabilities L is given by:

$$S(Y, L) = \sum_{k} \sum_{y} p(k, y) \log \frac{p(k, y)}{p(k) \cdot p(y)} \tag{6.36}$$

where k indexes the different class labels and y represents the image intensity. This can be calculated using histograms in a similar way to mutual information, each bin being incremented by the probability of being that tissue class.

6.4 Combined segmentation and atlas-to-image registration

More recent work by Ashburner et al. [15] and Pohl et al. [191] aims to register the subject to be segmented with the atlas, by incorporating the registration into

a Bayesian framework for segmentation using the EM algorithm. The objective is to find the registration that leads to a maximum log-likelihood (or expected log-posterior) in the segmentation.

6.4.1 Ashburner et al. (2005)

Ashburner et al. [15] integrate the registration of an intensity image to tissue class priors with the segmentation of that image. The segmentation is again modeled as a mixture of Gaussians and an EM approach is used. Additionally, they incorporate the bias field correction within the images. The overall aim of the method is then to maximise:

$$P(\mathbf{Y}, \boldsymbol{\beta}, R|\boldsymbol{\theta}) \tag{6.37}$$

where $\boldsymbol{\beta}$ represents the bias field parameters and R denotes the parameters of the registration. This is equivalent to minimising:

$$S = -\log P(\mathbf{Y}, \boldsymbol{\beta}, R|\boldsymbol{\theta}) \tag{6.38}$$

where the mixture parameters are given by $\boldsymbol{\theta} = \{\mu_k, \sigma_k^2, \gamma_k, k = 1, 2...K\}$. γ_k represents the prior probability of tissue k (which is dependent on the bias estimation). Modelling the problem as a mixture of Gaussians, with the additional provision for the bias field, $\rho(\beta)$, gives:

$$S = -\sum_j \log \left(\rho_j(\beta) \sum_k \frac{1}{\sqrt{(2\pi\sigma_k^2)}} \exp\left(-\frac{(\rho_j(\beta)y_j - \mu_k)^2}{2\sigma_k^2} \right) \cdot \pi_{j,k} \right) \tag{6.39}$$

Additionally, the priors used are spatially varying and these are to be deformed to match the image as the algorithm progresses.

The mixture parameters are optimised within an EM framework while holding the bias and deformation estimates fixed at their current "best" solutions. The bias is then optimised with the current mixture and deformation parameters fixed. The

deformations are optimised by taking the analytical derivative of the cost function (Equation 6.39) with respect to R and using the Gauss-Newton method [195]. The deformation model used is a low-dimensional warping algorithm based on combinations of basis functions [13].

6.4.2 Pohl et al. (2006)

Pohl et. [191] also propose a technique for combining segmentation with bias field correction and the registration of a prior atlas. This solves the MAP estimation problem of:

$$(\boldsymbol{\theta}^{(t+1)}, R^{(t+1)}) = \underset{\boldsymbol{\theta},R}{\operatorname{argmax}} \log P(\boldsymbol{\theta}, R|\mathbf{Y}) \tag{6.40}$$

With the addition of tissue labelling, $\mathbf{Z}$, indicating to which tissue class a voxel belongs:

$$(\boldsymbol{\theta}^{(t+1)}, R^{(t+1)}) = \underset{\boldsymbol{\theta},R}{\operatorname{argmax}} \log \left(\sum_{\mathbf{Z}} P(\boldsymbol{\theta}, R, \mathbf{Z}|\mathbf{Y}) \right) \tag{6.41}$$

The label map probabilities $\mathbf{P}(\mathbf{Z}|\mathbf{Y}, \boldsymbol{\theta}^{(t)}, R^{(t)})$ (where $(\boldsymbol{\theta}^{(t)}, R^{(t)})$ are the estimates of $\boldsymbol{\theta}$ and R at the previous iteration), are now incorporated as in Section 6.2.

$$\begin{aligned}(\boldsymbol{\theta}^{(t+1)}, R^{(t+1)}) &= \underset{\boldsymbol{\theta},R}{\operatorname{argmax}} \log \left(\sum_{\mathbf{Z}} \frac{P(\boldsymbol{\theta}, R, \mathbf{Z}|\mathbf{Y}) \cdot P(\mathbf{Z}|\mathbf{Y}, \boldsymbol{\theta}^{(t)}, R^{(t)})}{P(\mathbf{Z}|\mathbf{Y}, \boldsymbol{\theta}^{(t)}, R^{(t)})} \right) \\ &= \underset{\boldsymbol{\theta},R}{\operatorname{argmax}} \log E_{\mathbf{Z}|\mathbf{Y},\boldsymbol{\theta}^{(t)},R^{(t)}} \frac{P(\boldsymbol{\theta}, R, \mathbf{Z}|\mathbf{Y})}{P(\mathbf{Z}|\mathbf{Y}, \boldsymbol{\theta}^{(t)}, R^{(t)})}\end{aligned} \tag{6.42}$$

This is now in a form which can be maximised using the EM algorithm by maximising the lower bound of the above objective function:

$$(\boldsymbol{\theta}^{(t+1)}, R^{(t+1)}) = \underset{\boldsymbol{\theta},R}{\operatorname{argmax}} E_{\mathbf{Z}|\mathbf{Y},\boldsymbol{\theta}^{(t)},R^{(t)}} \left(\log \frac{P(\boldsymbol{\theta}, R, \mathbf{Z}|\mathbf{Y})}{P(\mathbf{Z}|\mathbf{Y}, \boldsymbol{\theta}^{(t)}, R^{(t)})} \right) \tag{6.43}$$

The following assumptions are made:

- Stationary image intensities. Y is therefore independent of the registration, R.

- The registration, R and the Gaussian parameters, $\boldsymbol{\theta}$, are independent.
- The Gaussian parameters, $\boldsymbol{\theta}$, are independent of the labels, $\mathbf{Z}$.

With these assumptions, and using Bayes's and conditional probability rules, Equation 6.43 simplifies to:

$$(\boldsymbol{\theta}^{(t+1)}, R^{(t+1)}) = \underset{\boldsymbol{\theta},R}{\operatorname{argmax}}\, E_{\mathbf{Z}|\mathbf{Y},\boldsymbol{\theta}^{(t)},R^{(t)}} \left(\log P(\mathbf{Y}|\mathbf{Z},\boldsymbol{\theta}) + \log P(R|\mathbf{Z}) + \log P(\boldsymbol{\theta})\right) \tag{6.44}$$

Assuiming statistically independent voxels, this yields:

$$\begin{aligned}(\boldsymbol{\theta}^{(t+1)}, R^{(t+1)}) = \underset{\boldsymbol{\theta},R}{\operatorname{argmax}} \sum_j \sum_k P(\mathbf{Z}_{k,j} = 1|\mathbf{Y}, \boldsymbol{\theta}^{(t)}, R^{(t)}) \cdot \\ (\log P(\mathbf{Y}_j|\mathbf{Z}_j, \boldsymbol{\theta}_j) + \log P(R|\mathbf{Z}_j) + \log P(\boldsymbol{\theta}))\end{aligned} \tag{6.45}$$

The posterior probabilities are once again calculated in the Expectation step in the same way as in the standard EM algorithm:

$$p(\mathbf{Z}_{j,l} = 1|Y, \boldsymbol{\theta}^{(t)} R^{(t)}) = \frac{p(Y_j|\mathbf{Z}_{j,l} = 1, \boldsymbol{\theta}) \cdot P(\mathbf{Z}_{j,l} = 1|R)}{P(Y_j|\boldsymbol{\theta}_j^{(t)}, R^{(t)})} = \frac{G_l(j) \cdot \pi_{jl}}{\sum_k G_k(j)\pi_{j,k}} := w_{j,l} \tag{6.46}$$

The parameters $\boldsymbol{\theta}$ and R updated in the maximisation step. Noting that each factor in Equation 6.44 depends on either R or $\boldsymbol{\theta}$, but never both, this can be separated into two independent equations:

$$\begin{aligned}R^{(t+1)} &= \underset{\boldsymbol{\theta},R}{\operatorname{argmax}} \sum_j \sum_k P(\mathbf{Z}_{j,l} = 1|\mathbf{Y}, \boldsymbol{\theta}^{(t)}, R^{(t)}) \cdot (P(\mathbf{Z}_{j,k} = 1, |R) + \log P(R)) \\ &= \underset{\boldsymbol{\theta},R}{\operatorname{argmax}} \sum_j \sum_k w_{j,k} \left(\log \pi_{j,k} + \log P(R)\right)\end{aligned} \tag{6.47}$$

$$\begin{aligned}\boldsymbol{\theta}^{(t+1)} &= \underset{\boldsymbol{\theta},R}{\operatorname{argmax}} \sum_j \sum_k P(\mathbf{Z}_{j,k} = 1|Y, \boldsymbol{\theta}^{(t)}, R^{(t)}) \cdot (P(\mathbf{Y}|\mathbf{Z}_k = 1, \boldsymbol{\theta}) + \log P(\boldsymbol{\theta})) \\ &= \underset{\boldsymbol{\theta},R}{\operatorname{argmax}} \sum_j \sum_k w_{j,k} \left(\log G_k(j) + \log P(\boldsymbol{\theta})\right)\end{aligned} \tag{6.48}$$

The Gaussian parameters are maximised using the standard method of solving the analytical derivative of Equation 6.48 with respect to each parameter. Since $w_{j,k}$ is fixed, maximising the registration paramters is equivalent to minimising the Kullback-Leibler divergence between the posterior distribution $w_{j,k}$ and the prior atlas π_k, if no prior information about the registration exists (that is, $P(R) = 0$).

6.5 Simultaneous segmentation and registration

Although the previous work attempts to register an image with a prior atlas to optimise the segmentation, it does not attempt to improve the registration between images. However, registration and segmentation would appear to be complementary processes. Segmentation allows ambiguous intensities to be defined as a certain class, which should aid registration. If subjects are aligned, it should be easier to segment the structures as there is more data available to make a decision.

6.5.1 Chen et al. (2004)

Chen et al. [38] pairwise register a target image, I_1, to a source (or floating) image I_2, using both the image intensities and existing probabilistic segmentations. The resulting transformation is then used to update the segmentation. When registered, the two images can be taken to be two observations of the same underlying scene (the segmentation), corrupted only by Gaussian noise, and therefore conditionally independent.

An initial segmentation of each image is found, together with the Gaussian distribution parameters of each tissue class (this could use, for example, the standard EM algorithm). The Gaussian parameters are held fixed throughout the process - it is only the segmentation of the target image which is updated. The initial segmentations are soft, or probabilistic, segmentations of each class, which is shown to be less sensitive to noise than hard segmentations. An iterative scheme is developed

which performs the following two steps:

1. Registration step: Set $\hat{\mathbf{T}} = \text{argmax}_{\mathbf{T}}\, P(\mathbf{T}|\hat{w}, I_1, I_2)$

2. Segmentation step: Set $\hat{w} = \text{argmax}_{w}\, P(w|\hat{\mathbf{T}}, I_1, I_2)$

where $\mathbf{T}$ indicates the current transformation and w represents the current soft segmentation. A similarity measure based on the source intensities and target segmentation is minimised during the registration step:

$$S_{REG} = -\sum_{j} log\left(\sum_{k}\left(G_{I_2,k}(y'_j)\cdot\hat{w}_{j,k}\right)\right) + E(T) \tag{6.49}$$

where $y'_j = I(\mathbf{T}(j))$ represents the intensity of the transformed voxel j, $w_{j,k}$ is the segmentation probability of class k at voxel location j, $E(\mathbf{T})$ is a regularisation term used to ensure a smooth transformation model, and $G_{I_2,k}(y)$ is given by:

$$G_{I_2,k}(y) = \frac{1}{\sqrt{2\pi\sigma^2_{I_2,k}}}\exp\left[-\frac{(y_{I_2}-\mu_{I_2,k})^2}{2\sigma^2_{I_2,k}}\right] \tag{6.50}$$

with y_{I_2} being the intensity of voxel j in image I_2. To conduct the registration itself, a B-spline transformation model is used as in [207].

The segmentation step then minimises:

$$S_{SEG} = -\sum_{j}\log\left(\sum_{k} G_{I_1,k}(y_j)\cdot w_{j,k}\right) - \sum_{j}\log\left(\sum_{k} G_{I_2,k}(y'_j)\cdot w_{j,k}\right) + E(p) \tag{6.51}$$

where $E(p)$ is a Markov Random Field model used to ensure spatial continuity between voxels, and $G_{I_1,k}(y_j)$ is the Gaussian of the target image for class k. This energy term is optimised by minimising the analytical derivative with respect to each p_k, using the projected gradient descent algorithm to constrain $\sum_k w_k = 1$.

6.5.2 Petrovic et al. 2006

A framework for combined image registration, segmentation and modelling of a 2D dataset is presented in [184]. This uses an iterative scheme to incrementally improve each of these, with the addition of modelling of partial volumes in the segmentation. The images are first warped to an initial affine reference space and initial estimates of the segmentations for each image obtained. These are used to calculate the most probable fraction f_{i,j,k_1,k_2} of each tissue in each voxel j (assuming at most two different tissues, k_1 and k_2, per voxel) in each image i:

$$f_{i,j,k_1,k_2} = \underset{f_{i,j,k_1,k_2}}{\operatorname{argmax}}\, p_{i,j,k_1,k_2}(g|f) \tag{6.52}$$

where $p_{i,j,k}(g) = \mathcal{N}(\mu_k, \sigma_k^2)$ is the distribution of pure tissue classes which follows a Gaussian distribution, and:

$$p_{i,j,k_1,k_2}(g|f) = \mathcal{N}(f\mu_{k_1} + (1-f)\mu_{k_2}, f\sigma_{k_1}^2 + (1-f)\sigma_{k_2}^2) \tag{6.53}$$

By estimating the fraction of each tissue type at each voxel for each image, reconstructions of each normalised image can be obtained. These reconstructions are then aligned to the current normalised population using an SSD intensity metric and a deformation model as in [52]. Furthermore, the optimum number of tissue classes is selected to be the one that results in a minimum description length of the modelled training set. How the use of a set of images, as opposed to considering each image individually, aids either the registration or the segmentation is not discussed in the paper. Additionally, the results presented are only of 2D data.

6.6 Combined groupwise segmentation-registration

Many techniques for segmentation rely on combining prior segmentations of comparable subjects. For example, [202, 105] have shown that fusing hard segmentations of

many subjects can lead to the better segmentation of a new subject. Alternatively, EM-based algorithms use prior atlases of affinely-aligned segmentations to aid in the segmentation of a new subject. These methods all rely on having an existing population of accurate hard segmentations. The problem of how to segment a new population of subjects is compounded by the fact that segmentations from a representative population may not exist. As discussed above, there has been a number of recent works aimed at non-rigidly aligning a standard prior atlas (such as the MNI set) with the intensity image to be segmented, in order to obtain priors which are more specific to, but not necessarily more representative of, the image. The aim of the work presented in the following sections is to build population-specific priors by combining groupwise registration and segmentation, and to use these priors to segment all individuals in the population. There are two aims for this section:

1. To use the population of aligned images to help in the segmentation process.
2. To use segmentations of the population to help in the registration process.
3. To create and use representative prior atlases that are non-rigidly aligned with the population.

6.6.1 Registration-based groupwise segmentation

A groupwise segmentation algorithm is first developed to segment individual subjects of a population, using information gained from the segmentation of the other subjects in the group. These subjects are initially non-rigidly aligned to a common average space of the population, so that corresponding voxel locations represent the same structures. It is possible to use the multi-channel EM algorithm with multiple aligned intensity images for segmentation of the common space. However, the need to calculate multi-dimensional Gaussian distributions once again creates dimensionality problems for large numbers of images, due to increased sparsity of the data (see Chapter 4). Instead of combining the intensities of individual images, we propose to combine their probabilistic segmentations, thus removing the need to estimate any

multi-dimensional functions.

In the algorithm presented here, the information provided by the other subjects in the population is instead used to create models for the probability, $\pi_{j,k}$, of a voxel j being a tissue class k. The simplest method to do this would be to take the mean of the individual posterior probability maps created at each iteration:

$$\pi_{j,k}^{(t+1)} = \frac{\sum_i p(z_{i,j,k} = 1 | y_{i,j}, \boldsymbol{\theta}_{i,j,k})}{n_I} = \frac{\sum_i w_{i,j,k}}{n_I} \tag{6.54}$$

However, if the population is small, or the original priors used are not unrepresentative of the population, it may not be prudent to completely disgard all the information from the initial priors. At each iteration, the model could be updated, instead of being completely recalculated. For example, the following formulation for updating the model could be used:

$$\pi_{j,k}^{(t+1)} = \lambda \left(\frac{\sum_i w_{i,j,k}}{n_I} \right) + (1 - \lambda)\, \pi_{j,k}^{(t)} \; \forall j, k \tag{6.55}$$

where λ is a weighting term which determines the influence of the previous model and the mean of the current posteriors. The values of $\pi_{j,k}^{(t+1)}$ need then to be normalised such that $\sum_k \pi_{j,k}^{(t+1)} = 1$.

A standard EM iteration can then be performed in order to maximise the log-likelihood function for each image, $i \in \{1, 2, ...I\}$, given this model:
Maximise:

$$S_i = \sum_j \log \left(\sum_k G_{i,k}(y_j) \cdot \pi_{j,k} \right) \tag{6.56}$$

Optimising this function with respect to the model $\pi_{j,k}$ directly will simply aim to increase all the prior probabilities subject constraint that $\sum_k \pi_{j,k} = 1$. Using this method, the increase of probabilities that should be increasing may not be reliably achieved. Instead, the function is optimised by using a single iteration of the EM

algorithm using the existing priors at first. At each subsequent EM iteration, a new model is created using the current posterior probabilities found by the previous iteration. The prior map for each tissue class therefore evolves with each iteration, and should get more population-specific as the confidence of the segmentations of the individual images improves. The full groupwise segmentation algorithm for a set of images aligned to a common space is therefore as follows:

1. Create model of probabilities, $\pi_{j,k}$, to be the mean of the tissue probabilities at each voxel location for each individual image using Equations 6.54 or 6.55. Initially, this model would require an input prior map for each image (for example, the MNI priors). As the input images are aligned, the same priors can be used for each image, resulting in an initial model equal to each of the individual priors.

2. For each image, perform one iteration of the EM algorithm

   (a) Parameter estimation: update Gaussian parameters and function based on current individual posterior probabilities for each image, i:

$$\mu_{ik} = \frac{\sum_j p(z_{i,j,k} = 1 | y_{i,j}, \boldsymbol{\theta}_{i,j,k}) \cdot y_{i,j}}{\sum_j p(z_{i,j,k} = 1 | y_{i,j}, \boldsymbol{\theta}_{i,j,k})} \tag{6.57}$$

$$\sigma^2_{i,k} = \frac{\sum_j p(z_{i,j,k} = 1 | y_{i,j}, \boldsymbol{\theta}_{i,j,k}) \cdot (y_{i,j} - \mu_{i,k})^2}{\sum_j p(z_{i,j,k} = 1 | y_{i,j}, \boldsymbol{\theta}_{i,j,k})} \tag{6.58}$$

$$G_{i,k}(j) = \frac{1}{\sqrt{2\pi\sigma^2_{i,k}}} \cdot \exp\left(-\frac{(y_{i,j} - \mu_{i,k})^2}{\sigma^2_{i,k}}\right) \tag{6.59}$$

   (b) Classification: update individual image probabilities based on new tissue class parameters and new model:

$$p(z_{i,j,l} = 1 | y_{i,j}, \boldsymbol{\theta}_{i,j,l}) = \frac{G_{i,l}(j) \cdot \pi_{j,l}}{\sum_{k=1} G_{i,k}(j) \cdot \pi_{j,k}} \tag{6.60}$$

3. Repeat (1)-(2) as often as necessary.

This creates, at each iteration, a model of priors for each class, specific to the population being studied, and in the same coordinate system as all of the subjects. The final segmentation can be transformed back into the image space using the inverse of the transformation which warps the original image into the common coordinate system.

6.6.2 Segmentation-based groupwise registration

The above algorithm requires the subjects to be aligned in the same common space, and so can be expected to work better the more well-aligned the subjects are. It also makes sense for this space to be as close to the individual subjects as possible - the average space as defined in Chapter 4 is therefore used. As shown in the same chapter, one of the best-performing methods in aligning a group of images to the average coordinate system uses the Kullback-Leibler divergence similarity measure. However, this in turn works best when accurate probabilistic segmentations of the group are available: a better segmentation leads to a better alignment and a better alignment leads to better segmentation. This suggests that combining the segmentation and registration processes together would be mutually beneficial.

The registration step proceeds as described in Chapter 4, using the Kullback-Leibler similarity metric on the current posterior probability maps. The transformation is found which minimises:

$$\sum_i \sum_j \sum_k p(z_{i,j,k} = 1 | y_{i,j}, \boldsymbol{\theta}_{i,j,k}) \log \left(\frac{p(z_{i,j,k} = 1 | y_{j,k}, \boldsymbol{\theta}_{i,j,k})}{\pi_{j,k}} \right) \tag{6.61}$$

where i is the image, j is the voxel and k is the tissue class index. For this, the model used should be the most representative of the whole population. The mean of the individual probabilities is therefore used:

$$\pi_{j,k} = \frac{\sum_i p(z_{i,j,k} = 1 | y_{i,j}, \boldsymbol{\theta}_{i,j,k})}{n_I} \tag{6.62}$$

where n_I is the number of images in the population.

6.6.3 Interleaved groupwise segmentation-registration

The segmentation and registration steps can be interleaved to give the full groupwise, segmentation-registration algorithm. For each iteration, the segmentation is estimated to give posterior probabilities for each image, $p(z_{i,j,k} = 1|y_{i,j}, \boldsymbol{\theta}_{i,j,k})$. These soft segmentations are then used to estimate the transformations using groupwise registration with the Kullback-Leibler divergence similarity measure. It should be noted however, that it is not necessary to use groupwise registration for this. Improving the pairwise registration using the Kullback-Leibler divergence, either to an average shape or to a chosen individual subject, should also help to improve the segmentation, as long as the alignment between subjects improves. Using a pairwise algorithm would increase the number of subjects that could potentially be used in this algorithm, which should in turn allow the creation of more representative priors.

6.6.4 A Bayesian approach to integrated groupwise segmentation-registration

An alternative approach is through a combined Bayesian formulation where the registration is modelled as one of the parameters of the segmentation, and solved for in a similar way to the Gaussian parameters. For a given location in a given image, the *maximum a-posteriori* (MAP) estimation is defined as follows:

$$(\boldsymbol{\theta}^{(t+1)}, R^{(t+1)}) = \underset{\boldsymbol{\theta},R}{\operatorname{argmax}} \log \left(\sum_{\mathbf{Z}} P(\boldsymbol{\theta}, R, \mathbf{Z}|\mathbf{Y}) \right) := \underset{\boldsymbol{\theta},R}{\operatorname{argmax}} f(\boldsymbol{\theta}, R, \mathbf{Z}|\mathbf{Y}) \quad (6.63)$$

This aims to maximise the probability of the Gaussian parameters, $\boldsymbol{\theta}$, the registration parameters, R, and the tissue labelling, $\mathbf{Z}$, given the image intensities, $\mathbf{Y}$. This is similar to the approach taken in [191]. However, the need to align all images to the average shape means that the image intensities, the Gaussian parameters and the

tissue labelling, all vary with the registration parameters. As shown in [191], and using the lower bound maximisation method described in Section 6.2.1, Equation 6.63 is equivalent to:

$$(\boldsymbol{\theta}^{(t+1)}, R^{(t+1)}) = \underset{\boldsymbol{\theta},R}{\operatorname{argmax}} E_{\mathbf{Z}|\mathbf{Y},\boldsymbol{\theta}^{(t)},R^{(t)}} \left(\log P(\mathbf{Y}|\mathbf{Z},\boldsymbol{\theta},R) + \log P(R|\mathbf{Z},\boldsymbol{\theta}) + \log P(\boldsymbol{\theta}|\mathbf{Z})\right) \tag{6.64}$$

where $E_{\mathbf{Z}|\mathbf{Y},\boldsymbol{\theta}^{(t)},R^{(t)}}$ denotes the expectation. Assuming the registration parameters do not depend on the Gaussian parameters, and that the Gaussian parameters do not depend on the tissue labelling, for statistically independent voxels, this yields:

$$\begin{aligned}
&(\boldsymbol{\theta}^{(t+1)}, R^{(t+1)}) = \underset{\boldsymbol{\theta},R}{\operatorname{argmax}} E_{\mathbf{Z}|\mathbf{Y},\boldsymbol{\theta}^{(t)},R^{(t)}} \left(\log P(\mathbf{Y}|\mathbf{Z},\boldsymbol{\theta},R) + \log P(R|\mathbf{Z}) + \log P(\boldsymbol{\theta})\right) \\
&= \sum_j \sum_k p(z_{j,k}=1|y_j,\boldsymbol{\theta}^{(t)},R^{(t)}) \cdot [\log p(y_j|z_{j,k}=1,\boldsymbol{\theta},R) + \log p(R|\mathbf{Z}) + \log p(\boldsymbol{\theta})] \\
&= \sum_j \sum_k p(z_{j,k}=1|y_j,\boldsymbol{\theta}^{(t)},R^{(t)}) \cdot \\
&[\log p(y_j|z_{j,k}=1,\boldsymbol{\theta},R) + \log p(z_{j,k}=1|R) + \log p(R) + \log p(\boldsymbol{\theta})]
\end{aligned} \tag{6.65}$$

Assuming uninformative priors for the registration and Gaussian parameters, this gives a final problem of:

$$(\boldsymbol{\theta}^{(t+1)}, R^{(t+1)}) = \underset{\boldsymbol{\theta},R}{\operatorname{argmax}} f(\boldsymbol{\theta}, R, \mathbf{Z}|\mathbf{Y})$$

$$= \underset{\boldsymbol{\theta},R}{\operatorname{argmax}} \sum_j \sum_k p(z_{j,k}=1|y_j,\boldsymbol{\theta}^{(t)},R^{(t)}) \cdot [\log p(y_j|z_{j,k}=1,\boldsymbol{\theta},R) + \log p(z_{j,k}=1|R)] \tag{6.66}$$

or

$$(\boldsymbol{\theta}^{(t+1)}, R^{(t+1)}) = \underset{\boldsymbol{\theta},R}{\operatorname{argmax}} \sum_j \sum_k w_{j,k}^{(t)} \cdot \left[\log G_k(y_j) + \log \pi_{j,k}^{(t)}\right] \tag{6.67}$$

This can also be solved using an EM framework. The expectation step calculates the posterior probabilities, $p(z_{j,k}=1|y_j,\boldsymbol{\theta}^{(t)},R^{(t)})$, and the maximisation step finds the values of $\boldsymbol{\theta}$ and R which maximise the new log-likelihood function (Equation 6.66). However, it can be seen that Equation 6.66 depends on both $\boldsymbol{\theta}$ and R. In

this case, a variant of the EM algorithm, the Expectation Conditional Maximisation (ECM) algorithm [165] can be used to calculate both the registration and segmentation parameters. The expectation step is the same as in Equation 6.29. In the maximisation step, one of the parameters is held fixed, while the value of the other which maximises the function is found. The new value of this parameter is then fixed, and the value of the first parameter is found to maximise the function:

$$(\boldsymbol{\theta}^{(t+1)}) = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, f(\boldsymbol{\theta}, R, \mathbf{Z}|\mathbf{Y})$$

$$(R^{(t+1)}) = \underset{R}{\operatorname{argmax}}\, f(\boldsymbol{\theta}^{(t+1)}, R, \mathbf{Z}|\mathbf{Y}) \tag{6.68}$$

Given this, [165] has proved that the following holds in the ECM:

$$f(\boldsymbol{\theta}^{(t+1)}, R) \geq f(\boldsymbol{\theta}, R) \tag{6.69}$$

$$f(\boldsymbol{\theta}^{(t+1)}, R^{(t+1)}) \geq f(\boldsymbol{\theta}^{(t+1)}, R) \tag{6.70}$$

The parameters $\boldsymbol{\theta}$ can be calculated as before, using Equations 6.31 and 6.32. The update of the registration parameters uses Equation 6.67 as the objective function to be maximised. This registration step can, again, either be done groupwise, to ensure the co-ordinate system is at the centre of the population, or pairwise to the current average estimate of the prior model.

6.6.5 Convergence criteria

At each iteration of both the groupwise segmentation and groupwise registrations methods, the model of priors is updated. Although at each iteration the log-likelihood (and therefore the theoretical segmentation) is improved for the model used, the use of different models renders the overall aim of maximising the log-likelihood for segmentation inappropriate. Instead, the optimum model needs be found.

From a Bayesian point-of-view, the best prior model is the one which is most probable given the data (the image intensities). For any given model M_x:

$$P(M_x|\mathbf{Y}) \propto P(\mathbf{Y}|M_x) \cdot \mathbf{P}(M_x) \tag{6.71}$$

The first term on the right-hand-side of the equation represents the *model evidence*. This implicitly penalises the error involved in using the model to predict an image, while the second term penalises complexity of the model. However, it is difficult to calculate the model evidence term:

$$P(\mathbf{Y}|M_x) = \int_{\phi} P(\mathbf{Y}|\phi, M_x) \cdot P(\phi|M_x) d\phi \tag{6.72}$$

where ϕ denotes the parameters that characterise the model. In this case, these can include the registration parameters and the posterior distributions. It is not feasible to calculate this term for all possible parameters.

Instead, a more intuitive approach is used, since the model represents the mean of the group of current segmentations. Given perfect segmentations, as the alignment of the images increases, the entropy of the model will decrease. Given perfect alignment, as the segmentation of each image converges, the entropy of the model will also decrease. The lowest model entropy will always occur when the images are segmented with greatest confidence and perfectly aligned. The aim of the segmentation-registration process is therefore to reduce the entropy of the model, and the process is terminated when this converges. The entropy of the prior model, M, at iteration t is given by:

$$H(M^{(t)}) = -\sum_{j}\sum_{k} \pi_{j,k}^{(t)} \log \pi_{j,k}^{(t)} \tag{6.73}$$

where $\pi_{j,k}$ gives the probability that voxel j is labelled as class k. This is equivalent to stopping the process when the complexity of the model converges, following the

principle of Occam's razor, which states that a model should be the simplest possible which fits the data.

6.6.6 Comparison of segmentation-registration methods

The interleaved and integrated approaches to combined groupwise segmentation-registration are shown diagramatically in Figures 6.3 and 6.4 respectively.

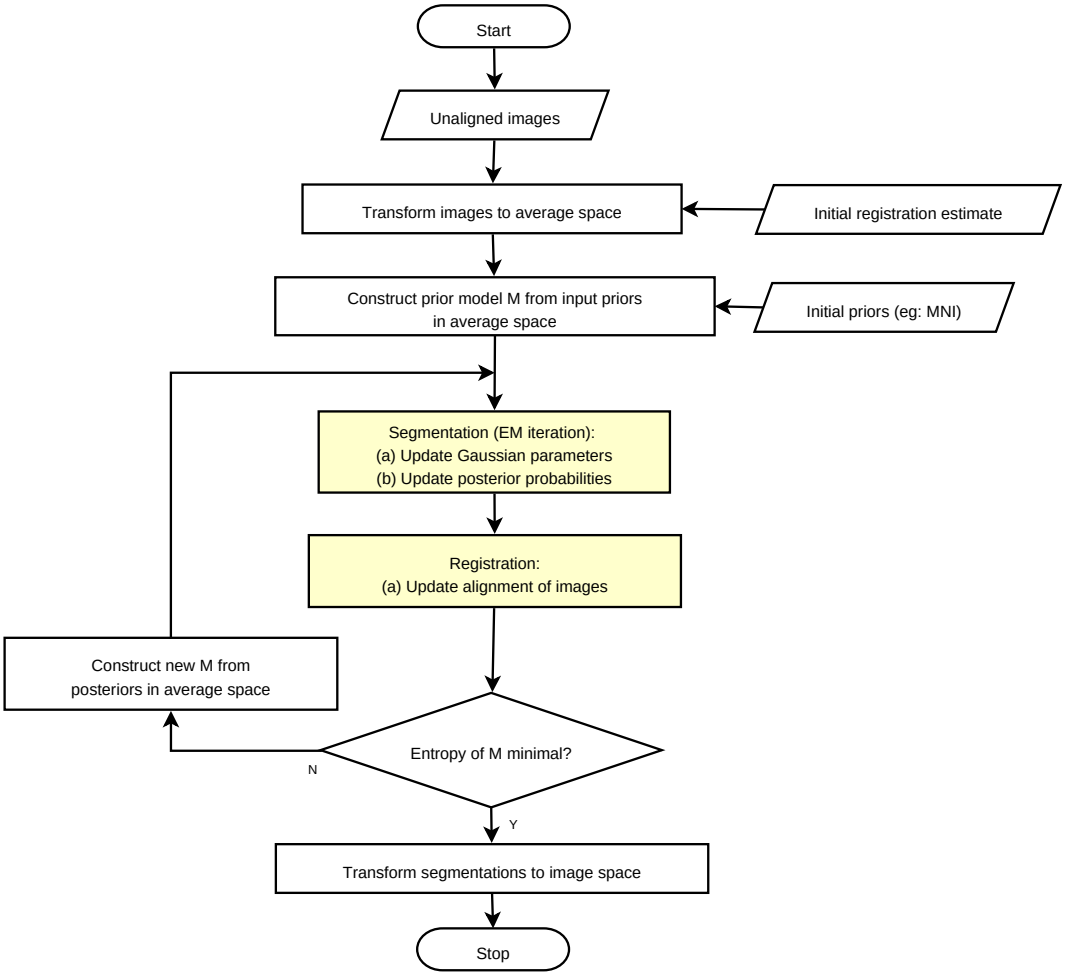


Figure 6.3: Flowchart showing the interleaved segmentation-registration algorithm.

Although both methods use the same groupwise segmentation technique to create increasingly specific and representative priors at every iteration, there are some fundamental differences. The interleaved segmentation-registration method iterates between:

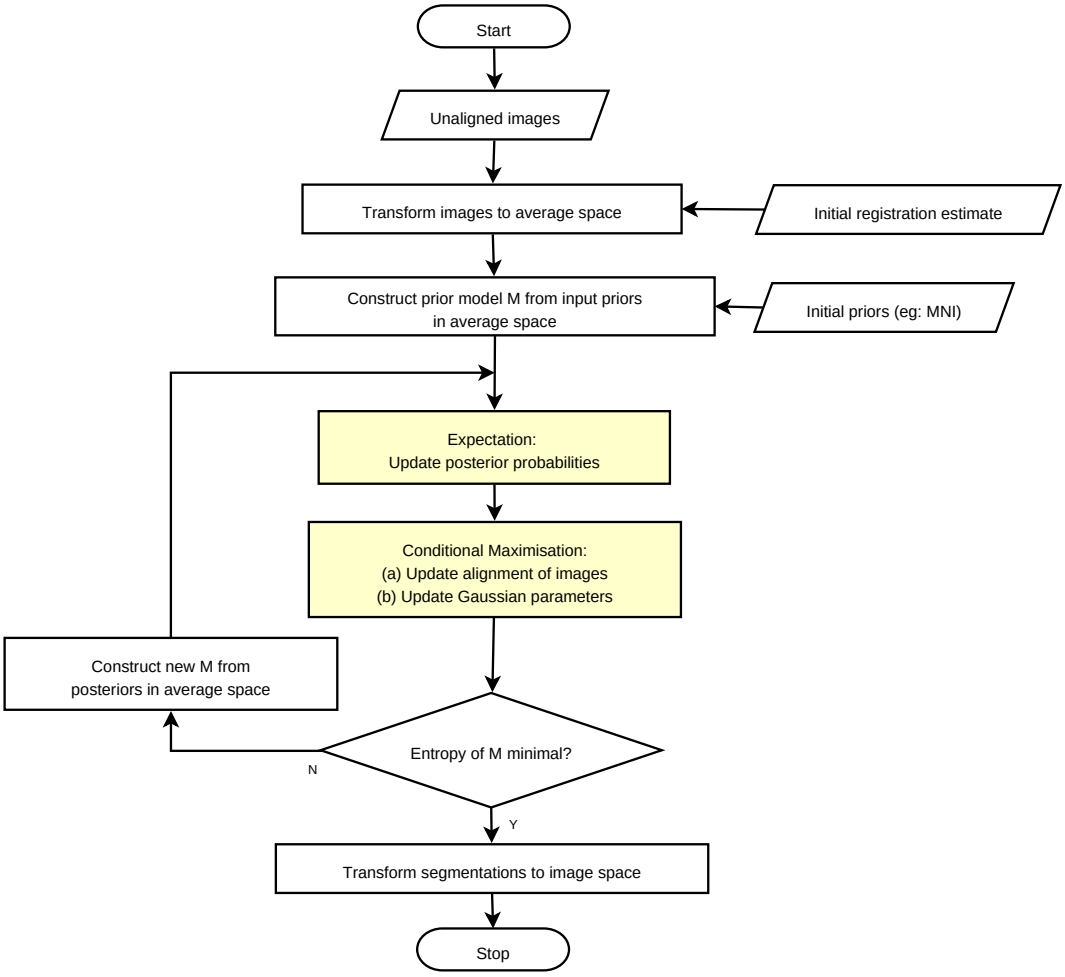


Figure 6.4: Flowchart showing the integrated (Bayesian) segmentation-registration algorithm.

1. *Segmentation*:

   (a) Maximisation: calculate Gaussian parameters using Equations 6.31 and 6.32.

   (b) Expectation: calculate posterior probabilities using Equation 6.29.

2. *Registration*: find the optimal transformation between the images by minimising Equation 6.61.

3. Repeat until convergence of entropy of prior model.

This has two aims: (a) to segment the images given the current alignment and (b) to align a population of subjects to their average shape, given their segmentations. However, it is not a specific requirement that improving the registration also improves the segmentation at every step.

The integrated approach instead aims to find the registration that leads to the best segmentation given the model. This consists of iterating between two steps (it is possible to start with either):

1. *Conditional Maximisation*: maximise Equation 6.67 with respect to:

   (a) Find the registration parameters which maximise Equation 6.67.

   (b) Find the Gaussian parameters which maximise Equation 6.67, given the updated registration parameters.

2. *Expectation*: calculate posterior probabilities using Equation 6.29.

3. Repeat until convergence of entropy of prior model.

However, improving the alignment of the group of images is not an explicit aim of the integrated method. Rewriting the objective function for this method (Equation 6.67), for an image i, gives:

$$\sum_j \sum_k w_{i,j,k} \left(\log G_k(y_j) \cdot w_{i,j,k} \right) - w_{i,j,k} \log \left(\frac{w_{i,j,k}}{\pi_{j,k}} \right) \tag{6.74}$$

The final term represents the KLD (Equation 6.61) between the posteriors and the model. The registration step of the integrated method aims to maximise all of Equation 6.74. However, as the first term depends only on the image under consideration, its maximisation may conflict with the second term which drives the alignment between images. It cannot, therefore, be expected that maximising the objective function of the integrated method will result in markedly or consistently improved registration between the images in the group.

In contrast, only the final term of Equation 6.74 is minimised in the registration stage of the interleaved method. In the segmentation stage, the Gaussian parameters are then updated to maximise the first term, given the updated registration (since only $G_k(y_j)$ is dependent on the Gaussian parameters). Updating the Gaussian parameters after updating the registration allows them to compensate for any reduction in the objective function caused by the change in registration.

6.7 Results

The algorithms above have been evaluated on synthetic 2D data and have also been used to segment populations of real 3D MR data of 22 one-year-old and 22 two-year-old preterm infants.

6.7.1 Artificially-deformed data

The same population of 2D synthetic data created in Chapter 4 has been used. The transformations to the average space, as well as the segmentations of the original images are therefore known. In addition, varying levels of Gaussian noise (with zero mean and standard deviation ranging from 0-5) were applied across the population.

6.7.1.1 Registration-based segmentation

First of all, it is necessary to test the premise that using the segmentations of a group of aligned subjects is preferable to trying to segment a single image using affinely-aligned priors. To do this, the sample population is transformed into the average space by the actual, known deformations and the groupwise segmentation algorithm is applied until convergence. For the initial input to the first iteration, the MNI priors are used.

To assess the criteria for the termination of the segmentation process, the negative log-likelihood is plotted at each iteration in Figure 6.5. This shows a decrease as the iterations increase. However, although the changes at later iterations become relatively small, the value does not actually converge. The entropy of the model is also plotted against the number of iterations in Figure 6.6. This converges after the fifth iteration. The models of priors as the algorithm progresses are shown in Figure 6.7. This shows how the models get sharper and more specific showing increased confidence in the segmentations. However, it can also be seen that some areas get increasingly misclassified, for example, there is a strong probability of areas surrounding the ventricles being grey matter, which is incorrect. These areas are particularly prone to partial-volume effects. As the alignment of the images increases, the misclassification of these areas becomes reinforced.

The segmentation results are compared to those obtained using the standard EM algorithm on each individual subject, using affinely-aligned MNI priors. For this population, the actual segmentations of all images in the population are known. To evaluate the accuracy of the segmentations produced using the methods described, the label consistency between the known and obtained segmentations are calculated for each image, as in Equation 4.8, Chapter 4. This evaluates the overlap between the segmentations. As can be seen from Figure 6.8, the groupwise segmentation algo-

rithm of aligned subjects outperforms the traditional EM algorithm on this dataset.

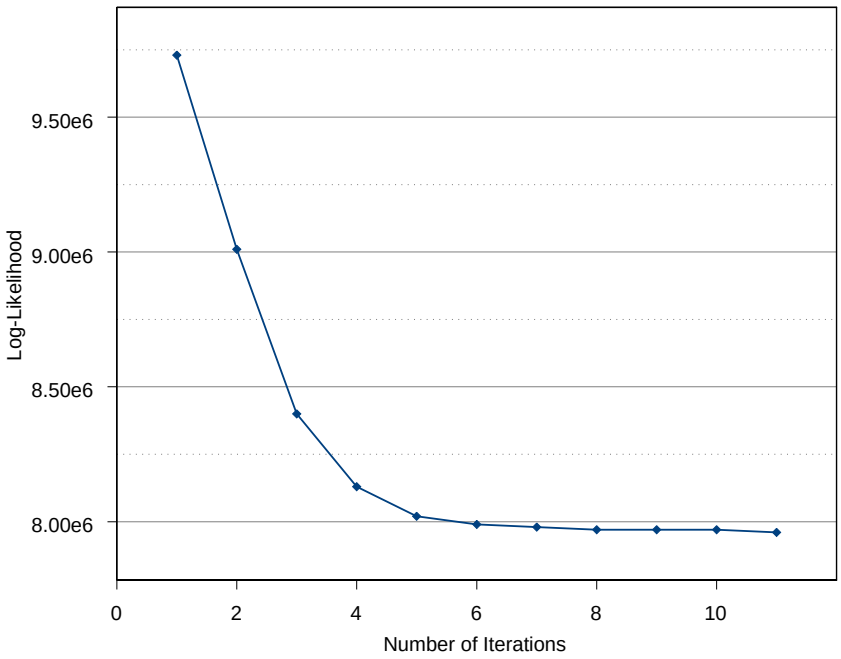


Figure 6.5: Evolution of log-likelihood with number of iterations for groupwise segmentation of perfectly aligned population.

6.7.1.2 Groupwise Segmentation-Registration

In this section, the interleaved and integrated segmentation-registration approaches are tested on the same dataset. An initial estimate of the registration is used (found using the groupwise registration approach described in Chapter 4), and the MNI priors are used as the initial prior probability map for each tissue class. Samples from the population together with their final segmentations using each method are shown in Figure 6.9. These have been compared to the segmentations obtained using the groupwise segmentation when all the images are already in the exact average space, and also to the segmentations obtained using the standard EM algorithm with the MNI priors. Additionally, the entropies of the prior models at each iteration are shown in Figure 6.10 for each method. The label consistencies at each iteration

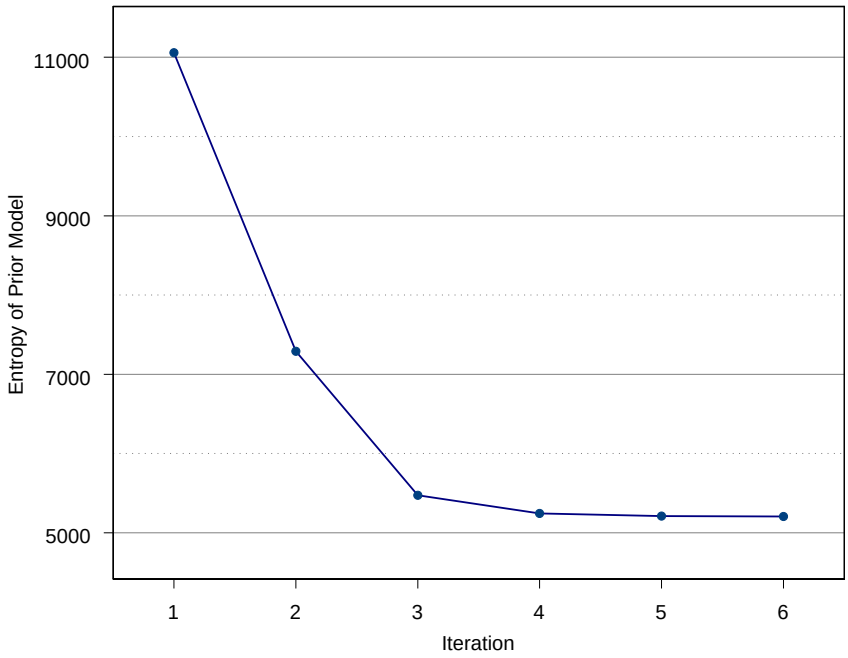


Figure 6.6: Evolution of entropy of prior model with number of iterations showing convergence at the fifth iteration for groupwise segmentation of a perfectly aligned population.

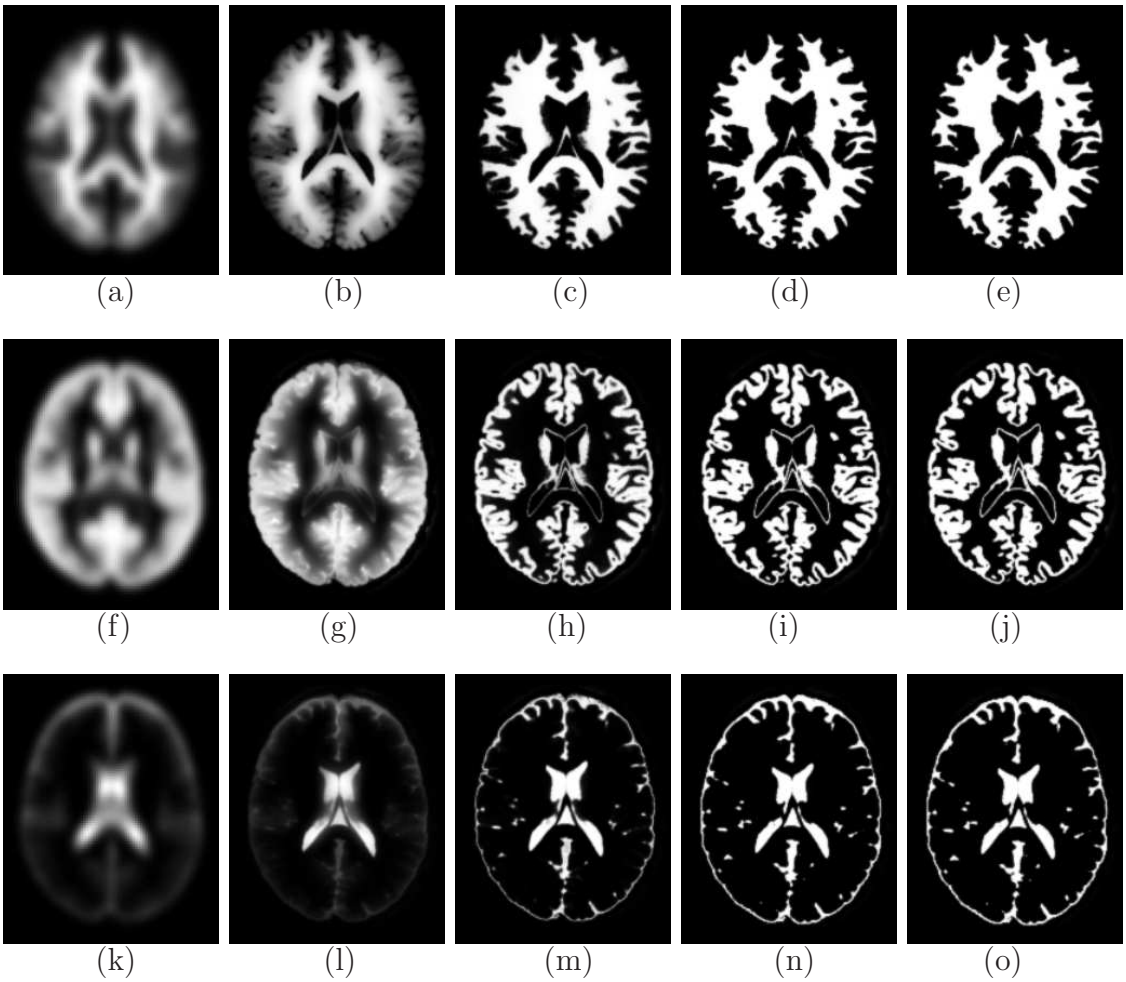


Figure 6.7: The evolution of prior models as the groupwise segmentation algorithm proceeds for WM (top row), GM (middle row) and CSF (bottom row). From left to right: the MNI priors, updated models at iterations 1, 3, 4 and 5 (convergence of model entropy).

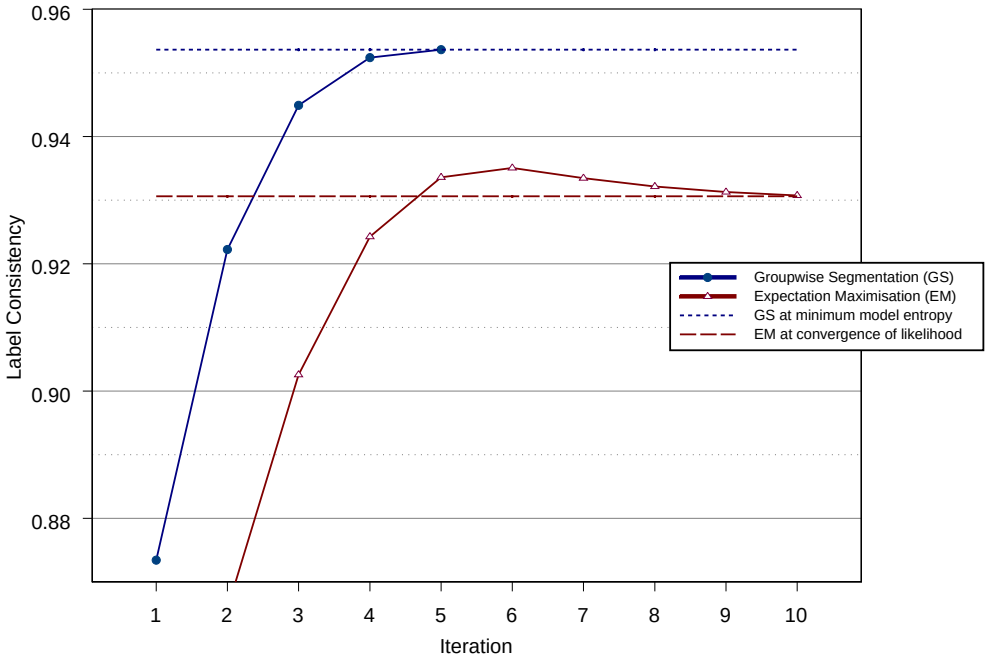


Figure 6.8: Average label consistency results of 100 images with groupwise segmentation using the actual, known transformation compared with single-subject EM (full lines). Average label consistency of final segmentation using groupwise segmentation and single-subject EM (dotted lines).

Method	BG	CSF	GM	WM
EM	0.976	0.779	0.871	0.922
Integrated	0.988	0.847	0.882	0.949
Interleaved	0.988	0.850	0.884	0.950
Groupwise	0.989	0.862	0.891	0.955

Table 6.1: Dice similarity results for each structure using each segmentation method.

and at the final point of the procedure are shown in Figure 6.11. Additionally, the Dice metric [72], as described in Chapter 4, is used to assess the alignment of each tissue class for each subject. These are shown in Table 6.1. The two groupwise segmentation-registration methods perform comparably well, and show a slight improvement on the accuracy of the segmentation over that of the standard EM algorithm alone. The difference between these methods is particularly noticeable for the CSF.

6.7.1.3 Effect on Registration

For the registration phase, both methods implemented groupwise registration with the constraint that the sum of all deformations should be equal to zero enforced. The average root-mean-squared (RMS) error of the displacements of the voxels is shown in Figure 6.12. As expected, the interleaved segmentation performs much better as maximising the alignment is a specific aim of this method. The integrated method has to balance maximisation of population alignment with the maximisation of an individual posterior to its segmentation. Comparing these results with Figure 4.11 of Chapter 4, it can be seen that interleaved method performs better than groupwise registration methods described in that chapter.

6.7.2 Groupwise segmentation of 3D MR images of one- and two-year-old infants born preterm

Populations of 22 preterm-born subjects imaged at one-year-old and again at two-years-old were aligned to their average shape at each time point, using the inverse

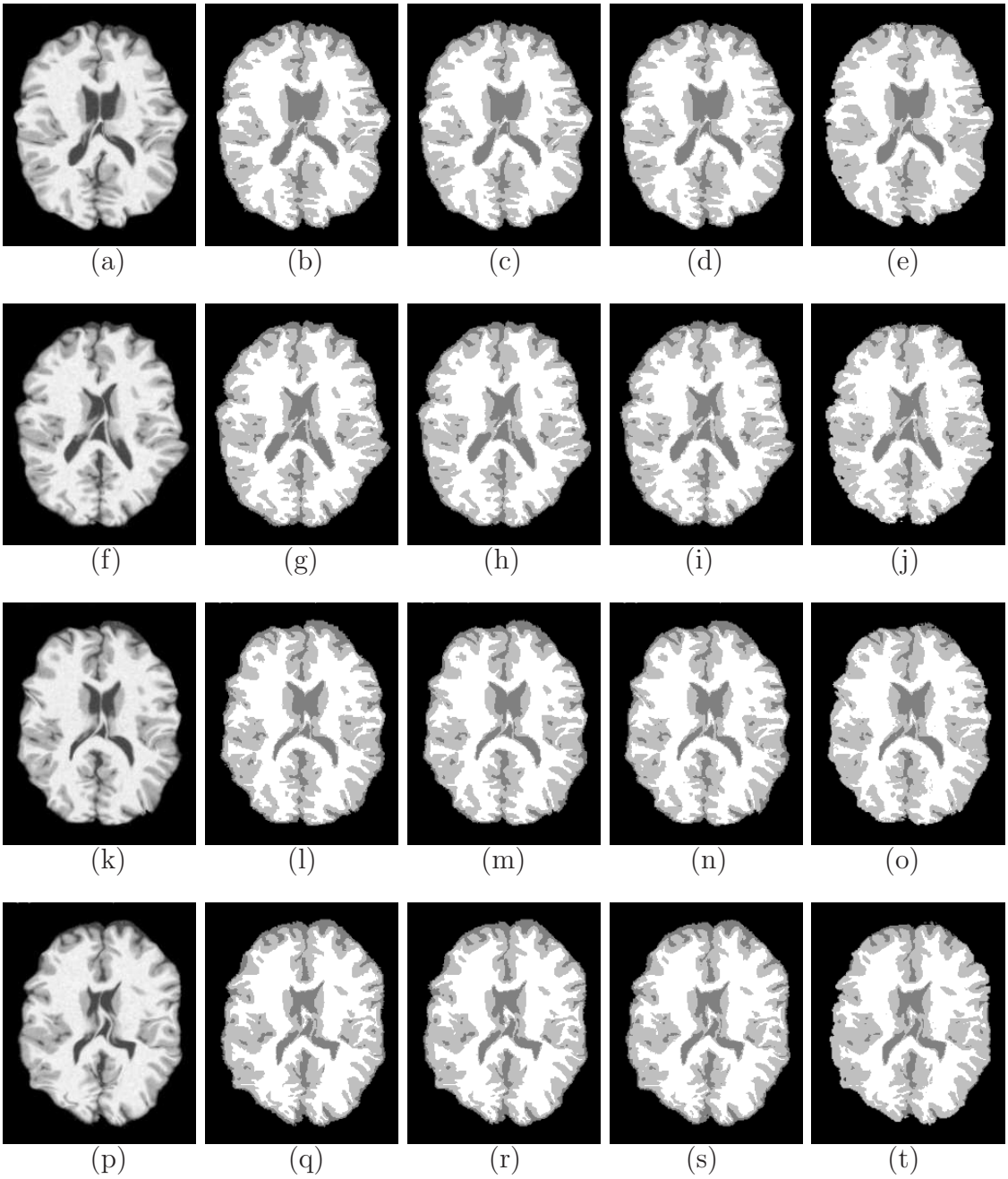


Figure 6.9: Sample images from the population (first column) and their respective segmentations: using interleaved segmentation-registration (second column), integrated segmentation-registration (third column), groupwise segmentation using the actual transformation (fourth column) and using the standard EM algorithm (final column).

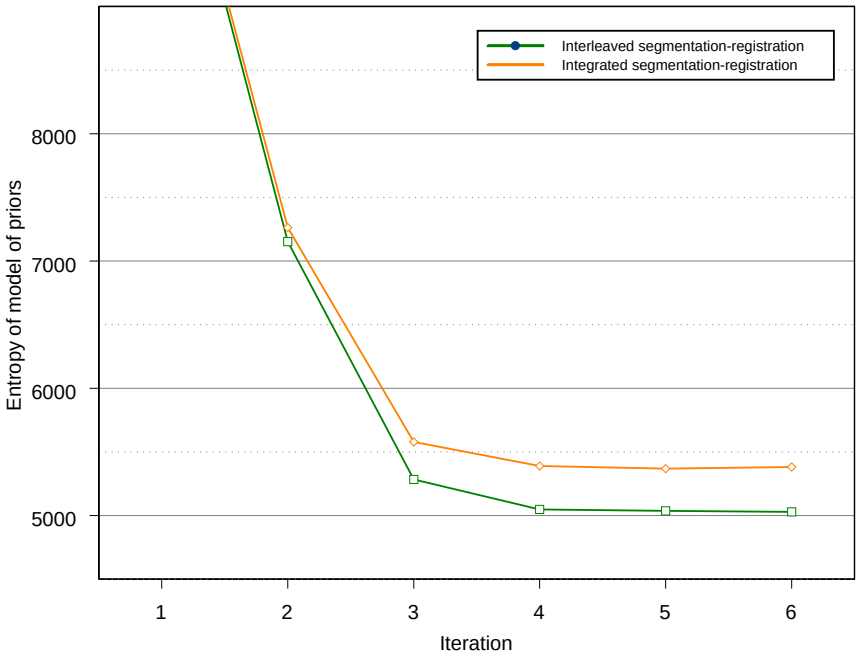


Figure 6.10: Evolution of entropies of prior models with number of iterations for interleaved and integrated groupwise segmentation-registration up to convergence.

of the mean of pairwise transformations, as described in Section 1.3 of Chapter 5. These populations consisted of all the subjects used in Chapter 5, plus an additional four others. The groupwise segmentation algorithm was run on each of these aligned populations until the entropy of the model of priors converged. In both cases, this occurred after six iterations. Samples of the segmentations obtained of the one-year-old population are shown in Figure 6.13, and of the two-years-old population in Figure 6.14. The MNI 305 priors were used as an initial input to the segmentation algorithm. After each iteration, the priors for the next iteration were recalculated to be the mean of the current soft segmentations of the population (i.e.: $\lambda = 1$). No update of the registration was used, however, this can easily be incorporated as in the 2D case. The intensity atlas of each population aligned to its average shape is shown in Figures 6.15 and 6.16, together with the WM, GM and CSF atlases and maximum probability estimate of the segmentation of the average shape at convergence. The evolution of the model of priors for the populations are shown in Figure

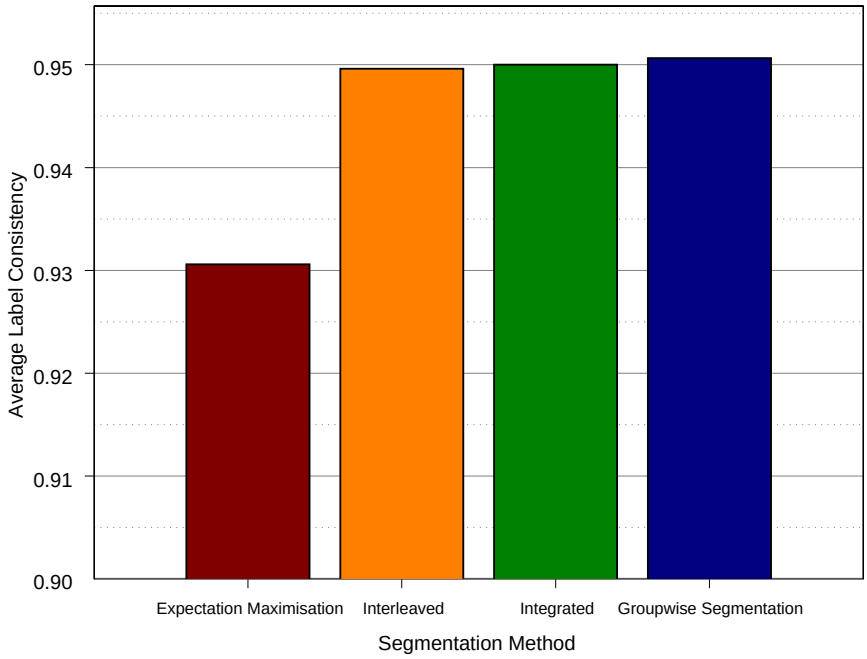


Figure 6.11: Average label consistency at end of process for each segmentation method.

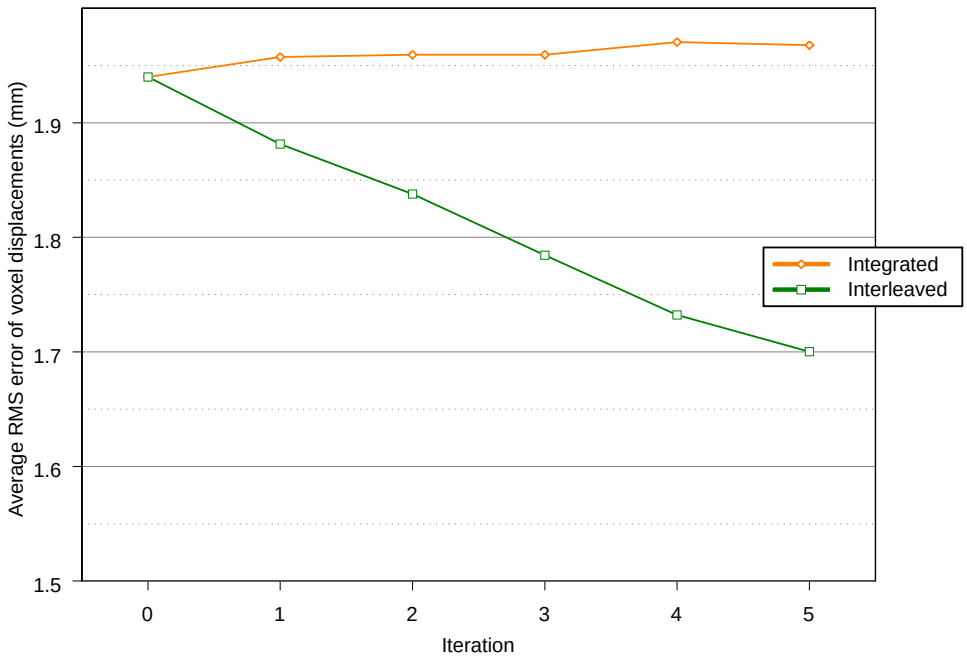


Figure 6.12: Average absolute voxel displacement error using the two groupwise segmentation-registration methods indicating registration accuracy.

6.17 and 6.18 for the one-year-old and two-year-old populations, respectively. The volume of WM and GM in each atlas has been found, and the quantitative growth of these tissues is given in Table 6.2.

	WM	GM
Volume at 1 year (cm^3)	278	536
Volume at 2 years (cm^3)	326	612
Volume change	+17%	+14%

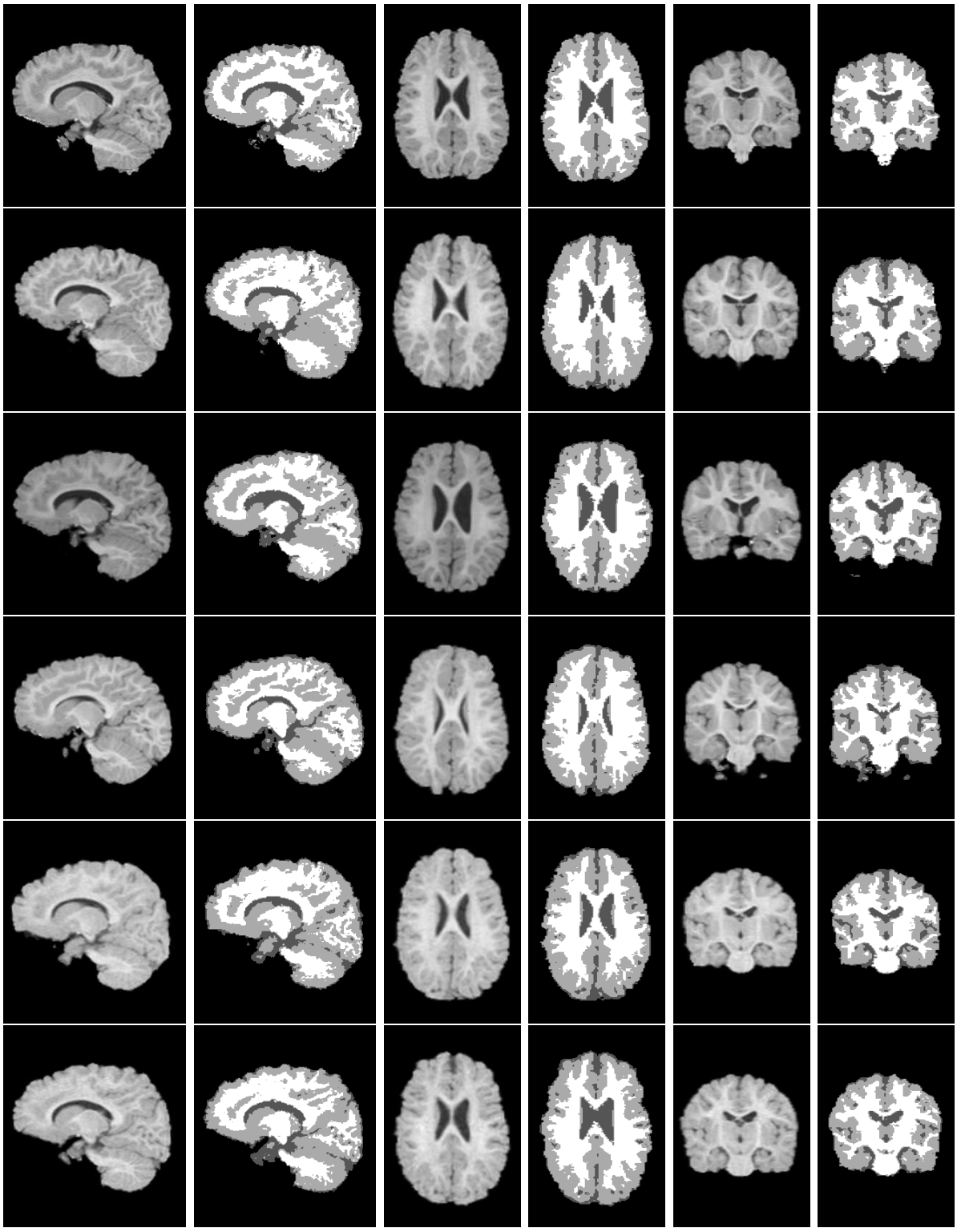
Table 6.2: Volumes of WM and GM at one- and two-years.

Using these methods has enabled the construction of both intensity and probabilistic tissue class atlases, as well as individual segmentations, of populations for which no existing standard atlas currently exists (to the best of our knowledge), and has allowed the quantification of volumetric differences between the groups.

6.7.3 Discussion

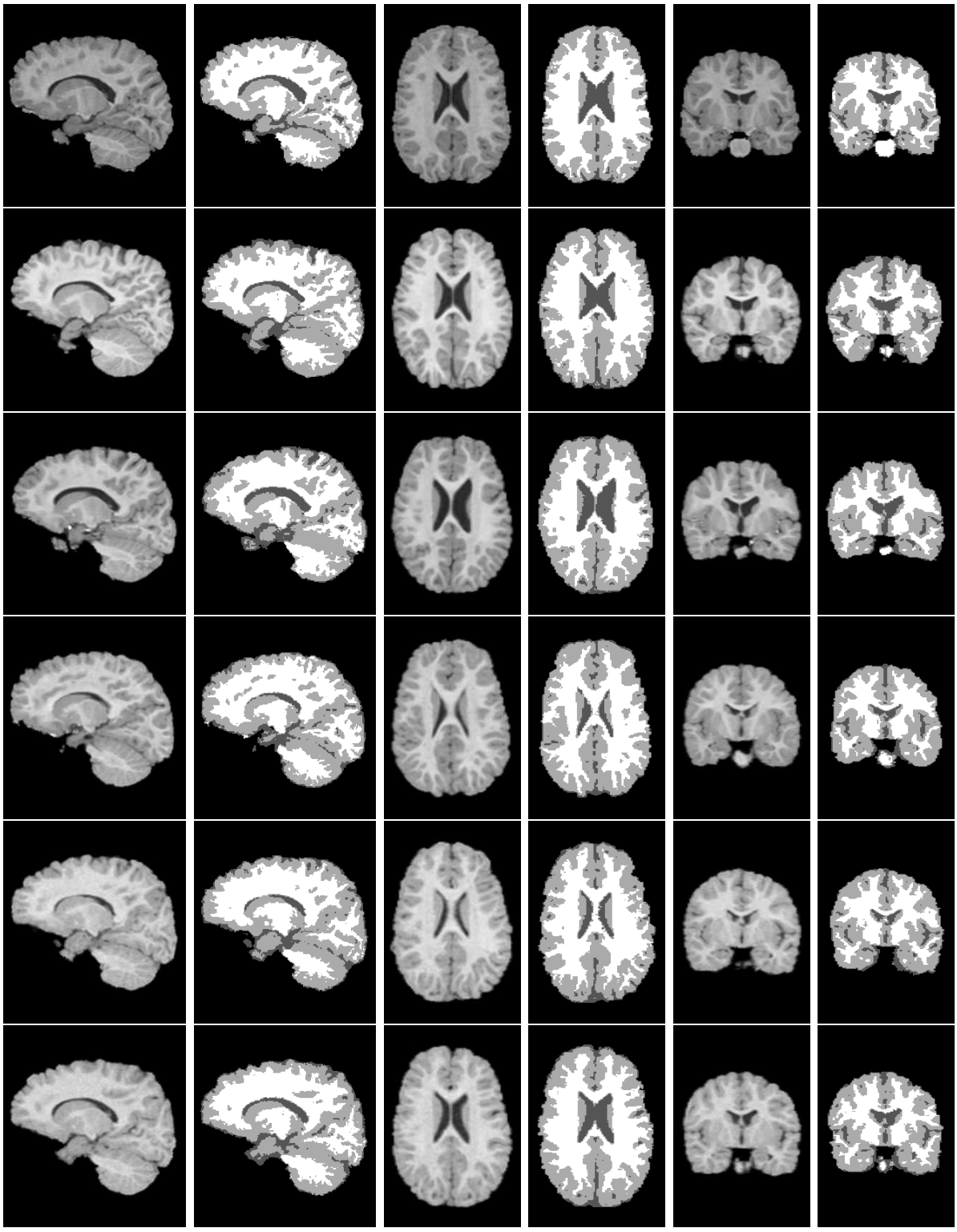
The groupwise segmentation and registration methods developed in this chapter and in Chapter 4 have been used to create average intensity, hard and soft segmentations of populations, representing their average shape. Additionally, the groupwise segmentation techniques have been used to segment the individual subjects of each population in their original image space. The combination of segmentation and registration processes allows an improvement in one process to aid the progression of the other.

On the simulated dataset, both combined segmentation-registration methods outperform the EM algorithm at segmentation of the original images, using the same initial input priors. The best segmentation occurs when the images are perfectly aligned, as obtained when the known transformations to the average space are applied. The integrated and interleaved segmentation-registration methods have a similar performance in terms of segmentation. However, only the interleaved method,



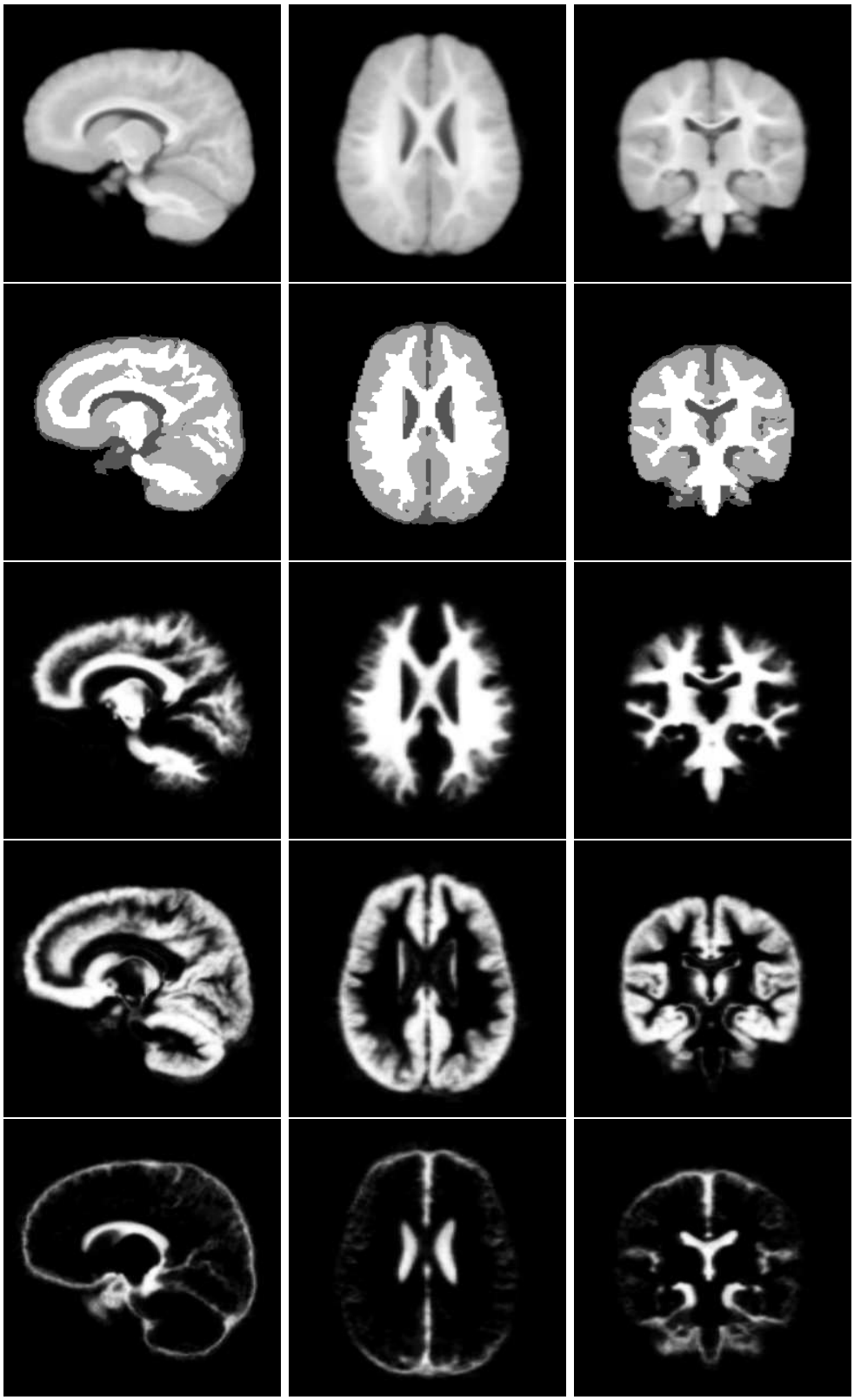
(a) (b) (c) (d) (e) (f)

Figure 6.13: Samples of segmentations of 3D MR subjects of 22 one-year-olds obtained using groupwise segmentation.



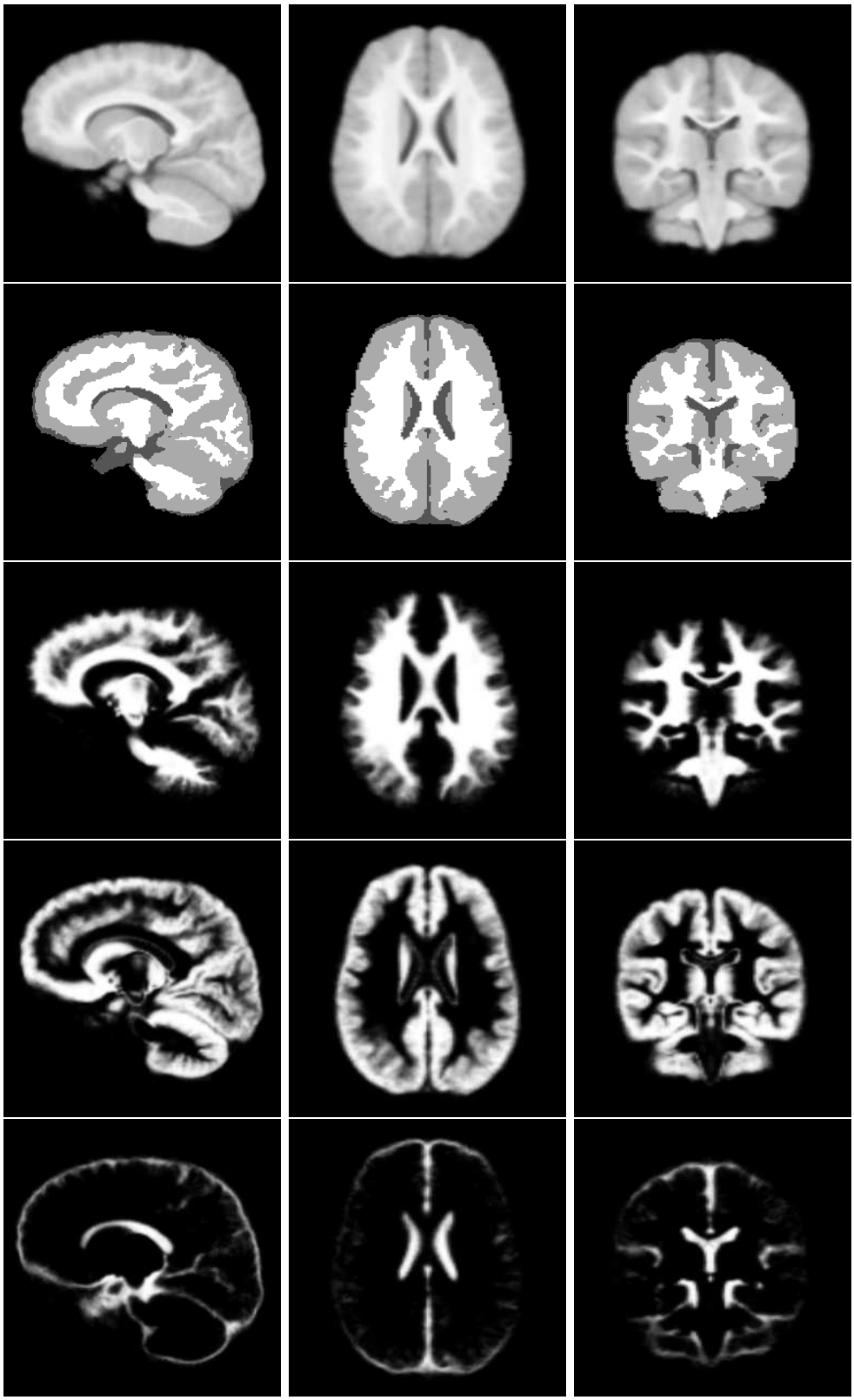
(a) (b) (c) (d) (e) (f)

Figure 6.14: Samples of segmentations of 3D MR subjects of 22 two-year-olds obtained using groupwise segmentation.



(a) (b) (c)

Figure 6.15: Average shape atlases of 22 one-year-olds. Top row: intensity atlas; second row: maximum probability estimate of segmentation of average shape; rows 3-5: WM, GM and CSF atlases in average space at convergence of groupwise segmentation algorithm.



(a) (b) (c)

Figure 6.16: Average shape atlases of 22 two-year-olds. Top row: intensity atlas; second row: maximum probability estimate of segmentation of average shape; rows 3-5: WM, GM and CSF atlases in average space at convergence of groupwise segmentation algorithm.

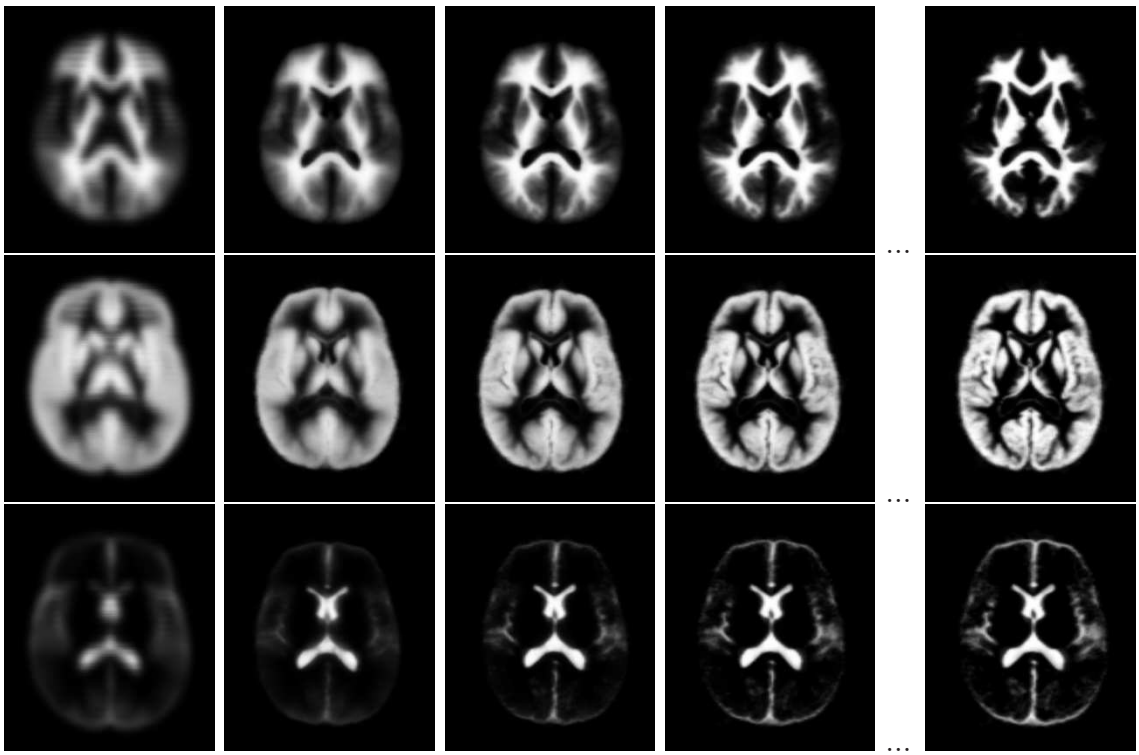


Figure 6.17: The evolution of prior models for the one-year-old population as the algorithm proceeds for WM (top row), GM (middle row) and CSF (bottom row). From left to right: the MNI priors, updated models at iterations 1-3 (b)-(d) and 6 (e) (convergence of model entropy).

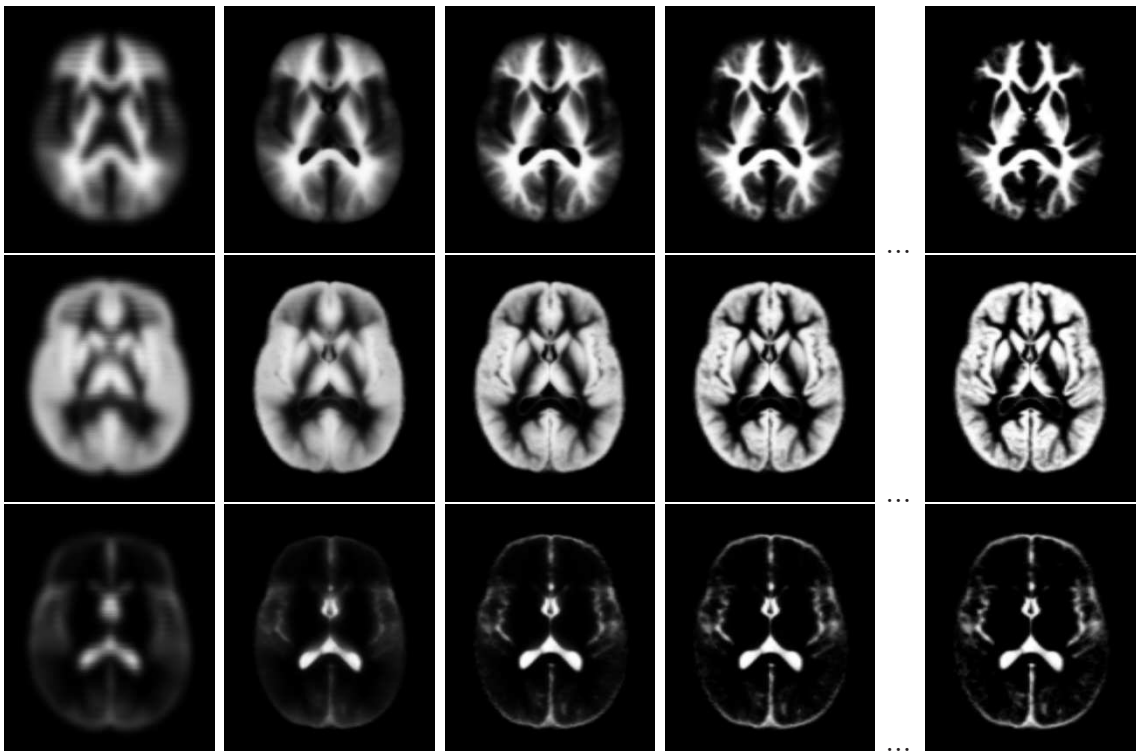


Figure 6.18: The evolution of prior models for the two-year-old population as the algorithm proceeds for WM (top row), GM (middle row) and CSF (bottom row). From left to right: the MNI priors, updated models at iterations 1-3 (b)-(d) and 6 (e) (convergence of model entropy).

which explicitly aims to maximise the registration improves the alignment of the images. The registration using this method outperforms that of any of the groupwise registration techniques described in Chapter 4 when starting with an initial approximation to the solution.

The groupwise segmentation algorithm has also been used to create average shape maximum probability and soft (probabilistic) segmentations of populations of 22 one-year-olds and 22 two-year-olds. The individual subjects of each population were also segmented. No update of the registration was used, although given the results on the simulated dataset, improving the registration is likely to lead to an improvement in the segmentations also.

The method of segmenting individual images in an average co-ordinate system makes sense if the transformation from the average space to the image space is known, and the inverse transformation from the average space to the image space can also be found. For the B-spline registration method, the inverse on a transformation can be calculated using a numerical scheme. However, errors in the inversion can occur if there is folding in the mesh generated by the registration. Any errors in the inversion process are likely to adversely affect the segmentation quality. Additionally, the segmentation quality is still dependent on the segmentation method used. In this chapter a simple EM algorithm was used. There are many possible extensions of the segmentation method that could increase the capability of segmenting MR images. The incorporation of image inhomogenieties (bias field correction) could be included within the Bayesian framework for segmentation, as in [191, 15, 116]. Additionally, the finite mixture model could be augmented by the inclusion of spatial information (as in [38]), through the use of Markov Random Fields (MRF), which model the brain tissue types as piecewise constant. This would use information from neighbouring voxels to estimate the probability of a given voxel being a given tissue class.

Methods to create more representative models could also be developed. Given the fact that different subjects will be registered to different extents, models specific to each image could be produced. For example, a given subject's model could be weighted to favour the probabilities given by subjects which are closer in alignment with that subject (these may not necessarily the best-registered subjects in the population). Furthermore, preference could be given to probabilities which are more certain. This could be done by weighting the effect of each model by the inverse of its entropy. This would help to alleviate the partial-voluming effects shown in Figure 6.7.

The segmentation methods developed in this chapter still require an initial estimate of the soft segmentation. In the experiments performed, the MNI 305 priors were used. However, this choice may bias the segmentation results, particularly when using the MNI 305 priors, created from segmentations of adult brain images, as the initial prior for child brain segmentation. Unbiased methods of creating initial estimates of the segmentation (such as unsupervised classifiers or deformable models described in Section 6.1.1), could therefore to be considered instead.

Chapter 7

Conclusions

This thesis has developed groupwise registration and segmentation methods for the construction of probabilistic and intensity-based atlases, representing the average shape of a population. Groupwise registration is used to determine the average shape of the population. This is taken to be the shape that requires least deformation from itself to other members of the population. By simultaneously registering all subjects in a population, the need to choose a reference subject has been eliminated from the registration procedure, thus removing any potential bias caused by this choice. Groupwise segmentation is used to determine the probabilistic segmentations of average shape atlases. Furthermore, by combining the segmentation with the registration, more accurate representations of shape and structure have been obtained.

The methods developed have been used to analyse the structure of the brain of preterm-born infants at varying time-points in their infancy. There is much clinical interest in studying this population due to the profound effect that preterm birth has on the developing brain and its significant long-term consequences. The infant brain displays vast changes in shape over the first few years of life. However, no standard intensity-based or probabilistic templates currently exist for either preterm or term-born infants, to the best of our knowledge.

The groupwise registration and segmentation techniques have been used to create anatomical atlases at term, one-year and two-years. At term, the intensity-based atlas of preterms has been compared to an atlas of term-born controls and quantitative differences between the two groups determined. Intensity-based and probabilistic atlases have additionally been constructed for the populations of one- and two-year-olds. This has enabled the growth of structures over this time period to be analysed and quantified.

7.1 Contributions

Chapter 4 developed a novel, unbiased, groupwise non-rigid registration algorithm for atlas construction. This involved the introduction of a constrained non-linear optimisation technique to constrain the sum of all deformations to be zero, and hence to constrain the atlas space to be the average of the population. Eliminating the need to choose a reference image in the registration process, eliminates the bias in the algorithm. Various metrics to assess the similarity of a group of images have been developed and evaluated. The methods have been tested on simulated 2D MR datasets and on real 3D adult MR data.

Chapter 5 applied the groupwise registration technique to construct average structural atlases of preterm and term-born neonates at term-equivalent age. This has enabled to visual comparison of the differences in size, shape and degree of myelination present in both groups. Additionally, deformation-based morphometry has been used to quantify volumetric differences between the two groups. The same methods have been applied to construct average atlases of 1- and 2-year old populations of preterm subjects and to analyse the growth of individual structures between these time-points.

Chapter 6 introduced novel methods for the segmentation of a population of images. Given an aligned population, a groupwise segmentation algorithm has been developed. Furthermore, the groupwise segmentation and groupwise registration algorithms have been combined to simultaneously segment and register a population of images. These methods have been evaluated on simulated MR data, showing that the combination of the two methods is benefits both. Average segmentations of the average shape of a population of one-year-olds and a population of two-year-olds have been created for white matter, grey matter and cerebro-spinal fluid tissue classes. Individual segmentations of all subjects of the population in their native space are also obtained.

7.2 Limitations and Future work

There are many interesting avenues to explore in extending this work, both in terms of algorithm development and in the application to neonatal image analysis.

7.2.1 Algorithm development

In constructing an average atlas of a population, the more subjects included, the more representative of the population the atlas should be. However, there are a number of limitations associated with processing very large amounts of data. Firstly, the memory requirements and processing times required may increase beyond that which can be handled on a single processor. An MPI implementation of the groupwise registration algorithm was developed in Chapter 4. However, this only distributes the processing of the algorithm and not the distribution of the actual data. Fortunately, due to the local control of B-splines, the registration technique developed is very suitable to parallelisation. Regions of images can be processed independently, and therefore stored across machines. However, as more images are used, the alignment of these images becomes more difficult due to the need to calculate the similarity of multiple images and the increased complexity of the search space. Optimisation

methods more suited to very large datasets and which are less sensitive to local optima, such as stochastic methods [221], could therefore be explored. Additionally, with increasing numbers of images, the registration of a single image becomes less important, which may result in some individual images being poorly registered.

An alternative is to try to reduce the computational complexity of the problem. One method for doing this is to use adaptive mesh refinement in the registration process [178]. Instead of uniformly subdividing the control point mesh at each level of the registration, localised measure of Mutual Information are used to determine the areas which require further deformation. Extra control points are only added in such areas. This process is repeated to adaptively refine the mesh, thus reducing the number of degrees of freedom.

The current method of creating the average atlas involves finding the arithmetic mean of the deformations. However, if outliers exist in the population, this mean could be unrepresentative. Robust statistics such as M-estimators [113] could be used instead. These aim to minimise a function corresponding to distances between the population, where very large distances are penalised. This would reduce the sensitivity of the final atlas to outliers.

When analysing growth, it is not always convenient to obtain enough intermediate scans for the growth to be small enough to be accurately captured via B-spline deformation fields. The use of registration methods that allow for larger deformations, such as geodesic flows [24], could additionally be investigated. These algorithms have the added benefit of providing a metric on the distance between the subjects being registered, and therefore could provide a metric on brain development.

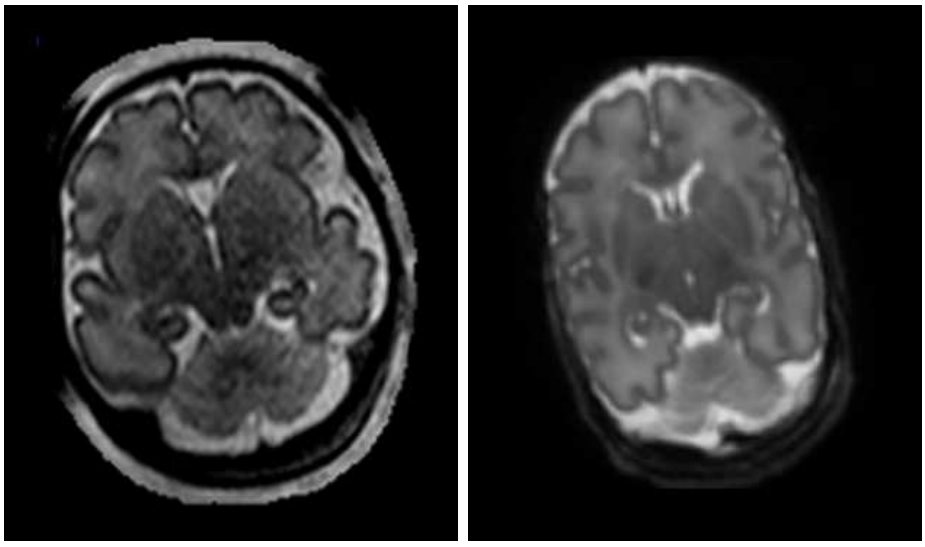
The segmentation algorithm used in this thesis is based on a simple implementation of the EM algorithm. There are many methods which have been developed

to work within an EM framework to improve the segmentation of real MR data and these could be incorporated into the algorithm described. These include methods for bias field correction [191, 15, 138], correction for partial volume effects [255, 76] and the use of neighbourhood information such as Markov random fields [38, 266] to aid segmentation. The current segmentation technique additionally still relies on the MNI 305 priors, albeit only for a single iteration. However, bias caused by this may still propagate through the segmentation as the algorithm proceeds. The use of unsupervised classifiers such as mixture models [75] could instead be investigated for the initial prior estimate.

7.2.2 Neonatal image anaylsis

The first time-point that preterm and term-born infants can be compared at is at term-equivalent age. However, by this time, large differences between the two groups are already significant. It would be beneficial to be able to understand why and when changes occur at earlier time-points, in order to potentially treat the preterm infant. This requires the comparison of the brain of preterm infants that of the fetus. However, obtaining usable fetal MR scans is difficult due to the movement of the fetus. Reconstruction of the scans is therefore needed to account for motion. This has been developed in [120]. Figure 7.1 shows a reconstructed image of a fetal brain at 33 weeks, compared with a preterm infant at the same equivalent age.

The interleaved groupwise registration and segmentation methods need to be used to analyse larger populations of data in order to obtain more accurate and more representative atlases of the populations. These can then be quantitively compared. It would, however, be necessary to obtain more intermediate scans between term and one-year due to the large amount of growth that occurs during this time. Tracking changes after two-years would also be interesting as the effects of preterm birth are known to extend into childhood and adolescence.



(a) (b)

Figure 7.1: Reconstructed fetal MR compared with preterm MR at equivalent age (33 weeks).

The shape variation of the populations could also be analysed. For example, using statistical modelling techniques such as principal components analysis (PCA) [55], which has previously been employed to detect shape differences between the hippocampus of normal subjects and those with depression [193].

7.3 Summary

This thesis has developed groupwise registration, groupwise segmentation and combined methods to create atlases representing the average shape and tissue structures of populations of neonates and infants. Intensity-based atlases have been constructed for populations of preterm and term-born neonates at term equivalent age. For populations of one- and two-year-old subjects, both intensity-based and probabilistic atlases have been constructed. These are populations for which no standard anatomical atlas currently exists, to the best of our knowledge. The atlases constructed have been used to quantify differences between preterm and term-born neonates at term equivalent age, and also to quantify the growth of structures be-

tween one- and two-years.

Publications

1. J. P. Boardman, K. Bhatia, S. Counsell, J. Allsop, O. Kapellou, M. A. Rutherford, A. D. Edwards, J. V. Hajnal, D. Rueckert. An evaluation of deformation-based morphometry in the developing human brain and detection of volumetric changes associated with preterm birth. *Sixth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'03)*, Lecture Notes in Computer Science, pages 697-704, Montreal 2003.

2. K. K. Bhatia, J. V. Hajnal, B. K. Puri, A. D. Edwards, D. Rueckert. Consistent groupwise non-rigid registration for atlas construction. *IEEE Symposium on Biomedical Imaging (ISBI'04)*, pages 908-911, Arlington 2004.

3. K. K. Bhatia, J. P. Boardman, J. V. Hajnal, A. D. Edwards, M. A. Rutherford, D. Rueckert. Groupwise atlas construction for the identification of neuroanatomical changes associated with preterm birth. *International Society of Magnetic Resonance in Medicince (ISMRM'05)*, Miami 2005.

4. P. Aljabar, K. K. Bhatia, J. V. Hajnal, J. Boardman, L. Srinivasan, M. Rutherford, L. Dyet, D. Edwards and D. Rueckert. Analysis of Growth in the Developing Brain Using Non-Rigid Registration. In *IEEE International Symposium on Biomedical Imaging (ISBI'06)*, pages 201-204, Arlington 2006.

5. P. Aljabar, K. K. Bhatia, M. Murgasova, J. V. Hajnal, J. P. Boardman, L. Srinivasan, M. A. Rutherford, L. E. Dyet, A. D. Edwards, D. Rueckert. Assessment of brain growth in early childhood using deformation based morphometry. *submitted to NeuroImage*.

6. K. K. Bhatia, J. V. Hajnal, D. Rueckert. Similarity metrics for groupwise non-rigid registration. *submitted to Tenth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'07).*

7. K. K. Bhatia, P. Aljabar, M. Murgasova, J. V. Hajnal, J. P. Boardman, L. Srinivasan, M. A. Rutherford, L. E. Dyet, A. D. Edwards, D. Rueckert. Groupwise combined segmentation and registration for atlas construction. *submitted to Tenth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'07).*

A PDF version of this thesis can be found at:
http://www.doc.ic.ac.uk/∼bhatia98/research/bhatia-thesis.pdf

Bibliography

[1] Confidential Enquiry into Maternal and Child Health (2005). Stillbirth, Neonatal and Post-natal Mortality 2000-2003, England, Wales and Northern Ireland.

[2] Statistical parametric mapping, http://www.fil.ion.ucl.ac.uk/spm.

[3] International consortium for brain mapping, http://www.loni.ucla.edu/icbm, 2002.

[4] M. Ajayi-Obe, N. Saeed, F. M. Cowan, M. A. Rutherford, and A. D. Edwards. Reduced development of cerebral cortex in extremely preterm infants. *Lancet*, 356:1162–1163, 2000.

[5] M. Alexa. Linear combination of transformations. In *SIGGRAPH'02*, pages 380–387. ACM Press, 2002.

[6] P. Aljabar, K. K. Bhatia, J.V. Hajnal, J.P. Boardman, L. Srinivasan, M. A. Rutherford, L.E. Dyet, A. D. Edwards, and D. Rueckert. Analysis of growth in the developing brain using non-rigid registration. In *IEEE International Symposium on Biomedical Imaging: Macro to Nano (ISBI'06)*, pages 201–204, 2006.

[7] P. Aljabar, K. K. Bhatia, M. Murgasova, J. V. Hajnal, J. P. Boardman, L. Srinivasan, M. A. Rutherford, L. E. Dyet, A. D. Edwards, and D. Rueckert. Assessment of brain growth in early childhood using deformation based morphometry. *NeuroImage (submitted)*.

[8] L. G. Apostolova, I. D. Dinov, R. A. Dutton, K. M. Hayashi, A. W. Toga, J. L. Cummings, and P. M. Thompson. 3D comparison of hippocampal atrophy in amnestic mild cognitive impairment and alzheimer's disease. *Brain*, 129(11):2867–2873, 2006.

[9] S. R. Arridge. Optical tomography in medical imaging. *Inverse Problems*, 15:R41–R93, 1999.

[10] S. R. Arridge and M. Schweiger. Image reconstruction in optical tomography. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 352:17–26, 1997.

[11] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Fast and simple calculus on tensors in the Log-Euclidean framework. In *Eigth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '05)*, Lecture Notes in Computer Science, pages 115–122. Springer, 2005.

[12] K. S. Arun, T. S. Huang, and S. D. Blonstein. Least square fitting of two 3-D point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):698–700, 1987.

[13] J. Ashburner and K. J. Friston. Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, 7:254–266, 1999.

[14] J. Ashburner and K. J. Friston. Voxel-based morphometry - the methods. *NeuroImage*, 11:805–821, 2000.

[15] J. Ashburner and K.J. Friston. Unified segmentation. *NeuroImage*, 26:839–851, 2005.

[16] T. Austin, A. P. Gibson, G. Branco, R. Md Yusof, S. R. Arridge, J. H. Meek, J. S. Wyatt, D. T. Delpy, and J. C. Hebden. Three dimensional optical imaging of blood volume and oxygenation in the neonatal brain. *NeuroImage*, 31:1426–1433, 2006.

[17] B. Avants and J. C. Gee. Shape averaging with diffeomorphic flows for atlas creation. In *IEEE International Symposium on Biomedical Imaging: Macro to Nano (ISBI'04)*, pages 595–598, 2004.

[18] R. Bajcsy and S. Kovacic. Multiresolution elastic matching. *Computer Vision, Graphics and Image Processing*, 46(1):1–21, 1989.

[19] R. Bansal, L. Staib, Z. Chen, A. Rangarajan, J. Knisely, R. Nath, and J. Duncan. A minimax entropy registration framework for patient setup verification in radiotherapy. *Computer Aided Surgery*, 4(6):287–304, 1999.

[20] J. L. Barron, D.J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, 1994.

[21] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Non-linear programming: theory and algorithms*. Wiley Interscience, 2006.

[22] M. S. Bazaraa and C. M. Shetty. *Nonlinear programming, theory and algorithms*. John Wiley, 1979.

[23] M. F. Beg and A. Khan. Computing an average anatomical atlas using LDDMM and geodesic shooting. In *IEEE International Symposium on Biomedical Imaging: Macro to Nano (ISBI'06)*, pages 1116–1119, 2006.

[24] M. F. Beg, M. I. Miller, A. Trouve, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*, 61(2):139–157, 2005.

[25] R. E. Bellman. *Adaptive Control Processes*. Princeton University Press, 1961.

[26] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57:289–300, 1995.

[27] A. T. Bhutta, M. A. Cleves, P. H. Casey, M. M. Cradock, and K. J. S. Anand. Cognitive and behavioral outcomes of school-aged children who were born preterm. *Journal of the Americal Medical Association*, 288:728–737, 2002.

[28] J.P. Boardman, K. Bhatia, S. Counsell, J. Allsop, O. Kapellou, M.A. Rutherford, A.D. Edwards, J.V. Hajnal, and D. Rueckert. An evaluation of deformation-based morphometry applied to the developing human brain and detection of volumetric changes associated with preterm birth. In *Sixth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '03)*, Lecture Notes in Computer Science, pages 697–704. Springer-Verlag, 2003.

[29] J.P. Boardman, S. J. Counsell, D. Rueckert, O. Kapellou, K. K. Bhatia, P. Aljabar, J. Hajnal, J. M. Allsop, M. A. Rutherford, and A. D. Edwards. Abnormal deep grey matter development following preterm birth detected using deformation-based morphometry. *Neuroimage*, 32(1):70–78, 2006.

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

[31] F. L. Bookstein. Thin-plate splines and the atlas problem for biomedical images. In *Information Processing in Medical Imaging: Proc. 12th International Conference (IPMI'91)*, number 511 in Lecture Notes in Computer Science, pages 326–342, 1991.

[32] F. L. Bookstein and W. D. K. Green. A feature space for edgels in images with landmarks. *Journal of Mathematical Imaging and Vision*, 3:231–261, 1993.

[33] N. Botting, A. Powls, R. W. Cooke, and N. Marlow. Attention deficit hyperactivity disorders and other psychiatric outcomes in very low birthweight children at 12 years. *J. Child Psychol. Psychiatry*, 38:931–941, 1997.

[34] J. W. Braams, J. Pruim, J. M. Freling, P. G. J. Nikkels, J. L. N. Roodenburg, G. Boering, W. Vaalburg, and A. Vermey. Detection of lymph node metastases of squamous-cell cancer of the head and neck with FDG-PET and MRI. *Journal of Nuclear Medicine*, 36:211–216, 1995.

[35] J. E. Brazy, V. Darrell, M. D. Lewis, M. H. Mitnick, and F. F. Jobsis. Non-invasive monitoring of cerebral oxygenation in preterm infants: preliminary observations. *Pediatrics*, 75:217–225, 1985.

[36] M. Bro-Nielsen and C. Gramkow. Fast fluid registration of medical images. In *Proc. Visualization in Biomedical Computing*, number 1131 in Lecture Notes in Computer Science, pages 267–276, 1996.

[37] B. Chance, E. Anday, S. Nioka, S. Zhou, L. Hong, K. Worden, C. Li, T. Murray, Y. Ovetsky, D. Pidikiti, and R. Thomas. A novel method for fast imaging of brain function, non-invasively, with light. *Optics Express*, 2:411–423, 1998.

[38] X. Chen, M. Brady, and D. Rueckert. Simultaneous segmentation and registration of medical image. In *Seventh Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '04)*, Lecture Notes in Computer Science, pages 663–670. Springer-Verlag, 2004.

[39] G. E. Christensen. *Deformable shape models for anatomy*. PhD thesis, Washington University, USA, 1994.

[40] G. E. Christensen and H. J. Johnson. Consistent image registration. *IEEE Transactions on Medical Imaging*, 20(7):568–582, 2001.

[41] G. E. Christensen and H. J. Johnson. Invertibiity and transitivity analysis for nonrigid image registration. *Journal of electronic imaging*, 12:106–117, 2003.

[42] G. E. Christensen, H. J. Johnson, and M. W. Vannier. Synthesizing average 3D anatomical shapes. *NeuroImage*, 32(1):146–158, 2006.

[43] G. E. Christensen, R. D. Rabbitt, and M. I. Miller. Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing*, 5(10):1435–1447, 1996.

[44] A. C. S. Chung, W. M. Wells III, A. Norbash, and W. E. L. Grimson. Multimodal image registration by minimising Kullback-Leibler distance. In *Fifth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '02)*, Lecture Notes in Computer Science, pages 525–532. Springer-Verlag, 2002.

[45] M. K. Chung, K. J. Worsley, T. Paus, C. Cherif, D. L. Collins, J. N. Giedd, J. L. Rapoport, and A. C. Evans. A unified statistical approach to deformation-based morphometry. *NeuroImage*, 14:595–606, 2001.

[46] C. A. Cocosco. Automatic generation of training data for brain tissue classification from MRI. Master's thesis, McGill University, Montreal, 2002.

[47] J. Cohen. *Statistical analysis for the behavioural sciences*. Academic Press, 1977.

[48] G. B. Coleman and H. C. Andrews. Image segmentation by clustering. *Proceedings of the IEEE*, 5:773– 785, 1979.

[49] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P Suetens, and G. Marchal. Automated multimodality image registration using information theory. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Information Processing in Medical Imaging*, pages 263–274, Dordrecht, 1995. Kluwer Academic Publishers.

[50] D. L. Collins and A. C. Evans. Animal: Validation and applications of nonlinear registration-based segmentation. *Intern. J. Pattern Recognit. Artif. Intell.*, 11:1271–1294, 1997.

[51] D.L. Collins, A.P. Zijdenbos, V. Kollokian, J.G. Sled, N.J. Kabani, C.J. Holmes, and A.C. Evans. Design and construction of a realistic digital brain phantom. *IEEE Transactions on Medical Imaging*, 17(3):463–468, 1998.

[52] T. Cootes, C. Twining, V. Petrovic, R. Schestowitz, and C.J. Taylor. Groupwise construction of appearance models using piece-wise affine deformations. In *Proc. 16th British Machine Vision Conference (BMVC'05)*, pages 879–888, 2005.

[53] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *Proc. 5th European Conference on Computer Vision (ECCV'98)*, volume 2, pages 484–498, 1998.

[54] T. F. Cootes, S. Marsland, C. J. Twining, K. Smith, and C. J. Taylor. Groupwise diffeomorphic non-rigid registration for automatic model building. In *Proc. 8th European Conference on Computer Vision (ECCV'04)*, pages 316–327, 2004.

[55] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active Shape Models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[56] S. J. Counsell, L. E. Dyet, D. J. Larkman, R. G. Nunes, J. P. Boardman, J. M. Allsop, J. Fitzpatrick, L. Srinivasan, F. M. Cowan, J. V. Hajnal, M. A. Rutherford, and A. D. Edwards. Thalamo-cortical connectivity in children born preterm mapped using probabilistic magnetic resonance tractography. *NeuroImage*, 34:896–904, 2007.

[57] S. J. Counsell, E. F. Maalouf, A. M. Fletcher, P. Duggan, M. Battin, H. J. Lewis, A. H. Herlihy, A. D. Edwards, G. M. Bydder, and M. A. Rutherford. MR imaging assessment of myelination in the very preterm brain. *American Journal of Neuroradiology*, 23:872–881, 2002.

[58] T. M. Cover and J. A. Thomas. *Elements of information theory.* John Wiley, 1991.

[59] W. R. Crum, O. Camara, D. Rueckert, K. K. Bhatia, M. Jenkinson, and D. L. G. Hill. Generalised overlap measures for assessment of pairwise and groupwise image registration and segmentation. In *Eigth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '05)*, number 3749 in Lecture Notes in Computer Science, pages 99–106. Springer, 2005.

[60] W. R. Crum, L. D. Griffin, D. L. G. Hill, and D. J. Hawkes. Zen and the art of medical image registration: correspondence, homology, and quality. *Neuroimage*, 20:1425–1437, 2003.

[61] J. G. Csernansky, S. Joshi, L. Wang, J. W. Haller, M Gado, J. P. Miller, U. Grenander, and M. I. Miller. Hippocampal morphometry in schizophrenia by high-dimensional brain mapping. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 95, pages 11406–11411, 1998.

[62] E. D'Agostino, F. Maes, D. Vandermeulen, and P. Suetens. Non-rigid atlas-to-image registration by minimization of class-conditional image entropy. In *Seventh Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '04)*, Lecture Notes in Computer Science, pages 745–753. Springer-Verlag, 2004.

[63] C. Davatzikos. Spatial transformation and registration of brain images using elastically deformed models. *Computer Vision and Image Understanding*, 66(2):207–222, 1997.

[64] C. Davatzikos, M. Vaillant, S. M. Resnick, J. L. Prince, S. Letovsky, and R. N. Bryan. A computerized approach for morphological analysis of the corpus callosum. *Journal of Computer Assisted Tomography*, 20(1):88–97, 1996.

[65] C. A. Davatzikos, J. L. Prince, and R. N. Bryan. Image registration based on boundary mapping. *IEEE Transactions on Medical Imaging*, 15(1):112–115, 1995.

[66] M. H. Davis, A. Khotanzad, D. P. Flaming, and S. E. Harms. A physics-based coordinate transformation for 3D image matching. *IEEE Transactions on Medical Imaging*, 16(3):317–328, 1997.

[67] L. S. de Vries, K. J. Rademaker, F. Groenendaal, P. Eken, I. C. van Haastert, W. P. Vandertop, R. Gooskens, and L. C. Meiners. Correlation between neonatal cranial ultrasound, MRI in infancy and neurodevelopmental outcome in infants with a large intraventricular haemorrhage with or without unilateral parenchymal involvement. *Neuropediatrics*, 29:180–188, 1998.

[68] J. Declerck, J. Feldmar, M. L. Goris, and F. Betting. Automatic registration and alignment on a template of cardiac stress and rest reoriented SPECT images. *IEEE Transactions on Medical Imaging*, 6:727–737, 1997.

[69] F. Dellaert. The expectation maximization algorithm. *Technical Report number GIT-GVU-02-20*, 2002.

[70] A. P. Dempster, N. M. Laird, and D. B.Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal Royal Statistical Society A*, 39:1–38, 1977.

[71] P.-F. D'Haese, V. Duay, T. E. Merchant, B. Macq, and B. M. Dawant. Atlas-based segmentation of the brain for 3-dimensional treatment planning in children with infratentorial ependymoma. In *Sixth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '03)*, Lecture Notes in Computer Science, pages 627–634. Springer-Verlag, 2003.

[72] L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[73] L. W. Doyle. Outcome at 5 years of age of children 23 to 27 weeks' gestation: refining the prognosis. *Pediatrics*, 108:134–141, 2001.

[74] J. Duchon. Interpolation des fonctions de deux variables suivant les principes de la flexion des plaques minces. *RAIRO Analyse Numerique*, 10:5–12, 1976.

[75] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley Interscience, 2001.

[76] G. Dugas-Phocion, M. A. Gonzlez Ballester, G. Malandain, C. Lebrun, and N. Ayache. Improved EM-based tissue segmentation and partial volume effect quantification in multi-sequence brain MRI. In *Seventh Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '04)*, Lecture Notes in Computer Science, pages 26–33. Springer-Verlag, 2004.

[77] M. Rutherford (Ed.). *MRI of the neonatal brain*. W. B. Saunders, 2002.

[78] A. D. Edwards, J. S. Wyatt, C. Richardson, D. T. Delpy, M. Cope, and E. O. R. Reynolds. Cotside measurement of cerebral blood flow in ill newborn infants by near infrared spectroscopy. *Lancet*, 2:770–771, 1988.

[79] H. C. Emsley, S. P. Wardle, D. G. Sims, M. L. Chiswick, and S. W. D'Souza. Increased survival and deteriorating developmental outcome in 23 to 25 week old gestation infant, 1990-4 compared with 1984-9. *Arch. Dis. Child. Fetal Neonatal Ed.*, 78:F99–104, 1998.

[80] A. C. Evans, D. L. Collins, S. R. Mills, E. D. Brown, R. L. Kelly, and T. M. Peters. 3D statistical neuroanatomical models from 305 MRI volumes. In *IEEE Nuclear Science Symposium and Medical Imaging Conference*, pages 1813–1817, 1993.

[81] J. Feldmar, J. Declerck, G. Malandain, and N. Ayache. Extension of the ICP algorithm to non-rigid intensity-based registration of 3D volumes. *Computer Vision and Image Understanding*, 66(2):193–206, 1997.

[82] R. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society, A*, 222:309–368, 1922.

[83] W. M. Fitch. Homology: a personal view on some of the problems. *Trends in Genetics*, 16:227–231, 2000.

[84] J. M. Fitzpatrick, D. L. G. Hill, and C. R. Maurer, Jr. Image registration. In Milan Sonka and J. Michael Fitzpatrick, editors, *Handbook of Medical Imaging*, volume 2, pages 447–513. SPIE Press, 2000.

[85] E. Fix and J. L. Hodges. Discriminatory analysis. nonparametric discrimination: Consistency properties. *Technical report, USAF School of Aviation and Medicine, Randolph Field*, 1951.

[86] N. C. Fox, W. R. Crum, R. I. Scahill, J. M. Stevens, J. C. Janssen, and M. N. Rossor. Imaging of onset and progression of alzheimer's disease with voxel-compression mapping of serial magnetic resonance images. *Lancet*, 358:201–205, 2001.

[87] N. C. Fox, E. K. Warrington, P. A. Freeborough, P. Hartikainen, A. M. Kennedy, J. M. Stephens, and M. N. Rossor. Presymptomatic hippocampal atrophy in alzheimer's disease : A longitudinal MRI study. *Brain*, 119(6):2001–2007, 1996.

[88] R. Gan, J. Wue, A. C. S. Chung, S. C. H. Yu, and W. M. Wells III. Multiresolution image registration based on kullback-leibler distance. In *Seventh Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '04)*, Lecture Notes in Computer Science, pages 599–606. Springer-Verlag, 2004.

[89] A. J. Gandhe, D. L. G. Hill, C. Studholme, D. J. Hawkes, C. F. Ruff, T. C. S. Cox, M. J. Gleeson, and A. J. Strong. Combined and three-dimensional ren-

dered multimodal data for planning cranial base surgery: a prospective evaluation. *Neurosurgery*, 35:463–470, 1994.

[90] F. Gardner, A. Johnson, P. Yudkin, U. Bowler, C. Hockley, L. Mutch, and U. Wariyar. Behavioural and emotional adjustment of teenagers in mainstream school who were born before 29 weeks gestation. *Pediatrics*, 114:676–682, 2004.

[91] J. C. Gee, C. Barillot, L. Le Briquer, D. R. Haynor, and R. Bajcsy. Matching structural images of the human brain using statistical and geometrical features. In *Proc. 3rd International Conference Visualization in Biomedical Computing (VBC'94)*, pages 191–204, Rochester, MN, 1994.

[92] J. N. Geidd, J. W. Snell, N. Lange, J. C. Rajapakse, D. Kaysen, A. C. Vaituzis, Y. C. Vauss, S. D. Hamburger, P. L. Kozuch, and J. L. Rapoport. Quantitative magnetic resonance imaging of human brain development: Ages 4-18. *Cerebral Cortex*, 6:551–559, 1996.

[93] X. Geng, D. Kumar, and G. E. Christensen. Transitive inverse-consistent manifold registration. In *Information Processing in Medical Imaging: Proc. 19th International Conference (IPMI'05)*, pages 468–479, 2005.

[94] C. R. Genovese, N. A. Lazar, and T. Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, 15(4):870–878, 2002.

[95] J. H. Gilmore, W. Lin, M. W. Prastawa, C. B. Looney, Y. S. K. Vetsa, R. C. Knickmeyer, D. D. Evans, J. K. Smith, R. M. Hamer, J. A. Lieberman, and G. Gerig. Regional gray matter growth, sexual dimorphism, and cerebral asymmetry in the neonatal brain. *Journal of Neuroscience*, 27(6):1255–1260, 2007.

[96] A. Goshtasby. Registration of images with geometric distortions. *IEEE Transactions on Geoscience and Remote Sensing*, 23(1):292–304, 1988.

[97] A. Gray and A. Moore. Rapid evaluation of multiple density models. In *Artificial Iintelligence and Statistics*, 2003.

[98] G. Grenander and M. I. Miller. Computational anatomy: An emerging discipline. *Quarterly of Applied Mathematics*, 56(4):617–694, 1998.

[99] W. Gropp, E. Lusk, and A. Skjellum. *Using MPI, portable parallel programming with the message-passing interface*. The MIT Press, 1999.

[100] A. Guimond, J. Meunier, and J-P. Thirion. Average brain models: a convergence study. *Computer vision and image understanding*, 77:192–210, 2000.

[101] M. Hack, D. J. Flannery, M. Schluchter, L. Cartar, E. Borawski, and N. Klein. Outcomes in young adulthood for very-low-birth-weight-infants. *New England Journal of Medicine*, 346:149–157, 2002.

[102] J. V. Hajnal, D. L. G. Hill, and D. J. Hawkes. *Medical Image Registration*. CRC Press, 2001.

[103] A. Hammers, R. Allom, M. J. Koepp, S. L. Free, R. Myers, L. Lemieux, T. N. Mitchell, D. J. Brooks, and J. S. Duncan. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Human Brain Mapping*, 19:224–247, 2003.

[104] J. C. Hebden, A. Gibson, R. Md Yusof, N. Everdell, E. M. C. Hillman, D. T. Delpy, S. R. Arridge, T. Austin, J. H. Meek, and J. S. Wyatt. Three-dimensional optical tomography of the premature infant brain. *Physics in Medicine and Biology*, 47:4155–4166, 2002.

[105] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage*, 33:115–126, 2006.

[106] D. L. G. Hill, J. V. Hajnal, S. M. Smith, T. Hartkens, and K. McLeish. A dynamic brain atlas. In *Fifth Int. Conf. on Medical Image Computing and*

Computer-Assisted Intervention (MICCAI '02), number 2488 in Lecture Notes in Computer Science, pages 532–539, Tokyo, Japan, 2002.

[107] D. L. G. Hill, D. J. Hawkes, J. E. Crossman, M. J. Gleeson, T. C. S. Cox, E. E. C. M. L. Bracey, A. J. Strong, and P. Graves. Registration of MR and CT images for skull base surgery using point-like anatomical features. *British Journal of Radiology*, 64:1030–1035, 1991.

[108] D. L. G. Hill, D. J. Hawkes, A. D. Castellano Smith, M. J. Gleeson, T. C. S. Cox, A. J. Strong, W.-L. Wong, C. F. Ruff, N. D. Kitchen, D. G. T. Thomas, J. E. Crossman, C. Studholme, A. J. Gandhe, S. E. M. Green, and G. P. Robinson. Accurate frameless registration of MR and CT images of the head: Applications in surgery and radiotherapy planning. *Radiology*, 191:447–454, 1994.

[109] E. T. M. Hille, A. L. den Ouden, S. Saigal, D. Wolke, M. Lambert, A. Whitaker, J. A. Pinto-Mari, L. Hoult, R. Meyer, J. F. Feldman, S. P. Verloove-Vanhorick, and N. Paneth. Preterm delivery. *The Lancet*, 360:1489–1497, 2002.

[110] S. R. Hintz, D. A. Benaron, A. M. Siegal, A. Zourabian, D. K. Stevenson, and D. A. Boas. Bedside functional imaging of the premature infant brain during passive motor activation. In *Journal of Perinatal Medicine*, volume 29, pages 335–343, 2001.

[111] S. R. Hintz, D. A. Benaron, J. P. van Houten, J. L. Duckworth, F. W. H. Liu, S. D. Spilman, D. K. Stevenson, and W.-F. Cheong. Stationary headband for clinical time-of-flight optical imaging at the bedside. *Photochemistry and Photobiology*, 68:361369, 1998.

[112] S. R. Hintz, W.-F. Cheong, J. P. van Houten, D. K. Stevenson, and D. A. Benaron. Bedside imaging of intracranial hemorrhage in the neonate using light:

comparison with ultrasound, computed tomography, and magnetic resonance imaging. *Pediatric Research*, 45:54–59, 1999.

[113] D. C. Hoaglin, F. Mosteller, and J. W. Tukey. *Understanding Robust and Exploratory Data Analysis*. Wiley Interscience, 2000.

[114] B. K. P. Horn and B. G. Schnuck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[115] P.S. Huppi, S. E. Maier, S. Peled, G. P. Zientara, P. D. Barnes, F. A. Jolesz, and J. J. Volpe. Microstructural development of human newborn cerebral white matter assessed in vivo by diffusion tensor magnetic resonance imaging. *Pediatric Research*, 44:584–590, 1998.

[116] W. M. Wells III, W. E. L. Grimson, R. Kikinis, and F. A. Jolesz. Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging*, 15(4):429–442, 1996.

[117] T. E. Inder, N. J. Anderson, C. Spencer, S. Wells, and J. J. Volpe. White matter injury in the premature infant: a comparison between serial cranial sonographic and MR findings at term. *American Journal of Neuroradiology*, 24:805–809, 2003.

[118] T. E. Inder, S. J. Wells, N. B. Mogridge, C. Spencer, and J. J. Volpe. Defining the nature of the cerebral abnormalities in the premature infant: a qualitative magnetic resonance imaging study. *Journal of Pediatrics*, 143:171–179, 2003.

[119] F. Ino, K. Ooyama, Takeuchi A., and Hagihara K. Design and implementation of parallel nonrigid image regsitration using off-the-shelf supercomputers. In *Sixth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '03)*, Lecture Notes in Computer Science, pages 327–334. Springer-Verlag, 2003.

[120] S. Jiang, H. Xue, A. Glover, M. A. Rutherford, and J. V. Hajnal. A novel approach to accurate 3D high resolution and high SNR fetal brain imaging. In *IEEE International Symposium on Biomedical Imaging: Macro to Nano (ISBI'06)*, pages 662– 665, 2006.

[121] A. Johnson, P. Townshend, P. Yudkin, D. Bull, and A. R. Wilkinson. Functional abilities at age 4 years of children born before 29 weeks of gestation. *British Medical Journal*, 306:1715–1718, 1993.

[122] H. J. Johnson and G. E. Christensen. Consistent landmark and intensity-based image registration. *IEEE Transactions on Medical Imaging*, 21(5):450–461, 2002.

[123] S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23:S151–S160, 2004.

[124] R. G. Miller Jr. *Simultaneous statistical inference (2nd edition)*. Springer-Verlag, 1981.

[125] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. J. Computer Vision*, 1(4):321–331, 1988.

[126] M. L. Kessler, S. Pitluck, P. Petti, and J. R. Castro. Integration of multimodality imaging data for radiotherapy treatment planning. *International Journal of Radiation Oncology, Biology, Physics*, 21:1653–1667, 1992.

[127] V. S. Khoo. MRI-magic radiotherapy imaging for treatment planning. *British Journal of Radiology*, 73:229–233, 1999.

[128] R. Kikinis, M. E. Shenton, D. V. Iosifescu, R. W. McCarley, P. Saiviroonporn, H. H. Hokama, A. Robatino, D. Metcalf, C. G. Wible, C. M. Portas, R. M. Donnino, and F. A. Jolesz. A digital brain atlas for surgical planning, model-driven segmentation, and teaching. *IEEE Transactions on Visualization and Computer Graphics*, 2(3):232–241, 1996.

[129] J. Kim, J. W. Fisher III, A. Tsai, C. Wible, A. S. Willsky, and W. M. Wells III. Incorporating spatial priors into an information theoretic approach for fMRI data analysis. In *Third Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '00)*, pages 62–71, 2000.

[130] N. Kovacevic, J. Chen, J. G. Sled, J. Henderson, and M. Henkelman. Deformation based representation of groupwise average and variability. In *Seventh Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '04)*, Lecture Notes in Computer Science, pages 615–622. Springer-Verlag, 2004.

[131] K. Kuban, U. Sanocka, A. Leviton, E. N. Allred, M. Pagano, O. Dammann, J. Share, D. Rosenfeld, M. Abiri, D. DiSalvo, P. Doubilet, R. Kairam, E. Kazam, M. Kirpekar, and S. Schonfeld. White matter disorders of prematurity: association with intraventricular hemorrhage and ventriculomegaly. the developmental epidemiology network. *Journal of Pediatrics*, 134:539–546, 1999.

[132] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[133] D. Kumar, X. Geng, G. E. Christensen, and M. W. Vannier. Characterizing shape differences between phantom image populations via multivariate statistical analysis of inverse consistent transformations. In *Biomedical Image Registration - WBIR*, pages 363–376, 2003.

[134] T. Kusaka, K. Kawada, K. Okubo, K. Nagano, M. Namba, H. Okada, T. Imai, K. Isobe, and S. Itoh. Noninvasive optical imaging in the visual cortex in young infants. *Human Brain Mapping*, 22:122–132, 2004.

[135] D. LeBihan, E. Breton, D. Lallemand, P. Grenier, E. Cabanis, and M. Laval-Jeantet. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology*, 161:401–407, 1986.

[136] S. Lee, G. Wolberg, K.Y. Chwa, and S. Y. Shin. Image metamorphosis with scattered feature constraints. *IEEE Transactions on Visualization and Computer Graphics*, 2(4):337–354, 1996.

[137] S. Lee, G. Wolberg, and S. Y. Shin. Scattered data interpolation with multilevel B-splines. *IEEE Transactions on Visualization and Computer Graphics*, 3(3):228–244, 1997.

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

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

[140] H. Lester and S. Arridge. A survey of hierarchical non-linear medical image registration. *Pattern Recognition*, 32(1):129–149, 1999.

[141] Y. Lin, A. Okumura, F. Hayakawa, K. Kato, T. Kuno, and K. Watanabe. Quantitative evaluation of thalami and basal ganglia in infants with periventricular leukomalacia. *Dev. Med. Child Neurol.*, 43(7):481–485, 2001.

[142] J. A. Little, D. L. G Hill, and D. J. Hawkes. Deformations incorporating rigid structures. *Computer Vision and Image Understanding*, 66(2):223–232, 1993.

[143] M. Litzkow, M. Livny, and M. Mutka. Condor - a hunter of idle workstations. In *Proceedings of the 8th International Conference of Distributed Computing Systems*, June 1988.

[144] T. Liu, A. W. Moore, and A. Gray. New algorithms for efficient high-dimensional nonparametric classification. *Journal of Machine Learning Research*, 7:11351158, 2006.

[145] P. Lorenzen, B. Davis, G. Gerig, E. Bullitt, and S. Joshi. Multi-class posterior atlas formation via unbiased Kullback-Leibler template estimation. In *Seventh Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '04)*, Lecture Notes in Computer Science, pages 95–102, Pittsburgh, USA, 2004.

[146] P. Lorenzen, M. Prastawa, B. Davis, G. Gerig, E. Bullitt, and S. Joshi. Multi-modal image set registration and atlas formation. *Medical Image Analysis*, 10(3):440–451, 2006.

[147] D.G. Luenberger. *Introduction to linear and nonlinear programming*. Addison-Wesley, 1973.

[148] E. F. Maalouf, P. J. Duggan, S. J. Counsell, M. A. Rutherford, F. Cowan, D. Azzopardi, and A. D. Edwards. Comparison of findings on cranial ultrasound and magnetic resonance imaging in preterm infants. *Pediatrics*, 107:719–727, 2001.

[149] E. F. Maalouf, P. J. Duggan, M. A. Rutherford, S. J. Counsell, A. M. Fletcher, M. Battin, F. Cowan, and A. D. Edwards. Magnetic resonance imaging of the brain in a cohort of extremely preterm infants. *Journal of Pediatrics*, 135:351–357, 1999.

[150] D. J. C. Mackay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.

[151] F. Maes, D. Vandermeulen, and P. Suetens. Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. *Medical Image Analysis*, 3(4):373–386, 1999.

[152] V. Magnotta, H. J. Bockholt, H. J. Johnson, G. E. Christensen, and N. C. Andreasen. Subcortical, cerebellar, and magnetic resonance based consistent brain image registration. *NeuroImage*, 19:233–245, 2003.

[153] J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–37, 1998.

[154] C. Malamateniou, S. J. Counsell, J. M. Allsop, J. A. Fitzpatrick, L. Srinivasan, F. M. Cowan, J. V. Hajnal, and M. A. Rutherford. The effect of preterm birth on neonatal cerebral vasculature studied with magnetic resonance angiography at 3 Tesla. *NeuroImage*, 32(3):1050–1059, 2006.

[155] R. Malladi, J. A. Sethian, and B. C. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 17:158–175, 1995.

[156] N. Marlow, B. L. Roberts, and R. W. Cooke. Motor skills in extremely low birthweight children at the age of 6 years. *Arch. Dis. Child*, 64:839–847, 1989.

[157] N. Marlow, D. Wolke, M. A. Bracewell, and M. Samara. Neurologic and development disability at six years of age after extremely preterm birth. *New England Journal of Medicine*, 352:9–19, 2005.

[158] S. Marsland, C.J. Twining, and C.J. Taylor. Groupwise non-rigid registration using polyharmonic clamped-plate splines. In *Sixth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '03)*, Lecture Notes in Computer Science, pages 771–779. Springer-Verlag, 2003.

[159] D. D. Maudgil, S. L. Free, S. M. Sisodiya, L. Lemieux, F. G. Woermann, D. R. Fish, and S. D. Shorvon. Identifying homologous anatomical landmarks on reconstructed magnetic resonance images of the human cerebral cortical surface. *Journal of Anatomy*, 193:559–571, 1998.

[160] J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, R. Zilles, K. Woods, T. Paus, G. Simpson, B. Pike, C. Holmes, L. Collins, P. Thompson, D. MacDonald, M. Iacoboni, T. Schormann, K. Amunts, N. Palomero-Gallagher, S. Geyer, L. Parson, K. Narr, N. Kabani, G. Le Goualher, J. Feidler, K. Smith,

D. Boomsma, P. H. Hulshoff, T. Cannon, R. Kawashima, and B. Mazoyer. A four dimensional probabilistic atlas of the human brain. *Journal of the American Medical Informatics Association*, 8(5), 2001.

[161] J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, R. Zilles, K. Woods, T. Paus, G. Simpson, B. Pike, C. Holmes, L. Collins, P. Thompson, D. MacDonald, M. Iacoboni, T. Schormann, K. Amunts, N. Palomero-Gallagher, S. Geyer, L. Parson, K. Narr, N. Kabani, G. Le Goualher, D. Boomsma, Cannon T., R. Kawashima, and B. Mazoyer. A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (ICBM). *The Royal Society*, 356:1293–1322, 2001.

[162] M. C. McCormic, K. Workman-Daniels, and J. Brooks-Gunn. The behavioural and emotional well-being of school-age children with different birth weights. *Pediatrics*, 97:18–25, 1996.

[163] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: A survey. *Medical Image Analysis*, 1:91–108, 1996.

[164] J. Meinguet. Multivariate interpolation at arbitrary points made simple. *Mathematik und Physik*, 30:292–304, 1979.

[165] X.-L. Meng and D. B. Rubin. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80(2):267–278, 1993.

[166] L. R. Ment, B. Vohr, W. Allan, M. Westerveld, K. H. Katz, K. C. Schneider, and R. W. Makuch. The etiology and outcome of cerebral ventriculomegaly at term in very low birth weight preterm infants. *Pediatrics*, 104(2):243–248, 1999.

[167] K. D. Merboldt, W. Hanicke, and J. Frahm. Self-diffusion NMR imaging using stimulated echoes. *Journal of Magnetic Resonance Imaging*, 64:479–486, 1985.

[168] C. L. Meyer, J. L. Boes, B. Kim, P. H. Bland, K. R. Zasadny, P. V. Kison, K. Koral, K. A. Frey, and R. L. Wahl. Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. *Medical Image Analysis*, 1(3):195–207, 1997.

[169] M.I. Miller, A. Banerjee, G.L.Christensen, S. Joshi, N. Khaneja, U. Grenander, and L. Matejic. Statistical methods in computational anatomy. *Statistical methods in medical research*, 6(3):267–299, 1997.

[170] T. Minka. Expectation-maximization as lower bound maximization, 1998.

[171] M. Murgasova, L. Dyet, A. D. Edwards, M. A. Rutherford, J.V. Hajnal, and D. Rueckert. Segmentation of brain MRI in young children. In *Eigth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '06)*, Lecture Notes in Computer Science, pages 687–694. Springer, 2006.

[172] R.M. Neal and G.E. Hinton. *Learning in Graphical Models*, chapter A view of the EM algorithm that justifies incremental, sparse, and other variants. Kluwer Academic Publishers, 2000.

[173] J. J. Neil, S. I. Shiran, R. C. McKinstry, G. L. Schefft, A. Z. Snyder, C. R. Almli, E. Akbudak, J. A. Aronovitz, J. P. Miller, B. C. Lee, and T. E. Conturo. Normal brain in human newborns: apparent diffusion coefficient and diffusion anisotropy measured by using diffusion tensor mr imaging. *Radiology*, 209:57–66, 1998.

[174] R. Nicolson, T. J. De Vito, C. N. Vidal, K. M. Hayashi, D. J. Drost, P. C. Williamson, N. Rajakumar, A. W. Toga, and P. M. Thompson. Detection and mapping of hippocampal abnormalities in autism. *Psychiatry Research*, 148(1):11–21, 2006.

[175] D. G. Nishimura. Time-of-flight MR angiography. *Magnetic Resonance in Medicine*, 14:194–201, 1990.

[176] N. Otaky, T. Paus, D. D'Avirro, D. Gutmans, D. MacDonald, Z. Caramanos, F. Tomaioulo, and A. C. Evans. Volumetric analysis of the human cingulate, paracingulate, and superior rostral sulci. *Society for Neuroscience Abstracts*, 21:154, 1995.

[177] N. Paneth, R. Rudelli, W. Monte, E. Rodriguez, J. Pinto, R. Kairam, and E. Kazam. White matter necrosis in very low birth weight infants: neuropathologic and ultrasonographic findings in infants surviving six days or longer. *Journal of Pediatrics*, 116:975–984, 1990.

[178] H. Park, P. H. Bland, K. K. Brock, and C. R. Meyer. Adaptive registration using local information measures. *Medical Image Analysis*, 8:465–473, 2004.

[179] E. Parzen. On estimation of a probability density function and mode. *Ann. Math. Stat.*, 33:10651076, 1962.

[180] T. Paus, F. Tomaiuolo, N. Otaky, D. MacDonald, M. Petrides, J. Atlas, R. Morris, and A. C. Evans. Human cingulate and paracingulate sulci: Pattern, variability, asymmetry and probabilistic map. *Cerebral Cortex*, 6:207–214, 1996.

[181] C. A. Pelizzari, G. T. Y. Chen, D. R. Spelbring, R. R. Weichelsbaum, and C.-T. Chen. Accurate three-dimensional registration of CT, PET and/or MR images of the brain. *Journal of Computer Assisted Tomography*, 13:20–26, 1989.

[182] D. Perperidis, R. Mohiaddin, and D. Rueckert. Spatial-temporal free-form registration of cardiac mr image sequences. *Medical Image Analysis*, 9:441–456, 2005.

[183] B. S. Peterson, A. W. Anderson, R. Ehrenkranz, L. H. Staib, M. Tageldin, E. Colson, J. C. Gore, C. C. Duncan, R. W. Makuch, and L.R. Ment. The etiology and outcome of cerebral ventriculomegaly at term in very low birth weight preterm infants. *Pediatrics*, 111(5):939–948, 2003.

[184] V. Petrovic, T. Cootes, C. Twining, and C. Taylor. Automatic framework for medical image registration, segmentation and modeling. In *Proc. of Medical Image Understanding and Analysis '06*, 2006.

[185] D.L. Pham and J.L. Prince. An adaptive fuzzy c-means algorithm for image segmentation in the presence of intensity inhomogeneities. *Pattern Recognition Letters*, 20(1):57–68, 1999.

[186] D.L. Pham, C. Xu, and J.L. Prince. Current methods in medical image segmentation. *Review of Biomedical Engineering*, 2:315–317, 2000.

[187] R. E. Piecuch, C. H. Leonard, B. A. Cooper, S. J. Kilpatrick, M. A. Schlueter, and A. Sola. Outcome of infants born at 24-26 weeks' gestatio: II neurodevelopmental outcome. *Obstet. Gynecol.*, 90:809–814, 1997.

[188] U. Pietrzyk, K. Herholz, A. Schuster, H.-M. v. Stockhausen, H. Lucht, and W.-D. Heiss. Clinical applications of registration and fusion of multimodality brain images from PET, SPECT, CT and MRI. *European Journal of Radiology*, 21:174–182, 1996.

[189] J. A. Pinto-Martin, A. H. Whitaker, J. F. Feldman, R. van Rossem, and N. Paneth. Relation of cranial ultrasound abnormalities in low-birthweight infants to motor or cognitive performance at ages 2, 6 and 9 years. *Dev. Med. Child Neurol.*, 41:826–833, 1999.

[190] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images:a survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004, 2003.

[191] K.M. Pohl, J. Fisher, W.E.L. Grimson, R. Kikinis, and W.M. Wells. A Bayesian model for joint segmentation and registration. *NeuroImage*, 31:228–239, 2006.

[192] M. Pokric, N. A. Thacker, M. L. J. Scott, and A. Jackson. The importance of partial voluming in multi-dimensional medical image segmentation. In *Fourth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '01)*, pages 1293–1294, 2001.

[193] J. A. Posener, L. Wang, J. L. Price, M. H. Gado, M. A. Province, M. I. Miller, C. M. Babb, and J. G. Csernansky. High-dimensional mapping of the hippocampus in depression. *American Journal of Psychiatry*, 160:83–89, 2003.

[194] M. Prastawa, J. H. Gilmore, W. Lin, and G. Gerig. Automatic segmentation of MR images of the developing newborn brain. *Medical Image Analysis*, 9:457466, 2005.

[195] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, England, 2 edition, 1992.

[196] J. A. Purdy. 3D treatment planning and intensity-modulated radiation therapy. *Oncology*, 13:155–168, 1999.

[197] J. C. Rajapakse, J. N. Geidd, C. DeCarli, J. W. Snell, A. McLaughlin, Y. C. Vauss, A. L. Krain, S. D. Hamburger, and J. L. Rapoport. A technique for single-channel MR brain tissue segmentation: Application to a pediatric sample. *Magnetic Resonance Imaging*, 14:1053–1065, 1996.

[198] A. Rao, R. Chandrashekara, G. I. Sanchez-Ortiz, R. Mohiaddin, P. Aljabar, J. V. Hanjnal, B. K. Puri, and D. Rueckert. Spatial transformation of motion and deformation fields using non-rigid registration. *IEEE Transactions on Medical Imaging*, 23(9):1065–1076, 2004.

[199] E. O. R. Reynolds, J. S. Wyatt, D. Azzopardi, D. T. Delpy, E. B. Cady, M. Cope, and S. Wray. New non-invasive methods for assessing brain oxygenation and haemodynamics. *Br. Med. Bull.*, 44:1052–1075, 1988.

[200] G. Rizzo, P. Pasquali, and M. Gilardi. Multimodal biomedical image registration: Use of a cross correlation technique. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 13, pages 219–220, 1991.

[201] C. S. Robertson, S. P. Gopinath, and B. Chance. Use of near infrared spectroscopy to identify traumatic intercranial hemotomas. *J. Biomed. Opt.*, 2:31–41, 1997.

[202] T. Rohlfing, R. Brandt, R. Manzel, and C. R. Maurer. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *Neuroimage*, 21(4):1428–1442, 2004.

[203] T. Rohlfing, R. Brandt, R. Menzel, D. B. Russakoff, and C. R. Maurer Jn. *The Handbook of Medical Image Analysis: Segmentation and Registration Models*, chapter Quo Vadis, Altas-Based Segmentation? Kluwer, 2005.

[204] J. B. Rosen. The gradient projection method for nonlinear programming. part i. linear constraints. *J. Soc. Indust. Appl. Math.*, 8(1):181–217, 1960.

[205] A. L. Rowland, T. Hartkens, M. Burns, J. V. Hajnal, D. Rueckert, and D. L. G. Hill. A grid-enabled medical image database. In *Seventh Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '04), DiDaMIC Workshop*, Lecture Notes in Computer Science. Springer-Verlag, 2004.

[206] D. Rueckert, A.F. Frangi, and J.A. Schnabel. Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. *IEEE Transactions on Medical Imaging*, 22(8):1014–1025, 2003.

[207] D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes. Non-rigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.

[208] M. A. Rutherford, F. M. Cowan, A. Y. Manzur, L. M. Dubowitz, J. M. Pennock, J. V. Hajnal, I. R. Young, and G. M. Bydder. Mr imaging of anisotropically restricted diffusion in the brain of neonates and infants. *Journal of Computer Assisted Tomography*, 15:188–198, 1991.

[209] N. Saeed, M. Ajayi-Obe, S. Counsell, J. V. Hajnal, and M. A. Rutherford. Convolution index computation of the cortex using image segmentation and contour following. In *Proc. of the 6th International Society for Magnetic Resonance in Medicine (ISMRM'98)*, page 2076, 1998.

[210] S. Saigal, L. den Ouden, D. Wolke, L. Hoult, N. Paneth, D. L. Streiner, A. Whitaker, and J. Pinto-Martin. School-age outcomes in children who were extremely low birth weight from four international population-based cohorts. *Pediatrics*, 112:943–950, 2003.

[211] C. H. Salmond, J. Ashburner, F. Vargha-Khadem, A. Connelly, D. G. Gadian, and K. J. Friston. Distributional assumptions in voxel-based morphometry. *NeuroImage*, 17:1027–1030, 2002.

[212] D. Sarrut and S. Clippe. Geometrical transformation approximation for 2D/3D intensity-based registration of portal images and CT scan,. In *Fourth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '01)*, pages 532–540, 2001.

[213] J.A. Schnabel, D. Rueckert, M. Quist, J.M. Blackall, A.D. Castellano-Smith, T. Hartkens, G.P. Penney, W.A. Hall, H. Liu, C.L. Truwit, F.A. Gerritsen, D.L.G. Hill, and D.J. Hawkes. A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. In *Medical im-*

age computing and computed-assisted intervention - MICCAI, pages 573–581, 2001.

[214] T. W. Sederberg and S. R. Parry. Free-form deformation of solid geometric models. In *13th Annual Conference on Computer Graphics and Interactive Techniques*, pages 151–160. ACM Press, 1986.

[215] J. A. Sethian and S. Osher. Fronts propagating with curvature dependent speed: algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79:12–49, 1988.

[216] J. Sijbers. *Signal and Noise Estimation From Magnetic Resonance Images*. PhD thesis, Universiteit Antwerpen, 1998.

[217] O. M. Skrinjar and H. Tagare. Symmetric, transitive, geometric deformation and intensity invariant nonrigid image registration. In *IEEE International Symposium on Biomedical Imaging: Macro to Nano (ISBI'04)*, pages 920–923, 2004.

[218] M. M. Slattery and J. J. Morrison. Preterm delivery. *The Lancet*, 360:1489–1497, 2002.

[219] J. G. Sled, A. P. Zijdenbos, D. L. Collins, and A. C. Evans. The impact of intensity non-uniformity on automated anatomical analysis of 3D MRI images. *NeuroImage*, 5:S399, 1997.

[220] J. G. Sled, A. P. Zijdenbos, and A. C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, 17:87–97, 1998.

[221] J. C. Spall. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. John Wiley, 2003.

[222] A. L. Stewart. Prediction of long-term outcome in high-risk infants: the use of objective measures of brain structure and function in the neonatal intensive care unit. *Bailliere's Clinical Obstetrics and Gynaecology*, 2:221–236, 1988.

[223] C. Studholme. Simultaneous population based image alignment for template free spatial normalisation of brain anatomy. In *Biomedical Image Registration - WBIR*, pages 81–90, 2003.

[224] C. Studholme and V. Cardenas. A template free approach to volumetric spatial normalisation of brain anatomy. *Pattern Recognition Letters*, 25(10):1191–1202, 2004.

[225] C. Studholme, V. Cardenas, N. Schuff, H. Rosen, B. Miller, and M. Weiner. Detecting spatially consistent structural differences in Alzheimer's and fronto-temporal dementia using deformation morphometry. In *Fourth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '01)*, pages 41–48, 2001.

[226] C. Studholme, D.L.G.Hill, and D.J. Hawkes. Multiresolution voxel similarity measures for mr-pet registration. In *Information Processing in Medical Imaging: Proc. 14th International Conference (IPMI'95)*, pages 287–298, 1995.

[227] C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86, 1998.

[228] C. Studholme, D.L.G. Hill, and D.J. Hawkes. Automated three-dimensional regisration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. *Medical Physics*, 24(1):71–86, 1997.

[229] P. Suetens. *Fundamentals of Medical Imaging*. Cambridge University Press, 2002.

[230] C. Svarer, K. Madsen, S.G. Hasselbalch, L.H. Pinborg, S. Haugbol, V.G. Frokjaer, S. Holm, O.B. Paulson, and G.M.Knudsen. MR-based automatic delineation of volumes of interest in human brain PET-images using probability maps. *Neuroimage*, 24(4):969–979, 2005.

[231] G. Taga, K. Asakawa, A. Maki, Y. Konishi, and H. Koizumi. Brain imaging in awake infants by near-infrared optical topography. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 100, pages 10722–10727, 2003.

[232] K. Takeda, Y. Nomura, H. Sakuma, T. Tagami, Y. Okuda, and T. Nakagawa. MR assessment of normal brain development in neonates and infants: Comparative study of T1 and diffusion-weighted images. *Journal of Computer Assisted Tomography*, 21:1–7, 1996.

[233] J. Talairach and G. Szikla. *Atlas d' Anatomie stereotaxique du Telencephalic: Etudes Anatomo-Radiologiques*. Masson and Cie, Paris, 1967.

[234] J. Talairach and P. Tournoux. *Co-Planar Stereotactic atlas of the human brain: 3-Dimensional proportional system: An approach to Cerebral imaging*. Georg Thieme Verlag, 1988.

[235] D. G. Taylor and M. C. Bushell. The spatial mapping of translational diffusion coefficients by the NMR imaging technique. *Physics in Medicine and Biology*, 30:345–349, 1985.

[236] D. Thain, T. Tannenbaum, and M. Livny. Distributed computing in practice: the condor experience. *Concurrency - Practice and Experience*, 17(2-4):323–356, 2005.

[237] P. Thevenaz and M. Unser. Optimization of mutual information for multiresolution image registration. *IEEE Transactions on Image Processing*, 9(12):2083–2099, December 2000.

[238] J-P. Thirion. Non-rigid matching using demons. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR'96)*, pages 245–251, 1996.

[239] J.-P. Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998.

[240] J.-P. Thirion and G. Calmon. Deformation analysis to detect and quantify active lesions in 3D medical image sequences. *IEEE Transactions on Medical Imaging*, 18:429–441, 1999.

[241] P. Thompson and A. W. Toga. A surface-based technique for warping three-dimensional images of the brain. *IEEE Transactions on Medical Imaging*, 15(4):402–417, 1996.

[242] P. M. Thompson, J. N. Giedd, R. P. Woods, D. MacDonald, A. C. Evans, and A. W. Toga. Growth patterns in the developing brain detected by continuum mechanical tensor maps. *Nature*, 404:190–193, 2000.

[243] P. M. Thompson, D. MacDonald, M. S. Mega, C. J. Holmes, A. C. Evans, and A. W. Toga. Detection and mapping of abnormal brain structure with a probabilistic atlas of cortical surfaces. *Journal of Computer Assisted Tomography*, 21(4):567–581, 1997.

[244] P. M. Thompson, M. S. Mega, R. P. Woods, R. E. Blanton, J. Moussai, C. I. Zoumalan, J. Aron, J. L. Cummings, and A. W. Toga. A probabilistic atlas of the human brain in Alzheimer's disease: emerging patterns of variability, assymmetry and degeneration. In *5th International Conference on Functional Mapping of the Human Brain*, 1999.

[245] P. M. Thompson and A. W. Toga. Detection, visualization and animation of abnormal anatomic structure with a deformable probabilistic brain atlas based on random vector field transformations. *Medical Image Analysis*, 1(4):271–294, 1997.

[246] P. M. Thompson and A. W. Toga. *Brain Warping*, chapter Anatomically-Driven Strategies for High-Dimensional Brain Image Warping and Pathology Detection, pages 311–336. Academic Press, 1998.

[247] P. B. Toft, H. Leth, B. Peitersen, H. C. Lou, and C. Thomsen. The apparent diffusion coefficient of water in gray and white matter of the infant brain. *Journal of Computer Assisted Tomography*, 20:1006–1011, 1996.

[248] J. Q. Trounce, D. Fagan, and M. I. Levene. Intraventricular haemorrhage and periventricular leucomalacia: ultrasound and autopsy correlation. *Arch. Dis. Child*, 61:1203–1207, 1986.

[249] J. Q. Trounce, D. Fagan, and M. I. Levene. Neonatal cranial ultrasound versus MRI and neurodevelopmental outcome at school age in children born preterm. *Arch. Dis. Child. Fetal Neonatal Ed.*, 90:489–493, 2005.

[250] J. Tucker and W. McGuire. Epidemiology of preterm birth. *British Medical Journal*, 329:675–678, 2004.

[251] C. Twining, T. Cootes, S. Marsland, R. Schestowitz, V. Petrovic, and C. Taylor. A unified information-theoretic approach to groupwise non-rigid registration and model building. In *Information Processing in Medical Imaging: Proc. 19th International Conference (IPMI'05)*, pages 1–14, 2005.

[252] C. Twining, S. Marsland, and C.J. Taylor. Groupwise non-rigid registration: the minimum description length approach. In *Proc. 15th British Machine Vision Conference (BMVC'04)*, 2004.

[253] M. van de Bor, G. L. Guit, A. M. Schreuder, J. Wondergem, and G. J. Vielvoye. Early detection of delayed myelination in preterm infants. *Pediatrics*, 84:407–411, 1989.

[254] K. van Leemput, F. Maes, F. Bello, D. Vandermeulen, A.C.F. Colchester, and P. Suetens. Automated segmentation of ms lesions from multi-channel MR im-

ages. In *Fifth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '02)*, Lecture Notes in Computer Science, pages 11–21. Springer-Verlag, 2002.

[255] K. van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. A unifying framework for partial volume segmentation of brain MR images. *IEEE Transactions on Medical Imaging*, 22(1):105–119, 2003.

[256] P. Viola and W.M. Wells III. Alignment by maximization of mutual information. In E. Grimson, S. Shafer, A. Blake, and K. Sugihara, editors, *International Conference on Computer Vision*, pages 16–23. IEEE Computer Society Press, 1995.

[257] M. P. Wand and M. C. Jones. *Kernel Smoothing*. Chapman and Hall, 1994.

[258] L. Wang, S. C. Joshi, M. I. Miller, and J. G. Cseransky. Statistical analysis of hippocampal asymmetry in schizophrenia. *NeuroImage*, 14(3):531–545, 2001.

[259] A. M. Weindling, M. J. Rochefort, S. A. Calvert, T. F. Fok, and A. Wilkinson. Development of cerebral palsy after ultrasonographic detection of periventricular cysts in the newborn. *Dev. Med. Child Neurol.*, 27:800–806, 1985.

[260] W. M. Wells, P. Viola, H. Atsumi, and S. Nakajima. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):33–51, 1996.

[261] N. S. Wood, N. Marlow, K. Costeloe, A. T. Gibson, and A. R. Wilkinson. Neurologic and delelopmental disability after extremely preterm birth. EPICure Study Group. *New England Journal of Medicine*, 343:378–384, 2000.

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

[263] R.P. Woods, S.T. Grafton, J.D.G. Watson, N.L. Sicotte, and J.C. Mazziotta. Automated image registration II: Intersubject validation of linear and non-linear models. *Journal of Computer Assisted Tomography*, 22:153–165, 1998.

[264] I. C. Wright, P. K. McGuire, J.-B. Poline, J. M. Travere, R. M. Murray, C. D. Frith, R. S. J. Frackowiak, and K. J. Friston. A voxel-based method for the statistical analysis of gray and white matter density applied to schizophrenia. *NeuroImage*, 2:244–252, 1995.

[265] C. W. Yoxall, A. M. Weindling, N. H. Dawani, and I. Peart. Measurement of cerebral venous oxyhaemoglobin saturation in children by near-infrared spectroscopy and parital jugular venous occlusion. *Pediatric Research*, 38:319–323, 1995.

[266] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20:45–57, 2001.

[267] A. Zijdenbos, A. Evans, F. Riahi, J. Sled, H.-C. Chui, and V. Kollokian. Automatic quantification of multiple sclerosis lesion volume using stereotactic space. In *Proc. 4th International Conference Visualization in Biomedical Computing (VBC'96)*, pages 439 – 448, Hamburg, Germany, 1996.

[268] R. A. Zimmerman, J. C. Haselgrove, Z. Wang, J. V. Hunter, M. C. Morriss, A. Hoydu, and L. T. Bilaniuk. Advances in pediatric neuroimaging. *Brain and Development*, 20:275–289, 1998.

[269] L. Zollei, E. Learned-Miller, E. Grimson, and W. Wells. Efficient population registration of 3D data. In *ICCV 2005, Computer Vision for Biomedical Image Applications*, 2005.