

Tracking longitudinal change using MR image data

Paul Aljabar

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

September 2007

Department of Computing
Imperial College London

Abstract

The physiology of the brain changes over time in ways that are increasingly better understood as new opportunities for *in vivo* analysis emerge. Such opportunities are made possible by the steady progress in image acquisition and analysis. Significant periods of change are exemplified by the rapid and complex growth in the early years and the subtle degeneration due to ageing or pathology. By characterising patterns of change for groups of subjects, information can be provided that may help in diagnosis and treatment or in understanding the nature of growth or the progress of a disease.

This thesis focuses mainly on the application of registration and structural segmentation techniques to serial image data in order to quantify and characterise such longitudinal changes. A framework is presented that combines growth estimates derived from longitudinal intra-subject registrations and average space atlases based on inter-subject registrations. An approach to structural segmentation is also developed in which classifiers are selected from a repository of atlases for use in label propagation and fusion.

These methods are used to characterise and quantify brain development in young children based on serial image data acquired at one and two years of age. It is shown how these methods can be used to investigate possible relationships between regional growth estimates and clinical outcome. The methods are applied to serial images acquired from a group of Alzheimer's disease patients and age-matched controls. The resulting longitudinal growth and structural segmentation data are shown to provide good discrimination between these clinical groups.

Acknowledgements

I am immensely grateful to Daniel Rueckert for his support, feedback and advice. I would also like to offer my sincere thanks Jo Hajnal and David Edwards for their incisive questions and infectious enthusiasm.

I must also acknowledge the constant encouragement and support given by my family, especially Jo, Owen and Leila. Thank you all!

Contents

1	Introduction	14
1.1	Neurological growth and degeneration	14
1.2	Acquisition of brain images	16
1.3	Analysis of brain images	18
1.3.1	Registration and segmentation	19
1.3.2	Longitudinal and cross-sectional studies	20
1.4	Contribution	23
1.5	Outline of thesis	24
2	Background	25
2.1	Introduction	25
2.2	Image registration	25
2.2.1	Voxel-based registration	26
2.2.2	Transformation models	28
2.2.3	Rigid transformations	29
2.2.4	Affine transformations	33
2.2.5	Non-rigid transformations	35
2.2.5.1	Background	36
2.2.5.2	The selected non-rigid transformation model	41
2.2.6	Multi-level free-form deformations	44
2.2.7	Combining the global and local components of a transformation	45

2.2.8	Similarity metrics	46
2.2.9	Interpolation	49
2.2.10	Optimisation	50
2.2.11	Regularisation	51
2.3	Segmentation	54
2.3.1	Manual segmentation	55
2.3.2	Automated segmentation	56
2.4	Atlas construction	59
2.5	Morphometry	64
2.5.1	Background	64
2.5.2	Morphometry for characterising longitudinal changes . . .	66
2.6	Conclusion	67
3	Registration-based methods for measuring longitudinal change	68
3.1	Registration and volume change estimation	69
3.1.1	Registration model choices	69
3.1.2	Registration parameter choices	70
3.1.3	FFD properties	73
3.1.3.1	Global and local coupling	73
3.1.3.2	Linearity of FFDs	74
3.2	Estimation of longitudinal change	75
3.3	Averaging volume change estimates	76
3.3.1	Change over a region of interest	77
3.3.2	Average change across subjects	78
3.4	Average space atlases via transformation averaging	80
3.4.1	Averaging transformations	83
3.4.1.1	Averaging global transformations	83
3.4.1.2	Averaging local transformations	85
3.4.1.3	Combining the global and local averages	86

3.4.2	Inverting transformations	87
3.5	Discussion	87
3.5.1	Tissue segmentation: Expectation maximisation.	88
3.5.2	Structural segmentation: Classifier fusion	90
3.6	Conclusion	90
4	Atlas-based segmentation, classifier fusion and classifier selection	92
4.1	Introduction	92
4.2	Measuring agreement of segmentations	93
4.2.1	Agreement on numerical data	93
4.2.2	Agreement on categories	94
4.3	Atlas-based segmentation	97
4.3.1	An example of atlas-based segmentation	99
4.4	Classifier fusion	103
4.4.1	An example of classifier fusion using the vote rule	106
4.4.2	The pattern classification context	108
4.4.3	An example of classifier fusion : multiple fusion rules	112
4.4.3.1	Comparison experiment	113
4.4.3.2	Implementation issues	114
4.5	Classifier selection	115
4.6	The motivation for classifier selection	116
4.7	Example data for classifier selection	117
4.8	Methods for classifier selection	119
4.8.1	Selection using similarity information	120
4.8.1.1	Query centric, one stage – Q1	122
4.8.1.2	Query centric, two stage – Q2	122
4.8.1.3	Reference centric, one stage – R1	123
4.8.1.4	Reference centric, two stage – R2	124

4.8.1.5	Hybrid selection and fusion – H	125
4.8.1.6	The computational burden	126
4.8.2	Selection using meta-information	127
4.9	Conclusion	128
5	An assessment of classifier selection schemes	130
5.1	Introduction	130
5.2	Data	131
5.3	Defining the ROI for similarity selection	133
5.4	A comparison of image similarity based schemes	134
5.5	Image similarity as a selection criterion	138
5.6	Detailed comparison with the fusion of random sets of classifiers .	140
5.7	The choice of similarity metric	143
5.8	Varying the number of classifiers	146
5.9	Comparing similarity based and age based selection	148
5.10	Discussion	152
5.11	Conclusion	156
6	Longitudinal change in early childhood	157
6.1	Growth in early childhood : Previous work	158
6.2	Image data	159
6.3	Experimental design	161
6.4	Individual registration results	162
6.5	Image segmentation	164
6.5.1	Expectation maximisation	164
6.5.2	Classifier fusion	165
6.6	Atlases	165
6.7	The effect of the choice of reference	170
6.8	Comparison of registration and segmentation based growth estimates	173

6.9	Regional volume changes and clinical information : Preliminary investigation	176
6.10	Discussion	179
6.11	Conclusion	182
7	Longitudinal change due to atrophy	184
7.1	Atrophy due to ageing and Alzheimer’s disease : background and previous work	185
7.2	Data	187
7.3	Mask generation	188
7.4	The effect of interpolation	193
7.5	Consistency of estimates of longitudinal change	197
7.6	Atlases	198
7.7	Group separation and classification	202
7.8	Discussion	206
7.9	Conclusion	207
8	Conclusion	208
8.1	Contribution	208
8.2	Limitations and future work	210
8.3	Summary	213
	Publications	204
	Bibliography	217
	Appendix	
A	Inverting Free-Form Deformations	242
A.1	Composition of FFDs	242
A.2	Estimating the inverse	243

Supplementary material (CD ROM)

List of Figures

1.1	T1-weighted images of a child at one and two years	19
1.2	Different types of alignment applied to serial images	21
1.3	Serial images acquired from an Alzheimer’s patient	21
1.4	Example of a structural segmentation	22
2.1	Schematic illustration of a target and source image	27
2.2	Subdivision of a free-form deformation	39
2.3	Reference space and average space atlases of the same cohort . . .	61
2.4	Schematic illustration of iteratively finding the centroid	63
3.1	The effect of the regularisation parameter, λ	72
3.2	Illustration of the effect of the similarity metric upon the smoothness of the transformation	72
3.3	Schematic illustration of the coupling of the global and local parts of a transformation	74
3.4	Schematic illustration of atlas construction	81
3.5	Flow chart illustrating the steps taken to average a set of transformations	82
3.6	Flow chart illustrating the steps taken when generating an average space atlas	82
3.7	Tissue maps obtained by expectation maximisation	89
3.8	Example of manual labels for an anatomic image	90
4.1	Example of an anatomy with a corresponding manual segmentation	98

4.2	Propagation of hippocampal labels to a query subject	100
4.3	Comparison of volumes from two classifiers: Scatter and Bland-Altman plots	101
4.4	Tanimoto and Dice overlaps for measuring agreement	102
4.5	Schematic illustration of a structural segmentation based on label propagation and fusion	105
4.6	Overlap accuracy plotted against number of fused classifiers: experimental data and a fitted model	107
4.7	Comparisons of segmentation accuracy obtained using different rules for the fusion of propagated label sets	114
4.8	Accuracy of segmentations obtained by separate fusions of all proper subsets of classifiers taken from a small repository of atlases . . .	118
4.9	Query centric classifier selection and fusion (one stage of registration)	122
4.10	Query centric classifier selection and fusion (two stages of registration)	123
4.11	Reference centric classifier selection and fusion (one stage of registration)	124
4.12	Reference centric classifier selection and fusion (two stages of registration)	125
4.13	Hybrid classifier selection and fusion	126
4.14	Classifier selection based on meta-information	128
5.1	An example dataset from the repository of manually labelled CMA images	132
5.2	The sub-cortical mask used for similarity based selection schemes	134
5.3	Accuracy comparisons for different similarity based selection schemes	136
5.4	Accuracy comparisons for different similarity based selection schemes: summary box-plots	137

5.5	Relationship between average Dice accuracy and the rank determined by similarity selection.	139
5.6	Detailed comparison of accuracy obtained using the Q1 scheme against random overlap distributions (subject 1)	141
5.7	Detailed comparison of accuracy obtained using the Q1 scheme against random overlap distributions (subject 2)	142
5.8	Detailed comparison of accuracy obtained using the Q1 scheme against random overlap distributions (subject 3)	143
5.9	Bar chart showing z-scores of the accuracy obtained for different structures using the Q1 scheme	144
5.10	The effect of the choice of similarity metric during the selection stage upon the resulting segmentation	146
5.11	The effect of fusing increasing numbers of ranked classifiers upon the segmentation accuracy of the resulting segmentation	147
5.12	A comparison of age selection and hybrid selection	150
5.13	The top ten classifiers for an 11 year-old subject using the hybrid similarity scheme.	150
5.14	The top ten classifiers for a 29 year-old subject using the hybrid similarity scheme.	151
5.15	The top ten classifiers for a 79 year-old subject using the hybrid similarity scheme.	151
6.1	Pre-processing an image: Non-uniformity correction and brain extraction	161
6.2	Schematic illustration of the images and registrations used in the framework for characterising longitudinal change	162
6.3	Images illustrating growth between one and two years of age (see accompanying CD ROM)	163
6.4	Difference images illustrating the different stages of alignment	163

6.5	Prior probability maps for EM tissue classification	165
6.6	Example of tissue labels generated by EM segmentation	166
6.7	Structural segmentation of an image based on classifier fusion . . .	167
6.8	Average space anatomical atlases of the set of childrens' images .	168
6.9	Average space growth map atlas for the cohort of children	169
6.10	Average space growth map atlas for the cohort of children overlaid onto the anatomy	169
6.11	Average space atlases of tissue densities for the set of childrens' images	170
6.12	Illustration of the average longitudinal (one to two year) deforma- tion for the cohort	171
6.13	Structural segmentation of the average space baseline image . . .	171
6.14	The effect of reference choice during the creation of average space atlases	172
6.15	The effect of reference choice during the creation of average space atlases: corresponding iso-intensity contours	173
6.16	The effect of reference choice: further corresponding iso-intensity contours	174
6.17	Segmentation-based tissue volumes at ages one and two	174
6.18	A comparison of tissue growth factors estimated by registration- and segmentation-based methods	175
6.19	Scatter plots of tissue growth factors showing outliers	177
6.20	Bland-Altman plots for the growth factor estimates produced by registration and by segmentation	177
6.21	Plot of cortical growth figures against the development quotient .	179
6.22	Plot of sub-cortical growth against gestational age	180
7.1	An example of a tissue segmentation for an Alzheimer's patient .	188
7.2	An example of a tissue segmentation for a control subject	189

7.3	An example of a structural segmentation for an Alzheimer’s patient	190
7.4	An example of a structural segmentation for a control subject . . .	190
7.5	The top-ranked classifiers for an Alzheimer’s patient	191
7.6	The top-ranked classifiers for a control subject	192
7.7	Registration-derived atrophy estimates: plotted separately accord- ing to the source image interpolation method	196
7.8	Baseline consistency of atrophy estimates: Bland-Altman plots . . .	198
7.9	Average space anatomical atlases for Alzheimer’s patients and matched controls	199
7.10	Average space Jacobian determinant atlases for Alzheimer’s pa- tients and matched controls	200
7.11	Overlays of anatomical and Jacobian determinant atlases for Alzheimer’s patients and matched controls	201
7.12	Illustration of the average longitudinal deformations for each of the Alzheimer’s and control groups	202
7.13	Box-plots of hippocampal volume distributions for AD and control subjects	203
7.14	Bar chart showing leave-one-out performance of classification of patients and controls for different experimental data	206

List of Tables

3.1	Parameters used for hierarchical FFD-based registration	71
4.1	Agreement of two raters	94
4.2	Outline of the computational costs of the different similarity-based selection schemes	127
5.1	Similarity based selection schemes: registration types, targets and their purpose	135
5.2	Detailed accuracy comparison of the Q1 and H schemes	137
5.3	Z-scores of the accuracy obtained for different structures using the Q1 scheme	144
5.4	Ages of top-ranked classifiers selected by the hybrid similarity scheme	152
6.1	Average tissue growth factors estimated by registration- and segmentation-based methods	176
6.2	Correlation coefficients of regional growth against clinical factors .	179
7.1	Estimates of age- and dementia-related atrophy: a comparison with published methods	197
7.2	Leave-one-out performance of linear discriminant classification of patients and controls	204
7.3	Combinations of data used for different linear discriminant experiments	205

7.4	Leave-one-out performance of classification of patients and controls for different experimental data	205
-----	---	-----

Chapter 1

Introduction

1.1 Neurological growth and degeneration

Like all organs of the body, the physiology of the brain changes over time and the structural complexity of the brain presents a challenge to the characterisation of this change. Changes in the brain can be rapid and dramatic, for example during fetal development and the early years [92, 120, 155], and it can be subtle and yet significant, as illustrated by atrophy due to ageing or neurodegeneration [185, 183, 62, 144, 94].

Changes can be viewed as occurring at different scales. At the microscopic scale, cells are created, destroyed and undergo biochemical changes. The process of myelination, in which individual neurons are coated in a lipid sheath, and which occurs up to the age of two years, is an example of change at the microscopic level [141].

At a macroscopic scale, there is the global change that acts in a uniform way across the brain region. For example, the heads of children undergo general growth in the early years in a way that can be approximated by a uniform global scaling.

Between these macroscopic and microscopic scales, individual structures can be observed to appear, grow or shrink over time. A good deal of brain volume-

try is carried out at this macroscopic scale, and advances in imaging techniques have allowed increasing numbers of such studies to be carried out *in-vivo*, using imaged representations of the anatomy. Changes in the brain over time, at the macroscopic level and above, are the main focus of the work presented in this thesis.

In the context of brain development in the early years, preterm birth has a significant effect on the developing brain, and infants born preterm commonly display neuropsychiatric problems during childhood [117, 178, 25]. This motivates the need to be able to characterise the patterns of typical and atypical brain development in infants and young children, something that is made increasingly feasible through the greater availability of brain images acquired during early childhood.

In the case of a condition such as Alzheimer's disease, a diagnosis is usually made based on psychological testing and the use of cognitive tests such as the mini mental state examination (MMSE [59]). Diagnoses of Alzheimer's disease can only be considered definitive if carried out by directly examining samples of the brain tissue, either obtained *post-mortem* or by brain biopsy, an invasive procedure with risk of complications [176]. It is, however, possible to obtain markers for the progression of Alzheimer's disease, using techniques based upon the acquisition and processing of serial brain images of Alzheimer's patients [96, 26, 172].

These examples illustrate the potentially useful role that can be played by the analysis of longitudinally acquired images of the brain. It can be used to inform clinical diagnosis and decisions on intervention and, by tracking disease progression, can inform studies aimed at measuring the efficacy of a particular drug or treatment.

1.2 Acquisition of brain images

There are many ways in which structural images of the brain can be acquired. Common modalities include magnetic resonance imaging, computed tomography, PET and SPECT imaging and ultrasound imaging. The work in this thesis will concentrate on the use of MR images.

Magnetic resonance imaging (MRI) uses a strong magnetic field created by a fixed magnet, a set of applied gradients generated by electromagnets and a series of radio frequency (RF) pulses to reconstruct images based on the spin properties of the nuclei of atoms such as hydrogen. The application of the RF pulses causes the alignments and phases of nucleic spins to vary in ways that allow inferences to be made about the spatial distributions of different types of molecules, for example water or contrast agents. The pulse sequences applied and the way the resulting signals are interpreted mean that MR scanners can produce a wide range of images that are adapted to the imaging of different aspects of the anatomy or physiology of the subject.

The strengths of the magnetic field in a MR scanner can be very high. Modern scanners typically use 1.5–3 Tesla magnets. In addition, acoustic noise levels can also be high, as the gradient coils are used; this necessitates the use of ear protection. Despite these considerations, MR imaging is considered to be non-invasive.

MR images aimed at identifying structure (for example ‘T1-’ or ‘T2-weighted’ images) are typically acquired at a spatial resolution of the order of 1mm and show good contrast differences between soft tissues within the brain, such as the difference between grey and white matter. MR scanners can also be used to generate images that measure the patterns of diffusion of water molecules in the brain. The technique, known as diffusion weighted imaging (DWI) [121], allows the estimation of the propensity of water to diffuse along particular directions. By measuring this propensity along a number of directions, a diffusion tensor can

be calculated that characterises the motion of water in all directions. Further processing using the diffusion tensor allows the estimation of physiological features such as the orientations of bundles of fibrous tissue, such as white matter, within the brain. Further processing can be used to identify patterns of connectivity between different parts of the brain [11].

Another specialised use of MR imaging is MR angiography (MRA), where the scanner is calibrated to highlight blood flow within the brain. This allows the vasculature of the brain to be visualised. It is possible to use a paramagnetic contrast agent during MRA acquisition which assists the identification of tissues with a relatively increased vasculature, such as tumours, or in the detection of aneurysms [16, 27]. Alternatively, an approach known as time-of-flight angiography can be used to generate images where relatively fast moving blood gives a higher signal; this allows vessels to be visualised.

A use of MR scanners known as functional MRI (fMRI) can also be used to provide maps that correspond to the level of blood oxygenation. Such blood oxygen level dependent (BOLD) signals can be regarded as being correlated with the levels of activity in different parts of the brain. By giving the subject different tasks to perform while they are being scanned, inferences can be made on which parts of the brain are associated with specific processes [97].

Computed tomography (CT) uses ionising radiation in the form of X-rays to generate volumetric representations of the anatomy. X-rays are passed through the patient in different directions and the variations in attenuation of the X-ray signal are used to calculate a (tomographic) image of the anatomy. CT has a high spatial resolution (less than 1mm is typical) and, compared with MR, is fast and inexpensive. However, CT is considered to be more invasive than MR due to the use of ionising radiation. Additionally, while CT provides excellent contrast differences between calcified tissue, such as bone, and soft tissue, its discrimination between different types of soft tissue is poor compared with that of MR imaging.

Ultrasound (US) imaging is an established and widely used method for obtaining images of many internal organs including the brain. US imaging is fast, inexpensive and can be carried out in real time at the patient's bedside. It is generally accepted as non-invasive. US is commonly used for fetal scanning, where tasks such as screening for abnormalities or assessing the fetal position can be quickly carried out. In cranial ultrasonography, the contrast allows for the ready identification of structures such as bone or the ventricles. Compared with MR, however, the contrast between different tissues is not as clear. The signal to noise ratio (SNR) is also low compared with CT and MR.

The resolution and contrast properties of structural MR images, such as T1- or T2-weighted volumes, along with their non-invasive acquisition makes them a good starting point for the tracking of longitudinal changes in the brain. The acquisition of structural MR images is relatively long, with times of 3–5 minutes being typical for T1- and T2-weighted images, and this clearly present a challenge, particularly when imaging young children, but these times are steadily being reduced.

Other modalities exist that are less well suited for identifying structural features in the brain. These include Positron emission tomography (PET) which measures the concentration of a radioactive tracer in the body and can be used to identify or detect tumours [184] and patterns of activity in the brain [30]. Optical imaging (topography or tomography) is a modality where tissues are distinguished based on their absorption and diffusion of near infra-red light. This can be useful in estimating blood oxygenation, haemodynamic changes in response to stimuli [32] or haemorrhages [7].

1.3 Analysis of brain images

Measuring longitudinal change is the main focus of this work and one way in which such change can be characterised is by analysing images of the same individual

acquired at different times. An illustration of the differences between serial scans of an individual is shown in Figure 1.1 for a child with images acquired at one and two years.

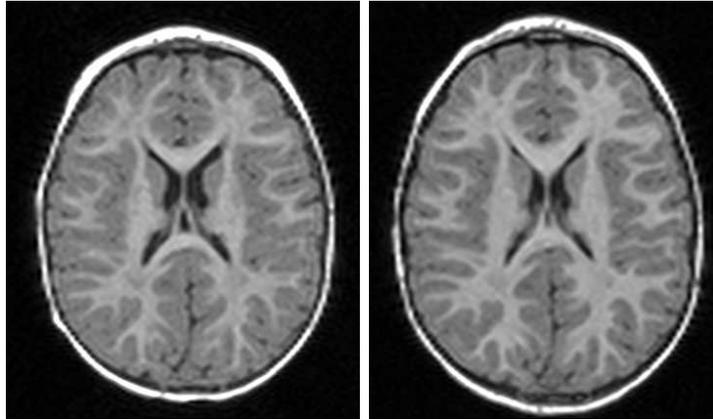


Figure 1.1: T1-weighted axial sectional images of a child acquired at one year (left) and two years old (right).

1.3.1 Registration and segmentation

One way of characterising the changes between serial scans of the head such as those in Figure 1.1 is to identify a transformation that aligns the two images. The process of finding such alignments, known as image registration, generates transformations that can be used to model the changes that take place during the interval between scans. These changes may be due to the growth of tissues or new structures, the degeneration of tissue or the growth of lesions or tumours. The process of registration provides correspondences between the images and these correspondences can be used to identify the changes between the serially acquired images and to track growth. This is illustrated in Figure 1.2 for the serial scans of the child in Figure 1.1. Along each row, the left hand image shows the one year scan and the middle image shows the two year scan after different degrees of alignment with the one year scan. The final image in each row shows the difference between the one year scan and the aligned two year scan. After rigid alignment, the difference image shows a band of bright and dark lines along

the cranial boundary. This gives an impression of the global size difference between the one and two year scans. After an affine alignment that accounts for these global changes, the cranial boundary differences have been reduced significantly but there remain differences within the image, especially at the ventricular boundaries. This is shown in the middle row of Figure 1.2. The bottom row shows how the application of a locally varying, non-rigid transformation removes most of the differences that remain after affine alignment. The longitudinal affine and non-rigid transformations can be said to have captured or encoded the structural changes that have taken place between the scans.

Another example of serial scans and the difference between them is shown for an Alzheimer’s disease patient in Figure 1.3. The serial scans in this figure have been rigidly aligned, and the most significant differences can be seen at the boundary of the ventricles, indicating their expansion between the scans, a phenomenon generally associated with atrophy of the surrounding tissues at this age.

In conjunction with the estimation of longitudinal change, the use of automated segmentation methods allows individual structures to be identified. The identified structures can be treated as ‘regions of interest’ for which specific estimates of growth or atrophy can be provided.

An example of a structural segmentation is shown in Figure 1.4 which shows a T1-weighted MR image for a subject that has been segmented into various structures. These structures were estimated by propagating segmentations from the images of other subjects and subsequently combining or fusing them.

1.3.2 Longitudinal and cross-sectional studies

Studies that use serially acquired images to estimate change over time are typically described as using longitudinal or prospective methods [28, 62, 65, 65, 72]. Such methods contrast with what are termed cross-sectional methods [20, 74, 77,

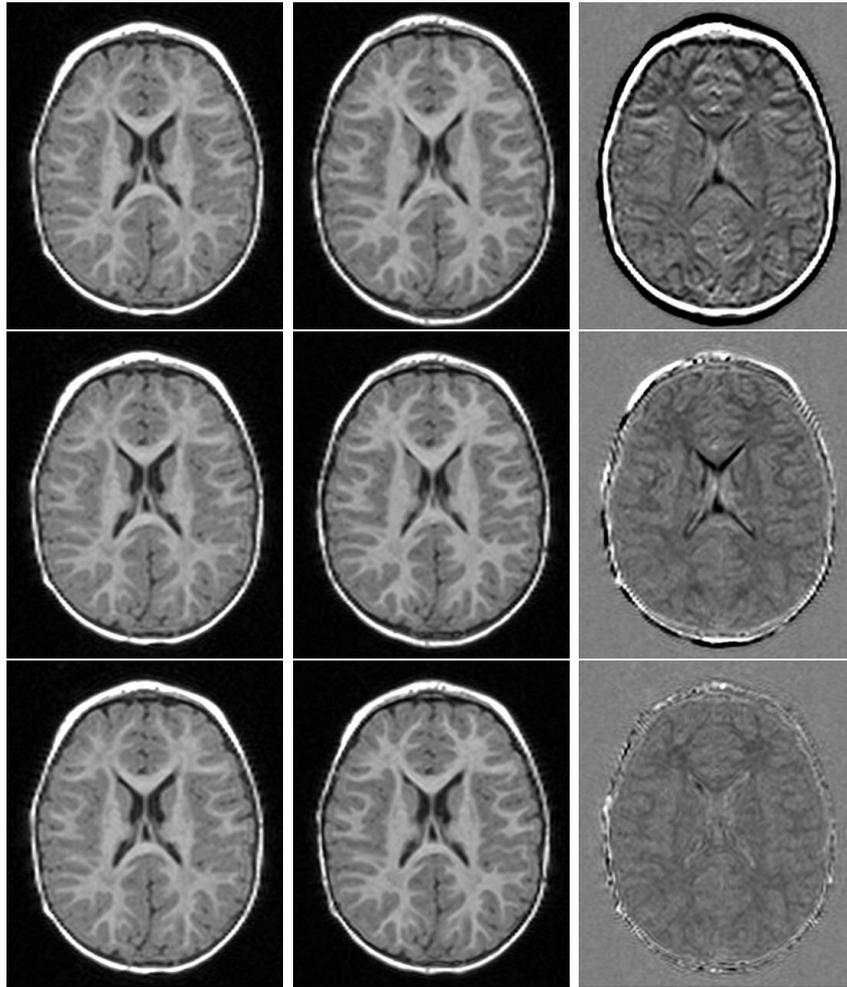


Figure 1.2: The middle column of images shows the two year scan of Figure 1.1 after varying degrees of alignment to the one year scan. Top row: After rigid alignment. Middle row: After correcting for global shape differences. Bottom row: After correcting for local differences with a non-rigid alignment. The first image in each row is the original year one scan. The last image shows the difference between the transformed year two scan and the year one scan.

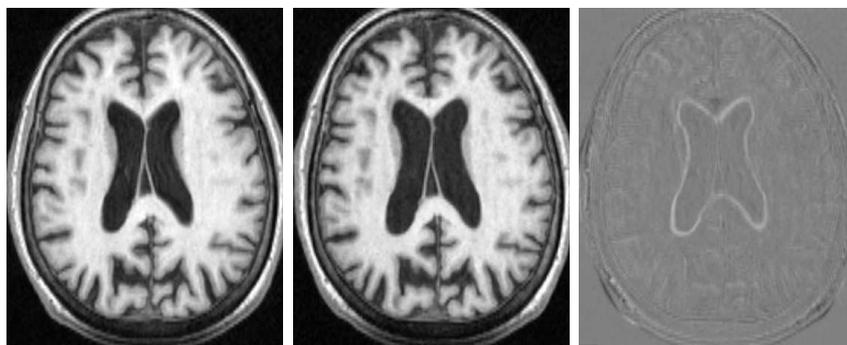


Figure 1.3: Scans acquired from a patient diagnosed with Alzheimer's disease. From left to right: A baseline scan; A follow-up scan acquired one year later and rigidly aligned with the baseline scan; The difference between the baseline and follow-up scans.

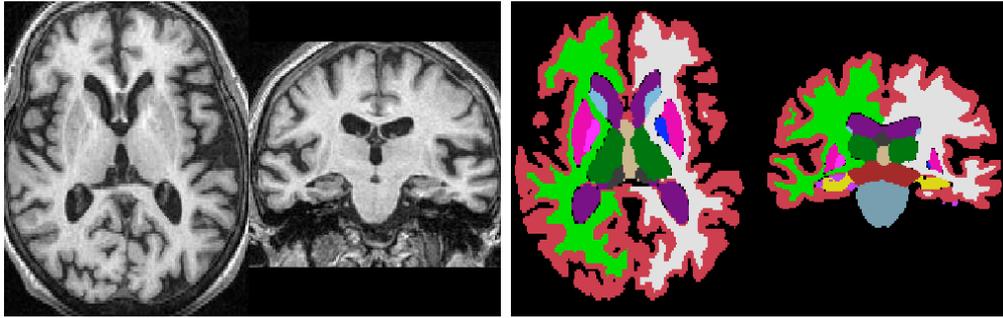


Figure 1.4: A T1-weighted MR image (left, axial and sagittal sections) and a corresponding structural segmentation (right). The labels in the right hand image were generated by combining multiple label sets that were propagated from different atlas images to the T1 image.

94], where the aim is often to identify differences between two groups based on a set of images acquired at the same time. Longitudinal methods rely on intra-subject registrations between the serially acquired images and cross-sectional methods use inter-subject registrations where one of the images is often a reference or a template image. Inter-subject registrations between a group of subjects and a reference image can also be used to provide atlases that can represent the group, containing the average of some feature derived from the subjects, for example by providing an estimate of the typical anatomy.

The work presented in this thesis draws upon information derived from both inter- and intra-subject registrations and from image segmentation. Using intra-subject registrations for serially acquired data, estimates of longitudinal change can be made. Inter-subject registrations and atlasing methods can then be used to characterise patterns of longitudinal change across subjects within a serially acquired cohort. By combining the information provided by registration with the information provided by segmentation, regional estimates of growth or atrophy can be made and the data from such regional estimates can, in turn, be explored in a clinical context.

1.4 Contribution

The thesis presents work on the combined use of registration based methods, to measure and characterise longitudinal change, and of segmentation methods to define regions of interest.

The contributions presented in this thesis can be divided into four areas, two are methodological and two are based on the application of these methods to specific types of data:

- A framework is presented for the combined longitudinal and cross-sectional analysis of patterns of longitudinal change within serially acquired data for a cohort of subjects. The patterns of longitudinal change are investigated through the combined use of longitudinal, intra-subject, registrations and inter-subject registrations that are used to generate average space atlases.
- A framework is developed for the use of a repository of labelled images (atlases) to generate segmentations in unseen query images. This framework treats the atlases as classifiers and combines their propagated labels to provide segmentations for the query subject. Novel methods for classifier selection are developed and analysed.
- Using these frameworks, atlases of brain morphometry and patterns of development are characterised for a group of young children based on scans acquired at one and two years of age. Quantitative estimates of growth between ages one and two derived from this serial data are presented.
- Patterns of brain atrophy are similarly investigated for a group of subjects with Alzheimer's disease and a group of matched controls. For this group of subjects, the combined use of growth estimates from registration and structural and tissue segmentations is shown to be a powerful discriminator between patients diagnosed with Alzheimer's disease and controls.

1.5 Outline of thesis

The remainder of the thesis begins with a description in Chapter 2 of the background and state of the art for the types of image analysis methods used in this work. This covers various aspects of image registration, segmentation and atlasing and a brief overview of morphometry is also given. Chapter 3 presents details on how registration can be used to assess longitudinal changes within a cohort. In particular, a novel framework is presented for characterising patterns of longitudinal change using a combination of serially acquired data for a cohort, information derived from longitudinal registrations and atlasing techniques. Chapter 4 presents a description of structural segmentation through the use of labelled atlas repositories, label propagation classifier fusion. This chapter also outlines novel techniques in which classifiers can be selected from an atlas repository in order to provide segmentations more appropriate for the query subject and to overcome the computational burden associated with the use of a large repository. In Chapter 5, various experiments are carried out to demonstrate the effectiveness of the classifier selection methods described in the Chapter 4. Using the methods presented in the earlier chapters, novel data are presented in Chapter 6 on the global and regionally specific brain growth in children between the ages of one year and two years. The same methods are also applied to serial scans acquired from a cohort of patients with Alzheimer's disease and a group of matched controls. The data, presented in Chapter 7 characterise patterns of atrophy for the groups and use measures of atrophy combined with structural data to show good discrimination between these clinical groups.

Chapter 2

Background

2.1 Introduction

The aim of this chapter is to provide a description of state-of-the-art methods that can be used for analysis of MR anatomical images of the brain along with a brief description of the associated background literature. These methods can be broadly categorised under the general headings of image registration, segmentation, morphometry and atlasing.

2.2 Image registration

In the following chapters, registration-based methods are used to generate atlases at different timepoints and to estimate longitudinal change from serially acquired data. They are also used as part of a segmentation method based on the propagation of multiple manual segmentations to a query image – a step that helps to identify specific structural areas and therefore helps to provide regionally specific estimates of growth or atrophy.

The following sections present background and description for the voxel-based registration method used in this work. More comprehensive reviews of image registration techniques in general can be found in many sources. Hajnal, Hill and

Hawkes [81], Zitová and Flusser [189] and Maintz and Viergever [116] provide useful starting points for an overview of this large and growing field.

2.2.1 Voxel-based registration

The goal of any registration method is to find a spatial correspondence between images of different subjects' anatomies, or images of the same subject acquired at different times. Given that registration methods are applied to images, the correspondence being found is between imaged representations of the anatomies rather than the anatomies themselves. An image of an anatomy can be viewed as a set of samples of a spatially varying signal that represents the response of the different tissues of the anatomy to the imaging modality. Usually, the samples are associated with a regular lattice of points in three dimensions, with each point being the centre of a cuboid shaped voxel (or a rectangular pixel in 2D). In MR images, the intensity associated with a voxel represents an estimate of the average response of the tissue located within it. The aim of voxel-based registration is to estimate the geometrical correspondence between two images based on the intensity values in each image – this contrasts with methods that require features to be first extracted from the images which are subsequently used during the registration.

The model of a typical voxel-based registration consists of a number of inter-related components: a transformation model; a regularisation method; a similarity metric; an optimisation method; and an interpolation method. Finding the correspondence between the images entails finding the best transformation under the model used. This is represented by the set of transformation parameters that provide optimal similarity between the images. It is possible, however, for optimal similarity to be achieved using transformations that are not plausible in relation to the application – for example, a registration of two MR brain scans of the same subject should not, in general, produce a transformation that breaks

topology. For this reason, voxel-based registration schemes often also incorporate a regularisation component, so that the optimised parameters represent a plausible transformation.

To clarify the terminology used in this work, one of the images being registered is termed the ‘target’ image and the other image is termed the ‘source’. After registration, the estimated transformation is used to map locations from the target image to the source image. This allows the intensities in the source image to be ‘pulled back’ to the target’s frame of reference, i.e. the source can be transformed to the target. This process is illustrated in Figure 2.1, where dots represent the locations of the voxel lattices of each image. A particular voxel location \mathbf{x} , in the target image, is mapped to a corresponding location $\mathbf{h}(\mathbf{x})$ in the source image. The intensity in the source image at $\mathbf{h}(\mathbf{x})$ can then be estimated and assigned to the location \mathbf{x} in the target reference frame. Repeating this process for all target voxels, a deformed version of the source can be estimated in the target reference frame. In general, the location in the source image, $\mathbf{h}(\mathbf{x})$, does not coincide with any of the source image’s voxel locations. For this reason, the intensity at $\mathbf{h}(\mathbf{x})$ needs to be estimated using some form of interpolation.

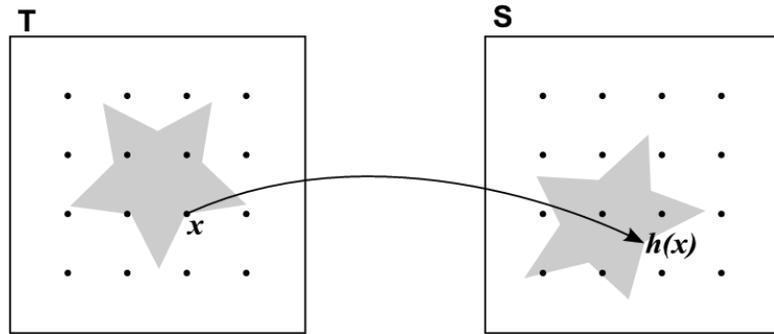


Figure 2.1: A target image T and a source image S . The dots represent the locations of the voxel lattices of each image. A transformation \mathbf{h} maps a location \mathbf{x} in T to a corresponding location $\mathbf{h}(\mathbf{x})$ in S . The source intensity at $\mathbf{h}(\mathbf{x})$ is estimated, ‘pulled back’ and associated with the target voxel at \mathbf{x} . Repeating this process for all target voxels allows the source image to be deformed onto the target frame of reference.

The following sections describe in more detail the aspects of the voxel-based registration method used in this work.

2.2.2 Transformation models

A number of different models can be used to represent transformations between a pair of images. The general aim is to obtain a parametric representation that allows any location in the target image to be mapped to the source image. There are many ways to parametrise transformations, but a broad categorisation of transformations can be given by their geometric properties. *Rigid* transformations preserve distances between points that are mapped from the target to the source. I.e. for two target points, \mathbf{x}_1 and \mathbf{x}_2 and a rigid transformation \mathbf{T} , $\|\mathbf{x}_1 - \mathbf{x}_2\| = \|\mathbf{T}(\mathbf{x}_1) - \mathbf{T}(\mathbf{x}_2)\|$. *Affine* transformations map straight lines to straight lines, preserving the collinearity of points in the target when transformed to the source. Furthermore, the ratios of distances between points are preserved under affine transformations. For three collinear points \mathbf{x}_1 , \mathbf{x}_2 and \mathbf{x}_3 the collinearity property can be expressed by

$$\mathbf{x}_3 = c\mathbf{x}_1 + (1 - c)\mathbf{x}_2 \quad ,$$

for some scalar c . This relationship is preserved under an affine transformation, i.e.

$$\mathbf{T}(\mathbf{x}_3) = c\mathbf{T}(\mathbf{x}_1) + (1 - c)\mathbf{T}(\mathbf{x}_2)$$

for the same scalar c . Transformations that represent a varying local deformation are known as *non-rigid* transformations. Under this distinction, the term ‘global’ is often used to describe a rigid or affine transformation, while the term ‘local’ is used to describe a non-rigid transformation.

The choice of type of transformation is affected by the application in which it is to be used. For example, if the images to be registered represent serially acquired scans of the head of a healthy adult, then a rigid transformation is appropriate as there is (in general) negligible change in the shape of the cranium in adults. However, the difference between two cranial scans of a child acquired at, say, one and two years of age would be better represented by an affine transformation as

there is a significant amount of global change. In both of these cases, there may be small scale differences that remain after the global transformation (rigid or affine) has been applied – this may necessitate a further non-rigid registration step to achieve an improved correspondence.

2.2.3 Rigid transformations

Under a rigid body transformation, angles and lengths are preserved and it can be represented by applying rotations and translations. In 3D, a general rotation can be decomposed into rotations about each of the coordinate axes and a translation can be parametrised by a single vector. In the 2D case, the images can be viewed as being embedded in a plane. The only rotations that need to be considered are rotations about an axis perpendicular to the plane.

Considering first, the two-dimensional case, the matrix representation $\mathbf{R}_z(\gamma)$ of an anti-clockwise rotation of the xy plane about the origin (z axis) by an angle γ can be derived from its effect on the unit vectors $(1, 0)^T$ and $(0, 1)^T$ as follows

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} \cos \gamma \\ \sin \gamma \end{pmatrix} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} -\sin \gamma \\ \cos \gamma \end{pmatrix}$$

giving

$$\mathbf{R}_z(\gamma) = \begin{pmatrix} \cos \gamma & -\sin \gamma \\ \sin \gamma & \cos \gamma \end{pmatrix}$$

Representing a rotation in this form allows it to pre-multiply a vector $\mathbf{x} = (x, y)^T$ representing a location in the Euclidean plane \mathbb{R}^2 .

A translation in the xy plane can be represented by a single vector $(a, b)^T$ whose components are the image of the origin under the translation. The effect of applying a rotation matrix \mathbf{R} and a translation vector \mathbf{v} to a point with co-

ordinates \mathbf{x} can be written as

$$\mathbf{x} \rightarrow \mathbf{R}\mathbf{x} + \mathbf{v} \tag{2.1}$$

It is possible to simplify this by appending an extra component to the vector \mathbf{x} and interpreting it as a set of homogeneous co-ordinates, i.e.

$$\begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

This requires the rotation matrix to be represented as

$$\begin{pmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and a translation to be represented as

$$\begin{pmatrix} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix}.$$

In block form, the rotation and translation can be written as

$$\begin{pmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \mathbf{I} & \mathbf{v} \\ \mathbf{0} & 1 \end{pmatrix},$$

where \mathbf{I} represents the identity matrix. Multiplying these two matrices allows a

single matrix to represent the effect of applying the rotation and the translation

$$\begin{pmatrix} \mathbf{I} & \mathbf{v} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{R} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R} & \mathbf{v} \\ 0 & 1 \end{pmatrix},$$

and the single matrix can be shown to be equivalent to the Euclidean co-ordinate representation in Equation 2.1 by noting that

$$\begin{pmatrix} \mathbf{R} & \mathbf{v} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R}\mathbf{x} + \mathbf{v} \\ 1 \end{pmatrix}.$$

This block representation for rigid transformations also applies to the three-dimensional case. Clearly, the translation vector needs to have three components, i.e.

$$\mathbf{v} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

Some choice is available for the representation of two-dimensional transformations, for example a rotation may be carried out first followed by a translation and vice versa. In three dimensions the situation is made more complex due to an increased number of choices. For example, in the case of three-dimensional rotations, a first rotation can be made about the x axis, and the next rotation can be made about the y axis, which can either be considered as a new y axis (i.e. affected by the previous rotation) or a *fixed* y axis. In the following, all rotation axes are considered to be fixed, i.e. unaffected by any previous rotations.

A rotation of γ about the z axis (from the x axis to the y axis in a right-handed

frame) in three-dimensional homogeneous coordinates becomes

$$\mathbf{R}_z(\gamma) = \begin{pmatrix} \cos \gamma & -\sin \gamma & 0 & 0 \\ \sin \gamma & \cos \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Similarly, rotations of α and β about the x and y axes can be represented respectively by

$$\mathbf{R}_x(\alpha) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha & 0 \\ 0 & -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and

$$\mathbf{R}_y(\beta) = \begin{pmatrix} \cos \beta & 0 & \sin \beta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \beta & 0 & \cos \beta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

In general, a 3D rotation about an axis through the origin can be effected by first rotating about the x axis, then the y axis then the z axis. This means that the single matrix representation

$$\begin{pmatrix} \mathbf{R} & \mathbf{v} \\ 0 & 1 \end{pmatrix}$$

still holds for the rigid transformation, where $\mathbf{R} = \mathbf{R}_z(\gamma)\mathbf{R}_y(\beta)\mathbf{R}_x(\alpha)$ and $\mathbf{v} = (a, b, c)^T$. This also shows that rigid transformations in three dimensions can be represented using 6 parameters.

2.2.4 Affine transformations

There may be a need to take into account differences in size between the images being registered, for example there may be a need to calibrate the scanner to prevent ‘drift’ in voxel sizes over time. If such size differences also need to be modeled, then scales and shears can be used in addition to rigid body motion. Under affine transformations, parallel lines are preserved but lengths and angles may be altered.

Scale factors can be applied along each of the coordinate axes, affecting coordinates along the chosen axis by different amounts. For example a scaling by a factor of s_y along the y axis is represented by

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

with similar representations for scales in the x and z directions.

Shears in 2D are readily defined by a line through the origin and a scalar value. Any 2D point is sheared by translating it parallel to the chosen line by an amount proportional to the point’s distance from the chosen line. In 3D, shears become more complex and can be characterised in more than one way. A 3D shear can be defined as a translation in one coordinate direction by an amount equal to a linear combination of the other two coordinate values. For example, the x coordinate could vary by a linear combination of the y and z coordinates

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} x + ay + bz \\ y \\ z \end{pmatrix}$$

A shear characterised in this way is known as a *beam* shear. Using the notation

of Chen et al. [35] the above beam shear can be written in terms of its parameters and the axis of interest as $S(x_b, a, b)$ with the suffix indicating a beam shear. This particular beam shear leaves points with $y = z = 0$ (the x axis) invariant.

A shear can also be defined as a translation affecting two coordinates, each by an amount proportional to the third, for example, with the y coordinate determining an xz translation, we have

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} x + ay \\ y \\ z + by \end{pmatrix}$$

This type of shear is known as a *slice* shear and the above example can be written as $S(y_s, a, b)$. This shear leaves points with $y = 0$ (i.e. the xz plane) invariant.

Despite it being possible to characterise shears in different ways, the representations are equivalent in the sense that the composition of three beam shears along orthogonal invariant axes can be re-written as a composition of three slice shears using orthogonal invariant planes. As an example, it can be shown that

$$S(z_b, e, f)S(y_b, c, d)S(x_b, a, b) \equiv S(x_s, c, e + fc)S(y_s, a, f)S(z_s, b - ad, d)$$

As before, it is possible to represent these shears as matrices acting upon homogeneous coordinates. For example, the slice shear above can be written in matrix form as

$$\begin{pmatrix} 1 & a & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & b & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Affine transformations can therefore be decomposed into a rigid portion (rotations and translation) and a set of scalings and shears. Rotations and scales have one parameter per axis, and a translation is specified with three parameters while

the shears, as described above, require two parameters per invariant axis/plane.

The total number of parameters needed to specify an affine transformation therefore appears to be 15 but these parameters are in fact coupled so that the effective number of parameters is fewer than 15. The reason for this can be seen in the form of the matrix representations for the transformations. Each transformation can be carried out by multiplying by a matrix of the form

$$\begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{23} & m_{33} & m_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

This type of matrix maps points represented in the form

$$\begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix}$$

to points of the same form, where \mathbf{x} represents a 3-D point. This mapping property requires that the final row of the matrix consists of zeros with a one in the final position. The product of two matrices of the above form is also of the this form, so the final matrix representing the affine transformation will be specified by 12 parameters.

2.2.5 Non-rigid transformations

The number of representations available for affine and rigid transformations are relatively few and even when different representations are used, the resulting transformations can be shown to be equivalent. For example, the same rotation could be represented by a matrix as described above or by using a quaternion representation [149].

After a pair of images have been aligned using a rigid or affine transforma-

tion, residual differences may remain that need to be accounted for by local or non-rigid deformations and there are many possible ways to represent non-rigid transformations. Furthermore, the properties of the non-rigid transformations may vary according to the model used to describe them. The following section describes some of the possible ways of modelling non-rigid transformations and the subsequent section describes the model used for the work presented in this thesis.

2.2.5.1 Background

Affine transformations can be used to account for global differences in location, size or shape between images, but there are occasions where the transformation needs to vary in its effect from one region of the image to another. An example application is given by inter-subject registration. The images of the brains of different subjects generally represent the same set of structures but the shapes of these structures vary in different ways in different parts of the brain. In order to provide a dense correspondence between two such images, the transformation needs to be characterised by more than the parameters used to define a global rigid or affine transformation – the effect of the transformation needs to vary locally across the image.

A function f on a domain Ω is said to be continuous if an infinitesimal change in the input produces an infinitesimal change in the output. The class of all continuous functions on a domain Ω is typically denoted as $C^0(\Omega)$. Functions in the class $C^0(\Omega)$ that also have continuous first derivatives are said to be continuously differentiable and define the class of functions $C^1(\Omega)$. The class $C^k(\Omega)$ encompasses functions f for which f and the derivatives $f^{(1)}, \dots, f^{(k)}$ are continuous and the class $C^\infty(\Omega)$ represents functions for which all derivatives are continuous. In mathematical terms, only functions in $C^\infty(\Omega)$ are described as smooth but in practical terms, functions in the class $C^2(\Omega)$ or even $C^1(\Omega)$ can be sufficiently smooth when modelling structures or transformations. When functions are used

to define mappings between coordinate systems further properties become relevant. A function f that maps points from a set X to a set Y is a homeomorphism if it is a bijection and both f and f^{-1} are continuous. The class of homeomorphisms preserves the topology of the spaces being mapped. If the underlying topology of the anatomy being imaged (for example in a scan of the brain) is assumed to be identical for two images, then a transformation between the images needs to be a homeomorphism. The stronger restriction that f be a bijection and that both f and f^{-1} are C^k continuous (i.e. k times differentiable) defines f as a C^k -diffeomorphism [29]. The term diffeomorphism is typically used for a C^∞ -diffeomorphism. Manifolds [29] related by a homeomorphism are said to be homeomorphic and, similarly, diffeomorphic manifolds are related by a diffeomorphism. It is possible for two manifolds to be homeomorphic (i.e. topologically equivalent) but not diffeomorphic, for example the surfaces of a sphere and a cube. Smooth manifolds are often used to represent the anatomical structures within images (see e.g. [99]) and if the anatomical structures represented within a pair of images are assumed to be smooth, then the transformation relating them is required to be a diffeomorphism of a suitable order. For this reason, diffeomorphic transformations are often used as a theoretical basis for the non-rigid registration of medical images [37, 78].

Non-rigid transformations can be represented as smooth displacement fields, which require a smooth assignment of vectors to each location in an image. For a 3D image, this leads to a representation that requires displacement vector components to be specified for each of the images voxels. The number of parameters needed to define such a transformation would then be three times the number of voxels in the image.

It is, however, possible to represent the number of parameters needed to represent a non-rigid transformation, by using a model or exploiting a property of the transformation, such as its smoothness. If each of a displacement field's separate components are viewed as scalar fields, then it is possible to use a set of smooth

basis functions to model each component. For example it is possible to use a limited number of discrete cosine functions, with varying frequencies, represent a non-rigid transform, each component of the transformation is then written as a linear combination of the basis functions [4]. A registration of two images in this case involves finding the optimal set of coefficients for the basis functions that align the images.

Another established technique for representing non-rigid warps is the use of a set of control points or knot points within the image. A vector is associated with each control point and a mechanism is then required that allows a smooth transformation to be globally defined based on the vectors associated with each of the control point locations.

The control points can be regularly spaced on a lattice within the image or they can be placed at an irregular set of locations. If the control points are irregularly spaced, they can be manually selected in the source and target images to represent homologous locations [166]. Alternatively, irregular control point locations can be automatically selected using a heuristic method [118, 147]. Control points can also be placed on a regular lattice that is aligned with the coordinate axes of the image. This is the approach used within the Free Form Deformation (FFD) model for non-rigid transformations [140].

It is possible to begin a registration with a small number of sparse control points and use a larger number of denser control points as the registration proceeds. In this way, the registration can begin by optimising for the low frequency components of the warp and later matches details at a smaller scale. Such a 'coarse-to-fine' strategy is popular, not least because it helps reduce the risk of a registration converging on a local optimum. As an example of this strategy, Shen et al. [147] who begin by (heuristically) selecting a small number control points, based on distinctive voxel correspondences, to start a registration and increase the number selected as the registration continues.

Using a FFD model, Schnabel et al. [143] begin a registration using a control

point lattice with a large control point spacing. The parameters for this initial lattice are optimised and the resulting transformation captures the non-rigid differences between the images at the largest scale. After optimising the initial lattice it is possible to re-represent the same FFD transformation using a subdivided lattice with half the control point spacing of the initial lattice [60]. The subdivided FFD can capture a finer level of detail and forms the starting point for a new optimisation step. An example of a FFD transformation and the result of subdividing it is illustrated in Figure 2.2.

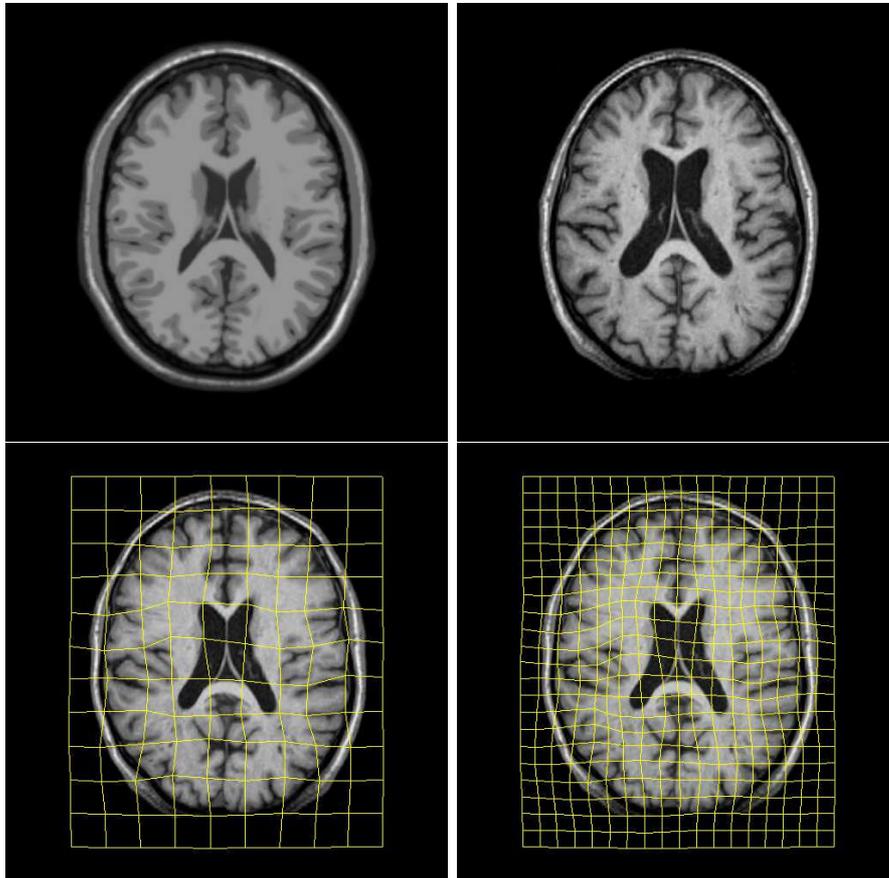


Figure 2.2: An example of the lattice subdivision of a FFD. Top left: A target image for a registration. Top right: The source image after affine alignment to the target. Bottom left: The source image after non-rigid FFD alignment. The control point spacing is 20mm and the deformed FFD grid is shown. Bottom right: The result of subdividing the 20mm FFD lattice, this FFD has a 10mm control point spacing and forms the starting estimate for the next step in the hierarchical registration.

For transformation models that use vectors at control points, the definition

of the transformation at general locations is essentially an interpolation problem. There is therefore a choice as to whether the transformation being represented should interpolate or approximate the vectors at each control point and this choice is linked to the method with which the transformation is defined globally.

For example, Twining et al. [166] use clamped plate splines to interpolate the control point vectors, these splines minimise the energy of the resulting deformation field which means smoothness of the field is ensured by the model chosen to globally represent the transformation. Free form deformations on the other hand approximate the control point vectors rather than interpolate them. This is achieved by convolving the control point vectors with a smoothing B-Spline kernel along each image dimension. Using a B-Spline kernel in this way leads to a smooth displacement field across the entire image. The B-Spline basis function is a piecewise continuous function which has implications for the smoothness and continuity of functions it is used to generate. For example a function generated from a fourth order (cubic) B-spline will have C^2 continuity at the knot-points and C^∞ continuity elsewhere (See e.g. [23]).

Alternative approaches to the representation of non-rigid transformations draw upon the use of physical models. For example, it is possible to represent the space of one of the images being registered as an elastic material that is being deformed towards the other image [10, 71, 63, 147]. In this case, the elastic energy derived from the deformation field also serves to constrain its smoothness. Another good example of the use of a physical model is represented by the use of fluid registration models [37, 66, 47]. Here, the equations that govern the flow of a fluid and under the influence of a force field are used to generate a smooth displacement field between two images. This can be carried out by estimating a series of time dependent smooth velocity fields and the displacement of a point is estimated by integrating its flow over the velocity fields [13].

The process of diffusion can also serve as a model for image registration, an example is presented by Thirion [163] who uses the concept of 'Maxwell's

Demons'. Pennec et al. [128] show how models that based on Maxwell's demons, with appropriate choices of parameters can be re-interpreted as examples of fluid or elastic registration depending on the parameters selected.

2.2.5.2 The selected non-rigid transformation model

The non-rigid deformations used in this work are represented using the free-form deformation (FFD) model, an approach introduced within the computer graphics community by Sederberg and Parry [145] and subsequently adapted for medical image registration by Rueckert *et al.* [140].

Free-form deformations belong within the category of spline-based methods but contrast with, for example, thin-plate splines [22, 166]. Approaches using thin plate splines typically require homologous points to be defined in both images and the displacements defined by these landmarks are used to parametrise the transformation. The landmarks may represent anatomically corresponding locations (e.g. [22]) or may be identified using a heuristic approach based on geometric properties derived from the images [147]. FFDs are typically parametrised using a regular grid of control points that do not have anatomical significance - these are sometimes termed quasi- or pseudo-landmarks. It is possible, however, to define FFDs based on grids other than a regular rectilinear lattice. For example, cylindrical meshes have been used to define FFDs for the registration of cardiac MR images [34]. Indeed FFDs have also been developed for grids with an arbitrary topology [33] that can be tailored to the anatomy being registered.

While FFDs are parametrised by the displacements at the control points, the displacements at general locations are defined using a B-spline kernel. The B-spline kernel has compact support [167] which implies that the displacement at a particular location depends only on those control points in a local neighbourhood and, conversely, a control point only affects a limited number of voxels. This contrasts with the radial basis functions used to define displacements in a thin-plate spline model. The (theoretically) infinite support of these basis functions means

that a control point’s effect extends over the whole field of view of the image. This can be computationally prohibitive when large numbers of control points are used as well as making the modelling of local deformations more difficult.

FFDs share properties with elastic models that can also use a regular grid of control points (see e.g. [70]) as well as irregular tetrahedral meshes of control points (see e.g. [131]). Displacements at control points for an elastic model are interpolated exactly with barycentric coordinates being used to estimate displacements at non-nodal locations. Like FFDs, the control points’ effect is restricted to a compact region, but FFDs, by contrast, approximate rather than interpolate the control point vectors (See e.g. [110] for a discussion on approximation and interpolation). For FFDs displacements in general are evaluated using a simple tensor product which contrasts with elastic models that require material properties to be modeled and typically use a finite element scheme to optimise the mesh displacements (see e.g. [70]). The use of a general tetrahedral mesh may also necessitate a topology correction step (see e.g. [131]).

The fact that FFDs are parametrised makes it possible to represent a deformation field more economically than within schemes that require a displacement field to be stored for each voxel in an image as can be the case, for example, in fluid schemes (see e.g. [37]). On the other hand, fluid schemes are capable of representing relatively large deformations, although a regriding scheme may be required if the Jacobians tend to become singular (see e.g. [37]). FFDs and elastic approaches by contrast are more suitable for modelling relatively small deformations - after pose and scale differences have been corrected for with an affine alignment step. It is, however, possible to carry out a series of successive FFD alignments with the resulting FFDs being composed to give a final displacement field [138]. This allows large displacements to be modeled although the resulting displacements themselves are no longer represented by a single FFD defined by a regular lattice.

A free-form deformation can be parametrised by a set of vectors $\{\Phi_{i,j,k}\}$.

Each vector is associated with one of a set of control points that are arranged on a regular $n_x \times n_y \times n_z$ lattice with spacings of δ_x , δ_y and δ_z along each dimension; the subscripts i, j, k index the location of a control point within the lattice.

By blending the control point vectors using a suitable basis function, a continuously varying displacement can be defined at each point of the target image domain. In this work, the control points are blended using the cubic B-spline basis functions:

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

and the local displacement at a location (x, y, z) is given by a tensor product over the control point vectors

$$\mathbf{T}_{local}(x, y, z) = \sum_{l=0}^3 \sum_{m=0}^3 \sum_{n=0}^3 B_l(r) B_m(s) B_n(t) \Phi_{i+l, j+m, k+n}. \quad (2.2)$$

Setting $i = \lfloor x/\delta_x \rfloor - 1$, $j = \lfloor y/\delta_y \rfloor - 1$ and $k = \lfloor z/\delta_z \rfloor - 1$ ensures that the set of control points $\{\Phi_{i+l, j+m, k+n}\}$ for $0 \leq i, j, k \leq 3$ are those in the neighbourhood of (x, y, z) that contribute to the B-spline summation. This reflects the compact support of the B-spline basis functions. The value of r , where the first basis function is evaluated is given by

$$r = x/\delta_x - \lfloor x/\delta_x \rfloor,$$

and similar expressions are used to derive values for s and t from the y and z coordinates.

2.2.6 Multi-level free-form deformations

When carrying out a registration using FFDs, it is possible to apply a hierarchical, coarse-to-fine approach. The first step involves optimising the parameters of a FFD with a large control point spacing in order to align the larger features of the images. After these parameters are optimised, it is possible to subdivide the lattice of control points using B-spline subdivision [60] and use the resulting control points to run a second registration step using a lattice with half the control point spacing. Proceeding in this way, more of the finer detail in the images is captured after each subdivision step.

In parallel with the subdivision of the control point spacings, it is also appropriate to blur and resample the images being registered, so that the scale of the features in the images is appropriate for the spacing of the control point lattice in the FFD. For example, an image may be blurred such that information relating to structures smaller than a certain size, say 20mm, is effectively removed. A FFD applied to such an image should have an appropriate control point spacing as this affects the degree of local control. In the above example, the use of a control point spacing of 10mm to 20mm would be appropriate for the scale of the images. The theory of scale space for images is a well established area of study (see e.g. [112]) and a discussion of multi-resolution approaches to optimising B-splines is given in Lee *et al.* [108] but a formal study of the interaction of image scales and multi-resolution B-splines is yet to be carried out. The images in this work were blurred and down-sampled to achieve the appropriate scale for the coarser control point spacings, with the degree of blurring and resampling resolution getting successively smaller at each step. Further details are given in section 3.1.2.

2.2.7 Combining the global and local components of a transformation

The registration of two images in this work was carried out in two main steps. First the global transformation (rigid or affine) was estimated, and this was then used as the starting point of the non-rigid registration. The global transformation itself can also be estimated in two steps, by first estimating the rigid transformation and using it as the starting point of the affine registration. The transition from rigid to affine registration is straightforward because the rigid transformation parameters resulting from the first step are a subset of the affine transformation parameters that are optimised in the second step. The affine registration therefore begins with initial translation and rotation parameters that are based on the rigid registration estimate and with scale and shear parameters set to one and zero respectively.

The global transformation between the images is obtained after the affine registration step and the subsequent non-rigid or local registration seeks to estimate the local residual displacements required to align the images after applying the global transformation. This means that, under the model used, the global and local components of the transformation were combined by addition. The addition of the global and local components introduces an interrelationship between them and this is discussed in more detail in section 3.1.3.1.

Let \mathbf{T}_{global} represent a global transformation and \mathbf{T}_{local} a local non-rigid displacement field. The global transformation can be represented by a translation vector \mathbf{d} and a transformation matrix \mathbf{M} . Matrix \mathbf{M} can either represent a rigid rotation using 3 parameters (rotations about each of the axes), or it can represent a 9 parameter affine transformation matrix \mathbf{M} encoding rotations, scales and shears. The global transformation is therefore represented as

$$\mathbf{T}_{global}(\mathbf{x}) = \mathbf{M}\mathbf{x} + \mathbf{d}. \quad (2.3)$$

The complete transformation \mathbf{T} that accounts for both global and local differences between a pair of images is modelled as the sum of these local and global components

$$\mathbf{T}(\mathbf{x}) = \mathbf{T}_{global}(\mathbf{x}) + \mathbf{T}_{local}(\mathbf{x}) = \mathbf{M}\mathbf{x} + \mathbf{d} + \mathbf{T}_{local}(\mathbf{x}) \quad (2.4)$$

for each location \mathbf{x} in the target image. Alternative methods for combining the global and local transformations can be used, for example they can be composed. The use of addition makes the transformations simpler to manipulate, for example the evaluation of Jacobians is easier (see section 3.2 for more detail).

2.2.8 Similarity metrics

When optimising parameters during a registration, some measure of the similarity of the images is required. The goal of the optimisation is to maximise the similarity of the images. Alternatively, if a measure of difference is used, the registration becomes a minimisation problem, but this can be converted into a maximisation problem by negating the difference measure and considering it to be a similarity metric. For example if $d(A, B|\mathbf{T})$ represents a measure of difference between two images A and B given a transformation \mathbf{T} then the aim of a registration is to minimise d and an alternative optimisation would be one that aims to maximise $-d$, i.e. $-d$ can be viewed as a similarity metric.

In order to calculate a similarity metric for a given pair of images, an estimate of the correspondence between them is needed. In this work, transformations map locations from a target image I_t to a source image I_s . If $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ represent the locations of the voxels in the target image, the corresponding source image locations under a transformation \mathbf{T} , from I_t to I_s , are $\{\mathbf{T}(\mathbf{x}_1), \dots, \mathbf{T}(\mathbf{x}_n)\}$. The set of intensity pairs $\{(I_t(\mathbf{x}), I_s(\mathbf{T}(\mathbf{x}_1))), \dots, (I_t(\mathbf{x}_n), I_s(\mathbf{T}(\mathbf{x}_n)))\}$ is then used in the calculation of a similarity metric during voxel-based registration. For the sake of clarity, let the set of corresponding target and source image intensity pairs be

represented more compactly by $\{(t_1, s_1), \dots, (t_n, s_n)\}$.

Many different similarity metrics are available. It is possible to use the sums of squared differences metric [82, 67], which penalises differences in intensity between corresponding voxels.

$$SSD = \frac{1}{n} \sum_{i=1}^n (t_i - s_i)^2$$

Using $-SSD$ as a similarity measure implies an assumption that the images should appear identical when correctly aligned, i.e. that the any differences in their intensities after alignment should only be due to noise.

Cross correlation is defined as

$$CC = \frac{\sum (t_i - \bar{t})(s_i - \bar{s})}{\sqrt{\sum (t_i - \bar{t})^2} \sqrt{\sum (s_i - \bar{s})^2}}.$$

where the summations are carried out over n voxel locations and \bar{t} and \bar{s} are the means of the target and source intensities respectively. Cross correlation is a general measure of statistical agreement but it has been used as a similarity measure [52]. Its use as a similarity measure implies the assumption that the intensities at corresponding locations of the images being registered have a linear relationship when they are aligned. Their dynamic ranges may differ, but an intensity scaling and shift can be used to map intensities at corresponding locations with, once again, any residuals between them being due to noise.

If the intensity pairs are viewed as entries in the joint histogram of the image pair, information theoretic measures can also be defined. With appropriate binning and normalisation, it is possible to provide estimates of the probability $p(t)$ of the occurrence of a particular intensity t in the target image, the probability $p(s)$ of intensity s occurring in the source and the probability $p(t, s)$ of the co-occurrence of these intensities (i.e. their occurrence at corresponding voxel locations).

Using these estimates of probability, it is possible to calculate the Shannon entropy [146] of the target image as

$$H(I_t) = - \sum_t p(t) \log(p(t))$$

with the entropy of the source image similarly defined. In the above expression, the summation ranges over all intensities t in the target image. The summation represents the expected uncertainty or information content of the image. An image that is completely uniform has $p(t) = 1$ for its single intensity and therefore has an information content of zero. An image consisting of only noise without any structure will, by contrast, have a high Shannon entropy. The joint entropy of the image pair is defined as

$$H(I_t, I_s) = - \sum_t \sum_s p(t, s) \log(p(t, s)).$$

Mutual information, which is defined as

$$MI = H(I_t) + H(I_s) - H(I_t, I_s).$$

A registration using MI as a similarity measure seeks a trade-off between maximising the marginal entropies of the images while reducing their joint entropy. MI has been in use as an image registration similarity metric since 1995 [169, 40] and was introduced mainly to register multimodality images, where the assumption of a linear relationship between corresponding intensities does not hold. For some pairs of images (e.g. PET and MR) there may not even be a functional relationship between the images.

Normalised mutual information is defined as

$$NMI = \frac{H(I_t) + H(I_s)}{H(I_t, I_s)},$$

i.e. it represents the ratio of the sum of the marginal entropies to the joint entropy in contrast to the difference represented by mutual information. Using a set of simulated examples, Studholme *et al.* [159] showed that *NMI* was more robust to variations in image overlap during the registration and also showed, using clinical data, that the registration accuracy was also robust to variations in the imaged field of view.

In practice, the values for the marginal and joint entropies can be estimated by constructing the joint histogram of the two images. If the intensities in each of the images are allocated to a series of bins, the joint histogram can be represented by a set of values $n(t, s)$ that count the number of co-occurrences of each of the (binned) intensity pairs (t, s) in the two images. The joint probability of $p(t, s)$ can then be estimated as

$$p(t, s) = \frac{1}{N} n(t, s)$$

where N represents the total number of samples in the histogram. The marginal histograms for each of the images separately can be estimated by summing the joint histogram along its rows or columns. The marginal probabilities $p(t)$ and $p(s)$ can, in turn, be similarly estimated. After estimation of the joint and marginal probabilities, the entropies can be calculated as described above. An alternative approach to the estimation of the joint and marginal probabilities is through the use of Parzen representations [127, 168].

2.2.9 Interpolation

The calculation of the similarity metrics described in Section 2.2.8 is based on the intensity correspondences obtained from the target voxel locations $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ and the source locations $\{\mathbf{T}(\mathbf{x}_1), \dots, \mathbf{T}(\mathbf{x}_n)\}$ under the transformation estimate \mathbf{T} . The source locations are clearly unlikely to coincide with the locations of the voxel centres in the source image. For this reason, the source intensities need to be interpolated from the sampled source values prior to evaluation of the similarity

metric.

There are various choices of interpolation method, with the simplest being nearest neighbour interpolation. Tri-linear interpolation is commonly used. More sophisticated interpolation schemes are available, such as piece-wise continuous cubic interpolation [105], B-spline interpolation [167] and sinc interpolation [161]. Lehmann et al. [110] provide a good overview of interpolation methods in medical image registration, and the performance of interpolators is discussed more theoretically in Thévenaz et al. [162].

From a registration perspective, the main issues underlying the choice of interpolation method centre on the trade-off between the performance cost of evaluating the interpolator and the quality it produces. For example, the calculations for linear interpolation are relatively fast but can affect the image in the same way as a low-pass filter. A converse example is given by sinc interpolation, which preserves the spectral content of the image but is very costly to compute.

2.2.10 Optimisation

The registrations carried out in this work optimised the similarity metric with respect to the transformation parameters using either downhill descent or steepest gradient descent (Strictly speaking, these should be called ‘uphill ascent’ and ‘steepest gradient ascent’ as the similarity measures are being maximised but expressing the problem as a minimisation is more common in the optimisation literature – see [58], for example).

The downhill method was used when optimising global (rigid or affine) transformations. With this approach, the transformation parameters are individually perturbed by a chosen step size and the similarity metric is re-evaluated for each perturbation. The parameter giving the biggest increase in similarity is then selected and the transformation is updated by modifying this parameter. This process is repeated until no further increase in similarity is achieved. The step

sizes were chosen empirically. A large initial step size was used and the transformation parameters optimised. This can then be repeated using successively smaller step sizes until an accurate rigid or affine alignment is obtained. For the relatively small number of parameters being optimised during rigid or affine registration, downhill descent represents a simple and fast method that generates transformations with no significant difference from those obtained by gradient based methods (such as gradient descent) without the overhead of calculating the gradient of the metric with respect to the transformation parameters.

Gradient descent was used to optimise the parameters during non-rigid registration (the components of the control point vectors). If C represents the similarity metric and Φ represents the collected components of the control point vectors, then gradient ascent simply optimises by stepping along the direction of maximum increase, i.e.

$$\Phi + \delta \nabla_{\Phi} C,$$

where δ represents the step size. This expression replaces the set of parameters Φ at each iteration step. Once again, successively smaller step sizes can be used to optimise in a coarse-to-fine fashion.

The gradient term $\nabla_{\Phi} C$ can be evaluated analytically for similarity metrics such as SSD , and it is also possible to obtain an analytic expression for the gradient of more complicated similarity metrics such as MI [90]. Alternatively, the similarity gradient can be estimated by simply using a finite difference method, for example by finding the central differences.

2.2.11 Regularisation

The relatively high dimension of non-rigid transformations means that a large number of such transformations can be used to align a pair of images. The existence of such multiple solutions makes the non-rigid registration of images an ill-posed problem. This leads to the notion of what may be described as the

plausibility of the transformation that relates the images that need to be registered, thereby restricting the number of possible solutions. For example, if it is known that the images represent anatomies with the same topology, then a plausible transformation should not alter the topology in any way. Typically, the smoothness of the transformation has been used as a measure of its plausibility [140, 13, 103] and therefore the process of providing smooth, and hence plausible, transformations has often been termed regularisation. The use of regularisation means that the under-determined nature of the registration problem becomes a better-determined or even a well-posed problem.

If a transformation \mathbf{T} represents the registration estimate between a target image A and a source image B , then the registration problem can be expressed in a Bayesian framework [125] as the requirement to maximise the posterior probability of the transformation given the image data:

$$P(\mathbf{T}|A, B) \propto P(A, B|\mathbf{T})P(\mathbf{T}),$$

where $P(A, B|\mathbf{T})$, the likelihood of the image data given the transformation, can be represented by the similarity of the images under the correspondence given by \mathbf{T} . The plausibility of the transformation is expressed by $P(\mathbf{T})$, its prior probability. There is, however, a trade-off between the aims of optimising for both the transformation plausibility and the image similarity. If smoothness is the desired property of plausible transformations, then the extreme case of the identity transformation represents the smoothest possible example but is likely to lead to a poor match between the images.

The deformation model can, by definition, provide a high degree of regularity in the final transformation. This is apparent in the case of FFD based transformations where the use of B-splines to interpolate the control point vectors gives a natural smoothness to the resulting transformation [140]. The use of clamped plate splines [166] also, by definition, minimises the bending energy of

the transformation that interpolates the displacements at the control points.

Additionally, a number of approaches for generating plausible transformation estimates have been provided in the literature. The broad categories of these approaches include those which explicitly model the material properties of the anatomies being registered, those which apply a processing step to the obtained deformations and methods which incorporate a term penalising implausible transformations during the optimisation step.

Soza *et al.* [156] estimate parameters for the mechanical properties of the tissue and directly incorporate them into the registration model. Karacali *et al.* [103] process a given deformation, which may be irregular and break topology, and estimate a similar deformation that is smooth and preserves topology. Rohlfing *et al.* [136] introduce *a priori* knowledge of the incompressibility of breast tissue in serial MR images and express this constraint as a penalty term with the similarity metric.

Using an additional term in the similarity metric or objective function that relates to the smoothness of the transformation is a typical approach due to the ease with which it can be evaluated from the deformation and incorporated into the voxel-based registration framework. If $C_{similarity}$ represents the similarity metric and C_{reg} represents a measure of the plausibility or regularity of the transformation, then the overall objective function C is represented by

$$C = C_{similarity} + \lambda C_{reg}, \quad (2.5)$$

where λ determines the relative contribution of the regularisation and similarity terms during the optimisation. An appropriate value of λ needs to be established. A value that is too low may lead to implausible transformations, while values that are too high may constrain the registration too much and prevent it from aligning the images sufficiently accurately.

As an example, in the large deformation diffeomorphic metric mapping (LD-

DMM) model, the transformation estimate is obtained by integrating a series of time varying velocity fields. By choosing an appropriate norm on the velocity vector fields and including it as a separate term in the cost function, it is possible to generate smooth displacement fields [13].

For the purposes of this work, the regularisation term used, C_{reg} , was the bending energy associated with the transformation [170, 140]. This energy is calculated using the second order derivatives of the transformation \mathbf{T} .

$$C_{reg} = -\frac{1}{|\Omega|} \int \int \int \left\| \frac{\partial^2 \mathbf{T}}{\partial x^2} \right\|^2 + \left\| \frac{\partial^2 \mathbf{T}}{\partial y^2} \right\|^2 + \left\| \frac{\partial^2 \mathbf{T}}{\partial z^2} \right\|^2 + \left\| \frac{\partial^2 \mathbf{T}}{\partial xy^2} \right\|^2 + \left\| \frac{\partial^2 \mathbf{T}}{\partial xz^2} \right\|^2 + \left\| \frac{\partial^2 \mathbf{T}}{\partial yz^2} \right\|^2 dx dy dz.$$

This expression represents a 3-D version of the bending energy associated with the deformation of thin 2-D plate. If a thin flat metal plate is viewed as occupying the xy plane and a series of displacements are applied in the z direction, then expression above measures the energy required to bend the plate to achieve the displacements. A linear or affine transformation has zero bending energy so that the above term is only relevant for non-rigid registrations. If the individual components of the transformation are viewed as scalar functions, then C_{reg} penalises high curvature in these functions. C_{reg} can also be viewed as sum of the norms of the Hessian matrices for the three scalar component functions.

2.3 Segmentation

There is a long history of identifying and delineating structural areas within the brain. The relatively recent introduction of non-invasive scanning techniques has allowed such segmentation to be carried out on imaged surrogates of the anatomy, as an alternative to histological methods. The ability to segment a structure or tissue type in the brain enables attention to be directed to specific regions, and this information can serve to support more general studies of pathology or func-

tion. Functional MRI studies have investigated how activations within different regions of the brain (as measured by blood oxygenation) can be correlated with specific tasks or activities [68]. The localisation of tasks to specific areas of the brain is, however, not new. For example, neuroanatomical correlates between different areas of the brain and specific aspects of linguistic performance have been the focus of many investigations in the form of studies of the effects of localised damage to various regions of the brain [113, 119].

In the context of the measurement of change through time, the ability to generate structural segmentations of the brain serves a useful purpose in that regions of interest can be isolated and the growth or atrophy within specific areas can be assessed. For example, hippocampal atrophy has been identified as a marker for Alzheimer’s disease [77, 94] and an increase in ventricular volume has been associated with multiple sclerosis [62]. An example related to neurodevelopment is the increase in the volume of myelinated white matter that takes place in first two years of life [141].

These examples motivate the identification of regionally specific measures of longitudinal change in the brain and the consequent need to identify the regions, something which can be achieved via segmentation.

2.3.1 Manual segmentation

Human raters can carry out manual segmentations on images of brain anatomies to identify structural and functional boundaries by assigning labels to the voxels of an image. This can be a difficult and time-consuming process but if the human rater is sufficiently experienced and the protocol that is applied is clear and detailed, then the levels of inter- or intra-rater agreement can be high enough for the resulting segmentations to be treated as a ‘gold standard’. Inter- and intra-rater agreements are often reported as intra-class correlation coefficients and, for manual segmentations, values of 0.93 and above are typical [21, 98] although these

figures clearly depend upon the training and experience of the expert raters.

The high quality of the segmentations manually produced by experts is clearly desirable but there is a trade-off against the level of effort required. Segmentations provided by experts have been used for many years and are especially important for structures that are difficult to segment automatically such as the hippocampus [95, 132]. If there are a large number of unseen images that need to be segmented, the effort of manual segmentation may become unfeasible and alternative, more automated approaches present themselves as a viable alternative. As well as being time-consuming, manual segmentation can be prone to errors that depend on human factors (e.g. inter- and intra- observer variation, drift in practice over time), the nature of segmentation protocols and acquisition effects (e.g. contrast characteristics, motion artefacts, scanner calibration issues etc.). In some cases, depending upon the application, deterministic and robust automated schemes may be preferred for their consistency and applicability to a wide range of data.

2.3.2 Automated segmentation

Numerous approaches have been developed to address the problem of automated image segmentation. In the context of general image processing, this involves the assignment of labels to the pixels or voxels of an image that share some feature or features. For example by labelling contiguous regions or areas similar in colour. In medical image processing the task is mainly to assign labels to the voxels within a particular structure or tissue type.

Noisy data represents a challenge to all automated segmentation methods and, for the segmentation of MR images in particular, effects such as poor contrast, inhomogeneity and partial voluming add to the difficulty of the task. A good image segmentation technique should be robust to such effects.

Unsupervised classifiers or clustering techniques are well established in the

general image processing community. The mean-shift [41] approach places a kernel over each point in the image and uses this kernel to evaluate a regionally specific mean. By moving in the direction of each mean iteratively, a maximum point is eventually reached and each starting point can be labelled according to which maximum it arrives at. The number of clusters found in such a scheme can vary and is dependent on the size of the kernel. Mean shift is a non-parametric technique that seeks to estimate a density. The feature space being considered, for example the image intensities, can be viewed as a probability density function and the local maxima of the this function, or its modes, represent locations that can be viewed as cluster centres.

Normalised cuts [148] represents another unsupervised approach where the pixels in an image are considered to be nodes in a weighted graph where the weights along each edge measure the ‘similarity’ of the corresponding pairs of pixels. Spectral clustering techniques applied to a matrix derived from this graph (the Laplacian) can be used to partition the image into two clusters of points where the within cluster similarity is high and the between cluster similarity is low. Each cluster corresponds to a region in the original image and repeated application of the technique can allow an arbitrary number of regions to be segmented.

An example of the use of an unsupervised method in medical image processing can be found in [129] where the a fuzzy c-means algorithm is used to segment the tissue types in MR brain images. This approach requires initial values for the mean intensities of each tissue types in order to begin the algorithm and the authors present automated methods for locating these values.

An approach that has found a good deal of success in tissue segmentation combines a parametric model for representing the tissue types and a supervised maximum likelihood for finding the optimal parameters. Such an approach requires training data in order to create prior probability estimates of the spatial variation of each tissue. Typically, such a framework models the tissue intensities

as a mixture of Gaussians and the expectation maximisation (EM) algorithm can be used to find estimates for the model parameters for each tissue in a given image [177, 109]. The EM algorithm alternates between estimating the model parameters for each tissue and updating the tissue membership probabilities for each tissue at each voxel.

Alternative approaches explicitly model the boundary of the object or structure being segmented. Active contours, or snakes, have been used to segment objects in 2-D images [104] where an initial estimate is driven by ‘forces’ derived from the image data (for example the gradient) to the boundary of an object. Typically, an internal force constrains the contour to be smooth.

Surfaces can be evolved in 3-D images to match the boundaries of structures. The brain extraction method of Smith [152] expands a sphere mesh, initialised at a location based on the image’s centre of gravity, outwards to meet the brain boundary using forces derived from the profile of intensities along the surface normals at each vertex. Han et al. [86] initialise an estimate of the cortical surface based on the boundary between grey and white matter found by tissue segmentation. A level set approach is then used to grow this surface to find the boundary of the cortex using forces derived from the tissue membership probabilities.

Models for the boundaries of structures can be trained using sets of example data where there is known point correspondence. Active appearance models (AAMs) represent the typical patterns of variation for the shape of a structure and the intensity patterns within or on the boundary of the structure [44]. After training such a model, the intensity patterns, or appearance, of an unseen image can be used to instantiate an instance of the model thereby providing an estimate of the structure boundary.

Finally, atlas-based segmentation is an approach that allows expert manual segmentations of images to be propagated to new images [93, 160, 49]. This clearly requires an estimate of the correspondence between the atlas or segmented image and the new image. This means that atlas-based segmentation requires a

registration step to allow the labels to be propagated.

The propagation of labels from multiple atlases, and their subsequent combination or fusion is a method that has shown good levels of accuracy, comparable with that of manual segmentation [135, 89]. Atlas-based segmentation and classifier fusion will be discussed in more detail in Chapter 4.

Methods that are trained in order to build an explicit model, such as AAM, require effort to be expended during this training stage. The resulting model is, however, reasonably compact and can easily be applied to new data. A single atlas-based segmentation method does not build an explicit model but uses an exemplar segmentation that represents the structure(s) for the whole population. The overhead in the application of atlas-based segmentation is mainly represented by the alignment step and this depends on the nature of the registration used. Multiple label propagation and fusion again does not build an explicit model and the bulk of its computational cost is in the registration of the multiple atlases to the query. There is an additional cost in the need to store the entire database of atlases. These costs are somewhat offset by the higher segmentation accuracy possible in comparison with other methods [89].

The generalisability of a model-based approach depends on how comprehensive the training data are in their representation of the population and on how much of the detail within the population is retained within the model. The generalisability of label propagation and fusion depends on the size and variability of the atlas database and on how well correspondence can be established with unseen subjects.

2.4 Atlas construction

Given a set of images for a group of subjects, it is possible to generate atlases to represent the cohort. The atlases can represent any of a number of features for the group, for example, the average anatomy, average tissue probability maps

or average patterns of longitudinal change. Atlases can help in determining the main clinical characteristics of the group, the main differences between clinical groups or the differences between the same group observed at different times. If an atlas represents a healthy population, it can be used to support diagnosis of disease when compared with an image for a particular individual.

One of the simplest methods for generating an atlas from a set of images is to register all the images to a chosen reference subject and use the resulting transformations to align all the images to the reference. An atlas can be created by averaging the values at corresponding voxels for the spatially normalised images.

As an example, using a set of scans acquired from a group of 25 two-year-old subjects, two atlases were created. The first atlas was created by registering all the subjects to a chosen reference and averaging the resulting images. The second atlas was generated to be in an ‘average space’ using the methods described in Section 3.4. Illustrations of these atlases are shown in Figure 2.3, where the bias of the ‘reference space’ atlas to the chosen reference is apparent, as are the differences between the reference space and average space atlases. For example variations in the boundary of the skull that are particular to the subject chosen as a reference (see Figure 2.3, top row) remain apparent in the atlas constructed by registering to the reference and averaging (see Figure 2.3, middle row). A similar comment applies to a slight asymmetry in reference subjects lateral ventricles. These particular features of the reference do not, however, remain in the average space atlas (see Figure 2.3, bottom row). Section 3.4 gives further details of the average space atlas creation. This illustrates the obvious drawback of atlases generated simply by registering to a reference and averaging. The atlas is likely to be biased to the chosen reference subject and this may confound any further data generated based upon the use of the atlas.

A number of methods have been proposed for the generation of ‘average-space’ or ‘template-free’ atlases. Guimond et al. [80] align all the images to a reference space using affine registrations. Working with the aligned images, a ‘demons’-

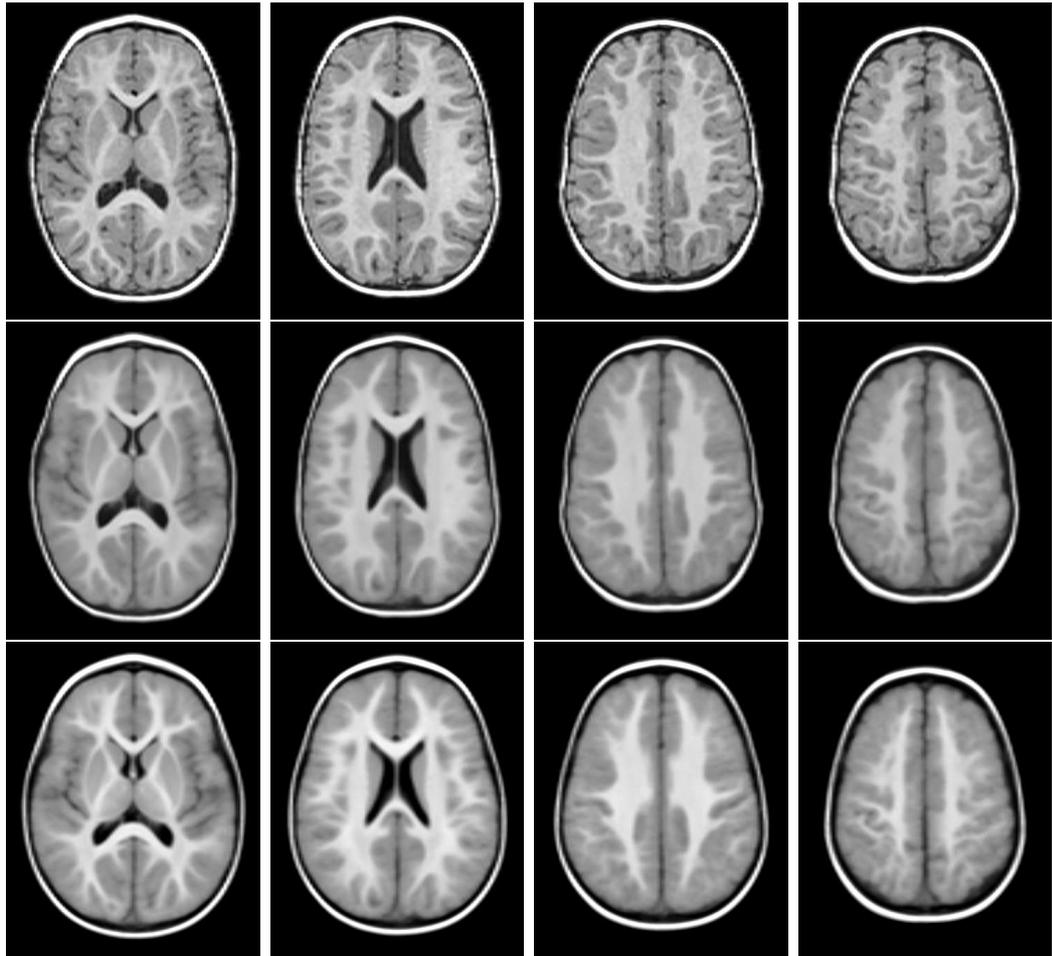


Figure 2.3: Atlasing a group of two-year-old subjects in reference and average space. Top row: Transverse slices taken from a reference subject. Middle: Slices of an atlas of the group created by transforming all subjects to the reference and averaging intensities. Bottom: An average space atlas of the same group of subjects.

based registration [128] step is carried out to correct for residual non-rigid differences between each image and the reference. The non-rigid registrations are then averaged and used to generate a new reference estimate and the process is repeated.

Wang et al. [173] affinely align a group of images for which sets of manual labels are available. They then compare the use of different images as a reference template from which labels are propagated to each of the other images. The fitness of a reference is then measured by how well it predicts the manual labels. The reference template can be an individual image in the group or an average constructed through the use of the average non-rigid pairwise transformations between the affinely aligned images.

Bhatia et al. [15, 14] use a similarity metric that measures the normalised mutual information of each of the images in an affinely aligned set with a fixed intensity reference. The images are then simultaneously non-rigidly aligned to a postulated average space, subject to the constraint that the sum of all the deformation fields is equal to zero.

Working with large deformation diffeomorphic metric mappings (LDDMM), Joshi et al. [101] simultaneously register a set of images using a cost function that penalises both intensity differences between the images and the total amount of deformation induced by the large deformation diffeomorphic mappings from the subjects to the atlas. The amount of deformation for a diffeomorphism is measured by integrating the norm of a differential operator applied to the time varying velocity fields that determine the diffeomorphism.

Approaches to generating average atlases can rely on first averaging the transformations between the different images from which the atlas is created. Methods used to average transformations can be analogous to the physical process of finding the centroid of a group of masses. There is more than one possible way to find a centroid iteratively, one possible method is illustrated in Figure 2.4. If $\mathbf{x}_1, \dots, \mathbf{x}_n$ represent locations for unit masses, then their centroid $\bar{\mathbf{x}}$ can be

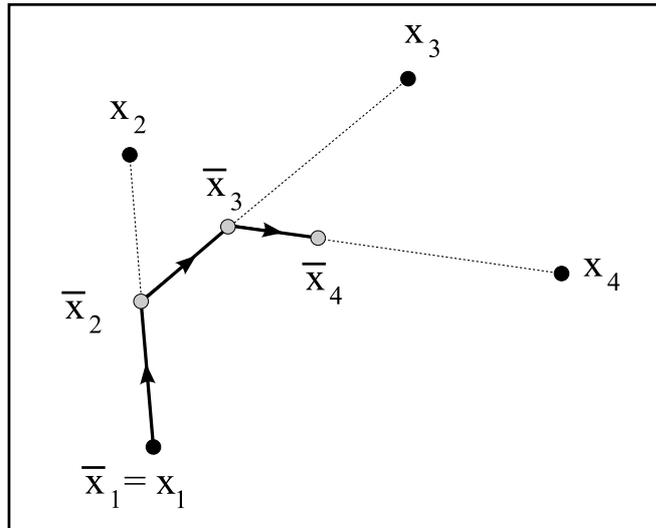


Figure 2.4: A schematic diagram of the process of iteratively calculating the location of the centroid for a group of equal masses. The centroid is initialised to be at the same location as one of the masses. Thereafter, its location can be updated by fractionally moving along the displacements to each of the remaining masses. The fraction moved diminishes with each iteration. Analogous methods have been used to provide the averages of sets of transformations between images for the purposes of atlas generation.

located using an iterative scheme as shown in Algorithm 1.

Algorithm 1 Iterative scheme for finding the centroid of a set of equal masses.

Input Locations $\mathbf{x}_1, \dots, \mathbf{x}_n$:

Output : $\bar{\mathbf{x}}$

$\bar{\mathbf{x}}_1 \leftarrow \mathbf{x}_1$

$i \leftarrow 1$

do

$\Delta = \mathbf{x}_{i+1} - \bar{\mathbf{x}}_i$

$\bar{\mathbf{x}}_{i+1} = \bar{\mathbf{x}}_i + \frac{1}{i+1}\Delta$

$i \leftarrow i + 1$

while ($i < n$)

The approach known as Woods matrix averaging [180, 181] finds the average of a set of affine transformations in an analogous way to the method described for iteratively finding the centroid. The points shown in Figure 2.4 in this case represent images and a line between two points represents the affine transformation between the images. Traversing a fraction of a line between two points becomes finding a root of the affine transformation. It would be desirable, given a set

of images, that the average space reached be the same regardless of the starting image and the subsequent traversal through the images. Another desirable property is that the affine transformations from the estimated average space to the input images give the identity when composed. Given that affine transformation matrices do not, in general, commute, it seems unlikely that the results will be exact although the results presented by in Woods *et al.* [180, 181] and in subsequent work suggest that the method provides a reasonable approximation. Avants [8] presents a similar scheme for the averaging of non-rigid transformations. Woods [179] presents further algorithms that can be used to average non-rigid transformations using their Jacobians. Because each Jacobian tensor represents a local affine transformation matrix. The complete average transformation is then characterised by the aggregation of the averages of the Jacobian matrices across the region of interest.

Beg *et al.* [12] present an averaging method based on the concept of ‘geodesic shooting’. The authors present an argument that large deformation diffeomorphic metric mappings (LDDMM) can be completely characterised by the initial ($t = 0$) velocity vector field and by a ‘speed’ term. Within such a framework, they further argue that the transformation between two images can then be represented by an initial ‘momentum’ and that the momenta representing a number of transformations can be linearly averaged.

2.5 Morphometry

2.5.1 Background

The ability to characterise physiological features such as the size and shape of neuro-anatomical structures or the quantity or density of different types of tissue at different locations is clearly of use in that it provides additional information that clinicians may use during diagnosis or treatment. In this context, a wide

range of methods for analysing image data, encompassing registration, segmentation and atlasing techniques, fall under the general heading of ‘morphometry’. The different approaches can be placed under the broad categories of voxel-based morphometry (VBM), deformation-based morphometry (DBM) and tensor-based morphometry (TBM).

A typical application of VBM [6, 75] aims to find differences between two groups of subjects. The images for all the subjects are aligned to a reference image. The registrations used for this alignment typically correct for affine differences and have a low-dimensional non-rigid component. The images are then segmented into different tissue types, and one of the tissue types, typically grey matter, is used to calculate a statistic (e.g. a t-statistic) on a per-voxel basis using the information on group membership. In this manner, regions in the brain where there are statistically significant differences in tissue density between the groups can be located.

VBM therefore uses the information present in the images after they have been spatially normalised. DBM and TBM, by contrast, focus attention on the transformations that align the images to the reference. If all the images being studied are registered to a reference using a high-dimensional non-rigid registration step, then they will be made to resemble the reference closely and the variation within the group is encoded in the transformations mapping the subjects to reference.

The terminology is not entirely consistent in the literature but DBM can be used to describe approaches that apply statistical techniques to the displacement fields directly, for example by using a general linear model and multivariate statistics [69].

If information based on derivatives of the non-rigid transformations is used, for example by calculating the Jacobian tensor, then the term TBM can be used to describe the approach [38].

The term DBM is, however, also used in studies where the Jacobian tensor is calculated, for example in a longitudinal study of atrophy [28] and in a cross-

sectional study of preterm infants and controls [20].

2.5.2 Morphometry for characterising longitudinal changes

When using images to identify changes over time, some way of ordering the images in time is needed. This necessitates attaching some form of temporal marker to the images. In most cases, this is simply the age of the subject at the time of the scan, but the marker can also represent a stage during disease progression or treatment; for example, images can be acquired longitudinally, before, during and after a particular intervention.

When using image data to measure longitudinal change, more than one experimental design is possible, depending on how the data are acquired:

- In a cross-sectional study that tracks longitudinal change,¹ the data acquired generally represent a snapshot of a number of individuals, where each subject is associated with a particular stage or time-point. For example, a group of children with a range of ages can be scanned and their ages recorded. Subsequently, measurements (e.g. volumes) derived from the imaged data can be related to the childrens' ages to provide a representation of the longitudinal change. The main confounding factor in cross-sectional studies that measure longitudinal change is the influence of inter-subject variability on the tissue or structure volumes being estimated. Variations between subjects may affect any subsequent estimates of change that are made.
- A longitudinal study in which each subject is scanned at two or more time-points is less susceptible to the effect of inter-subject variability, because each subject effectively acts as its own control, and changes, such as those due to growth or degeneration, are more easily isolated. Furthermore, if se-

¹A different sense of the term 'cross-sectional' is used to describe studies carrying out a static (i.e. non-time dependent) comparison of two or more groups. The focus of this work, however, is longitudinal change and the meaning should be clear from the context.

rial scans are acquired for subjects in distinct clinical groups (for example one group receiving a treatment and a group of controls), then the longitudinal data can be used to make estimates of change that can then be compared across groups in a mixed design experiment that incorporates both longitudinal and cross-sectional features.

The simplest longitudinal study is one where scans are acquired at two comparable timepoints. For example, scans might be acquired at specific ages for a cohort of subjects. In this case, as well as obtaining estimates of longitudinal change for particular individuals, it is possible to generate atlases for the cohort at each timepoint. This can help to identify general patterns of change over the interval studied for the cohort as a whole.

2.6 Conclusion

This chapter has presented a review of the methods that can be used to analyse image data in general and for the purpose of characterising longitudinal change in particular. Descriptions have been given of methods that can be used to register and to segment images along with techniques for atlasing images and characterising longitudinal change.

Registration, segmentation and atlasing are components that will form parts of later parts of this thesis. Chapters 4 and 5 use registrations to propagate manual labels to unseen images in the context of atlas-based segmentation and classifier fusion. Chapter 3 will give further detail of a registration-based framework that can be used to characterise longitudinal change. This framework incorporates longitudinal registrations for measuring growth and inter-subject registrations for generating average space atlases. This framework is applied to different sets of data in Chapters 6 and 7 where information derived from structural segmentations and tissue segmentations is also used.

Chapter 3

Registration-based methods for measuring longitudinal change

The ability to assess and quantify structural changes in the brain over time is clearly useful in a clinical setting. Examples of such change are growth during childhood, degeneration due to ageing or pathology or changes due to therapy for a particular condition. For any of these examples, identifying patterns of longitudinal change for a cohort, either throughout the brain or by region, can provide clinically useful information. Such information can be used to track growth and development or it can serve as a bio-marker for the degree of pathology, for example during a clinical trial [51].

In finding the patterns of variation in the longitudinal change for a cohort, assessments relating to individual subjects can be made. This can help to determine, for example, whether a child's brain is developing as expected, to decide if a particular clinical intervention is appropriate, or to assess whether a patient with a neurodegenerative condition is responding to a particular drug or treatment.

The remainder of this chapter describes methods that can be used to assess and quantify longitudinal change and to create atlases in a longitudinally tracked cohort. These methods are based on a variety of approaches. Intra-subject longitudinal registrations are used to estimate change for individuals. Inter-subject

registrations are used to generate atlases of a cohort at different timepoints. The atlases are created in an ‘average space’, using averages of inter-subject transformations between the subjects in a cohort and a particular reference. These methods are compared and combined with segmentation techniques to provide either validation or further estimates of change.

3.1 Registration and volume change estimation

Before describing some of the choices taken for the registration models and parameters, a brief description of the data is presented to provide some context. The registrations carried out in this work for volume change estimation were applied to two sets of serially acquired T1-weighted MR images. One set was acquired from a group of children at one and two years of age. A second set consists of baseline and follow-up scans acquired from a group of Alzheimer’s disease patients and age-matched controls. The childrens’ image data was reconstructed to have an in-plane resolution of 1 mm^2 and a slice thickness of 1.6 mm . The images used in the Alzheimer’s study were reconstructed with an in-plane resolution of 0.9375 mm^2 and a slice thickness of 1.5 mm .

3.1.1 Registration model choices

For a given pair of images, a registration can be carried out in two steps by first estimating a global transformation which is then used as a starting point for a non-rigid registration step. The choice of global transformation between the images depends on the type of images being registered. For example, if the images represent serial scans of an adult subject, then a rigid transformation is appropriate, as it is reasonable to assume that there is no change in global head shape between acquisitions. If the images represent scans of different subjects, or longitudinal scans of young infants where global shape changes are plausible, then a 12 parameter affine transformation is more appropriate.

The local non-rigid registrations were carried out using a B-spline based free-form deformation (FFD) model [140] in a hierarchical coarse-to-fine manner [143] as described in Section 2.2.6. As well as speeding up the optimisation process, the use of this hierarchical coarse to fine approach when optimising the non-rigid transformation parameters reduces the chance of being trapped by local minima of the objective function [143].

Linear interpolation was used for the source image and normalised mutual information [159] was used as the similarity metric. The regulariser used was the transformation bending energy [170, 140]. This combination of choices was found to be reasonably robust and appropriate for both inter-subject and intra-subject registrations presented in this work. Further discussion of related parameter choices is given in the next section.

3.1.2 Registration parameter choices

The hierarchical coarse-to-fine optimisation of the FFD registrations was carried out with successive control point spacings of 20 mm , 10 mm , 5 mm and 2.5 mm . As an example, a FFD with a control point spacing of 20 mm was first estimated and was then subdivided to generate an initial estimate for a 10 mm spacing FFD. This was in turn optimised and the process repeated.

In tandem with the subdivision of the FFDs, the images were also blurred with a Gaussian kernel and down-sampled. The extent of the downsampling and blurring was based on the native resolution of the image and varied for each FFD control point spacing. During the final optimisation stage the images were registered in their native resolution and without resampling. The kernel sizes and down-sampling factors used are shown in Table 3.1, where r represents the smallest dimension of image voxel size. For example, an image with a smallest voxel dimension of 1 mm , being registered with a FFD having a 20 mm control point spacing, would be blurred with a Gaussian kernel of 1 mm and resampled

to an isotropic voxel size of $2 \times 1 = 2mm$. This example relates to the first row in Table 3.1.

Control point spacing	Down-sampling factor	Kernel size
20mm	$2r$	r
10mm	$1.5r$	$0.75r$
5mm	-	$0.5r$
2.5mm	-	-

Table 3.1: Image down-sampling factors and sizes of blurring kernel used during hierarchical optimisation of the FFDs. Dashes indicate that either down-sampling or blurring was not applied, i.e. the original image data was used.

In order to control the relative contribution of the regularisation term (see Equation 2.5), a choice of the weighting value of $\lambda = 0.01$ was made. Clearly it is desirable for the estimated transformation to be smooth, but if the value of λ is too high, the transformation will become too constrained. In other words there is a trade-off between the smoothness and the accuracy of the transformation.

As an experiment to test the effect of the regularisation parameter λ , the images for a subject scanned at one and two years were segmented using an expectation maximisation (EM) [109, 123] based algorithm to give tissue estimates for white and grey matter. The images were registered using different values of λ . Based on the resulting transformations, the average tissue overlap (measured by the Dice coefficient [50]) was found between the year one tissues and the transformed year two tissues.

The overlap values found after the different choices of λ are shown in Figure 3.1. The overlap deteriorates for the highest values of λ (≥ 1), reflecting the trade-off between smoothness and accuracy. Choosing the value of λ that gives the most regularisation prior to the drop in accuracy gives the value of 0.01. It should be noted, however, that the use of B-spline basis functions to weight the control points of a free-form deformation means that there is an intrinsic level of smoothing already present in the transformation model.

Another factor that can affect the smoothness of the transformation is the choice of similarity metric. For example, the use of *SSD* or *CC* was found to

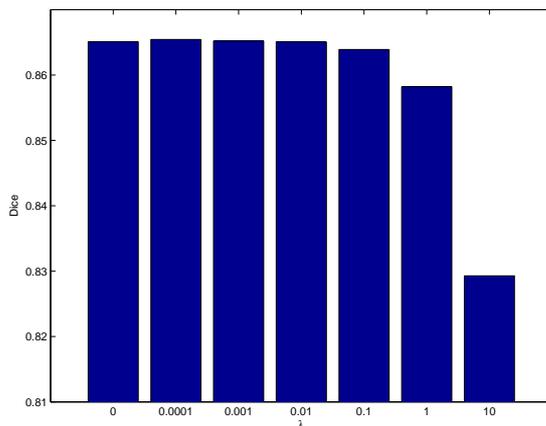


Figure 3.1: The effect of the choice of regularisation parameter λ . For various settings of λ , a longitudinal (intra-subject) registration was carried out for a subject. The resulting transformations were then used to align tissue segmentations in each of the images. The average overlap of grey and white matter was calculated and is shown on the vertical axis.

be more likely to produce transformations which are less smooth. Examples of deformations produced by *SSD*, *CC* and *NMI* are illustrated in Figure 3.2. All parameters apart from the similarity metric were the same for these registrations and the value of λ was 0.01. It can be seen that *SSD* produces deformations which are the least smooth, *NMI* the smoothest and *CC* is intermediate.

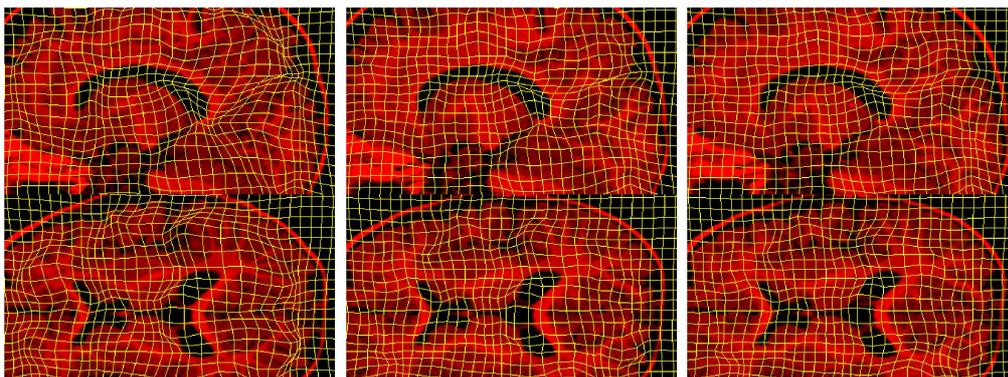


Figure 3.2: An illustration of the different amounts of transformation smoothness produced by different similarity measures during an inter-subject registration. Left to right: *SSD*, *CC*, *NMI*.

3.1.3 FFD properties

Two properties of the FFD transformation model used are relevant to the calculation of average transformations used later in this chapter (Section 3.4.1.2). These are the coupling of the global and local components of the transformation and the additive properties of FFDs.

3.1.3.1 Global and local coupling

After the global and non-rigid registration steps, the resulting global transformation \mathbf{T}_{global} and the local displacement field \mathbf{T}_{local} are combined as described in Equation 2.4, which is repeated here:

$$\mathbf{T}(\mathbf{x}) = \mathbf{T}_{global}(\mathbf{x}) + \mathbf{T}_{local}(\mathbf{x}) = \mathbf{M}\mathbf{x} + \mathbf{d} + \mathbf{T}_{local}(\mathbf{x}).$$

Combining the global and local parts of the transformation by addition creates a dependence of the local displacement field on the global transformation. This is illustrated schematically in Figure 3.3, where the local displacements needed to map the left image to the one on the right are shown as dashed arrows. The top pair of images are related by an identity global transformation and the bottom pair of images are related by a 90 degree anti-clockwise rotation. If the displacement field for the bottom pair of images is multiplied by the inverse of their global rotation, the displacement field for the top pair is obtained.

The identity

$$\mathbf{T}(\mathbf{x}) = \mathbf{M}\mathbf{x} + \mathbf{d} + \mathbf{T}_{local}(\mathbf{x}) = \mathbf{M}(\mathbf{x} + \mathbf{M}^{-1}\mathbf{T}_{local}(\mathbf{x})) + \mathbf{d},$$

shows that applying the global transformation followed by a local displacement of $\mathbf{T}_{local}(\mathbf{x})$ is equivalent to applying a local displacement field of $\mathbf{M}^{-1}\mathbf{T}_{local}(\mathbf{x})$ followed by the global transformation.

The displacement field represented by $\mathbf{M}^{-1}\mathbf{T}_{local}$, (that is associated with the

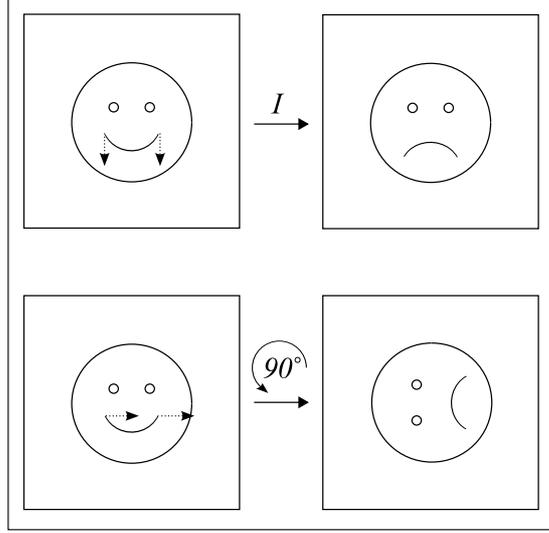


Figure 3.3: The coupling of the global and local parts of a transformation. The displacements required to match each left hand image to the right hand image are shown as dashed arrows. The top images are related by an identity global transformation while the bottom pair are related by a 90 degree rotation.

pair of images with an identity global transformation) can be viewed as a version of \mathbf{T}_{local} that has been ‘de-coupled’ from its associated global transformation \mathbf{M} .

3.1.3.2 Linearity of FFDs

Because \mathbf{T}_{local} is a linear sum of the control point vectors, there is an additive property with respect to different sets of control points. If two FFDs $\mathbf{T}_{\{\Phi\}}$ and $\mathbf{T}_{\{\Gamma\}}$ are defined on the same lattice locations with control point vectors $\{\Phi_{i,j,k}\}$ and $\{\Gamma_{i,j,k}\}$, then

$$\mathbf{T}_{\{\Phi\}} + \mathbf{T}_{\{\Gamma\}} = \mathbf{T}_{\{\Phi+\Gamma\}}$$

i.e. the sum of the displacement fields can be represented by a single FFD with control point vectors $\{\Phi_{i,j,k} + \Gamma_{i,j,k}\}$.

A further consequence of this linearity is that the transformation obtained by multiplying all the control points of a FFD by a fixed matrix has the same effect as multiplying all the displacement vectors by the same matrix. Using the same

notation as above, for a fixed matrix \mathbf{M} ,

$$\mathbf{T}_{\{\mathbf{M}\Phi\}}(\mathbf{x}) = \mathbf{M}\mathbf{T}_{\{\Phi\}}(\mathbf{x})$$

for all locations \mathbf{x} .

3.2 Estimation of longitudinal change

In order to characterise change in individual subjects using longitudinal data, it is possible to perform intra-subject registrations using the baseline and follow-up scans as target and source images respectively. If the transformation $\mathbf{T} = (T_x, T_y, T_z)^T$ maps target (baseline) to source (follow-up) locations, the Jacobian operator D can be applied to the transformation,

$$D\mathbf{T} = \begin{pmatrix} \frac{\partial T_x}{\partial x} & \frac{\partial T_x}{\partial y} & \frac{\partial T_x}{\partial z} \\ \frac{\partial T_y}{\partial x} & \frac{\partial T_y}{\partial y} & \frac{\partial T_y}{\partial z} \\ \frac{\partial T_z}{\partial x} & \frac{\partial T_z}{\partial y} & \frac{\partial T_z}{\partial z} \end{pmatrix}.$$

The Jacobian gives the best local linear approximation to the transformation. In effect, it is an extension of the one-dimensional derivative operator. If \mathbf{v} represents a small perturbation applied to \mathbf{x} , then

$$\mathbf{T}(\mathbf{x} + \mathbf{v}) \approx \mathbf{T}(\mathbf{x}) + D\mathbf{T}(\mathbf{x}) \cdot \mathbf{v}$$

and the volume change induced by the transformation in an infinitesimally small region around a target location \mathbf{x} is given by the determinant of the Jacobian at \mathbf{x} , which can be denoted $J(\mathbf{x})$. From Equation 2.4, and using the linearity of

Jacobian operator,

$$\begin{aligned} J(\mathbf{x}) &= \det(D\mathbf{T}(\mathbf{x})) \\ &= \det(D\mathbf{T}_{global}(\mathbf{x}) + D\mathbf{T}_{local}(\mathbf{x})). \end{aligned} \quad (3.1)$$

Using the notation of Equation 2.3, $D\mathbf{T}_{global}(\mathbf{x}) = \mathbf{M}$. $D\mathbf{T}_{local}(\mathbf{x})$ can be calculated analytically using the derivatives of the B-spline basis functions. Making the separate components of the local displacement described by Equation 2.2 explicit gives

$$\mathbf{T}_{local}(x, y, z) = \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix} = \sum_{l=0}^3 \sum_{m=0}^3 \sum_{n=0}^3 B_l(r)B_m(s)B_n(t)\Phi_{i+l,j+m,k+n}.$$

The partial derivative $\frac{\partial T_x}{\partial y}$, for example, can be obtained as

$$\frac{\partial T_x}{\partial y} = \sum_{l=0}^3 \sum_{m=0}^3 \sum_{n=0}^3 B_l(r)B'_m(s)B_n(t)\Phi_x$$

where, for clarity, Φ_x represents shorthand for the x component of the vector $\Phi_{i+l,j+m,k+n}$. The derivatives of the B-spline basis functions (such as $B'_m(s)$ above) can be evaluated analytically, because closed form expressions of the basis functions are available.

3.3 Averaging volume change estimates

Given a transformation between images, it is possible to estimate a volume change map in the space of each target image using the Jacobian determinants, $J(\mathbf{x})$. The context in which values of $J(\mathbf{x})$ are aggregated or averaged for statistical purposes generally fall under two main headings: they can be used to define volume change for a particular region of interest (ROI) within the image of a

single subject; or they can be averaged, at a particular anatomical location, across multiple subjects. The methods used to average the Jacobian determinant values depend upon which context applies.

3.3.1 Change over a region of interest

If the target and source images represent baseline and follow-up scans for an individual subject, then the map of $J(\mathbf{x})$ represents estimates of tissue volume change during the interval between scans at each voxel. Let Ω represent a target image region of interest (ROI) or structure for which an estimate of change is required, then an estimate of the volume $|\mathbf{T}(\Omega)|$ is needed. This represents the volume of Ω after the transformation is applied and is given by

$$|\mathbf{T}(\Omega)| = \sum_{\mathbf{w} \in \mathbf{T}(\Omega)} 1 \approx \sum_{\mathbf{v} \in \Omega} J(\mathbf{v}) \quad (3.2)$$

where \mathbf{v} and \mathbf{w} represent target and source voxel locations respectively. This estimate represents a discretised version of the change of variable formula for integration.

It is convenient to measure the change for a structure as a growth or atrophy factor $\frac{|\mathbf{T}(\Omega)|}{|\Omega|}$, and this is equivalent to calculating the arithmetic mean of Jacobian determinants over the ROI defined by Ω

$$\frac{|\mathbf{T}(\Omega)|}{|\Omega|} \approx \frac{\sum_{\mathbf{v} \in \Omega} J(\mathbf{v})}{\sum_{\mathbf{v} \in \Omega} 1} = \frac{1}{|\Omega|} \sum_{\mathbf{v} \in \Omega} J(\mathbf{v}). \quad (3.3)$$

This method can be used to find growth factors for segmented structures and only requires one of the images (the baseline) to be segmented.

The Jacobian determinants of a transformation should all be finite and greater than zero for a transformation that does not fold, tear or introduce a reflection. A calculated growth factor of one represents no change in volume at a location, a value greater than one represents expansion. A growth factor with a value less

than one represents shrinking.

So far it has been assumed that Ω is defined by a binary mask, a voxel \mathbf{v} is either in Ω or not, and similarly for $\mathbf{T}(\Omega)$. If, however, a spatially varying probability map $p(\mathbf{v})$ represents membership of the structure or tissue class (e.g. white matter), where $0 \leq p(\mathbf{v}) \leq 1$ for voxels \mathbf{v} in the image domain V , then Equation (3.3) can be modified to become

$$\frac{|\mathbf{T}(\Omega)|}{|\Omega|} \approx \frac{\sum_{\mathbf{v} \in V} p(\mathbf{v})J(\mathbf{v})}{\sum_{\mathbf{v} \in V} p(\mathbf{v})}. \quad (3.4)$$

In this instance, the change for the structure is represented by a weighted linear average of the Jacobian determinant values. If the distribution of $p(\mathbf{v})$ is bi-modal, consisting only of values of zero and one, then Equation 3.4 reduces to Equation 3.3. The application of segmentation methods such as the EM algorithm, to brain images produces probabilistic tissue maps that can be used in conjunction with Equation 3.4.

3.3.2 Average change across subjects

If serial scans at the same or similar timepoints are available for a number of subjects, then longitudinal registrations may be carried out and volume change maps estimated for each individual. It is a natural extension to combine the individual volume change maps in a common coordinate system in order to identify general patterns of change.

Let $\{J_1, \dots, J_n\}$ represent a set of spatially normalised Jacobian determinant maps for a group of n subjects over a particular interval. For a given voxel location \mathbf{v} in the common coordinate system, the values of $J_i(\mathbf{v})$ for $i = 1, \dots, n$ represent a set of multiplicative scale factors for the volume of the voxel at \mathbf{v} . Given this multiplicative nature, a natural choice for the average Jacobian determinant value

at \mathbf{v} , $\bar{J}(\mathbf{v})$ is given by the geometric mean of the individual estimates,

$$\bar{J}(\mathbf{v}) = \left(\prod_{i=1}^n J_i(\mathbf{v}) \right)^{\frac{1}{n}}. \quad (3.5)$$

By choosing the geometric mean to represent a set of Jacobian determinant values at a location, the resulting value can still be interpreted as a local scale factor that is applied to infinitesimal volume neighbourhoods. This is because the volume obtained by successively applying the Jacobian determinants J_i for $i = 1, \dots, n$ is equivalent to applying the geometric mean n times.

The choice of the geometric mean is equivalent to using the arithmetic mean for the values of the log Jacobian. In their work on the use of deformation based morphometry for studying the effects of alcoholism using inter- and intra-subject registration, Rohlfing et al. [137] apply statistical methods to log-Jacobian maps that have been spatially normalised to a reference. They also demonstrate that any per-voxel statistics derived from such maps (e.g. z-scores, t-statistics) are independent of the choice of reference space. This lends support to the choice of the geometric mean as an average for Jacobian determinants from multiple subjects.

A further consideration that needs to be taken into account are variations in the length of the interval over which the scans were acquired for different individuals. For example, if the interval studied represents the year between the first and second birthdays for a group of children, then each subject's scans are obtained at times close to either birthday with some variation.

To account for this variation, the days between scans for the i th individual, d_i , may be used to correct the Jacobian determinant values prior to averaging across subjects as described in Equation 3.5. This can also be carried out geometrically:

$$J_i(\mathbf{v}) \leftarrow (J_i(\mathbf{v}))^{\frac{365}{d_i}},$$

which represents an annualised rate of change. This approach to temporal normalisation of the Jacobian determinants assumes an exponential model for change of the Jacobian determinant over time or, equivalently, a linear model of change in the log determinant. Rohlfing et al.[137] also apply this approach to temporal normalisation.

3.4 Average space atlases via transformation averaging

Given a cohort of subjects with scans available at two or more timepoints, it is possible to generate atlases to represent the cohort at each timepoint. Such atlases can help to characterise volume change for the population as a whole, to estimate region-specific patterns of change or simply as an aid to visualisation.

Anatomical atlases for a group of subjects can be created for the images based on their inter-subject transformations. This study combines methods used for the averaging of non-rigid deformations [80, 139] with estimates of the averages of the global affine transformations – in this way the average space atlases represent the group at a local level as well as in terms of global shape and size.

The inter-subject registrations all use a reference subject as the target image but the possible bias towards the reference may be reduced using the average of inter-subject transformations. This process is schematically illustrated in Figure 3.4.

Let $\{I_1, \dots, I_n\}$ represent the images for all subjects and I_{ref} represent the reference subject's image. Let $\{\mathbf{T}_1, \dots, \mathbf{T}_n\}$ denote the inter-subject transformations (with I_{ref} as target) and let the result of averaging these transformations be denoted $\bar{\mathbf{T}}$. In order for the reference to be included in the atlas, it can be assumed that I_{ref} is included in the set $\{I_1, \dots, I_n\}$ and that the corresponding transformation is the identity. As with the input transformations, I_{ref} is the tar-

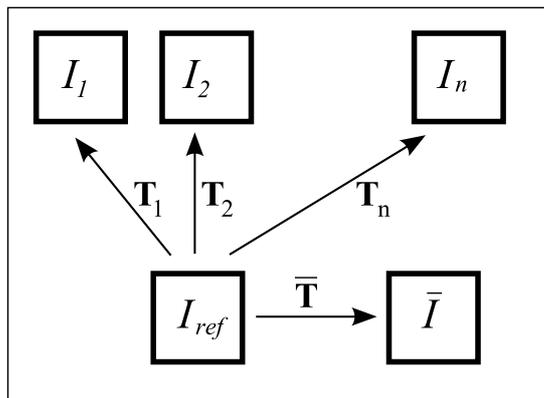


Figure 3.4: Atlas construction: The numbered arrows show the transformations between the reference and the remaining subjects. These transformations are averaged to create $\bar{\mathbf{T}}$. The compositions of $\bar{\mathbf{T}}^{-1}$ with the individual transformations \mathbf{T}_i , $i = 1, \dots, n$, were used to create an atlas in the average space.

get image for $\bar{\mathbf{T}}$. It will, however, map locations in I_{ref} to a postulated average space image \bar{I} . The composition of $\bar{\mathbf{T}}^{-1}$ with each of $\{\mathbf{T}_1, \dots, \mathbf{T}_n\}$ can be used to map locations in \bar{I} to each of the individual subjects' images, i.e. the mapping of a location \mathbf{v} in the average space image to the corresponding location in the image of the i th subject is represented by

$$\mathbf{v} \rightarrow \mathbf{T}_i(\bar{\mathbf{T}}^{-1}(\mathbf{v})).$$

This allows intensity values in the images $\{I_1, \dots, I_n\}$ to be pulled back to voxel locations in \bar{I} prior to averaging. This method for generating atlases is similar to the approach used by Guimond et al. [80] and Rueckert et al. [139] that were applied using local deformations. The approach presented here differs in that the global affine component of the transformations is also included in the averaging step, i.e. information about the average shape in global terms is incorporated into the atlases.

The analysis pipeline, showing the steps taken during the generation of average space atlases, is shown in Figure 3.6. The input to this pipeline is represented by the data to be atlased, for example Jacobian determinant maps, tissue probability maps or anatomies, and a set of transformations between the average space atlas

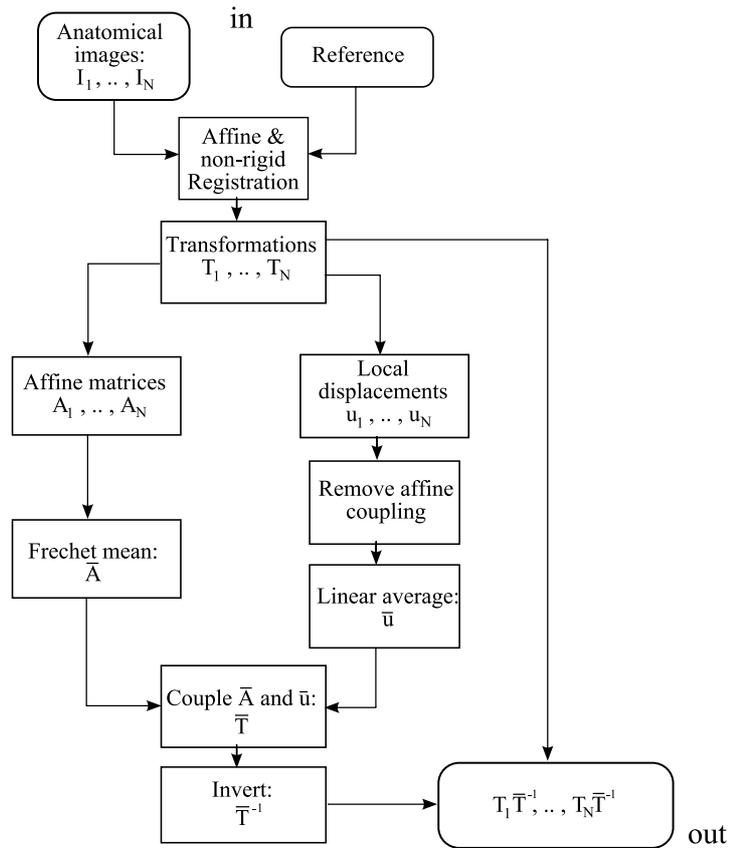


Figure 3.5: A diagram to illustrate the steps taken to find transformations from an average space atlas to the images in a cohort. Details of the individual steps are described in Sections 3.4.1 and 3.4.2.

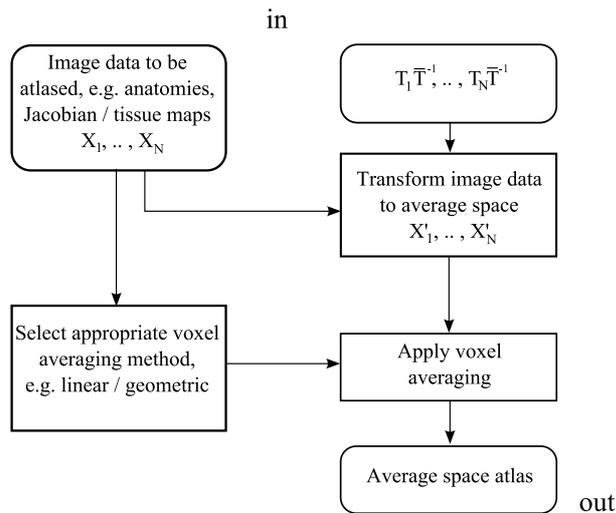


Figure 3.6: A diagram to illustrate the steps taken to produce an average space atlas for a given set of images. The input images can represent different types of data, for example anatomies or Jacobian maps. The transformations $T_1 \bar{T}^{-1}, \dots, T_n \bar{T}^{-1}$ map locations in the average space to each of the input images. See Figure 3.5 for details of the generation of these transformations.

and the individual subjects.

Details of the steps taken to generate the transformations between the average space atlas and the images in the cohort are illustrated in Figure 3.5 and are described in Sections 3.4.1 and 3.4.2.

3.4.1 Averaging transformations

The averaging of the inter-subject transformations during atlas creation was carried out separately for the global and local components of the transformations. The resulting averages of the global and local components were subsequently combined to create the average overall transformation \bar{T} .

3.4.1.1 Averaging global transformations

Using the same notation as that used above, a set of transformations $\{\mathbf{T}_1, \dots, \mathbf{T}_n\}$ relate each subject's image to the reference - one of these transformations can be assumed to be an identity transformation from the reference to itself. For a given subject i , \mathbf{T}_i is represented by

$$\mathbf{T}_i(\mathbf{x}) = \mathbf{T}_{global,i}(\mathbf{x}) + \mathbf{T}_{local,i}(\mathbf{x}) = \mathbf{M}_i\mathbf{x} + \mathbf{d}_i + \mathbf{T}_{local,i}(\mathbf{x}), \quad (3.6)$$

a subject-specific version of Equation 2.4 where the global transformation is determined by affine matrix \mathbf{M}_i and translation \mathbf{d}_i . The affine matrix \mathbf{M}_i can be decomposed as $\mathbf{M}_i = \mathbf{R}_i\mathbf{A}_i$, where \mathbf{R}_i is a rotation matrix and \mathbf{A}_i represents scales and shears.

If the full set of global parameters $\{\mathbf{M}_i, \mathbf{d}_i\}_{i=1,\dots,n}$ are used in the averaging process, then the resulting atlas will also reflect the average position and orientation of the images (relative to the reference). This position and orientation information is independent of the anatomical and longitudinal change information that the cohort encapsulates, so only the matrices $\{\mathbf{A}_i\}_{i=1,\dots,n}$ are used in the averaging process, providing an estimate of the average with respect to global

scales and shears. The restriction of the averaging to the global scales and shears at this stage implies the average space atlases produced have the same position and orientation as the reference subject.

Affine transformation matrices cannot be averaged directly because they do not occupy a linear space. A simple example in 2 dimensions can be seen if linear averaging is applied to two matrices representing the identity and 90° clockwise rotation about the origin:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Intuitively, the average of these matrices should simply represent a 45° rotation. The linear average of these matrices, however, represents a scaling along each axis by a factor of $\frac{1}{\sqrt{2}}$ followed by a clockwise rotation of 45°. This is seen in the following decomposition:

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{-1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

which demonstrates that the linear average is inappropriate in this case.

The averaging of affine matrices requires a different approach that takes into account the manifold that they occupy. Previously, approaches for averaging elements of non-linear spaces have been presented for data representing transformations [24, 1], data representing tensors [3] and, more recently, in a machine vision context, on the space of symmetric covariance matrices of features extracted for the identification of humans in images [165].

The average $\bar{\mathbf{A}}$ of the matrices $\{\mathbf{A}_i\}$ is given by their Frechét (or intrinsic) mean [64] and can be calculated using Algorithm 2. In this algorithm, the log operation maps matrices from their (non-linear) manifold to vectors in the tangent space centred on the current estimate of the average. The vectors are averaged

linearly in the tangent space and the result is mapped back to the manifold using the exponential map. The resulting location is then used as the updated estimate of the mean and the process is repeated until the average of the tangent space discrepancies between the estimate of the average and the data becomes negligible. The discrepancy average is represented by the norm of $\frac{1}{n} \sum_{i=1}^n \log(\bar{\mathbf{A}}^{-1} \mathbf{A}_i)$ (or $\log(\mathbf{A}')$) in Algorithm 2.

Algorithm 2 Find the Frechét mean of a set of matrices.

Input : Matrices $\{\mathbf{A}_i\}_{i=1\dots n}$

Output : Frechét mean $\bar{\mathbf{A}}$

$\bar{\mathbf{A}} = \mathbf{I}$

do

$\mathbf{A}' = \exp\{\frac{1}{n} \sum_{i=1}^n \log(\bar{\mathbf{A}}^{-1} \mathbf{A}_i)\}$

$\bar{\mathbf{A}} = \bar{\mathbf{A}} \mathbf{A}'$

while ($\|\log(\mathbf{A}')\| > \epsilon$)

3.4.1.2 Averaging local transformations

Let the control point vectors for the local displacement field $\mathbf{T}_{local,i}$ be denoted $\{\Phi_{u,v,w,i}\}$ where u, v and w index the position of the control point within the lattice and i indexes the subject. It is worth noting that, since the transformations $\mathbf{T}_{local,i}$ for $i = 1, \dots, n$ all share the same target image (the reference subject), they also share the same locations for their control point lattices. This means that the indices u, v and w range over the same values for all subjects.

Equation (3.6) shows the individual transformations expressed as the sum of a global transformation and a local FFD displacement field $\mathbf{T}_{local,i}$. As described in Section 3.1.3.1, this implies a coupling between the control point vectors and the global affine transformation. Prior to averaging the local transformations, the effect of the corresponding global transformations upon their control point vectors needs to be removed. The effect of the global transformation is removed by pre-multiplying each of the control point vectors by the Jacobian of the corresponding

inverse affine transformation to obtain

$$\{\Phi'_{u,v,w,i}\} = \{\mathbf{M}_i^{-1}\Phi_{u,v,w,i}\}.$$

The linearity of the FFD with respect to its control points (Section 3.1.3.2) means that pre-multiplying all the control points by \mathbf{M}_i^{-1} has the same effect as multiplying the displacements of the FFD globally by the same matrix. The affine-corrected control points $\{\Phi'_{u,v,w,i}\}$ are linearly averaged over all the subjects for $i = 1, \dots, n$ to produce $\{\bar{\Phi}'_{u,v,w}\}$

$$\bar{\Phi}'_{u,v,w} = \frac{1}{n} \sum_{i=1}^n \Phi'_{u,v,w,i}.$$

This makes use of the additive property of free-form deformations that have the same lattice locations that was described in Section 3.1.3.2. This implies that the free-form deformation generated by the average control point vectors $\bar{\Phi}'$ gives the equivalent displacement over all locations as the average of the displacements of the input free-form deformations. Despite the simplicity of this approach, it has been shown that linearly averaging local displacement fields in this way produces atlases that are reasonably robust to the initial choice of reference [80, 139].

3.4.1.3 Combining the global and local averages

Let $\bar{\mathbf{T}}'_{local}$ denote the free-form deformation generated by the average of the affine corrected control point vectors $\bar{\Phi}'$. The FFD $\bar{\mathbf{T}}'_{local}$ is not coupled to an affine transformation (or, equivalently, it is coupled with the identity matrix). It needs to be coupled with the previously calculated affine transformation, $\bar{\mathbf{A}}$, before being used for the complete (global and local) average transformation. The local component $\bar{\mathbf{T}}_{local}$ of the average complete transformation has control points $\{\bar{\Phi}_{u,v,w}\}_i$ given by

$$\{\bar{\Phi}_{u,v,w}\} = \{\bar{\mathbf{A}} \cdot \bar{\Phi}'_{u,v,w}\}$$

Finally, given the average affine matrix and average local deformation field, the average transformation $\bar{\mathbf{T}}$ of the input transformations $\{\mathbf{T}_1, \dots, \mathbf{T}_n\}$ is again expressed as a sum, i.e.

$$\bar{\mathbf{T}}(\mathbf{x}) = \bar{\mathbf{A}}\mathbf{x} + \bar{\mathbf{T}}_{local}(\mathbf{x})$$

3.4.2 Inverting transformations

The inversion of the global affine component of \mathbf{T}_{av} is straightforward, but the local FFD component is not analytically invertible. We used a numeric method to invert the FFD by starting a lattice of control points in the space of the reference image (which has the same position and orientation as the average space atlas). For each control point location (x, y, z) an estimate of the displacement $\mathbf{T}_{av}^{-1}(x, y, z)$ was obtained using a Newton method. Once all the displacements are obtained for the control point locations, a set of B-spline FFD components are calculated for each control point using the inverse filtering method described by Unser [167]. Further details of the method used to invert a transformation consisting of global affine and local FFD component can be found in Appendix A.

3.5 Discussion

The previous sections have presented a framework that encompasses a set of techniques for measuring longitudinal change. When measuring such change, there is a need for some form of validation or consistency estimation against which the estimates of change can be assessed.

Estimates of longitudinal growth or degeneration can be validated in a number of ways. Clearly, for a particular structure, it is possible to provide what are described as ‘gold standard’ estimates for segmentations of that structure at each timepoint. These could be provided by an expert clinician applying an appropriate manual segmentation protocol to the image data or, more invasively, by histological segmentation. If manual segmentations are generated for each time-

point, then an estimate of the structure’s longitudinal change can be made. This estimate should be highly accurate given the high levels of accuracy provided by expert manual segmentations.

The drawback of such a method is clearly the amount of time and effort needed to generate manual segmentations, whether on images or histologically. For assessing automatically generated volume change estimates for a large number of subjects and/or structures, this approach becomes impractical. A more straightforward approach to providing comparison estimates is therefore via automated image segmentation techniques, which are easier to apply on a large scale. In recent years, advances in automated segmentation have led to significant improvements in consistency and accuracy, in some cases reaching levels comparable to those of manual raters [88].

As well as providing a comparison for estimates of volume change, segmentation methods can be used to define regions of interest that can, in turn, provide regionally specific estimates of volume change.

Two segmentation methods were used to provide figures that could be compared and combined with estimates of volume change in the brain that were obtained from registration: Expectation maximisation and classifier fusion. Both techniques are briefly described in the following sections.

3.5.1 Tissue segmentation: Expectation maximisation.

It is possible to generate segmentations of the different brain tissues using expectation maximisation. Expectation maximisation (EM) is a well established method for segmenting structures in brain MR images [109, 123]. A typical application of EM assumes that the intensities of the different tissues within a brain image are modelled by a mixture of Gaussians. It is common to provide a set of prior probability maps for the tissues or structures being segmented and the EM algorithm is then used to find the maximum *a posteriori* (MAP) values of the parameters

for the Gaussian mixture.

The application of EM to brain images allows the generation of tissue probability maps for each subject in a cohort. An illustration of the tissue probability maps obtained by an EM segmentation is shown in Figure 3.7. If this is done at multiple timepoints, then estimates of change in each tissue can be obtained, either for individuals or for the cohort as a whole if atlases for the tissues are created by aggregating individual subjects' tissue maps.

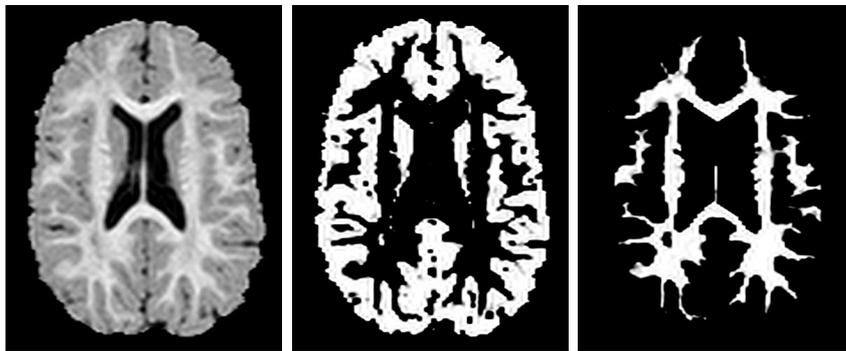


Figure 3.7: An illustration of tissue maps obtained by EM segmentation. Left to right: A transverse view of the anatomy of a subject; An estimate of the grey matter tissue density; A corresponding white matter estimate.

While EM based methods have a proven track record in classifying brain tissues within images acquired from healthy adults, challenges can be presented by images acquired from other types of subjects. For example, the brain tissue responses during MR scanning for children up to approximately two years differ from those of adults. This can confound EM based segmentation methods. Another example is given by the brain MR images of elderly subjects who have enlarged ventricles compared with the majority of the population. In such cases a typical set of tissue priors may not represent the individual well. It is, however, possible to improve the results of an EM segmentation by providing the algorithm with more specific prior probability maps of the tissues that are representative of the population being studied. For example, the prior maps can be generated by taking a manual atlas for a particular subject and propagating this atlas to each of the remaining subjects to act as an initial estimate for EM segmentation [123].

3.5.2 Structural segmentation: Classifier fusion

An alternative segmentation technique that can provide comparison estimates of volume change relies on the use of repositories of images that have been structurally labelled. An example of a set of manual labels for an anatomical image is illustrated in Figure 3.8. Labels from the repository can be propagated and

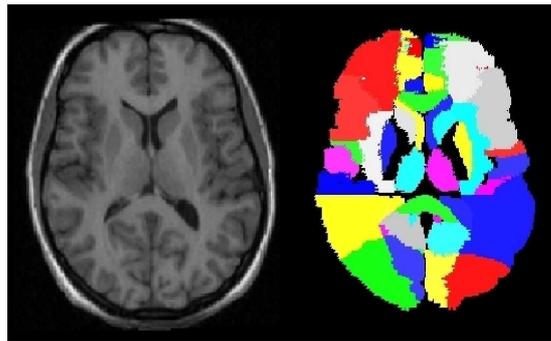


Figure 3.8: An example of manual labelling for an anatomic image. It is possible to segment query or unseen images using methods that rely on registration of anatomic images and subsequent propagation of manual labels.

fused or combined to provide a structural segmentation for a query image. This approach has been termed classifier fusion [135, 88]. Further details of classifier fusion will be given in Chapters 4 and 5.

3.6 Conclusion

This chapter has described the ways in which registration can be used to identify patterns of longitudinal change in a cohort of serially scanned subjects. Atlasing methods were described that can be used, for example, to generate atlases of anatomy or volume change, and to identify patterns of change for a cohort. The atlasing methods rely on the averaging of inter-subject registrations and create an average space representation of the cohort via a particular reference. Segmentation methods that can provide a benchmark for comparing registration based volume change figures were briefly described. Further details of structural segmentation through classifier fusion will be presented in Chapters 4 and 5.

Experiments based on the techniques given in this chapter will be presented in Chapters 6 and 7 using data from young children and a group of elderly subjects consisting of patients with Alzheimer's disease and healthy controls.

Chapter 4

Atlas-based segmentation, classifier fusion and classifier selection

4.1 Introduction

This chapter describes how atlas-based segmentation can be used to provide segmentations for unseen images given a manually labelled atlas. The process of classifier fusion, where labels from multiple atlases are propagated to an unseen image and combined is described along with its advantages over atlas-based segmentation using a single atlas. Finally, novel schemes for the selection of atlases from a large repository are presented. The selection of atlases, particularly when the size of the repository is large, can be used as an initial step prior to classifier fusion. As well as reducing the computational burden, selection of appropriate atlases for a given query can provide better quality segmentations.

Initially, background material is presented on ways in which the agreement of raters or automated classifiers can be measured. A review of the pattern classification context to classifier fusion is also given later in the chapter.

4.2 Measuring agreement of segmentations

When generating segmentations for an anatomical structure, a way of measuring the level of success is needed. If a gold standard segmentation is available and a particular method is used to provide an alternative segmentation (either manually or automatically) then the agreement of the segmentations gives a measure of the method's accuracy. Alternatively, estimates of a structure may be generated by two different segmentations, neither of which is considered a gold standard. The agreement of the two segmentations then provides a measure of precision.

Measuring agreement between structural segmentation methods falls under the more broad heading of assessing the agreement between raters and a rough distinction can be made among agreement measures depending on whether the raters provide numerical or categorical data.

4.2.1 Agreement on numerical data

If the rater(s) being assessed provide numerical measures then statistical approaches can be applied to provide measures of agreement. This can be done in the context of structural segmentation methods by applying the method and calculating the volumes of the extracted structures. This can be appropriate in cases where a structure's volume can be directly related to a pathology, for example the volume of the hippocampus in Alzheimer's. If the volumes for a structure are estimated by two raters then agreement can be assessed using any of a number of statistical techniques such as Bland-Altman plots [26], Pearson correlation or intra-class correlation [9]. Agreement based on extracted numerical measures such as volumes, however, ignores spatial information so it is possible for different raters to agree on the volume of a structure even if the corresponding segmentations are in widely differing locations. For the purposes of this work, agreement is assessed using methods that incorporate location information.

4.2.2 Agreement on categories

For pairs of raters that provide categorical classifications, the simplest case is a binary classification where each rater identifies an object as belonging to (1) or not belonging (0) to a particular class. A further distinction can be made depending on whether or not one of the raters can be considered to be a ‘gold standard’ measurement. For example, a non-invasive test could be used to predict the presence or absence of a condition which is later confirmed or denied by surgery. In such an example it is possible to categorise the positive and negative results as being ‘true’ or ‘false’. This is illustrated on the left in Table 4.1 (The presentation of the information based on classifications in this way is sometimes known as a ‘confusion matrix’). If one of the raters is considered a gold standard then measures such as sensitivity and specificity are commonly used:

$$\text{sensitivity} = \frac{TP}{TP + FN}, \quad \text{specificity} = \frac{TN}{FP + TN}.$$

i.e. sensitivity measures the chance that a positive test result is correct and specificity measures the chance that a negative test result is correct.

Alternatively, if neither rater is considered to be a gold standard measurement, then the remaining option is simply to count the number of instances of agreement or disagreement as shown on the right in Table 4.1 where, for example, the number of instances that the first rater assesses as belonging to the class and the second rater does not is denoted as n_{10} .

		Gold Standard	
		0	1
Rater	0	TN	FN
	1	FP	TP

		R_2	
		0	1
R_1	0	n_{00}	n_{01}
	1	n_{10}	n_{11}

Table 4.1: Agreement of two raters. Left: One of the raters is considered a ‘gold standard’ and the assignments of the second rater can be labelled as true and false positives and true and false negatives. Right: Both raters are on an equal footing and counts are made of the number of instances of agreement and disagreement on the class assignments.

MR image based structural segmentation methods can be viewed as falling into the latter category of rater; in most instances, a gold standard is unavailable and consistency across methods is the only available option. In this case, a structural segmentation method can assign 0 or 1 to a voxel to indicate membership of the structure. If the counts of agreement and disagreement shown in Table 4.1 are available then it is possible to calculate a measure of agreement known as the kappa statistic [43]. This is a measure of the observed probability of agreement p_o adjusted by the probability of agreement by chance, p_c ,

$$\kappa = \frac{p_o - p_c}{1 - p_c}.$$

It can be shown that the kappa coefficient can be expressed in terms of the agreement figures n_{xy} as

$$\kappa = \frac{2(n_{00}n_{11} - n_{01}n_{10})}{(n_{00} + n_{01})(n_{01} + n_{11}) + (n_{00} + n_{10})(n_{10} + n_{11})}.$$

If the raters described by Table 4.1 represent two methods for giving a structural segmentation where a value of 1 represents the presence of the structure, then the number of voxels for which both raters agree on the absence of the structure, n_{00} , will tend to dominate. Zijdenbos et al. showed [188] that under these circumstances, using the notation above,

$$\lim_{n_{00} \rightarrow \infty} \kappa = \frac{2n_{11}}{n_{01} + 2n_{11} + n_{10}}.$$

Assuming that raters R_1 and R_2 provide binary segmentation estimates for a structure, and making a change in notation, it is possible to represent the set of voxels labelled as part of the structure by R_1 as A and the set of voxels labelled as part of the structure by R_2 as B . This means that it is possible to identify n_{11} with $|A \cap B|$, $n_{11} + n_{10}$ with $|A|$ and $n_{01} + n_{11}$ with $|B|$ in the expression above to

obtain the following representation of the limiting value of the kappa coefficient

$$\lim_{n_{00} \rightarrow \infty} \kappa = \frac{2|A \cap B|}{|A| + |B|} = d$$

where d is the Dice coefficient [50] (also known as the similarity index).

Another simple measure of the overlap of two binary structures is given by the Tanimoto overlap (also known as the Jaccard similarity) which, for two segmentation estimates A and B , is given by the ratio of the overlap and the union

$$t = \frac{|A \cap B|}{|A \cup B|}.$$

Both the Dice and Tanimoto overlap measures give zero for no overlap and one for complete agreement and they can be shown to be related by the formula

$$d = \frac{2t}{1+t}. \quad (4.1)$$

This represents a monotonically increasing function on the interval $[0, 1]$ so that the ordering of agreement measures using either overlap is preserved. For the purposes of this study, agreement between labels will be reported using the Dice overlap.

If the segmentations produced consist of a series of categorical labels then it is possible to aggregate the overlap measures for each label by averaging. The method used by Heckemann et al. [88] provides a simultaneous segmentation of multiple structures for a query subject. If the query subject has gold standard labels available, Dice overlaps for all the structures can be calculated. Heckemann et al. then used the linear average of all these Dice overlaps to obtain a summary figure for the accuracy of the segmentation.

Crum et al. [45] also present a framework for aggregating overlaps over multiple structures. This framework introduces weights that can be associated with the overlaps for particular structures. For example it is possible to assign higher

weights to the overlaps calculated for smaller structures, on the principle that a good overlap for a small structure is less likely to be obtained. Crum et al. also allow for the possibility of more than two subjects, i.e. the aggregation can also be evaluated over all pairwise overlaps between subjects as well as over all structures studied. Finally, Crum et al. incorporate operations on fuzzy sets (see e.g. [53]) that allow overlaps and their aggregation to be implemented in cases where the segmentations produced are probabilistic or ‘soft’.

4.3 Atlas-based segmentation

A straightforward method of utilising a segmentation that is considered accurate, for example a manual segmentation created by an expert rater, is to use it to generate a segmentation for an unseen query image. For the purposes of this part of the thesis, the term ‘atlas’ shall be used to refer to a pair of images: a scan of the anatomy and a corresponding manual segmentation¹. An example of an anatomical image and its corresponding manual labelling is illustrated in Figure 4.1.

Given an atlas image pair, the MR image can be registered to the anatomical image of an unseen query subject. The resulting transformation can then be used to transform the atlas labels to the space of the query and the transformed labels can then be treated as a segmentation estimate for the query. The atlas that is propagated may represent a single segmented individual [93, 160, 49], or, alternatively, an average of multiple segmentations can be produced prior to propagation [134] or treated as a probabilistic prior for further segmentation [114].

¹The term ‘atlas’ has been used in earlier chapters to denote an average representation for a group of images. While this is ambiguous, the sense in which the term is used should be clear from the context.

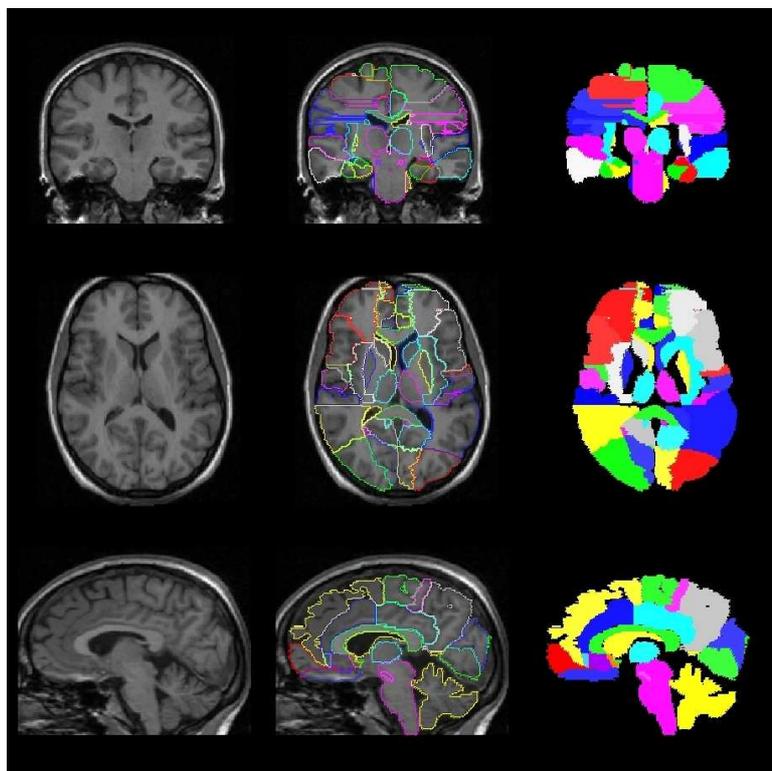


Figure 4.1: An example illustrating an anatomical image and a corresponding manual labelling. The first column shows the anatomy, the final column shows the manual segmentation of various structures. An overlay is shown in the middle column.

4.3.1 An example of atlas-based segmentation

In order to illustrate the process of atlas propagation, and to demonstrate the use of the types of agreement measure described in Section 4.2, an experiment was carried out based on hippocampal segmentation.

The data used for the experiment is a set of 30 T1-weighted MR brain images, each with a corresponding manual segmentation. Each image was manually segmented into 83 anatomical structures using a protocol published by Hammers et al. [83, 84]. The images represented scans of 30 healthy volunteers (15 women and 15 men). The median age of the women (men) was 31 (30) years and the range was 20-54 (20-53) years. The coronal T1-weighted 3-D volumes were obtained using a 1.5 Tesla GE Signa Echospeed scanner (GE, Milwaukee, WI). An inversion recovery FSPGR sequence was used ($TE = 4.2\text{ms}$, $TR = 15.5\text{ms}$) to give 124 1.5mm slices covering the whole brain with a voxel size of $0.9375 \times 0.9375 \times 1.5\text{mm}$. The images were resliced using sinc interpolation to give images with isotropic 0.9375^3mm voxels where the line between the anterior and posterior commissures (AC-PC line) defined the horizontal plane and the midline defined the sagittal plane. Further details describing the subjects and the MR acquisition can be found in [84].

For the purposes of this exemplar experiment, only the hippocampus labels were used. Two of the subjects' images were separated and treated as atlases. These were propagated to provide segmentation estimates for the hippocampuses of the remaining 28 subjects, which were treated as query subjects. Let the two atlases, which can also be viewed as classifiers, be denoted A and B. In order to provide a hippocampus estimate for a query, an atlas' MR image was registered to that of the query and the resulting transformation was used to propagate the atlas manual label to the query. In this way, two hippocampus estimates were generated for each of the 28 query subjects, one from atlas A and one from atlas B. An example of this is shown in figure 4.2 where the results of propagating the

hippocampus label from each of the two classifiers are shown overlaid on the query anatomy. Also shown is the manual label for the query subject for comparison.

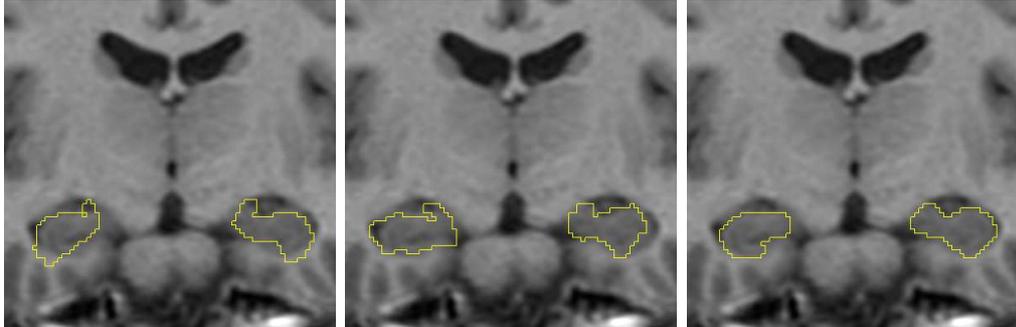


Figure 4.2: The left and middle images show propagations of hippocampus labels from two atlases to a query subject. The final image shows the manual hippocampus labels for the query subject for comparison.

In some studies, only the volumes provided by the classifiers are of interest. In this case, a simple comparison of the volumes provided by the propagation of atlases A and B can be achieved through a scatter plot of their resulting volume estimates (Figure 4.3, left). It is possible to discern a relative bias between the two classifiers in this figure: the estimates from classifier A tend to be consistently lower than those given by classifier B. Such a bias is more readily seen if the data are presented in a Bland-Altman plot, where the mean volumes given by the classifiers is plotted against their difference. Such a plot is given in Figure 4.3 (right) where the general disagreement between the two classifiers is shown by the high number of points that are below the horizontal axis corresponding to a difference of zero.

The Pearson correlation between the volume estimates provided by the volume measurements in Figure 4.3 is 0.76 and the intra-class correlation is 0.52, which would be generally considered a low estimate of agreement between the two classifiers.

Further aspects of agreement of the two classifiers can be explored if the gold standard of the target structure is used. This is represented by the manual hippocampus labels for the query subjects. For each estimate of a hippocampus

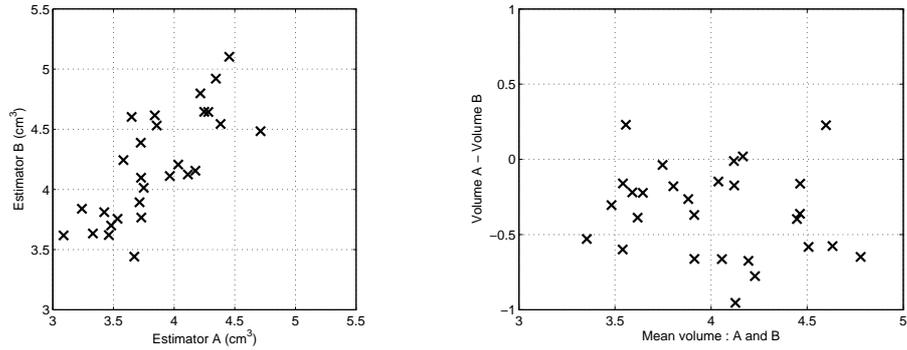


Figure 4.3: Left : A scatter plot comparing the volumes of hippocampuses in 28 subjects given by two classifiers based on atlas propagation, A and B. Right: A Bland-Altman plot of the data provided by the same classifiers. The bias of one classifier relative to the other is more apparent in a Bland-Altman plot.

label by classifiers A and B, its accuracy can be measured using the overlap with the query subject’s manual label. The agreement as measured by the Tanimoto and Dice overlaps are shown in shown in Figure 4.4

The overlap measures shown in Figure 4.4 provide specific information on the performance of each classifier for each of the individual query subjects. The relative performance of each classifier on an individual query subject is preserved irrespective of whether it is measured by Tanimoto or Dice overlap. This reflects the monotonic relationship between Dice and Tanimoto overlaps described in Equation 4.1.

The accuracy and agreement data for the classifiers A and B show that the different atlases can provide segmentations with different accuracy for the same subject and that the same classifier’s performance can vary depending on the query subject being segmented.

This reflects the types of error that can affect the process of atlas propagation. Errors may arise due to inaccuracies in the correspondence estimate between the atlas and the query and the process of transforming and reslicing the atlas to the target voxel lattice may also introduce error. Atlas-based segmentation can also be confounded by errors in the original atlas labelling or if the atlas used is anatomically unrepresentative of the query image to be segmented, for

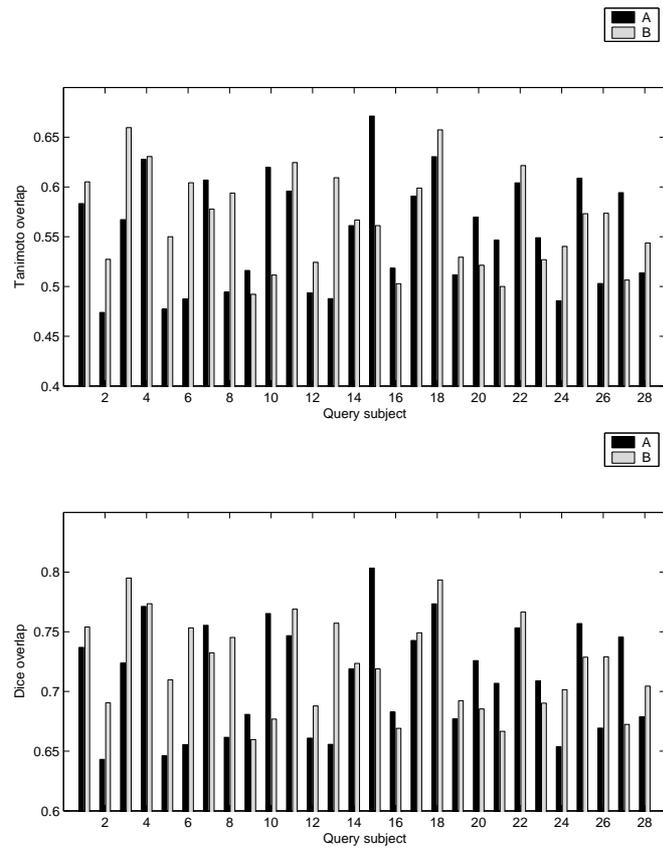


Figure 4.4: Tanimoto overlaps (Top) and Dice overlaps (Bottom) for the the hippocampuses segmentations of 28 subjects provided by propagation of atlases A and B.

example if they do not share the same topology. Errors in the original atlas cannot be removed even if the correspondence found with the query subject is highly accurate.

To help overcome the effect of various types of errors in the propagation of an individual atlas labelling, it is possible to propagate sets of labels from multiple atlases to the query. After propagation, they can be treated as separate classifiers and fused to form a single consensus segmentation estimate.

4.4 Classifier fusion

If a number of atlases are available, it becomes possible to propagate multiple sets of labels to the query subject. The propagated segmentations can then be treated as raters or classifiers and methods for combining or fusing the output from groups of classifiers can be applied. The main benefit of classifier fusion is that the effect of errors associated with any single atlas propagation can be reduced in the process of combination.

Such methods have an established history in the broader pattern recognition community; a good overview of different methods for classifier fusion is presented by Kittler et al. [106]. The number of possible schemes for combining or fusing classifiers are many and varied and depend on the nature of data being fused and the application context. The fusion can be viewed as taking place at different levels. The focus of this work is on the low-level fusion of classifiers that assign structural labels to individual voxels in an image and only the simplest fusion schemes will be used.

Mid level fusion can describe the process of fusing a feature or features extracted from low-level data. An example might be the averaging of volumes calculated from separate structural segmentations to provide a volume estimate for a particular structure. Fusion at a higher level might be exemplified by the use of features extracted from low-level data to provide multiple clinical assessments

of the same subject. The high level fusion would then combine these assessments into a single clinical label or diagnosis.

In the context of medical image segmentation, Warfield et al. [175, 174] present methods where classifiers are weighted prior to combination as part of an expectation maximisation framework (See section 4.4.2). A simpler approach is to treat the labels assigned to each voxel as ‘votes’ and to categorise the voxel according to a simple majority. Rohlfing et al. [135] used a database of images of bee brains to show that fusing segmentations according to a simple vote or majority rule is robust and accurate compared with, for example, producing segmentations from the propagation of an average shape atlas, or by propagating an individual atlas, selected according to its similarity to the query image. The vote rule has also been shown to perform well relative to other fusion approaches in a more general pattern recognition context [106].

The main advantage of propagating multiple segmentations and fusing them is in reducing the effect of errors associated with any single propagated segmentation. For example, if multiple atlases are registered to a query subject, then a correspondence error during propagation of the segmentation for a particular atlas is less likely to affect the final query segmentation estimate if it is combined with multiple other propagated segmentations. A schematic diagram illustrating the process of aligning a set of atlases to a query image, propagating their labels and fusing them is shown in Figure 4.5.

Heckemann et al. [89, 88] present a series of experiments using a set of atlases to investigate the precision and accuracy of structural brain segmentation based upon label propagation and classifier fusion. The atlases consisted of brain MR images for 30 subjects with corresponding manual segmentations created according to the protocol published by Hammers et al. [83]. An example of an anatomical image and a corresponding manual labelling from this data set was previously shown in Figure 4.1.

Adopting a ‘leave-one-out’ approach, it was possible to treat each of the

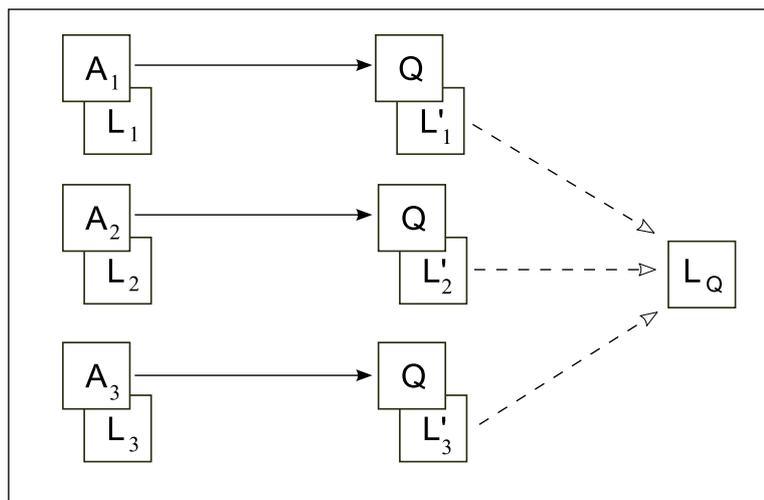


Figure 4.5: Schematic representation of segmentation based on label propagation and fusion. A set of atlas anatomical images A_i are registered to the query anatomy Q . Using the resulting transformations, the corresponding atlas labels L_i are propagated to the query. The transformed labels L'_i are then fused to create an estimate of the query labelling L_Q .

anatomical images as a query subject and to estimate its structural segmentation by label propagation and fusion, using some or all of the remaining subjects. The query subject's own manual segmentation could then be used as a gold standard with which to assess the accuracy of the estimated segmentation. Precision could be assessed by measuring the agreement of segmentations produced by different subsets of atlases.

It was shown that, as the number of classifiers increased, the precision and accuracy of the resulting segmentations increased (as measured by overlap with the manual segmentation). With the assumption that the Dice values are normally distributed, it was shown that the Dice overlap as a function of the number of classifiers was well modelled by the relation

$$d(n) = a - \frac{b}{\sqrt{n}}, \quad (4.2)$$

where n represents the number of atlases used as classifiers, d represents the Dice overlap of the resulting segmentation and the manual labels, and a and b are a pair of constants to be determined where $0 \leq a \leq 1$ and $0 < b$.

There is a caveat on the assumption of a normal distribution for the Dice overlaps. In theory, this is not possible given that the Dice coefficient is bounded between zero and one. However, tests for normality on Dice values obtained in practice, which are typically well away from the bounds, show that the normality assumption can still be reasonable [88, 89].

The relation in Equation 4.2 models overlaps as monotonically increasing for increasing numbers of classifiers up to an asymptotic limit determined by the constant a . The constant b determines the rate at which the overlap increases. As more classifiers are used, the overlap values increase because random errors, associated with individual atlas-to-query propagations, are increasingly cancelled out. There remains, however, a systematic bias that cannot be removed by using larger numbers of classifiers. This is represented by the difference between the asymptotic value a and one.

The systematic error, the size of which is represented by the parameter a , can arise for a number of reasons. For example, the anatomy of the query subject may represent a variation that is not represented by any of the atlases. Alternatively, there may be a limit to the correspondence accuracy that can be achieved: for example, the registrations may optimise the affine transformation only which may be insufficient for local small scale alignment.

4.4.1 An example of classifier fusion using the vote rule

In order to give an illustration of the accuracy properties of classifier fusion given by Equation 4.2, a second experiment was carried out on the same set of atlases used to illustrate atlas-based segmentation in Section 4.3.1.

Each subject in the group of 30 atlases was treated as a ‘leave-one-out’ subject and a number of the other subjects were randomly selected. The hippocampus labels from the randomly selected atlases were propagated to the left out subject and fused using the vote rule to obtain a classifier fusion segmentation estimate.

Voxels where there was a tied vote were decided randomly. The overlap of the segmentation estimate with the subject’s manual labels was then used as a measure of the segmentation accuracy.

This process was repeated ten times for each of the thirty subjects and by using different numbers of randomly chosen classifiers. For each number of classifiers, the Dice overlaps from the segmentation estimates across all subjects and repetitions were averaged. This gives an indication of how the accuracy of the segmentations varies according to the number of randomly selected atlases. This relationship is shown in Figure 4.6.

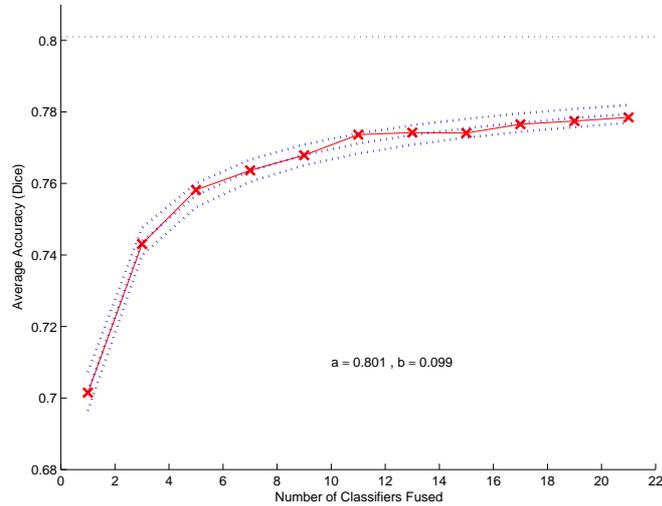


Figure 4.6: The average overlap achieved after repeated segmentations of hippocampuses using various numbers of random classifiers. The overlap is plotted against the number of classifiers fused. The central dashed line shows the Dice overlaps predicted by the model represented in Equation 4.2. The upper and lower dashed lines show the predictions based on the 95% confidence limits for the parameters a and b .

Each cross in Figure 4.6 represents the average accuracy of 300 segmentations (30 subjects \times 10 repetitions). The model parameters for the relationship between accuracy and the number of classifiers (Equation 4.2) were estimated based on the average overlap data. The dashed lines in the figure show the values predicted by this model and by the models obtained using the 95% confidence limits for the parameters a and b . It can be seen that the model predicts the behaviour well.

4.4.2 The pattern classification context

In a general context, pattern classification is applied to many different types of data such as speech, biometrics, astronomy, etc. A typical pattern classification model is described in Duda et al. [54] as comprising a series of different steps:

- Sensing
- Segmentation
- Feature extraction
- Classification
- Post processing

Atlas propagation and label fusion does not have an obvious place within such a schema although, at the simplest level, it is possible to argue that the process of propagating an atlas's labels to a query strictly belongs to the segmentation step in the list. Feature extraction and classification could then be interpreted as higher level steps, for example after a query subject's structures are segmented, geometric or volumetric features can be extracted and a classification of the subject to a clinical condition could then follow.

Such a view does not, however, fully describe the process of propagating multiple label sets and fusing them and it does not fully represent the role of finding geometric correspondence (registration).

If a propagated label set is considered to be a classifier in the sense of the above list, then it can be considered as a spatial function in formal terms. Let there be M atlases available, $\{A_1, \dots, A_M\}$ and N query subjects to be segmented, $\{Q_1, \dots, Q_N\}$ and let there be K classes or structures $\{c_1, \dots, c_K\}$ (including background) to be delineated in the query images. The particular combination of atlas A_i and query Q_j gives rise to a classifier function h_{ij} that is dependent on A_i , Q_j and the estimate of correspondence between them. The classifier function,

h_{ij} , can be viewed as a vector valued function

$$h_{ij} : \mathbf{x} \rightarrow (p_1, \dots, p_K)^T$$

where p_1, \dots, p_K are the spatial membership probabilities that h_{ij} assigns to the query image voxel at location \mathbf{x} for each of the K classes, i.e. $\sum_{i=1}^K p_i = 1$.

Viewing a classifier in this way implies that, in the context of the list above, the feature it is operating on is simply a spatial location of the query image.

Making the spatial dependency of the membership probabilities explicit gives the classifier the following form

$$h_{ij}(\mathbf{x}) = (p_1(\mathbf{x}), \dots, p_K(\mathbf{x}))^T.$$

The dependence of the membership probability functions on the combination of atlas A_i and query subject Q_j can also be made explicit by adding a superscript

$$h_{ij}(\mathbf{x}) = (p_1^{ij}(\mathbf{x}), \dots, p_K^{ij}(\mathbf{x}))^T \quad (4.3)$$

although this will be dropped if the meaning is clear.

Such a definition allows h_{ij} to describe a probabilistic or ‘soft’ classifier for the structures in the query. For a hard labelling, the one of the components p_1, \dots, p_K will have a value of one and the remainder will be zero. In this case, the vector $(p_1, \dots, p_K)^T$ represents a basis vector in K -dimensions. If the component with a value of one is indexed by s where $1 \leq s \leq K$ then this basis vector can be denoted e_s which means h_{ij} can be re-written as

$$h_{ij} : \mathbf{x} \rightarrow e_s \quad , \quad 1 \leq s \leq K. \quad (4.4)$$

In practice, a soft classifier can be obtained from a set of labels propagated to a query image through the use of an interpolator. For example, partial volume

interpolation [115] weights can be used to generate the vectors $(p_1, \dots, p_K)^T$ at each query voxel location. A soft classifier can be converted into a hard classifier by finding the maximum component among the p_1, \dots, p_K for each voxel. Alternatively, nearest neighbour interpolation can be used when propagating the labels to the query image.

After all the atlases have been propagated to query Q_j , we obtain M classifiers h_{1j}, \dots, h_{Mj} . These are, in turn, combined to give a final classification result. Let c_t represent the final label assignment, where $c_t \in \{c_1, \dots, c_K\}$. The classifiers can be combined using a variety of rules, examples of which include the sum rule, the median rule, the max rule and the majority vote rule:

Sum rule

Using the same notation as that used in Equation 4.3, the sum rule involves assigning label c_t to the voxel at location \mathbf{x} where

$$t = \operatorname{argmax}_{s \in \{1, \dots, K\}} \left(\sum_{i=1}^M p_s^{ij}(\mathbf{x}) \right). \quad (4.5)$$

The sum rule can be viewed as a type of majority vote rule where the individual classifiers make fractional votes for each label.

Median rule

The sum rule may also be described as the mean rule because the result of combining classifiers according to Equation 4.5 does not change if it is normalised by the number of classifiers, i.e. if c_t is determined by

$$t = \operatorname{argmax}_{s \in \{1, \dots, K\}} \left(\frac{1}{M} \sum_{i=1}^M p_s^{ij}(\mathbf{x}) \right).$$

This can be extended to the median to produce a new rule where label c_t is assigned to the voxel at location \mathbf{x} where

$$t = \operatorname{argmax}_{s \in \{1, \dots, K\}} \left(\operatorname{med}_{i=1}^M p_s^{ij}(\mathbf{x}) \right). \quad (4.6)$$

In other words, the median probability value (across all contributing atlases) is calculated for each of the classes and the class with the highest median probability is assigned to the voxel. The use of median averaging in this rule means that outlier probability values have a reduced impact on the final classification compared with the mean or sum rule.

Max rule

The max rule assigns label c_t at location \mathbf{x} where

$$t = \operatorname{argmax}_{s \in \{1, \dots, K\}} \left(\max_{i=1}^M p_s^{ij}(\mathbf{x}) \right). \quad (4.7)$$

This rule assigns the class corresponding to the highest probability value across all classifiers and classes. Such an assignment is, in a sense, diametrically opposed to the assignment given by the median rule above where the effect of high probability values is reduced.

Majority vote rule

If the h_{ij} represent hard classifiers as described in Equation 4.4, then the expression

$$\sum_{i=1}^M h_{ij}(\mathbf{x})$$

provides a K -vector at each voxel location where the components represent the votes for each of the labels $\{c_1, \dots, c_K\}$. The majority vote rule can then be expressed as : Assign label c_t at location \mathbf{x} where

$$t = \operatorname{argmax}_{s \in \{1, \dots, K\}} \left(e_s \cdot \sum_{i=1}^M h_{ij}(\mathbf{x}) \right). \quad (4.8)$$

These rules are described in a Bayesian context by Kittler et al. [106] who showed that the sum rule outperforms the others based on a series of experiments. It is worth noting that the majority vote rule can be viewed as a special case of the sum rule, where the classifiers provide a hard labelling at each voxel.

Other more complex rules and algorithms for combining segmentation estimates are clearly possible and are considered beyond the scope of this work. A notable example is the STAPLE framework described by Warfield et al. [175]. This framework treats multiple classifiers as observed estimates of a postulated (and unobserved) true estimate of the segmentation of a structure. The framework includes sensitivity and specificity parameters that measure the performance or reliability of each classifier with respect to the postulated true segmentation.

The reliability measures of the classifiers can then be taken into account when combining their segmentations in order to estimate the true segmentation. The classifiers each assign probability for a given label to be present at a given location. The overall probability estimate of the label at a voxel is given by a weighted average of the probabilities assigned by the different classifiers with the weight for each classifier determined by its reliability measure. An expectation maximisation approach is applied to iteratively update the parameters describing classifier performance and the segmentation estimate of the label or labels being segmented. Warfield et al. extend this process to apply to more general scalar fields associated with the segmentations produced by different classifiers, for example in combining distance transforms obtained from multiple binary segmentations [174].

4.4.3 An example of classifier fusion : multiple fusion rules

The following sections present experimental data to give a demonstration of the different rules for fusing classifiers described above. The accuracy of the different rules is assessed and implementation issues are also discussed.

4.4.3.1 Comparison experiment

Using the same 30 manually labelled images described in the examples of atlas-propagation and vote rule classifier fusion above (Hammers et al. [83]), an experiment was carried out to assess the relative performance of the different simple fusion rules described in Section 4.4.2.

Once again, each subject was treated as a query subject on a leave-one-out basis and the propagated hippocampus labels for all the remaining subjects were then used to provide an estimate of the query subject's hippocampus. The query subject's manual labels were then used to assess the accuracy of each estimate using the Dice overlap measure.

The 30 T1-weighted MR images were all registered to each other in a pair-wise manner. This allowed, for a given query subject, the hippocampus labels for the remaining 29 subjects to be propagated to the space of the query. The propagated label sets were then fused according to the different classifier combination rules described in Section 4.4.2 and the Dice overlaps with the query's manual labels found for each rule.

The results after treating each of the subjects as a query in turn are summarised in Figure 4.7. In this figure, four bars represent the performance of each of the max, mean, median and vote rules in providing a hippocampus estimate for each subject. It can be seen that the max rule performs worse than the other rules on a consistent basis. There is also little to separate the other three rules (mean, median and vote) in terms of the accuracy that each one achieves. The average (S.D.) Dice coefficients across all subjects for each of the fusion methods used were: Max 0.688 (0.047), mean 0.779 (0.034), median 0.780 (0.033) and vote 0.780 (0.033).

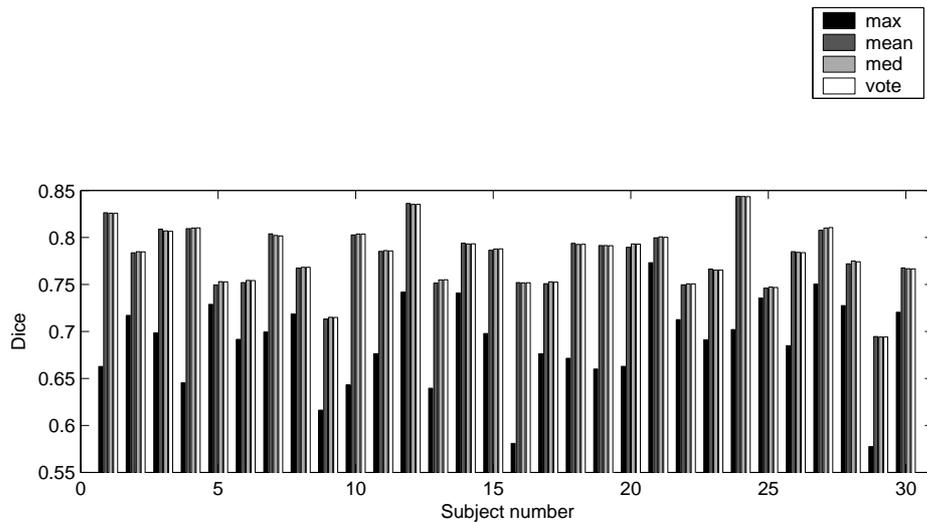


Figure 4.7: A comparison of the different fusion rules applied to the segmentation of the hippocampus for 30 subjects on a leave-one-out basis. The horizontal axis shows the number of each subject and each group of bars represents the Dice scores achieved by the different fusion rules in estimating the subject’s hippocampus. The subject’s manual label was used as a gold standard to assess the accuracy of each segmentation. For each subject, the hippocampus labels for all remaining subjects were propagated and fused using different rules. The shading of each bar indicates the rule used.

4.4.3.2 Implementation issues

In practical terms, and in the context of image segmentation using propagated labels, the computational costs associated with the different fusion rules vary. An implementation of the vote rule requires a single label per classifier to be associated with each query image voxel prior to fusion. For classifiers containing labels for multiple structures, this means that an atlas can be propagated to the query in a single transformation step using nearest neighbour interpolation. An implementation of the mean or median rule requires a vector of probabilities per query voxel for each classifier prior to fusion. This implies that a separate transformation step needs to be carried out for each structural label in the propagated label sets. A binary mask is needed for each label and this needs to be propagated using an interpolation scheme that can represent fractional membership of the propagated structure (linear interpolation, for example).

After propagation of the labels from each of the classifiers, the vote rule re-

quires the counts of each label to be evaluated prior to the final decision. The mean rule require the calculation of the mean probability for each label for each voxel. Using the resulting vector of mean probabilities, a final decision can then be made. The computational cost of the median rule will generally be higher than for the mean rule because a the probabilities of the labels need to be sorted before the final decision is made.

As these aspects of the implementation of the fusion rules suggest, the computational cost associated with the propagation of labels and subsequent fusion is lower for using the vote rule than it is for the other rules discussed in this section. This relative ease of computation, combined with its comparable performance in terms of segmentation accuracy makes it the preferred choice for this work.

4.5 Classifier selection

As increasing numbers of MR images have steadily become more available over recent years, the creation and maintenance of databases or repositories of atlases consisting of MR images with corresponding reliable structural segmentations (manual or otherwise) has become more feasible. A good example is the Internet Brain Segmentation Repository [31]. It has therefore been a natural consequence to use such expert annotations to assist in providing automatic segmentations of query or unseen images.

As described in the previous chapter, the propagation of manual segmentations and their fusion provides a reliable method for obtaining structural segmentations in a query image. There are, however, issues that arise when the available repository becomes large.

4.6 The motivation for classifier selection

The size of the atlas repository can have an effect on the steps involved in the propagation and fusion of labels as well as the quality of the final segmentation.

The results of Section 4.4.1 show that the average segmentation accuracy achieved by fusing random groups of classifiers increases monotonically with the number fused up to an asymptotic value for large numbers of classifiers. A natural question is whether the asymptotic level reached by fusing large numbers of random sets of labels is the best that can be achieved. Is it possible to equal this accuracy, or even exceed it, with smaller numbers of selected classifiers (as opposed to random classifiers)?

One type of problem that may be encountered with the fusion of large numbers of classifiers relates to variation across the population of the structure(s) being segmented. This variation may mean that the same structure may have qualitatively different variants across the population. In these circumstances, choosing increasing numbers of classifiers will result in a segmentation that is close to the mean shape of the whole population and yet the query subject (and others with the same structure variant) may not be well represented by such a mean shape. If a structure is manifested in more than one variant, then, for a given query subject, only a proper subset of the atlases in the repository is appropriate to use – atlases from subjects who occupy the same cluster as the query with respect to the variants of the structure. In a more general context, some work has been carried out to apply clustering techniques to anatomical MR images [17] but this remains at an early stage.

Anatomical variations can also affect the registrations used to generate correspondence between atlas and query. If the query subject anatomy is dissimilar to that of an atlas, then it is possible for the registration to fail in regions where the anatomies differ, especially if the anatomies have differing topologies. The converse is also true, a subject that is anatomically similar to an atlas will be

well segmented by propagation not simply because they share anatomical features, but also because these shared anatomical features lead to a better correspondence estimate from registration.

Another consequence of using a large number of classifiers is that the resulting segmentation is more likely to be a smooth shape because it will approximate the mean of a large sample where small scale differences are averaged out. The structure being segmented may, however, be badly represented by a smooth shape, and may be better represented with a shape having higher curvature.

For these reasons, the propagation of a selection of classifiers that are appropriate for the query subject would appear preferable to the selection of an arbitrarily large number of classifiers. Wu et al. investigated methods for optimal selection of a single template for atlas-based segmentation [182]. The work described in this chapter contrasts with this by covering methods for the selection of multiple atlases for subsequent fusion.

On a practical level, the propagation and fusion of a large number of classifiers to a query may represent a prohibitive computational burden. If the alignment step is carried out via a registration, the number of registrations needed increases linearly with both the number of atlases and the number of query subjects. Therefore, a method for selecting classifiers from the repository becomes a desirable option when the size of the repository becomes large.

4.7 Example data for classifier selection

To illustrate the benefits of classifier selection, five subjects from the dataset described earlier (Sections 4.3.1, 4.4.3, 4.4.1) were chosen to represent a repository and all proper subsets of this small repository were used to provide a segmentation estimate for the hippocampus of a sixth subject which is treated as the query. A small number was chosen for the repository in order to easily enumerate all subsets of classifiers.

For subsets consisting of a single classifier the segmentation estimate is obtained by simply propagating its hippocampus label to the query subject. Where a subset consisted of more than one classifier the propagated labels were fused using the vote rule with tied voxels being decided randomly.

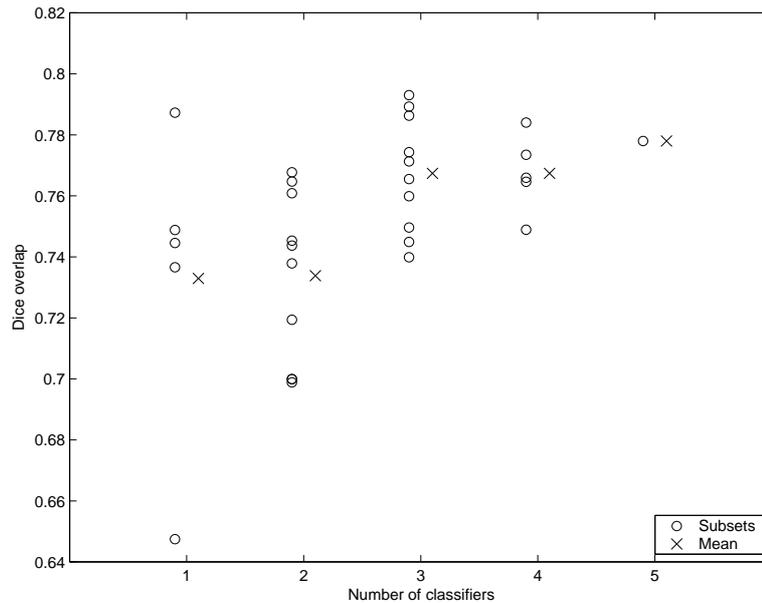


Figure 4.8: The circles show overlap accuracy values of segmentations derived by propagating fusing all 31 proper subsets taken from a small repository of five atlases. The query subject was fixed throughout. The size of each fused subset is shown on the horizontal axis. The mean overlaps for all subsets of the same size are shown by \times symbols.

Figure 4.8 shows that, while the general trend (represented by the mean values) is for accuracy to improve with increasing numbers of fused classifiers, there can be subsets of classifiers smaller than the whole repository that can perform very well. Indeed, for this example, one of the classifiers performs better as a single propagated atlas than does the fusion of all five propagated atlases (see top left of Figure 4.8).

This small example serves to show that, for a given query subject and a given structure, there may be merit in restricting the classifiers used in a fusion scheme to a subset selected from the repository.

4.8 Methods for classifier selection

Theoretically, the identification of an optimal selection of classifiers for a query is possible by carrying out an exhaustive search. Given a repository of M atlases, this would entail the use of each of the $2^M - 1$ proper subsets of the repository as classifiers for a particular query. The resulting fused segmentations would then need to be evaluated according to some criterion and the classifier subset giving the best performance could then be identified. Leaving aside the question of how the segmentations could be evaluated, such a scheme is clearly impractical for a large repository and alternative, more heuristic approaches would be preferable.

In the context of label propagation and fusion, a straightforward approach would be to create a method for assessing each of the atlases in a repository in terms of some measure of ‘suitability’ or ‘fitness’ as classifiers for a given query. If the fitness of the classifiers can be numerically represented then this allows the repository to be ranked with respect to the given query. The classifier selection can then be made by choosing a particular number of the top-ranked classifiers.

An obvious approach to ranking the repository atlases based on the image information they contain is to evaluate an image similarity metric between each atlas and the query. Alternatively, information pertaining to the subjects themselves, i.e. meta-information can be used to rank the repository atlases. For example, if the query belongs to a particular clinical group, then atlases in the repository in the same clinical group might be selected. Clearly this latter approach requires that the meta-information be available for both the repository and the query subjects.

Approaches to classifier selection based on image similarity will be described in Section 4.8.1 and approaches for selecting atlases based on the use of meta-information are described in Section 4.8.2.

4.8.1 Selection using similarity information

The atlases in a repository can be ranked in terms of the similarity of each of the atlas images to the query image according to a chosen similarity metric. Selection of classifiers based on image similarity requires an estimate of the spatial correspondence between the atlases and the query – i.e. a registration step of some form is required.

The rationale for using image similarity to rank the classifiers is that the success of a classifier fusion method in generating a segmentation is to a great extent dependent on obtaining a good correspondence estimate between each atlas and the query. The use of similarity gives an estimate of the success of a registration and hence the potential suitability of an atlas as a classifier. The use of a similarity metric for registration needs to be distinguished from its use for selection. For registration, the similarity metric is used to determine the optimal transformation parameters in an iterative fashion. For selection purposes, the similarity metric is evaluated once *post hoc* on the query and atlas images. Furthermore, the similarity metric used for selection does not need to be the same metric that was used for registration.

When defining an image similarity based method to rank the classifiers, a number of considerations need to be taken into account:

- Registration type: The registrations applied can be affine or non-rigid, and non-rigid registrations can be carried out either at a fine or a coarse scale. It is also possible to make the distinction between registrations carried out for selection purposes and registrations carried out for label propagation prior to fusion.
- Registration target: While the ultimate target for a segmentation is the space of the query subject image to be segmented, it is possible to use an intermediate space for selection or propagation. For example, a fixed reference image can be chosen and the atlas images, along with the query image,

can be spatially normalised to the reference. Combinations of selection and fusion can take place using images registered to a reference or to the query's native space. An advantage of registrations between the repository images and the reference is that they can be carried out 'off-line'.

- Selection similarity metric: Clearly a number of image similarity metrics are available when evaluating the similarity between an atlas and the query. These were described in Section 2.2.8.
- Choice of region of interest (ROI): The similarity metric can be evaluated using the whole of the region where the query image overlaps with an atlas. Alternatively, a ROI can be defined, and the calculation of similarity can be restricted to voxels within the ROI. This is appropriate if the structure for which a segmentation is required is small relative to the field of view of the whole image, for example the hippocampus.

Given the various considerations listed above, there are many possible experimental designs that could be defined for the selection of classifiers based on image similarity. Focus will be restricted, however, on five particular experimental arrangements of registrations, selections, propagations and fusion that capture the main aspects of the considerations listed above. Each arrangement will be described as a 'scheme' generally, and as an 'image similarity based scheme' specifically, given the use of similarity metrics during the selection stage:

- Query centric, one stage (Q1)
- Query centric, two stage (Q2)
- Reference centric, one stage (R1)
- Reference centric, two stage (R2)
- Hybrid (H)

4.8.1.1 Query centric, one stage – Q1

This selection and fusion scheme represents a minor modification to the simple propagation and fusion of labels described by Figure 4.5, Section 4.4. Under the Q1 scheme, all atlases in the repository are registered directly to the query image. This is illustrated schematically in the left hand diagram of Figure 4.9 where solid lines represent registrations.

After this first registration step, the similarity between the aligned repository images and the query is evaluated and used to rank the repository images. The top-ranked images from the repository are then selected and, using the transformations obtained from the registrations, their labels are propagated to the query and combined. This is illustrated in the right hand diagram of Figure 4.9 where the dashed lines indicate the propagation of labels.

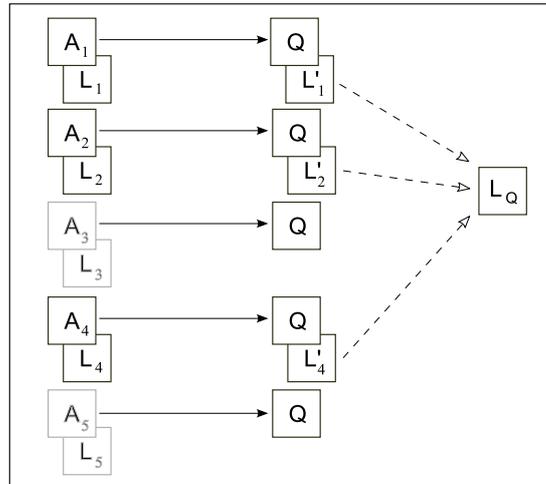


Figure 4.9: Query centric, one stage – Q1. All the repository anatomical images A_i are registered to the query subject Q . The similarity between each of the aligned anatomies and the query is used to decide their ranks. The labels L_i for the top-ranked atlases are selected and propagated. The resulting transformed labels L'_i are fused to create the query label estimate L_Q .

4.8.1.2 Query centric, two stage – Q2

This scheme represents a variation of the Q1 scheme where the registrations of atlas images to the query take place in two stages. An initial registration is

made to the query followed by selection of the atlases. After selection, a second registration is carried out that maps the selected atlases to the query. An affine or coarse non-rigid transformation can be optimised during the initial registration prior to selection and a ‘finer’ transformation with a higher number of parameters can be optimised after selection.

After the second registration step, the labels from the selected atlases are propagated to the query and fused. The Q2 scheme is illustrated in Figure 4.10.

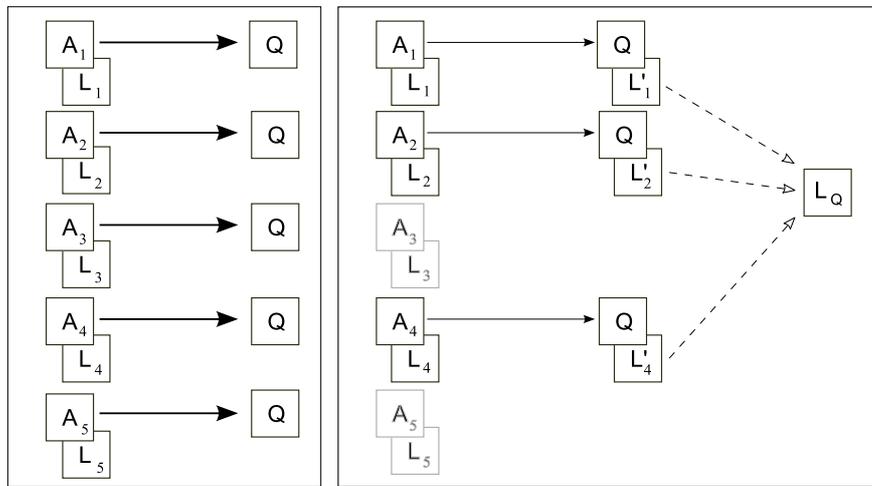


Figure 4.10: Query centric, two stage – Q2. Left: All the anatomical images A_i in the repository are registered to the query image Q . The similarities of the aligned images are used to rank the atlases. Right: The top-ranked atlases undergo a second, finer, registration step to the query and their corresponding labels L_i are propagated. The transformed labels L'_i are fused to estimate the query labelling L_Q

4.8.1.3 Reference centric, one stage – R1

A scheme based on registrations to a reference image represents an easing of the computational burden compared with query centric schemes. In the R1 scheme, all the atlas images are registered with a single reference image. When a segmentation is required for a given query subject, the query image is also aligned with the reference. The similarity measurements are then made between the atlas images and the query image after alignment in the space of the reference image. The resulting similarity values are then used to rank the repository atlases and

the labels for the top-ranked atlases are propagated to the reference space and fused. This scheme is illustrated in Figure 4.11. The resulting estimate of the query segmentation is in the space of the reference image. Depending on the application, this may be sufficient but a further transformation step may be required to map the query segmentation back from the reference to the original query image.

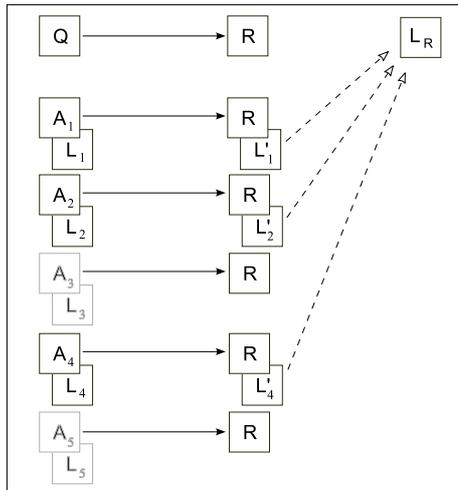


Figure 4.11: Reference centric, one stage – R1. The atlas anatomic images A_i , along with the query image Q , are registered with the reference image R . The similarities between the reference-aligned query and each of the reference-aligned atlas anatomies are used to rank the repository. The labels from the top-ranked atlases L_i are transformed to the space of the reference generating L'_i . The propagated labels are fused to generate a reference space estimate L_R of the labelling of the query subject.

4.8.1.4 Reference centric, two stage – R2

The reference centric, two stage scheme represents a modification of the R1 scheme that mirrors the modification of scheme Q1 to become Q2. The R2 scheme begins with an initial registration of all the repository atlases to a fixed reference image. The query image is also registered to the reference. After the initial registration step, the similarities between the aligned query and each of the aligned atlas anatomies are used to rank the repository. The top-ranked atlases are then registered a second time to the reference and the resulting transformations are

used to propagate their label sets. The second registration step can optimise a finer, higher dimensional transformation than that carried out in the first. As with the R1 scheme, the labels of the selected atlases are then propagated to the reference and fused. This scheme is illustrated in figure 4.12.

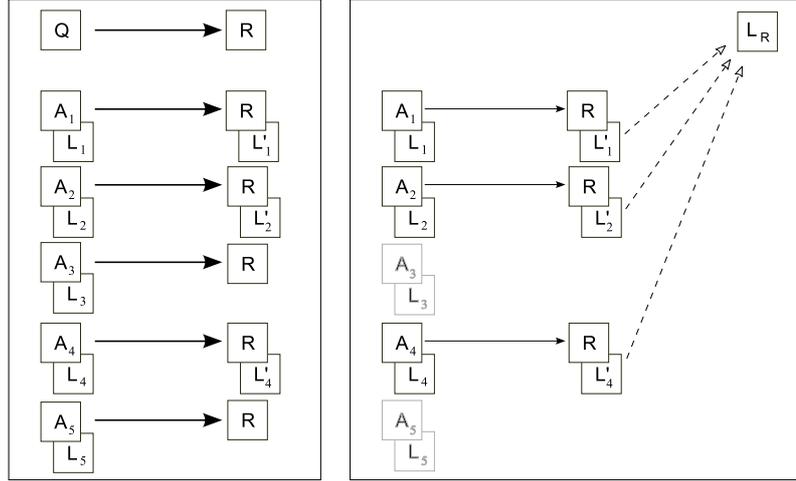


Figure 4.12: Reference centric, two stage – R2. Left: All the atlas anatomic images A_i and the query image Q are registered to the reference image R . The similarities between the aligned query and each of the aligned atlas anatomies are used to rank the repository. Right: The top-ranked atlases are selected and undergo a second registration step. Their labels L_i are then propagated to the space of the reference image. The transformed labels L'_i and are fused to generate a reference space estimate L_R of the labels for the query.

4.8.1.5 Hybrid selection and fusion – H

The hybrid scheme combines aspects from both reference centric and query centric schemes. Under the hybrid scheme, all the repository atlases are registered with the reference image. The query image is also registered with the reference. Similarity measurement and selection is made based upon the repository and query images after alignment to the reference as described in schemes R1 and R2. The selected repository images are then registered directly to the query image. The resulting transformations are used to propagate the labels for the selected atlases direct to the query where they are combined. The hybrid scheme is illustrated in Figure 4.13.

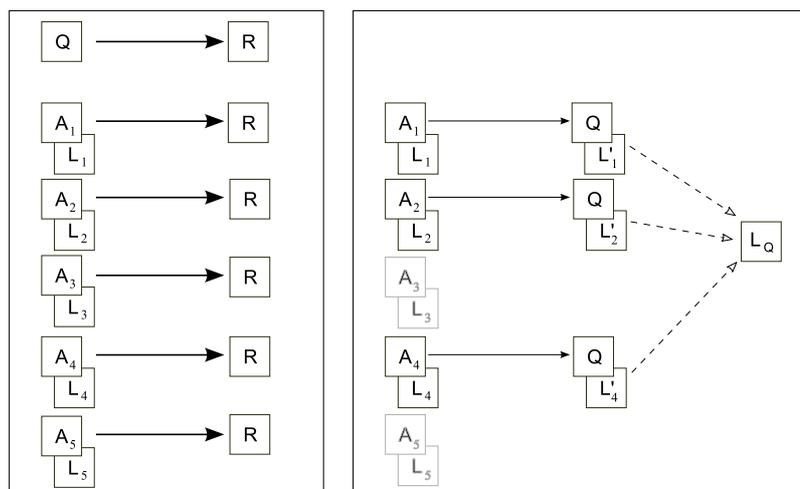


Figure 4.13: Hybrid classifier selection and fusion – H. Left: All the atlas anatomic images A_i and the query image Q are registered to the reference image R . The similarities between the aligned query and each of the aligned atlas anatomies are used to rank the repository. Right: The top-ranked atlases are selected and registered to directly to the query subject. The labels L_i of the selected atlases are propagated to the query. The propagated labels L'_i are fused to generate an estimate L_Q of the query image labelling.

4.8.1.6 The computational burden

The majority of the computation required for label propagation and fusion is taken up by the need to register images and in order to describe the computational burden associated with the various schemes described above, the distinction needs to be made between ‘on-line’ and ‘off-line’ registrations.

For a scheme such as Q1, none of the registrations can be carried out until the query subject is identified; all of the registrations therefore need to be done on-line. For the R1 scheme, only the registration between the query and the reference needs to be done on-line; the registrations of the atlas subjects can all be carried out beforehand (‘off-line’), which may represent a practical saving in computational terms.

A scheme such as Q2 requires all registrations to be carried out on-line but the computational burden may be eased by the fact that the initial registrations have fewer parameters to optimise, making them faster, while the registrations with the greater number of parameters are only carried out for the selected atlases. A

similar comment applies to the R2 scheme.

In terms of computational requirement, the hybrid scheme represents a compromise between the most demanding query centric schemes and the least demanding reference centric schemes.

If the number of atlases in the repository is M , the number of query subjects is N and the number of selected classifiers propagated and fused for each segmentation is S , then the number of off-line and on-line registrations for each scheme is summarised in Table 4.2.

registration cost	Off-line		On-line	
	low	high	low	high
Q1				MN
Q2			MN	SN
R1		M	1*	
R2	M	M	1*	
H	MN*			SN

Table 4.2: An outline of the computational costs of the different similarity-based selection schemes. There is a distinction made between coarse or affine registrations that are described as ‘low cost’, and finer non-rigid registrations that are described as ‘high cost’. Registrations are also divided into those that can be carried out off-line and those that need to be carried out on-line, i.e. once the query subject’s image is given. Asterisks indicate a choice of whether to carry out low or high cost registrations.

4.8.2 Selection using meta-information

It is fairly routine for general and clinical information to be collected from subjects during scanning, especially in large scale studies. Such information includes gender, age, handedness, information relating to clinical conditions, scores in cognitive tests, etc. This will be referred to as meta-information. If meta-information is also available for a query subject, then the atlases in the repository can be chosen according to how well the corresponding subjects match the query subject on some aspect of the information. For example, the repository subjects who are closest in age to the query can be selected for classifier fusion.

One advantage of selection based on meta-information is that it can be carried out independently of the image data and no pre-processing or alignment of the images is required, which represents a reduction in computation. Additionally, the selection of classifiers by meta-information may result in better quality segmentations – a query subject with a specific clinical condition may be better segmented by classifiers sharing the condition. Selection and fusion based on meta-information is illustrated in Figure 4.14.

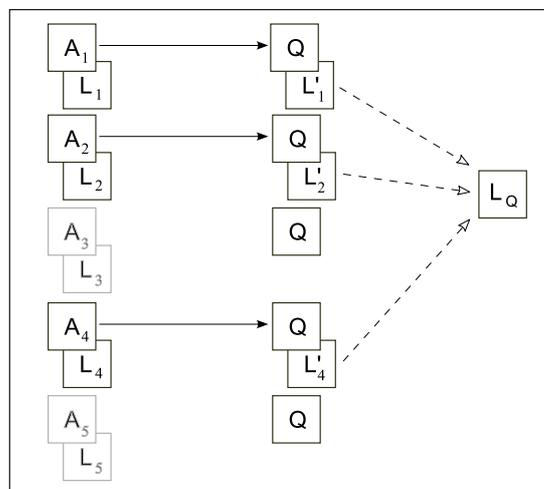


Figure 4.14: Meta-information selection – M. A subset of the anatomical images A_i in the repository is selected and registered to the query image Q . Selection takes place based on meta-information such as age or condition. The labels L_i for the selected atlases are transformed to the query. The propagated labels L'_i are fused to create an estimate L_Q of the query labels.

4.9 Conclusion

This chapter has presented an overview of methods that can be used to provide structural segmentations in images. The ability to isolate specific regions is useful in the context of assessing longitudinal change as it can be used to identify regional patterns of growth or degeneration.

Methods of assessing the agreement and accuracy of segmentation methods have also been discussed.

The methods presented are based on the use of registration to propagate

manual segmentations from atlases to a target query, followed by the use of a simple rule to combine the segmentations. Various simple rules for classifier combination have been presented, discussed in a pattern recognition context and compared in terms of performance.

The use of large repositories of labelled atlases raises particular issues with respect to segmentation methods based on label propagation and fusion and motivates the idea of classifier selection. Different classifier selection schemes have been presented and described which are based on image similarity or on meta-information. The next chapter will present data from experiments aimed at appraising the different classifier selection schemes.

Chapter 5

An assessment of classifier selection schemes

5.1 Introduction

The previous chapter presented descriptions of different methods that can be used to select classifiers from a repository as part of the process of providing a structural segmentation for a query image. After selection, the scheme used to generate the segmentation can vary, but attention will be restricted to the propagation of labels and their fusion via the vote rule, a method that has been shown to be robust and accurate [88, 89, 135, 106].

The two main categories for selection of classifiers are selection based on image similarity and selection based on meta-information. A further subdivision of image similarity selection described schemes that were query centric, reference centric or hybrid and with one or two stages of registration. Although the number of possibilities can increase rapidly, the particular set of selection schemes described in Section 4.8 will be the focus of this chapter.

This chapter presents data from a set of experiments that aim to establish whether classifier selection is useful and what might constitute a good implementation. The different parts of the chapter address different aspects of the selections

schemes. Section 5.2 describes the data used in the various experiments using the selection schemes. Section 5.3 describes the process used to define the region of interest. Section 5.4 assesses how the different selection schemes compare with each other. Section 5.5 makes an assessment of the suitability of image similarity as a selection criterion for classifiers prior to fusion. Section 5.6 presents stochastic experiments to identify how image similarity selection performs relative to the segmentations provided by randomly selected classifiers. The effect of the choice of the selection similarity metric is tested in Section 5.7. Section 5.8 assesses how the quality of resulting segmentations is affected by the fusion of different numbers of labels drawn from the ranked set of classifiers. Finally, a comparison between image similarity selection and selection based on age as available meta-information is made in Section 5.9.

5.2 Data

T1-weighted MR brain images acquired from 275 male and female subjects aged between 4 and 83 years were made available by the Centre for Morphometric Analysis (CMA, Massachusetts General Hospital, Charlestown, MA). The images were acquired from multiple centres and subsequently various cortical and sub-cortical structures were manually delineated within each image. An example of an anatomical image and the corresponding manual labels are shown in Figure 5.1.

The set of structures that were present in all the images were: lateral ventricle, caudate, putamen, accumbens, pallidum, thalamus, amygdala, hippocampus and brainstem.

With these data, it was possible to carry out experiments on a leave-one-out basis. For a given query subject, an estimated segmentation can be generated via classifier selection, label propagation and fusion. Subsequently, the segmentation estimate can be compared with the query subject's manual labels which are treated as a gold standard.

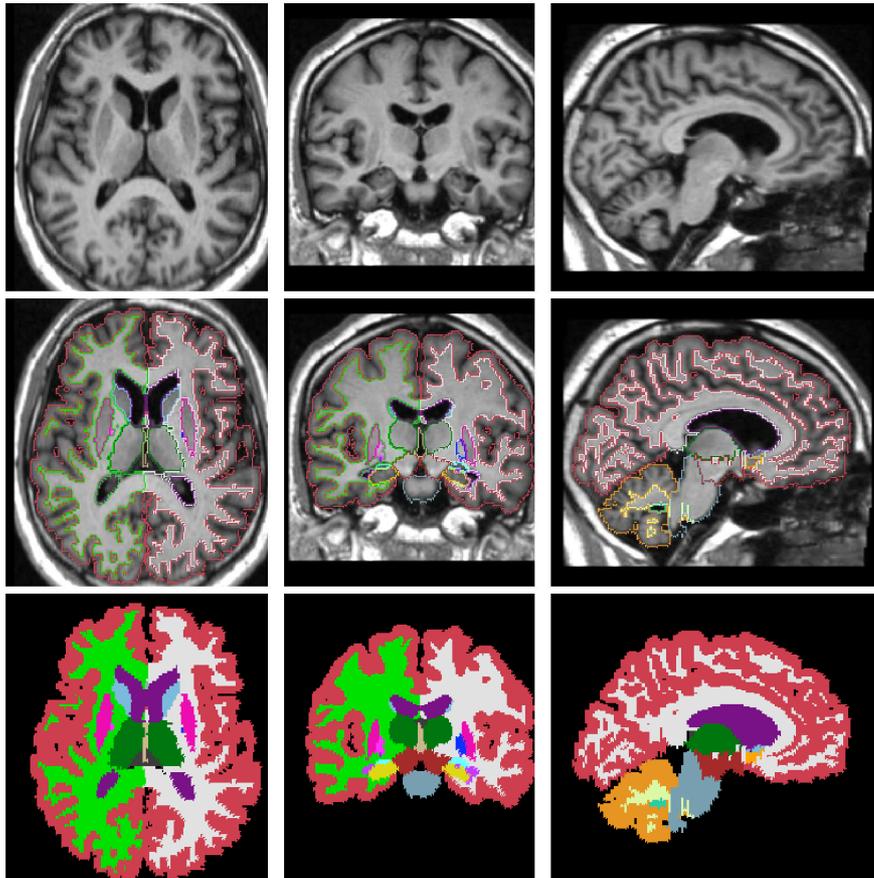


Figure 5.1: An example dataset from the repository of CMA images. Top to bottom: A T1-weighted MR image. Contours of the manual labels overlaid on the anatomy. The manual labels.

As well as the CMA data, a reference image was used to define a standard space where a selection scheme required it. The image used as a reference was the single subject atlas available from the Montreal Neurological Institute [91] (MNI, McGill University, Quebec, Canada).

5.3 Defining the ROI for similarity selection

When finding the similarity of a pair of aligned images, the region over which voxels contribute to the similarity metric needs to be defined. For the purposes of this work, a mask of the sub-cortical structures was used throughout in order to restrict the similarity metric calculation to the neighbourhood of the sub-cortical structures. Other masks are clearly possible, for example masks created for specific single structures, but a single mask was used for simplicity of presentation.

The first step was to identify a mask for the sub-cortical structures in each of the individual images in the CMA dataset based on their manual labels. The sub-cortical structures selected were the ventricles (lateral, inferior, third, fourth), thalamus, caudate, putamen, pallidum, brainstem, hippocampus, amygdala, nucleus accumbens, and the ventral diencephalon. The mask from each individual image was transformed to the MNI reference image using the transformation derived from an affine registration. The union of all the transformed masks was then found in the space of the reference image and dilated twice to generate the collective sub-cortical mask for the cohort. An illustration of the mask is shown in Figure 5.2.

The mask generated in this way can be used for similarity measurements between images that have been aligned to the reference – this corresponds to the R1, R2 and H schemes described in the previous chapter. For experiments where similarity comparisons need to be made in the native space of the query subject (i.e. schemes Q1 and Q2), the collective mask was transformed back to the query image with the inverse of the affine transformation that was used during the mask

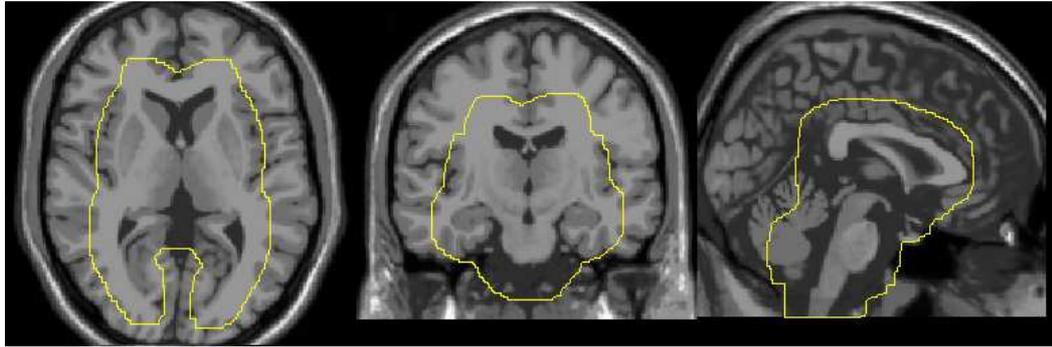


Figure 5.2: The sub-cortical mask used for similarity based selection schemes. This is shown in contour form over the MNI single subject atlas. The mask defines the region over which the similarity of an image pair is calculated.

creation stage.

5.4 A comparison of image similarity based schemes

Experiments were carried out to compare the quality of the results produced by each of the image similarity based selection schemes.

For a given query subject, the remaining 274 atlases were treated as the repository and each of the selection schemes was used to identify its best classifiers in the repository for subsequent fusion. The selection was made using normalised mutual information (as this was the similarity metric used during registration) and the region used for similarity measurement was the sub-cortical ROI described in Section 5.3. After ranking based on each of the schemes, the number selected for subsequent fusion was 20 for each experiment.

For practical reasons, the number of experiments carried out for each scheme differed depending on whether the similarity measurements were made in the space of the reference image or in the native space of the query. Where the similarity measurements were made between images aligned to the reference (schemes R1, R2 and H), segmentation estimates for all 275 of the subjects could be made. The query centric schemes (Q1, Q2), however, require the alignment of all the images in the repository to each query image prior to selection. Carrying out query

centric selection on a leave-one-out basis for many images was consequently prohibitive. Three subjects with different ages were therefore selected for testing the query centric schemes.

The registrations carried out for the schemes either optimised an affine transformation (12 parameters) or a non-rigid transformation. Non-rigid registrations were initialised with the affine transformation estimate and optimised free-form deformations for the local displacements using successive control point spacings of 20mm, 10mm and 5mm (See Sections 2.2.7 and 2.2.6). The different combinations of target subject and type of registration are listed in Table 5.1.

Scheme	Stage	Target	Type	Purpose
Q1	1	Query	Non-rigid	Selection and label propagation
Q2	1	Query	Affine	Selection
	2	Query	Non-rigid	Label propagation
R1	1	Reference	Non-rigid	Selection and label propagation
R2	1	Reference	Affine	Selection
	2	Reference	Non-rigid	Label propagation
H	1	Reference	Affine	Selection
	2	Query	Non-rigid	Label propagation

Table 5.1: The combinations of registration types, targets and their uses for the comparison of the similarity based selection schemes.

Differences between the selection schemes imply differences in the way in which the final assessment of segmentation accuracy can be made. For schemes where the fusion step is carried out in the query image’s native space (schemes Q1, Q2, H), the accuracy of the resulting segmentation can be measured by the overlap with the query’s manual labels. If, however, the fusion step takes place in the space of the reference image (schemes R1, R2), the resulting segmentation and the query’s manual labels differ by the transformation between query and reference. For this reason, the accuracy of segmentations estimated by reference centric schemes R1 and R2 are given by their overlaps with query manual labels propagated non-rigidly to the space of the reference.

Summary results for the different similarity based selection schemes are shown in Figure 5.3 where the average Dice overlaps achieved for each scheme are dis-

played. A separate summary of the performance of the schemes is shown in

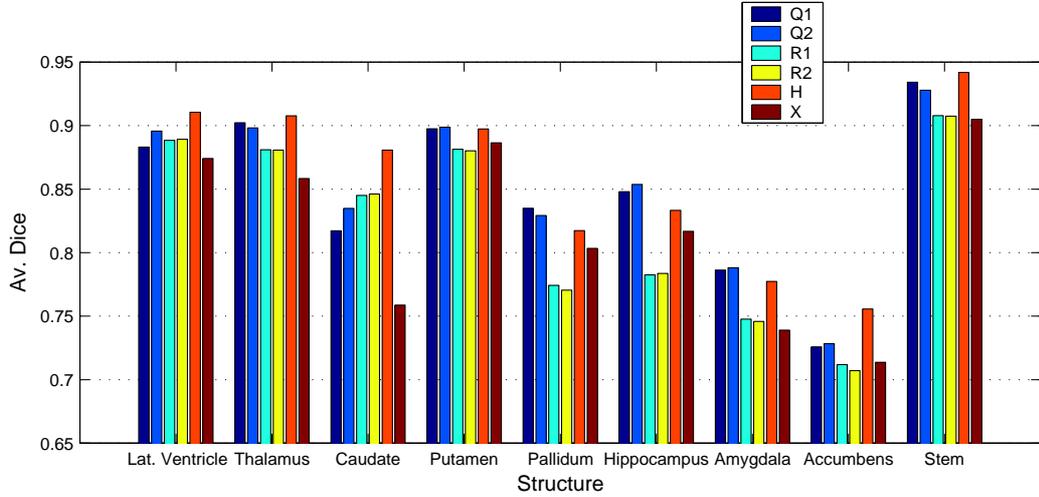


Figure 5.3: A comparison of different similarity based selection schemes using average Dice overlaps for different structures. The structures are listed on the horizontal axis and, for each structure, groups of bars indicate the performance of each scheme – left-right structures are combined. Within each group of bars, the overlaps achieved for the Q1, Q2, R1, R2 and H schemes are listed from left to right followed by the average overlap achieved without selection, i.e. by fusing random sets of 20 classifiers (X).

Figure 5.4, where the Dice overlaps for each scheme across all the structures is shown in the form of a box plot.

The data in Figures 5.3 and 5.4 suggest that schemes producing segmentation estimates in query space tend to perform better than reference centric schemes. This was confirmed by a t-test comparing overlap accuracy obtained by schemes Q1, Q2 and H (combined mean 0.8538) on one hand and schemes R1 and R2 (combined mean 0.8202) on the other ($p < 0.0001$, 2-tailed test, unequal variances). The best performing schemes based on the box-plots shown in Figure 5.4 are compared in more detail in Table 5.2. The results produced by hybrid selection also appear to be slightly better than those produced by query centric schemes Q1 and Q2 although the results should be treated with some caution as the averages are based on different numbers of experiments. The mean Dice overlap for hybrid selection was 0.854 and the combined mean for schemes Q1 and Q2 was 0.848. In a t-test, the difference between these means was not sig-

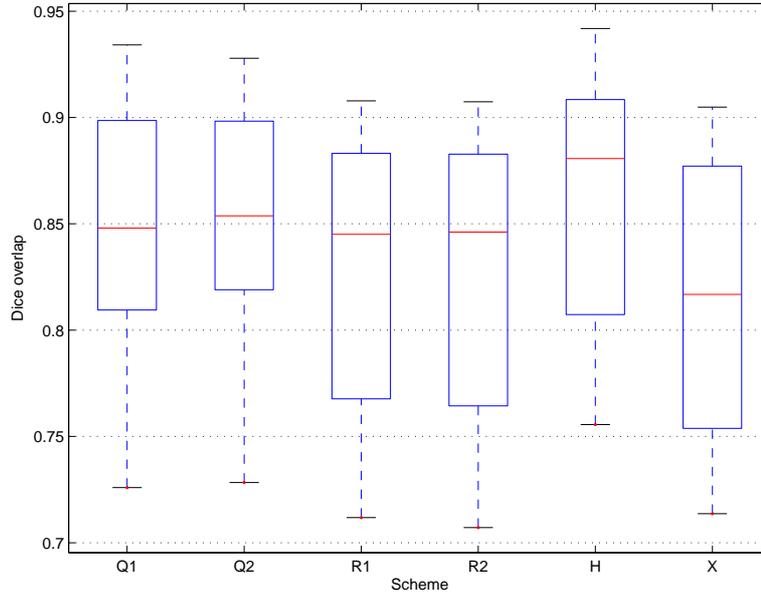


Figure 5.4: A box-plot to compare the different schemes for classifier selection and fusion. This chart represents a summary of the data shown in Figure 5.3 where the results across all structures for each scheme are represented by a box and whisker diagram. The last box and whisker plot shows the distribution of overlaps obtained without selection i.e. by fusing random sets of 20 classifiers.

Structure	Scheme Q1	Scheme H	No selection
Lateral ventricle	0.881 (0.09)	0.914 (0.04)	0.866 (0.09)
	0.901 (0.05)	0.911 (0.04)	0.882 (0.05)
Thalamus	0.898 (0.01)	0.908 (0.02)	0.854 (0.02)
	0.906 (0.01)	0.909 (0.02)	0.862 (0.02)
Caudate	0.826 (0.06)	0.883 (0.03)	0.747 (0.11)
	0.839 (0.04)	0.879 (0.03)	0.766 (0.09)
Putamen	0.901 (0.03)	0.898 (0.02)	0.887 (0.03)
	0.900 (0.03)	0.898 (0.02)	0.882 (0.03)
Pallidum	0.852 (0.03)	0.819 (0.05)	0.803 (0.03)
	0.837 (0.04)	0.818 (0.05)	0.800 (0.02)
Hippocampus	0.846 (0.01)	0.832 (0.04)	0.808 (0.03)
	0.852 (0.02)	0.837 (0.04)	0.825 (0.03)
Amygdala	0.794 (0.05)	0.778 (0.06)	0.749 (0.05)
	0.797 (0.05)	0.776 (0.06)	0.728 (0.09)
Accumbens	0.767 (0.09)	0.765 (0.07)	0.726 (0.08)
	0.723 (0.07)	0.751 (0.07)	0.698 (0.07)
Brainstem	0.934 (0.01)	0.941 (0.01)	0.903 (0.02)

Table 5.2: Detailed comparison of the overlaps achieved after selection using schemes Q1 and H. For each paired structure, the left and right overlaps are shown on successive rows. For reference, the average overlap achieved by fusing random sets of 20 classifiers are given in the final column.

nificant at the 5% level. The results also suggest that there is little difference in performance between one stage and two stage schemes, i.e. Q1 (mean 0.8503) and Q2 (mean 0.8458) performed very similarly and could not be separated by a t-test at $p = 5\%$. This was also the case for schemes R1 (mean 0.8210) and R2 (mean 0.8195).

5.5 Image similarity as a selection criterion

The experiments described in section 5.4 compared schemes for the selection of classifiers that differed in various aspects such as the space chosen for selecting classifiers, the type and number of registrations used and the space in which the final segmentation estimate is made. A common feature of all the schemes was the use of image similarity as the criterion on which classifiers were ranked prior to selection – in each case, NMI was used as the similarity measure. It is natural to ask, therefore, whether image similarity represents a suitable selection criterion with respect to the ultimate aim of obtaining accurate segmentations.

If two images, X and Y , are being registered, and their similarity is $sim(X, Y)$, then the aim is to maximise $sim(X, Y)$ with respect to the parameters of a transformation between them; the only images that are considered are X and Y . The use of similarity for selection differs, however, in that similarities between a query image and multiple other subject images are considered. If Q represents a query image and classifier atlases are represented by A and B , then a desirable property of similarity selection would be that it determines the best of A and B as classifiers for Q . This means, for example, that if $sim(Q, A) > sim(Q, B)$ then A should provide a more accurate atlas segmentation of Q . It is possible, however, that this may be confounded by variations in the contrasts and quality of the atlases and variations in scan protocol. The CMA data were acquired from multiple centres and represent a diverse range of subjects – it is unlikely that this large number of images were acquired with the same scanning protocols.

In order to make a preliminary assessment of the suitability of image similarity as a selection criterion, an experiment was carried out to determine the relationship between the ranks of classifiers as given by image similarity and their performance (as individual classifiers) in segmenting the query image. The 275 T1 images and their corresponding labels were all affinely aligned to the MNI reference image. Each subject in the data set was treated as a query in turn and the ranks of the remaining atlases were found based on their similarity with the query T1 image. The accuracy of the remaining images as classifiers for a structure by was estimated from their Dice overlaps with the query’s manual labels. Repeating this process for all subjects allows the average accuracy to be calculated for the classifiers for each rank. Plots of the average Dice overlap against the rank of the classifier are shown for the hippocampus and the lateral ventricle in figure 5.5. These plots indicate that highly ranked classifiers are associated with a higher level of accuracy for both these structures. Although no label fusion is carried out in this experiment, the correlation between classifier rank determined by similarity and accuracy seems to justify the use of similarity as a basis for selection of classifiers prior to fusion (Correlation coefficients: Lateral ventricle, -0.9488 ; Hippocampus, -0.8765).

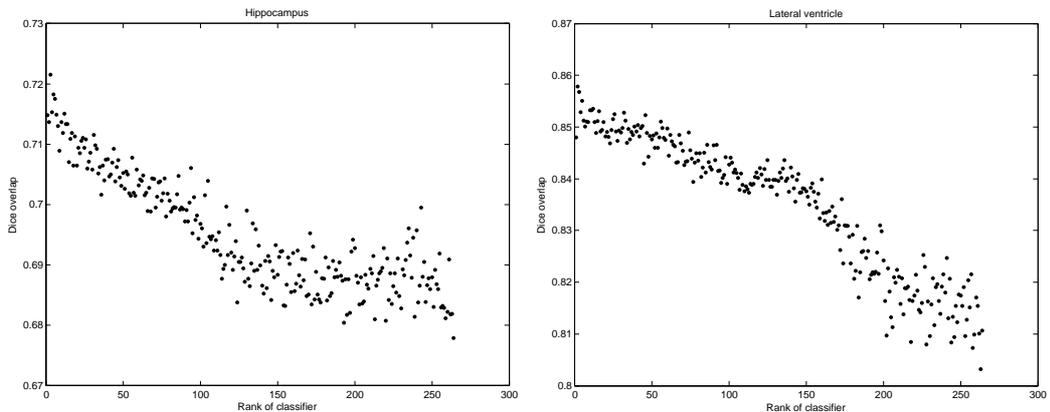


Figure 5.5: The relationship between average Dice accuracy obtained by individual classifiers and their rank as determined by image similarity with the query. The results for the hippocampus are shown on the left and those for the lateral ventricle are shown on the right.

5.6 Detailed comparison with the fusion of random sets of classifiers

As described in Section 5.4, the average accuracy values for the query centric schemes were based on fewer subjects ($N = 3$) than for the reference centric and hybrid schemes ($N = 275$). This was due to the higher computational burden associated with query centric selection and fusion. For this reason, a further, more detailed, investigation of the accuracy of the query centric schemes is presented.

Three individual subjects were selected in order to assess the query centric schemes. The only criterion for the selection of these subjects was that their ages should represent the range of ages in all the data. The subjects are simply referred to as subjects 1, 2 and 3 and the ages at scan were: Subject 1: 11.6 years. Subject 2: 29 years. Subject 3 : 79 years. For these subjects, a set of experiments was carried out in order to compare in more detail the accuracy of segmentations obtained after classifier selection with the accuracy of segmentations obtained by simply fusing random sets of classifiers.

The selection scheme used to generate the segmentations for each individual subject was the query centric, one stage scheme. Randomly selected groups of 20 subjects were selected from the remaining 274 subjects and their labels were transformed non-rigidly to the selected individual and fused to generate a segmentation for comparison. The process of random selection and fusion of labels was repeated 1000 times for each of the selected individuals.

The Dice overlaps of each segmentation produced by the fusion of a random group of subjects were then evaluated for the sub-cortical structures. The overlaps were calculated between the randomly fused labels and the gold standard manual labels. This allowed an estimate of the distribution of Dice overlaps to be made for each structure and for each of the selected individuals.

The accuracy of segmentations produced by the Q1 selection and fusion scheme (as measured by overlaps with the gold standard manual labels) can then be com-

pared against the distributions of overlaps based on the fusion of random groups.

The distributions of Dice overlaps from randomly fused sets of labels are shown for the different structures in Figure 5.6. As an example, focusing on the data for the left pallidum in Figure 5.6, the random overlap distribution has a median Dice overlap of approximately 0.78. The limits of the inter-quartile range are 0.77 and 0.79, the overlaps for half the random segmentations lie between these limits. The whiskers for the box-plot of this distribution, representing the extremes of the distribution, are at 0.72 and 0.84. The Dice overlap obtained from the single segmentation of the left pallidum obtained after Q1 selection is plotted as a circle and can be observed slightly above the whisker at the upper end of the random distribution.

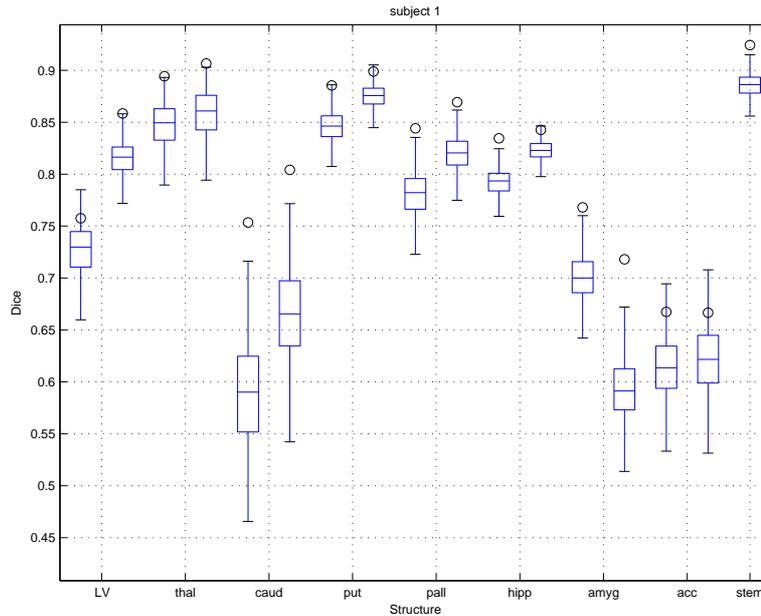


Figure 5.6: An assessment of the overlaps for different structures for subject 1 obtained after applying the Q1 scheme. The overlap obtained for each structure (plotted as a circle) can be compared against the distribution of overlaps obtained from the fusion of 1000 sets of 20 labels from randomly chosen subjects.

Plots that assess the gold standard overlaps of the Q1 scheme for subjects 2 and 3 are shown in a similar way in Figures 5.7 and 5.8. For all the cases, the accuracy of the segmentation obtained by Q1 selection and fusion is higher than the average overlap obtained by fusing random classifier sets. In most cases,

the Q1 segmentation estimate exceeds the 75th percentile of the random overlap distributions. In many cases it exceeds, the outlier limits defined by the whiskers in each box-plot.

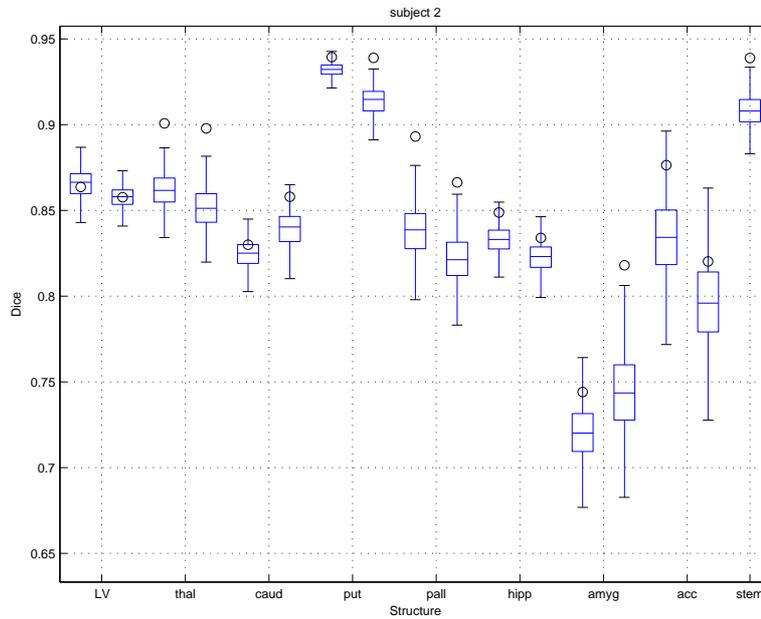


Figure 5.7: An assessment of the gold standard overlaps obtained after applying the Q1 scheme for subject 2. The overlap obtained for each structure (plotted as a circle) can be compared against the distribution of overlaps obtained from the fusion of 1000 sets of 20 labels from randomly chosen subjects.

A very small number of the overlaps achieved after the fusion of random sets of labels fell outside the ranges defined by the whiskers of each box-plot in Figures 5.6, 5.7 and 5.8. These were not plotted for readability. Across all structures and subjects, the whiskers defined (on average) the 0.007% and 99.998% percentiles of the random distributions of overlaps.

A slightly different way of using the distributions of random overlaps to assess the quality of the overlaps achieved after selection is through the use of z-scores. For a given subject and structure, the mean and standard deviation of the distribution of random Dice overlaps can be calculated. The gold standard overlap achieved after selection can be converted into a z-score based on this mean and standard deviation. The z-score measures the signed difference between an overlap achieved after classifier selection and the mean of the random distribution.

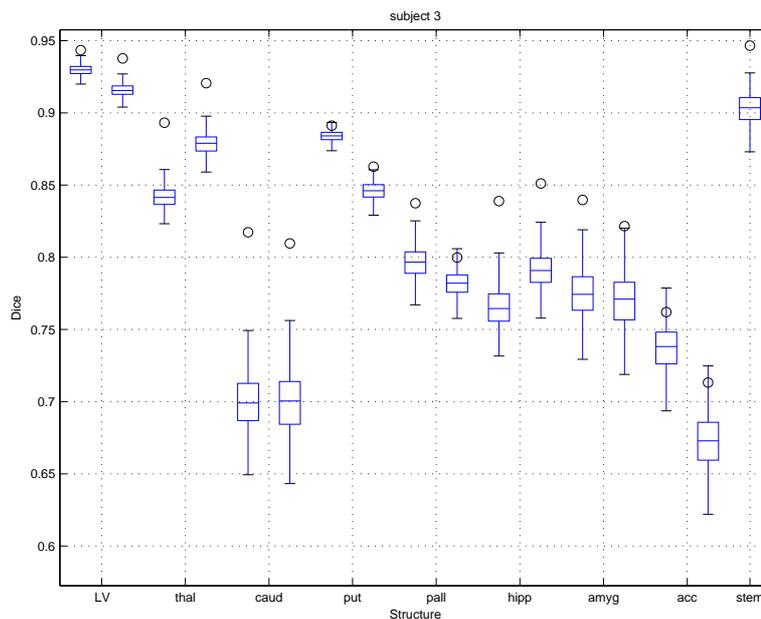


Figure 5.8: An assessment of the gold standard overlaps obtained after applying the Q1 scheme for subject 3. The overlap obtained for each structure (plotted as a circle) can be compared against the distribution of overlaps obtained from the fusion of 1000 sets of 20 labels from randomly chosen subjects.

The difference is expressed as the number of standard deviations above / below the mean. Clearly, a large positive z-score indicates that the overlap achieved significantly out-performs the overlaps achieved from randomly fused labels.

A summary of the z-scores for each of the selected subjects and structures is shown in Table 5.3 and the same information is also shown pictorially in Figure 5.9, where the left and right figures for each of the structures (except the brainstem) are shown separately. These data show that the overlaps achieved by segmentations after classifier selection will, in general, perform much better than overlaps achieved by fusing labels randomly. The average z-score achieved overall is 2.91 : this represents a cumulative percentage of 99.8% for normal distributions.

5.7 The choice of similarity metric

This section presents data to illustrate how the choice of similarity metric used during the selection of classifiers affects the quality of the final segmentations.

	Subject			Average
	1	2	3	
Lateral Ventricle	1.70	-0.05	4.21	1.96
Thalamus	2.21	3.87	6.31	4.13
Caudate	3.22	1.17	5.54	3.31
Putamen	2.48	2.38	2.37	2.41
Pallidum	2.83	3.31	2.89	3.01
Hippocampus	2.76	1.63	5.00	3.13
Amygdala	3.73	2.26	3.22	3.07
Accumbens	1.58	1.43	1.78	1.59
Brainstem	3.43	3.10	4.20	3.58
Average	2.66	2.12	3.95	2.91

Table 5.3: Z-scores achieved by segmentations produced by fusion of labels after classifier selection (Scheme Q1).

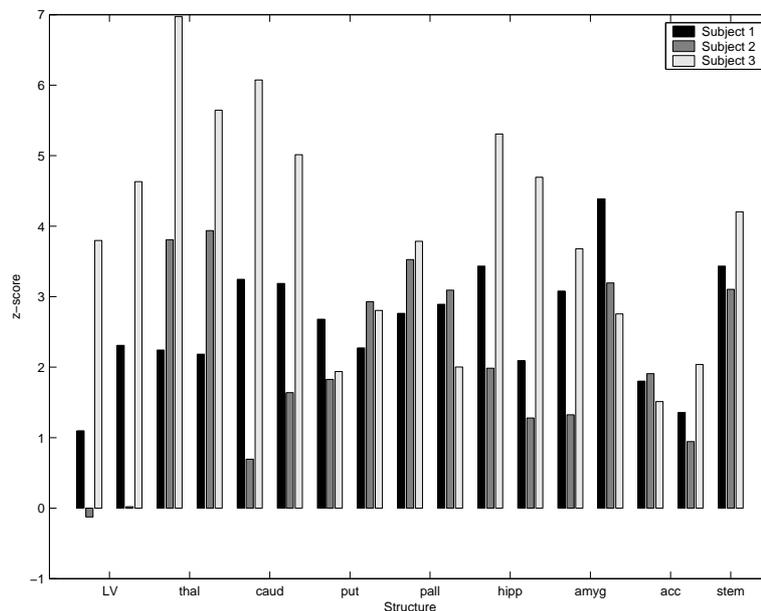


Figure 5.9: Z-scores achieved by segmentations produced by fusion of labels after classifier selection (Scheme Q1). For paired structures, separate bars are shown for the left and right structures. The abbreviations listed on the horizontal axis are for the following structures : Lateral ventricle (LV) thalamus (thal), caudate (caud), putamen (put), pallidum (pall), hippocampus (hipp), amygdala (amyg), accumbens (accum), brainstem (stem).

During all the results presented so far, normalised mutual information (NMI) was the metric used to determine the similarity of subjects for the various selection schemes. This choice was made because this metric was used during all registrations. It therefore seemed reasonable to assume that it would correlate with the degree of success of each registration. Because classifier fusion relies on good estimates of correspondence between atlases and query images, the measure of registration success should correlate with the quality of the resulting segmentation. Conversely, an atlas that is badly aligned with a query image will not make a good contribution in a label propagation and fusion scheme.

The NMI metric was compared against the results obtained from selection using cross correlation (CC), a non-information theoretic measure (see Section 2.2.8). The selection scheme that was used for the comparison was the hybrid scheme (H). This meant that the repository images were all aligned to the MNI reference image and selections were made in reference space of the top 20 classifiers for each subject. The selections were made twice, once with each of the similarity metrics. It should be stressed, however, that the registrations of each of the subjects to the reference were not repeated. For both sets of selections, the registrations based on NMI were used for the alignment and subsequent label propagation.

The average Dice overlaps obtained after hybrid selections using NMI and CC are compared in Figure 5.10. These results show that the accuracy of results based on CC selection are slightly lower than those based on NMI but that there is very little difference in the accuracy of the final segmentations based on selection with either metric. The results were obtained after 275 leave-one-out experiments for each selection metric. T-tests were applied to the sets of data represented by the averages in Figure 5.10. None of these tests showed a significant difference between the results of using each metric at the $p = 0.01$ significance level.

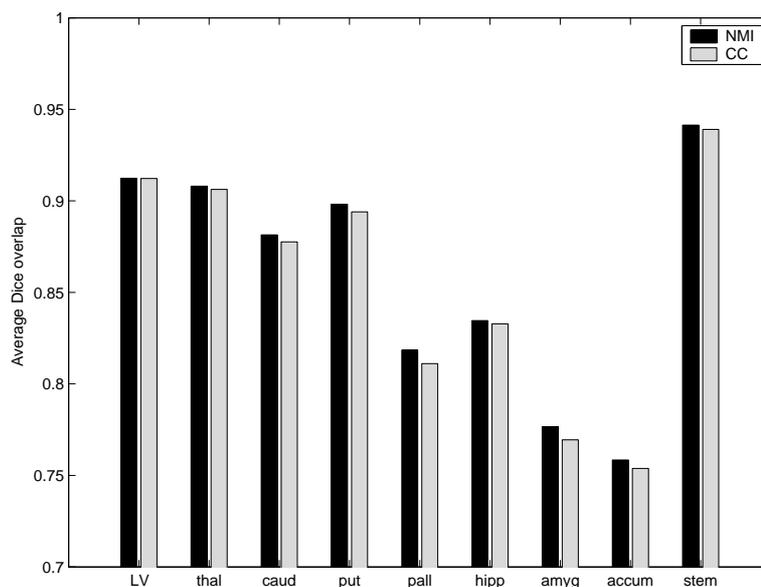


Figure 5.10: A comparison of the effect of the choice of metric during the classifier selection stage. Average Dice overlaps for hybrid selection and fusion of various structures. For paired structures, the averages of the left and right instances are combined. The abbreviations listed on the horizontal axis are for the following structures : Lateral ventricle (LV) thalamus (thal), caudate (caud), putamen (put), pallidum (pall), hippocampus (hipp), amygdala (amyg), accumbens (accum), brainstem (stem).

5.8 Varying the number of classifiers

When applying a similarity based selection scheme, the atlases in the repository are ranked according to their similarity with the query image. A decision then needs to be made on the number of classifiers to use for the next stage, label propagation and fusion, that gives the final segmentation estimate. This section presents data from experiments that aim to address this question.

The experiments were carried out using the images for subjects 1, 2 and 3 described in Section 5.6 and by applying the Q1 selection scheme. For a given query image, the repository atlases ($N = 274$) were ranked by their similarity to the query. Subsequently, increasing numbers of the *ordered* classifiers were selected and fused to provide separate estimates of the query segmentation; in other words, the label sets combined to provide a segmentation after selecting k classifiers is represented by the label sets used after selection of $k - 1$ classifiers

with the inclusion of the k^{th} classifier in the ordered list. The accuracy of each segmentation estimate was assessed using the Dice overlap with the query image's corresponding manual label.

The combined left-right average accuracy for various structures for all three query subjects chosen are shown in Figure 5.11 by the solid blue lines. Each chart shows how the accuracy of the resulting structure varies as increasing numbers of the ordered classifiers are selected. The average overlap accuracy for all structures and all subjects is shown in the bottom right of Figure 5.11.

For comparison, the accuracy achieved by fusing random sets of classifiers, with varying numbers in each, is shown in each chart as a dashed line.

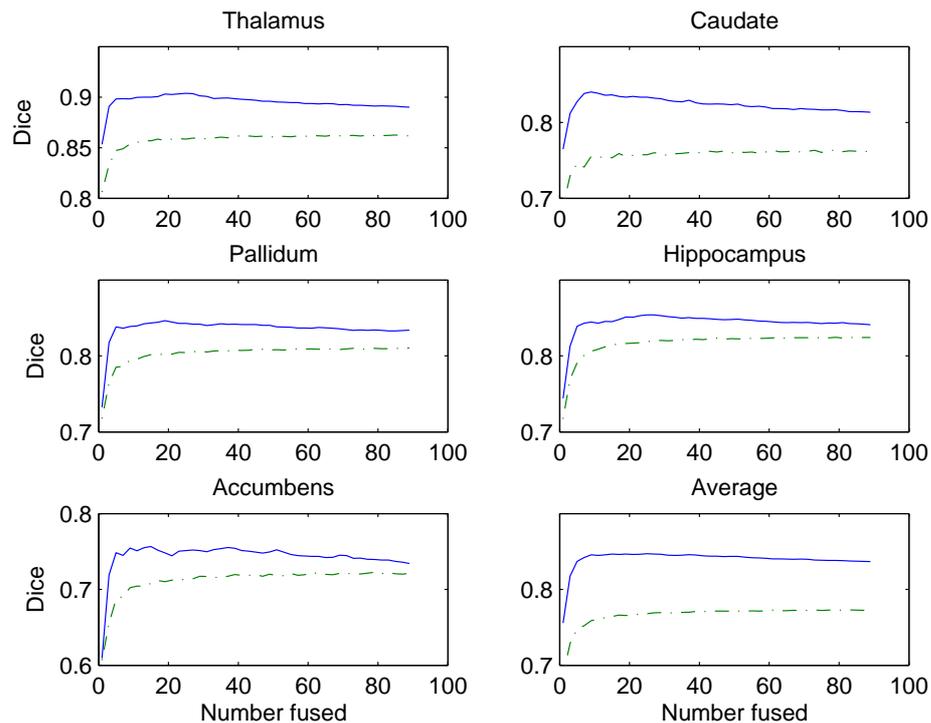


Figure 5.11: The segmentation accuracy for various structures by fusing increasing numbers of ranked classifiers. The vertical axes show average Dice with the manual labels. The horizontal axes show the number fused. The solid lines show overlap for classifiers chosen from the ranked list. For comparison, the dashed lines show overlap for random sets of classifiers. Data for five structures (combined left-right) are shown. The average over all structures is shown in the plot at the bottom right.

The data shown in Figure 5.11 can be a little noisy, as demonstrated, for

example, by the overlaps for the accumbens (a small structure). The general pattern, however, is that there is a sharp initial increase in overlap accuracy up to a maximum level followed by a gradual decline.

This contrasts with the overlaps achieved by fusing random sets of classifiers; after a sharp initial increase, these show a continued monotonic increase but always remain below the accuracy level used by fusing ranked classifiers. Clearly as the number fused approaches the size of the repository the accuracy of ranked and random classifier fusion should converge to the same level, which would be the accuracy obtained by using the population average as the segmentation estimate.

The number of ranked classifiers needed to give the highest accuracy varies for the different structures. The overlaps for the caudate reach a maximum for about 8 classifiers while the maximum for the hippocampus is reached after the selection of the top 25 classifiers on average. The average overlaps across all structures in the bottom right of the figure show a fairly flat section of the highest overlap values for between 15 and 25 classifiers. This suggests that the choice of 20 classifiers is appropriate for general segmentations. If, however, the segmentation of a particular structure is required, the number of classifiers used from the ranked atlases could then be adapted specifically to suit the structure.

5.9 Comparing similarity based and age based selection

A comparison was made for all 275 subjects in the repository of the accuracy of segmentations generated after selection based on age and selection based on similarity. The similarity selection scheme used was the hybrid selection scheme, where images are aligned to the reference before similarity comparison. In order to select classifiers by age, the twenty nearest subjects to a given query were simply chosen and fused.

The performance of each method for each subject was measured by the mean overlap over the sub-cortical structures. After calculating the means, a single figure was estimated for each subject representing the signed difference between the schemes: age selection accuracy minus hybrid selection accuracy.

The resulting values are shown in Figure 5.12 where the points for each subject are plotted using different markers simply to indicate their approximate age group. Points lying above the horizontal axis represent subjects for whom age selection represents an improvement over hybrid selection and *vice-versa*. For subjects in the middle age group (represented by crosses) the distribution of the differences suggests that there is negligible difference between the schemes. For the youngest subjects (represented by dots) there are some subjects for whom age selection appears to produce a slightly worse result although the differences are small. The mean difference for the younger subjects was -0.0055 (S.D. 0.0057) and the maximum change was -0.0199. A single sample two-sided t-test identified the mean as significantly different from zero ($p < 0.0001$) with the 95% confidence limits for mean as $\{-0.0065, -0.0045\}$. The oldest subjects (represented by circles) were more varied than the other two groups, but contained the subjects for whom age selection made the most improvement. Again, the differences are small with a maximum increase in Dice accuracy of 0.0276. The mean increase in Dice for the older subjects was 0.007 while the standard deviation for Dice increases in the older group was 0.011. The standard deviations for the younger and middle groups was 0.0057 and 0.0035 respectively. The mean increase for the older group was different from zero in a statistically significant sense ($p < 0.005$, two-sided t-test) and the 95% confidence interval for the mean was $\{0.0023, 0.0108\}$. In general, the small values for differences in performance between age-based selection and hybrid similarity based selection suggest that the performance of both schemes is comparable.

To give a qualitative impression of the classifiers selected, three subjects were chosen from the young, middle and older aged groups and the top 10 classifiers

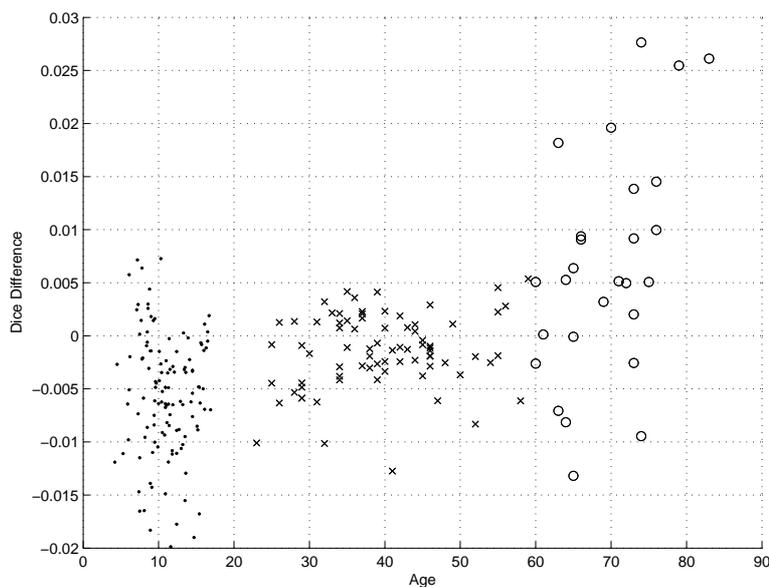


Figure 5.12: A comparison of age-based selection and similarity-based (hybrid) selection. The horizontal axis shows the age of the subjects, subjects in different age groups are plotted with different symbols. The vertical axis shows the improvement that age selection gives over the hybrid scheme – ‘improvement’ can be negative as well as positive.

based on the hybrid similarity selection scheme were identified. Slices from each of the top-ranked classifiers for each subject are shown in Figures 5.13 , 5.14 and 5.15. Figure 5.13 shows the classifiers selected for an 11 year-old subject, Figure 5.14 a 29 year-old and Figure 5.15 a 79 year-old.

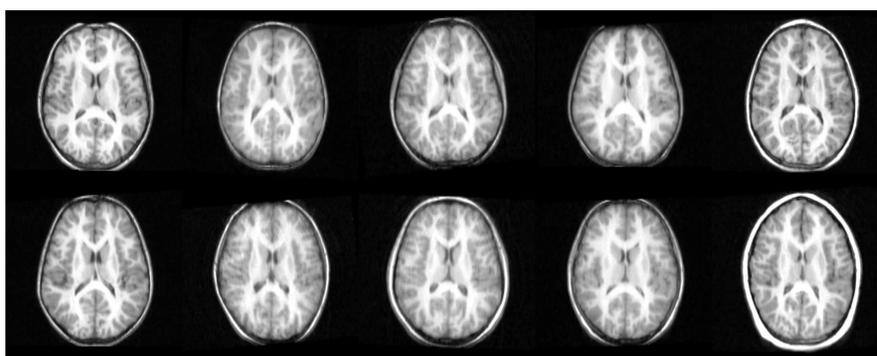


Figure 5.13: The top ten classifiers for an 11 year-old subject using the hybrid similarity scheme.

The ages of the top ten classifiers for each of these three subjects are shown in Table 5.4. The table shows that, although the classifiers were selected using an image similarity based scheme, the ages of the best classifiers match the age

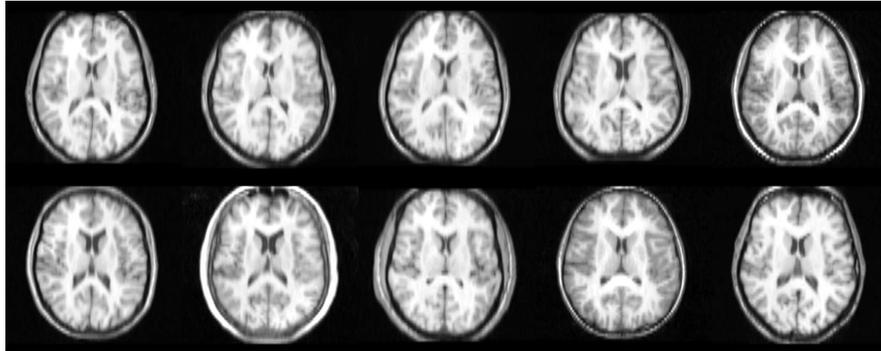


Figure 5.14: The top ten classifiers for a 29 year-old subject using the hybrid similarity scheme.

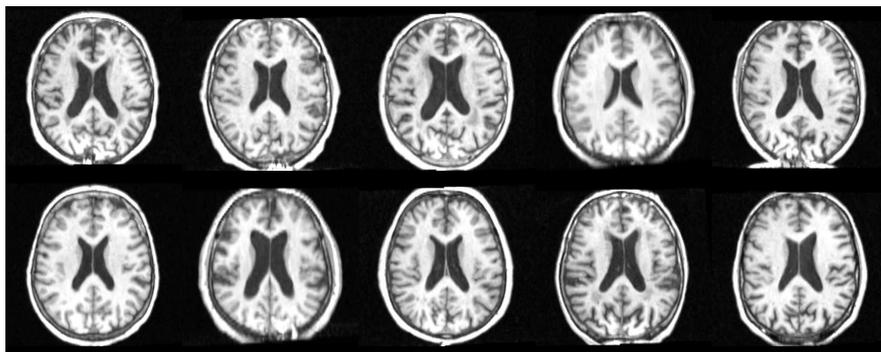


Figure 5.15: The top ten classifiers for a 79 year-old subject using the hybrid similarity scheme.

of the query subject well. This suggests that there is a good deal of agreement between selection on similarity and selection on age and gives a possible reason for the similar segmentation accuracy provided by age-based selection and by hybrid similarity selection as shown in Figure 5.12.

Rank of classifier	Age of query subject		
	11	29	79
1	13.6	29	73
2	4.5	52	72
3	10.1	25	76
4	9	43	31
5	8.6	16.5	75
6	8	25	65
7	9.5	55	63
8	11.3	34	73
9	11.9	12.7	83
10	7.5	55	65
mean (SD)	9.4 (2.55)	34.72 (15.74)	mean 67.6 (14.20)

Table 5.4: The ages of the top 10 classifiers selected using the hybrid similarity scheme for three different query subjects. The three subjects were chosen to cover the age range of the data.

5.10 Discussion

A comparison of the accuracy achieved by the different similarity based selection schemes has been presented in Section 5.4. Differences in computational burden associated with the schemes mean that there are some differences in the numbers of experiments carried out with each scheme prior to comparison. The number of possible experiments that could be carried out for the query centric schemes was limited by practical considerations but a detailed comparison for the Q1 scheme was carried out against distributions of overlaps obtained by fusions of random groups of classifiers. Additionally, the accuracy measurements for the reference centric schemes were made against a transformed version of the manual labels due to the segmentation estimates being made in the reference space.

Notwithstanding these differences, it is still possible to obtain an impression

of the relative performance of the different similarity based selection schemes. The best performing schemes, Q1 and H, were compared against each other in more detail (see Table 5.2). The overlap measures achieved by both these schemes are comparable with or better than those achieved by previously published automated methods [88] and are comparable with previous manual segmentation methods [57, 157].

The hybrid scheme has a significantly lower computational overhead compared with query centric schemes. The results presented suggest that the accuracy of the hybrid scheme is at least comparable with the query centric schemes, something that makes it a desirable alternative.

The lower accuracy values of segmentations produced by reference centric schemes could be due to more than one cause and further work would need to be done to isolate their effects. The likely reasons for this may relate to errors in the transformation between the query and the reference, this transformation is needed to produce the accuracy figures for the scheme by propagating the query labels to the reference, so that it can act as a gold-standard. Moreover, there are essentially two registrations involved in the construction and assessment of the accuracy of a reference centric segmentation. The registrations from the atlases to the reference and the registration from the query to the reference. This means that such a scheme is likely to contain more errors due to inaccuracies in registration compared with other schemes. The lower accuracy values could also be due to errors in the segmentations themselves.

Comparing the performance of the reference centric schemes R1 and R2, the results suggest that there is little accuracy lost when adopting a two stage scheme (R2) instead of a one stage scheme (R1). This would imply that the computationally cheaper option (two stage) is generally worth pursuing as the majority of the registrations carried out in the two stage schemes optimise a lower number of parameters – these are the registrations used during the selection stage.

In section 5.5 an initial assessment of the suitability of image similarity as a

selection criterion prior to label fusion was made. The results suggest that the rank of an atlas in terms of its image similarity with the query correlates with its accuracy as a classifier. This gives some justification for the use of image similarity for selection. Additionally, based on a visual inspection, the images showing the top-ranked atlases for three query subjects in Figures 5.13 , 5.14 and 5.15 indicate that the anatomies of the classifiers selected based on image similarity are indeed similar to those of the query subjects. In future work, selection based on geometric features can be explored and compared with selection based on image similarity. Geometric features can be extracted from the transformations between the images in the atlas pool and a reference image and those that are closest to the features obtained from a query-reference transformations can be used as a basis for selection. A similar approach was proposed in [42] where a metric defined between transformations to a reference space were used to select the most similar template to a query image.

The comparisons of the Q1 scheme accuracy figures against those achieved by fusing randomly chosen groups of classifiers (Section 5.6) suggest that the use of a similarity selection scheme produces a higher level of accuracy than chance alone.

The effect of the choice of similarity metric was briefly assessed in Section 5.7 and the results suggest a very slight advantage in using the same metric for selection as that used for registration (NMI in this case) but that the differences in performance were not significant.

The accuracy figures achieved by the fusion of different numbers of classifiers presented in Section 5.8 indicate that simply using larger and larger numbers of classifiers (after ranking) actually produces lower levels of accuracy in the resulting segmentation. As the number of classifiers fused increases, the accuracy achieved tends to rise quickly to a maximum and subsequently decline gradually. The number of classifiers required for the maximal accuracy varied according to structure, but the fusion of approximately 20 classifiers produces the near

maximal accuracy when the results are averaged across structures. A modification to the scheme, when segmentations for a specific structure are required, would be to identify the optimal number of classifiers for that structure in advance prior to the use of classifier selection.

The decline in the overlap values as increasing numbers of classifiers are used may be explained by the tendency of the resulting fused segmentation to the mean of the structure across the cohort. This mean segmentation may not represent the query subject as well as the segmentation provided by a relatively smaller number of ranked classifiers.

Another general implementation choice was that of the ROI used for making similarity comparisons. This represented a mask that covered all the sub-cortical regions for all subjects. Again, a structure-specific implementation could be carried out using a mask tailored to the structure that is required for segmentation.

In Section 5.9 the results of a comparison of the accuracy achieved after selection based on age and selection based on similarity (scheme H) were presented. These results suggest that there were no significant differences in the performance of the schemes. For some subjects, selection on age may even produce a very slightly worse result although the differences are very small. It is noted, however, that the subjects for whom age selection gave the most improvement were more likely to be among the oldest subjects. This may be due to the scans acquired from older subjects being more likely to have a significantly different appearance from the rest of the cohort. Interestingly, when comparing against the Dice overlaps generated by the fusion of random groups of subjects (Section 5.6), the subject for whom the accuracy of segmentations after classifier selection showed the greatest improvement over random selection (Table 5.3) was the oldest subject (age 79).

It should be noted that the accuracy of scheme H is comparable with scheme Q1 which, in turn, performs significantly better than the fusion of random subjects. The comparable levels of accuracy achieved by scheme H and age based

selection suggests that age based selection should also significantly outperform random selection.

5.11 Conclusion

This chapter has presented various experiments to compare different schemes for classifier selection and fusion and to test their accuracy against the levels of accuracy achieved by fusing random groups of classifiers.

The results suggest that the accuracy levels achieved by the query centric and hybrid schemes are significantly higher than would be achieved by the fusion of random sets of classifiers. The good performance of the hybrid scheme in particular, along with its comparatively low computational burden, makes it an attractive scheme for producing accurate structural segmentations. Additionally, the comparable performance of age based selection and the hybrid scheme indicate the potential that meta-information based selection has in generating good segmentations.

An interesting area for future work would be to extend the schemes presented to allow adaptive weighting of the different classifiers during the fusion stage. For example, such a scheme might use a local measure of the similarity between the classifying atlas and the query subject and, instead of using a simple majority vote rule at each voxel, the votes could be weighted by some measure of the classifiers' local agreement in the neighbourhood of the voxel.

Structural segmentations using the methods described in this chapter are used to help define regions of interest in a dataset consisting of Alzheimer's disease patients and age-matched controls. These data are presented in Chapter 6 and 7.

Chapter 6

Longitudinal change in early childhood

In the previous chapters a framework was presented for characterising longitudinal change in serially acquired data for a cohort. Methods for structural segmentation were also described. These methods rely on the use of label propagation and classifier fusion.

In this chapter, these methods, along with EM based tissue segmentations, are applied to serially acquired images to characterise longitudinal change in a group of children for the interval between one and two years of age.

There is a particular need to characterise and quantify brain development in this vulnerable population: extreme preterm delivery potentially disrupts brain growth and many children born prior to 26 weeks' gestational age develop neurological disabilities [117] and neurocognitive problems that can continue into adolescence [25].

Experiments were carried out aiming to provide estimates of global and regional growth within an exemplar group of 25 infants who were born preterm. Volumetric magnetic resonance (MR) imaging data were acquired at both one and two years of age for this group. As well as estimating growth, average space atlases at each timepoint were also generated to represent anatomy, tissue maps

and regional growth across the population. Growth estimates from individual registrations and from atlas based estimates are compared and shown to be consistent with estimates obtained from conventional tissue segmentation.

6.1 Growth in early childhood : Previous work

Most brain growth occurs during the first two years of life, much occurring in utero prior to birth at 40 weeks' gestational age. A full understanding of human brain development in general, therefore, must include this early period of rapid development. However, most previous neuroinformatic studies of human brain growth have focused on older children and adolescents [73, 164, 155].

Approximately 5% of children are born prematurely and, among those who survive, the primary impact of preterm birth is neurological. Impaired neurological development in preterm born children has been identified in infants [178] and in six-year old children [117]. Neuropsychiatric problems are common in the teenage years following preterm birth [25]. This clearly highlights the need for characterising and quantifying the patterns of brain development in infants and children.

Many previous studies have used cross-sectional designs to assess brain growth. These have included, for example, Huppi et al. [92], who estimated brain volumes of neonates with gestational ages of 29 to 41 weeks. Using images of similarly aged infants, Nishida et al.[126] used a semi-automated method for regional volumetric analysis. Giedd et al. [74] estimated growth for children and adolescents from four years, comparing in the context of a linear model. Matsuzawa et al.[120] applied more complex models to tissue volumes derived from cross-sectional data of children up to 10 years of age.

While cross-sectional studies can be used to identify general trends in longitudinal development, inter-subject variability can act as an additional source of error, given that the data are obtained from different subjects at various time-

points. Longitudinal studies are more costly and time consuming due to the need for serial acquisitions but have an advantage in that each subject effectively acts as its own control, reducing the confounding effect of inter-subject variability. Further discussion of the distinction between developmental inferences drawn from cross-sectional and longitudinal studies can be found in [107].

Growth studies based on longitudinal data include the models of white matter (WM) and grey matter (GM) growth presented by Giedd et al. [73] for children and adolescents aged between 4 and 20 years. Thompson et al.[164] tracked the development of the total brain, corpus callosum, ventricles and caudate in a group of six children scanned repeatedly between the ages of 3 and 15. Gerig et al.[72] reported regional growth estimates using atlas-based approaches for longitudinal data of eight subjects with baseline and follow-up scans at 2 and 4 years. A non-volumetric assessment of growth using longitudinal data is presented in Sowell et al.[155], where cortical thickness measurements are estimated for children scanned at 2 year intervals between ages 5 and 11. A critical review of literature relating to MRI studies of the developing brain can be found in Durston et al.[55].

Boardman et al.[20, 19, 18] used maps of the Jacobian determinant to identify volumetric group differences in deep grey matter between preterm born children and term born controls. These types of approaches, where derivatives of the deformation field are used to characterise regional differences (for example in surface area or volume), are sometimes described as Tensor Based Morphometry (TBM) [48, 39, 111]. Such approaches contrast with Voxel Based Morphometry (VBM) [6] where, after all images have been spatially normalised, the main focus of study is difference in tissue composition at individual voxels.

6.2 Image data

The subjects studied consisted of 25 preterm born children (9 female, 16 male) who were scanned at one and two years of age. The mean gestational age at

birth was 27.7 weeks (SD 2.2), the mean corrected ages at scan were 54.0 weeks (SD 5.8) for the one year (baseline) scans and 106.4 weeks (SD 4.4) for the two year (follow-up) scans. Scan ages were corrected for gestational age by recording the ages of the children relative to their due dates at 40 weeks gestation rather than the actual premature birth dates. The mean interval between corresponding scans was 52.4 weeks (SD 7.1).

Young children are more likely to move during MR scanning which increases the likelihood of motion artefacts in the resulting images. The children used in this study were born preterm so it was possible to obtain ethical permission to sedate them (using chloral hydrate) during scanning on the grounds that there is clinical information to be gained from imaging them. Ear protection was also used (Natus MiniMuffs, Natus Medical Inc, San Carlos, CA). The scans acquired for all children were T1-weighted MR volumes. Seven subjects' images were acquired using a Marconi 0.5 T Apollo scanner, TR/TE = 23ms/6ms, flip angle = 30°. The images for the remaining subjects were acquired using a 1.0T HPQ system (Philips Medical Systems, Cleveland, Ohio), TR/TE = 23ms/6ms, flip angle = 35°. All images were reconstructed with voxel dimensions of $1 \times 1 \times 1.6\text{mm}^3$. For each subject, both the baseline and follow-up images were acquired on the same scanner. The images were pre-processed initially by extracting the brain regions and by correcting for MR non-uniformity. Brain masks were created for the images using 'BET' [152] and the non-uniformity correction was carried out using 'N3' [150].

To illustrate these pre-processing steps, a slices taken from an original scan and the non-uniformity corrected image are shown in Figure 6.1. This figure also shows the difference image along with the result of applying the brain extraction.

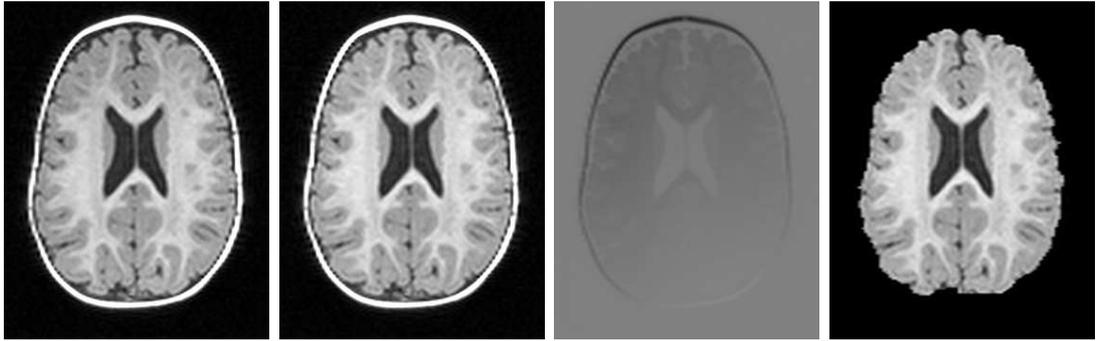


Figure 6.1: Illustrations to show the pre-processing steps applied to the image data. From left to right: A slice taken from an acquired image. The corresponding slice taken from the non-uniformity corrected image. The difference between the original and non-uniformity corrected images. The result of applying the brain extraction step.

6.3 Experimental design

With the aim of using the data described in the previous section to quantify growth for individual subjects during the second year of life, registration based growth estimates were calculated over the whole brain and these were also transformed to a common coordinate system and aggregated to produce a population average.

In order to obtain registration based growth estimates, for individuals or for the cohort, two types of registration were carried out. A longitudinal registration was performed for each subject using the baseline and follow-up scans. The transformations obtained from these intra-subject registrations can be used to provide estimates of growth for each subject by calculating the Jacobian determinant. Cross-sectional (inter-subject) registrations were carried out within each timepoint by randomly selecting a reference subject and registering the remaining subjects within the timepoint to the reference. The cross-sectional transformations can be used to generate atlases at each time-point in the space of the reference subject. It is, however, possible to generate ‘average space’ atlases within each time-point by using the average cross-sectional transformation to the reference. These atlases can represent anatomy, particular structures or regional growth for the population. A schematic diagram for the different registrations

and the atlases is illustrated in Figure 6.2.

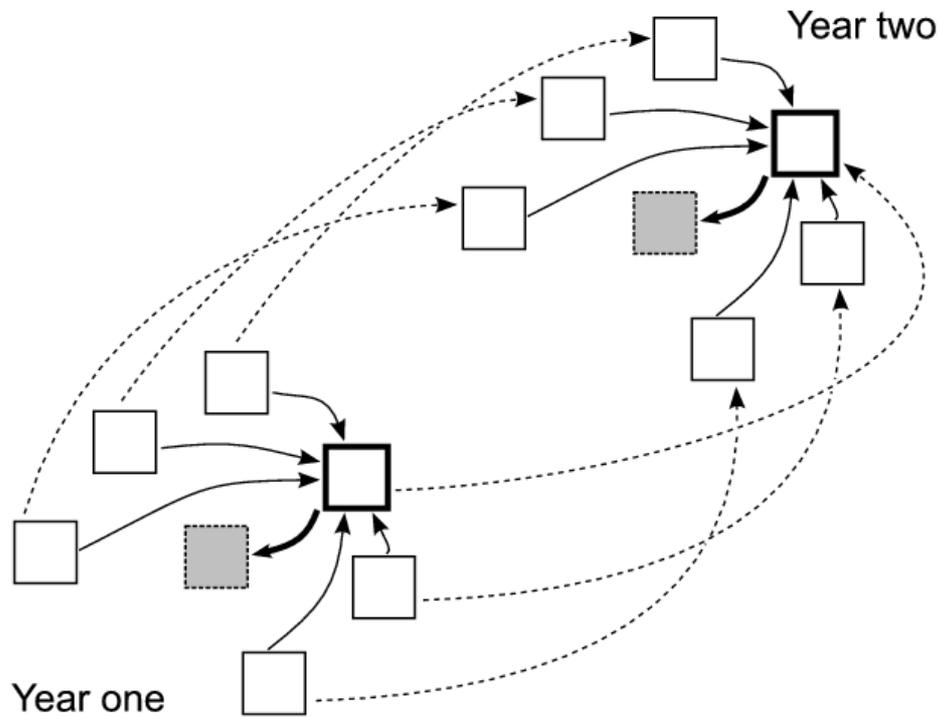


Figure 6.2: Schematic illustration of the images used and registrations performed. The subject images are represented by the clear squares with the reference subject shown in bold. The year one (baseline) and year two (follow-up) images for each subject are registered to obtain longitudinal transformations (shown as dashed arrows). Within each timepoint, the subjects are registered to the reference (solid arrows) and the average transformation (bold arrows) is used to map reference locations to the average space (grey square). Atlases were generated in the average space which could represent anatomy, growth or tissue class across the population.

6.4 Individual registration results

The longitudinal registrations were carried out using the baseline scan as the target or fixed image and the follow-up scan as the source or deformed image. The baseline and follow-up images for an individual are shown in Figure 6.3, along with an illustration of the transformed follow-up image after non-rigid registration (which also shows the deformed lattice of control points used to transform the image).

Images showing the difference between the baseline and follow-up images are

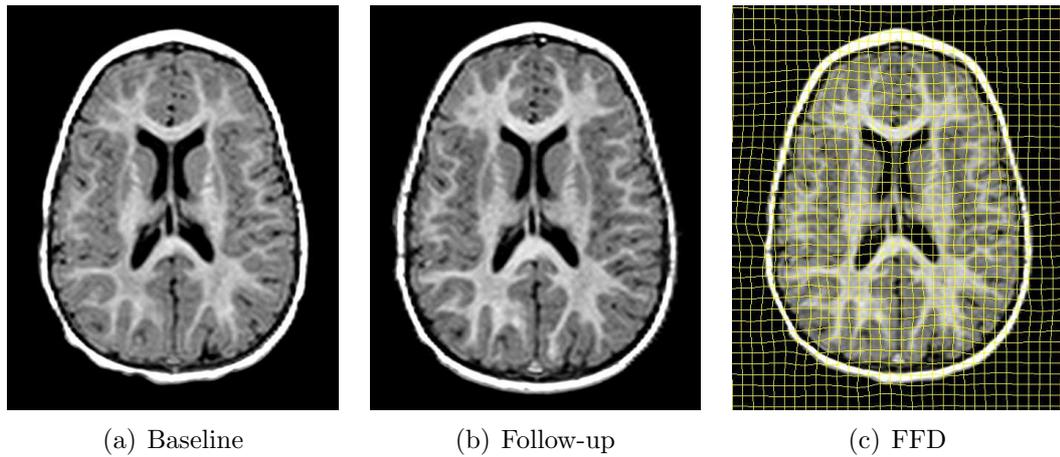


Figure 6.3: The left and middle images show the baseline (year one) and follow-up (year two) images for a particular subject. The right image shows the follow-up image after non-rigid registration along with the grid used for the free-form deformation. *An animation depicting the growth from one to two years for the subjects studied is available on the accompanying CD ROM.*

shown in Figure 6.4. Figure 6.4(a) shows the difference image after a rigid alignment. Figures 6.4(b) and 6.4(c) show the difference image after affine and non-rigid registration respectively.

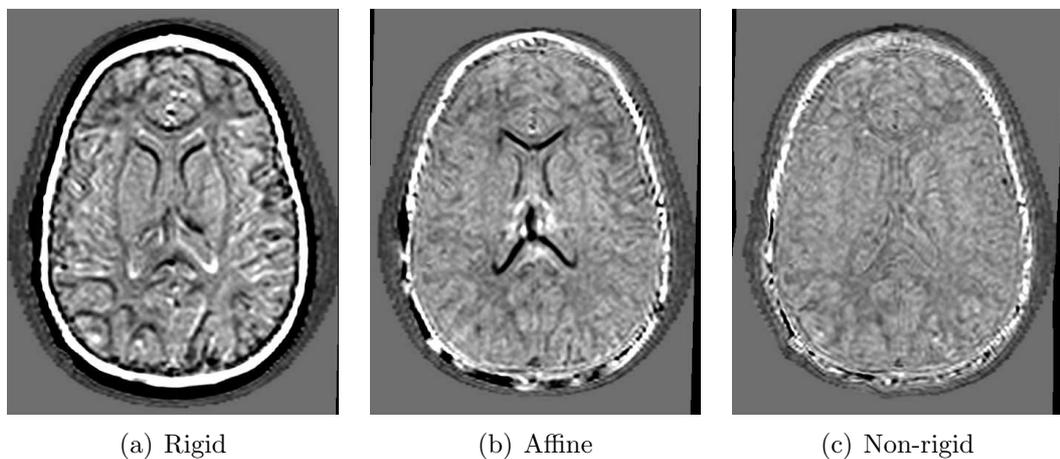


Figure 6.4: Difference images between the baseline and follow-up images after a rigid alignment and after the affine and non-rigid registration steps. These images are of the same subject shown in Figure 6.3, i.e. 6.4(a) shows the difference between Figures 6.3(a) and 6.3(b).

As described in Section 3.1.2, a weighting value of 0.01 was chosen for the regularisation parameter λ (Equation 2.5). The effect of this choice of regularisation parameter was evaluated by finding the percentage of voxels with a negative

Jacobian determinant for each of the various transformations. Assuming that there are no differences in the topologies of the anatomies represented in the images, the ideal transformation should not tear or fold the underlying space, i.e. the Jacobian determinant values should be positive and finite. Based on this assumption, longitudinal registrations for the data studied should have no regions with negative Jacobians while these may exist for inter-subject registrations. This is because the growth in the interval studied is unlikely to break topology while there may be topological variation between subjects.

For the longitudinal transformations, between the baseline and follow-up images, two of the twenty-five subjects' transformations contained negative determinants and these occurred in less than $10^{-3}\%$ of the voxels. For the cross-sectional transformations used to create the atlases, the average percentage of voxels with negative Jacobian determinants was 0.03% (SD 0.06%, maximum 0.3%). The average transformations used to create the average space atlases did not generate any negative Jacobian determinant values.

6.5 Image segmentation

6.5.1 Expectation maximisation

A particular subject was selected and manual segmentations of various structures were outlined by a clinical expert. The manual segmentations for each of the main tissue classes (grey and white matter) and for cerebro-spinal fluid (CSF) were propagated to each of the remaining subjects using an affine transformation and blurred. The transformed images were used as prior probability maps for an EM segmentation for each subject [123]. Validation results for this approach can be found in [124].

An example of the priors for a particular subject are shown in Figure 6.5. Examples of the segmentations produced by the EM algorithm are shown in



Figure 6.5: Prior probability maps for use in an EM tissue classification for a subject. These maps were generated by transforming a manual segmentation for to the target subject using an affine transformation and blurring.

Figure 6.6.

6.5.2 Classifier fusion

A repository of 30 image pairs consisting of T1-weighted images and manual segmentations was used to provide classifiers for structural segmentations. Each atlas contained labels for 83 structures delineated according to a published protocol by Hammers et al. [85] (see Section 4.3.1).

For a given query subject, the 30 annotated T1- images were registered with the subject's baseline scan and the resulting transformations were used to propagate the corresponding manual segmentations of each of the pairs to the space of the subject. Each propagated manual segmentation is treated as an individual classifier for the subject. A simple vote rule was used to combine the classifiers with the label receiving the majority vote at a voxel being assigned to that voxel. An example of a structural segmentation for an individual subject is shown in Figure 6.7.

6.6 Atlases

Using the methods described in Section 2.4, average space atlases of the anatomy across the group were generated for each time-point. These are illustrated in

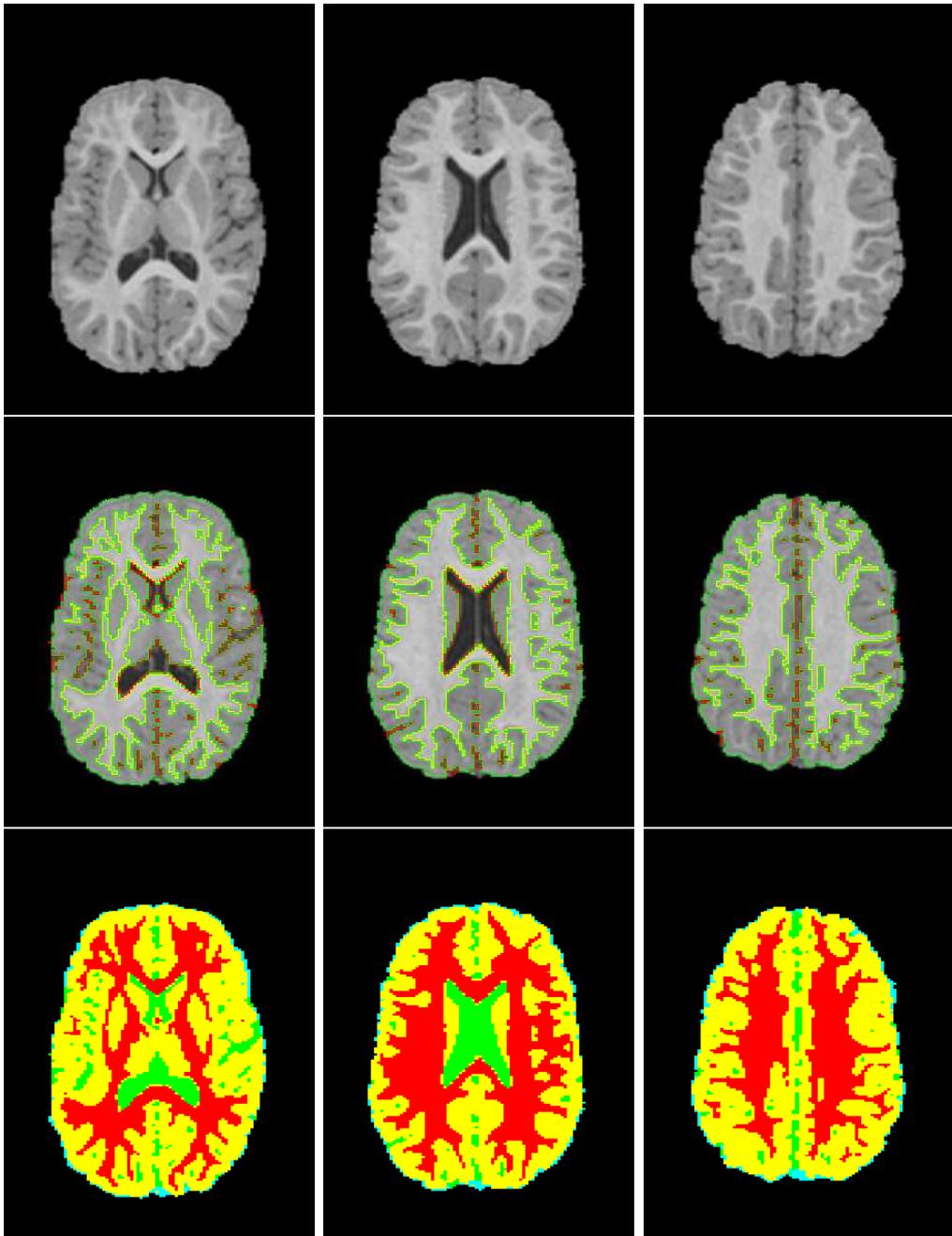


Figure 6.6: Transverse slices of tissue labels generated by EM segmentation for an individual subject. The top row shows slices from the original anatomy. The bottom row shows the labels. The middle row shows the label boundaries overlaid onto the anatomy.

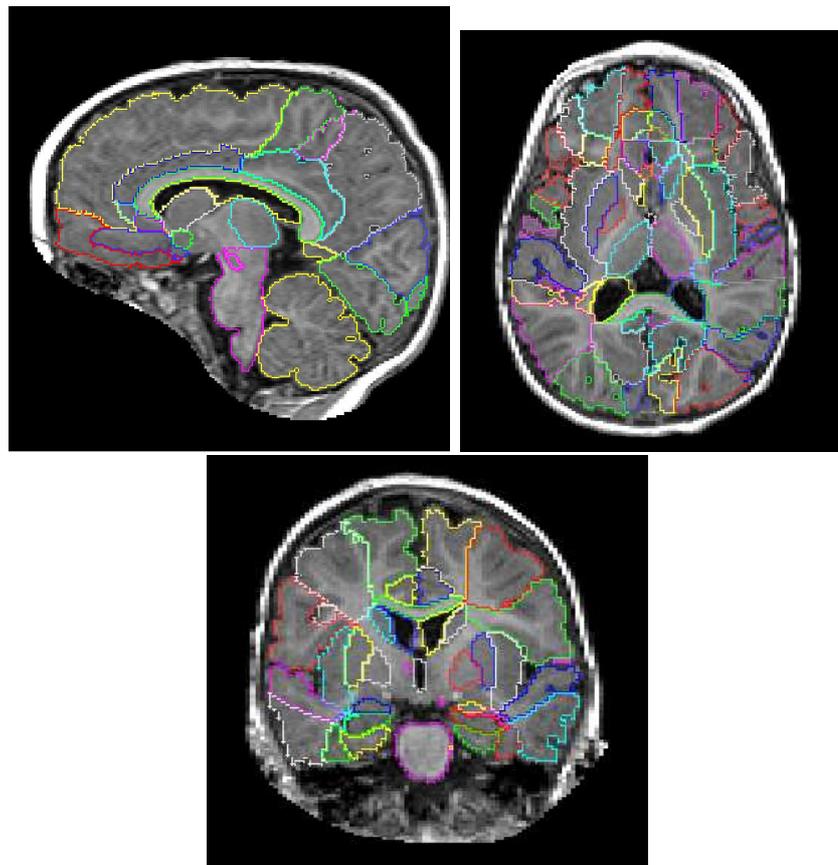


Figure 6.7: An example of a structural segmentation for an individual based on the propagation of labels from a set of classifiers to the subject and subsequent fusion using the vote rule.

Figure 6.8.

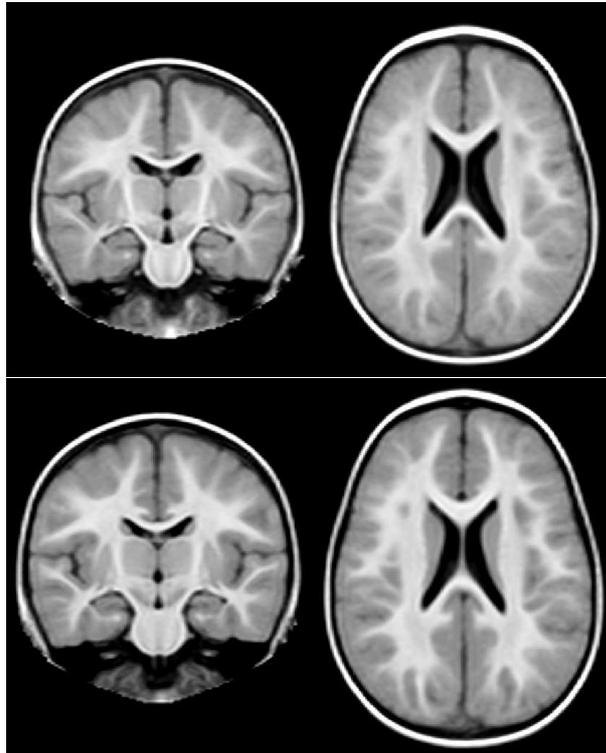


Figure 6.8: The average space anatomical atlases of the cohort based on the one year scans (top) and the two year scans (bottom).

Similarly, in the same coordinate system, an atlas of growth is shown in Figure 6.9 and this growth atlas is overlaid onto the anatomy of the year one average in Figure 6.10. Atlases of the tissue maps across the population at each timepoint are shown in Figure 6.11.

In order to obtain an estimate of the average deformation that is produced by the registration of the year one and year two images, the individual longitudinal registrations for all the subjects were transformed to the space of the year one average space atlas. The method used to map the individual deformation fields to the atlas is described in Rao et al. [133]. After spatial alignment of the deformation fields, their control points were averaged to produce the average deformation field as described in Section 3.4.1. The longitudinal global affine transformation that was combined with the average longitudinal deformation field was the affine map relating the average space atlases at one and two years. This was carried

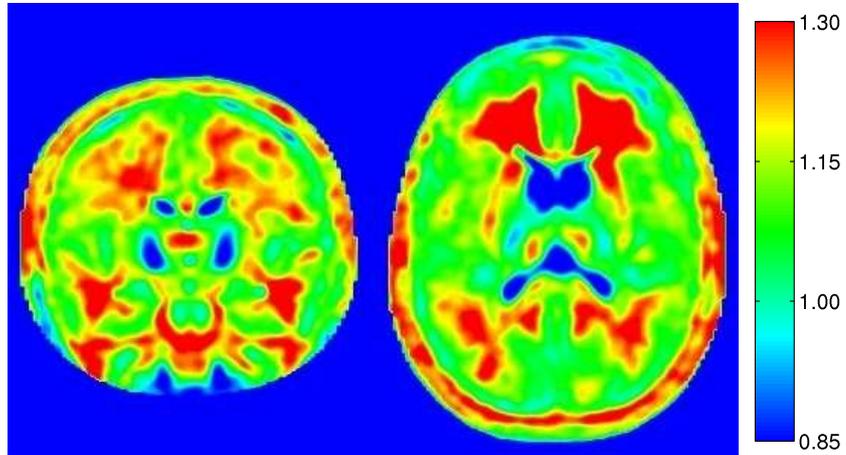


Figure 6.9: The growth map atlas based on all individual growth maps for the one to two year old data. This is in the space of the year 1 (baseline) average space atlas. The growth figures displayed are based on both the global and longitudinal transformations (see Equation 3.1).

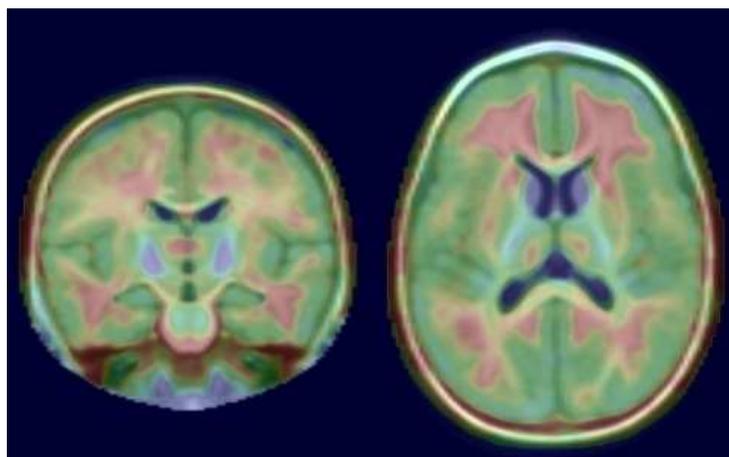


Figure 6.10: The growth map atlas for the one to two year old longitudinal data overlaid onto the average space anatomical atlas for the scans acquired at age one.

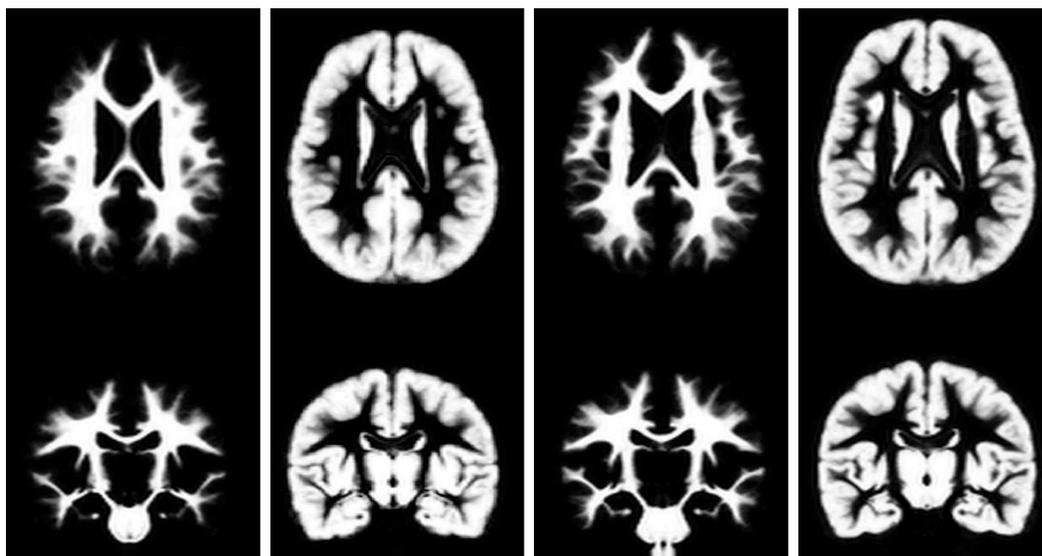


Figure 6.11: The average space atlases of the tissue maps. From left to right: WM year 1, GM year 1, WM year 2, GM year 2.

out as described in Section 3.4.1.3. An illustration of the average deformation is shown in Figure 6.12 where the deformations have been enlarged by 50% in order to give a sense of the patterns of change across the group. The colour scheme applied to the average deformation field is based on the associated Jacobian determinants with the hotter colours representing regions of growth and the colder colours representing contraction (see also Figures 6.9 and 6.10).

An average space atlas was also created from the fused segmentation estimates for the structural labels. The individual fused segmentations for the subjects were used in a second classifier fusion step where each baseline segmentation was propagated to the baseline average space and, once again, combined according to the vote rule. This provides an average space atlas estimate for the structural segmentations which is illustrated in Figure 6.13.

6.7 The effect of the choice of reference

In order to assess the influence of the choice of reference subject used during the construction of the average space atlas, two average space anatomical atlases were generated using two different reference subjects. Each atlas has the same

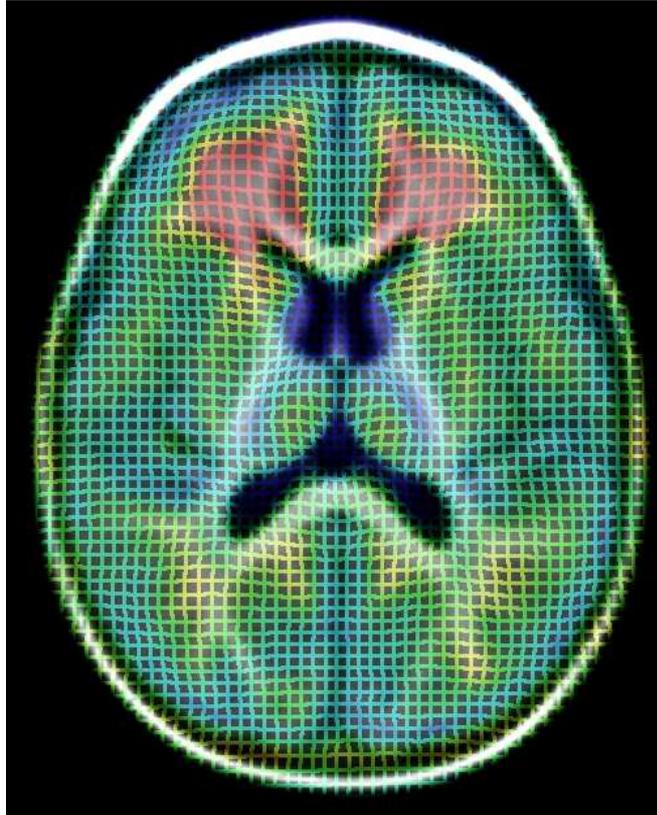


Figure 6.12: An illustration of the average longitudinal deformation across the cohort. The individual longitudinal transformations were mapped to the space of the anatomical one year atlas and linearly averaged. The colouring represents the size of the Jacobian determinant in the same manner as that shown in Figures 6.9 and 6.10.

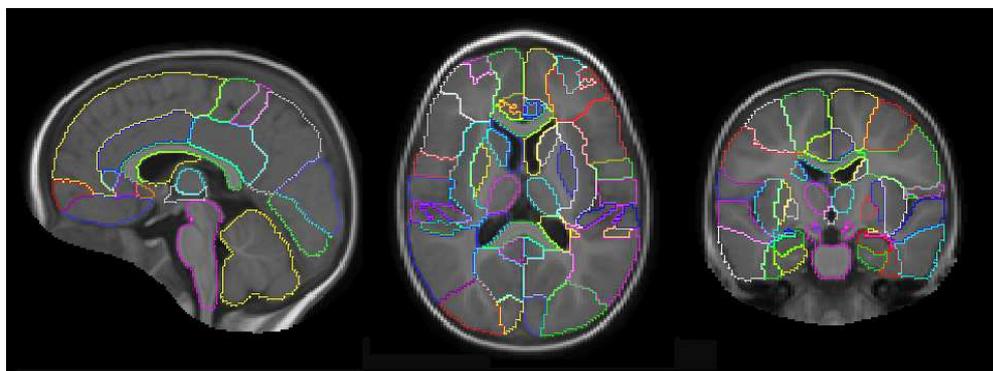


Figure 6.13: The structural segmentation of the average space baseline image. This segmentation was generated by propagation of the structural segmentations of each of the subjects' baseline segmentations to the average space year-one atlas. After propagation, a second fusion step using the vote rule was applied to generate the final structural atlas.

orientation as its corresponding reference (with respect to translation and rotation parameters), so a rigid transformation was used to align the atlases. These atlases are shown Figure 6.14.

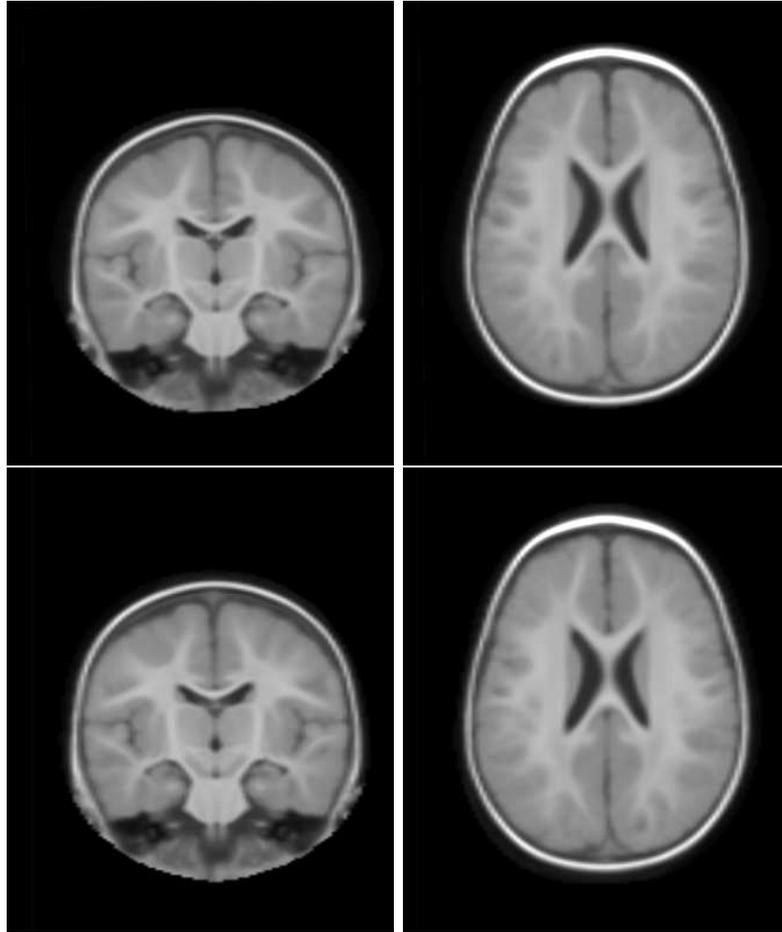


Figure 6.14: Two average space atlases generated from the same subjects. The atlases were generated using two different reference subjects, i.e. two different choices for the subject labelled I_{ref} in Figure 3.4. The atlases were rigidly aligned using the rigid transformation relating the corresponding reference subjects.

The similarity of these atlases can be shown by overlaying their iso-intensity contours. Figure 6.15 shows two sets of iso-intensity contours that correspond to the same intensity taken from each of the atlases after rigid alignment. Further iso-intensity contours derived from the average space atlases are also shown in Figure 6.16. There are some slight differences between corresponding contours from the different atlases but they generally agree and demonstrate the robustness of the process of generating the average space atlases with respect to the choice

of reference.

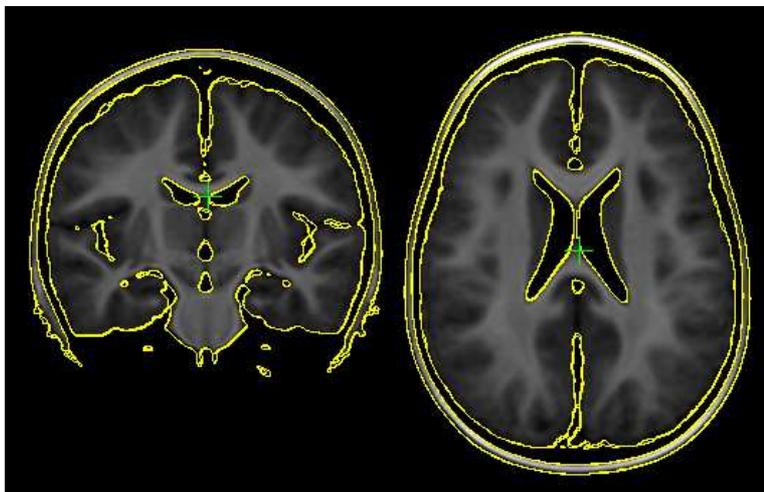


Figure 6.15: Two corresponding iso-intensity contours taken from each of the pair of average space atlases shown in Figure 6.14. These atlases were generated using two different reference subjects, i.e. two different choices for the subject labelled I_{ref} in Figure 3.4. The atlases were rigidly aligned using the rigid transformation relating the corresponding reference subjects. Further iso-intensity contours using different values are shown in Figure 6.16.

6.8 Comparison of registration and segmentation based growth estimates

The tissue segmentations estimated via EM can be used to provide volume estimates for grey and white matter for each subject at each timepoint. These are plotted in Figure 6.17. In this figure, the baseline and follow-up volume estimates for a particular tissue that correspond to the same subject are joined with a dashed line. This is simply to show corresponding volumes and not intended to suggest that the growth is linear. The mean volumes of grey matter for the cohort were $665.7 (69.7) \text{ cm}^3$ at age one and $746.7 (81.9) \text{ cm}^3$ at age two, where the figures in brackets indicate standard deviations. The corresponding volumes of white matter at ages one and two were $198.7 (18.6) \text{ cm}^3$ and $254.8 (26.7) \text{ cm}^3$.

The segmentation algorithm provides maps for grey matter, white matter and cerebro-spinal fluid. Since the focus of this work is on quantifying tissue

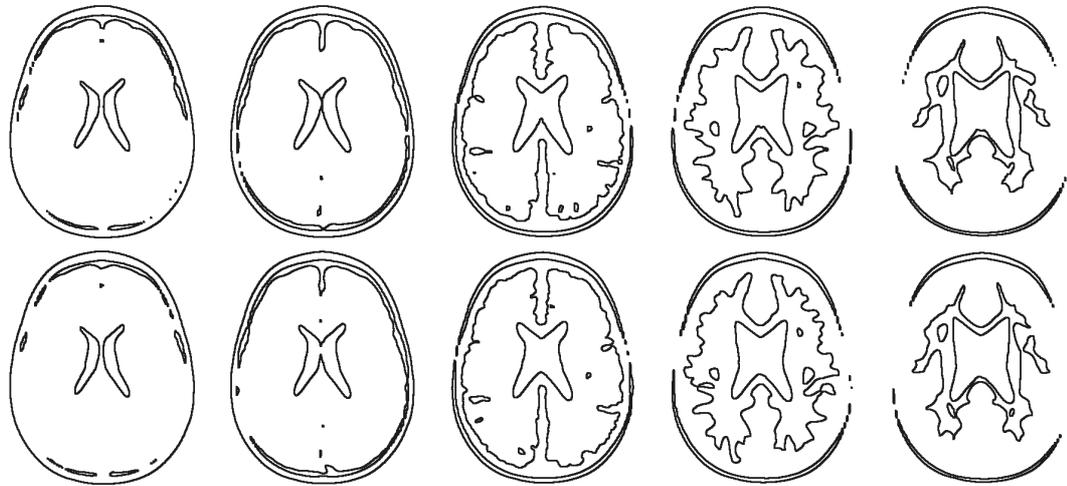


Figure 6.16: Further iso-intensity contours corresponding to the same levels taken from two average space atlases. The atlases were generated using two different reference subjects. Each row corresponds to one of the atlases and each column corresponds to an iso-intensity level (see Figure 6.15).

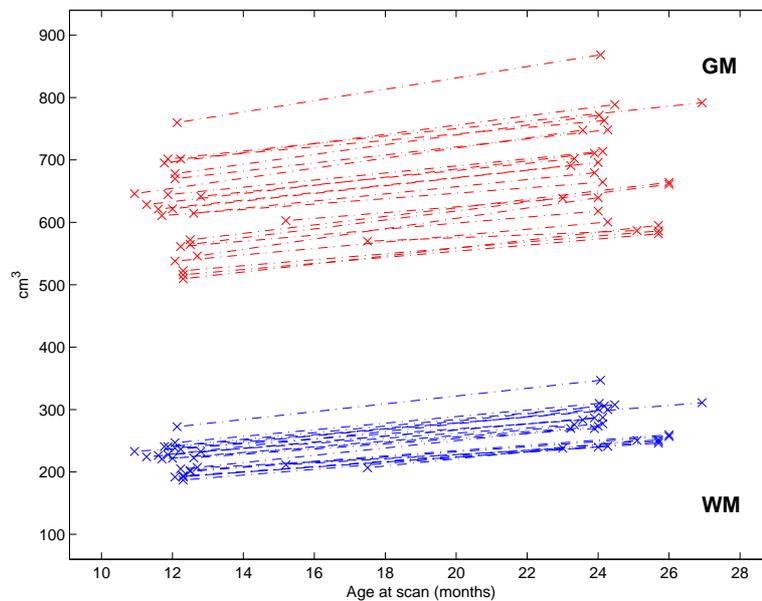


Figure 6.17: Tissue volumes at ages one and two based on the segmentation results. The baseline and follow-up volumes of each tissue for an individual are joined with a dashed line to show the correspondences between the estimates.

growth, only the grey and white matter maps have been used to estimate growth figures. Using the segmentation results, the growth factor for a tissue class for an individual can be estimated by finding the ratio of the volumes given by the tissue segmentation algorithm within the baseline and follow-up images. Alternatively, using the registration results, the growth factor for a tissue can be estimated using the Jacobian determinant map and the baseline tissue segmentation (see Equation (3.4)).

In order to assess consistency, the growth factors estimated by registration for each tissue class were compared against the growth factors obtained from segmentation. This comparison between growth factors provided by registration and segmentation is illustrated in Figure 6.18 and shows good agreement between the two methods. Table 6.1 gives a summary of the growth factors obtained by both methods for grey and white matter.

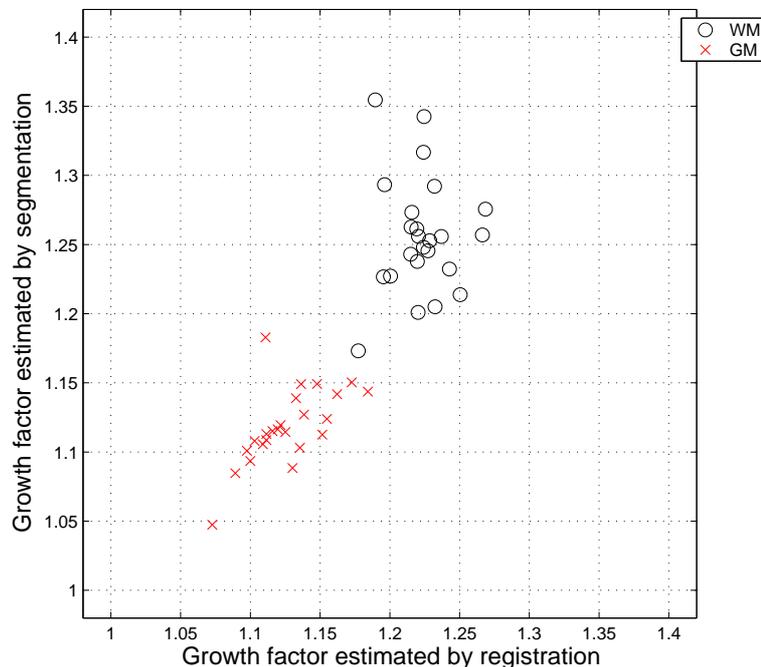


Figure 6.18: GM and WM growth factor estimates for all subjects. This chart shows the general agreement between estimates of growth provided by registration and those provided by segmentation. The registration-based figures are derived using Equation (3.4) and the segmentation-based figures are the ratio of the volumes of each tissue at the two timepoints. The mean absolute difference (MAD) between the growth factor estimates based on registration and segmentation were 0.042 for white matter and 0.017 for grey matter.

Growth factors	Grey Matter	White Matter
Segmentation	1.12 (0.030)	1.28 (0.049)
Registration	1.11 (0.035)	1.26 (0.046)

Table 6.1: Growth factor estimates obtained by segmentation and registration. The segmentation estimates were obtained from the ratio of volumes of each tissue at each age and averaged across subjects. The registration estimates were obtained by integration of Jacobian determinants as described in Equation 3.4. Standard deviations are shown in brackets.

A further assessment of consistency is given in scatter plots that compare registration and segmentation estimates of the growth of grey and white matter in more detail - these are shown in Figure 6.19. In Figure 6.19 subjects for whom the difference in registration and segmentation based estimates represented outliers are highlighted. Outliers were identified by considering the upper and lower quartiles of the signed differences between the registration and segmentation based estimates. The outlier limits were set at $1.5 \times$ the inter-quartile range (IQR) above the upper quartile and $1.5 \times$ IQR below the lower quartile¹. If the outlier subjects are excluded, the mean absolute differences (MAD) for the growth factor estimates given by the two methods are 0.037 for white matter and 0.014 for grey matter. Bland Altman plots for the inlier subjects are shown for the growth estimates of each tissue in Figure 6.20. No systematic bias is apparent in these plots.

6.9 Regional volume changes and clinical information : Preliminary investigation

For some of the subjects in the group, data relating to clinical outcome were also available. The available clinical data were gestational age (GA), birth-weight (BW) and the development quotient (DQ). These data were available for 17 of the 25 subjects. The development quotient was obtained by carrying out neuro-

¹see e.g. <http://mathworld.wolfram.com/Outlier.html>

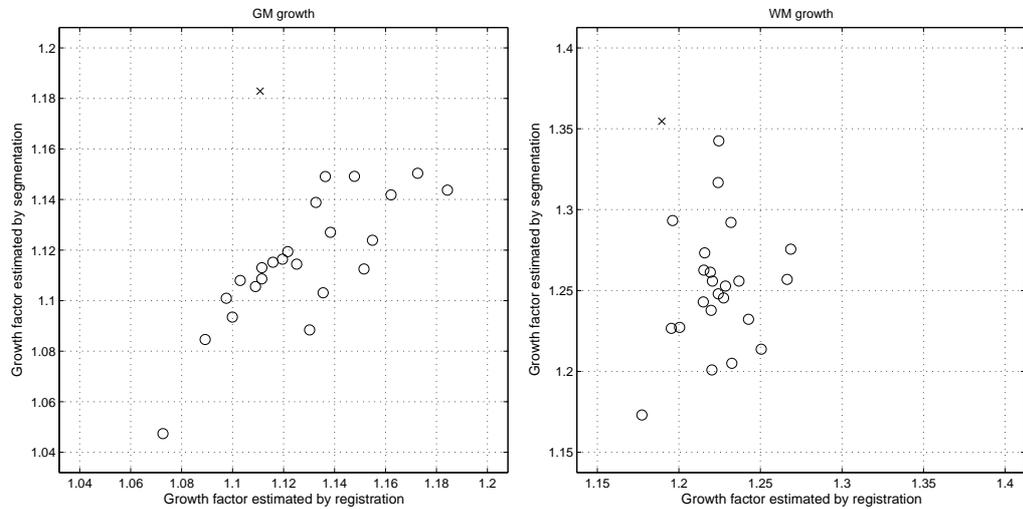


Figure 6.19: Scatter plots of tissue growth factors showing outliers. The upper and lower quartiles were found for the differences between the segmentation- and registration-based methods and outlier limits were set at $1.5 \times \text{IQR}$ above the upper quartile and $1.5 \times \text{IQR}$ below the lower quartile where IQR represents the inter-quartile range. The factors for these subjects are shown as crosses, the factors for the remaining subjects are shown as circles.

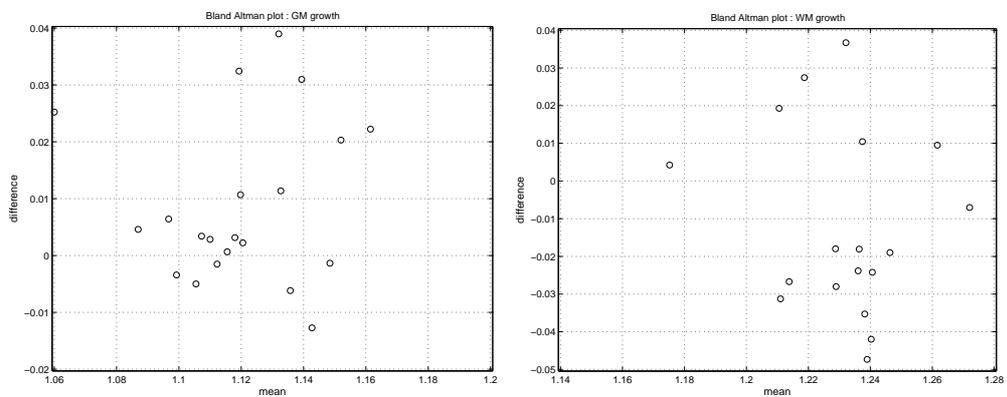


Figure 6.20: Bland-Altman plots for the growth factor estimates produced by registration and by segmentation.

developmental assessment using the Griffiths Mental Development Scales [79] at 2 years corrected age. The Griffiths Mental Development Scales provide an overall developmental quotient (DQ) with sub-scales assessing skill areas (locomotor, personal-social, hearing and speech, eye and hand co-ordination, performance). The mean (SD) DQ score for the general population is 100(12).

For these subjects, an initial exploration of possible relations between growth and clinical data was made. The availability of structural segmentation estimates for each subject allows the growth figures for particular regions to be estimated.

The growth figures for the subjects with clinical data were separated into sub-cortical and cortical growth. This was made possible by aggregating structural segmentations to provide cortical and sub-cortical masks and integrating the Jacobians over these masks.

The structures that were grouped to form a sub-cortical mask were : hippocampus, amygdala, brainstem, caudate nucleus, accumbens, putamen, thalamus, pallidum, corpus callosum, lateral ventricle and substantia nigra. The cortical and cerebellar grey and white matter were grouped to form cortical masks.

In order to mitigate against possible sources of error, the regional growth figures obtained were filtered prior to further analysis. The growth figures obtained by the two methods of registration and segmentation were compared for each subject. This was carried out separately for the cortical and sub-cortical regions. Subjects with relatively high degree of inconsistency between the two growth estimation methods were treated as outliers and removed. This was carried out by finding the upper and lower quartiles of the signed differences in the method estimates and setting limits at $1.5 \times \text{IQR}$ above the upper quartile and $1.5 \times \text{IQR}$ below the lower quartile. The outliers were identified separately using the cortical and sub-cortical growth estimates and combined to form a single group before removal.

After removing the outliers for each region's growth estimates, correlation coefficients were calculated for the growth figures against each of the clinical

	Cortical growth	Sub-cortical growth
DQ	0.66	0.18
GA	0.34	0.38
BW	0.10	0.34

Table 6.2: Correlation coefficients calculated for regional growth estimates (Cortical and sub-cortical) against clinical factors (gestational age - GA, birth weight - BW, development quotient - DQ).

factors available. These figures are summarised in Table 6.2. The results indicate correlation between cortical growth and the development quotient ($r^2 = 0.66$) and between sub-cortical growth and gestational age ($r^2 = 0.38$). The specific data for these correlations are shown in Figures 6.21 and 6.22. Although the number of available data points is low, it is of interest that there may be links among these different factors.

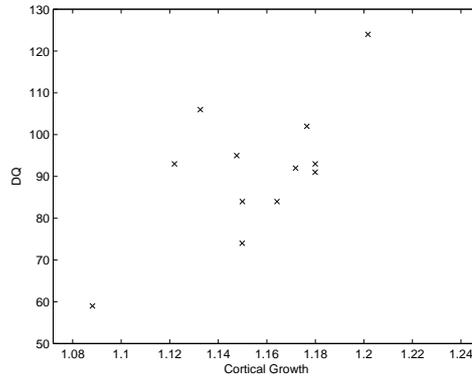


Figure 6.21: A plot of the growth figures for the cortical region against the development quotient. There appears to be a weak positive correlation between these factors ($r^2 = 0.66$).

6.10 Discussion

The results presented in this chapter show how deformation based morphometry via the registration of longitudinal data can be used to generate quantitative estimates of brain growth for a cohort and for individual subjects during the second year of life. The data to which these methods were applied are images

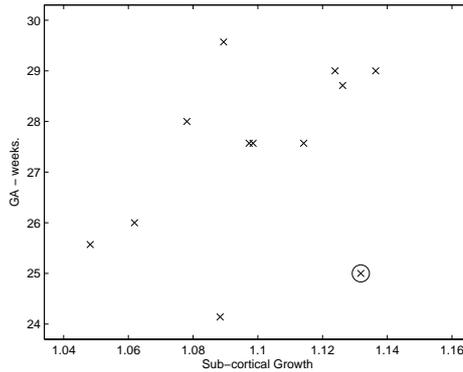


Figure 6.22: A plot of the growth figures for the sub-cortical region against gestational age. There appears to be a weak positive correlation between these factors ($r^2 = 0.38$). If the circled subject is treated as an outlier, the correlation coefficient rises to $r^2 = 0.63$.

of preterm born children. Data for term born controls is relatively scarce (as it is difficult to obtain ethical permission for sedation during scanning) but the techniques used to estimate growth can be readily transferred to such data.

Using Jacobian determinant maps that represent growth on a per-voxel basis, growth estimates for white and grey matter for each individual and for the cohort as a whole have been calculated. Among the 25 infants studied, there was an increase of around 11 – 12% in grey matter volume and an increase of 26 – 28% in white matter (see Figure 6.19 and Table 6.1).

Comparison between tissue class growth estimates obtained via registration and from tissue segmentation shows excellent agreement on grey matter growth but higher variability in the estimates for white matter growth. This suggests a possible increased susceptibility to error in white matter estimates which could be attributed to one or more of a variety of factors: For example, the relatively smaller volume of white matter, changes due to myelination or partial volume effects. Additionally, there are fewer features in white matter tissue compared with grey matter where images typically display stronger edges, for example within the sulci and gyri.

The volumes of each tissue at each timepoint, as shown in Figure 6.17, illustrate the correspondences between baseline and follow-up volumes. Without

making any assumptions about the best model for the growth (e.g. linear, non-linear etc.), these correspondences suggest that development at this age follows what is described as a ‘parallel trajectory’ by Kraemer et al. [107]. Allometric studies of brain development (for example [102] and [186]) are made possible by these data, and the DBM approach has been successfully used for studies of more immature brains, and thus a combination of the DBM approach with automatic brain labelling holds the promise of a large scale neuroinformatic understanding of brain development during this period of rapid growth.

In particular, the preliminary data shown in Section 6.9 show the interesting possibilities for exploration in this area. Based on the small number of subjects for which the data were available, correlations between clinical factors such as DQ and GA and regional growth estimates may be present and should provide pointers for future studies with larger numbers of subjects.

A method based on inter-subject registrations and transformation averaging has been used to generate average space atlases for the cohort at each timepoint. These atlases can represent anatomy, tissue class maps, structural segmentations or growth across the cohort at each timepoint. In general, across the population, the pattern of change is mainly represented by relatively higher growth within white matter, especially for the anterior white matter region (see Figure 6.10) and a general contraction in the ventricular region which can be observed, for example, by comparing the one and two year images in Figure 6.8.

A visual comparison of two average space atlases generated using two different reference subjects (Figure 6.15) shows them to be in close agreement. This suggests that the bias towards the chosen reference has been significantly reduced. The approach used to create the atlases is similar to that used by Guimond et al. [80] and Rueckert et al. [139] who both showed that the atlases generated are reasonably robust to the choice of reference when the transformations being averaged are the local deformation fields. The work described in this section incorporates the global affine transformations between subjects into the averaging

process and the results indicate that the robustness to the choice of reference remains.

Longitudinal transformations have been used to estimate growth using the Jacobian determinants and an illustration of the average longitudinal deformation field for the cohort was given in Figure 6.12. There are many other properties of the longitudinal deformation fields that can be extracted and subsequently used to answer questions about the patterns of development. For example, the curl or divergence of the fields can be calculated or estimates of the main modes of variation in the fields can be made [139]. These options were beyond the scope of this work but should provide a very interesting area for future research.

6.11 Conclusion

This chapter has presented examples that show how patterns of longitudinal change can be characterised by combining methods based on registration and region of interest identification. Identifying regions of interest can be implemented via general tissue segmentations or by structural segmentations achieved via atlas propagation and fusion. These methods have been applied successfully to characterise patterns of growth in young children between the ages of one and two.

The combination of growth data and structural segmentation allows the identification of patterns of growth in particular regions. The ability to identify regional change estimates and the availability of outcome or clinical data mean that higher levels of analysis are made possible. For example, Section 6.9 presented data that suggest some correlation between estimates of cortical and sub-cortical growth and clinical factors in the group of young children – this may represent an interesting area for further study. This suggests that further analysis using larger datasets and more powerful statistical methods (e.g factor analysis), or methods based on data-mining, may shed further light on the patterns of physiological

development in young children that may inform clinical diagnosis and decision making.

Chapter 7

Longitudinal change due to atrophy

The previous chapter presented data on the patterns of growth for a cohort of children scanned at one and two years of age. The work presented in this chapter describes a second set of experiments on identifying longitudinal change that was carried out using serial data acquired from a group of patients with Alzheimer's disease (AD) and a group of age-matched controls. The experimental set-up for these data shares many similarities with that described in the experiments on growth in early childhood. The use of longitudinal registrations for growth estimation and inter-subject registrations for atlasing is the same as that described in Section 6.3 and illustrated in Figure 6.2. Longitudinal change can once again be estimated by the integration of Jacobian determinants using suitable masks.

The main differences between the two sets of experiments are represented by the amount of change over the interval studied and by the availability of more distinct clinical classifications for the subjects. The amount of change in the brain due to atrophy in the elderly in general and Alzheimer's patients in particular is much smaller than that due to growth in the early years. The subtle nature of the changes in the elderly presents a challenge to schemes that aim to measure it. The subjects studied in this section also fall into much more distinct clinical

categories than those for the children studied in earlier sections. The subjects are clearly defined as (AD) patients or controls. The younger subjects described in the previous chapter had clinical information that generally belonged to a continuous scale (e.g. gestational age or birth weight) with no obvious discrete boundary. The categorisation of the elderly subjects into patients and controls allows experiments based on group separation and classification to be carried out using information obtained from processing their image data.

In the following sections, a description of the background and previous work in this area is given and the data and experiments carried out using the images acquired from the AD patients and the controls are presented.

7.1 Atrophy due to ageing and Alzheimer's disease : background and previous work

The incidence of Alzheimer's disease (AD) is generally found among older sections of the population, typically 65 years of age and above. Around 60% of the estimated 24 million worldwide cases of dementia are attributed to AD [56, 122] and increases in longevity mean these figures are likely to rise. AD is characterised at the histological level by the build up of plaques of a misfolded protein known as beta amyloid and the development of aggregates of proteins known as neurofibrillary tangles within neurons. These cellular effects of AD generally occur for a significant period prior to the diagnosis of the disease, which is typically identified through its cognitive and behavioural effects. The microscopic changes due to AD can be observed at a larger scale as a general atrophy of brain tissue, which means that there is a role for medical imaging in the study and management of AD: it can provide supplementary data to inform clinical decisions and assessments of the efficacy of treatments.

The use of serially acquired data allows for estimates of atrophy to be made

using fewer subjects than would be required by a ‘snap-shot’ or cross-sectional study. Serial acquisitions compensate for the inter-subject variability.

Using serially acquired images, the most straightforward way of detecting and measuring atrophy is for experts to manually delineate the brain tissue in each of the scans acquired from a subject. The difference between the delineated tissue masks then represents a measure of the atrophy that has occurred between scans; if M_1 represents a baseline tissue mask, and M_2 is the follow-up tissue mask, then the voxels used to estimate atrophy are represented by the difference between the union and intersection of the masks, $(M_1 \cup M_2) \setminus (M_1 \cap M_2)$.

The difficulty with such an approach lies partly in the time-consuming nature of producing expert outlines of brain tissue and also in the natural inter- and intra-rater variability of the raters.

An approach known as the Boundary Shift Integral (BSI) was presented by Freeborough and Fox [65] to address this problem. BSI focuses on the voxels in the difference between the masks and determines if they truly represent atrophy by taking their intensities in each of the baseline images into account. If the voxel represents atrophy then, in T1-weighted images, the transition of its intensity from the baseline to the follow-up image will be from a high value to a low value. By integrating the signed intensity transitions for voxels in the difference, an estimate of the atrophy can be made. BSI is a well-established technique. Variations of the technique continue to be explored [144]. It has also been used within clinical trials [61].

A method known as SIENA (Structural Image Evaluation using Normalisation of Atrophy, Smith et al. [154]) estimates atrophy in a similar way. It uses an automated technique to estimate a mesh-based representation of the brain boundary in the baseline and estimates the amount of atrophy by the movement required to align the meshes. A development of this method by Wang et al. [171] incorporates partial volume information in the voxels at the boundary.

A different approach is to register the baseline and follow-up images non-

rigidly and use the Jacobian determinants of the resulting transformation to estimate the change in brain volume between scans by integrating the determinants over the region of the baseline mask. Work by Boyes et al. [26] presented measurements of atrophy determined by Jacobian integration of longitudinal non-rigid registrations in comparison with BSI. Both methods estimated comparable levels of atrophy.

The progression of AD typically starts earlier in the hippocampus [151, 142]. The hippocampus has therefore been the focus of many studies that use serial data to investigate change in this particular structure [172, 46].

Ashburner et al. [5] present a good general overview of morphometric methods that can be used for the analysis of healthy and diseased brains.

7.2 Data

The MIRIAD (Minimum Interval Resonance Imaging in AD) data set, generated at the Dementia Research Centre (Institute of Neurology, UCL, UK), includes T1-weighted MR scans acquired from 58 subjects. Thirty nine subjects were diagnosed with Alzheimer’s disease and had a mean age of 69.3 years at first scan date (SD 7.0). Nineteen age-matched controls (mean age 69.7, SD 7.2) also took part in the study. The Local Research Ethics Committee granted approval for the study and the subjects gave written informed consent.

Volumetric MR scans were acquired coronally on the same 1.5T Signa Unit (GE Medical Systems, Milwaukee) using an inversion recovery (IR)-prepared spoiled GRASS sequence. Imaging parameters were as follows: TE, 6.4 ms; TI, 650 ms; TR 3000 ms; bandwidth 16 kHz; $256 \times 256 \times 124$ matrix; $240 \times 240 \times 186$ mm field of view (FOV).

During the baseline visit, two scans were acquired sequentially (i.e. with the subject remaining in the scanner) and a single scan was acquired during each of two follow-up visits at six months and one year after the baseline scan date.

7.3 Mask generation

When using free-form deformations (FFDs) to measure subtle longitudinal changes such as atrophy, the mask over which the Jacobian determinants are to be integrated needs to be carefully defined. For the purposes of this work, segmentations of the main tissue types were carried out for all the subjects using expectation maximisation (EM) [177, 109] and structural segmentations were carried out using classifier fusion.

The output from the EM algorithm was used to create whole brain masks of grey and white matter tissue. The probability maps of grey and white tissue were added together and thresholded at 50%. Examples of the tissue masks can be seen for an Alzheimer’s subject in Figure 7.1 and for a control subject in Figure 7.2.

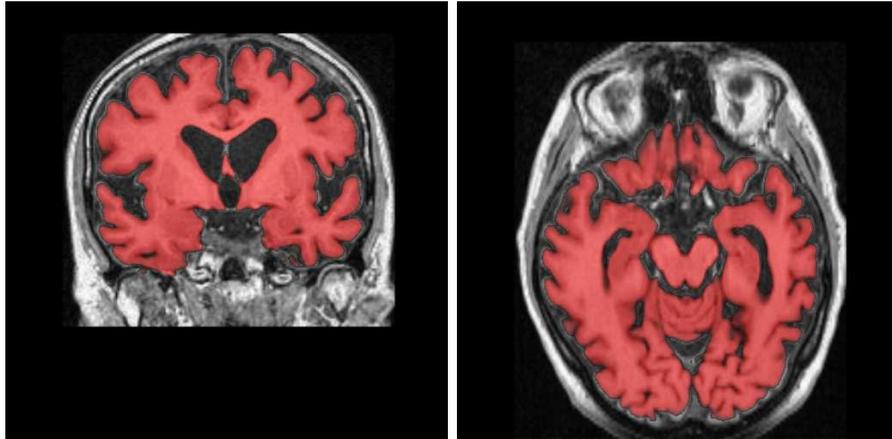


Figure 7.1: An example of a tissue segmentation for an Alzheimer’s patient.

The CMA dataset was used to provide structural classifications of the MIRIAD subjects. As described in Section 5.2, this dataset consists of 275 T1-weighted MR images acquired from male and female subjects with a range of ages. Each image was manually segmented into various structures by an expert rater. The CMA images and their corresponding segmentations were treated as the repository of atlases. For each query subject in the MIRIAD dataset, a set of classifiers was selected from the repository using the hybrid similarity selection scheme (H) described in Section 4.8.1.

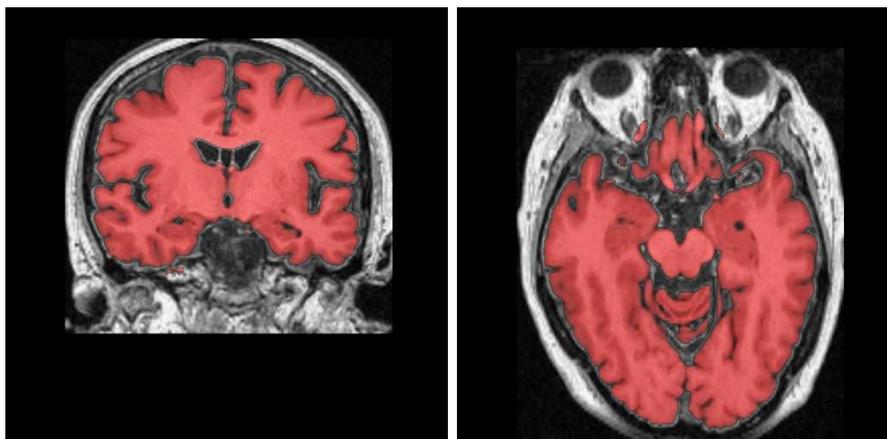


Figure 7.2: An example of a tissue segmentation for a control subject.

Under the hybrid scheme, the repository images and the query were all spatially normalised to a reference (the MNI single subject image) and the similarity of each repository image with the query was calculated. The similarity metric used was normalised mutual information (NMI) and the region over which this was calculated was a mask encompassing the sub-cortical structures as described in Section 5.3. The top 20 subjects from the repository ranked by NMI were then registered directly to the query subject, and the resulting transformations were used to propagate their labels to the query prior to fusion.

Examples of structural segmentations are illustrated for an Alzheimer’s patient in Figure 7.3 and for a control subject in Figure 7.4. Figures 7.5 and 7.6 illustrate the 15 top-ranked classifiers for the same control subject and Alzheimer’s subject shown in Figures 7.3 and 7.4.

In order to generate a mask for use in integration of Jacobian determinants, the grey and white tissue mask defined by the EM segmentation was generated and the brainstem was then removed. The estimate of the brainstem was taken from the structural segmentation based on classifier fusion.

The downward extent of the brainstem towards the neck region means that deformations exhibited by the FFD in this area can be affected by the deformations that aim to adjust for differences in the relative cranial and neck positions apparent in the images being registered. It was felt that this would lead to arte-

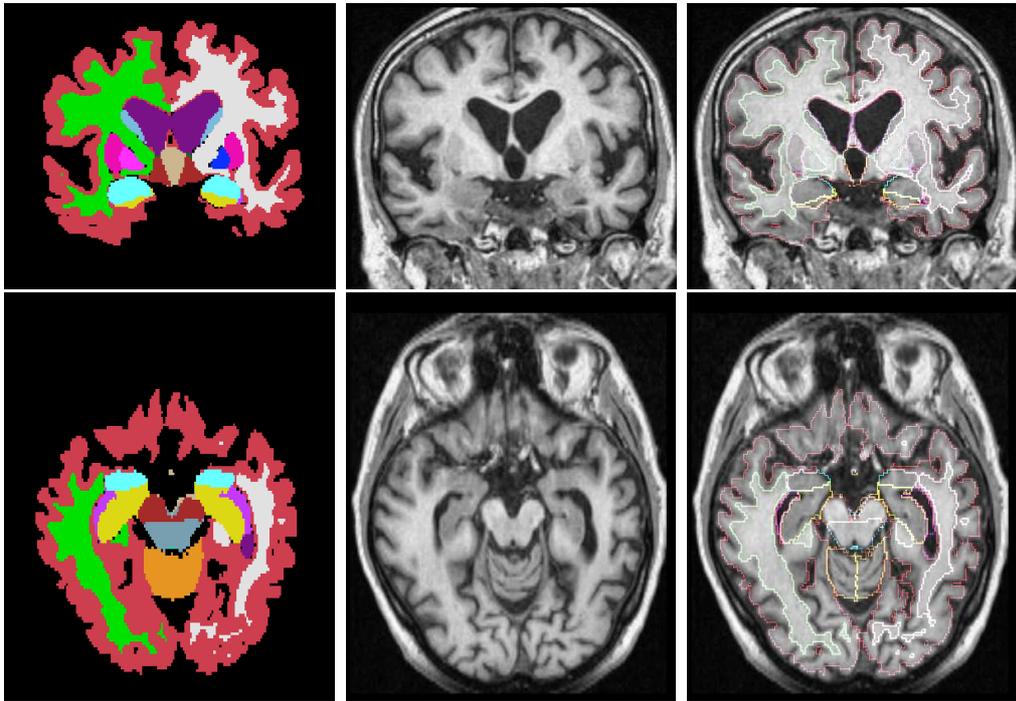


Figure 7.3: An example of a structural segmentation for an Alzheimer's patient.

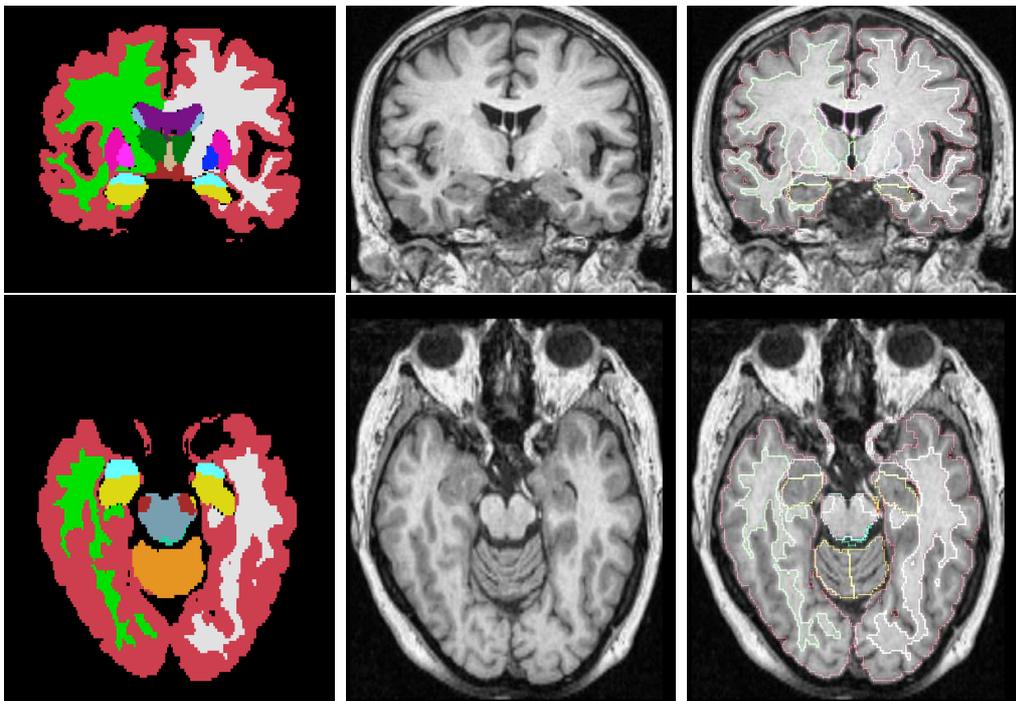


Figure 7.4: An example of a structural segmentation for a control subject.

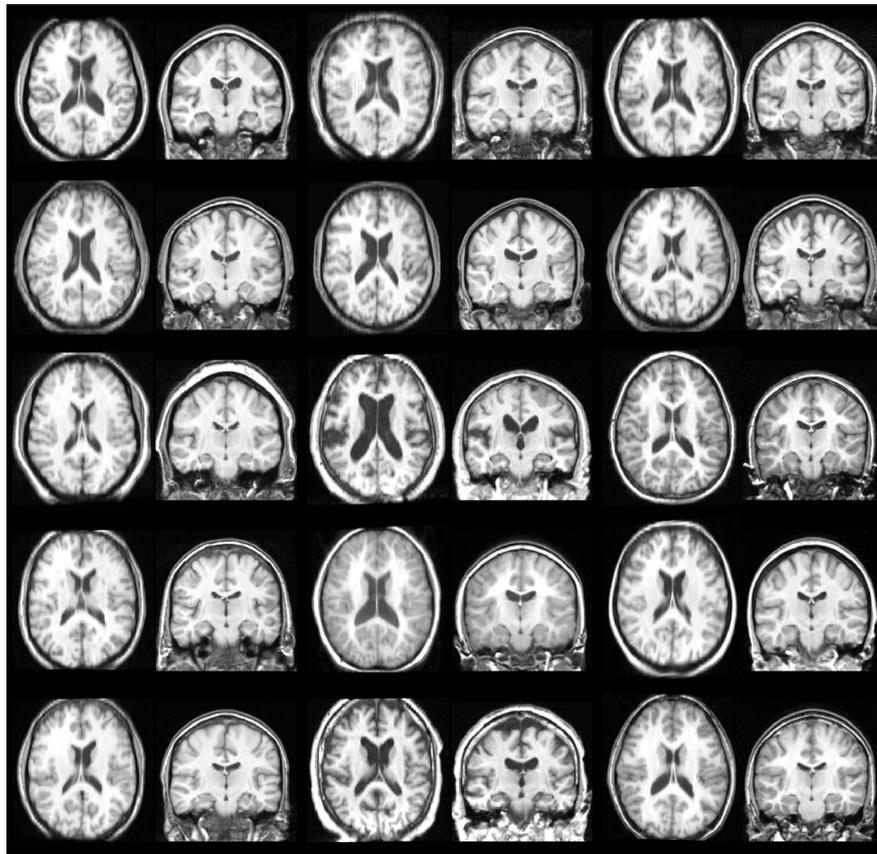


Figure 7.5: The top 15 ranked classifiers for the Alzheimer's patient shown in Figure 7.3. These images were obtained after affine alignment of each classifier subject to the MNI single subject atlas. Each row shows three classifier subjects with a transverse and coronal view of each.

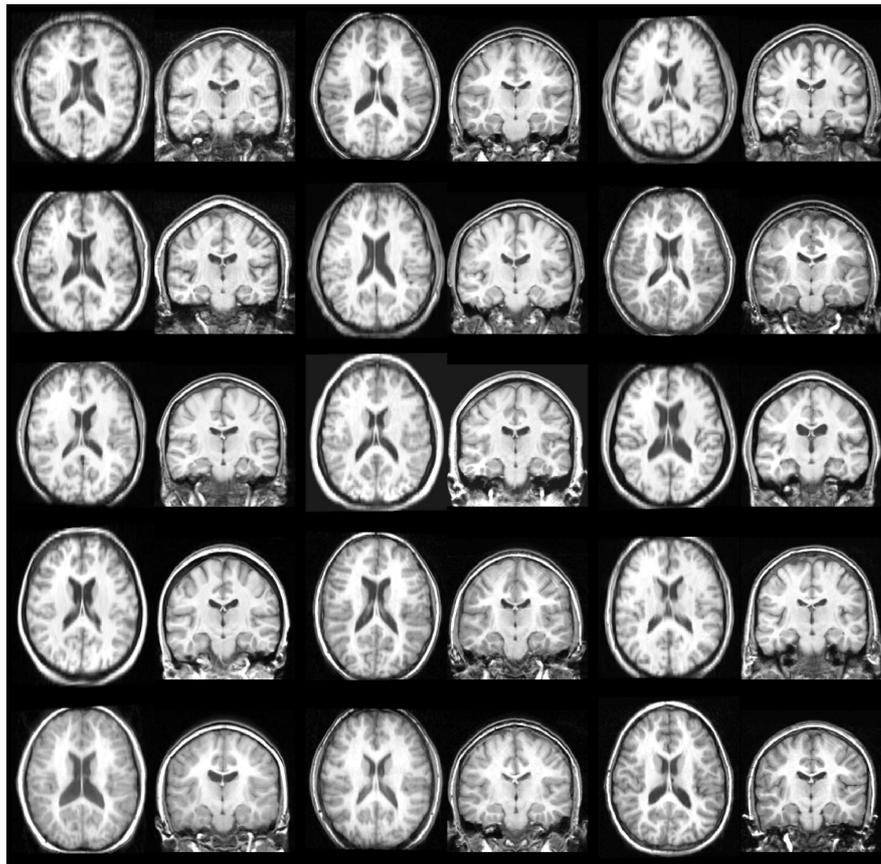


Figure 7.6: The top 15 ranked classifiers for the control subject shown in Figure 7.4. These images were obtained after affine alignment of each classifier subject to the MNI single subject atlas. Each row shows three classifier subjects with a transverse and coronal view of each.

factual estimates of volume change and therefore the brainstem was removed from the mask. The resulting mask for Jacobian integration therefore comprised the sub-cortical structures, white matter, cortical grey matter and the cerebellum.

7.4 The effect of interpolation

The subtle nature of the changes in the brain due to atrophy makes them harder to detect with accuracy. As described in Section 2.2.9, there is a need to choose an interpolation scheme during the registration of an image pair. In previous work, it was shown that the choice of interpolation method can affect the quantity of artefactual deformation generated by a registration [2]. For this reason, the longitudinal registrations carried out between each subject's baseline scan and one year follow-up scan were carried out twice with different interpolators. The interpolation schemes used were standard tri-linear interpolation and a B-spline based scheme known as cardinal interpolation.

Linear interpolation can be achieved by a separable convolution of the samples at the image lattice by a kernel that is defined as

$$\phi_{linear}(x) = \begin{cases} 1 - |x| & |x| < 1 \\ 0 & |x| \geq 1 \end{cases}$$

A standard B-spline interpolator can have varying orders, for example the order 3 B-spline interpolator, $\beta^3(x)$, is defined as a repeated convolution of the order zero B-spline kernel [167]

$$\beta^3(x) = \beta^0(x) * \beta^0(x) * \beta^0(x) * \beta^0(x)$$

where

$$\beta^0(x) = \begin{cases} 1 & |x| < 0.5 \\ 0.5 & |x| = 0.5 \\ 0 & \text{otherwise} \end{cases}$$

The standard B-spline interpolator is known as an ‘approximating’ interpolator because it does not actually interpolate the data at the sample locations but it does have the desirable property of compact support which makes its computation easier. It is possible, however, for a given image, to generate a set of coefficients at the same sample locations that will provide a high quality interpolator to the original image when convolved with a standard B-spline kernel. The generation of these coefficients is described in [162] and [167] who describe the resulting scheme as the cardinal interpolation.

The quality of an interpolator is often described by its spectral properties, i.e. how it filters different parts of the spectrum of an image. The ideal interpolator which does not affect any of the frequencies in a band-limited signal is known as the sinc interpolator. It is known that the linear interpolator acts as a low pass filter which, in effect, blurs the image being interpolated. The cardinal interpolator has been shown to approximate the sinc interpolator [162] with the quality of the approximation improving with the order of the underlying B-spline kernel used. This excellent result means that the application of a compact support B-spline kernel to a set of coefficients derived from the image can approximate a universal support kernel (sinc) applied to the original image data. In the following, the term B-spline will be used to refer to the cardinal interpolator.

After registering each of the baseline images to the one year follow-up images, each resulting longitudinal FFD was used to estimate the rate of atrophy for the subject. The Jacobian determinants of the transformation were integrated over the brain mask for the subject that was generated using the method described in Section 7.3. The availability of two baseline images for each subject meant that there were two longitudinal transformations from which to estimate atrophy

over the one year interval. The atrophy estimates provided by each of these transformations were then averaged to provide a more accurate estimate.

The resulting atrophy estimates are shown in Figure 7.7 as a set of box plots, with a separate plot for each combination of subject group and interpolation scheme. The crosses shown in this plot represent outlier values. A value is defined as an outlier if it is outside the inter-quartile range by a value more than $1.5\times$ the distance between the upper and lower quartiles. As can be seen in Figure 7.7, the use of a B-spline interpolator during registrations gives a slightly improved separation between the groups with respect to measured atrophy. The improved group separation shown by the B-spline based figures can also be observed by applying a t-test to the figures obtained by each interpolator. Using the atrophy figures based on linear interpolation, the t-statistic for the AD subjects and controls is 3.23 ($p = 0.0021$, 56 degrees of freedom, two tailed). The t-statistic using the B-spline based figures is 3.78 ($p = 3.8 \times 10^{-4}$).

The mean values obtained from these atrophy estimates, excluding the outliers, are shown in Table 7.1. In this table, the estimates for linear and B-spline based FFD registrations are shown alongside estimates of atrophy for the same group of subjects recently published by Smith et al. [153]. This work compared the atrophy figures produced by SIENA (Structural Image Evaluation, using Normalisation, of Atrophy [154]) and BSI (Boundary Shift Integral [65]).

It can be seen in Table 7.1 that, using all four methods, there is more variability among the Alzheimer's subjects than among the controls, the standard deviations of the atrophy estimates are in the range 0.35-0.62 for control subjects while the SD range for AD subjects is 0.95-1.34.

Linear interpolation during FFD registration seems to produce generally higher estimates of atrophy for all subjects compared to the other three methods. It gave the highest average estimate of atrophy for the control subjects (0.98%) and the second highest value for the AD subjects (2.09%). Atrophy estimates obtained using B-spline interpolation with FFD registration are generally closer to the fig-

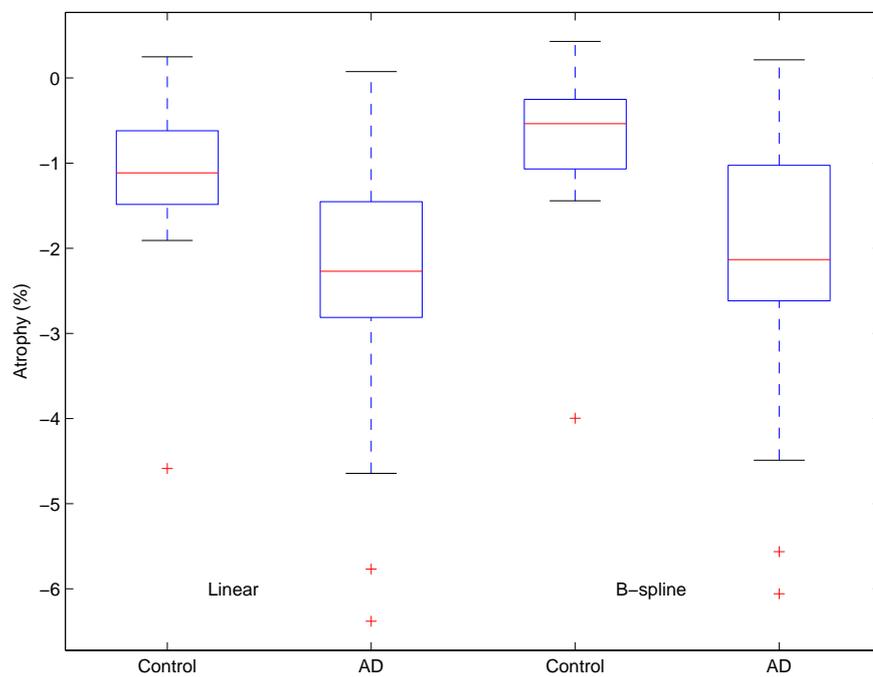


Figure 7.7: Atrophy estimates derived from longitudinal FFD transformations. The left hand pair of plots represent figures derived from registrations using linear interpolation, the right hand pair derive from registrations using B-spline interpolation. The crosses represent outlier values (see text). For both interpolation methods, the Jacobian values of the resulting transformations were integrated over the same brain-mask estimate (see Section 7.3).

ures produced by BSI and SIENA than the figures based on linear interpolation.

Method	Mean atrophy, % (S.D.)	
	Controls	AD
FFD Linear	-0.98 (0.62)	-2.09 (1.05)
FFD B-spline	-0.56 (0.55)	-1.87 (1.09)
BSI	-0.38 (0.35)	-1.80 (0.95)
SIENA	-0.53 (0.45)	-2.43 (1.34)

Table 7.1: Values of atrophy among AD subjects and controls estimated using different methods. FFD Linear and FFD B-spline represent integrated Jacobian determinants following non-rigid registration using linear and B-spline interpolation of the source image. BSI represents the ‘Boundary Shift Integral’ and SIENA represents ‘Structural Image Evaluation, using Normalisation, of Atrophy’ (see text). The BSI and SIENA figures were recently published in [153] and are based on the same image data that is described in this chapter.

In summary, the data presented in this section suggest that B-spline interpolation during FFD registration produces better quality estimates of atrophy for this group of subjects; there is better separation between the clinical groups and the results match more closely those published previously using different estimation methods. For this reason, the results reported in subsequent sections will be based on B-spline interpolation during the registrations.

7.5 Consistency of estimates of longitudinal change

The availability of two scans acquired during the first visit means that an assessment of the consistency accuracy of atrophy measurements can be made. Separate measurements of atrophy can be obtained based on registrations between each of the baseline scans and the follow-up images.

The separate measurements of atrophy for both the six month and the twelve month intervals are shown in Figure 7.8 in the form of Bland-Altman plots. In these plots, no discernible bias of inconsistency in the measurements is apparent.

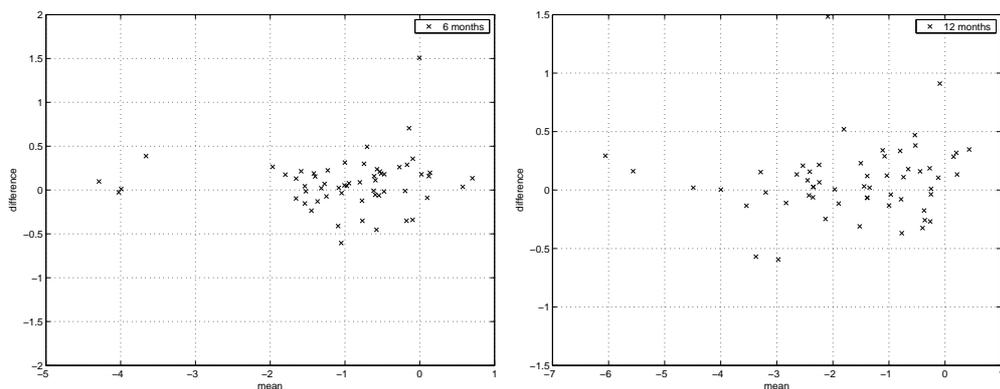


Figure 7.8: Bland-Altman plots to show the level of baseline consistency. The six month and one year estimates for atrophy were estimated twice using the two baseline images as the target of each longitudinal registration. The left hand plot represents the six month data and the right hand plot represents the one year data.

7.6 Atlases

Using the methods described in Section 2.4, a set of average space atlases were generated for the subjects in the MIRIAD data set. The atlases were generated separately for the patient and control groups.

An illustration of the anatomical atlases obtained can be seen in Figure 7.9. In this figure, the first and third rows represent slices taken from the atlas obtained from the Alzheimer’s subjects. Rows two and four of the images show the corresponding slices in the atlas of control subjects. After atlasing, physiological differences between the groups remain apparent. The ventricles appear larger for the AD atlas and the cortical sulci appear deeper and wider. In the first images in row 1 and row 2, more parahippocampal cerebrospinal fluid is apparent in the AD slice compared with the control slice.

Atlases of volume change estimated from the longitudinal Jacobian determinant maps were also created for each group. These are illustrated in Figure 7.10. The slices of the volume change atlases in this figure correspond to the slices shown in the first column of images in Figure 7.9. The different characteristics of the patterns of volume change between the groups is apparent in these volume change maps. The AD subjects clearly show greater levels of expansion within

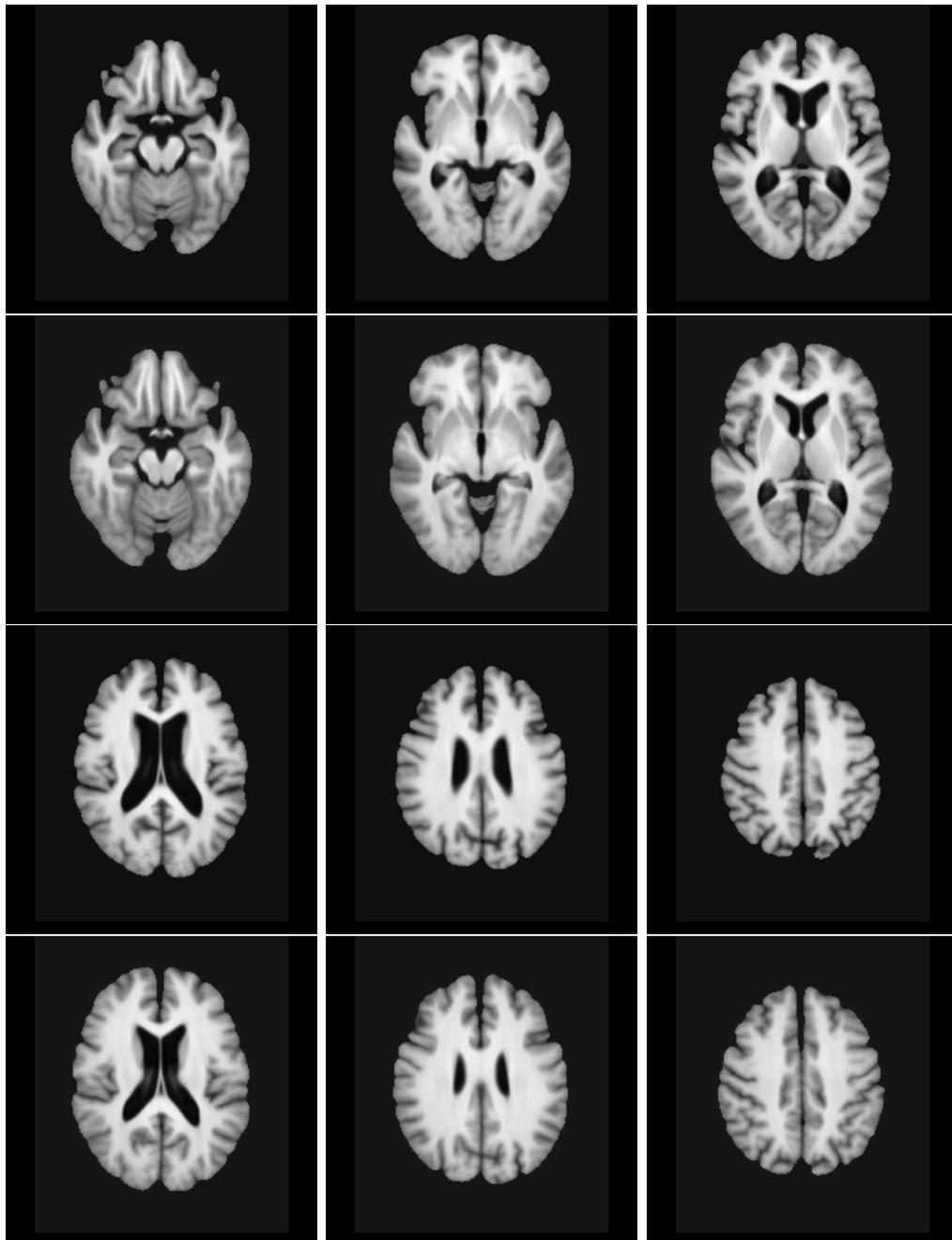


Figure 7.9: Slices taken from average anatomical space atlases for a set of subjects with Alzheimer's disease and a set of matched controls. Rows one and three show the slices taken from the atlas of AD subjects, rows two and four show the corresponding slices taken from the controls subjects' atlas.

the ventricles and regions of contraction within the cortical tissue are also visible. A region of expansion is also visible at the boundary of the ventricles in the control subjects but this is much smaller in extent than the corresponding regions in the AD volume change atlas. Also apparent in the top left of Figure 7.10 is a region of CSF expansion at the edge of hippocampus and amygdala in the AD volume change atlas.

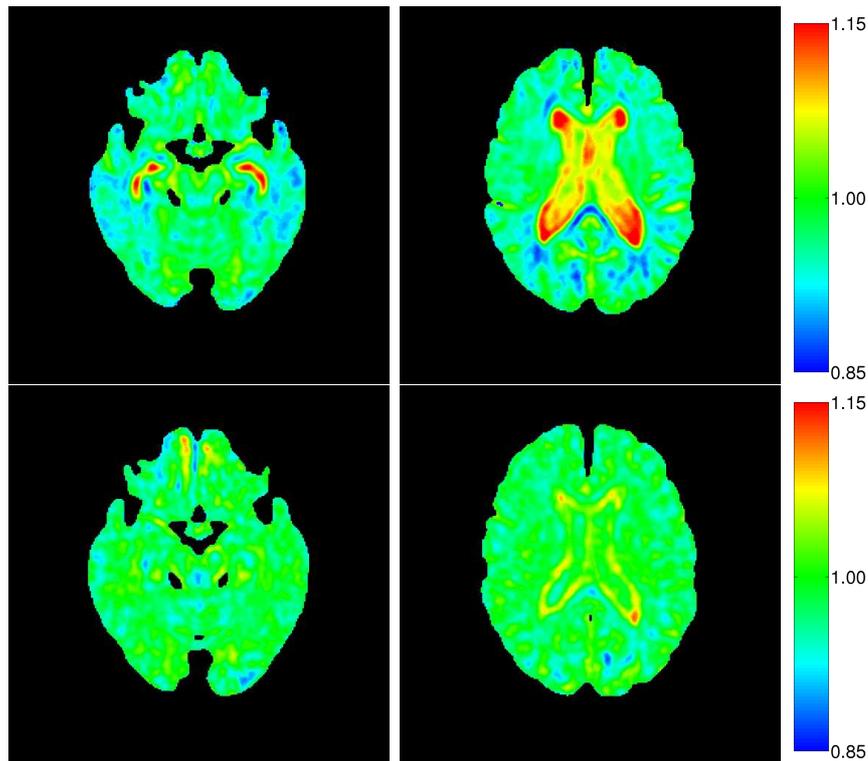


Figure 7.10: Slices taken from average space Jacobian determinant atlases for a set of subjects with Alzheimer’s disease and a set of matched controls. Row one shows slices taken from the Alzheimer’s atlas, row two shows the corresponding controls’ atlas slices. These slices correspond to those in the first column of images in Figure 7.9.

In order to obtain a clearer picture of the patterns of volume change shown in the Jacobian determinant atlases, they are also shown overlaid onto the anatomical atlases in Figure 7.11.

The longitudinal FFD displacement fields for each subject estimated by registration of the baseline and one-year follow-up images were spatially normalised [133] to the average space atlas for each group and averaged as described in sec-

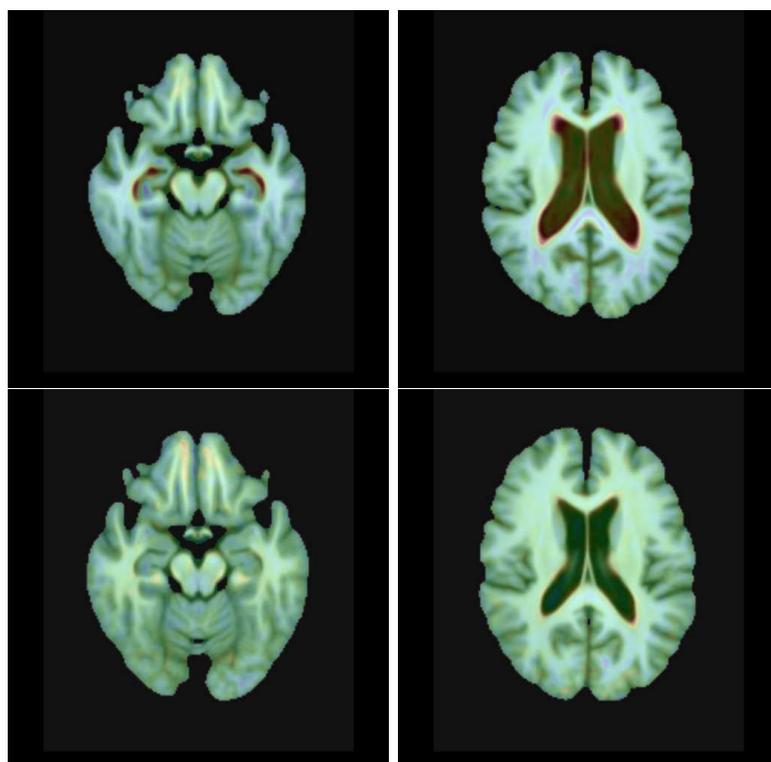


Figure 7.11: Slices taken from average space Jacobian determinant atlases for a set of subjects with Alzheimer's disease and a set of matched controls. The images are overlaid onto the corresponding anatomies and the slices correspond to those shown in Figures 7.10 and 7.9.

tion 3.4.1. The average deformations within each of the Alzheimer’s and control groups are shown separately in figure 7.12. These deformations have been scaled by a factor of 5 for easier visualisation.

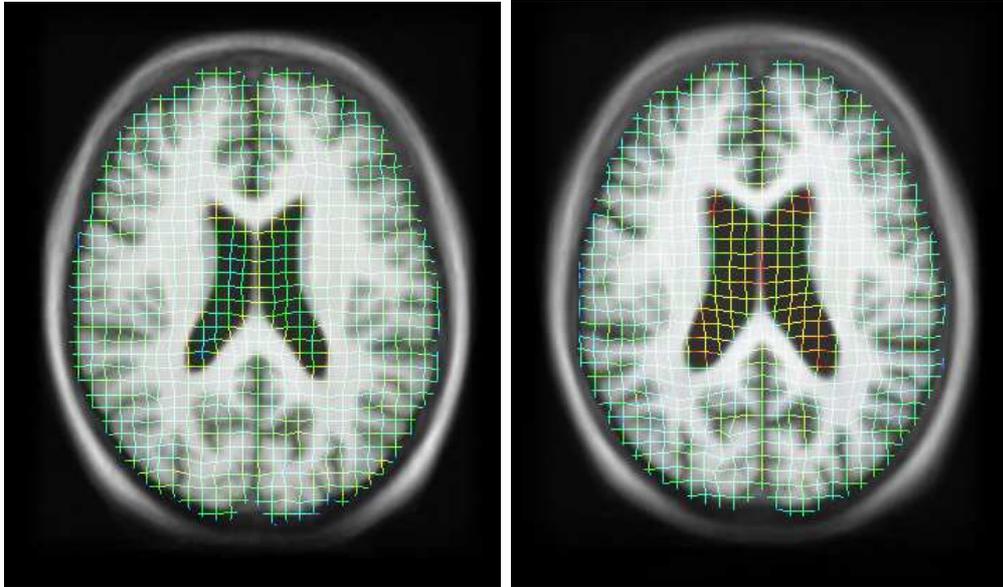


Figure 7.12: An illustration of the average longitudinal transformations for the MIRIAD data. Left the average longitudinal transformation for the control subjects overlaid onto their average space anatomical atlas. Right the corresponding image for the AD patients. The individual longitudinal transformations for each subject were mapped to the space of the corresponding baseline atlas for the group and linearly averaged. The colours indicate regions of expansion or contraction in the same way as shown in Figure 7.10.

7.7 Group separation and classification

As discussed in Section 7.4, it is possible to separate the groups based on the degree of atrophy measured by FFD volume changes (see Figure 7.7). There are other possible features or measurements on which to attempt the separation of groups of Alzheimer’s patients and age-matched controls.

For example, previous work has shown that the hippocampus is one of the first structures affected by Alzheimer’s disease with respect to volume loss [77, 94]. In order to investigate this within the MIRIAD data, estimates for the hippocampal volumes of each subject were made using the structural segmentations of the

images generated by classifier fusion.

Structural segmentations were available for all four images acquired for each subject. Normalisation for global scales and shears was carried out by affine alignment of the segmentations to the MNI single subject template. This may mean that the measurement of the volumes of structures is no longer accurate but the group separation exercise can still be carried out using the size of each structure normalised by global size. In order to mitigate for possible errors in the separate estimates for each subject, the four sets of hippocampal volumes for each subject were averaged to produce a final estimate for the combined left right volume.

The resulting hippocampus volumes for all subjects are summarised as box plots in Figure 7.13. Applying a t-test to these volume estimates gives a t-statistic of 4.75 ($p = 1.46 \times 10^{-5}$, 56 degrees of freedom, two tailed). This shows that the normalised hippocampal volumes give slightly better group separation than the estimates of atrophy derived from B-spline interpolated FFD registrations ($t = 3.78$, $p = 3.8 \times 10^{-4}$).

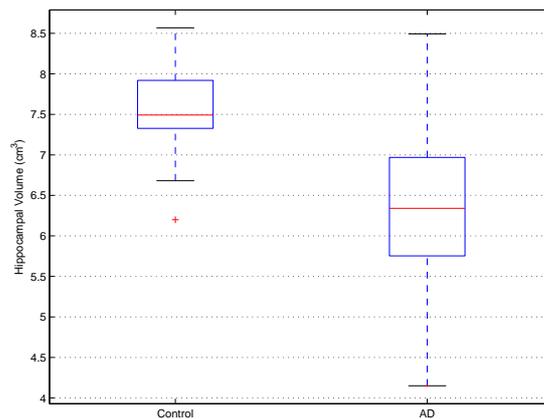


Figure 7.13: Box-plots to show the distributions of hippocampal volumes for the control subjects ($N = 19$) and the AD subjects ($N = 39$). The figures represent combined left and right hippocampal volumes and were estimated by averaging the values obtained from segmentations of the four scans for each subject. These figures give better separation of the group compared with that given by FFD based atrophy estimates (see Figure 7.7).

The hippocampal volume estimates shown in Figure 7.13 can be used to train

a supervised classifier on a ‘leave-one-out’ basis, i.e. the hippocampal data and the clinical groups (i.e. AD or control) from all but one of the subjects are used to train a classifier which is then used to predict the clinical group of the left out subject. If a Fisher linear discriminant [54] is used as the classifier and is trained with the hippocampal volume data, the resulting performance in the leave-one-out experiments is shown in the confusion matrix in Table 7.2.

		Clinical group		
		Control	AD	
Prediction	Control	16	9	Sensitivity : 76.9%
	AD	3	30	Specificity : 84.2%
				Overall : 79.3%

Table 7.2: The performance of a Fisher linear discriminant in classifying the clinical groups of the MIRIAD subjects on a leave-one-out basis. The classifier was trained using the hippocampal volumes shown in Figure 7.13

Using the figures in Table 7.2, the performance of a linear classifier based on hippocampal volume is represented by a sensitivity of 76.9%, a specificity of 84.2% and an overall correct classification rate of 79.3%.

The process of training a classifier was repeated using different input data. These data were combinations that were selected from the following:

- Hippocampal volumes
- Atrophy estimates over the year
- Grey and white matter proportions

The hippocampal volumes and atrophy estimates were derived as described above. The grey and white matter ratios were calculated using the EM [177, 109] tissue segmentations, where the volume of each tissue was expressed as a proportion of the total volume of grey matter, white matter and CSF for each subject.

Five sets of experiments were carried out, varying the data with which the linear classifier was trained. The experiments are coded with the letters A–E and the combinations of data used for each experiment are listed in Table 7.3, which

shows, for example, that experiment B used the hippocampal volumes and the one year atrophy figures to train the classifier.

Data	Experiment				
	A	B	C	D	E
GM/WM tissue proportion			1	1	1
Hippocampal volume	1	1		1	1
One year atrophy (GM & WM)		1			1

Table 7.3: Combinations of data used for different linear discriminant experiments. The experiments are coded A–E and the data used to train the classifier in each experiment is indicated by the 1s in each column.

The ‘leave-one-out’ classifier performance for the different combinations of training data was calculated for each of the experiments in the same way as described earlier (see Table 7.2). The resulting values for sensitivity, specificity and overall classification rate are shown in Table 7.4 and in Figure 7.14. These values show that the use of data from multiple sources to train the linear classifier gives improved classification performance, e.g. the results of experiment B are an improvement upon those of experiment A, showing that the use of hippocampal volumes and atrophy estimates is better than hippocampal volumes alone. The use of hippocampal volumes (A) gave identical performance to the use of grey and white tissue proportions (C), but their combination (D) resulted in significantly improved classification rates. The use of all the data during training led to the best performing classifier (experiment E).

Experiment	Sensitivity	Specificity	Overall
A	0.769	0.842	0.793
B	0.821	0.842	0.828
C	0.769	0.842	0.793
D	0.872	0.947	0.897
E	0.897	0.947	0.914

Table 7.4: The performance of classifiers applied on a ‘leave-one-out’ basis to the MIRIAD data. The classifiers used in experiments A–E were trained with different combinations of data as described in Table 7.3. The data shown here are also shown in Figure 7.14.

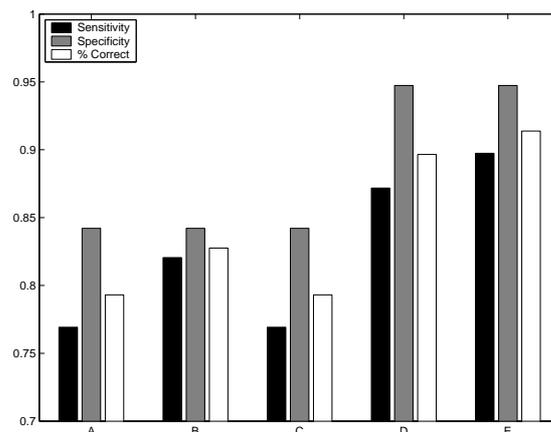


Figure 7.14: The data from Table 7.4 shown as a bar chart. (Codes: A = Hippocampus volume, B = Hippocampus volume + Atrophy, C = Tissue proportions, D = Hippocampus volume + Tissue proportions, E = Hippocampus volume + Tissue proportions + Atrophy)

7.8 Discussion

This work presented above has investigated the application of various techniques to characterise patterns of change and explore group differences in the MIRIAD data set – a group of Alzheimer’s patients and a group of age-matched controls. The methods applied have been the registration based methods for estimating growth described in Chapter 3 along with tissue segmentation via expectation maximisation and structural segmentation using classifier selection and fusion.

Average space atlases of the patient and control groups have been generated using the same methods applied to the images acquired from young children described in Chapter 6. The contrasting patterns of atrophy in the AD and control subjects have been represented via the generation of average space atlases of Jacobian determinant maps derived from longitudinal transformations of individual subjects.

As was the case for the data acquired from subjects at one and two years, the longitudinal registrations of the subjects and the resulting transformations should provide an interesting area of future work where further features can be extracted directly from the transformations and analysed.

The data acquired via the application of different automated processes to the

images in the MIRIAD cohort – atrophy estimates from registration, structural estimates from fusion and tissue estimates via EM – have been shown to be very powerful in the separation of the patient and control groups and in providing good classification performance during leave-one-out experiments. In this context, the use of data acquired from multiple processing steps to train classifiers has resulted in more powerful classifier discrimination.

Registration based methods clearly have a role in identifying patterns of change in longitudinal data acquired from elderly subjects and patients with neuro-degenerative diseases. The successful incorporation of data derived from structural segmentations achieved via classifier selection and fusion shows the potential of such structural segmentation methods in helping to characterise and measure the changes associated with atrophy.

7.9 Conclusion

This chapter has presented data on the patterns of atrophy due to AD related dementia and healthy age-matched controls based on serially acquired images. The techniques applied were the same as those used to identify patterns of growth in young children in Chapter 6. These techniques allow the regional and global patterns of atrophy to be identified. This is of direct relevance to the study of neurological atrophy in general and the progression of Alzheimer’s disease in particular.

Chapter 8

Conclusion

This thesis has presented work on measuring and characterising patterns of longitudinal changes in the brain. Registration- and segmentation-based approaches have been used in investigating such patterns of longitudinal change. As well as focusing on patterns of volume change over time for individuals and for groups of subjects, segmentations have been used to explore longitudinal change on a regional basis.

8.1 Contribution

This section discusses the contributions presented in this work and divides them into methodological and application-related categories:

- Methodology:
 - A framework for characterising longitudinal change based on serial data has been presented. Within this framework, estimates of longitudinal change for individuals are identified using intra-subject registrations. Additionally, cross-sectional inter-subject registrations are used to create transformations that map locations from the space of individual subjects' images to a postulated 'average space'. These

transformations can be used to generate average space atlases of longitudinal change based on the individual longitudinal registrations.

The transformation framework linking images with the average space can also be used to generate atlases representing other features of a cohort; anatomical atlases, for example, or atlases of the density for a particular tissue type.

- Segmentations of structures have been carried out using a classifier fusion approach and methods for the selection of classifiers from a large repository have been developed and assessed. Such segmentations allow regional estimates of growth to be derived from the per-voxel growth maps generated by longitudinal registration.

The classifier selection approaches are based on image similarity metrics or on meta-data relating to the subjects. The segmentation estimates were assessed for accuracy using the manual labels for query subjects as gold-standard validation estimates. Experiments have demonstrated the accuracy of segmentations resulting from the fusion of limited numbers of selected classifiers; this accuracy significantly exceeds that achieved by fusing random sets of classifiers. Additionally the levels of accuracy obtained are comparable with those of previously published automated approaches [88] and with manual segmentation methods [57, 157].

- Applications

- The methods for measuring longitudinal change, atlas generation and segmentation have been applied effectively to serially acquired brain image data from a group of children. The growth data for this group have been shown to be reasonably consistent with growth estimates based on tissue segmentation volumes, especially for the estimates of grey matter growth. Gold-standard estimates of growth, that might,

for example, be estimated from manual segmentations of successive serial images, are unavailable. However, the consistency of estimates derived by different methods (segmentation- and registration-based) provides some support for these automatically obtained growth figures. The use of structural segmentation allowed the growth of different regions based on the childrens' image data to be separately estimated. An initial investigation of regionally specific longitudinal change has indicated the possibility of an association between the growth in different regions of the brain and clinical data, for example developmental quotients or gestational age.

- The methods for measuring longitudinal change, atlas generation and segmentation have also been applied to serially acquired brain image data from a group consisting of Alzheimer's patients and age-matched controls. The data obtained from an analysis of these images show clear differences in the patterns of atrophy between Alzheimer's patients and control subjects. By using information based on static estimates of structure or brain volume, combined with longitudinal volume change data, excellent separation of the patient and control groups was achieved and classifiers trained on these features performed very well.

8.2 Limitations and future work

The longitudinal transformations estimated in this work have mainly been used for the generation of Jacobian determinant maps. These have also been aggregated to obtain average space volume change atlases for the cohorts studied. It would be possible, however, to extract further information from these longitudinal transformations. For example, geometric information such as the curl of the displacement field or the stress tensor. Further work can also be done by carrying out a full spatial normalisation of the longitudinal displacement fields for each

subject [133]. This work has presented the average of the spatially normalised displacement fields for illustrative purposes. This can be further extended by, for example, an investigation of the statistical properties of these transformations [139]. If the patterns of variation for the transformations are estimated (for example by applying a principal components analysis) then the subjects' longitudinal transformations may be given a low dimensional parameterisation. Such a representation may be explored for possible associations with clinical data.

Further work can also be carried out to improve the consistency of white matter growth estimates in young children, as achieved by registration- and segmentation-based methods. The variation in white matter estimates of change was higher than for the grey matter estimates. This can be due to one or more of a number of reasons. For example, the comparatively smaller volume of white matter may have been more significantly affected by partial volume effects during segmentation. Alternatively, relatively fewer features in white matter (compared with, say, the folding patterns in cortical grey matter) may make it inherently more difficult to align during registration.

An improved tissue segmentation method, should assist this process. The tissue segmentations presented in this work were derived from a simple EM-based optimisation of a Gaussian mixture model of the tissue types, using propagated manual atlases as priors. Improvements to this model are possible, for example, by using a Markov random field to incorporate context information [187]. Furthermore, it is possible to carry out the registration and the segmentation of images simultaneously [36, 130]. A recently presented approach, addressing difficulties specific to white matter registration, uses a similarity metric based on diffusion tensor data as well as the intensities of structural MR images [158]. Such an approach may help generate more consistent estimates of white matter change in groups such as the children studied in this work.

There is also a choice of the underlying transformation model. In this work, registrations were carried out using a free-form deformation model blended with

B-spline basis functions. Different models may be explored, for example a fluid model that allows for large deformations [37, 100] may be appropriate for large scale changes such as growth in the early years.

On a related issue, an interesting area for further study would be an investigation of growth, based on serially acquired images, for the interval between birth and the first year. This represents a particular challenge because the changes that occur during this period are more pronounced than between one and two years of age shown in this work. An additional confound is the degree of myelination that takes place during the first year of growth [141]. This affects the response that tissues have to MR signals and makes the contrast properties of neonatal brain images very different from those acquired from older children, thus making registrations of images across this age gap more challenging.

Further work may also be carried out on the investigation of regional changes in volume in relation to clinical outcome. In this thesis, a coarse partitioning of growth data into cortical and sub-cortical regions has been presented for the one- and two-year old subjects. The availability of a finer scale structural segmentations, however, makes an investigation of growth or atrophy at a smaller scale more feasible. The use of a larger number of smaller regions, in which growth figures are obtained and compared with clinical outcomes, means that a larger number of subjects would be necessary for statistical conclusions to be drawn with confidence. This also means that there would be a proliferation of data from which meaningful associations could be derived.

If, for a group of subjects, a set of variables representing clinical outcomes are available along with volume change data from a number of regions, the amount of data available for study can increase combinatorially. With such large amounts of data, following a hypothesis-based approach becomes more difficult and techniques used in the data-mining community [87] present themselves as useful alternatives. For example, factor analysis [76] may be used to search for a small number of latent factors that may be viewed as underlying or explaining growth

or atrophy for groups of multiple structures. An exploration of associations between a limited number of such factors and clinical outcome data would then be more feasible than one based on separate volume change estimates for a large number of structures.

In relation to the fusion of different classifiers when creating segmentations for query subjects, alternative, more sophisticated methods of fusing the classifiers represent another area that could be further investigated. Approaches such as STAPLE [175] may be adapted into the classifier selection and fusion framework used in this thesis. Alternatively, the vote rule used in this work may be extended to become spatially dependent on the regional similarity between classifiers and the query subject. Under such a scheme, classifiers' votes on the label for a particular voxel could be weighted by some measure similarity each one has with the query subject in the neighbourhood of the voxel.

8.3 Summary

This thesis has presented work that aims to characterise and explore patterns of longitudinal change for populations for which serial image data are available. Combinations of methods have been applied and a registration based framework has been used to estimate volume change data for individuals and to generate atlases of volume change and other features of the population in an average space. These techniques have been used to characterise growth or atrophy patterns within serial data acquired for both young and old subjects. A framework for the selection of atlases prior to fusion has been developed and has been successfully applied to data acquired from a set of Alzheimer's patients and age matched controls. Some exploration of association between growth data in young children and clinical outcome has also been carried out as well as the use of volume change data and structural segmentations for distinguishing between clinical groups.

Publications

1. A. Rao, P. Aljabar and D. Rueckert. Hierarchical statistical shape analysis and prediction of sub-cortical brain structures. *Medical Image Analysis*, Elsevier, 12(1), pp. 55–68, February 2008.
2. P. Aljabar, K.K. Bhatia, M. Murgasova, J.V. Hajnal, J.P. Boardman, L. Srinivasan, M. Rutherford, L.E. Dyet, A.D. Edwards and D. Rueckert. Assessment of brain growth in early childhood using deformation based morphometry. *NeuroImage*, Elsevier, 39(1), pp. 348–358, January 2008.
3. P. Aljabar, R. Heckemann, A. Hammers, J.V. Hajnal and D. Rueckert. Classifier selection strategies for label fusion using large atlas databases. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2007*, Brisbane, Lecture Notes in Computer Science, 4791, Springer, pp. 523–531, 2007.
4. K.K. Bhatia, P. Aljabar, J.P. Boardman, L. Srinivasan, M. Murgasova, M.A. Rutherford, J.V. Hajnal, A.D. Edwards and D. Rueckert. Groupwise combined segmentation and registration for atlas construction. *Medical Image Computing and Computer-Assisted Intervention - MICCAI, 2007*, Brisbane, Lecture Notes in Computer Science, 4791, Springer, pp. 532–540, 2007.
5. A. Hammers, R. Heckemann, M. Koepp, J. Duncan, J.V. Hajnal, D. Rueckert and P. Aljabar. Automatic detection and quantification of hippocampal atrophy on MRI in temporal lobe epilepsy: a proof-of-principle study. *NeuroImage*, Elsevier, 36(1), pp. 38–47, May 2007.

6. 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. Quantifying brain development in early childhood using segmentation and registration. (In Proceedings) *SPIE Medical Imaging*, San Diego (CA), February 2007.
7. 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 Proceedings) *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI 2006)*, Arlington (VA.), 2006.
8. 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*, Elsevier, Volume 33(1), pp. 115–126, October 2006.
9. J.P. Boardman, S.J. Counsell, D. Rueckert, O. Kapellou, KK Bhatia, P. Aljabar, J.V. Hajnal, J.M. Allsop, M.A. Rutherford and AD Edwards. Abnormal Deep Grey Matter Development Following Preterm Birth Detected using Deformation Based Morphometry. *NeuroImage*, Elsevier, Volume 32(1), pp. 70–78, August 2006.
10. D. Rueckert, P. Aljabar, R.A. Heckemann, J.V. Hajnal and A Hammers. Diffeomorphic Registration using B-Splines. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2006*. Lecture Notes in Computer Science, 4191, Springer, pp. 702–709, 2006.
11. R.A. Heckemann, J.V. Hajnal, P. Aljabar, D. Rueckert and A Hammers. Multiclassifier fusion in human brain MR segmentation: modelling convergence. *Medical Image Computing and Computer-Assisted Intervention*

- *MICCAI 2006*. Lecture Notes in Computer Science, 4191, Springer, pp. 815–822, 2006.
12. R.G. Boyes, D. Rueckert, P. Aljabar, J.L. Whitwell, J.M. Schott, D.L.G. Hill and N.C. Fox. Cerebral Atrophy Measurements using Jacobian Integration: Comparison with the Boundary Shift Integral. *NeuroImage*, Elsevier, Volume 32(1), pp. 159–169, August 2006.
 13. P. Aljabar, J.V. Hajnal, R.G. Boyes and D. Rueckert. Interpolation Artefacts in Non-rigid Registration. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2005*. Lecture Notes in Computer Science, 3750, Springer, pp. 247–254 2005.
 14. P. Aljabar, D. Rueckert and J.V. Hajnal. Interpolation Artefacts in Non-Rigid Image Registration. (In Proceedings) *International Society for Magnetic Resonance in Medicine (ISMRM'04)*, Kyoto, Japan, 2004.
 15. A. Rao, R. Chandrashekhara, G.I. Sanchez-Ortiz, R. Mohiaddin, P Aljabar, J.V. Hajnal, B.K. Puri and D. Rueckert. Spatial Transformation of Motion and Deformation Fields using Non-Rigid Registration. *IEEE Transactions in Medical Imaging*, Volume 23(9), pp. 1065–1076, 2004.

This thesis is available in electronic format at

<http://www.doc.ic.ac.uk/~pa100/thesis>

Bibliography

- [1] M. Alexa. Linear combination of transformations. *ACM Trans. Graph.*, 21(3):380–387, 2002.
- [2] P. Aljabar, J.V. Hajnal, R.G. Boyes, and D. Rueckert. Interpolation artefacts in non-rigid registration. In *Eighth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '05)*, volume 3750 of *Lecture Notes in Computer Science*, pages 247–254, 2005.
- [3] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Fast and simple calculus on tensors in the log-Euclidean framework. In *Eighth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '05)*, volume 3749 of *Lecture Notes in Computer Science*, pages 115–122, 2005.
- [4] J. Ashburner. *Computational Neuroanatomy*. PhD thesis, University College London, 2000.
- [5] J. Ashburner, J.G. Csernansky, C. Davatzikos, N.C. Fox, G.B. Frisoni, and P.M. Thompson. Computer-assisted imaging to assess brain structure in healthy and diseased brains. *Lancet Neurology*, 2:79–88, 2003.
- [6] J. Ashburner and K.J. Friston. Voxel-based morphometry – the methods. *NeuroImage*, 11(6):805–821, 2000.
- [7] T. Austin, A.P. Gibson, G. Branco, R. Yusof, S.R. Arridge, J.H. Meek, J. S. Wyatt, D.T. Delpy, and J.C. Hebden. Three dimensional optical imag-

- ing of blood volume and oxygenation in the neonatal brain. *NeuroImage*, 31(4):1426–1433, 2006.
- [8] B.B. Avants and J.C. Gee. Shape averaging with diffeomorphic flows for atlas creation. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI), Arlington, VA, April 2004*, pages 595–598, 2004.
- [9] E.H. Aylward, Q. Li, N.A. Honeycutt, A.C. Warren, M.B. Pulsifer, P.E. Barta, M.D. Chan, P.D. Smith, M. Jerram, and G.D. Pearlson. MRI volumes of the hippocampus and amygdala in adults with Down’s syndrome with and without dementia. *Am J Psychiatry*, 156(4):564–568, 1999.
- [10] R. Bajcsy and S. Kovacevic. Multiresolution elastic matching. *Computer Vision, Graphics, and Image Processing*, 46(1):1–21, 1989.
- [11] P.J. Basser and C. Pierpaoli. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *Journal of Magnetic Resonance*, 111(3):209–219, 1996.
- [12] M.F. Beg and A. Khan. Computing an average anatomical atlas using lddmm and geodesic shooting. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI), Arlington, VA, April 2006*, pages 1116–1119, 2006.
- [13] M.F. Beg, M.I. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*, 61(2):139–157, 2005.
- [14] K.K. Bhatia, J.V. Hajnal, A.D. Edwards, M.A. Rutherford, and D. Rueckert. Groupwise atlas construction for the identification of neuroanatomical changes associated with preterm birth. In *Proceedings International Society for Magnetic Resonance in Medicine (ISMRM), Miami, USA, 2005*, page 643, 2005.

- [15] K.K. Bhatia, J.V. Hajnal, B.K. Puri, A.D. Edwards, and D. Rueckert. Consistent groupwise non-rigid registration for atlas construction. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI), Arlington, VA, April 2004*, pages 908–911, 2004.
- [16] D. Birchall and N. Botan. Volume rendered MR angiography for the detection of small intracranial aneurysms. *British Journal of Neurosurgery*, 17(4):366–367, 2003.
- [17] D.J. Blezek and J.V. Miller. Atlas stratification. In *Ninth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '06)*, Lecture Notes in Computer Science, pages 712–719, 2006.
- [18] J.P. Boardman, S.J. Counsell, O. Kapellou, D. Rueckert, J.V. Hajnal, K.K. Bhatia, P. Aljabar, M.A. Rutherford, J.M. Allsop, and A.D. Edwards. White matter abnormality is associated with volume reduction in deep grey nuclei following preterm birth. *Pediatric Research*, 58:362–362, 2005.
- [19] J.P. Boardman, S.J. Counsell, O. Kapellou, D. Rueckert, M.A. Rutherford, J. Stark, K.K. Bhatia, P. Aljabar, J.M. Allsop, J.V. Hajnal, and A.D. Edwards. Neural systems affected by preterm birth. In *Proceedings International Society for Magnetic Resonance in Medicine (ISMRM), Miami, USA, 2005*, page 299, 2005.
- [20] J.P. Boardman, S.J. Counsell, Daniel Rueckert, O. Capellou, Kanwal Bhatia, P. Aljabar, J.V. Hajnal, J. 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.
- [21] A.L. Bokde, S.J. Teipel, R. Schwarz, G. Leinsinger, K. Buerger, T. Moeller, H.J. Möller, and H. Hampel. Reliable manual segmentation of the frontal, parietal, temporal, and occipital lobes on magnetic resonance images of

- healthy subjects. *Brain research, Brain research protocols*, 14(3):135–45, 2005.
- [22] 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.
- [23] C. De Boor, editor. *A Practical Guide to Splines*. Springer, 1978.
- [24] M.N. Bossa and S. Olmos. Statistical model of similarity transformations: Building a multi-object pose. In *Computer Vision and Pattern Recognition Workshop*, page 59, 2006.
- [25] N.F. Botting. *Psychological and educational outcome of very low birthweight children at 12 years*. PhD thesis, University of Liverpool, 1997.
- [26] R.G. Boyes, D. Rueckert, P. Aljabar, J. Whitwell, J.M. Schott, D.L.G. Hill, and N.C. Fox. Cerebral atrophy measurements using jacobian integration: Comparison with the boundary shift integral. *NeuroImage*, 32(1):159–169, 2006.
- [27] L.M. Brubaker, E. Bullitt, C. Yin, T. Van Dyke, and W. Lin. Magnetic Resonance Angiography Visualization of Abnormal Tumor Vasculature in Genetically Engineered Mice. *Cancer Research*, 65(18):8218–8223, 2005.
- [28] V.A. Cardenas, C. Studholme, S. Gazdzinski, T.C. Durazzo, and D.J. Meyerhoff. Deformation-based morphometry of brain changes in alcohol dependence and abstinence. *NeuroImage*, 34(3):879–887, 2006.
- [29] M. Do Carmo, editor. *Differential Geometry of Curves and Surfaces*. Prentice Hall, 1976.
- [30] R.E. Carson, M.E. Daube-Witherspoon, and P. Herscovitch. *Quantitative functional brain imaging with positron emission tomography*. Academic Press, San Diego, 1998.

- [31] Center for Morphometric Analysis (CMA). Internet Brain Segmentation Repository (IBSR). <http://www.cma.mgh.harvard.edu/ibsr/>.
- [32] 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.
- [33] R. Chandrashekara, R. Mohiaddin, R. Razavi, and D. Rueckert. Non-rigid image registration with subdivision lattices: Application to cardiac MR image analysis. In *Tenth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '07)*, volume 4791, pages 335–342, 2007.
- [34] R. Chandrashekara, R. H. Mohiaddin, and D. Rueckert. Analysis of myocardial motion and strain patterns using a cylindrical B-spline transformation model. In *Surgery Simulation and Soft Tissue Modelling (IS4TM '03)*, Lecture Notes in Computer Science, pages 88–99, 2003.
- [35] B. Chen and A.E. Kaufman. 3D volume rotation using shear transformations. *Graphical Models*, 62(4):308–322, 2000.
- [36] X. Chen, M. Brady, and D. Rueckert. Simultaneous segmentation and registration for medical image. In *Seventh Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '04)*, volume 3216 of *Lecture Notes in Computer Science*, pages 663–670, 2004.
- [37] 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.
- [38] M.K. Chung, S. Robbins, K.M. Dalton, S. Wang, A.C. Evans, and R.J. Davidson. Tensor-based cortical morphometry via weighted spherical har-

- monic representation. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pages –, 2006.
- [39] M.K. Chung, K.J. Worsley, T. Paus, D.L. Collins C. Cherif, J.N. Giedd, J.L. Rapoport, and A.C. Evans. A unified statistical approach to deformation-based morphometry. *NeuroImage*, 14(3):595–606, 2001.
- [40] 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.
- [41] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.
- [42] O. Commowick and Grégoire Malandain. Efficient selection of the most similar image in a database for critical structures segmentation. In *Tenth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '07)*, pages 203–210, 2007.
- [43] R.J. Cook. *Kappa*, pages 2160–2166. Wiley, 1998.
- [44] T.F. Cootes, G.J. Edwards, and C.J.Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.
- [45] W.R. Crum, O. Camara, and D.L.G. Hill. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Transactions on Medical Imaging*, 25(11):1451–1461, 2006.

- [46] W.R. Crum, R.I. Scahill, and N.C. Fox. Automated hippocampal segmentation by regional fluid registration of serial MRI: Validation and application in alzheimer's disease. *NeuroImage*, 13(5):847–855, 2001.
- [47] E. D'Agostino, F. Maes, D. Vandermeulen, and P. Suetens. A viscous fluid model for multimodal non-rigid image registration using mutual information. In *Fifth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '02)*, volume 2489 of *Lecture Notes in Computer Science*, pages 541–548, 2002.
- [48] 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*, 20:88–97, 1996.
- [49] P.F. D'Haese, V. Duay, T.E. Merchant, B.M. 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)*, pages 627–634, 2003.
- [50] L.R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26:297, 1945.
- [51] B.C. Dickerson and R.A. Sperling. Neuroimaging biomarkers for clinical trials of disease-modifying therapies in Alzheimer's disease. *NeuroRx*, 2(2):348–360, 2005.
- [52] J.P. Didon and F. Langevin. Fast 3D registration of MR brain images using the projection correlation registration algorithm. *Medical and Biological Engineering and Computing*, 36(1):107–111, 1998.
- [53] D. Dubois and H. Prade, editors. *Fundamentals of Fuzzy Sets*. New York: Kluwer, 2000.

- [54] R.O. Duda, P.E. Hart, and D.G. Stork, editors. *Pattern Classification*. Wiley, 2001.
- [55] S. Durston, H.E. Hulshoff Pol, B.J. Casey, J.N. Giedd, J.K. Buitelaar, and H. van Engeland. Anatomical MRI of the developing human brain: What have we learned? *Journal of the American Academy of Child and Adolescent Psychiatry*, 40(9):1012–1020, 2001.
- [56] C.P. Ferri, M. Prince, C. Brayne, H. Brodaty, L. Fratiglioni, M. Ganguli, K. Hall, K. Hasegawa, H. Hendrie, Y. Huang, A. Jorm, C. Mathers, P.R. Menezes, E. Rimmer, and M. Scazufca. Global prevalence of dementia: a Delphi consensus study. *The Lancet*, 366(9503):2112–2117, 2006.
- [57] B. Fischl, D.H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A.M. Dale. Whole brain segmentation: Automated labeling of neuroanatomical structure in the human brain. *Neuron*, 33(3):341–355, 2002.
- [58] R. Fletcher, editor. *Practical methods of optimization*. Wiley, 1980.
- [59] M.F. Folstein, S.E. Folstein, and P.R. McHugh. "Mini-mental state". a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3):189–98, 1975.
- [60] D.R. Forsey and R.H. Bartels. Hierarchical B-spline refinement. In *Proceedings of the 15th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1988*, pages 205–212, New York, NY, USA, 1988. ACM.
- [61] N.C. Fox, R.S. Black, S. Gilman, M.N. Rossor, S.G. Griffith, L. Jenkins, M. Koller, and for the AN1792(QS-21)-201 Study Team. Effects of Abeta

- immunization (AN1792) on MRI measures of cerebral volume in Alzheimer disease. *Neurology*, 64(9):1563–1572, 2005.
- [62] N.C. Fox, R. Jenkins, S.M. Leary, V.L. Stevenson, N.A. Losseff, W.R. Crum, R.J. Harvey, M.N. Rossor., D.H. Miller, and A.J. Thompson. Progressive cerebral atrophy in MS: A serial study using registered, volumetric MRI. *Neurology*, 54(4):807–812, 2000.
- [63] A.F. Frangi, D. Rueckert, J.A. Schnabel, and W.J. Niessen. Automatic 3D ASM construction via atlas-based landmarking and volumetric elastic registration. In *Information Processing in Medical Imaging: Proc. 17th International Conference (IPMI'01)*, Lecture Notes in Computer Science, pages 78–91, Davis, CA, July 2001.
- [64] M. Frechét. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'institut Henri Poincaré*, 10(4):215–310, 1948.
- [65] P.A. Freeborough and N.C. Fox. The boundary shift integral: An accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE Transactions on Medical Imaging*, 16(5):623–629, 1997.
- [66] P.A. Freeborough and N.C. Fox. Modeling brain deformations in Alzheimer disease by fluid registration of serial 3D MR images. *Journal of Computer Assisted Tomography*, 22(5):838–843, 1998.
- [67] K.J. Friston, J. Ashburner, J.B. Poline, C. Frith, J.D. Heather, and R.S.J. Frackowiak. Spatial registration and normalization of images. *Human Brain Mapping*, 3(3):165–189, 1995.
- [68] K.J. Friston, P. Jezzard, and R. Turner. Analysis of functional MRI time-series. *Human Brain Mapping*, 1(2):153–171, 1994.
- [69] C. Gaser, I. Nenadoc, B.R. Buchsbaum, E.A. Hazlett, and M.S. Buchsbaum. Deformation-based morphometry and its relation to conventional

- volumetry of brain lateral ventricles in MRI. *NeuroImage*, 13:1140–1145, 2001.
- [70] J. Gee and D. Haynor. Numerical methods for high dimensional warps. In *Brain Warping*. Academic Press, 1998.
- [71] J. Gee, M. Reivich, and R. Bajcsy. Elastically deforming 3D atlas to match anatomical brain images. *Journal of Computer Assisted Tomography*, 17(2):225–236, 1993.
- [72] G. Gerig, B. Davis, P. Lorenzen, Shun Xu, M. Jomier, J. Piven, and S.C. Joshi. Computational anatomy to assess longitudinal trajectory of brain growth. In *Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT)*, pages 1041–1047, June 2006.
- [73] J.N. Giedd, N.O. Jeffries, F.X. Castellanos, H. Liu, A. Zijdenbos, T. Paus, A.C. Evans, and J.L. Rapoport. Brain development during childhood and adolescence: a longitudinal MRI study. *Nat Neurosci*, 2(10):861–863, Oct 1999.
- [74] J.N. Giedd, J.W. Snell, N.Lange, J.C. Rajapakse, B.J. Casey, A.C.Vaituzis P.L. Kozuch, Y.C. Vauss, S.D. Hamburger, and D.K.and J.L. Rapoport. Quantitative magnetic resonance imaging of human brain development ages 4-18. *Cerebral Cortex*, 6:551–560, 1996.
- [75] C. Good, I. Johnsrude, J. Ashburner, R. Henson, K. Friston, and R. Frackowiak. A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage*, 14(1):21–36, 2001.
- [76] R.L. Gorsuch, editor. *Factor Analysis (2nd ed.)*. Erlbaum, Hillsdale, NJ, 1983.

- [77] K.M. Gosche, J.A. Mortimer, C.D. Smith, W.R. Markesbery, and D.A. Snowdon. Hippocampal volume as an index of Alzheimer neuropathology: Findings from the Nun study. *Neurology*, 58(10):1476–1482, 2002.
- [78] U. Grenander and M.I. Miller. Computational anatomy: An emerging discipline. *Quarterly Of Applied Mathematics*, 56:617–694, 1998.
- [79] R. Griffiths and M. Huntley, editors. *The Griffiths Mental Development Scales: from birth to 2 years. Manual: the 1996 revision*. Association for Research in Infant and Child Development, 1996.
- [80] A. Guimond, J. Meunier, and J.P. Thirion. Average brain models: a convergence study. *Computer vision and image understanding*, 77:192–210, 2000.
- [81] J.V. Hajnal, D.J. Hawkes, and D.L.G. Hill, editors. *Medical Image Registration*. CRC Press, 2001.
- [82] J.V. Hajnal, N. Saeed, E.J. Soar, A. Oatridge, I.R. Young, and G.M. Bydder. A registration and interpolation procedure for subvoxel matching of serially acquired MR-images. *Journal of Computer Assisted Tomography*, 19(2):289–296, 1995.
- [83] A. Hammers, R. Allom, M.J. Koeppe, 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.
- [84] A. Hammers, Chen C-H, L. Lemieux, R. Allom, S. Vossos, S.L. Free, R. Myers, D.J. Brooks, J.S. Duncan, and M.J. Koeppe. Statistical neuroanatomy of the human inferior frontal gyrus and probabilistic atlas in a standard stereotaxic space. *Human Brain Mapping*, 28:34–48, 2007.

- [85] A. Hammers, M.J. Koepp, S.L. Free, M. Brett, M.P. Richardson, C. Labbe, V.J. Cunningham, D.J. Brooks, and J. Duncan. Implementation and application of a brain template for multiple volumes of interest. *Human Brain Mapping*, 15(3):165–174, Mar 2002.
- [86] X. Han, C. Xu, D. Tosun, and J.L. Prince. Cortical surface reconstruction using a topology preserving geometric deformable model. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pages 213–220, 2001.
- [87] D.J. Hand, P. Smyth, and H. Mannila. *Principles of data mining*. MIT Press, Cambridge, MA, 2001.
- [88] 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(1):115–126, Oct 2006.
- [89] R.A. Heckemann, J.V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers. Multiclassifier fusion in human brain MR segmentation: Modelling convergence. In *Ninth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '06)*, volume 4191, pages 815–822, 2006.
- [90] G. Hermosillo, C. Chef d’Hotel, and O.D. Faugeras. Variational methods for multimodal image matching. *International Journal of Computer Vision*, 50(3):329–343, 2002.
- [91] C.J. Holmes, R. Hoge, L. Collins, R. Woods, A.W. Toga, and A.C. Evans. Enhancement of mr images using registration for signal averaging. *Journal of Computer Assisted Tomography*, 22(2):324–333, 1998.
- [92] P.S. Huppi, S. Warfield, R. Kikinis, P.D. Barnes, G.P. Zientara, F.A. Jolesz, M.K. Tsuji, and J.J. Volpe. Quantitative magnetic resonance imaging of

- brain development in premature and mature newborns. *Annals of Neurology*, 43(2):224–235, 1998.
- [93] D.V. Iosifescu, M.E. Shenton, S.K. Warfield, R. Kikinis, J. Dengler, F.A. Jolesz, and R.W. McCarley. An automated registration algorithm for measuring MRI subcortical brain structures. *NeuroImage*, 6(1):13–25, 1997.
- [94] C.R. Jack Jr, D.W. Dickson, J.E. Parisi, Y.C. Xu, R.H. Cha, P.C. O’Brien, S.D. Edland, G.E. Smith, B.F. Boeve, E.G. Tangalos, E. Kokmen, and R.C. Petersen. Antemortem MRI findings correlate with hippocampal neuropathology in typical aging and dementia. *Neurology*, 58(5):750–757, 2002.
- [95] C.R. Jack Jr, R.C. Petersen, Y.C. Xu, S.C. Waring, P.C. O’Brien, E.G. Tangalos, G.E. Smith, R.J. Ivnik, and E. Kokmen. Medial temporal atrophy on MRI in normal aging and very mild Alzheimer’s disease. *Neurology*, 49:786–794, 1997.
- [96] C.R. Jack Jr., M.M. Shiung, J.L. Gunter, P.C. O’Brien, S.D. Weigand, D.S. Knopman, B.F. Boeve, R.J. Ivnik, G.E. Smith, R.H. Cha, E.G. Tangalos, and R.C. Petersen. Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. *Neurology*, 62(4):591–600, 2004.
- [97] P. Jezzard, P.M. Matthews, and S.M. Smith, editors. *Functional MRI : An Introduction to Methods*. OUP, 2001.
- [98] J.P. John, L. Wang, A.J. Moffitt, H.K. Singh, M.H. Gado, and J.G. Csernansky. Inter-rater reliability of manual segmentation of the superior, inferior and middle frontal gyri. *Psychiatry Research: Neuroimaging*, 148(2-3):151–163, 2006.
- [99] S. Joshi, U. Grenander, and M. Miller. On the geometry and shape of brain sub-manifolds. *International Journal of Pattern Recognition and Artifi-*

- cial Intelligence: Special Issue on Processing of MR Images of the Human*, 11(8):1317–1343, 1997.
- [100] S.C. Joshi. *Large Deformation Diffeomorphisms and Gaussian Random Fields for Statistical Characterization of Brain SubManifolds*. PhD thesis, Dept. of Electrical Engineering, Sever Institute of Technology, Washington University, 1997.
- [101] S.C. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23(S1):151–160, 2004.
- [102] O. Kapellou, S.J. Counsell, N. Kennea, L.E. Dyet, N. Saeed, J. Stark, E. Maalouf, P. Duggan, M. Ajayi-Obe, J.V. Hajnal, J.M. Allsop, J.P. Boardman, M.A. Rutherford, F. Cowan, and A.D. Edwards. Abnormal cortical development after premature birth shown by altered allometric scaling of brain growth. *PLoS Medicine*, 3(8):e265, 2006.
- [103] B. Karaçali and C. Davatzikos. Estimating topology preserving and smooth displacement fields. *IEEE Transactions on Medical Imaging*, 23(7):868–880, 2004.
- [104] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1987.
- [105] R.G. Keys. Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoustics, Speech, And Signal Processing*, 29(6):1153, 1981.
- [106] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998.

- [107] H.C. Kraemer, J.A. Yesavage, J.L. Taylor, and D. Kupfer. How can we learn about developmental processes from cross-sectional studies, or can we? *Am J Psychiatry*, 157(2):163–171, 2000.
- [108] S. Lee, G. Wolberg, and S. Yong Shin. Scattered data interpolation with multilevel B-splines. *IEEE Transactions on Visualization and Computer Graphics*, 3(3):228–244, 1997.
- [109] 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.
- [110] T.M. Lehmann, C. Gönner, and K. Spitzer. Survey: Interpolation methods in medical image processing. *IEEE Transactions on Medical Imaging*, 18(11):1049–1075, 1999.
- [111] A.D. Leow, A.D. Klunder, C.R. Jack Jr, A.W. Toga, A.M. Dale, M.A. Bernstein, P.J. Britson, J.L. Gunter, and C.P. Ward. Longitudinal stability of MRI for mapping brain change using tensor-based morphometry. *NeuroImage*, 31(2):627–640, Jun 2006.
- [112] T. Lindeberg. Scale-space theory: a basic tool for analyzing structures at different scales. *Journal of Applied Statistics*, 21(1):225–270, 1994.
- [113] O. Longe, B. Randall, E.A. Stamatakis, and L.K. Tyler. Grammatical categories in the brain: The role of morphological structure. *Cerebral Cortex*, 17(8):1812–1820, 2007.
- [114] M. Lorenzo-Valdes, G.I. Sanchez-Ortiz, A.G. Elkington, R.H. Mohiaddin, and D. Rueckert. Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm. *Medical Image Analysis*, 8(3):255–265, Sep 2004.

- [115] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997.
- [116] J. Maintz and M. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.
- [117] N. Marlow, D. Wolke, M.A. Bracewell, and M. Samara. Neurologic and developmental disability at six years of age after extremely preterm birth. *New England Journal of Medicine*, 352(1):9–19, 2005.
- [118] 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)*, volume 2879 of *Lecture Notes in Computer Science*, pages 771–779. Springer-Verlag, 2003.
- [119] W.D. Marslen-Wilson and L.K. Tyler. Morphology, language and the brain: the decompositional substrate for language comprehension. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481):823–836, 2007.
- [120] J. Matsuzawa, M. Matsui, T. Konishi, K. Noguchi, R.C. Gur, W. Bilker, and T. Miyawaki. Age-related volumetric changes of brain gray and white matter in healthy infants and children. *Cerebral Cortex*, 11(4):335–342, 2001.
- [121] S. Mori and P.B. Barker. Diffusion magnetic resonance imaging: its principle and applications. *The Anatomical record*, 257(3):102–109, 1999.
- [122] S.G. Mueller, M.W. Weiner, L.J. Thal, R.C. Petersen, C.R. Jack, W. Jagust, J.Q. Trojanowski, A.W. Toga, and L Beckett. The Alzheimer’s Disease

- Neuroimaging Initiative (ADNI). *Journal of the Alzheimer's Association*, 1(1):55–66, 2005.
- [123] M. Murgasova, L.E. Dyet, A.D. Edwards, M.A. Rutherford, J.V. Hajnal, and D. Rueckert. Segmentation of brain MRI in young children. In *Ninth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '06)*, volume 4190 of *Lecture Notes in Computer Science*, pages 687–694, 2006.
- [124] M. Murgasova, L.E. Dyet, A.D. Edwards, M.A. Rutherford, J.V. Hajnal, and D. Rueckert. Segmentation of brain MRI in young children (in press). *Academic Radiology*, pages –, 2007.
- [125] M. Nielsen, P. Johansen, A.D. Jackson, and B. Lautrup. Brownian warps: A least committed prior for non-rigid registration. In *Fifth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '02)*, pages 557–564, 2002.
- [126] M. Nishida, N. Makris, D.N. Kennedy, M. Vangel, B. Fischl, K.S. Krishnamoorthy, V.S. Caviness, and P.E. Grant. Detailed semiautomated MRI based morphometry of the neonatal brain: Preliminary results. *NeuroImage*, 32(3):1041–1049, Sept 2006.
- [127] E. Parzen. On estimation of a probability density function and mode. *Annals Math Statistics*, 33:1065–1076, 1962.
- [128] X. Pennec, P. Cachier, and N. Ayache. Understanding the "demon's algorithm": 3D non-rigid registration by gradient descent. In *Second Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '99)*, volume 1679 of *Lecture Notes in Computer Science*, pages 597–605, 1999.

- [129] 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.
- [130] 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(1):228–239, 2006.
- [131] G. Postelnicu, L. Zollei, R. Desikan, and B. Fischl. Geometry driven volumetric registration. In *Information Processing in Medical Imaging: Proc. 20th International Conference (IPMI'07)*, Lecture Notes in Computer Science, pages 675–686, 2007.
- [132] J.C. Pruessner, M. Li, W. Serles, and et al. Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: Minimizing the discrepancies between laboratories. *Cerebral Cortex*, 10(4):433–442, 2000.
- [133] A. Rao, R. Chandrashekara, G.I. Sanchez-Ortiz, R. Mohiaddin, P. Aljabar, J.V. Hajnal, B.K. Puri, and D. Rueckert. Spatial transformation of motion and deformation fields using nonrigid registration. *IEEE Transactions on Medical Imaging*, 23(9):1065–1076, 2004.
- [134] A. Rao, G.I. Sanchez-Ortiz, R. Chandrashekara, M. Lorenzo-Valdés, R. Mohiaddin, and D. Rueckert. Construction of a cardiac motion atlas from mr using non-rigid registration. In *FIMH*, volume 2674 of *Lecture Notes in Computer Science*, pages 141–150, 2003.
- [135] T. Rohlfing, R. Brandt, R. Menzel, and C.R. Maurer Jr. 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.

- [136] T. Rohlfing, C.R. Maurer Jr., D.A. Bluemke, and M.A. Jacobs. Volume-preserving non-rigid registration of MR breast images using free-form deformation with an incompressibility constraint. *IEEE Transactions on Medical Imaging*, 22(6), 2003.
- [137] T. Rohlfing, E.V. Sullivan, and A. Pfefferbaum. Deformation-based brain morphometry to track the course of alcoholism: Differences between intra-subject and inter-subject analysis. *Psychiatry Research: Neuroimaging*, 146(2):157–170, 2006.
- [138] D. Rueckert, P. Aljabar, R.A. Heckemann, J.V. Hajnal, and A. Hammers. Diffeomorphic registration using B-splines. In *Ninth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '06)*, volume 4191, pages 702–709, 2006.
- [139] D. Rueckert, A.F. Frangi, and J.A. Schnabel. Automatic construction of 3D statistical deformation models of the brain using nonrigid registration. *IEEE Transactions on Medical Imaging*, 22(8):1014–1025, 2003.
- [140] 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.
- [141] M.A. Rutherford, editor. *MRI of the neonatal brain*. W.B. Saunders, 2002.
- [142] R.I. Scahill, J.M. Schott, J.M. Stevens, M.N. Rossor, and N.C. Fox. Mapping the evolution of regional atrophy in Alzheimer’s disease: Unbiased analysis of fluid-registered serial MRI. *Proceedings of the National Academy of Sciences*, 99(7):4703–4707, 2002.
- [143] 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 Wiro J. Niessen and Max A. Viergever, editors, *Fourth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '01)*, volume 2208 of *Lecture Notes in Computer Science*, pages 573–581, 2001.
- [144] J.M. Schott, S.L. Price, C. Frost, J.L. Whitwell, M.N. Rossor, and N.C. Fox. Measuring atrophy in Alzheimer disease: A serial MRI study over 6 and 12 months. *Neurology*, 65(1):119–124, 2005.
- [145] T.W. Sederberg and S.R. Parry. Free-form deformation of solid geometric models. In K. Robinson, editor, *Proceedings of the 13th annual conference on Computer graphics and interactive techniques (SIGGRAPH), Dallas, USA, August, 1986*, pages 151–160, 1986.
- [146] C.E. Shannon. Communication in the presence of noise. In *Proceedings Institute of Radio Engineers*, pages 10–21, 1949.
- [147] D. Shen and C. Davatzikos. HAMMER: Hierarchical attribute matching mechanism for elastic registration. *IEEE Transactions on Medical Imaging*, 21(11):1421–1439, 2002.
- [148] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.
- [149] K. Shoemake. Animating rotation with quaternion curves. *SIGGRAPH*, 19(3):245–254, 1985.
- [150] J.G. Sled, A.P. Zijdenbos, and A.C. Evans. A non-parametric method for automatic correction of intensity non-uniformity in MRI data. *IEEE Transactions on Medical Imaging*, 17(1):87–97, 1998.

- [151] A.D. Smith. Imaging the progression of Alzheimer pathology through the brain. *Proceedings of the National Academy of Sciences*, 99(7):4135–4137, 2002.
- [152] S.M. Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143–155, Nov 2002.
- [153] S.M. Smith, A. Rao, N. De Stefano, M. Jenkinson, J.M. Schott, P.M. Matthews, and N.C. Fox. Longitudinal and cross-sectional analysis of atrophy in Alzheimer’s disease: Cross-validation of BSI, SIENA and SIENAX. *NeuroImage*, 36(4):1200–1206, 2007.
- [154] S.M. Smith, N. De Stefano, M. Jenkinson, and P.M. Matthews. Normalised accurate measurement of longitudinal brain change. *Journal of Computer Assisted Tomography*, 25(3):466–475, 2001.
- [155] E.R. Sowell, P.M. Thompson, C.M. Leonard, S.E. Welcome, E. Kan, and A.W. Toga. Longitudinal mapping of cortical thickness and brain growth in normal children. *Journal of Neuroscience*, 24(38):8223–8231, Sep 2004.
- [156] G. Soza, R. Grosso, C. Nimsy, G. Greiner, and P. Hastreite. Estimating mechanical brain tissue properties with simulation and registration. In *Seventh Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI ’04)*, volume 3217, pages 276–283, 2004.
- [157] R. Spinks, V.A. Magnotta, N.C. Andreasen, K.C. Albright, S. Ziebell, P. Nopoulos, and M. Cassell. Manual and automated measurement of the whole thalamus and mediodorsal nucleus using magnetic resonance imaging. *NeuroImage*, 17(2):631–642, 2002.
- [158] C. Studholme. Incorporating DTI data as a constraint in deformation tensor morphometry between T1 MR images. In *Information Processing in Medical*

- Imaging: Proc. 20th International Conference (IPMI'07)*, volume 4584 of *Lecture Notes in Computer Science*, pages 223–232. Springer, 2007.
- [159] 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, 1999.
- [160] 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.
- [161] N.A. Thacker, A. Jackson, D. Moriarty, and E. Vokurka. Improved quality of re-sliced MR images using re-normalized sinc interpolation. *Journal of Magnetic Resonance Imaging*, 10(4):582–588, 1999.
- [162] P. Thévenaz, T. Blu, and M. Unser. Interpolation revisited. *IEEE Transactions on Medical Imaging*, 19(7):739–758, 2000.
- [163] J.P. Thirion. Image matching as a diffusion process: An analogy with maxwell’s demons. *Medical Image Analysis*, 2(3):243–260, 1998.
- [164] 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.
- [165] O. Tuzel, F. Porikli, and P. Meer. Human detection via classification on riemannian manifolds. In *Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.
- [166] C.J. Twining and S. Marsland. Constructing diffeomorphic representations of non-rigid registrations of medical images. In *Information Processing in Medical Imaging: Proc. 18th International Conference (IPMI'03)*, Lecture Notes in Computer Science, pages 413–425, 2003.

- [167] M. Unser. Splines : A perfect fit for signal and image processing. *IEEE Signal Processing Magazine*, pages 22–38, November 1999.
- [168] P. Viola. *Alignment by Maximization of Mutual Information*. A. I. technical report no. 1548, Massachusetts Institute of Technology, 1995.
- [169] P. Viola and W.M. Wells III. Alignment by maximization of mutual information. In Grimson, editor, *Proc. 5th International Conference on Computer Vision (ICCV'95)*, pages 16–23. IEEE Computer Society Press, 1995.
- [170] G. Wahba, editor. *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, 1990.
- [171] D.M. Wang and D.M. Doddrell. MR image-based measurement of rates of change in volumes of brain structures part i: Method and validation. *Magnetic Resonance Imaging*, 20:27–40, 2002.
- [172] L. Wang, J.S. Swank, I.E. Glick, M.H. Gado, M.I. Miller, J.C. Morris, and J.G. Csernansky. Changes in hippocampal volume and shape across time distinguish dementia of the alzheimer type from healthy aging. *NeuroImage*, 20(2):667–682, 2003.
- [173] Q. Wang, D. Seghers, E. D’Agostino, F. Maes, D. Vandermeulen, P. Suetens, and A. Hammers. Construction and validation of mean shape atlas templates for atlas-based brain image segmentation. In *Information Processing in Medical Imaging: Proc. 19th International Conference (IPMI'05)*, Lecture Notes in Computer Science, pages 689–700, 2005.
- [174] S.K. Warfield, K.H. Zou, and W.M. Wells III. Validation of image segmentation by estimating rater bias and variance. In *Ninth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '06)*, volume 4191, pages 839–847, 2006.

- [175] S.K. Warfield, K.H. Zou, and W.M. Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7):903–921, 2004.
- [176] J.D. Warren, J.M. Schott, N.C. Fox, M. Thom, T. Revesz, J.L. Holton, F. Scaravilli, D.G.T. Thomas, G.T. Plant, P. Rudge, and M.N. Rossor. Brain biopsy in dementia . *Brain (Advance Access)*, page awh543, 2005.
- [177] W.M. Wells, 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.
- [178] N.S. Wood, N. Marlow, K. Costeloe, A.T. Gibson, and A.R. Wilkinson. Neurologic and developmental disability after extremely preterm birth. EPI-Cure study group. *New England Journal of Medicine*, 343(6):429–30, 2000.
- [179] R.P. Woods. Characterizing volume and surface deformations in an atlas framework: theory, applications, and implementation. *NeuroImage*, 18(3):769–788, 2003.
- [180] R.P. Woods, S.T. Grafton, C.J. Holmes, S.R. Cherry, and J.C. Mazziotta. Automated image registration: I. general methods and intrasubject, intramodality validation. *Journal of Computer Assisted Tomography*, 22(1):139–152, 1998.
- [181] 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 nonlinear models. *Journal of Computer Assisted Tomography*, 22(1):153–165, 1998.

- [182] M. Wu, C. Rosano, P. Lopez-Garcia, C.S. Carter, and H.J. Aizenstein. Optimum template selection for atlas-based segmentation. *NeuroImage*, 34(4):1612–1618, 2007.
- [183] H. Yamaura, M. Ito, K. Kubota, and T. Matsuzawa. Brain atrophy during aging: A quantitative study with computed tomography. *Journal of Gerontology*, 35:492–498, 1980.
- [184] S. Younes-Mhenni, M.F. Janier, L. Cinotti, J. C. Antoine, F. Tronc, V. Cottin, P.J. Ternamian, P. Trouillas¹, and J. Honnorat. FDG-PET improves tumour detection in patients with paraneoplastic neurological syndromes. *Brain*, 127(10):2331–2338, 2004.
- [185] L.M. Zatz, T.L. Jernigan, and A.J. Ahumada. Changes on computed cranial tomography with aging: intracranial fluid volume. *American Journal of Neuroradiology*, 3:1–11, 1982.
- [186] K. Zhang and T.J. Sejnowski. A universal scaling law between gray matter and white matter of cerebral cortex. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 97:5621–5626, 2000.
- [187] 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.
- [188] A.P. Zijdenbos, B.M. Dawant, R.A. Margolin, and A.C. Palmer. Morphometric analysis of white matter lesions in MR images: method and validation. *IEEE Transactions on Medical Imaging*, 13:716–724, 1994.
- [189] B. Zitová and J. Flusser. Image registration methods: a survey. *Image Vision Comput.*, 21(11):977–1000, 2003.

Appendix A

Inverting Free-Form Deformations

This section presents a description of the methods used to estimate the inverse of transformations that use free-form deformations (FFDs). This inversion is used as a stage in the generation of average space atlases as described in Chapter 3.

A.1 Composition of FFDs

Prior to a discussion of the inversion of transformations involving FFDs, the basic properties of compositions of such transformations are described in this section.

Let $f(\mathbf{x}) : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ represent a transformation mapping a location \mathbf{x} from the space of one image to another such that $f(\mathbf{x})$ is parametrised by a translation \mathbf{t}_f , an affine matrix \mathbf{M}_f (encoding rotation, scales and shears) and a local displacement field u_f represented by a FFD

$$f(\mathbf{x}) = \mathbf{M}_f \mathbf{x} + \mathbf{t}_f + u_f(\mathbf{x}).$$

Let a second transformation g be similarly characterised

$$g(\mathbf{x}) = \mathbf{M}_g \mathbf{x} + \mathbf{t}_g + u_g(\mathbf{x}).$$

It is possible to combine the effect of the affine matrix and the translation in a single homogeneous transformation matrix but, for ease of presentation, they are kept separate here.

Let $h = f \circ g$, the composition of the transformations, be characterised in the same way. This implies that

$$\begin{aligned}
\mathbf{M}_h \mathbf{x} + \mathbf{t}_h + u_h(\mathbf{x}) &= f(g(\mathbf{x})) \\
&= f(\mathbf{M}_g \mathbf{x} + \mathbf{t}_g + u_g(\mathbf{x})) \\
&= \mathbf{M}_f(\mathbf{M}_g \mathbf{x} + \mathbf{t}_g + u_g(\mathbf{x})) + \mathbf{t}_f + u_f(\mathbf{M}_g \mathbf{x} + \mathbf{t}_g + u_g(\mathbf{x})) \\
&= \mathbf{M}_f \mathbf{M}_g \mathbf{x} + \mathbf{M}_f \mathbf{t}_g + \mathbf{t}_f + \mathbf{M}_f u_g(\mathbf{x}) + u_f(\mathbf{M}_g \mathbf{x} + \mathbf{t}_g + u_g(\mathbf{x}))
\end{aligned}$$

which allows the following identifications

$$\begin{aligned}
\mathbf{M}_h &= \mathbf{M}_f \mathbf{M}_g \\
\mathbf{t}_h &= \mathbf{M}_f \mathbf{t}_g + \mathbf{t}_f \\
u_h(\mathbf{x}) &= \mathbf{M}_f u_g(\mathbf{x}) + u_f(\mathbf{M}_g \mathbf{x} + \mathbf{t}_g + u_g(\mathbf{x})) \\
\text{or } u_h(\mathbf{x}) &= \mathbf{M}_f u_g(\mathbf{x}) + u_f(g(\mathbf{x}))
\end{aligned}$$

A.2 Estimating the inverse

Let g be a transformation with known parameters mapping locations in image X to image Y . A transformation $f(\mathbf{x})$ is required such that it acts as the inverse of g , i.e. $f(g(\mathbf{x})) = \mathbf{x}$. Let the lattices of control points for the FFD components of $f(\mathbf{x})$ and g be Ω_f and Ω_g respectively.

$$\Omega_g = \{\mathbf{x}_{ijk} \in X \mid 0 \leq i \leq l_g, 0 \leq j \leq m_g, 0 \leq k \leq n_g\}$$

$$\Omega_f = \{\mathbf{y}_{ijk} \in Y \mid 0 \leq i \leq l_f, 0 \leq j \leq m_f, 0 \leq k \leq n_f\}$$

As f and g are inverses of each other, it is a reasonable assumption that the

global transformation of $h = f \circ g$ should be the identity, i.e.

$$\mathbf{x} = \mathbf{M}_h \mathbf{x} + \mathbf{t}_h = \mathbf{I} \mathbf{x} + \mathbf{0}$$

which gives

$$\begin{aligned} \mathbf{M}_h = \mathbf{I} &\Rightarrow \mathbf{M}_f = \mathbf{M}_g^{-1} \\ \mathbf{t}_h = \mathbf{0} &\Rightarrow \mathbf{t}_f = -\mathbf{M}_f \mathbf{t}_g = -\mathbf{M}_g^{-1} \mathbf{t}_g \end{aligned}$$

Let the lattice of control point locations for f in image Y be defined as above, then it is possible to iterate over each of the lattice points \mathbf{y}_{ijk} , and estimate where it corresponds to in image X . This can be done using a numerical implementation of a Newton method for solving non-linear equations (*Numerical Recipes in C*). This method provides an approximation \tilde{f} of g^{-1} at the lattice locations \mathbf{y}_{ijk}

$$f(\mathbf{y}_{ijk}) = g^{-1}(\mathbf{y}_{ijk}) \approx \tilde{f}(\mathbf{y}_{ijk})$$

$f(\mathbf{y}_{ijk})$ is represented by

$$f(\mathbf{y}_{ijk}) = \mathbf{M}_f \mathbf{y}_{ijk} + \mathbf{t}_f + u_f(\mathbf{y}_{ijk})$$

which gives an estimate for $u_f(\mathbf{y}_{ijk})$ as

$$u_f(\mathbf{y}_{ijk}) \approx \tilde{f}(\mathbf{y}_{ijk}) - \mathbf{M}_f \mathbf{y}_{ijk} - \mathbf{t}_f.$$

which, in turn, can be expressed in terms of the parameters of g , which are known

$$u_f(\mathbf{y}_{ijk}) \approx \tilde{f}(\mathbf{y}_{ijk}) - \mathbf{M}_g^{-1} \mathbf{y}_{ijk} + \mathbf{M}_g^{-1} \mathbf{t}_g$$

The values $u_f(\mathbf{y}_{ijk})$ represent estimates for the *displacements* at the lattice locations for the local displacement field u_f . The remaining step is to estimate values for the coefficient vectors at each control point that are needed to parametrise

u_f . This can be achieved using the method described by Unser [167] where the values of a signal are known at a set of knot locations and coefficients need to be calculated that will provide a spline that interpolates the signal at the knots. The method makes use of the z-transform and is implemented as a very efficient filtering process.

The remaining choice for the inverse transformation estimate is in the placement of the control points of f , $\{\mathbf{y}_{ijk}\}$. This was simply determined once a control point spacing was decided. If δ represents the distance between adjacent control points, then the control point lattice was determined as the maximal lattice that could be placed within the image volume Y that is simultaneously centred in the volume and aligned with the voxel axes.