Imperial College London
Department of Computing

Segmentation of brain MRI during early childhood

Maria Murgasova

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Computing of the Imperial College London and
the Diploma of Imperial College, October 2008

Abstract

The objective of this thesis is the development of automatic methods to measure the changes in volume and growth of brain structures in prematurely born infants. Automatic tools for accurate tissue quantification from magnetic resonance images can provide means for understanding how the neurodevelopmental effects of the premature birth, such as cognitive, neurological or behavioural impairment, are related to underlying changes in brain anatomy. Understanding these changes forms a basis for development of suitable treatments to improve the outcomes of premature birth.

In this thesis we focus on the segmentation of brain structures from magnetic resonance images during early childhood. Most of the current brain segmentation techniques have been focused on the segmentation of adult or neonatal brains. As a result of rapid development, the brain anatomy during early childhood differs from anatomy of both adult and neonatal brains and therefore requires adaptations of available techniques to produce good results.

To address the issue of anatomical differences of the brain during early childhood compared to other age-groups, population-specific deformable and probabilistic atlases are introduced. A method for generation of population-specific prior information in form of a probabilistic atlas is proposed and used to enhance existing segmentation algorithms.

The evaluation of registration-based and intensity-based approaches shows the techniques to be complementary in the quality of automatic segmentation in different parts of the brain. We propose a novel robust segmentation method combining the advantages of both approaches. The method is based on multiple label propagation using B-spline non-rigid registration followed by EM segmentation.

Intensity inhomogeneity is a shading artefact resulting from the acquisition process, which significantly affects modern high resolution MR data acquired at higher magnetic field strengths. A novel template based method focused on correcting the intensity inhomogeneity in data acquired at higher magnetic field strengths is therefore proposed.

The proposed segmentation method combined with proposed intensity inhomogeneity correction method offers a robust tool for quantification of volumes and growth of brain structures during early childhood. The tool have been applied to 67 T1-weigted images of subject at one and

two years of age.

Acknowledgements

I would like to express my thanks to:

- My supervisor Prof. Daniel Rueckert for giving me the opportunity to work on this exciting project, for his invaluable guidance, support and feedback without which completion of these thesis would not be possible
- Prof. David Edwards and Dr. Mary Rutherford for giving direction and medical insights
- Dr. Leigh Dyet for the manual segmentation on which the results of these thesis are built
- Prof. Jo Hajnal for the data and advice on Physics of MRI
- My colleagues Dr. Paul Aljabar, Dr. Kanwal Bhatia, Dr. Serena Counsell, Dr. James Boardman, Ioannis Gousias and Dr. Mustafa Anjari for helping with various aspects of my research
- EPSRC grant MIAS-IRC for providing funding for my research
- My parents, Zuzi and Vincent for their support, encouragement and patience

Dedication

Dedicated to Zuzana

Contents

Abstract **3**

Acknowledgements **5**

1 Introduction **33**

1.1 Brain development in prematurely born children 33

1.2 Imaging of the developing brain 35

1.3 Challenges in segmenting brain MRI 38

1.3.1 Intensity distribution in brain MRI 38

1.3.2 Intensity inhomogeneity 39

1.4 Contribution of the thesis 40

1.5 Outline of the thesis 41

2 Segmentation methods for brain MRI **43**

2.1 Analysis of brain MRI 43

2.2 Registration-based segmentation 44

2.2.1 Image registration 45

2.2.2 Similarity measures 46

2.2.3 Rigid and affine registration 47

2.2.4 Non-rigid registration 48

2.2.5 B-spline non-rigid registration 49

2.2.6 B-spline registration-based segmentation 51

2.3 Intensity-based methods 53

2.3.1 C-means clustering 54

2.3.2 kNN classification 55

2.3.3 Entropy minimisation 58

2.4 EM-based segmentation 60

2.4.1 Optimization criteria 61

2.4.2 Probabilistic framework for image segmentation 62

2.4.3 Optimization via Expectation Maximization 63

2.4.4 Segmentation via EM 65

2.4.5 Mixture of Gaussians as a model for MR image intensities 67

2.5 Prior information for EM-based segmentation 69

2.5.1 Probabilistic atlases 69

2.5.2 Aligning the probabilistic atlas 70

2.5.3 Incorporating the probabilistic brain atlas into EM segmentation 71

2.5.4 Prior information for segmentation of brain structures 73

2.5.5 Including neighbourhood information in EM-based segmentation 75

2.6 Brain extraction 76

2.7 Segmentation of brain MRI during early childhood 77

2.8 Comparison of registration-based and intensity-based approaches 79

2.9 Conclusion 79

3 Bias correction of brain MRI 80

3.1 Intensity inhomogeneity 80

3.2 Bias correction methods 81

3.2.1 C-means clustering 82

3.2.2 N3 method 82

3.2.3 Entropy minimization 83

3.3 Bias field estimation via EM 84

3.3.1 Parametric bias field estimation via the ML principle 84

3.3.2 Non-parametric bias field estimation via MAP principle 87

3.4 Simultaneous segmentation, registration and bias correction 88

3.4.1 Incorporating the registration of the probabilistic atlas into the EM segmentation algorithm 89

3.4.2 Minimization of objective function using iterated conditional modes . . . 91

3.5 Template-based bias correction 93

3.6 Comparison of the methods 95

3.7 Conclusion 96

4 Segmentation of brain MRI during early childhood 97

4.1 Introduction 97

4.1.1 Datasets 98

4.1.2 Pre-processing 99

4.1.3 Brain tissues vs. brain structures 100

4.1.4 Manual segmentation of the reference subject 101

4.1.5 Evaluation using manual segmentations 102

4.1.6 Implementation 103

4.2 Registration-based segmentation 103

4.2.1 Deformable atlas and registration-based segmentation 104

4.2.2 Non-rigid registration 105

4.2.3 Optimizing registration parameters 106

4.3 EM-based segmentation 107

4.3.1 Probabilistic atlas 108

4.3.2 Creating a population-specific atlas for EM segmentation 110

4.3.3 Improvement of EM brain tissue segmentation using population-specific probabilistic atlas 112

4.3.4 EM segmentation of brain structures 116

4.4 Combination of registration-based and EM-based segmentation 117

4.4.1 Creating a subject-specific atlas for EM segmentation 117

4.4.2 Multiple registration of training images to a new subject 119

4.4.3 Algorithm for combined registration-based and EM-based segmentation . 121

4.5 Comparison of the methods 124

4.6 Calculating volume and growth of brain structures during early childhood 127

4.6.1 Measuring volumes of brain structures 127

4.6.2 Measuring growth of brain structures 131

4.6.3 Measuring growth of brain tissues 132

4.7 Discussion 134

4.8 Conclusion 136

5 Intensity inhomogeneity correction of brain MR images 138

5.1 Processing brain MRI in presence of strong intensity inhomogeneity 138

5.2 Template-based bias correction 142

5.3 B-spline model for the bias field 142

5.4 Bias field estimation by weighted least squares 144

5.5 Outlier estimation 148

5.6 Smoothness constraint for B-spline functions 153

5.7 Incorporating the smoothness constraint into the bias field estimation 157

5.8 Implementing the bias field estimation 158

5.9 Intensity matching 161

5.10 Template-based bias correction algorithm 165

5.11 Conclusion 167

6 Intensity inhomogeneity correction and segmentation of brain MRI 168

6.1 Data and Materials 169

6.2 Determining the parameters for template-based bias correction 169

6.2.1 Estimating parameter λ 170

6.2.2 Two-dimensional search for the parameters λ and L 174

6.3 Registration-based segmentation 180

6.3.1 Determining the registration parameters 180

6.3.2 Interleaved versus sequential registration 182

6.3.3 Registration-based segmentation algorithm in the presence of strong intensity inhomogeneity 183

6.4 EM segmentation with a population-specific probabilistic atlas 184

6.4.1 Registration of template image during bias correction and EM segmentation 184

6.4.2 EM segmentation in the presence of strong intensity inhomogeneity . . . 186

6.5 Combined registration-based and EM-based segmentation 188

6.6 Comparison of the methods 190

6.7 Measuring volumes of brain structures 194

6.8 Discussion 196

6.9 Conclusion 199

7 Conclusion 200

7.1 Contribution 200

7.2 Discussion 202

7.3 Limitations and future work 203

7.4 Summary 205

Publications **205**

Bibliography **206**

List of Tables

4.1 Correspondence between brain tissues and brain structures. 100

4.2 Parameters for the non-rigid registration. All sizes are given in millimeters. . . . 106

4.3 Agreement of the manual segmentation with the results of the registration-based segmentation with different parameters given in table 4.2 as measured with the Dice overlap metric. The last column shows the approximate running times of the registration algorithm. 107

4.4 Parameters for non-rigid registration. All sizes are given in millimeters. 107

4.5 Agreement of the manual segmentations with the results of the EM segmentation into WM, GM and CSF using different probabilistic atlases measured with Dice metric. 113

4.6 The proportion of subcortical GM structures caudate and thalamus classified as GM using EM segmentation into WM, GM and CSF with different probabilistic atlases. 113

4.7 Agreement of the manual segmentation with the results of the EM segmentation of brain structures using a population-specific atlas for the age of two years measured with Dice metric. 117

4.8 Agreement of the manual segmentations with the results of the multiple label propagation with different levels of precision of the subject-specific probabilistic atlas measured with Dice metric. The last column shows an example of the approximate running times for registration of a training image to the image to be segmented. 120

4.9 Agreement of the manual segmentations with the results of the combined registration-based and EM-based segmentation with different levels of precision of the subject-specific probabilistic atlas measured with Dice metric. The last column shows an example of the approximate running times for registration of a training image to the image to be segmented. 120

4.10 Parameters for the non-rigid registrations during the second step of the creation of the subject-specific atlases. All sizes are given in millimeters. 121

4.11 Overlap between manual and automatic segmentation of different segmentation methods on four two year old subjects measured with Dice metric. 124

4.12 Volumes of the brain structures at the age of one year given in cm^3. 128

4.13 Volumes of the brain structures at the age of two years given in cm^3. 128

4.14 Growth factors between the age of one and two years. 131

4.15 Mean growth factors for WM and GM between one and two years estimated by combined segmentation and compared to the results obtained by Aljabar *et al.* [1] using EM segmentation with a population-specific probabilistic atlas and deformation based morphometry. 133

6.1 Conversion table between a number of control points L at each dimension and B-spline control point spacing of the model for bias field. 170

6.2 Parameters of the non-rigid registration: The columns of the table are the B-spline control point spacing at a given level of registration, resolution of the subsampled image, standard deviation of Gaussian blurring used before subsampling and length of steps when optimizing the non-rigid transformation. r denotes the largest dimension of a voxel in mm, in either the target or source image. 171

6.3 Parameters for the registration at four resolution levels. The columns of the table are B-spline control point spacing at a given resolution level, resolution of the subsampled image, variance of Gaussian blurring used before subsampling and the step size of the finite difference approximation of the derivatives during the optimization of the transformation. r denotes the largest dimension of a voxel in mm, in either target or source image. 175

6.4 Parameters for the non-rigid registration. All sizes are given in millimeters. . . . 181

6.5 Agreement of the manual segmentation with the results of the registration-based segmentation with different parameters given in table 6.4 as measured with the Dice overlap. 181

6.6 Optimal choice of parameters for non-rigid registration of 3T images. All sizes are given in millimeters. 182

6.7 Agreement of the manual segmentation with the results of the registration-based segmentation obtained using interleaved and sequential registration and bias correction algorithm as measured with the Dice overlap. 183

6.8 Performance of EM segmentation with population-specific atlas on four subjects. 187

6.9 Registration parameters for creating subject-specific atlases for 3T images. All sizes are given in millimeters. 189

6.10 Performance of EM segmentation with population-specific atlas on four 3T MRI (three at the age of two years and one at the age of one year). Two different groups of training images were used: 35 1T images corrected by N3 [78] (combined 1T) and 35 3T images corrected by proposed template-based bias correction method (combined 3T). 189

6.11 Performance of different segmentation methods on brain MRI of three subjects at two years of age and one subject at one year of age in the presence strong intensity inhomogeneity. 190

6.12 Mean volumes and standard deviation of brain structures calculated from 1T and 3T data. Results given in cm^3 194

6.13 Mean relative volumes and standard deviation of brain structures calculated as a percentage of the whole brain volume from 1T and 3T data. 195

List of Figures

1.1 (a) T2-weighted MRI of an infant born at 26 weeks, imaged at 40 weeks gestational age; (b) T2-weighted MRI of an infant born and imaged at 40 weeks gestational age. The black arrow in (a) points to a large area filled with cerebrospinal fluid in prematurely born infant, while in full-term infant this space is filled with brain tissue, see the black arrow in (b). 34

1.2 T1-weighted MR images of developing brain from top to bottom: baby at birth, one year old child, two year old child and adult. 37

1.3 Histogram of a bias-corrected 1T T1-weighted MRI of a brain of a two year old child. Histograms of the tissue classes are based on manual segmentation. GM and CSF distributions slightly differ from Gaussian due to partial volume effect and presence of other tissues, such as vessels, in CSF. 39

1.4 Histogram of 1T T1-weighted MRI of a neonatal brain with inverted WM/GM contrast due to large proportion of non-myelinated WM 39

1.5 T1-weighted MRI scanned at different magnetic field strengths: (a) 0.5T with virtually no bias; (b) 1T with slightly visible bias; (c) 3T with strong bias. First row: original image; second row: corrected image; third row: bias field; fourth row: histogram of the original image; fifth row: histogram of the corrected image. 42

2.1 Comparison of different types of registrations. Top row: the source image with iso-contours registered to target image with (a) rigid registration; (b) affine registration; (c) non-rigid registration [73]. Bottom row: The target image with overlayed iso-contours of source image registered with (d) rigid registration; (e) affine registration; (f) non-rigid registration [73]. 52

2.2 Gaussian mixture model - fitting the mixture of Gaussians to the normalized image histogram. 68

2.3 The MNI probabilistic atlas [34]: (a) the template; brain tissue prior probability maps: (b) WM; (c) GM; (d) CSF. 69

2.4 Probability maps from ICBM152 probabilistic atlas [58]. From left to right: GM, WM and CSF. 70

4.1 Removing background voxels: (a) MRI of a two year old child; (b) image with adjusted contrast to show the background noise; (c) final results of the background removal procedure. 99

4.2 Manual segmentation of brain structures for the two years old reference subject contains these 11 structures: CSF, cortex, caudate, putamen, substantia nigra, cerebellar GM, thalamus, pallidum, brainstem, WM and cerebellar WM. 101

4.3 The anatomies of the brain in adulthood and during early childhood differ significantly in the subcortical area, with the internal capsule being larger in adults than in children: (a) the adult brain; (b) the MNI305 brain atlas (WM probability map); (c) the CCHMC5-9 brain atlas (WM probability map); (d) the brain of a two year old child; (e) misclassification of subcortical GM in a two year old child when the MNI305 atlas is used; (f) misclassification of subcortical GM in a two year old child when the CCHMC5-9 atlas is used. 109

4.4 Histogram of the tissue intensity distributions for WM, GM and thalamus (a subcortical GM structure) based on a manual segmentation of brain MRI of a two year old child. The intensity distribution of the thalamus overlaps with those of WM and GM. 110

4.5 Creating a population-specific atlas: A single manual segmentation of a reference subject is transferred to a population of training subjects via non-rigid registration. All segmentations are then aligned to the reference subject via affine registration and averaged. 111

4.6 Comparison of the EM segmentations with different probabilistic atlases. (a) MRI of a two year old child. (b) Manual segmentation. WM probability maps for different probabilistic atlases: (c) MNI305 atlas; (d) CCHMC5-9 atlas; (e) population-specific atlas for the age of two years. EM segmentation into WM, GM and CSF using: (f) MNI305 atlas; (g) CCHMC5-9 atlas; (h) population-specific atlas for the age of two years. The arrow points at internal capsule showing different widths depending on the probabilistic atlas used. 114

4.7 The amount of misclassification in the thalamus depends on the type of the probabilistic atlas used. (a) MRI of a two year old child. (b) Manual segmentation. EM segmentation using (c) MNI305 atlas; (d) CCHMC5-9 atlas; (e) population-specific atlas for the age of two years. 115

4.8 Comparison of different probabilistic atlases for the area of the brain shown in fig. 4.7. The probability maps for WM (the first row) and GM (the second row). (a) MNI305 atlas; (b) CCHMC5-9 atlas; (c) population-specific atlas for the of age two years. 115

4.9 Segmentation of a brain MRI of a two year old child: (a) the original MRI; (b) EM-based segmentation into 11 anatomical structures; (c) EM-based segmentation into three tissue classes. 116

4.10 Creation of a subject-specific atlas: A manual segmentation of a reference subject is transferred to a population of subjects by non-rigid registration. All the segmentations are then aligned with the new subject to be segmented using non-rigid registration. In the final step the resulting registrations are averaged to create a subject-specific probabilistic atlas. 118

4.11 Comparison of EM-based and combined segmentations of brain structures: (a) EM segmentation with population-specific atlas; (b) Combined segmentation with subject-specific atlas with precision 10mm; (c) Combined segmentation with subject-specific atlas with precision 5mm; (d) Combined segmentation with subject-specific atlas with precision 2.5mm. Probability maps for WM in (e) population-specific atlas created with affine registration; (f) subject-specific atlas created with 10mm non-rigid registration; (g) subject-specific atlas created with 5mm non-rigid registration; (h) subject-specific atlas created with 2.5mm non-rigid registration. 123

4.12 (a) Brain MRI of a two year old child. Automatic segmentation of brain structures using the three segmentation methods: (b) Registration-based segmentation; (c) EM-based segmentation; (d) Combined segmentation.
The arrows from top to bottom: 1. small CSF-filled space fails to be segmented by Registration-based segmentation (b), improved segmentation by EM segmentation (c), well delineated by combined segmentation. 2. Thin plate of WM fails to be segmented by registration-based segmentation (b), improved segmentation by EM segmentation (c), well delineated by combined segmentation. 3. Substantia nigra approximately delineated by registration-based segmentation (b), failed segmentation by EM segmentation (c), well delineated by combined segmentation. 4. Boundary of the cerebellum correctly delineated by registration-based segmentation (b), failed segmentation by EM segmentation (c), correctly delineated by combined segmentation. 125

4.13 (a) Brain MRI of a two year old child severely affected by focal lesions. Automatic segmentation of brain structures using the three segmentation methods: (b) Registration-based segmentation; (c) EM-based segmentation; (d) Combined segmentation. The EM-based segmentation with an affinely aligned probabilistic atlas fails to adapt to the enlarged CSF-filled space. 126

4.14 Mean volumes of brain structures at one and two years. The error bars are showing the standard deviation. 129

4.15 Relative mean volumes of brain structures at one and two years, given as the percentage of the whole brain volume. The error bars are showing the standard deviation. 129

4.16 Mean volumes of brain structures at one and two years. The error bars are showing the standard deviation. 130

4.17 Relative mean volumes of brain structures at one and two years, given as the percentage of the whole brain volume. The error bars are showing the standard deviation. 130

4.18 Mean growth factors from one to two years for brain structures. The error bars are showing the standard deviation. 131

4.19 Changes in the volume of WM and GM between the age of one and two years. . 132

4.20 (a) Relative volumes of WM and GM tissues shown as a function of age. (b) Average relative volumes of WM and GM tissues at approximate ages of one and two years. The percentage of WM increases, while the percentage of GM decreases between ages one and two years. 133

4.21 Normalized histograms brain structures. The intensities are brighter for subcortical GM structures (caudate, putamen, thalamus, pallidum) as well as subcortical WM structure corpus callosum, compared to cortical GM and WM, respectively. 135

4.22 (a) The original MR image of a two year old child. Segmentation of WM: (b) EM with population-specific atlas; (c) SPM5 [7] with non-rigidly aligned standard adult atlas; Parts of thalamus and pallidum are misclassified as WM (e.g. in the region in black circle). 136

5.1 A EM-based segmentation algorithm yields poor results in the presence of strong intensity inhomogeneity: (a) a segmentation of an image with strong intensity inhomogeneity; (b) an improved segmentation after the intensity inhomogeneity has been corrected. The arrows point at some of the locations where the segmentation has been improved after intensity inhomogeneity correction. The bias field has been corrected with a novel template-based method described later in this chapter. 139

5.2 Non-rigid registration yields poor results in the presence of strong intensity inhomogeneity: (a) a target image with strong intensity inhomogeneity; (b) a source image with iso-contours; (c) the target image with strong intensity inhomogeneity with overlayed iso-contours of the source image registered by non-rigid registration without bias correction; (d) the source image registered by non-rigid registration without bias correction; (e) the bias corrected target image with overlayed iso-contours of the source image registered by non-rigid registration after bias correction; (f) the source image registered by non-rigid registration after bias correction. Note the boundaries of subcortical structures (see the arrow): Misaligned in (c); Correctly aligned in (e). The B-spline control point spacing at the final resolution level of non-rigid registration is 5 mm. The bias field has been corrected with a novel template-based method described later in this chapter. 141

5.3 Template-based intensity inhomogeneity correction using least square fitting: (a) the original MRI with intensity inhomogeneity; (b) the MRI after intensity inhomogeneity correction with least square fitting; (c) the MRI after intensity inhomogeneity correction with weighted least square fitting; (d) the intensity difference between the original MRI and the template MRI; (e) the bias field obtained by least square fitting; (f) the bias field obtained by weighted least square fitting with the weights calculated to reduce the influence of the outliers. The control point spacing in the B-spline model used is approximately 30mm and a smoothness constraint (explained later in this chapter) with a weighting coefficient 2.5 is also applied during both, least-square and weighted least-square fit. 147

5.4 The weight function: The x-axis represents the intensities of the difference image $Y' - X$. The function shown in grey is the estimated intensity distribution $P(y' - x|\gamma, \beta)$ of the image $Y' - X$ rescaled to $< 0, \frac{1}{\sigma_0^2} >$. The function shown in black is the weight function $w(y' - x)$. Both functions have been calculated using parameters $\sigma_1 = 3\sigma_0$, $c_0 = 0.6$, and $c_1 = 0.4$. 151

5.5 The influence of the B-spline control point spacing on the weighted least square fit: (a) the original image; The corrected image and the corresponding bias field obtained with the B- spline control point spacing: (b) 150mm; (c) 75mm; (d) 50mm; (e) 40mm; (f) 30mm; (g) 25mm; (h) 20mm. 152

5.6 The influence of the smoothness constraint on the weighted least square fit with a B-spline control point spacing of 15mm. The corrected image and the corresponding bias field with the weighting coefficient: (a) $\lambda = 0$; (b) $\lambda = 0.15$; (c) $\lambda = 1.5$; (d) $\lambda = 15$. 159

5.7 Intensity matching: (a) original image; (b) reference image transformed by adjusting intensity mean and variance to original image; (c) reference image transformed by least square fitting in coordinate y results in poorer match. Intensity histograms of the images are shown in (d), (e), and (f). 2D intensity histogram of original image and aligned reference image with linear transformation obtained by: (g) adjusting intensity mean and variance; (h) linear regression in coordinate y. The linear transformation obtained by adjusting intensity mean and variance (d) provides more precise fit. 164

6.1 Segmentation results following template-based bias correction using different values of λ. All images were bias corrected with $L = 6$, corresponding to control point spacing of approximately 30 mm. The value of λ is shown on the horizontal axis. The agreement with the manual segmentation as calculated by the Dice metric is shown on the vertical axis. 172

6.2 Registration-based segmentation using the parameters from table 6.2. Before segmentation the images were bias corrected with parameters λ shown on horizontal axis and L shown as a series of graphs. The template was registered with the images for the purposes of bias correction with a control point spacing 5mm. The agreement with the manual segmentation as calculated by the Dice metric is shown on vertical axis. 175

6.3 EM-based segmentation. Before segmentation the images were bias corrected with parameters λ shown on horizontal axis and L shown as a series of graphs. The template was registered with the images for the purposes of bias correction with a control point spacing 5mm. The agreement with the manual segmentation as calculated by the Dice metric is shown on vertical axis. 176

6.4 Registration-based segmentation with parameters from table 6.3. Before segmentation the images were bias corrected with the parameters λ shown on the horizontal axis and L shown as a series of graphs. The template was registered with the images for the purposes of bias correction with a control point spacing of 5mm. The agreement with the manual segmentation as calculated by the Dice metric is shown on the vertical axis. 178

6.5 EM-based segmentation. Before segmentation the images were bias corrected with the parameters λ shown on the horizontal axis and L shown as a series of graphs. The template was registered with the images for the purposes of bias correction with a control point spacing of 5mm. The agreement with the manual segmentation as calculated by the Dice metric is shown on the vertical axis. . . . 178

6.6 The influence of the parameter λ on the bias correction. Values of $\lambda = 0.1$ results in a darker centre of the brain compared to the tissue closer to the skull and less contrast in the subcortical area than values $\lambda = 3$: (a) $\lambda = 0.1$; (b) $\lambda = 3$. . . . 179

6.7 The influence of the parameter λ on the EM-based segmentation of WM. First row: Image with overlayed segmentation. Second row: Prior probability maps for WM with overlayed segmentation. (a) $\lambda = 0.1$; (b) $\lambda = 3$. Value $\lambda = 0.1$ results in a brighter tissue closer to the skull compared to the center of the brain. WM close to skull can be therefore still recognized by EM segmentation due to very high intensity values, even though the prior probabilities in the location are low. Examples pointed out by the arrows. 179

6.8 EM-based segmentation. The images were bias corrected with parameters $L = 6$ and $\lambda = 2$ after different registrations of the template images shown on the horizontal axis. After bias correction, the population-specific probabilistic atlas was aligned (a) for each bias corrected image separately; (b) using the same registration with image corrected after rigid alignment with the template. The agreement of EM segmentation with manual segmentation calculated by Dice metric is shown on the vertical axis. 185

6.9 Comparison of segmentation methods on 3T data, sagittal view. (a) 3T MRI of two years old child; (b) registration-based segmentation; (c) EM segmentation with population-specific atlas; (d) Combined EM and registration based segmentation. Arrows from top to bottom: 1. CSF space imprecisely delineated by registration-based segmentation, undetected by EM segmentation, well delineated by combined segmentation. 2. Substantia nigra correctly delineated by registration-based segmentation and combined segmentation, incorrectly by EM segmentation. 3. Part of the brain not segmented by EM segmentation due to imprecise alignment of the probabilistic atlas, correctly delineated by combined segmentation, less precisely by registration-based segmentation. 191

6.10 Comparison of EM segmentation with a population-specific atlas and combined segmentation with a subject-specific atlas: Detail of the sagittal slice shown in fig. 6.9 demonstrates how better prior information for CSF improves the segmentation: (a) EM segmentation with population-specific probabilistic atlas; (b) combined segmentation with subject-specific probabilistic atlas; (c) prior for CSF in population-specific atlas; (d) prior for CSF in subject-specific atlas. Segmentation of CSF space fails with the EM segmentation due to insufficient prior probability for CSF in the location, while the combined method classifies CSF in this area correctly thanks to the more precise subject-specific atlas. 192

6.11 Comparison of EM segmentation with population-specific atlas and combined segmentation with subject-specific atlas, coronal view. Segmentations shown in the first row. Probability maps for WM (second row), GM (third row) and CSF (fourth row). (a) EM segmentation with population-specific probabilistic atlas; (b) Combined segmentation with subject-specific probabilistic atlas. 193

6.12 Comparison of mean volumes of brain structures obtained from 3T and 1T data. 195

6.13 Comparison of mean relative volumes of brain structures obtained from 3T and 1T data. 195

6.14 Comparison of mean volumes of brain structures obtained from 3T and 1T data. 196

6.15 Comparison of mean relative volumes of brain structures obtained from 3T and 1T data. 196

6.16 Bias correction of brain MRI of a two year old subject scanned at 1T magnetic field strength. First row: the original image. Second row: Image corrected by N3 and the estimated bias field. Third row: Image corrected by our proposed method and the estimated bias field. 198

Chapter 1

Introduction

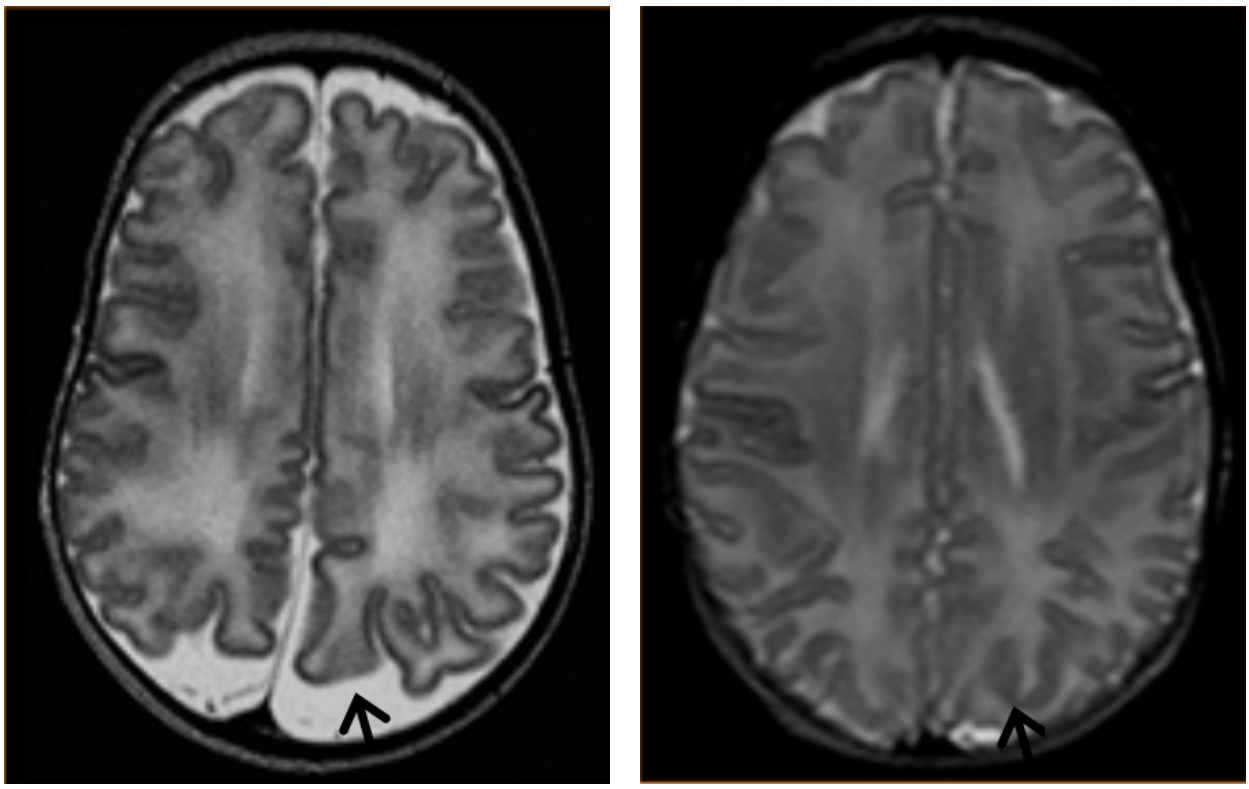
A normal pregnancy usually lasts between 38 to 42 weeks and babies born within this period are referred to as full-term infants. Premature birth, defined as a delivery of a baby before 37 completed weeks of pregnancy, affects 5-7% of all children [89]. Over the last 20-30 years, developments in neonatal medicine improved the outcomes of prematurely born infants. However, the effects of premature birth still extend into later life [57], most likely as a result of the profound effect of premature birth on brain development.

The objective of this thesis is the development of automatic methods to measure the changes in volume and growth of brain structures in prematurely born infants as well as in full-term infants. Automatic tools for accurate tissue quantification can provide the means for understanding how the neurodevelopmental effects of the premature birth are related to underlying changes in brain anatomy [1]. Understanding these changes forms a basis for the development of suitable treatments to improve the outcomes of premature birth.

1.1 Brain development in prematurely born children

Approximately 40% of survivors of premature birth suffer cognitive, neurological or behavioral impairment. About 12% of extremely premature infants (born before 28 weeks of gestation) develop cerebral palsy and 50% have some form of disability at 6 years of age [57].

In comparison to term born infants, preterm infants at term have reduced cortical and subcortical grey matter volume, enlarged fluid-filled spaces and white matter abnormalities [53], [44], see fig. 1.1.



(a) (b)

Figure 1.1: (a) T2-weighted MRI of an infant born at 26 weeks, imaged at 40 weeks gestational age; (b) T2-weighted MRI of an infant born and imaged at 40 weeks gestational age. The black arrow in (a) points to a large area filled with cerebro-spinal fluid in prematurely born infant, while in full-term infant this space is filled with brain tissue, see the black arrow in (b).

The major predictors of altered cerebral volumes are gestational age at birth and the presence of cerebral white matter injury. Infants with significantly reduced cortical and subcortical grey matter volumes and increased cerebro-spinal fluid volume at term-equivalent age exhibit moderate to severe neurodevelopmental disability at one year of age [43]. Other studies have correlated reductions in tissue volume during adolescence with neurocognitive and behavioural problems [82], [2], [62]. While the neural correlates for cerebral palsy are well described [91], the cerebral abnormalities underlying more subtle problems such as minor motor impairment or inattention remain unclear.

To improve the neurodevelopmental outcome in prematurely born children, it is crucial to

understand the nature of brain abnormalities that underlie these disabilities, as well as factors (such as white matter injury) involved in pathological development. Imaging of the developing brain therefore plays a key role in measuring and visualizing these brain abnormalities.

1.2 Imaging of the developing brain

Magnetic resonance (MR) imaging is a non-invasive medical imaging technique providing high resolution 3D volumetric data with high intensity contrast between soft tissues. It is therefore very well suited for analyzing human brain anatomy. MR imaging uses a powerful magnetic field to align the nuclear magnetization of hydrogen atoms in water in the body. A sequence of magnetic fields systematically alters the alignment of this magnetization, causing the hydrogen nuclei to produce a rotating magnetic field detectable by the scanner. MR imaging can be used to image not only the anatomy of the brain, but also the vasculature as well as microstructure of brain tissues. For example, diffusion tensor MR imaging is used to visualize white matter tracts and MR angiography detects the blood flow in the brain.

Magnetic fields, unlike X-rays used in computer tomography or gamma rays in positron emission tomography, are not harmful to living cells, making MR imaging a non-invasive technique, suitable for imaging neonates and children. Cranial ultrasonography is another non-invasive technique suitable for soft tissue widely used in neonatal imaging. However, ultrasound waves are poorly conducted through the bones and therefore cranial ultrasonography cannot be performed once the fontanel (gaps between the bones of the cranium) has been closed, usually between ages 3 months and one year. Unlike in MR imaging, the resolution and contrast between soft tissues provided by cranial ultrasonography are not sufficient for detecting subtle changes in brain anatomy correlated with neurocognitive and behavioral impairment.

MR imaging allows the distinction of different tissue types in the brain, e.g. white matter (WM), grey matter (GM) and cerebro-spinal fluid (CSF), based on their intensities. In T1-weighted MR images (MRI) the intensity distributions of the tissues in the normal adult brain are as follows: WM has the highest intensity, GM has a medium intensity and CSF has the lowest

intensity (see fig. 1.2, bottom row). In T2-weighted MR images the intensity distributions are inverted with CSF having the highest intensity and WM the lowest intensity.

The appearance of fetal and neonatal brains in MRI differs significantly from a mature adult brain. Fetal and neonatal T1-weighted MRI are characterized by an inverted contrast of WM and GM (see fig. 1.2, top row). As the brain matures, the darker intensities of WM present in the MRI of the fetal brain gradually increase, eventually exhibiting a bright intensity pattern on T1-weighted MRI. This is caused by a decrease in both T1 and T2 times as the water content decreases and myelin sheath forms around the WM tracts. Myelin is an insulating layer made up of protein and fatty substances that allows rapid and efficient transmission of impulses along the nerve cells. The process of myelination, or formation of the myelin, starts before birth and is most rapid during the first two or three years of life. Myelination appears to occur earliest in the tracts connecting the cerebellum and brainstem, which can be already identified at the age of three months. By the age of one year, all major white matter tracts including the corpus callosum, subcortical white matter, and the internal capsule are well defined. However, the refined configuration of the adult brain is not attained until early adolescence [41], [74]. As a result, WM and GM of very young children can be very difficult to distinguish in MRI, due to age and location dependent WM/GM contrast.

By the age of one or two years the process of myelination is almost complete and the tissue contrast is very similar to the contrast in the adult brain (see fig. 1.2, second and third row). Compared to neonates, the cortical folding has increased resulting in a more complicated shape of WM and cortical GM than in neonates. Even though tissue contrast is already adult-like, brain structures in young children have different shapes and sizes [97]. Also, the smaller size of the brains structures together with the limited resolution of MRI result in an increased number of voxels containing two or more tissue classes, causing mixing of tissue intensities.

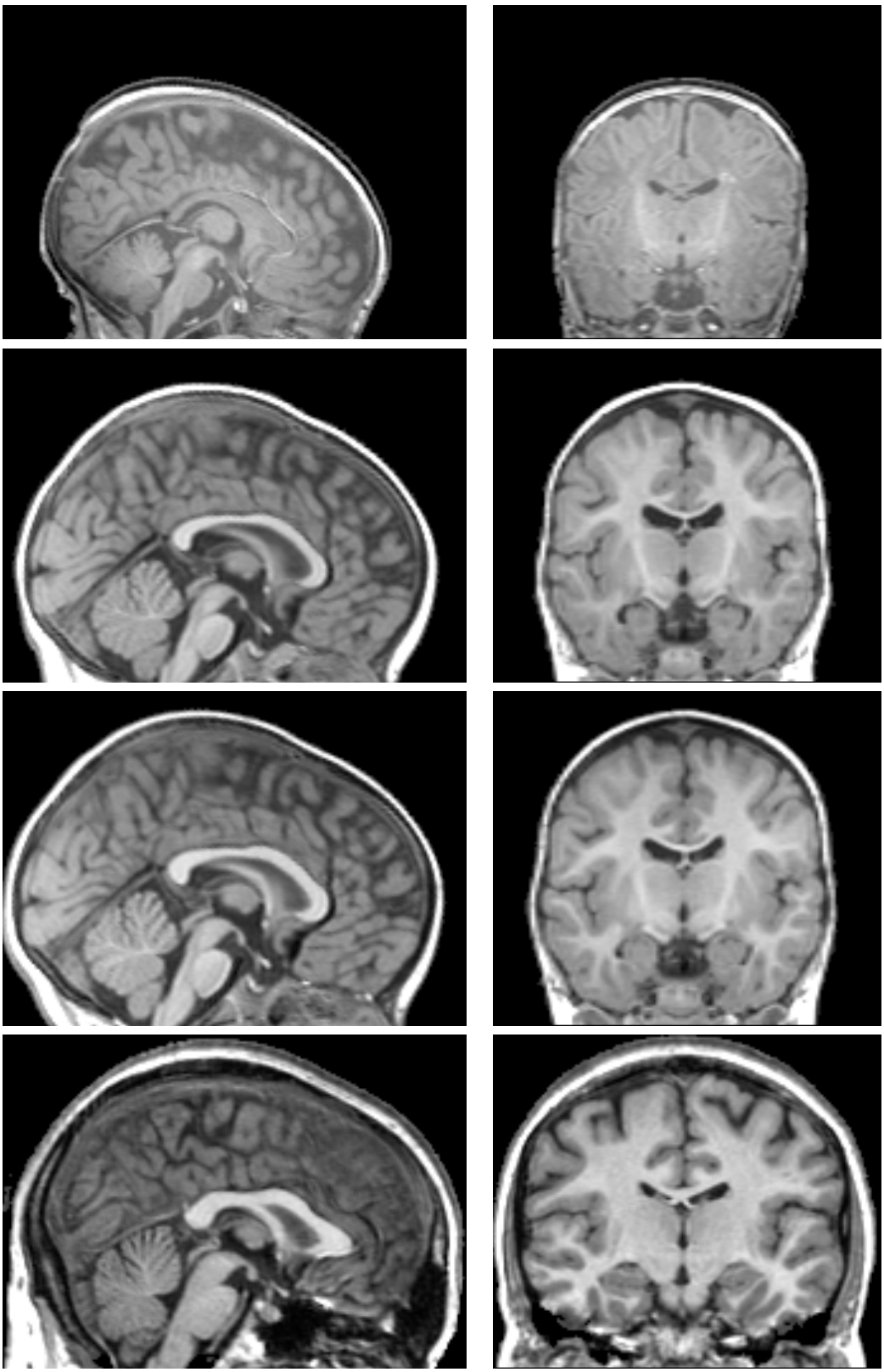


Figure 1.2: T1-weighted MR images of developing brain from top to bottom: baby at birth, one year old child, two year old child and adult.

1.3 Challenges in segmenting brain MRI

Even though the three basic tissue classes are mostly visually distinguishable on MRI, the automatic segmentation of the brain tissues is a non-trivial problem. In addition to the fact that intrinsic properties of brain tissues (e.g. relaxation times T1, T2) result in within-tissue variation of intensities, MRI are also corrupted by artefacts during the image acquisition, such as patient motion and noise, or mixing of intensities due to more than one tissue present at the voxel, often referred to as *partial volume effect* (PVE). Perhaps the most challenging artefact is the *intensity inhomogeneity*, also called the *bias field*. The bias field is a smooth low-frequency multiplicative artefact caused by inhomogeneity of magnetic field during scanning process, altering intensities across the image [77], [78].

1.3.1 Intensity distribution in brain MRI

If we neglect the influence of the bias field, the histogram of brain MRI, after removing the background voxels, will have three main peaks corresponding to the three main tissue classes, see fig. 1.3. If we also neglect within tissue intensity variation (which is a reasonable assumption if regions such as the cortical area of the adult brain are being segmented), then the intensities inside the brain can be considered to be a piecewise constant function, corrupted by Rician noise and PVE. In practice the problem is often simplified by assuming the Gaussian distribution for the noise, as it is a good approximation of the Rice distribution if the mean intensity is not close to zero. PVE causes the tissue intensity distributions to diverge from a Gaussian distribution, as can be seen from the histogram in fig. 1.3.

The situation is more complicated for neonatal images where the within-tissue intensity variability cannot be neglected due to the process of myelination. The inverted WM/GM contrast of the neonatal T1 MRI (see fig. 1.2, top row) can be seen on histogram in fig. 1.4.

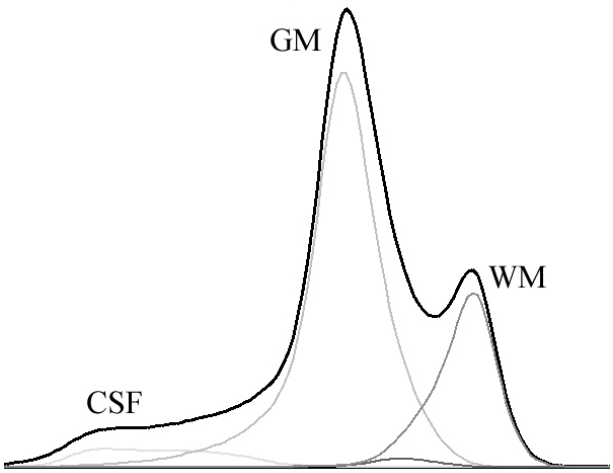


Figure 1.3: Histogram of a bias-corrected 1T T1-weighted MRI of a brain of a two year old child. Histograms of the tissue classes are based on manual segmentation. GM and CSF distributions slightly differ from Gaussian due to partial volume effect and presence of other tissues, such as vessels, in CSF.

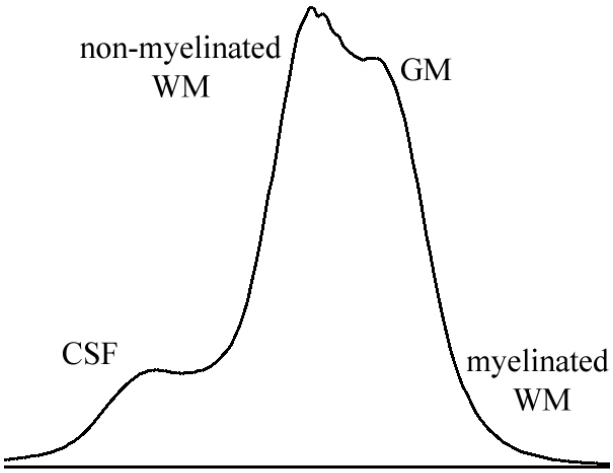


Figure 1.4: Histogram of 1T T1-weighted MRI of a neonatal brain with inverted WM/GM contrast due to large proportion of non-myelinated WM

1.3.2 Intensity inhomogeneity

The degree of inhomogeneity of the magnetic field during scanning strongly depends on strength of the magnet used, see fig. 1.5. Data scanned at 0.5T exhibit virtually no bias field and consequently, this artefact can be simply neglected. These days MRI are usually acquired on scanners with a magnetic field strength between 1 and 3T producing an inhomogeneity strong enough to cause problems for intensity-based segmentation algorithms.

Currently, the state-of-the-art magnetic field strength for new MR scanners is 3T, as the higher magnetic field allows acquiring higher resolution images with better contrast in shorter scanning time. Therefore developing good quality bias correction techniques becomes even more important than before.

1.4 Contribution of the thesis

The main goal of the work presented in this thesis is to provide a robust segmentation tool for measuring volumes of brain structures in prematurely born children, so that underlying changes in anatomy resulting from impaired brain development can be identified and eventually treated. To achieve this, robust segmentation techniques during different stages of brain development are required.

Brain MRI segmentation techniques have been intensively researched in last decade (see the following chapter for more details) and the results of this extensive work form the basis for development of new tools. However, most of this research has been focused on adult brains. Recently, some techniques developed specially for neonatal segmentation have been proposed (e.g. [69], [100]). The work presented in this thesis is focused on the segmentation of the brain during early childhood, which differs from both previously mentioned categories. Possible adaptations of current state-of-the-art methods for this problem will therefore be investigated alongside with proposing novel solutions. An important aspect of this thesis is the segmentation of subcortical GM structures, as it has been shown that changes occur in this area as a result of premature birth [13].

The contributions in this thesis can be summarized as follows:

1. Population specific atlases are develped to improve the performance of existing tools for segmentation of adult brains on images of brain during early childhood.

2. A novel robust method combining registration-based and intensity-based segmentation approaches is developed. The method outperforms both state-of-the-art approaches.

3. A novel algorithm is proposed to correct MRI with strong bias field.

4. The novel segmentation and bias correction methods are applied to segment the brain structures in children at the age of 1 to 2 years. These automatic segmentations provide the basis for the quantification of volumes of the brain structures and assessment of their growth.

1.5 Outline of the thesis

The remainder of this thesis is organized as follows: The second chapter provides a general review of segmentation techniques. Intensity-based and registration-based segmentation approaches for brain MRI are described in detail. The third chapter reviews advanced bias correction techniques. Chapter 4 compares the performance of intensity-based and registration-based approaches for the task of brain MRI segmentation during early childhood. The influence of prior information on the resulting segmentation is investigated and a method for creating prior probabilistic atlases for brains during early childhood is developed. A novel robust segmentation method combining intensity- and registration-based approaches is proposed and shown to outperform both of the investigated state-of-the-art methods. The proposed method is then used to calculate volume and growth of brain structures from one to two years of age. Chapter 5 presents a novel template-based bias correction method specially developed for correction of strong intensity inhomogeneity. In chapter 6 we will apply the novel bias correction method to MRI scanned at the field strength of 3T. The new method enables the segmentation approaches described in chapter 4 to produce excellent segmentation results on these challenging datasets. The chapter 7 summarizes the contributions and results presented in this thesis and discusses the directions for the future work.

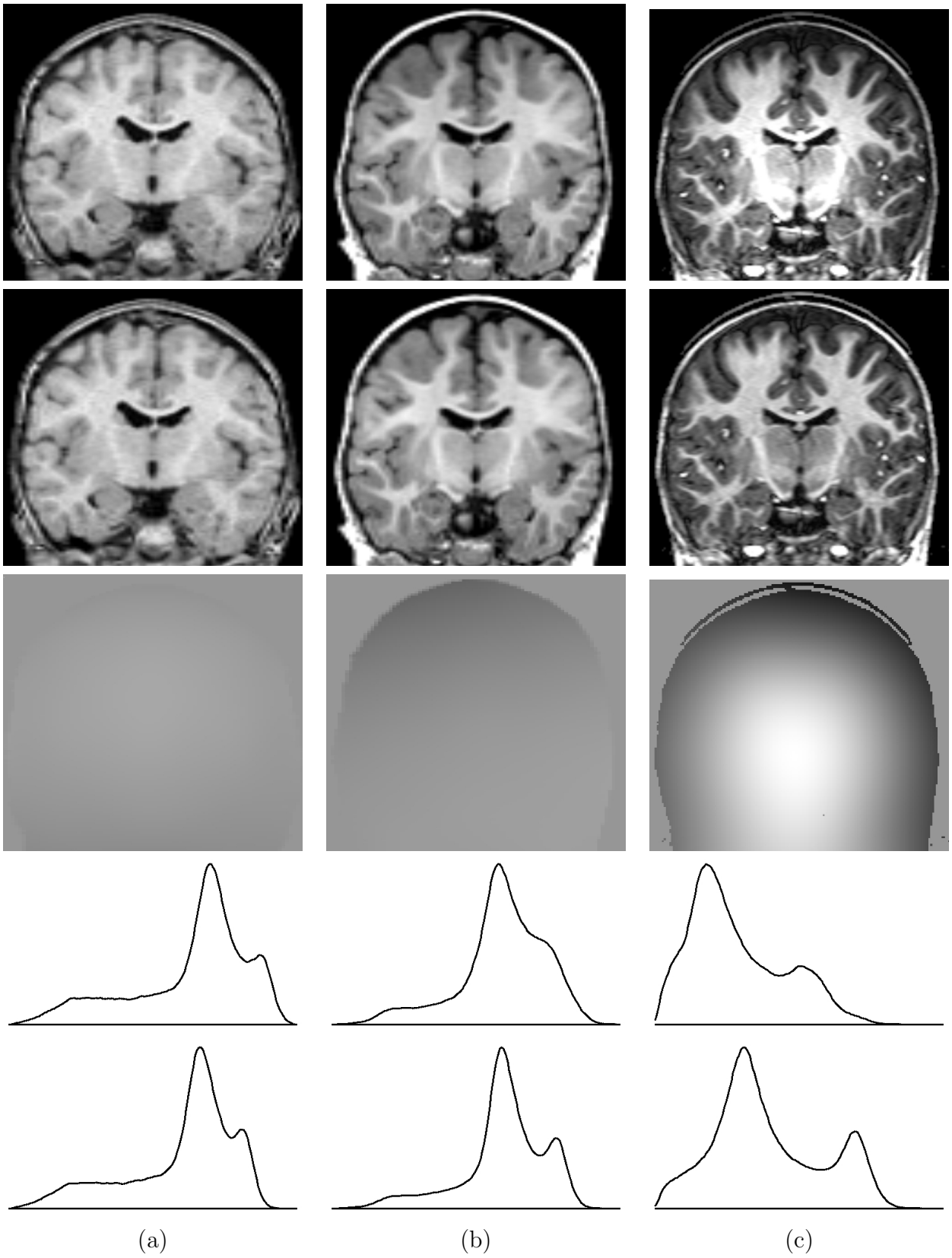


Figure 1.5: T1-weighted MRI scanned at different magnetic field strengths: (a) 0.5T with virtually no bias; (b) 1T with slightly visible bias; (c) 3T with strong bias. First row: original image; second row: corrected image; third row: bias field; fourth row: histogram of the original image; fifth row: histogram of the corrected image.

Chapter 2

Segmentation methods for brain MRI

2.1 Analysis of brain MRI

In the previous chapter we described the challenges for the automatic quantification of brain structures in MRI of the developing brain. In this chapter we will describe registration-based and intensity-based methods for the segmentation of brain structures and tissues which serve as a basis for calculating features such as their volumes and/or shapes.

In *registration-based methods*, often referred to as *atlas-based segmentation methods*, a deformable atlas is elastically warped to an image and the tissue or anatomical labels are then transferred. The main advantage of these methods is the possibility to propagate any brain structure available in the atlas without any additional cost. The performance of the registration-based methods is directly dependent on quality of the registration used.

In contrast, *intensity-based methods* classify intensities of individual voxels usually only into WM, GM and CSF. A more detailed classification is not possible, unless spatial priors are included in classification process, as the intensity profiles of different brain structures overlap. Even separation of the three main tissue classes based on intensity itself requires incorporating tools for dealing with artefacts in MRI, such as noise, intensity inhomogeneity, partial volume, as well as overlap in intensities of brain and non-brain tissue and natural within tissue intensity

variation.

Brain atlases play crucial role in developing reliable segmentation tools. A *deformable atlas* is a labelled anatomical template of the brain, which can be non-rigidly aligned with the image to be segmented [17]. The intensity-based segmentation methods can be enhanced with prior knowledge of brain anatomy in form of a *probabilistic atlas* consisting of a template image and a set of probability maps for each tissue class. Each probability map contains prior probabilities of a tissue class occurring at each location. Probabilistic atlases are usually created by averaging large number of aligned manual segmentations [34], [58]. Probabilistic atlases can be affinely or non-rigidly aligned with the image to be segmented.

In addition to registration-based and intensity-based methods, there are a number of alternative brain segmentation approaches. These approaches include surface-based methods, such as level sets and active contour/deformable models [39], [101], active shape models [23] and active appearance models [22]. Active shape models [23] search to match a shape model defined by landmarks and constrained by learning the modes of variation to the features such as edges identified in the image. Active appearance models [22] seek to find the best match for the model of the texture and shape of the object in the image. Both of these approaches can be combined to improve the robustness of the resulting segmentation [63]. These methods require prior training of the model as well as landmark placing and finding correspondences, which is a difficult problem. Moreover, the models are used mainly for the segmentation of structures with simple topology. Closer description of these methods is out of scope of this thesis.

2.2 Registration-based segmentation

Image registration involves finding the deformation or transformation of images so that the corresponding features are brought into spatial alignment. If a segmentation is available for one of those images, the registration process can serve as segmentation process, transferring the segmentation from one image to another. The original segmentation is then referred to as *deformable atlas* and is usually obtained using one or more manual segmentations.

The alignment of the images is usually initialized using global linear registration, which can be rigid or affine [38]. A rigid transformation is a 6-parameter transformation composed of translation and rotation. If scaling and skewing is allowed, we obtain a 12-parameter affine transformation. A rigid registration is sufficient for the alignment of images of the same subject (*intrasubject registration*) if the object of interest does not deform. This is a reasonable assumption for images of the brain if these are acquired at the same stage of brain development. However, if the task is to match images belonging to either different subjects (*intersubject registration*) or the same subject at different stages of brain development (e.g. growth in children, changes related to ageing, or atrophy due to disease), a non-rigid registration of the images is required to obtain satisfactory results.

2.2.1 Image registration

The registration problem seeks to find an alignment of a *source* image I_S and a *target* image I_T to maximize their similarity. An image I is a function $I : R^3 \rightarrow R$, which defines intensity values for each location in three dimensional Euclidean space. The alignment of the images can then be defined as a transformation $\mathcal{T}_{I_T \rightarrow I_S} : R^3 \rightarrow R^3$, $u = \mathcal{T}_{I_T \rightarrow I_S}(v)$ (or shorter notation $\mathcal{T}$) which maps a point v in target image I_T to its corresponding point u in source image I_S. In mathematical notation, the aim is to find a transformation $\mathcal{T}$, which maximizes a *similarity measure* C_{sim}:

$$\hat{\mathcal{T}} = \arg\max_{\mathcal{T}} C_{sim}(I_T, I_S \circ \mathcal{T})$$

where $I_S \circ \mathcal{T}$ is defined by $I_S \circ \mathcal{T}(v) = I_S(\mathcal{T}(v))$.

Registration methods are not constrained to matching voxel intensities. Methods exist to match surfaces and points (landmarks) in the images. However, surface-based and point-based registration methods are out of scope of this thesis.

2.2.2 Similarity measures

There are different similarity measures available. The most simple way of defining such a measure is to use *sum of square differences*. Let us denote $(t_1, ..., t_n)$, $t_i = I_T(v_i)$ the n voxels of the target image I_T, and $(s_1, ..., s_n)$, $s_i = I_S(\mathcal{T}(v_i))$ the n voxels of the aligned source image $I_S \circ \mathcal{T}$. Then the sum of squared differences similarity measure can be expressed as

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

This measure assumes that the images have the same intensity distributions and is therefore suitable only when the same image acquisition protocol has been used.

If this assumption is relaxed to deal with linear changes of the intensities in the two images, the *cross-correlation* can be used as a similarity measure:

$$C_{CC} = \frac{\sum_{i=1}^{n}(t_i - \bar{t})(s_i - \bar{s})}{\sqrt{\sum_{i=1}^{n}(t_i - \bar{t})^2 \sum_{i=1}^{n}(s_i - \bar{s})^2}}$$

where $\bar{t}$ denotes the mean voxels intensity in the source and target images.

When different image acquisition parameters or even different image modalities are used to generate the target and source image, the intensity relationship is generally not linear. In these cases it is possible to use an information-theoretical approach: The *entropy* H of the image I_T can be calculated

$$H(I_T) = -\sum_{t \in \mathcal{I}} P(t) \log P(t)$$

where $\mathcal{I}$ is the set of intensity values occurring in image I_T. The probability $P(t)$ denotes the probability of observing intensity t in image I_T and can be estimated from the histogram of the image. Other methods for estimating this probability distribution include techniques such as Parzen windowing [95]. Similarly, the *joint entropy* H of images I_T and I_S can be calculated as

$$H(I_T, I_S) = -\sum_{(t,s) \in \mathcal{I}'} P(t, s) \log P(t, s).$$

Here $P(t, s)$ is the joint probability of observing intensity t in the target image and intensity s at the corresponding location in the source image. Again, this joint probability can be estimated from the joint histogram of images I_T and I_S.

Using these definitions, a similarity measure called *mutual information*

$$C_{MI} = H(I_T) + H(I_S) - H(I_T, I_S)$$

can be defined [95], [55]. A modified version of this similarity measure which is more robust to variations in image overlap and field of view is *normalized mutual information* [85]:

$$C_{NMI}(I_T, I_R) = \frac{H(I_T) + H(I_R)}{H(I_T, I_R)}$$

2.2.3 Rigid and affine registration

In the most simple case, the transformation $\mathcal{T}$ belongs to the space of affine transformations defined as follows:

$$\begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{14} \\ a_{31} & a_{32} & a_{33} & a_{14} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ 1 \end{pmatrix}$$

where $u = (u_1, u_2, u_3)$ and $v = (v_1, v_2, v_3)$. The matrix representing the affine transformation can be decomposed to

$$\begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & s_3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & k_1 & k_2 & 0 \\ 0 & 1 & k_3 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where $\mathbf{R} = (r_{ij})_{ij}$ is an orthogonal rotation matrix, which can be represented by three Euler angles, $(t_1, t_2, t_3)^T$ is a translation vector, s_i are the scaling parameters and k_i are the skew-

ing parameters. An affine transformation has 12 degrees of freedom (DOF) corresponding to translation, rotation, scaling and skewing parameters, while rigid transformation has 6 DOF consisting rotation and translation parameters, while scaling and skewing is not allowed, resulting in $s_1 = s_2 = s_3 = 1$ and $k_1 = k_2 = k_3 = 0$.

The aim of the registration algorithm is to find these 12 or 6 parameters, depending on type of registration required. To do this, optimization methods, such as gradient descent or Powell's method can be used [70]. Difficulties can arise because these optimization methods find only local maxima. To improve chances of finding global maxima, a multi-resolution framework is often employed. The optimal parameters are first found for a sub-sampled image and refined for higher resolution at the next level. One such algorithm based on normalized mutual information and gradient descent optimization was developed by Studholme *et al.* [85]. Another example of an affine registration method is FLIRT [46], [45]. This robust method is based on correlation ratio similarity measure and Powell's optimization.

2.2.4 Non-rigid registration

The non-rigid registration algorithms are typically based either on physical models for transformation (such as elastic [76] or fluid deformation models [18], [26], [27]), or a linear combination of smooth basis functions [6] or free-form deformations [73]. A general review of registration techniques can be found in [103], [24].

One such technique, which uses a combination of discrete cosine transforms, was developed by Ashburner *et al.* [6] to align brain images of different subject to a space defined by a template image. The registration based on this deformation model is rather coarse and therefore not suitable when precise alignment of images is required, such as when registration-based segmentation is being performed. This algorithm is relatively quick and inexpensive and is used as a preprocessing step for different brain image analysis procedures, such as intensity-based segmentation [4], [7]. The similarity metric used in this algorithm is C_{SSD} and the transformations parameters are found using a Gauss-Newton optimization. The algorithm works only on assumption that the images have the same intensity distribution. The algorithm

also incorporates a prior constraint on the deformation in form of "membrane energy". This registration method is implemented in software package SPM99 and SPM2 [80].

2.2.5 B-spline non-rigid registration

An example of a free-form deformation based non-rigid registration is the B-spline registration algorithm by Rueckert *et al.* [73].

Let us denote $\mathcal{T}_{\theta,\phi}$ the non-rigid transformation modelled as a sum of a global affine transformation $\mathcal{A}_\theta$ given by 12 parameters θ and a local non-linear transformation $\mathcal{B}_\phi$ defined by control points ϕ:

$$\mathcal{T}_{\theta,\phi}(v) = \mathcal{A}_\theta(v) + \mathcal{B}_\phi(v)$$

The local component $\mathcal{B}_\phi$ is represented using uniform tensor-product 3D cubic B-splines, defined as

$$\mathcal{B}_\phi(v) = \sum_{p=0}^{3}\sum_{q=0}^{3}\sum_{r=0}^{3} B_p(t_1)B_q(t_2)B_r(t_3)\Phi_{i_1+p,i_2+q,i_3+r}$$

where $\Phi_{i,j,k}$ are B-spline control points defined on regular $L_1 \times L_2 \times L_3$ lattice with spacing of $\delta_1, \delta_2, \delta_3$ along each dimension and $B_0, ..., B_3$ are cubic B-spline basis functions:

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

Parameters t_1, t_2, t_3 are defined as $t_1 = v_1/\delta_1 - i_1, t_2 = v_2/\delta_2 - i_2, t_3 = v_3/\delta_3 - i_3$ where $v = (v_1, v_2, v_3)$ and indexes $i_1 = \lfloor v_1/\delta_1 \rfloor, i_2 = \lfloor v_2/\delta_2 \rfloor, i_3 = \lfloor v_3/\delta_3 \rfloor$. To constrain the local transformation to be smooth, the following penalty term can be introduced:

$$C_{smooth}(\mathcal{T}) = \frac{1}{V} \int_0^{X_1} \int_0^{X_2} \int_0^{X_3} \left[\left(\frac{\partial^2 \mathcal{T}}{\partial v_1^2} \right)^2 + \left(\frac{\partial^2 \mathcal{T}}{\partial v_2^2} \right)^2 + \left(\frac{\partial^2 \mathcal{T}}{\partial v_3^2} \right)^2 \right.$$
$$\left. + \left(\frac{\partial^2 \mathcal{T}}{\partial v_1 v_2} \right)^2 + \left(\frac{\partial^2 \mathcal{T}}{\partial v_2 v_3} \right)^2 + \left(\frac{\partial^2 \mathcal{T}}{\partial v_1 v_3} \right)^2 \right] du_1\, du_2\, du_3 \tag{2.1}$$

where V denotes the volume of the image domain $\{(v_1, v_2, v_3) | 0 \leq v_1 < X_1, 0 \leq v_2 < X_2, 0 \leq v_3 < X_3\}$ and $X_i = \delta_i(L_i - 1)$.

The similarity measure used for this non-rigid registration is normalized mutual information C_{NMI} [85]. The overall cost function C consists of the similarity measure and the regularization term and can be expressed as follows:

$$C(\theta, \phi) = -C_{sim}(I_T, I_S \circ \mathcal{T}_{\theta,\phi}) + \lambda C_{smooth}(\mathcal{B}_\phi)$$

Note that the regularization term is always zero for affine transformation and therefore $C_{smooth}(\mathcal{T}_{\theta,\phi}) = C_{smooth}(\mathcal{B}_\phi)$. The weighting parameter λ provides the tradeoff between the image alignment and smoothness of the transformation.

When minimizing the cost function C, the algorithm first proceeds by optimizing the global affine transformation before the local B-spline transformation is optimized. The optimization method used is based on gradient descent. The registration is performed in multiresolution framework. The resolution of the B-spline transformation is refined by halving the spacing between the B-spline control points and consequently inserting new B-spline control points. The algorithm can be summarized as follows:

Non-rigid registration by Rueckert *et al.*

Calculate optimal affine transformation parameters $\hat{\theta}$ by maximizing normalized mutual infor-

mation:

$$\hat{\theta} = \arg\max_{\theta} C_{NMI}(I_T, I_S \circ \mathcal{A}_\theta)$$

Initialize B-spline control points ϕ

For each resolution level r

Calculate the gradient vector of cost function C with respect to ϕ at current estimate $\phi^{(r,0)}$

$$\nabla C = \left.\frac{\partial C(\hat{\theta}, \phi)}{\partial \phi}\right|_{\phi=\phi^{(r,0)}}$$

while $\|\nabla C\| > \varepsilon$ **do**

Recalculate the control points

$$\phi^{(r,m+1)} = \phi^{(r,m)} + \delta \frac{\nabla C}{\|\nabla C\|}$$

Recalculate the gradient ∇C

end while

Increase the control point resolution by subdividing from $\phi^{(r,m+1)}$ to $\phi^{(r+1,0)}$

Increase image resolution

end for

An example of a non-rigid registration of brain MRI using this algorithm is shown in fig. 2.1c,f.

2.2.6 B-spline registration-based segmentation

In this thesis we use the B-spline based registration by Rueckert *et al.* to perform the registration-based segmentation. This method has been shown to be especially suitable for registration-based segmentation because of its flexibility and accuracy compared to other registration methods [25]. Let S_X be the segmentation of a template image X. A new image Y can be segmented

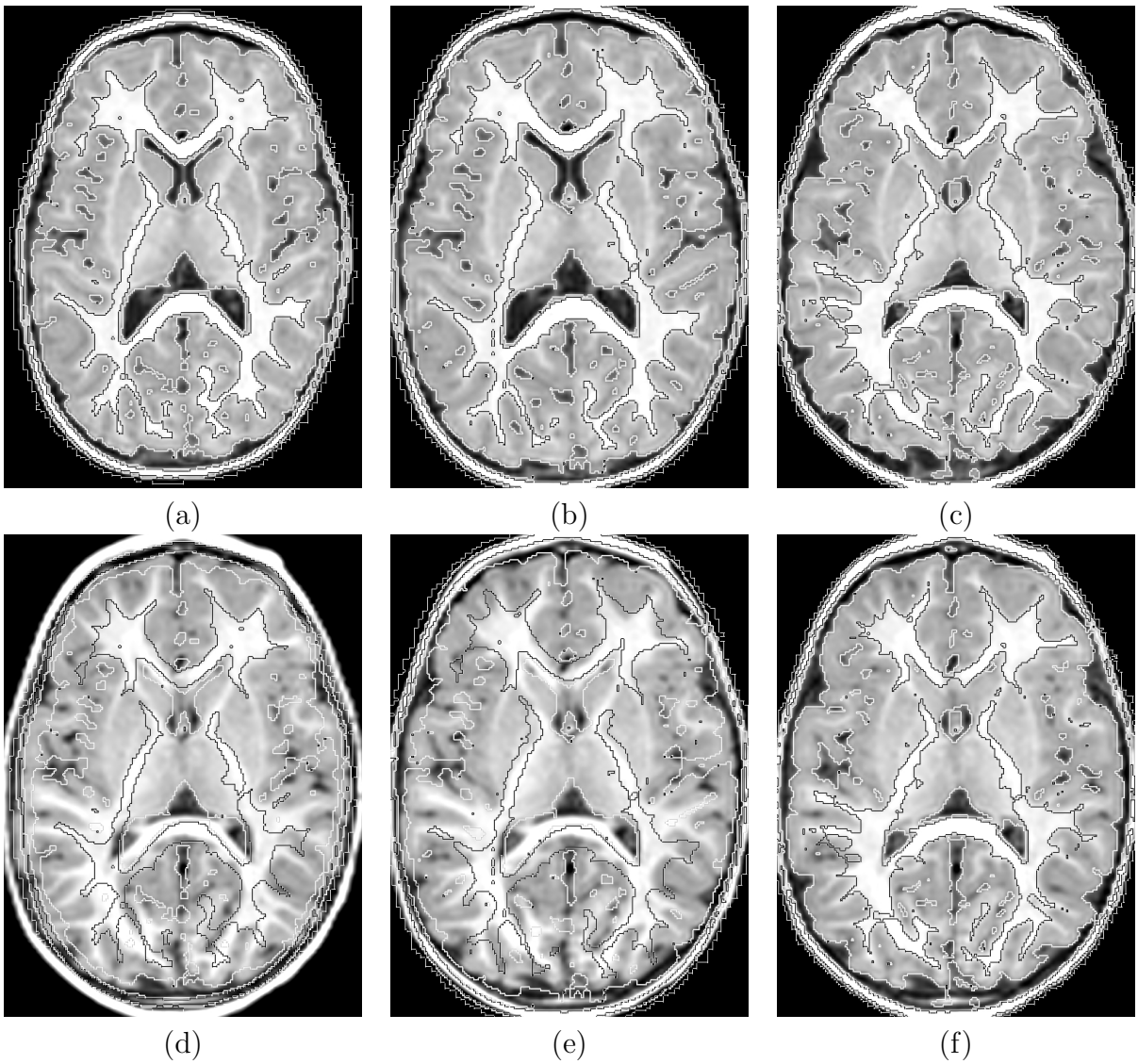


Figure 2.1: Comparison of different types of registrations. Top row: the source image with iso-contours registered to target image with (a) rigid registration; (b) affine registration; (c) non-rigid registration [73]. Bottom row: The target image with overlayed iso-contours of source image registered with (d) rigid registration; (e) affine registration; (f) non-rigid registration [73].

by transferring the segmentation from the template image using non-rigid registration. Let $S_{rbs}(Y)$ denote registration-based segmentation of Y. Then

$$S_{rbs}(Y) = S_X \circ \mathcal{T}_{Y \to X} \tag{2.2}$$

The robustness of registration-based segmentation can be improved if a number of segmented images or atlases are used. This was shown by Heckemann *et al.* [40] where manual segmentation from 29 images $X_1, .., X_{29}$ were transferred by non-rigid registration [73] to target image

Y and the vote rule for decision fusion was used to obtain the final segmentation:

$$S_{rbs-multi}(Y(v)) = \arg\max_{l} \|\{i|S_{X_i}(\mathcal{T}_{Y\rightarrow X_i}(v)) = l\}\| \tag{2.3}$$

where $\|.\|$ denotes the number of elements in a set. With this multiple label propagation, Gousias *et al.* [37] segmented 33 brain MRI of two year old children into 83 regions interests, using 30 manual segmentations of adult brains. Validation results show very high similarity index (SI) despite the differences in the shape and size between adult and child brains. However, this method is dependent on having a number of high-quality atlases, which might not always be available in practice.

2.3 Intensity-based methods

Intensity-based methods are used to classify individual voxels in brain MRI based on their intensity. Three main tissue classes, WM, GM and CSF, can be distinguished based on intensity. The intensity based classification methods make use of standard classifiers, such as a Gaussian Mixture Model (GMM), C-means or k-NN [32]. These methods will be described in more detail in the following sections. However, in order to obtain satisfactory results using these techniques, the following procedures are usually carried out:

Brain extraction: Non-brain tissues such as fat, skull, or neck have intensities overlapping with intensities of brain tissues. Therefore, the brain has to be extracted before intensity-based methods can be used.

Bias correction: The bias field, which causes intensity inhomogeneity, has to be calculated and image intensities have to be corrected to achieve homogeneous intensities within tissue classes.

In addition, several constraints are usually integrated into intensity-based segmentation algorithms:

Prior information: Misclassifications due to overlaps in tissue intensity distributions can be reduced by incorporating prior knowledge of the brain anatomy. Spatial prior information can also reduce the impact of noise.

Neighbourhood information: Incorporating neighbourhood information helps to give preference to spatially homogenous regions in the resulting segmentation and therefore reduces the impact of noise and misclassifications due to PVE.

In the remainder of this section we describe methods based on C-means clustering and k-NN clustering. The methods based on the Expectation-Maximization framework (GMM and its extensions) will be presented in detail in the next section, as they form the basis for our work. Subsequent sections of this chapter describe incorporating prior information, neighbourhood information and brain extraction into intensity-based methods. The bias correction step will not be addressed in this chapter, as it will be described in detail in chapter 3.

In the following sections Y will denote the observed image consisting of n voxels, with intensities $Y = (y_1, ..., y_n)$. Our aim is to classify voxels into K clusters, representing the main tissue classes, based on their intensity.

2.3.1 C-means clustering

Let v_k be a centroid of cluster k and u_{ik} the membership, or a degree of belonging of location i to class k, with $\sum_{k=1}^{K} u_{ik} = 1$. Fuzzy C-means clustering [33] iteratively optimizes the objective function

$$F = \sum_{i=1}^{n} \sum_{k=1}^{K} (u_{ik})^q \|y_i - v_k\|^2$$

using the following algorithm:

Fuzzy C-means clustering

Initialize centroids v_k^0

Iterate between following two steps:

1. Calculate class membership for each voxel:

$$u_{ik}^{(m)} = \frac{\|y_i - v_k^{(m-1)}\|^{-\frac{2}{q-1}}}{\sum_{k=1}^{K} \|y_i - v_k^{(m-1)}\|^{-\frac{2}{q-1}}}$$

2. Recalculate centroids:

$$v_k^{(m)} = \frac{\sum_{i=1}^{n} (u_{jk}^{(m)})^q y_i}{\sum_{i=1}^{n} (u_{jk}^{(m)})^q}$$

until

$$\max_{ik} \|u_{ik}^{(m)} - u_{ik}^{(m-1)}\| < \varepsilon$$

In the previous algorithm m denotes the number of iterations and $\|.\|$ the Euclidian distance. The parameter q defines the degree of fuzziness. For $q = 1$ algorithm reduces to K-means clustering [54]. The value most often used in practice is $q = 2$. In this case the membership value is proportional to inverse of distance from the class centroid.

The use of C-means clustering for brain segmentation was first proposed by Pham and Prince [65]. The clustering is initialized by finding the three centroids from the image histogram. The probability density function (PDF) of the image intensities is non-parametrically estimated using a Gaussian kernel. The width of the kernel is the smallest value for which exactly three local maxima of the estimated PDF are found. The advantage of this method is that it is fully automatic and fast. It also corrects the bias field during the iterative segmentation process. However, it does not provide any mechanism for incorporating prior information into the segmentation process. The brain has to be extracted in pre-processing step in order to produce good results.

2.3.2 kNN classification

The k-Nearest Neighbour (kNN) classification rule is a technique for non-parametric supervised pattern classification. The classifier is *supervised*, which means that it requires a set of patterns with known classification for training. Unlike C-means classification, it is especially suitable if

a large number of training data is available.

Given a training data set P consisting of M prototype patterns - vectors of dimension D, and corresponding correct classification of each prototype into one of the K classes, the unknown pattern vector $\mathbf{v}$ is classified as class j if the largest number of the k closest prototypes are from class j. The D-dimensional pattern vectors are called *features*, the D-dimensional feature domain is called the *feature space* and the constant k can be chosen to suit the particular problem. The distance measure d has to be defined in the feature space to be able to calculate the distance between patterns and choose the k nearest prototype patterns for each new pattern to be classified.

The most widely used distance measure is the Euclidian distance measure

$$(d(\mathbf{v}, \mathbf{p}))^2 = \sum_{f=1}^{D} (v_f - p_f)^2$$

where d is the distance, v_f is the value of the pattern vector $\mathbf{v}$ in feature f, p_f is the value of the prototype vector $\mathbf{p}$ in feature f.

Let $l_{\mathbf{p}}$ denote the label assigned to each prototype $\mathbf{p} \in P$ and $l_1, ..., l_n$ the labelling of image Y. Let P_i denote the set of k closest prototypes to a voxel at location i and feature vector $\mathbf{v}_i$. Then the kNN classification of the image is defined as follows:

kNN classification

1. For each location i find set P_i of k nearest prototypes to voxel with feature vector $\mathbf{v}_i$:

$$P_i \subset P, \|P_i\| = k, \forall \mathbf{p} \in P : (d(\mathbf{p}, \mathbf{v}_i) < \max_{\mathbf{p}' \in P_i} d(\mathbf{p}', \mathbf{v}_i)) \Rightarrow (\mathbf{p} \in P_i)$$

2. Calculate the labelling of image Y

$$l_i = \arg\max_{l} \|\{l_{\mathbf{p}} = l | \mathbf{p} \in P_i\}\|$$

The features for the kNN classifier should be chosen so that they provide the best possible contrast between the classes. For brain MRI the most obvious features are the image intensities y_i for each voxel.

The k-NN classification approach was adapted for brain MRI segmentation by Warfield *et al.* [93]. In addition to image intensities, Warfield used spatial localization of brain structures (classes) in form of a non-rigidly registered template as an additional feature to enhance the classification process. The segmentation is then calculated in an iterative process by interleaving the segmentation refinement with updating the non-rigid alignment to the template. This procedure requires manual selection of large number of training samples for each tissue class to train the kNN classifier. The method is therefore not fully automatic and results depend on particular choice of the training set.

kNN classification by Warfield *et al.*

---

1. For each tissue class j manually select the prototypes to train the classifier to recognize tissue intensity PDF

2. Align deformable atlas with the image

3. **Repeat** for a fixed number of iterations

   a. Identify intensity and spatial features for each voxel using current alignment of the atlas

   b. Update the labelling using kNN classification

   c. Update the non-rigid alignment of deformable template with current labelling of the image

---

Cocosco *et al.* [19] developed a method for the robust selection of training samples to make the kNN classification process fully automatic. In his method, the feature vector consist only of voxel intensities of multichannel MRI. Firstly, a probabilistic atlas (see fig. 2.3) is affinely aligned with the image and samples with high prior probabilities are chosen for each class. A minimum spanning tree (MST) is then constructed in feature space. The training samples are then selected by pruning procedure: long edges in the MST in feature space are removed until there are four distinct clusters; samples with initial labelling consistent with the result of pruning are used as a training set. The final segmentation is then obtained using the kNN classification. This method is reported to deal well with anatomies which differ from the probabilistic atlas. However, it does not deal with the problem of natural intensity variation within each tissue class. Both methods require the bias to be corrected in a pre-processing step.

2.3.3 Entropy minimisation

Another approach, proposed by Tasdizen *et al.* [88], is to minimize the entropy associated with a set of K tissue PDFs. This results in an iterative segmentation process which interleaves the non-parametric estimation of PDF for intensity of each tissue class with estimation of a hard segmentation. In the segmentation step each voxel is labelled with the class with highest probability. The method is made more robust by estimating a 7-dimensional PDF of each tissue class for a 7-neighbourhood rather single voxel intensities.

Let N_i denote the random variable for the 7-neigbourhood of the voxel i and $\mathbf{n}_i$ the vector of voxel intensities in the neighbourhood N_i. Let $P(N_i = \mathbf{n}_i | Z_i = k)$ be the PDF of observing the neighbourhood N_i given that the center voxel i has been assigned tissue label k. Then the entropy H to be minimized is defined as:

$$H = -\sum_{k=1}^{K} \int_{\mathbb{R}^m} P(N = \mathbf{n}|Z = k) \log P(N = \mathbf{n}|Z = k) d\mathbf{n}$$

where integration is performed over a domain $\mathbb{R}^m$ of the random vector $\mathbf{n}$. The minimization can be achieved by the following algorithm:

Entropy minimization by Tasdizen *et al*

Initialize segmentation $(z_1, ...z_n)$ by creating the hard segmentation from affinely aligned probabilistic atlas.

Repeat

1. Estimate the PDF for each tissue using Parzen windowing and the current segmentation estimate:

$$P^{(m)}(N_i = \mathbf{n}_i | Z_i = k) = \frac{1}{|A_{ik}^{(m-1)}|} \sum_{j \in A_{ik}^{(m-1)}} G(\mathbf{n}, \mathbf{n}_i, \sigma)$$

   where $A_{ik}^{(m-1)}$ is a randomly chosen set of voxels assigned to tissue k at iteration $(m-1)$ and $G(\mathbf{n}, \mathbf{n}_i, \sigma)$ is 7-dimensional Gaussian distribution with mean $\mathbf{n}_i$ and variance σ.

2. Update segmentation by assigning each voxel i to the class with highest likelihood

$$z_i^{(m)} = \operatorname*{argmax}_k P^{(m)}(N_i = \mathbf{n}_i | Z_i = k)$$

until

$$\frac{H^{(m-1)} - H^{(m)}}{H^{(m)}} < \varepsilon$$

$$H^{(m)} = -\sum_{k=1}^{K} \frac{1}{T_k^{(m)}} \sum_{\forall i : z_i^{(m)} = k} \log P^{(m)}(N = \mathbf{n} | Z = k)$$

where $T_k^{(m)}$ denotes the number of voxels classified as belonging to tissue k and $H^{(m)}$ is an approximation of entropy H.

The obvious advantage of this method is robustness to noise and no bias towards a parametric model of the intensity distribution. A bias field estimation can be also interleaved with the segmentation process. The method can be used to estimate spatially variable non-parametric tissue PDFs, by choosing the random samples $A_{ik}^{(m-1)}$ from a Gaussian distribution centered at location i. For a spatially constant PDF a global sample set $A_k^{(m-1)}$ can be formed. The main disadvantage is the slow computational speed.

2.4 EM-based segmentation

One of the most successful and widely used brain segmentation approaches was first proposed over a decade ago when Wells *et al.* [96] used Bayesian model to estimate segmentation and bias field through the EM algorithm [29]. Different variations of EM-based segmentation methods have been successfully implemented in several software packages used in the medical imaging community (EMS:[48], [49]; SPM: [4], [7]; FAST: [102]; FreeSurfer: [36]; 3DSlicer: [96], [67];). All these methods implement a segmentation and bias correction in the EM framework. They also include various additional improvements, such as non-rigid alignment of atlas [7], including neighbourhood information in form of Markov Random Fields (MRF) [102], [49] and partial volume estimation [47], [50]. The general framework for EM-based segmentation can be summarized as follows:

EM approach for brain segmentation

E-step:

Estimate the soft segmentation given the current estimate of model parameters. This may include the use of neighbourhood statistics in form of a MRF as well as partial volume estimation.

M-step:

Estimate the model parameters which can consist of a combination of the following steps:

1. Estimate the intensity distribution parameters for each tissue class.

2. Estimate the bias correction parameters.

3. Estimate the registration parameters for alignment of probabilistic atlas with the image.

This section gives a thorough theoretical background for EM-based segmentation, as the bias correction algorithm proposed in chapter 5 directly builds on this framework.

2.4.1 Optimization criteria

Let $\mathbf{y} = (y_1, ..., y_n)$ denote the vector of intensities in the image Y consisting of n voxels. Then the image function Y can be viewed as a random variable with realization $\mathbf{y}$. Let us also assume that we have a model of the image Y, represented with parameters Φ. These parameters could include bias field and/or intensity distribution parameters.

In the *Bayesian* probabilistic framework, as proposed by Wells [96] and later adapted by Zhang [102] and Pohl [67], the *Maximum A Posteriori* (MAP) criteria is optimized to find the segmentation of brain MRI:

$$\hat{\Phi} = \arg\max_i P(\Phi|Y) \tag{2.4}$$

Maximum likelihood (ML) is an alternative probabilistic optimization criteria used by Van Leemput [48]:

$$\hat{\Phi} = \arg\max_i P(Y|\Phi) \tag{2.5}$$

The term $P(Y|\Phi)$ is called the *likelihood*. The relationship between both criteria can be seen if Bayes' theorem is used:

$$P(\Phi|Y) = \frac{P(Y|\Phi)P(\Phi)}{P(Y)}$$

The criteria is optimized with respect to parameters Φ, so $P(Y)$ acts as a constant. It has therefore no influence on equation 2.4 and the MAP criteria is equivalent to

$$\hat{\Phi} = \arg\max_i P(Y|\Phi)P(\Phi) \tag{2.6}$$

It is now clear that MAP is closely related to ML. Unlike ML, MAP contains a regularization term $P(\Phi)$, also called the *prior probability*. This term enables us to incorporate prior information for model parameters Φ.

To simplify the problem, independence of the voxels is usually assumed [96], [48], [7], [67]. The random variable Y representing the image intensities can be viewed as n-dimensional random variable $Y = (Y_1, ..., Y_n)$. Each voxel intensity y_i is a realization of random variable Y_i. The assumption that random variables $Y_1, ..., Y_n$ are independent allows us to express the likelihood

as follows:

$$P(Y|\Phi) = \prod_{i=1}^{n} P(Y_i|\Phi)$$

As we can see from this equation, calculating the likelihood in practice means multiplying a large number of non-negative numbers smaller than one and consequently the resulting value is too small for practical purposes. It is more convenient to work with *log-likelihood* because the product is transformed to a sum. The logarithmic function $\log(x)$ is increasing and continuous and therefore the solution of eq. 2.5 can be found as maximum of the log-likelihood $L(\Phi)$:

$$L(\Phi) = \log P(Y|\Phi) \tag{2.7}$$

The log-likelihood of the observed image $L(\Phi)$ is a suitable objective function for the segmentation problem and can be further expressed as

$$L(\Phi) = \log \prod_{i=1}^{n} P(Y_i|\Phi) = \sum_{i=1}^{n} \log P(Y_i|\Phi)$$

The alternative objective function $L_{MAP}(\Phi)$ based on MAP principle is suitable when constraints need to be imposed on the parameter space and can be expressed as:

$$L_{MAP}(\Phi) = \log P(Y|\Phi)P(\Phi) = L(\Phi) + \log P(\Phi) \tag{2.8}$$

2.4.2 Probabilistic framework for image segmentation

Segmenting the image Y means separating the n voxels into K classes, or in other words, assigning a label from the set $\{1, 2, ..., K\}$ to each voxel. Let $\mathbf{z} = (z_1, ..., z_n)$ be the unknown labeling of the image Y. Then $Z = (Z_1, ..., Z_n)$ is a n-dimensional random variable and each label z_i is a realization of random variable Z_i. Now we can introduce the segmentation Z into the optimization criteria using the total probability theorem [59]:

$$P(Y_i|\Phi) = \sum_{k=1}^{K} P(Y_i|Z_i = k, \Phi)P(Z_i = k|\Phi) \tag{2.9}$$

The relationship between the image Y and the unknown segmentation Z can be expressed using Bayes' theorem:

$$P(Z_i = k|Y_i, \Phi) = \frac{P(Y_i|Z_i = k, \Phi)P(Z_i = k|\Phi)}{P(Y_i|\Phi)} \tag{2.10}$$

The *posteriors* $P(Z_i = k|Y_i, \Phi)$ are an estimate of the segmentation inferred from the image Y and model parameters Φ, while the *likelihood* $P(Y_i = y_i|Z_i = k, \Phi)$ is the probability of obtaining the voxel intensity y_i from estimated segmentation and model parameters. *Priors* $P(Z_i = k|\Phi)$ can be used to incorporate knowledge obtained from training data and $P(Y_i|\Phi)$ is a normalizing constant, as given by equation 2.9.

The objective function $L(\Phi)$ can now be written as

$$L(\Phi) = \sum_{i=1}^{n} \log \sum_{k=1}^{K} P(Y_i|Z_i = k, \Phi)P(Z_i = k|\Phi) \tag{2.11}$$

2.4.3 Optimization via Expectation Maximization

The expectation maximization algorithm (EM) [29] is a general technique for finding maximum likelihood parameter estimates in problems with incomplete data and is therefore suitable for solving segmentation problems. In case of the segmentation problem, the *incomplete* or *observed data* are voxel intensities Y and unknown segmentation Z constitutes *hidden data*. The combined data (Y, Z) are called the *complete* data.

The *complete log-likelihood* is the log-likelihood from the complete data Y, Z

$$L_c(\Phi) = \log P(Y, Z|\Phi)$$

The log-likelihood $L(\Phi) = \log P(Y|\Phi)$ based on the incomplete data Y is also called *incomplete*

log-likelihood. The expectation maximization algorithm searches for a maximum likelihood estimate of the parameters Φ through maximizing the estimate of complete log-likelihood.

Definition of the EM algorithm

Iterate between expectation step (E-step) and maximization step (M-step) **until convergence**:

E-step: Calculate a function $Q(\Phi|\Phi^{(m)})$ based on current parameter estimate $\Phi^{(m)}$:

$$Q(\Phi|\Phi^{(m)}) = E_Z(L_c(\Phi)|Y, \Phi^{(m)}) = E_Z(\log P(Y, Z|\Phi)|Y, \Phi^{(m)})$$

M-step: Find parameters $\Phi^{(m+1)}$ that maximize the function $Q(\Phi|\Phi^{(m)})$

$$\Phi^{(m+1)} = \underset{\Phi}{\operatorname{argmax}}\, Q(\Phi|\Phi^{(m)})$$

The expected value of the complete log-likelihood $E_Z(\log P(Y, Z|\Phi)|Y, \Phi^{(m)})$ is the *conditional expected value* defined as follows:

$$E_A(f(A)|B) = \sum_{\forall a} P(A = a|B) f(a)$$

Here a is assumed to be a realization of a discrete random variable A and B is another random variable.

The function $Q(\Phi|\Phi^{(m)})$ represents the expected value of log-likelihood from the complete data X and defines a lower bound to the log-likelihood $L(\Phi)$ (eq. 2.7). Maximizing the lower bound leads to increasing of the log-likelihood $L(\Phi)$. The lower bound is tightened to $L(\Phi)$ at every iteration until its maximum converges to a local maximum of the log-likelihood [60]. The sufficient conditions for convergence to maximum likelihood estimate can be found in [29] and [98].

2.4.4 Segmentation via EM

The EM algorithm can be used to solve the segmentation problem. If the independence of voxel intensities and labels is assumed, the complete log-likelihood can be further expressed as:

$$L_c(\Phi) = \log P(Y, Z|\Phi) = \log P(Y|Z, \Phi) + \log P(Z|\Phi) =$$

$$= \sum_{i=1}^{n} \log P(Y_i|Z_i, \Phi) + \sum_{i=1}^{n} \log P(Z_i|\Phi) =$$

$$= \sum_{i=1}^{n} \log P(Y_i|Z_i, \Phi)P(Z_i|\Phi) = \sum_{i=1}^{n} \log P(Y_i, Z_i|\Phi)$$

To be able to calculate the expected value over Z we need to amend the definition of the hidden data. Instead of simply assigning the label k we will now consider Z_i to take values from the set of k-dimensional unit vectors $\{\mathbf{e}_1, ..., \mathbf{e}_K\}$ where $Z_i = \mathbf{e}_k = (0, ..., 0, 1, 0, ..., 0)$ means that i^{th} voxel belongs to tissue k. Let us denote $\mathbf{z}_i = (z_{i1}, ..., z_{iK})$. Then for $\mathbf{z}_i = \mathbf{e}_k$ $\log P(Y_i = y_i, Z_i = \mathbf{e}_k|\Phi)$ can be expressed as $\sum_{j=1}^{K} z_{ij} \log P(Y_i = y_i, Z_i = \mathbf{e}_j|\Phi)$ and the complete log likelihood becomes:

$$L_c(\Phi) = \sum_{i=1}^{n} \sum_{j=1}^{K} z_{ij} \log P(y_i, Z_i = \mathbf{e}_j|\Phi) = \sum_{i=1}^{n} \mathbf{z}_i^T \mathbf{V}(y_i|\Phi)$$

where vector $\mathbf{V}(y_i|\Phi) = (\log P(y_i, Z_i = \mathbf{e}_1|\Phi), ..., \log P(y_i, Z_i = \mathbf{e}_K|\Phi))$ is now constant in variable $\mathbf{z}_i$. We can now calculate the function $Q(\Phi|\Phi^{(m)})$:

$$Q(\Phi|\Phi^{(m)}) = E_Z(L_c(\Phi)|\mathbf{y}, \Phi^{(m)}) = E_Z(\sum_{i=1}^{n} \mathbf{z}_i^T \mathbf{V}(y_i|\Phi)|\mathbf{y}, \Phi^{(m)}) =$$

$$= \sum_{i=1}^{n} E(\mathbf{z}_i|\mathbf{y}, \Phi^{(m)})^T \mathbf{V}(y_i|\Phi)$$

According to the definition of the conditional expected value $E(\mathbf{z}_i|\mathbf{y}, \Phi^{(m)})$, this can be further expressed as

$$E(\mathbf{z}_i|\mathbf{y}, \Phi^{(m)}) = \sum_{j=1}^{K} P(Z_i = \mathbf{e}_j|y_i, \Phi^{(m)})\mathbf{e}_j = \sum_{j=1}^{K} p_{ij}^{(m+1)} \mathbf{e}_j$$

where $p_{ij}^{(m+1)} = P(Z_i = \mathbf{e}_j | y_i, \Phi^{(m)})$ is a soft assignment of voxel i to tissue j at $(m+1)^{st}$ iteration. Finally, the function $Q(\Phi|\Phi^{(m)})$ can be expressed as

$$Q(\Phi|\Phi^{(m)}) = \sum_{i=1}^{n}\sum_{j=1}^{K} p_{ij}^{(m+1)} \log P(y_i, Z_i = \mathbf{e}_j|\Phi) =$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{K} p_{ij}^{(m+1)} \log P(y_i|Z_i = \mathbf{e}_j, \Phi) + \sum_{i=1}^{n}\sum_{j=1}^{K} p_{ij}^{(m+1)} \log P(Z_i = \mathbf{e}_j|\Phi)$$

It is now obvious that for the segmentation problem, calculating of $Q(\Phi|\Phi^{(m)})$ in the E-step is equivalent to calculating soft assignment according to Bayes rule (eq. 2.10) and that the M-step maximizes the resulting expression for $Q(\Phi|\Phi^{(m)})$. In many cases the M-step can be performed by direct partial differentiation resulting in a system of linear equations. The EM algorithm for image segmentation can be summarized as follows:

Segmentation via EM

E-step:

Calculate the probability maps (soft segmentation) $\mathbf{p}_j^{(m+1)} = \{p_{ij}, i = 1, ..., n\}$, $j = 1, ..., K$ given the observed voxel intensities $\mathbf{y}$ and parameter estimate $\Phi^{(m)}$:

$$p_{ij}^{(m+1)} = \frac{P(y_i|Z_i = \mathbf{e}_j, \Phi^{(m)})P(Z_i = \mathbf{e}_j|\Phi^{(m)})}{\sum_{k=1}^{K} P(y_i|Z_i = \mathbf{e}_k, \Phi^{(m)})P(Z_i = \mathbf{e}_k|\Phi^{(m)})}$$

M-step:

Estimate the parameters $\Phi^{(m+1)}$ based on probability maps $\mathbf{p}_j^{(m+1)}$ and the observed voxel intensities $\mathbf{y}$

$$\Phi^{(m+1)} = \underset{\Phi}{\operatorname{argmax}}\, Q(\Phi, \Phi^{(m)})$$

$$Q(\Phi, \Phi^{(m)}) = \sum_{i=1}^{n}\sum_{j=1}^{K} p_{ij}^{(m+1)} (\log p(y_i|Z_i = \mathbf{e}_j, \Phi) + \log p(Z_i = \mathbf{e}_j|\Phi))$$

2.4.5 Mixture of Gaussians as a model for MR image intensities

The classical model for the segmentation of brain MRI assumes that the intensity distribution of each tissue class is Gaussian. The M-step of the EM algorithm simplifies further if a Gaussian distribution is assumed for the likelihood of observing intensity y_i for a given tissue class k:

$$P(y_i|Z_i = \mathbf{e}_k, \Phi) = G(y_i, \mu_k, \sigma_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \mathrm{e}^{\frac{(y-\mu_k)^2}{2\sigma_k^2}}$$

The likelihood of observing intensity y_i can then be expressed using the total probability theorem:

$$P(y_i|\Phi) = \sum_{k=1}^{K} G(y_i, \mu_k, \sigma_k) P(Z_i = \mathbf{e}_k|\Phi)$$

This assumption leads to significant simplification of the lower bound $Q(\Phi, \Phi^{(m)})$:

$$Q(\Phi, \Phi^{(m)}) = \sum_{i=1}^{n} \sum_{j=1}^{K} p_{ij}^{(m+1)} \left(\frac{(y_i - \mu_k)^2}{2\sigma_k} - \log \sqrt{2\pi}\sigma_k + \log p(Z_i = \mathbf{e}_j|\Phi) \right)$$

If we further assume that the prior probability that the voxel i belongs to tissue class k is spatially constant

$$p(Z_i = \mathbf{e}_k|\Phi) = c_k$$

we obtain the widely used *Gaussian mixture model* (GMM). The equation for the likelihood of observing intensity y_i simplifies as follows:

$$P(y_i|\Phi) = \sum_{k=1}^{K} G(y_i, \mu_k, \sigma_k) c_k$$

The parameters Φ represent unknown means μ_k and variances σ_k of the Gaussian PDFs for the intensities of each tissue class, as well as the mixture coefficients c_k

$$\Phi = \gamma \equiv (\mu_1, \sigma_1, c_1, ..., \mu_K, \sigma_K, c_K)$$

The task of finding the parameters Φ of Gaussian mixture model can be intuitively described as fitting the Gaussians to the image histogram (fig. 2.2).

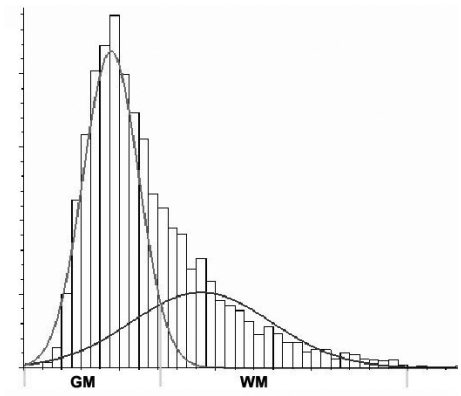


Figure 2.2: Gaussian mixture model - fitting the mixture of Gaussians to the normalized image histogram.

The intensity distribution parameters can now be found by optimizing the maximum likelihood:

$$\hat{\Phi} = \arg\max_{\Phi} \sum_{i=1}^{n} \log \sum_{k=1}^{K} G(y_i, \mu_k, \sigma_k) c_k$$

This can be done via the EM algorithm. The lower bound becomes:

$$Q(\Phi, \Phi^{(m)}) = \sum_{i=1}^{n} \sum_{j=1}^{K} p_{ij}^{(m+1)} \left(\frac{(y_i - \mu_k)^2}{2\sigma_k} - \log \sqrt{2\pi}\sigma_k + \log c_k \right)$$

The updating equations for parameters μ_k, σ_k, c_k in M-step can now be obtained by setting the partial derivatives of Q in μ_k, σ_k and c_k to zero.

Gaussian mixture model estimation via the EM algorithm

E-step:

$$p_{ij}^{(m+1)} = \frac{G(y_i, \mu_j^{(m)}, \sigma_j^{(m)}) c_j^{(m)}}{\sum_{k=1}^{K} G(y_i, \mu_k^{(m)}, \sigma_k^{(m)}) c_k^{(m)}}$$

M-step:

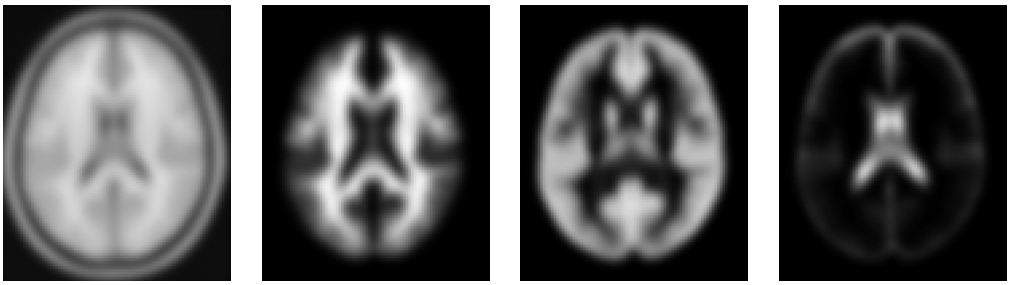
$$
\begin{aligned}
\mu_j^{(m+1)} &= \frac{\sum_{i=1}^{n} y_i p_{ij}^{(m+1)}}{\sum_{i=1}^{n} p_{ij}^{(m+1)}} \\
(\sigma_j^{(m+1)})^2 &= \frac{\sum_{i=1}^{n} (y_i - \mu_j^{(m+1)})^2 p_{ij}^{(m+1)}}{\sum_{i=1}^{n} p_{ij}^{(m+1)}} \\
c_j^{(m+1)} &= \frac{1}{n} \sum_{i=1}^{n} p_{ij}^{(m+1)}
\end{aligned}
$$

2.5 Prior information for EM-based segmentation

2.5.1 Probabilistic atlases

Any automatic segmentation of brain MRI should be robust and precise in order to be practically and clinically usable. It is therefore desired to improve the accuracy and robustness of the segmentation methods by using prior knowledge of the brain anatomy. In case of EM-based methods a probabilistic atlas is often used to initialize and constrain the segmentation process.

The standard *probabilistic atlas* consists of a template and three tissue probability maps for WM, GM and CSF [34], see fig. 2.3. It is obtained by aligning a number of segmented subjects followed by averaging of anatomical images to obtain the anatomical template and averaging the segmentation to obtain the tissue probability maps. The probabilistic atlas then describes the anatomical variability of the brain.



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

Figure 2.3: The MNI probabilistic atlas [34]: (a) the template; brain tissue prior probability maps: (b) WM; (c) GM; (d) CSF.

The probabilistic atlas which is most commonly used in intensity-based algorithms is the MNI305 atlas from Montreal Neurological Institute [34]. This atlas was created from 305 subjects by linear alignment into Talairach coordinate system [87]. Later, the International Consortium for Brain Mapping has affinely registered 152 scans to the MNI305 atlas to create the ICBM152 probabilistic atlas [58]. The probability maps of the ICBM152 atlas are shown in fig. 2.4.

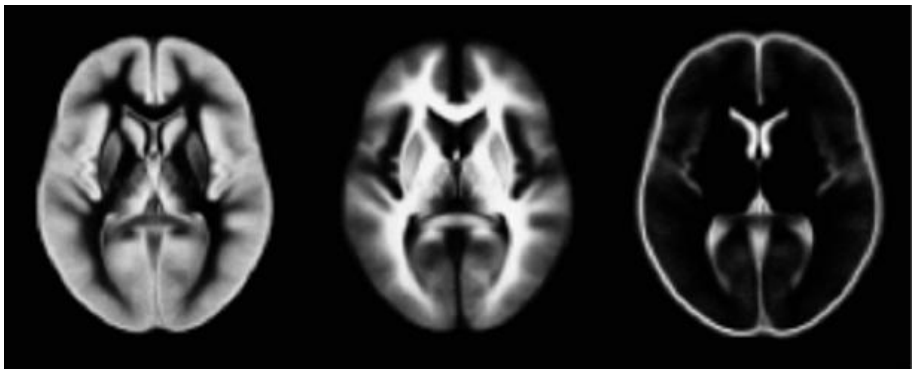


Figure 2.4: Probability maps from ICBM152 probabilistic atlas [58]. From left to right: GM, WM and CSF.

2.5.2 Aligning the probabilistic atlas

Before a probabilistic atlas can be used as a prior knowledge, it has to be registered with the image to be segmented. An aligned probabilistic atlas can be used as a good initial estimate of the segmentation, which is especially important for EM-based methods, as EM algorithm is guaranteed to converge to local, not global, maxima. In addition, most EM-based methods, [49], [4], [7], [67] use the probabilistic atlas to constrain the segmentation process as well. Therefore the correct alignment of the probabilistic atlas is crucial for a successful and accurate segmentation.

The traditional way of aligning the probabilistic atlas with the image is to register the anatomical template with the image to be segmented. Unfortunately, an affine alignment may not be sufficient if the brain anatomy in question differs from the average atlas anatomy significantly. Pohl *et al.* therefore suggest aligning the atlas using non-rigid registration [68]. However, in

their later work Pohl reports difficulties in registering anatomical template with the image to be segmented using standard registration methods [66]. D'Agostino developed a special similarity measure for registering probabilistic maps directly to the new image [28]. Recently, several methods have been developed which aim to overcome this problem by iteratively refining the segmentation and non-rigid registration of the probabilistic atlas at the same time. Ashburner developed a method for simultaneous segmentation, bias correction and non-rigid registration of a probabilistic atlas [7]. This approach has been proven to be successful in practice and is part of the software package SPM5.

2.5.3 Incorporating the probabilistic brain atlas into EM segmentation

Let us denote p_{ij}^{atlas} the probability that i^{th} voxel belongs to j^{th} tissue class. We can initialize the EM algorithm in the E-step of the first iteration with prior information from the atlas as follows:

$$p_{ij}^{0} = p_{ij}^{atlas}$$

After initialisation, the GMM can be used to calculate the segmentation. However, the classic GMM does not produce satisfactory results for the brain segmentation. Even if non-brain tissues are extracted in pre-processing step and only three brain tissues are left in the image, the noise will affect the resulting segmentation. Therefore the GMM works well only on high-contrast images with low levels of noise. A more robust algorithm can be obtained if the probabilistic atlas is used not only to initialize but also spatially constrain the segmentation process. Consequently, the voxels are classified using not only intensity information but also information about their spatial position.

Van Leemput *et al.* ([49], [48]) proposed an approach which amends the GMM by using the atlas as prior information at each iteration. The prior information is fixed and does not change:

$$p(Z_i = \mathbf{e}_j|\Phi) = p_{ij}^{atlas}$$

With this assumption the updating equations for the modified GMM resulting from EM optimization become:

EM segmentation by Van Leemput

E-step:

$$p_{ij}^{(m+1)} = \frac{G(y_i, \mu_j^{(m)}, \sigma_j^{(m)}) p_{ij}^{atlas}}{\sum_{k=1}^{K} G(y_i, \mu_k^{(m)}, \sigma_k^{(m)}) p_{ik}^{atlas}}$$

M-step:

$$\begin{aligned} \mu_j^{(m+1)} &= \frac{\sum_{i=1}^{n} y_i p_{ij}^{(m+1)}}{\sum_{i=1}^{n} p_{ij}^{(m+1)}} \\ (\sigma_j^{(m+1)})^2 &= \frac{\sum_{i=1}^{n} (y_i - \mu_j^{(m+1)})^2 p_{ij}^{(m+1)}}{\sum_{i=1}^{n} p_{ij}^{(m+1)}} \end{aligned}$$

In the GMM the mixture weights c_k are changed at each iteration to reflect the proportion of the image volume classified as the kth tissue type. In Van Leemput's model these weights are derived from the atlas, but they vary with the position of the voxel. Ashburner *et al.* [4] combines both approaches:

EM segmentation by Ashburner (SPM99)

E-step:

$$p_{ij}^{(m+1)} = \frac{G(y_i, \mu_j^{(m)}, \sigma_j^{(m)}) c_{ij}^{(m)}}{\sum_{k=1}^{K} G(y_i, \mu_k^{(m)}, \sigma_k^{(m)}) c_{ik}^{(m)}}$$

M-step:

$$\begin{aligned} \mu_j^{(m+1)} &= \frac{\sum_{i=1}^{n} y_i p_{ij}^{(m+1)}}{\sum_{i=1}^{n} p_{ij}^{(m+1)}} \\ (\sigma_j^{(m+1)})^2 &= \frac{\sum_{i=1}^{n} (y_i - \mu_j^{(m+1)})^2 p_{ij}^{(m+1)}}{\sum_{i=1}^{n} p_{ij}^{(m+1)}} \\ c_{ij}^{(m+1)} &= \frac{p_{ij}^{atlas} \sum_{l=1}^{n} p_{lj}^{(m+1)}}{\sum_{l=1}^{n} p_{lj}^{atlas}} \end{aligned}$$

Very often multi-spectral brain MR images are available, such as T1-weighted, T2-weighted and PD images. When correctly aligned, multi-spectral information can enhance the segmentation process and reduce the impact of artefacts such as noise or bias fields.

Let $\mathbf{y}_i = (y_{i1}, ..., y_{iR})$ denote the image intensities of R different channels of a multi-spectral MR image. Let us denote $\boldsymbol{\mu}_k = (\mu_{k1}, ..., \mu_{kR})$ where μ_{kr} denotes the mean intensity of the kth tissue in channel r and $\mathbf{\Sigma_k}$ the covariance matrix for all the channels of kth tissue class. We will assume a multivariate Gaussian mixture PDF for the image, in parallel with the single channel version:

$$G(\mathbf{y}_i, \boldsymbol{\mu}_k, \mathbf{\Sigma}_k) = \frac{1}{\sqrt{2\pi}^R |\mathbf{\Sigma}_k|^{\frac{1}{2}}} \mathrm{e}^{-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu}_k)^T \mathbf{\Sigma}_k^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_k)}$$

Van Leemput *et al.*([49], [48]) describes the multi-channel version of their algorithm:

Multichannel EM segmentation by Van Leemput *et al.*

E-step:

$$p_{ij}^{(m+1)} = \frac{G(\mathbf{y}_i, \boldsymbol{\mu}_j^{(m)}, \mathbf{\Sigma}_j^{(m)}) p_{ij}^{atlas}}{\sum_{k=1}^{K} G(\mathbf{y}_i, \boldsymbol{\mu}_k^{(m)}, \mathbf{\Sigma}_k^{(m)}) p_{ik}^{atlas}}$$

M-step:

$$\begin{aligned} \mu_{jr}^{(m+1)} &= \frac{\sum_{i=1}^{n} y_{ir} p_{ij}^{(m+1)}}{\sum_{i=1}^{n} p_{ij}^{(m+1)}} \\ (\Sigma_j^{(m+1)})_{rs}^2 &= \frac{\sum_{i=1}^{n} (y_{ir} - \mu_{jr}^{(m+1)})(y_{is} - \mu_{js}^{(m+1)}) p_{ij}^{(m+1)}}{\sum_{i=1}^{n} p_{ij}^{(m+1)}} \end{aligned}$$

2.5.4 Prior information for segmentation of brain structures

EM-based methods are traditionally used to segment brain into WM, GM and CSF because the boundaries of this tissues mostly correspond to intensity boundaries visible on MRI. However, if a prior probabilistic atlas is created for other anatomical structures, such as thalamus, caudate, putamen or cerebellum, this approach can be extended to segment more brain structures. The

probability maps for different structures can be created by averaging a number of affinely aligned manual segmentation in the same way as done for the three tissue classes. The main difficulty in segmenting different subcortical structures lies in the lack of intensity contrast between those structures and therefore the segmentation becomes crucially dependent on the correct alignment of the atlas with the image to be segmented. Pohl [66] reports that non-rigid registration has too many degrees of freedom to correctly align the subcortical grey matter brain structures. In his work he suggested the use of a structure-dependent, partially affine registration.

Fischl [36] takes the idea of using prior information further and creates a probabilistic atlas not only for the spatial location of the structures but also for the relative position of neighbouring structures. Let v_i be a location of a voxel with intensity y_i and N_i denote the 6-neighbourhood of v_i. Let v_j be a location of a voxel in 6-neighbourhood N_i. Let $I_1, ..., I_M$ denote training images with attached segmentations $S_1, ..., S_M$. Let $p^{atlas}_{jl|ik}$ denote the probability that tissue l is present at location v_j given that tissue k is present at location v_i. These values are calculated from the aligned manual segmentations $S_1, ..., S_M$ using following equation:

$$p^{atlas}_{jl|ik} = \frac{\|\{a|S_a(v_j) = l \wedge S_a(v_i) = k\}\|}{\|\{a|S_a(v_i) = k\}\|}$$

The aim is to find the MAP estimate of anisotropic (dependent on direction) MRF segmentation $\mathbf{z} = \{z_1, ..., z_n\}$. As it is not tractable to directly solve this problem, the approximate solution is iteratively calculated using Iterated Conditional Modes [10], until the segmentation does not change anymore:

$$z_i^{(m+1)} = \arg\max_k G(y_i, \mu_k, \sigma_k) p^{atlas}_{ik} \prod_{j;v_j \in N_i} p^{atlas}_{jz_i^{(m)}|ik}$$

In this method [36], Fischl et al. include spatially varying means and variances for the PDF of each structure in the prior probabilistic atlas. However, this can be problematic to do in practice, as these parameters will vary with the image acquisition protocol.

2.5.5 Including neighbourhood information in EM-based segmentation

In the previous section we showed how including neighbourhood information in the probabilistic atlas can enhance the segmentation of the brain into structures. In a similar fashion, the neighbourhood information can be used to eliminate the influence of noise and PVE on the resulting segmentation, as proposed by Van Leemput [49], and later also Zhang [102].

To achieve this, the EM-based segmentation by Van Leemput from section 2.5.3 can be adapted. Instead of using prior information from an aligned probabilistic atlas p_{ik}^{atlas}, a prior dependent on the labelling of neighbouring voxels $p_{ik|l_{N_i}}$ is introduced. Here l_{N_i} denotes a labelling of voxels in the neighbourhood N_i. The E-step in the EM algorithm then changes to:

$$p_{ik}^{(m+1)} = \frac{G(y_i, \mu_k^{(m)}, \sigma_k^{(m)}) p_{ik|l_{N_i}}^{(m+1)}}{\sum_{j=1}^{K} G(y_i, \mu_j^{(m)}, \sigma_j^{(m)}) p_{ij|l_{N_i}}^{(m+1)}}$$

The calculation of prior $p_{ik|l_{N_i}}^{(m+1)}$ can be achieved using the Markov Random Field (MRF) framework. A random field $Z = (z_1, ..., z_n)$ is a *Markov Random Field* if

$$P(z_i = k | \{z_1, ..., z_n\} \setminus z_i) = P(z_i = k | N_i) = p_{ik|l_{N_i}}$$

According to the Hammersley-Clifford theorem [9], an MRF can equivalently be characterized by a Gibbs distribution

$$p_{ik|l_{N_i}} = \frac{\mathrm{e}^{-u_{ik|l_{N_i}}}}{\sum_{j=1}^{K} \mathrm{e}^{-u_{ij|l_{N_i}}}}$$

where $u_{ik|l_{N_i}}$ are energy functions. Van Leemput [49] suggests the following calculation of the energy functions:

$$u_{ik|l_{N_i}}^{(m+1)} = \sum_{j \in N_i} \sum_{l=1}^{K} h_{kl}^{(m+1)} p_{jk}^{(m)}$$

where h_{kl} are elements of a $K \times K$ neighbourhood matrix $\mathbf{H}$, with K denoting the number of

tissues. Van Leemput et al. [49] finds a least square fit of **H** using the current approximation of the soft segmentation, which results in extra step within the E-step of EM algorithm. A different approach was adopted by Zhang et al. [102], who uses ICM [10] to estimate the hard segmentation followed by calculating the neighborhood probabilities. Unlike Fischl [36], Van Leemput [49] and Zhang [102] use an isotropic, or direction independent, MRF.

Using MRFs reduces the impact of noise on resulting segmentation and makes the segmentation algorithms more robust. It also helps avoid misclassifications due to partial volume effect - e.g. voxels on the boundary of WM and CSF in T1-weighted images can be misclassified as GM, as they have the same intensities.

Van Leemput et al. [49] suggested the use of both – the neighbourhood information and prior from the probabilistic atlas – to update the soft segmentation as:

$$p_{ik}^{(m+1)} = \frac{G(y_i, \mu_k^{(m)}, \sigma_k^{(m)}) p_{ik}^{atlas} p_{ik|l_{N_i}}^{(m+1)}}{\sum_{j=1}^{K} G(y_i, \mu_j^{(m)}, \sigma_j^{(m)}) p_{ij}^{atlas} p_{ij|l_{N_i}}^{(m+1)}}$$

2.6 Brain extraction

Brain extraction is a segmentation of the head MRI into brain and non-brain tissues. Since the intensities of brain and non-brain tissues overlap, an intensity-based brain extraction is not possible and spatial information must be integrated. The simplest method for brain extraction is to use prior information. A deformable template can be registered with an image and non-brain tissue is then removed by transferring the brain mask from the template, e.g. in [100]. Alternatively, the probabilistic atlas can be used as a spatial prior for one or several tissue classes for non-brain tissue during intensity-based segmentation, for example in [48]. Brain extraction using a probabilistic atlas is usually not very accurate and can cause misclassifications around the brain boundary.

An alternative method for extracting the brain is BET (brain extraction tool) [79], which is part of the publicly available software package FSL. This method finds the centre of gravity of

the brain and then inflates a sphere until the brain boundary is found. It has been proven to work in practice on good-quality T1 and T2 weighted images of the brain.

2.7 Segmentation of brain MRI during early childhood

To our knowledge there has been very little research published on the segmentation of brain MRI during early childhood. Even though by one or two years of age the brain is already similar to the adult brain, there are still significant differences is shape, size and appearance of the anatomical structures. Wilke *et al.* [97] point out that the anatomy of 5-18 year old children's brains differs from adult brains. Consequently, probabilistic atlases based on adult brains are not suitable as prior information for the segmentation of pediatric brains. Wilke *et al.* created new pediatric probabilistic atlases by segmenting pediatric brain MRI using SPM99 [80]. The segmentations were then averaged to obtain the probability maps for WM, GM, CSF as well as a brain mask. The resulting atlases [64] are publicly available. Recently, Gousias *et al.* [37] showed, that excellent segmentations of two year old brains can be achieved by multiple label propagation of manual segmentations from adult brains, using B-spline non-rigid registration by Rueckert *et al.* [73].

Several researchers attempted to develop specialized methods for neonatal segmentation. The semi-automatic kNN-based method by Warfield *et al.* [93] (see sec. 2.3.2) has been applied to adult as well as neonatal brain images. This method requires a deformable atlas. Zöllei in her thesis [104] shows that an unbiased population-specific probabilistic atlas created by non-rigid group-wise registration from population of pre-term children can significantly improve results of this method on neonatal images of pre-term subjects. Another method for group-wise construction of population-specific atlases have been proposed by Bhatia *et al.* [12], [11].

Prastawa *et al.* [69] combined several methods for the segmentation of adult brain MRI into a neonatal segmentation algorithm. The algorithm is initialized with a quasi-probabilistic atlas derived from a single subject manual segmentation. The initial tissue intensity distribution is then estimated by a procedure inspired by Cocosco [19] and used to initialize the EM-based

segmentation and bias correction by Van Leemput [48]. The resulting segmentation serves as an input to MST-based method by Cocosco [19] (see sec. 2.3.2), which identifies samples to estimate non-parametric tissue intensity distributions by Parzen windowing. The final segmentation is calculated using non-parametric intensity distributions and prior probabilities from the atlas.

The semi-automatic method by Weisenfeld *et al.* [94] estimates the intensity distributions of six tissue types (non-myelinated WM, myelinated WM, cortical GM, basal ganglia (subcortical GM), CSF and extra-cerebral tissue) from an interactively labelled training set of samples. The segmentation is then calculated via Maximum Likelihood estimation. The method uses an unbiased probabilistic atlas created by non-rigid group-wise registration by Zöllei [104].

In his method for the cortical segmentation of neonatal brains, Xue *et al.* [100] address the difficulties in segmenting neonatal brains directly. First, the brain is extracted and subcortical tissues removed by registration-based label propagation by Rueckert [73]. Next, an EM-based segmentation algorithm is used to obtain the segmentation of cortex. Since the neonatal brain is rapidly developing, it is not possible to create a suitable probabilistic atlas for all neonatal images. A quasi probabilistic atlas is therefore created for each subject by k-means clustering and blurring. The EM segmentation by Van Leemput [49] is then performed. A MRF is used to avoid misclassifications on the CSF-GM boundary, where voxels exhibit the same intensities as in WM, due to the reversed GM/WM contrast. Instead of estimating tissue intensity distributions for the whole brain, the within-tissue intensity variability is addressed by splitting the brain into homogeneous regions using a 4D k-means algorithm, in which features consists of the voxel position and voxel intensity. The intensity distributions are then estimated for each region separately throughout the segmentation process.

2.8 Comparison of registration-based and intensity-based approaches

Intensity-based and registration-based methods are often complementary in succeeding or failing in certain areas of the brain. While intensity-based methods are flexible enough to segment complicated cortical WM and GM regions, registration-based techniques are limited by the type of transformation and the degrees of freedom of the transformation used and therefore often not flexible enough to capture complex cortical folding.

On the other hand, the natural tissue intensity variation in different areas of the brain as well as overlaps in intensity distributions of different tissue classes are a significant source of error in intensity-based methods. An automatic segmentation cannot match the performance of a human expert unless guided by prior information. The correct alignment of prior information with image is difficult to achieve as well.

In addition, Fischl *et al.* [36] point out that different central brain structures have different intensity distributions. Subcortical GM is characterized by brighter intensities than cortical GM, and is therefore likely to be misclassified as WM. Ren *et al.* [71] reports such misclassification in adult brain using the EM-based segmentation method FAST [102]. However, non-rigid registration-based label propagation [73] succeeds in this area of the brain [25].

2.9 Conclusion

In this chapter we described several state-of-the-art methods for segmentation of brain MRI. We compared registration-based and intensity-based approaches. To develop a suitable method for segmenting the brain MRI during early childhood, we aim to combine the two approaches so that we can use the advantages of both. The performance of the methods presented in this chapter will only be satisfactory if intensity inhomogeneity of MRI has been corrected. The following chapter describes and compares various bias correction methods.

Chapter 3

Bias correction of brain MRI

3.1 Intensity inhomogeneity

The bias field is a low-frequency spatially varying shading artifact causing a slow variation of intensities in MRI within tissue of the same physical properties. It arises from variations in the sensitivity of the reception coil, spatial inhomogeneity of the excitation field and the interaction between the human body and the magnetic field [78], [77], [21]. This problem substantially affects MRI acquired with modern high-field MR scanners (3T and higher).

The results of intensity-based segmentation methods can be severely affected by this intensity inhomogeneity of the MRI. Various methods have therefore been proposed to correct the distorted intensities in MRI. The inhomogeneity is usually modelled as low frequency multiplicative field [96], [78], [65], [56], [84], [51].

If the intensities of MRI are logarithmically transformed, the multiplicative bias field becomes additive. Formally, the additive property of the bias field can be expressed as follows: Let $T = (t_1, ..., t_n)$ be the observed intensities of the original image and $T' = (t'_1, ..., t'_n)$ the ideal intensities. Let $B = (b_1, ..., b_n)$ denote the bias field. Then the degradation effect at each voxel can be expressed as

$$t_i = t'_i b_i$$

Let $Y = (y_1, ..., y_n)$ and $Y' = (y'_1, ..., y'_n)$ denote respectively the log-transformed observed and ideal intensities. The logarithmic transformation changes the multiplicative bias field to an additive bias field:

$$y_i = y'_i + b_i \tag{3.1}$$

Even though most state-of-the-art bias correction methods rely on this simple multiplicative model to represent the bias field, there are limitations to the correctness of this model. The model is consistent with the variations arising form the sensitivity of the receiver coil, however, non-uniformity due to induced currents and spatial inhomogeneity of the excitation field depend on the geometry and electromagnetic properties of the subject as well as the pulse sequence and coil polarization [78], resulting in a more complicated relationship between the true and measured intensities in MRI. In spite of these difficulties, the multiplicative low-frequency model is widely used in practice to model the distortions of intensities in MRI.

3.2 Bias correction methods

The low-frequency nature of the bias field can be exploited to estimate this artefact. The earliest methods proposed to correct the bias field are based on low-pass filtering [15], [20]. However, these approaches also remove the low-frequency component of the true image data and as a result additional artifacts can be introduced in the image. In contrast, early surface fitting methods depended on the manual labelling of WM voxels which were then used to reconstruct the bias field in form of a parametric surface. The disadvantage of this approach is the need for manual interaction. The surface fitting methods can be made fully automatic, if they are interleaved with automatic segmentation of brain tissues, e.g. in [5], [65], [86], [48]. Low pass filtering can be also improved if coupled with segmentation [96]. Other approaches include maximizing the high-frequency content of the image [78], minimization of the image entropy [56], or fitting of the histogram of the local neighbourhood to global histogram of the image [75]. Relatively recently, bias correction methods based on the use of a registered template image have emerged [51], [84]. Some of these methods are reviewed in detail in the following

sections.

3.2.1 C-means clustering

Pham and Prince [65] integrated the bias field correction step into their C-means segmentation method. The bias field is estimated at each iteration of the segmentation and C-means clustering is performed for the current estimate of the intensity-corrected image. There is no explicit prior model for the bias field, but the objective function is extended to include a regularization term which imposes smoothness on the bias field:

$$F = \sum_{i=1}^{n}\sum_{k=1}^{K}(u_{ik})^q\|y_i - b_i v_k\|^2 + \lambda_1 \sum_{i=1}^{n}\sum_{j=1}^{3}(\Delta^j * B)_i^2 + \lambda_2 \sum_{i=1}^{n}\sum_{j=1}^{3}\sum_{k=1}^{3}(\Delta^j * \Delta^k * B)_i^2$$

Here u_{ik} denotes the membership function and Δ^j denotes the finite difference operator along j^{th} dimension of the image. $(\Delta^j * B)_i$ refers to the ith voxel of the image obtained by convolution of bias field B with the kernel Δ^j. Minimizing this objective function results in extra bias correction step in the C-means classification algorithm, see section 2.3.1.

3.2.2 N3 method

N3 [78] is a non-parametric bias correction method, developed by Sled et al. to correct the bias field in pre-processing step. In contrast to previous method no segmentation is performed during the bias correction.

The additive nature of the bias field in the logarithmically transformed image

$$y = y' + b$$

results in the following relationship for intensity distributions:

$$P_y(y) = P_{y'}(y) * P_b(y) = \int P_b(y-b)P_{y'}(b)db$$

The non-uniformity distribution P_b can be viewed as blurring of the ideal intensity distribution $P_{y'}$. The distribution $P_{y'}$ can be therefore found by sharpening the measured distribution P_y. Sled et al. proposes to achieve this by iteratively deconvolving narrow Gaussian distributions from the intensity distribution of the image corrupted by the bias field. Given the distribution P_b and the measured distribution P_y, the distribution $P_{y'}$ can be obtained by deconvolution as follows:

$$\tilde{P}_{y'} = \frac{\tilde{P}_b^*}{|\tilde{P}_b|^2 + Z^2}\tilde{P}_y$$

where $\tilde{P}$ denotes the Fourier transform of P, * denotes the complex conjugate and Z is a constant. The bias corrected image intensities are then estimated as

$$E(y'|y) = \frac{\int y' P_b(y - y') P_{y'}(y') dy'}{\int P_b(y - y') P_{y'}(y') dy'}$$

The bias field is estimated by

$$b = y - E(y'|y)$$

followed by smoothing, which is achieved by fitting a regularized B-spline surface.

An implementation of N3 is publicly available and widely used especially on MRI data with weaker bias fields.

3.2.3 Entropy minimization

A similar idea of sharpening the intensity distribution of the image has been used by Mangin [56]. Since the bias field blurs the distribution of each tissue class, it consequently increases the entropy of the intensity distribution of the image. The aim of the method is therefore to find a smooth multiplicative field that minimizes the entropy of the corrected image. The optimal bias field is then found by minimizing a functional F consisting of the entropy of the corrected image, a regularization term to impose smoothness on the bias field and a quadratic term penalizing the difference of means of the original and corrected images:

$$F = \lambda_1 H(\hat{I}') + \lambda_2 C_{smooth} + \lambda_3 (\mu(I) - \mu(\hat{I}'))^2$$

Here I denotes the original image without logarithmic transformation of the intensities, $\hat{I}'$ denotes the current estimate of corrected image, H is the entropy estimated from normalized image histogram, and $\mu(I)$ denotes image mean. The smoothness constraint C_{smooth} represents the membrane energy

$$C_{smooth} = \sum_i \sum_{j, v_j \in N_i} (\log b_i - \log b_j)^2$$

where v_i denotes the spatial position of a voxel, b_i the bias at the position v_i and N_i denotes the 6-neighbourhood of the voxel v_i.

The functional is optimized using a fast annealing schedule. During the optimization only a piecewise linear function is used to approximate the bias field. The final estimate is then modelled using a cubic spline representation.

3.3 Bias field estimation via EM

The estimation of the bias field via the EM algorithm was first proposed by Wells et al. [96] as a part of the segmentation algorithm. If estimates of the segmentation and tissue intensity distributions are available an estimate of the bias corrected image can be calculated and the bias field is then obtained from the difference of the log-transformed real and estimated image. This section describes the EM-based bias correction [96] and different variations of this algorithm as proposed by [48], [7] and [67].

3.3.1 Parametric bias field estimation via the ML principle

Thanks to the additive nature of the model (eq. 3.1), the bias field estimation can be easily incorporated into the EM segmentation algorithm (sec. 2.5.3). The likelihood of observing a voxel intensity y_i in the presence of the bias field can be expressed as

$$P(y_i | Z_i = e_k, \gamma, \beta) = G(y_i - b_i, \mu_k, \sigma_k)$$

The low frequency characteristics of the bias field B can be modelled as a linear combination of smooth basis functions $\Psi_l(v)$

$$b_i = \sum_{l=1}^{L} d_l \Psi_l(v_i)$$

where v_i denotes the 3D position of voxel i and $\beta \equiv \mathbf{d} = (d_1, ..., d_L)$ denotes the bias field parameters. The basis functions can be polynomials, splines or any other smooth functions.

The maximum likelihood estimate of the bias field parameters $\mathbf{d}$ can be found by setting the partial derivative of the lower bound $Q(\Phi, \Phi^{(m)})$ to zero:

$$\frac{\partial}{\partial d_j}(Q(\Phi, \Phi^{(m)})) = 0$$

This approach was taken by Van Leemput et al. in [48]:

EM segmentation with bias field correction by Van Leemput

E-step:

$$p_{ij}^{(m+1)} = \frac{G(y_i - \sum_{l=1}^{L} d_l^{(m)} \Psi_l(v_i), \mu_j^{(m)}, \sigma_j^{(m)}) p_{ij}^{atlas}}{\sum_{k=1}^{K} G(y_i - \sum_{l=1}^{L} d_l^{(m)} \Psi_l(v_i), \mu_k^{(m)}, \sigma_k^{(m)}) p_{ik}^{atlas}}$$

M-step:

1. Gaussian distribution parameters estimation

$$\begin{aligned} \mu_j^{(m+1)} &= \frac{\sum_{i=1}^{n} (y_i - \sum_{l=1}^{L} d_l^{(m)} \Psi_l(v_i)) p_{ij}^{(m+1)}}{\sum_{i=1}^{n} p_{ij}^{(m+1)}} \\ (\sigma_j^{(m+1)})^2 &= \frac{\sum_{i=1}^{n} (y_i - \sum_{l=1}^{L} d_l^{(m)} \Psi_l(v_i) - \mu_j^{(m+1)})^2 p_{ij}^{(m+1)}}{\sum_{i=1}^{n} p_{ij}^{(m+1)}} \end{aligned}$$

2. Bias correction

$$(\mathbf{d}^{(m+1)})^T = (\mathbf{F}^T \mathbf{W}^{(m+1)} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{W}^{(m+1)} R^{(m+1)} \tag{3.2}$$

$$\mathbf{F} = \begin{pmatrix} \Psi_1(v_1) & \Psi_2(v_1) & . & . & . & \Psi_L(v_1) \\ \Psi_1(v_2) & \Psi_2(v_2) & & & & \Psi_L(v_2) \\ . & & . & & & . \\ . & & & . & & . \\ . & & & & . & . \\ \Psi_1(v_n) & \Psi_2(v_n) & . & . & . & \Psi_L(v_n) \end{pmatrix}$$

$$\mathbf{W}^{(m+1)} = \begin{pmatrix} \sum_{k=1}^{K} w_{1k}^{(m+1)} & 0 & . & . & . & 0 \\ 0 & \sum_{k=1}^{K} w_{2k}^{(m+1)} & & & & 0 \\ . & & . & & & . \\ . & & & . & & . \\ . & & & & . & . \\ 0 & 0 & . & . & . & \sum_{k=1}^{K} w_{nk}^{(m+1)} \end{pmatrix}$$

$$w_{ik}^{(m+1)} = \frac{p_{ik}^{(m+1)}}{(\sigma_k^{(m+1)})^2}$$

$$\tilde{y}_i^{(m+1)} = \frac{\sum_{k=1}^{K} w_{ik}^{(m+1)} \mu_{ik}^{(m+1)}}{\sum_{k=1}^{K} w_{ik}^{(m+1)}}$$

$$R = \begin{pmatrix} y_1 & - & \tilde{y}_1^{(m+1)} \\ & . & \\ & . & \\ & . & \\ y_n & - & \tilde{y}_n^{(m+1)} \end{pmatrix}$$

The bias correction step can be interpreted as follows: The estimated soft segmentation and Gaussian distribution parameters can be used to reconstruct the image estimate $\tilde{Y} = (\tilde{y_1}, ..., \tilde{y_n})$ which is not corrupted by the bias field. When subtracted from the observed image, the *residual image* R is calculated. From the residual image the bias field is estimated. The matrix $\mathbf{F}$ represents the discretized geometry of the bias field and $\mathbf{W}$ is an inverse covariance matrix.

The ML estimation of B-spline control points $\mathbf{d}$ results in a weighted least squares estimate of $\mathbf{d}$ from the residual image R with weights on the diagonal of the matrix $\mathbf{W}$ and modified normal equations:

$$\mathbf{F}^T\mathbf{W}^{(m+1)}\mathbf{F}(\mathbf{d}^{(m+1)})^T = \mathbf{F}^T\mathbf{W}^{(m+1)}R^{(m+1)}$$

3.3.2 Non-parametric bias field estimation via MAP principle

The bias field can be also estimated in the EM framework using the MAP principle, as proposed by Wells *et al.* [96]. In this method the Gaussian distribution parameters are assumed to be known (they are estimated from the histogram in a preprocessing step) and the EM algorithm is only used to estimate the bias field. As the MAP criteria is closely related to the ML criteria (eq. 2.6, 2.8), there exists a lower bound $Q_{MAP}(\Phi, \Phi^{(m)})$ to $L_{MAP}(\Phi)$ (eq. 2.8), as shown in [66]:

$$Q_{MAP}(\Phi, \Phi^{(m)}) = Q(\Phi, \Phi^{(m)}) + \log P(\Phi)$$

In Wells' method [96] the parameters Φ directly represent the bias field B: $\Phi \equiv \beta \equiv B = (b_1, ..., b_n)$. The prior term for the parameters $\log P(\Phi)$ is included in the optimisation function, and therefore a n-dimensional zero mean Gaussian distribution

$$P(\Phi) = P(B) = G(B, \mathbf{0}, \mathbf{\Sigma}_B)$$

can be assumed for the bias field, thus eliminating the need for the parametric model. The equation for the bias field estimation step will then become

$$(B^{(m+1)})^T = (\mathbf{W}^{(m+1)} + \mathbf{\Sigma}_B^{-1})^{-1}\mathbf{W}^{(m+1)}R^{(m+1)}$$

This equation differs from (eq. 3.2) by including the smoothness constraint $\mathbf{\Sigma}_B^{-1}$ and setting $\mathbf{F}$ to an identity matrix, as no parametric model for the bias field is assumed.

Let us further define the *weighted residual image* $\bar{R}^{(m+1)}$

$$\bar{R}^{(m+1)} = \mathbf{W}^{(m+1)} R^{(m+1)}$$

which means that

$$\bar{R}_i^{(m+1)} = \sum_{k=1}^{K} \frac{p_{ik}^{(m+1)}(y_i - \mu_k^{(m+1)})}{\sigma_k^{(m+1)}}$$

The bias field estimation step can be then simplified to

$$(B^{(m+1)})^T = \mathbf{H}\bar{R}^{(m+1)}$$

where $\mathbf{H}$ is a linear operator defined by:

$$\mathbf{H} \equiv (\mathbf{W}^{(m+1)} + \mathbf{\Sigma}_B^{-1})^{-1}$$

In practice the linear operator H can be approximated by a linear low-pass filter. Wells et al. uses the following efficient implementation which incorporates the weights into the low-pass filtering:

$$b_i^{(m+1)} = \frac{[\mathbf{F}\bar{R}^{(m+1)}]_i}{[\mathbf{F}\mathbf{W}^{(m+1)}\mathbf{1}]_i} \tag{3.3}$$

where $\mathbf{F}$ is a low-pass filter and $\mathbf{1} = (1, ..., 1)^T$.

3.4 Simultaneous segmentation, registration and bias correction

The correct alignment of the prior information in form of probabilistic atlases with the image to be segmented is crucial for successful intensity-based segmentation. Recently developed methods aim to achieve better registration of these probabilistic atlases to the image by gradually refining the registration during segmentation process. In the following, the registration parameters will be denoted as α, the bias field parameters as β, the means and variances of the

Gaussian intensity distribution as γ, the image intensities as Y and the segmentation as Z.

3.4.1 Incorporating the registration of the probabilistic atlas into the EM segmentation algorithm

The registration of a probabilistic atlas can be incorporated into the EM framework. Pohl *et al.* [67] extends Wells' MAP segmentation and bias correction framework (described in section 3.3.2) to include the registration parameters as unknowns which must be estimated during the segmentation. In this method the Gaussian distribution parameters are assumed to be known and unknown parameters $\Phi = (\beta, \alpha)$ consist of the bias field parameters β and the registration parameters α. The aim is to maximize MAP of both sets of parameters (β, α):

$$(\hat{\beta}, \hat{\alpha}) = \arg\max_{\beta,\alpha} P(\beta, \alpha|Y)$$

It has been shown in [66] that this can be optimized in the EM framework through iterative maximization of the lower bound:

$$(\beta^{(m+1)}, \alpha^{(m+1)}) = \arg\max_{\beta,\alpha} E_Z(\log P(\beta, \alpha, Z|Y)|Y, \beta^{(m)}, \alpha^{(m)})$$

The expression $\log P(\beta, \alpha, Z|Y)$ can be decomposed into

$$\log P(\beta, \alpha, Z|Y) = \log P(Y|Z, \beta, \alpha)) + \log P(\alpha|Z, \beta) +$$

$$+ \log P(\beta|Z) + \log P(Z) - \log P(Y)$$

The likelihood of observing the image Y given segmentation Z and bias field β does not depend on the registration of the prior information α. We can also assume that the registration α does not depend on the bias field β and can be fully inferred from the segmentation Z. In addition, the bias field β does not depend on segmentation Z. The last two terms are constant in parameters (β, α) and therefore can be omitted in the maximization process. This results in

the following simplification:

$$(\beta^{(m+1)}, \alpha^{(m+1)}) = \arg\max_{\beta,\alpha} E_Z(\log P(Y|Z,\beta) + \log P(\alpha|Z) + \log P(\beta)|Y, \beta^{(m)}, \alpha^{(m)})$$

Similar to function $Q(\Phi, \Phi^{(m)})$ in section 2.4.4, this equation can be rewritten as:

$$(\beta^{(m+1)}, \alpha^{(m+1)}) = \arg\max_{\beta,\alpha} \sum_{i=1}^{n} \sum_{j=1}^{K} p_{ij}^{(m+1)} (\log(P(Y|Z_i = e_j, \beta)) +$$

$$+ \log P(\alpha|Z_i = e_j) + \log P(\beta))$$

The terms in this expression are always dependent either on α or β, but never on both, so the bias field and registration parameters can be estimated separately. If we further observe that maximizing $\log P(\alpha|Z)$ is equivalent to maximizing $\log P(Z|\alpha) + \log P(\alpha)$, we can see that the segmentation can be achieved by following algorithm:

Simultaneous segmentation and registration by Pohl *et al.*

E-step:

$$p_{ij}^{(m+1)} = \frac{P(y_i|Z_i = e_j, \beta^{(m)}) P(Z_i = e_j|\alpha^{(m)})}{\sum_{k=1}^{K} P(y_i|Z_i = e_k, \beta^{(m)}) P(Z_i = e_k|\alpha^{(m)})}$$

M-step:

$$\alpha^{(m+1)} = \underset{\alpha}{\operatorname{argmax}} \sum_{i=1}^{n} \sum_{j=1}^{K} p_{ij}^{(m+1)} \log P(Z_i = e_j|\alpha) + \log P(\alpha)$$

$$\beta^{(m+1)} = \underset{\beta}{\operatorname{argmax}} \sum_{i=1}^{n} \sum_{j=1}^{K} p_{ij}^{(m+1)} (\log P(y_i|Z_i = e_j, \beta) + \log P(\beta))$$

The estimation of the bias field parameters β is described in section 3.3.2. The registration parameters can represent any kind of registration - rigid, affine or non-rigid. Note, that the expression $P(Z_i = e_j|\alpha)$ represents the prior probability values from the registered probabilistic

atlas and $P(\alpha)$ the prior constraints for registration. To find the registration parameters α, numerical optimization methods have to be employed as it is usually not possible to find a closed-form expression for the registration estimation step.

3.4.2 Minimization of objective function using iterated conditional modes

Instead of optimizing all parameters in the EM framework, Ashburner [7] uses Iterated Conditional Modes (ICM)[10] to unify segmentation, registration and bias correction in a probabilistic framework.

The task is to optimize a global objective function F:

$$(\alpha, \beta, \gamma) = \arg\max_{\alpha,\beta,\gamma} \log P(Y, \alpha, \beta|\gamma) = \arg\max_{\alpha,\beta,\gamma} F(\alpha, \beta, \gamma)$$

where α are the registration parameters, β the bias field parameters and γ the Gaussian mixture parameters. The ICM optimization can be described as follows:

Iterated Conditional Modes for simultaneous segmentation, registration and bias correction

---

Repeat until convergence:

Segmentation step:

Hold α and β constant, and maximize F with respect to γ

Bias correction step:

Hold γ and α constant, and maximize F with respect to β

Deformation of the probabilistic atlas:

Hold γ and β constant, and maximize F with respect to α

end

The objective function F is modelled in a similar fashion as in approaches from previous sections. However, the original image intensities, rather than log-transformed intensities, are used. Maximization of the objective function can be expressed as follows:

$$(\alpha, \beta, \gamma) = \arg\max_{\alpha,\beta,\gamma} F(\alpha, \beta, \gamma) = \arg\max_{\alpha,\beta,\gamma} \log P(Y, \alpha, \beta|\gamma) =$$

$$= \arg\max_{\alpha,\beta,\gamma} \log(P(Y|\alpha, \beta, \gamma) + \log P(\alpha) + \log P(\beta))$$

The terms $\log P(\alpha)$ and $\log P(\beta)$ enable the incorporation of prior constraints on the registration and bias field parameters. Both are defined using a Gaussian distribution: $P(\alpha) = G(\alpha, \mathbf{0}, \mathbf{C}_\alpha)$ and $P(\beta) = G(\beta, \mathbf{0}, \mathbf{C}_\beta)$.

The term $\log P(Y|\alpha, \beta, \gamma)$ is the log-likelihood and can be expressed by eq. 2.11. Priors $P(z_i = e_k|\alpha, \beta, \gamma)$ are a combination of mixture coefficients c_k and prior probabilities from the registered probabilistic atlas $p_{ik}^{atlas}(\alpha)$:

$$P(z_i = e_k|\alpha, \beta, \gamma) = P(z_i = e_k|\alpha, \gamma) = \frac{c_k p_{ik}^{atlas}(\alpha)}{\sum_{j=1}^{K} p_{ij}^{atlas}(\alpha)}$$

The likelihood $\log P(y_i|z_i = e_k, \alpha, \beta, \gamma)$ represents the intensity distribution of tissue class k and is modelled by Gaussian PDF, extended by including the bias field $\rho_i(\beta)$ in the model:

$$P(y_i|z_i = e_k, \alpha, \beta, \gamma) = P(y_i|z_i = e_k, \beta, \gamma) = \frac{1}{\sqrt{2\ pi}\frac{\sigma_k}{\rho_i(\beta)}} \exp(\frac{(y - \frac{\mu_k}{\rho_i(\beta)})^2}{2(\frac{\sigma_k}{\rho_i(\beta)})^2})$$

Optionally, the model for the intensity distribution of a tissue class can be extended to mixture of Gaussian PDFs, so that the distribution is not restricted to Gaussian shape.

The resulting objective function F is maximized using ICM. The segmentation step is performed by the EM algorithm, while the registration and bias correction step are optimized using a Levenberg-Marquardt scheme

$$\theta^{(m+1)} = \theta^{(m)} - \left(\frac{\partial^2 F}{\partial \theta^2} + \lambda I\right)^{-1} \left(\frac{\partial F}{\partial \theta}\right)$$

which requires the calculation of first and second derivatives of F with respect to the parameters α and β. The parameter λ regulates stability versus speed of convergence of the scheme.

This framework for the simultaneous segmentation, registration and bias correction has been implemented in publicly available software package SPM5 [81]. The advantage of this implementation is that non-rigid registration of the probabilistic atlas is performed, thus improving the accuracy of the segmentation and reconstruction of the corrected image, and consequently robustness of the bias correction. The method is suitable for segmentation and bias correction of MRI data with strong bias field.

3.5 Template-based bias correction

The majority of bias correction methods rely on the assumption that either tissue intensities fall within a discrete set of classes, or that the intensity distribution should be most peaked when tissue intensities have been corrected. However, the underlying cell structure of different brain regions induces a variation of the observed intensities in MRI, especially in subcortical structures such as thalamus. In addition, brain tissues are known to exhibit intensity changes due to illness and aging [84].

The development of accurate registration methods for brain MRI offers an alternative approach to correcting the intensity inhomogeneity, which may overcome the previously mentioned problems. If an image corrupted by a bias field can be correctly aligned with a template image without bias field, it can help to separate the anatomically induced intensity variation from the variation imposed by the acquisition process.

Lewis *et al.* [51] proposed such a template-based bias correction technique to correct differential intensity inhomogeneity in longitudinal studies. The technique removes the difference of the bias fields of a pair of images to enable their further comparison. Longitudinal studies deal with the same patient scanned at different time points, so only a rigid registration is required. In addition, the MRI data is assumed to be obtained by the same imaging protocol, and therefore

no intensity matching is required. The assumed model for the image formation is

$$y = y'b + n$$

where n denotes the noise. The log-transformed intensity differences of images X and Y then become

$$\log(y) - \log(x) = \log\left(\frac{y'b_y - n_y}{x'b_x - n_x}\right) =$$

$$= \log(y') - \log(x') + \log(b_y) - \log(b_x) + \log\left(1 + \frac{n_y}{yb_y}\right) - \log\left(1 + \frac{n_x}{xb_x}\right)$$

The last two terms represent the original additive noise of the system. $\log(y') - \log(x')$ represent the difference of the true intensities of the image, and given that the images come from the same subject, the difference contains only residual registration error and therefore will have only small scale structure in a scale space sense. The term $\log(b_y) - \log(b_x)$ represents the differential bias field which is assumed to be dominant at larger scales compared to other contributing factors. If a median filter is applied to this differential image, it will remove the Gaussian noise as well as small-scale structure, if a suitable size of a filtering kernel is chosen. Therefore

$$\log b_y - \log b_x \approx \text{median}(\log y - \log x)$$

resulting in

$$\frac{b_y}{b_x} \approx \exp(\text{median}(\log y - \log x))$$

Studholme *et al.* [84] proposes a similar technique for inter-subject bias correction of scans acquired by the same imaging protocol. First, a template image is chosen and WM is manually segmented. The WM intensities are then fitted with a B-spline, to extrapolate the estimate of intensity inhomogeneity to other regions of the brain.

The bias corrected template is then aligned with the image by B-spline non-rigid registration by Rueckert [73]. In the next step, both the aligned template and the image to be corrected are smoothed with Gaussian kernel. Let $\mathcal{T}$ denote the transformation from image Y to template X. Let $\mathcal{M}$ be a binary mask for identification of brain volume and $\mathcal{G}$ an isotropic Gaussian

filter. Then both image and template can be filtered as follows:

$$X_g(v) = \frac{\sum_{u \in U} \mathcal{M}(v-u)\mathcal{G}(v-u,\sigma)X(\mathcal{T}(v))}{\sum_{u \in U} \mathcal{M}(v-u)\mathcal{G}(v-u,\sigma)}$$

$$Y_g(v) = \frac{\sum_{u \in U} \mathcal{M}(v-u)\mathcal{G}(v-u,\sigma)Y(v)}{\sum_{u \in U} \mathcal{M}(v-u)\mathcal{G}(v-u,\sigma)}$$

where Y_g and X_g denote the blurred image and template, respectively, and U is the domain of the discretized and truncated zero-mean Gaussian kernel. The bias field can then be estimated as

$$b(v) = \frac{Y_g(v)}{X_g(v)}$$

and image Y can be corrected by

$$Y'(v) = \frac{Y(v)}{b(v)}$$

The standard deviation σ of the Gaussian kernel was experimentally estimated by calculating the root mean square intensity difference (RMS) between the automatically corrected image Y' and manually corrected image Y'_m:

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y'(v_i) - Y'_m(v_i))^2}$$

The best results were achieved with approximately Gaussian with $\sigma = 20$mm at FWHM.

3.6 Comparison of the methods

There have been numerous techniques for correction of bias field proposed in literature, however, the number of comparative studies is relatively small. Recent reviews of different bias correction methods [42], [8] point to the study by Arnold *et al.* [3] as the most comprehensive recent comparative study. According to this study, the N3 algorithm [78] exhibits the most stable performance, outperforming most of the other techniques under most circumstances. On the other hand, the good performance of EM-based bias correction techniques becomes unstable

if only small bias field is present. It has to be noted though, that the EM-based technique evaluated in this study is the one implemented in SPM99 [5] and that a new version in SPM5 [7] has been developed in the meantime. Both reviews conclude that none of the methods can be considered the best under all conditions, but N3 method performs consistently well under most circumstances. In contrast, Studholme reports, that his template-based method [84] performs better than N3 if a suitable filtering kernel width is chosen.

There is a lack of studies in literature comparing the recent bias correction methods, such as template-based bias correction or simultaneous segmentation, registration and bias correction. In addition, recent trend in MR acquisition is to use stronger magnetic fields, such as 3T, usually resulting in stronger bias fields than in earlier MRI. Due to these developments in MR technology, new studies are needed to offer up-to-date comparison of performance of current bias correction methods the recently acquired data. The promising approaches suitable for bias correction of MRI with stronger bias fields include methods based on simultaneous segmentation, registration and bias-correction, such as SPM5 [7] or the template-based bias correction based on non-rigid registration, such as the one by Studholme [84].

3.7 Conclusion

In this chapter we described the most important state-of-the-art bias correction methods. The correction of intensity inhomogeneity is an important step for the efficient segmentation and registration of brain MRI. In line with the conclusions of the comparative studies of bias correction methods we will use N3 method for images with lower field strengths. However, in our experiments, N3 did not prove efficient on MRI exhibiting very strong bias field, such as 3T images scanned with an MP-RAGE imaging sequence [14]. We will therefore develop a novel robust template based bias correction method efficient for correcting these challenging datasets.

Chapter 4

Segmentation of brain MRI during early childhood

4.1 Introduction

The aim of this chapter is the development of segmentation tools for MRI of the brain during early childhood. A novel method for the segmentation of brain structures combining registration-based and EM-based approaches is proposed. The evaluation of the proposed method shows consistent results superior to both, registration-based and EM-based approaches. The method is applied for the quantification of volumes and growth of brain structures at one and two years of age.

First, we aim to adapt state-of-the-art methods for the segmentation of adult brain MRI to MRI during early childhood. To achieve this, the non-rigid registration by Rueckert *et al.* [73] is chosen to perform registration-based segmentation in sec. 4.2. To implement an intensity-based approach, the EM-based segmentation by Van Leemput *et al.* [49] is chosen and adapted in section 4.3. We will show that prior information created specifically for this age-group is necessary to achieve good segmentation results. Therefore, a method for creating population-specific probabilistic atlases will be proposed.

Second, the registration-based and EM-based approach are combined in a novel method described in sec. 4.4. The robustness of this method is achieved by combining multiple registration-based label propagation to create a subject-specific atlas which is then used as prior information by the subsequent EM-based segmentation.

Sec. 4.5 compares the registration-based, EM-based and our proposed combined approach. We demonstrate that the registration-based segmentation performs better in the subcortical area of the brain while the intensity-based segmentation tends to be very successful in the cortical area. Our proposed combined method exploits the advantages of both approaches and consequently outperform them.

Finally, we will demonstrate the robustness of the proposed segmentation method for the quantification of the volumes and growth of brain structures during early childhood.

In this chapter we focus on the segmentation of images with negligible or weak bias field so that the issue of bias correction does not need to be addressed directly. The development of segmentation tools for the brain MRI in the presence of strong bias fields is the focus of chapters 5 and 6.

4.1.1 Datasets

The data used in this chapter are T1-weighted brain MRI acquired using a 1.0T HPQ system (Philips Medical Systems), TR/TE=23ms/6ms, flip angle = 30 degrees. The images were reconstructed with voxels dimensions $1.6\times1.035\times1.035$ mm^3. The images exhibit only a weak bias field, see fig. 1.5b and 4.1a, which does not affect the visual evaluation of the images, however is strong enough to cause problems for any intensity-based segmentation algorithm.

Subjects were imaged at one and two years of age (36 scans at one year and 33 scans at two years). Out of these 69 scans 32 belong to 16 subjects scanned at both time-points. One of the subjects with scans at the age of one and two was randomly chosen as the reference subject. The remaining 15 subjects with scans at both time-points were used to calculate volumes and growth of brain structures in sec 4.6. All subjects excluding the reference subject (35 scans

at one year and 32 scans at two years of age) were used as training images to create the population-specific and subject-specific atlases.

4.1.2 Pre-processing

To prepare the data for further processing, we performed the following pre-processing steps: First, the orientation of the data was aligned with orientation of MNI305 probabilistic atlas [34] using only a 90° and 180° rotation. To avoid the need for resampling, a sequence of reflections equivalent to the rotations needed was used to rearrange the voxels.

Next, the background voxels were removed with following procedure: The image is segmented into background and non-background voxels using a Gaussian Mixture Model (see sec. 2.4.5). To avoid removing the voxels inside the head, the segmentation is followed by a series of dilations and erosions of the non-background tissue to close the surface of the head. Then, a region-growing procedure starting at the eight corners of the image is used to find all the voxels outside the head. Removing the background voxels is vital for speeding up all the segmentation methods. This procedure is different from brain extraction, such as the one performed by BET [79], as we aim to keep the scalp in the images. The skull is an important feature for achieving good registration results at the brain boundary. The result of this background removal procedure is illustrated in fig. 4.1. Finally, after background removal, the intensities of the remaining voxels are bias corrected using the N3 method [78].

(a) (b) (c)

Figure 4.1: Removing background voxels: (a) MRI of a two year old child; (b) image with adjusted contrast to show the background noise; (c) final results of the background removal procedure.

4.1.3 Brain tissues vs. brain structures

The standard probabilistic atlases, such as MNI305 [34], ICBM152 [58] and CCHMC5-9 [97], consist of probability maps for the three main brain tissues - WM, GM and CSF. In the reminder of these thesis, the segmentation into these three tissue classes will be referred to as *segmentation of brain tissues*. However, when assessing the development of the brain in prematurely born children, it is not sufficient to quantify the volumes of brain tissues. Very premature birth (at under 32 weeks of gestation) has been linked to atypical development of GM structures, such as thin cortex with simplified gyri and reduction of volume in cerebellum and subcortical GM structures (eg. basal ganglia: caudate, putamen, thalamus and pallidum) [13], [52], [16].

Our aim is to choose a suitable set of brain structures important for the assessment of developmental changes in premature brains, which can be segmented by intensity-based as well as registration-based methods. Our deformable and probabilistic atlases will be therefore created for the following 11 structures: CSF, cortex, caudate, putamen, substantia nigra, cerebellar GM, thalamus, pallidum, brainstem, WM and cerebellar WM. The segmentation of the brain into these 11 structures will be referred to as *segmentation of brain structures*. In addition to their relevance to development of premature brains, these 11 structures exhibit the natural differences in their intensity distributions. Structures such as amygdala or hippocampus were not considered, as their intensities do not differ from intensities of cortical GM and intensity-based methods are not suitable for their delineation.

When comparing the segmentation methods proposed in this thesis with segmentation algorithms which use standard probabilistic atlases, we will merge the 11 structures to the three brain tissue classes as shown in table 4.1:

Brain tissues	Brain structures
WM	WM, brainstem, cerebellar WM
GM	cortex, cerebellar GM, caudate, putamen, nigra, thalamus, pallidum
CSF	CSF

Table 4.1: Correspondence between brain tissues and brain structures.

4.1.4 Manual segmentation of the reference subject

Creating new population-specific deformable and probabilistic atlases requires a large number of manual segmentations. However, this is often difficult to achieve in practice. The manual segmentation of a whole brain MRI is very laborious and costly, requiring several full working days or even weeks of an experienced human rater.

To develop the segmentation tools for brain MRI during early childhood, we were able to obtain one manual segmentation of the reference subject scanned at two years of age, containing 11 segmented structures: CSF, cortex, caudate, putamen, substantia nigra, cerebellar GM, thalamus, pallidum, brainstem, WM and cerebellar WM, see fig. 4.2. The segmentation contains elements from the discrete set of labels $\{0, 1, ..., 11\}$ with 0 denoting the non-brain tissue and background voxels.

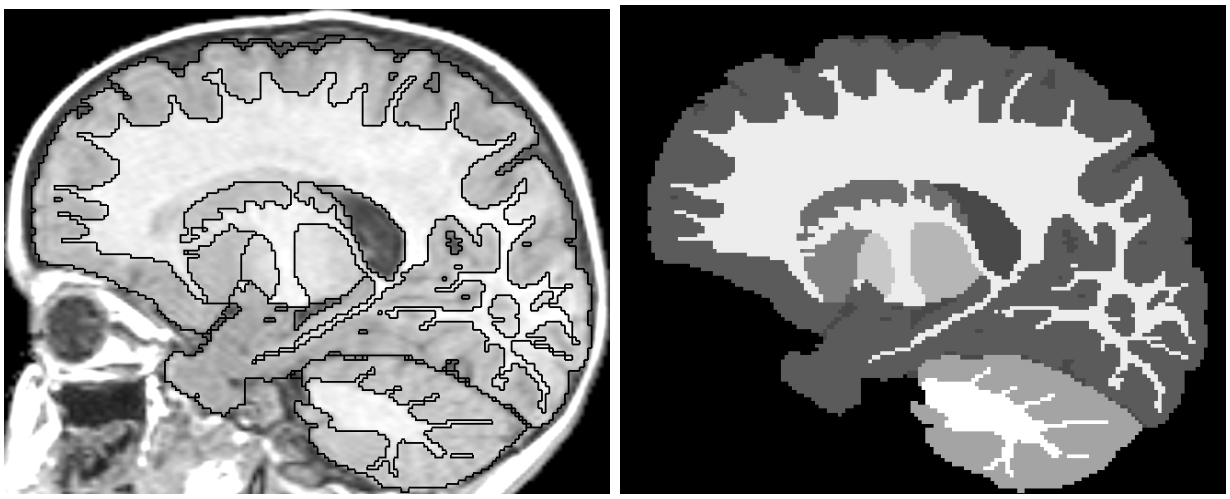


Figure 4.2: Manual segmentation of brain structures for the two years old reference subject contains these 11 structures: CSF, cortex, caudate, putamen, substantia nigra, cerebellar GM, thalamus, pallidum, brainstem, WM and cerebellar WM.

This manual segmentation will be used as a deformable atlas, as described in sec. 2.2, to segment the brain MRI at two years of age. In the following sections we will also propose methods for creating population-specific and subject-specific probabilistic atlases using only a single manual segmentation.

4.1.5 Evaluation using manual segmentations

To evaluate the algorithms for the segmentation of brain structures proposed later in this chapter, we need additional manual segmentations of several subjects. Four more two years old subjects were therefore partially segmented, with the WM and cortical GM being segmented on 6-8 slices, and a complete segmentation of caudate and thalamus.

The results of the automatic segmentation methods will be evaluated throughout this chapter against these four manual segmentations using the Dice metric [30]

$$D = \frac{2 * |T_{gt} \cap T_{seg}|}{|T_{gt}| + |T_{seg}|} \tag{4.1}$$

where T_{gt} and T_{seg} denote the set of samples of a given tissue type of the ground truth (in our evaluation the manual segmentation) and the automatic segmentation, respectively.

This measure of accuracy can be also described in terms of true/false positives/negatives. For each structure l, the *true positives* (TP) are voxels labelled as belonging to the structure l in both the manual and automatic segmentation, the *false positives* (FP) are labelled as structure l in automatic, but not in manual segmentation, the *false negatives* (FN) are labelled as structure l in manual, but not in automatic segmentation, and finally the *true negatives* (TN) do not belong to structure l neither in the manual nor in automatic segmentation. Given these definitions, the Dice metric can be equivalently expressed as

$$D = \frac{2 * |\mathrm{TP}|}{2 * |\mathrm{TP}| + |\mathrm{FP}| + |\mathrm{FN}|}$$

When evaluating the segmentation using standard probabilistic atlases, only the three brain tissue classes are classified by the automatic segmentation. Consequently, the performance of the method in subcortical GM structures cannot be measured using Dice metric. However, we can evaluate the proportion of the subcortical GM structure in the manual segmentation classified as GM in automatic segmentation. Let T_{gt}^{sGM} denote the set of voxels belonging to a subcortical GM structure in the manual segmentation and T_{seg}^{GM} the set of voxels belonging

to GM in the automatic segmentation. The automatic segmentation of the subcortical GM structure is not known, therefore the following metric is used to evaluate the performance of the automatic segmentation:

$$D' = \frac{|T_{gt}^{sGM} \cap T_{seg}^{GM}|}{|T_{gt}^{sGM}|} \tag{4.2}$$

which is equivalent to

$$D' = \frac{|\mathrm{TP}|}{|\mathrm{TP}| + |\mathrm{FN}|}$$

This measure is used because the number of false positives is not known. The measure D' is often referred to as *sensitivity* of the method.

To provide the ground truth for evaluation of the algorithms for segmentation of brain tissues, the brain structure segmentations of cortex and cortical WM were manually adjusted to include all WM and GM tissue, respectively.

4.1.6 Implementation

The algorithms described in this and the following chapters were implemented as a part of IRTK software [72], which is implemented in C++ and runs on operating system Linux. The software contains libraries for image registration, processing and visualization. The segmentation and bias correction algorithms described in the chapters 4, 5 and 6 were implemented to create a brain segmentation module. If a large number of images has to be processed, the parallel system Condor at Department of Computing, Imperial College London, can be used to perform all the processes at the same time on a large number of computers.

4.2 Registration-based segmentation

In this section we investigate the performance of the registration-based segmentation for brain MRI during early childhood. Our registration method of choice, the non-rigid registration by Rueckert *et al.* [73] described in sec. 2.2.5, uses a highly flexible hierarchical B-spline model for

modelling local deformations, which allows detailed inter-subject matching of the brain MRI and is therefore especially suitable for registration-based segmentation of the brain.

4.2.1 Deformable atlas and registration-based segmentation

The manual segmentation of the brain structures of a two year old reference subject is used as a deformable atlas. Let us denote the reference subject X and manual segmentation of the reference subject S_X. The non-rigid alignment of the image to be segmented Y with the reference image X is denoted $\mathcal{T}_{Y \to X}$.

The registration-based segmentation of image Y, denoted $S_{rbs}(Y)$ can be obtained by eq. 2.2 using nearest neighbour interpolation:

$$S_{rbs}(Y)(v_i) = S_X(\text{round}(\mathcal{T}_{Y \to X}(v_i)))$$

where round(u) is a function finding the closest voxel to the location u. In our experience, better results can be obtained if the binary label map for each structure l

$$S_{X,l}(u_i) = \begin{cases} 1 & \text{if } S_X(u_i) = l \\ 0 & \text{if } S_X(u_i) \neq l \end{cases} \tag{4.3}$$

is transformed and resampled using linear interpolation. The label for each voxel is then calculated by choosing the structure with the highest value:

$$S_{rbs}(Y)(v_i) = \max_l S^{lin}_{X,l}(\mathcal{T}_{Y \to X}(v_i))$$

where function $S^{lin}_{X,l}$ denotes the linear interpolation of image $S_{X,l}$.

4.2.2 Non-rigid registration

Our aim is to find a highly accurate alignment of two brain MRI to achieve a good registration-based segmentation. The images have already been transformed to the same orientation in the pre-processing step and background voxels have been removed.

During the registration process, background voxels are ignored. This helps to reduce computational complexity and ignores intensities outside the head when evaluating the image similarity. However, this also means that the boundaries of the head are not necessarily well matched during the registration process. Therefore we do not ignore background voxels which are close to the head. This is achieved by extending the binary head mask using four dilations.

To initialize the alignment, an affine registration algorithm [85] is used. This registration uses normalized mutual information as a similarity measure. The affine registration is performed via gradient descent optimization at three resolution levels with isotropic voxels of size 6.4mm, 3.2mm and 1.6mm.

After affine registration, a non-rigid registration is performed at four resolution levels, with B-spline control point spacings of 20mm, 10mm, 5mm and 2.5mm. At each resolution, the image is first blurred and then subsampled to avoid local minima and optimize the speed of the algorithm. The quality and speed of the non-rigid registration depends on various parameters used at each resolution level: voxel size, kernel size of Gaussian blurring represented by standard deviation and step size when performing gradient descent. These parameters need to be determined to provide an optimal tradeoff between the quality and speed of registration-based segmentation.

The optimization at each resolution level starts by subdividing the B-spline transformation from previous level to halve the control point spacing. The gradient descent optimization is then performed four times with the step size for the finite difference approximation of the gradient halved every time.

4.2.3 Optimizing registration parameters

Our aim is to find the optimal tradeoff between the speed and the performance of the registration-based segmentation. If low resolution images are used at levels with larger control point spacings, the computations at these levels can be speeded up from hours to minutes without affecting the final results. We have compared four different strategies for the choice of the parameters. The four sets of parameters are shown in table 4.2. The agreement of registration-based segmentations with manual segmentations obtained when using those sets of parameters and measured with Dice metric (eq. 4.1) are shown in table 4.3.

Parameter set #1

control point spacing	image resolution	image blurring	finite difference step size
20	full	0	8
10	full	0	8
5	full	0	8
2.5	full	0	8

Parameter set #2

control point spacing	image resolution	image blurring	finite difference step size
20	full	0	4
10	full	0	4
5	full	0	4
2.5	full	0	4

Parameter set #3

control point spacing	image resolution	image blurring	finite difference step size
20	3	1.5	8
10	2	1	4
5	full	0.5	2
2.5	full	0	1

Parameter set #4

control point spacing	image resolution	image blurring	finite difference step size
20	3	1.5	8
10	2	1	8
5	full	0.5	8
2.5	full	0	8

Table 4.2: Parameters for the non-rigid registration. All sizes are given in millimeters.

The results in table 4.3 show, that downsampling the images in the first two resolution levels does not affect the segmentation results. Downsampling the first two resolution levels reduces the time needed for performing the registration by 30-40%. Additionally, a step size of 8mm

Parameter set	WM	Cortex	Caudate	Thalamus	Time
#1	0.83 ± 0.02	0.85 ± 0.01	0.86 ± 0.04	0.89 ± 0.01	11h
#2	0.82 ± 0.02	0.84 ± 0.01	0.86 ± 0.05	0.89 ± 0.01	10h
#3	0.76 ± 0.02	0.80 ± 0.01	0.82 ± 0.09	0.88 ± 0.02	6h
#4	0.83 ± 0.02	0.85 ± 0.01	0.86 ± 0.04	0.89 ± 0.01	7h

Table 4.3: Agreement of the manual segmentation with the results of the registration-based segmentation with different parameters given in table 4.2 as measured with the Dice overlap metric. The last column shows the approximate running times of the registration algorithm.

seems to produce good results. This finding suggests, that to obtain high quality registration of two brains, relatively large deformations are required even at the finest control point resolution to reflect the complexity of cortical folding. The final choice of registration parameters in shown in table 4.4.

control point spacing	image resolution	image blurring	finite difference step size
20	3	1.5	8
10	2	1	8
5	full	0.5	8
2.5	full	0	8

Table 4.4: Parameters for non-rigid registration. All sizes are given in millimeters.

4.3 EM-based segmentation

The EM algorithm is the most widely used framework for intensity-based segmentation of brain MRI. There are many variants of the EM-based segmentation framework. We chose the segmentation algorithm proposed by Van Leemput et al. [49] and as described in section 2.5.3. The original method incorporates bias correction and a MRF to include neighbourhood information, however, we will omit both, since in our case the images have been bias corrected in the pre-processing step by N3 method [78] and have good contrast-to-noise ratio. The EM segmentation algorithm proposed by Van Leemput et al. [49] is suitable for our purpose because it automatically estimates the means and variances of the Gaussian tissue PDF during the segmentation process, unlike the method by Wells [96], which requires prior knowledge of these parameters. To achieve the best possible performance of the EM segmentation method,

the brain volumes in all images are extracted using the Brain Extraction Tool (BET) [79], before the segmentation algorithm is executed.

4.3.1 Probabilistic atlas

The EM segmentation algorithm requires a probabilistic atlas to be registered by affine registration in the pre-processing step. The atlas commonly used for this purpose is the MNI305 probabilistic atlas [34] created from 305 brain MRI of healthy young adults acquired on a 1.5T MR scanner and aligned with a 9-parameter transformation (three rotations, three translations and three scalings), see fig. 4.3b.

As we are trying to segment brain MRI during early childhood, we have also tested the EM segmentation with the CCHMC5-9 probabilistic atlas [97], created from 49 brain MRI of children between five and nine years of age acquired on a 3T MR scanner. To create this atlas Wilke *et al.* aligned all the images into stereotaxic space with a 12-parameter affine transformation and segmented the images with SPM99 [4] using the MNI305 probabilistic atlas. A new probabilistic atlas was then created by averaging the resulting segmentations. The images were then segmented a second time using the newly obtained probabilistic atlas, to provide segmentation which is less biased towards prior information corresponding to anatomy of an adult brain. The segmentations obtained at the second pass were then averaged to create the CCHMC5-9 probabilistic atlas, see fig. 4.3c.

We use both atlases to segment the brain MRI during early childhood into four tissue classes - WM, GM, CSF and non-brain tissue. In our preliminary experiments we found that when either of these two atlases is used as a prior, the EM algorithm significantly overestimates WM in the central structures of the brain, see fig. 4.3e and f. A significant number of voxels are misclassified, for example in the thalamus, which is characterized by higher intensities on T1-weighted MRI than cortical GM. The intensity histogram calculated from a manual segmentation of a brain MRI of a two year old child (fig. 4.4) shows that the intensity distribution within the thalamus overlaps with the intensity distributions of WM and GM. A correct segmentation of this structure based on intensity alone is thus impossible. Instead prior information from a

probabilistic atlas must be used to segment this structure correctly. However, the anatomy of the adult brain differs from the brain during early childhood in this area. The internal capsule, which lies between subcortical GM structures, is thinner in young children than in adults, see the annotation in fig. 4.3a and d, and therefore the amount of WM in this area is overestimated (fig. 4.3e), if a probabilistic atlas derived from adult brains is used. Similarly, the prior information from the pediatric atlas CCHMC5-9 [97] resembles the anatomy of the adult brain, rather than anatomy of the brain during early childhood and thus leads to similar problems (fig. 4.3f).

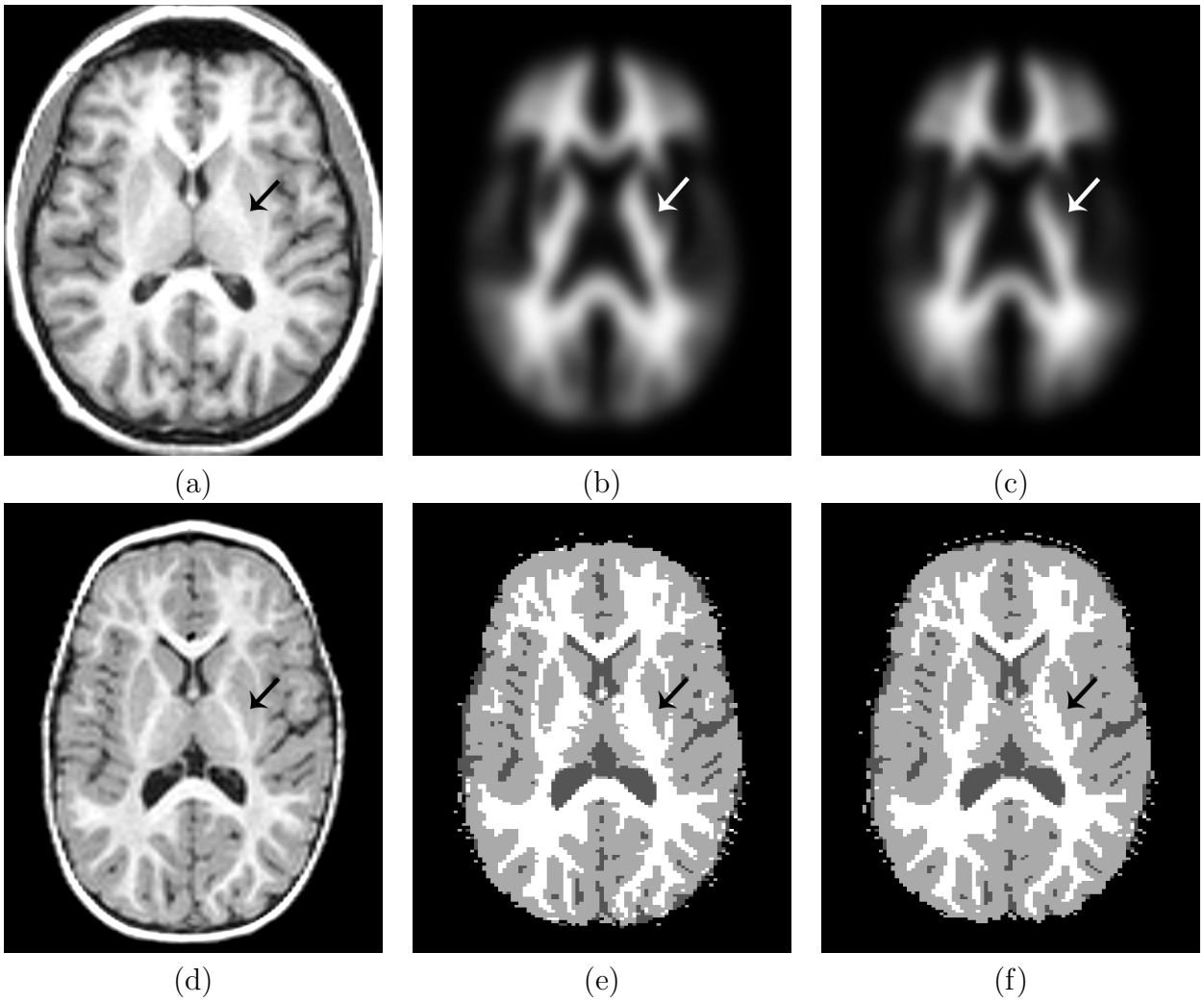


Figure 4.3: The anatomies of the brain in adulthood and during early childhood differ significantly in the subcortical area, with the internal capsule being larger in adults than in children: (a) the adult brain; (b) the MNI305 brain atlas (WM probability map); (c) the CCHMC5-9 brain atlas (WM probability map); (d) the brain of a two year old child; (e) misclassification of subcortical GM in a two year old child when the MNI305 atlas is used; (f) misclassification of subcortical GM in a two year old child when the CCHMC5-9 atlas is used.

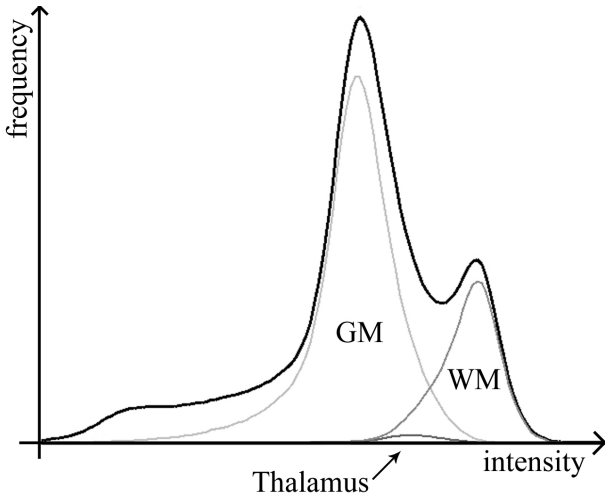


Figure 4.4: Histogram of the tissue intensity distributions for WM, GM and thalamus (a subcortical GM structure) based on a manual segmentation of brain MRI of a two year old child. The intensity distribution of the thalamus overlaps with those of WM and GM.

4.3.2 Creating a population-specific atlas for EM segmentation

To be able to segment the subcortical GM correctly, we have developed a technique to create a probabilistic atlas specifically for a population of subjects. When creating population-specific atlases, the major challenge is obtaining a sufficient number of manual segmentations since the manual delineation of the images is an extremely time-consuming process. Our method uses a single manual segmentation of a reference subject and a group of training images from a given population to build a population-specific probabilistic atlas, making the creation of probabilistic atlases for different age groups practically achievable.

To obtain good segmentations for a group of training subjects, we use a non-rigid registration-based segmentation instead of manual delineation. The results in table 4.3 show, that if suitable parameters are used, this method can produce a good quality segmentation for subcortical GM structures, where the brain during early childhood differs from the adult brain.

Let us consider the binary images of the manual segmentations of each structure $S_{X,1}, ..., S_{X,K}$ for the reference image X, as defined by eq. 4.3. Let us denote the group of training images $\{X_1, ..., X_q\}$ and the non-rigid transformations of the reference subject to the training images as $\mathcal{T}_{X_i \to X}$, $i = 1, ..., q$. Then the registration-based segmentations of the training images using

the manual segmentation of the reference image can be defined as

$$S_{rbs}(X_i, l) = S^{lin}_{X,l} \circ \mathcal{T}_{X_i \to X} \tag{4.4}$$

where $S^{lin}_{X,l}$ denotes the linear interpolation of the binary segmentation $S_{X,l}$. Once all the training images have been segmented, a probabilistic atlas can be created by aligning the training images with the affine registration to the space of the reference subject and averaging the segmentations. It is important to use only the global affine registration in this step to preserve the variation in shape of the brain structures. Let $\mathbf{p}^{atlas}_l$ denote the probabilistic atlas for structure l and $\mathcal{A}_{X \to X_i}$ the affine transformation of a training image X_i to reference image X. Then, in mathematical notation, the construction of the probabilistic atlas can be written as

$$\mathbf{p}^{atlas}_l(u_i) = \frac{1}{q} \sum_{i=1}^{q} S^{lin}_{X,l}(\mathcal{T}_{X_i \to X}(\mathcal{A}_{X \to X_i}(u_i)))$$

or in shorter notation

$$\mathbf{p}^{atlas}_l = \frac{1}{q} \sum_{i=1}^{q} S^{lin}_{X,l} \circ \mathcal{T}_{X_i \to X} \circ \mathcal{A}_{X \to X_i} \tag{4.5}$$

An overview of the proposed method is shown in Fig. 4.5.

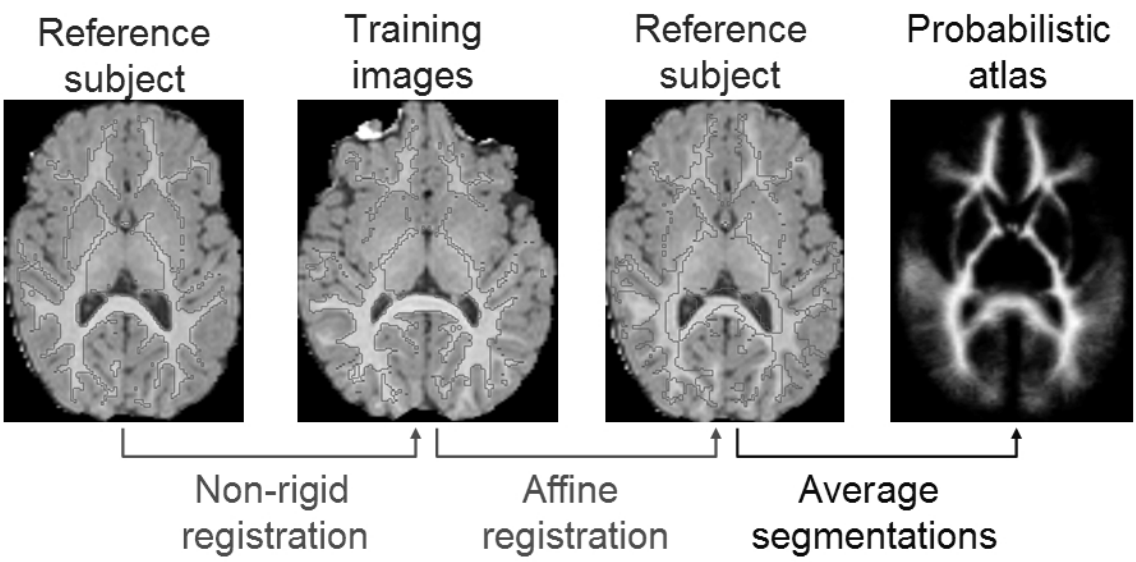


Figure 4.5: Creating a population-specific atlas: A single manual segmentation of a reference subject is transferred to a population of training subjects via non-rigid registration. All segmentations are then aligned to the reference subject via affine registration and averaged.

The probabilistic atlas transformed to the space of the image to be segmented Y using an affine transformation $\mathcal{A}_{Y\to X}$ serves then as the prior information $\mathbf{p}_l^{atlas}(Y)$ for EM-based segmentation of the image Y:

$$\mathbf{p}_l^{atlas}(Y) = \mathbf{p}_l^{atlas} \circ \mathcal{A}_{Y\to X}$$

Creating a good probabilistic atlas unbiased to the reference subject requires a good registration-based segmentation of the training images. In addition, the speed of these registrations is not crucial, as they are performed only once. We therefore choose the parameters given in table 4.4 for the registrations $\mathcal{T}_{X_i\to X}$, as derived in section 4.2.

4.3.3 Improvement of EM brain tissue segmentation using population-specific probabilistic atlas

Standard probabilistic atlases such as the MNI305 and CCHMC5-9 contain only probability maps for WM, GM, CSF and the brainmask. For comparison with these atlases we first created a population-specific atlas containing only these four probability maps. The atlas was created using the manually segmented two year old reference image and another 35 two year old subjects were used as training images described in sec. 4.1.1. The population-specific probabilistic atlas for two year old children was created using eq. 4.5. The probability map for WM is shown in fig. 4.6e. The white arrows points to internal capsule, showing one of the main differences of this population-specific atlas compared to the other two atlases, see fig. 4.6f,g. Fig. 4.7 shows a substantial reduction of the misclassification of voxels in the thalamus when the population-specific atlas is used to perform EM segmentation of a two year old subject. This improvement reflects the prior information from the probabilistic atlases shown in fig. 4.8. The segmentations of the whole axial slice are compared in fig. 4.6f-h. Even though there are visible differences in the area of the internal capsule, the overall performance of the EM segmentation using the different probabilistic atlases appears rather similar.

The quantitative evaluation (see table 4.5) confirms the conclusions obtained by visual inspections of the images. The overall improvement of Dice overlap in the segmentation of WM

and GM when using the population-specific atlas is only 0.02 - 0.03 compared MNI305 atlas, and only 0.01 compared to the CCHMC5-9 atlas. To highlight the more substantial improvement in subcortical area, we measured the proportion of volumes of caudate and thalamus that was classified as GM (eq. 4.2), as there are no automatic segmentations available for these structures (table 4.6). When the population-specific atlas was used, the fraction of volume of caudate segmented as GM increased by 0.07 compared to MNI305 atlas and 0.02 compared to CCHMC5-9 atlas, while improvement in thalamus is 0.31 and 0.22 respectively. The difference in performance in the caudate and thalamus is expected, as the caudate has relatively homogeneous intensity profile, while the voxel intensities in the thalamus vary greatly based on spatial location.

Atlas	WM	GM
MNI305	0.84 ± 0.01	0.87 ± 0.01
CCHMC5-9	0.86 ± 0.01	0.88 ± 0.01
Population-specific	0.87 ± 0.02	0.89 ± 0.01

Table 4.5: Agreement of the manual segmentations with the results of the EM segmentation into WM, GM and CSF using different probabilistic atlases measured with Dice metric.

Atlas	Caudate	Thalamus
MNI305	0.84 ± 0.05	0.51 ± 0.03
CCHMC5-9	0.89 ± 0.03	0.60 ± 0.04
Population-specific	0.91 ± 0.03	0.82 ± 0.02

Table 4.6: The proportion of subcortical GM structures caudate and thalamus classified as GM using EM segmentation into WM, GM and CSF with different probabilistic atlases.

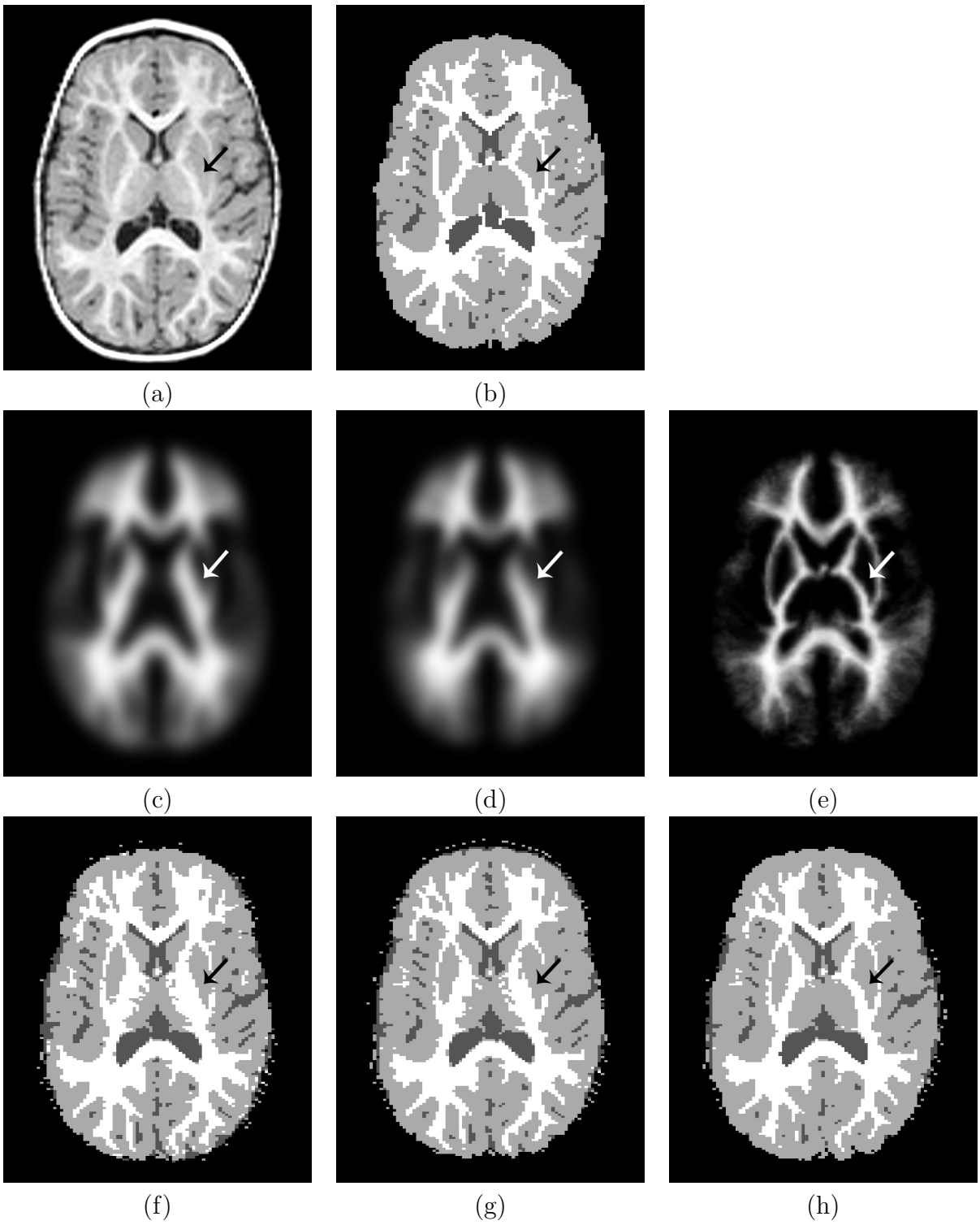


Figure 4.6: Comparison of the EM segmentations with different probabilistic atlases. (a) MRI of a two year old child. (b) Manual segmentation. WM probability maps for different probabilistic atlases: (c) MNI305 atlas; (d) CCHMC5-9 atlas; (e) population-specific atlas for the age of two years. EM segmentation into WM, GM and CSF using: (f) MNI305 atlas; (g) CCHMC5-9 atlas; (h) population-specific atlas for the age of two years. The arrow points at internal capsule showing different widths depending on the probabilistic atlas used.

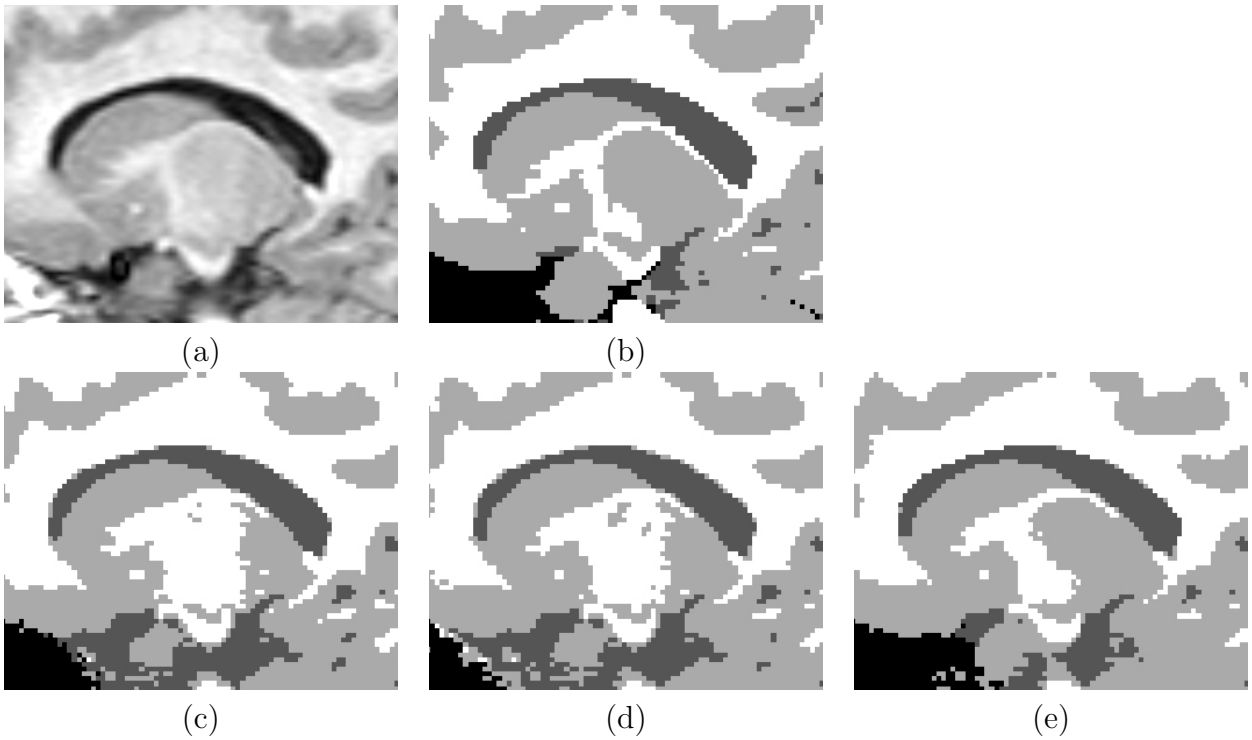


Figure 4.7: The amount of misclassification in the thalamus depends on the type of the probabilistic atlas used. (a) MRI of a two year old child. (b) Manual segmentation. EM segmentation using (c) MNI305 atlas; (d) CCHMC5-9 atlas; (e) population-specific atlas for the age of two years.

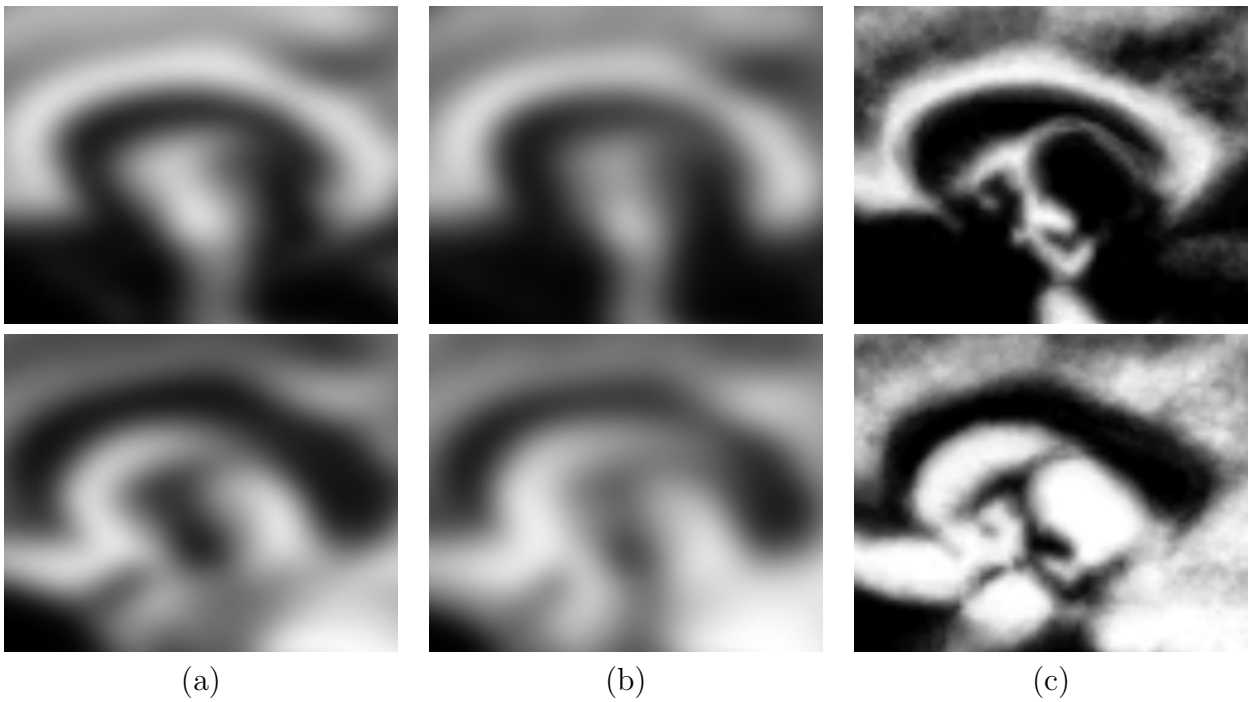
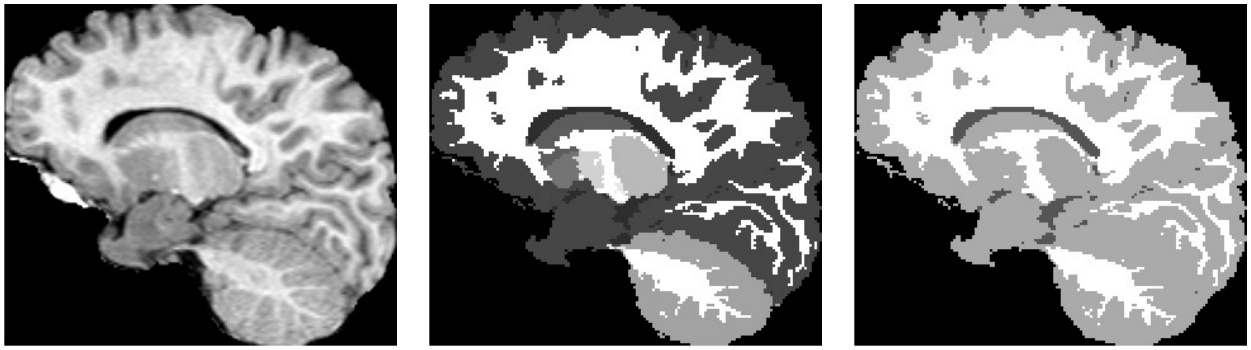


Figure 4.8: Comparison of different probabilistic atlases for the area of the brain shown in fig. 4.7. The probability maps for WM (the first row) and GM (the second row). (a) MNI305 atlas; (b) CCHMC5-9 atlas; (c) population-specific atlas for the of age two years.

4.3.4 EM segmentation of brain structures

The advantage of our method for creating probabilistic atlases is that any number of structures present in the manual segmentation of the reference subject can be used for the construction of the probabilistic atlas. In our case, the manual segmentation of a two year old subject was used to create a population-specific probabilistic atlas for the age of two years, containing 11 brain structures, as mentioned in sec. 4.1.4.

The results of the EM segmentation of these brain structures are shown in fig. 4.9b. For comparison, the segmentation of brain tissues is shown in fig. 4.9c.



(a) (b) (c)

Figure 4.9: Segmentation of a brain MRI of a two year old child: (a) the original MRI; (b) EM-based segmentation into 11 anatomical structures; (c) EM-based segmentation into three tissue classes.

The EM segmentation of the brain structures with the population-specific atlas for the age of two years was evaluated on four manual segmentations using the Dice metric, as described in sec. 4.1.5. The results are shown in table 4.7. We compared the segmentation with and without extracting the brain volume using BET [79] before the segmentation process. The results show, that using BET improves the segmentation results, especially for cortical GM. If the brain is not extracted prior to the segmentation process, many misclassifications occur around the boundary of the brain due to presence of non-brain tissue with intensities similar to GM.

method	WM	Cortex	Caudate	Thalamus
EM without BET	0.87 ± 0.02	0.87 ± 0.02	0.85 ± 0.04	0.86 ± 0.01
EM with BET	0.87 ± 0.02	0.89 ± 0.01	0.85 ± 0.04	0.87 ± 0.01

Table 4.7: Agreement of the manual segmentation with the results of the EM segmentation of brain structures using a population-specific atlas for the age of two years measured with Dice metric.

4.4 Combination of registration-based and EM-based segmentation

In the previous sections we described the registration-based and EM-based segmentation approaches using the population-specific deformable or probabilistic atlases. The registration-based approach performs well in the subcortical area of the brain, where shapes of the brain structures are relatively simple and the variability of the shapes is relatively small. In the cortical area, where WM and GM have complex shapes and there is no obvious one-to-one correspondence between two brains, the registration is rather error-prone and the EM-based segmentation is more suited. In contrast, the affine registration of a probabilistic atlas with an image is rarely sufficient to provide a good alignment between subcortical structures due to local shape differences, thus misguiding the segmentation process. The combination of both approaches is therefore desirable to produce results superior to both segmentation methods. In this section we propose a novel algorithm for segmentation of brain MRI using a combination of registration-based and EM-based approaches: Instead of creating a probabilistic atlas in the space of the reference subject, a subject-specific atlas is created for each new subject, providing prior information with a high degree of agreement with the anatomy of the brain to be segmented.

4.4.1 Creating a subject-specific atlas for EM segmentation

In line with the previous sections, let us denote the segmented reference image X and the corresponding manual segmentation S_X. The training images $X_1, ..., X_q$ are segmented using registration-based segmentation, producing segmentations $S_{rbs}(X_i)$, $i = 1, ..., q$. To provide

prior information adjusted to the anatomy of a subject better than affinely registered probabilistic atlas, we propose to register the training images directly to the new image using non-rigid registration denoted $\mathcal{T}_{Y \to X_i}$. The segmentations are then averaged to produce a subject-specific probabilistic atlas. This can be written in mathematical notation

$$\mathbf{p}_l^{atlas}(Y)(v_i) = \frac{1}{q} \sum_{j=1}^{q} S_{X,l}^{lin}(\mathcal{T}_{X_i \to X}(\mathcal{T}_{Y \to X_i}(v_i)))$$

or in shorter notation

$$\mathbf{p}_l^{atlas}(Y) = \frac{1}{q} \sum_{j=1}^{q} S_{X,l}^{lin} \circ \mathcal{T}_{X_i \to X} \circ \mathcal{T}_{Y \to X_i} \tag{4.6}$$

The resulting subject-specific probabilistic atlas can be used as a prior for EM-based segmentation without any further alignment, as opposed to population-specific atlas created in previous section. A schematic overview of this approach is shown in fig. 4.10.

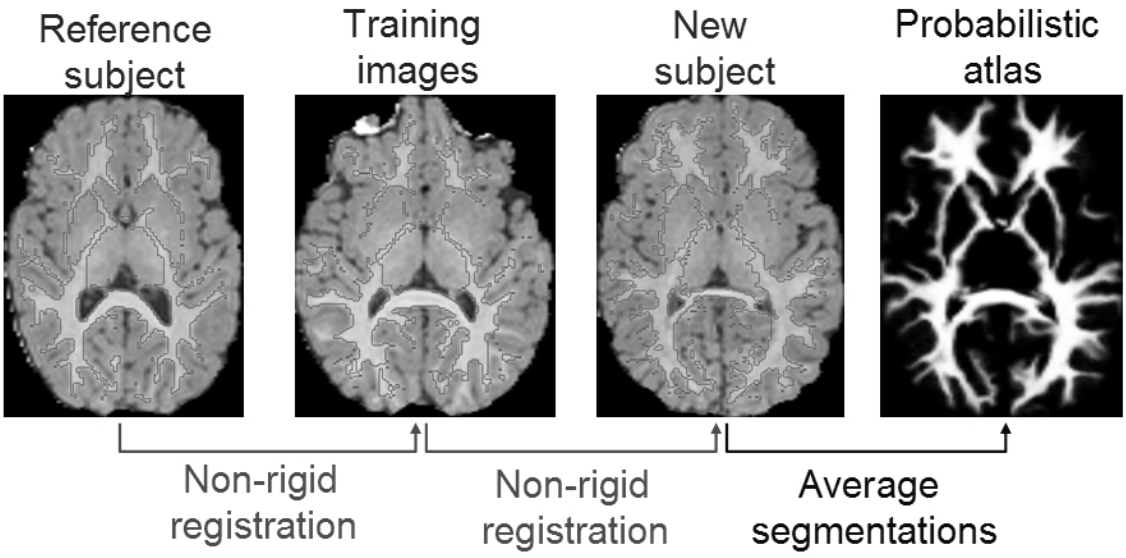


Figure 4.10: Creation of a subject-specific atlas: A manual segmentation of a reference subject is transferred to a population of subjects by non-rigid registration. All the segmentations are then aligned with the new subject to be segmented using non-rigid registration. In the final step the resulting registrations are averaged to create a subject-specific probabilistic atlas.

A subject-specific atlas is closely related to multiple label propagation described in sec. 2.2.6, eq. 2.3. Calculating a hard segmentation from a subject-specific atlas is equivalent to label fusion, with registration based segmentations of training images being used instead of man-

ual segmentations. Our segmentation based on multiple-label propagation can be therefore calculated using the following modified equation:

$$S_{rbs-multi}(Y(v_i)) = \arg\max_l \left\{ \mathbf{p}_l^{atlas}(Y)(v_i) \right\} \tag{4.7}$$

where $\mathbf{p}_l^{atlas}(Y)$ is defined using equation eq. 4.6.

4.4.2 Multiple registration of training images to a new subject

When calculating the transformations $\mathcal{T}_{X_i \to X}$ the speed of the registration is not crucial as these registrations are performed only once. Thus, we have used the set of parameters shown in table 4.4. However, the non-rigid registrations $\mathcal{T}_{Y \to X_i}$ have to be performed for each new subject Y and therefore computational speed is an issue which must be considered. Our aim is to find optimal trade off between speed and accuracy of registrations $\mathcal{T}_{Y \to X_i}$ while not compromising accuracy of the segmentation results.

Let us define the *precision of the subject specific atlas* as the final B-spline control point spacing of registrations $\mathcal{T}_{Y \to X_i}$. To find the optimal trade-off between speed and accuracy of the combined method, three subject specific atlases with different precisions were created for each subject:

Precision 10mm: The registrations $\mathcal{T}_{Y \to X_i}$ are performed with the first two resolution levels from table 4.4, 20mm and 10mm.

Precision 5mm: The registrations $\mathcal{T}_{Y \to X_i}$ are performed with the first three resolution levels from table 4.4, 20mm, 10mm and 5mm.

Precision 2.5mm: The registrations $\mathcal{T}_{Y \to X_i}$ are performed with all resolution levels from table 4.4, 20mm, 10mm, 5mm and 2.5mm.

The subject-specific atlases with these precisions were created using one reference subject (sec. 4.1.4) and 35 training images (sec. 4.1.1) using eq. 4.6.

The quality of the resulting atlases can be evaluated if the segmentation of the subjects is calculated using eq. 4.7 and compared to manual segmentations. The evaluation results on four manual segmentations presented in table 4.8 show, that it is vital to use precision 2.5mm to obtain good label-fusion results.

Atlas precision	WM	Cortex	Caudate	Thalamus	Time
10mm	0.76 ± 0.03	0.82 ± 0.01	0.83 ± 0.06	0.88 ± 0.01	0.5h
5mm	0.83 ± 0.02	0.85 ± 0.01	0.85 ± 0.06	0.89 ± 0.00	2.5h
2.5mm	0.85 ± 0.02	0.86 ± 0.01	0.88 ± 0.03	0.89 ± 0.00	7h

Table 4.8: Agreement of the manual segmentations with the results of the multiple label propagation with different levels of precision of the subject-specific probabilistic atlas measured with Dice metric. The last column shows an example of the approximate running times for registration of a training image to the image to be segmented.

The vote rule for creation of the segmentation from the subject-specific atlas can be replaced by EM segmentation proposed by Van Leemput [49] and presented in section 2.5.3, with the subject-specific atlas used as a prior. We hypothesized that lower precision of the atlas would be sufficient, as the EM segmentation would capture details of the brain anatomy based on intensity. The evaluation results of EM segmentation with subject-specific atlas with different precisions on four manual segmentations are shown in table 4.9.

Atlas precision	WM	Cortex	Caudate	Thalamus	Time
10mm	0.87 ± 0.02	0.89 ± 0.01	0.87 ± 0.02	0.88 ± 0.01	0.5h
5mm	0.88 ± 0.02	0.89 ± 0.01	0.87 ± 0.02	0.88 ± 0.01	2.5h
2.5mm	0.88 ± 0.02	0.90 ± 0.01	0.87 ± 0.02	0.88 ± 0.01	7h

Table 4.9: Agreement of the manual segmentations with the results of the combined registration-based and EM-based segmentation with different levels of precision of the subject-specific probabilistic atlas measured with Dice metric. The last column shows an example of the approximate running times for registration of a training image to the image to be segmented.

The evaluation results confirm that unlike in case of label fusion, the performance of the EM segmentation with subject-specific atlas drops only marginally if the final control point spacing for registrations $\mathcal{T}_{Y \to X_i}$ is increased from 2.5mm to 10mm. However, the computational speed increases sharply, speeding each registration $\mathcal{T}_{Y \to X_i}$ from several hours to only about 30 minutes. However, this is not the case when using label fusion, where precision 10mm brings substandard

results, especially in WM. Therefore EM segmentation with subject-specific atlas is superior to label-fusion.

For practical applications we chose EM segmentation with subject-specific probabilistic atlas with precision 10mm. Our final choice of parameters for the registrations $\mathcal{T}_{Y \to X_i}$ is shown in table 4.10. The cluster computing system Condor available at the Department of Computing, Imperial College London, was used to calculate all 35 registrations at the same time, making the algorithm computationally efficient enough to be practically usable.

control point spacing	image resolution	image blurring	finite difference step size
20	3	1.5	8
10	2	1	8

Table 4.10: Parameters for the non-rigid registrations during the second step of the creation of the subject-specific atlases. All sizes are given in millimeters.

4.4.3 Algorithm for combined registration-based and EM-based segmentation

In previous sections we proposed a method for creating a subject-specific probabilistic atlas and found the optimal registration parameters for efficient performance of the number of registration involved. Such subject-specific probabilistic atlas forms a suitable prior for EM-based segmentation, resulting in robust segmentation algorithm:

Combination of registration-based and EM-based segmentation

Calculate the registration-based segmentations $S_{rbs}(X_i, l)$ of training images $X_1, ..., X_q$ using a manual segmentation S_X of the reference subject X and B-spline non-rigid registrations $\mathcal{T}_{X_i \to X}$ with parameters in table 4.4:

$$S_{rbs}(X_i, l) = S^{lin}_{X,l} \circ \mathcal{T}_{X_i \to X}$$

where $S^{lin}_{X,l}$ is a linear interpolation of a binary manual segmentation of structure l in the reference image X.

For each new subject Y:

1. Transfer the registration-based segmentations $S_{rbs}(X_i, l)$ of the training images $X_1, ..., X_q$ to image Y using the non-rigid registrations $\mathcal{T}_{Y \to X_i}$ with parameters in table 4.10 and calculate the subject-specific atlas $\mathbf{p}_l^{atlas}(Y)$ as follows:

$$\mathbf{p}_l^{atlas}(Y) = \frac{1}{q} \sum_{i=1}^{q} S_{rbs}(X_i, l) \circ \mathcal{T}_{Y \to X_i}$$

2. Perform EM segmentation as described in sec. 2.5.3 using subject-specific probabilistic atlas created in step 1.

3. Calculate the hard segmentation for each voxel by choosing the label of the tissue with highest posterior probability.

Examples of the segmentations using the subject-specific probabilistic atlases as well as the probability maps for those atlases are shown in fig. 4.11.

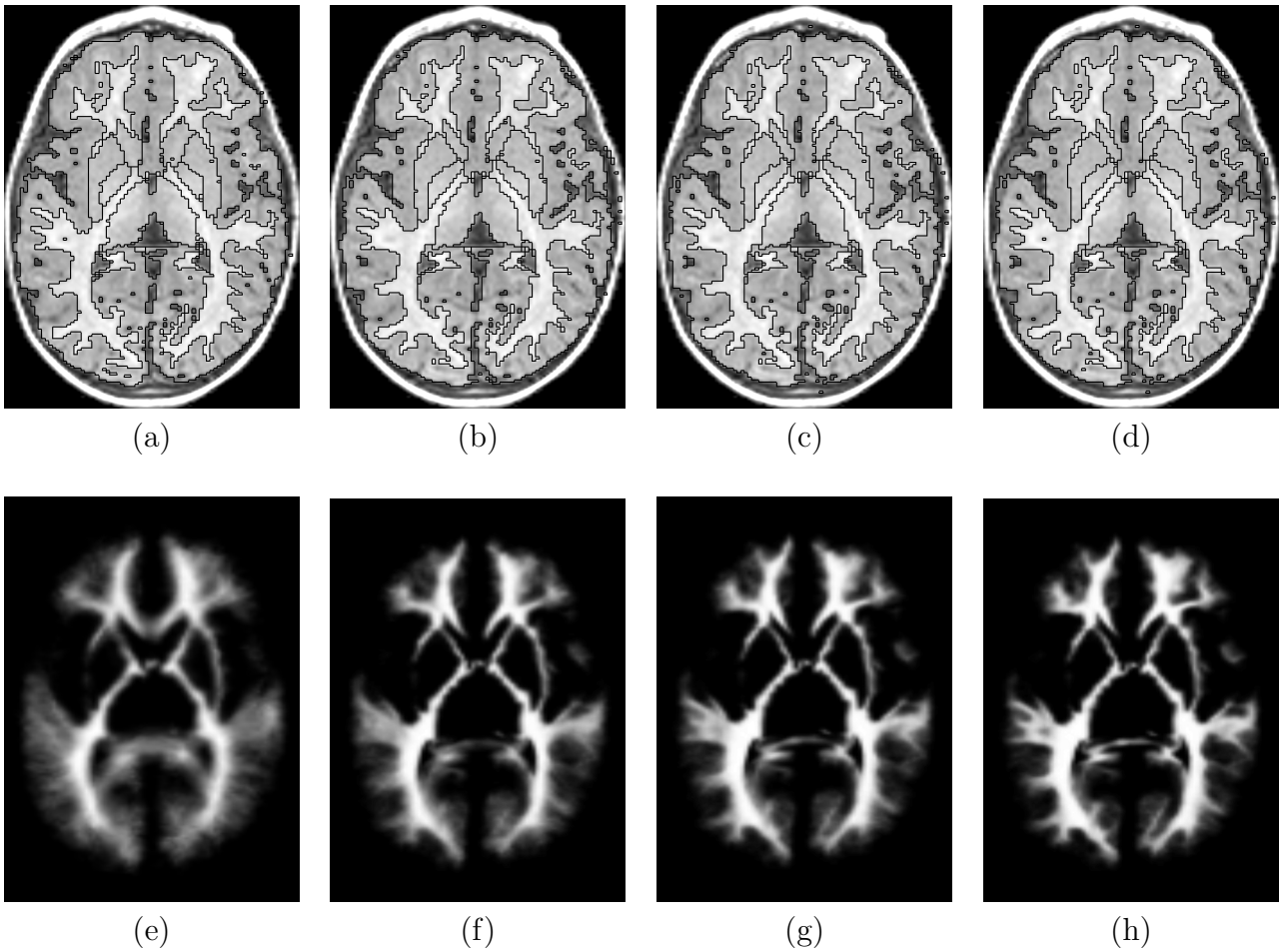


Figure 4.11: Comparison of EM-based and combined segmentations of brain structures: (a) EM segmentation with population-specific atlas; (b) Combined segmentation with subject-specific atlas with precision 10mm; (c) Combined segmentation with subject-specific atlas with precision 5mm; (d) Combined segmentation with subject-specific atlas with precision 2.5mm. Probability maps for WM in (e) population-specific atlas created with affine registration; (f) subject-specific atlas created with 10mm non-rigid registration; (g) subject-specific atlas created with 5mm non-rigid registration; (h) subject-specific atlas created with 2.5mm non-rigid registration.

4.5 Comparison of the methods

In this section we compare the performance of the three methods for the segmentation of the brain structures:

1. Registration-based segmentation with parameters from table 4.4 (sec. 4.2).
2. EM-based segmentation with population-specific atlas for two years of age, with the brain extracted by BET in the pre-processing step (sec. 4.3).
3. Combined registration-based and EM-based segmentation with the subject-specific atlas with precision 10mm (sec. 4.4).

In previous sections, all three methods were evaluated by measuring the agreement with the manual segmentations of four two year old subject using the Dice metric. In this section we summarize the evaluation results in table 4.11. The comparison shows that the aim of exploiting advantages of both, registration-based and EM-based segmentation, was achieved in the combined segmentation method. This method matches the performance of the EM-based segmentation in WM and GM, while performance in subcortical structures is comparable to registration-based segmentation, thus rendering the combined method superior to both original approaches.

Method	WM	Cortex	Caudate	Thalamus
Registration-based	0.83 ± 0.02	0.85 ± 0.01	0.86 ± 0.04	0.89 ± 0.01
EM-based	0.87 ± 0.02	0.89 ± 0.01	0.85 ± 0.04	0.87 ± 0.01
Combined	0.87 ± 0.02	0.89 ± 0.01	0.87 ± 0.02	0.88 ± 0.01

Table 4.11: Overlap between manual and automatic segmentation of different segmentation methods on four two year old subjects measured with Dice metric.

The advantages of the combined method can be also observed visually, see fig. 4.12. The black arrows highlights details in the image where the three segmentation methods perform differently. The small detail in the cortical area of the brain, the thin WM structures and small CSF-filled spaces, are mostly undetected by registration-based segmentation, improved by EM segmentation and best detected by combined segmentation. The boundary between

cortex and cerebellum, as well as the substantia nigra, is incorrectly delineated by the EM segmentation, as the affine registration of the probabilistic atlas is not flexible enough to align all smaller structures correctly. The registration-based and combined segmentations detect the boundaries of both structures correctly.

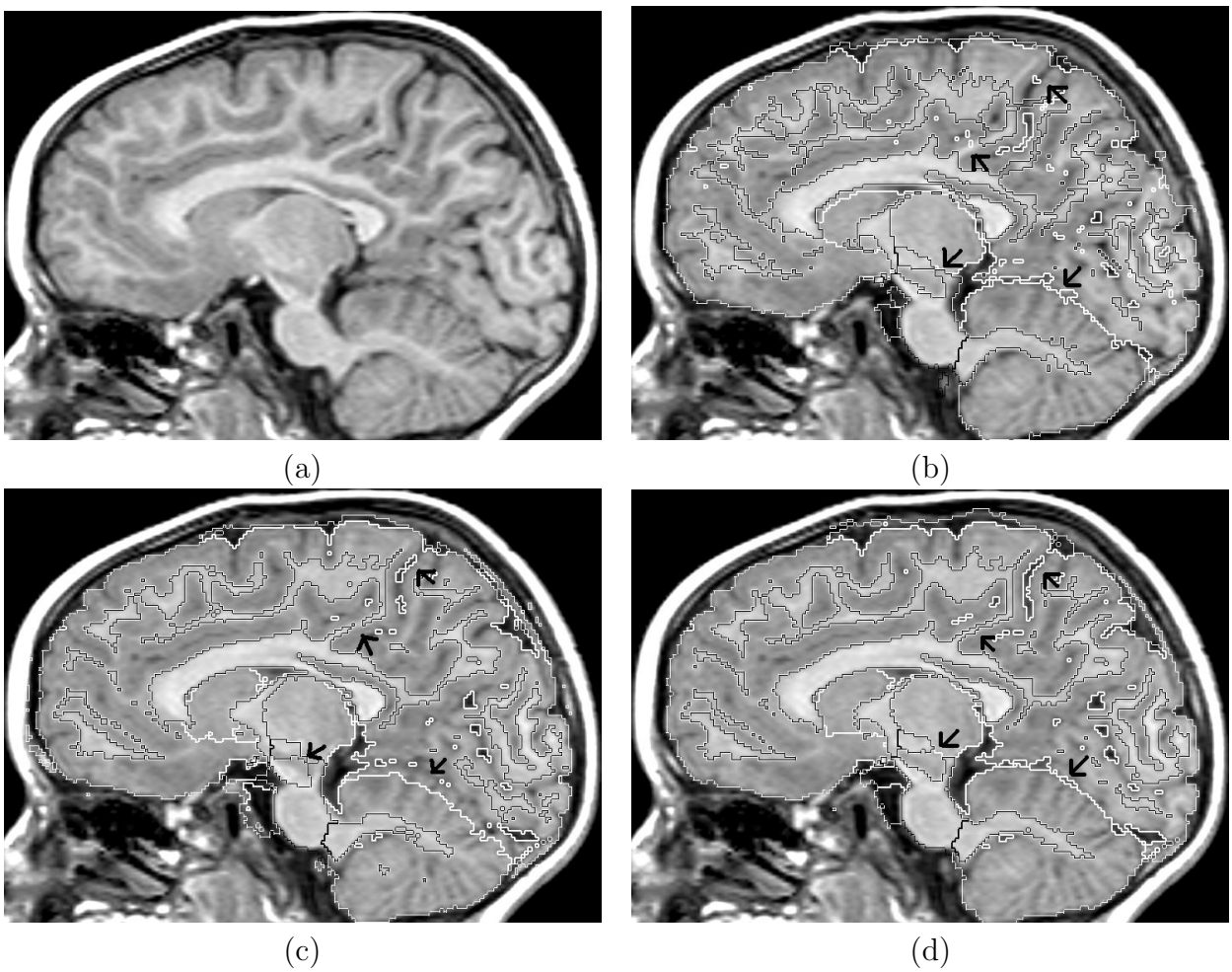


Figure 4.12: (a) Brain MRI of a two year old child. Automatic segmentation of brain structures using the three segmentation methods: (b) Registration-based segmentation; (c) EM-based segmentation; (d) Combined segmentation.
The arrows from top to bottom: 1. small CSF-filled space fails to be segmented by Registration-based segmentation (b), improved segmentation by EM segmentation (c), well delineated by combined segmentation. 2. Thin plate of WM fails to be segmented by registration-based segmentation (b), improved segmentation by EM segmentation (c), well delineated by combined segmentation. 3. Substantia nigra approximately delineated by registration-based segmentation (b), failed segmentation by EM segmentation (c), well delineated by combined segmentation. 4. Boundary of the cerebellum correctly delineated by registration-based segmentation (b), failed segmentation by EM segmentation (c), correctly delineated by combined segmentation.

As mentioned before, the EM-based segmentation is sensitive to the correct alignment of the probabilistic atlas. If only affine registration is used to align the probabilistic atlas to the image, the EM-based segmentation can fail if the shapes of the brain structures are significantly different from those in the probabilistic atlas. An example of a brain severely affected by focal lesions is shown in fig. 4.13. The EM segmentation completely fails to delineate the brain structures in this image (fig. 4.13c). However, the registration-based segmentation can accommodate for differences in brain shape and performs better (see fig. 4.13b). The same is true for the combined segmentation, see fig. 4.13d. Since the subject-specific atlas is produced by non-rigid registration it can better adapt to differences in the shape of the brains.

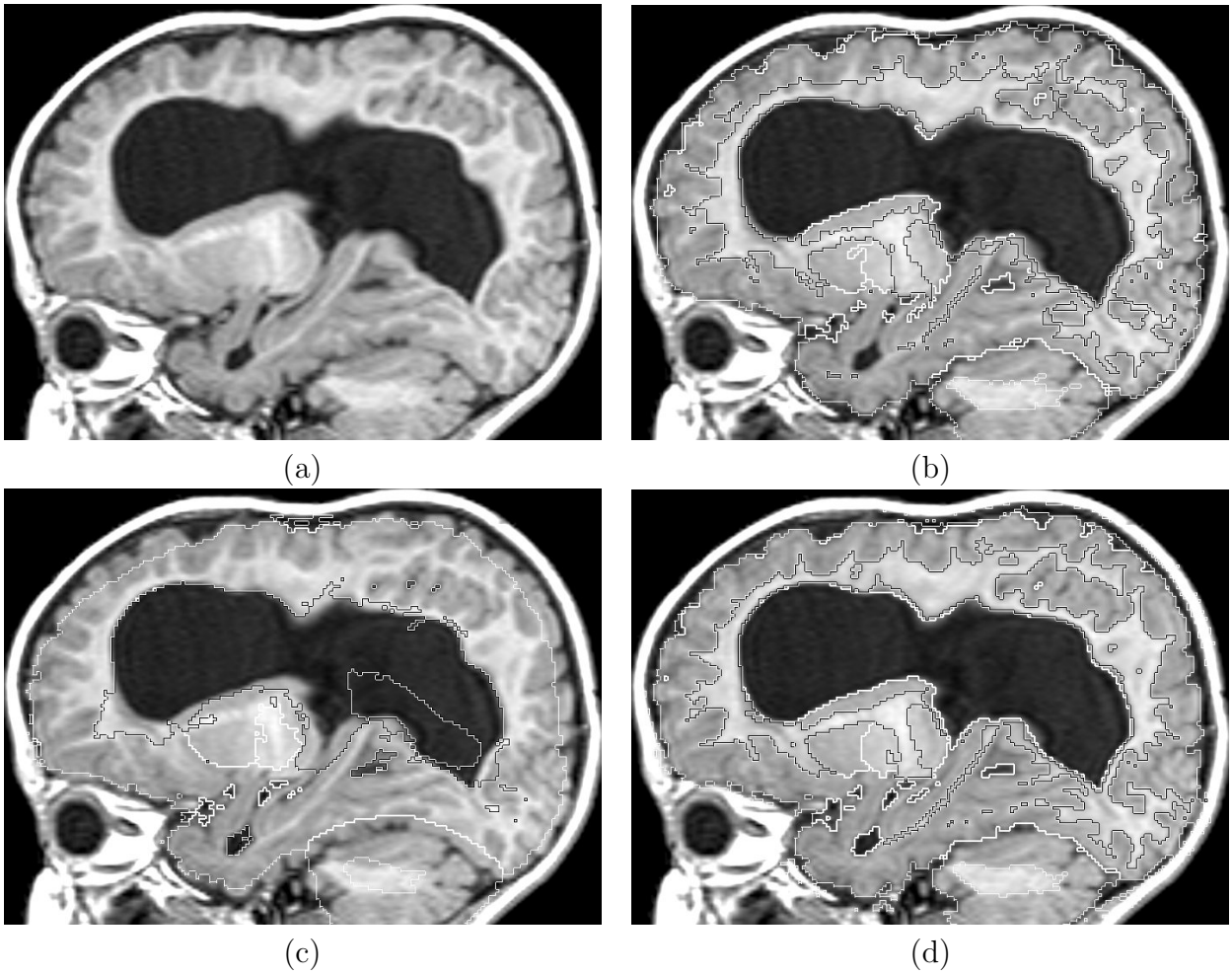


Figure 4.13: (a) Brain MRI of a two year old child severely affected by focal lesions. Automatic segmentation of brain structures using the three segmentation methods: (b) Registration-based segmentation; (c) EM-based segmentation; (d) Combined segmentation. The EM-based segmentation with an affinely aligned probabilistic atlas fails to adapt to the enlarged CSF-filled space.

4.6 Calculating volume and growth of brain structures during early childhood

In this section we demonstrate how the automatic segmentation methods can be used to determine the growth of brain structures during early childhood. To do this, brain MRI of 15 subjects with MR scans both at the age of one and two years were analyzed. All the subjects were born prematurely, with gestational ages at birth between 24 and 29 weeks. Ideally, the measurements should be compared to volumes of brain structures of full-term subjects. However we were not able to perform such experiments due to unavailability of MR scans of control subjects, which are very difficult to obtain as scanning of healthy children is not generally allowed.

To calculate the volumes and growth of 10 structures (excluding CSF), the combined registration-based and EM-based segmentation (sec. 4.4) was used, as it was shown to perform best during the evaluation in the previous section. The subject-specific atlases at two years were calculated using 35 training images of brain MRI at two years of age and one manual segmentation of the two years old reference subject. The reference subject was also scanned at one year of age and the manual segmentation was transferred to this earlier scan using non-rigid registration. We assumed that the registration of the same brain at different time-points would yield better results than the registration of the brains of different subjects. A registration-based segmentation of the reference subject at one year of age was then used to create the subject-specific atlases for scans of the 15 subjects at one year of age using 32 training images of the brain scanned at one year of age (see description of the data in sec. 4.1.1).

4.6.1 Measuring volumes of brain structures

The volumes of the brain structures were calculated by counting the number of voxels with the given label and multiplying by the volume of a voxel, $1.6 \times 1.035 \times 1.035 = 1.714\text{mm}^3$. The resulting volumes are shown in tables 4.12 and 4.13. The average volumes of the brain structures obtained from the 15 subjects are shown in fig. 4.14 and 4.16. When all the volumes are

corrected for overall brain volume, the resulting values can be interpreted as percentage of the brain volume. The mean relative sizes of the brain structures are shown in fig. 4.15 and 4.17.

Subject	Cortex	Caudate	Putamen	Nigra	Crb GM	Thalamus	Pallidum	Brainstem	WM	Crb WM
1	537.20	6.07	7.91	0.77	77.73	10.05	2.26	11.66	150.40	10.66
2	546.10	5.73	8.34	0.79	81.70	10.16	2.25	11.55	145.90	11.27
3	577.40	4.34	7.12	0.79	75.76	9.58	2.05	12.04	165.90	9.155
4	617.20	5.96	8.33	0.82	97.16	10.91	2.22	12.82	174.20	13.34
5	601.50	4.94	8.15	0.85	101.60	10.27	2.16	14.04	172.20	12.78
6	607.50	5.33	8.07	0.79	89.74	10.42	2.57	13.58	182.40	12.03
7	670.90	6.78	10.24	0.86	102.50	11.55	2.71	14.99	179.50	12.63
8	593.80	4.43	7.91	0.77	83.33	9.85	2.26	12.46	181.30	10.13
9	642.70	5.55	8.51	0.85	89.70	10.40	2.35	13.09	180.60	11.69
10	554.20	4.78	7.55	0.62	54.83	8.62	1.94	8.67	138.60	5.34
11	653.40	5.90	9.43	0.84	100.80	11.99	2.45	14.27	194.40	13.24
12	696.20	7.08	10.24	0.88	117.70	12.99	2.79	15.94	212.10	16.28
13	633.90	7.04	10.39	0.90	92.63	11.58	2.70	14.60	195.50	12.56
14	659.20	7.18	10.59	0.86	101.30	11.48	2.65	13.85	193.60	14.45
15	576.60	6.44	9.41	0.83	90.60	10.04	2.51	12.27	173.70	11.36
mean	611.19	5.84	8.81	0.81	90.47	10.66	2.39	13.06	176.02	11.80
std dev	47.91	0.95	1.14	0.07	14.79	1.10	0.26	1.76	19.86	2.48

Table 4.12: Volumes of the brain structures at the age of one year given in cm^3.

Subject	Cortex	Caudate	Putamen	Nigra	Crb GM	Thalamus	Pallidum	Brainstem	WM	Crb WM
1	620.70	7.23	8.85	0.87	92.27	11.66	2.59	14.11	207.50	13.40
2	620.30	6.77	9.60	0.86	96.11	11.69	2.79	14.17	206.90	13.22
3	615.10	4.43	7.40	0.86	84.05	10.31	2.24	13.64	192.30	10.12
4	689.80	6.71	9.67	0.93	113.40	12.35	2.75	16.14	240.40	15.19
5	646.20	5.58	9.17	0.99	112.80	11.23	2.55	16.81	225.20	15.18
6	662.80	6.01	9.22	0.92	100.60	11.49	3.02	16.10	240.90	14.24
7	744.40	7.55	10.65	0.98	116.60	12.92	2.81	18.21	240.70	14.38
8	653.70	4.76	8.53	0.85	88.84	11.11	2.62	14.61	233.40	12.25
9	699.70	6.25	9.84	0.87	98.58	11.77	2.88	15.05	247.40	13.60
10	604.30	5.43	8.20	0.68	59.95	9.61	2.23	10.23	190.50	6.71
11	723.80	6.87	10.80	0.99	115.10	13.77	3.00	17.08	262.00	16.31
12	780.70	7.97	11.48	1.07	135.90	14.42	3.59	19.54	296.70	19.40
13	710.30	8.15	11.76	1.06	107.40	12.92	3.41	17.98	261.20	14.20
14	730.60	8.15	12.41	0.99	119.60	12.83	3.34	17.03	257.50	17.42
15	629.00	7.33	9.86	0.95	99.57	11.04	2.79	14.20	227.10	13.43
mean	675.43	6.61	9.83	0.92	102.72	11.94	2.84	15.66	235.31	13.94
std dev	54.52	1.18	1.38	0.10	17.98	1.28	0.39	2.30	28.65	2.95

Table 4.13: Volumes of the brain structures at the age of two years given in cm^3.

All the subject used for these calculations were very preterm (defined as born before 32 weeks of gestation) and reduced volumes of GM structures such as cortex, cerebellum or subcortical GM structures are therefore expected [16], [52], [13]. An example can be seen in subject number 10 with extremely small cerebellum, forming only approximately 60% of average cerebellum size of the group and 70% after adjustment for the brain volume at both one and two years of age.

Even though the exact cause of this reduction is not known, this subject was born very early at 24 weeks of gestation, with very low birth weight only 610g. Additionally, the baby suffered intraventricular haemorrhage and diffuse white matter injury.

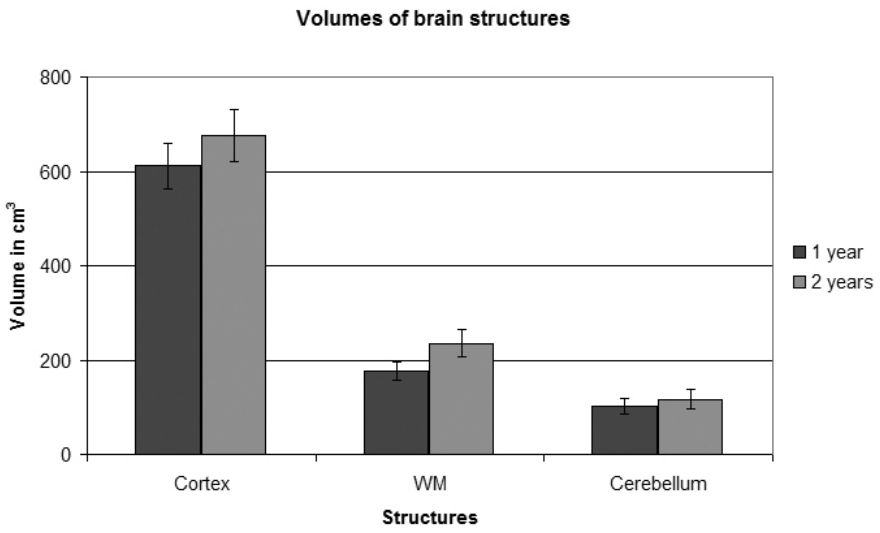


Figure 4.14: Mean volumes of brain structures at one and two years. The error bars are showing the standard deviation.

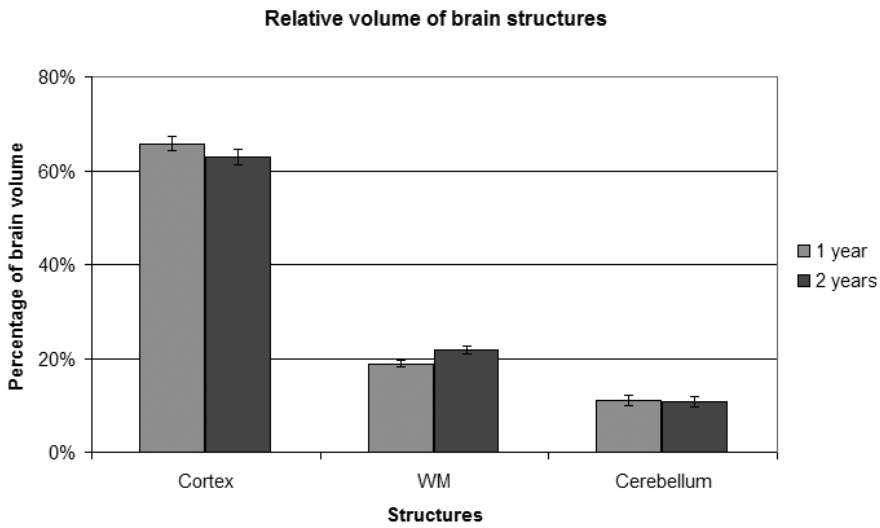


Figure 4.15: Relative mean volumes of brain structures at one and two years, given as the percentage of the whole brain volume. The error bars are showing the standard deviation.

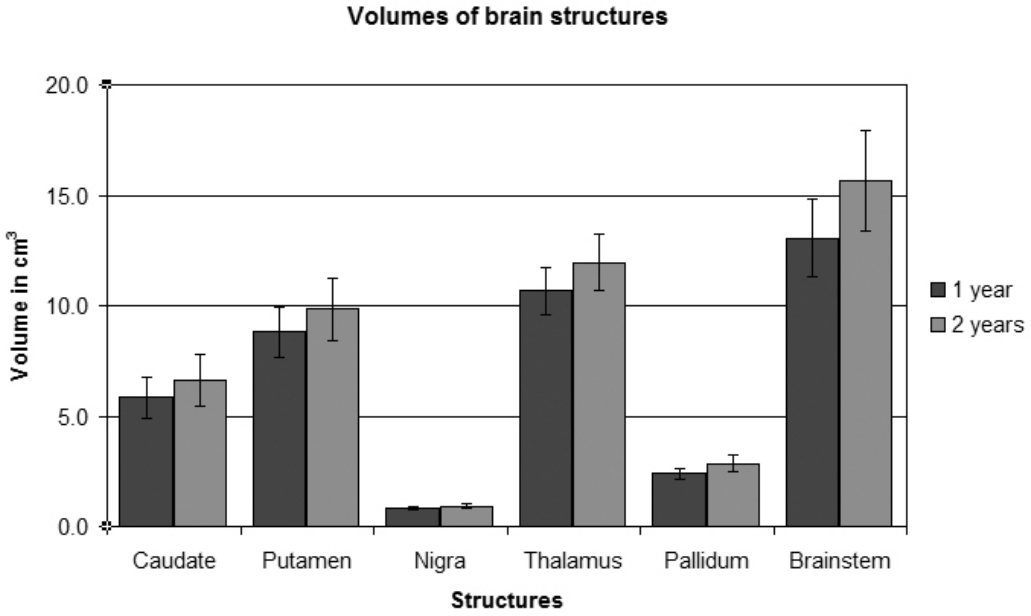


Figure 4.16: Mean volumes of brain structures at one and two years. The error bars are showing the standard deviation.

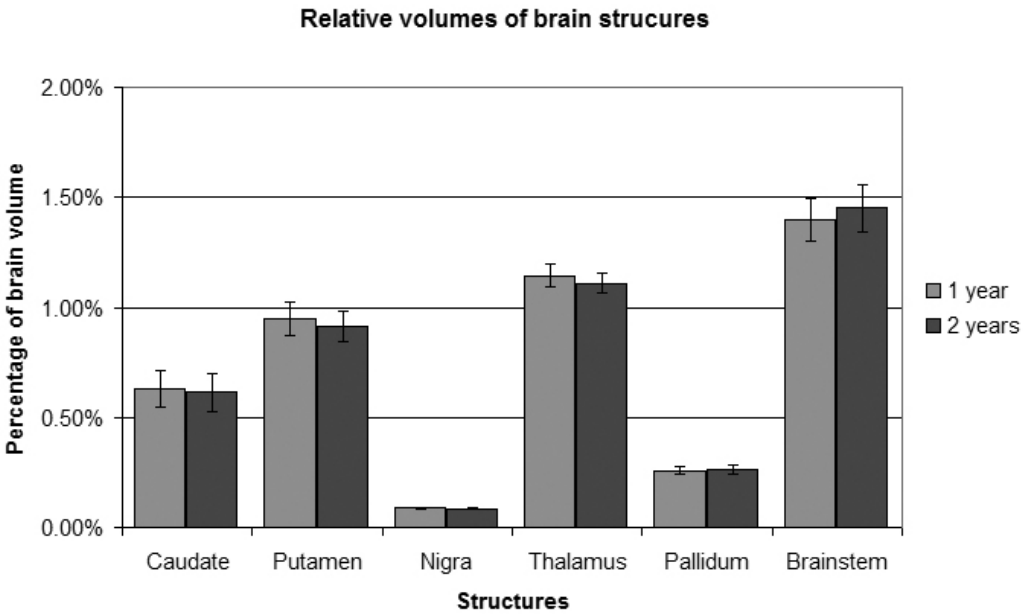


Figure 4.17: Relative mean volumes of brain structures at one and two years, given as the percentage of the whole brain volume. The error bars are showing the standard deviation.

4.6.2 Measuring growth of brain structures

The growth of the brain structures between one and two years of age was calculated for each of the 15 subjects by dividing the volume at two years by the volume at one year. These growth factors are shown in table 4.14. Average growth factors for each structure are shown in fig. 4.18.

Subject	Cortex	Caudate	Putamen	Nigra	Crb GM	Thalamus	Pallidum	Brainstem	WM	Crb WM
1	1.16	1.19	1.12	1.13	1.19	1.16	1.15	1.21	1.38	1.26
2	1.14	1.18	1.15	1.09	1.18	1.15	1.24	1.23	1.42	1.17
3	1.07	1.02	1.04	1.10	1.11	1.08	1.09	1.13	1.16	1.11
4	1.12	1.13	1.16	1.13	1.17	1.13	1.24	1.26	1.38	1.14
5	1.07	1.13	1.12	1.17	1.11	1.09	1.18	1.20	1.31	1.19
6	1.09	1.13	1.14	1.16	1.12	1.10	1.18	1.19	1.32	1.18
7	1.11	1.11	1.04	1.14	1.14	1.12	1.04	1.21	1.34	1.14
8	1.10	1.08	1.08	1.10	1.07	1.13	1.16	1.17	1.29	1.21
9	1.09	1.13	1.16	1.03	1.10	1.13	1.23	1.15	1.37	1.16
10	1.09	1.14	1.09	1.10	1.09	1.11	1.15	1.18	1.37	1.24
11	1.11	1.17	1.15	1.18	1.14	1.15	1.22	1.20	1.35	1.23
12	1.12	1.13	1.12	1.22	1.15	1.11	1.29	1.23	1.40	1.19
13	1.12	1.16	1.13	1.18	1.16	1.12	1.26	1.23	1.34	1.13
14	1.11	1.14	1.17	1.14	1.18	1.12	1.26	1.23	1.33	1.21
15	1.09	1.14	1.05	1.15	1.10	1.10	1.11	1.16	1.31	1.18
mean	1.11	1.13	1.11	1.13	1.13	1.12	1.19	1.20	1.34	1.18
std dev	0.02	0.04	0.05	0.05	0.04	0.02	0.07	0.04	0.06	0.04

Table 4.14: Growth factors between the age of one and two years.

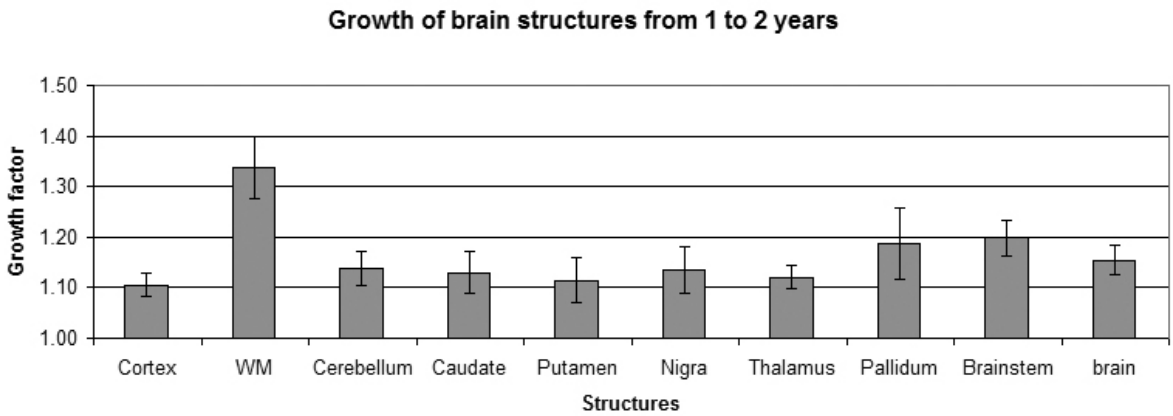


Figure 4.18: Mean growth factors from one to two years for brain structures. The error bars are showing the standard deviation.

4.6.3 Measuring growth of brain tissues

The growth of WM and GM tissue volumes for all 15 subjects is shown in fig. 4.19, with the age at the time of the scan on the horizontal axis and the volume of WM and GM on the vertical axis. The values for the WM and GM volumes from the same subject are drawn with a line to visualize the growth. The WM tissue is composed of WM, cerebellar WM and brainstem, while GM tissue is composed of cortex, cerebellar GM and five subcortical GM structures (caudate, putamen, substantia nigra, thalamus and pallidum), as previously presented in table 4.1.

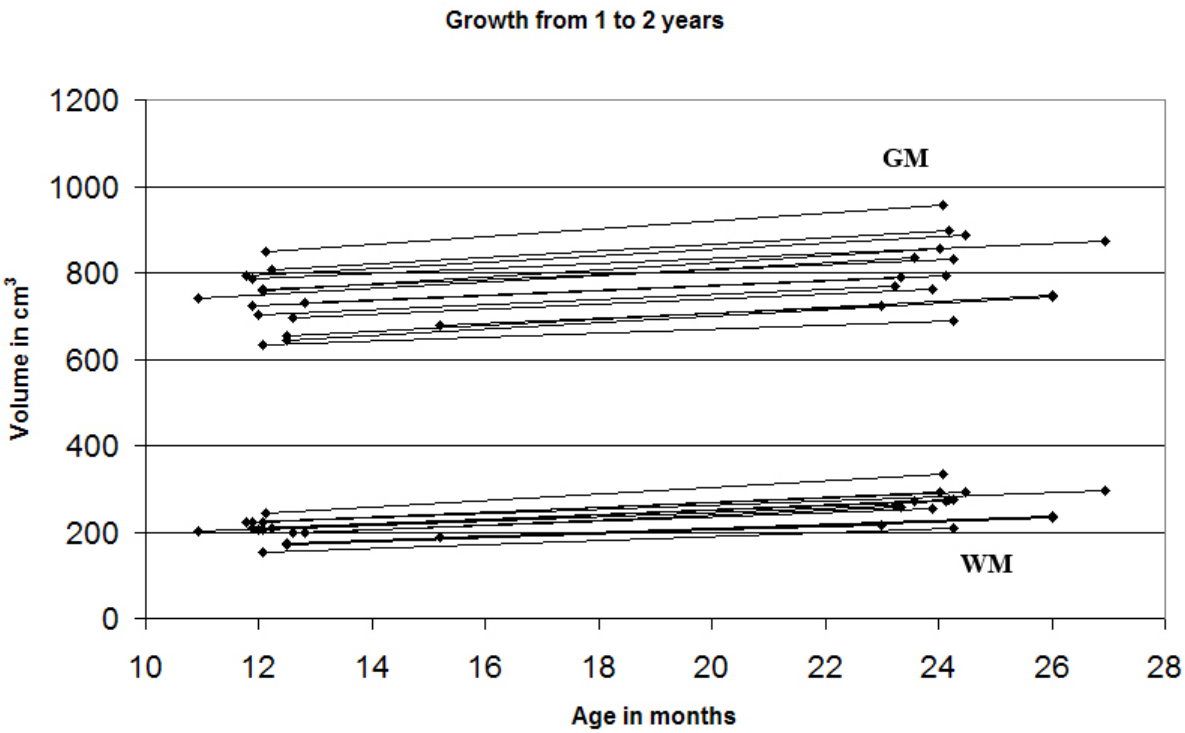


Figure 4.19: Changes in the volume of WM and GM between the age of one and two years.

The average growth factors for WM and GM are shown in table 4.15 and compared with results of the study by Aljabar *et al.* [1], who computed the average growth factors of WM and GM from one to two years from 29 subjects using deformation-based morphometry and EM-segmentation with a population-specific atlas. The 15 subjects used in our experiments were also used in Aljabar's study.

method	WM	GM
Combined segmentation	1.32 ± 0.06	1.11 ± 0.02
EM segmentation (Aljabar)	1.28 ± 0.05	1.12 ± 0.03
Deformation-based morphometry (Aljabar)	1.26 ± 0.05	1.11 ± 0.03

Table 4.15: Mean growth factors for WM and GM between one and two years estimated by combined segmentation and compared to the results obtained by Aljabar *et al.* [1] using EM segmentation with a population-specific probabilistic atlas and deformation based morphometry.

Finally, the relative volume of WM and GM is shown in fig. 4.20 as a function of the age at the time of the scan. The plot shows the increase of the proportion of WM and decrease of the proportion of GM between ages one to two years.

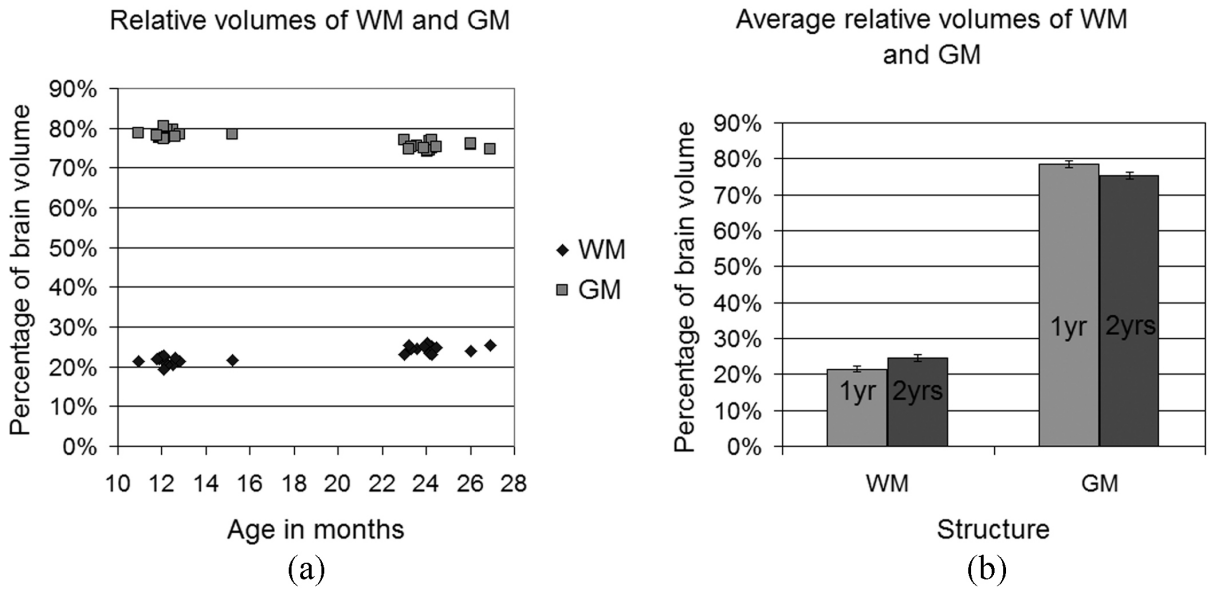


Figure 4.20: (a) Relative volumes of WM and GM tissues shown as a function of age. (b) Average relative volumes of WM and GM tissues at approximate ages of one and two years. The percentage of WM increases, while the percentage of GM decreases between ages one and two years.

4.7 Discussion

In this chapter we examined registration-based and intensity-based approaches for the segmentation of brain MRI of one and two year old children. We also proposed a novel segmentation method which combines the advantages of these two approaches. In the proposed method, the registration-based segmentation is used to transfer a manual segmentation of a reference subject to a set of training images and these segmentations are then transferred to a new subject using non-rigid registration and averaged. The resulting subject-specific probabilistic atlas is in precise alignment with the new image thanks to the flexibility of the non-rigid registration and robustness of multiple-registration approach. The boundaries of the structures are then refined using EM-based segmentation.

To develop robust tools for the segmentation of brain MRI during early childhood, we selected the non-rigid registration algorithm proposed by Rueckert [73] and the EM-based segmentation algorithm by Van Leemput [48]. However, the findings presented in this chapter as well as developed methodology could be generalized to other similar registration and intensity-based segmentation techniques.

In sec. 4.5 we concluded, that in registration-based segmentation the incorrect matching of cortical WM and GM remains a consistent source of errors. In his comparative study, Crum *et al.* [25] conclude that all tested non-rigid registration approaches struggled to match small structures with complicated shapes correctly. In particular the large inter-subject variability of the cortical anatomy remains a severe challenge for registration algorithms. In this study, the non-rigid registration based on B-splines [73] outperforms other approaches such as fluid registration in cortical regions of the brain.

In sec. 4.3.1 we have shown that intensity-based segmentation is prone to classification errors in subcortical region. The reason for this stems from the theoretical model adopted in current intensity-based approaches: The intensities within each tissue class are considered to be homogeneous and only corrupted by Gaussian noise. In practice, the tissue intensity is spatially variable in some areas, especially in the central brain structures, as illustrated in fig. 4.21.

Classification errors resulting from the variation of the intensity pattern within WM and GM were reported by [71] where the brain MRI of adults were segmented using FAST [102]. We observed similar misclassifications in the brain MRI during early childhood in our experiments. The classification error in these approaches often depends on the quality of alignment of the probabilistic atlas and the correction of the intensity inhomogeneity.

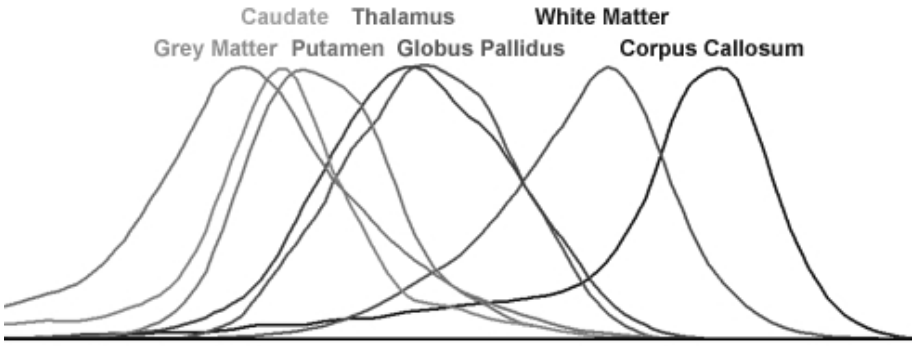
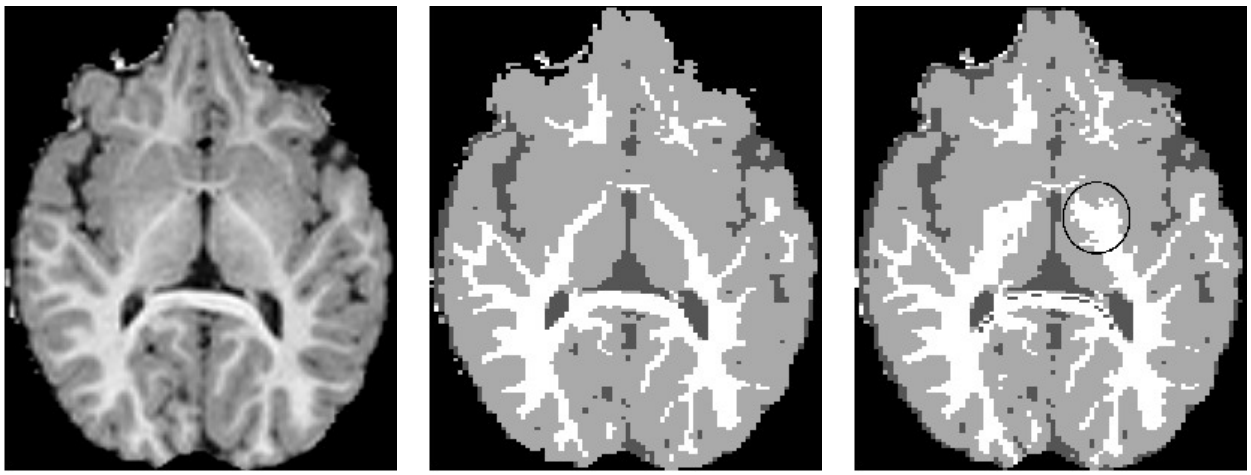


Figure 4.21: Normalized histograms brain structures. The intensities are brighter for subcortical GM structures (caudate, putamen, thalamus, pallidum) as well as subcortical WM structure corpus callosum, compared to cortical GM and WM, respectively.

When the probabilistic atlas is registered using affine alignment, classification errors will occur in subjects which differ from the average anatomy. It has been suggested to use non-rigid alignment of the probabilistic atlas to solve this problem [68], [7]. Unfortunately, if the flexibility of the registration increases, the potential for registration errors increases as well. Even though SPM5 [7] uses non-rigid alignment of probabilistic atlas of the adult brain, it misclassifies the central brain structures in young children similarly to EMS [48] (see fig. 4.22c). In spite of the fact that the non-rigid alignment of the probabilistic atlas does not seem to solve the misclassifications in subcortical GM, it is still necessary when dealing with anatomy substantially different from the anatomy in the atlas. This is therefore a promising area for future developments for segmentation algorithms of brain MRI. However, fig. 4.22c suggests that it might be still necessary to use population-specific atlases to guide the segmentation process correctly.

In our novel combined registration-based and EM-based segmentation method, we successfully addressed the causes of misclassification in standard state-of-the-art segmentation approaches. To address the issue of intensity variation within brain structures, we proposed a method for



(a) (b) (c)

Figure 4.22: (a) The original MR image of a two year old child. Segmentation of WM: (b) EM with population-specific atlas; (c) SPM5 [7] with non-rigidly aligned standard adult atlas; Parts of thalamus and pallidum are misclassified as WM (e.g. in the region in black circle).

creating spatial prior corresponding very well the new subject anatomy, referred to as subject-specific atlas. As the subject-specific atlas is created using multiple non-rigid registrations with the new subject, robust non-rigid alignment of the prior with the image is achieved. The final EM segmentation step ensures a good quality classification of the complicated cortical structures based on intensity. These attributes render it as potentially suitable tool for robust evaluation of growth patterns during early childhood as we demonstrated in the growth study presented in sec. 4.6.

4.8 Conclusion

The aim of this chapter was to develop a robust tool for segmentation of brain images during early childhood. We developed a novel algorithm based on combination of two state-of-the-art approaches, B-spline registration-based and EM-based segmentation. The usability of the method was demonstrated by calculating volumes and growth of brain structures during early childhood.

The limitations of this study include the limited number of manual segmentations of the brain MRI during early childhood used in the evaluation. We used only four images of brain at

two years of age segmented in 6-8 slices to evaluate the methods. However, the results of this evaluation were consistent and confirmed by visual inspection of larger number of subject, as well as consistent volume and growth measurements.

In this chapter we did not address the issue of intensity inhomogeneity. However, the quality of bias correction affects the quality of segmentation results, especially for the modern MRI scanned at higher field strengths. Chapters 5 and 6 therefore address this issue in more detail.

Chapter 5

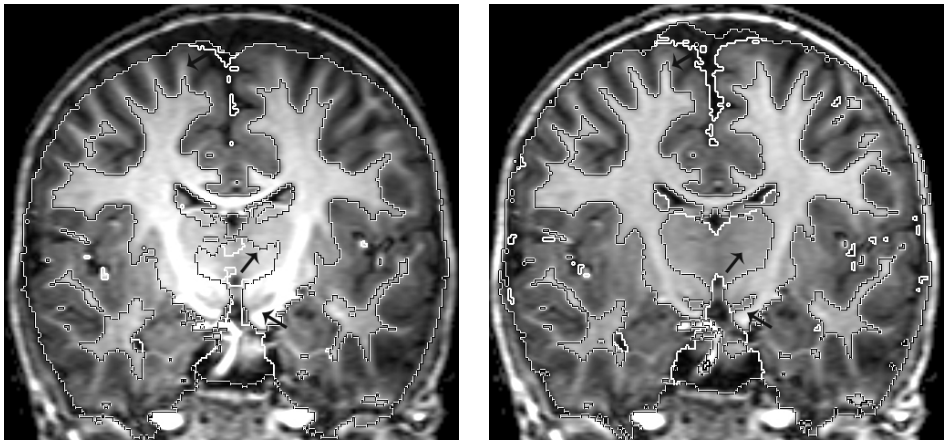
Intensity inhomogeneity correction of brain MR images

5.1 Processing brain MRI in presence of strong intensity inhomogeneity

In the previous chapter we have developed a methodology for the segmentation of brain MRI during early childhood. We have shown that this methodology can be used to assess the morphology of the brain during early childhood as well as changes or growth during this phase of brain development. The data used in the previous chapter was acquired using a MR scanner with a relatively low field strength (1 Tesla). In these datasets the effect of intensity inhomogeneity is relatively weak and can be corrected using standard techniques such as N3 [78]. However, MR scanners with higher field strength are becoming increasingly widespread and the effect of intensity inhomogeneity becomes much stronger at higher field strength. The data used in this chapter have been acquired using a 3 Tesla MR scanner which is now the standard scanner for imaging the brain during early childhood at Hammersmith Hospital. In these images the contrast resulting from the slowly varying bias field is larger than the contrast between tissue classes. In our experience, such strong intensity inhomogeneity cannot be appropriately

corrected using methods such as N3 [78].

A good intensity inhomogeneity correction is essential for obtaining an accurate segmentation. Since the amount of intensity inhomogeneity across different tissue classes directly correlates with the performance of intensity-based segmentation methods, such segmentation methods can be used to asses the performance of different intensity inhomogeneity correction methods. Figure 5.1 shows the effect of strong intensity inhomogeneity on intensity-based segmentation methods such as the EM-based segmentation described in the previous chapter.



(a) (b)

Figure 5.1: A EM-based segmentation algorithm yields poor results in the presence of strong intensity inhomogeneity: (a) a segmentation of an image with strong intensity inhomogeneity; (b) an improved segmentation after the intensity inhomogeneity has been corrected. The arrows point at some of the locations where the segmentation has been improved after intensity inhomogeneity correction. The bias field has been corrected with a novel template-based method described later in this chapter.

The presence of intensity inhomogeneity does not only have a detrimental effect on the image segmentation. For example, voxel-based similarity measures for non-rigid registration are also sensitive to intensity inhomogeneity. Whereas the registration algorithm of Rueckert et al. [73] performs very well on images with weak intensity inhomogeneity, it often fails on images with strong intensity inhomogeneity as it is guided by the gradient of the bias field rather than by intensity gradients resulting from the brain anatomy. Figure 5.2 shows the effect of the strong intensity inhomogeneity on the non-rigid registration algorithm.

In this chapter we will develop a novel template-based intensity inhomogeneity correction method. The proposed method assumes that a template image without bias is available. First,

the template image not containing any bias field is aligned with the image to be corrected. To avoid misregistration due to the bias field, we align the template and the image only with affine registration. The voxel pairs in both images are then classified as inliers or outliers, depending whether they are likely to belong to the same tissue type. The classification is performed within the EM framework and interleaved with the estimation of the bias field. Voxel pairs with higher probabilities of belonging to the same tissue are assigned higher weights during the bias field estimation.

When estimating the bias field, the key property that helps to distinguish influence of the bias field from other artefacts in the MRI is its low-frequency nature. The low frequency of the B-spline model for the bias field can be ensured in two different ways: An *intrinsic* low-frequency nature of the model can be achieved using large B-spline control points spacing, which is equivalent to small number of B-spline control points. This is similar to approach proposed by Van Leemput [48], who uses polynomial basis functions (see sec. 3.3.1) to model the bias field. Alternatively, the *extrinsic* low-frequency nature can be modelled by imposing a smoothness constraint on the bias field during optimization process. In this way more B-spline control points can be used while preserving the low-frequency nature of the bias field, making the model less biased towards the shape of the basis functions. This approach is similar to the bias field model in SPM5 [7], see section 3.4.2, which uses discrete cosine basis functions.

The advantage of the proposed method is that a closed-form solution for the bias field estimation can be obtained if all the image intensities have been logarithmically transformed, thanks to using B-spline as the model for bias field. Moreover, the template-based bias correction can be used in conjunction with image registration, thus producing a registration algorithm which is robust even in the presence of strong intensity inhomogeneity.

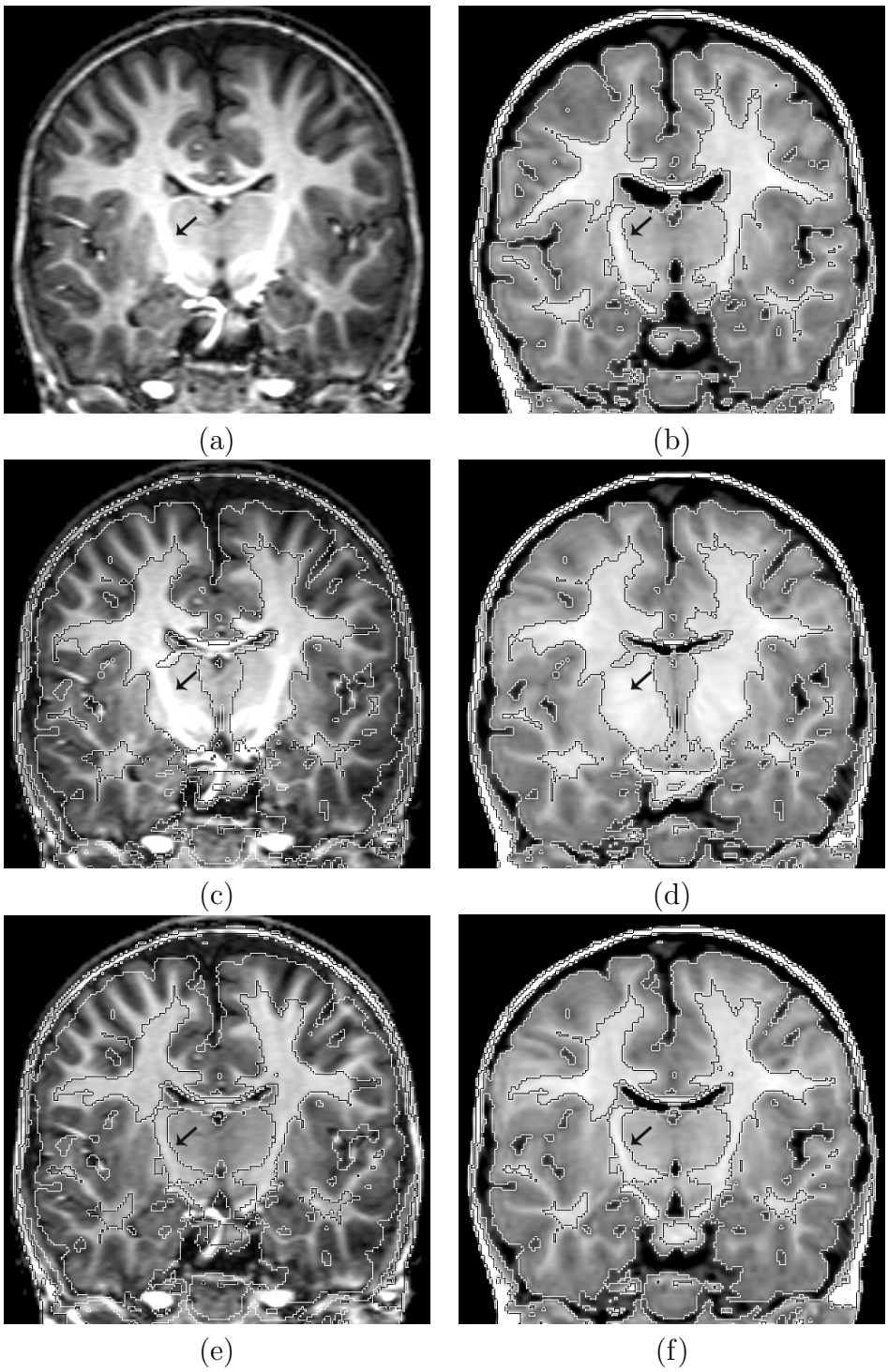


Figure 5.2: Non-rigid registration yields poor results in the presence of strong intensity inhomogeneity: (a) a target image with strong intensity inhomogeneity; (b) a source image with iso-contours; (c) the target image with strong intensity inhomogeneity with overlayed iso-contours of the source image registered by non-rigid registration without bias correction; (d) the source image registered by non-rigid registration without bias correction; (e) the bias corrected target image with overlayed iso-contours of the source image registered by non-rigid registration after bias correction; (f) the source image registered by non-rigid registration after bias correction. Note the boundaries of subcortical structures (see the arrow): Misaligned in (c); Correctly aligned in (e). The B-spline control point spacing at the final resolution level of non-rigid registration is 5 mm. The bias field has been corrected with a novel template-based method described later in this chapter.

5.2 Template-based bias correction

Let Y be an image with intensity inhomogeneity and X a template image with no intensity inhomogeneity aligned with the image Y. Let us assume that image X and intensity inhomogeneity corrected image Y' have the same underlying intensity distribution. Later, we will describe a method for matching the image intensity distributions using linear regression if this assumption is not fulfilled. Let us further assume that the intensities of both images X and Y have been logarithmically transformed. Then $Y = Y' + \beta$, where β represents the bias field which causes the intensity inhomogeneity.

Let $v_i \in \mathcal{V}$ denote the location of a voxel in both of the aligned images X and Y with intensities $x_i = X(v_i)$ and $y_i = Y(v_i)$. Similarly, let us denote $y'_i = Y'(v_i)$ and $b_i = \beta(v_i)$.

A key property of the bias field is its low frequency characteristics. Let $\mathcal{B}$ be a linear low-pass function, such that $\mathcal{B}(\beta) = \beta$ and $\mathcal{B}(X) = \mathcal{B}(Y')$. The consequence of the first property is, that the function $\mathcal{B}$ does not remove low frequencies contained in the bias field. If we assume, that the registration error and anatomical variation of the images X and Y is contained in the higher frequencies, then those can be removed by applying function $\mathcal{B}$, as indicated by the second property. Now the bias field can be expressed as

$$\beta = \mathcal{B}(\beta) = \mathcal{B}(Y - Y') = \mathcal{B}(Y) - \mathcal{B}(Y') = \mathcal{B}(Y) - \mathcal{B}(X) = \mathcal{B}(Y - X)$$

The bias field can be therefore obtained by low-pass filtering of the difference image $Y - X$.

5.3 B-spline model for the bias field

The **B-spline function** of a degree n is a C^{n-1} continuous function with $L - 1$ polynomial segments of degree n. Each point $S(u)$ of the B-spline function S is a linear combination of $L + n - 1$ control points $d_{-\lfloor \frac{n-1}{2} \rfloor}, ..., d_{L+\lfloor \frac{n}{2} \rfloor - 1}$ given by

$$S(u) = \sum_{i=-\lfloor \frac{n-1}{2} \rfloor}^{L+\lfloor \frac{n}{2} \rfloor -1} d_i \mathcal{N}_i^n(u)$$

where $\mathcal{N}^n_{-\lfloor \frac{n-1}{2} \rfloor}(u), ..., \mathcal{N}^n_{L+\lfloor \frac{n}{2} \rfloor -1}(u)$ are the **B-spline basis functions** defined over a sequence of $L + 2n - 2$ knots $u_{-n+1}, ..., u_{L+n-2}$. The function $S(u)$ is defined over $< u_0, u_{L-1} >$. The general definition of the B-spline basis functions can be found in [35]. If the knots are evenly distributed, for example $u_i = i$, the B-spline function is referred to as **uniform B-spline**.

In practice uniform cubic B-splines (of a degree $n = 3$) are most often used because of their increased flexibility compared to quadratic and linear B-splines and because of their computational simplicity compared to higher-order B-splines. Let us define a uniform cubic B-spline $S^3(u)$ with control points $d_{-1}, ..., d_L$ and knot sequence $u_i = i$, $i = -2, ..., L + 1$ by

$$S^3(u) = \sum_{i=-1}^{L} d_i \mathcal{N}_i(u)$$

where $u \in < 0, L - 1 >$ and basis functions $\mathcal{N}_i(u) = \mathcal{N}_i^3(u)$ given by

$$\mathcal{N}_i(u) = \begin{cases} \frac{1}{6}(-t^3 + 3t^2 - 3t + 1) & \text{for} \quad i = l - 1 \\ \\ \frac{1}{6}(3t^3 - 3t^2 + 4) & \text{for} \quad i = l \\ \\ \frac{1}{6}(-3t^3 + 3t^2 + 3t + 1) & \text{for} \quad i = l + 1 \\ \\ \frac{1}{6}t^3 & \text{for} \quad i = l + 2 \end{cases}$$

where $l = \lfloor u \rfloor$ and $t = u - l$.

The 1D B-spline functions can be easily extended to higher dimensions. In MRI, the bias field must be defined as a function in 3D and can be therefore modelled as a **3D tensor-product**

B-spline function:

$$S^3(u_1, u_2, u_3) = \sum_{j=-1}^{L_1} \sum_{k=-1}^{L_2} \sum_{l=-1}^{L_3} d_{jkl} \mathcal{N}_j(u_1) \mathcal{N}_k(u_2) \mathcal{N}_l(u_3)$$

Without any loss of generality the theory in the following sections only deals with 1D B-splines. If a bijective function $\mathcal{I}$: $N^3 \rightarrow N$ is defined to order the 3D indices (j, k, l) in one dimension, and 1D B-spline basis function $\mathcal{N}_i(u)$ is replaced by its 3D counterpart $\mathcal{N}_{\mathcal{I}(j,k,l)}(u) = \mathcal{N}_j(u_1)\mathcal{N}_k(u_2)\mathcal{N}_l(u_3)$, where $u = (u_1, u_2, u_3)$, we obtain the following:

$$S^3(u) = \sum_{i=0}^{L'} d_i \mathcal{N}_i(u)$$

where $i = \mathcal{I}(j, k, l)$ and $L' = (L_1 + 2)(L_2 + 2)(L_3 + 2)$.

5.4 Bias field estimation by weighted least squares

Assuming that the image Y with the bias and template image X without the bias have been registered and the intensities have been matched, the bias field can be estimated from the difference image $Y - X$ via a least square fit of a B-spline function. Let $R = (r_1, ..., r_n)$ be the data to be approximated by a B-spline function $S(u) = \sum_{i=-1}^{L} d_i \mathcal{N}_i(u)$ and let $\mathbf{d} = (d_{-1}, ..., d_L)$ denote a vector of B-spline control points. Then the least square approximation of the data R by a B-spline function $S(u)$ can be obtained by minimizing the sum of square differences between the data points and the function:

$$\hat{\mathbf{d}} = \arg\min_{\mathbf{d}} \sum_{i=1}^{n} (r_i - S(v_i))^2$$

where v_i represents the location of data point r_i. This objective function can be minimized by setting its first partial derivatives in $d_{-1}, ..., d_L$ to zero. The resulting system of linear equations can be written in the matrix form. Let $\mathbf{N}$ be the $n \times (L+2)$ matrix with elements $\mathbf{N}_{ij} = \mathcal{N}_j(v_i)$.

The least square solution can be calculated as follows:

$$\mathbf{d} = (\mathbf{N}^T\mathbf{N})^{-1}\mathbf{N}^T R$$

A significant disadvantage of least square fitting is the sensitivity to outliers. The larger the distance of the outlier from the estimated function, the larger its influence will be on the resulting fit. To address this problem, robust statistics in the form of weighted least squares fitting [31] can be used to obtain more robust results. In the weighted least square fitting the objective function is modified as follows:

$$\hat{\mathbf{d}} = \arg\min_{\mathbf{d}} \sum_{i=1}^{N} w_i (r_i - S(v_i))^2$$

where $w_1, ..., w_n$ are called **weights**. Let $\mathbf{W}$ denote a $n \times n$ matrix with elements $\mathbf{W}_{ii} = w_i$ and $\mathbf{W}_{ik} = 0$ for $i \neq k$. Then the weighted least square fit can be calculated as follows:

$$\mathbf{d} = (\mathbf{N}^T\mathbf{W}\mathbf{N})^{-1}\mathbf{N}^T\mathbf{W}R$$

Using the control points resulting from the least square fit, the bias field is determined as follows: $b_i = \sum_{j=-1}^{L} d_j \mathcal{N}_j(v_i)$. This can be rewritten in the matrix form:

$$\beta = \mathbf{N}\mathbf{d} = \mathbf{N}(\mathbf{N}^T\mathbf{W}\mathbf{N})^{-1}\mathbf{N}^T\mathbf{W}R$$

Let us define the linear low-pass function $\mathcal{B}$ as follows:

$$\mathcal{B}(R) = \mathbf{N}(\mathbf{N}^T\mathbf{W}\mathbf{N})^{-1}\mathbf{N}^T\mathbf{W}R$$

We now show that we defined a suitable low-pass function, as required in section 5.2: $\mathcal{B}$ is a linear function, as it is in the form $\mathcal{B}(R) = \mathbf{M}R$, where $\mathbf{M}$ represents a matrix, so obviously

$\mathcal{B}(aA + bB) = \mathbf{M}(aA + bB) = a\mathbf{M}A + b\mathbf{M}B = a\mathcal{B}(A) + b\mathcal{B}(B)$. It also holds that $\mathcal{B}(\beta) = \beta$:

$$\begin{aligned}\mathcal{B}(\beta) &= \mathbf{N}(\mathbf{N}^T\mathbf{W}\mathbf{N})^{-1}\mathbf{N}^T\mathbf{W}\beta \\ &= \mathbf{N}(\mathbf{N}^T\mathbf{W}\mathbf{N})^{-1}\mathbf{N}^T\mathbf{W}\mathbf{N}\mathbf{d} = \mathbf{N}\mathbf{d} = \beta\end{aligned}$$

The last property of the low-pass function $\mathcal{B}$ that should hold is the ability remove the frequencies encoding the difference between the image and the template $\mathcal{B}(Y') = \mathcal{B}(X)$:

$$\mathcal{B}(Y') = \mathcal{B}(Y - \beta) = \mathcal{B}(Y) - \mathcal{B}(\beta) = \mathcal{B}(Y) - \beta = \mathcal{B}(Y) - \mathcal{B}(Y - X) = \mathcal{B}(Y - Y + X) = \mathcal{B}(X)$$

In practice we do not have any exact way how to determine which frequencies in the difference images result from the corruption of the bias field and which are the result of the imperfect alignment or residual anatomical differences between the image and the template. We will therefore experimentally determine the exact model for the bias field in chapter 6.

Determining suitable weights for the least-squares estimation of the bias field is a crucial requirement for obtaining good results. Ideally, we would like to assign low weights to voxels likely to be misaligned and higher weights to voxels where the same anatomy or tissue is present. In the following section we will demonstrate how the weights can be estimated in a probabilistic framework using the EM algorithm. An example of the proposed template-based intensity inhomogeneity correction using least square fitting and weighted least square fitting is shown in fig. 5.3.

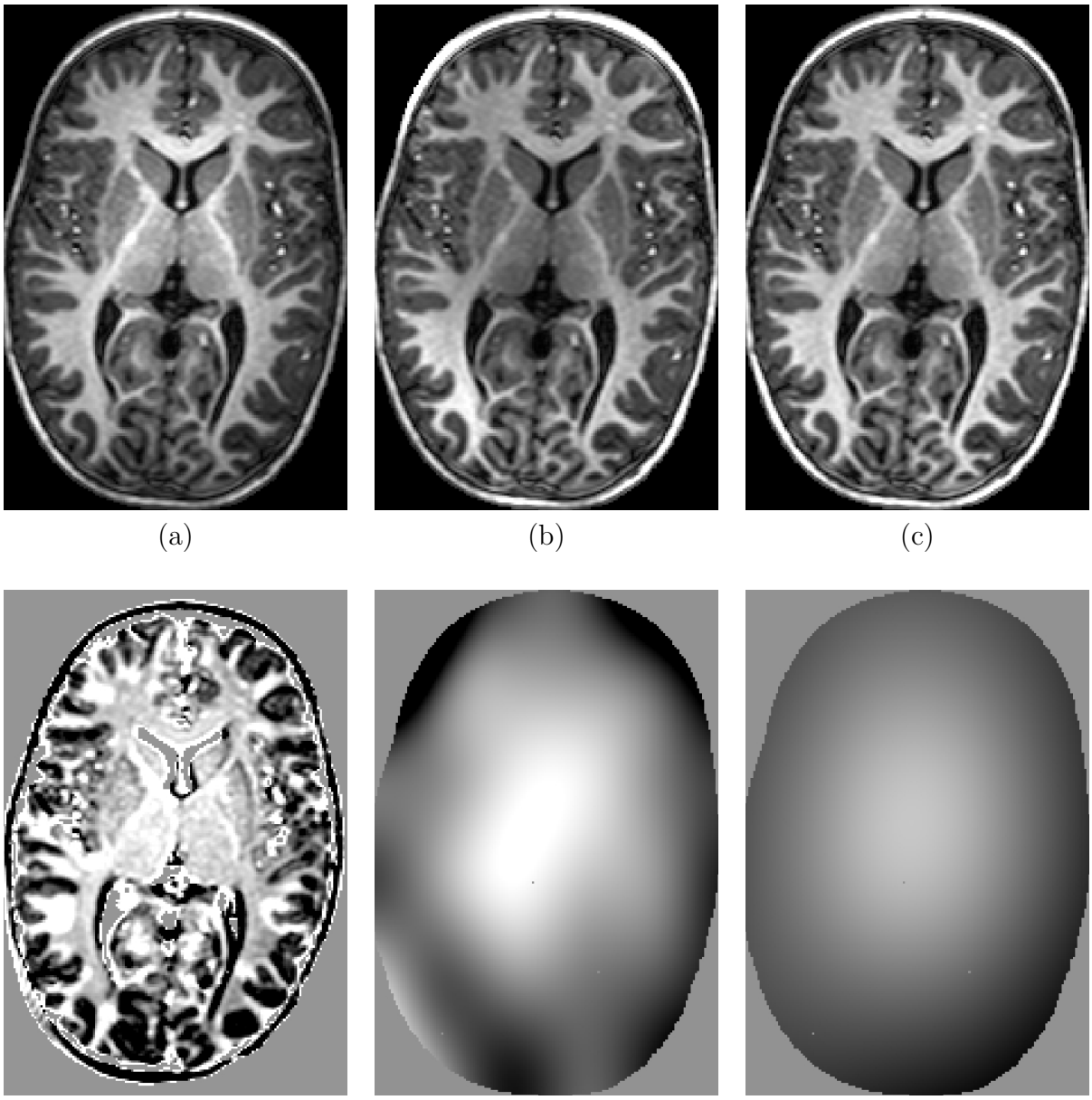


Figure 5.3: Template-based intensity inhomogeneity correction using least square fitting: (a) the original MRI with intensity inhomogeneity; (b) the MRI after intensity inhomogeneity correction with least square fitting; (c) the MRI after intensity inhomogeneity correction with weighted least square fitting; (d) the intensity difference between the original MRI and the template MRI; (e) the bias field obtained by least square fitting; (f) the bias field obtained by weighted least square fitting with the weights calculated to reduce the influence of the outliers. The control point spacing in the B-spline model used is approximately 30mm and a smoothness constraint (explained later in this chapter) with a weighting coefficient 2.5 is also applied during both, least-square and weighted least-square fit.

5.5 Outlier estimation

If the image Y and template image X were perfectly aligned, the difference image $Y - X$ would consist only of bias field and noise, assuming that the intensity distributions of the images X and Y' have been correctly matched. However, in practice the alignment is never perfect. If the corresponding voxels in the location of alignment error contain different tissues, the intensity difference does not correspond only to bias field, but also the intensity difference between the tissue classes. Those voxels are therefore labelled as *outliers*.

The problem of estimating outliers can be viewed as a classification problem. In general, large positive and negative values in the difference image $Y' - X$ are likely to correspond to misaligned voxels (outliers). Differences in intensities due to the presence of the bias field are in general smaller and will be iteratively removed during estimation process. Correctly aligned voxels, also referred to as *inliers*, will produce values close to zero. The PDF of the intensities in the difference image $Y' - X$ can be therefore modelled as a mixture of two zero-mean Gaussians, one with a smaller variance σ_0 modelling the correctly aligned voxels, or inliers, and the other one with a larger variance σ_1 for the misaligned voxels, or outliers. This simple model helps to avoid the classification of different tissue classes and thus keep the algorithm simple and generally applicable. It would be possible to extend the model to include the tissue class PDFs and resulting intensity characteristics of the difference image.

The inliers and the outliers in the difference image can be classified within the EM framework while simultaneously estimating the bias field. This is done by optimizing maximum likelihood:

$$(\hat{\gamma}, \hat{\beta}) = \arg\max_{\gamma,\beta} \log P(Y - X|\gamma, \beta)$$

where γ denotes the Gaussian mixture parameters $(\sigma_0, \sigma_1, c_0, c_1)$. As there is no prior information available for spatial position or volume of outliers and inliers, we assume a spatially constant prior for each tissue class $P(z_i = k) = c_k$. The log-likelihood can be expressed as follows:

$$\log P(Y - X|\gamma, \beta) = \sum_{i=1}^{n} \log \sum_{k=0}^{1} c_k G(y_i - x_i - b_i, 0, \sigma_k)$$

Using the EM algorithm the maximum likelihood can be optimized in two steps:

E-step:

$$\hat{p_{ik}} = \frac{\hat{c_k} G(y_i - x_i - \hat{b_i}, 0, \hat{\sigma_k})}{\sum_l \hat{c_l} G(y_i - x_i - \hat{b_i}, 0, \hat{\sigma_l})}$$

M-step:

$$(\hat{\gamma}, \hat{\beta}) = \arg\max_{\gamma,\beta} \sum_{i=1}^{n} \sum_{k=0}^{1} \hat{p_{ik}} (\log c_k + \log G(y_i - x_i - b_i, 0, \sigma_k)) \tag{5.1}$$

where ˆ denotes the current estimate. Equation 5.1 can be split into two separate equations: The first equation

$$(\hat{c_0}, \hat{c_1}) = \arg\max_{c_0,c_1} \sum_{i=1}^{n} \sum_{k=0}^{1} \hat{p_{ik}} \log c_k$$

yields

$$\hat{c_k} = \sum_{i=1}^{n} \frac{\hat{p_{ik}}}{n}$$

The objective function in the second equation

$$(\hat{\sigma_0}, \hat{\sigma_1}, \hat{\beta}) = \arg\max_{\sigma_0,\sigma_1,\beta} \sum_{i=1}^{n} \sum_{k=0}^{1} \hat{p_{ik}} \log G(y_i - x_i - b_i, 0, \sigma_k)$$

can be further expressed as

$$-n \log \sqrt{2\pi} - \sum_{k=0}^{1} \log \sigma_k \sum_{i=1}^{n} \hat{p_{ik}} - \frac{1}{2} \sum_{i=1}^{n} (\sum_{k=0}^{1} \frac{\hat{p_{ik}}}{\sigma_k^2})(y_i - x_i - b_i)^2$$

The estimation of (σ_0, σ_1) cannot be separated from the estimation of the bias field β. We will therefore use the current estimate of the parameters to keep the problem computationally tractable. (σ_0, σ_1) can be thus obtained by

$$(\hat{\sigma_0}, \hat{\sigma_1}) = \arg\max_{\sigma_0,\sigma_1} (- \sum_{k=0}^{1} \log \sigma_k \sum_{i=1}^{n} \hat{p_{ik}} - \frac{1}{2} \sum_{i=1}^{n} (\sum_{k=0}^{1} \frac{\hat{p_{ik}}}{\sigma_k^2})(y_i - x_i - \hat{b_i})^2)$$

which results in

$$\hat{\sigma_k} = \frac{\sum_{i=1}^{n} \hat{p_{ik}}(y_i - x_i - \hat{b_i})^2}{\sum_{i=1}^{n} \hat{p_{ik}}}$$

Similarly, the bias field β can be estimated from

$$\hat{\beta} = \arg\max_{\beta}(-\frac{1}{2}\sum_{i=1}^{n}(\sum_{k=0}^{1}\frac{\hat{p_{ik}}}{\hat{\sigma_k}^2})(y_i - x_i - b_i)^2)$$

and this is equivalent to the weighted least square fit

$$\hat{\beta} = \arg\min_{\beta}\sum_{i=1}^{n} w_i(y_i - x_i - b_i)^2 \tag{5.2}$$

with the weights

$$w_i = \sum_{k=0}^{1}\frac{\hat{p_{ik}}}{\hat{\sigma_k}^2}$$

The bias field can then be estimated as shown in the previous section:

$$\hat{\beta} = \mathbf{N}(\mathbf{N}^T\mathbf{W}\mathbf{N})^{-1}\mathbf{N}^T\mathbf{W}(Y - X)$$

Fig. 5.4 shows how the weights are assigned to voxels of the difference image $Y' - X$ depending on their intensities. The histogram of the image is approximated as a mixture of two zero-mean Gaussians:

$$P(y_i' - x_i|\gamma, \beta) = c_0 G(y_i' - x_i, 0, \sigma_0) + c_1 G(y_i' - x_i, 0, \sigma_1)$$

and the weights can be calculated as a function of the intensities $y_i' - x_i$ and parameters γ:

$$w_i = w(y_i' - x_i) = \frac{\frac{1}{\sigma_0^2}c_0 G(y_i' - x_i, 0, \sigma_0) + \frac{1}{\sigma_1^2}c_1 G(y_i' - x_i, 0, \sigma_1)}{c_0 G(y_i' - x_i, 0, \sigma_0) + c_1 G(y_i' - x_i, 0, \sigma_1)}$$

During the entire estimation we assume $\sigma_0 < \sigma_1$ and consequently $w(y' - x)$ represents a suitable weight function suppressing the influence of the outliers on the resulting bias field estimation.

The results of the estimation of the bias field with weighted least square fit and various B-spline control point spacings are shown in fig. 5.5. While a control point spacing of 150mm produces a too smooth bias field estimate and the corrected image still contains visible residual

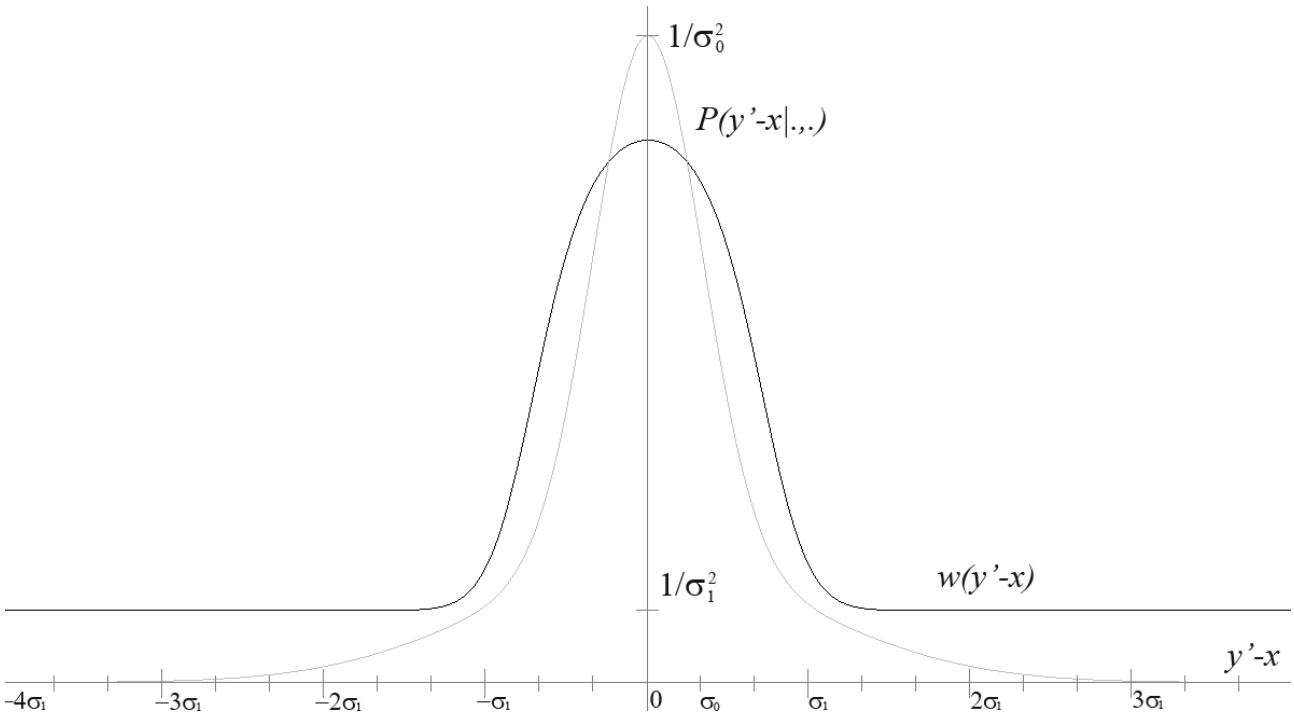


Figure 5.4: The weight function: The x-axis represents the intensities of the difference image $Y' - X$. The function shown in grey is the estimated intensity distribution $P(y' - x|\gamma, \beta)$ of the image $Y' - X$ rescaled to $< 0, \frac{1}{\sigma_0^2} >$. The function shown in black is the weight function $w(y' - x)$. Both functions have been calculated using parameters $\sigma_1 = 3\sigma_0$, $c_0 = 0.6$, and $c_1 = 0.4$.

bias, a control point spacing between 70mm and 40mm produces a visually more plausible bias correction. Control point spacings below 40mm produce bias fields containing higher frequencies, introducing new intensity inhomogeneities into the corrected image.

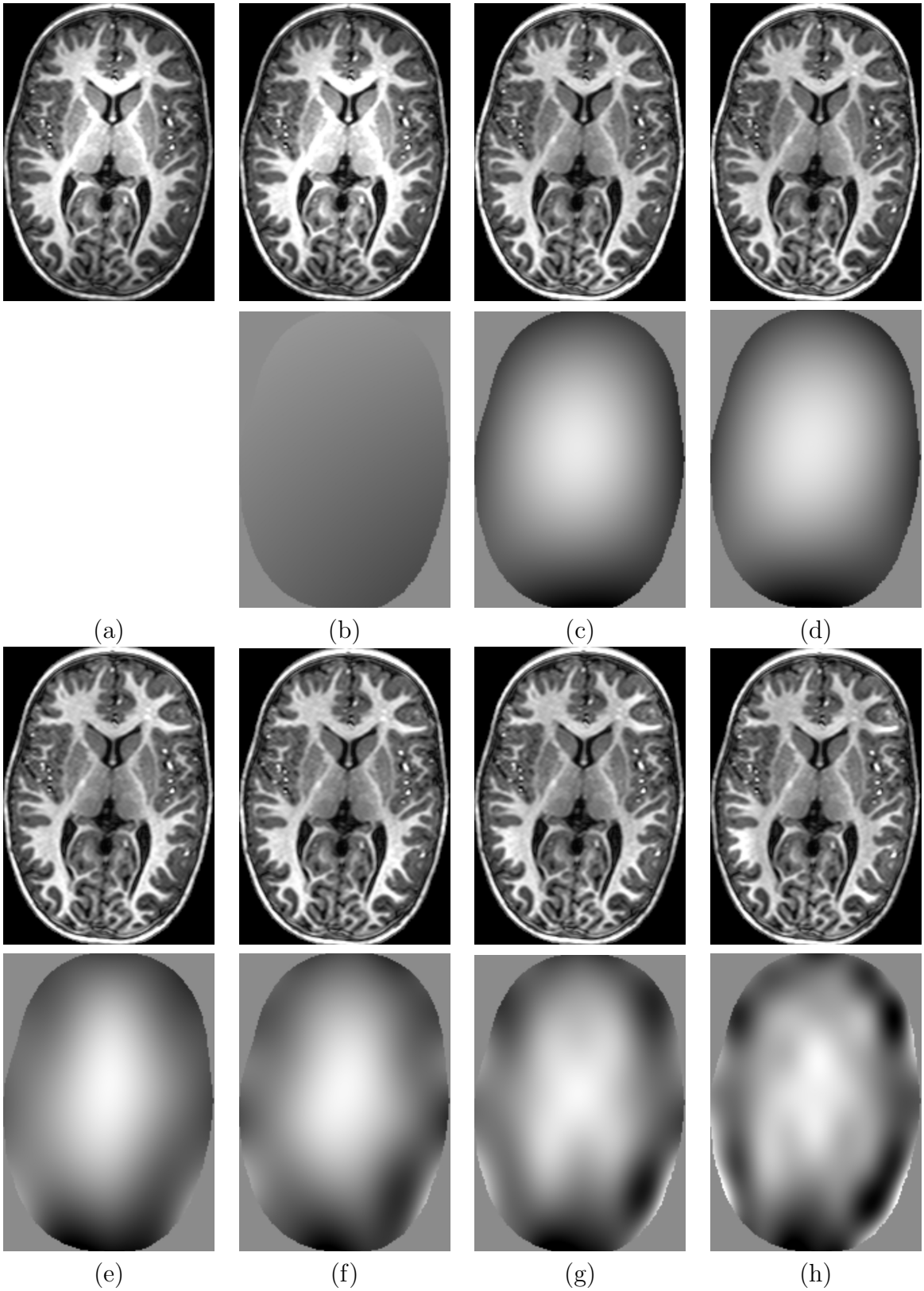


Figure 5.5: The influence of the B-spline control point spacing on the weighted least square fit: (a) the original image; The corrected image and the corresponding bias field obtained with the B- spline control point spacing: (b) 150mm; (c) 75mm; (d) 50mm; (e) 40mm; (f) 30mm; (g) 25mm; (h) 20mm.

5.6 Smoothness constraint for B-spline functions

The model for the bias field should consist of low frequency components, but at the same time it should have enough flexibility to model bias fields realistically. Rather than enforcing a low frequency bias field implicitly by using B-splines with a large control point spacing, it may be better to use a smaller control point spacing for the bias field model. However, this may lead to bias fields which are not smooth enough. Therefore we have decided to add an explicit smoothness constraint such as the one minimizing the bending energy introduced in section 2.2.5 (equation 2.1). The general form of such penalty term has been described by Wahba [92]. This quantity is a measure of the smoothness of a transformation representing the 3D counterpart of 2D bending energy of a thin plate of metal. This smoothness constraint was numerically implemented in [72] to constrain the B-spline registration [73]. Here we exploit the uniform B-spline model of the bias field and calculate this expression exactly.

Let us denote the six terms of the smoothness constraint C_{abc}:

$$C_{200} = \int_0^{X_1} \int_0^{X_2} \int_0^{X_3} \left(\frac{\partial^2 S(\frac{v_1}{\delta_1}, \frac{v_2}{\delta_2}, \frac{v_3}{\delta_3})}{\partial v_1^2} \right)^2 dv_1 dv_2 dv_3$$

$$C_{020} = \int_0^{X_1} \int_0^{X_2} \int_0^{X_3} \left(\frac{\partial^2 S(\frac{v_1}{\delta_1}, \frac{v_2}{\delta_2}, \frac{v_3}{\delta_3})}{\partial v_2^2} \right)^2 dv_1 dv_2 dv_3$$

$$C_{002} = \int_0^{X_1} \int_0^{X_2} \int_0^{X_3} \left(\frac{\partial^2 S(\frac{v_1}{\delta_1}, \frac{v_2}{\delta_2}, \frac{v_3}{\delta_3})}{\partial v_3^2} \right)^2 dv_1 dv_2 dv_3$$

$$C_{110} = \int_0^{X_1} \int_0^{X_2} \int_0^{X_3} \left(\frac{\partial^2 S(\frac{v_1}{\delta_1}, \frac{v_2}{\delta_2}, \frac{v_3}{\delta_3})}{\partial v_1 \partial v_2} \right)^2 dv_1 dv_2 dv_3$$

$$C_{101} = \int_0^{X_1} \int_0^{X_2} \int_0^{X_3} \left(\frac{\partial^2 S(\frac{v_1}{\delta_1}, \frac{v_2}{\delta_2}, \frac{v_3}{\delta_3})}{\partial v_1 \partial v_3} \right)^2 dv_1 dv_2 dv_3$$

$$C_{011} = \int_0^{X_1} \int_0^{X_2} \int_0^{X_3} \left(\frac{\partial^2 S(\frac{v_1}{\delta_1}, \frac{v_2}{\delta_2}, \frac{v_3}{\delta_3})}{\partial v_2 \partial v_3} \right)^2 dv_1 dv_2 dv_3$$

The smoothness constraint C_{smooth} is then

$$C_{smooth} = \frac{1}{V}(C_{200} + C_{020} + C_{002} + C_{110} + C_{101} + C_{011})$$

$$V = X_1 X_2 X_3$$

Note, that $S(\frac{v_1}{\delta_1}, \frac{v_2}{\delta_2}, \frac{v_3}{\delta_3})$ denotes a reparametrized bias field function with domain $< 0, X_1 > \times < 0, X_2 > \times < 0, X_3 >$, while the original parameters $u_i = \frac{v_i}{\delta_i}$ run over domain $< 0, L_1 > \times < 0, L_2 > \times < 0, L_3 >$.

To calculate the smoothness constraint, we need to compute the first and second derivatives of the B-spline functions. **The first derivative of a B-spline function** can be expressed as

$$\frac{\partial}{\partial u} \sum_{i=-1}^{L} d_i \mathcal{N}_i(u) = \sum_{i=0}^{L} (d_i - d_{i-1}) \mathcal{N}_i^2(u)$$

$$\mathcal{N}_i^2(u) = \begin{cases} \frac{1}{2}(t^2 - 2t + 1) & \text{for} \quad i = l \\ \frac{1}{2}(-2t^2 + 2t + 1) & \text{for} \quad i = l + 1 \\ \frac{1}{2}t^2 & \text{for} \quad i = l + 2 \end{cases}$$

with $l = \lfloor u \rfloor$ and $t = u - l$, where $\mathcal{N}_i^2(u)$ is a B-spline basis function of the second degree. Similarly, **the second derivative of a B-spline function** can be expressed as

$$\frac{\partial^2}{\partial u^2} \sum_{i=-1}^{L} d_i \mathcal{N}_i(u) = \sum_{i=0}^{L-1} (d_{i+1} - 2d_i + d_{i-1}) \mathcal{N}_i^1(u)$$

$$\mathcal{N}_i^1(u) = \begin{cases} 1 - t & \text{for} \quad i = l \\ t & \text{for} \quad i = l + 1 \end{cases}$$

with $l = \lfloor u \rfloor$ and $t = u - l$, where $\mathcal{N}_i^1(u)$ is a B-spline basis function of the first degree.

If the B-spline functions are reparametrized with $v = \delta u$, the following holds:

$$\frac{\partial}{\partial v}\sum_{i=-1}^{L} d_i \mathcal{N}_i(\frac{v}{\delta}) = \frac{1}{\delta}\sum_{i=0}^{L}(d_i - d_{i-1})\mathcal{N}_i^2(\frac{v}{\delta})$$

$$\frac{\partial^2}{\partial v^2}\sum_{i=-1}^{L} d_i \mathcal{N}_i(\frac{v}{\delta}) = \frac{1}{\delta^2}\sum_{i=0}^{L-1}(d_{i+1} - 2d_i + d_{i-1})\mathcal{N}_i^1(\frac{v}{\delta})$$

Let us define

$$\begin{aligned}
\Delta_i^0 &= d_i \\
\Delta_i^1 &= d_i - d_{i-1} \\
\Delta_i^2 &= \Delta_{i+1}^1 - \Delta_i^1 = d_{i+1} - 2d_i + d_{i-1}
\end{aligned}$$

This can be easily extended into 3D:

$$\begin{aligned}
\Delta_{ijk}^{000} &= d_{ijk} \\
\Delta_{ijk}^{100} &= d_{ijk} - d_{i-1,j,k} \\
\Delta_{ijk}^{200} &= \Delta_{i+1,j,k}^{100} - \Delta_{ijk}^{100} = d_{i+1,j,k} - 2d_{ijk} + d_{i-1,j,k} \\
\Delta_{ijk}^{110} &= \Delta_{i,j,k}^{010} - \Delta_{i-1,j,k}^{010} = d_{ijk} - d_{i,j-1,k} - d_{i-1,j,k} + d_{i-1,j-1,k}
\end{aligned}$$

and similarly for all the other permutations (a, b, c) in Δ_{ijk}^{abc}, where $(a + b + c) \leq 2$. Let us further define matrices $\mathbf{H}^a$, $a = 1, 2, 3$ with elements h_{ij}^a:

$$\begin{aligned}
h_{ij}^3 &= \textstyle\int_0^{L-1} \mathcal{N}_i^3(u)\mathcal{N}_j^3(u)du,\ i, j = -1, ..., L \\
h_{ij}^2 &= \textstyle\int_0^{L-1} \mathcal{N}_i^2(u)\mathcal{N}_j^2(u)du,\ i, j = 0, ..., L \\
h_{ij}^1 &= \textstyle\int_0^{L-1} \mathcal{N}_i^1(u)\mathcal{N}_j^1(u)du,\ i, j = 0, ..., L-1
\end{aligned}$$

Again, the reparametrization results in $h_{ij}^a(v) = \delta h_{ij}^a$. Given the previous definitions, each term of the smoothness constraint can be expressed as

$$\begin{aligned}
C_{200} &= \tfrac{\delta_1\delta_2\delta_3}{\delta_1^4}\textstyle\sum_{i_1=0}^{L_1-1}\sum_{i_2=-1}^{L_2}\sum_{i_3=-1}^{L_3}\sum_{j_1=0}^{L_1-1}\sum_{j_2=-1}^{L_2}\sum_{j_3=-1}^{L_3}\Delta_{i_1i_2i_3}^{200}\Delta_{j_1j_2j_3}^{200}h_{i_1j_1}^1 h_{i_2j_2}^3 h_{i_3j_3}^3 \\
C_{110} &= \tfrac{\delta_1\delta_2\delta_3}{\delta_1^2\delta_2^2}\textstyle\sum_{i_1=0}^{L_1}\sum_{i_2=0}^{L_2}\sum_{i_3=-1}^{L_3}\sum_{j_1=0}^{L_1}\sum_{j_2=0}^{L_2}\sum_{j_3=-1}^{L_3}\Delta_{i_1i_2i_3}^{110}\Delta_{j_1j_2j_3}^{110}h_{i_1j_1}^2 h_{i_2j_2}^2 h_{i_3j_3}^3
\end{aligned}$$

and so on. The exact calculation of h_{ij}^a in the matrices $\mathbf{H}^1$, $\mathbf{H}^2$, $\mathbf{H}^3$ can be easily implemented, as it involves the integration of a product of two polynomials. The matrices for $L = 5$ are listed

below:

$$\mathbf{H}^1 = \frac{1}{6}\begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 1 & 4 & 1 & 0 & 0 \\ 0 & 1 & 4 & 1 & 0 \\ 0 & 0 & 1 & 4 & 1 \\ 0 & 0 & 0 & 1 & 2 \end{pmatrix}$$

$$\mathbf{H}^2 = \frac{1}{120}\begin{pmatrix} 6 & 13 & 1 & 0 & 0 & 0 \\ 13 & 60 & 26 & 1 & 0 & 0 \\ 1 & 26 & 66 & 26 & 1 & 0 \\ 0 & 1 & 26 & 66 & 26 & 1 \\ 0 & 0 & 1 & 26 & 60 & 13 \\ 0 & 0 & 0 & 1 & 13 & 6 \end{pmatrix}$$

$$\mathbf{H}^3 = \frac{1}{5040}\begin{pmatrix} 20 & 129 & 60 & 1 & 0 & 0 & 0 \\ 129 & 1208 & 1062 & 120 & 1 & 0 & 0 \\ 60 & 1062 & 2369 & 1191 & 120 & 1 & 0 \\ 1 & 120 & 1191 & 2416 & 1191 & 120 & 1 \\ 0 & 1 & 120 & 1191 & 2369 & 1062 & 60 \\ 0 & 0 & 1 & 120 & 1062 & 1208 & 129 \\ 0 & 0 & 0 & 1 & 60 & 129 & 20 \end{pmatrix}$$

To simplify the notation, let us rewrite the smoothness constraint in matrix form: First of all, let us define matrices $\mathbf{H}^{abc}$ with elements

$$\mathbf{H}^{abc}_{\mathcal{I}(i_1 i_2 i_3),\mathcal{I}(j_1 j_2 j_3)} = h^a_{i_1 j_1} h^b_{i_2 j_2} h^c_{i_3 j_3}$$

and matrix $\mathbf{\Delta}^{abc}$ so that

$$\mathbf{e}_{\mathcal{I}(i_1 i_2 i_3)} \mathbf{\Delta}^{abc} \mathbf{d} = \Delta^{abc}_{i_1 i_2 i_3}$$

where $\mathbf{e}_i$ is a unit vector with value 1 at ith position and zeros elsewhere and $\mathcal{I}$ is a bijec-

tive function defining an ordering of the 3D indices in one dimension. Then the smoothness constraint can be rewritten in the matrix form as:

$$C_{smooth} = \mathbf{d}^T \mathbf{H} \mathbf{d} \tag{5.3}$$

where $\mathbf{H}$ is defined as

$$\begin{aligned}\mathbf{H} &= \frac{\delta_1\delta_2\delta_3}{V}\left(\frac{1}{\delta_1^4}(\mathbf{\Delta}^{200})^T\mathbf{H}^{133}\mathbf{\Delta}^{200} + \frac{1}{\delta_2^4}(\mathbf{\Delta}^{020})^T\mathbf{H}^{313}\mathbf{\Delta}^{020} + \frac{1}{\delta_3^4}(\mathbf{\Delta}^{003})^T\mathbf{H}^{331}\mathbf{\Delta}^{002}\right.\\ &+ \left.\frac{1}{\delta_1^2\delta_2^2}(\mathbf{\Delta}^{110})^T\mathbf{H}^{223}\mathbf{\Delta}^{110} + \frac{1}{\delta_1^2\delta_3^2}(\mathbf{\Delta}^{101})^T\mathbf{H}^{232}\mathbf{\Delta}^{101} + \frac{1}{\delta_2^2\delta_3^2}(\mathbf{\Delta}^{011})^T\mathbf{H}^{322}\mathbf{\Delta}^{011}\right)\end{aligned}$$

5.7 Incorporating the smoothness constraint into the bias field estimation

The smoothness constraint can be incorporated into the EM algorithm through the MAP principle:

$$(\hat{\gamma}, \hat{\beta}) = \arg\max_{\gamma,\beta} \log P(\gamma, \beta | Y - X) = \arg\max_{\gamma,\beta} \log P(Y - X | \gamma, \beta) P(\gamma, \beta)$$

We do not assume any prior for the parameters γ, so $P(\gamma, \beta) = P(\beta)$. We can constrain the shape of the bias field by employing the smoothness constraint for the B-spline control points:

$$P(\beta) = P(\mathbf{d}) = G(\mathbf{d}, 0, \frac{1}{\lambda}\mathbf{H}^{-1}) = \frac{1}{(2\pi)^{\frac{L+2}{2}}|\frac{1}{\lambda}\mathbf{H}^{-1}|} \exp\left(-\frac{\lambda}{2}\mathbf{d}^T\mathbf{H}\mathbf{d}\right)$$

where G is a multivariate Gaussian distribution with zero mean and covariance matrix $\frac{1}{\lambda}\mathbf{H}^{-1}$. Note, that the B-spline function is defined in 1D for notational simplicity and the theory can be easily extended to 3D as demonstrated in sec. 5.3.

The prior $P(\beta)$ is correctly defined only if matrix $\mathbf{H}$ is regular and the inverse $\mathbf{H}^{-1}$ exists.

However, the smoothness matrix $\mathbf{H}$, as defined in previous section, is singular. The theoretical model can therefore be amended to define the Gaussian distribution for a vector $\mathbf{d}'$ of $L'+2$ suitable linear combinations of control points $\mathbf{d}$, with covariance matrix $\frac{1}{\lambda}(\mathbf{H}')^{-1}$ so that $\mathbf{H}'$ is a regular matrix and $\mathbf{d}^T\mathbf{H}\mathbf{d} = (\mathbf{d}')^T\mathbf{H}'\mathbf{d}'$:

$$P(\beta) = P(\mathbf{d}') = G(\mathbf{d}', 0, \frac{1}{\lambda}\mathbf{H}'^{-1}) = \frac{1}{(2\pi)^{\frac{L'+2}{2}}|\frac{1}{\lambda}\mathbf{H}'^{-1}|}\exp\left(-\frac{\lambda}{2}\mathbf{d}'^T\mathbf{H}'\mathbf{d}'\right)$$

Maximizing $\log P(\beta)$ is equivalent to minimizing $\frac{\lambda}{2}\mathbf{d}'^T\mathbf{H}'\mathbf{d}' = \frac{\lambda}{2}\mathbf{d}^T\mathbf{H}\mathbf{d}$ and the whole objective function for finding the parameters β can then be written, using the eq. 5.2, as

$$\hat{\beta} = \arg\min_{\beta}\left(\sum_{i=1}^{n} w_i(y_i - x_i - b_i)^2 + \lambda\mathbf{d}^T\mathbf{H}\mathbf{d}\right)$$

Setting the first partial derivatives in each d_j to zero results in the following system of equations:

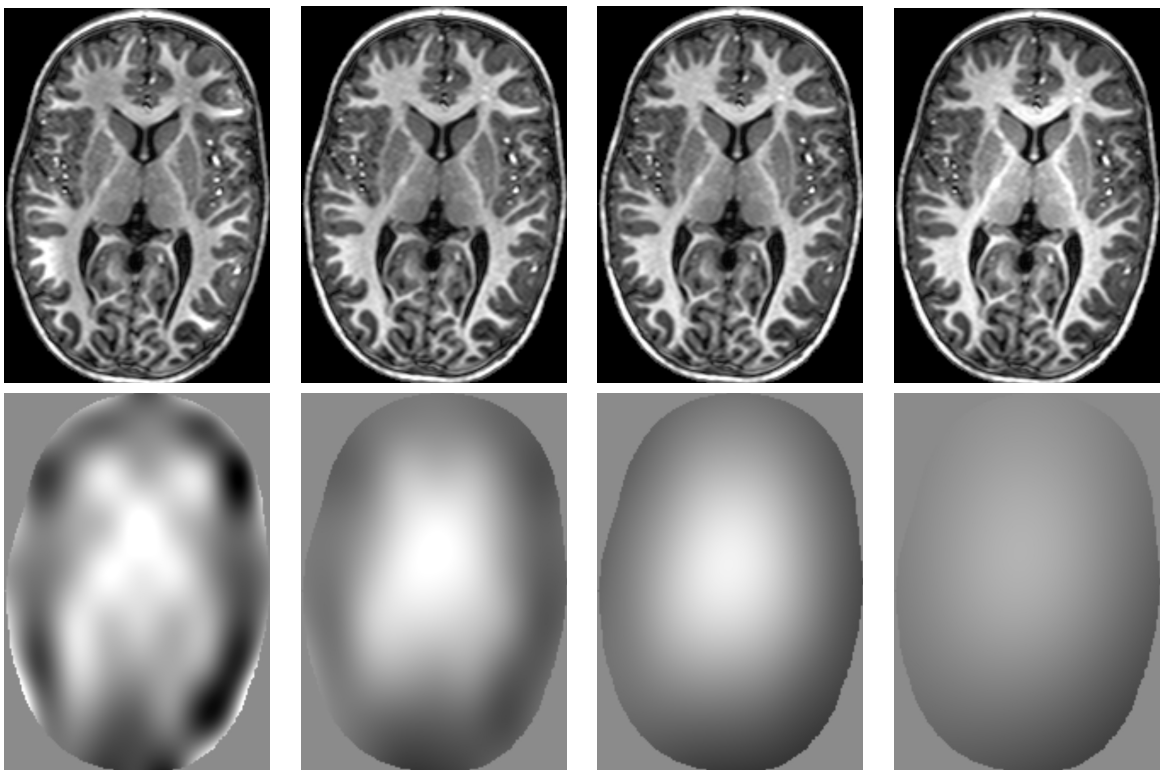
$$\hat{\mathbf{d}} = (\mathbf{N}^T\mathbf{W}\mathbf{N} + 2\lambda\mathbf{H})^{-1}\mathbf{N}^T\mathbf{W}(Y - X) \tag{5.4}$$

$$\hat{\beta} = \mathbf{N}\hat{\mathbf{d}}$$

The influence of the smoothness constraint and the weighting coefficient λ on the results is demonstrated in fig. 5.6.

5.8 Implementing the bias field estimation

Care needs to be taken when implementing the bias field estimation algorithm as defined in equation 5.4: Not all the voxels in brain MRI are suitable for the estimation of the bias field. For example, the air outside the head tends to have intensity values that are close to zero and as the bias field is multiplicative, the change in the intensity of these background voxels is negligible. Consequently, the noise has a higher magnitude than the bias field in this location and including these background voxels introduces numerical problems into the bias field estimation algorithm. We have therefore decided to exclude background voxels from the



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

Figure 5.6: The influence of the smoothness constraint on the weighted least square fit with a B-spline control point spacing of 15mm. The corrected image and the corresponding bias field with the weighting coefficient: (a) $\lambda = 0$; (b) $\lambda = 0.15$; (c) $\lambda = 1.5$; (d) $\lambda = 15$.

bias field estimation. Unfortunately, this often results in the matrix $\mathbf{N}^T\mathbf{W}\mathbf{N}$ being singular. The matrix $\mathbf{N}^T\mathbf{W}\mathbf{N} + 2\lambda\mathbf{H}$ is also not guaranteed to be regular. Therefore, we use singular value decomposition [70] to calculate the solution of the expression below:

$$(\mathbf{N}^T\mathbf{W}\mathbf{N} + 2\lambda\mathbf{H})\hat{\mathbf{d}} = \mathbf{N}^T\mathbf{W}(Y - X)$$

The process can be made more numerically stable if we impose boundary conditions for the B-spline function. We opt for linear boundary conditions given by equations

$$\begin{aligned} d_{-1} &= 2d_0 - d_1 \\ d_L &= 2d_{L-1} - d_{L-2} \end{aligned}$$

in the 1D case. This condition ensures, that the three control points d_{-1}, d_0, d_1, as well as d_{L-2}, d_{L-1}, d_L lie on a straight line. Consequently, the B-spline curve begins and ends in control

points d_0 and d_{L-1} and in the proximity of those points resembles a straight line.

The **3D cubic B-spline with linear boundary conditions** is defined by control points d_{ijk}, $i, j, k = 0, ..., L-1$ and remaining control points are determined using recurrent application of the following equations:

$$\begin{aligned}
d_{-1,j,k} &= 2d_{0,j,k} - d_{1,j,k} \\
d_{i,-1,k} &= 2d_{i,0,k} - d_{i,1,k} \\
d_{i,j,-1} &= 2d_{i,j,0} - d_{i,j,1} \\
d_{L,j,k} &= 2d_{L-1,j,k} - d_{L-2,j,k} \\
d_{i,L,k} &= 2d_{i,L-1,k} - d_{i,L-2,k} \\
d_{i,j,L} &= 2d_{i,j,L-1} - d_{i,j,L-2}
\end{aligned}$$

For example, the control point $D_{-1,-1,-1}$ can be determined as follows:

$$\begin{aligned}
d_{-1,-1,-1} &= 2d_{0,-1,-1} - d_{1,-1,-1} \\
&= 2(2d_{0,0,-1} - d_{0,1,-1}) - (2d_{1,0,-1} - d_{1,1,-1}) \\
&= 2(2(2d_{0,0,0} - d_{0,0,1}) - (2d_{0,1,0} - d_{0,1,1})) \\
&\quad -(2(2d_{1,0,0} - d_{1,0,1}) - (2d_{1,1,0} - d_{1,1,1})) \\
&= 8d_{0,0,0} - 4(d_{1,0,0} + d_{0,1,0} + d_{0,0,1}) \\
&\quad +2(d_{1,1,0} + d_{1,0,1} + d_{0,1,1}) - d_{1,1,1}
\end{aligned}$$

Additionally, the calculation of the full matrix $\mathbf{N}^T\mathbf{W}\mathbf{N}$ at each iteration is computationally expensive and we therefore often subsample the image to a lower resolution for the purposes of estimating the bias field.

The calculation of the smoothness matrix can also take a significant amount of computing time, depending on the number of control points involved. Fortunately, this matrix can be pre-calculated and remains the same unless the number or position of the control points is changed. We therefore pre-calculate this matrix only once.

5.9 Intensity matching

In the previous sections we had assumed that images X and Y have the same underlying intensity distributions. However, this is usually not the case in practical applications. Therefore the intensity distribution of the template image X has to be adjusted to match the intensity distribution of the image Y before the bias correction can be performed. The simplest way to perform this intensity matching is to use linear regression. In general, different imaging protocols result in non-linear relationships of the tissue intensities in MRI and therefore the linear regression cannot produce exact matching. However, a good approximation can still be achieved, as the pairs of voxels of the aligned images are mostly present in two clusters centered around mean WM and GM intensities, see fig. 5.7 d, e.

Let $X = (x_1, ..., x_n)$ and $Y = (y_1, ..., y_n)$ be the intensities of the registered images X and Y. We are trying to find a linear intensity transformation ι from the image X to the image Y in form

$$y = \iota(x) = ax + b$$

The objective function for a least square fit is

$$(\hat{a}, \hat{b}) = \arg\min_{a,b} \sum_{i=0}^{n} (y_i - ax_i - b)^2$$

A better solution, as shown in fig. 5.7, can be obtained by minimizing the least square error in both images x and y:

$$(\hat{a}, \hat{b}) = \arg\min_{a,b} \sum_{i=0}^{n} ((y_i - at_i - b)^2 + (x_i - t_i)^2) \tag{5.5}$$

$$t_i = \arg\min_{t} ((y_i - at - b)^2 + (x_i - t)^2) \tag{5.6}$$

These two equations minimize the sum of square distances of the intensity pair (x_i, y_i) from the line $y = ax + b$. Setting the derivative in t of the objective function in eq. 5.6 to zero results in

$$t_i = \frac{x_i + ay_i - ab}{a^2 + 1}$$

and eq. 5.5 becomes

$$(\hat{a}, \hat{b}) = \arg\min_{a,b} \frac{1}{a^2+1} \sum_{i=0}^{n} (y_i - ax_i - b)^2 \tag{5.7}$$

Setting the first derivative in b of the objective function in eq. 5.7 to zero results in

$$b = E(Y) - aE(X) \tag{5.8}$$

where E is the expected value defined as

$$E(X) = \frac{1}{n} \sum_{i=1}^{n} x_i$$

If we further define the variance and covariance

$$Var(X) = \frac{1}{n} \sum_{i=1}^{n} (x_i - E(X))^2$$

$$Cov(X, Y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - E(X))(y_i - E(Y))$$

then equations 5.7 and 5.8 yield

$$\hat{a} = \arg\min_{a} \frac{1}{a^2+1} (a^2 Var(X) - 2aCov(X, Y) + Var(Y))$$

The differentiation of this objective function gives the solution

$$a_{+,-} = \frac{Var(Y) - Var(X) \pm \sqrt{(Var(Y) - Var(X))^2 + 4Cov(XY)^2}}{2Cov(XY)} \tag{5.9}$$

where $a = a_+$ if $Cov(XY) > 0$ and $a = a_-$ if $Cov(XY) < 0$.

An important property of this solution to be taken into account is its sensitivity to any scaling of the image intensities

$$\iota(X) \neq \iota(cX)$$

where c is a constant. This means, that rescaling the intensities of the template image X will result in different intensity matched template. To tackle this problem, we can scale the

template image X before the matching to have the same mean and variance as image Y. This results in considerable simplification of the equation 5.9. Given the re-scaled image X', so that $E(X') = E(Y)$ and $Var(X') = Var(Y)$, the solution to the least square problem eq. 5.7 will become

$$a = \begin{cases} 1 & \text{if} \quad Cov(X'Y) > 0 \\ -1 & \text{if} \quad Cov(X'Y) < 0 \end{cases}$$

$$b = E(Y)(1-a)$$

To obtain the final solution, the image X' needs to be found. Let us assume $X' = rX + s$. Then

$$E(X') = rE(X) + s = E(Y)$$

$$Var(X') = r^2 Var(X)$$

From this it follows that $X' = \sqrt{\frac{Var(Y)}{Var(X)}}(X - E(X)) + E(Y)$ and the final intensity transformation ι will become

$$\iota(X) = \begin{cases} \sqrt{\frac{Var(Y)}{Var(X)}}(X - E(X)) + E(Y) & \text{if} \quad Cov(XY) > 0 \\ -\sqrt{\frac{Var(Y)}{Var(X)}}(X - E(X)) + E(Y) & \text{if} \quad Cov(XY) < 0 \end{cases}$$

Fig. 5.7 shows the improvement of this simple method for matching means and variances of the images over the linear regression in coordinate y. However, other suitable solutions to the problem of matching intensities of two brain MRI could be proposed and possibly improve the results of the template-based bias correction. Particularly, it would be desirable to incorporate a method which can capture non-linear relationship between the intensities of brain MRI acquired using different imaging protocols.

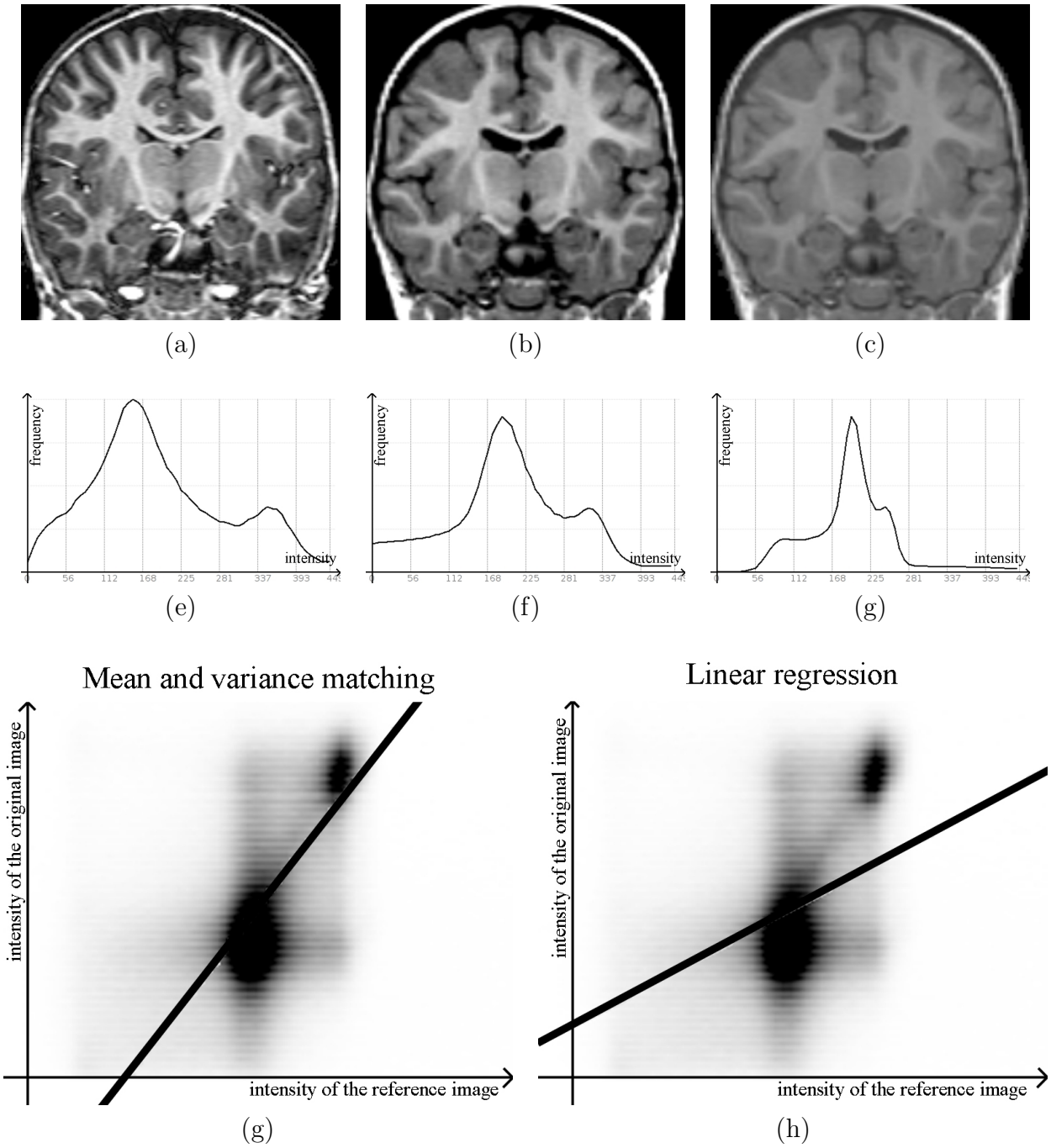


Figure 5.7: Intensity matching: (a) original image; (b) reference image transformed by adjusting intensity mean and variance to original image; (c) reference image transformed by least square fitting in coordinate y results in poorer match. Intensity histograms of the images are shown in (d), (e), and (f). 2D intensity histogram of original image and aligned reference image with linear transformation obtained by: (g) adjusting intensity mean and variance; (h) linear regression in coordinate y. The linear transformation obtained by adjusting intensity mean and variance (d) provides more precise fit.

5.10 Template-based bias correction algorithm

In this section we will summarize the intensity inhomogeneity correction algorithm which we have developed in this chapter. Unlike in the previous sections, here X and Y denote the images in original position and with original intensities, before any alignment or the logarithmic transformation was performed.

When calculating the bias field in the EM framework, logarithmically transformed intensities of template image X and image to be segmented Y need to be used. However, the logarithmic transformation reduces the contrast between higher intensities (usually GM and WM in T1-weighted images), thus making the registration and matching less sensitive to GM/WM contrast. Registration and intensity matching are therefore better performed with original image intensities.

Template-based bias correction

Register the template image X to the image $Y = (y_1, ..., y_n)$, $y_i = Y(v_i)$ using transformation $\alpha = \mathcal{T}_{Y \to X}$ by maximizing normalized mutual information to obtain the aligned template $\alpha X = X \circ \mathcal{T}_{Y \to X}$ where $\alpha X(v_i) = X(\mathcal{T}_{Y \to X}(v_n))$.

Find the linear transformation ι to match the intensity of aligned template image αX and the image Y to obtain intensity matched aligned template $X' = \iota(\alpha X)$, $x'_i = X'(v_i) = \iota(\alpha X(v_i))$:

$$x'_i = \begin{cases} \sqrt{\frac{Var(Y)}{Var(\alpha X)}}(\alpha X(v_i) - E(\alpha X)) + E(Y) & \text{if} \quad Cov(\alpha XY) > 0 \\ -\sqrt{\frac{Var(Y)}{Var(\alpha X)}}(\alpha X(v_i) - E(\alpha X)) + E(Y) & \text{if} \quad Cov(\alpha XY) < 0 \end{cases}$$

Initialize:

$$
\begin{aligned}
m &= 0 \\
c_k^{(0)} &= 0.5,\ k = 0, 1 \\
\sigma_0^{(0)} &= 0.25 Var(\log Y - \log X') \\
\sigma_1^{(0)} &= 4 Var(\log Y - \log X') \\
w_i^{(0)} &= 1,\ i = 1, ..., n \\
b_i^{(0)} &= 0,\ i = 1, ..., n
\end{aligned}
$$

Repeat

$$
\begin{aligned}
m &= m + 1 \\
p_{ik}^{(m)} &= \frac{c_k^{(m-1)} G(\log y_i - \log x'_i - b_i^{(m-1)}, 0, \sigma_k^{(m-1)})}{\sum_l c_l^{(m-1)} G(\log y_i - \log x'_i - b_i^{(m-1)}, 0, \sigma_l^{(m-1)})},\ i = 1, ..., n,\ k = 0, 1 \\
c_k^{(m)} &= \sum_{i=1}^{n} \frac{p_{ik}^{(m)}}{n},\ k = 0, 1 \\
\sigma_k^{(m)} &= \frac{\sum_{i=1}^{n} p_{ik}^{(m)} (\log y_i - \log x'_i - b_i^{(m-1)})^2}{\sum_{i=1}^{n} p_{ik}^{(m)}},\ k = 0, 1 \\
w_i^{(m)} &= \sum_{k=0}^{1} \frac{p_{ik}^{(m)}}{(\sigma_k^{(m)})^2},\ i = 1, ..., n \\
\mathbf{d}^{(m)} &= (\mathbf{N}^T \mathbf{W}^{(m)} \mathbf{N} + 2\lambda \mathbf{H})^{-1} \mathbf{N}^T \mathbf{W}^{(m)} (\log Y - \log X'), \\
b_i^{(m)} &= \sum_{j=-1}^{L} d_j^{(m)} \mathcal{N}_j(v_i),\ i = 1, ..., n \\
L^{(m)} &= -\sum_{i=0}^{n} \log \sum_{k=0}^{1} c_k^{(m)} G(\log y_i - \log x'_i - b_i^{(m)}, 0, \sigma_k^{(m)}) + \lambda (\mathbf{d}^{(m)})^T \mathbf{H} \mathbf{d}^{(m)}
\end{aligned}
$$

until $\frac{L^{(m-1)} - L^{(m)}}{L^{(m)}} < 0.001$

Calculate the bias corrected image Y', $y'_i = Y'(v_i) = y_i \mathrm{e}^{-b_i}$

5.11 Conclusion

In the presence of strong bias field, both segmentation and registration techniques often fail due to the distortion of the image intensities resulting from this artefact. While intensity-based tissue classification is disrupted by blurring between the intensity distributions, registration of the two images can be guided by the gradient of the bias field rather then intensity features resulting from the anatomy of the brain.

In this chapter we described a novel template-based algorithm for bias correction of brain MRI. The advantage of this algorithm is its ability to deal with strong bias fields. In addition, the bias field is calculated by comparing the image with a template image without bias, thus the technique is not biased towards any simplified theoretical model of brain tissue intensities. The method can be used as a preprocessing step for tissue classification. It is also ideal to be used in conjunction with image registration, as the bias field is estimated by comparing the image to be corrected with the registered template image. Another advantage of the method is that it does not require precise registration of the images thanks to the outlier detection and use of weighted least squares to estimate the bias field.

To achieve good bias correction results, a suitable B-spline control point spacing and weighting coefficient λ have to be experimentally determined. Chapter 6 deals with the practical application of this bias correction method to enhance segmentation and registration of MRI with strong intensity inhomogeneity.

Chapter 6

Intensity inhomogeneity correction and segmentation of brain MRI

The aim of this chapter is to develop an optimal strategy for the segmentation of brain MRI during early childhood using the segmentation techniques developed in chapter 4 and the intensity inhomogeneity correction technique developed in chapter 5. In particular we will focus on data acquired on a 3T MR scanner in which the intensity inhomogeneity is far more severe than in data acquired on a 1T MR scanner used in chapter 4.

This chapter is organized as follows: First, the optimal parameters for the template based bias correction are determined. Secondly, the three segmentation methods described in chapter 4 (registration-based segmentation, EM-based segmentation with population-specific atlas and combined registration-based and EM-based segmentation) combined with the template-based bias correction are evaluated and compared. Finally, we demonstrate the application of the novel combined segmentation method in conjunction with the novel template bias correction method for volumetric measurements of brain structures.

6.1 Data and Materials

The images segmented in this chapter are T1-weighted MRI of 35 two years old subjects with three of the subject scanned also at one year of age. The images were acquired on a Phillips 3T scanner using MP RAGE imaging sequence [14] and reconstructed with voxel size $0.8\times0.8\times0.8$ mm^3. Imaging at higher magnetic field strengths, such as 3T, enables images to be acquired with higher resolution. However, it also results in significant increase of intensity inhomogeneity.

Four of those images (three images of two year old children and one image of an one year old child) were manually segmented, including a complete manual segmentation of the caudate and thalamus, and a partial manual segmentation of WM and GM on 7-8 slices. The performance of the automated segmentation methods is measured by agreement with these manual segmentations using the Dice metric [30].

The image used as a template for the bias correction as well as the deformable atlas for the registration based segmentation and the combined registration-based and EM-based segmentation in this chapter is the one introduced in chapter 4: It consists of a 1T brain MRI of a two year old child after bias correction with N3 [78] and a manual segmentation of 11 structures.

The probabilistic atlases used for the EM-based segmentation are the population-specific atlases at one year and two years of age, constructed from 1T images as described in chapter 4.

6.2 Determining the parameters for template-based bias correction

There is a large range of possible values for the parameters of the template-based bias correction algorithm described in the previous chapter. These parameters include the B-spline control point spacing for modelling the bias bield and the weighting parameter λ for the penalty term which controls the smoothness of the bias field. The optimal values may depend on the shape

and frequency of the bias field as well as the level of misregistration between template image and image to be corrected. A fundamental difficulty arises from the fact that the ground truth for the bias corrected image is not known. Therefore we will pursue an approach which measures the performance of the bias correction method by its influence on the performance of the subsequent registration-based and EM-based segmentation methods.

In the following, we will use the number of control points L as a parameter rather than control point spacing. Note, that the number of control points for a bias field represented with cubic B-splines is denoted as $(L_1+2)\times(L_2+2)\times(L_3+2)$. If the boundary conditions are applied, the number of control points reduces to $L_1 \times L_2 \times L_3$. We set the number of control points to equal in all directions $L = L_1 = L_2 = L_3$, as this simplifies the calculations. However, in general the number of control points in different directions may vary. The relationship between the number of control points L in one dimension and the corresponding B-spline control spacing is shown in table 6.1.

L	2	3	4	5	6	7	9	12
spacing in mm - x	129	64	43	32	26	21	16	12
spacing in mm - y	189	94	63	47	38	31	23	17
spacing in mm - z	144	72	48	36	29	24	18	13

Table 6.1: Conversion table between a number of control points L at each dimension and B-spline control point spacing of the model for bias field.

To perform the template-based bias correction, both images are subsampled to a resolution of $3\times3\times3$ mm^3. This reduces the run-time of the bias correction algorithm from approximately 1h to several minutes.

6.2.1 Estimating parameter λ

To initialize our search for the optimal parameters, we aim to obtain an approximate estimate of optimal weighting coefficient λ for the smoothness constraint. In this experiment we used the simplest alignment of the template image in the form of a rigid registration and a B-spline control point spacing of approximately 30 mm ($6 \times 6 \times 6$ mesh of B-spline control points with

$L = 6$, $\delta_1 = 26$, $\delta_2 = 38$, $\delta_3 = 29$). The values of λ to be tested were determined as follows: $\lambda = 10^0, 10^2, 10^4, 10^6, 10^8$.

First, the template image is registered to the four test images. All test images are then corrected using the different values of λ. Secondly, the EM-based segmentation and the registration-based segmentation are carried out on all four bias-corrected images. The non-rigid registration is optimized for speed and is performed at three different resolutions levels with the parameters shown in table 6.2.

Control point spacing	image resolution	blurring	finite difference step size
20mm	3r	2r	8r
10mm	2r	r	4r
5mm	r	0.5r	2r

Table 6.2: Parameters of the non-rigid registration: The columns of the table are the B-spline control point spacing at a given level of registration, resolution of the subsampled image, standard deviation of Gaussian blurring used before subsampling and length of steps when optimizing the non-rigid transformation. r denotes the largest dimension of a voxel in mm, in either the target or source image.

The results are evaluated on four manual segmentations to determine the value of λ resulting in the best performance of those two segmentation methods. The results are shown at fig. 6.1. The segmentation agreement measured by the Dice metric [30] shows increases in performance until λ reaches value 10^4. If λ is further increased, the performance of segmentation methods sharply drops.

It is important to note, that the smoothness constraint as defined in eq. 5.3 depends on the parameterization of the B-spline model for the bias field. As a result, the optimal value of λ varies with the units in which we measure the control point spacing δ_1, δ_2, δ_3. In other words, working in millimeters would yield a different value of λ than working in centimeters. Therefore if we want to find the generally applicable values for parameter λ, we have to demonstrate this dependence of λ on the units used for B-spline control point spacing.

To understand this dependence, we can determine how the value of λ has to change if we multiply the size of the image at each dimension by a constant c while keeping the original number of control points fixed. The new size of the image in dimension i is $X'_i = cX_i$ and

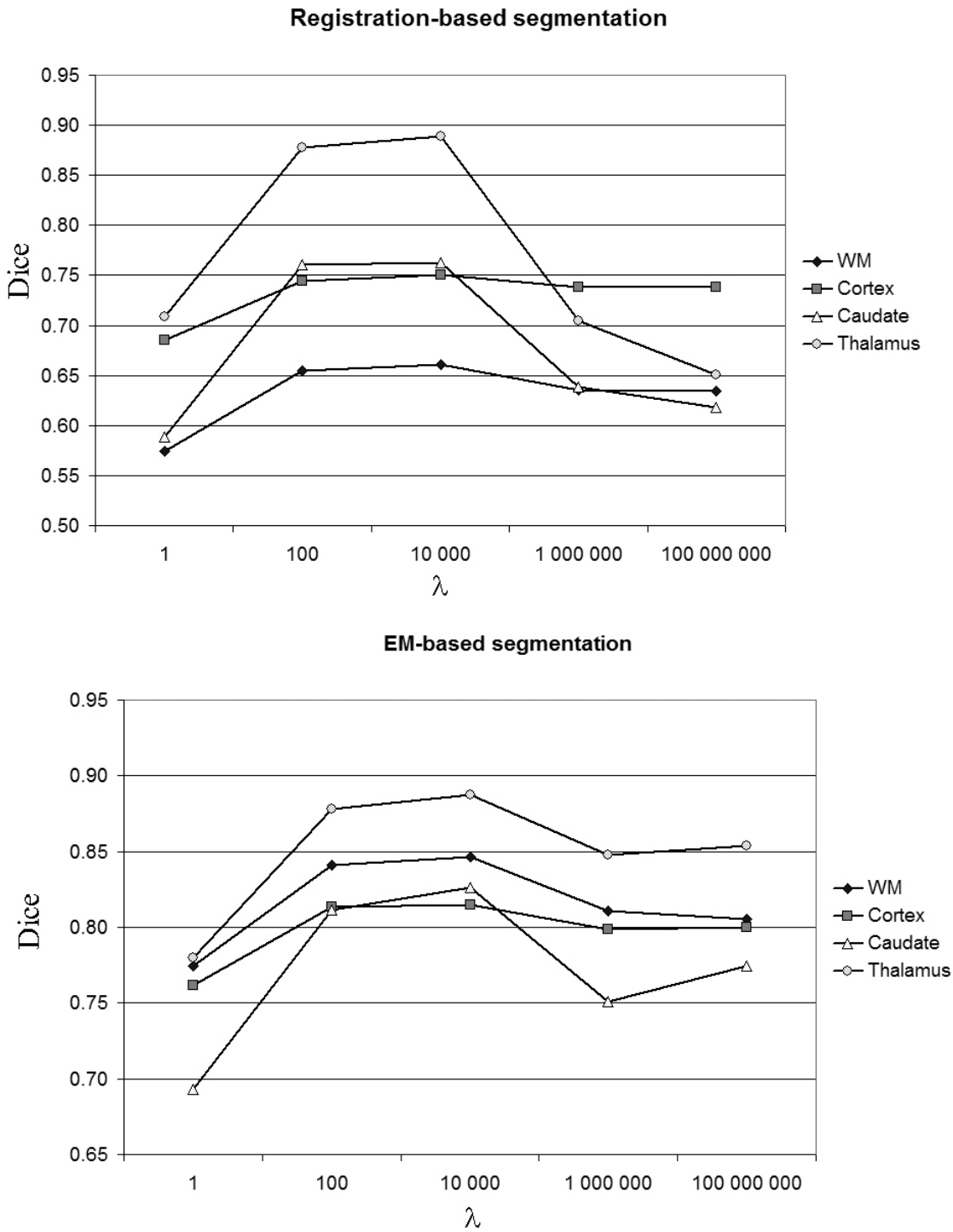


Figure 6.1: Segmentation results following template-based bias correction using different values of λ. All images were bias corrected with $L = 6$, corresponding to control point spacing of approximately 30 mm. The value of λ is shown on the horizontal axis. The agreement with the manual segmentation as calculated by the Dice metric is shown on the vertical axis.

consequently the new control point spacing is $\delta'_i = c\delta_i$ since the relationship $X_i = (L_i - 1)\delta_i$ holds. Let v'_i be the new parameterization of the bias field, with $v'_i = cv_i$. The objective

function is in the form:

$$\log P(\beta, \gamma|Y) = \log P(Y|\beta, \gamma) + \lambda C_{smooth}$$

With the change of the parameterization the smoothness constraint C_{smooth} will change to C'_{smooth} and consequently the weighting coefficient λ has to change to preserve the same weighting in the objective function. Therefore:

$$\lambda C_{smooth} = \lambda' C'_{smooth}$$

Let us denote

$$F_{ij}(v_1, v_2, v_3) = \left(\frac{\partial^2 S(\frac{v_1}{\delta_1}, \frac{v_2}{\delta_2}, \frac{v_3}{\delta_3})}{\partial v_i \partial v_j} \right)^2$$

Then

$$\begin{aligned} F'_{ij}(v'_1, v'_2, v'_3) &= \left(\frac{\partial^2 S(\frac{v'_1}{\delta'_1}, \frac{v'_2}{\delta'_2}, \frac{v'_3}{\delta'_3})}{\partial v'_i \partial v'_j} \right)^2 \\ &= \left(\frac{\partial^2 S(\frac{v_1}{\delta_1}, \frac{v_2}{\delta_2}, \frac{v_3}{\delta_3})}{\partial v_i \partial v_j} \frac{\partial v_i}{\partial v'_i} \frac{\partial v_j}{\partial v'_j} \right)^2 \\ &= \frac{1}{c^4} F_{ij}(v_1, v_2, v_3) \end{aligned}$$

It follows that

$$\begin{aligned} \frac{1}{V'} C'_{abc} &= \frac{1}{X'_1 X'_2 X'_3} \int_0^{X'_1} \int_0^{X'_2} \int_0^{X'_3} F'_{ij}(v'_1, v'_2, v'_3) dv'_1 dv'_2 dv'_3 \\ &= \frac{1}{c^3 X_1 X_2 X_3} \int_0^{X_1} \int_0^{X_2} \int_0^{X_3} \frac{1}{c^4} F_{ij}(v_1, v_2, v_3) \frac{dv'_1}{dv_1} \frac{dv'_2}{dv_2} \frac{dv'_3}{dv_3} dv_1 dv_2 dv_3 \\ &= \frac{1}{c^3 X_1 X_2 X_3} \int_0^{X_1} \int_0^{X_2} \int_0^{X_3} \frac{1}{c^4} F_{ij}(v_1, v_2, v_3) c^3 dv_1 dv_2 dv_3 \\ &= \frac{1}{c^4 V} C_{abc} \end{aligned}$$

Obviously

$$C'_{smooth} = \frac{1}{c^4} C_{smooth}$$

and consequently

$$\lambda' = \lambda \frac{C_{smooth}}{C'_{smooth}} = c^4 \lambda$$

Practically, this means, that if we change units from millimeters to centimeters, resulting in $c = 0.1$, then $\lambda_{cm} = \lambda_{mm} 10^{-4}$. We determined that the optimal λ when working in millimeters is around value 10^4. Changing units to centimeters would bring optimal value of λ around 1, which is a good user-friendly option. From now on, we will therefore assume that we are working in centimeters, which is equivalent to working in millimeters with the objective function:

$$\log P(\beta, \gamma | Y) = \log P(Y | \beta, \gamma) + 10^4 \lambda C_{smooth}$$

.

6.2.2 Two-dimensional search for the parameters λ and L

The performance of the bias correction was also calculated for other numbers of B-spline control points with $L = 3, 6, 9$. Fig. 6.2 shows the influence of the bias correction parameters L and λ on registration-based segmentation. Similarly, fig. 6.3 shows the influence of bias correction parameters L and λ on EM-based segmentation. These results show improvement of the segmentation results if L is increased from 3 to 6. However, further increases of L do not seem to produce any consistent improvement.

To choose the optimal value of λ we refined our search by testing values closer to the optimal value $\lambda = 1$ as determined in the previous section. Tested values were $\lambda = 0.1, 0.5, 1, 2, 3, 5, 10, 20$. The figures 6.2 and 6.3 show that the performance of both segmentation methods drops when λ reaches a value of 5, except for a marginal increase in the cortical overlap in the EM segmentation. We will concentrate of values of $\lambda = 0.1, 0.5, 1, 2, 3$ for which we have performed a slower and more accurate registration-based segmentation using the parameters shown in table 6.3.

The results in fig. 6.4 show that all values of λ in the range $< 0.1, 3 >$ result in similar performance. Fig. 6.5 shows the results for the EM-based segmentation already presented in

Control point spacing	image resolution	blurring	finite difference step size
20mm	4r	2r	10r
10mm	2r	r	5r
5mm	r	0.5r	2.5r
2mm	-	-	2.5r

Table 6.3: Parameters for the registration at four resolution levels. The columns of the table are B-spline control point spacing at a given resolution level, resolution of the subsampled image, variance of Gaussian blurring used before subsampling and the step size of the finite difference approximation of the derivatives during the optimization of the transformation. r denotes the largest dimension of a voxel in mm, in either target or source image.

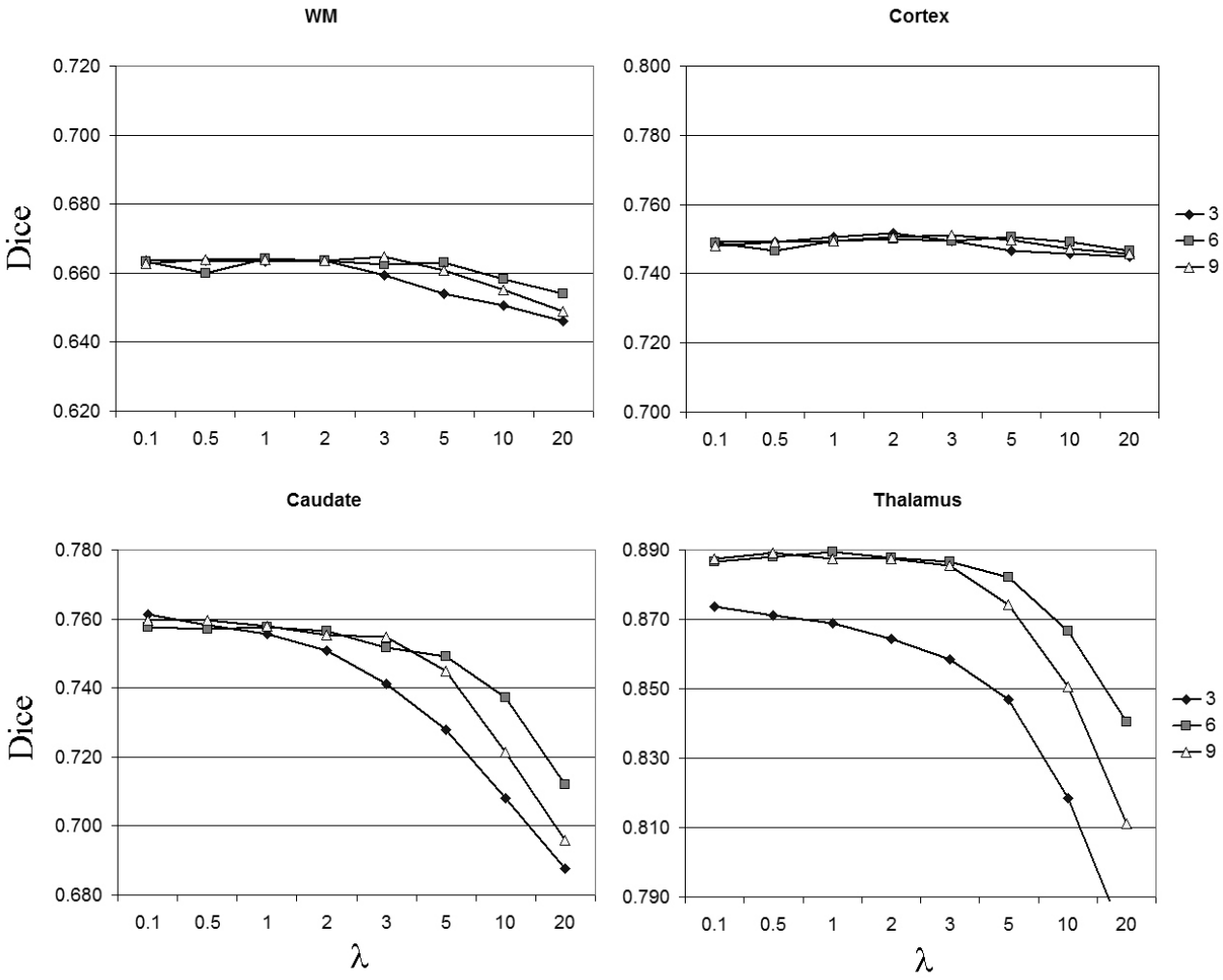


Figure 6.2: Registration-based segmentation using the parameters from table 6.2. Before segmentation the images were bias corrected with parameters λ shown on horizontal axis and L shown as a series of graphs. The template was registered with the images for the purposes of bias correction with a control point spacing 5mm. The agreement with the manual segmentation as calculated by the Dice metric is shown on vertical axis.

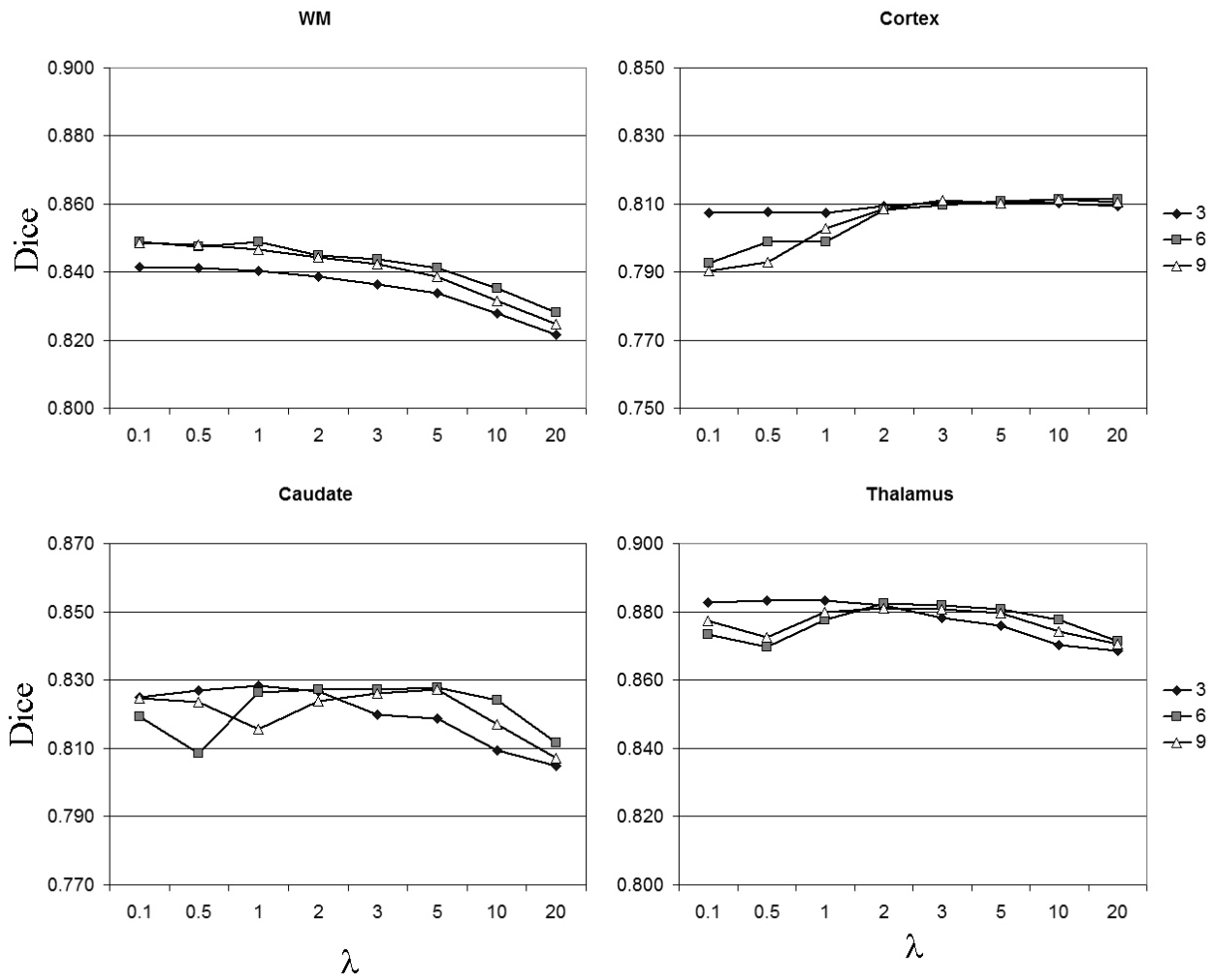


Figure 6.3: EM-based segmentation. Before segmentation the images were bias corrected with parameters λ shown on horizontal axis and L shown as a series of graphs. The template was registered with the images for the purposes of bias correction with a control point spacing 5mm. The agreement with the manual segmentation as calculated by the Dice metric is shown on vertical axis.

fig. 6.3, but in a form which is equivalent to fig. 6.4. The Dice overlap for cortex, caudate and thalamus is best for values $\lambda = 2, 3$. The performance for WM slightly decreases with increasing λ. However, the reason for this is not an improved bias correction. Fig. 6.6 shows that for lower values of lambda, the centre of the brain is darker than the tissues close to the cortex. Even though this is not a visually appealing result of the bias correction, it improves the EM-based segmentation of WM close to the cortical region as shown in fig. 6.7. The inconsistent performance of the EM-based segmentation for caudate and thalamus with lower

values of lambda has a similar reason: Lower values of λ result in a darker centre of the brain producing less contrast between WM and subcortical GM. The results therefore depend on the quality of the registration of the probabilistic atlas rather than on the intensity patterns of central brain structures.

To conclude, we observe, that the results of the EM-based segmentation are also influenced by other factors, not only the homogeneity of intensities within each tissue class. In addition, the registration-based segmentation has a relatively low sensitivity towards any residual bias fields and therefore is not a good measure for fine-tuning of the parameter values for the bias correction. However, the bias corrected images for values $\lambda = 2, 3$ and $L = 6, 9$ seem to produce good segmentation results for both methods and are visually plausible. For our further experiments we choose parameters $\lambda = 2$ and $L = 6$, even though there are other parameter combinations which can produce comparable bias correction results.

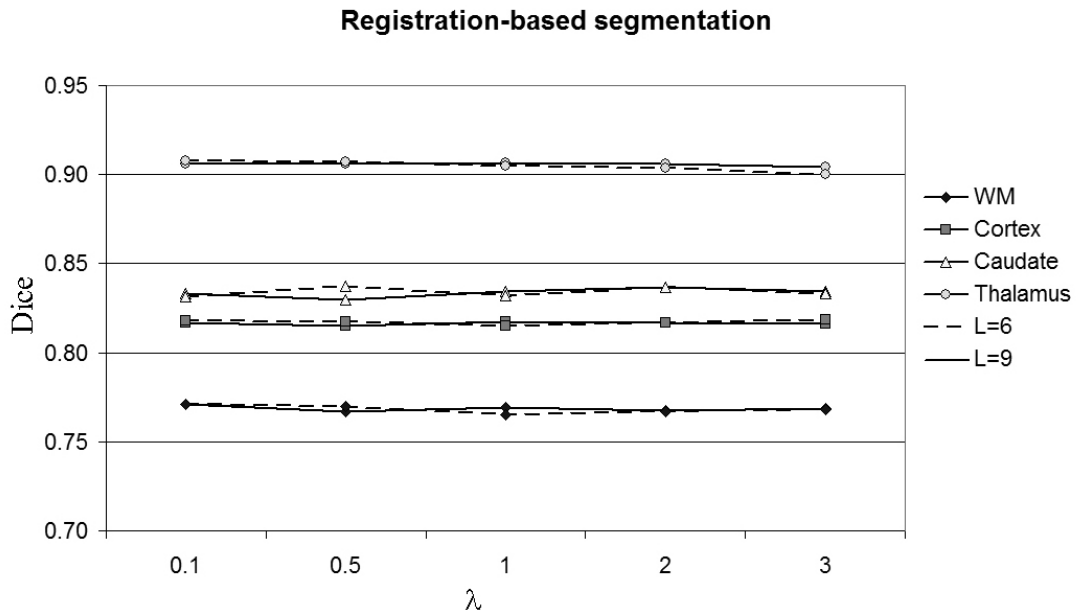


Figure 6.4: Registration-based segmentation with parameters from table 6.3. Before segmentation the images were bias corrected with the parameters λ shown on the horizontal axis and L shown as a series of graphs. The template was registered with the images for the purposes of bias correction with a control point spacing of 5mm. The agreement with the manual segmentation as calculated by the Dice metric is shown on the vertical axis.

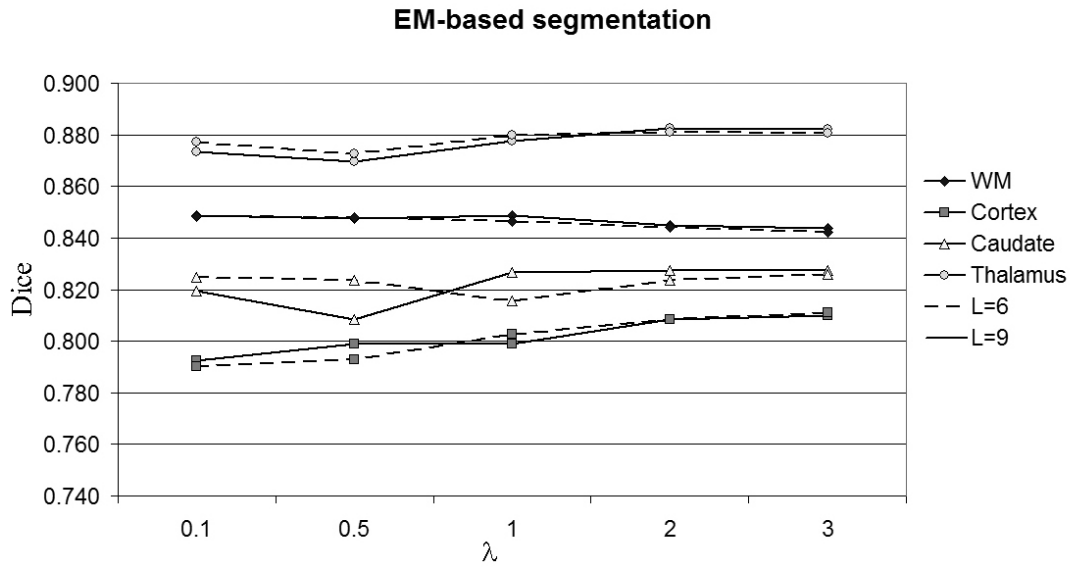
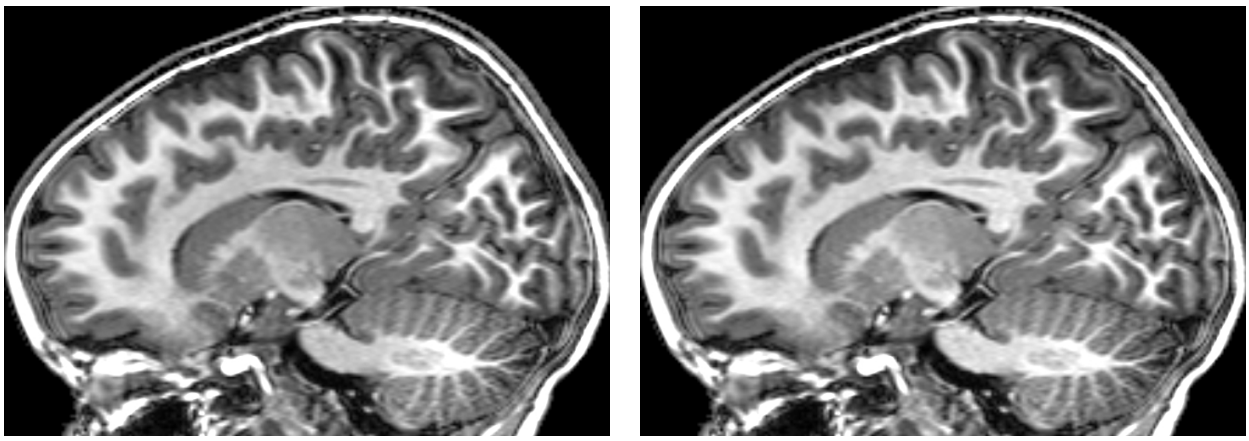
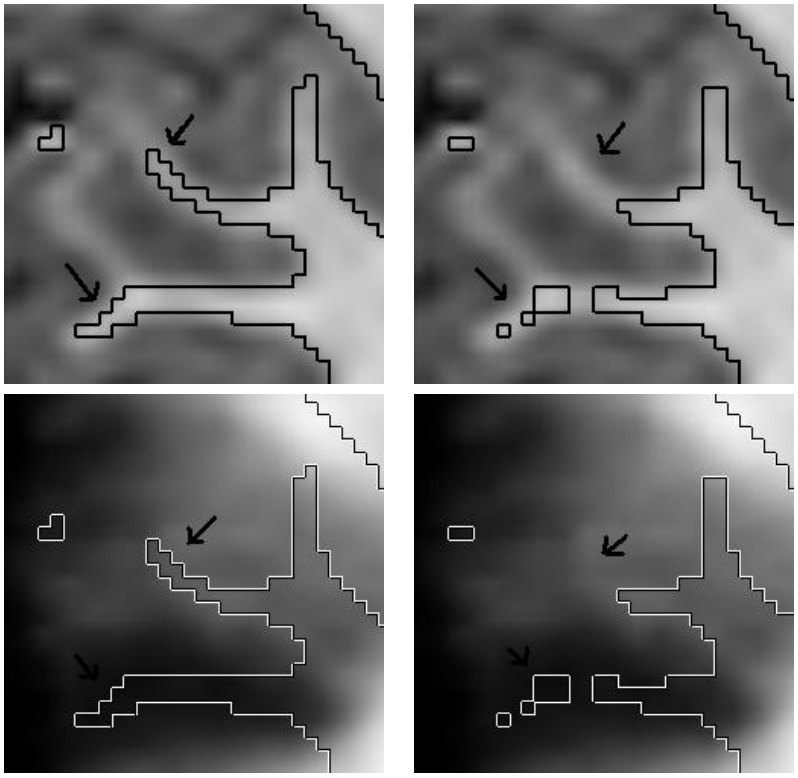


Figure 6.5: EM-based segmentation. Before segmentation the images were bias corrected with the parameters λ shown on the horizontal axis and L shown as a series of graphs. The template was registered with the images for the purposes of bias correction with a control point spacing of 5mm. The agreement with the manual segmentation as calculated by the Dice metric is shown on the vertical axis.



(a) (b)

Figure 6.6: The influence of the parameter λ on the bias correction. Values of $\lambda = 0.1$ results in a darker centre of the brain compared to the tissue closer to the skull and less contrast in the subcortical area than values $\lambda = 3$: (a) $\lambda = 0.1$; (b) $\lambda = 3$.



(a) (b)

Figure 6.7: The influence of the parameter λ on the EM-based segmentation of WM. First row: Image with overlayed segmentation. Second row: Prior probability maps for WM with overlayed segmentation. (a) $\lambda = 0.1$; (b) $\lambda = 3$. Value $\lambda = 0.1$ results in a brighter tissue closer to the skull compared to the center of the brain. WM close to skull can be therefore still recognized by EM segmentation due to very high intensity values, even though the prior probabilities in the location are low. Examples pointed out by the arrows.

6.3 Registration-based segmentation

In the previous section we have determined suitable parameters for the template-based bias correction. Our parameters of choice were $L = 6$ for the number of B-spline control points for the model of bias field in each dimension and $\lambda = 2$ as a weighting coefficient for the smoothness constraint for the bias field. In this section, the registration parameters for 3T data will be determined, similarly as in sec. 4.2 for 1T data. In addition, we will combine the non-rigid registration and template-based bias correction to perform registration-based segmentation in the presence of strong intensity inhomogeneity.

To find the optimal strategy for the coupling the registration-based segmentation with template-based bias correction for MRI with strong bias fields, we need to determine the level of alignment of the template image with the image to be corrected to achieve the best possible bias correction. Moreover, it has to be investigated, whether performing bias correction several times during registration process can improve the final outcome of the algorithm.

6.3.1 Determining the registration parameters

The parameters for the non-rigid registration providing an optimal trade-off between speed and performance of the registration-based segmentation were determined similarly to sec. 4.2. We tested four different strategies for the choice of the registration parameters given in table 6.4. The non-rigid registrations using these four sets of parameters were performed to align the 1T bias corrected template image and the four manually segmented 3T images. After rigid and affine alignment, the 3T images were intensity-corrected using the template-based bias correction with parameters $L = 6$ and $\lambda = 2$. For simplicity, no bias correction was performed after non-rigid alignment was started. In the next section we will investigate whether updating of the bias field estimate during non-rigid alignment can improve the registration results.

The evaluation results for different parameter values are presented in table 6.5. The results show that a constant step size of 8mm is necessary to achieve good alignment at all resolution levels. However, subsampling to isotropic voxels sizes of 3.2 mm, 1.6 mm in resolution levels

20 mm and 10 mm respectively, and subsampling to isotropic voxel sizes of 3.2 mm, 2.4 mm, 1.6mm in resolution levels 20 mm, 10 mm and 5 mm yields the same results at final control point resolution of 2.5mm. Therefore the computational speed of the registration can be increased without reducing the quality of the final alignment. In the remainder of the chapter, the parameters given in table 6.6 will be used.

Parameter set #1

control point spacing	image resolution	image blurring	finite difference step size
20	3.2	1.6	8
10	1.6	0.8	8
5	full	0.4	8
2.5	full	0	8

Parameter set #2

control point spacing	image resolution	image blurring	finite difference step size
20	3.2	1.6	8
10	1.6	0.8	6.4
5	full	0.4	4.8
2.5	full	0	4

Parameter set #3

control point spacing	image resolution	image blurring	finite difference step size
20	3.2	1.6	8
10	2.4	1.2	8
5	1.6	0.8	8
2.5	full	0	6.4

Parameter set #4

control point spacing	image resolution	image blurring	finite difference step size
20	3.2	1.6	8
10	2.4	1.2	8
5	1.6	0.8	8
2.5	full	0	8

Table 6.4: Parameters for the non-rigid registration. All sizes are given in millimeters.

Parameter set	WM	Cortex	Caudate	Thalamus
#1	0.79 ± 0.03	0.83 ± 0.02	0.85 ± 0.02	0.90 ± 0.01
#2	0.77 ± 0.03	0.82 ± 0.01	0.84 ± 0.04	0.90 ± 0.01
#3	0.79 ± 0.03	0.83 ± 0.02	0.84 ± 0.04	0.90 ± 0.01
#4	0.79 ± 0.03	0.83 ± 0.02	0.86 ± 0.03	0.91 ± 0.01

Table 6.5: Agreement of the manual segmentation with the results of the registration-based segmentation with different parameters given in table 6.4 as measured with the Dice overlap.

control point spacing	image resolution	image blurring	finite difference step size
20	3.2	1.6	8
10	2.4	1.2	8
5	1.6	0.8	8
2.5	full	0	8

Table 6.6: Optimal choice of parameters for non-rigid registration of 3T images. All sizes are given in millimeters.

6.3.2 Interleaved versus sequential registration

Simultaneous algorithms for segmentation, registration and bias correction have been very popular in recent years [7], [67], [99], following the idea, that if those processes are iterated at the same time, better results can be achieved. We therefore investigated, whether simultaneous registration and template-based bias correction would produce better results than a simple sequential approach. Let Y denote the image containing a bias field and X the template image with no bias. We tested the following two algorithms:

Sequential registration and bias correction

1. Register image Y with template image X with affine registration $\mathcal{A}_{Y\to X}$
2. Find bias corrected image Y' using template-based bias correction (sec. 5.10) with registered template $\alpha X = X \circ \mathcal{A}_{Y\to X}$
3. Register the corrected image Y' with the template image X with non-rigid registration $\mathcal{T}_{Y'\to X}$ using the parameters in table 6.6 and initial affine alignment $\mathcal{A}_{Y\to X}$.

Interleaved registration and bias correction

1. Register image Y with template image X using affine registration $\mathcal{A}_{Y\to X}$
2. For each resolution level 20 mm, 10 mm, 5 mm and 2.5 mm and parameters in table 6.6
   a. Find bias corrected image Y' using template-based bias correction (sec. 5.10) of image Y with template αX aligned at previous level of registration

b. Refine registration $\alpha = \mathcal{T}_{Y' \to X}$ at current resolution level using latest corrected image Y'.

The comparison of these two algorithms on the four manually segmented images shows that the sequential and interleaved algorithms produces registrations of the same quality, see table 6.7. It seems that therefore there is no need for the simultaneous algorithm and that the simple sequential approach is sufficient.

parameters	WM	Cortex	Caudate	Thalamus
sequential	0.79 ± 0.03	0.83 ± 0.02	0.86 ± 0.03	0.91 ± 0.01
interleaved	0.79 ± 0.03	0.83 ± 0.02	0.86 ± 0.03	0.91 ± 0.01

Table 6.7: Agreement of the manual segmentation with the results of the registration-based segmentation obtained using interleaved and sequential registration and bias correction algorithm as measured with the Dice overlap.

6.3.3 Registration-based segmentation algorithm in the presence of strong intensity inhomogeneity

In the previous sections we found, that the simple sequential approach is sufficient for the registration and bias correction of a pair of images. Thanks to the outlier detection within the EM framework, the affine alignment of the template image X with the image Y is sufficient for optimal bias correction. We also derived optimal parameters for the best trade-off between speed and performance of registration-based segmentation. The strategy for the registration-based segmentation in presence of strong bias fields can now be summarized as follows:

Registration-based segmentation and template-based bias correction

---

1. Given image Y containing a bias field and template image X, perform an affine registration $\mathcal{A}_{Y \to X}$ and find the corrected image Y' using template-based bias correction (sec. 5.10) with the aligned template $\alpha X = X \circ \mathcal{A}_{Y \to X}$.

2. Find a non-rigid registration $\mathrm{T}_{Y' \to X}$ of images X and Y' using the parameters given in table 6.6 and the initial affine alignment $\mathcal{A}_{Y \to X}$.

3. Compute the registration-based segmentation $S_{rbs}(Y)$ of the image Y using binary manual segmentations $S_{X,l}$ of each structure l in the template image X using the following equation:

$$S_{rbs}(Y)(v_i) = \max_l S^{lin}_{X,l}(\mathcal{T}_{Y' \to X}(v_i))$$

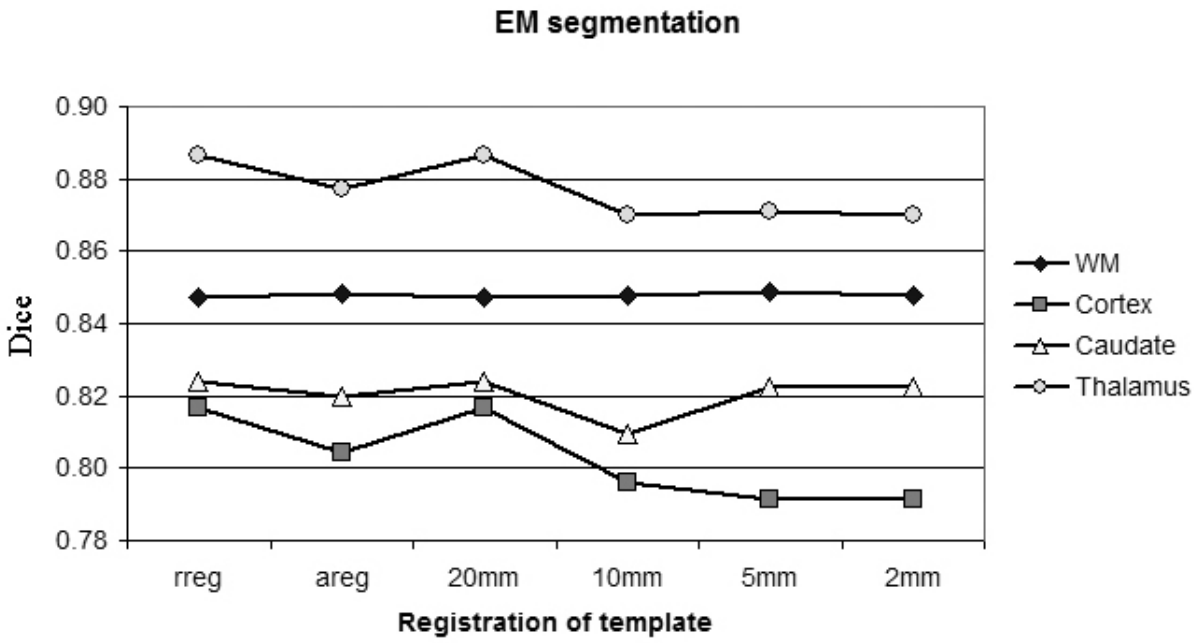
where $S^{lin}_{X,l}$ denotes the linear interpolation of image $S_{X,l}$.

6.4 EM segmentation with a population-specific probabilistic atlas

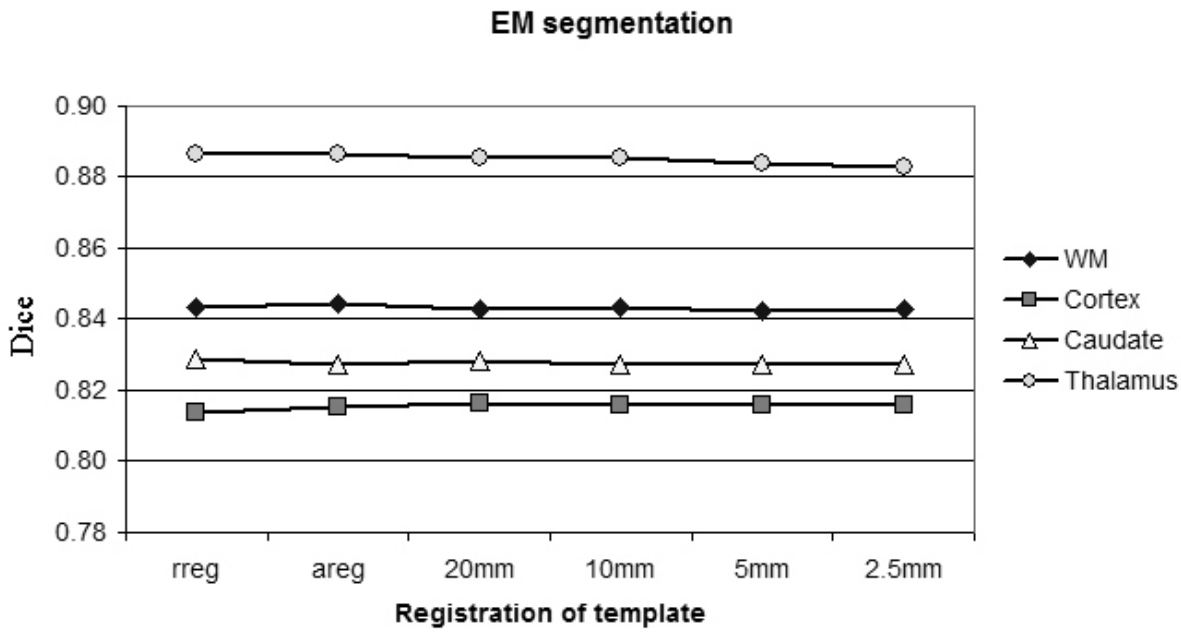
In chapter 4 we have assumed that the bias field was removed in the pre-processing step and consequently, we were able to perform the EM segmentation with a population-specific atlas without interleaving the bias field correction within the EM framework. In this section we will investigate the best strategy for combining the EM segmentation with the template-based bias correction to create an algorithm for EM-based segmentation in the presence of strong bias fields. This algorithm forms an alternative to the simultaneous segmentation and bias correction a proposed in [49] and [7].

6.4.1 Registration of template image during bias correction and EM segmentation

If we are able to correct the images in the pre-processing step, the EM segmentation (section 2.5.3) can be performed in the same way as is section 4.3. The aim of this section is to find the optimal template-based bias correction algorithm to serve as a pre-processing bias correction step for the EM segmentation. We already determined suitable parameters for the proposed bias correction algorithm ($L = 6$ and $\lambda = 2$). In this section we would like to investigate whether the precision of the alignment of the template image with the image to be corrected while correcting the bias, has any influence on the results of the EM segmentation.



(a)



(b)

Figure 6.8: EM-based segmentation. The images were bias corrected with parameters $L = 6$ and $\lambda = 2$ after different registrations of the template images shown on the horizontal axis. After bias correction, the population-specific probabilistic atlas was aligned (a) for each bias corrected image separately; (b) using the same registration with image corrected after rigid alignment with the template. The agreement of EM segmentation with manual segmentation calculated by Dice metric is shown on the vertical axis.

To find the influence of the precision of the alignment of the template image with the image to be corrected during the bias correction on the resulting segmentation, the image Y was corrected after the following levels of alignment α of the template image: rigid, affine, 20 mm, 10 mm, 5 mm and 2.5 mm alignment of the template image. Then, the EM segmentation was performed on all versions of the bias corrected images Y'_α.

We performed the following two experiments: First, the probabilistic atlas was affinely aligned with each image Y'_α separately, followed by EM segmentation with a population specific probabilistic atlas as described in sec. 4.3.2. To eliminate the influence of the differences of the probabilistic atlas alignment, in the second experiment we used the affine alignment of the probabilistic atlas with image Y'_{rigid} to perform the EM segmentation on all images Y'_α. The evaluations of both experiments on the four test images are shown in fig. 6.8.

The results of neither of these two experiments suggest that there is any advantage in improving the alignment of the template image with the image Y during bias correction for the purposes of the EM segmentation. Fig. 6.8a shows random variations in the performance of the method which is caused by small differences in the affine alignment of the probabilistic atlas resulting from small differences in the bias correction results with different alignments of template image with image Y. This influence is eliminated if the same affine transformation of the probabilistic atlas is used for all images Y'_α, as shown in fig. 6.8b. Consequently it is revealed, that there is no correlation between the level of alignment of the template image during bias correction and performance of EM segmentation.

Therefore, in the following sections, the template based bias correction will be performed only once, after affine alignment of the template with the image.

6.4.2 EM segmentation in the presence of strong intensity inhomogeneity

The EM segmentation algorithm with the population specific atlas can be summarized as follows:

EM segmentation and template-based bias correction

1. Given image Y containing a bias field and template image X, perform an affine registration $\mathcal{A}_{Y\to X}$ and find the corrected image Y' using template-based bias correction (sec. 5.10) with the aligned template $\alpha X = X \circ \mathcal{A}_{Y\to X}$.

2. Find the optimal affine registration $\mathcal{A}_{Y'\to A}$ of the population-specific probabilistic atlas A with the bias-corrected image Y'.

3. For each structure l transform the probability maps $\mathbf{p}_l^{atlas}$ from the atlas A to the space of image Y:

$$\mathbf{p}_l^{atlas}(Y) = \mathcal{A}_{Y'\to A}(\mathbf{p}_l^{atlas})$$

4. Initialize the EM segmentation of Y' with the probability maps $\mathbf{p}_l^{atlas}(Y)$ and calculate the soft segmentations $\mathbf{p}_l^{(m)}$ where m is the last iteration of the EM algorithm (sec. 2.5.3).

5. Calculate a hard segmentation

$$S_{em}(Y)(v_i) = \max_l \mathbf{p}_l^m(v_i) = \max_l p_{il}^{(m)}$$

Experimental results for this algorithm are shown in table 6.8. To compare the results, the EM algorithm with the population-specific atlas was also applied to the original images without correcting the bias field, resulting in a significant drop in performance. Unlike in section 4.3, the brain was not extracted with BET, as the performance of BET was very poor on the data used in this chapter. This is probably caused by the poor contrast between GM and CSF.

method	WM	Cortex	Caudate	Thalamus
EM without bias correction	0.79 ± 0.02	0.80 ± 0.02	0.67 ± 0.17	0.82 ± 0.08
EM with bias correction	0.84 ± 0.02	0.83 ± 0.02	0.82 ± 0.03	0.88 ± 0.02

Table 6.8: Performance of EM segmentation with population-specific atlas on four subjects.

6.5 Combined registration-based and EM-based segmentation

The combined segmentation algorithm proposed in section 4.4 was also used to segment the 3T images in this chapter. Let X be the template image with attached binary manual segmentations $S_{X,l}$ for each brain structure l and $X_1, ..., X_q$ the training images corrupted by a bias field. Let Y be the image to be segmented corrupted by a strong bias field. Then the combined registration-based and EM-based segmentation with template-based bias correction can be summarized as follows:

Combined registration-based and EM-based segmentation with template-based bias correction

1. For each training image X_i corrupted by a bias field, perform an affine registration $\mathcal{A}_{X_i \to X}$ with the template image X and find the corrected image X_i' using the template-based bias correction (sec. 5.10) with aligned template $\alpha X = X \circ \mathcal{A}_{X_i \to X}$.

2. Find the optimal non-rigid registrations $\mathcal{T}_{X_i' \to X}$ of images X and X_i' using parameters given in table 6.6 and initial affine alignment $\mathcal{A}_{X_i \to X}$.

3. Calculate the soft registration-based segmentations $S_{X,l}^{lin} \circ \mathcal{T}_{X_i' \to X}$ of the training images, where $S_{X,l}^{lin}$ denotes linear interpolation of the binary image $S_{X,l}$.

4. For each new image Y corrupted by a bias field

   a. Perform affine registrations $\mathcal{A}_{Y \to X}$ of the template image X with the image Y corrupted by a bias field and find the corrected image Y' using the template-based bias correction (sec. 5.10) with aligned template $\alpha X = X \circ \mathcal{A}_{Y \to X}$.

   b. Peform registration-based segmentations $\mathcal{T}_{Y' \to X_i'}$ of the corrected training images and the corrected image Y', using parameters in table 6.9.

   c. For each structure l calculate the probability maps $\mathbf{p}_l(Y)$ of subject specific atlas using

equation

$$\mathbf{p}_l(Y) = \frac{1}{q}\sum_{i=1}^{q} S_{X,l}^{lin} \circ \mathcal{T}_{X_i' \to X} \circ \mathcal{T}_{Y' \to X_i'}$$

d. Initialize the EM segmentation of Y' with the probability maps $\mathbf{p}_l(Y)$ and calculate the soft segmentations $\mathbf{p}_l^{(m)}$ where m is the last iteration of EM algorithm (sec. 2.5.3).

e. Calculate the hard segmentation

$$S_{rbs-em}(Y)(v_i) = \max_l \mathbf{p}_l^m(v_i) = \max_l p_{il}^{(m)}$$

---

control point spacing	image resolution	image blurring	finite difference step size
20	3.2	1.6	8
10	2.4	1.2	8

Table 6.9: Registration parameters for creating subject-specific atlases for 3T images. All sizes are given in millimeters.

This algorithm was performed using 35 training images of two year old children imaged using a 3T scanner and one manually segmented template image of a two year-old child imaged with a 1T scanner and corrected with N3 [78]. Experimental results on four manually segmented images (three subjects scanned at two years and one subject scanned at one year) are shown in table 6.10. For comparison the 35 1T training images used in chapter 4 were also used to create subject specific atlases. These images were not corrected with template-based bias correction, as they were corrected with N3 [78] before. The results show that slightly better results can be obtain when 3T images instead of 1T images are used as training images.

method	WM	Cortex	Caudate	Thalamus
Combined 1T	0.84 ± 0.02	0.84 ± 0.02	0.83 ± 0.02	0.90 ± 0.01
Combined 3T	0.84 ± 0.02	0.84 ± 0.02	0.85 ± 0.02	0.91 ± 0.01

Table 6.10: Performance of EM segmentation with population-specific atlas on four 3T MRI (three at the age of two years and one at the age of one year). Two different groups of training images were used: 35 1T images corrected by N3 [78] (combined 1T) and 35 3T images corrected by proposed template-based bias correction method (combined 3T).

6.6 Comparison of the methods

In this section we compare the performance of the segmentation methods for brain MRI with strong intensity inhomogeneity in a similar fashion to sec. 4.5. The following three methods for the segmentation of brain structures are evaluated:

1. Registration-based segmentation, also called atlas-based segmentation using the parameters in table 6.6, described in sec. 6.3.3.
2. EM-based segmentation with a population-specific atlas for two years of age, described in sec. 6.4.2.
3. Combined registration-based and EM-based segmentation with a subject-specific atlas with precision 10mm, described in sec. 6.5.

In previous sections, all three methods were validated by measuring the agreement with manual segmentations of four 3T subjects containing strong intensity inhomogeneity (three subjects at two years of age and one subject at one year of age) using the Dice overlap metric. The summary of the results is shown in table 6.11. The improvement of the combined method, compared to the registration-based and EM-based segmentation, is very similar to the improvement achieved on 1T data in sec. 4.5. Again, the combined method matches the performance of the EM-based segmentation in WM and GM, while performance in subcortical structures is comparable to that of the registration-based segmentation, thus rendering the combined method superior to both original approaches.

method	WM	Cortex	Caudate	Thalamus
Registration-based	0.79 ± 0.03	0.83 ± 0.02	0.86 ± 0.03	0.91 ± 0.01
EM-based	0.84 ± 0.02	0.83 ± 0.02	0.82 ± 0.03	0.88 ± 0.02
Combined	0.84 ± 0.02	0.84 ± 0.02	0.85 ± 0.02	0.91 ± 0.01

Table 6.11: Performance of different segmentation methods on brain MRI of three subjects at two years of age and one subject at one year of age in the presence strong intensity inhomogeneity.

Fig. 6.9 shows a visual comparison of the results obtained using the three methods. While the registration-based segmentation does not delineate all the boundaries precisely, it appears

correct overall, see fig. 6.9b. The EM segmentation has a tendency to leave out small details and incorrectly locate small structures, such as substantia nigra, misaligned in the probabilistic atlas, see fig. 6.9c. The combined segmentation correctly delineates these boundaries. Moreover, it is able to capture most of the small detail, see fig. 6.9d.

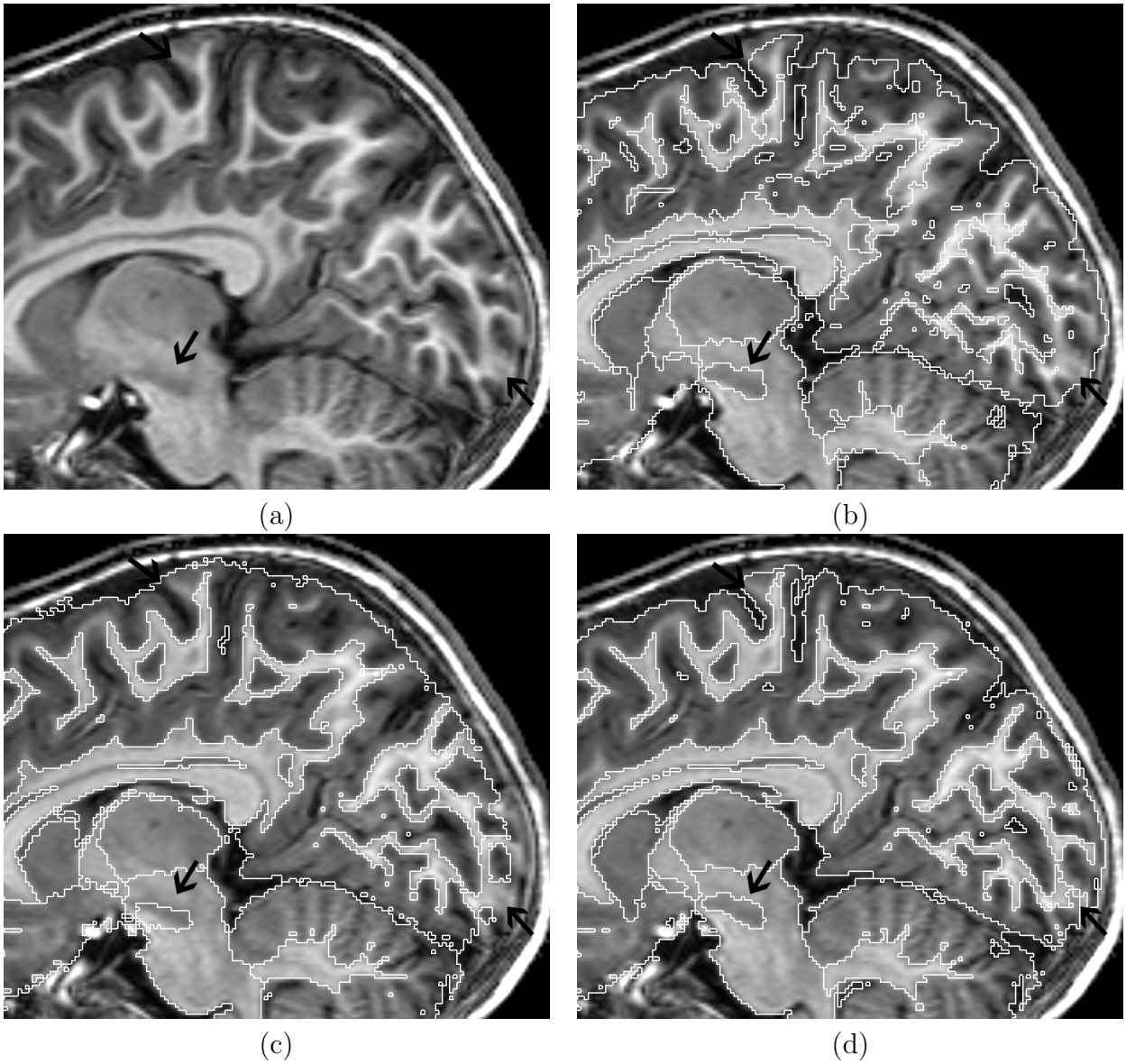
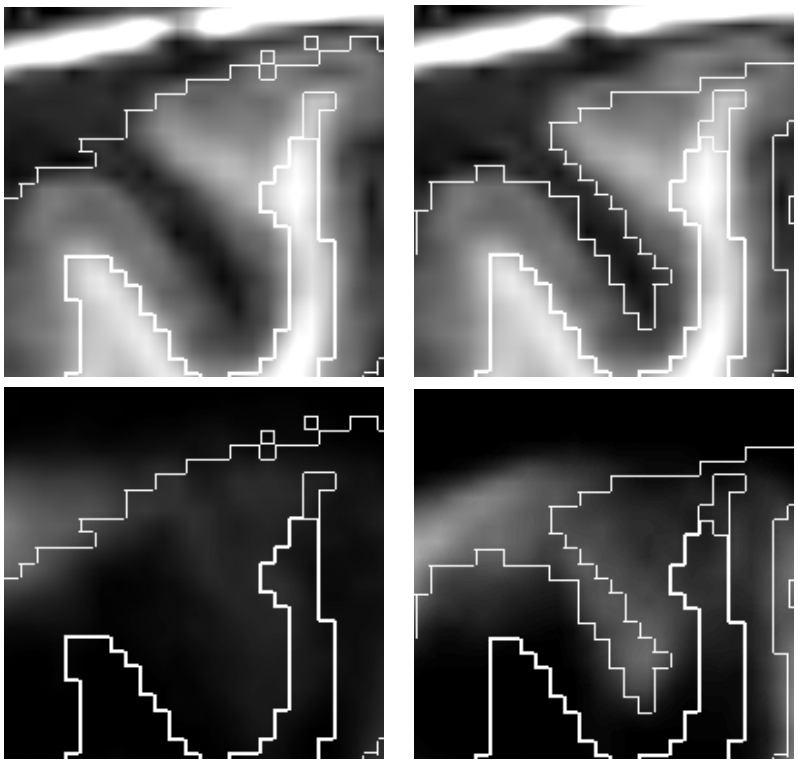


Figure 6.9: Comparison of segmentation methods on 3T data, sagittal view. (a) 3T MRI of two years old child; (b) registration-based segmentation; (c) EM segmentation with population-specific atlas; (d) Combined EM and registration based segmentation. Arrows from top to bottom: 1. CSF space imprecisely delineated by registration-based segmentation, undetected by EM segmentation, well delineated by combined segmentation. 2. Substantia nigra correctly delineated by registration-based segmentation and combined segmentation, incorrectly by EM segmentation. 3. Part of the brain not segmented by EM segmentation due to imprecise alignment of the probabilistic atlas, correctly delineated by combined segmentation, less precisely by registration-based segmentation.

Fig. 6.10 shows an example of the advantage of the subject-specific atlas over the population-specific atlas: While CSF is undetected by the EM segmentation due to an insufficient prior for CSF in this area, the combined segmentation can detect the CSF thanks to the prior individually adapted particularly for this brain. Fig. 6.11 shows the probability maps for WM, GM and CSF in the population-specific and subject-specific atlas.



(a) (b)

Figure 6.10: Comparison of EM segmentation with a population-specific atlas and combined segmentation with a subject-specific atlas: Detail of the sagittal slice shown in fig. 6.9 demonstrates how better prior information for CSF improves the segmentation: (a) EM segmentation with population-specific probabilistic atlas; (b) combined segmentation with subject-specific probabilistic atlas; (c) prior for CSF in population-specific atlas; (d) prior for CSF in subject-specific atlas. Segmentation of CSF space fails with the EM segmentation due to insufficient prior probability for CSF in the location, while the combined method classifies CSF in this area correctly thanks to the more precise subject-specific atlas.

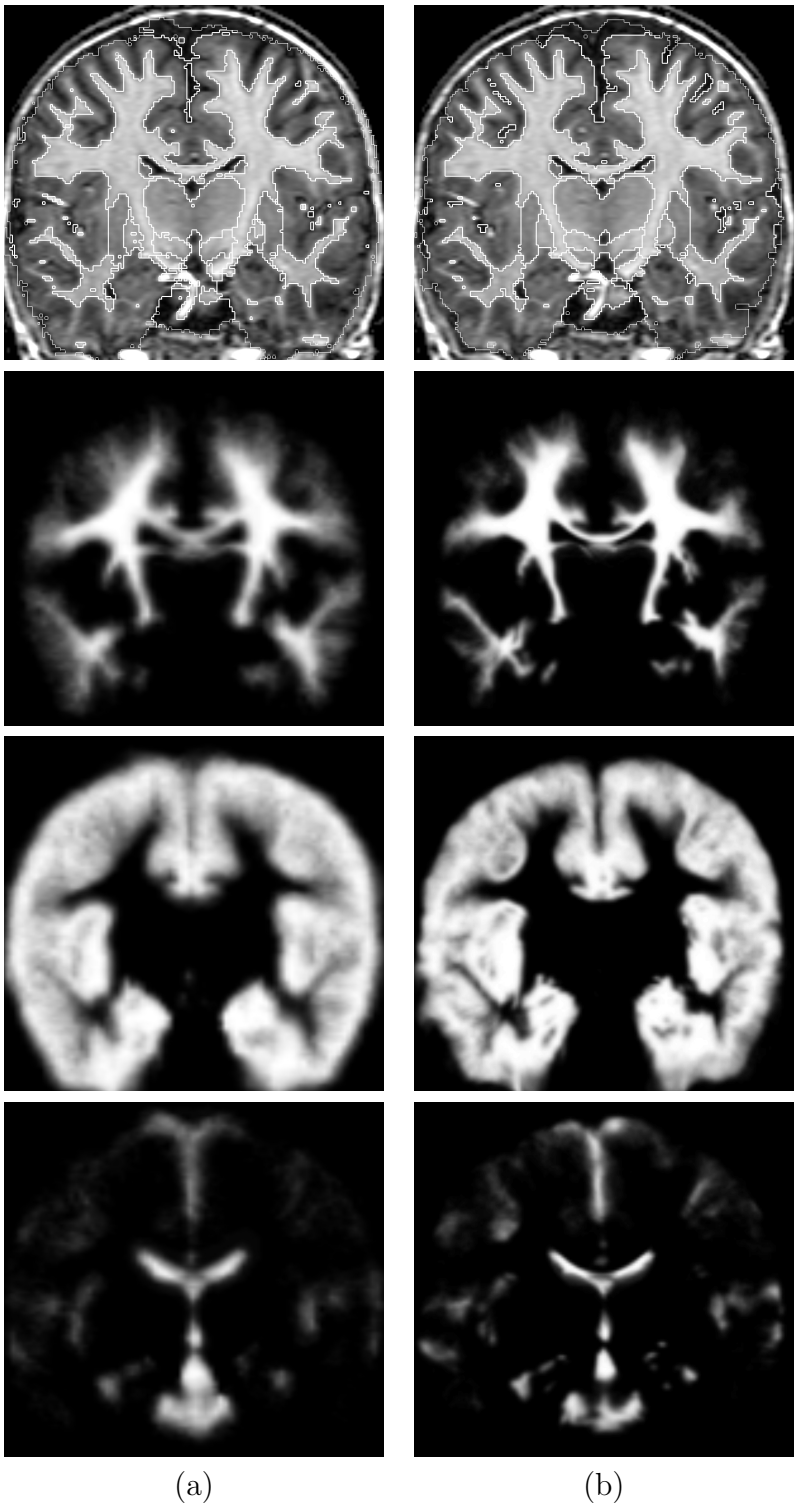


Figure 6.11: Comparison of EM segmentation with population-specific atlas and combined segmentation with subject-specific atlas, coronal view. Segmentations shown in the first row. Probability maps for WM (second row), GM (third row) and CSF (fourth row). (a) EM segmentation with population-specific probabilistic atlas; (b) Combined segmentation with subject-specific probabilistic atlas.

6.7 Measuring volumes of brain structures

In this section we will demonstrate that the proposed combined EM-based and registration-based segmentation method in conjunction with the template-based bias correction, as described in sec. 6.5 is a suitable tool for robust calculation of volumes of brain structures at two years of age for the MRI with strong bias field. We will compare the results the measurements obtained from 1T data in section 4.6.

The volumes of 9 brain structures (WM, GM, cerebellum, brainstem and five subcortical GM structures - caudate, putamen, nigra, thalamus and pallidum) were measured using the proposed segmentation method combining registration-based and EM-based approaches (sec. 6.5). CSF was excluded from the calculations and cerebellar WM and GM were merged together. The volumes were measured on 35 subjects at two years of age scanned on a 3T MR scanner and compared to measurements obtained from 15 two years old subjects scanned at 1T and calculated in section 4.6. None of the subjects scanned at 3T scanner was identical with a subject scanned at 1T scanner. Like the subjects used in chapter 4, all the subjects used in this chapter are preterm. The results are presented in table 6.12 and visualized in fig. 6.12 and 6.14. The mean volumes obtained using 1T and 3T images show good agreement. The measurements agree even better if adjusted for the size of the brain. The structure sizes expressed as percentage of the whole brain volume, here referred to as relative volumes, are presented in table 6.13 and visualized in fig. 6.13 and 6.15.

	Cortex	Caudate	Putamen	Nigra	Cerebellum
1T	675 ± 55	6.61 ± 1.18	9.83 ± 1.38	0.92 ± 0.10	103 ± 18
3T	715 ± 68	7.02 ± 0.87	10.85 ± 0.90	0.93 ± 0.11	125 ± 10
	Thalamus	Pallidum	Brainstem	WM	
1T	11.94 ± 1.28	2.84 ± 0.39	15.66 ± 2.30	235 ± 29	
3T	12.90 ± 1.16	3.29 ± 0.29	16.94 ± 1.78	234 ± 30	

Table 6.12: Mean volumes and standard deviation of brain structures calculated from 1T and 3T data. Results given in cm^3

	Cortex	Caudate	Putamen	Nigra	cerebellum
1T	63.75 ± 1.57%	0.62 ± 0.09%	0.92 ± 0.07%	0.09 ± 0.01%	9.62 ± 1.01%
3T	63.55 ± 1.17%	0.62 ± 0.05%	0.97 ± 0.07%	0.08 ± 0.01%	11.11 ± 0.69%
	Thalamus	Pallidum	Brainstem	WM	
1T	1.12 ± 0.04%	0.27 ± 0.02%	1.47 ± 0.11%	22.13 ± 0.84%	
3T	1.15 ± 0.08%	0.29 ± 0.02%	1.51 ± 0.12%	20.71 ± 1.12%	

Table 6.13: Mean relative volumes and standard deviation of brain structures calculated as a percentage of the whole brain volume from 1T and 3T data.

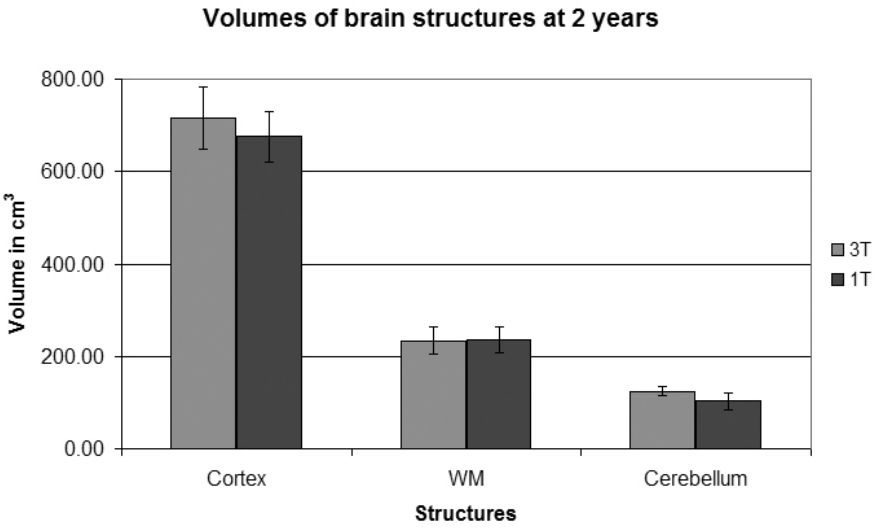


Figure 6.12: Comparison of mean volumes of brain structures obtained from 3T and 1T data.

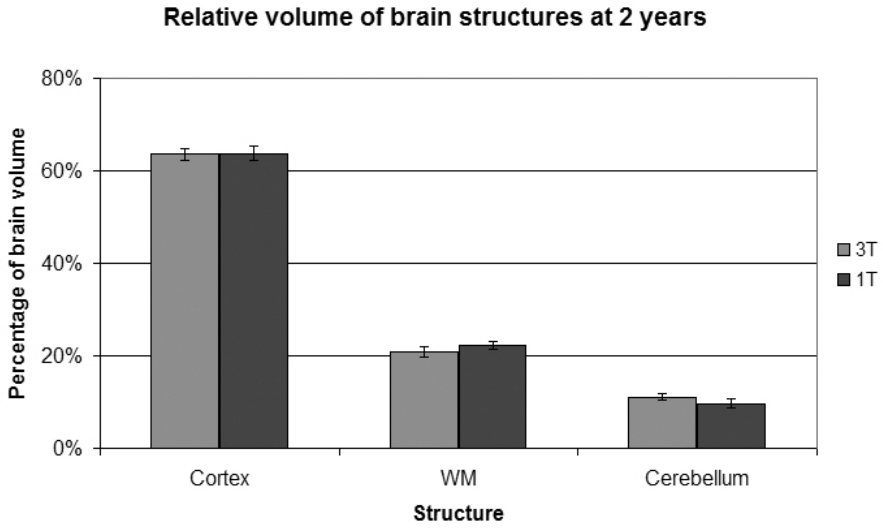


Figure 6.13: Comparison of mean relative volumes of brain structures obtained from 3T and 1T data.

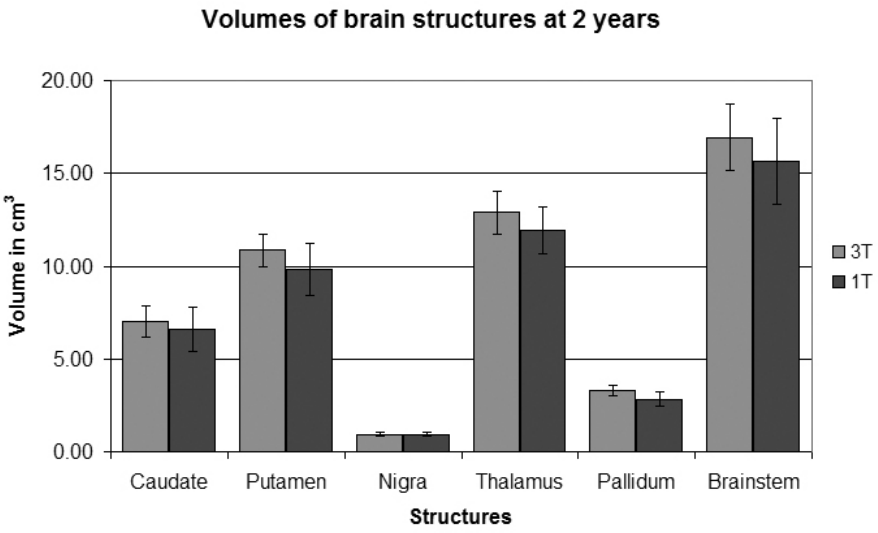


Figure 6.14: Comparison of mean volumes of brain structures obtained from 3T and 1T data.

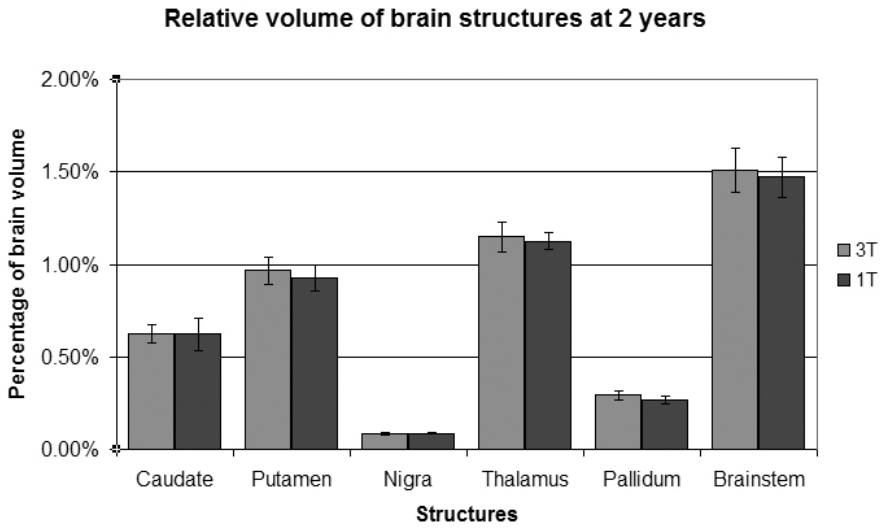


Figure 6.15: Comparison of mean relative volumes of brain structures obtained from 3T and 1T data.

6.8 Discussion

In this chapter we have developed strategies for the segmentation of brain MRI during early childhood in presence of strong intensity inhomogeneity using the combination of a novel template-based bias correction method proposed in chapter 5 and the segmentation tools proposed in chapter 4.

The parameters for the novel template-based bias correction method were experimentally determined to correct brain MRI obtained on a 3T MR scanner using a MP-RAGE imaging sequence [14]. These images exhibit a very strong bias field characterized by central brightening of the image [21]. We were able to obtain good automatic segmentation results for these challenging images when the parameters $L = 6$ and $\lambda = 2$ were used. This was experimentally tested on four manually segmented MRI. In addition our results have shown good agreement with volume measurements of brain structures obtained from 3T and 1T images. However, we were not able to validate the quality of bias correction directly, as there is no ground truth available for the bias field.

The proposed bias correction method was applied to images obtained using the same imaging sequence. Therefore, the question arises, whether the segmentation tools described in this chapter are generally applicable. If the 1T images used in chapter 4 are bias corrected using the bias correction parameters determined in this chapter, plausible results can be obtained as shown in fig. 6.16, which compares the bias correction with N3 and our proposed method. This suggests, that the method is stable in the presence of weak bias fields and there is no need to change the parameters. We hypothesize, that the artefacts of the imperfect alignment of template image with image to be segmented have a similar frequency given by size of the brain and its anatomy, independent of the bias field. Therefore the constraints on the model for the bias field do not need to be changed depending on the bias field strength and frequency. However, more investigation has to be performed before the general applicability of the method can be established.

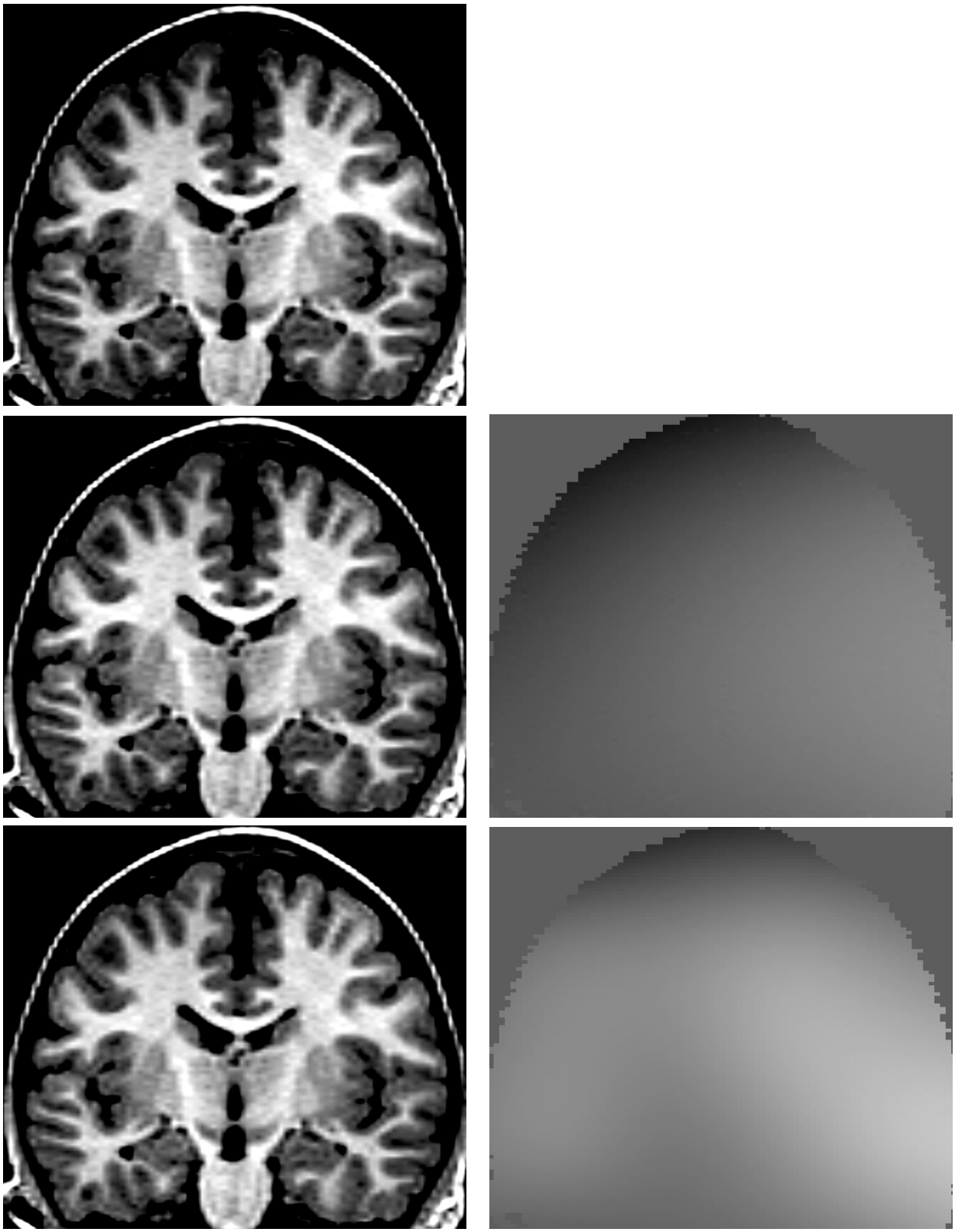


Figure 6.16: Bias correction of brain MRI of a two year old subject scanned at 1T magnetic field strength. First row: the original image. Second row: Image corrected by N3 and the estimated bias field. Third row: Image corrected by our proposed method and the estimated bias field.

6.9 Conclusion

The aim of this chapter was to develop tools for the robust segmentation of brain structures during early childhood in the presence of strong intensity inhomogeneity. The segmentation tools developed in chapter 4 were improved by including the novel robust template-based bias correction method proposed in chapter 5. The bias field was modelled using 3D tensor product B-spline. To model the low-frequency characteristic of the bias field, suitable constraints in the form of B-spline control point spacings and weighting coefficients for the smoothness constraint representing the bending energy of the field were experimentally determined.

The novel segmentation method proposed in chapter 4 and the novel template-based bias correction method proposed in chapter 5 with parameters determined in chapter 6 were combined to produce a segmentation tool robust to the variability in the shape of the brain as well as the frequency and strength of the bias field. The volume measurements of brain structures in 50 subject at two years of age (15 scanned at 1T and 35 at 3T magnetic field strength) show good consistency, rendering this tool suitable for quantifying changes in growth of WM and GM tissues, as well as subcortical GM structures.

Chapter 7

Conclusion

This thesis has presented the development of segmentation tools for brain structures during early childhood. Such tools are required to quantify the changes in brain development of prematurely born children compared to full-term children. In this chapter we will summarize the main contributions of this thesis. We will also discuss the limitations of the work presented here and directions for future research.

7.1 Contribution

The contributions of these thesis can be divided into four main categories:

1. **Development of population-specific probabilistic atlases of the brain during early childhood.** To achieve correct segmentation of the brain during early childhood, we created population-specific probabilistic atlases of the brain for ages one and two years. We demonstrated that due to anatomical differences of the brain during early childhood and during adulthood, age-specific atlases are necessary to obtain results usable for practical applications.

2. **Development of tools for segmentation of brain structures during early childhood.** To develop a robust method for segmentation of brain MRI we combined registration-

based and intensity-based approaches. The adaptations of available techniques have been investigated and a novel tool combining both approaches has been proposed. Important aspect of this methodology is ability to segment all the brain structures required by an application as opposed to only the three main brain tissue classes classified by many standard tools. The robustness of the method is achieved by multiple label propagation using non-rigid registration to create prior information for intensity-based segmentation.

3. **Development of tools for bias correction of brain MRI.** Currently, brain MRI are usually acquired using scanners with higher magnetic field strengths, often resulting in strong intensity inhomogeneity present in the image. To make the segmentation tool robust to this artefact, a template-based bias correction technique has been developed. The important advantage of this technique is ability to deal with strong bias field. In addition, the bias field is calculated by comparing with a template image without bias, thus the technique is not biased towards any simplified theoretical model of brain tissue intensities. The optimal strategies for coupling of this bias correction technique with segmentation tools were also proposed.

4. **Application of the developed tools to quantification of volume and growth of brain structures from one to two years of age.** The quantitative analysis of brain structure volumes and growth was performed using the novel segmentation tool combining registration-based and intensity based approaches. The results for MRI of one and two year old subjects with weak bias exhibited excellent agreement, thus confirming the robustness of the tool. The novel segmentation method has also been coupled with the novel bias correction technique to obtain a segmentation tool robust to strong bias field. The technique was applied to MRI of two years old subjects with strong bias field, producing volume measurements consistent with measurements obtained from MRI with weak bias field.

7.2 Discussion

The problem of the segmentation of brain MRI has been extensively studied by many researchers in recent years. In spite of all this effort there is still the need for improvement before the methods can be successfully adapted to different applications. An example of such application demonstrated in our work is processing brain MRI during early childhood, such as at one or two years of age.

While adapting the segmentation methods to this age-group, we developed novel methodology, which is not restricted for use on pediatric subjects. More evaluation needs to be performed to determine, whether this methodology would outperform standard tools developed for adult brains. However, we suspect that the methodology presented in this thesis would not be sufficient to deal with special challenges in processing fetal and neonatal brain MRI, as those images exhibit different intensity characteristics compared to children and adults as a result of on-going development.

The recent trend in development of brain segmentation tools has been guided by the idea that performing the segmentation, registration and bias correction simultaneously would improve the performance of the segmentation methods. The evaluation presented in this thesis suggest a surprising conclusion, that this is not necessarily so. During our experiments we observed only marginal, if any improvement of interleaved compared to sequential approaches. On the contrary to the common belief, our experiments seem to support the idea, that performing all the processes separately can actually make the final approach more robust against converging to suboptimal local extrema. An example of such event can be anticipated during simultaneous segmentation and non-rigid registration of the probabilistic atlas in the subcortical region. The brighter characteristics of the subcortical GM could be dealt with by shrinking the prior for GM and extending prior for WM by non-rigid registration to fit the voxel intensities into Gaussian intensity distributions for WM and GM. This would result in misclassification of subcortical GM as WM. Separating this two processes can prevent convergence to such suboptimal solution.

7.3 Limitations and future work

There are many limitations of the current work that give opportunity for future research to improve performance and robustness of the developed tools. The areas for potential improvement can be categorized as follows:

1. **Segmentation:**

   (a) **Creating unbiased probabilistic atlases:** Probabilistic atlases not biased towards the reference subject could be developed by calculating the average brain [11]. The template image representing average brain anatomy rather than average intensity image could be registered with the image to be segmented to provide robust non-rigid alignment of the resulting probabilistic atlas. This might remove the need for multiple label propagation and thus improve the efficiency of the proposed segmentation method.

   (b) **Partial volume estimation:** The current segmentation tools usually fail to classify the small detail due to complex cortical folding. The methods for partial volume estimation, such as [75], [47], are therefore needed. The challenge in addressing this problem lies in the fact, that small volumes of tissue are often coupled with insufficient prior for their detection.

2. **Bias correction:**

   (a) **The B-spline model:** The major drawback of using B-spline model for the bias field is computational speed when estimating the bias field parameters. Subsampling is therefore needed to obtain bias correction algorithm that finishes within a few minutes. The estimation of B-spline control points could be made more efficient by developing suitable filtering techniques which can incorporate weights, such as in [96], [90].

   (b) **The EM framework:** There are other frameworks commonly used in medical imaging, such as information-theoretic approach [95], [56]. The evaluation of differ-

ent approaches is needed to determine which framework is the most suitable for the task of estimating the bias field.

(c) **The outlier detection:** We used the weight function for estimating bias by weighted B-splines resulting from the use of EM framework. However, it is possible, that more suitable weight function could be estimated for the purpose of bias correction.

3. **The evaluation:**

   (a) **Validation of segmentation techniques:** The major limitation of work presented in these thesis is insufficient number of manually segmented subjects to obtain quantitative measurements of overlap between manual and automatic segmentation. More thorough validation of the methodology presented would be therefore desired.

   (b) **Protocol for manual segmentation:** Another important limitation is the quality of manual segmentations. It is a generally recognized problem that manual delineation is an error-prone process. Many boundaries of brain structures are not visible on MRI and protocol for manual segmentation is therefore needed to guide the human rater. In this thesis we did not perform any validation of the quality of manual segmentations. In addition, no formal protocol for the manual delineation of the brain structures was defined.

   (c) **Evaluation of bias correction technique:** We were not able to validate the proposed bias correction technique because of unavailability of the ground truth. The exact evaluation of bias correction techniques can be performed only if measurements of the bias field are obtained during the acquisition process [83], [61]. The comparative study using images with measured bias field would be desirable to perform thorough validation of the proposed bias correction technique.

4. **The applications:** The purpose of the application of the developed tools for calculating of the volumes and growth of brain structures of one and two year old children presented in this thesis is to demonstrate the potential usability of the method. However, the proposed methodology can now be used for medical experiments to quantify the volumes and growth of brain structures during early childhood on large number of subject and

between different populations of subject, to obtain statistically significant differences in the brain development of the populations.

7.4 Summary

In this thesis we developed a tool for segmentation of brain structures from MRI during early childhood. The proposed methodology is robust to variation in size and shape of brain structures as well as the type and strengths of the bias field present in the MRI. The main motivation for the work presented in this thesis have been the need for quantification of changes in development of the brain in premature children compared to full term children. The quantitative measurements of volumes and growth of brain structures exhibit consistency suggesting suitability of the tool for the research of the impact of premature birth on brain development.

Publications

1. P. Aljabar, K.K. Bhatia, M. Murgasova, J.V. Hajnal, J.P. Boardman, L. Srinivasan, M.A. Rutherford, L.E. Dyet, A.D. Edwards and D. Rueckert. Assessment of brain growth in early childhood using deformation based morphometry. *NeuroImage*, Elsevier, 39(1), pp. 348–358, January 2008.

2. M. Murgasova, L.E. Dyet, A.D. Edwards, M.A. Rutherford, J.V. Hajnal, D. Rueckert. Segmentation of brain MRI in young children. *Academic Radiology*, 14(11), pp. 1350–66, November 2007.

3. 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.

4. M. Murgasova, L. Dyet, A.D. Edwards, M.A. Rutherford, J.V. Hajnal, D. Rueckert. Robust Segmentation of Brain MRI Using Combination of Registration and EM-based Methods. *Proceedings of the Spring Conference on Computer Graphics - SCCG* 2007, Budmerice, Slovakia, pp. 115–122.

5. M. Murgasova, L. Dyet, A.D. Edwards, M.A. Rutherford, J.V. Hajnal, D. Rueckert. Segmentation of Brain MRI in Young Children. *Medical Image Computing and Computer-Assisted Intervention - MICCAI* 2006, Copenhagen, Lecture Notes in Computer Science, 4190, Springer, pp. 687–694, 2006.

Bibliography

[1] P. Aljabar, K.K. Bhatia, M. Murgasova, J.V. Hajnal, J.P. Boardman, L. Srinivasan, M.A. Rutherford, L.E. Dyet, A.D. Edwards, and D. Rueckert. Assessment of brain growth in early childhood using deformation-based morphometry. *NeuroImage*, 39(1):348–358, 1 January 2008.

[2] M. Allin, H. Matsumoto, A.M. Santhouse, C. Nosarti, M.H. Al Asady, and A.L. Stewart. Cognitive and motor function and the size of the cerebellum in adolescents born very pre-term. *Brain*, 124:60–66, 2004.

[3] J.B. Arnold, J.S. Liow, K.A. Schaper, J.J. Stern, J.G. Sled, D.W. Shattuck, A.J. Worth, M.S. Cohen, R.M. Leahy, J.C. Mazziotta, and D.A. Rottenberg. Qualitative and quantitative evaluation of six algorithms for correcting intensity nonuniformity effects. *NeuroImage*, 13(5):931–943, 2001.

[4] J. Ashburner. *Computational Neuroanatomy*. PhD thesis, University College London, 2000.

[5] J. Ashburner and K.J. Friston. MRI sensitivity correction and tissue classification. *NeuroImage*, 7(4):S107, 1998.

[6] J. Ashburner and K.J. Friston. Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, 7(4):254–266, 1999.

[7] J. Ashburner and K.J. Friston. Unified segmentation. *NeuroImage*, 26:839–851, 2005.

[8] B. Belaroussi, J. Milles, S. Carme, Y.M. Zhu, and H. Benoit-Cattin. Intensity non-uniformity correction in MRI: Existing methods and their validation. *Medical Image Analysis*, 10(2):234–246, April 2006.

[9] J. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of Royal Statistical Society, series B*, 36(2):192–326, 1974.

[10] J. Besag. On the statistical analysis of dirty pictures. *Journal of Royal Statistical Society*, 48(3):259–302, 1986.

[11] 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. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2007, Brisbane*, pages 532–540. Lecture Notes in Computer Science, 4791, Springer, 2007.

[12] K.K. Bhatia, J.V. Hajnal, B.K. Puri, A.D. Edwards, and D. Rueckert. Consistent groupwise non-rigid registration for atlas construction. In *IEEE Symposium on Biomedical Imaging (ISBI)*, pages 908–911, 2004.

[13] J. P. Boardman, S.J. Counsell, D. Rueckert, O. Capellou, K. 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*, 2006.

[14] M. Brant-Zawadzki, G.D. Gillan, and W.R. Nitz. MP RAGE: a three-dimensional, T1-weighted, gradient-echo sequence–initial experience in the brain. *Radiology*, 182(3):769–775, 1992.

[15] B.H. Brinkmann, A. Manduca, and R.A. Robb. Optimized homomorphic unsharp masking for MR grayscale inhomogeneity correction. *IEEE Transactions on Medical Imaging*, 17(2):161–171, April 1998.

[16] S.S. Brunssen and G.J. Harry. Diffuse white matter injury and neurologic outcomes of infants born very preterm in the 1990s. *Obstertic, Gynecologic, and Neonatal Nursing*, 36(4):386–395, 2007.

[17] G.E. Christensen, R.D. Rabbit, and M.I. Miller. A deformable neuroanatomy textbook based on viscous fluid mechanics. In *Conference on Information Sciences and Systems*, pages 211–216, 1993.

[18] 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.

[19] C.A. Cocosco, A.P. Zijdenbos, and A.C. Evans. A fully automatic and robust brain MRI tissue classification method. *Medical Image Analysis*, 7(4):513–527, 2003.

[20] M.S. Cohen, R.M. DuBois, and M.M Zeneih. Rapid and effective correction of RF inhomogeneity for high field magnetic resonance imaging. *Human Brain Mapping*, 10(4):204 – 211, July 2000.

[21] C.M. Collins, W. Liu, W. Schreiber, Q.X. Yang, and M.B. Smith. Central brightening due to constructive interference with, without, and despite dielectric resonance. *Journal of Magnetic Resonance Imaging*, 21(2):192–6, 2005.

[22] T.F. Cootes, G.J. Edward, and C.J. Taylor. Active appearance models. *Proceedengs of 5th European Conference on Computer Vision - Volume II*, 1407:484–498, 1998.

[23] T.F. Cootes, A. Hill, C.J. Taylor, and J. Haslam. The use of active shape models for locating structures in medical images. *Image and Vision Computing*, 12(6):355–366, 1994.

[24] W.R. Crum, T. Hartkens, and D.L.G. Hill. Non-rigid image registration: theory and practice. *Br J Radiol*, 77:S140–S153, 2004.

[25] W.R. Crum, D. Rueckert, M. Jenkinson, D. Kennedy, and S.M. Smith. A framework for detailed objective comparison of non-rigid registration algorithms in neuroimaging. In *MICCAI (1)*, pages 679–686, 2004.

[26] 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.

[27] E. DAgostino, F. Maes, D. Vandermeulen, and P. Suetens. A viscous fluid model for multimodal non-rigid image registration using mutual information. In *In Fifth Int. Conf. on Medical Image Computing and Computer- Assisted Intervention (MICCAI 02)*, page 541548, 2002.

[28] E. D'Agostino, F. Maes, D. Vandermeulen, and P. Suetens. Non-rigid atlas-to-image registration by minimization of class-conditional image entropy. In *Proc. of the 7th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention, Part I*, pages 745–753, 2004.

[29] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM-algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38, 1977.

[30] L.R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26:297–302, 1945.

[31] N.R. Draper and H. Smith. *Applied Regression Analysis.* John Wiley & Sons, 1998.

[32] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification.* John Wiley & Sons Inc, second edition, 2001.

[33] J. C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact, well separated clusters. *Journal of Cybernetics*, 3:32–57, 1974.

[34] A.C. Evans, D.L. Collins, S.R. Mills, E.D. Brown, R.L. Kelly, and T.M. Peters. 3D statistical neuroanatomical models from 305 MRI volumes. *Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record.*, pages 1813–1817 vol.3, 31 Oct-6 Nov 1993.

[35] G. Farin. *Curves and Surfaces for CAGD, a practical guide.* Academic Press, 2002.

[36] B. Fischl, D. Salat, E. Busa, M. Albert, M. Dietrich, C. Haselgrove, A. van der Kouwe, R. Kiliany, D. Kenedy, S. Klaveness, A. Mondtilo, N. Makris, B. Rosen, and A. Dale. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neurotechnique*, 2002(January 31):341–355, 1998.

[37] I.S. Gousias, D. Rueckert, R.A. Heckemann, L.E. Dyet, J.P. Boardman, A. D. Edwards, and A. Hammers. Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest. *NeuroImage*, 40(2):672–84, 2008.

[38] J.V. Hajnal, D.L.G. Hill, and D.J. Hawkes. *Medical Image Registration*. CRC Press, 2001.

[39] X. Han, C. Xu, and J. L. Prince. A topology preserving level set method for geometric deformable models. *IEEE Trans. Pattern Analysis Machine Intelligence*, 25(6):755–768, 2003.

[40] R.A. Heckermann, J.V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage*, 2006.

[41] B.A. Holland, D.K. Haas, D. Norman, M. Brant-Zawadzki, and T.H. Newton. MRI of normal brain maturation. *American Journal of Neuroradiology*, 7:201–208, 1986.

[42] Z. Hou. A Review on MR Image Intensity Inhomogeneity Correction. *International Journal of Biomedical Imaging*, 2006:article ID 49515, 2006.

[43] T.E. Inder, S.K. Warfield, H. Wang, P.S. Huppi, and J.J. Volpe. Abnormal cerebral structure is present at term in premature infants. *N Engl J Med*, 352(1):9–19, 2005.

[44] T.E. Inder, S.J. Wells, N.B. Mogridge, C. Spencer, and J.J. Volpe. Defining the nature of the cerebral abnormalities in the premature infant: a qualitative magnetic resonance imaging study. *The Journal of Pediatrics*, 143(2):171–179, 2003.

[45] M. Jenkinson, P.R. Bannister, J.M. Brady, and S.M. Smith. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2):825–841, 2002.

[46] M. Jenkinson and S. M. Smith. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2):143–156, 2001.

[47] N. Joshi and M. Brady. A non-parametric mixture model for partial volume segmentation of MR images. In *British Machine Vision Conference*, 2005.

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

[49] 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.

[50] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. A unifying framework for partial volume segmentation of brain MR images. *IEEE Transactions on Medical Imaging*, 22(1):105–119, 2003.

[51] E.B. Lewis and N.C. Fox. Correction of differential intensity inhomogeneity in longitudinal MR images. *Neuroimage*, 23(1):75–83, 2004.

[52] C. Limperopoulos, J.S. Soul, K. Gauvreau, P.S. Huppi, S.K. Warfield, H. Bassan, R.L. Robertson, J.J. Volpe, and A.J. du Plessis. Late gestation cerebellar growth is rapid and impeded by premature birth. *Pediatrics*, 115(3):688–695, 2005.

[53] E.F. Maalouf, P.J. Duggan, M.A. Rutherford, S.J. Counsell, A.M. Fletcher, M. Battin, F. Cowan, and A.D. Edwards. Magnetic resonance imaging of the brain in a cohort of extremely preterm infants. *Pediatrics*, 135(3):351–7, 1999.

[54] J.B. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability 1*, pages 281–297. University of California Press, 1967.

[55] 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):187–98, 1997.

[56] J. Mangin. Entropy minimization for automatic correction of intensity nonuniformity. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA'00)*, pages 162–169, 2000.

[57] N. Marlow, D. Wolke, M.A. Bracewell, M. Samara, and the EPICure Study Group. Neurologic and developmental disability at six years of age after extremely preterm birth. *Pediatrics*, 115(2):286–94, 2005.

[58] J. Mazziota, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T. Paus, G. Simpson, B. Pike, C. Holmes, L. Collins, P. Thompson, D. MacDonald, M. Iacoboni, T. Schormann, K. Amunts, N. Palomero-Gallager, S. Geyer, L. Parsons, K. Narr, N. Kabani, G. Le Goualher, D. Boomsma, T. Cannon, R. Kawashima, and B. Mazoyer. A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (ICBM). *Philos Trans R Soc Lond B Biol Sci*, 356(1412):1293–1322, August 2001.

[59] W. Mendenhall, R.J. Beaver, B.M. Beaver, and T. Brooks. *Introduction to Probability and Statistics*. Academic Internet Publishers Incorporated, 2005. page 159.

[60] T. Minka. Expectation-maximization as lower bound maximization. Technical Report, MIT, 1998. www.microsoft.com/en-us/um/people/minka/papers/em.html.

[61] P.A. Narayana, W.W. Brey, M.V. Kulkarni, and C.L. Sievenpiper. Compensation for surface coil sensitivity variation in magnetic resonance imaging. *Magn. Reson. Imag.*, 6(3):271274, 1988.

[62] C. Nosarti, M.H. Al Asady, S. Frangou, A.L. Stewart, L. Rifkin, and R.M. Murray. Adolescents who were born very preterm have decreased brain volumes. *Brain*, 125(Pt 7):1616–23, 2002.

[63] B. Patenaude. *Bayesian Statistical Models of Shape and Appearance for Subcortical Brain Segmentation.* PhD thesis, University of Oxford, 2007.

[64] CCHMC Pediatric Brain Templates. http://www.irc.cchmc.org/software/pedbrain.php.

[65] D. Pham and J. Prince. An adaptive fuzzy c-means algorithm for image segmentation in the presence of intensity inhomogeneities. In *'Proc. SPIE Medical Imaging 1998: Image Processing'*, volume 3338, pages 555–563, 1998.

[66] K.M. Pohl. *Prior Information for Brain Parcellation.* PhD thesis, Massachusetts Institute of Technology, 2005.

[67] 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.

[68] K.M. Pohl, W.M. Wells, A. Guimond, K. Kasai, M.E. Shenton, R. Kikinis, W.E.L. Grimson, and S.K. Warfield. Incorporating non-rigid registration into expectation maximization algorithm to segment MR images. In *Proc. of the 5th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, volume 2488 of *Lecture Notes in Computer Science*, pages 564–572, 2002.

[69] M. Prastawa, J.H. Gilmore, W. Lin, and G. Gerig. Automatic Segmentation of Neonatal Brain MRI. In *Proc. of the 7th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention, Part I*, pages 10–17, 2004.

[70] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press, third edition edition, 2007.

[71] J. Ren, B. Sneller, D. Rueckert, J. Hajnal, and D. Hill. A comparison of tissue type labelings of direct classification and segmentation propagation techniques for MR brain images. In *Proc. of Medical Image Understanding and Analysis '05*, pages 55–58, 2005.

[72] D. Rueckert and J. Schnabel. Image Registration Toolkit (IRTK), 1999-2008. www.doc.ic.ac.uk/˜dr/software.

[73] 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.

[74] M. Rutherford. *MRI of the Neonatal Brain*. Saunders Ltd., 2001.

[75] D.W. Shattuck, S.R. Sandor-Leahy, K.A. Schaper, D.A. Rottenberg, and R.M. Leahy. Magnetic resonance image tissue classification using a partial volume model. *Neuroimage*, 13(5):856–876, 2001.

[76] D. Shen and C. Davatzikos. Hammer: Hierarchical attribute matching mechanism for elastic registration. *IEEE Transactions on Medical Imaging*, 21(11):14211439, 2002.

[77] A. Simmons, P.S. Tofts, G.J. Barker, and S.R. Arridge. Sources of intensity nonuniformity in spin echo images at 1.5 T. *Magnetic Resonance in Medicine*, 32(1):121–8, 1994.

[78] J.G. Sled, A.P. Zijdenbos, and A.C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, 17(1):87–97, 1998.

[79] S.M. Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143–155, 2002.

[80] Statistical Parametric Mapping. www.fil.ion.ucl.ac.uk/spm.

[81] Statistical Parametric Mapping: SPM5. www.fil.ion.ucl.ac.uk/spm/software/spm5.

[82] A.L. Stewart, L. Rifkin, P.N. Amess, V. Kirkbride, J.P. Townsend, and D.H. Miller. Brain structure and neurocognitive and behavioural function in adolescents who were born very preterm. *Lancet*, 353(9165):1653–1657, 1999.

[83] R. Stollberger and P. Wach. Imaging of the active B1 field in vivo. *Magn. Reson. Med.*, 35:246251, 1996.

[84] C. Studholme, V. Cardenas, E. Song, F. Ezekiel, A. Maudsley, and W. Weiner. Accurate template-based correction of brain MRI intensity distortion with application to dementia and aging. *IEEE Transactions on Medical Imaging*, 23(1):99–110, 2004.

[85] C. Studholme, D.L.G. Hills, and D.J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86, 1999.

[86] M. Styner, C. Brechbhler, G. Szkely, and G. Gerig. Parametric estimate of intensity inhomogeneities applied to MRI. *IEEE Transactions on Medical Imaging*, 19:153–165, 2000.

[87] J. Talairach and P. Tournoux. *"Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System - an Approach to Cerebral Imaging"*. Thieme Medical Publishers, New York, 1988.

[88] T. Tasdizen, S.P. Awate, R.T. Whitaker, and N.L. Foster. MRI tissue classification with neighborhood statistics: A nonparametric, entropy-minimizing approach. In *Proc. of the 8th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention, Part II*, Lecture Notes in Computer Science, pages 517–525, 2005.

[89] J. Tucker and W. McGuire. Epidemiology of premature birth. *British Medical Journal*, 329:675–678, 2004.

[90] M. Unser, A. Aldroubi, and M. Eden. B-Spline signal processing: Part I—Theory. *IEEE Transactions on Signal Processing*, 41(2):821–833, February 1993. IEEE Signal Processing Society's 1995 best paper award.

[91] L.S. De Vries, I.L. Van Haastert, K.J. Rademaker, C. Koopman, and F. Groenendaal. Ultrasound abnormalities preceding cerebral palsy in high-risk preterm infants. *Pediatrics*, 144(6):815–20, 2004.

[92] G. Wahba. *Spline Models for observational data.* Society for Industrial Applied Mathematics, 1990.

[93] S.K. Warfield, M.Kaus, F.A. Jolesz, and R. Kikinis. Adaptive template moderated spatially varying statistical classification. In *Proc. of the 1st Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, pages 431–438, 1998.

[94] N. Weisenfeld, A. Mewes, and S.K. Warfield. Segmentation of newborn brain MRI. In *Proceedings of IEEE International Symposium on Biomedical Imaging*, pages 766–769, 2006.

[95] W. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51, 1996.

[96] W.M. Wells III, W.E.L. Grimson, R. Kikinis, and F.A. Jolesz. Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging*, 15(4):429–442, 1996.

[97] M. Wilke, V.J. Schmithorst, and S.K. Holland. Normative pediatric brain data for spatial normalization and segmentation differs from standard adult data. *Magnetic Resonance in Medicine*, 50(4):749–757, 2003.

[98] C.F.J. Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1):95–103, 1983.

[99] Ch. Xiaohua. *Simultaneous Segmentation and Registration of Medical Images*. PhD thesis, University of Oxford, 2005.

[100] H. Xue, L. Srinivasan, S. Jiang, M. Rutherford, A.D. Edwards, D. Rueckert, and J. Hajnal. Automatic Cortical Segmentation in the Developing Brain. In *Information Processing in Medical Imaging (IPMI 2007)*, 2007.

[101] P.A. Yushkevich, J. Piven, C. H. Hazlett, G.R. Smith, Sean. Ho, J.C. Gee, and G. Gerig. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*, 31(3):1116–1128, 2006.

[102] 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(1):45–57, 2001.

[103] B. Zitová and J. Flusser. Image registration methods: a survey. *Image Vision Comput.*, 21(11):9771000, 2003.

[104] L. Zöllei. *A Unified Information Theoretic Framework for pair- and Group-wise Registration of Medical Images.* PhD thesis, MIT, 2006.