Tissue Segmentation of the Cerebellum from MR Images

by

Keri Woods

supervised by

Patrick Marais and Ernesta Meintjes

A thesis submitted in fulfillment
of the requirements for the degree of

Master of Science

and is all my own work

The Department of Computer Science
University of Cape Town

November 21, 2011

Acknowledgements

I would like to acknowledge the following people who have helped me considerably during my thesis.

I would like to thank Patrick Marais, who I was fortunate enough to have as my supervisor. He has consistently been available to offer advice and guidance throughout this long journey. He showed considerable patience and was generous with the time he offered me.

To Ernesta Meintjes, who co-supervised my work, and who initiated this study, I would like to thank for the time and expertise she shared with me. I would also like to thank her for arranging funding, for her enthusiasm for the subject and for providing MR images for this thesis.

I would like to thank the NRF for the generous funding I received from them.

To my brother, Drew, many thanks for the technical guidance and encouragement he offered me.

I would also like to thank J.P. Longmore for his help, and the many discussions we had about the study. His advice was invaluable.

To Mom and Dad, who are always supportive and encouraging, many thanks for reading through my dissertation, despite it not being in their line of expertise.

Abstract

The cerebellum is known to be very susceptible to prenatal exposure to alcohol and people affected by Fetal Alcohol Spectrum Disorder (FASD) show a decrease in cerebellar volume. Unfortunately, there seems to be little work that focuses specifically on cerebellum tissue segmentation. We *set out to produce a method, as automated as possible, that segments the cerebellum into grey and white matter, capturing as much of the fine detail as possible.*

The key features of our segmentation approach include modeling the tissue class intensities as a Gaussian mixture model and using the iterative Expectation-Maximisation (**EM**) algorithm with a Markov Random Field (MRF) prior to estimate the model parameters in order to perform an initial segmentation. Using this segmentation information, we remove the voxels likely to be partial volume voxels, as outliers. Ignoring the outliers, we then calculate model parameters based on the initial segmentation. Using these parameter estimates, we then classify the voxels into different tissue classes.

We also investigated the idea of *bootstrapping* a tissue segmentation algorithm by applying the algorithm to the whole brain to estimate the parameters, and then using these parameters to classify the voxels of the cerebellum. We found that bootstrapping worked well with the "**EM** algorithm" and the "**EM** algorithm with MRF", but offered no significant improvement to our outlier algorithm.

We then compared the performance of our outlier algorithm with and without bootstrapping to the "**EM** algorithm" and the "**EM** algorithm with MRF" on simulated Brainweb data and real Internet Brain Segmentation Repository data. On the simulated data, our algorithms only performed best in segmentation at low noise levels, but both our algorithms outperformed the others in parameter estimation, with the version without bootstrapping performing the best. On real data, our method without bootstrapping performed significantly better ($p < 0.05$) than the "**EM** algorithm" or the "**EM** algorithm with MRF". Our method with bootstrapping estimated the parameters the best. In conclusion, our algorithm manages to pick up more of the fine structures of the cerebellum than the conventional "**EM** algorithm with MRF" and should enable more accurate studies of cerebellar tissue distribution.

Contents

Chapter 1

Introduction

The cerebellum, or small brain, is in the inferior posterior portion of the brain and is divided into two hemispheres. The outside of the cerebellum is highly convoluted and it contains both white and grey matter, with the white matter forming branchlike structures inside the grey matter. Recent developments in neurobiology suggest that the cerebellum plays a more fundamental role than previously thought in a human being's ability to learn, to plan and in temporal sequencing [96]. These developments have resulted in a renewed interest in studying the cerebellum, particularly through the modality of Magnetic Resonance Imaging (MRI).

Studies show that the cerebellum is very susceptible to prenatal exposure to alcohol and that people affected by Fetal Alcohol Spectrum Disorder (FASD) show a decrease in cerebellar volume [91]. Unfortunately, there is little work that focuses specifically on cerebellum tissue segmentation. Some techniques used to segment the cerebrum have been applied to the cerebellum, although with limited success, since they generally fail to capture fine-scale cerebellar detail [73].

It is for these reasons that we *set out to propose an automated method to segment the cerebellum into grey and white matter, capturing as much of the fine detail of the cerebellum as possible.* Specifically, this dissertation develops a framework for the tissue segmentation of the cerebellum from MR images. It is our hope that the segmentation techniques developed in this work will be suitable for clinical use in South African studies of the effect of FASD on the cerebellum.

1.1 Segmentation of the Cerebellum

Image segmentation is the process of dividing an image into homogeneous regions. The aim of segmentation is to simplify the image into something that is more meaningful and easier to analyse. In the case of brain tissue segmentation, segmentation involves assigning each voxel (volumetric pixel) to either grey matter (GM), white matter (WM) or cerebrospinal fluid (CSF). The segmentation of medical images into key anatomical structures is useful for diagnosis, treatment planning and therapy evaluation, for 3D visualisation of structures for surgical planning and to compare normal and abnormal structures. The segmentation of brain tissue into white matter, grey matter and cerebrospinal fluid has received much attention in the literature [140]. Measurements of the amount and spatial distribution of these tissue types can be used to support diagnosis for degenerative brain illnesses, such as Alzheimer's, and for assessing the progress or remission of these diseases. It can also be used to characterise the severity of the diseases. While tissue segmentation of the *cerebrum* is a well investigated area, segmentation of the *cerebellum* into tissue classes has not been widely addressed in the literature [46]. Because of the lack of work done in this field, it is an exciting area to explore.

Medical images are often manually segmented, but this is a very time-consuming task and impractical for large amounts of data. Manual segmentation is also highly subjective and therefore not reproducible. It is because of these issues that there is a need for automatic segmentation. However, the development of an automatic segmentation algorithm is not an easy task. Although MRI is the best method of imaging the brain, its finite resolution means that there are often many *partial volume* voxels, which are voxels containing a mixture of tissue types. In particular, a structure like the cerebellum with fine structures will have a large proportion of partial volume voxels. Other difficulties in the segmentation of MRI data include noise, motion artifacts and inhomogeneities caused by the field bias. These difficulties are especially problematic if trying to segment tissue types based only on individual voxel intensities.

1.2 A Framework for Cerebellum Segmentation

In order to address the difficulties mentioned above, this thesis presents a cerebellum segmentation algorithm. The key features of our approach include modeling the tissue class intensities as a Gaussian mixture model and using the iterative Expectation-Maximisation (**EM**) algorithm with a Markov Random Field (MRF) prior to estimate the model param-

eters in order to perform an initial segmentation. Using this segmentation information, we remove the voxels likely to be partial volume voxels as outliers. Due to its many fine structures, the cerebellum has a large proportion of *partial volume voxels*, making this a particularly important problem to address. Ignoring the outliers, we calculate model parameters based on the initial segmentation. With the partial volume voxels removed, the parameter estimation is generally more accurate than it would be with the partial volume voxels included. Using these parameter estimates, we classify the voxels into different tissue classes.

As mentioned above, many of the cerebellum's voxels are *partial volume* voxels, making it more difficult to segment than the whole brain. In order to try and overcome this problem, we investigate the idea of bootstrapping [1] a tissue segmentation algorithm by applying the algorithm to the whole brain to estimate the parameters, and then using these parameters to classify the voxels of the cerebellum. Bootstrapping can be combined with our outlier algorithm, or most other tissue segmentation algorithms that rely on parameter estimation, such as the "**EM** algorithm" or the "**EM** algorithm with MRF".

1.3 Results

We found that bootstrapping worked well with the "**EM** algorithm" and the "**EM** algorithm with MRF", but offered no significant improvement to our outlier algorithm.

We compared the performance of our outlier algorithm with and without bootstrapping to the "**EM** algorithm" and the "**EM** algorithm with MRF" on simulated Brainweb data and real Internet Brain Segmentation Repository data. On the simulated data, our algorithms only performed best in segmentation at low noise levels, but both our algorithms outperformed the others in parameter estimation, with the version without bootstrapping performing the best.

On real data, our method without bootstrapping performed statistically significantly better than the "**EM** algorithm" or the "**EM** algorithm with MRF". Our method with bootstrapping estimated the parameters the best, followed by our method without bootstrapping. Our algorithm manages to pick up more of the fine structures of the cerebellum than the "**EM** algorithm with MRF".

[1] For our purposes, the term bootstrap refers to a process whereby parameters are obtained from the whole brain which ensures that subsequent cerebellum segmentation is more robust.

1.4 Thesis Outline

The outline of the thesis is as follows:

Chapter 2 provides general background information about various aspects relevant to the rest of the thesis. We briefly introduce the cerebellum and Fetal Alcohol Spectrum Disorder (FASD) and describe the basics behind the Magnetic Resonance Imaging process. Next we discuss general image segmentation methods. Finally, we introduce brain tissue segmentation and briefly discuss the difficulties associated with it.

Chapter 3 reviews the literature regarding the methods used for brain tissue segmentation. Emphasis is placed on the *Gaussian mixture model*, as much of the literature deals with methods involving that model.

Chapter 4 explains the framework that we have developed and motivates the reasons for making various choices about the system. It explains how we *bootstrap* our system with information from the whole brain in order to provide a different segmentation method with the intention that this method will provide better segmentation results.

Chapter 5 looks at the performance of three different methods: the simple "**EM** algorithm", the "**EM** algorithm with an MRF prior" and our outlier method. In an attempt to improve the segmentation of the cerebellum, we run the algorithms on the whole brain first and then use that information to segment the cerebellum. We perform our testing on simulated data, provided by Brainweb and real data provided by the Internet Brain Segmentation Repository. We use the simulated data to demonstrate the complexity of cerebellum segmentation and to motivate various choices we made in our algorithm design. We look at the effect of the MRF parameters on the segmentation performance and see how our algorithm performs in comparison to the other algorithms. Using the real data, we more rigorously tested how our algorithm performs in comparison to the other algorithms.

Finally, Chapter 6 presents the conclusions of this work.

Chapter 2

Background

This chapter provides general background information about various aspects relevant to the rest of the thesis. We briefly introduce the cerebellum and Fetal Alcohol Spectrum Disorder (FASD) and describe the basics behind the Magnetic Resonance Imaging process. Next we discuss general image segmentation methods. Finally, we introduce brain tissue segmentation and briefly discuss the difficulties associated with it.

2.1 Cerebellum

The cerebellum is in the inferior posterior portion of the brain and is divided into two hemispheres. Figure 2.1 shows its position in the brain. The outside of the cerebellum is highly convoluted and it contains both white and grey matter, with the white matter forming branchlike structures inside the grey matter.

Typically, it is known to be involved in the coordination of somatic motor activity, the regulation of muscle tone, and being concerned with mechanisms that influence and maintain equilibrium [11, 58]. However, there has recently been a major shift in the understanding of the role of the cerebellum in the nervous system. There is recent evidence that indicates that the cerebellum is involved in a wider range of functions, such as learning, planning and temporal sequencing [96]. It is because of this recent realization that studying the cerebellum has become very important.

One of the aims of this thesis is to segment cerebellar white and grey matter on MR images, capturing as much of the fine branching structure as possible.

Looking at the literature, there is little work that focuses specifically on cerebellum tissue segmentation. Some techniques used to segment the cerebrum have been applied to the cerebellum, although with limited success, since they usually fail to capture fine-

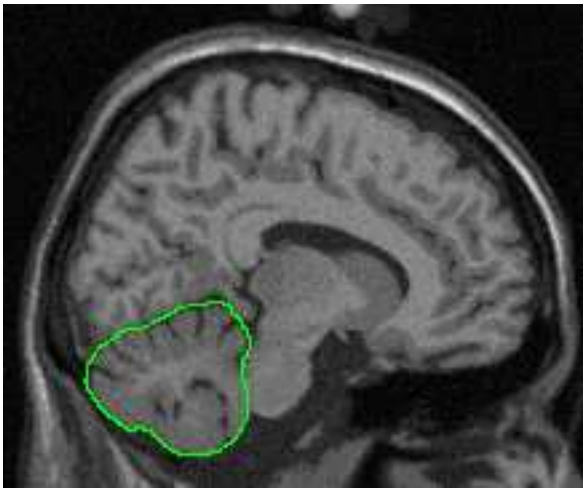


Figure 2.1: The cerebellum and the rest of the brain. The cerebellum is shown outlined in green.

scale cerebellar detail [73]. In spite of this fact, it is helpful to review methods used for brain tissue segmentation, since many of the same principles apply. When adapting these algorithms to better suit the cerebellum, a few key issues need to be considered. The method of segmenting the cerebellum needs to be able to capture finer structure. Furthermore, although brain tissue is usually segmented into grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF), the cerebellum contains only GM and WM.

2.2 Fetal Alcohol Spectrum Disorder

Fetal Alchol Spectrum Disorder (FASD) describes the range of lifelong effects that can occur in an individual whose mother drank alcohol during pregnancy. These effects can be physical, mental or behavioural [115] and are especially common in the Western Cape, making it a very relevant area of research in South Africa.

The most severe form of FASD is Fetal Alcohol Syndrome (FAS), which is characterised by distinctive craniofacial dysmorphology (short palpebral fissues, thin upper lip and flat philtrum), small head circumference and growth retardation [95]. FAS is associated with a range of problems, such as learning difficulties, attention and memory deficits. There is no agreement with regard to the neurobiological changes that are related to prenatal alcohol exposure, however since there is a wide range of behaviours that are disrupted, there are likely to be many changes in the brain. These include biochemical, physiological and neuroanatomical changes [59].

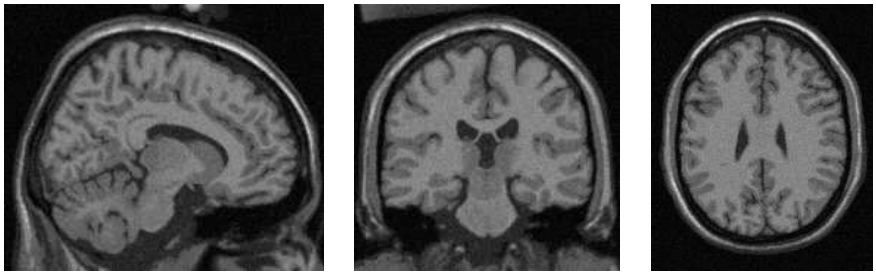


Figure 2.2: Examples of MR Images in three orthogonal planes: sagital, coronal and axial

2.3 Magnetic Resonance Imaging (MRI)

Magnetic Resonance Imaging (MRI) is a non-invasive medical imaging technique that is used to image the soft tissue in the body. Different tissues have different magnetic responses, causing different signal intensities in the resultant MR image. MRI has many advantages. It is a safe and non-invasive technique which creates high resolution images with good contrast (a difference in brightness) between different soft tissues, such as grey and white matter. There is no radiation exposure. Because of this, it has been widely used in brain research, as well as for diagnosis. It is a very flexible technique: the contrast can vary depending on the way the image is acquired, which allows for a great range of diagnostic options. The disadvantage of MRI is that it is very expensive. Examples of MR images are shown in Figure 2.2.

Pooley [109] provides a good introduction to MRI and some of the concepts are summarised below.

The main magnetic field is the major part of the magnetic imaging system. This magnetic field is created by a large electrical current flowing through wires in a loop.

The human body contains large amounts of hydrogen, in both fat and water. These hydrogen nuclei each contain one proton. Each of the positively charged protons spins around its axis, creating a tiny magnet. Usually, the hydrogen proton axes in the body are randomly oriented, without a predominance in any direction. Consequently, their "spins" cancel each other out, so no net magnetic field is created.

When the patient is exposed to a strong magnetic field, some of the hydrogen protons in their body will align in the same direction as the main magnetic field, while others align in the opposite direction. There will be slightly more protons aligned with the magnetic field, resulting in a *net magnetisation* that is parallel to the main magnetic field. We call this direction the *longitudinal direction*[1], which is usually along the bore of the magnet,

[1]Net magnetism aligned with the longitudinal direction is known as *longitudinal magnetism*.

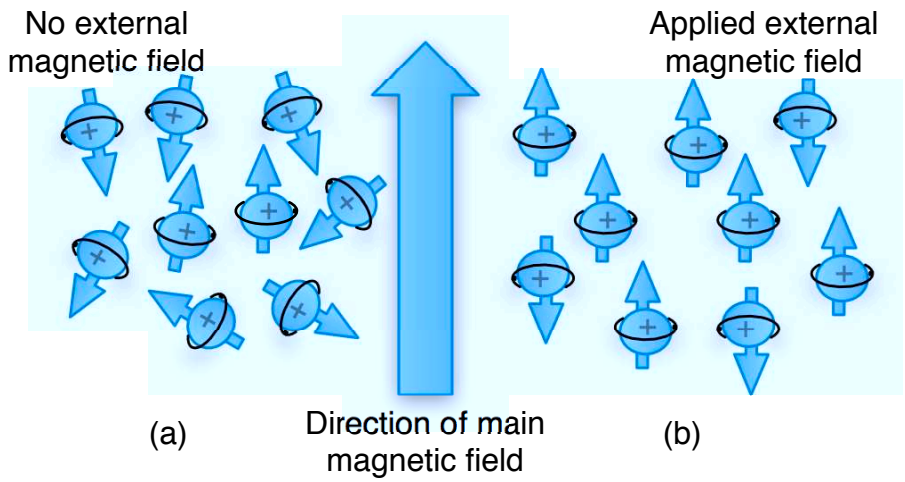


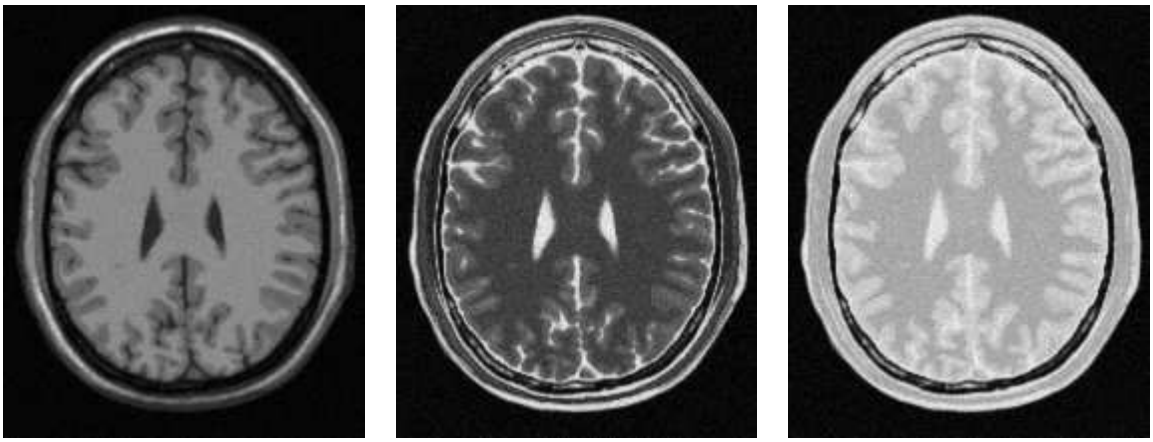
Figure 2.3: (a) Protons initially randomly oriented resulting in no net magnetic field. (b) When exposed to a strong magnetic field, many protons align in the same direction as the magnetic field, resulting in a net magnetisation parallel to the main magnetic field.

running from the head to the foot (or vice versa) of the person. This process is shown in Figure 2.3.

To create an MR image, the radiofrequency (RF) transmit coil in the MR system transmits RF energy for a short time. This transmitted RF energy must be equal to the energy difference between the parallel and anti-parallel states in order for energy transfer to take place. The protons in the body absorb energy from the pulse, which causes the equilibrium magnetisation to be disturbed. After terminating the RF pulse, the system returns to its equilibrium magnetisation by emitting energy. This energy is the measured MR signal.

T1-weighted Images There are various MR imaging protocols - for example *T1-weighted*, *T2-weighted* and *proton density-weighted* - each exploiting different tissue properties to produce images with different contrast. Since the images used in this project are T1-weighted, the creation of this contrast will be explained.

All MR images are generated by repeatedly applying RF pulses to upset the equilibrium magnetisation and allowing time (called repetition time TR) between successive RF pulses for signal readout and the magnetisation to relax back to its equilibrium value. When using short TRs, tissues with short T1 (eg white matter) will have relaxed almost fully, yielding maximum signal after the next RF pulse. By contrast, long T1 tissues (eg CSF) will not have relaxed substantially, yielding less signal after the next RF pulse. In this way, the contrast in the image will arise predominantly due to differences in T1 between tissues, producing a T1-weighted image. In T1-weighted brain images, white



(a) T1-weighted (b) T2-weighted (c) Proton density

Figure 2.4: Different modalities of MR images

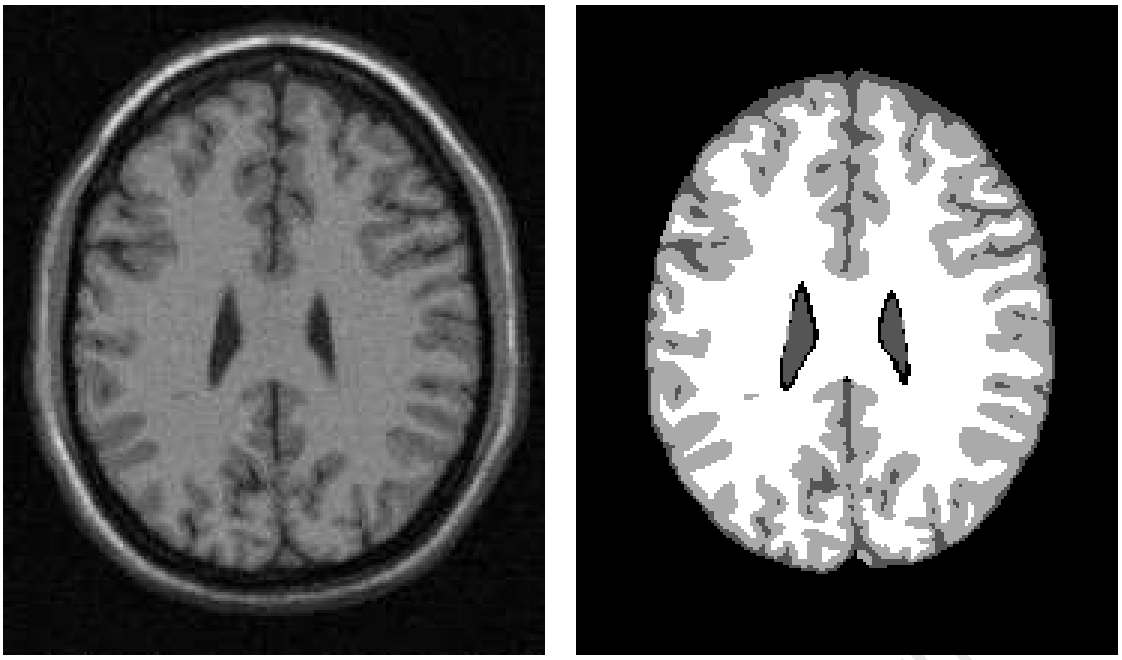
matter will be the brightest, grey matter intermediate, and CSF dark.

When considering imaging the brain, white matter has a very short T1 time and relaxes rapidly, CSF a long T1 time and relaxes slowly and grey matter has intermediate T1 time and relaxation rate. Since white matter has the shortest T1 time, its longitudinal magnetism is the greatest and a large number is recorded in the voxel located at that position in the image. This large number corresponds to a brighter intensity in the image.

T1-weighted images are used in this project as they provide good contrast between the grey and white matter in the brain.

Multi-spectral Images When only one MR image of an object is created, it is referred to as a single-spectral image. T1-weighted images are an example of this and are what we segment. When MR images with different contrast of the same object are created, they are referred to as a multi-spectral image, see Figure 2.4. For a given scanning time, the voxel sizes that can be achieved for single-spectral images are smaller than for multi-spectral images. However, one voxel of a multi-spectral image contains more information than one voxel of a single-spectral image [111].

When one needs precise and accurate measurements, single-spectral images are well suited, due to their smaller voxel size. This is especially true when dealing with the fine structures in the cerebellum.



(a) Original image (b) Segmented image

Figure 2.5: (a) Original MR brain image and (b) the same image segmented into tissue classes.

2.4 Image Segmentation Techniques

Image segmentation is the process of dividing an image into homogeneous regions. This is equivalent to finding the boundaries between the regions. The aim of segmentation is to simplify the image into something that is more meaningful and easier to analyse. In the case of brain tissue segmentation, segmentation involves assigning each voxel to either grey matter (GM), white matter (WM) or cerebrospinal fluid (CSF), or a combination of these classes (see the section on the partial volume effect). Figure 2.5 shows an example of tissue segmentation performed on an MR image of the brain.

Segmentation results can either be *hard* or *soft*. Soft (or *fuzzy*) segmentation allows pixels to belong to more than one class, with varying degrees of membership, keeping more information than hard segmentation. Hard (*crisp*) segmentation forces a pixel to belong to a single class.

The are many different segmentation techniques and in medical image segmentation, and looking through the literature, usually not just a single method is used, but a combination of a number of them.

This section briefly describes various image segmentation techniques. For simplicity, we will assume that we are looking at 2D images, with a single feature, intensity. This assumption of 2D images is dropped when we look at volumetric brain tissue segmenta-

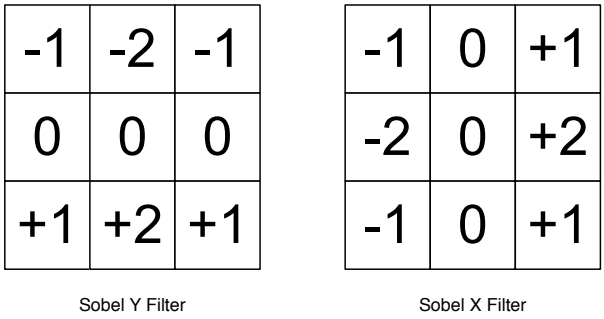


Figure 2.6: Sobel filters

tion.

If one is interested in knowing more about these techniques, there are a number of useful references [106, 85, 51, 148, 28, 16, 125, 86].

Edge Detection

Edge detection involves the detection and localisation of sharp changes in intensity, which should correspond to boundaries between different regions of the image. Most edge detection techniques involve three operations: smoothing, differentiation and edge labeling [155]. Smoothing reduces noise and helps to ensure robust edge detection. However, this smoothing process can also result in information loss. Differentiation is the process of computing the derivatives in order to locate sharp intensity changes. Edge labeling localises edges and suppresses false edges. Two examples of edge detection methods are Marr Hildreth [62] and Canny edge detection [23].

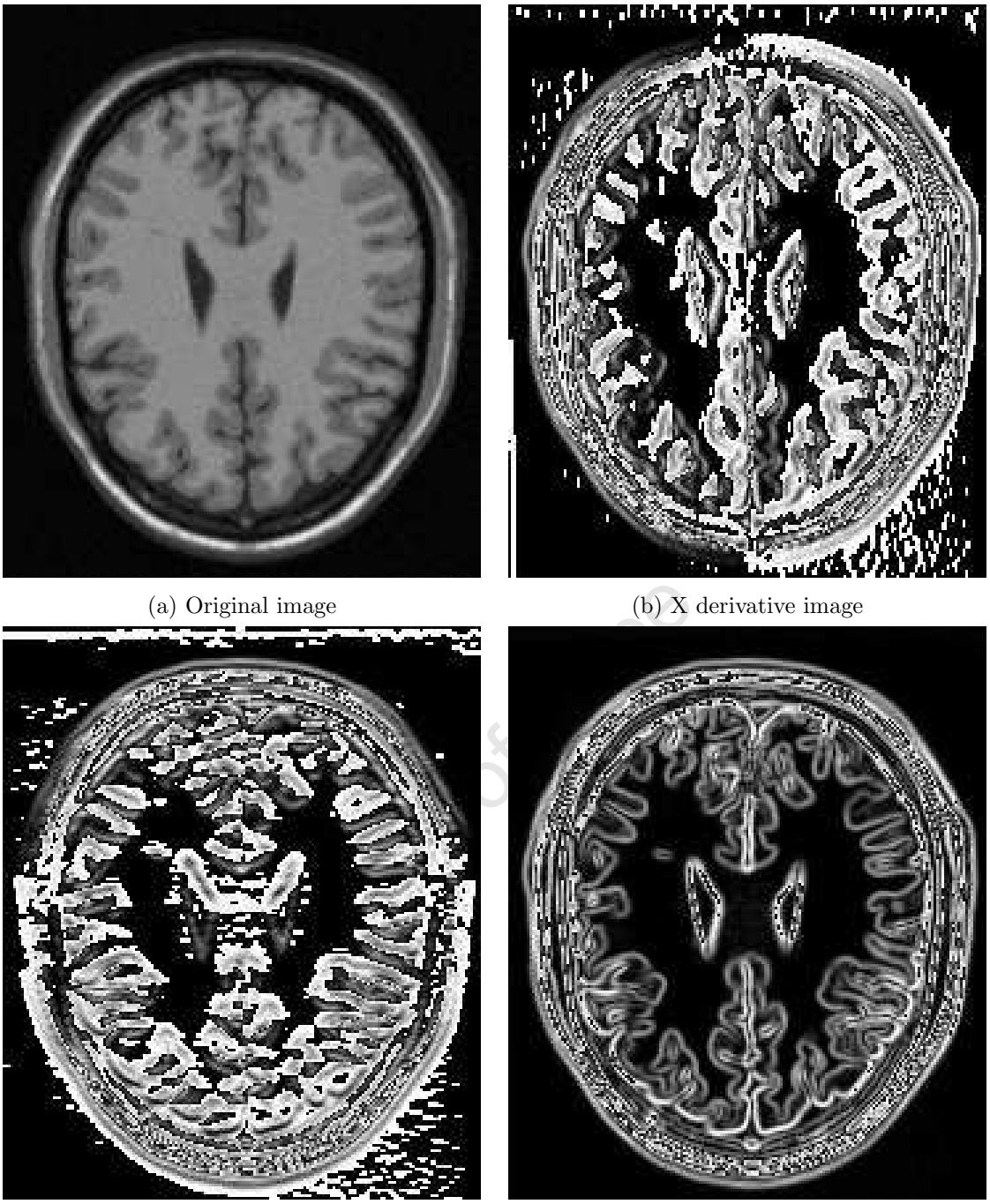
A Sobel filter, seen in Figure 2.6, is an operator that can perform differentiation on an image.

The original image 2.7a is convolved with each of the Sobel filters to create an X derivative image 2.7b and a Y derivative image 2.7c. These two images are combined to form a magnitude image 2.7d, which gives the strength of the edges. The magnitude image is then thresholded and all pixels with values greater than the threshold are chosen as edges.

A detailed discussion of edge detection methods can be found in Davis [40]. .

Histogram Thresholding

In its original form, a single threshold T is chosen to create a binary partition of the image. The aim of *thresholding* is to choose an intensity value to separate the desired regions, where each region corresponds to a mode of the histogram. All pixels with



(a) Original image

(b) X derivative image

(c) Y derivative image

(d) Magnitude image

Figure 2.7: Edge detection with a Sobel filter

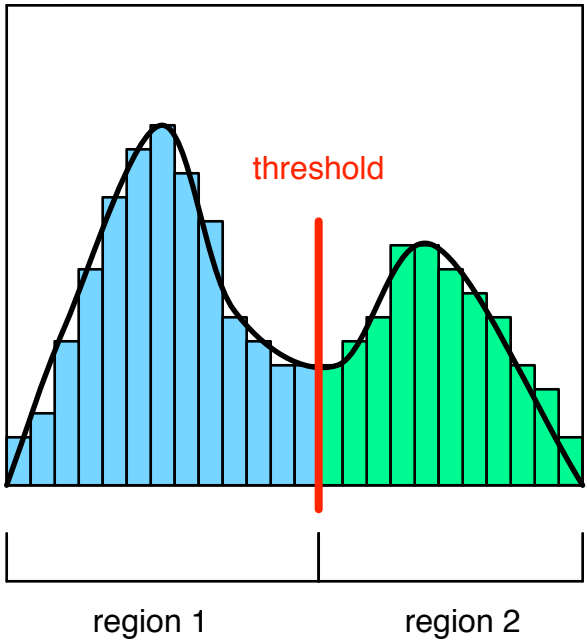


Figure 2.8: Histogram Thresholding

intensities greater than T are assigned to one class, and the rest of the pixels are assigned to the other class. This is illustrated in Figure 2.8. The choice of the threshold value is critical, and the segmentation results are dependent upon it. Approaches for choosing a threshold value include taking the threshold to be the intensity half way between the peaks of each histogram mode, or choosing the threshold to be the local minimum of the valley between them. Thresholding is a simple approach and can be very effective when there is good contrast between regions. However, thresholding in its simplest form does not take into account any spatial characteristics and so is sensitive to noise and inhomogeneities, which corrupt the histogram [106]. Smoothing can be applied in an attempt to improve thresholding results [85], but thresholding rarely produces a perfect result [86], and is often used as an initial step in the segmentation process, rather than on its own.

For images with more than two homogeneous regions, multi-thresholding, where multiple thresholds are used, has been developed [119]. In this case, each region is defined by upper and lower intensity thresholds, and includes all pixels with intensities lying between those two thresholds. Other variations used to improve results include information based on local intensities or connectivity.

Fuzzy thresholding is a generalisation of hard thresholding [86], where the segmentation result specifies the degree to which a pixel belongs to a region.

A detailed discussion of image thresholding techniques can be found in Sahoo *et*

al[119].

.

Region growing and related methods

Region growing and related techniques have been widely used in medical image segmentation [112, 108, 113, 65, 41]. Region growing attempts to identify homogeneous regions based on the characteristics of the pixels within the region. In its simplest form, a seed point is manually selected for each region. Each seed point becomes a growth point and pixels adjacent to it are considered. A pixel is added if it meets a certain predefined homogeneity criterion. This process is continued, with each new pixel becoming a new growth point, until no more pixels can be added. The homogeneity criterion is very important, but it can be difficult to establish suitable homogeneity criteria to achieve the desired segmentation. This region growing process is illustrated in Figure 2.9.

Region growing has the very desirable property of forming connected regions, which is what is desired from segmentation. A major disadvantage of the simple region growing algorithm is that seed points need to be manually selected to initialize the algorithm. Improvements to simple region growing [63, 102, 2] have dealt with this problem. Variations include fuzzy region growing [135], preserving topology between the initial region and the extracted region [88], and including edge constraints with the homogeneity criteria. Region growing can be sensitive to noise [105] and the partial volume effect can cause disconnected regions to connect. Because of these problems, region growing is usually not used alone in medical image segmentation but rather used in combination with other image processing techniques.

A few closely related algorithms are *region splitting*, *region merging* and the *split-and-merge* technique [90]. Region merging starts with each pixel being a region. Adjacent regions are merged if they meet the homogeneity criterion. Region splitting begins with all image pixels allocated to a single region. If this region does not meet the homogeneity criterion, it is divided into equal sized regions. The process is then repeated with each of the new regions until each region meets the homogeneity criterion. The split-and-merge technique is a combination of the previous two algorithms. These three methods all have the advantage of not requiring seed points.

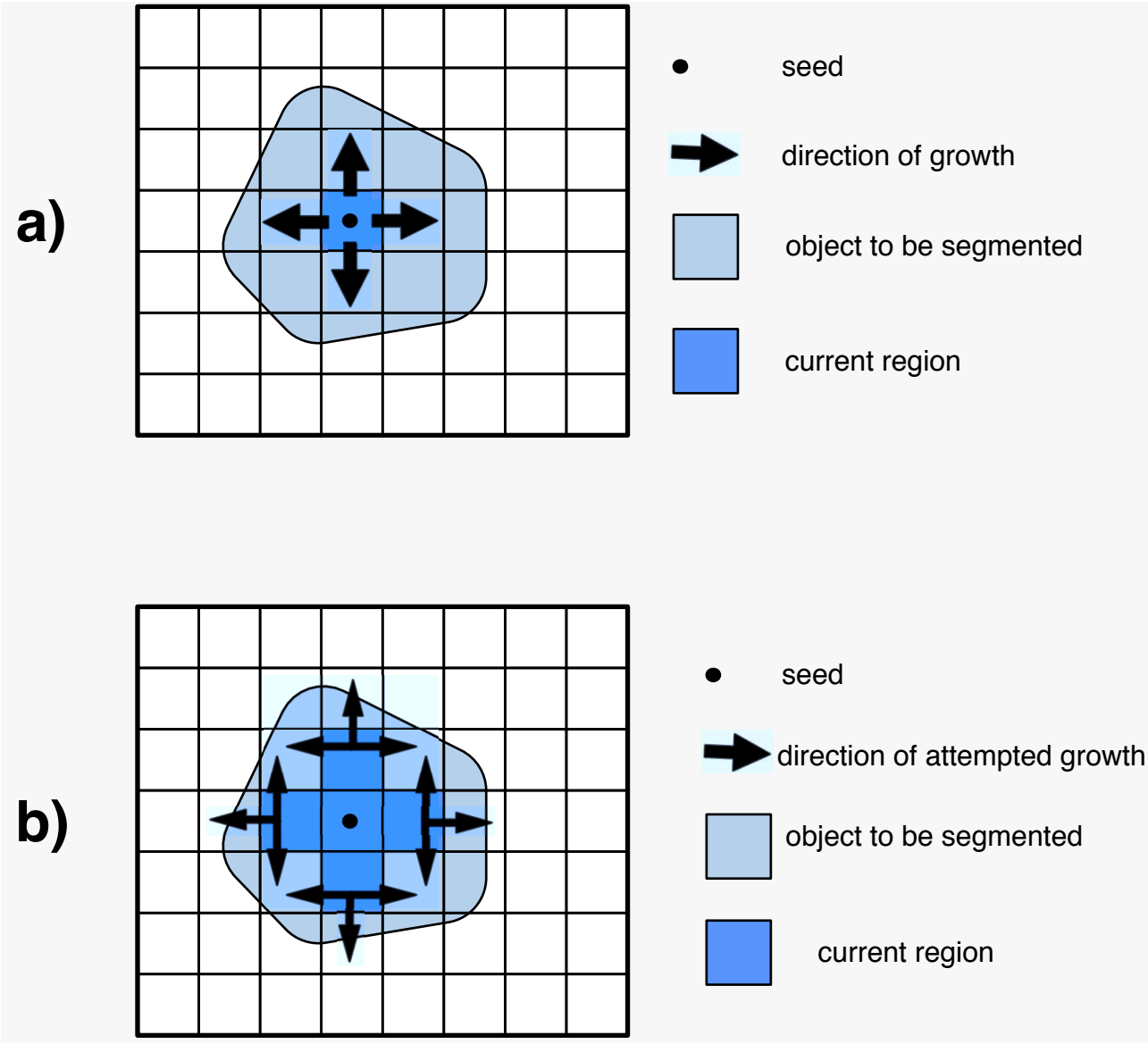


Figure 2.9: During the region growing process, a region initially starts as a single seed and attempts to grow outwards (a). When new pixels are added to the region, they also attempt to grow outwards (b).

Classifiers

Classifiers are a *supervised* segmentation technique [106], requiring training data which has been manually segmented. The idea is to divide the image into regions with known labels, based on how the intensities are partitioned in the training data. This assumes that the image being segmented is similar enough to the training data. For example, if the classifier is trained using images of normal anatomy, it may fail when segmenting an image of abnormal anatomy.

The simplest classifier is the *nearest neighbour* classifier, where each pixel is assigned to the same class as the pixel in the training data that is closest in intensity to it. A modification of this is the *k-nearest neighbour* classifier, where a pixel is classified in the same region as the majority of the k pixels nearest to its intensity in the training data. The *Parzen window* classifier [101] is another similar, well-known method.

Unlike the previous methods, the *maximum likelihood* (ML) or *Bayes classifier* [116] is a *parametric classifier*, assuming that the data follows a finite Gaussian mixture distribution. The training data is used to calculate the parameters for the Gaussian distributions, which are then used to segment the new data. This process will be explained in more detail later, where *ML clustering*, the unsupervised equivalent to this method is discussed.

Classifiers are computationally efficient once the initial training data has been classified. They have the advantage of automatically providing labels for the different regions. A major disadvantage is that the training data has to be manually segmented, which is a very slow, labour intensive process. The training data is only applicable to a small range of images, so many different training data sets need to be created.

Classifiers, in their simplest form, do not consider any spatial interactions between pixels, and therefore may perform poorly on images with high noise.

Clustering

Clustering based segmentation methods are the *unsupervised* counterparts to classifiers. They require no training data, but instead train themselves by iterating between calculating a group property (for example, mean intensity) for each region and classifying the pixels into the region whose group property most closely matches the pixel's intensity.

Probably the best known clustering algorithm is the *k-means* clustering algorithm, first used by MacQueen [87]. This algorithm iteratively calculates the mean intensity of each region and classifies the pixels into the region whose mean is closest to their intensity. This algorithm is a *crisp* segmentation method, meaning that a pixel can only belong

to a single region. *Fuzzy c-means clustering* [15] is the soft-segmentation equivalent to the k-means algorithm. Fuzzy c-means allows a pixel to belong to more than one region. Another commonly used clustering technique is the *Expectation-Maximisation* (**EM**) algorithm, which is discussed in detail in section 3.1.

Clustering has a major advantage over classifiers in that they do not need manually segmented training data. They do, however, need to be properly initialised, either with an initial segmentation or appropriate parameters. If the initialisation is poor, the algorithm might not produce an optimal segmentation. Another disadvantage is that in their original form, they do not include any spatial modelling. This is overcome in the adaptive k-means algorithm [152, 38, 141], which adds local adaptivity to the k-means algorithm. Similarly, *Markov Random Fields* [76] can be used for spatial modelling with the **EM** algorithm, which is discussed in more detail in Section 4.2. Due to their iterative nature, clustering algorithms are computationally expensive [110].

Deformable models

Snakes are an early form of deformable models. They are also known as *active contours* and were proposed by Kass and Witkin [74]. They are *closed parametric curves* that deform iteratively under the influence of *internal* and *external forces*. The idea is that by minimising the *energy function*, the sum of *internal* and *external forces*, one ends up with the curve delineating the object of interest.

The internal forces are *stretching* and *bending forces*, which enforce smoothness and continuity constraints on the curve. The external forces pull the curve towards the edges in the image.

The contour needs to be initialised near the boundary of the object of interest in order for it to converge to the correct boundary. This is a major disadvantage of the original snake model, as manual interaction is required to place the initial contour, as well as to choose appropriate parameters.

Snakes are more robust to noise than edge detection methods, due to their smoothness and continuity constraints, which cause gaps in edges to be filled and spurious edges to be ignored. Another advantage of snakes is that they directly generate closed parametric curves.

Keleman *et al* [75] and Caselles [24] give brief backgrounds on snakes. Initially, snakes were only used in 2D, but they were extended into 3D active surfaces [31, 129]. The initial snakes had a problem of being unable to converge to concave boundaries, but this was

overcome with *pressure forces* [32] and modified external forces [150].

Due to their smoothness constraints, snakes have difficulty conforming to deeply convoluted boundaries. This makes them unsuitable for brain tissue segmentation.

McInerney and Terzopoulos [92] wrote a survey on deformable models in medical image analysis which is a useful reference.

Atlas Based Segmentation

Atlas based segmentation is a way of incorporating *a priori* knowledge about anatomy into the segmentation algorithm. An atlas is either the segmented image of a particular part of a single person, or a composite image created from segmented, co-registered images from many different people. This composite version of the atlas creates a probabilistic atlas and provides information about population variability [148].

When performing atlas based segmentation, the atlas (a segmented and labelled image) is registered with the image to be segmented: a transformation is found that maps the reference image (the atlas) to the target image (the image to be segmented), using a process known as *atlas warping*. Once this transformation has been found, the region labels can be projected on to the target image.

A problem with atlas based segmentation is that it is difficult and computationally expensive to determine a robust and accurate registration. This is due to anatomical variability and so this type of segmentation is better suited to segment structures that are relatively stable over the population. However, if an atlas is available and a mapping can be efficiently found, atlases can significantly improve the accuracy of segmentation results [127, 70, 7].

Atlases do not need to be used entirely on their own, but can also be incorporated into other segmentation algorithms. For example, they are often used to initialise other segmentation methods.

Atlas based segmentation would not be suitable for use in our case, as they usually do not perform well on abnormal brains and we will be considering people with FASD. It is also not known how similar the anatomy of the cerebellum of different people are, further making an atlas based technique unsuitable.

2.5 Brain Tissue Segmentation

Brain tissue segmentation involves segmenting the brain into grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF). Here, we assume that the entire brain has already been segmented from its surroundings. Hard tissue segmentation involves assigning each voxel a label, usually GM, WM or CSF. Sometimes other classes are considered: GM + WM; GM + CSF; WM + CSF; WM + GM + CSF. In soft classification, each voxel is associated with 3 values: the fraction α_k of each tissue type k that it contains. Accordingly, $0 \leq \alpha_k \leq 1$ and $\Sigma_{k=1}^{3} \alpha_k = 1$. When dealing with the segmentation of the cerebellum, we consider only two tissue classes, GM and WM.

Reviewing the literature, one can see that little work has been done involving the tissue segmentation of the cerebellum. Sometimes the cerebellum is included in the tissue segmentation of the whole brain, but even when the PVE (partial volume effect) is accounted for, the segmentation can fail to pick up its fine structures [139]

The results of brain tissue segmentation are used for numerous applications, such as visualisation, anatomical studies, predictive models and the study of brain disorders. The study of many brain disorders, such as Alzheimer's, epilepsy, multiple sclerosis and schizophrenia, has been aided by tissue segmentation. Segmentation allows for quantitative volume analysis, which can aid in diagnosis, help assess the severity of the disease and help evaluate response to drug therapy [117, 122, 44, 56, 79, 54].

Many brain disorders, including FASD, have an effect on the structure of the brain [6]. We are interested in how FASD affects the structure of the cerebellum. Tissue segmentation of the cerebellum is a good place to start when trying to determine how FASD affects the cerebellar structure.

Tissue segmentation of the cerebellum differs from tissue segmentation of the brain in two major ways. Firstly, the cerebellum only has two pure tissue classes, GM and WM; and secondly, the structures of the cerebellum are much finer. Some brain tissue segmentation methods may be applicable to the segmentation of the cerebellum, but these differences need to be taken into account.

Automatic Segmentation In many clininal studies, tissue segmentation is performed manually or strongly supervised by an expert [83]. This is a very time-consuming task and impractical for large amounts of data. Zhang *et al* [154] consider this to be one of the biggest obstacles in the effective use of MRI.

Manual segmentation results often display large inter- and intra-observer variabil-

ity, which means that they are not reproducible. This decreases the significance of the analyses and effects that rely on the segmentation [138].

Ideally, an automatic tissue segmentation method should be able to process large quantities of data in a reasonable amount of time and to produce a segmentation that is reproducible and correlates well with the segmentation produced by experts.

2.6 Difficulties with Brain Tissue Segmentation

Despite the fact that there has been much focus on brain tissue segmentation, it still remains a challenging task [21]. Anatomical structures, particularly in the brain, are often non-rigid and have complex shapes, which are difficult to identify. To add to the difficulty, organ shapes show much variance even between normal individuals, making it difficult for explicit shape models to fully capture the anatomical variation [73]. Although MR images provide high spatial resolution and good tissue contrast - making them the best method of imaging the brain [61] - their finite resolution means that there are often many *partial volume* voxels [48, 140]. Other difficulties in the segmentation of MRI data include inhomogeneities caused by the field bias, noise and motion artifacts [73]. These difficulties are especially problematic if trying to segment tissue types based only on individual voxel intensities [126].

Partial Volume Effect When dealing with brain tissue segmentation, the partial volume effect (PVE) occurs when more than one type of pure tissue (GM, WM and CSF) occupies a voxel. This is due to the limited resolution of MR images and the complexity of tissue or organ boundaries. The voxels containing more than one tissue type are known as *mixels*[27] and the intensity value of a mixel depends linearly on the proportion of each pure tissue in the voxel [54]. This is known as *partial volume averaging*. Mixels are said to belong to mixed classes or mixture classes: classes containing more than one tissue type. The partial volume effect is illustrated in Figure 2.10.

We assume there are K' different tissue types, with K of them being pure tissues types and $K' - K$ being mixed tissue types.

The PVE occurs mainly along the border regions between the tissues and can cause fine structures to be lost [83]. Because of the fine structures in the cerebellum, it is particularly important that we deal with the PVE.

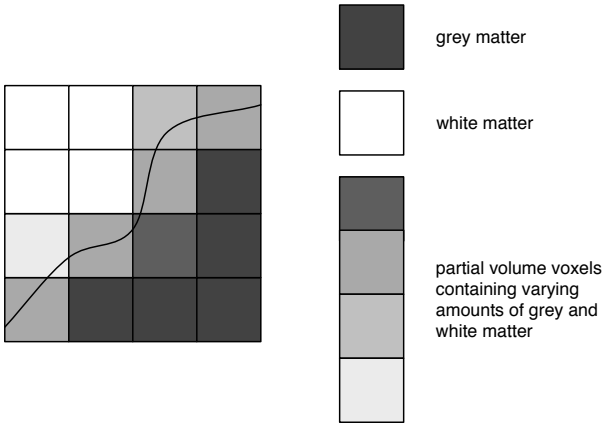


Figure 2.10: The partial volume effect

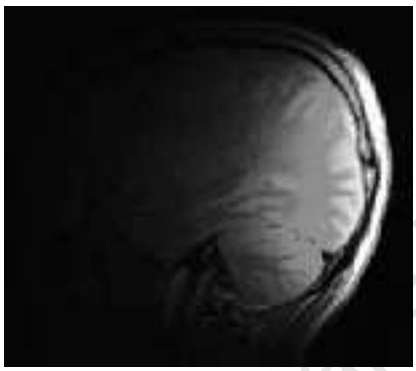


Figure 2.11: An example of bias field inhomogeneities

Ignoring the PVE and performing hard tissue segmentation introduces significant errors in quantitative measurements [97].

Bias Field Inhomogeneities A *bias field* is a smoothly spatially varying intensity inhomogeneity that can occur when the MR data is acquired. This is seen as a smooth increase or decrease in intensity across the MR image and is due to inhomogeneities in the magnetic fields of MR systems. Bias fields usually do not affect the visual diagnosis process by humans, but can interfere with automatic tissue segmentation [137, 10]. Figure 2.11 shows what bias field inhomogeneities can look like if the image is badly affected.

We are not going to address the problem of inhomogeneities in this thesis. Due to its small size and compact shape, the segmentation of the cerebellum will be less affected by inhomogeneities than the segmentation of the cerebrum. For more information about bias field inhomogeneities and ways to deal with them, see Sled *et al* [128].

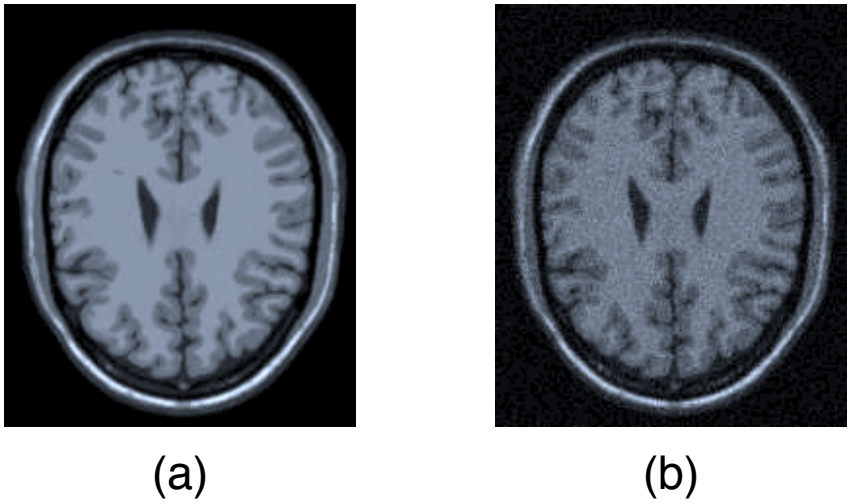


Figure 2.12: (a) Image with no noise added. (b) Image with noise added.

Noise All MR images are affected by random noise [83], which causes the images to be less than perfect. This noise is caused by time and equipment limitations [61]. Noise causes tissue classes to overlap in the image intensity histogram [154], making tissue segmentation more difficult. Figure 2.12 shows the effect of noise on an MR image.

Closing Comments Edge detection based techniques tend to be more susceptible to noise than intensity based techniques [142], making intensity based methods better suited to tissue segmentation. A class of these intensity based methods, clustering based techniques are particularly suitable for tissue segmentation. Clustering based methods work well when the means of the various classes are well separated [67]. This is the case for brain tissue segmentation, where the means of the intensities of the tissue classes are relatively far apart, making clustering based methods an ideal choice.

Because of the fine structures in the cerebellum, a large proportion of the voxels will be partial volume voxels, making the partial volume effect very important to address when segmenting the cerebellum.

Various ways to address the segmentation difficulties described here are discussed in the next chapter, which reviews brain tissue segmentation methods.

Chapter 3

Brain Tissue Segmentation

This chapter reviews the literature regarding the methods used for brain tissue segmentation. Emphasis is placed on the *Gaussian mixture model*, as much of the literature deals with methods involving that model.

3.1 Gaussian Mixture Model

Looking through the literature, the model most commonly used in brain tissue segmentation is the *Gaussian mixture model (GMM)*. Nurmi [100] provides a good introduction to mixture models. This statistical parametric model assumes that the distribution of the intensities of a given tissue class can be represented as a Gaussian probability distribution. This has been shown by Gudbjartsson and Patz [57] to be a valid assumption for pure classes.

A mixture model is a probabilistic modeling tool that is useful for statistical data clustering, which includes the clustering of pixels or voxels according to intensity during image segmentation [49], and in particular brain tissue segmentation.

When dealing with brain tissue segmentation in a Gaussian mixture model framework, the simplest model ignores the partial volume effect and bias field inhomogeneities, and assumes no spatial correlation and deals with only single-spectral MR images (i.e. each voxel is associated with just one intensity value). In this model, voxel intensities are assumed to have come from a mixture of a number of *populations* (in this case the tissue classes) [107]. Each voxel is assumed to belong to one of the tissue classes and each pure tissue class is given its own distribution.

Suppose there are K different tissue classes making up the model. Each of these tissue classes has a different intensity distribution, called its *component density*. This

density function gives the probability of a voxel having a specific intensity, given that it belongs to tissue class k. Often, these functions are parametric with each tissue class k characterized by parameter θ_k.

When segmenting brain tissue from MR images, it is common to assume that each tissue class has a Gaussian distribution of intensities [138, 137, 122]. We assume that component k has a Gaussian distribution. Since we are dealing with single-spectral data, this will be a univariate Gaussian function, G with parameters $\theta_k = \{\mu_k, \sigma_k\}$, where μ_k is the mean intensity and σ_k is the standard deviation. This can easily be adapted to deal with multi-spectral data by using a multi-variate Gaussian [98, 99, 138, 137].

Further, suppose we have an image with N voxels, where voxel i has an observed intensity of x_i. If one picks a random voxel in the image, the probability of it belonging to tissue class k is w_k. These w_k's are the mixing weights of the model, representing the prior probability of a random voxel belonging to tissue class k, when the only information available about the data is the weights of each component. Since the w_k's are probabilities and each voxel must belong to a tissue class, we have the following constraints: $0 \leq w_k \leq 1$ for all k and $\sum_{k=1}^{K} w_k = 1$. We can combine all the parameters of the model into one vector $\theta = \{w_1, ..., w_K, \theta_1, ..., \theta_K\}$.

Thus, the probability distribution function of the mixture can be written as:

$$p(x_i|\theta) = \sum_{k=1}^{K} w_k p(x_i|\theta_k) \tag{3.1}$$

This represents the probability of voxel i having observed intensity x_i given the mixing weights and the Gaussian parameters. This model does not depend on the position of a voxel, so the same probability would hold for any voxel with the intensity x_i.

We assume that each tissue class follows a Gaussian distribution $G(., \mu, \sigma)$, the parameter θ_k is comprised of μ_k and σ_k, the mean and standard deviation of the distribution for tissue class k.

To give a tissue class label to each voxel, we assign a scalar value to each voxel, assuming it belongs entirely to that tissue class. We need to fit the Gaussian mixture to a normalised histogram of the image data. This requires finding parameters for each Gaussian component such that the model fits the histogram as closely as possible. Each voxel is then assigned to the tissue class to which it is most likely to belong.

Various different approaches to estimate parameters have been used to estimate these parameters, and these are discussed in the next subsection.

The aim is to determine the mean and standard deviation of each tissue class. Once

these parameters are known, segmentation can be performed. In a simple case, each voxel can be assigned to the tissue class to which it has the highest probability of belonging.

The model presented above is highly simplified and would not provide accurate segmentation. Various modifications applied in the literature to improve this model can be roughly divided into the following categories: i) Parameter Estimation, ii)Noise, iii) Inclusion of Spatial Information, iv) Accounting for the Partial Volume Effect, v) Inhomogeneities, vi) Underlying Model, vii) Parameter Initialisation.

Parameter Estimation

When using the Gaussian Mixture Model to segment an MR image into tissue classes, we need to estimate both the class labels and the parameters of the model, making it an incomplete-data problem [154].

The method most commonly used for parameter estimation with the Gaussian mixture model is the iterative *Expectation-Maximisation (**EM**) algorithm* [43]. The **EM** algorithm was first used in the context of tissue segmentation by Wells *et al* [146] and subsequently used for tissue segmentation by numerous others [98, 99, 122, 138].

The **EM** algorithm tries to optimise the parameters of the Gaussian distributions so that they fit the histogram of the image as closely as possible. To do this, two steps are iteratively interleaved. Starting with an initial set of parameters, the **E** step computes the posterior probabilities of each voxel belonging to each tissue class. Given these new posterior probabilities, the **M** step computes the parameters of the Gaussian distributions. These new parameters are then used in the next **E** step. This iterative procedure continues until the algorithm converges. Given normal conditions, the **EM** algorithm is guaranteed to converge to a local optimum, though this is not necessarily the global optimum [43]. The **EM** algorithm is summarised in Figure 3.1.

The **EM** algorithm is widely used with the Gaussian Mixture Model, but it does have some drawbacks. Richard *et al* [114] mention some of the limitations of the **EM** algorithm. The **EM** algorithm is sensitive to the initialisation of the model parameters. If the initial estimates are too far from the actual values, the algorithm could converge to an incorrect result. Even if the parameters are well initialised, there still is the chance that the algorithm will be slow to converge; or that it will converge to a local maximum, missing the global maximum and producing undesirable results.

Many variations of the **EM** algorithm have been used to solve for the parameters during tissue segmentation. Many of these variations involve speeding up the running

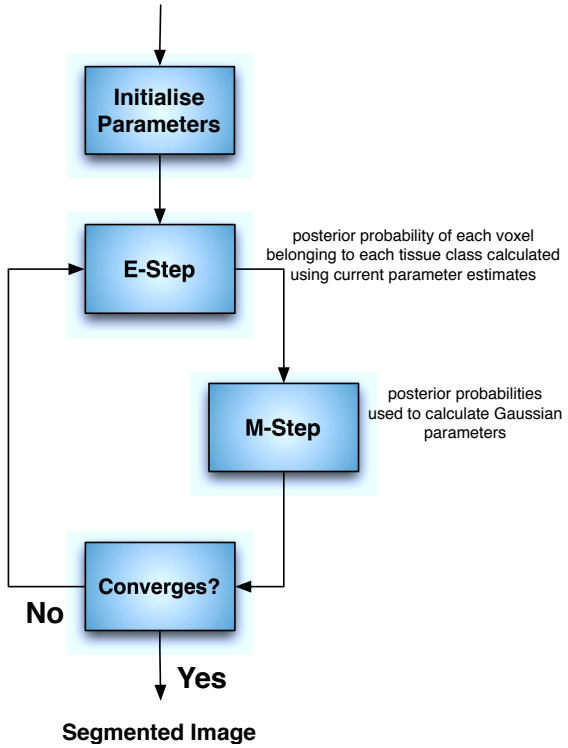


Figure 3.1: A summary of the **EM** algorithm.

time [35].

Zhang *et al* [153] deal with the problem of the partial volume effect by assigning a vector of values (known as a PVE vector) to each voxel (see "Accounting for the partial volume effect" below for more details), instead of just assigning the voxel to a single tissue class. The estimation of the parameters of the model and the PVE vectors is a rather difficult task. This is dealt with by modifying the **E**-step of the **EM** algorithm to use a greedy method similar to the *iterated conditional modes* (ICM) method [14] and a sampling method similar to the *Metropolis algorithm*, a stochastic sampling method. At each voxel, a random PVE vector is generated. Only if this vector yields an improvement in local energy, (the energy associated with the MRF, see "Inclusion of spatial information" before), is it kept as the voxel's new PVE vector.

The **EM** algorithm is guaranteed to converge to a stationary point of the likelihood function when used with Gaussian mixture models [43, 149]. However, it is not guaranteed to converge to the global maximum and suboptimal results can be obtained. The initialisation of the parameters $\theta^{(0)}$ of the **EM** algorithm has a significant effect on whether or not optimal results are obtained [107, 142]. As mentioned previously, this sensitivity to initialisation is one of the major drawbacks of the **EM** algorithm and papers have

been written to address this problem [18, 93]. To deal with the fact that the **EM** algorithm can be slow to converge [122], modified versions have been devised to address this problem [107], including deterministic improvements [84] and stochastic modifications [25, 42, 145].

Van Leemput *et al* [137] include an explicit parametric model for the bias field in their model and use a *3-Step EM algorithm*, with the additional step involving bias field parameter estimation. They use a Generalized **EM** algorithm: at each step they increase, rather than maximise the likelihood.

Richard *et al* [114] perform distributed EM algorithms, in order to cope with spatially varying means and standard deviations caused by intensity inhomogeneities. The image is divided up into partitions and a local control agent is assigned to each partition, where it performs a local EM algorithm on that partition of the image. Since each EM algorithm operates on only a a small partition of the image, it is more prone to modelling errors than if it had been performed over the whole image. To overcome this drawback, parameters estimated by local EM algorithms are compared with parameters interpolated from neighbouring partitions. The estimated parameters are only accepted if the difference between the estimated parameters and the interpolated parameters is less than a certain threshold.

As well as the **EM** algorithm, other methods of parameter estimation are used. Santago and Gage [120] use a tree annealing algorithm [19], based on the principles of simulated annealing, in order to estimate the parameters of the GMM. A drawback of this method is that the time-complexity increases very fast with increasing variables in the problem.

Another method commonly used is *Genetic algorithms* (GA). It is an optimization technique that was developed by John Holland [64]. It mimics natural selection, allowing an algorithm to adapt. Solutions are represented by a population of individual chromosomes, usually represented as binary strings. A chromosome is made up of genes, each of which can represent a particular characteristic. Each individual in the population is evaluated and given a fitness score based on how well they solve the particular problem. The higher the individual's fitness score, the greater their probability of breeding. Breeding creates the next generation through crossover and mutation. Crossover combines the chromosomes of two individuals, creating a new individual which is unlike either of the parents. Mutation, which occurs only a small percent of the time, randomly alters a new individual's chromosome. Since the more optimal individuals have a greater chance of breeding, the population tends to evolve and reach an optimal solution [130, 4, 50, 17].

Pernkopf and Bouchaffra [104] proposed a genetic-based expectation-maximization (GA-EM) algorithm for learning the parameters of Gaussian mixture models. This method combines both genetic algorithms and the **EM** algorithm into a single procedure. The use of GAs enables the GA-EM algorithm to explore the search space more thoroughly than a regular **EM** algorithm, enabling the algorithm to escape from local optimal solutions. This makes the algorithm less sensitive to its initialisation. Experiments show that the GA-EM algorithm outperforms the **EM** algorithm.

Schroeter *et al* [122] added a component with a uniform distribution to their GMM, to be discussed in the section section "Underlying model". This additional component greatly impaired the convergence of the **EM** algorithm, which necessitated the use of a different parameter estimation method. They used a genetic algorithm using a floating point representation, known as a Genetic Algorithm in Continuous Space (GACS) or Real-Coded Genetic Algorithm (RCGA). They found that the GA is less sensitive to the initial conditions than the **EM** algorithm is. This GA uses a flat crossover operator which, according to Tohka *et al* [132], causes premature convergence of the algorithm.

In order to overcome the initialisation problems of the **EM** algorithm, Tohka *et al* [132] also propose a method based on RCGAs to solve the parameter estimation problem globally. In order to avoid the problem of premature convergence, they use a blended crossover operation. They also introduce a new permutation operator to reduce the size of the search space.

Noise

In the Gaussian mixture model, noise is assumed to have a Gaussian distribution. In reality, scanner noise has a Rician distribution [22], but Gudbjartsson and Patz [57] have shown that the Rician distribution approximates a Gaussian distribution if the signal to noise ratio is high enough. In the case of T1-weighted scans, such as the ones we are dealing with, the signal to noise ratio is usually high enough for it to be a valid assumption. This assumption helps in simplifying the analytical treatment of the segmentation problem [72].

There are two main models used to deal with noise: tissue independent and tissue specific. In the physically motivated tissue independent model, the noise behaves the same way no matter what tissue type is being imaged. In this case, the standard deviation of all tissue types is the same [120, 79]. In the tissue specific model, the noise behaves differently depending on the tissue type being imaged. In this case, the standard deviations of the

tissue types are not in general the same, and so the standard deviations of the pure tissue types are modelled as independent parameters [35, 117, 111].

According to Zhang *et al* [153], the assumption of tissue independent noise might not be valid, as some tissue types (such as CSF) normally have much larger variances that others. Though not motivated by the physics of MR acquisition [35], the tissue dependent model gives more flexibility to the model and allows it to adapt to other types of artifacts.

Inclusion of Spatial Information

In brain MR images, there tends to be much overlap in the intensities of different tissue types [114]. Because of this, it is important to use spatial modelling in order to achieve good segmentation results. Van Leemput *et al* [140] demonstrate how the use of spatial information improves segmentation results and show that it is often indispensable for robust parameter estimation.

In the simple Gaussian mixture model, voxels are classified based only on their intensity. However, the tissue classes of the voxels surrounding a particular voxel should affect the decision regarding the tissue class of that voxel. For example, if there is one pale voxel surrounded by much darker grey matter voxels it is likely that the pale voxel actually belongs to grey matter and is pale due to noise, as shown in Figure 3.2.

Markov Random Field (MRF) models are frequently used with the **EM** algorithm to segment the cerebrum into its tissue types [35]. In most cases, the addition of an MRF to the **EM** algorithm improves results [35].

MRF models are used in computer vision to encode contextual constraints into the prior probability in a Bayesian framework [82]. They were introduced into computer vision by Geman and Geman in their seminal paper [52] regarding image restoration.

MRFs have been widely used in medical image processing and are the most common method of incorporating spatial correlation into the tissue segmentation process. They provide a method of smoothing homogeneous regions while still preserving edges. When an MRF is included, the segmentation is usually less susceptible to noise [140]. MRFs also introduce *a priori* knowledge by penalising unlikely tissue transitions [44].

Cuadra *et al* [35] demonstrate the fact that no matter what partial volume model is used (or whether one was used at all), methods including an MRF spatial prior outperforms their counterparts without a spatial prior.

An MRF is defined with respect to a *neighbourhood system*. The *neighbours* of a certain voxel are the voxels close to it, where "close" is defined in the particular appli-

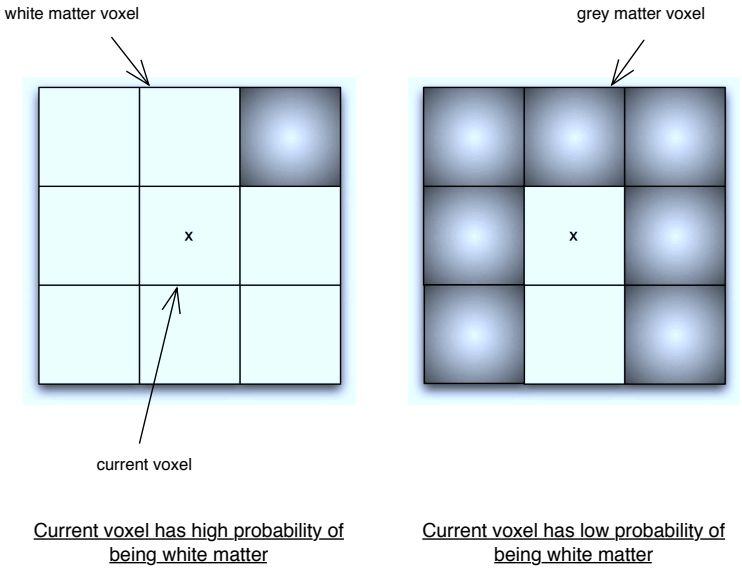


Figure 3.2: The basic idea behind the MRF

cation. Usually a first order neighbourhood - where directly adjacent to the voxel are its neighbours - or a second order neighbourhood - where the first order neighbourhood and the diagonals are the voxel's neighbours - is used. One of the ways in which the implementation of MRFs differ is the choice of neighbourhood system to be used. Some examples of different neighbourhood systems are shown in Figure 3.3.

By the Hammersley-Clifford theorem, an MRF can by characterised by a Gibbs distribution. Therefore, we can write:

$$p(y) = \frac{1}{Z} e^{-\beta U(y)} \tag{3.2}$$

where Z is a normalising constant, β is the neighbourhood parameter and $U(y)$ is the energy function.

In image segmentation, a simplified form of the energy function is usually used [154]:

$$U(y) = \sum_{i=1}^{N} \sum_{j \in \mathcal{N}_i} V_{ij}(y_i, y_j) \tag{3.3}$$

where $\mathcal{N}_i$ is the set of neighbours of voxel i and $V_{ij}(y_i, y_j)$ models the interaction between two neighbours.

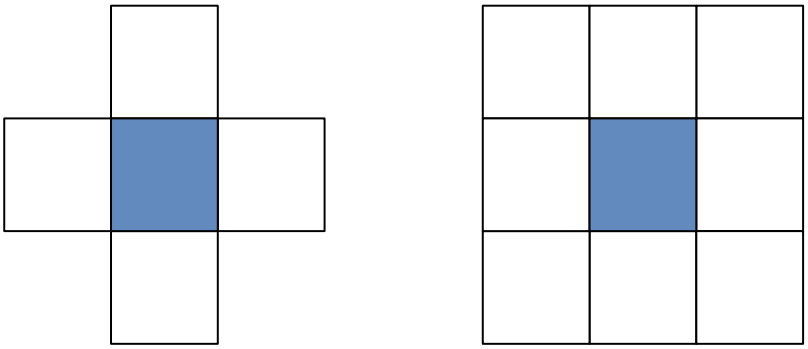


Figure 3.3: Possible neighbourhood systems

Since the prior probability is a part of what we maximise, we can use it to change the probability of a voxel belonging to a particular class. In particular, we want the probability of a voxel belonging to tissue class k to increase with the number of neighbours with that tissue type. The following function has been widely used [126, 98]:

$$V_{ij}(y_i, y_j) = \frac{\delta(y_i, y_j)}{d(i,j)} \tag{3.4}$$

where

$$\delta(y_i, y_j) = \begin{cases} -2, & \text{if } y_i = y_j \\ +1, & \text{otherwise} \end{cases}$$

and $d(i,j)$ is the spatial distance between voxels i and j. If larger neighbourhoods are used, this limits the influence of distant voxels.

This model can be extended to take partial volume (PV) classes into account [98, 35]:

$$\delta(y_i, y_j) = \begin{cases} -2, & \text{if } y_i = y_j \\ -1, & \text{if PV classes } y_i \text{ and } y_j \text{ share a constituent tissue type} \\ +1, & \text{otherwise} \end{cases}$$

Noe *et al* [99] extend equation 3.4 by *incorporating anatomical information from class template images*, which contain prior class probabilities of each voxel in the image. One class template image is needed for each pure tissue class and these images need to be registered to the image requiring segmentation. They assume that registration is not

perfect and so limit the influence of the class template images. Without this limitation, errors in the registration process could be detrimental to the segmentation. The following equation shows their extension:

$$V_{ij}(y_i, y_j) = \frac{\delta(y_i, y_j) - \alpha Q_i(k)}{d(i,j)} \tag{3.5}$$

where $0 \leq Q_i(k) \leq 1$ is the probability that the $i'th$ voxel belongs to class k according to the relevant class template.

α is a constant the determines the influence of the class templates to the segmentation. Noe *et al* set $\alpha = 2$. This value slightly favours neighbourhood information over spatial class information, so that even if the registration is completely wrong, the MRF can still include some useful information.

Class templates only provide prior class probabilities for pure tissue classes. While we do not go into the detail here, Noe *et al* create additional PV class templates by combining the pure class templates.

Noe *et al* found that including ideal class template information significantly improves segmentation performance, while still limiting the error as the prior information gets worse. The disadvantages of this method are that suitable class template images need to be available and that registration of the class template images is computationally expensive.

Instead of using a distance function, Rajapakse *et al* [111] weight the interactions with first and second order neighbours differently, by including two different β parameters, β_1 and β_2, corresponding to the first- and second- order neighbours respectively. They also take into account single-site neighbourhoods. If voxel i is a single-site neighbourhood currently belonging to tissue class k, the potential contributed by that voxel is α_k, where α_k is a real constant. These additions add an additional $K + 1$ MRF parameters.

Van Leemput *et al* [138] have a more complicated weighting function. Because voxels usually are not cubic, they weight the neighbouring voxels in the same slice differently to neighbouring voxels in adjacent slices. They use two KxK matrices as weighting parameters, with one matrix for voxels in the same slice and one for neighbouring voxels in the adjacent slice. They solve for these parameters during parameter estimation, which is a big advantage over trial and error as a method of selecting parameter values.

Ruan *et al* [117] perform an additional classification step on mixed classes, where each PV voxel is assigned to a pure tissue class. They do this by adding an extra energy component, which uses multifractal dimension information [81] to include structural

information.

Instead of using an MRF to include spatial information, Greenspan *et al* [56] take a very different approach in their *Constrained Gaussian Mixture Model*. Each tissue type is represented by many Gaussian components, and spatial information is included with intensity as a feature for each region. This provides a more localised analysis than an MRF and their algorithm tends to be more robust to noise. They have shown that their algorithm performs better than an EM algorithm with an MRF for medium to high noise levels. The algorithm is discussed in more detail below in the section "Underlying Model".

More recently, Bricq *et al* [21] use a *Hidden Markov Chain (HMC) model* instead of an MRF to take into account neighbourhood information. They use the HMC model instead of an MRF because it is less computationally intensive and does not need any estimation. The Markov chain uses a Hilbert-Peano path to visit each voxel.

Accounting for the Partial Volume Effect

It is very important to account for the partial volume effect. This was shown in Ruan *et al* [117] by demonstrating that when fitting the sum of Gaussian functions of pure tissue classes to the histogram of the MR image, the difference, according to the D'Agostino-Pearson K^2 test of normality [37], is significant. However, when 5 Gaussians were used, taking into account the partial volume effect by including 2 mixed classes, the difference is no longer significant. Cuadra *et al* [35] showed similar results. They found that when using a Gaussian mixture model for tissue segmentation, methods taking into account the partial volume effect outperformed methods that considered only pure tissues, whether or not spatial priors are used. Niessen *et al* [97] showed that consistently misplacing PV voxels can result in volume errors of up to 60% for $1mm^3$ resolution images. The problem would be even worse if lower resolution images were used.

A common simplifying assumption when dealing with the PVE is to assume that at most, two tissue classes can occur together in a single voxel. Sometimes it is also assumed that CSF and WM never occur together. This is often necessary either to make the problem tractable, or to increase the computational efficiency. This is a valid assumption, because all three tissue classes occur together in the same voxel so seldomly that the assumption has an insignificant effect [35].

Cuadra *et al* [35] discuss the two main methods of dealing with mixed tissues in the Gaussian mixture model:

Type 1 PV Segmentation Model is the more simplistic model, where separate

classes are created to represent mixed tissue classes, and these classes are also modelled by independent Gaussian distributions. This model is similar to the model above, except that both pure and mixed classes are now considered. In this case, $5 \leq K' \leq 7$, depending on the assumptions made concerning the possible combinations of tissues in a voxel.

Though this method outperforms methods only taking pure tissue classes into account, Cuadra *et al* [35] showed that this model is too simplistic and that PV distributions are not properly modelled by a Gaussian function.

The problem with this method is that is assumes that when 2 tissues mix, they always mix in the same proportion, which does not happen in reality.

Type 2 PV Segmentation Model is a more realistic model that was proposed by Santago *et al* [120, 121] and has been used extensively [98, 99, 126, 79, 140, 54]. In this model, pure tissue intensities are modelled by Gaussian distributions, while mixture tissues are represented as a linear combination of intensities associated with K possible pure tissue classes;

$$x_i = \sum_{k=1}^{K} \alpha_{i,k} G(\mu_k, \sigma_k), \tag{3.6}$$

$$\text{and } \sum_{k=1}^{K} \alpha_{i,k} = 1 \tag{3.7}$$

where $\alpha_{i,k}$ is the proportion of tissue type k in voxel i.

Equation (3.6) describes the partial volume averaging that is found in these PV voxels.

Assuming that a voxel can contain at most pure tissue classes, a voxel containing a mixed class k comprised of pure classes k_1 and k_2 has probability distribution function:

$$p(x|k,\alpha) = \frac{1}{\sigma_k(\alpha)\sqrt{2\Pi}} e^{\frac{-(x-\mu_k(\alpha))^2}{2\sigma_k^2(\alpha)}}, \tag{3.8}$$

where α is the fraction of k_1 in the mixture voxel.

The mean and variance in the above equation are determined by the model parameters of the pure tissues k_1 and k_2:

$$\mu_k(\alpha) = \alpha\mu_{k_1} + (1-\alpha)\mu_{k_2} \tag{3.9}$$

$$\sigma_k^2(\alpha) = \alpha^2\sigma_{k_1}^2 + (1-\alpha)^2\sigma_{k_2}^2 \tag{3.10}$$

The probability distribution for the whole mixed tissue k is:

$$p(x|k) = \int_0^1 p(x|k,\alpha)p(\alpha|k)d\alpha \tag{3.11}$$

The choice for the function $p(\alpha|k)$ is important, as discussed by Cuadra *et al* [35]. We apply the common assumption of a uniform distribution for α [35], ie. $p(\alpha|x_i) = 1$, for $0 \leq \alpha \leq 1$. With this assumption, the previous equation becomes:

$$p(x|k) = \int_0^1 p(x|k,\alpha)d\alpha \tag{3.12}$$

This distribution is, in general, not Gaussian and its shape depends on the parameters $\theta_k = \{\mu_k, \sigma_k\}$, which is discussed in more detail by Cuadra *et al* [35]. There is no known closed form for this integral, and so this equation needs to be numerically calculated.

There are fewer parameters to estimate in this model than in the Type 1 PV segmentation model, as μ_k and σ_k are only estimated for pure tissue classes, while w_k is still estimated for both pure and mixed classes. An advantage of this method is that once these parameters are found, the $\alpha_{i,k}$ can be calculated and used to find a soft classification, giving the proportions of each tissue type present in each voxel [98].

Cuadra *et al* [35] showed that though in general this model outperforms the Type 1 PV Tissue Segmentation model, the percentage of voxels correctly classified as mixture tissue is still poor - far worse than the percent of voxels correctly classified as a pure tissue. This shows that the PV distribution is still not being properly modelled in Type 2 PV Tissue Segmentation. The problem with this model is that it is assumed that all α occur with equal likelihood, which is not realistic [139].

Using simulation, Ruan *et al* [117] showed that in practical situations using T1-weighted MR images, equation (3.12) can be approximated with a Gaussian distribution.

Type 3 PV segmentation, used by Van Leemput *et al* [140, 139], takes a very different approach. The original MR image $\tilde{X}$, containing the partial volume effect, is considered a downsampling of another higher resolution image X, which contains no partial volume effect. Each of the voxels of the original image can be divided into M subvoxels, each of them corresponding to a voxel in the higher resolution image. Each of

these subvoxels contains only pure tissue. Let J_i be the set of M subvoxel indices j that are subdivided from voxel i.

The underlying label image Y of the higher-resolution intensity image indicates the non-mixed tissue type to which each subvoxel site j belongs:

$$Y = \{y_j : 1 \leq j \leq M.N, 1 \leq y_j \leq K\} \tag{3.13}$$

.

The high-resolution, non-mixed intensity image $X = \{x_j : 1 \leq j \leq M.N\}$ is considered to be generated from Y by drawing a sample from the probability distribution $f(X|Y, \theta_X)$. They assume that the intensity of each subvoxel j, is conditionally independent from the intensity of the other subvoxels, given its tissue label y_j and that it follows a normal distribution:

$$f(X|Y, \theta_X) = \prod_j f(x_j|y_j, \theta_X) = \prod_j G(x_j, \theta_{y_j}) \tag{3.14}$$

The observed image $\tilde{X} = \{\tilde{x}_i : 1 \leq i \leq N\}$ is obtained by downsampling X. The observed intensity $\tilde{x}_i$ in voxel i is the sum of the intensities x_j of the subvoxels underlying it: $\tilde{x}_i = \sum_{j \in J_i} x_j$.

This method is advantageous, in that it automatically provides subvoxel level segmentation and does not assume that all mixing proportions are equally probable as is done in Type 2 PV segmentation

This discrete PV segmentation method would be accurate if the number of subvoxels were infinite, which in practice is not attainable [47]. Eremina *et al* [47] extend this method to overcome this disadvantage, by using a continuous set of fractions rather than discrete labels. They found that this resulted in a more accurate estimation of fractional composition of the voxels.

In order to deal with the partial volume effect, Desco *et al* [44] use logistic regression to calculate the posterior probabilities.

Assuming that partial volume voxels have been identified, González Ballester *et al* [54] attempt to estimate the fractional tissue content within each voxel.

Manjón *et al* [89] deal with partial volume voxels in a very different way, treating them as outliers for the pure tissue distributions. This is advantageous as the partial volume voxels do not have to be modelled. Ideally, all partial volume voxels will have been removed and all remaining voxels will consist of only pure tissues, which can be modelled with Gaussian distributions.

Underlying Model

The typical Gaussian Mixture Model used in brain tissue segmentation includes just three tissue classes: cerebrospinal fluid (CSF), grey matter (GM) and white matter (WM).

Initially, more classes were used. Van Leemput *et al* [137] included 6 classes in their model, 3 for the brain tissue classes, 2 for non-brain classes and one class for the background signal.

Rajakapse *et al* [111] claim to be the first to use just the three brain tissue classes. They were able to do this because they extracted the brain from the head scan prior to tissue class segmentation.

Schroeter *et al* [122] add an extra class to the Gaussian mixture model in order to make the estimation scheme *more robust to the presence of noise*. The idea behind this is that most of the observations that are not able to be captured by the other Gaussian classes, will be able to be modelled by this class of outliers. They assume that this class of outliers has a uniform distribution. They found that when an image is badly corrupted by noise, the addition of this class results in improved accuracy.

The *Constrained Gaussian Mixture Model* developed by Greenspan *et al* [56] uses multiple Gaussian components to represent each tissue type, in order to capture the complex spatial layout of the brain. Each voxel has an intensity and 3 spatial features (the X, Y and Z co-ordinates). The means and standard deviations of these features for each Gaussian component are still estimated by the **EM** algorithm, but the intensity parameters are constrained such that all Gaussian components belonging to the same tissue type have the same intensity features. This method has the advantage of combining global intensity modelling with localised spatial modelling and results in a segmentation that is smoother and less granulated and is resistant to noise.

Parameter Initialisation

The results obtained from the **EM** algorithm are known to be sensitive to the initial parameter estimates [114, 49] because of their local nature [35]. This makes the initial choice of parameters important for the final results.

Wells *et al* [146] use an interactive method to initialise the parameters.

Van Leemput *et al* [138, 137] use a digital brain atlas to initialise the model parameters. The atlas they use contains spatially varying prior probability maps for the location of each tissue class. The atlas is registered with the data to be segmented and provides a prior estimation of the tissue class label of each voxel. From this, initial values for the

parameters are estimated. The advantage of this method over the one above, is that it removes the need for user intervention, making the segmentation fully automated and the results reproducible and objective. However, the registration process may result in poor initialisation in abnormal brains. It also relies on such an atlas being available, which might limit the algorithm's practical use.

Cuadra *et al* [35] found that the means are the parameters most sensitive to initialisation. They initialised the means using a prior k-means classification. They initialised the weightings w_k as $\frac{1}{K}$ and the standard deviations as 5, a small number not equal to zero.

Based on experimentation, Noe and Gee [98] found that to make the algorithm robust to initialisation, one needs to specify a sufficiently large class variance. They chose initial mean intensity values equally spaced between the minimum and maximum intensity values found in the image:

$$\mu_k = \frac{k}{K+1}(\text{maximum intensity} - \text{minimum intensity}) + \text{minimum intensity} \quad (3.15)$$

They set the variances of the tissue classes to the range of the image intensity:

$$\sigma_k = \text{maximum intensity} - \text{minimum intensity} \quad (3.16)$$

3.2 Other Methods

Because the Gaussian mixture model is by far the most used method for brain tissue segmentation, much of this chapter has looked at the variations in that model. However, some methods do take totally different approaches and a few of these will be briefly mentioned in this section.

An alternative to the Gaussian mixture model is to use a nonparametric method, where the analytic expression of the probability density function of each tissue class is unknown [5, 30]. The argument for the use of these methods, rather than assuming that the PDFs are Gaussian distributed, is that a departure of the real data from the Gaussian assumption can greatly decrease the accuracy of the segmentation. Cuadra *et al* [35] discusses some nonparametric methods. In some cases these methods perform equally or even better than methods based on a Gaussian mixture model.

The k-means algorithm [87] has been modified to be used for brain tissue class segmentation. Ahmed *et al* [3] use a modified fuzzy c-means algorithm for bias field estimation

and the segmentation of MR data. They modify the objective function of the standard fuzzy c-means algorithm to account for the spatially varying bias field inhomogeneities, and to enable the segmentation of a voxel to be affected by the voxels in its neighbourhood. Incorporating the neighbourhood information is useful for dealing with the noise found in MR images. Pham and Prince [105] extend a 3D adaptive fuzzy c-means algorithm to model bias field inhomogeneities.

Grau *et al* [55] use a watershed transform with the inclusion of prior information. They also combine the watershed transform with atlas registration, through the use of markers.

3.3 Cerebellum Segmentation

There is very little previous work that focuses on the tissue segmentation of the cerebellum. The only article we found that specifically focused on the cerebellum was one by Datta *et al* [39]. The segmentation method they used was the EM algorithm combined with an MRF prior (HMRF algorithm). They compared the segmentation results when the cerebellum was segmented in isolation, to the results obtained when the cerebellum was segmented with the rest of the brain. They found that when segmenting real images, better results were obtained when the cerebellum was segmented in isolation. The reason they give for the differences in results is that the intensities of the tissue types of the cerebellum are not representative of the intensities of the same tissue types in the cerebrum.

When the same experiment was performed on Brainweb images[1], they found no difference in the results. The reason for this was that Brainweb assumes that the tissue intensities in the cerebellum are the same as the tissue intensities of the rest of the brain.

In their cerebellum segmentation algorithm, Datta *et al* [39] did not address the problem of the partial volume effect and so we introduced a new model for cerebellum tissue segmentation. Since the Gaussian mixture model using an **EM** algorithm for parameter estimation has been widely and successfully used for brain tissue segmentation, we based our segmentation method on this model.

[1]See chapter 5 for a brief introduction to Brainweb

Chapter 4

Framework for Cerebellar Tissue Segmentation

In this chapter we explain the framework that we have developed and motivate the reasons for making various choices about the system. Though our aim is to segment the cerebellum, our system can segment the full brain into three tissue classes.

We explain how we *bootstrap* our system with information from the whole brain, in order to provide a different segmentation method, with the intention that this method will provide better segmentation results.

As mentioned previously in section 2.6 under the section on Bias Field Inhomogeneities, our system ignores bias field inhomogeneities.

4.1 Gaussian Mixture Model and EM Algorithm

Clustering based segmentation algorithms are well suited to situations where the means of the different classes are distant in feature space. This is the case in MR brain images, where the means of the intensities of the different tissue classes are well separated [67]. Looking through the literature, the clustering algorithm most commonly used for brain tissue segmentation is the **EM** algorithm. Since it is widely used with much success for brain tissue segmentation, we use the **EM** algorithm in our segmentation.

When considering the model behind the **EM** algorithm, we wanted to initially create a simple system, and so ignored complications such as the partial volume effect and bias field inhomogeneities, and did not include any spatial information. We assumed that the intensities of the tissue classes are Gaussian distributed. Even though the distributions are in fact Rician [22], the common practice is to assume that they are Gaussian distributed

[35, 122, 111]. This assumption simplifies calculations, and it has been shown that for the signal to noise ratios found in T1-weighted MR images, it is a reasonable assumption. The fact that we are considering a finite mixture model further supports the use of the **EM** algorithm, since the **EM** algorithm is the standard way of finding the parameters for finite mixture models [142].

We did not constrain the model by forcing the standard deviations of different tissue classes to be the same. Though we have more parameters to estimate, it is more accurate not to constrain them. Zhang *et al* [153] argued that in reality the standard deviations of some tissue types are greater than others, so assuming that all standard deviations are equal is not realistic. Even if the standard deviations are equal, our unconstrained model will still be able to capture this occurrence.

Since the cerebellum occupies a relatively small portion of the image, we expect that it will be less affected by bias field inhomogeneities than if the whole brain was being considered. Because of this, we did not deal with any of the complications arising from bias field inhomogeneities. If the inhomogeneities were in fact a problem, we could have applied a filter to the image beforehand in an attempt to remove their effect, but this was not necessary.

After the initialisation of the parameters (the mean ($\mu_k^{(0)}$), standard deviation ($\sigma_k^{(0)}$) and weighting ($w_k^{(0)}$) for each tissue class), we iterate between the **E** and **M**-steps until convergence is reached, as shown in Figure 4.1.

During the t'th **E**-step, the posterior probability for each voxel of belonging to each tissue class is calculated, using the parameters from the t'th iteration:

$$p^{(t)}(k|x_i) = \frac{p(x_i|k)w_k^{(t)}}{\sum_{k'=1}^{K} p(x_i|k')w_{k'}^{(t)}} \tag{4.1}$$

and

$$p(x_i|k) = G(x_i, \mu_k, \sigma_k) = \frac{1}{\sqrt{2\Pi}\sigma_k} e^{-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}} \tag{4.2}$$

where: $G(x_i, \mu_k, \sigma_k)$ is the k'th Gaussian distribution with mean μ_k and standard deviation σ_k.

During the $(t+1)$'th **M**-step, each voxel is assigned to the tissue class to which it has the highest posterior probability of belonging. Using this new classification, the $(t+1)$'th mean, standard deviation and prior probability of each tissue class is calculated, using the t'th posterior probabilities, according to the equations.

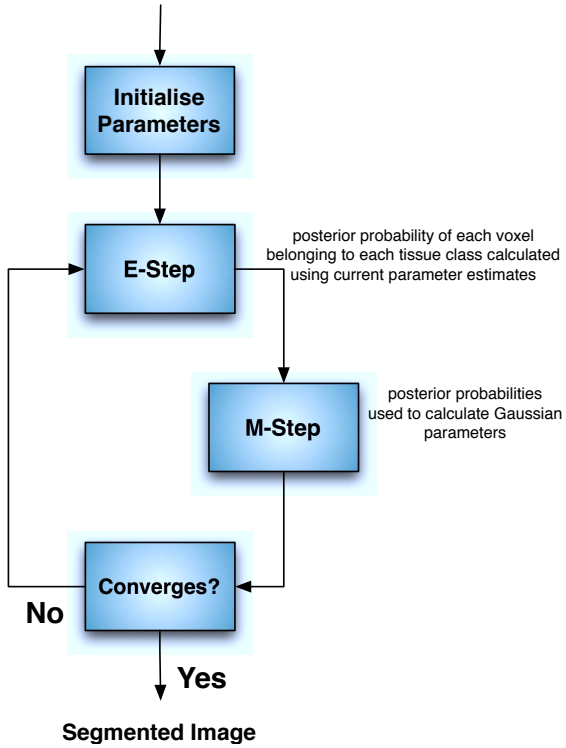


Figure 4.1: A summary of the **EM** algorithm.

$$\mu_k^{(t+1)} = \frac{\sum_{i=1}^{N} p^{(t)}(k|x_i)x_i}{\sum_{i=1}^{N} p^{(t)}(k|x_i)}, \tag{4.3}$$

$$\sigma_k^{(t+1)} = \sqrt{\frac{\sum_{i=1}^{N} p^{(t)}(k|x_i)(x_i - \mu_k^{(t+1)})^2}{\sum_{i=1}^{N} p^{(t)}(k|x_i)}}, \text{ and} \tag{4.4}$$

$$w_k^{(t+1)} = \frac{\sum_{i=1}^{N} p^{(t)}(k|x_i)}{N}. \tag{4.5}$$

Convergence is reached when all the means and all the standard deviations change by less than ϵ between iterations. Once convergence is reached, each voxel is assigned to the tissue class to which it has the highest probability of belonging.

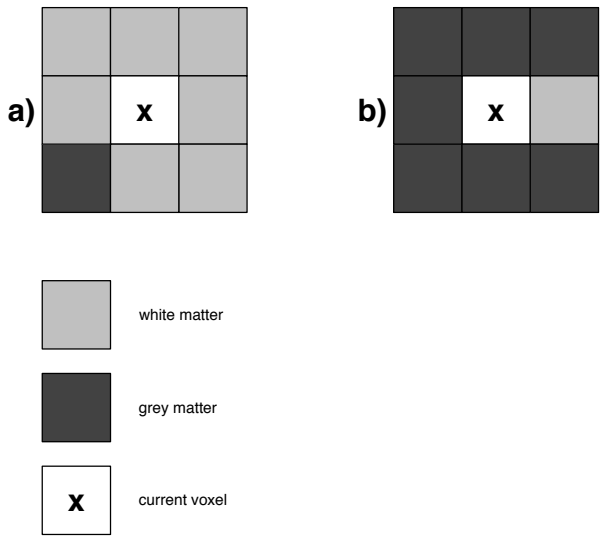


Figure 4.2: The idea behind the inclusion of spatial information into the segmentation process. (a)Current voxel more likely to be white matter. (b)Current voxel more likely to be grey matter.

4.2 Including Spatial Information: MRF

One of the major drawbacks of clustering algorithms, such as the standard **EM** algorithm, is that they do not include any spatial information. As a result of this, there could be regions that are not spatially compact, which is one of the major desirable qualities of a good segmentation. For example, in an MR image, one voxel of white matter could appear much darker than the other white matter voxels around it, due to noise. If ignoring spatial information, this voxel could quite likely be assigned to grey matter, which would be incorrect. However, taking spatial information into account would increase the probability of the voxel being classified correctly as white matter. This principle is illustrated in Figure 4.2.

Including an MRF prior into the model is the standard way of introducing spatial information into the **EM** algorithm [118]. Cuadra *et al* [35] considered various different methods of brain tissue segmentation, including looking at the effect of adding an MRF prior. In all cases, the MRF improved the segmentation results. Because of this, we decided to use an MRF to incorporate spatial information into our model.

In brain MR images, even though the means of different tissue types are well separated [67], there tends to be much overlap in their intensities [114] due to noise, bias field inhomogeneities and the PVE [30]. Because of this, it is important to use spatial modelling in order to achieve good segmentation results.

Going back to the **E**-step of the **EM** algorithm, the calculation of the posterior prob-

abilities involves an MRF. Recall that when calculating energy ($U(y)$) in the MRF the following formula is used:

$$U(y) = \sum_{i=1}^{N} \sum_{j \in \mathcal{N}_i} V_{ij}(y_i, y_j) \tag{4.6}$$

The energy of a voxel is calculated by adding the results of the MRF function ($\delta(y_i, y_j)$) between the voxel and each of its neighbours.

When calculating the energy between 2 neighbours, we use the δ function:

$$\delta(y_i, y_j) = \begin{cases} -2, & \text{if } y_i = y_j \\ +1, & \text{otherwise} \end{cases} \tag{4.7}$$

This function is calculated for each neighbour in a second order neighbourhood, so diagonals are included as neighbours. Since we are assuming that all voxels contain only pure tissue, we do not have to take PV classes into account as done by others [98]. We ignore the distance function used in the calculation of $V_{ij}(y_i, y_j)$ in equation 3.4; so all neighbours are weighted equally.

Once we have performed this segmentation using the **EM** algorithm with MRF, most of the pure voxels should have been correctly classified, though not necessarily the mixed ones. Because of this, we want to try and remove all the partial volume voxels and treat them as outliers.

4.3 Dealing with the PVE: Outliers

When performing brain tissue segmentation, it is very *important to consider the partial volume effect*. Niessen *et al* [97] showed that ignoring the partial volume effect can cause volume measurement errors in the range of $20 - 60\%$. The consideration of this effect is even more important when dealing with the cerebellum, as, due to the presence of many fine structures, it will have an even larger proportion of partial volume voxels. This will result in even higher error rates if the PVE is ignored.

The partial volume effect is one of the major artifacts that can cause the Gaussian parameters to be incorrectly estimated [122]. One way of dealing with the PVE is including the PV tissue classes as part of the model. However, Cuadra *et al*[35] demonstrate that the PVE has not been perfectly modelled, even though methods that do account for mixture classes outperform methods that only consider pure Gaussian classes.

Cocosco *et al* [30] argue that the Gaussian model assumption can be poor, due to the intrinsic biological heterogeneity within the tissue classes, as well as bias field inhomogeneities and the partial volume effect. Over the small compact volume of the cerebellum, biological heterogeneity and bias field inhomogeneities should not be an issue, but the PVE cannot be ignored. In order to keep to the assumption of the Gaussian model, voxels containing mixed tissues need to be removed.

Instead of trying to model them, we aim to *remove partial volume voxels as outliers* of the pure tissue distributions. This is done because the PV voxels corrupt the pure tissues' Gaussian distributions. Once the PV voxels are removed, we should be left with an unpolluted mixture of Gaussians and be able to more accurately calculate the parameters of the model. The more accurate parameter estimates should result in more accurate segmentation of the voxels. It is preferable to classify too many voxels as outliers rather than to classify too few voxels as outliers; we want all the partial volume voxels classed as outliers. It is not important if some outliers are not actually PV voxels. If we remove pure voxels as outliers, this should have little negative impact on the parameter estimation, provided there are enough remaining to be able to accurately estimate parameters. However, if too many PV voxels remain, this could cause inaccurate parameter estimation.

Our approach of removing outliers from the estimation process, is supported by Schroeter *et al*'s [122] statement, that data not belonging to any of the Gaussian components can strongly bias the estimation of the parameters. Though they did not include PV voxels as outliers in their outlier rejection scheme, they found that the removal of outliers from the parameter estimation process improved the parameter estimates.

The method we use to find outliers is similar to that used by Manjón *et al* [89], though we have modified it. The outlier detection process needs to be initialised with a fairly accurate hard segmentation of the image. Manjón *et al* used an incremental k-means method, because it is not as sensitive to parameter initialisation as the **EM** algorithm or the original k-means algorithm. In order to provide a more accurate initial segmentation, we used the mixture of Gaussians model with the **EM** algorithm and MRF prior to perform parameter initialisation.

The first step in outlier detection is *context trimming*, which involves classifying all boundary voxels as outliers, as demonstrated in Figure 4.3. Each voxel with at least one neighbour belonging to a class different from its own is classified as an outlier. The motivation for this is the fact that most partial volume voxels are found on the boundaries between different tissue classes.

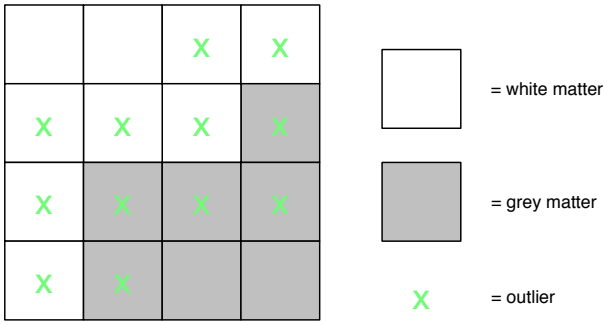


Figure 4.3: Boundary voxels become outliers

The next step, *gradient trimming*, is motivated by the fact that voxels with a high local gradient are associated with a rapid change in intensity between neighbouring voxels. These voxels are likely to be partial volume voxels, or voxels that are particularly badly affected by noise. In order to calculate the gradient information, we ran 2D X- and Y-Sobel filters over each slice to get the derivatives along each axis in the plane. These derivatives were combined to create a magnitude image. The voxels were ranked according to their values in the magnitude image.

We chose a fraction λ, $0 \leq \lambda \leq 1$ and classify the λN (N is the total number of voxels) voxels with the highest gradient as outliers. This differed slightly from Manjón *et al*, as they looked at each tissue class separately and set the λN_c (N_c = number of voxels in tissue class c) voxels in each tissue class as outliers. We found that $\lambda = 0.1$ worked well, which is very different to Manjón *et al* who used $\lambda = 0.95$.

Parameter Estimation with Outliers Excluded Once we have selected the outliers, they are excluded from the next step, using a mask. Most of the PV voxels should have been removed, and predominantly pure voxels should remain. Since the **EM** algorithm with MRF prior segments pure voxels well, all non-outlier voxels are assumed to be correctly segmented. Using the tissue class labels assigned to the voxels during the **EM** algorithm, we calculate the mean and standard deviation of each tissue class. These calculations should not be distorted by the presence of outliers.

Final Classification of Voxels Next, the outliers have to be classified. They can be either assigned to a single tissue class, or be given an estimated fraction of each tissue class they contain.

In order to create a hard segmentation, we perform the **E**-step of the **EM** algorithm, calculating the posterior probabilities, using the new (more accurate) parameters just

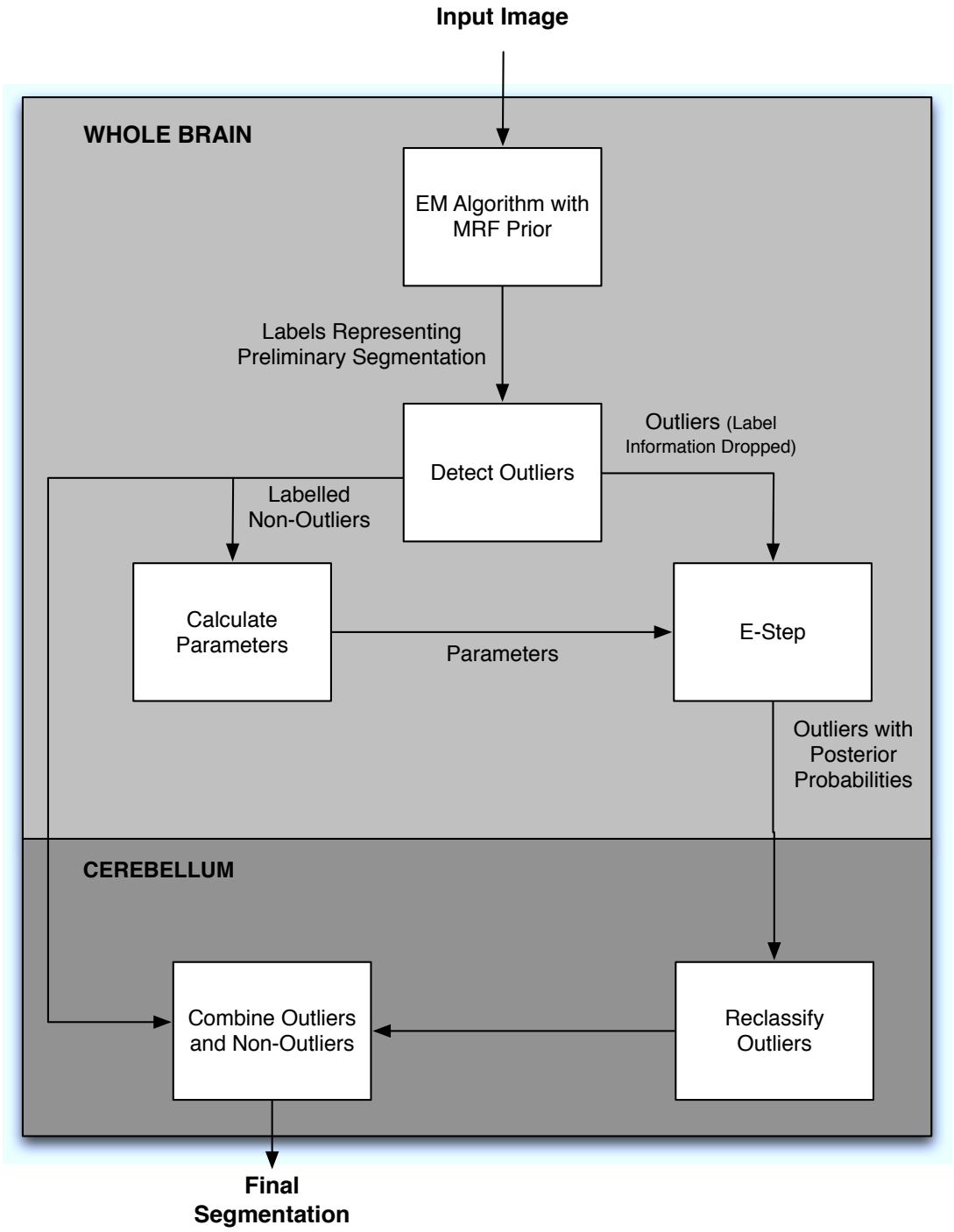


Figure 4.4: Our bootstrapped framework for cerebellum tissue segmentation.

calculated. This is done *without* using an MRF prior. Most of these outliers are at the boundaries between tissue types, therefore we do not want to use an MRF to encourage them to be similar to their neighbours. Effectively, the MRF is being weighted heavily for the non-outliers - mainly interior voxels - and is given a zero weighting at the edges. Once the parameters have been calculated, the voxels can then be assigned to tissue classes, in the same way as previously.

4.4 Overview of the Basic Segmentation System

To summarise, our model for brain tissue segmentation treats partial volume voxels as outliers and assumes pure tissues have Gaussian distributions. First, we ignore partial volume effects, run the algorithm on the whole brain and find the model parameters using the **EM** algorithm combined with an MRF prior. Using these parameters, we create an accurate hard segmentation of the image. Using this segmentation, we choose outliers, which are voxels most likely to be partial volume voxels, based on the method used by Manjón *et al* [89]. Once the outliers are removed, we estimate the parameters from the remaining voxels. Using these parameters, we can then reclassify the outliers to a specific tissue class.

4.5 Bootstrapping our System with Tissue Information from the Cerebrum

Because of its numerous fine structures, many of the cerebellum's voxels will be partial volume voxels, which are more likely to be misclassified. Because of this, we can use the whole brain to estimate the parameters of the tissue classes and then use this information to classify the voxels of the cerebellum. The bootstrapping process is shown in Figure 4.4.

Chapter 5

Analysis and Results

In this chapter, we look at the performance of three different methods: the simple "**EM** algorithm", the "**EM** algorithm with an MRF prior" and our outlier method. In an attempt to improve the segmentation of the cerebellum, we run the algorithms on the whole brain first and then use that information to segment the cerebellum.

We perform our testing on simulated data, provided by Brainweb[1] and real data provided by the Internet Brain Segmentation Repository. We use the simulated data to demonstrate the complexity of cerebellum segmentation, and to motivate various choices we made in our algorithm design. We look at the effect of the MRF parameters on the segmentation performance and see how our algorithm performs in comparison to the other algorithms. Using the real data, we tested more rigorously how our algorithm performs in comparison to the other algorithms.

5.1 Evaluation of Segmentation Results

Numerically evaluating a volumetric segmentation usually involves comparing it to a ground truth image. Many different metrics have been used for this evaluation, as discussed by Warfield *et al* [144]. A simple approach involves comparing the volume of the object that the algorithm has segmented with the volume of the object in the ground truth image. This is not a very accurate measure, as two objects with different shape and location can have the same volume [68, 143].

Measures of overlap, such as the *Dice Similarity Measure* (DSM)[45] and the *Tanimoto Coefficient*, (also known as the *Jaccard Similarity Coefficient*) [69] are commonly used for evaluation of brain tissue segmentation.

[1]Brainweb is discussed in more detail is section 5.2.1

The Tanimoto Coefficient for class k is defined as the ratio of the number of voxels that both the ground truth (g_k) and the image being segmented (s_k) classify as class k, to the total number of voxels classified as class k by either the ground truth or the segmented image:

$$T_k = \frac{|s_k \cap g_k|}{|s_k \cup g_k|} \tag{5.1}$$

where $|s_k \cap g_k|$ is the number of voxels assigned to class k by both the ground truth and the segmented image and $|s_k \cup g_k|$ is the number of voxels assigned to class k by either the ground truth or the segmented image.

The Tanimoto Coefficient is a global performance metric and has been criticized for having a lack of local sensitivity to error [71]. In spite of this fact, it has become the standard metric used for MRI segmentation analysis.

The Dice Similarity Measure is a similar measure to the Tanimoto Coefficient and is also widely used. The Dice metric for class k is the ratio of twice the number of voxels assigned to class k by both the ground truth and the segmented image, to the sum of number of voxels assigned to class k by the ground truth and the number of voxels assigned to k by the segmented image:

$$D_k = \frac{2|s_k \cap g_k|}{|s_k| + |g_k|} \tag{5.2}$$

where $|s_k|$ and $|g_k|$ are the number of voxels assigned to class k by the segmented image and the ground truth respectively.

The Dice metric has been criticized for depending on the size and shape complexity of the object being segmented. Since most errors occur on the boundaries, small objects are penalised and have a lower value than larger objects [56].

Since they are similar measures, the Tamimoto and Dice metrics tend to follow the same trends. However the Tanimoto metric for class k is always less than or equal to the Dice metric for class k.

A problem with all these measures is that they give separate metrics for each tissue class. If we simply want a measure of performance, we can use the *Pergood* metric[36], which is the fraction of the voxels correctly classified:

$$\text{Pergood} = \frac{\text{Number Of Voxels Correctly Classified}}{\text{Total Number Of Voxels Segmented}} \tag{5.3}$$

The *class mean error metric* (CME) is a metric often used with the **EM** algorithm to measure how well the algorithm estimates the model parameters. It is the average

difference between the estimated mean and the true mean.

$$\text{CME} = \frac{\sum_{i=1}^{K} |\mu_{i(est)} - \mu_{i(true)}|}{K} \tag{5.4}$$

When calculating this measure, one needs to know the ideal means. When using Brainweb data, the ideal means can be calculated from the ideal hard segmentation. However, some of the voxels included in a particular class will be partial volume voxels. This can skew the calculation, especially when focusing on the cerebellum which has a high proportion of partial volume voxels.

When we calculated the ideal means, we wanted to include only voxels which were pure tissue voxels. In the Brainweb data we used, none of the voxels contained a single tissue class only, so we considered a voxel a pure tissue voxel if it contained 99% or more of one tissue type. This same procedure was followed by Tohka [133].

5.2 Testing on Brainweb Images

In this section, we use simulated Brainweb [29, 33, 77, 78, 1] images to demonstrate the complexity of cerebellum segmentation, to motivate choices we made in designing our algorithm and to investigate the effect of the MRF parameters. We then also used the images to compare the performance of our algorithm with the "**EM** algorithm" and the "**EM** algorithm with MRF prior".

There are two different categories of Brainweb images. For the first class, we have one phantom image, with simulated images at various noise levels. For the second class we have 20 other brains, each with just a single simulated image. Because we wanted to test at different noise levels, we used to first class to demonstrate the complexity of cerebellum segmentation, to investigate the effect of MRF parameters on performance and to motivate the choices we made in our algorithm design. We would have liked to use the second class, Brainweb's 20 new "normal brains" [8, 9] for performance comparison, as just one sample at each noise level is not enough data to validate our algorithm, but the phantom images contained 2^3 times more voxels than the simulated MR image.

5.2.1 Input

The input to our system is the collection of voxels belonging to either the whole brain or the cerebellum, depending on the problem requirements. We only consider voxels within this region, and mask out all other voxels and ignore them.

For each Brainweb simulated image, we obtained the whole brain outline from the corresponding Brainweb ground truth segmentation. We included all voxels assigned as grey matter, white matter or CSF as part of the brain. Unfortunately, Brainweb does not provide a segmentation outline for the cerebellum, so we used Freesurfer to generate this segmentation.

Testing on Simulated Data The quantitative assessment of the segmentation of real data is very difficult, since a ground truth segmentation is usually either not available or very time consuming to create manually. For this reason, simulated data is usually used when testing segmentation algorithms.

We used images from Brainweb, which provides realistic simulated brain MR image volumes. The Brainweb MRI simulator starts from a digital phantom and performs accurate modelling of the imaging process based on Bloch equations. This makes the data well-suited to validation, since an ideal segmentation is known. The simulator allows various parameters to be changed, such as imaging modality, slice thickness, noise and intensity non-uniformity levels, creating different MR image volume simulations. Because the first set of simulations are all based on the same digital phantom, this enables testing of the sensitivity of a particular algorithm to the change of an imaging parameter. Brainweb also provides ideal fuzzy tissue membership, which is essential for the testing of PVC estimation and impossible to obtain from real data.

Unfortunately, simulated data cannot perfectly capture all the complexity of real data. Since the aim of segmentation algorithms is to process real data, it is not enough to validate the model on only simulated data. Therefore tests on real data are also necessary and will be addressed later.

Brainweb Data as an Approximation to Real Data Cuadra [34] found that a typical noisy MR image corresponds to the mean of Brainweb images with 5% and 7% noise.

The noise of real MRI data has been found to be Rician [35] and the noise in Brainweb simulated images has Rician statistics. Brainweb defines the "percent noise" to be the percent ratio of the white Gaussian noise to the signal for a reference tissue.

5.2.2 The Complexity of Cerebellum Segmentation

This thesis was motivated by the need for a method to handle the fine structures of the cerebellum. Here we demonstrate the difficulty of cerebellum segmentation by running

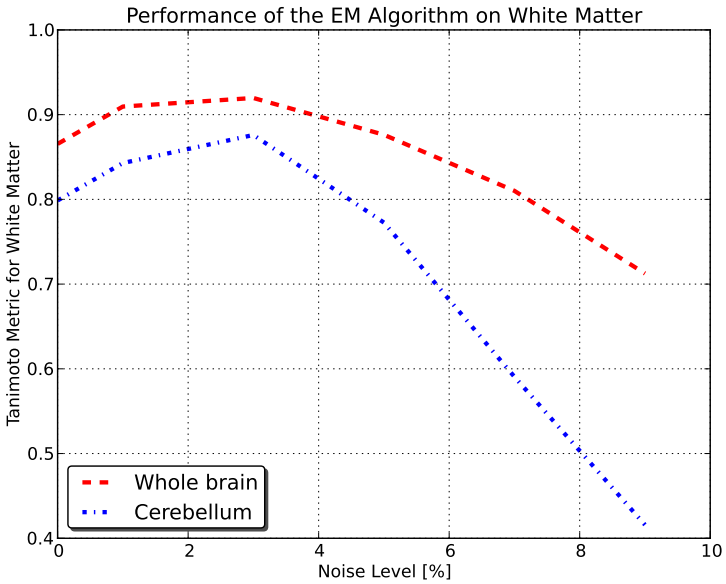


Figure 5.1: Performance of the "EM algorithm" on the white matter on the whole brain and on the cerebellum. The "EM algorithm" segments the whole brain more successfully than it segments the cerebellum, showing the complexity of cerebellum segmentation.

the "**EM** algorithm" on the whole brain (including the cerebellum) and running the "**EM** algorithm" on just the cerebellum.

As can be seen in Figure 5.1, when segmenting white matter, the "**EM** algorithm" performs better on the whole brain than it does on the cerebellum at all noise levels, but especially at higher noise levels. This occurs because the fine white matter structures in the cerebellum cause many of the voxels to be PV voxels, which are more likely to be misclassified. Even though the cerebellum is also included in the whole brain, there are relatively few cerebellum WM voxels, and so a few extra voxels classified incorrectly would have a larger effect on the cerebellum metric than it would have on the whole brain metric. Furthermore, fewer voxels means that a single voxel classified incorrectly during the **E**-step of the "**EM** algorithm" would have a larger effect on the parameter estimation in the **M**-step.

While the performance of the "**EM** algorithm" when segmenting both the whole brain and the cerebellum decreases at noise levels above 3%, it decreases much more quickly for the cerebellum and the graphs rapidly diverge. This occurs for similar reasons: as noise increases, additional PV voxels misclassified in the cerebellum have a larger effect on the cerebellum metric. The difference between segmentation quality of the whole brain and of the cerebellum is even more pronounced when an "**EM**-algorithm with an MRF prior"

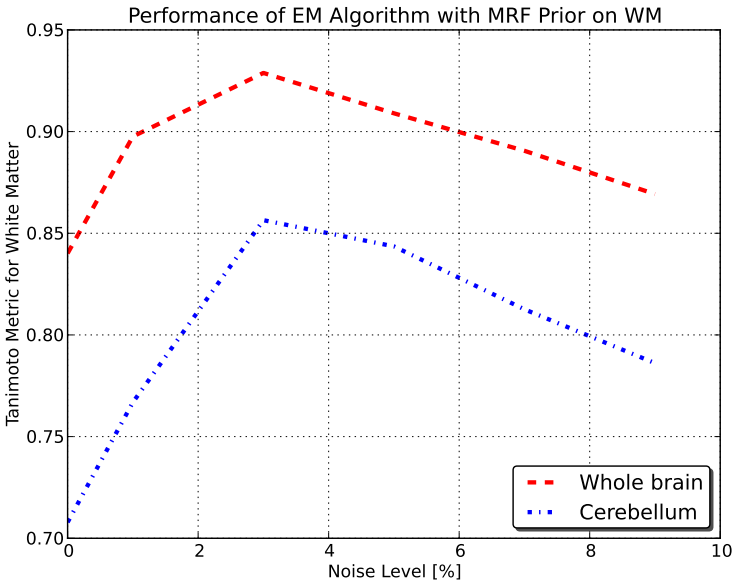


Figure 5.2: Performance of the "EM algorithm with MRF prior" on the white matter on the whole brain and on the cerebellum. The "EM algorithm with MRF prior" segments the whole brain more successfully than it segments the cerebellum, showing the complexity of cerebellum segmentation.

is used, as is common practice [118, 131]. This is shown is Figure 5.2.

It is interesting to note, however, that these graphs do not diverge and are nearly equidistant the whole way. The fine structures of the cerebellum are poorly suited to the MRF prior, as it encourages clumping of similar tissue types, which will cause the fine structures of the cerebellum to be lost. It is for this reason that the performance difference between whole brain and the cerebellum is substantial. Because the MRF causes the fine structures to be lost even at a low noise level, increasing noise does not have much further effect on the fine structures and so the graphs do not diverge any further.

Next, we motivate various choices we made in our algorithm design. First we look at the decision to *bootstrap* our algorithm by obtaining parameters from the *whole brain* and then we investigate the decision to use an MRF with the **EM** algorithm in our initial segmentation.

5.2.3 Bootstrapping the Algorithm

Since the whole brain is segmented more accurately than the cerebellum, one would expect the parameter estimates obtained when segmenting the whole brain to be more accurate than the parameters obtained from cerebellum segmentation. Because of the

large number of partial volume voxels in the cerebellum, fitting the Gaussians to the data is more difficult, and one would expect the estimation of the parameters to be less accurate.

One could use the parameters obtained from the whole brain to classify the voxels of the cerebellum. Since the parameter estimation for the whole brain is more accurate, it should provide more accurate segmentation of the cerebellum. However, the mean intensity of each tissue type in the whole brain might not be representative of that tissue type in the cerebellum, as shown in Datta *et al* [39].

As mentioned previously, the cerebellum has a large proportion of partial volume voxels. Partial volume voxels make parameter estimation difficult, and so we decided to bootstrap the segmentation of the cerebellum with parameter estimates of the whole brain.

We obtained the parameter estimates by running the "**EM** algorithm with MRF prior" on the whole brain. We then used these parameters to classify the voxels of the cerebellum, by running the **E**-step of the "**EM** algorithm without MRF prior" on the cerebellum, using the parameter estimates obtained from the whole brain, providing the posterior probabilities of each tissue class for each voxel. Next we assigned each voxel to the tissue class with the highest posterior probability. It is important to note that we ran the **E**-step *without the MRF prior*, as we did not want the fine structures of the cerebellum to be lost. White matter fine structures are likely to be boundary voxels and therefore will be assigned as outliers. If using an MRF prior, the MRF encourages the white matter voxels of the fine structures to be assigned to grey matter like the tissues surrounding them, instead of considering only intensity information and classifying them as white matter. The results we obtained for grey matter using the "**EM** algorithm with MRF prior" are shown in Figure 5.3.

As can be seen, bootstrapping the algorithm with parameters from the whole brain improves classification of the cerebellum. The difference is most pronounced at low noise levels: this may be because at low noise levels the grey matter of the cerebellum is poorly approximated by a Gaussian. The **EM** algorithm tries to fit Gaussians to the data, but performs poorly because of the poor approximation of grey matter to a Gaussian. As the noise level increases, Rician noise is added, which approximates Gaussian noise under certain conditions, as mentioned previously. This means that the grey matter intensities become more Gaussian, and so they can be more easily classified at higher noise levels and segmentation performance improves. Similar results are obtained when using the "**EM** algorithm" without the MRF, see figure 5.4.

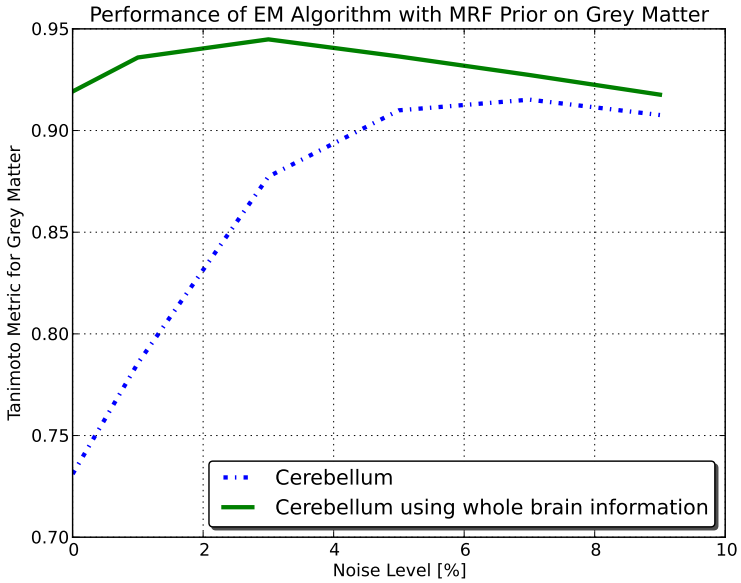


Figure 5.3: Comparison of cerebellum segmentation performance on grey matter of the "EM algorithm with MRF prior", with and without bootstrapping. Bootstrapping offers a substantial performance increase, especially at low noise levels.

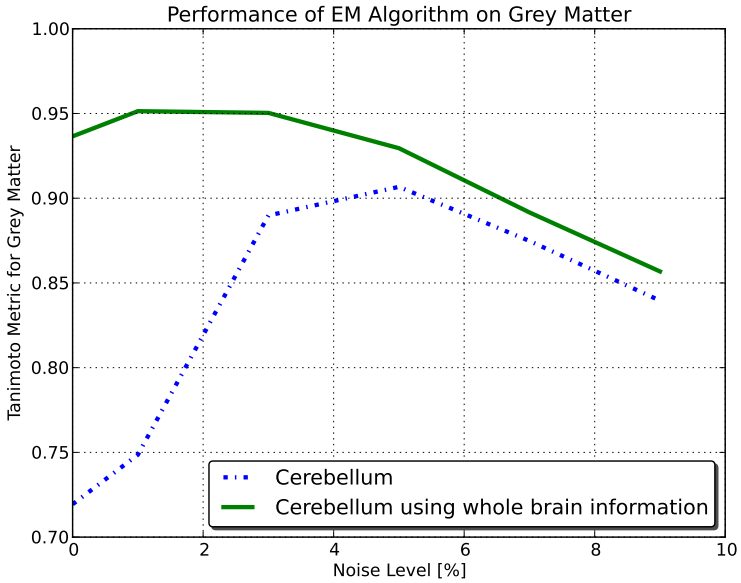


Figure 5.4: Comparison of cerebellum segmentation performance on grey matter of the "EM algorithm" without MRF prior, with and without bootstrapping. Bootstrapping offers a substantial performance increase, especially at low noise levels.

Bootstrapping provides a significant segmentation performance increase when using both the "**EM** algorithm" and the "**EM** algorithm with MRF prior". This justifies our decision to bootstrap our algorithm with parameters from the whole brain.

Next, we look at the decision of using an MRF prior with the **EM** algorithm for our initial segmentation.

5.2.4 Deciding When to Use the MRF

Here we investigate the role of the MRF prior in improving the performance of the **EM** algorithm when segmenting the whole brain. This will determine whether it is worthwhile using an MRF to calculate the whole brain parameters.

We use both the "**EM** algorithm" and the "**EM** algorithm with an MRF prior" to segment the whole brain and compare the results. As can be seen from Figure 5.5, for all noise levels of 3% and higher the "**EM** algorithm with an MRF prior" outperforms the "**EM** algorithm" without an MRF prior. Since the noise level of real data can be considered to lie between 5% and 7%, the "**EM** algorithm with an MRF prior" should outperform the "**EM** algorithm" without an MRF prior on real data. The same result was obtained by Cuadra *et al* [35]. This is an expected result, since at high noise levels it is more difficult to classify the data based on intensity alone. The MRF adds spatial information to the **EM** algorithm, and encourages clumps of similar types of tissue, producing better segmentation.

Because there is a clear advantage to using an MRF with the **EM** algorithm, we decided to follow this approach for our initial segmentation, before selecting outliers. However, the use of an MRF prior brings an added complication - the choice of the MRF parameters. In the next section we look at the effect that the β parameter has on the segmentation performance.

5.2.5 The Effect of Beta and Beta Tuning

The parameters chosen for the MRF make a large difference in the segmentation results we obtain. The parameters of the MRF are β and Z, where β determines how heavily the MRF is weighted, and Z is a normalising constant. We found that varying Z has little effect on the segmentation obtained, while varying β has a pronounced effect on the segmentation obtained. The larger β, the higher the weighting of the MRF and the smoother the segmentation results. Our aim is to find a value of β large enough to benefit from using the MRF, but small enough that fine detail is not lost.

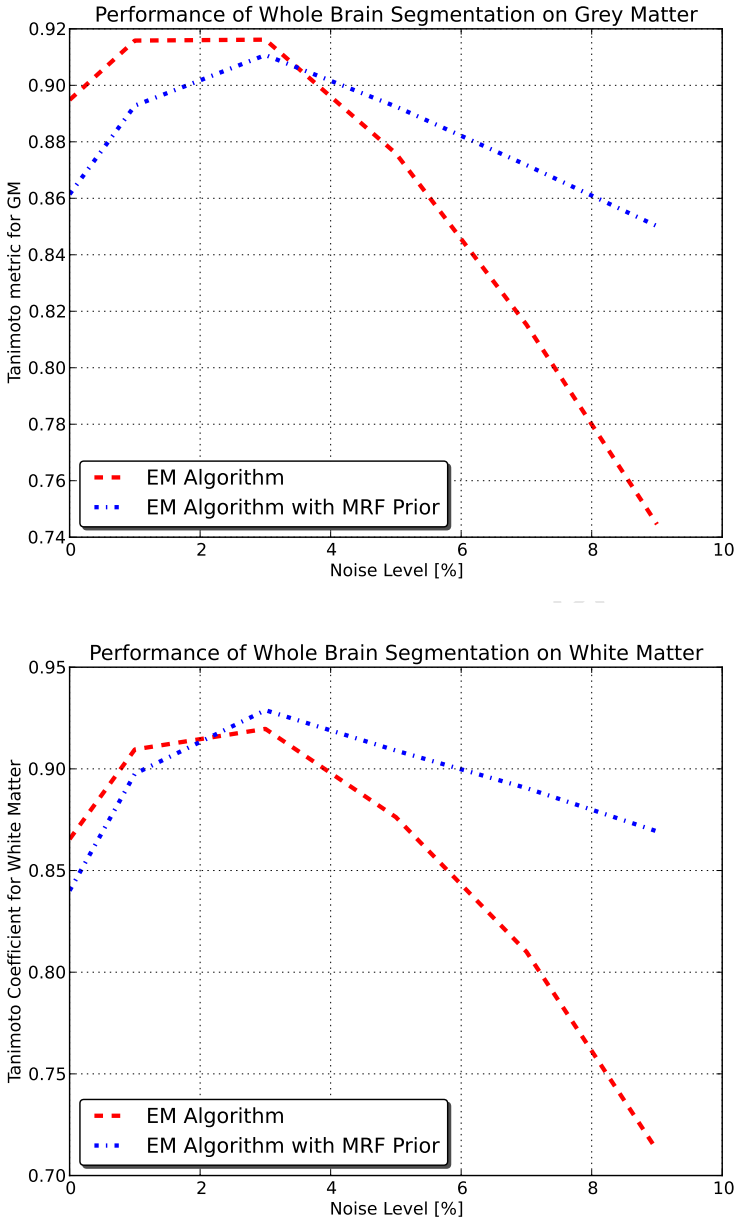
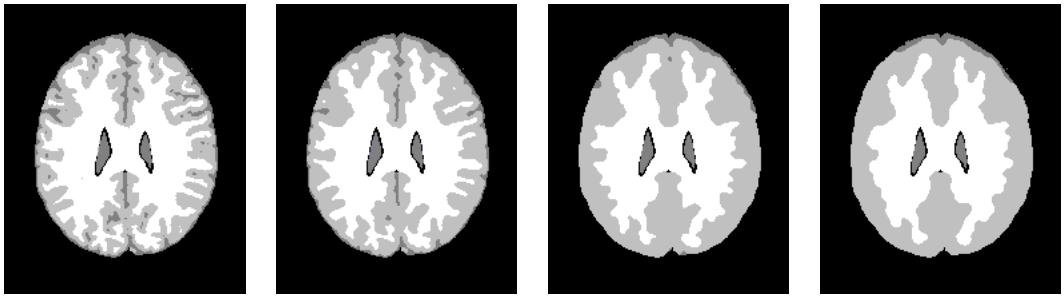


Figure 5.5: Comparison of the performance of the EM Algorithm with and without MRF Prior. The MRF prior provides an improvement in performance at higher noise levels.



(a) $\beta = 2$ (b) $\beta = 5$ (c) $\beta = 10$ (d) $\beta = 20$

Figure 5.6: The effect of varying β on the segmentation result. The larger β, the higher the weighting of the MRF and the smoother the segmentation results.

We segmented a single brain image[2] multiple times using the "**EM** algorithm with an MRF prior". For each segmentation, a different value of β was chosen. Corresponding slices from the segmentations are shown in Figure 5.6.

We segmented the cerebellum at different noise levels using the "**EM** algorithm with an MRF prior". At each noise level, we tried the segmentation with different values of β and calculated the Pergood measure to evaluate performance. The results are shown in Figure 5.7.

As can be seen, at different noise levels, different values of β work better - no value is clearly better than all others for all images. This demonstrates one of the major challenges of using an MRF - the need for parameter tuning.

There seems to be no clear relationship between noise levels and β, so other features of the image must influence the optimum β value as well as noise level.

Of the values of β that we tried, $\beta = 3$ worked best for low noise levels and $\beta = 1.5$ and $\beta = 2$ worked best for higher noise levels. The range of noise in real data lies between noise levels of 5% and 7% and in this region $\beta = 1.5$ and $\beta = 2$ performed the best. Because of this, we investigated values of β lying between these 2 values.

Figure 5.8 shows the performance of the algorithms with β in that range. As can be seen, the performance of all the algorithms is similar. We chose $\beta = 2$ for the rest of this investigation, as it performed the best over a wide range of noise levels, including part of the noise range found in real data. This does not guarentee an optimal β for other images, as noise level is not the only feature of MR images that changes between different images.

To further test the effect of β and noise level, we ran the "**EM** algorithm with MRF

[2]The whole brain was used instead of the cerebellum as it is easier to demonstrate the result of changing β on the whole brain.

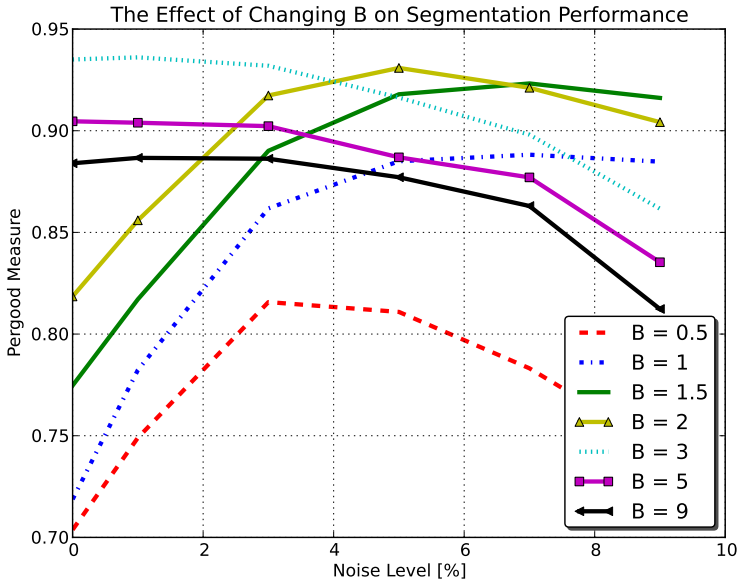


Figure 5.7: The effect of β (0.5-9) on segmentation performance. At different noise levels, different values of β work better - no value is clearly better than all others for all images.

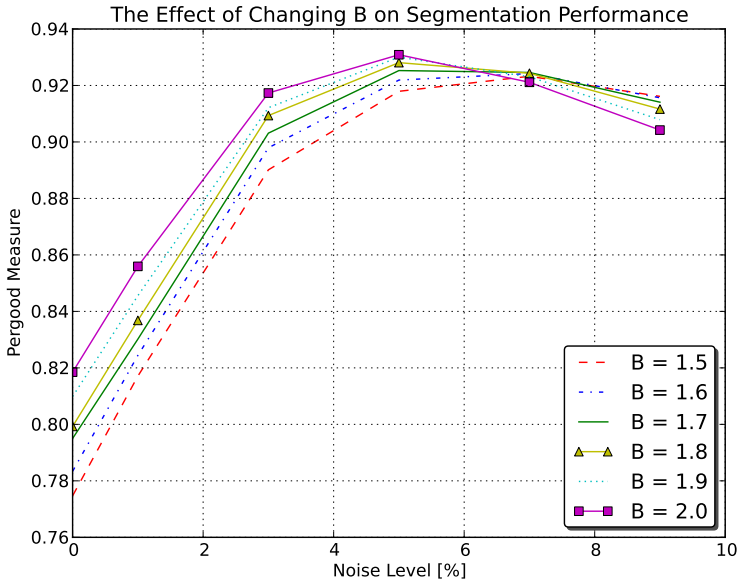


Figure 5.8: The effect of β (1.5-2) on segmentation performance. The performance of all the algorithms is similar.

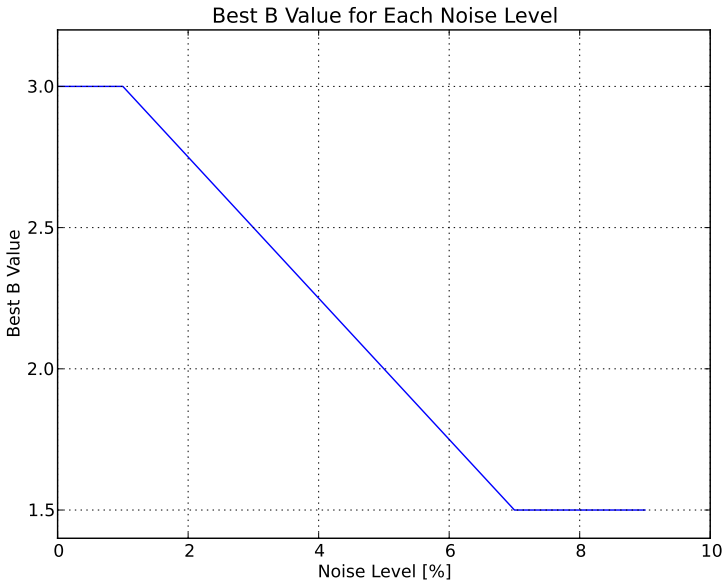


Figure 5.9: The effect of β on segmentation performance. We ran the "EM algorithm with MRF prior" once for each β value in the range $[0.5, 9]$ at 0.5 intervals. At each noise level we plotted which β value performed the best.

prior" once for each β value in the range $[0.5, 9]$ at 0.5 intervals. At each noise level we recorded which β value performed the best and graphed our results, which are shown in Figure 5.9.

We see that at noise levels 0 and 1, the optimal β value is 3. As the noise level increases, the optimal β linearly decreases, until it stabilises at $\beta = 1.5$. This means that as the noise level increases, the MRF should be weighted less, until β reaches 1.5 and then it should not be weighted any lower.

5.2.6 Algorithm Performance Comparison

We compared the performance of our algorithm with various other methods: the simple "**EM** algorithm" without an MRF prior, the "**EM** algorithm with an MRF prior" (used to segment the cerebellum in a recent paper [39]) and our algorithm with and without bootstrapping. The principle measures of performance that we used were the Tanimoto Coefficient and Pergood measure, as discussed in section 5.1.

We wanted to compare algorithms with parameter initialisations where they perform well. Since we have an ideal segmentation, we can estimate the ideal parameters, and we initialised the means and standard deviations fairly close to their ideals. To initialise the prior probabilities, we assumed that each tissue type occurs with equal frequency.

To create an initial segmentation, which is needed for the methods using an MRF prior, we used this information to calculate the posterior probabilities and from there, assigned each voxel to the tissue class with the highest posterior probability.

Alternatively, we could have randomly initialised the parameters and run the algorithm a few times and then selected the best results, to try and overcome the effect that parameter initialisation has on the segmentation results.

Figure 5.10 shows the performance of the algorithms on grey and white matter. Our methods outperform the other methods at low noise levels, but perform worse than the "**EM** algorithm with MRF prior" at higher noise levels. For grey matter, our algorithms perform best up to a noise level of about 5%, while for white matter our methods perform the best only up to a noise level of 3%. Little improvement is obtained by bootstrapping our outlier algorithm with information from the whole brain, and sometimes this may even lead to a performance reduction. It must also be noted that the computation time of our algorithm without bootstrapping is much less than the bootstrapped version, making the version without bootstrapping seem the preferred choice.

Instead of having a separate measure of performance for each tissue class, we can look at the Pergood measure, (see Figure 5.11) which combines the results of all tissue classes into one measure.

The trends here are very similar to the graph of the Tanimoto Coefficient for grey matter. This is expected, since most of the voxels are grey matter, and so it will have a larger effect on the percentage of correctly classified voxels.

If we look at the class mean error graph, Figure 5.12, we see a very different result. Both versions of our algorithm estimate the means better than any of the other methods. One of the main reasons for using outliers is to obtain better parameter estimates, and our algorithm achieved this. It seems to have managed to remove most of the PV voxels as outliers before calculating the parameters, resulting in more accurate parameter estimates. This suggests that the main weakness of our algorithm is the assignment of voxels to tissue classes, given the parameter estimates. Based on these results, our algorithm performed very well in terms of parameter estimation.

Our method without bootstrapping generally performs better than the version with bootstrapping. An explanation could be because the means of the tissue classes of the whole brain are not representative of that of the cerebellum, but this is known not to be the case for Brainweb data [39].

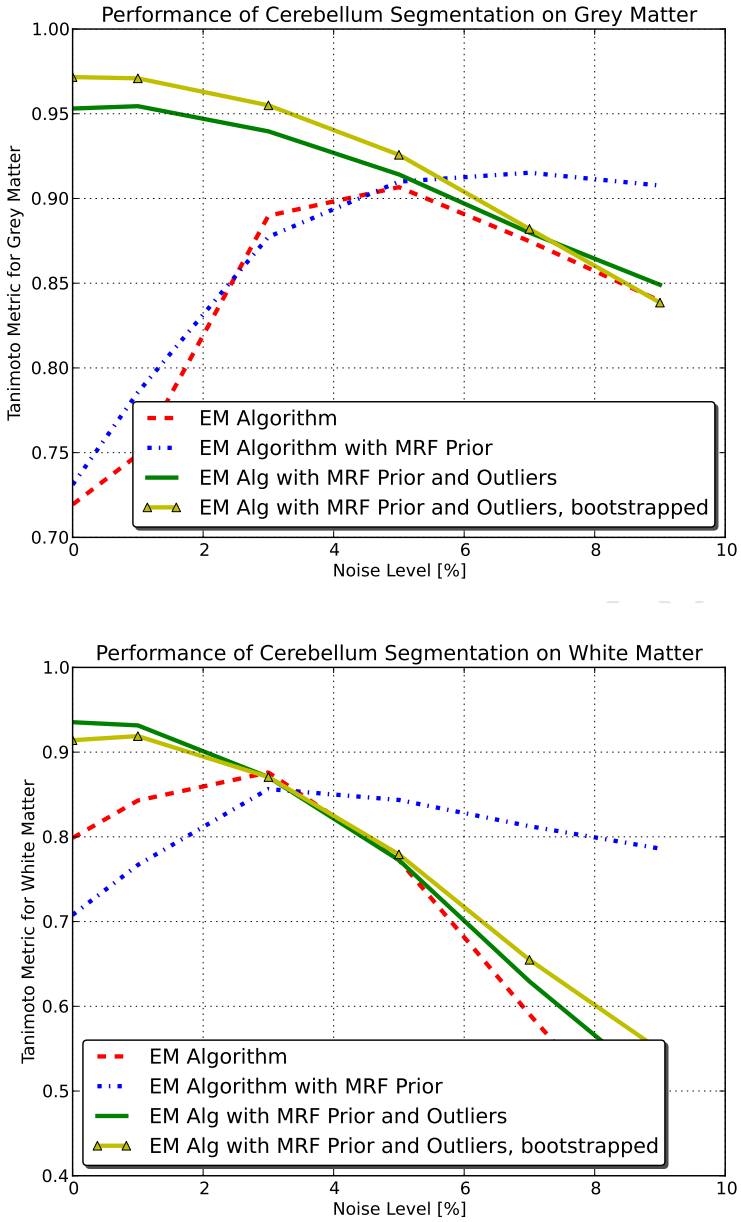


Figure 5.10: Comparison of performance of the different algorithms segmenting the cerebellum from simulated data rated with the Tanimoto Coefficient. Our methods outperform the other methods at low noise levels, but perform worse than the "**EM** algorithm with MRF prior" at higher noise levels. Little improvement is obtained by bootstrapping our outlier algorithm with information from the whole brain, and sometimes this may even lead to a performance reduction.

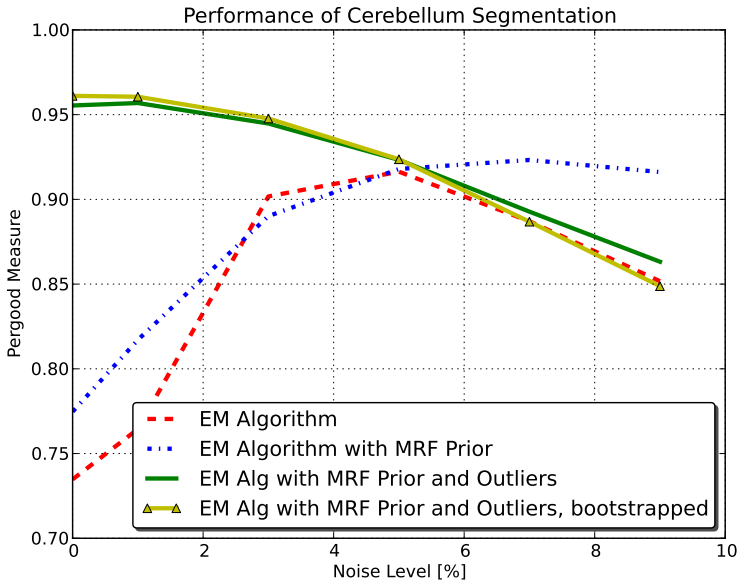


Figure 5.11: Comparison of performance of the different algorithms segmenting the cerebellum from simulated data rated with the Pergood Measure. The trends are very similar to the graph of the Tanimoto coefficient for grey matter.

5.2.7 Summary

We demonstrated the complexity of the cerebellum segmentation by showing how both the "**EM** algorithm" without MRF prior and the "**EM** algorithm with MRF prior" performed better on the whole brain white matter than they did on the cerebellum white matter. The difference was particularly large when using the MRF, which shows how the naive use of an MRF prior is poorly suited to the cerebellum with its fine white matter structures.

We found that bootstrapping the cerebellum segmentation with parameters from the whole brain caused a significant increase in performance when using both the "**EM** algorithm" and the "**EM** algorithm with MRF prior". This supports our decision to bootstrap our algorithm. We found that the "**EM** algorithm with MRF prior" outperformed the "**EM** algorithm" for all noise levels of 3% and higher, which includes the noise range found in real data, supporting our decision to use the "**EM** algorithm with MRF prior" for our initial segmentation.

We then demonstrated how larger MRF β parameter values lead to the smoother segmentation results. β values that are too high result in loss of fine detail in the segmentation. Conversely, if the β value is too low, the benefit of the MRF prior is not realised. We found that no single value of β was optimal at all noise levels, which demon-

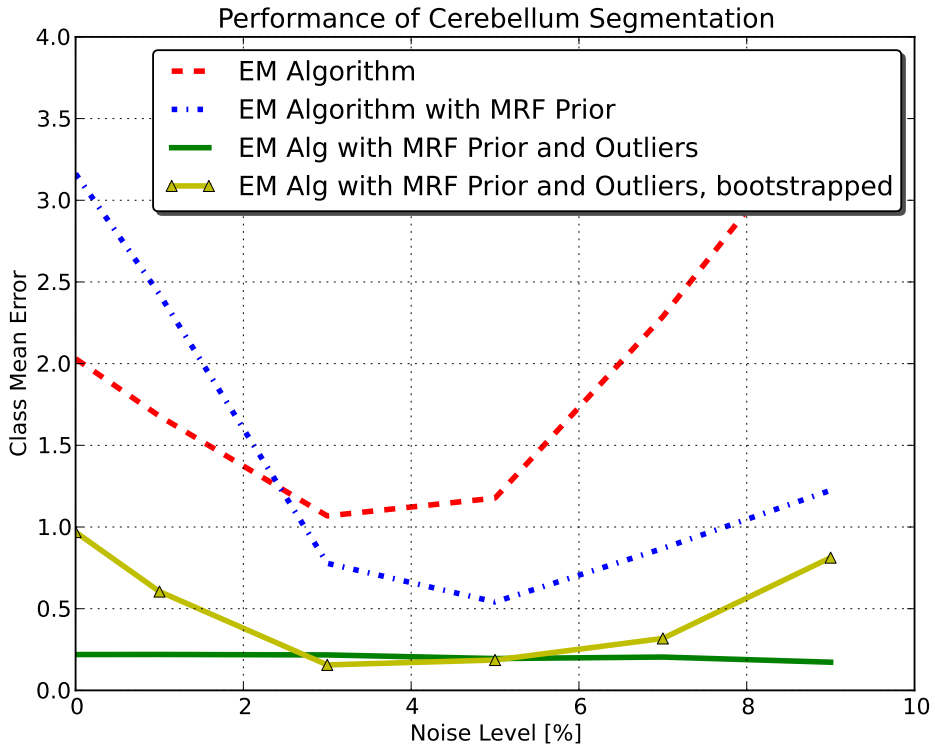


Figure 5.12: Performance on simulated data: success of the different algorithms at finding the correct means. Both versions of our algorithm estimate the means better than any of the other methods. One of the main reasons for using outliers is to obtain better parameter estimates, and our algorithm achieved this.

strates one of the major challenges of using an MRF: choosing the optimum value of β.

We found no clear relationship between noise levels and β, so other features of the image other than noise level must influence the optimum β. We chose to use $\beta = 2$ for our experiments, as it performed best over a wide range of noise levels, including part of the range of noise levels found in real images. However, it is not guaranteed to be the optimal β for all images.

When looking at the Tanimoto measure or Pergood measure, our algorithm outperforms the other algorithms we compared it with at low noise level, but performs worse than the "**EM** algorithm with MRF prior" at high noise levels. Little improvement is obtained by bootstrapping our algorithm with information from the whole brain, and sometimes it may even lead to a performance reduction.

Finally, we examined the class mean errors of the various algorithms. Our method with bootstrapping far outperformed all other methods. Both our methods actually improved performance as the noise level increased. Since the main reason for removing the outliers is to obtain better parameter estimates, our algorithm is successful in this regard.

In the next section, we test our algorithms on real data.

5.3 Testing on Real Data: Internet Brain Segmentation Repository

As mentioned previously, it is not sufficient to test a segmentation method on simulated data alone. In this section, we test our algorithm on real data. We obtained this data from Massachusetts General Hospital's Internet Brain Segmentation Repository (IBSR) [66]. These 18 normal brain data sets are available along with manually-guided segmentation results which can serve as a ground truth. These segmentation results were created using semi-automated segmentation techniques, which took a trained expert many hours to complete. Although these results cannot be considered to be completely accurate, they are still a good way to compare segmentation methods.

We ran the "**EM** algorithm", the "**EM** algorithm with MRF", our outlier method and our bootstrapped outlier method on each of the 18 normal subjects' MR brain images provided by the IBSR. It is important to note that we initialised the parameters the same way for all methods and all subjects, and did not try to optimise the parameter

initialisation. Better results could have been obtained by all methods by tuning the initial parameters. We initialised the means to 50 and 100, and all the standard deviations to 20.

We initially visually inspected the segmentation output for the cerebellum to get an overview of how the algorithms performed and then quantitatively evaluated the segmentation results using the various metrics.

5.3.1 Visual Inspection

We roughly gauged the segmentation quality by looking at the segmentation output on various slices of the cerebellum. Figure 5.13 shows various slices of the brain of subject 18.

Our outlier method managed to pick up more of the fine structures of the cerebellum than did the "**EM** algorithm with MRF prior".

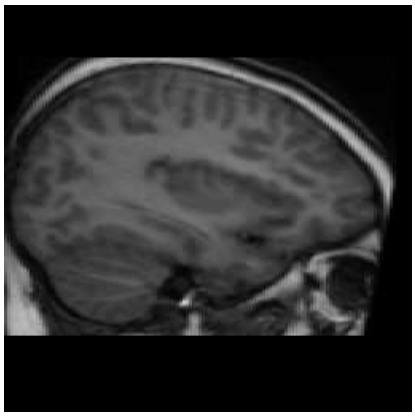
5.3.2 Tanimoto Metrics

We first looked at the Tanimoto metric of each of the tissue classes separately. The results are shown in Figures 5.14 and 5.15.

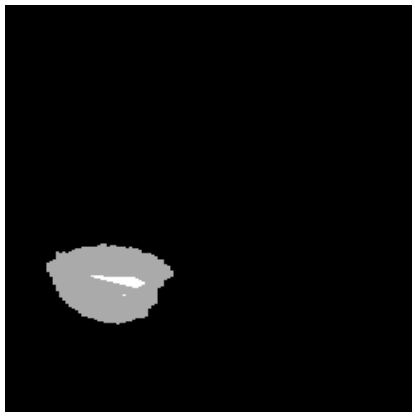
For grey matter, our outlier algorithm without bootstrapping performed the best most of the time. The only exception was subject 4, where our outlier algorithm with boostrapping performed slightly better. For most of the samples, our method with bootstrapping performed slightly worse than our method without bootstrapping. One notable exception, however, was subject 11, where our method with bootstrapping performed much worse than any of the other algorithms. A reason for this could be that for that individual the mean intensity of the grey matter in the cerebellum is very different to the intensity of the mean of the grey matter in the whole brain.

The "**EM** algorithm" and the "**EM** algorithm with MRF prior" performed quite erratically. Sometimes, their performance was comparable to our algorithms, but sometimes they performed much worse. We surmise that this is because they are more sensitive to parameter initialisation than our methods are.

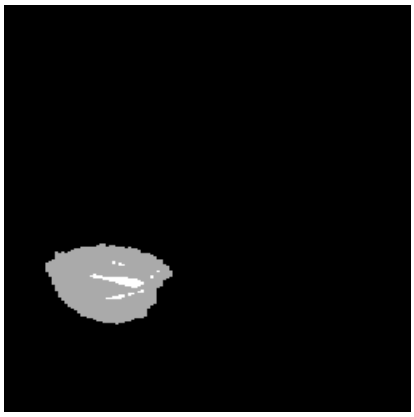
For white matter, once again our outlier method without bootstrapping tended to perform the best in general, however, this time there were a few more exceptions. It was occasionally outperformed by our method with bootstrapping and twice by the "**EM** algorithm with MRF prior" and once by the "**EM** algorithm". It performed particularly poorly on subject 11, where its Tanimoto metric was 0. However, for this subject, the



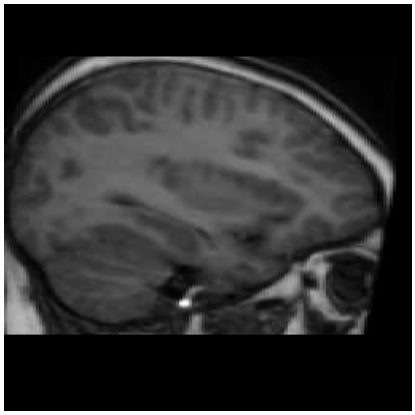
(a) Original Image



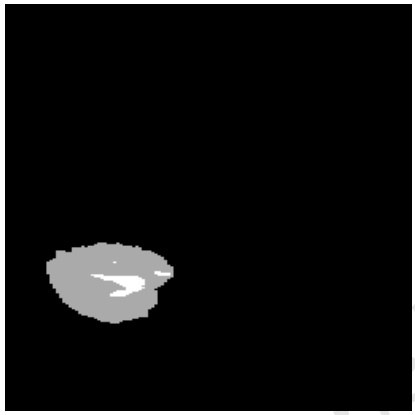
(b) MRF segmentation



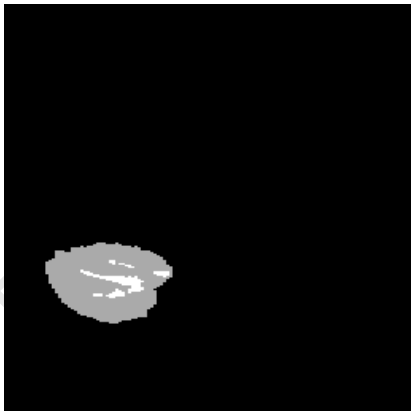
(c) Outlier segmentation



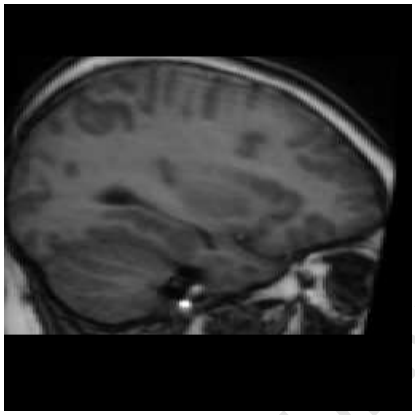
(d) Original Image



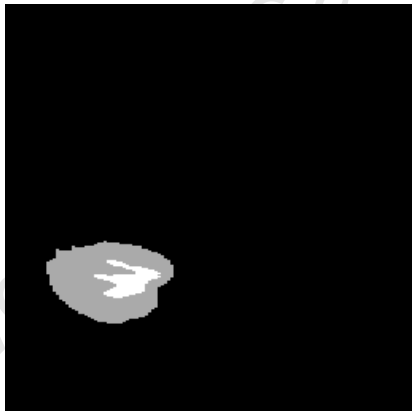
(e) MRF segmentation



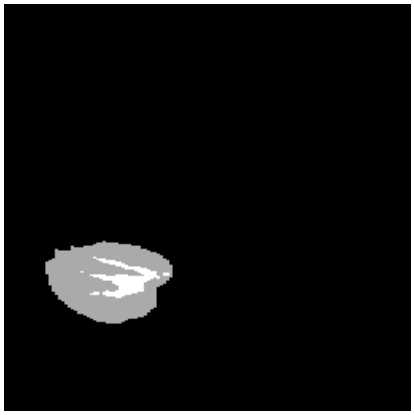
(f) Outlier segmentation



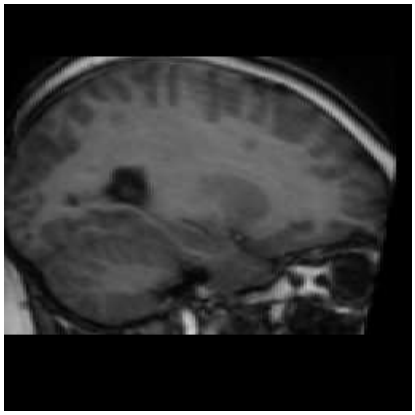
(g) Original Image



(h) MRF segmentation

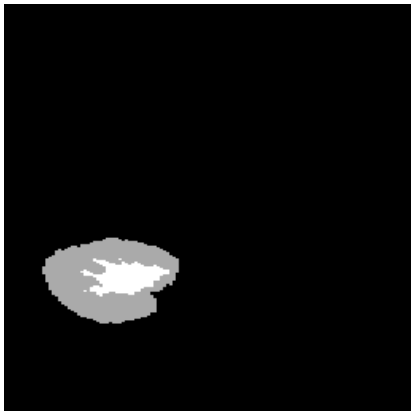


(i) Outlier segmentation

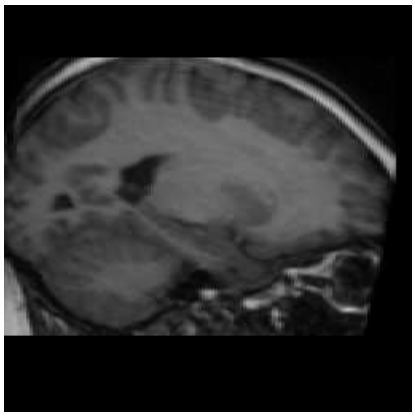


(j) Original Image

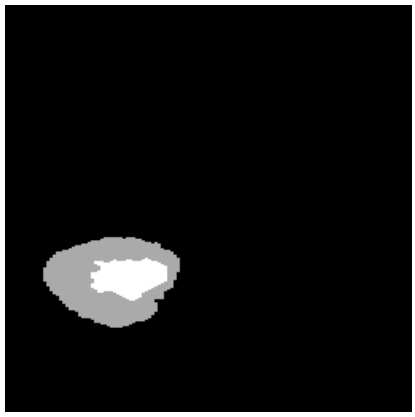
(k) MRF segmentation



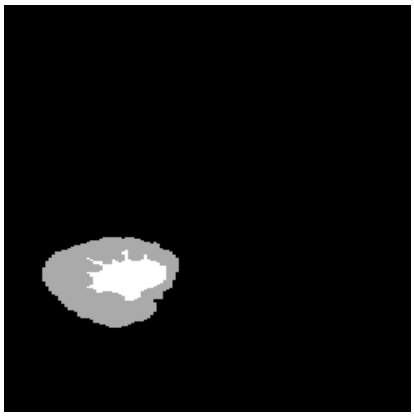
(l) Outlier segmentation



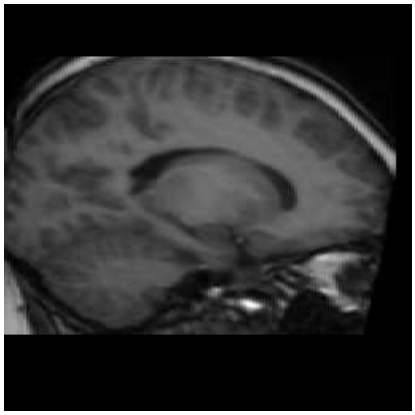
(a) Original Image



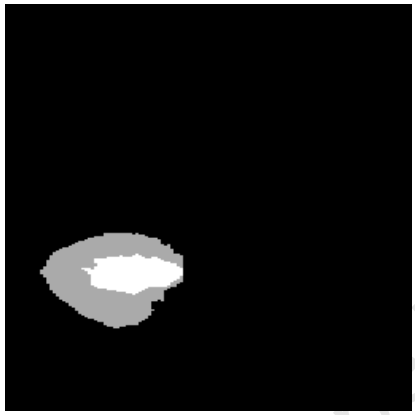
(b) MRF segmentation



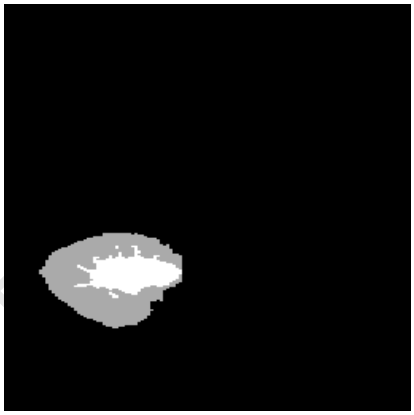
(c) Outlier segmentation



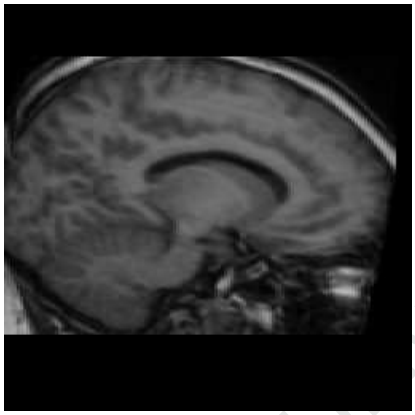
(d) Original Image



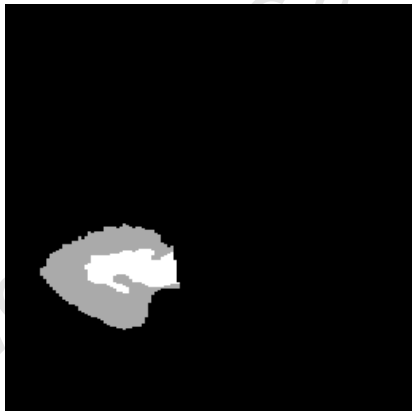
(e) MRF segmentation



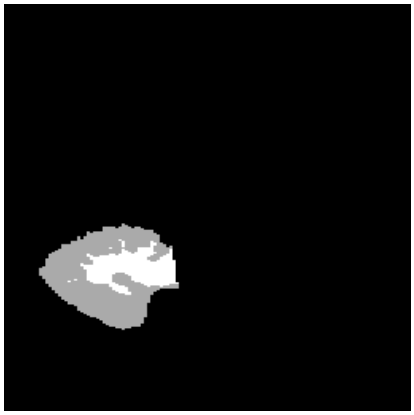
(f) Outlier segmentation



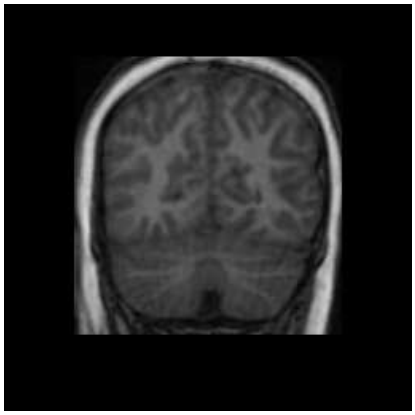
(g) Original Image



(h) MRF segmentation



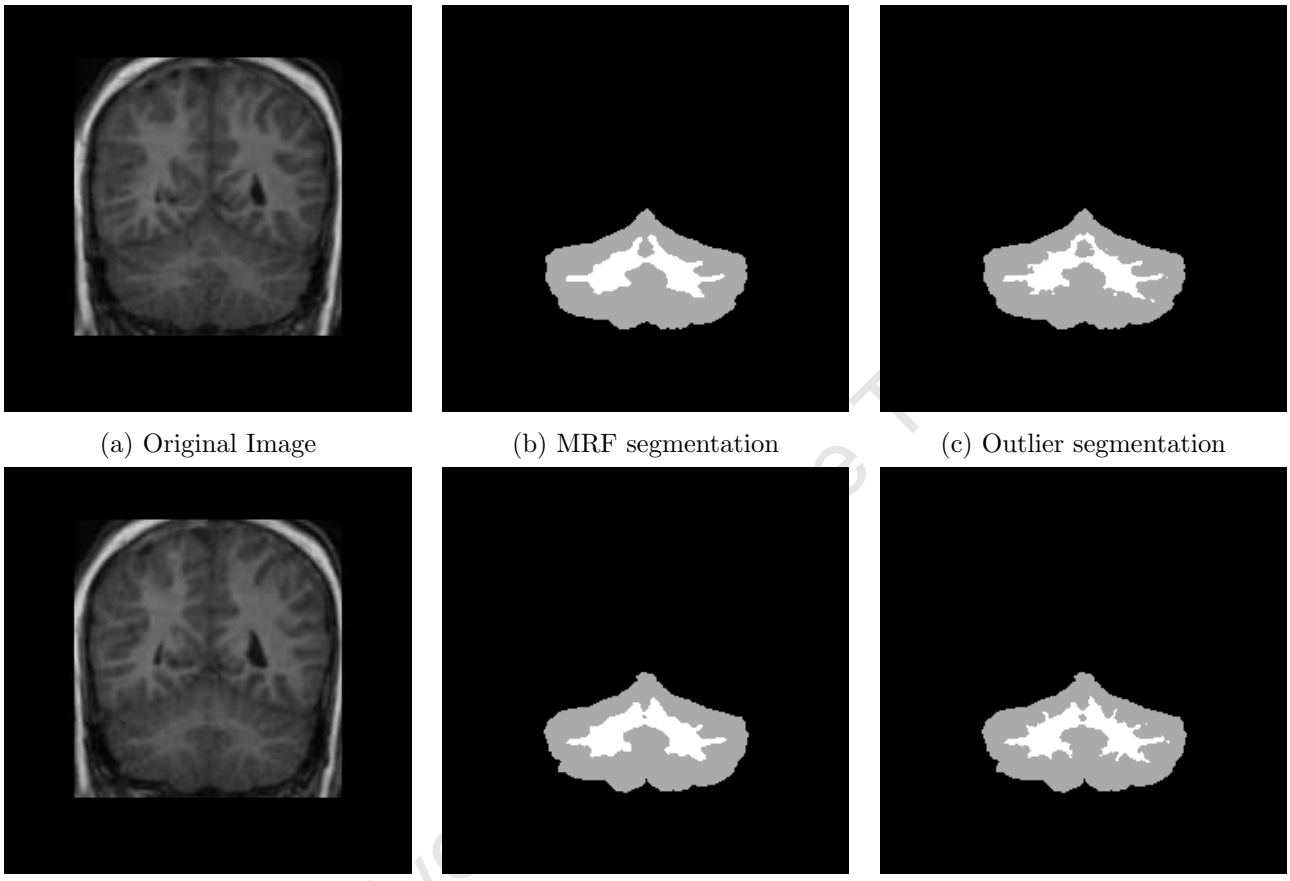
(i) Outlier segmentation



(j) Original Image

(k) MRF segmentation

(l) Outlier segmentation



(a) Original Image (b) MRF segmentation (c) Outlier segmentation

(d) Original Image (e) MRF segmentation (f) Outlier segmentation

Figure 5.13: Comparison between the segmentation of different slices of the brain of subject 18 by the "EM algorithm with MRF" and with our outlier method/

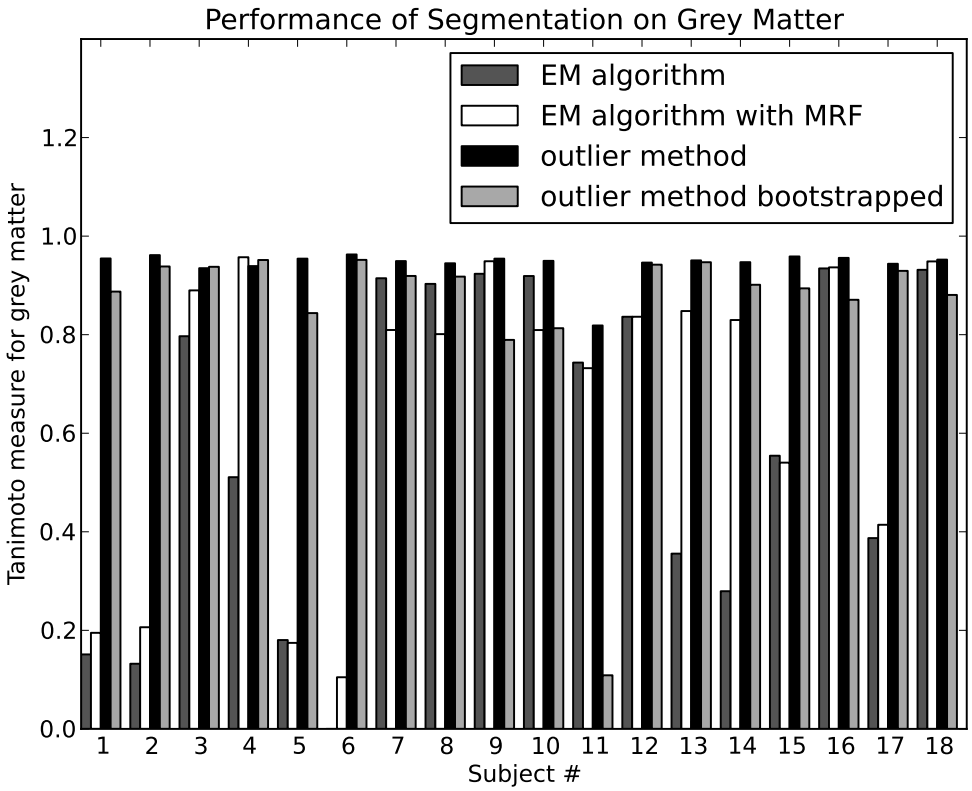


Figure 5.14: Comparison of performance of the different algorithms segmenting the grey matter of the cerebellum from real data rated with the Tanimoto Measure. Our outlier algorithm without bootstrapping performed the best most of the time. For most of the samples, our method with bootstrapping performed slightly worse than our method without bootstrapping. The "EM algorithm" and the "EM algorithm with MRF prior" performed quite erratically. Sometimes, their performance was comparable to our algorithms, but sometimes they performed much worse.

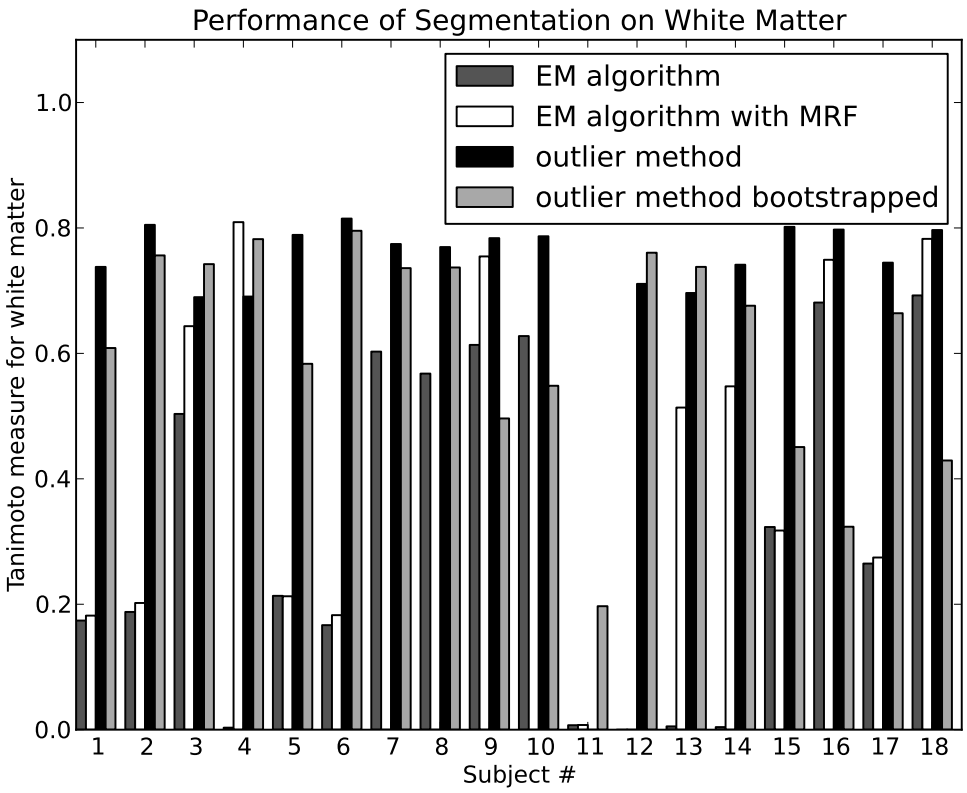


Figure 5.15: Comparison of performance of the different algorithms segmenting the white matter of the cerebellum from real data rated with the Tanimoto Measure. Our outlier method without bootstrapping tended to perform the best in general.

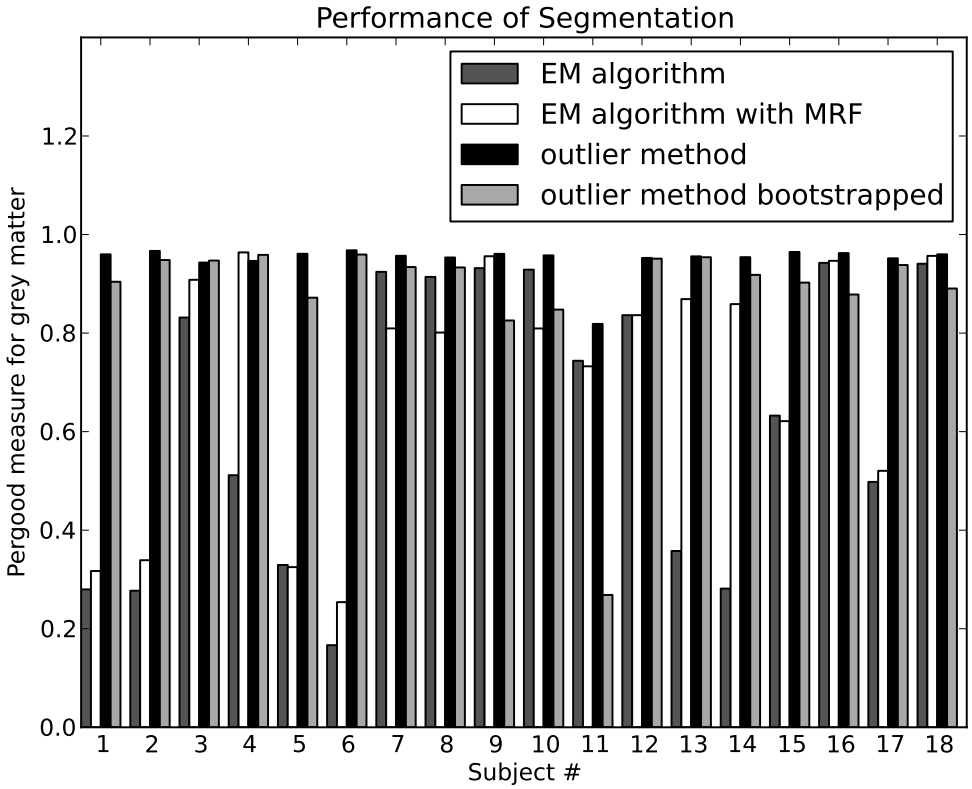


Figure 5.16: Comparison of performance of the different algorithms segmenting the cerebellum from real data rated with the Pergood Measure. Our outlier algorithm without bootstrapping performed the best most of the time. For most of the samples, our method with bootstrapping performed slightly worse than our method without bootstrapping. The "EM algorithm" and the "EM algorithm with MRF prior" performed quite erratically. Sometimes, their performance was comparable to our algorithms, but sometimes they performed much worse.

"**EM** algorithm with MRF prior" and the "**EM** algorithm" achieved metrics only very slightly higher. This poor performance could be explained by poor initial parameters. In this case, the outlier method with bootstrapping performed much better than any of the other algorithms, due to its bootstrapping with information from the whole brain.

5.3.3 Pergood Measure

In order to have just one measure to compare the methods, we look at the Pergood measure (Figure 5.16).

As before, it follows the trends of the Tanimoto metric for grey matter, which is expected because the grey matter dominates. It is worth noting that none of the Pergood values are 0, whereas the Tanimoto metric on grey matter is 0 for the "**EM** algorithm"

for subject 5. The Tanimoto metric was 0, as no voxels were classified as grey matter and therefore the intersection between the segmented voxels belonging to grey matter and the ground truth voxels belonging to grey matter was empty. Because the white matter voxels were correctly classified, the Pergood measure is not 0.

5.3.4 Class Mean Error

Looking at the class mean error (Figure 5.17), our outlier algorithm with bootstrapping performs the best. This is interesting, as it estimates the cerebellum parameters better from the whole brain than any of the other algorithms do, using just the cerebellum. This suggests that the means of the tissue classes of the cerebellum are quite close to the means of the tissue classes of the whole brain, which differs from the recent findings of Datta *et al* [39].

Our outlier algorithm with bootstrapping also has the least erratic results - performing well all of the time. The "**EM** algorithm" and the "**EM** algorithm with MRF prior" perform the most erratically. In particular, they estimate the parameters for subject 11 very poorly, whereas our algorithms do a much better job of it.

It is important to note that all the ideal means were calculated based on the *hard segmentation*. Some of these voxels are actually partial volume voxels, containing more than one tissue type. The cerebellum in particular has a relatively large proportion of partial volume voxels, which could skew the tissue class means, causing inaccurate class mean error metrics.

Our algorithm without bootstrapping performed on average next best. It is interesting to notice that when it came to parameter estimation, our method with bootstrapping performed better, while with the actual segmentation performance, our algorithm without bootstrapping performed better. As seen with the Brainweb data, better parameter estimation and better segmentation do not always correspond. If the data were perfectly Gaussian, we would expect better parameter estimation to correspond with better segmentation results. The fact that they do not correspond could indicate that the data is not perfectly Gaussian. The fact that parameter estimation and segmentation correspond better for the Brainweb data than for the IBSR could indicate that the real data is less Gaussian than the simulated data.

As with the results shown using Brainweb data, the use of outliers resulted in better parameter estimation, which was the primary reason for using outliers.

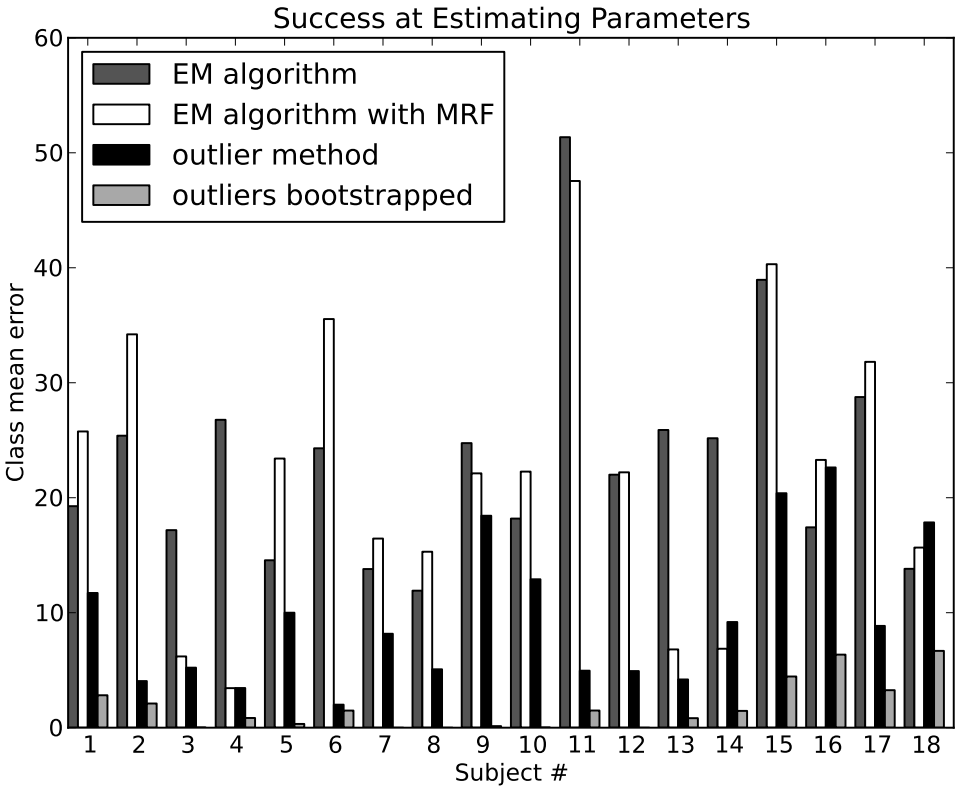


Figure 5.17: Performance on real data: success at finding the correct means of the tissue classes of the cerebellum. Our outlier algorithm with bootstrapping performs the best and also has the least erratic results.

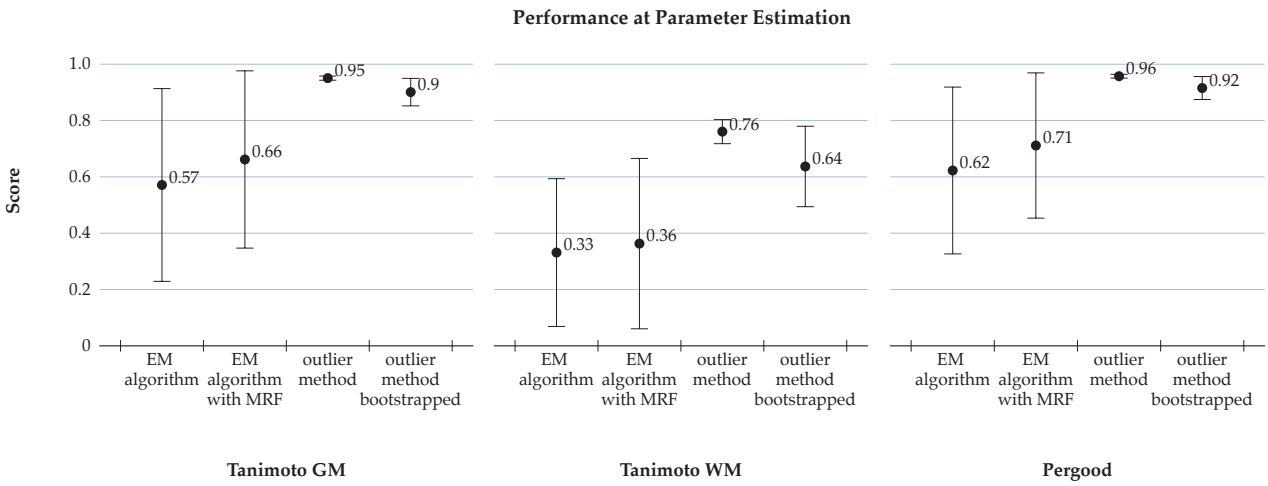


Figure 5.18: Means and standard deviations of the segmentation performance metrics on real data. For all measures that evaluate segmentation quality, our method without bootstrapping performed the best, followed by our method with bootstrapping. In addition, for all the measures except the Tanimoto measure for white matter, our method without bootstrapping had a very small standard deviation, much smaller than that of any of the other methods.

5.3.5 Overall Performance Comparison

In order to get an overall picture of how the different algorithms performed, we calculated the means and standard deviations (over all the samples) for each measure. These can be seen in Figures 5.18 and 5.19 and are tabulated in Appendix A. From this, we can see that for all measures that evaluate segmentation quality, our method without bootstrapping performed the best, followed by our method with bootstrapping. In addition, for all the measures except the Tanimoto measure for white matter, our method without bootstrapping had a very small standard deviation, much smaller than that of any of the other methods. This means that the results obtained are consistently good over all the samples. This is confirmed by looking at Figures 5.16 and 5.14: for those measures our method without bootstrapping achieves greater than 0.8 for all samples. In comparison, our method with boostrapping performs badly for subject 11 and the "**EM** algorithm" and "**EM** algorithm with MRF" perform poorly on numerous occasions, most noticeably for subject 6.

As mentioned, our method without bootstrapping had a larger standard deviation for the Tanimoto measure for white matter than for the other measures. By looking at Figure 5.15, we can see that the reason for this was that it performed poorly on white matter for subject 11. It is this single sample that causes the larger standard deviation.

When we look at the class mean error graph, Figure 5.19, we see that once again both

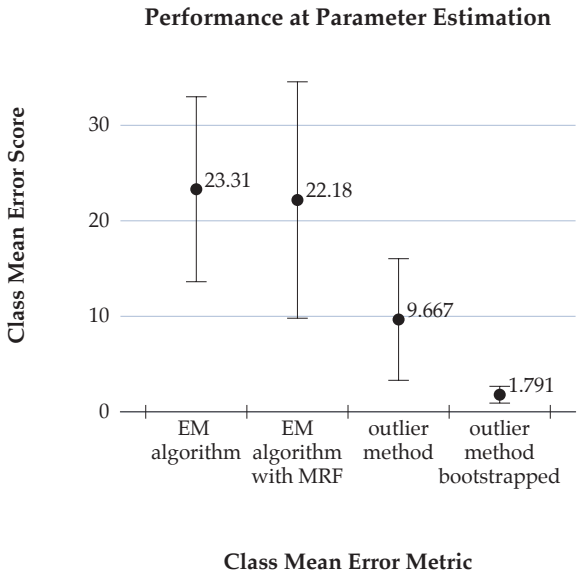


Figure 5.19: Means and standard deviations of the Class Mean Errors on real data. Both our methods perform better than the other two methods, but this time the bootstrapped version of our algorithm performs better than the version without bootstrapping.

our methods perform better than the other two methods, but this time the bootstrapped version of our algorithm performs better than the version without bootstrapping. This makes sense, since one of the aims of bootstrapping was to improve parameter estimation and this demonstrates that it succeeded in this regard. Our method with boostrapping also has a very low standard deviation, so it consistently estimates the parameters very well. What is unexpected is that while our method with bootstrapping estimates the parameters better than the version without bootstrapping, the version without bootstrapping segments the data more successfully.

Paired Difference Test

We cannot say that because our method performed better than the "**EM** algorithm with MRF" for the subjects that we looked at, it will usually perform better. We need to test that its performance was not better due to randomness: we need to show that it performs statistically *significantly* better. In order to do this, we perform a *paired difference test* for statistical significance. The aim is to disprove the null hypothesis that our method does not perform significantly better than the "**EM** algorithm with MRF" ie. that the means of the 2 sets of segmentation results aren't significantly different. We perform a t-test on the Pergood results and find that our results are statistically better at a significance level of 5%.

	Grey Matter	White Matter
Initial Means	50	100
True Means for Subject 6	84	112
True Means for Subject 12	32	57

Figure 5.20: Table of initial and true means

5.3.6 Zero Tanimoto Measures

When running our experiments, a Tanimoto measure of zero was occasionally obtained. A Tanimoto measure of zero *for a particular tissue class* occurs when there is a zero intersection for that tissue class between the ground truth and segmented image. This occurred because no voxels were assigned to that tissue class. This does not mean that *every* voxel was incorrectly assigned, but that no voxels were correctly assigned to that particular tissue class - and so the Pergood measure is not zero.

Two examples of zero Tanimoto measures when using the "**EM** algorithm" are Subject 6 and Subject 12. For Subject 6, the Tanimoto measure for grey matter is zero and for Subject 12, the Tanimoto measure for white matter is zero. To explain this, we need to look at the true mean values of the tissue classes and the values at which the means were initialised, as shown in Figure 5.20.

For Subject 6, there was an empty intersection for grey matter because the "**EM** algorithm" assigned no voxels as belonging to grey matter. Without going into the mathematics of the **EM** algorithm, the true mean for grey matter, 84, is much closer to the initial mean value for white matter, 100, than it is to the initial value for the mean of grey matter, 50. This means that voxels with intensities close to that of the true mean of grey matter will have a higher posterior probability for white matter than for grey matter (see Equation 4.1). Since most of the grey matter voxels will have intensities close to the mean of grey matter, they will have low posterior probabilities for grey matter, (and the white matter voxels will have an even lower posterior probability for grey matter). Because of the low posterior probabilities for grey matter, when the mixture weight for grey matter is updated, it will be low (see Equation 4.5). This low mixture weight for grey matter will then keep the posterior probabilities for grey matter low, which will in turn keep the next iteration's mixture weight for grey matter low. Therefore, even though the mean for white matter gets updated (it ended up being 87, which is close to the ideal of 84), the posterior probabilities for grey matter stay low and so few voxels (in this case none) get assigned to grey matter.

This is an example how *poor initialization* of parameters for the "**EM** algorithm" can result in the algorithm converging to a solution that is very far from optimal. It also shows how success at parameter estimation and success at segmentation do not always correspond.

For Subject 12, the opposite occurred: the true mean for white matter is much closer to the initial mean for grey matter than it is to the initial mean for white matter. The posterior probabilities for white matter and the mixture weight for white matter remain low and so no voxels are assigned to white matter, resulting in a zero intersection for white matter and a zero Tanimoto measure for white matter.

5.3.7 Summary

On real data, our method without bootstrapping performed statistically significantly better than the "**EM** algorithm" or the "**EM** algorithm with MRF" at a confidence level of 95%. Our method with bootstrapping estimated the parameters the best, followed by our method without bootstrapping. Our algorithm manages to pick up more of the fine structures of the cerebellum than the "**EM** algorithm with MRF".

Chapter 6

Conclusion

The segmentation of medical images into key anatomical structures is a very important process. The segmentation of brain tissue into white matter, grey matter and cerebrospinal fluid has received much attention in the literature. While the tissue segmentation of the *cerebrum* is a well investigated area, the segmentation of the *cerebellum* into tissue classes has not been addressed to any notable degree.

Medical images are often manually segmented, but this is a very time-consuming task and impractical for large amounts of data. Manual segmentation is also highly subjective and therefore not reproducible. It is because of these issues that there is a need for (semi-)automatic segmentation, ideally with as little user interaction as possible. Although MR images provide high spatial resolution and good tissue contrast - making them the best method of imaging the brain for tissue segmentation - their finite resolution means that there are often many *partial volume* voxels. Other difficulties in the segmentation of MRI data include inhomogeneities caused by the field bias, noise and motion artifacts. These difficulties are especially problematic if trying to segment tissue types based only on individual voxel intensities.

The aim of this thesis was to produce a method that segments the cerebellum into grey and white matter, capturing as much of the fine detail as possible.

Most of the work on brain tissue segmentation involves the use of a Gaussian Mixture Model, where it is assumed that each tissue's intensities are Gaussian distributed. The iterative Expectation-Maximisation (**EM**) algorithm is then used to fit Gaussians to the histogram of the data set by finding parameter estimates of the Gaussians. Often, a Markov Random Field (MRF) prior is used with the **EM** algorithm in order to include spatial information.

Due to its many fine structures, the cerebellum has a large proportion of partial

volume voxels. Manjón *et al* [89] deal with the partial volume voxels in a novel way - they attempt to remove them as outliers, so they are not involved in the parameter estimation process. This leads to better parameter estimates.

We designed an algorithm that combines the above-mentioned principles with the idea of *bootstrapping* the algorithm with information from the whole brain. Because of the fine structures in the cerebellum, a large proportion of the cerebellum's voxels will be partial volume voxels. This means that once the outliers are removed, there may be too few voxels left to accurately estimate the tissue class parameters. In order to try to overcome this, we decided to run our algorithm first on the whole brain to obtain parameters estimates, and then use those parameter estimates to classify the voxels of the cerebellum. Our algorithm runs the "**EM** algorithm with MRF prior" on the whole brain and based on this classification, we select outliers in a similar way to Manjón *et al*. We then ignore the outliers, and based on the class they were assigned to by the **EM** algorithm, we use the remaining voxels to calculate the mean and standard deviation of each tissue class. Using these parameter estimates, we run the **E**-step of the **EM** algorithm on the cerebellum to calculate the posterior probability of each voxel (including the outliers) belonging to each tissue class. Each voxel is then assigned to the tissue class with the highest posterior probability.

6.1 Findings and Implications

We tested our algorithm on simulated data provided by Brainweb and real data provided by the Internet Brain Segmentation Repository. Using the simulated data to demonstrate the complexity of cerebellum segmentation, to motivate various choices we made in our algorithm design, and to look at the effect of the MRF parameters on the segmentation performance and see how our algorithm performs in comparison to the other algorithms.

We demonstrated the complexity of the cerebellum segmentation by showing how both the "**EM** algorithm" without MRF prior and the "**EM** algorithm with MRF prior" performed better on the whole brain white matter than they did on the cerebellum white matter, due to the fine white matter structures of the cerebellum. The difference was particularly large when using the MRF, which shows how the use of an MRF prior is poorly suited to the cerebellum with its fine white matter structures.

Next, we motivated various algorithm design choices we made. We found that bootstrapping the cerebellum segmentation with parameters from the whole brain caused a significant increase in performance when using both the "**EM** algorithm" and the "**EM**

algorithm with MRF prior". This supports our decision to bootstrap our algorithm. We found that the "**EM** algorithm with MRF prior" outperformed the "**EM** algorithm" for all noise levels of 3% and higher, which includes the noise range found in real data, supporting our decision to use the "**EM** algorithm with MRF prior" for our initial segmentation.

The parameters chosen for the MRF make a large difference in the segmentation result we obtain. Recall that β determines how heavily the MRF is weighted. We demonstrated how the larger the β parameter, the smoother the segmentation results. β values that are too high result in loss of fine detail in the segmentation, while if the β value is too low, the benefit of the MRF prior is not realised. We segmented a cerebellum with different levels of noise added with the "**EM** algorithm with MRF prior" multiple times with different values of β and found that no single value of β worked better than all others at all noise levels, which demonstrates one of the major challenges of using an MRF: choosing the optimum value of β. We chose to use $\beta = 2$ for the rest of our investigation, as it worked best for a wide range of noise levels, including some of the noise levels within the range found in real data. We found that there was not a clear relationship between β and the noise level, so we concluded that other factors must influence the optimal β value and not just noise level.

Using the Tanimoto measure, we compared our algorithm to the "**EM** algorithm", the "**EM** algorithm with MRF prior" and our algorithm without bootstrapping. Our algorithm outperforms all other algorithms at low noise level, but performs worse than the "**EM** algorithm with MRF prior" at high noise levels. Little improvement is obtained by bootstrapping our algorithm with information from the whole brain, and sometimes it may even lead to a performance reduction. This may be because the mean tissue class intensities of the cerebellum are not representative of that of the whole brain. Next we compared the algorithms using the Pergood measure. The trends were very similar to that of the Tanimoto measure for grey matter.

We then looked at the class mean errors of the various algorithms. Our method with bootstrapping far outperformed all other methods. Both our methods actually improved performance as the noise level increased. The main reason for removing the outliers is to obtain better parameter estimates, and it seems as if our algorithm succeeded in doing that.

Using the real data, we visually compared the performance of our segmentation algorithm to that of the "**EM** algorithm with MRF" and numerically compared the performance of our algorithm with the various other algorithms we investigated. On real data,

our method without bootstrapping performed statistically significantly better at segmentation at a confidence level of 95% than the "**EM** algorithm" or the "**EM** algorithm with MRF". Our outlier algorithm without bootstrapping performed the most consistently, while the "**EM** algorithm" and the "**EM** algorithm with MRF" performed the most erratically. This could be because our outlier algorithm is less sensitive to parameter initialisation than the other algorithms.

Our method with bootstrapping estimated the parameters the best, followed by our method without bootstrapping. This is unexpected, since their performance rating at segmentation was the other way around. This demonstrates how segmentation quality does not always correspond to accurate parameter estimation. Our algorithm manages to pick up more of the fine structures of the cerebellum than the "**EM** algorithm with MRF prior".

Though much of the fine detail of the cerebellum cannot be picked up by MR images, it is likely that as technology improves in the future, more of the fine detail will be able to be imaged. With our algorithm's ability to pick up fine structures, hopefully it will be able to exploit the improvement in technology.

Our improved segmentation of the cerebellum, where the finer details of the neuroanatomy can be shown, could have important clinical applications. It could contribute to the early detection and enable the monitoring of the progression of neuropathological conditions, such as Fetal Alcohol Spectrum Disorder, and other pathologies affecting the cerebellum. As there has been relatively little focus on the cerebellum, our segmentation algorithm should enhance further research and studies of this important brain structure.

Appendix A

	EM alg	EM alg with MRF	outliers	outliers bootstrapped
Tanimoto GM	0.581 ± 0.334	0.666 ± 0.306	0.943 ± 0.0319	0.857 ± 0.193
Tanimoto WM	0.313 ± 0.266	0.343 ± 0.305	0.718 ± 0.184	0.612 ± 0.173
Pergood	0.629 ± 0.289	0.713 ± 0.250	0.950 ± 0.033	0.880 ± 0.158
Class Mean Error	23.305 ± 9.683	22.178 ± 12.377	9.667± 6.368	1.791 ± 2.144

Figure 6.1: Means and standard deviations of the performance measures of the various algorithms on IBSR data

Bibliography

[1] Brainweb: Simulated brain database. http://www.bic.mni.mcgill.ca/brainweb/, 2006.

[2] R. Adams and L. Bischof. Seeded region growing. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(6):641–647, 1994.

[3] M. Ahmed, S. Yamany, N. Mohamed, A. Farag, and T. Moriarty. A modified fuzzy c-means algorithm for bias field estimation andsegmentation of MRI data. *IEEE Transactions on Medical Imaging*, 21(3):193–199, 2002.

[4] M. Alfonseca. Genetic algorithms. In *ACM SIGAPL APL Quote Quad*, volume 21, pages 1–6. ACM, 1991.

[5] U. Amato, M. Larobina, A. Antoniadis, and B. Alfano. Segmentation of magnetic resonance brain images through discriminant analysis. *Journal of neuroscience methods*, 131(1-2):65–74, 2003.

[6] S. Archibald, C. Fennema-Notestine, A. Gamst, E. Riley, S. Mattson, and T. Jernigan. Brain dysmorphology in individuals with severe prenatal alcohol exposure. *Developmental Medicine & Child Neurology*, 43(3):148–154, 2001.

[7] X. Artaechevarria, A. Muñoz-Barrutia, and C. Ortiz-de Solorzano. Combination strategies in multi-atlas image segmentation: Application to brain mr data. *Medical Imaging, IEEE Transactions on*, 28(8):1266–1277, 2009.

[8] B. Aubert-Broche, A. Evans, and L. Collins. A new improved version of the realistic digital brain phantom. *NeuroImage*, 32(1):138–145, 2006.

[9] B. Aubert-Broche, M. Griffin, G. Pike, A. Evans, and D. Collins. Twenty new digital brain phantoms for creation of validation image data bases. *Medical Imaging, IEEE Transactions on*, 25(11):1410–1416, 2006.

[10] S. Aylward and J. Coggins. Spatially Invariant Classification of Tissues in MR Images. *Proceedings of the Third Conference on Visualization in Biomedical Computing*, 1994.

[11] K. Barrett. *Ganong's review of medical physiology*. McGraw-Hill Medical, 2010.

[12] M. Basu. Gaussian-based edge-detection methods-a survey. *IEEE Transactions on Systems, Man and Cybernetics, Part C*, 32(3):252–260, 2002.

[13] P. Bazin and D. Pham. Topology-preserving tissue classification of magnetic resonance brain images. *Medical Imaging, IEEE Transactions on*, 26(4):487–496, 2007.

[14] J. Besag. On the Statistical Analysis of Dirty Pictures, JR Statist. *Soc. B*, 48(3):259–302, 1986.

[15] J. Bezdek, R. Ehrlich, et al. Fcm: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2-3):191–203, 1984.

[16] J. Bezdek, L. Hall, and L. Clarke. Review of MR image segmentation techniques using pattern recognition. *Med. Phys*, 20(4):1033–1048, 1993.

[17] B. Bhanu, S. Lee, and J. Ming. Adaptive image segmentation using a genetic algorithm. In *IEEE Transactions on Systems, Man and Cybernetics*, volume 25, pages 1543–1567, December 1995.

[18] C. Biernacki, G. Celeux, and G. Govaert. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics and Data Analysis*, 41(3-4):561–575, 2003.

[19] G. Bilbro and W. Snyder. Optimization of functions with many minima. *Systems, Man and Cybernetics, IEEE Transactions on*, 21(4):840–849, 1991.

[20] J. A. Bilmes. A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. *Technical Report of the International Computer Science Institute, Berkeley, CA*, 1998.

[21] S. Bricq, C. Collet, and J. Armspach. Unifying framework for multimodal brain MRI segmentation based on Hidden Markov Chains. *Medical Image Analysis*, 2008.

[22] R. Brouwer, H. Hulshoff Pol, and H. Schnack. Segmentation of MRI brain scans using non-uniform partial volume densities. *NeuroImage*, 49(1):467–477, 2010.

[23] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[24] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.

[25] G. Celeux and J. Diebolt. The sem algorithm: a probabilistic teacher algorithm derived from the em algorithm for the mixture problem. *Computational Statistics Quarterly*, 2(1):73–82, 1985.

[26] Z. Chi, H. Yan, and T. Pham. *Fuzzy algorithms: with applications to image processing and pattern recognition*. World Scientific Pub Co Inc, 1996.

[27] H. Choi, D. Haynor, and Y. Kim. Partial volume tissue classification of multichannel magnetic resonance images-a mixel model. *IEEE Transactions on Medical Imaging*, 10(3):395–407, 1991.

[28] L. Clarke, R. Velthuizen, M. Camacho, J. Heine, M. Vaidyanathan, L. Hall, R. Thatcher, and M. Silbiger. MRI Segmentation: Methods and Applications. *Magnetic Resonance Imaging*, 13:343–343, 1995.

[29] C. Cocosco, V. Kollokian, K. Remi, G. Pike, and A. Evans. Brainweb: Online interface to a 3D MRI simulated brain database. In *NeuroImage*. Citeseer, 1997.

[30] C. Cocosco, A. Zijdenbos, and A. Evans. A fully automatic and robust brain MRI tissue classification method. *Medical Image Analysis*, 7(4):513–527, 2003.

[31] I. Cohen, L. Cohen, and N. Ayache. Using deformable surfaces to segment 3-d images and infer differential structures. *CVGIP: Image understanding*, 56(2):242–263, 1992.

[32] L. Cohen. On active contour models and balloons. *CVGIP: Image understanding*, 53(2):211–218, 1991.

[33] D. Collins, A. Zijdenbos, V. Kollokian, J. Sled, N. Kabani, C. Holmes, and A. Evans. Design and construction of a realistic digital brain phantom. *Medical Imaging, IEEE Transactions on*, 17(3):463–468, 2002.

[34] M. Cuadra. Atlas-based segmentation and classification of magnetic resonance brain images. *Signal Process. Inst., Swiss Federal Inst. Technol.*, 2003.

[35] M. Cuadra, L. Cammoun, T. Butz, O. Cuisenaire, and J. Thiran. Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images. *IEEE Transactions on Medical Imaging*, 24(12):1548–1565, 2005.

[36] M. Cuadra, B. Platel, E. Solanas, T. Butz, and J. Thiran. Validation of Tissue Modelization and Classification Techniques in T1-Weighted MR Brain Images. *Lecture Notes in Computer Science*, pages 290–297, 2002.

[37] R. DAgostino, A. Belanger, and R. DAgostino Jr. A suggestion for using powerful and informative tests of normality. *The American Statistician*, 44(4):316–321, 1990.

[38] C. Darken and J. Moody. Fast adaptive k-means clustering: some empirical results. In *Neural Networks, 1990., 1990 IJCNN International Joint Conference on*, pages 233–238. IEEE, 1990.

[39] S. Datta, G. Tao, R. He, J. Wolinsky, and P. Narayana. Improved cerebellar tissue classification on magnetic resonance images of brain. *Journal of Magnetic Resonance Imaging*, 29(5):1035–1042, 2009.

[40] L. Davis. A survey of edge detection techniques. *Computer Graphics and Image Processing*, 4(3):248–270, 1975.

[41] S. Dellepiane and F. Fontana. Extraction of intensity connectedness for image processing. *Pattern Recognition Letters*, 16(3):313–324, 1995.

[42] B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the em algorithm. *The Annals of Statistics*, 27(1):94–128, 1999.

[43] A. Dempster, N. Laird, D. Rubin, et al. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38, 1977.

[44] M. Desco, J. Gispert, S. Reig, A. Santos, J. Pascau, N. Malpica, and P. Garcia-Barreno. Statistical segmentation of multidimensional brain datasets. *Proceedings of SPIE*, 4322:185, 2001.

[45] L. Dice. Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3):297–302, 1945.

[46] P. Dokládal, I. Bloch, M. Couprie, D. Ruijters, R. Urtasun, and L. Garnero. Topologically controlled segmentation of 3D magnetic resonance images of the head by using morphological operators. *Pattern Recognition*, 36(10):2463–2478, 2003.

[47] D. Eremina, X. Li, W. Zhu, J. Wang, and Z. Liang. Investigation on an EM Framework for Partial Volume Image Segmentation. In *Proc. of SPIE Vol*, volume 6144, pages 61444D–1.

[48] A. Ferreira da Silva. A Dirichlet process mixture model for brain MRI tissue classification. *Medical Image Analysis*, 11(2):169–182, 2007.

[49] M. Figueiredo and A. Jain. Unsupervised Learning of Finite Mixture Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 381–396, 2002.

[50] S. Forrest. Genetic algorithms. *ACM Computing Surveys*, 28(1):77–80, March 1996.

[51] K. Fu and J. Mui. A survey on image segmentation. *Pattern Recognition*, 13(1):3–16, 1981.

[52] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian distribution of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

[53] M. González Ballester, A. Zisserman, and M. Brady. Segmentation and measurement of brain structures in MRI including. *Medical Image Analysis*, 4:189–200, 2000.

[54] M. González Ballester, A. Zisserman, and M. Brady. Estimation of the partial volume effect in MRI. *Medical Image Analysis*, 6(4):389–405, 2002.

[55] V. Grau, A. Mewes, M. Alcaniz, R. Kikinis, and S. Warfield. Improved watershed transform for medical image segmentation using prior information. *IEEE Transactions on Medical Imaging*, 23(4):447–458, 2004.

[56] H. Greenspan, A. Ruf, and J. Goldberger. Constrained Gaussian mixture model framework for automatic segmentation of MR brain images. *IEEE Transactions on Medical Imaging*, 25(9):1233–1245, 2006.

[57] H. Gudbjartsson and S. Patz. The Rician distribution of noisy MRI data. *Magnetic Resonance in Medicine*, 34(6):910–914, 1995.

[58] A. Guyton and J. Hall. *Textbook of medical physiology*, volume 9. Saunders Philadelphia, PA, 1991.

[59] D. Hamilton, P. Kodituwakku, R. Sutherland, and D. Savage. Children with fetal alcohol syndrome are impaired at place learning but not cued-navigation in a virtual morris water task. *Behavioural brain research*, 143(1):85–94, 2003.

[60] R. Haralick and L. Shapiro. Image segmentation techniques. *Computer vision, graphics, and image processing*, 29(1):100–132, 1985.

[61] K. Held, E. Kops, B. Krause, W. Wells III, R. Kikinis, and H. Muller-Gartner. Markov random field segmentation of brain MR images. *IEEE Transactions on Medical Imaging*, 16(6):878–886, 1997.

[62] E. C. Hildreth. Theory of edge detection. 207:187–217, 1980.

[63] S. Hojjatoleslami and J. Kittler. Region growing: A new approach. *Image Processing, IEEE Transactions on*, 7(7):1079–1084, 1998.

[64] J. Holland. Adaption in Natural and Artificial Systems, 1975.

[65] T. Hong and A. Rosenfeld. Compact region extraction using weighted pixel linking in a pyramid. *IEEE transactions on pattern analysis and machine intelligence*, 6(2):222–229, 1984.

[66] M. G. Hospital. Internet brain segmentation repository, 2009.

[67] G. Hu and Mageras. Survey of Recent Volumetric Medical Image segmentation Techniques.

[68] D. Iosifescu, M. Shenton, S. Warfield, R. Kikinis, J. Dengler, F. Jolesz, and R. McCarley. An Automated Registration Algorithm for Measuring MRI Subcortical Brain Structures. *Neuroimage*, 6(1):13–25, 1997.

[69] P. Jaccard. The Distribution of Flora in the Alpine Zone. *New Phytologist*, 11:37–50, 1912.

[70] M. Janani and M. Kavithadevi. A survey on image segmentation using decision fusion method. *Global Journal of Computer Science and Technology*, 11(10), 2011.

[71] J. Jimenez-Alaniz, V. Medina-Banuelos, and O. Yanez-Suarez. Data-driven brain MRI segmentation supported on edge confidence and a priori tissue information. *IEEE Transactions on Medical Imaging*, 25(1):74–83, 2006.

[72] N. Joshi and M. Brady. A non-parametric model for partial volume segmentation of mr images. In *Proceedings BMVC*, volume 5, pages 919–928. Citeseer, 2005.

[73] T. Kapur, W. Grimson, W. Wells, and R. Kikinis. Segmentation of brain tissue from magnetic resonance images. *Medical Image Analysis*, 1(2):109–127, 1996.

[74] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988.

[75] A. Kelemen, G. Szekely, and G. Gerig. Elastic model-based segmentation of 3-D neuroradiological data sets. *IEEE Transactions on Medical Imaging*, 18(10):828–839, 1999.

[76] R. Kindermann, J. Snell, and A. M. Society. *Markov random fields and their applications*. American Mathematical Society Providence, RI, 1980.

[77] R. Kwan, A. Evans, and G. Pike. An extensible MRI simulator for post-processing evaluation. In *Visualization in Biomedical Computing*, pages 135–140. Springer, 1996.

[78] R. Kwan, A. Evans, and G. Pike. MRI simulation-based evaluation of image-processing and classification methods. *Medical Imaging, IEEE Transactions on*, 18(11):1085–1097, 2002.

[79] D. Laidlaw, K. Fleischer, and A. Barr. Partial-volume Bayesian classification of material mixtures in MR volume data using voxel histograms. *IEEE Transactions on Medical Imaging*, 17(1):74–86, 1998.

[80] S. Lee, S. Yoon Chung, and R. Park. A comparative performance study of several global thresholding techniques for segmentation. *Computer Vision, Graphics, and Image Processing*, 52(2):171–190, 1990.

[81] J. Levy Vehel, P. Mignot, and J. Berroir. Texture and multifractals: New tools for image analysis. *Rapport de Recherche-Institut National de Recherche en Infomatique et en Automatique*, 1992.

[82] S. Li. Markov random field models in computer vision. *Lecture Notes in Computer Science*, 801:361–370, 1994.

[83] A. Liew and H. Yan. Current Methods in the Automatic Tissue Segmentation of 3D Magnetic Resonance Brain Images. *Current Medical Imaging Reviews*, 2(1):91–103, 2006.

[84] T. Louis. Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(2):226–233, 1982.

[85] L. Lucchese and S. Mitra. Colour Image Segmentation: A State-of-the-Art Survey. *Proceedings-Indian National Science Academy Part A*, 67(2):207–222, 2001.

[86] V. H. M. Sonka and R. Boyle. *Image Processing, Analysis and Machine Vision, Second Edition*. Brooks/Cole, Thomson Asia Pte Led, United State of America.

[87] J. MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, page 14. California, USA, 1967.

[88] J. Mangin, V. Frouin, I. Bloch, J. Régis, and J. López-Krahe. From 3D magnetic resonance images to structural representations of the cortex topography using topology preserving deformations. *Journal of Mathematical Imaging and Vision*, 5(4):297–318, 1995.

[89] J. Manjón, J. Tohka, G. Garcia-Marti, J. Carbonell-Caballero, J. Lull, L. Marti-Bonmati, and M. Robles. Robust MRI brain tissue parameter estimation by multistage outlier rejection. *Magnetic Resonance in Medicine*, 59(4), 2008.

[90] I. Manousakas, P. Undrill, G. Cameron, and T. Redpath. Split-and-merge segmentation of magnetic resonance medical images: Performance evaluation and extension to three dimensions. *Computers and Biomedical Research*, 31(6):393–412, 1998.

[91] J. T. L. Mattson, S. N. and E. P. Riley. MRI and prenatal alcohol exposure. *Alcohol Health & Research World*, 18(1):49–52, 1994.

[92] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: a survey. *Medical image analysis*, 1(2):91–108, 1996.

[93] P. McKenzie and M. Alder. *Initializing the EM algorithm for use in Gaussian Mixture Modelling*. University of Western Australia, 1993.

[94] G. McLachlan and T. Krishnan. *The EM algorithm and extensions*. LibreDigital, 2008.

[95] E. Meintjes, J. Jacobson, C. Molteno, J. Gatenby, C. Warton, C. Cannistraci, H. Hoyme, L. Robinson, N. Khaole, J. Gore, et al. An fmri study of number processing in children with fetal alcohol syndrome. *Alcoholism: Clinical and Experimental Research*, 34(8):1450–1464, 2010.

[96] R. Moretti. *The Cerebellum and the Reading Process*. Nova Science Publishers, 2003.

[97] W. Niessen, K. Vincken, J. Weickert, B. Romeny, and M. Viergever. Multiscale Segmentation of Three-Dimensional MR Brain Images. *International Journal of Computer Vision*, 31(2):185–202, 1999.

[98] A. Noe and J. Gee. Partial volume segmentation of cerebral MRI scans with mixture model clustering. *Proceedings of the 17th International Conference on Information Processing in Medical Imaging (IPMI01)*, pages 423–430.

[99] A. Noe, S. Kovacic, and J. Gee. Segmentation of cerebral mri scans using a partial volume model, shading correction, and an anatomical prior. *Proc. SPIE (Medical Image Processing)*, 2001.

[100] P. Nurmi. Mixture models. *Helsinki Institute for Information Technology*, 2004.

[101] E. Parzen. On the estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3):10651076, 1962.

[102] T. Pavlidis and Y. Liow. Integrating region growing and edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3):225–233, 1990.

[103] T. Peli and D. Malah. A study of edge detection algorithms. *Computer Graphics and Image Processing*, 20(1):1–21, 1982.

[104] F. Pernkopf and D. Bouchaffra. Genetic-Based EM Algorithm for Learning Gaussian Mixture Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1344–1348, 2005.

[105] D. Pham and J. Prince. Adaptive fuzzy segmentation of magnetic resonance images. *IEEE Transactions on Medical Imaging*, 18(9):737–752, 1999.

[106] D. Pham, C. Xu, and J. Prince. A survey of current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2(315-337):1, 2000.

[107] F. Picard. An introduction to mixture models. *Univ. dEvry, France UMR CNRS*, 8071, 2007.

[108] R. Pohle and K. Toennies. Segmentation of medical images using adaptive region growing. In *Proc. SPIE Medical Imaging*, volume 4322, pages 1337–46. Citeseer, 2001.

[109] R. Pooley. Fundamental Physics of MR Imaging 1. *RadioGraphics*, 25(4):1087–1099, 2005.

[110] W. Pratt. *Digital image processing: PIKS inside*. John Wiley & Sons, Inc. New York, NY, USA, 2001.

[111] J. Rajapakse, J. Giedd, and J. Rapoport. Statistical approach to segmentation of single-channel cerebral MR images. *IEEE Transactions on Medical Imaging*, 16(2):176–186, 1997.

[112] S. Raya. Low-level segmentation of 3-D magnetic resonance brain images-arule-based system. *IEEE Transactions on Medical Imaging*, 9(3):327–337, 1990.

[113] M. Rhodes. An algorithmic approach to controlling search in three-dimensional image data. *ACM SIGGRAPH Computer Graphics*, 13(2):134–142, 1979.

[114] N. Richard, M. Dojat, and C. Garbay. Distributed Markovian segmentation: Application to MR brain scans. *Pattern Recognition*, 40(12):3467–3480, 2007.

[115] E. Riley and C. McGee. Fetal alcohol spectrum disorders: an overview with emphasis on changes in brain and behavior. *Experimental Biology and Medicine*, 230(6):357–365, 2005.

[116] I. Rish. An empirical study of the naive bayes classifier. In *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, volume 3, pages 41–46, 2001.

[117] S. Ruan, C. Jaggi, J. Xue, J. Fadili, D. Bloyet, and C. Greyc-Ismra. Brain tissue classification of magnetic resonance images usingpartial volume modeling. *IEEE Transactions on Medical Imaging*, 19(12):1179–1187, 2000.

[118] A. Ruf, H. Greenspan, and J. Goldberger. Tissue classification of noisy mr brain images using constrained gmm. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005*, pages 790–797, 2005.

[119] P. Sahoo, S. Soltani, A. Wong, and Y. Chen. A survey of thresholding techniques. *Computer Vision, Graphics, and Image Processing*, 41(2):233–260, 1988.

[120] P. Santago and H. Gage. Quantification of MR brain images by mixture density and partialvolume modeling. *IEEE Transactions on Medical Imaging*, 12(3):566–574, 1993.

[121] P. Santago and H. Gage. Statistical models of partial volume effect. *IEEE Transactions on Image Processing*, 4(11):1531–1540, 1995.

[122] P. Schroeter, J. Vesin, T. Langenberger, and R. Meuli. Robust parameter estimation of intensity distributions for brainmagnetic resonance images. *IEEE Transactions on Medical Imaging*, 17(2):172–186, 1998.

[123] N. Senthilkumaran and R. Rajesh. Edge detection techniques for image segmentation-A survey of soft computing approaches. *International Journal of Recent Trends in Engineering*, 1(2):250–254, 2009.

[124] M. Sezgin and B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13:146, 2004.

[125] N. Sharma and L. Aggarwal. Automated medical image segmentation techniques. *Journal of Medical Physics/Association of Medical Physicists of India*, 35(1):3, 2010.

[126] D. Shattuck, S. Sandor-Leahy, K. Schaper, D. Rottenberg, and R. Leahy. Magnetic Resonance Image Tissue Classification Using a Partial Volume Model. *Neuroimage*, 13(5):856–876, 2001.

[127] F. Shi, Y. Fan, S. Tang, J. Gilmore, W. Lin, and D. Shen. Brain tissue segmentation of neonatal mr images using a longitudinal subject-specific probabilistic atlas. In

Proceedings-Society of Photo-Optical Instrumentation Engineers, volume 7259. NIH Public Access, 2009.

[128] J. Sled, A. Zijdenbos, and A. Evans. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *Medical Imaging, IEEE Transactions on*, 17(1):87–97, 1998.

[129] D. Terzopoulos, A. Witkin, and M. Kass. Symmetry-seeking models and 3d object reconstruction. *International Journal of Computer Vision*, 1(3):211–221, 1988.

[130] S. Thede. An introduction to genetic algorithms. In *Journal of Computing Sciences in Colleges*, volume 20, pages 115–123, October 2004.

[131] J. Tohka, I. Dinov, D. Shattuck, and A. Toga. Brain MRI tissue classification based on local Markov random fields. *Magnetic Resonance Imaging*, 28(4):557–573, 2010.

[132] J. Tohka, E. Krestyannikov, I. Dinov, D. Shattuck, U. Ruotsalainen, and A. Toga. Genetic algorithms for finite mixture model based tissue classification in brain mri. In *Proc. of European Medical and Biological Engineering Conference, IFMBE Proceedings*, volume 11, pages 4077–4082. Citeseer, 2005.

[133] J. Tohka, A. Zijdenbos, and A. Evans. Fast and robust parameter estimation for statistical partial volume models in brain MRI. *Neuroimage*, 23(1):84–97, 2004.

[134] V. Torre and T. Poggio. On edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (2):147–163, 1986.

[135] J. Udupa and S. Samarasekera. Fuzzy connectedness and object definition: theory, algorithms, and applications in image segmentation. *Graphical Models and Image Processing*, 58(3):246–261, 1996.

[136] K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Transactions on Medical Imaging*, 20(8):677–688, 2001.

[137] 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.

[138] 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.

[139] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. A Statistical Framework for Partial Volume Segmentation. In *Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 204–212. Springer-Verlag London, UK, 2001.

[140] 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.

[141] B. Vemuri, S. Rahman, and J. Li. Multiresolution adaptive k-means algorithm for segmentation of brain mri. *Image Analysis Applications and Computer Graphics*, pages 347–354, 1995.

[142] J. Vergés Llahí et al. Color constancy and image segmentation techniques for applications to mobile robotics. 2005.

[143] S. Warfield, R. Mulkern, C. Winalski, F. Jolesz, and R. Kikinis. An Image Processing Strategy for the Quantification and Visualization of Exercise-Induced Muscle MRI Signal Enhancement. *Journal of Magnetic Resomance Imaging*, 11:525–531, 2000.

[144] S. Warfield, K. Zou, and W. Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *Medical Imaging, IEEE Transactions on*, 23(7):903–921, 2004.

[145] G. Wei and M. Tanner. A monte carlo implementation of the em algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, pages 699–704, 1990.

[146] W. Wells III, W. Grimson, R. Kikinis, and F. Jolesz. Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging*, 15(4):429–442, 1996.

[147] J. Weszka. A survey of threshold selection techniques. *Computer Graphics and Image Processing*, 7(2):259–265, 1978.

[148] D. Withey and Z. Koles. A review of medical image segmentation: methods and available software. *International Journal of Bioelectromagnetism. v10 i3*, pages 125–148.

[149] C. Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, pages 95–103, 1983.

[150] C. Xu and J. Prince. Snakes, shapes, and gradient vector flow. *Image Processing, IEEE Transactions on*, 7(3):359–369, 1998.

[151] L. Xu and M. Jordan. On Convergence Properties of the EM Algorithm for Gaussian Mixtures. *Neural Computation*, 8(1):129–151, 1996.

[152] M. Yan and J. Karp. Segmentation of 3d brain mr using an adaptive k-means clustering algorithm. In *Nuclear Science Symposium and Medical Imaging Conference, 1994., 1994 IEEE Conference Record*, volume 4, pages 1529–1533. IEEE, 1994.

[153] Y. Zhang, J. Brady, and S. Smith. An HMRF-EM Algorithm for Partial Volume Segmentation of Brain MRI FMRIB Technical Report TR01YZ1. Technical report, Technical report, Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, 2001.

[154] 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.

[155] D. Ziou and S. Tabbone. Edge Detection Techniques - an Overview. *Pattern Recognition and Image Analysis*, 8:537–559, 1998.