IMPROVING THE SUB-CORTICAL GM SEGMENTATION USING EVOLUTIONARY HIERARCHICAL REGION MERGING

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF MIDDLE EAST TECHNICAL UNIVERSITY

BY

MUSTAFA ULAŞ ÇİFTÇİOĞLU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
IN
THE DEPARTMENT OF MEDICAL INFORMATICS

JUNE 2011

Approval of the Graduate School of Informatics

_____

Prof. Dr. Nazife Baykal

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

_____

Assist. Prof. Dr. Didem Gökçay

Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

_____

Assist. Prof. Dr. Didem Gökçay

Supervisor

Examining Committee Members

Assist. Prof. Dr. Alptekin Temizel (METU, WBL)_____

Assist. Prof. Dr. Didem Gökçay (METU, MI)_____

Prof. Dr. Metehan Çiçek (Ankara U.) _____

Assist. Prof. Dr. Yeşim Aydın Son (METU, MI) _____

Assist. Prof. Dr. İlkay Ulusoy (METU, EEE)_____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Mustafa Ulaş Çiftçioğlu

Signature :

ABSTRACT

IMPROVING THE SUB-CORTICAL GM SEGMENTATION USING EVOLUTIONARY HIERARCHICAL REGION MERGING

Çiftçioğlu, Mustafa Ulaş
M.Sc., Department of Medical Informatics
Supervisor : Assist. Prof. Dr. Didem Gökçay

June 2011, 119 pages

Segmentation of sub-cortical Gray Matter (GM) structures in magnetic resonance brain images is crucial in clinic and research for many purposes such as early diagnosis of neurological diseases, guidance of surgical operations and longitudinal volumetric studies. Unfortunately, the algorithms that segment the brain into 3 tissues usually suffer from poor performance in the sub-cortical region. In order to increase the detection of sub-cortical GM structures, an evolutionary hierarchical region merging approach, abbreviated as EHRM, is proposed in this study. Through EHRM, an intensity based region merging is utilized while merging is allowed to proceed among disconnected regions. Texture information is also incorporated into the scheme to prevent the region merging between tissues with

similar intensity but different texture properties. The proposed algorithm is tested on real and simulated datasets. The performance is compared with a popular segmentation algorithm, which is also intensity driven: the FAST algorithm [1] in the widely used FSL suite. EHRM is shown to make a significant improvement the detection of sub-cortical GM structures. Average improvements of 10%, 36% and 22% are achieved for caudate, putamen and thalamus respectively. The accuracy of volumetric estimations also increased for GM and WM. Performance of EHRM is robust in presence of bias field. In addition, EHRM operates in O(N) complexity. Furthermore, the algorithm proposed here is simple, because it does not incorporate spatial priors such as an atlas image or intensity priors. With these features, EHRM may become a favorable alternative to the existing brain segmentation tools.

Keywords: subcortical, region merging, MRI, brain, segmentation.

ÖZ

KORTEKS ALTI GRİ MADDE BÖLÜTLEMESİNİ EVRİMSEL HİYERARŞİK BÖLGE KAYNAŞTIRMASI KULLANARAK GELİŞTİRMEK

Çiftçioğlu, Mustafa Ulaş
Yüksek Lisans, Sağlık Bilişimi
Tez Yöneticisi : Yrd. Doç. Dr. Didem Gökçay

Haziran 2011, 119 sayfa

Manyetik rezonans beyin görüntülerinde korteks-altı gri madde yapıların bölütlenmesi, klinikte ve araştırmada, nörolojik hastalıkların erken tanısı, cerrahi operasyonların yönlendirilmesi ve longitudinal hacimsel çalışmalar gibi birçok amaç için çok önemlidir. Malesef beyni 3 dokuya bölütleyen algoritmalar genellikle korteks-altı bölgedeki zayıf performanstan zarar görmektedir. Bu çalışmada, korteks-altı gri madde yapıların tespitini artırmak için, EHRM olarak kısaltılan evrimsel hiyerarşik bölge kaynaştırması yaklaşımı önerilmektedir. EHRM ile birlikte, intensite temelli bir bölge kaynaştırması, kaynaşmanın bağlantısız bölgeler arasında da ilerlemesine izin verilerek, yararlanılmıştır. Örgü bilgise de intensite açısından benzer fakat örgü özellikleri farklı dokular arasındaki kaynaşmaların engellenmesi için şemaya dahil edilmiştir. Önerilen algoritma gerçek ve simule edilmiş veri setlerinde test edilmiştir. Performans, yaygınca kullanılan FSL paketindeki intensite dayalı popüler bir bölütleme algoritması olan FAST

algoritması[1] ile karşılaştırılmıştır. EHRM'in korteks-altı gri madde yapıların tespitinde anlamlı bir iyileştirme yaptığı gösterilmistir. Kaudat, putamen ve talamus için sırasıyla 10%, 36% ve 22% ortalama iyileşme sağlanmıştır. Gri madde ve beyaz madde için hacimsel kestirim doğruluğu da artmıştır. EHRM'in performansı, manyetik alan sapması varlığında dayanıklıdır. Ek olarak, EHRM O(N) kompleksitede çalışmaktadır. Ayrıca, burda önerilen algoritma basittir çünkü atlas görüntü gibi uzaysal önbilgi veya intensite önbilgisi dahil etmemektedir. Bu özelliklerle, EHRM mevcut beyin bölütleme araçlarına uygun bir alternatif haline gelebilir.

Anahtar Kelimeler: korteks-altı, bölge kaynaştırması, Manyetik Rezonans Görüntüleme, beyin, bölütleme

To My Family and Merve

ACKNOWLEDGMENTS

I consider myself very lucky to have an supervisor like Assist. Prof. Dr. Didem Gökçay, who means more than an academic supervisor to me. I would like to express my deepest thanks to her for her guidance, patience, encouragement and confidence to me. Her style of thinking and rich perspective contributed not only to this study but also to my understanding and view as a graduate student who wishes to pursue an academic career.

I would like to thank very much to the administrive staff like Ali Kantar and Sibel Gülnar and friends at Informatics Institute for all their support. I would also like to thank to Merve and friends in Ankara for their compaionship during this study.

I would like to appreciate my family, Rukiye Çiftçioğlu, Muhsin Çiftçioğlu and Ertuğrul Necdet Çiftçioğlu, for their love, guidance and encouragement in every stage of my life.

I would like to express my special thanks to the Scientific and Technological Research Council of Turkey (TÜBİTAK) for their generous scholarship which allowed me to concentrate my studies in a full time manner, without corcerning financial issues.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

MRI	Magnetic Resonance Imaging
RF	Radio Frequency
WM	White Matter
GM	Gray Matter
CSF	Cerebrospinal Fluid
EHRM	Evolutionary Hierarchical Region Merging
MRF	Markov Random Field
FMM	Finite Mixture Model
EM	Expectation Maximization
MAP	Maximum a Posteriori
SNR	Signal-to-Noise Ratio
CCA	Cellular Coevolutionary Algorithm
FSL	FMRIB Software Library
PVE	Partial Volume Effect
pdf	probability density function
TH	Texture Heterogeneity
SK	Spatial K-means

CHAPTER 1

INTRODUCTION

Magnetic Resonance Imaging (MRI) has changed the era of medical imaging in clinics and research beginning from late 1980s. Its non-invasiveness, high soft tissue contrast, high spatial resolution are the major reasons why MRI became so popular. Today, MRI is not only utilized for anatomical imaging, but also used for functional imaging (functional MRI (fMRI)), the imaging of movements of water molecules on biological tissues (diffusion MRI, diffusion tensor imaging (DTI)), real time imaging (Real Time MRI), guiding interventional operations (interventional MRI), imaging blood vessels (MR Angiography (MRA)).

MRI depends on the detection of the net magnetic field created by the nuclear spins. The spins, which are randomly oriented in the absence of an external magnetic field, can produce a detectable net magnetic field when a strong DC magnetic field is applied. The manipulation of this magnetic field with gradient magnetic fields and Radio Frequency (RF) excitation pulses enables the detection of signals dependent on the tissue properties. The flexibility in the application of gradient fields and RF excitation, together with the several properties of tissues, enables the usage of MRI for different applications.

Brain, which is the center of the nervous system, is composed of soft tissues (see Appendix A for main brain structures). MR is a very suitable modality for brain imaging due to its high soft tissue contrast. Also its non-invasiveness allows MR to be used more freely in comparison to other popular modalities like Computed Tomography (CT) which produces ionizing radiation. Therefore, the usage of MR for research purposes is possible and the clinical usage does not have

certain limitations such as time of scanning and maximum number of scans due to radiation exposure.

High resolution and good contrast anatomic MR brain images contain rich information but the utilization of this rich information is limited if the images are only inspected visually by the radiologists as in traditional radiology. However, this rich information can be utilized when different tissues and structures can be analyzed separately. The separation of the brain into parts of interest enables accurate volumetric analysis, better 3-D visualization of anatomy, the detection of abnormalities and detection of changes in longitudinal studies. All this information can then be utilized for many purposes, such as early diagnosis of neurological diseases, guiding the interventions, measuring the progress of therapies, finding markers of diseases, mapping functions to anatomical structures and understanding the anatomy of the brain.

Separation of the brain into desired regions can be performed manually by following manual segmentation procedures. The quality of such a segmentation can be very good but this procedure requires high labor and time. The wide utilization of these segmentations is not feasible due to these stated limitations. In order to increase the utilization of MR images, many semi-automated and fully automated procedures are proposed beginning from early 1990s.

The automated algorithms can work quite successfully for some cases but their performance can degrade especially when the properties of the images are different than the images that are used to train or optimize the algorithm. Generalization is an important problem for brain segmentation algorithms because the MR devices, MR sequences and sequence parameters are not standard. Also noise, bias field, inter-subject anatomical variability and tissue property variability among subjects change the properties of the MR images. All these factors are challenges for the development of accurate automated brain segmentation algorithms. The only property that can be accepted as common in MR brain images are the order of the intensity levels of tissue for a certain weighted image. For

instance, in a conventional T1 weighted anatomical brain image, the intensity levels of WM tend to be higher than GM and the intensity levels of GM is expected to be higher than CSF.

Although the robustness of algorithms is an important issue, satisfactory results can be usually obtained in certain brain areas. The cortical regions, which have good contrast in most cases are segmented successfully the algorithms such as FSL FAST [1]. The effect of bias field can be removed by incorporating the bias field into the segmentation model, which improves the segmentation quality in cortical regions. Bias field is the change of intensity levels in different locations of an image due to the practical problems related with MRI hardware. The inhomogeneity formed by the receiver and RF excitation coils is the main reason of bias field. On the other hand, segmentation in the sub-cortical region of the brain is usually not satisfactory for the algorithms that segment the brain into only 3 tissues, CSF, GM and WM. The anatomical complexity of the sub-cortical region is the major reason for the poor performance. The tissue properties of some sub-cortical GM structures are different than the cortical GM. GM structures like putamen, thalamus and globus pallidus contain WM fibers, with close intensities to WM intensities. Therefore, the contrast is lower and the boundaries between tissues are smooth. The segmentation of such structures becomes a real challenge for the algorithms that segment the whole brain into 3 tissues. Significant parts of these GM structures are segmented as WM.

There are also algorithms like [2] which segment to the brain into each anatomical structure separately. These approaches can detect the sub-cortical GM structures much better but they have some disadvantages. The algorithms are more complex and harder to implement. Their computational demand and running time is quite high (on the order of several hours). They often necessitate the usage of strong computers. Also, the incorporation of spatial priors can be a challenge for the robustness when images which have quite different anatomy than the training images are tried to be segmented.

1.1 Objective of The Study

The sub-cortical GM structures have particular functions in the brain. The volume, the shape or other properties of these structures can be used as diagnostic markers for some diseases [2]. Unfortunately, poor performance of the general purpose segmentation algorithms limits the utilization of segmentation in the sub-cortical region.

In this study, the aim is to develop an algorithm to improve the segmentation in sub-cortical region and to increase the detection of sub-cortical GM structures. The desired properties of such an algorithm can be listed as follows:

- Robustness to
    - Inter-subject anatomical variability
    - Bias field
    - Noise
    - Different scanners, sequences and sequence parameters
- Feasible computation time
- Applicability to images with any weighting
- Operability with multispectral images
- Improvement in the detection sub-cortical GM structures
- Better volumetric estimations

To realize these objectives, an evolutionary hierarchical region merging (EHRM) algorithm, which only operates in the sub-cortical region, is proposed. The algorithm does not necessitate any preprocessing steps and utilizes only intensity information. The texture information is also incorporated into the framework to control region merging. The algorithm does not incorporate any prior information

from training images, which makes the algorithm insensitive to inter-subject anatomical variability.

1.2 Outline of The Thesis

In the 2nd chapter, major approaches to the brain segmentation problem are presented. The FAST[1] and Freesurfer[2] approaches are analyzed in more detail because in our study, FAST is used for comparison of performance and Freesurfer is used as the gold standard for the real dataset for which the manual segmentation is not available. In Chapter 3, the proposed algorithm is explained in detail. Also the evolutionary segmentation algorithm proposed by Veenman et al. [3], which is the starting point for the proposed algorithm, is presented and discussed. The results are presented in Chapter 4, and operation of the algorithm is illustrated. The comparison of performance of EHRM with FAST on 3 datasets is also presented in this section. In Chapter 5, the conclusions are stated and a discussion on future work is presented.

CHAPTER 2

OVERVIEW ON AUTOMATIC SEGMENTATION OF BRAIN

Magnetic Resonance (MR) became a very popular and widely used modality for brain imaging starting from late 1980s due to its non-invasiveness, higher soft tissue contrast and higher spatial resolution. Segmentation is one of the most critical processing steps which enables the usage of rich information in MR images. Manual segmentation can form a highly reliable segmentation but it requires high labor and time. In addition, as spatial resolution increases with the development of MR technology, fully manual segmentation will become infeasible for many applications. As a result, the development of semi or fully automated segmentation algorithms is reinforced.

There are many approaches proposed in literature that handle the brain segmentation problem. Due to the enormous amount of literature that has accumulated over the last 20 years, it is very hard to find comprehensive review articles. In [4] and [5], useful reviews of this field can be found.

If the approaches are classified with an integrative view, two aspects of the algorithms can be considered: 1. purpose, 2. the computational approach.

In this section, a brief overview of the current brain segmentation approaches in terms of the stated 2 aspects is given. Then, 2 popular algorithms used for comparison in the current study are analyzed in more detail, which are FSL FAST[1], and the Freesurfer software sub-cortical segmentation algorithm[2].

2.1 Classification According to Purpose

The type of information that is desired from a brain image depends on the particular application. Therefore, the segmentation algorithms change accordingly with their purpose. While some approaches intend to classify the brain image into 3 tissue types, CSF, GM and WM, some others aim to detect specific anatomical structures separately. For instance, Chupin et al. [6] proposed an algorithm for the automatic segmentation of hipoocampus and amygdala, which are neighboring structures placed in the temporal lobe. The volumetric changes at these structures can be important indicators for Alzheimer's disease and epilepsy even at the beginning stages of the diseases. The method first constructs a probabilistic atlas from images with manually segmented hippocampus and amygdala structures. This probabilistic atlas provides the necessary spatial and intensity information for the construction of initial estimates of 2 structures after the registration of the atlas space to the subject's native space. Then the anatomical landmarks at the borders are iteratively updated to optimize the boundaries of the structures via smooth deformations. Specific algorithms are also proposed for the segmentation of brain tumors in [7] and [8]. These specific approaches are out of the scope of the current study, since our interest lies at the level of the whole brain.

2.1.1 Tissue Classification Algorithms

These algorithms aim to segment the brain into the 3 main type tissue types, CSF, GM and WM aiming to calculate the corresponding volumes. This type of volumetric information is associated with some diseases, abnormalities and aging. Segmentation into a few tissue types is relatively simpler than detection of specific structures because the usage of spatial information and nonlinear registration techniques is not necessitated. Therefore, these methods are usually easier to implement and faster to run.

The algorithms proposed by Zhang et al. [1], Wells et al. [9], Shattuck et al. [10], Yi et al. [11], Ashburner and Friston [12], Tohka et al. [13], Mayer and Greenspan [14], Xuan et al. [15], Kapur et al. [16], Shen et al. [17] and Greenspan et

al [18] are only a few approaches falling into this category. Different aspects of some of these methods are analyzed in the related section.

The approach proposed in this study falls in this category, aiming to segment the sub-cortical region into 3 tissues, with special emphasis on improving the detection of sub-cortical GM structures.

2.1.2 Structure Segmentation Algorithms

These approaches aim to segment the whole brain not just into 3 basic tissues but also to some specific anatomical structures of interest. Sub-cortical GM structures are the target structures most of the time because each have a different function. Changes in particular structures can be diagnostic tools for neurological diseases. Therefore, these structures must be separated to be able to conduct an individual analysis. The models falling into this category are quite complex because they should combine the registration and segmentation steps in a single framework. They are hard to implement and their computational complexity is higher.

Their performance is usually better than the algorithms segmenting the brain into 3 tissues because they incorporate more information to the segmentation. The approaches proposed by Pham and Prince [19], Patenaude et al. [20], Fischl et al. [2], Zhou and Rajapakse [21], Corso et al. [22], Tu et al.[23], Barra and Boire [24], Sabuncu et al. [25], Morra et al. [26] are some of such approaches for the segmentation of sub-cortical structures.

The information regarding the gyri and sulci on the cortical surface is crucial for some applications. Therefore, cortical segmentation algorithms are also proposed. For instance, Fischl et al. [27] proposed a cortical parcellation algorithm to segment the cortical structures with an anisotropic Markov Random Field (MRF) Model. This algorithm is also the basis for the parcellation method in the widely used Freesurfer software [28].

2.2 Classification According to Computational Approach

Many approaches from Image Processing, Pattern Recognition and other fields are utilized for automatic brain segmentation. A brief overview and some example studies are presented in this section.

2.2.1 Finite Mixture Models (FMM)

FMM's are one of the most popular approaches for the segmentation of 3 tissues. These approaches model the intensity histogram of the brain as a collection of 3 components, which represent the contribution of each tissue. It is assumed that each tissue has a probability density function and a weighting, which represents its volume in the brain. The probability density functions can be either modeled by some parametric distribution or by nonparametric distributions derived from the image itself or an atlas brain. Often, the tissues are modeled with a Gaussian distribution as in [1] and [12].

Ashburner and Friston [12] proposed a segmentation approach that is unified with the registration of atlas to incorporate spatial information. The tissues are modeled with Gaussian distributions. Starting with random initialization for the tissue parameters, the bias field, atlas deformation parameters and the tissue parameters are iteratively updated by keeping 2 of them constant during the update of the 3rd one. The bias field is modeled by a small number of parameters used as the coefficients of smooth basis functions. The deformations on the probabilistic atlas are also modeled with discrete cosine transform (DCT) basis functions [59]. The simplicity in these models makes the optimization process much easier. This algorithm is accessible as a part of the MATLAB based Statistical Parametric Mapping (SPM) software [29].

2.2.2 Fuzzy Logic

Fuzzy C means is utilized for the segmentation of brain into 3 tissues. Pham and Prince [19] proposed the adaptive fuzzy segmentation algorithm that incorporates the bias field into the fuzzy c means clustering algorithms. In the traditional fuzzy c means algorithm, the cost function is formulated as the sum of squared differences of voxels to the class centroids weighted with the membership for each class. The bias field is incorporated with a multiplicative model and also regularization terms are used to constrain the bias field as a smooth field. After the initial estimations of centroids are performed, the memberships of each voxel are calculated and then the bias field is estimated. With an iterative approach, the bias field and the memberships are updated to find the bias field and final membership levels of the voxels. Shen et al. [17] utilized the fuzzy c means algorithm with the incorporation of local neighborhood information. The intensity similarity and the distance of the neighbors, which regularize the smoothness of segmentation, are incorporated to the cost function as a multiplicative component. The weighting of these two local features are determined by an Artificial Neural Network.

Fuzzy logic is also used for sub-cortical structure segmentation in the method proposed by Zhou and Rajapakse [21]. Three types of information are extracted first from the manually segmented brains to form fuzzy maps of these features. These are intensity, spatial location and the relative location of sub-cortical structures. The fuzzy atlas maps are linearly registered to a new brain by a 12 parameter affine transform for the standardization of anatomical coordinates. Then the 3 fuzzy features of voxels are fused to a single membership value for each structure. A maximum membership classification and then thresholding are used to finalize the segmentation. The approach seems very similar to the Freesurfer approach to be explained in Section 2.4. The former uses fuzzy logic framework while the latter uses probability and parametric distributions utilizing the same features.

2.2.3 Graphical Models

Several graphical models are utilized in the brain segmentation problems. Markov Random Field (MRF) is very popular for the incorporation of neighborhood information. For instance, FAST [1] uses an isotropic MRF to impose a smoothness expectation on the voxel labels. In Freesurfer sub-cortical segmentation algorithm, MRF carries the information regarding the relative orientation of the anatomical structures.

A hierarchical graphical model is also proposed for the segmentation of sub-cortical structures by Corso et al. [22]. The merging of voxels and regions forms a hierarchy of regions in the image. Then the overall energy function dependent on the voxel intensity and local neighborhood information is minimized efficiently using iterations called graph shifts. The shift which decreases the energy most is performed and the effected parts are updated. This iterative produce provides an efficient way to find a feasible solution of the segmentation.

2.2.4 Morphological Image Processing

Some basic image processing tools are combined to provide a framework for brain tissue segmentation. In the study proposed by Xuan et al. [15], Canny edge detection is performed first. Then region growing is performed iteratively to obtain a supervoxel representation of the image. Then these small supervoxels are merged to form larger regions. The outputs of edge detection and region merging are joined to eliminate the false boundaries and correct the boundaries of regions.

2.2.5 Shape Based Methods

Shape based models try to detect the structures by the finding their boundary. FSL FIRST [20] is a recent approach that uses deformable models for sub-cortical structure segmentation. Initially, two steps of linear registrations are performed. The first registration is on the whole brain and the second one is restricted to the sub-cortical region derived from the atlas. Manually labeled images are used to model the properties and the variability of the surface structures. The quality of fit of the

surface to the image is maximized by iteratively deforming the surface. The method is accessible by operating the related function from the command line as a part of FSL.

2.2.6 Clustering Approaches

Top-down or bottom-up clustering approaches are utilized in brain segmentation. In the adaptive mean shift algorithm proposed by Mayer and Greenspan [14], the coordinate and intensity features of voxels are clustered. First, the density estimation is performed by a Parzen window Kernel density estimator, which is then used to find the dense regions of the feature space. The dense points are expected to represent the pieces of the brain. The number of regions is decreased with region merging dependent on the pairwise Mahalanobis distances. As the final step, a derivative of K means clustering is performed to obtain the 3 tissue segmentation.

Another interesting approach is proposed by Greenspan et al. [18]. The image is assumed to be composed of smooth pieces which can be modeled by Gaussians. Both intensity and spatial information is modeled are Gaussian distributions. The initialization is performed by intensity thresholding. Then connected component analysis produces the regions. The minimum ellipsoids covering the regions are determined. Then the ellipsoids are checked for homogeneity. The regions which cover other tissue voxels more than a predefined threshold and high number of outlier voxels are divided to smaller regions. This splitting is a K means clustering dependent only to the spatial information. When the splits are finished, the segmentation is obtained. The approach aims to provide smoothness and robustness to the noise.

2.3 FSL FAST Algorithm

FAST is an unsupervised algorithm which uses the Hidden Markov Random Field (HMRF) model. It is easily accessible as a part of the FSL software [30] and

segments the brain into 3 tissues. This method is analyzed in detail because the method proposed in the current study is compared with FAST.

2.3.1 The Image Model

The tissues are assumed to be generated from Gaussian distributions. There is no information about the tissue labels at the beginning. Therefore, the parameters of these distributions and the bias field are not known a priori. The image voxels are considered as random variables in a markov random field. The states of the random variables are the class labels and they are not directly observed. However, these states can be estimated through the intensity information in the image, which makes the model a Hidden Markov Random Field (HMRF) model.

The reason behind the utilization of Markov random field is the incorporation of neighborhood label information. The main assumption is that neighboring voxels tend to contain similar labels, derived from the smoothness of brain tissues. This assumption is quite reasonable for most brain structures, especially when the resolution is high. With the incorporation of neighborhood information, the likelihood of voxel labels are modeled as the multiplication of two terms: 1. an intensity based likelihood derived from the tissue parameters and 2. a probability term derived from the clique potentials between the neighboring voxel pairs. If the neighbors have different labels, the change in the clique potentials decreases the likelihood of that segmentation. Although such a segmentation cannot compensate for the neighboring label differences by high intensity likelihood terms, the segmentation is forced to be smoother. This kind of modulation is especially important for the segmentation of images with high noise.

The algorithm uses global class parameters for each tissue so the bias field should be removed during the segmentation. The bias removal method proposed by Guillemaud and Brady [31] is incorporated into the segmentation. In this method, the bias field is modeled as a multiplicative field. The likelihood of bias field is composed of 2 components. The first component is the likelihood derived from the image intensities which is a measure of the bias field to represent the variations in

different parts of the image. The second term is a regularization term which imposes smoothness constraint on the bias field.

2.3.2 Initialization of the Model

As the first step, the image is segmented into 3 classes neglecting the effect of bias field. A thresholding method proposed by Otsu [32] is used for the optimal initialization of the segmentation. This thresholding method aims to minimize the intraclass variances while maximizing the interclass variances. The Gaussian distribution parameters of the tissues are then calculated from the thresholded image. This initial segmentation is also utilized for the initial estimation of bias field distribution.

2.3.3 Optimization

After the initial labels and bias field distribution are obtained, the bias field removed image is constructed. Then the posteriori probability of the segmentation is maximized. The posterior probability of the segmentation is modeled as the multiplication of the posterior probability of individual voxels, assuming that the posterior probabilities of the voxels are independent. The maximization is conducted by updating the class labels of voxels. The Iterated Conditional Modes method proposed by Besag [33] is utilized to update the class labels. At the end of this maximization, the segmentation is expected to be smoother due to the contribution of clique potentials. After the labels are updated, the tissue parameters are updated with the local Maximum Likelihood estimations using the Expectation Maximization (EM) algorithm. The bias field is also updated by the EM algorithm to find its Maximum a Posteriori (MAP) estimate. By updating the class labels, tissue parameters and the bias field in an iterative manner, the algorithm converges to a local maximum point of the segmentation probability function.

The smoothness level of the segmentation can be adjusted by the weighting of the clique potentials. When the Signal-to-Noise Ratio (SNR) is high and the images are sharp, smoothness can be disadvantageous due to the elimination of

small anatomical details by the effect of clique potentials. However, in noisy images, the clique potentials play an important role in suppressing the effect of noise. The smoothness can be regulated by the related parameter if the fast command is operated from the command line.

2.3.4 Performance

The segmentations are quite smooth and especially successful in the cortex because there is usually sufficient contrast between the tissues. However, the segmentations of sub-cortical GM structures like thalamus and putamen are very poor. Main reasons for this are the modeling of all GM structures with a single Gaussian distribution and weak boundaries between GM and WM in the sub-cortical region. The intensity levels of sub-cortical GM structures are within the range of the cortical GM and WM intensity levels, therefore the intensity based likelihood is not much likely to contribute to the segmentation of these structures as GM. In addition, if the initial label estimates of these structures are WM after the thresholding, the bias field may be modeled accordingly to compensate the intensity levels of these structures to approach to the intensity levels of WM. The problem gets worse in such cases and the possibility of correcting WM labels into GM becomes almost impossible.

The first thresholding step is not robust, which can often produce unacceptable segmentation results. The quality of brain extraction is very critical at this point. The inclusion of non-brain tissue which can be regarded as a separate peak in the histogram is problematic for the initial thresholding step. Although this problem is not observed in the test images presented in this study, the case is observed in some cases where the peaks of tissue components are highly fused.

In terms of time, for a conventional scan with 1mm resolution cubic voxels, it takes about 10-15 minutes on a desktop with a 4GB RAM and a 4 GHz processor. In terms of flexibility, the algorithm also operates with multispectral images. The performance often improves as compared to the single weighted case because the

missing contrast information in one weighting can be compensated with the information from another weighting.

2.4 Freesurfer Software Sub-cortical Segmentation Algorithm

The method proposed by Fischl et al. [2] is a popular sub-cortical segmentation algorithm and is available in the widely known Freesurfer software, which is capable of doing different analyses, from cortical surface analysis to sub-cortical structure segmentation. The performance of the method is shown to be comparable with manual labeling and the volumetric derivations from the segmentations are shown to detect the differences with control and Alzheimer Disease groups in some sub-cortical structures and ventricles. Due to its successful segmentation, this method is used to form the gold standard segmentations for the real data where the manual segmentations are not available. This method also depends on a Markov Random Model as FSL FAST algorithm.

2.4.1 Construction of Probabilistic Atlas

The intensity information is itself not sufficient to discriminate the different anatomical structures because of the high overlaps in the intensity histogram. Therefore, spatial information is also necessary.

Manually segmented images are collected and aligned linearly to each other first because spatial information is useless when the anatomical spaces of images are different. Then for each voxel location, the probability of each structure is calculated by simply finding the percentages of observations of the structures in that anatomical coordinate. Also the intensity distributions for each class are modeled by Gaussian distributions at each voxel location. The neighborhood information is also taken into consideration with the MRF framework. The anisotropic modeling of MRF enables to encode the information regarding the relative orientations of anatomical structures, which is quite consistent among different individuals. The utilization of relative orientations of structures is very crucial for reducing the complexity of the problem and better representation of structures.

2.4.2 Formulation of the Optimization Problem

The objective function is formulated as a Bayesian probability of the segmentation labels for a given image. Three kinds of information determine the probability: the intensity likelihood of a voxel for a structure, the spatial likelihood that determines the probability of a class being in a certain atlas coordinate and the label relationships between the 1st order neighbor voxels with direction information. The incorporation of higher order neighborhoods would have been preferred but this is regarded as infeasible due to memory requirements.

2.4.3 Linear Registration

The atlas should be aligned with the new image before voxel labeling can start. To minimize the effect of the anatomical differences between the atlas and the new brain, the points of the atlas where the prior probability of a tissue is higher than 0.9 are sampled first. The error function minimization is only conducted for the error of these sample points. A 12 parameter affine transformation is conducted with using a Davidson – Fletcher – Powell numerical optimization. When the optimum transformation is reached, a final regularization including all samples of the images is performed to increase the match between the anatomic structure borders of 2 images. The method is later improved with a structure specific registration procedure [34]. Each structure is registered linearly to the atlas separately so that the intensity standardization is conducted pairwise, rather than a global intensity standardization. This is intended to compensate the changes in intensity characteristics at different scanners. It is shown that the robustness of the algorithm on segmentation of images acquired at different types of scanners from the training images are significantly improved.

2.4.4 Initialization

No label information is present at the beginning. Therefore, the initial labels are derived by maximizing the multiplication of individual spatial and intensity

likelihoods of voxels, neglecting the anisotropic MRF term. All voxels are independently labeled in an efficient way.

2.4.5 Optimization

The optimization strategy is similar to the FSL FAST's strategy. After initial labels are available, the probability of the whole segmentation is maximized through the update of necessary labels with the Iterated Conditional Modes algorithm [33]. High number of labels can be regarded as a serious challenge for optimization but the number of possible anatomical classes in a certain coordinate is often limited. Therefore the optimization is tractable. At the end of iterations, the labels converge and the segmentation is obtained. Small checks are performed to finalize the segmentation, by checking violation of neighborhood information and reassigning the small structures which are disconnected from the main body of that structure.

2.4.6 Performance

The algorithm is expected to operate very successfully for the subjects that are similar with the training set of the atlas. Even in the absence of sufficient contrast and in the presence of high noise levels, the structures can be segmented quite successfully with the incorporation of rich prior information. However, the performance of the algorithm on subjects having significantly altered anatomy due to diseases is questionable because spatial priors can reduce the robustness of the algorithm for such cases.

The computational demand of this algorithm is inevitably high. In the Freesurfer manual, it is reported that the segmentation step can last up to 15 hours [35]. For the test images having 1mm cubic voxels, the computation takes about 3 hours in a on a desktop with a 4GB RAM and a 4 GHz processor.

In this section we have presented a restricted segmentation survey suitable for our target application. We chose to present methods selectively regarding to their wide use in the neuroimaging community. In addition, among the methods that we presented, the raw intensity information is used rather than a sophisticated feature

vector derived from multi-spectral or higher-level intensity-based data such as gradient or Laplacian. Needless to say, there is a plethora of methods for brain segmentation. Curious readers are referred to [4] or [5] for a wider review.

CHAPTER 3

EHRM ALGORITHM

Region based segmentation is one of the major segmentation approaches in image processing literature. In these approaches, each region is modeled as a group of interconnected pixels. These regions are formed directly by utilizing region features, differing from approaches depending on the detection of boundary between regions. Region shrinking, region growing, region splitting and region merging are some of the major strategies for obtaining the resultant regions.

In region shrinking, a target region is obtained by the iterative elimination of boundary pixels depending on some criteria. It seems less common compared to other methods. In a study [36], the region shrinking is utilized in a motion segmentation algorithm. On the other hand, region growing algorithms start with some initial seed regions or points. Then neighboring regions or points are joined to the regions subject to some criteria. This approach and its derivatives are widely used in biomedical applications, like automatic abdominal MRI segmentation [37]. The seed selection step may be either manual or automatic.

Region splitting algorithms aim to obtain the target regions by iteratively dividing larger regions into smaller regions by a split criterion like the one proposed in [38]. The region merging algorithms work in the opposite way by iteratively merging regions to obtain the final regions based on a merging criterion. A study utilizing the textural information of regions is presented in [39]. Split and merge techniques can be used together as presented in [40].

In the current study, the goal is to segment the sub-cortical brain region into 3 tissue types, CSF, GM and WM. There exists several disconnected regions belonging to the same tissue type. Also, whole sub-cortical region is to be segmented, rather than some portion of the image as in tumor or lesion segmentation. In addition, the sizes of the individual regions may change dramatically, from very small regions having a few voxels to regions having in the order of 50000 – 100000 voxels in a conventional MR scan having 1mm x 1mm x 1mm resolution. In such conditions, it is very hard to utilize region shrinking or growing approaches because they are more suitable when the number of regions is relatively lower. Manual seed selection is not practical. In addition, automatic selection of seed points is not straightforward but problematic when the number of regions is high and many small regions are present.

The presence of many small regions and the high volume of 3-D data of anatomic images are important problems in the implementation of region splitting algorithms in the current situation. On the other hand, region merging seems as a more suitable way to deal with the problem because it does not require the selection of any seed points and many regions having quite different sizes, arbitrary shapes and orientations can be implemented in a computationally more efficient way.

The method proposed by Veenman et al. [3] provided an inspiration for the method proposed in this study. The regions are formed with operations similar to region merging transferring border elements in an evolutionary framework. Region merging is performed according to the joint variance resulting from a possible merging of the two neighboring regions under the maximum variance threshold constraint. As the second operation, border pixel transfer aims to improve the region boundaries. In this thesis, the evolutionary region algorithm [3] is modified to operate in the sub-cortical region in a computationally efficient, fully automatic, generalized way with minimum assumptions.

The simple flowchart of the proposed method is shown in Figure 3.1. Initially, an atlas brain whose sub-cortical region is already defined is linearly

registered to the target brain. Therefore sub-cortical region of the target brain can be isolated. Evolutionary Hierarchical Region Merging (EHRM) approach is then utilized in a local framework by region merging operations only between neighboring regions. When the critical point is reached, where the regions are considered to reach a certain level of maturity, the local nature of EHRM is turned to a global framework, allowing region merging operations also between disconnected regions. When EHRM is terminated, the regions are classified into 3 tissues, CSF, GM and WM, forming the final segmentation.

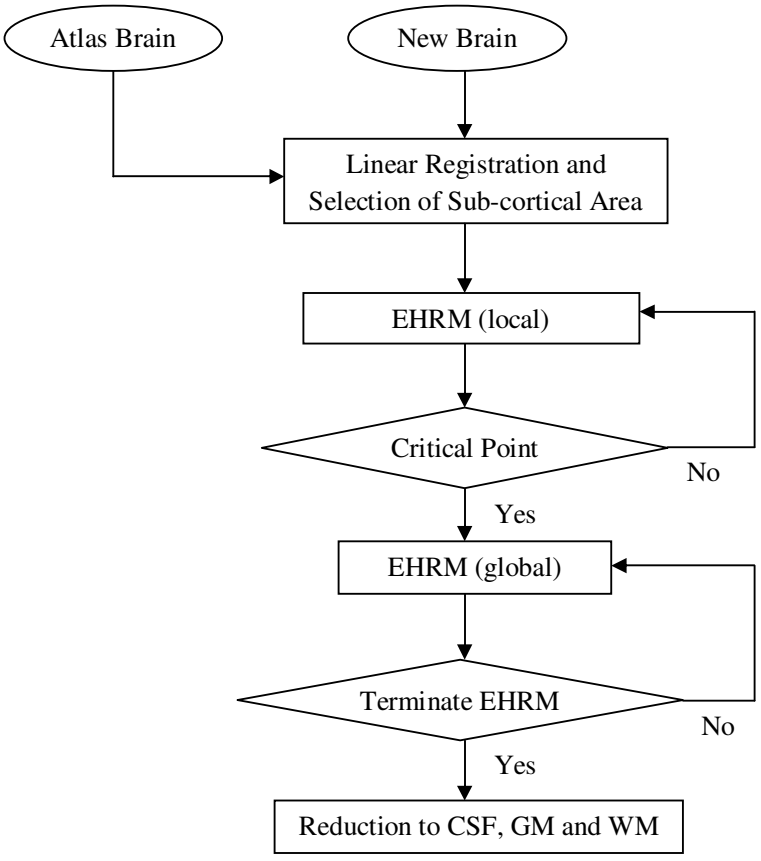


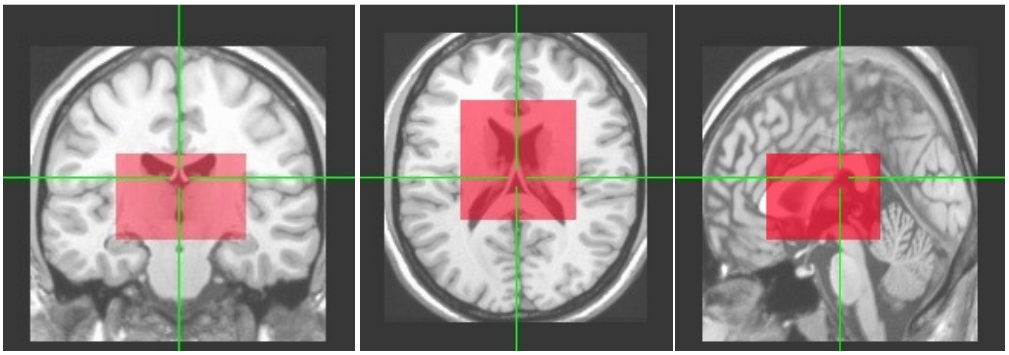
Figure 3.1 Flowchart of the proposed algorithm.

In this chapter, the selection of sub-cortical area is explained first. Then the evolutionary method proposed in [3] is presented briefly. The novel EHRM

approach for conducting sub-cortical region segmentation is proposed and explained in detail. At the end, the final step reducing the regions to 3 tissues is discussed.

3.1 Selection of Sub-cortical Brain Region

The well-known ICBM template [41] is used for defining the region containing the sub-cortical structures. A rectangular prism is roughly defined to include these structures by leaving some margin for the anatomic variability among subjects. The target structures are sub-cortical GM structures, putamen, thalamus, caudate, globus pallidus, etc, which are segmented poorly by most algorithms. It is expected that this region includes all parts of the structures of interest. The defined region can be seen in Figure 3.2.



(a) (b) (c)

Figure 3.2 The sub-cortical region in 3 planes on the ICBM template.(a) a coronal slice, (b) an axial slice, (c) a sagittal slice.

When a new brain is to be segmented, the ICBM template is first aligned linearly to this brain. The linear registration is performed by the FLIRT tool [42] of FSL software. FLIRT performs a 12 parameter affine transformation on the image. The defined sub-cortical region is transformed with the same parameters so that it can also represent the sub-cortical region of the new brain. Visual investigation across all test brains, indicate that all of the structures of interest lie in the defined sub-cortical region. The bounding box which defines the extracted sub-cortical area is defined as follows:

- The anterior limit is set as the most anterior part of anterior horn of corpus callosum.
- The posterior margin is the posterior of 4th ventricle.
- The left and right margins are determined by the most lateral parts of putamen. Few more slices (2-3 slices) are added to compensate inter-subject anatomical variability.
- The inferior limits are determined by the most inferior parts of putamen and globus pallidus. Similarly, few more slices are added.
- The superior margin is determined according to the most superior part of caudate. Similarly, few more slices are also added.

3.2 A Cellular Coevolutionary Algorithm for Image Segmentation

The algorithm (CCA) was proposed [3] to form a generalized framework for image segmentation especially when the number of clusters in the image is unknown a priori. Homogeneous regions are aimed to be formed where the level of segmentation is controlled by the user-set variance threshold parameter which is the only user-set parameter. The segmentation task was formulated as an optimization problem as:

$$C_{opt} = \min_{C} \sum_{c_i \in C} \text{var}iance(c_i) x (\#\_of\_pixels(c_i)) \quad (3.1)$$

subject to the constraints

$$\forall c_i : \quad \text{var}iance(c_i) < \sigma_{threshold}^2 \quad (3.2)$$

$$\forall c_i, c_j, i \neq j \quad \text{var}iance(c_i \cup c_j) \geq \sigma_{threshold}^2 \quad (3.3)$$

where C represents the set of all clusters c_i in the image. The objective is to obtain the set of clusters minimizing (3.1) subject to (3.2) and (3.3).

Each pixel start as a region at the beginning and a set of operations between regions are defined to optimize the segmentation. The regions try to perform these

operations with their neighbor regions in an evolutionary framework which decreases the computational load while giving the chance of escape from local minimum points of the segmentation problem. In the following sections, the evolution process and the operations between regions are explained. Lastly, the advantages and disadvantages of the approach are examined.

3.2.1 Evolution Process

The operation of the algorithm is divided into separate parts called “epoch” analogous to “iteration” in optimization problems. In 1 epoch:

- Each region has a chance of becoming active to try operations with its neighbors.
- If the region is active in an epoch, it first tries to perform “region merge” operation. If it cannot perform, it tries “border element transfer”. If the region can do one of these operations, the necessary updates are done.
- The evolutionary nature of the algorithm is implemented within the probability of a region for becoming active in an epoch.
- The activation probability of a region is determined by the success of that region on performing operations in a certain number of preceding trials. It is assumed that a region which is not able to perform operations in the last trials has approached to its natural borders. The activation probability of these regions should be decreased to prevent unnecessary trials which increase the computational cost. On the other hand, the regions which can perform operations with a high success rate are assumed to be in an intermediate state. The activation probabilities should be increased to fasten the convergence of these regions to their final state.
- In an epoch, a region can become active if its activation probability is larger than a random activation threshold sampled from a uniform distribution between [0,1]. The activation probabilities of regions have a basal probability even when no successful operations are performed in the last trials. The presence of basal probability gives the algorithm the chance of

escaping local minimum points of the segmentation. The activation probability can be formulated as

$$P_{act} = (1 - P_{basal}) \times P_{succ\_avg} + P_{basal} \quad (3.4)$$

where P_{act}, P_{basal}, P_{succ_avg} represent activation probability, basal probability of activation and the success ratio at a predefined number of last trials respectively.

- The operation of the algorithm may be terminated either when no region can perform any operation for a certain number of epochs or when the total number of epochs reach a predefined number of epochs.

3.2.2 Region Merge

If a region is active in an epoch, it first tries "Region Merge" with its neighbors. The variance of each possible each merge is calculated. The one with lowest variance which also satisfy the variance constraint in (3.2) is performed. The region merge is illustrated in Figure 3.3 and formulated in (3.5) and (3.6).

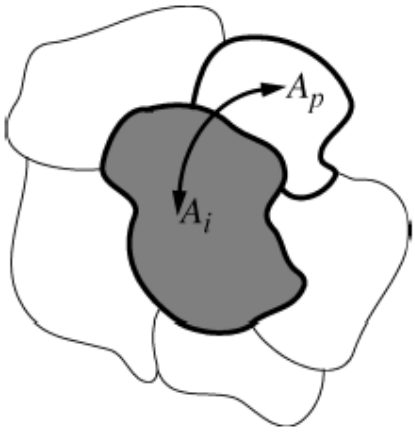


Figure 3.3 Illustration of "region merge" for regions A_i and A_p. The variance of each possible merge is calculated. Figure is taken from [3].

$$\text{Re}\,gion\_Merge = (c_i \cup (\min_{c_j}(\text{var}\,iance(c_i \cup c_j)))) \quad (3.5)$$

subject to

$$\text{var}iance(\text{Re}\,gion\_Merge) \le \sigma_{threshold}^{\,2} \qquad (3.6)$$

The region merges satisfying (3.5) and (3.6) are performed. Then, the neighborhood information of the regions in the affected area is updated.

If a region cannot perform “Region Merge”, it tries a “Border Element Transfer”.

3.2.3 Border Element Transfer

The aim of “Border Element Transfer” is to decrease the cost function (3.1) to be optimized by transferring pixels between neighboring regions. The pixels of neighboring regions which are also neighboring to the boundary pixels of the active region are candidates for “Border Element Transfer”. Only some of these randomly selected points are selected for practical purposes regarding computational load. The active regions try to decrease the overall cost function by taking one of these pixels from their neighbors. The effect of possible pixel transfers to the cost function is calculated and the one which decreases the cost function most is performed. If there is none decreasing the function, “Border Element Transfer” is not performed. The operation is illustrated in Figure 3.4 and formulated as (3.7), (3.8) and (3.9).

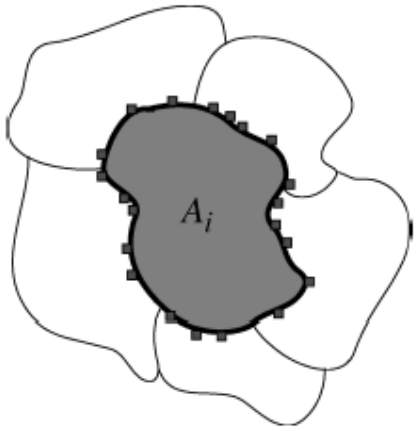


Figure 3.4 Illustration for “Border Element Transfer”. Some of the border pixels are sampled and tested for transfer to region A_i. Figure is taken from [3].

$$v_1 = (n_{c_i} + 1) \times \text{var}\,iance(c_i \cup \{p_j\}) + (n_{c_j} - 1) \times \text{var}\,iance(c_j - \{p_j\}) \quad (3.7)$$

$$v_2 = (n_{c_i}) \times \text{var}\,iance(c_i) + (n_{c_j}) \times \text{var}\,iance(c_j) \quad (3.8)$$

$$\Delta v = v_1 - v_2 \quad (3.9)$$

where n denotes the number of pixels in a region, p_j is a pixel that is tested for transfer to region A_i. If the change of this operation on the cost (Δv) is negative, the transfer is listed as a candidate. Among all candidates, the one which decreases the cost function most is performed. If there is none, the regions are not modified. The necessary updates on the regions are conducted if a transfer is performed.

After a "Border Element Transfer", two issues regarding the definition and constraints of the problem should be taken into consideration. A region is defined as a set of interconnected pixels. After a pixel is transferred from a region, that region divides into 2 regions if the pixel is a cut vertex. It is investigated whether a pixel is a cut vertex and if so 2 new regions are created replacing the old region.

After a transfer is performed, the variances of the regions alter. An additional operation on the regions which begin to violate the variance constraint in (3. 2) is done. To satisfy the constraint, the pixel of the region which has the feature with largest distance to the region mean feature is found. This pixel is removed from the region and a new region from this pixel is created. After the operation, the variance constraint is satisfied.

3.2.4 Advantages and Disadvantages of CCA

CCA aims to provide a general framework for image segmentation. The advantages of the algorithm can be listed as follows:

- The algorithm requires only one user-set parameter, the variance threshold. Although the problem of automatic or adaptive determination of the parameter is not solved in [3], it is favorable for an algorithm to keep the number of such parameters as low as possible. Each parameter introduces a

risk for automation and robustness. There are also some more parameters in the algorithm but up to a certain extent they are more related with computational efficiency, rather than the performance of the segmentation.

- The evolutionary framework is very useful for the efficient operation of such a region merging algorithm. Unnecessary trials are reduced.
- There is no requirement for preprocessing the image. Although preprocessing may improve the performance in many cases, it can also distort the image and its automation may introduce errors.

The algorithm and the proposed implementation [3] have the following limitations and disadvantages:

- The variance feature is used as the criterion for merging and the evaluation of within region similarity. Variance is a statistical measure that requires the calculation of the squared distance of each sample in the set to the mean value. Therefore, the cost of variance calculation increases linearly with the number of pixels in the region. Although the properties of images differ a lot, when regions get larger as epochs advance, the average number of neighbors of a region is expected to increase in general. Since the total number of pixels in the image remain constant, the total computation required for testing possible merges in one epoch is expected to increase unless the activation probabilities of regions is not decreased significantly.
- The "Border Element Transfer" operation samples the border pixels and calculates the change in the variances of regions to capture whether there is a decrease in the cost function. When the regions get larger, the number of regions decreases and the number of pixels in a region increase. At the same, the number of border pixels in that region also increases, but with a slower rate roughly proportional to the square root of the increase in the number of pixels when we consider the number of pixels and the number of border pixels analogical with area and circumference. This assumption is valid in 2 dimensional images and realistic when the borders of regions are smooth and simple.

The number of trials decreases accordingly but the increase in the sizes of regions increases the computation time of variance calculation. At this point, the border pixel sampling strategy plays an important role. If the number of sampled border pixels is proportional to the square root of the total number of border pixels as in the current implementation [3], the total computational load of the trials for one region increases roughly linearly with pixel number. Due to the linear decrease of region number with average region size, the computational cost of border element transfer stays similar as epochs advance.

- The variance measure is not capable of representing a statistical difference between two regions. Therefore some very large homogeneous regions can merge with relatively smaller regions having different intensity characteristics than the large region because their union obeys the variance constraint. This type of measure restricts the possible sizes of regions. Small regions have a risk of joining unnecessarily to other large regions. While the sizes of target regions are comparable in some applications, the region sizes may vary significantly in many applications.
- Variance measure is a very sensitive measure to outliers. Especially, when the borders between regions are very sharp, the intensity of pixels in the border can be very different from the mean, so they dramatically affect the variance, causing a misleading representation of region variance. This situation imposes great problems for regions having a relatively high border to area ratio.

There are also important issues to be considered when the algorithm is to be extended for performing 3D image segmentation:

- If there is a region as a sphere in a 3D image, the number of voxels increases with the 3^{rd} power of the radius whereas the number of border voxels increases with the 2^{nd} power of the radius. Therefore, it is obvious that the ratio of border to the number of elements in a region increases compared to

the 2D case. Unless the border sampling strategy is changed, this increases the computational load of border element transfer operation, making the algorithm more greedy as epochs advance. It is fair to expect that the algorithm should become computationally more efficient as epochs advance because the algorithm carries higher level information with smaller number of regions. However, computational time increases if the 2D implementation is directly extended.

- Thin regions contain a large number of border voxels which is a great challenge for the utilization of variance information.
- Due to increase in the dimension, more complex neighbor structures can form and the number of trials for region merging is expected to increase which increases the computational load as epochs advance.

3.3 Issues Regarding Segmentation of Sub-cortical Region

The aim of the current study is to develop an algorithm that can provide a generalized and robust framework for sub-cortical region segmentation. The challenges and their possible solutions are discussed in this section.

3.3.1 Bias Field Inhomogeneity

Bias field inhomogeneity, also known as intensity inhomogeneity, is one of the most important problems for threshold based segmentation because the intensity levels within the same tissue type may differ significantly in the whole brain. The most prominent causes for this artifact are practical problems related with receiver and RF coils. The inhomogeneity in the receiver coil sensitivity and flip angle profile causes a spatially varying distribution of intensity levels in the image. Therefore, most widely known segmentation algorithms operating at whole brain try to remove bias field prior to final voxel classification. The main assumption of such algorithms is that the bias field is a spatially changing smooth field which can be modeled as a multiplicative or additive component. For instance, the bias field is modeled and updated iteratively together with the segmentation in the FAST

algorithm [1]. The reader can find a detailed review and analysis of bias field correction algorithms in [43].

When the sub-cortical region is considered, which is a relatively small part about 10-15% of the whole brain in the current study, the maximum possible intensity variability due to bias field is limited. In practice, sharper changes in bias field are more observed in regions closer to the coils in the outer part of the brain. Sub-cortical region is located at the inner part of the head, which is more distant to the RF and receiver coils compared to cortical regions. Therefore, sharper changes due to bias field in sub-cortical area are less expected.

In the current region merging algorithm, the regions are represented with their mean intensity. When the regions get larger, the effect of bias field on mean intensity of regions decreases because large regions contain voxels that are from a wider spatial profile, decreasing the effect of bias field by the averaging out voxel intensities. Also, the texture constraint may prevent the merging of regions that have different texture characteristics while having similar intensity mean due to bias field. Therefore, the bias field is not a serious problem for the proposed algorithm and its effects can be neglected in the sub-cortical area.

Some bias field removal algorithms are available in most widely known software packages like FSL and can be utilized prior to segmentation. However, the modeling of bias field as a slowly changing field may corrupt the texture property in the sub-cortical region because these smooth anatomical changes may be perceived as a part of bias field and removed. Therefore, the texture information derived from a bias field removed image may be misleading. There are very smooth boundaries, textural and anatomical changes in some parts of sub-cortical region, especially in thalamus and internal capsule. The output of the bias removal algorithms may corrupt such anatomical information and complicate the segmentation process.

3.3.2 Partial Volume Effects

Partial volume effect(PVE) is observed when more than one tissue types are present within a voxel. Therefore, it is generally observed in the boundaries between tissues. This type of voxels tend to have intensity values between the intensities of tissues they are composed of, depending on their proportion.

In a region merging algorithm, the mean intensity of the region may change significantly due to PVE especially when the border to volume ratio of a region is high. The intensities of PVE voxels are not informative for the intensity of a certain tissue type so it is better to eliminate the effect of these voxels for the determination of region statistics. If the regions in the image are large, these effects may be less prominent but for the segmentation of sub-cortical region, the target regions may be very small and most of the voxels of a region may be on the border. Therefore, the exclusion of PVE voxels becomes inevitable in the current problem.

3.3.3 Smooth Boundaries

The sub-cortical region has a very complex structure which contains highly interconnected structures of different tissue types. In some main sub-cortical GM structures like thalamus, putamen and globus pallidus, which have low contrast with WM, the boundaries are very smooth. If the region merging criterion is fixed in all stages of the algorithm, it should be large enough for the segmentation of regions while keeping over-segmentation at a considerable level. In this case, the algorithm may fail to preserve the smooth boundaries in the image especially when the noise level is significant.

The flow of the algorithm should be modified to preserve weak boundaries. The solution is to start with a more restrictive merging criterion. As the regions get larger, the criterion can be increased iteratively forming regions in a hierarchical manner. As the algorithm operates, the regions get larger and carry a higher level of information representing more meaningful regions of the anatomy.

3.3.4 Disjoint Structures of Same Tissue Type

The aim of the sub-cortical segmentation is to classify the voxels into 3 classes. There are structures which are composed of the same tissue type although there are spatially disconnected. If the merging of the regions is only performed with the neighboring regions, it is impossible to connect disjoint structures of the same tissue type. Therefore, the algorithm must also allow the merging of disconnected regions. Allowing this type of mergings from the beginning of the algorithm results in the clustering of intensities which does not impose any spatial connectivity of voxels. In this case, the output image is a intensity thresholded image similar to the output of histogram based approaches.

The stated problem can be solved with a hybrid approach. Up to a certain stage, the algorithm must merge the neighboring voxels of same tissue types. After a certain level of maturation is achieved in the formation of local regions, the regions should be allowed to merge with other similar regions removing the neighborhood criterion to merge.

3.3.5 Region Features

The intensity is the prominent feature for a region because it is a function of tissue parameters. Intensity can be a sufficient feature for the separation of regions which have high contrast but in the main structures of sub-cortical region, the contrast between GM and WM is relatively low. Therefore, the usage of intensity as the only feature may not be sufficient to prevent the merging of structures having low contrast.

Additional features can be incorporated into merging criterion in this case. These features should be informative and robust to variations in scanners, scan protocols, scan parameters, artifacts, noise and inter-subject variability of anatomy. A feature having an anatomical meaning is a good candidate for the necessary robust feature. It is also known that the sub-cortical GM structures contain various nuclei and WM in a dispersed manner, making their tissue properties different than the

cortical GM. The heterogeneity in the tissue of these structures can be an informative texture feature. If this texture feature can be extracted reliably, it can be valuable to control the merges, preventing the merging of regions having similar intensities but different texture properties.

For the efficient operation of the algorithm, the features should be easily calculated to test all possible merges efficiently. It would be very advantageous if the features of the union of two regions can be estimated by using the individual statistics of regions. Mean intensity is already such a feature that can be estimated with the number of voxels and mean intensity of individual regions.

3.3.6 Termination of Segmentation

Incorporating the texture information in region merging can be quite useful for the separation of structures having low contrast. However, as the threshold criterion for region merging is increased in the late stages of the algorithm, the risk of merging regions of different tissues increases. In the algorithm, a single threshold for merging is used to separate the tissue throughout the whole sub-cortical region. It is also known that the intensity level differences between different tissues may vary significantly. For instance, the sub-cortical GM has a low contrast with WM while GM has a relatively better contrast with CSF in general. In this case, using a merging threshold to obtain 3 regions at the end of segmentation has a high risk of merging regions of different tissues. Therefore, the increase in threshold must be limited or a certain minimum number of regions should be preserved to eliminate false merges. This results in over-segmentation of the image, having more than 3 regions.

The over-segmented image formed at the end of region merging process should then be reduced to 3 classes to obtain the final segmentation. This final classification must also be a robust approach.

3.4 EHRM

The developed algorithm, Evolutionary Hierarchical Region Merging (EHRM), aims to provide a generalized framework for sub-cortical region segmentation. The issues discussed in the preceding section are taken into account with a robust and computationally efficient approach. The details of the algorithm are explained in detail in this section.

3.4.1 Region Features

There are two features controlling the flow of region merging process. These are intensity and texture heterogeneity.

3.4.1.1 Intensity

As previously stated, the most informative feature of a region regarding its tissue type is the intensity. The intensity feature, denoted by "M", is defined as the mean intensity of all voxels in the region as:

$$M = \frac{1}{n} \sum_{v_i \in R} I(v_i) \qquad (3.10)$$

where n, v, R and I represent the number of voxels in the region, a voxel, all voxels of the region and intensity respectively.

In larger regions, the effect of PVE voxels in the M feature of the region may corrupt the intrinsic M of the region especially when the surface to volume ratio of the region is high. Therefore, the border voxels are discarded from the calculation of M. This type of elimination is only performed in intermediate and large regions. These types of regions have a number of voxels more than a certain threshold, denoted by "n_inter". In the default operation, the border voxels are eliminated if the number of interior voxels which are not a part of the boundary is above 25% of the n_inter parameter. For a scan with a conventional resolution (1x1x1mm), this number is about 50 voxels, which is expected to generate a statistically reliable M feature in the vicinity of border voxels.

3.4.1.2 Texture Heterogenity

Texture heterogeneity is expected to represent the differences in texture between the tissues having low contrast. The variance of intensity values of voxels in a region is one of the basic features that can be considered first. However there are some practical problems that can be listed as follows:

- The PVE voxels tend to have slightly different intensities than the mean intensity. Therefore, their effect on variance feature may dominate the feature, corrupting the intrinsic tissue heterogeneity. Also the surface to volume ratio has a strong effect on the parameter. The nature of boundaries in terms of smoothness also affects the variance.
- Some regions are larger and some are smaller. The spatial distance between voxels within the same large region may increase very much. Although such voxels are from the same tissue type, their intensity may change due to bias field inhomogeneity and regional intensity changes within a tissue. In the presence of such effects, larger regions inevitably have larger variance, which is a misleading texture heterogeneity feature.

The local image properties can provide a much more robust estimation of texture properties because they are less affected by bias field and regional intensity level changes within a tissue type. One of the most widely used texture features are known as the Haralick texture features proposed in [44] which are derived from the co-occurrence matrix. Although these 14 features carry rich textural information, the incorporation of so many features in the region merging criteria complicates the decision of region merging process. The separate testing of each feature is infeasible because it is very difficult to match a high number number of features even for the right and left of the same structure. All texture features can be reduced in a single feature but in this case the training of this feature would be another issue. Also the normalization and fusion of so many features is not straightforward to automate because each one carries a different type of information. Manually segmented images as training sets can be used to fuse these features and determine the texture

criterion but it is very hard to generalize these criteria due to variability in MR images like noise, bias field, resolution, sequence, sequence parameters, inter-subject anatomic differences, etc.

At this stage, a single robust feature which also carries anatomical information seems as the best way to represent and test texture heterogeneity in the tissues. Local image properties are used to realize this texture feature. The proposed feature is a modification of the angular second moment from Haralick texture features. Now, let us consider the 6 neighborhood system in a 3D image as shown in Figure 3.5.

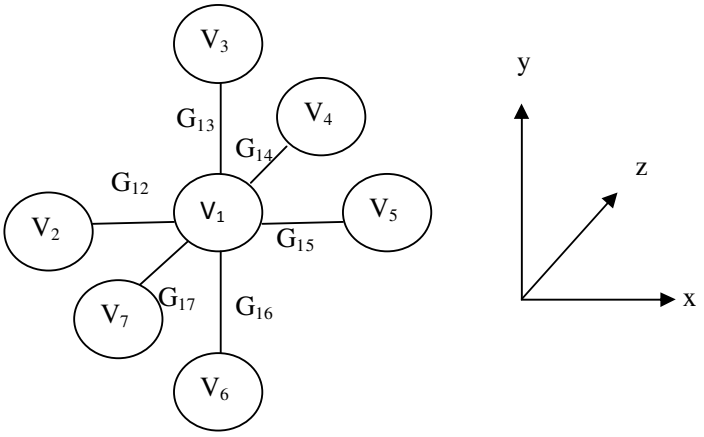


Figure 3.5 The neighborhood system in a 3D image. 2 neighboring voxels in each direction, resulting at 6 total neighbors. "V" denotes the voxel and "G" represents the gap between two voxels.

The gap "G" between two neighboring voxels is defined as the absolute intensity difference between these voxels as

$$G_{12} = \left| I(V_1) - I(V_2) \right| \qquad (3.11)$$

To calculate the text heterogeneity feature, all gaps which are defined between the interior voxels of a region are calculated and collected in a histogram. The effect of border voxels are removed from the calculation because they may have different intensities, forming gaps which have high values. This inevitably produces

a bias in the calculation of texture heterogeneity. It is expected the distribution of all gaps is a normal distribution. In fact, the normal distribution is one sided because the absolute value of the intensity differences between neighboring voxels are taken. The normal distribution is represented with its 2 parameters, the mean and the standard deviation (or variance). The peak of the distribution is expected to be 0 ideally because smaller gaps are more frequently observed whereas larger gaps are less frequent. If the histogram of gaps is a wider distribution, the variance of the modeled normal distribution is higher. Then it can be inferred that the texture heterogeneity increases as the gap distribution gets more distributed.

This measure of wideness of a distribution can be estimated by the variance but the direct variance calculation would be very sensitive to outliers in this case. Therefore, the feature should estimate the variance in a robust manner. A feature which reflects the sum of squared probabilities of the probability density function of the gaps can be used as measure of the wideness of the histogram, which can then be utilized as a heterogeneity feature for a large region. In heterogeneous regions, the neighboring voxels tend to have more variable intensities, which can be considered as a clue for local tissue heterogeneity, vice versa. The proposed texture feature, TH, is expressed in (3.12) and Figure 3.6 illustrates the calculation for a sample histogram.

$$TH = (1 / (\sum_{i=0}^{G_{max}} p(i)^2)) \qquad (3.12)$$

where G_{max} is the maximum gap in the region and the p represents the probability values at the probability density function. The TH feature is higher for wider histograms because the likelihood values are smaller due to the wide distribution, decreasing the squared sum. The situation is reverse for the narrow histogram. The proposed TH feature is expected to be robust to outliers in the histogram, which have small likelihood values. The calculated texture feature is roughly linearly proportional to the standard deviation for a normal distribution. The proof is in Appendix B.

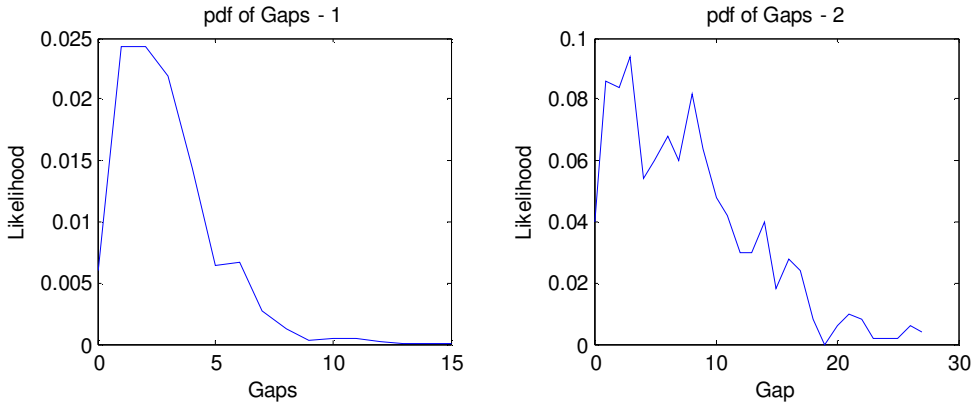


Figure 3.6 Calculation of texture heterogeneity feature. The probability density functions (pdf's) for 2 gap distributions which are sampled from a normal distribution of 5 and 10 standard deviation are plotted. The absolute values of the gaps are taken. The THs of the distributions are calculated as 9.1 and 16.8 respectively, roughly preserving the initial standard deviation ratios.

The variance of the gaps can also be calculated and used as the feature but it would be more sensitive large gaps in the region. Most of these large gaps may represent the effect of noise or transitions from strong boundaries rather than the anatomical texture information. Although border voxels are excluded, the transition effect of some strong boundaries may be realized also in interior voxels near to the boundary. It is expected that as the number of gaps in a region increase, the reliability of the texture information also increases. There is a gap number threshold for regions to be represented with their texture information and this threshold is discussed in the Region Models section. Figure 3.7 illustrates the gaps for texture analysis in a sample region in a 2D space.

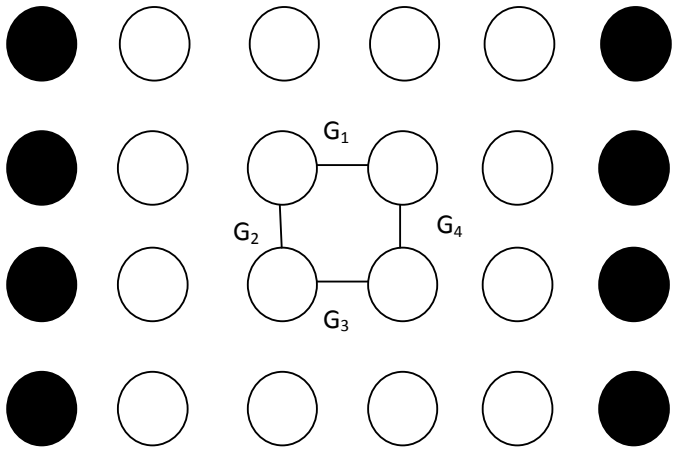


Figure 3.7 The illustration of gaps in a sample region. Black voxels belong to neighbor regions.

During the calculation of gap histogram, a uniform noise between [-1,1] interval is added to the gaps to prevent the possible biases that may result from the quantization of intensity to integer values. This operation is especially useful when the total steps of intensity values are relatively lower in the image.

3.4.2 Region Models

In the original algorithm proposed in [3], there is a single region model and every region is expected to obey the variance constraint. The variance constraint is tested for determining legitimate regions. So all regions have the same duties and roles regardless of their size and the level of information they carry.

In the problem of sub-cortical region segmentation, the sizes of regions and the proportions of tissues vary significantly: a small region may only represent a small portion of an anatomical structure whereas a larger region may represent a major part of a structure. Therefore, the anatomic regions are separated into 3 categories according to the level of information: small regions, intermediate regions and large regions.

3.4.2.1 Small Regions

Small regions are regions with number of voxels below a certain threshold. This threshold may depend significantly on the resolution of the MR image, therefore the threshold voxel number should be better defined as a ratio of all voxels in the sub-cortical region. This ratio value is one of the parameters of the algorithm, denoted as "vol_rat". The default value of this parameter is 0.1%. In practice where the sub-cortical region analyzed has a volume around 200mm^3, the volume threshold is about 0.2mm^3, which is significantly smaller than the volume of all main sub-cortical structures. The reader can find a detailed volumetric analysis of many brain structures in [45] and [46]. Small regions are only represented with their intensity values.

3.4.2.2 Intermediate Regions

Intermediate regions have a number of voxels more than the threshold defined above. Also it is checked whether reliable texture information can be derived from an intermediate region. When an intermediate region also carries reliable texture information, it is named as a large region. The determination of whether an intermediate region is also a large region or not is conducted in the following way:

- The texture feature of the region is calculated. If the number of gaps in the image is larger than a threshold, the region can be represented with its texture feature and is called a large region. Otherwise, it stays as an intermediate region. In the default operation of the algorithm, this threshold is kept the same with the threshold of voxels to be an intermediate region for simplicity and the reduction of parameter number. This assumption is realistic because the voxel number and gap number both increase as the regions get larger.
- The intermediate region is represented with the mean of all its voxels in the beginning. During texture extraction, the number of interior voxels is calculated. If they are above a certain limit, they are used for mean

calculation, discarding the effect of border voxels. 25% of the voxel threshold to be an intermediate region is set to be the necessary number of interior voxels for the elimination of border voxels, which corresponds to a number about 50 voxels for a conventional scan of 1x1x1mm resolution.

3.4.2.3 Large Regions

Large regions are intermediate regions which can also be represented with texture information. The criterion to be a large region is discussed in the preceding section. Large regions are generally formed by the tissues with a significant thickness. These are expected to be the regions where significant anatomical structures are formed.

3.4.3 Region Merging Process

In our study, only region merging among the operations proposed in [3], is utilized. The border element transfer operation is not used because it can be computationally very greedy in a 3D image segmentation problem. Also the hierarchical framework for the merging criterion is expected to decrease inappropriate merges, decreasing the need for border element transfer significantly. At the end of epochs, the following operations are conducted:

- The activation probabilities of regions are updated.
- The updated regions in the last epoch are checked to find out whether new intermediate and large regions are formed.
- After reaching the critical point texture feature standardization is performed.

Each voxel starts as a region at the beginning of segmentation. Up to a certain point that is called the critical point of the algorithm:

- Regions are only allowed for merging with their neighbors.
- If one of the merging regions is a small region, the threshold denoted by “thres_1” is used as the criterion. If both regions are intermediate or large, the merging is governed by a more restrictive threshold denoted by

"thres_2". These thresholds are explained in the Region Merging Criteria Section.

- At the critical point of the algorithms, the regions are expected to reach a certain level of maturity.
- Both thres_1 and thres_2 are increased as the epochs advance. At the beginning of the algorithm, a threshold limit denoted by "thres_limit" is calculated from the histogram of the image.

thres_1 becomes equal to the thres_limit as epochs advances. At this point, thres_1 and the thres_2 are kept constant for some epochs to maturate the regions. The volume of sub-cortical region possessed by intermediate regions is calculated. When the regions maturate, the rate of region merging decreases. The critical point is accepted as the epoch where the volume increase in all intermediate regions compared to the previous epoch is lower than the volume of the threshold for being an intermediate region. This is an indicator that no new intermediate regions are formed. Therefore, it is assumed that a certain degree of maturity is achieved.

After the critical point is reached:

- Large regions are allowed to perform merges with other intermediate and large regions that are not their neighbors.
- thres_1 is limited to "thres_limit" and "thres_2" is continued to increase.
- The merges are continued until one of the termination conditions for the algorithm is satisfied. Termination is explained in the related section.

3.4.4 Evolution

The evolutionary framework of the original algorithm [3] is used with the incorporation of neighborhood intensity information in addition to the success parameter.

3.4.4.1 Neighborhood Intensity Information

In an image having strong boundaries, intensity based region merging algorithms can be quite successful by following the edges between regions. However, when contrast is lower and significant noise is also present in the image, it becomes challenging to track the boundaries because regions of different tissues may have quite similar intensity. Some measures should be taken to prevent merging of such inappropriate regions. Although some commonly used edge detection algorithms like Canny edge detector [47] uses Gaussian windows or other approaches for smoothing the image prior to the detection and tracing the boundary voxels to improve their performance, smoothing operation has its own risks. The automation and the effect of smoothing depend on the application. Stronger and simple shaped boundaries can be detected quite successfully but the situation is worse in weaker and complex shaped boundaries which are among the greatest challenges of sub-cortical segmentation. In addition, the smoothing operation has a strong potential to distort the texture information in the image, which is used as a second criterion to measure region similarity prior to region merging. Therefore, smoothing operations are not suitable for the current problem.

Most of the problems regarding region merging are expected in the edges where boundaries may be weak and the PVE is also present. The prevention of region merging of such boundary voxels can eliminate the problems. Although we do not know the exact edges of an image prior to segmentation, we can derive some features which can give clues about whether a voxel or a region may be on a boundary. At the beginning of the algorithm where each voxel is also a region, regions on the boundary generally have neighbors which have quite different intensities because intensity transitions on borders are expected. In other words, the gaps between the neighboring voxels are expected to be larger. The average of gaps of a voxel may be a measure reflecting the probability of a voxel to be on the boundary.

As the epochs advance, the regions get larger and the average of gaps feature should be extended so that a similar measure for regions can be utilized. Therefore, the average of the intensity differences between a region and its neighbors can be calculated. For smaller regions, this measure is meaningful but not representative for large regions which have achieved a certain level of maturity because they inevitably become adjacent to regions of other tissue. The feature can be a regarded as a derivative of gradient, therefore its probabilistic effect is referred as "$P_{gradient}$", representing that it is the gradient based probability component of activation. The intensity in different images may change a lot, so the effective value of this gradient feature should be selected with respect to the merging threshold, thres_1 because we are trying to modulate the activity of small regions.

For a region R, Pgradient is formulated in (3.13).

$$P_{gradient} = \exp((\frac{1}{n} \sum_{R_i \in N(R)} |I(R) - I(R_i)|) / thres\_1) \qquad (3.13)$$

where n is the number of neighbors of R, N is the set of neighbors, I is intensity.

3.4.4.2 Success Parameter

The definition of fitness, which is a measure of success, is as defined in the original paper [3]. The fitness is the average success of a region to perform a region merging operation in a predefined number of last trials. The fitness based probability Psucc_avg is calculated as in (3.14).

$$P_{succ\_avg} = \frac{1}{n_{tr}} \sum_{tr_i \in L} s(tr_i) \qquad (3.14)$$

where n_{tr} is the number of trials to be considered for fitness, tr_i is a trial and L is the set of last n_{tr} trials. The success function s is defined in (3.15):

$$s(tr) = \begin{cases} 1 & if \quad successful \\ 0 & if \quad unsuccessful \end{cases} \qquad (3.15)$$

In the current implementation, n_{tr} is 5. When the number of trials of a region is lower than 5, the average of present regions is taken into consideration as P_{succ_avg}.

3.4.4.3. Determination of Activation

In the default of the algorithm, it is adjusted that a region with a volume of 0.01% of the sub-cortical region volume (10% of size threshold for intermediate regions) is affected equally from gradient feature and the success parameter to become active in an epoch. The 0.01% of sub-cortical volume is about 20 voxels in a scan having 1mm cubic voxels. At the beginning of the algorithm, the feature should fully dominate the activation probability. This weighting between the current feature and the success parameter is exponentially regulated by the size of the region. The effect of this feature is decayed exponentially to 0 with the size of the region. The activation probability of a region can be formulated as in (3.16).

$$P_{act} = (1 - P_{basal}) \times (w \times P_{gradient} + (1 - w) \times P_{succ\_avg}) + P_{basal} \qquad (3.16)$$

where the probability terms P_{act}, P_{basal}, $P_{gradient}$ and P_{succ_avg} represent the activation, basal, gradient based and average success based probabilities respectively. w is the weighting between the Pgradient and Psucc_avg.

$$w = \exp(-(V_R - 1 \times V_{voxel}) / V_{0.01\%}) \qquad (3.17)$$

where V is volume. The volume of one voxel is subtracted from the volume of the region because when a region has 1 voxel at the beginning of the algorithm there is no success information of the region. As epochs advance, a region may still be composed of 1 voxel with unsuccessful merging trials. This success based information is neglected in such cases but it does not seem a big issue because such a voxel already has a low $P_{gradient}$ value due to its neighbors having significantly different intensities. The success parameter would also try to decrease the activation probability but it is expected to be low in any case. When the size of a region becomes 1/3 of the threshold for being an intermediate region, w becomes 0.1. If the region gets larger from this point, w decreases further. At the same time, the number

of neighbors of the region increases and the computational load for calculating $P_{gradient}$ also increases. Therefore, the effect of gradient feature is ignored when the size of the region becomes 1/3 of the intermediate region voxel threshold and only the success parameter affects the activation probability.

A different random number for each region is generated from a uniform distribution in the [0,1] interval. If the activation probability is larger the random number of the region, the region is allowed to try a region merging operation in that epoch. If not, the region is passive in that epoch. With a new epoch, each region is also randomly activated to try region merging operation. In the long run, regions which are more successful at their region merging trials are more active, vice versa. As some regions get maturate, they perform less merging because they achieve their natural borders. Becoming unsuccessful in the last trials cause these regions to be inactive. Success based control activation is important for the reducing unnecessary trials for region merging, which in turn improves the computational efficiency. The presence of basal probability decreases the probability of the algorithm to be stuck in local minimum points of the problem. The basal probability is 0.2 in this implementation so there exists at least one expected trial for any region in 5 epochs. At each epoch, a random order of regions is formed for the activation.

3.4.5 Region Merging Criteria

There are 2 different criteria for determining appropriate merges in the region merging process. These are intensity similarity and texture similarity criteria and they are discussed separately in the following sections. At the end, the selection process among the candidate merges is explained.

3.4.5.1 Intensity Similarity

As previously stated, the most prominent feature of a tissue is its intensity. Voxels of same tissue type tend to have similar intensities, vice versa. However, due to anatomical variability within the tissues and several problems that decrease the

contrast between tissues, the regions of different tissue type can have quite similar intensities. There are 2 intensity thresholds, denoted by “thres_1” and “thres_2”.

Intensity Threshold “thres_1”

This threshold is used as a criterion for merges when one or both of the regions are small. It is increased as the epochs advance up to a certain limit denoted by thres_limit. The initialization, limitation and update of thres_1 are explained in the following sections.

Initialization of thres_1

Although the anatomy of the subcortcial region is quite complex, the number of interior voxels is still significantly higher than the number of border voxels in high resolution anatomic scans. At the beginning stages of the algorithm, the merges performed by the interior voxels are more desired because the formation and preservation of the natural boundaries in the images are quite difficult when the border voxels are also performing merges with their neighbors. In the ideal case, larger regions should begin to form in the middle of a tissue distant to the boundaries. Then these large regions should merge with the border regions to form better representation of tissues and anatomic structures.

The regions merge with one of their neighbors which has the closest intensity that the threshold permits. Therefore, the initialization of the threshold should be related to the intensity difference of neighboring voxels. First, for each voxel, the absolute intensity differences to all its neighbors are calculated. The minimum differences of all individual voxels are selected and collected in a histogram. The initialization of thres_1 is performed using the information derived from this histogram. The straightforward and simple way for initialization is to determine a percentage allowing some of the candidate merges while rejecting others.

The selection of this percentage is not straightforward, it is rather conducted in an intuitive manner. However, there are a few issues to be considered during the selection. This percentage should be flexible enough to allow the merging of interior

voxels but also restrictive enough to prevent the merging of border voxels. The percentage can be very restrictive to allow low rate of merging to reduce inappropriate merges but in this case there is a risk of decreasing the activation probabilities of most voxels due to the initial activation function. The rate of merging is also to be low, decreasing the amount of progress in an epoch. For each region, activation probability update and activation trial are conducted at each epoch. In each epoch, if the number of regions is higher, the computational load of these standard operations increases. Therefore, the number of regions should be significantly decreased in the first few epochs for computational efficiency. In the current implementation, the percentage for thres_1 initialization denoted by "thres_1_init_perc"is selected as 75 to achieve a high success of merges.

The 75^{th} percentile of this histogram is calculated and set as the initial value for thres_1. thres_1_init, the initial value of thres_1, also determines the activation probabilities of all voxels at the beginning. Although the properties of images may differ a lot, it is observed that about half of all voxels have an activation probability larger than 0.5, which are more probable to be active in the first epoch. The success percentage of merges are about 80-85%, larger than 75%, which is expected because the border voxels which have a lower chance of performing merges are more deactivated as compared to interior voxels.

Restriction of thres_1

The number of target regions at the final segmentation is 3 but it is very hard to achieve the final segmentation at the end of region merging procedure as previously discussed. The intensity differences between the pairs of tissues may not be distributed homogeneously in the intensity spectrum in most cases. Due to the single global intensity threshold, there is high risk of merging two tissues into a single class and leaving a tissue in 2 classes even when the image quality is high. Therefore, thres_1 should be restricted to prevent such cases.

The determination of thres_1 is not straightforward. The algorithm should allow reducing the image into a feasible number of representative regions and

decrease the probability of merges between different tissues. In order to investigate this, an ideal image is considered. In this ideal image, the volumetric ratios of 3 tissues are the same and the intensity probability density functions of 3 tissues are well separated as in Figure 3.8. The mean intensities of tissues are 100, 200 and 300 for CSF, GM and WM respectively. The standard deviation is 20 for all classes. This kind of an image is an ideal scheme for many segmentation algorithms like the ones modeling the histogram as a mixture of Gaussians.

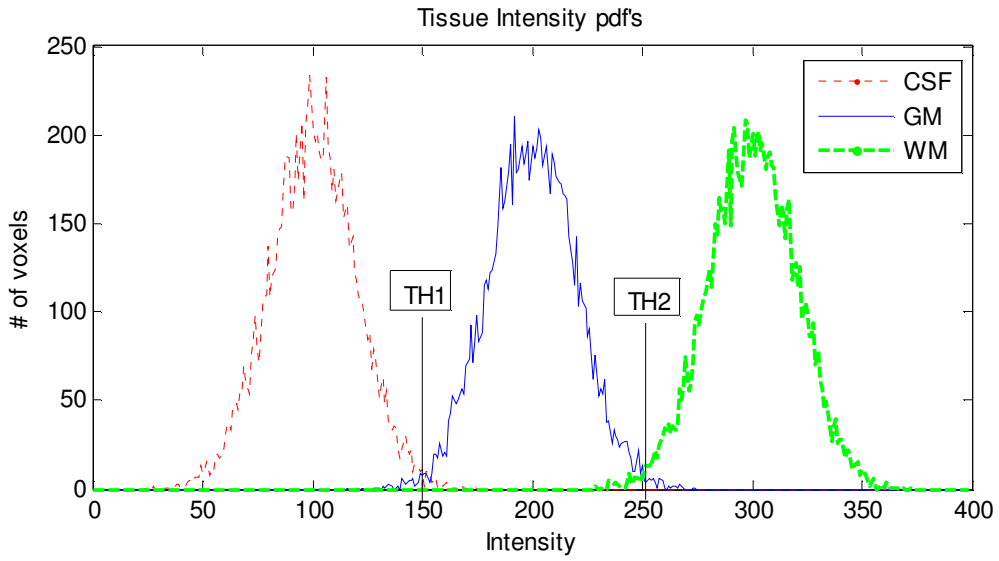


Figure 3.8 The probability density functions of 3 tissues in the ideal image stated.

In this image, TH1 and TH2 thresholds, which are in the middle of neighboring tissue means are the thresholds minimizing the misclassification rate for a possible threshold based voxel classification. If a region merging algorithm is utilized in this case, the restriction of thres_1 to the difference between to the tissue means to the nearest of TH1 and TH2 seems reasonable to reduce the inappropriate merges. This idea to determine a reasonable limitation to thres_1 can be implemented as follows:

- Calculate the 16.67^{th}, 50^{th} and 83.33^{th} percentiles of the histogram. These 3 values may be considered as a rough estimation for the mean values of 3 tissues. Denote these by m_1, m_2 and m_3.

- Then find the amount of 2 separations, s_{1-2} and s_{2-3}, between the 3 tissues.
- TH1 and TH2 are expected to be in the middle of tissue means. So the distance of TH1 and TH2 to the nearest tissue means is the half of the corresponding separations.
- Among the distances of TH1 and TH2 to the nearest tissue means, select the smallest one by dividing the smaller separation by 2. This value is the limit for thres_1.

This restriction is derived from the case where some assumptions are made for an ideal image. This limit should also be investigated for generalized cases. The volumetric ratios of tissues are not similar and the shape of probability density functions may change a lot in most cases. The pdf's can significantly overlap and the intensity levels of tissues may not be distributed homogeneously in the intensity spectrum.

When the volumetric ratios are not similar, this restriction is expected to be tighter because the minimum of the calculated separations becomes smaller than the difference of real mean values between 2 tissues. In the case of inhomogeneous intensity spectrum, which is frequently observed, the intensity levels of GM are much closer to WM than CSF. The current solution is robust to this problem, by taking using the minimum of the separations. When there is significant overlap between different tissues, the restriction becomes even tighter because the intensity levels get closer. In the stated 3 cases, the threshold choice seems to be safe to prevent the merging of different tissues having considerable contrast.

The threshold restriction can be regulated by a parameter if desired by the user. The regulation coefficient is denoted by "thres1_limit_coef" and linearly scales the histogram derived thres1_limit. In the default implementation, thres1_limit_coef is set to 1. The determination of thres1_limit is illustrated in a sample histogram as in Figure 3.9.

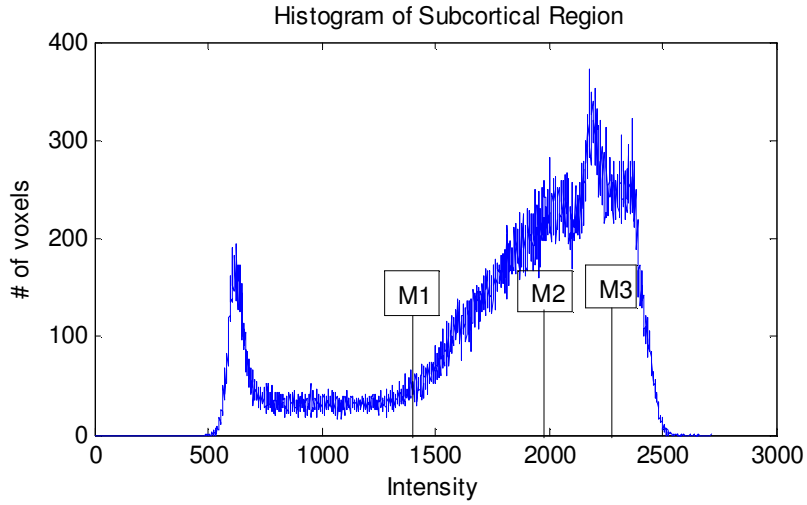


Figure 3.9 Histogram of sub-cortical region of a sample brain.

In Figure 3.9, the 16.67th, 50th and 83.33th percentiles are represented with M1, M2 and M3 respectively. These are the estimates of tissue means for the ideal stated before. The thres1_limit is calculated from M1, M2 and M3 as in Eq. (3.18).

$$thres1\_\lim it = \frac{1}{2}\min((M2 - M1),(M3 - M2)) \qquad (3.18)$$

Update of thres_1

The algorithm starts with the initial value of thres_1. As the region merge in the first few epochs, the success rate of merges "succ_merge" and the activation rates "act" of the regions decrease because the neighboring regions of a region tend to have more different intensities as the regions get larger. The succ_merge and act are defined in (3.19) and (3.20).

$$succ\_merge = n_{succ} / n_{active} \qquad (3.19)$$

$$act = n_{act} / n_{total} \qquad (3.20)$$

These two measures are defined for one epoch of the algorithm. The succ_merge is defined as the number of regions that performs region merging to the number of regions that are active in that epoch. The act parameter is the ratio of

active regions to the number of regions at the beginning of the epoch. These two measures carry information regarding the operation of the algorithm, indicating the rate of decrease in the total number of regions as the epoch advances. Therefore, a threshold called "succ_percentage" is applied on the succ_merge measure to perform the update of thres_1. After the update, thres_1 is increased, allowing new merges and increasing the activation probabilities of regions where neighborhood intensity information is taken into consideration.

About half of the active regions in an epoch are expected to initiate successful merges. By leaving some margin, the threshold for update is set to be 40% success in the current implementation. If succ_percentage is below 0.4, thres_1 is increased in an exponential manner. The total number of updates is defined according to the basal probability value for activation. When this probability is 0.2, each region is expected to perform at least 1 trial in 5 epochs. The number of updates, n_update is selected as 10, the twice of 5 epochs. Therefore, each region is expected to be active at least twice, decreasing the probability of the algorithm to be stuck in local minimum points. The update of thres_1 is expressed in (3.21) and (3.22).

$$thres\_1_{epoch+1} = \begin{cases} thres1_{epoch} \times thres1\_rate & if \quad succ\_merge < succ\_percentage \\ thres1_{epoch} & if \quad succ\_merge \geq succ\_percentage \end{cases} \tag{3.21}$$

$$thres1\_rate = \exp\left(\frac{1}{n\_update} \ln(thres1\_\lim it / thres\_1\_init)\right) \tag{3.22}$$

thres_1 is increased exponentially to thres1_limit according to the success of merges in the last epoch. When critical point is reached, large regions are allowed to merge with other large and intermediate in addition to their neighbors. After the critical point, thres_1 is kept constant.

Intensity Threshold "thres_2"

The intensity threshold thres_2 is used to determine the intensity similarity between either intermediate or large regions. This is a more restrictive threshold

than thres_1 to prohibit the inappropriate merges between larger regions. The initialization, limitation and update of thres_2 are discussed in the following sections.

Initialization of thres_2

There are no intermediate or large regions at the beginning and most likely in the first few epochs of the algorithm. Therefore thres_2 becomes necessary only when intermediate regions begin to form. Naturally, thres _2_init, the initial value of thres_2, is set to the initial value of thres_1, thres_1_init.

Limitation of thres_2

The final limit for thres_2 denoted by "thres_2_limit" is the same with the limit of thres_1. However, thres_2 is a more restrictive threshold due to its update strategy.

Update of thres_2

The update of thres_2 is conducted differently before and after the critical epoch of the algorithm. Up to the critical point, thres_2 is set to the average of the initial value of thres_1 and the current thres_1 value as expressed in (3.23).

$$thres\_2_{epoch} = \frac{1}{2}(thres\_1\_init + thres\_1_{epoch}) \qquad (3.23)$$

The reasoning behind this update strategy is that thres_2 should not begin with a smaller value than the initial value of thres_1 because it is not likely that the intensity differences between regions decreases as the regions get larger. thres_2 gets more flexible as the epochs advance but it is always more restrictive than thres_1 with a certain degree.

After critical point where the regions are expected to achieve a certain level of maturity, thres_1 is kept constant. The merging between 2 regions between 2 intermediate or large regions is more dominant in the algorithm. The ratio of allowable merges governed by thres_2 to all possible merge trials governed by

thres_2 is calculated. If this ratio denoted by "ratio_thres2" is very low, the thres_2 should be increased. The threshold for this ratio is selected as 10% because this ratio should be low in general to prevent merges of the regions of different tissues. Since there are 3 target tissues, if the ratio_thres2 is larger than 33.33%, this indicates a high risk of a merge between different tissues even in the ideal image where the volumetric ratios of tissues are the same and contrast is high. Therefore, 10% threshold is expected to operate the algorithm with appropriate merging, keeping ratio_thres2 between 10-20% in general. The algorithm is desired to perform the most favorable merges at each epoch, thereby reducing the accumulated errors to the final of the segmentation. The ratio_thres2 is calculated in (3.24).

$$ratio\_thres2 = \frac{tr2_{legal}}{tr2_{all}} \qquad (3.24)$$

A region first tests its intensity similarity to all the regions that it is allowed to merge. The results of each intensity similarity test governed by thres_2 are collected during an epoch. The total number of all trials "$tr2_{all}$" and the number of trials which satisfy the thres_2 constraint "$tr2_{legal}$" are calculated. Their ratio in (3.24) gives information about whether thres_2 should be updated. The update rule is expressed in (3.25), (3.26) and (3.27).

$$thres\_2_{epoch+1} = \begin{cases} thres\_2_{epoch} * thres\_2\_scale & if \quad ratio\_thres2 < 0.1 \\ thres\_2_{epoch} & if \quad ratio\_thres2 \geq 0.1 \end{cases} \qquad (3.25)$$

$$thres\_2\_scale = \exp\left(\frac{1}{n_{update}} \ln(thres\_2\_\lim it / thres\_2\_critical)\right) \qquad (3.26)$$

$$thres\_2\_critical = \frac{1}{2}(thres\_2\_init + thres\_2\_\lim it) \qquad (3.27)$$

The update of thres_2 after the critical point is conducted quite similarly with the update of thres_1 before the critical point. thres_2 is increased up to its limit if none of the termination conditions is satisfied.

3.4.5.2 Texture Similarity Criteria

When the critical epoch is reached, first all intermediate regions are checked to see whether they satisfy the criterion for being a large region. Then large regions are determined and they are given the permit of trying to merge with all other intermediate and large regions. The merging between 2 large regions is tested also according to texture, in addition to the intensity. It is desired that different tissues having low contrast can be discriminated according to their texture information. If the textures of 2 large regions are not similar, the merging is not allowed even if the intensity similarity criterion is satisfied.

The texture feature defined before is very sensitive to noise and other properties of the image like the intensity levels, the sharpness, etc. Therefore, the texture features of all large regions are first standardized to reduce the bias of additive noise in the image. Then these standardized features are eligible for testing texture similarity. The test is similar to the statistical hypothesis test for the comparison of the variances of two different distributions. The texture similarity test, texture feature standardization, the initialization and update of the texture similarity threshold "TH_thres" is explained in the following parts.

Texture Similarity Test

The texture feature TH is discussed to be a measure of standard deviation of the magnitude of gaps in a large region. Thus TH^2 can be used as the variance of magnitude of gaps. It is expected that tissues with similar texture should have similar TH^2, vice versa. In a brain image, some regions are quite homogeneous for conventional MR image resolutions. However, some tissues are quite heterogeneous at these resolutions. This heterogeneity may be regarded as an additive noise on the image but in fact it is a component of the anatomy, which should be regarded as a part of the image to be acquired. The heterogeneity of tissues lie in a spectrum, where the spectrum is affected by the resolution of image, sharpness, additive noise, selection of scan parameters and any preprocessing if performed.

TH^2 is a measure of the variance of a gap distribution, therefore variance equality hypothesis seems a reasonable candidate for testing texture heterogeneity similarity. When 2 large regions try to merge, first the thres_2 is used to test their intensity similarity. Then the F statistics which corresponds to the ratio of the larger TH^2 to the smaller TH^2 is calculated to obtain a measure of texture similarity as in (3.28).

$$F = \frac{TH^{2}{}_{larger}}{TH^{2}{}_{smaller}} \qquad (3.28)$$

In the statistical test, the table F statistics with the present degrees of freedom and significance level are found. Then the decision is made by comparing the test F statistics to the compared F statistics.

In the current problem, the TH^2 values are inevitably corrupted to a certain level due to a bias due to the artifacts, noise and other properties of the scan. Therefore, it is a rough measure of tissue heterogeneity. High number of gaps does not increase the reliability of the feature after a certain level because the feature is already corrupted. In this case, the statistical testing same with the F value does not seem appropriate. The calculated F statistics itself represents the amount of texture similarity.

The F statistics is compared with a threshold denoted by "TH_thres", which is valid for all tests in the same epoch. It is assumed the textures of regions are similar if F is lower than thres_var.

Texture Feature Standardization

The texture feature is standardization is necessary for the robustness of the algorithm especially to various levels of additive noise and sharpness. If a region is considered in an image, the measured variance of the intensities in the image can be modeled as in (3.29). This modeling assumes that the noise and the variability in the image are uncorrelated. The calculated variance is the sum of variances of the noise and real variability in the region.

$$\sigma^2_{calc} = \sigma^2_{noise} + \sigma^2_{real} \qquad (3.29)$$

When the noise level changes, the calculated variance of the region also changes. In the current problem, no information regarding the noise is present. Only the TH values of the large regions are known. It is sufficient to standardize these TH values among themselves because their ratios are important. In statistics, a constraint for performing some hypothesis tests is the homogeneity of variances among groups. The homogeneity of variance states the levels of variance of distributions should be similar. For the distributions which can be accepted to have sufficient number of samples (about 25-30), the F statistics is around 2 for the 0.05 significance level. This fact is also stated in [48] as a rule of thumb for determining the homogeneity of variance.

This rule can be utilized to make a standardization of texture features of large regions. A histogram from the TH^2 values of regions weighted with their volume is constructed. It is assumed that the ratio of the median to the minimum of this histogram should be adjusted to 2. With this strategy, the difference between the middle and the bottom of the texture heterogeneity feature is set to be in the border of significant difference.

A sample histogram is shown in Figure 3.10. The bottom and median values are observed. The TH features of all large regions are standardized as in (3.30) and (3.31).

$$TH_{add} = (TH^2_{median} - 2 \times TH^2_{bottom}) \qquad (3.30)$$

$$TH^2_{standardized} = TH^2_{calc} + TH_{add} \qquad (3.31)$$

After this standardization, the ratio of the middle and the bottom TH^2 is set to 2. In the current example TH^2_{median}=192.07, TH^2_{bottom}= 134.50 and TH_{add}=-76.91. It can be considered that the effect of noise is removed. This removal enables the texture features to spread to the texture spectrum more uniformly, allowing for discrimination between the textures of regions.

This update is repeated at the beginning of each epoch after the critical point. Ideally, this standardization is totally robust to additive noise.

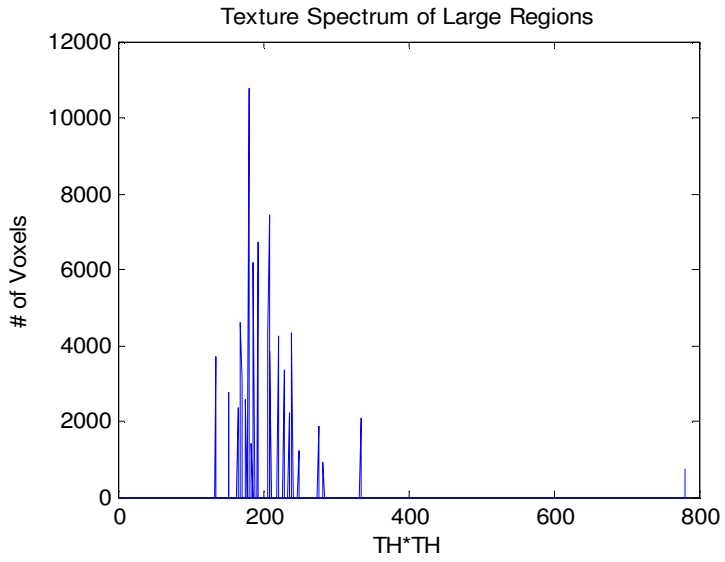


Figure 3.10 Texture spectrum of large regions.

TH_thres Initialization

The initial default value for TH_thres is set to 1.5, which is a quite restrictive at the beginning to reduce inappropriate merges.

The Update of TH_thres

If TH_thres is very restrictive, even the merges between the same tissues may be prevented. TH_thres is accordingly increased with a scalar multiplier coefficient if it is very restrictive. To measure the restrictiveness of TH_thres, the success of texture similarity tests are used similar to the success of thres_2. The ratio of success in texture tests are defined in (3.32).

$$ratio\_TH = \frac{TH\_trial_{legal}}{TH\_trial_{all}} \qquad (3.32)$$

It is necessary to note that texture test is performed for only the pair of large regions satisfying the intensity similarity coefficient. If the success ratio is lower

than a threshold, the TH_thres should be increased to allow more merges. This ratio is selected as 25%. If the tissues in the brain are divided roughly to 2 according to their texture, this ratio should be lower than 50% in general. A threshold of 25% seems appropriate as a safe threshold for the problem. The updates are conducted with 5% increases when necessary, which is an increment that makes TH_thres about 2 at the end of 10 updates. Maximum allowable value of TH is set to 2. The update rule is stated in (3.33).

$$TH\_thres_{epoch+1} = \begin{cases} TH\_thres_{epoch} * 1.03 & if \quad ratio\_TH < 0.2 \\ TH\_thres_{epoch} & if \quad ratio\_TH \geq 0.2 \end{cases} \tag{3.33}$$

If the total number of texture similarity test are smaller than 5, the reciprocal of 25%, no update is performed because it may not be right to decide an update with low number trials and insufficient statistics. Also inappropriate merges can be observed when the TH_thres is increased a lot.

3.4.5.3 The merging process

When one region tries to perform region merging with its neighbors, there may be more than legal 1 merge candidate. In this case, the most appropriate merge should be determined. A measure of cost "Cost_merge" is calculated for each legal merge. The Cost_merge is determined by the degree of intensity similarity of the 2 regions. Cost_merge for a candidate merge is equal to the ratio of the intensity difference of 2 regions to the threshold which governs the merging, either thres_1 or thres_2. Therefore, costs of all legal candidate merges are normalized to 1.

After the costs are calculated for each possible merge, the one with the lowest cost is performed. Since thres_1 is more flexible than thres_2, the merges containing small regions are encouraged. After the merging is decided, the necessary updates listed below should be performed:

- The mean intensities of regions are weighted with their volume to give the mean intensity of the resultant region.

- The voxels, volume, neighbors and the list containing the intensities of neighbor regions are updated. The similar information of the affected neighbors is also updated.
- If 2 large regions merge, the resultant TH^2 feature is calculated according the volume weighted sum of TH^2 of the regions. This is a rough assumption. At the end of the epoch, the exact value of TH^2 is calculated.
- If 1 of the 2 regions is a large region, the TH^2 of the large region is the feature of the new region. If the volume of the resulting of region exceeds the volume of the large region at the last update by 10%, the exact TH^2 is calculated at the end of the epoch. If not, texture feature is kept as the feature of the large region.
- If one of the regions is small and the other is intermediate, it is checked whether the number of gaps can exceed the threshold for texture calculation. Each new voxel can increase the number of gaps up to 3 so the number of maximum gaps that can be present in the region is calculated. The number of gaps and voxels of the intermediate region in the last texture calculation is used. If this number is below the threshold number of gaps, texture calculation is not performed. If there is a chance of satisfying the threshold, texture is searched at the end of the epoch and the regions become a large region if it exactly satisfies the criterion.
- The success history of regions is updated. This information is used to update the activation probabilities at the end of each epoch.

3.4.6 Termination Conditions

At the end of the algorithm, it is desired that the regions of the same tissue are merged as much as possible while the merges of regions from different tissues are prevented as much as possible. Therefore, 2 criteria are defined to finalize the region merging process. These are listed as follows:

1. A minimum number of large regions are left in the segmentation. This is a parameter denoted by “n_region” and set as 10 in the default

implementation. The number of large regions should reach 150% of the threshold (15) once so that this rule is active. This is a precaution for the cases where the number of large regions is about 10 or lower than 10 when the critical point is reached. If the number of large regions cannot reach 150% of the threshold, this rule becomes active when thres_2 reaches to its limit.

2. After thres_2 has reached to its limit, a predefined maximum number of epochs is allowed to proceed before the algorithm terminates. This number of epochs is also equal to the number of epochs which allow at least one activation for any region, 5 in the current implementation derived from the basal probability.

These 2 conditions are checked at the end of each epoch after the critical point. If one of the conditions are satisfied, the algorithm is terminated.

3.5 Reduction of Segmentation to Three Tissues

When the region merging process is terminated, the image contains a predefined number of large regions, 10 in the default implementation. However, the desired segmentation is composed of 3 tissues.

In most unsupervised brain segmentation algorithms, the brain histogram is assumed to be summation of distribution of 3 tissues. Most of the variability within a tissue is assumed to be due to the anatomy of the tissues. However, PVE and other image artifacts are the most prominent factors creating within tissue variability. For example, most sub-cortical GM structures have a different tissue composition compared to cortical GM, due to significant WM diffused homogeneously or heterogeneously within GM. This is the main cause of difference among different anatomic structures categorized as GM. In the sub-cortical region, the bias field is less effective so the intrinsic tissue intensities should be quite similar among different positions in the sub-cortical region. Therefore, the intrinsic mean values of the tissues are estimated first in the current approach. The intrinsic intensity term is

used to refer to the mean of the tissues which is composed of only one of the 3 tissue types.

In the PVE voxel modeling, the voxels are expected to have intensities weighted with the volume ratios and intrinsic intensities of the tissues composing the tissues. Therefore, the aim is to classify the regions to the tissues whose intrinsic intensity is closest to the intensity of the region.

The estimation of intrinsic tissue intensities is explained first. Then the determination of thresholds and the finalization of segmentation are presented.

3.5.1 Estimation of Intrinsic Tissue Intensity

In a typical T1 or T2 weighted brain image, the intensity levels of tissues are in an order. For a T1 image, WM has a higher intensity level than GM and GM has a higher intensity level than CSF even when the contrast is poor.

The tissues of the anatomic structures in the sub-cortical region consist of homogeneous parts and heterogeneous parts. The interior parts of lateral ventricles are expected to be a quite homogeneous part of CSF tissue. WM in the superior parts of sub-cortical regions surrounding the superiors parts of lateral ventricles are expected to be quite homogenous, composed of only WM tissue. The presence of such homogeneous regions can be useful during the calculation of intrinsic tissue intensities.

To implement this idea, first the histogram of the image formed by the region merging process is constructed. To prevent the increase in region number due to small regions, only the intermediate regions which have a volume ratio above 0.1% of all sub-cortical volume is filtered from the histogram.

The filtered histogram is divided into 3 tissues with 2 candidate thresholds. With the constraint that each tissue should have at least one region, all possible threshold combinations are listed. If 15 regions are present in the filtered histogram, the total number of combinations is 13x12/2=78. For each combination, the intrinsic

tissue characteristic of each tissue is estimated. The intensity of each region of a tissue is smoothed with Gaussian windows. This step aims to reduce the effect of accumulation of intensities into a single value. In a discrete histogram as in Fig 3.11, the median may not be so sensitive to the volumetric distribution of other regions of the same tissue. The standard deviation of this Gaussian kernel is as in (3.34). This standard deviation is derived from the threshold limit, which governs the merging. With this implementation, expected 95% of the voxels in a region are within the threshold limit neighborhood of the mean, which is a reasonable assumption. Then all intensities of that tissue are collected in a separate distribution. Then for the middle tissue, the median value is selected as the intrinsic tissue. For the tissues with lowest and highest intensities, it is assumed that approximately half of the tissue may be affected from PVE. Therefore, the remaining half is assumed to be homogeneous part of the tissue. The median of the homogeneous part of the tissue is assumed to be the intrinsic intensity of the tissue. Therefore, for the tissue with lowest intensity, 25^{th} percentile is set as the intrinsic intensity whereas 75^{th} percentile is selected as the intrinsic intensity of the tissue with highest intensity. The assumption that half of the tissues are affected by PVE may not be so realistic because of the changes in scanning and inter-subject variations. But the intensity values are expected to be quite robust because slight changes in the percentage of the percentile are not expected to change the calculated tissue intensity dramatically. For instance, if 30% of the tissue with lowest intensity is affected from PVE, 35^{th} percentile is expected to give the intrinsic tissue intensity. However, the 25^{th} and 35^{th} percentile is not expected to be so different because it probably lies in a dense region in the histogram.

$$\sigma_{\ker nel} = thres1\_\lim it / 2 \qquad (3.34)$$

The median of the collection of the smoothed intensities is accepted as the intrinsic intensity of the middle tissue. Median is selected as the measure because it seems a more robust estimator than mean. There are parts of the middle tissue which affected by the PVE with both tissues. The median likely corresponds to the parts where the tissue is homogeneous.

After all the intrinsic intensities are available, the squared error for that threshold combination is calculated as in (3.35). The threshold combination which gives the minimum error is selected as the optimal one. This type of error may not be reliable if the number of regions in the filtered histogram is high. In that case, it is expected that the middle mean is placed on the average of the low and high intrinsic tissue intensities to minimize error. However, a small number of regions is present in our histogram, therefore the estimation is quite reliable.

$$error = \sum_{R_i \in H_{filt}} Vol(R_i) \times (I(R) - I(Tissue(R_i))^2 \qquad (3.35)$$

A sample filtered histogram and calculated optimal tissue intrinsic intensities are shown in Figure 3.11.

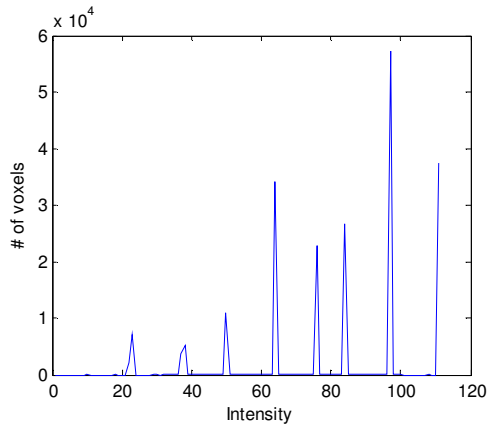


Figure 3.11 The filtered histogram of sub-cortical region for a T1 weighted image. The means of CSF, GM and WM are 22.7, 71.5 and 109.4. The CSF-GM and GM-WM thresholds are calculated as 47.1 and 90.5 accordingly.

3.5.2 Finalization of Segmentation

After the optimal intrinsic tissue intensities are calculated, the 2 thresholds are determined. They are calculated as in (3.36), (3.37), (3.38) and (3.40).

$$Threshold_{1-2} = I(Tissue_1) \times Rat_{1-2} + I(Tissue_2) \times (1 - Rat_{1-2}) \qquad (3.36)$$

$$Threshold_{2-3} = I(Tissue_2) \times Rat_{2-3} + I(Tissue_3) \times (1 - Rat_{2-3}) \quad (3.37)$$

$$Rat_{1-2} = 0.5 \quad (3.38)$$

$$Rat_{2-3} = 0.5 \quad (3.39)$$

The thresholds are set as midway intensities of tissues. This strategy comes from the following model. It is assumed that the intensity levels of regions are mostly determined by the volumetric ratios of tissues forming it and the effect of smoothing on the boundaries between tissues. This modeling is formulated in (3.40).

$$I(R_i) = (\sum_{n=1}^{3} Vol\_ratio(Tissue_n) \times I(Tissue_n)) + Smoothing \quad (3.40)$$

The effect of smoothing is decreased by the ignorance of border voxels but it can still have a bias. It is not possible to estimate the exact effect of smoothing on the intensity of a region. Therefore, it is neglected. The region is assigned to the closest tissue in terms of intensity. The 2 ratios defined in (3.38) and (3.39) can be changed to regulate the final classification. This procedure is very simple and does not require the reoperation of the whole algorithm. This final segmentation can be performed in a very negligible time, about few seconds. This type of regulation may be important when thin structures which can be affected significantly by smoothing and PVE is tried to be enhanced.

CHAPTER 4

RESULTS

In this chapter, the performance of the proposed EHRM algorithm on simulated and real MR brain images are presented. The performance is compared with the FSL FAST algorithm[1], which is currently one of the most widely used segmentation algorithms. FAST models the segmentation as a Hidden Markov Random Field Model with the utilization of Expectation Maximization algorithm for the estimation of tissue class parameters. Further information about FAST algorithm is available in the section 2.3. The comparison is performed with some of widely known performance measures like Jaccard Index [50] and Detection Ratio.

First, the operation stages of the algorithm are presented with the results of a sample brain image. Then the performance of the algorithm on 3 different datasets are given and discussed.

4.1 Operation Stages of the Algorithm

The operation of the algorithm with the default parameters for a sample brain is explained in this section.

The sub-cortical region of the brain is first determined by the linear registration of the template brain to the new brain. The transformation of the predefined sub-cortical region with the same linear registration parameters corresponds to the sub-cortical region of the new area. An axial slice chosen

approximately mid-way through the sub-cortical region in the inferior-to-superior direction is shown in Figure 4.1.

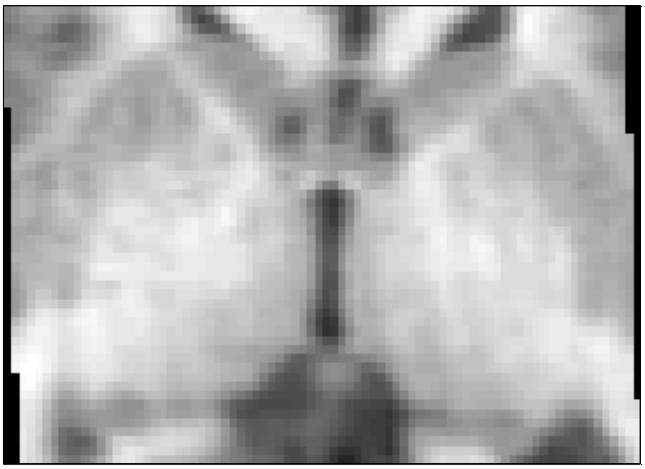


Figure 4.1 An axial slice from sub-cortical region.

The initial threshold is determined from the histogram containing minimum pairwise differences among neighboring voxels. The histogram is plotted in Figure 4.2. The initial threshold, 75th percentile of the histogram is calculated as 3.

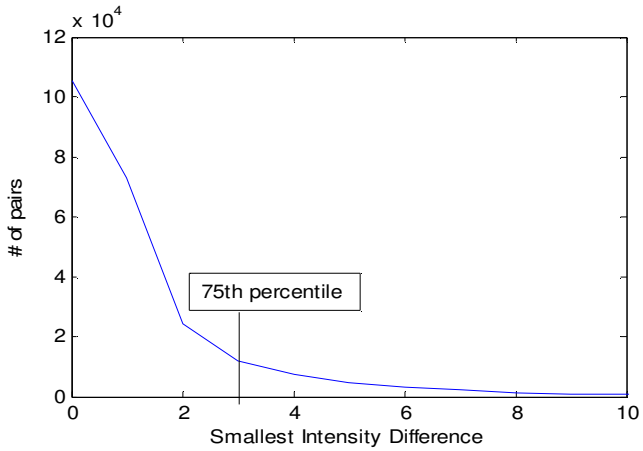


Figure 4.2 Histogram of minimum intensity differences between neighboring voxels.

The threshold limits are also determined at the beginning, extracted from the intensity histogram of the sub-cortical region. The histogram is shown in Figure 4.3. The 16.67th, 50th an 83.33th percentiles are calculated as 60, 86 and 106 respectively. Then the threshold limit is determined as 10 ((106-86)/2).

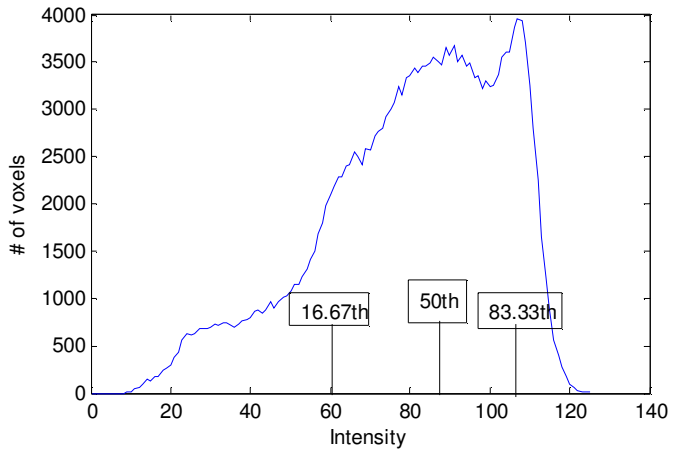


Figure 4.3 Histogram of sub-cortical region.

When the algorithm reaches to the critical point, it is expected that the regions achieve a certain level of maturity. The image at the critical point is displayed in Figure 4.4. It is seen that there are significant regions representing the major parts of anatomic structures. At the critical point, there are generally about 40-60 intermediate regions in the image.

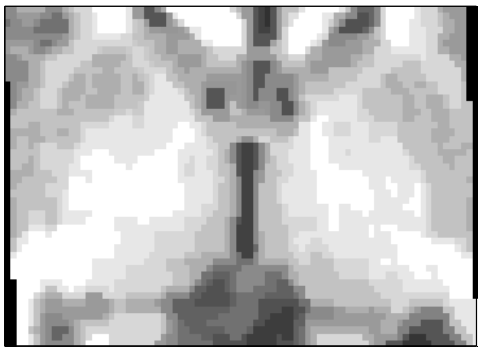


Figure 4.4 Image at the critical point.

After the critical point, the large regions are determined. Their texture heterogeneity features are calculated and standardized. The texture heterogeneity map of the image is displayed on Figure 4.5. The voxels which do not belong to a large region are represented with black color. It is easily seen that the regions which are adjacent to strong boundaries tend to have higher heterogeneity values. This observation is mainly due to the smoothing and PVE in the image. Although the boundary voxels of a region are excluded from the texture analysis, the smoothed effects of strong boundaries can be observed also in interior voxels of the image near to the boundary. Such smoothing deviates the intensity of effected interior voxels and is perceived as texture heterogeneity.

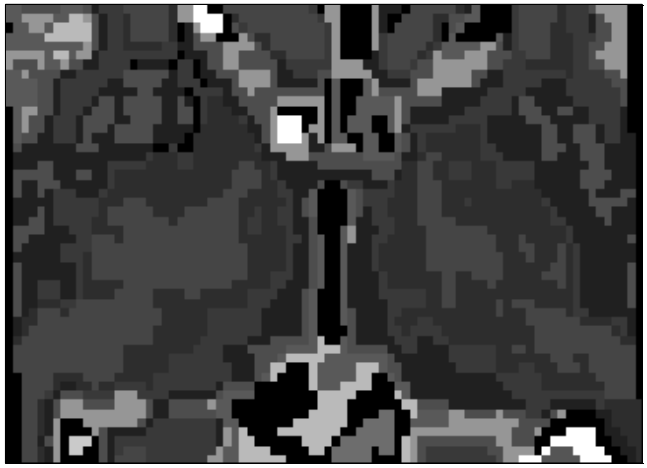


Figure 4.5 Texture map at the critical point.

When visually inspected, it is observed that the smoothing effect mostly diminishes after 2 layers of voxels. But the removal of 2 layers from the surface of regions does not seem feasible because it would roughly lead to a 4 layer of decrease in the tissue thickness in texture analysis. The anatomy of the sub-cortical region is quite complex and sub-cortical structures are not thick enough to allow for reliable extraction of texture features after removal of 2 layers. For a conventional scan with

1x1x1mm resolution, it may be very hard to achieve a certain number of voxels and gaps to derive a reliable texture feature.

It should be noted that the texture features may not be reliable in many cases and may not contribute much to the region merging process. However, this is not a great risk because texture is used only as a second criterion for the region merging. If texture criterion is excluded, the intensity similarity would be the only determinant. The texture may only be preventive, rather than facilitative, denying some merges between large regions.

At the end of the region merging process, the image in Figure 4.6 is obtained. It is observed that the image is quantized to intensity levels quite successfully. Some weak boundaries can also be detected and preserved. The related texture map is given in Figure 4.7.

Figure 4.6 Regions at the end of region merging.

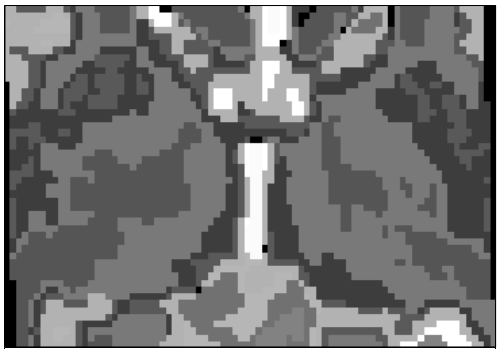


Figure 4.7 Texture map at the end of region merging.

The image in Figure 4.6 should be reduced to 3 classes according to its histogram. First, the intrinsic tissue intensity means are calculated from the histogram in Figure 4.8. The estimated tissue means for CSF, GM and WM are 22.7, 71.5 and 109.4 respectively. In the default operation, the 2 thresholds for final segmentation are selected as the averages of the related tissue means. The CSF-GM and GM-WM thresholds are calculated as 47.1 and 90.5 respectively.

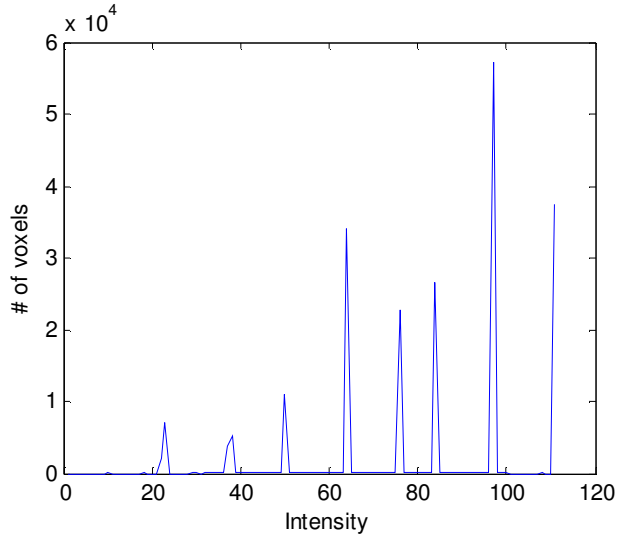
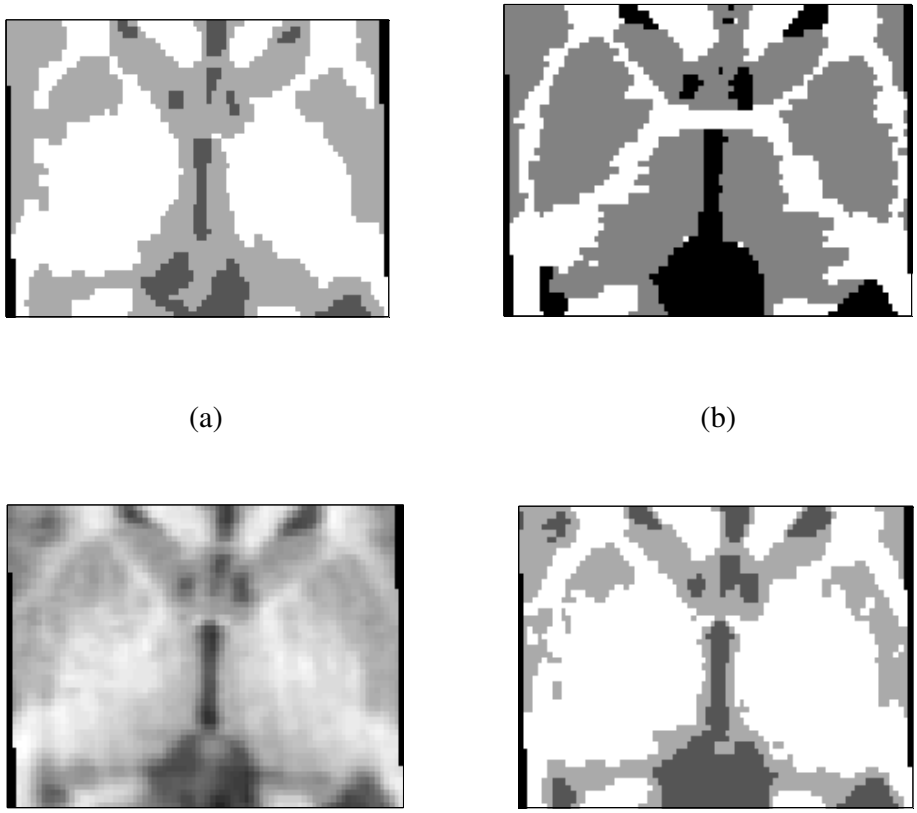


Figure 4.8 Histogram at the end of region merging.

The final results of segmentation are displayed in Figure 4.9 together with the raw image, FAST segmentation and the reference segmentation which is accepted as the gold standard. Slight improvements in putamen and thalamus can be observed.

The segmentation of CSF is not problematic due to the sufficient contrast most of the time. There may be subtle errors at CSF boundaries due to the smoothing and PVE. The PVE effect in the boundaries of lateral ventricles and WM are frequently detected as GM because the boundary voxels have intermediate intensity.



(a) (b)

(c) (d)

Figure 4.9 The results of the segmentation. (a) is the segmented image, (b) is the manually segmented image, (c) is the raw image, (d) is the FAST segmentation.

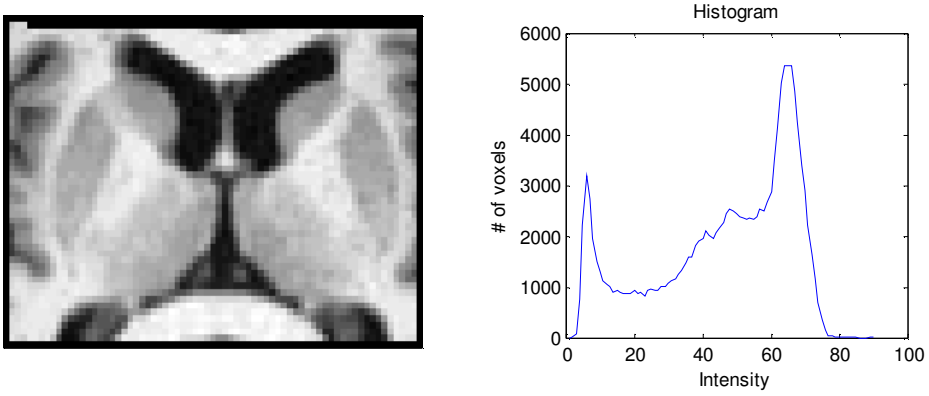
The segmentation of WM is also problematic in certain areas. The first one is the internal capsule, which separates the putamen and globus pallidus from the thalamus. The boundary of internal capsule with thalamus is very smooth. In addition, thalamus also has a gradual intensity decrease from the internal capsule to the ventricle due to the spatially changing anatomical characteristics, smoothing effect and PVE. Therefore, the boundary between internal capsule and thalamus may not be placed to the exact boundary, resulting in underestimation of the thalamus frequently. The shape of the boundary may be very complicated which is significantly shaped by the noise, due to the insufficient contrast. The boundary between the globus pallidus and the internal capsule is a very weak boundary and the intensity of globus pallidus is usually closer to the internal capsule rather than the putamen. So it is quite impossible to detect the globus pallidus. It usually merges to the internal capsule. The boundary between the putamen and the globus pallidus is more evident and enables the detection of the putamen boundary. The boundary of putamen with internal capsule is less evident in the posterior parts of putamen which increases false segmentations. In terms of texture, internal capsule is more heterogeneous than most other WM tissues like corpus callosum. Also the neighbors of the internal capsule, especially thalamus and globus pallidus have a heterogeneous GM tissue, which is a challenge for the utilization of texture heterogeneity in the region merging process.

Brain stem region also turns out to be a problematic region for segmentation. Brain stem, which is a mixture of WM and GM, is segmented as WM in manual segmentations. The tissue is also quite heterogeneous. Therefore, some parts of the brain stem are segmented as GM by the proposed algorithm.

The most problematic GM structures are the thalamus, putamen and globus pallidus. As previously discussed, it is very hard to detect the globus pallidus. On the other hand, the major parts of putamen and thalamus are usually segmented quite acceptably although the boundaries with the internal capsule are not detected precisely.

The cortical GM tissue, insula and caudate have a relatively higher contrast with other tissues and they are segmented quite successfully.

The performance of the algorithm is investigated and compared with a simple quantizer. The image shown in Figure 4.10 is used for testing.



(a) (b)

Figure 4.10 An axial slice from the test image (a) and its histogram (b).

The EHRM algorithm is operated for this image and the image at Figure 4.11(b) is obtained at the end of merging, before final reduction to 3. The corresponding histogram is in Figure 4.11(d). 7 prominent regions are observed at the end of region merging. The quantizer is set to divide the intensity range also to 7. The output of the algorithm and the digitizer is displayed with a sample slice and histogram of both images in Figure 4.11.

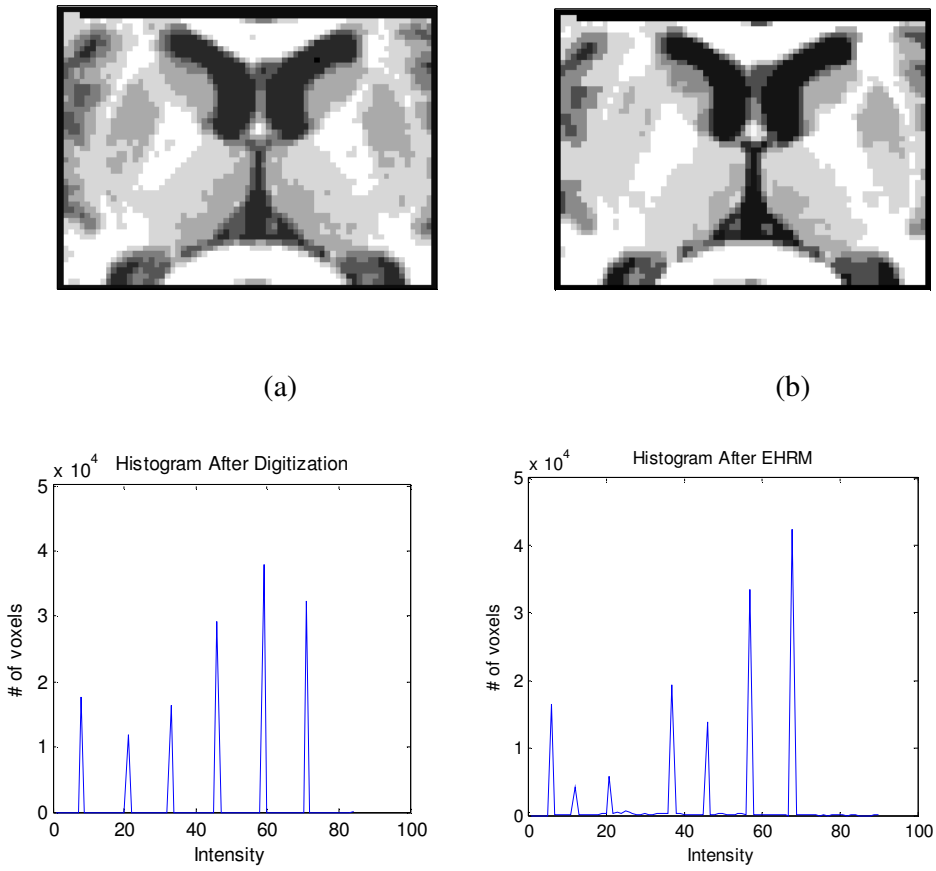


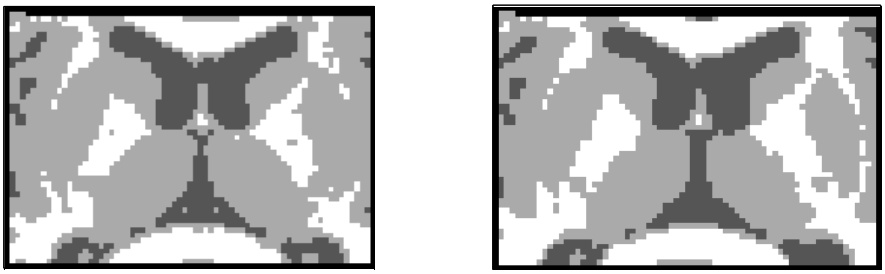
Figure 4.11 Performance of a digitizer and EHRM. (a) is the image of the digitizer, (b) is the image at the end of region merging, (c) is the digitizer histogram, (d) is the histogram at the end of region merging[1]. There is also a very small peak at the histogram of digitizer about 84 intensity value, which is hard to see.

It is observed that the digitizer cannot impose connectivity constraint as the region merging algorithm. Therefore the boundaries at the image may not be smooth. Also the effect of noise can be seen as spots in the middle of regions. A simple digitizer can also split an anatomic structure into 2 because of the hard

[1]There exists a very small peak at the histogram of digitizer about 84 intensity value, which is hard to see.

transition points in the quantization process. This is observed for both internal capsules that separate thalamus and putamen. Also the detection of external capsule at the right side is handled better with EHRM approach. The connectivity of WM is achieved better. Region merging is more adaptive and is more likely to form coherent representations of anatomical structures and smoother boundaries. It is expected to be more robust to noise and bias field in comparison to a simple intensity quantization process.

Both images are reduced to 3 classes with the proposed approach. The segmentations are displayed in Figure 4.12. It is observed the WM structures are detected much better by the EHRM approach. The internal capsules and right external capsule are detected much more accurately. The superior performance of region merging compared to simple quantization can also be seen in the quantitative measures. The Jaccard Index (JI), defined in (4.1), for WM is 0.417 in digitization and 0.519 for EHRM. The reason for this is the significant underestimation of WM in quantization. The estimated volume for WM is 0.223 in digitization whereas EHRM estimates as 0.294, which is still an underestimation but much closer the actual volumes of WM. Similar observations are also observed for GM JI and volumetric estimations. The segmentation quality of CSF is affected negligibly.



(a) (b)

Figure 4.12 Final segmentations after quantization (a) and EHRM (b).

4.2 Performance Analysis on the Real IBSR Dataset

The algorithm is operated on the 18 high resolution T1 weighted brain images which have manual segmentations provided by the Center for Morphometric Analysis at Massachusetts General Hospital and are available at http://www.cma.mgh.harvard.edu/ibsr/. The details of the scan and the manual segmentation procedure are available on the stated website. Same images are also segmented with the FSL FAST algorithm for comparison. Before operating FAST, brain extraction is performed using the FSL BET [49] tool. The extractions are visually optimized with the modulation of f and g parameters of the BET algorithm. Another algorithm, a simple k-means algorithm which incorporates coordinates to the segmentation, is also implemented and operated on the same images. The details of this k-means algorithm, which is abbreviated as SK, is given in Appendix C. The purpose of implementing SK is to compare the performance of EHRM with a known simple approach which can be quite successful for many segmentation problems.

The performance of the algorithms are compared in terms of Jaccard Index(JI) [50] for individual tissues and the success of volumetric ratio estimation; caudate, thalamus and putamen detection ratio. Also the robustness of EHRM for the 2 major parameters is discussed.

4.2.1 GM Analysis

Jaccard Index(JI) [50], also known as Jaccard Similarity Coefficient, is a popular measure for the segmentation quality assessment. It is defined by the ratio of number of elements in the intersection set to the number of elements in the union set of 2 sets. It is formulated as in (4.1). In the formula, the segmentation result of an algorithm is taken as the first set A and the second set B is the manually segmented image. In case of a large overlap, the Jaccard index appoaches to 1, and in case of a large mismatch, JI approaches 0. This metric turns out to be somewhat problematic for the IBSR dataset, because the provided manual segmentations can include some anatomical details which are not apparent from the raw image intensities. Some parts, which do not have GM intensity appearance, is labeled as a part of GM at

manual segmentations because GM structures are expected as such locations. Therefore, this property of manual segmentations produces a bias in JI metrics.

$$JI = \frac{|A \cap B|}{|A \cup B|} \qquad (4.1)$$

The JI for GM tissue segmentation with EHRM is calculated and compared with the FSL FAST [1] and SK algorithm. The volumetric ratio of GM in total sub-cortical area of the EHRM, FAST, SK and the manual segmentations are also presented. The results for GM tissue are in Table 4.1.

Table 4.1 JI for and Volumetric Ratios for GM.

Subject No	EHRM GM Volume	EHRM GM JI	FAST GM Volume	FAST GM JI	SK GM Volume	SK GM JI	GM Volume
1	0.535	0.566	0.295	0.484	0.376	0.532	0.438
2	0.484	0.593	0.304	0.517	0.370	0.531	0.378
3	0.502	0.660	0.333	0.531	0.477	0.609	0.478
4	0.587	0.622	0.348	0.548	0.577	0.555	0.485
5	0.384	0.505	0.286	0.453	0.353	0.471	0.413
6	0.439	0.577	0.280	0.467	0.357	0.466	0.395
7	0.399	0.533	0.321	0.490	0.347	0.512	0.389
8	0.498	0.523	0.324	0.539	0.383	0.549	0.378
9	0.371	0.540	0.272	0.481	0.264	0.461	0.331
10	0.323	0.517	0.269	0.504	0.289	0.477	0.326
11	0.402	0.578	0.336	0.540	0.283	0.526	0.377
12	0.438	0.599	0.331	0.552	0.411	0.556	0.399
13	0.507	0.611	0.354	0.545	0.462	0.581	0.523
14	0.438	0.624	0.360	0.596	0.464	0.616	0.464

Table 4.1 (Continued)

15	0.653	0.569	0.354	0.540	0.404	0.394	0.497
16	0.658	0.680	0.312	0.495	0.457	0.575	0.497
17	0.517	0.660	0.319	0.511	0.395	0.518	0.506
18	0.460	0.565	0.304	0.476	0.473	0.575	0.486
Mean	0.477	0.585	0.317	0.515	0.397	0.528	0.431
Std. Dev.		0.051		0.037		0.057	

The JI of GM tissue for EHRM is 7.0% better than the JI of FAST algorithm and this difference is statistically significant ($p<0.00001$). The volumetric ratio estimation of EHRM for GM is also superior to the performance of FAST algorithm at 5% confidence interval ($p=0.002$). It is observed that EHRM is more successful for the detection of GM tissue and the volumetric estimation. EHRM slightly overestimates the GM whereas FAST underestimates GM significantly.

EHRM is also superior to SK in terms of GM JI with 5.7%. Their volumetric estimation performance is similar. EHRM slightly overestimates GM whereas SK slightly underestimates.

4.2.2 Detection of Caudate, Putamen and Thalamus

The performance of the algorithm on the detection of caudate, putamen and thalamus structures is calculated. The calculated metric is the ratio of segmented GM volume in the structure to whole structure volume. The results are tabulated in Table 4.2. The detection ratio (DR) for a structure S can be expressed as in (4.2).

$$DR(S) = \frac{\left|S \cap Tissue(S)\right|}{\left|S\right|} \qquad (4.2)$$

Table 4.2 Detection Ratios (DR) for Caudate (CA), Putamen (PU) and Thalamus (TH).

Subject No	EHRM DR CA	FAST DR CA	SK DR CA	EHRM DR PU	FAST DR PU	SK DR PU	EHRM DR TH	FAST DR TH	SK DR TH
1	0.958	0.765	0.835	0.945	0.101	0.279	0.837	0.403	0.623
2	0.972	0.864	0.836	0.845	0.324	0.633	0.873	0.503	0.560
3	0.965	0.879	0.915	0.858	0.446	0.837	0.743	0.459	0.665
4	0.981	0.826	0.952	0.978	0.646	0.908	0.863	0.412	0.761
5	0.843	0.826	0.835	0.259	0.047	0.062	0.555	0.378	0.633
6	0.961	0.902	0.868	0.790	0.153	0.083	0.655	0.374	0.688
7	0.964	0.942	0.898	0.896	0.683	0.608	0.587	0.435	0.468
8	0.985	0.930	0.903	0.978	0.808	0.835	0.835	0.445	0.538
9	0.973	0.932	0.924	0.799	0.441	0.145	0.598	0.367	0.373
10	0.946	0.923	0.911	0.616	0.607	0.502	0.437	0.316	0.353
11	0.973	0.915	0.766	0.808	0.750	0.517	0.717	0.652	0.493
12	0.931	0.841	0.579	0.855	0.457	0.445	0.668	0.437	0.729
13	0.972	0.871	0.904	0.936	0.607	0.817	0.681	0.479	0.631
14	0.928	0.885	0.928	0.754	0.617	0.743	0.592	0.554	0.754
15	0.971	0.914	0.651	1.000	0.429	0.732	0.964	0.456	0.440
16	0.975	0.809	0.778	0.994	0.385	0.790	0.887	0.303	0.687
17	0.941	0.729	0.530	0.729	0.281	0.432	0.649	0.388	0.372
18	0.824	0.713	0.840	0.569	0.324	0.744	0.418	0.310	0.674
Mean	0.948	0.859	0.825	0.812	0.450	0.562	0.698	0.426	0.580
Std. Dev.	0.045	0.070	0.123	0.185	0.220	0.272	0.155	0.088	0.135

Although the average detection ratios for all these 3 sub-cortical GM structures are higher for EHRM algorithm, we would like to note the dramatic gain obtained by using EHRM for putamen and caudate. The results show that on the average, the EHRM is superior than FAST about 36.2% for putamen detection ($p<0.00001$) and 27.2% better for thalamus detection ($p<0.00001$). For caudate structure, the detection ratio is approximately 8.9% better than FAST algorithm($p<0.00001$).

Although SK seems more successful than FAST in average DR values of 3 structures, EHRM is still superior to SK in terms of DR in all 3 structures.

4.2.3 CSF and WM Analysis

The JI for CSF and WM tissues are calculated and compared with the FSL FAST and SK algorithms. The results for CSF are presented in Table 4.3. The volumetric ratio of CSF in total sub-cortical area of the EHRM, FAST, SK and the manual segmentations are presented below.

Table 4.3 JI for and Volumetric Ratios for CSF.

Subject No	EHRM CSF Volume	EHRM CSF JI	FAST CSF Volume	FAST CSF JI	SK CSF Volume	SK CSF JI	CSF Volume
1	0.156	0.606	0.162	0.625	0.118	0.660	0.108
2	0.083	0.690	0.109	0.660	0.088	0.683	0.080
3	0.049	0.551	0.084	0.511	0.066	0.507	0.051
4	0.112	0.578	0.129	0.593	0.118	0.547	0.089
5	0.084	0.627	0.124	0.676	0.099	0.666	0.102
6	0.125	0.707	0.176	0.786	0.132	0.686	0.165
7	0.118	0.655	0.150	0.596	0.119	0.677	0.099
8	0.209	0.740	0.241	0.721	0.193	0.752	0.190
9	0.106	0.781	0.134	0.765	0.102	0.758	0.114

Table 4.3 (Continued)

10	0.191	0.832	0.228	0.857	0.206	0.858	0.214
11	0.091	0.707	0.133	0.605	0.089	0.703	0.090
12	0.118	0.667	0.158	0.639	0.119	0.683	0.121
13	0.173	0.574	0.196	0.593	0.170	0.567	0.128
14	0.109	0.696	0.149	0.663	0.116	0.649	0.109
15	0.147	0.398	0.115	0.530	0.206	0.252	0.080
16	0.068	0.605	0.092	0.633	0.135	0.374	0.074
17	0.112	0.706	0.170	0.692	0.106	0.662	0.131
18	0.090	0.622	0.167	0.688	0.147	0.622	0.130
Mean	0.119	0.652	0.151	0.657	0.129	0.628	0.115
Std. Dev.		0.097		0.087		0.141	

The average JI for EHRM is slightly lower for than the FAST for CSF tissue. A 2 tailed paired t test is conducted and this difference is found not to be significant for 5% confidence level ($p=0.7013$).

If the volumetric ratio estimation of EHRM and FAST is compared, the performance of EHRM is superior to FAST ($p=0.0131$). The errors of the algorithms for volumetric estimation are taken as the ratio of volume error to the whole sub-cortical volume while conducting the statistical test. However, it should also be noted that the CSF is slightly underestimated in the manual segmentations, which might have produced some bias towards EHRM in the statistical testing.

The JI for CSF is better for EHRM when compared with the SK algorithm. The SK algorithm seems to slightly overestimate CSF if the possible errors in the manual segmentation are ignored. The EHRM estimates the CSF volume better than SK. Similar analysis is shown for WM and the results are shown in Table 4.4.

Table 4.4 JI for and Volumetric Ratios for WM.

Subject No	EHRM WM Volume	EHRM WM JI	FAST WM Volume	FAST WM JI	SK WM Volume	SK WM JI	WM Volume
1	0.309	0.549	0.544	0.647	0.507	0.635	0.454
2	0.433	0.661	0.587	0.702	0.542	0.671	0.543
3	0.449	0.681	0.584	0.660	0.457	0.637	0.471
4	0.301	0.563	0.523	0.652	0.305	0.483	0.426
5	0.532	0.639	0.590	0.642	0.548	0.610	0.485
6	0.436	0.668	0.544	0.648	0.511	0.617	0.440
7	0.483	0.658	0.530	0.677	0.534	0.661	0.512
8	0.294	0.519	0.435	0.683	0.424	0.650	0.433
9	0.523	0.703	0.594	0.714	0.634	0.715	0.555
10	0.486	0.689	0.502	0.689	0.505	0.662	0.460
11	0.507	0.698	0.530	0.717	0.628	0.718	0.533
12	0.444	0.690	0.511	0.714	0.469	0.654	0.480
13	0.321	0.586	0.450	0.615	0.368	0.576	0.349
14	0.453	0.668	0.491	0.685	0.420	0.645	0.427
15	0.200	0.456	0.531	0.647	0.390	0.501	0.423
16	0.274	0.575	0.596	0.614	0.408	0.632	0.429
17	0.371	0.620	0.512	0.584	0.499	0.553	0.363
18	0.451	0.594	0.529	0.589	0.380	0.575	0.383
Mean	0.404	0.623	0.532	0.660	0.474	0.622	0.454
Std. Dev.		0.070		0.041		0.064	

The WM JI of FAST is statistically superior to EHRM (p=0.0214). The WM volumetric ratio estimations of EHRM is slightly better than FAST but is not statistically significant (p=0.5909).

The JI for WM is similar for EHRM and the SK algorithms. The SK algorithm seems to slightly overestimate WM but volumetric estimations are closer to actual values in average when compared to EHRM algorithm.

4.2.4 Robustness of EHRM to Initial Parameters

The 2 crucial parameters of the algorithm, which are also static (not adaptive), are the volumetric ratio threshold for determining intermediate regions (vol_rat) and the minimum number of large regions (n_region). A single brain, which is the 1st image of IBSR dataset, is operated with several combinations of the 2 parameters to test robustness against these two parameters.

6 different values for vol_rat: 0.0003, 0.0006, 0.001 (default value), 0.0015, 0.002, 0.003 and 3 different values for n_region: 7, 10(default value), 13 are used to operate the algorithm for 18 different parameter combinations.

4.2.4.1 JI and Volumetric Estimation Analysis

The effect of different parameters on JI and volumetric ratios are investigated for tissue classes separately. The JI and volumetric ratio of CSF is shown in Figure 4.13.

The parameters do not seem to have a significant effect on the results. A significant pattern cannot be observed on the operation of the algorithm. Very subtle variations are found, probably due to the stochastic nature of the evolutions.

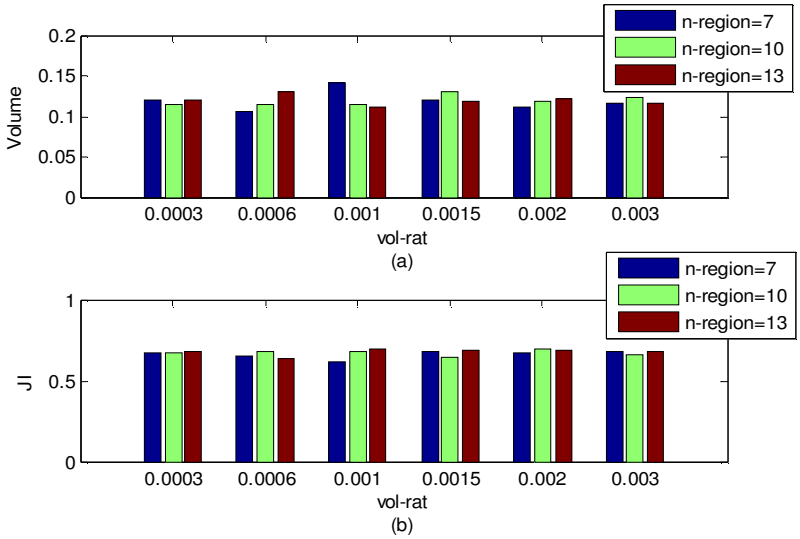


Figure 4.13 Volume and JI for CSF for different parameter combinations. (a) the volumetric estimations, (b) JI.

Similar analysis is also conducted for GM and WM tissues. The results are displayed in Figure 4.14 and 4.15.

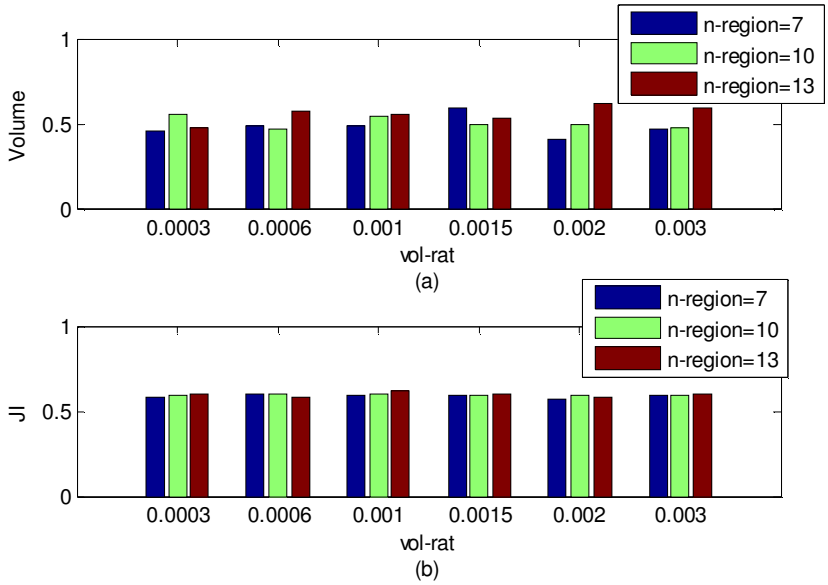


Figure 4.14 Volume and JI for GM for different parameter combinations. (a) the volumetric estimations, (b) JI.

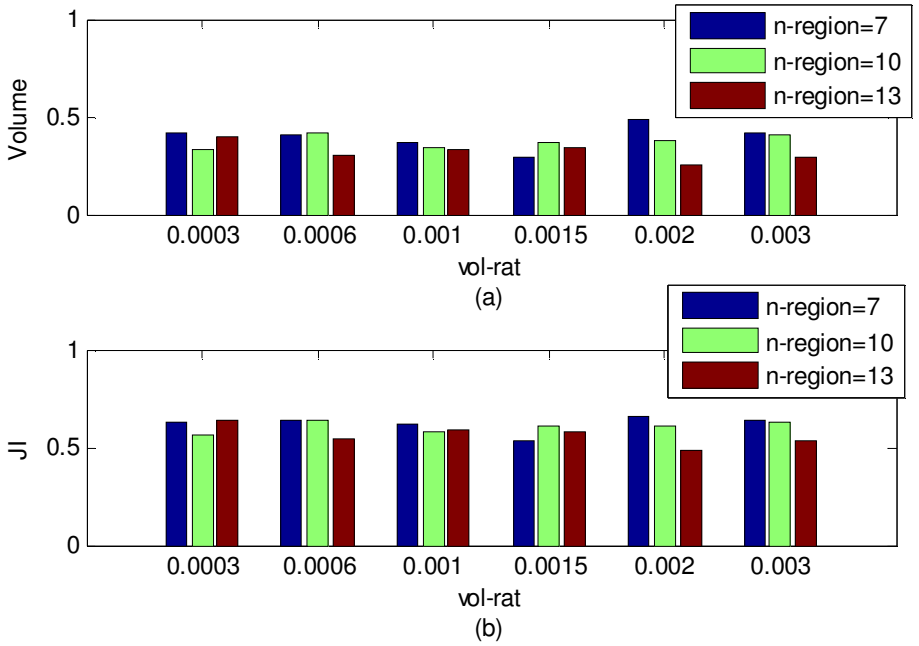


Figure 4.15 Volume and JI for WM for different parameter combinations. (a) the volumetric estimations, (b) JI.

For n_region 7 and 13, the volumetric ratios for GM and WM do not look very stable when the vol_rat parameter is high. This is probably due to the regions which have intensity values in the middle of GM and WM intensities. Such regions can be segmented differently each time due to the stochastic nature of the algorithm. But for medium vol_rat parameters, the JI is quite stable for especially GM. On the other hand, n_region parameter does not seem to affect the performance of GM and CSF significantly, but the JI for WM is lower for the high n_region value in general. The medium n_region and vol_rat parameters seem more stable, therefore the default parameters vol_rat=0.001 and n_region=10 seem reasonable. In any case, the JI for GM is stable.

4.2.4.2 Caudate, Putamen and Thalamus Analysis

On another front, the effect of these 2 parameters on the detection of caudate, putamen and thalamus are investigated. The results of the analysis are displayed on Figures 4.16, 4.17 and 4.18.

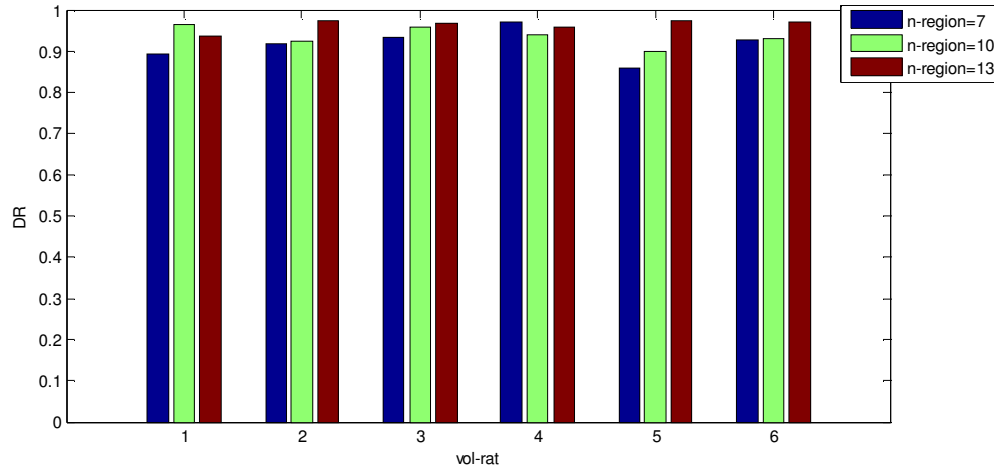


Figure 4.16 Detection Ratio of caudate for different parameter combinations.

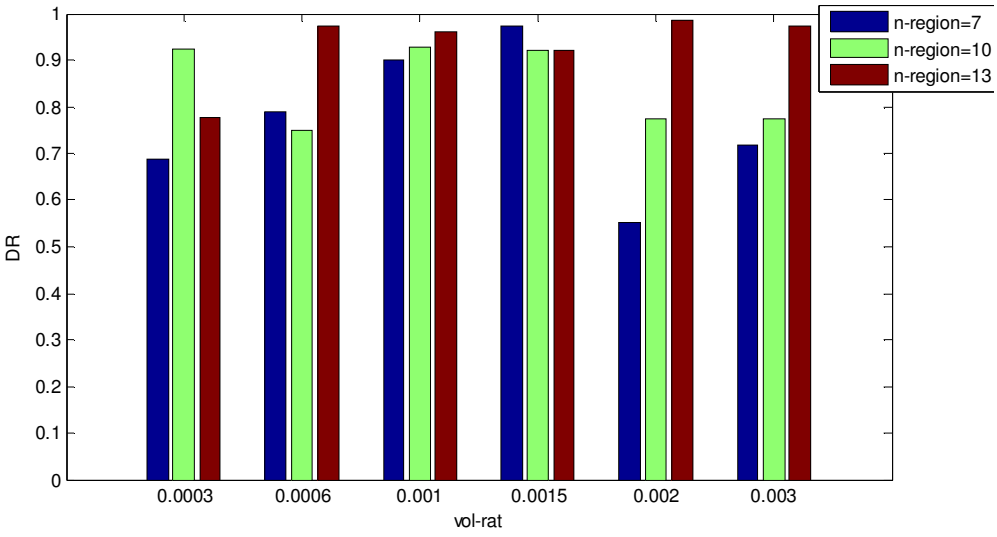


Figure 4.17 Detection Ratio of putamen for different parameter combinations.

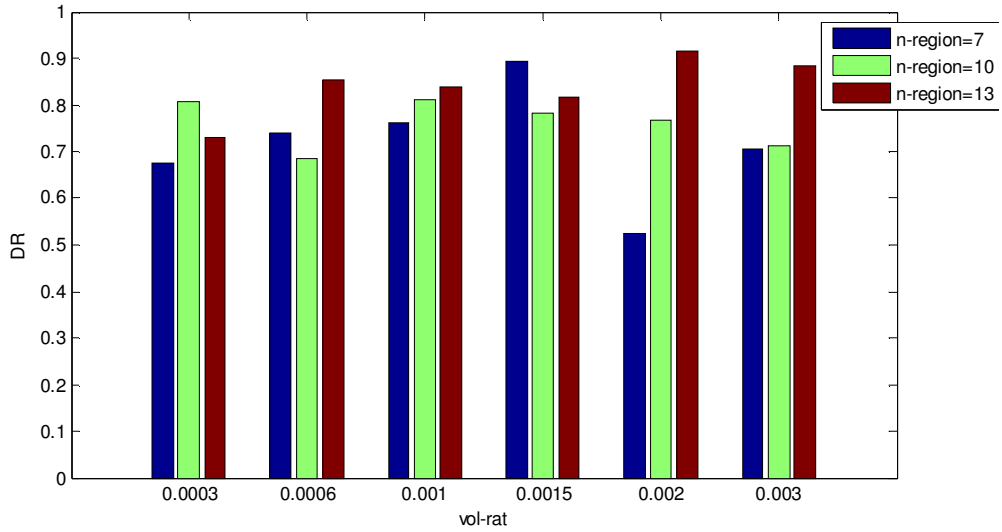


Figure 4.18 Detection Ratio of thalamus for different parameter combinations.

The detection of caudate is quite stable because it usually has a good contrast with WM. However, the detection for putamen and thalamus are sensitive to the parameters. The performance is especially more sensitive at higher vol_rat values. The default parameters turn out as a good choice.

4.2.5 Randomness Analysis

EHRM is a stochastic algorithm due to its evolutionary nature. Two types of stochastic processes are present: 1. The order of activation of regions in an epoch, 2. The random threshold that the determines whether a particular region can be active at a certain epoch. Due to these 2 processes, the segmentation results differ when the experiment is repeated with the same input image.

It is desired that the outputs of EHRM are quite stable so that a single operation of the algorithm is sufficient. EHRM algorithm is operated 10 times with the same input image. The variability of the JI and volumetric ratios of tissues are displayed in Figures 19 and 20. It is observed that the JI of 3 tissues are quite stable with small deviations. The volumetric estimations seem less stable. Although the CSF volume is stable, GM and WM does not seem so stable. This is probably due to the regions

which lie near the GM – WM intensity threshold. Such regions are classified differently in different cases. Therefore, deviations in volume are observed. However, the JI values are more stable because such regions contain voxels of both tissues. In any case, they cause similar error on the segmentation, making JI values stable. It can be concluded that the stochastic nature of the algorithm is not an important issue when the JI values are considered.

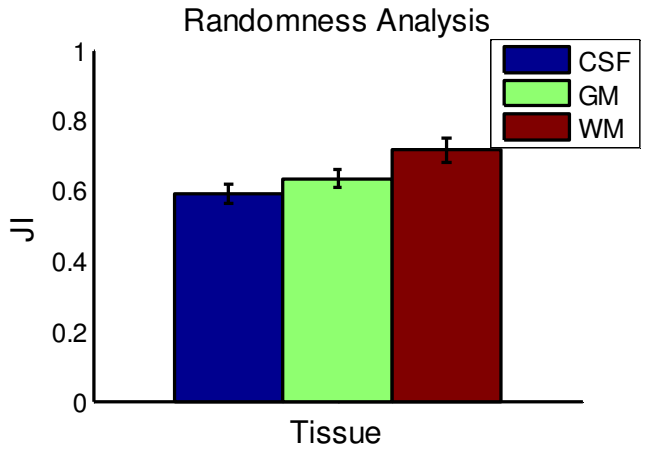


Figure 4.19 The randomness of JI for 3 tissues. The error bars show the average and standard deviation of JI.

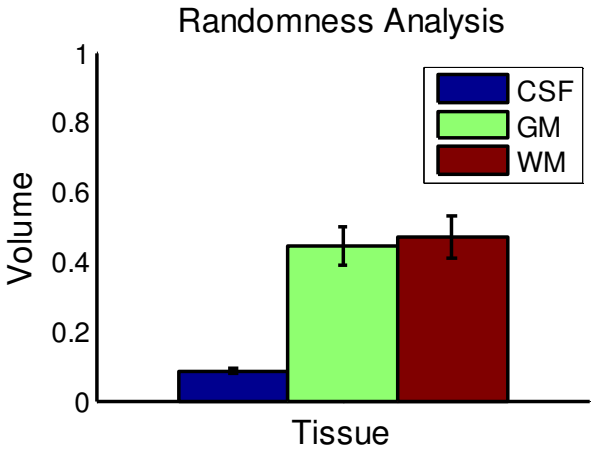


Figure 4.20 The randomness of Volumetric Estimations for 3 tissues. The error bars show the average and standard deviation of volumetric estimations.

4.3 Performance Analysis on the Real Dataset

The algorithm is operated on the 17 high resolution normal T1 weighted brain images. The images are taken with Siemens Magnetom Vision MR Device. Standard mprage sequence is used and scan settings are flip angle=12^0, TE=4ms, TR=9.7ms. The resolution is 1mm x 1mm x 1mm. No manual segmentation is available. Therefore the segmentations produced by the Freesurfer software [2] are taken as the gold standard to evaluate the segmentation. The performance of this algorithm is shown to be comparable to manual segmentation [2]. All anatomical structures are grouped in their tissue types to perform the JI and volumetric analysis. Same images are also segmented with the FSL FAST and SK algorithms. Before operating FAST, the brain extraction is taken from the Freesurfer software. The extractions are visually checked to prevent any possible errors.

Similar to the evaluation in IBSR dataset, the performance of two algorithms are compared in terms of Jaccard Index (JI) for individual tissues and the success of volumetric ratio estimation; caudate, thalamus and putamen detection ratio.

4.3.1 GM Analysis

The JI for GM tissue is calculated and compared with the FSL FAST and SK algorithms. The volumetric ratio of CSF in total sub-cortical area of the EHRM, FAST, SK and the manual segmentation are also presented. The results for GM tissue are in Table 4.5.

Table 4.5 JI for and Volumetric Ratios for GM.

Subject No	EHRM GM Volume	EHRM GM JI	FAST GM Volume	FAST GM JI	SK GM Volume	SK GM JI	GM Volume
1	0.498	0.660	0.342	0.590	0.344	0.542	0.441
2	0.390	0.585	0.305	0.563	0.364	0.497	0.406
3	0.352	0.567	0.321	0.573	0.436	0.559	0.407

Table 4.5 (Continued)

4	0.479	0.622	0.320	0.556	0.423	0.552	0.415
5	0.504	0.619	0.338	0.590	0.383	0.520	0.401
6	0.365	0.601	0.282	0.568	0.360	0.555	0.343
7	0.368	0.577	0.264	0.504	0.322	0.498	0.363
8	0.519	0.509	0.244	0.449	0.445	0.449	0.398
9	0.385	0.607	0.285	0.557	0.301	0.527	0.377
10	0.361	0.578	0.350	0.600	0.455	0.592	0.415
11	0.408	0.587	0.302	0.545	0.412	0.574	0.388
12	0.440	0.655	0.347	0.618	0.368	0.539	0.423
13	0.385	0.600	0.282	0.508	0.427	0.627	0.415
14	0.429	0.569	0.336	0.555	0.341	0.466	0.442
15	0.446	0.607	0.278	0.530	0.397	0.537	0.372
16	0.391	0.574	0.342	0.582	0.493	0.583	0.424
17	0.435	0.616	0.276	0.533	0.379	0.540	0.361
Mean	0.421	0.596	0.307	0.554	0.391	0.539	0.400
Std. Dev.		0.035		0.041		0.045	

The JI of GM tissue for EHRM is 4.2% higher than the JI of FAST algorithm and this difference is also statistically significant ($p<0.001$). The volumetric ratio estimation of EHRM for GM is also better than the performance of FAST algorithm ($p<0.0001$). It is observed that EHRM is more successful for the detection of GM tissue and the volumetric estimation. EHRM slightly overestimates the GM whereas FAST underestimates GM significantly.

EHRM is also superior to SK in terms of GM JI with 5.7%. Their volumetric estimation performance is similar. EHRM slightly overestimates GM whereas SK slightly underestimates.

4.3.2 Detection of Caudate, Putamen and Thalamus

The performance of the algorithm on the detection of caudate, putamen and thalamus structures is calculated. The detection ratio is calculated as in Table 4.6.

Table 4.6 Detection Ratios (DR) for Caudate (CA), Putamen (PU) and Thalamus (TH).

Subject No	EHRM DR CA	FAST DR CA	SK DR CA	EHRM DR PU	FAST DR PU	SK DR PU	EHRM DR TH	FAST DR TH	SK DR TH
1	0.945	0.777	0.778	0.791	0.196	0.164	0.645	0.324	0.296
2	0.901	0.794	0.792	0.365	0.188	0.283	0.401	0.318	0.319
3	0.837	0.820	0.907	0.364	0.281	0.769	0.259	0.270	0.335
4	0.947	0.758	0.690	0.733	0.147	0.480	0.625	0.331	0.619
5	0.949	0.830	0.649	0.915	0.385	0.404	0.677	0.371	0.441
6	0.924	0.832	0.875	0.638	0.249	0.732	0.455	0.373	0.387
7	0.915	0.838	0.849	0.617	0.250	0.406	0.544	0.331	0.515
8	0.942	0.711	0.855	0.742	0.071	0.498	0.599	0.240	0.460
9	0.895	0.790	0.814	0.547	0.229	0.392	0.536	0.359	0.397
10	0.698	0.821	0.612	0.233	0.265	0.489	0.305	0.372	0.747
11	0.843	0.795	0.818	0.633	0.233	0.711	0.490	0.337	0.708
12	0.891	0.715	0.405	0.486	0.140	0.268	0.504	0.334	0.297
13	0.876	0.699	0.812	0.493	0.088	0.684	0.414	0.319	0.696
14	0.817	0.708	0.392	0.348	0.111	0.094	0.314	0.288	0.275
15	0.956	0.837	0.819	0.854	0.331	0.662	0.708	0.356	0.664
16	0.717	0.750	0.825	0.301	0.192	0.794	0.338	0.355	0.558

Table 4.6 (Continued)

17	0.938	0.798	0.825	0.757	0.150	0.710	0.723	0.400	0.705
Mean	0.882	0.781	0.748	0.577	0.206	0.502	0.502	0.334	0.495
Std. Dev.	0.078	0.049	0.152	0.207	0.085	0.219	0.148	0.040	0.168

Although the average detection ratios for all 3 sub-cortical GM structures are higher for EHRM algorithm, there is a dramatic gain introduced for the detection of putamen by the use of EHRM. The results indicate that the EHRM is superior than FAST about 37.1% in average for putamen detection ($p<0.00001$). The performance of EHRM is also better than FAST in thalamus detection with 16.8% difference in average($p<0.001$). For caudate structure, the detection ratio is 10.1% better than FAST algorithm ($p<0.001$).

Although SK seems more successful than FAST in average DR values of 3 structures, EHRM is still superior to SK in terms of DR in especially caudate and putamen. The thalamus performance is similar.

4.3.3 CSF and WM Analysis

The JI for CSF and WM tissues are calculated and compared with the FSL FAST and SK algorithms. The results for CSF are presented in Table 4.7. The volumetric ratio of CSF in total sub-cortical area of the EHRM, FAST, SK and the manual segmentation are also presented.

Table 4.7 JI for and Volumetric Ratios for CSF.

Subject No	EHRM - CSF Volume	EHRM-CSF JI	FAST - CSF Volume	FAST - CSF JI	SK CSF Volume	SK CSF JI	CSF Volume
1	0.095	0.629	0.110	0.671	0.092	0.577	0.113
2	0.148	0.732	0.179	0.786	0.156	0.720	0.176

Table 4.7 (Continued)

3	0.052	0.562	0.069	0.624	0.060	0.506	0.074
4	0.092	0.613	0.100	0.652	0.104	0.563	0.110
5	0.116	0.702	0.136	0.735	0.114	0.664	0.133
6	0.079	0.672	0.094	0.720	0.090	0.662	0.096
7	0.062	0.564	0.083	0.631	0.081	0.569	0.080
8	0.135	0.714	0.153	0.757	0.130	0.685	0.151
9	0.084	0.646	0.110	0.705	0.093	0.648	0.101
10	0.060	0.551	0.080	0.600	0.064	0.472	0.084
11	0.047	0.541	0.064	0.625	0.092	0.530	0.070
12	0.083	0.616	0.098	0.672	0.091	0.573	0.110
13	0.084	0.619	0.108	0.690	0.088	0.598	0.111
14	0.064	0.489	0.108	0.669	0.106	0.578	0.108
15	0.091	0.665	0.112	0.698	0.096	0.661	0.098
16	0.097	0.605	0.137	0.728	0.127	0.654	0.143
17	0.046	0.518	0.059	0.589	0.099	0.436	0.069
Mean	0.084	0.614	0.106	0.680	0.099	0.594	0.107
Std Dev.		0.070		0.056		0.078	

The average JI for EHRM is lower for than the FAST for CSF tissue with 6.4%($p<0.00001$). Although the difference is statistically strong, the average volumetric differences of 2 algorithms are 2.2% of sub-cortical volume, which is not very dramatic. The similarity of the models for the Freesurfer software and FAST, which both depend on MRF models, may be an important factor for the perfect similarity between the Freesurfer and FAST for CSF tissue. Another reason for the superior performance of FAST may be due to the smoothness imposed by the MRF

model, which helps especially in CSF – WM boundaries. EHRM can segment the border which have intermediate intensities as GM, while FAST better handle such boundaries with the MRF model.

If the volumetric ratio estimation of EHRM and FAST is compared, the performance of FAST is superior to EHRM ($p<0.0001$). The MRF model of FAST may be an important reason for the superior performance but in any case the volumetric estimation error is 1.75% in average more for EHRM, which is not very critical.

The JI for CSF is better for EHRM when compared with the SK algorithm. The SK algorithm seems to slightly underestimate CSF and is more successful than EHRM in average. Similar analysis is conducted for WM and the results are presented in Table 4.8.

Table 4.8 JI for and Volumetric Ratios for WM.

Subject No	EHRM - WM Volume	EHRM-WM JI	FAST – WM Volume	FAST - WM JI	SK WM Volume	SK WM JI	WM Volume
1	0.407	0.699	0.548	0.712	0.564	0.683	0.446
2	0.462	0.685	0.516	0.697	0.480	0.615	0.418
3	0.596	0.723	0.610	0.736	0.504	0.667	0.519
4	0.429	0.686	0.580	0.714	0.472	0.650	0.475
5	0.379	0.656	0.526	0.722	0.502	0.645	0.466
6	0.556	0.770	0.623	0.781	0.550	0.735	0.561
7	0.570	0.752	0.654	0.744	0.597	0.707	0.557
8	0.347	0.517	0.602	0.667	0.425	0.521	0.451
9	0.531	0.751	0.605	0.753	0.606	0.730	0.522
10	0.579	0.723	0.570	0.741	0.482	0.679	0.500
11	0.545	0.722	0.633	0.736	0.496	0.688	0.543

Table 4.8 (Continued)

12	0.478	0.744	0.555	0.747	0.541	0.674	0.467
13	0.530	0.728	0.609	0.695	0.485	0.727	0.474
14	0.507	0.677	0.556	0.684	0.553	0.609	0.450
15	0.463	0.714	0.611	0.741	0.507	0.675	0.530
16	0.512	0.701	0.521	0.703	0.380	0.609	0.433
17	0.520	0.748	0.665	0.761	0.523	0.721	0.570
Mean	0.495	0.706	0.587	0.726	0.510	0.667	0.493
Std Dev.		0.058		0.030		0.055	

The WM JI of FAST is statistically superior to EHRM with 2.0% ($p=0.05$). The WM volumetric ratio estimations of EHRM is quite better than FAST ($p<0.0001$).

The JI of WM is better for EHRM when compared with SK. The SK algorithm seems to slightly overestimate WM and EHRM performs a more successful volumetric estimation in average.

4.4 Robustness to Bias Field and Noise

The effect of bias field and the Signal to Noise Ratio (SNR) in MR images differ for different scanners, sequences, acquisition parameters, resolution, RF and receiver coils. Therefore, robustness to bias field and noise is a crucial property for a brain segmentation algorithm.

Different bias field distributions and noise levels can be added to MR images using simulations. Brainweb database [51] provides such simulations for an anatomical image model with different bias field and noise levels [52]-[55]. The manual segmentation of the image is also available, which allows the researchers to

test the performance of segmentation algorithms for various bias field and noise levels.

3 levels of bias field and 3 levels of additive Gaussian noise are analyzed. EHRM algorithm is operated for the 2 sets of 3 images. In one set, the noise level is same and the bias field level is different among images. Similarly in the other set, the bias field is same and the noise levels are different. The algorithm is operated 10 times on a single image and the avarage performances are evaluated. The purpose of conducting repeated experiments is to reduce the bias due to the randomness of the evolutionary algoritm.

The manual segmentation of the images for 3 tissues, rather than specific anatomical structures, is available so quantitative analysis is performed for the JI and volumetric estimations of these 3 tissues.

Various bias field and noise levels are represented with percentages. The percentage of the bias field represents the percentage of intensity decrease in the parts of the image which have the most underestimated intensities compared to intensity of the parts which have the most overestimated intensities. For instance, in the presence of 20% bias field, a voxel which is represented with an intensity value of 800 at the brightest part of an image is represented with 640 intensity in the parts which are most affected from the bias field. On the other hand, the percentage of noise determines the standard deviation of the additive Gaussian noise. The value of standard deviation is determined by calculating the desired percentage of the maximum intensity in the noiseless image. In an image where the highest intensity is 500, 5% noise corresponds to a standard deviation of 25.

The volumetric estimations and JI of both algorithms for CSF, GM and WM tissues are displayed in Figures 4.19, 4.20 and 4.21 respectively. When various noise levels are analyzed, the bias field level is set to 20%. When various bias fields are analyzed, the noise level is set to 3%. The volumes of tissues are represented with the ratio of tissue volumes to the volume of whole sub-cortical region. The error plots show the average and standard deviation of the JI and volumetric estimations.

It is observed the volumetric estimations of CSF are quite insensitive to noise and bias field. The JI of CSF decreases with increasing noise and bias field as expected but the reduction of performance does not seem so dramatic.

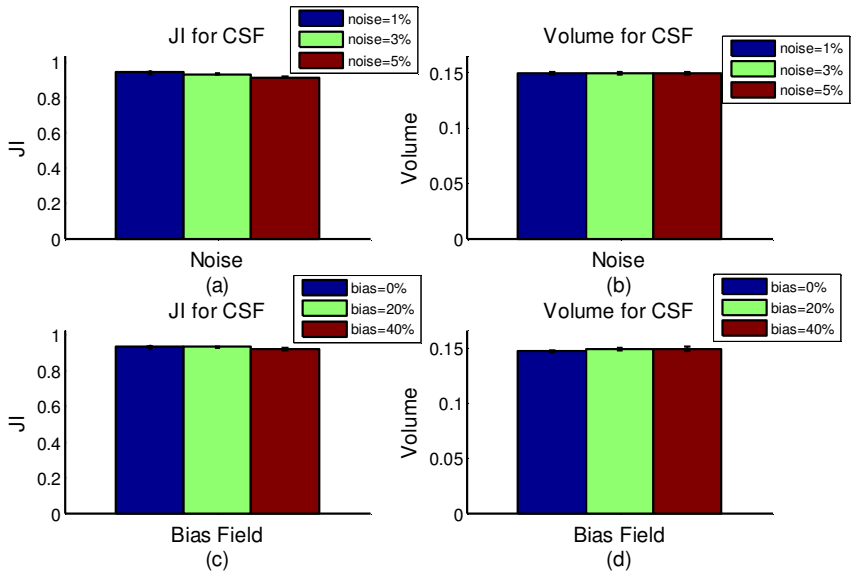


Figure 4.19 JI and Volumetric estimation for CSF at different bias field and noise levels. (a) and (b) are JI and volumetric estimations for various noise levels. (c) and (d) JI and volumetric estimations for various bias field levels.

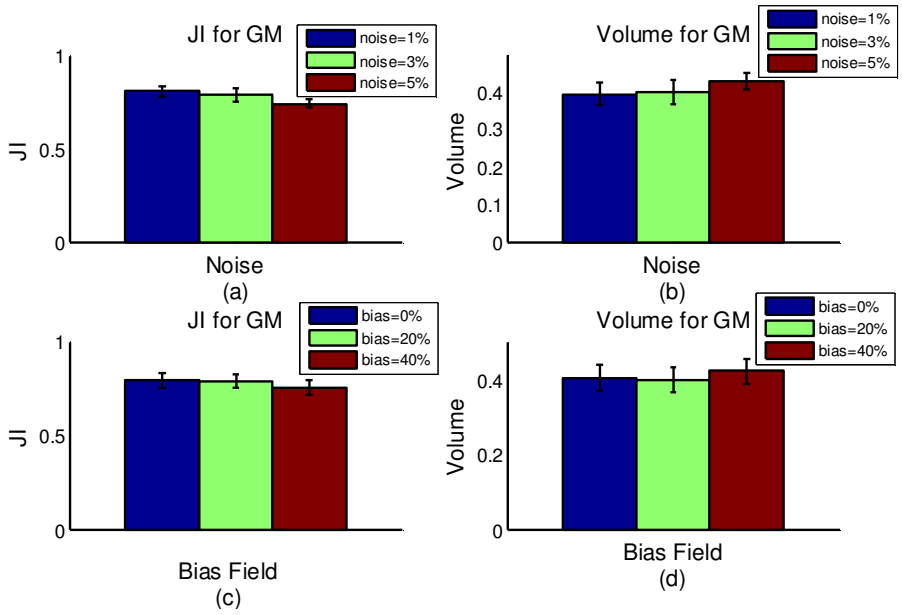


Figure 4.20 JI and Volumetric estimation for GM at different bias field and noise levels. (a) and (b) are JI and volumetric estimations for various noise levels. (c) and (d) JI and volumetric estimations for various bias field levels.

It is observed the volumetric estimations of GM are quite insensitive to noise and bias field. Although there are some changes, these small changes does not seem significant when the error plots are considered. The JI of GM decreases with increasing noise and bias field as expected but the reduction of performance is not extreme.

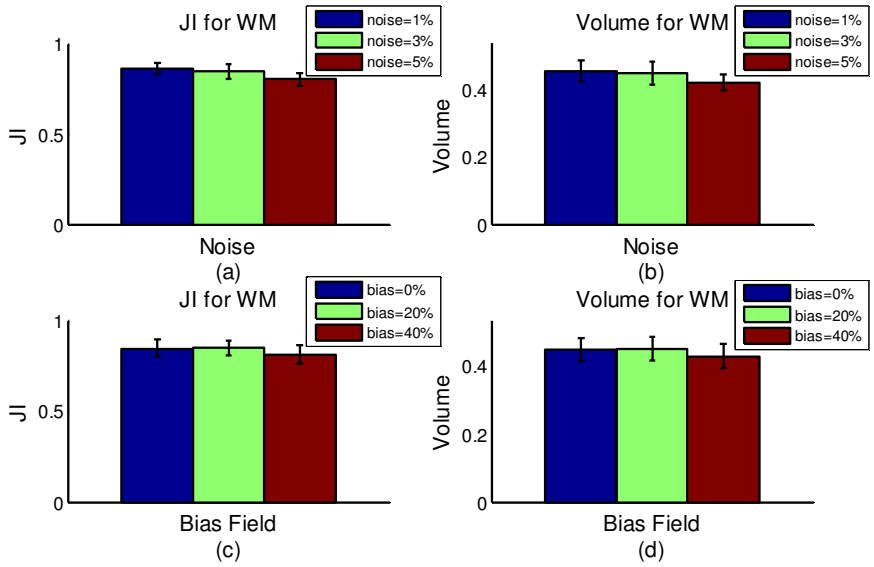


Figure 4.21 JI and Volumetric estimation for WM at different bias field and noise levels. (a) and (b) are JI and volumetric estimations for various noise levels. (c) and (d) JI and volumetric estimations for various bias field levels.

It is observed the volumetric estimations of WM seem insensitive to bias field and slightly sensitive to noise. Although there are some changes, these small changes does not seem significant when the error plots are considered. The JI of WM decreases slightly with increasing noise and bias field as expected but the changes are not so dramatic.

The JI values show that the bias field does not have a strong affect on JI. It is seen that bias field is not a serious problem for EHRM even if no bias field removal is performed. This is expected because the sub-cortical region is a small part of the brain, which limits the amount of bias field variation from one location of sub-cortical region to another location. However, noise seems to be a more serious problem compared to bias field. The JI values decrease with increasing noise but the changes are not so dramatic. In any case, the volumetric estimations are not affected much from bias field and noise. The estimations are quite robust.

The behavior of EHRM to various noise and bias field conditions is investigated with repeated experiments. It seems that the performance of EHRM is not significantly affected by various bias field and SNR.

4.5 Complexity Analysis of EHRM

Our algorithm operates by the region merging operations conducted between regions. At the beginning of the algorithm, each voxel is a region. Regardless of the image resolution, each voxel, except the border voxels in the image, has 6 neighbors. Therefore, the number of region merging trials and the number of successful region merges are linearly proportional to the number of voxels in the image (O(n)). As epochs advance, the regions get larger, the number of regions decreases but the average number of neighbors for a region increases. The number of decrease in regions is faster than the increase of average neighbor number so the total number of trials in an epoch tends to decrease as epochs advance. The cost of each region merging trial is the same regardless of the sizes of the regions if the effect of texture feature is neglected, which is incorporated for a small number of trials which can be neglected in total trials. Although the cost of necessary updates after a single merge increases as epochs advance, the number of total merges in a single epoch decrease with a faster rate. Therefore, the computational complexity of epochs decreases with increasing iterations. In this case, the complexity of the first epoch, which has the highest number of trials and region merges, determines the complexity of the entire approach. Theoretically, the complexity of the algorithm is directly proportional to the number of voxels in the image (O(n), n is the number of voxels). Since the algorithm operates only on the extracted sub-cortical box, $n << M$, where n depicts the number of voxels within this sub-cortical box, and M depicts the number of voxels within the entire image.

The operation time of EHRM takes about 30 minutes in the MATLAB [56] implementation for a conventional resolution of 1mm cubic voxels. Observations on images with different resolutions demonstrated that the computation time is roughly linear with the total number of voxels, as expected from the complexity analysis.

The algorithms are operated on an ordinary netbook having a 1 GB of RAM and a 1.73 GHz processor. The operation time in MATLAB is actually misleading to compare with other methods which are implemented in faster programming languages. The presence of nested loops in EHRM is a major reason for the current operation time. MATLAB is known to be very slow for operating loops, therefore the implementation of EHRM in a faster programming language is expected to produce a running time of approximately a few minutes.

In comparison, FAST operates at the whole brain with similar resolution about 10-15 minutes in a desktop with a 4GB RAM and a 4GHz processor, which is obviously a much more powerful computer than the netbook that EHRM is operated. Therefore, it seems that running time of EHRM algorithm on the sub-cortical region is acceptable.

CHAPTER 5

CONCLUSION AND FUTURE WORK

The segmentation of sub-cortical GM structures is very challenging because the anatomy of this region is more complex and the contrast is lower. Structures like thalamus, putamen and globus pallidus, which are termed as a part of GM tissue, have a tissue composition different than the cortical GM. Sub-cortical GM tissue is not pure and also contains WM fibers. Therefore, the intensities of these composite tissues approach to the intensity levels of WM, causing a reduction in contrast. Modeling of all GM tissue in the brain as a single class may not be an appropriate representation of sub-cortical GM. Therefore, the algorithms which segment the brain into 3 tissues is usually not successful in these structures and segment them as a part of WM tissue. This necessitates the handling of sub-cortical segmentation in a different way from the rest of the brain. There are also other approaches that segment each anatomical structure separately but these methods have high computational demand and impose high spatial constraints which can limit their wide range usage.

In this study, an evolutionary hierarchical region merging (EHRM) algorithm is proposed to improve the segmentation quality in the sub-cortical GM structures in MR brain images. The image segmentation algorithm proposed by Veenman et al. [3] is used as a starting point for EHRM. In this evolutionary segmentation algorithm, region merging and border pixel transfer operations are utilized depending on the intensity variance feature of separate regions. These operations are

conducted iteratively to reach to the final regions. An evolutionary framework is used to reduce the computation time by preventing the number of unsuccessful trials which do not contribute to the segmentation process. The border pixel transfer operation, which significantly increases computational complexity, is not utilized in EHRM. Instead, region merging in a hierarchical framework is proposed as a single operation to guide the segmentation. The hierarchical approach allows gradual formation of regions and is expected to represent the anatomical structures more coherently. Also, the necessity of border pixel transfer is reduced because the regions maturate gradually. Region merging in EHRM simplifies the image quite successfully to a certain number of regions.

The performance of the algorithm is presented in a comparative manner with the FAST algorithm in 3 different datasets: Brainweb simulated data with various noise and bias field conditions, the IBSR manually segmented dataset [57], and a real MRI dataset where manual segmentation is not available.

Although some findings are questionable due to the errors in the ground truth segmentation within IBSR, some important improvements of EHRM approach can be clearly observed. The detection of sub-cortical GM structures like thalamus, caudate and putamen are analyzed quantitatively in 2 datasets. In both datasets, the performance of EHRM is significantly superior to FAST. Average improvements of 10%, 36% and 22% are achieved for caudate, putamen and thalamus respectively. Also improvement in the JI of GM tissue is observed in both datasets. However, the amount of improvement in overall sub-cortical GM JI is less because of the good performance of FAST in structures other than the target GM structures, where smoothness imposed by the MRF model is useful.

EHRM is also superior to SK in many aspects. The JI and DR values are higher for all tested cases. The volumetric performance is also similar. Therefore, it is realized that EHRM is more successful than widely known k-means algorithm which also incorporates spatial information.

EHRM is also more successful than FAST in the estimation of tissue volumes. GM is consistently underestimated and WM is consistently overestimated by FAST. Whereas EHRM can slightly overestimate GM but it is shown to make more accurate estimations especially for GM volume. The robustness of EHRM approach to the initialization of 2 important parameters (the size threshold for intermediate regions and the number of regions desired at the end of region merging) is questioned. It is observed that the performance is quite insensitive to the initialization of these parameters although some stability problems are observed when the threshold parameter guiding the hierarchy of regions is high.

The robustness of the algorithm to various bias field and noise conditions are examined on the datasets provided by the Brainweb database [51]. The algorithm seems to be sufficiently robust to the bias field and noise variations.

The running time of the algorithm is quite feasible because it depends only on the raw intensity information and simple operations within the evolutionary framework. If a successful segmentation of the whole brain is desired, EHRM algorithm can be fused with another algorithm that operates successfully in the cortical region. In this hybrid framework, both algorithms should be operated. Then the sub-cortical segmentation of EHRM and cortical segmentation of another algorithm can be combined to yield better segmentation. The only problem with this approach is the discontinuity in the segmentation in the borders of the extracted sub-cortical region.

In future studies, some issues need to be further investigated and studied to increase the performance, robustness and usability of the EHRM algorithm: 1. A more reliable and robust texture heterogeneity feature should be implemented. 2. The current method for the determination of threshold limit can be modified or replaced with another approach. The threshold limit for intensity similarity criterion is observed to be important when the segmentations are visually inspected. This limit is sometimes so flexible that the sub-cortical GM and WM combine with region merging. The final classification to 3 tissues may not be able to fix this error

and the segmentation quality decreases inevitably. Therefore, the threshold limit should be set more conservatively. 3. Some details of the final classification step are intuitive and its robustness is questionable. Although no serious problem is observed for the tested images, the performance can increase a little bit more with a better approach. 4. The integration of EHRM with other algorithms is an important issue. To obtain a smooth transition between cortical and sub-cortical segmentations, the transition regions can be segmented in a smooth manner with the contribution of both segmentations. A new approach for this fusion may be developed. 5. A cluster index can be proposed to determine the optimum number of regions that should be left at the end of region merging process. This issue is actually searched but the optimal cluster number depends on the dispersion measure of regions. However, it is very difficult to extract reliable dispersion measures that are not affected from PVE. A reliable cluster index would help the algorithm to operate with a sound background.

In this study, we proposed a unique method that is fast and effective for segmenting the sub-cortical region which is problematic for whole brain segmentation algorithms. Significant improvements are observed in the detection of three major sub-cortical GM structures, caudate, putamen and thalamus. In addition, EHRM does not necessitate any preprocessing step. Only the intensity information is utilized and no spatial or intensity priors are incorporated. Therefore, EHRM proposes a general framework for the sub-cortical segmentation problem, which can increase the detection of sub-cortical GM structures. The proposed method can be easily fused with plenty of successful approaches, which perform whole brain segmentation.

REFERENCES

[1] 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*, vol. 20, pp. 45–57, 2001.

[2] B. Fisch, D.H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrov, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A.M. Dale, "Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, pp. 341–355, 2002.

[3] C. J. Veenman, M.J.T. Reinders, and E. Backer, "A cellular coevolutionary algorithm for image segmentation," *IEEE Transactions on Image Processing*, vol. 12, pp. 304-316, 2003.

[4] M.A. Balafar, A.R. Ramli, M.I. Saripan, and S. Mashohor, "Review of brain MRI image segmentation methods," *Artificial Intelligence Review*, vol. 33, pp. 261-274, 2010.

[5] A. W. Liew, and H. Yan, "Current Methods in the Automatic Tissue Segmentation of 3D Magnetic Resonance Brain Images," *Current Medical Imaging Reviews*, vol. 2, pp. 1-13, 2006.

[6] M. Chupin, A. Hammers, R.S.N. Liu, O. Colliot, J. Burdett, E. Bardinet, J.S. Duncan, L. Garnero, and L. Lemieux, "Automatic segmentation of the hippocampus and the amygdala driven hybrid constraints: Method and validation," *Neuroimage*, vol. 46, pp. 749-761,2009.

[7] M.C. Clark, L.O. Hall, D.B. Goldgof, R. Velthuizen, M.R. Murtagh, and M.S. Silbiger, "Automatic Tumor Segmentation Using Knowledge-Based Techniques," *IEEE Transactions on Medical Imaging*, vol. 17, pp. 187-201, 1998.

[8] M.B. Cuadra, C. Pollo, A. Bardera, O. Cuisenaire, J. Villemure, and J. Thiran, "Atlas-Based Segmentation of Pathological MR Brain Images Using a Model of Lesion Growth," *IEEE Transactions on Medical Imaging*, vol. 23, pp. 1301-1314, 2004.

[9] W.M. Wells, W.L. Grimson, R. Kikinis, and F.A. Jolesz, "Adaptive segmentation of MRI data," *IEEE Transactions on Medical Imaging*, vol. 15, pp. 429–442, 1996.

[10] D.W. Shattuck, S.R. Sandor-Leahy, K.A. Schaper, D.A. Rottenberg, and R.M. Leahy, R.M., "Magnetic resonance image tissue classification using a partial volume model," *Neuroimage*, vol. 13, pp. 856–876, 2001.

[11] Z. Yi, A. Criminisi, J. Shotton, and A. Blake, "Discriminative, Semantic Segmentation of Brain Tissue in MR Images," *in Proceedings of MICCAI*, vol. 1, pp. 558-565, 2009.

[12] J. Ashburner, and K.J. Friston, "Unified segmentation," *Neuroimage*, vol. 26, pp. 839–851, 2005.

[13] J. Tohka, I.D. Dinov, D.W. Shattuck, and A.W. Toga, "Brain MRI tissue classification based on local Markov random fields," *Magnetic Resonance Imaging*, vol. 28, pp. 557- 573, 2010.

[14] A. Mayer, and H. Greenspan, "An Adaptive Mean-Shift Framework for MRI Brain Segmentation," *IEEE Transactions on Medical Imaging*, vol. 28, pp. 1238–1250, 2009.

[15] J. Xuan, T. Adali, and Y. Wang, "Segmentation of Magnetic Resonance Brain Image: Integrating Region Growing and Edge Detection," *in Proceedings of International Conference on Image Processing (ICIP)*, vol. 3, pp. 544-547, 1995.

[16] T. Kapur, W.E.L. Grimson, W.M. Wells, and R. Kikinis, "Segmentation of brain tissue from magnetic resonance images," *Medical Image Analysis*, vol. 1, pp. 109-127, 1996.

[17] S. Shen, W. Sandham, M. Granat, and A. Sterr, "MRI Fuzzy Segmentation of Brain Tissue Using Neighborhood Attraction With Neural-Network Optimization," *IEEE Transactions on Information Technology in Biomedicine*, vol. 9, pp. 459-467, 2005.

[18] H. Greenspan, A. Ruf, and J. Goldberger, "Constrained Gaussian Mixture Model Framework for Automatic Segmentation of MR Brain Images," *IEEE Transactions on Medical Imaging*, vol. 25, pp. 1233–1245, 2006.

[19] D.L. Pham, and J.L. Prince, "Adaptive Fuzzy Segmentation of Magnetic Resonance Images," *IEEE Transactions on Medical Imaging*, vol. 18, pp. 737–752, 1999.

[20] B. Patenaude, S.M. Smith, D.N. Kennedy, and M. Jenkinson, "A Bayesian model of shape and appearance for subcortical brain segmentation," *Neuroimage*, vol. 56, pp. 907-922, 2011.

[21] J. Zhou, and J.C. Rajapakse, "Segmentation of subcortical brain structures using fuzzy templates," *Neuroimage*, vol. 28, pp. 915-924, 2005.

[22] J.C. Corso, Z. Tu, A.L. Yuille, and A.W. Toga, "Segmentation of Sub-cortical Structures by the Graph-Shifts Algorithm," *in Proceedings of IPMI*, vol. 1, pp. 183-197, 2007.

[23] Z. Tu, K.L. Narr, P. Dollar, I. Dinov, P.M. Thompson, and A.W. Toga, "Brain Anatomical Structure Segmentation by Hybrid Discriminative/Generative Models," *IEEE Transactions on Medical Imaging*, vol. 27, pp. 495-508, 2008.

[24] V. Barra, and J. Boire, "Automatic Segmentation of Subcortical Brain Structures in MR Images Using Information Fusion," *IEEE Transactions on Medical Imaging*, vol. 20, pp. 549-558, 2001.

[25] M.R. Sabuncu, B.T.T. Yeo, K.V. Leemput, B. Fischl, and P. Golland, "A Generative Model for Image Segmentation," *IEEE Transactions on Medical Imaging*, vol. 29, pp. 1714-1729, 2010.

[26] J.H. Morra, Z. Tu, L.G. Apostolova, A.E. Green, A.W. Toga, and P.M. Thompson, P.M. "Automatic Subcortical Segmentation Using a Contextual Model," *in Proceedings of MICCAI 2008 Workshop on Computational Anatomy and Physiology of the Hippocampus (CAPH)*, pp. 97-104, 2008.

[27] B. Fischl, A. van der Kouwe, C. Destrieux, F. Halgren, F. Segonne, D.H. Salat, E. Busa, L.J. Seidman, J. Goldstein, D. Kennedy, V. Caviness, N. Makris, B. Rosen, and A.M. Dale, "Automatically Parcellating the Human Cerebral Cortex," *Cerebral Cortex*, vol. 14, pp. 11-22, 2004.

[28] Athinoula A. Martinos Center for Biomedical Imaging (2005). FreeSurfer. [Online]. Available: http://surfer.nmr.mgh.harvard.edu/

[29] Wellcome Trust Centre for Neuroimaging (2011). Statistical Parametric Mapping. [Online]. Available: http://www.fil.ion.ucl.ac.uk/spm/.

[30] FMRIB Analysis Group (2008). FMRIB Software Library. (2008). [Online]. Available: http://www.fmrib.ox.ac.uk/fsl/.

[31] R. Guillemaud, and J.M. Brady, "Estimating the bias field of MR images," *IEEE Transactions on Medical Imaging*, vol. 16, pp. 238–251, 1997.

[32] N. Otsu, "A threshold selection method from gray-level histogram," *IEEE Transactions on Systems Man and Cybernetics*, vol. 9, pp. 62–66, 1979.

[33] J. Besag "On the statistical analysis of dirty pictures (with discussion)," *Journal of Royal Statistical Society*, ser. B, vol 48, pp. 259–302, 1986.

[34] X. Han, and B. Fischl, "Atlas Renormalization for Improved Brain MR Image Segmentation Across Scanner Platforms," *IEEE Transactions on Medical Imaging*, vol. 26, pp. 479-486, 2007.

[35] Athinoula A. Martinos Center for Biomedical Imaging (2007). Freesurfer Tutorial. [Online]. Available: http://surfer.nmr.mgh.harvard.edu/fswiki.

[36] Z. Li, F. Huang, and Y. Liu, "A Method of Motion Segmentation Based on Region Shrinking," *in Proceedings of IDEAL*, pp. 275-282, 2006.

[37] J. Wu, S. Poehlman, M.D. Noseworthy, and M.V. Kamath, M.V, "Texture feature based automated region growing in abdominal MRI segmentation," *Journal of Biomedical Science and Engineering*, vol. 2, pp. 1-8, 2009.

[38] R. Ohlander, K. Price, and D.R. Reddy, "Picture Segmentation Using A Recursive Region Splitting Method," *Computer Graphics and Image Processing*, vol. 8, pp. 313-333, 1978.

[39] F. Calderero, and F. Marques, "Region Merging Techniques Using Information Theory Statistical Measures," *IEEE Transactions on Image Processing*, vol. 19, pp. 1567- 1586, 2010.

[40] S.L. Horowitz, and T. Pavlidis, "Picture Segmentation by a Tree Traversal Algorithm," *Journal of the Association for Computing Machinery*, vol. 23, pp. 368-388, 1976.

[41] Laboratory of Neuro Imaging (2008). ICBM Template. [Online]. Available: http://www.loni.ucla.edu/Atlases/Atlas_Detail.jsp?atlas_id=5.

[42] M. Jenkinson, and S. Smith, "A global optimisation method for robust affine registration of brain images," *Medical Image Analysis*, vol. 5, pp. 143-156, 2001.

[43] U. Vovk, F. Pernus, and B. Likar, "A Review of Methods for Correction of Intensity Inhomogeneity in MRI. *IEEE Transactions on Medical Imaging*, vol. 26, pp. 405-421, 2007.

[44] R.M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 6, pp. 610-621, 1973.

[45] K.B. Walhovd, A.M. Fjell, I. Reinvang, A. Lundervold, A.M. Dale, D.E. Eilertsen, B.T. Quinn, D. Salat, N. Makris, and B. Fischl, "Effects of age on volumes of cortex, white matter and subcortical structures," *Neurobiology of Aging*, vol. 26, pp. 1261-1270, 2005.

[46] K.B. Walhovd, L.T, Westlye, I. Amlien, T. Espeseth, I. Reinvang, N. Raz, I. Agartz, D.H. Salat, D.N. Greve, B. Fischl, A.M. Dale, and A.M. Fjell, "Consistent neuroanatomical age-related volume differences across multiple samples," *Neurobiology of Aging*, vol. 32, pp. 916-932, 2011.

[47] J. Canny "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 679-698, 1986.

[48] R. Alpar, Spor Bilimlerinde Uygulamalı İstatistik. Ankara: Nobel Yayın Dağıtım, 2006.

[49] S.M. Smith "Fast Robust Automated Brain Extraction," *in Proceedings of Human Brain Mapping*, vol. 17, pp. 143-155, 2002.

[50] P. Jaccard "The distribution of the flora of the alpine zone," *New Phytologist*, vol. 11, pp. 37-50, 1912.

[51] McConnell Brain Imaging Center (2006). Brainweb: Simulated Brain Database. [Online]. Available: http://mouldy.bic.mni.mcgill.ca/brainweb/

[52] C.A. Cocosco, V. Kollokian, R.K.-S. Kwan, and A.C. Evans, "BrainWeb: Online Interface to a 3D MRI Simulated Brain Database," *in International Conference on Functional Mapping of the Human Brain*, Copenhagen, May 1997.

[53] R.K.-S. Kwan, A.C. Evans, and G.B. Pike, "MRI simulation-based evaluation of image-processing and classification methods", *IEEE Transactions on Medical Imaging*, vol. 18, pp. 1085-1097, 1999.

[54] R.K.-S. Kwan, A.C. Evans, and G.B. Pike, "An Extensible MRI Simulator for Post-Processing Evaluation", *Lecture Notes in Computer Science*, vol. 1131, pp. 135-140, 1996.

[55] D.L. Collins, A.P. Zijdenbos, V. Kollokian, J.G. Sled, N.J. Kabani, C.J. Holmes, and A.C. Evans, "Design and Construction of a Realistic Digital Brain Phantom," *IEEE Transactions on Medical Imaging*, vol. 17, pp. 463-468, 1998.

[56] MathWorks (2011). MATLAB – The Language of Technical Computing. [Online]. Available: http://www.mathworks.com/products/matlab/.

[57] Center for Morphometric Analysis (2009). Internet Brain Segmentation Repository. [Online]. Available: http://www.cma.mgh.harvard.edu/ibsr/.

[58] Wikipedia, The Free Encylopedia (2011). Gaussian Integral. [Online]. Available: http://en.wikipedia.org/wiki/Gaussian_integral.

[59] J. Ashburner, and K.J. Friston, "Nonlinear spatial normalization using basis functions", *in Proceedings of Human Brain Mapping*, vol. 7, pp. 254–266, 1999.

APPENDICES

APPENDIX A: BRAIN ANATOMY ON A SAMPLE SLICE

The brain is composed of 3 main tissues, which are CSF, GM and WM. The outer part of the brain is called the cortex. The inner region, where a plenty of GM structures are present, is called the sub-cortical region. The main structures of the brain, which are mentioned in the thesis, are shown on a sample axial slice in Figure A.1 taken from [2]. The tissue type of structures is also indicated in parenthesis.

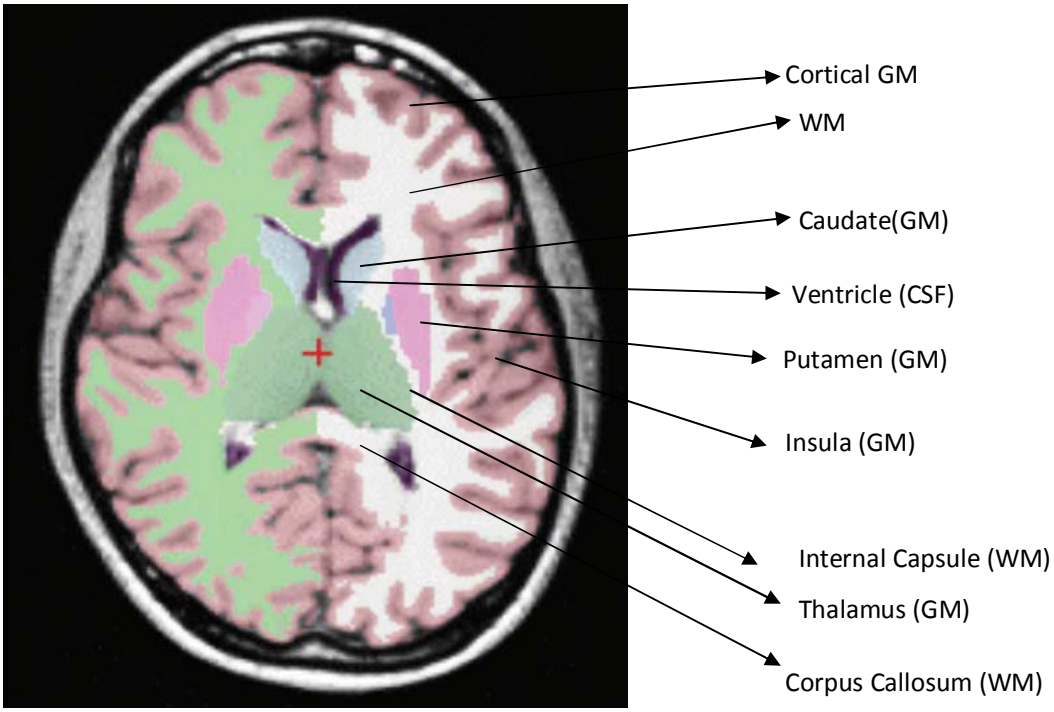


Figure A.1 An axial brain slice with labeled structures. Figure taken from [2].

APPENDIX B: TEXTURE FEATURE DERIVATION

The texture feature TH is shown to be a linear function of variance for a Gaussian distribution. The derivation is based on the generalized form of Gaussian integral [58].

The TH for discrete random variables are expressed as below.

$$TH = (1/(\sum_{i=0}^{G_{max}} p(i)^2)) \quad \text{(B.1)}$$

The continuous form of TH can be expressed as follows:

$$TH = (1/(\int_0^{\infty} p(i)^2 di)) \quad \text{(B.2)}$$

For the right side of a Gaussian distribution with zero mean, TH can be written as

$$TH = (1/(\int_0^{\infty} (\frac{1}{\sqrt{2\pi\sigma^2}} e^{-i^2/2\sigma^2})^2 di)) = (1/(\int_0^{\infty} \frac{1}{2\pi\sigma^2} e^{-i^2/\sigma^2} di)) \quad \text{(B.3)}$$

The general form of Gaussian integral can be calculated as follows:

$$\int_{-\infty}^{\infty} ae^{-(x+b)^2/c^2} dx = ac\sqrt{\pi} \quad \text{(B.4)}$$

The effect of variance to the TH feature can be analytically calculated using this generalization.

Substituting $a = \frac{1}{2\pi\sigma^2}$, $b = 0$ and $c = \sigma$, the dependence of TH feature on the standard deviation σ can be determined.

$$TH \propto (\frac{1}{ac\sqrt{\pi}}) = (\frac{2\pi\sigma^2)}{\sigma\sqrt{\pi}}) = (\frac{2\pi\sigma}{\pi}) \quad \text{(B.5)}$$

It is derived that the TH value is a linear function of the standard deviation σ.

APPENDIX C: SK ALGORITHM

A spatial k-means algoritm, abbreviated as SK, is implemented to compare with the proposed EHRM algorithm. It is a simple algorithm which incorporates spatial coordinate features in addition to the intensity features.

The steps of the SK algorithm can be listed as follows:

- Form the 4D (1 intensity + 3 spatial coordinate features) data vectors.
- Normalize each feature to [0,1] range using a linear transformation.
- There are 3 classes for intensity feature, which represent individual tissues. Roughly assume 3 classes for each individual spatial features. Therefore, the number of clusters are set to 3x3x3x3=81.
- With 81 target clusters and 4D features, operate the k-means algorithm.
- Obtain the class labels and class centers when the k-means algorithm converges.

After 81 classes are left, reduce 81 classes to 3 tissues using the same methodology as the reduction implemented in the EHRM algorithm. A slight difference is that smoothing of the histogram peaks are not conducted because the information on the dispersion of the classes is not available. The number of classes are quite high compared to the EHRM, therefore the absence of smoothing is not expected to change the results dramatically.