MR IMAGE CONTRAST SYNTHESIS FOR
CONSISTENT SEGMENTATION

by

Snehashis Roy

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

January, 2013

© Snehashis Roy 2013

All rights reserved
Abstract

Magnetic resonance (MR) is a noninvasive imaging modality that has been widely used to image the human brain. Many image processing algorithms, such as segmentation and registration, when applied to MR images, provide insights about brain tissues that are used to further the understanding of normal aging, as well as the detection and progression of diseases like Multiple Sclerosis and Alzheimer’s Disease. Brain image segmentation divides brain images into several major tissues, e.g., cerebro-spinal fluid, gray matter and white matter. Most segmentation techniques use image intensities as primary features. However, unlike computed tomography, MR intensities are not calibrated to represent a specific tissue property and they vary widely in both range and distribution, depending on the physical properties of the scanners and the pulse sequences used to image the brain. Image processing results are usually affected by inconsistencies due to these variations.

The research presented here primarily focuses on normalizing scans acquired in a variety of scanners and pulse sequences to achieve consistency in their segmentations, so that a large pool of data from various sources can be consistently analyzed for fur-
ABSTRACT

ther understanding. First, we improve an existing fuzzy C-means tissue segmentation algorithm and then develop a Rician model based segmentation method to account for the tissue intensity variations arising from the differences in image acquisition protocols. We show that more consistent segmentations can be achieved from two popular pulse sequences, namely spoiled gradient recalled (SPGR) and magnetization prepared rapid gradient echo (MPRAGE). Then we develop an Magnetic resonance Image Example based Contrast Synthesis (MIMECS) technique, that uses sparsely distributed patches from multi-contrast atlases to normalize images acquired from a variety of scanners and pulse sequences. It is also used to synthesize new tissue contrasts, similar to simulating new pulse sequences, for the purpose of consistent and improved segmentation if there is any missing scans, usually seen in the case of large longitudinal studies. Next, an image normalization algorithm is developed for the purpose of consistent normalization and segmentation across a longitudinal dataset. All the normalization methods developed are pre-processing steps to any segmentation algorithm and differ from their predecessors in two ways, they are not dependent on any particular segmentation algorithm and they do not require any scanner specific parameters. Segmentation results on both normal and diseased subjects show superior performance compared to existing methods.

Primary Reader: Dr. Jerry L. Prince
Secondary Reader: Dr. Trac D. Tran, Dr. Mounya Elhilali
ABSTRACT
Acknowledgments

This dissertation brings to a close a journey which started in August 2006, when I attempted this uphill task. It has been a long arduous journey and I am greatly thankful to a lot of people for helping me complete it.

I thank Dr. Jerry L. Prince, my research advisor, for his encouragement, guidance and support. It has been an honor to work with him. I really appreciate his intellectual help and financial support throughout the course of this research.

I express my sincere thank to Aaron Carass for his untiring patience and valuable suggestions on my thesis. His feedback as well as our endless discussions on my work kept me motivated and striving. I would also like to thank Dr. Trac D. Tran and Dr. Mounya Elhilali for patiently reading my thesis and serving in my thesis committee. I would like to thank Dr. Susan Resnick for providing me with the data without which the research would not have been possible.

I am grateful to all the past and present members of the Image Analysis and Communications Lab. Dr. Pierre-Louis Bazin, Dr. Dzung Pham, Dr. Emi Murano, Dr. Sarah Ying, Dr. Lotta Ellingsen and Dr. Ying Bai gave me many useful pieces
ACKNOWLEDGMENTS

advices during stages of my PhD study. I thank Dr. Harsh Agarwal for many helpful discussions and also for being my good friend. I thank John Bogovic, Navid Shiee, Amod Jog and Min Chen for their encouragement and inspiration.

Finally, I want to thank my parents, my bulwarks, who offered me endless support, guidance and advice and being there for me whenever I needed. I thank my wife who was always by my side all these years to share the smooth stretches and smooth out the kinks of graduate life. To them I dedicate this thesis.
Contents

Abstract i

Acknowledgments v

List of Tables xii

List of Figures xiii

1 Introduction 1

2 Background 7

2.1 MR Image Contrast 7

2.2 Statistical modeling of Image Intensities 10

2.3 Image Classification Models 12

2.3.1 Gaussian Mixture Model 13

2.3.2 Fuzzy C-means 19

2.4 Summary 22
## 3 Methods for Consistent Image Segmentation

3.1 Introduction .................................................. 23

3.2 Fuzzy C-means with Variable Compactness .................. 26

3.2.1 Previous Works ........................................... 27

3.2.2 Compactness Parameters .................................. 28

3.2.3 Integration with FANTASM .............................. 35

3.2.4 Estimation of Compactness Parameters ....................... 36

3.2.5 Segmentation Consistency .............................. 38

3.2.6 Discussion ................................................. 40

3.3 Consistent Segmentation using a Rician Classifier .......... 42

3.3.1 Motivation ................................................ 43

3.3.2 Method ................................................ 46

3.3.3 Validation ............................................... 51

3.3.3.1 Brainweb Phantom Validation ......................... 51

3.3.3.2 IBSR Validation .................................. 54

3.3.4 Comparison with a Gaussian model ....................... 56

3.3.4.1 Segmentation consistency .......................... 56

3.3.4.2 Cortical surface consistency ...................... 58

3.3.5 Comparison with other methods ......................... 61

3.3.6 Discussion and Future Work .......................... 64

3.4 Summary ................................................... 67
CONTENTS

7 Appendix 160

Bibliography 163

Vita 184
List of Tables

2.1 Average $T_1$, $T_2$ and relative $P_D$ values of CSF, GM and WM at 1.5T and 3T ................................................................. 9

3.1 Percent misclassification rate for SPGR/MPRAGE with Ground Truth for two simulated phantoms .............................................. 39

3.2 Jaccard Coefficients, averaged over all three classes, between SPGR and MPRAGE hard segmentations for five normal subjects .......... 40

3.3 Dice coefficients between the ground truth and hard segmentations of tissue classes, averaged over three inhomogeneity levels, are shown for different noise levels of Brainweb phantoms ..................................... 53

3.4 Dice coefficients of GM, WM, and a volume weighted average (WA) between manual segmentations and hard segmentations obtained by FAST, FANTASM, Freesurfer FCM based segmentation (Freesurfer), SPM, and RiCE ......................................................... 55

3.5 Dice comparison of Rician and Gaussian mixture models on BLSA data .... 56

3.6 CRUISE surface Differences between Gaussian and Rician models. . . 61

3.7 Dice coefficients between segmentations of a $T_1$-w SPGR and $T_1$-w MPRAGE acquisitions for RiCE, FAST, FANTASM, Freesurfer, and SPM segmentations. ......................................................... 63

3.8 Surface differences of the consistency experiment. .......................... 65

4.1 Dice coefficients (mean ± standard deviation) between the segmentations of the original and synthetic images ........................................... 103

4.2 Mean KL distances between histograms of the original/synthetic images of BIRN data ................................................................. 103

4.3 Mean KL distances between histograms of the original/synthetic images for MBIRN data ............................................................... 108

4.4 Dice coefficients (mean ±standard deviation) between segmentations of original scans, registered and transformed versions, and MIMECS synthesis for BIRN data. ......................................................... 109
5.1 Dice coefficients of hard segmentations obtained of images before and after normalization, comparing our method with histogram matching and a landmark based matching.

5.2 Dice coefficients of hard segmentations between the original MPRAGE acquisition and the synthetic ones as well as the original SPGR acquisitions are shown for 20 BLSA subjects [1]. Our method is compared with histogram matching and a landmark based synthesis [2].

5.3 Average Dice coefficients of hard segmentations between the MPRAGE acquisition and the synthetic ones as well as the original SPGR acquisitions for 5 BIRN [3] subjects.

6.1 The table shows the (mean ± std) $R^2$ of linear fits of the tissue volumes obtained from hard segmentations of the original image and also the hard segmentations based on three normalization approaches: histogram matched; landmark based [4]; and our approach (Longitudinal). The $R^2$ values are based on an average of seven subjects.
List of Figures

1.1 Four different acquisitions of the same subject under different scanners and pulse sequences ................................................. 2
2.1 $T_1$-w, $T_2$-w and $P_D$-w images. ............................................. 10
2.2 Acquisition of MR images as a complex data .............................. 11
2.3 Rician distributions for various $v$ and $\sigma$. ............................... 11
2.4 An MR image and its segmentation into three major tissues, CSF, GM and WM. ................................................................. 12
2.5 Skull-stripping and inhomogeneity correction of an MR image ....... 14
2.6 Gaussian mixture model segmentation of an MR image .............. 17
3.1 SPGR and MPRAGE acquisitions of a subject, their hard segmentations by FANTASM, and the corresponding histograms ............ 27
3.2 FCM memberships $u_{jk}$'s for fixed values of $v_k$'s for $q = 2$ and $q = 3$. ................................. 30
3.3 (a) The simplex of $u$ and (b) its trajectory for different values of $y_j$. 31
3.4 Trajectories of $u$ and $\bar{u}$ for different values of $p$. .................. 33
3.5 FCMVC memberships for different values of $p$. ........................ 34
3.6 Comparison between hard segmentation of real MPRAGE and SPGR images. ................................................................. 41
3.7 (a) The histogram (solid black) of an inhomogeneity corrected MPRAGE image (shown inset), overlapped with a Gaussian (dotted blue) and Rician (dashed dot) fitting. (b) CSF, GM, WM distributions as obtained from the Rician (dot) and Gaussian (solid colored) fit. .............. 45
3.8 Comparison of RiCE (2nd row) with ground truth (1st row) on true hard segmentation and fuzzy membership functions of a Brainweb phantom with 3% noise .................................................. 52
3.9 Comparison of hard segmentations of a normal IBSR subject using 5 different methods ...................................................... 54
3.10 Rician and Gaussian fitting of histograms for an SPGR and an MPRAGE scan of the same subject ........................................ 57
LIST OF FIGURES

3.11 CRUISE outer surface comparison between the Rician model and a comparable Gaussian one. ........................................... 59
3.12 Comparison of the consistency in CRUISE inner cortical surface between the Rician model and a Gaussian one ......................... 60
3.13 Hard segmentations of a pair of SPGR and MPRAGE acquisition of a BLSA subject, obtained using 5 different algorithms. ............ 62
3.14 Surface difference between the inner and outer CRUISE cortical surfaces generated from a pair of SPGR and MPRAGE scans of the same subject .......................................................... 64

4.1 The top row shows four different acquisitions of the same subject under different scanners and pulse sequences, while the bottom row shows the corresponding intensity histograms. The scale on the horizontal axis is arbitrary. ................................................................. 71
4.2 Phase transition diagram of a random Gaussian matrix .................. 75
4.3 (a) Atlas $T_1$-w SPGR and (b) its corresponding $T_1$-w MPRAGE. (c) A subject $T_1$-w SPGR scan and (d) its $T_1$-w MPRAGE image. The atlas SPGR is deformably registered (using SyN [5]) to the subject SPGR. This deformation is applied to the atlas MPRAGE to obtain (e) a synthetic subject MPRAGE. .......................................................... 80
4.4 Effect of sparsity of $x(j)$ on the $C_2$ contrast: The left most column shows a subject’s SPGR (top) and MPRAGE (bottom) acquisitions. The second, third, and fourth columns of the top row show synthetic MPRAGEs generated using another portion of the subject’s MPRAGE as the atlas. The synthetic MPRAGEs were generated using $\lambda$ values of 0.05, 0.80, and 0.95, respectively. The plot shows the MPRAGE synthesis error vs. the average sparsity of all $x(j)$’s, averaged over all non-zero voxels. The average sparsity scale is on the top of the plot while the sparsity regularization parameter, $\lambda$, is plotted on the bottom axis. ................................................................. 92
4.5 Phase transition diagram of a $A_1$, constructed using patches from a real SPGR image .......................................................... 93
4.6 The left column shows the Brainweb phantom with 0% noise, used as both $C_1$ and $C_2$ contrast atlas. The next four columns show the subject images, reconstructed images using the atlas, baseline noise and residual noise after reconstruction, respectively, for two noise levels in the two rows. ................................................................. 96
4.7 Plots of the baseline and residual noise vs. the noise levels available in the Brainweb phantom. The left plot shows when $C_1$ and $C_2$ atlases are both the same $T_1$-w image. The plot on the right shows when $C_1$ and $C_2$ are $T_1$-w and $T_2$-w images respectively. ................................. 97
4.8 The top row shows the $T_1$-w and $T_2$-w phantoms from Brainweb, used as the $C_1$ and $C_2$ contrast atlases. The second and third rows show the corresponding subject images, with their synthetic MIMECS result, along with baseline and residual noise images for Brainweb noise levels of 3% and 5%, respectively.  

4.9 The top row shows an example of a repeat scan from the Kirby21 dataset and the corresponding difference image. The middle row shows the synthetic version of scan #1 using scan #2 as both $C_1$ and $C_2$ atlas contrasts, as well as the difference image between the synthetic result and both of scans #1 and #2, respectively. The bottom row shows another synthetic result, this time synthesizing scan #2 with scan #1 as the $C_1$ and $C_2$ atlas contrasts. The scale is for the difference images.  

4.10 The top row shows two example images of repeat scans from the Kirby21 dataset, as well as their corresponding synthetic images. The bottom row shows the hard segmentations of the original images and their synthetic counterpart.  

4.11 Intensity normalization result on BIRN data  

4.12 Intensity normalization on MBIRN data: The top row shows $C_1$ and $C_2$ atlas images with 20° and 30° flip angles, respectively. The second row shows four other subjects with 20°irc flip angles and the corresponding synthetic result with 30°irc flip angles, generated by MIMECS, is shown in the third row.  

4.13 Segmentation result on BIRN data: The top two rows show the original scans of one subject from four different scanners and their corresponding segmentations. The bottom two rows show the MIMECS synthetic versions and their segmentations.  

4.14 The top row shows an MPRAGE scan of a normal subject, used as both $C_1$ and $C_2$ contrast atlases in MIMECS and also used for the registration based synthesis. The second, third, and fourth rows show two OASIS subjects in each column, the atlas image registered and transferred to be in the subject space and the MIMECS result.  

5.1 Mean squared errors (MSE) between atlas ($\alpha_A = 90°$) and subject images ($\alpha_S = 30°, 45°, 60°, 75°$) are shown for various noise levels.  

5.2 Top row shows the atlas ($\alpha_A = 90°$) and the subject ($\alpha_S = 30°$). Middle row shows the histogram matched, landmark based normalized and patch based normalized images. Bottom row shows their differences with the atlas.  

5.3 Top row shows the subject ($\alpha = 20°$), atlas ($\alpha = 30°$) and the corresponding normalized image. Bottom row shows a plot of the histograms of the above three images.
LIST OF FIGURES

5.4  (a) Atlas $C_1$ contrast ($\alpha = 90^\circ, n = 0\%$) and (b) $C_2$ contrast volume. (c) Subject $C_1$ contrast ($\alpha = 45^\circ, n = 3\%$), (d) original subject $C_2$ contrast ($n = 3\%$) image and (e) synthetic subject $C_2$ contrast image. (f) Plot of baseline and residual noise between atlas and synthetic $C_2$ contrast images for $\alpha = 45^\circ$ and $\alpha = 60^\circ$. .............................. 138

5.5 Top row shows the original SPGR, MPRAGE, the patch based synthetic MPRAGE and the landmark based synthetic MPRAGE contrast of a BLSA subject [1], bottom row shows their [6] segmentation. .............................. 140

5.6 Atlas and subject SPGRs along with the corresponding segmentations .............................. 143

5.7 Landmark based and patch based synthesis of the atlas and subject SPGRs, along with the corresponding segmentations. .............................. 144

6.1 Five time points of a normal subject, along with the hard segmentations. The plots of relative volumes of three tissues (CSF, GM and WM) are also shown in the bottom row. .............................. 149

6.2 (a)-(c) shows the first, fourth, and sixth year of a normal subject. Original intensities of a voxel, the center voxel of the blue box in (a), are plotted in (d) as a blue line, while an AR(1) fit of the intensities is shown as a red line. .............................. 151

6.3 A hidden Markov model with normalized patches $x_i^{(t)}$ as hidden states, while the observed patches $y_i^{(t)}$'s are the output. The transition from the $(t−1)^{th}$ time-point to the $t^{th}$ time-point is an AR(1) process. .............................. 154

6.4 (a) Three time-points (1$^{st}$, 4$^{th}$ and 9$^{th}$) of a subject and the corresponding normalized images are shown. (b) After segmentation, CSF, GM, and WM relative volumes are plotted with respect to age. .............................. 159
Chapter 1

Introduction

Magnetic resonance imaging (MRI) is one of the most popular noninvasive imaging modalities to image and analyze the structure of human brain. MRI has the capability to distinguish soft tissues and also lacks the risk of radiation exposure, like computed tomography (CT). Postprocessing of MR brain images, particularly image segmentation [7–11], has been used for many scientific purposes such as furthering our understanding of normal aging [1,12], disease progression [13,14] and prognosis [15]. When large multi-site and multi-center studies are involved (e.g., [3, 16, 17]), two major difficulties in carrying out consistent image processing across all the available imaging data often emerge.

1. Due to scanner and pulse sequence differences, the image intensities in the most fundamental anatomical acquisitions, typically $T_1$-weighted ($T_1$-w) spoiled gradient recalled (SPGR) or magnetization prepared rapid gradient echo (MPRAGE)
CHAPTER 1. INTRODUCTION

Figure 1.1: Four different acquisitions of the same subject under different scanners and pulse sequences

pulse sequence, do not share a common scale, resulting in inconsistencies among the processed results [10,18–20].

2. Some additional tissue contrasts that might have been acquired at one site for a local study, such as T2-weighted (T2-w) or Fluid Attenuated Inversion Recovery (FLAIR) contrasts, might not have been acquired at another site, resulting in entirely missing tissue contrasts.

Unlike CT, it is a fundamental problem in MRI that the image intensities do not have any specific numeric meaning, and their values differ with the pulse sequence parameters, the specific implementation of the pulse sequence, the scanner manufac-
CHAPTER 1. INTRODUCTION

turer, and the scanner’s calibration [21]. As an example, the intensity histograms are evidently quite different for two common structural imaging pulse sequences, SPGR and MPRAGE, of the same subject on the same scanner (Fig. 1.1(a)-(b)) and using the same pulse sequence on different scanners (Fig. 1.1(c)-(d)). This lack of consistency in the image intensity scale arising from the variations of pulse sequences and scanners causes inconsistencies in subsequent image segmentation results. Our primary contribution in this thesis is to develop algorithms to account for these inconsistencies as well as to address the missing tissue contrasts.

The research presented here follows two different overall strategies to account for these two issues. First, we sought improvements to standard classification approaches. In particular, we improved [10] upon an existent segmentation method and also develop a Rician mixture model based Bayesian segmentation algorithm [20,22]. Second, we propose patch based methods building on approaches of sparse priors and dictionary learning. Specifically, we develop a Magnetic resonance IMage Example based Contrast Synthesis (MIMECS) method to normalize scans acquired from diverse sources as well as synthesize missing contrasts. Finally we propose an intensity normalization algorithm to normalize scans so that the segmentation consistency is maintained longitudinally. Thus the primary contributions of this thesis are,

1. We have improved upon a Fuzzy C-means based segmentation method called FANTASM [6]. FANTASM is used to segment $T_1$-w brain MR images into different tissue classes, e.g. cerebro-spinal fluid (CSF), gray matter (GM) and
CHAPTER 1. INTRODUCTION

white matter (WM). It produces membership functions of the three tissues that are further used to reconstruct cortical surfaces [23]. Fuzzy C-means (FCM) [7] is an iterative clustering algorithm that confers fuzzy memberships to image intensities based on their distances from some intensity “centroids”. However, FCM does not inherently take into consideration the relative size of the clusters. To account for this, we propose a novel improvement of FCM by introducing “compactness” parameters into the framework and extend it as a brain image segmentation algorithm, called FANTASM with Variable Compactness (FANTASM-VC). While FANTASM was validated and successfully used to segment SPGR scans [24,25], FANTASM-VC is shown to provide the same accuracy while improving consistency between segmentations of SPGR and MPRAGE contrast scans by introducing the cluster size parameter.

2. Although FCM is a popular method to segment brain MR images, it does not possess any probabilistic interpretation. Most generative model based image segmentation approached use mixture of Gaussians to model the image histogram and determine probabilities of tissues belonging to a certain tissue type. However, MR image intensities follow a Rician distribution [26-28] irrespective of the pulse sequence used. Hence, we propose a Rician mixture model based classification algorithm called Rician classifier using EM (RiCE) [20] to segment MR images and show that by using a better model of the image intensities, segmentations become more consistent across different pulse sequences (namely,
CHAPTER 1. INTRODUCTION

3. As the image intensities are highly variable across scanners and pulse sequences, instead of improving segmentation methods, we propose a pre-processing normalization technique (MIMECS [19]) that does not depend on any particular segmentation, making it unbiased for a range of MR segmentation methods. MIMECS is an atlas based patch-matching technique that uses sparsely distributed image patches from a multi-contrast atlas to normalize as well as synthesize images across different sites, scanners and pulse sequences, addressing the two issues of having inconsistent image segmentation, as described earlier, simultaneously. MIMECS is different from existing techniques in two ways. First, it does not require any imaging parameters, which are often difficult to find or estimate. Second, although it is an atlas based method, it does not require any subject to atlas registration. It has been shown to produce consistent segmentations across scans of both normal and diseased subjects acquired from various scanners with different image sequences and imaging parameters.

4. As the sparse prior based MIMECS algorithm is based on discriminative models, we also provide a generative framework for the normalization as well as the synthesis method. We build upon the idea of coherent point drift [29] and propose point cloud matching in a Bayesian framework and show that it produce consistent segmentations across scanner and different pulse sequences.
CHAPTER 1. INTRODUCTION

5. Although MIMECS is used to normalize scans across sites and scanners, it does not take into account any information about the age of the subjects. However, longitudinal studies often suffer from an instability of longitudinal changes in segmentation results. Although all the scans are usually performed with the same scanners and pulse sequences, the resultant segmentations often lack a desired level of longitudinal smoothness due to the noise variation across scans and the inherent biological changes in the brain. We develop a longitudinal normalization algorithm that takes into account these two factors. It is different from existing longitudinal methods in two ways. First, it is not tied to any particular segmentation method, thus it can be used as a pre-processing step to many brain processing pipelines. Second, it does not require any subject to atlas registration, thus being faster than existing methods.
Chapter 2

Background

In this chapter, we provide a brief description of MR image intensity models and a few image segmentation and normalization techniques that are essential for a better understanding of this thesis. First, we describe the MR image acquisition process and a statistical modeling of the image intensities. Then a brief description of two image segmentation models, namely Fuzzy C-means and Gaussian mixture model, are provided, which is going to be used to describe FANTASM-VC and RiCE in Chapter 3.

2.1 MR Image Contrast

Magnetic resonance images are obtained using the nuclear magnetic properties of the hydrogen atoms $^1\text{H}_1$. Hydrogen atoms are abundant in the brain due to the
CHAPTER 2. BACKGROUND

large content of water molecules. As a hydrogen atom contains an odd number of protons, it has a spin, which is exploited to obtain images. One of the fundamental MR properties is the density of the hydrogen atoms in the particular tissue volume, or voxel, and is called the “proton density”, denoted by $P_D$. Each hydrogen nucleus contains a microscopic magnetic field around it. However, in a given volume, there is no net magnetic field as the randomly aligned microscopic fields cancel each other out. When it is influenced by a large static magnetic field, usually denoted $B_0$, the net magnetization vector of all the nuclei align toward the direction of $B_0$ and its magnitude is directly proportional to the magnitude of $B_0$ as well $P_D$ [30].

During imaging, a sample of the tissue is excited by another magnetic field $B_1$, usually perpendicular to the direction of the static field $B_0$. As a result, the net magnetization vector of the tissue deviates from the direction of $B_0$ and “precesses” around it in a pre-determined way based on the Bloch equations. There are two components of the net magnetization. The component along the $B_0$ is called longitudinal magnetization and the one perpendicular to it is called transverse magnetization. It can be shown that for a particular tissue, these two magnetization vectors follow exponential decay with time constants $T_1$ and $T_2$, called the longitudinal and transverse relaxation times. As any time varying magnetic field through a coil induces electrical signal, the time-varying magnetization of the tissue in a voxel produces electrical signals that are represented by the image intensities. The longitudinal and transverse time constants along with the proton density comprises the three fundamental MR
CHAPTER 2. BACKGROUND

Table 2.1: Average $T_1$, $T_2$ and relative $P_D$ values of CSF, GM and WM at 1.5T and 3T

<table>
<thead>
<tr>
<th>Tissue Type</th>
<th>Relative $P_D$</th>
<th>$T_2$(ms)</th>
<th>$T_1$(ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.5T</td>
<td>3T</td>
<td>1.5T</td>
</tr>
<tr>
<td>White Matter</td>
<td>0.61</td>
<td>0.61</td>
<td>67</td>
</tr>
<tr>
<td>Gray Matter</td>
<td>0.69</td>
<td>0.0.69</td>
<td>77</td>
</tr>
<tr>
<td>Cerebrospinal Fluid</td>
<td>1.00</td>
<td>1.00</td>
<td>280</td>
</tr>
</tbody>
</table>

properties of any tissue. For human brains, the average $T_1$, $T_2$, and $P_D$ values for three different tissues, CSF, GM, and WM, are shown in Tab. 2.1 for two $B_0$ values [30,31].

One advantage of MR imaging is that these three fundamental properties can easily be translated to image intensities by applying suitable time-varying magnetic fields, called pulse sequences. As each tissue has its own representative $T_1$, $T_2$, and $P_D$ values, the imaging contrast obtained from each of them are quite different, as seen in the corresponding $T_1$-w, $T_2$-w, and $P_D$-w images in Fig. 2.1. Different contrasts are suitable for delineating different tissues. As we are interested in segmenting the primary tissue classes in the brain, $T_1$-w images provide the best contrast. For $T_1$-w contrast, different kind of pulse sequences also provide variation in tissue contrast. For example, SPGR sequences provide a better contrast between CSF and GM while MPRAGE sequences provide better contrasts between GM and WM (Fig. 2.1).
2.2 Statistical modeling of Image Intensities

MR images are usually acquired as complex data. The real and imaginary parts are acquired in separate in-phase and quadrature phase channels (Fig. 2.2). Then they are combined to get magnitude images, that are used for any further processing. We assume that each channel is corrupted with uncorrelated additive Gaussian noise, having zero mean and the same variances $\sigma^2$ [27,28]. Let $A_R$ and $A_I$ be the noise-free real and imaginary components at a particular voxel, and $y$ be the observed noisy value. If the noise in both the channels are i.i.d. $\mathcal{N}(0,\sigma^2)$, it can be shown that the observed magnitude intensity $y$ follows a Rician distribution [32],

$$f(y|\nu,\sigma) = \frac{y}{\sigma^2} e^{-\frac{y^2 + \nu^2}{2\sigma^2}} I_0 \left( \frac{\nu y}{\sigma^2} \right), y \geq 0, \sigma > 0,$$

(2.1)
CHAPTER 2. BACKGROUND

where $v = \sqrt{A_R^2 + A_I^2}$ and $I_p$ is the $p^{th}$ order Bessel function of the first kind. The Rice distribution has two parameters, $v$ and $\sigma$, with means and variances given by

\[
\text{mean} = \sigma \sqrt{\pi/2} L_{1/2} \left( -\frac{v^2}{2\sigma^2} \right),
\]

\[
\text{variance} = 2\sigma^2 + v^2 - \frac{\pi\sigma^2}{2} L_{1/2}^2 \left( -\frac{v^2}{2\sigma^2} \right),
\]

(2.2)

where $L_{1/2}$ is the Laguerre polynomial.

The plots of $f(y|v, \sigma)$ are shown in Fig. 2.3(a) for different values of $v$ and $\sigma$.
CHAPTER 2. BACKGROUND

Clearly for high $\frac{v}{\sigma} \geq 3$, denoting a high value of signal-to-noise ratio (SNR), the Rice distribution looks like a Gaussian distribution (Fig. 2.3(a)). It can be shown that the Rice distribution can be approximated by a Gaussian one with mean $\sqrt{v^2 + \sigma^2}$ and variance $\sigma^2$ for large $\frac{v}{\sigma}$, although a closer look at the two distributions at $\frac{v}{\sigma} = 3$ shows that the Gaussian approximation is slightly biased. Most statistical image segmentation techniques use the Gaussian distribution to model image intensities since the math is cleaner and the equations are simpler.

2.3 Image Classification Models

![Original Image](image1.png) ![3-class Segmentation](image2.png)

Figure 2.4: An MR image and its segmentation into three major tissues, CSF, GM and WM.

Many image classification algorithms use the above mentioned statistical model
CHAPTER 2. BACKGROUND

on the image intensities. In this section, we describe two popular classification models, namely Gaussian mixture model and Fuzzy C-means. We are primarily interested in classifying soft tissues in the brain, namely CSF, GM and WM, as shown in Fig. 2.4. However, before applying any segmentation technique to an image, several pre-processing steps are necessary. Two of the pre-processing steps are skull-stripping and inhomogeneity correction. All images are first stripped of any skull, fat or dura using a hybrid registration-segmentation skull-stripping technique [33, 34]. Then the images are corrected for any intensity inhomogeneity, that mostly arises from the inhomogeneity present in the $B_0$ field. As most of the image segmentation algorithms rely on intensities, any shading artifact present in the images inevitably decreases the performance of the segmentation. We use a non-parametric non-uniformity correction tool N3 [35] that assumes the inhomogeneity to be a smoothly varying multiplicative field and estimates it using splines. An example of the skull-stripping and the inhomogeneity correction is shown in Fig. 2.5.

2.3.1 Gaussian Mixture Model

Once an image is stripped of any non-brain tissues and corrected for inhomogeneities, many image classification algorithms use a generative model for the intensities and estimates the classification based on the resulting probabilities. As mentioned earlier, MR image intensities approximately follow Gaussian distributions. Thus to segment three primary tissues from an image, a Gaussian mixture model (GMM)
CHAPTER 2. BACKGROUND

Figure 2.5: Skull-stripping and inhomogeneity correction of an MR image

approach is often used. The probability of observing an image intensity $y_j$ at voxel $j$ is given by a $K$-class mixture model having means $\mu_k$ and variances $\sigma_k^2$ with prior probabilities $\pi_k$,

$$P(y_j|\sigma_k, \mu_k, \pi_k) = \sum_{k=1}^{K} \frac{\pi_k}{\sigma_k} \phi \left( \frac{y_j - \mu_k}{\sigma_k} \right), \quad \phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}. \quad (2.3)$$

$\pi_k$ are the prior probability of $y_j$ belonging to the $k^{th}$ class and $\sum_{k=1}^{K} \pi_k = 1$, $\mu_k$ and $\sigma_k^2$ are the means and variances of the $k^{th}$ class and $\Theta = \{\mu_k, \sigma_k, \pi_k\}, \ k = 1, \ldots, K$. In our case, $K = 3$. If the probability can be estimated, the posterior probability of having $y_j$ belonging to the $k^{th}$ class can simply be obtained using Bayes’ rule, which will provide a classification of $y_j$. The posterior probabilities as well as the parameter set $\Theta$ are found efficiently using Expectation-Maximization (EM) [36].

To cast the problem of finding $\Theta$ and the posterior probabilities into an EM framework, we first define $z_{jk}$ as the indicator function of $y_j$ belonging to the $k^{th}$ class. Thus $z_{jk}$ is the hidden underlying true classification of the tissues, $\sum_{k=1}^{K} z_{jk} =$
CHAPTER 2. BACKGROUND

∀ \ j, \ z_{jk} \in \{0,1\}. \ With \ this \ definition, \ the \ mixture \ model \ in \ Eqn. \ 2.3 \ can \ be \ reformulated \ using \ \ z_{jk}'s \ as,

\[ P(y_j, z_j) = \prod_{k=1}^{3} \left[ \pi_k \phi \left( \frac{y_j - \mu_k}{\sigma_k} \right) \right]^{z_{jk}}, \] (2.4)

where \ z_j = [z_{j1}, \ldots, z_{jK}] \ is \ a \ 1 \times K \ vector \ of \ indicator \ function \ of \ y_j \ and \ \phi \ is \ defined \ in \ Eqn. \ 2.3. \ Now \ the \ segmentation \ problem \ becomes \ an \ estimation \ problem \ where \ the \ estimates \ of \ the \ underlying \ segmentation \ \ z_j's \ are \ calculated \ based \ on \ the \ observed \ intensities \ \ y_j. \ The \ segmentations \ can \ be \ computed \ if \ \Theta \ is \ known, \ while \ \Theta \ can \ be \ estimated \ if \ \ z_j's \ are \ known. \ Thus \ the \ EM \ algorithm \ can \ be \ used \ to \ find \ the \ maximum \ likelihood \ estimates \ of \ the \ parameters \ \Theta \ by,

\[ \hat{\Theta} = \text{argmax}_{\Theta} \sum_{j \in \Omega} \sum_{z_j} \log P(y_j, z_j|\Theta), \] (2.5)

where \ \Omega \ is \ the \ image \ domain. \ The \ EM \ algorithm \ iteratively \ estimates \ the \ underlying \ true \ segmentation \ \ z_{jk}'s \ based \ on \ the \ current \ estimate \ of \ \Theta, \ and \ then \ updates \ \Theta \ based \ on \ the \ estimate \ of \ \ z_{jk}. \ This \ can \ be \ described \ as \ a \ two \ step \ process:

- **E Step**: To find new update \( \Theta^{(m+1)} \) at the \( m \)th iteration, we compute,

\[ Q(\Theta^{(m+1)}|\Theta^{(m)}) = E \left[ \log P(Z|\Theta^{(m+1)})|y, \Theta^{(m)} \right] \] (2.6)

- **M Step**: Find new estimation \( \Theta^{(m+1)} \) based on the previous estimation of parameters \( \Theta^{(m)} \) using the following equation,

\[ \Theta^{(m+1)} = \text{argmax}_{\Theta^{(m+1)}} Q(\Theta^{(m+1)}|\Theta^{(m)}), \] (2.7)
where \( \mathcal{Z} = \{ z_{jk} : j \in \Omega, k = 1 \ldots K \} \) is the true underlying segmentation of the whole image. The algorithm terminates if the difference between log-likelihoods of successive iterations drops below a certain threshold. It has been shown that the EM algorithm is guaranteed to increase the likelihood, but the final convergence depends heavily upon its initialization. If the algorithm is not initialized near the true maximum, it may find a local optimum, so the EM is often initialized using some prior information about \( \Theta \).

It can be shown that at the \( m^{th} \) iteration, the E-step gives the estimate of the posteriors as,

\[
 w_{jk}^{(m)} = E(z_{jk} | y_j, \Theta^{(m)}) = P(z_{jk} | y_j, \Theta^{(m)}) = \frac{\pi_k^{(m)} \phi \left( \frac{y_j - \mu_k^{(m)}}{\sigma_k^{(m)}} \right)}{\sum_{\ell=1}^{K} \pi_{\ell}^{(m)} \phi \left( \frac{y_j - \mu_{\ell}^{(m)}}{\sigma_{\ell}^{(m)}} \right)} 
\]

The M-step provides the update equations for \( \Theta^{(m+1)} \) as

\[
\pi_k^{(m+1)} = \frac{1}{N} \sum_{j \in \Omega} w_{jk}^{(m)} \\
\mu_k^{(m+1)} = \frac{\sum_{j \in \Omega} w_{jk}^{(m)} y_j}{\sum_{j \in \Omega} w_{jk}^{(m)}} \\
\sigma_k^{(m+1)} = \sqrt{\frac{\sum_{j \in \Omega} w_{jk}^{(m)} (y_j - \mu_k^{(m+1)})^2}{\sum_{j \in \Omega} w_{jk}^{(m)}}} 
\]

\( N = |\Omega| \) is the total number of voxels. The algorithm is said to converge if \( \max_k |w_{jk}^{(m+1)} - w_{jk}^{(m)}| < \delta \) for some small \( \delta \).
CHAPTER 2. BACKGROUND

It has been shown that EM algorithm always converges to a local maxima, although it is sensitive to the initialization. With appropriate initialization, the posteriors \( w_{jk} \) for three classes, CSF, GM, and WM are shown in Fig. 2.6 for a \( T_1 \)-w MPRAGE image. Using the converged values of \( \Theta \), the fitted probability density function (pdf) of the image intensities are also shown in Fig. 2.6 in green, closely resembling the histogram of the image (blue).

![Original Image](image1)

![Gaussian Fit](image2)

CSF  GM  WM

Figure 2.6: Gaussian mixture model segmentation of an MR image

There have been several modifications to the model proposed in Eqn. 2.4. To impose spatial coherence in the segmentation, spatially varying priors \( \pi_{jk} \)'s have been
proposed instead of fixed $\pi_k$'s. Biologically, the underlying segmentation $Z$ should be locally smooth. The local smoothness is often captured by introducing a Markov random field (MRF) on the segmentations $z_j$'s \cite{37,38}, which is also a spatial smoothness criteria on the prior probabilities $\pi_{jk}$'s. No spatial relationship was imposed on them in Eqn. 2.4 and they are assumed to be unknown parameters. Under the MRF assumption, the prior probability at every voxel depends on the segmentation of the neighborhood voxels. Defining $z_{N_j}$ as the underlying segmentation of a neighborhood $N_j$ of the $j^{th}$ voxel, $\pi_k$ in Eq. 2.4 is changed to a spatially varying prior $f_{\text{MRF}}(z_{jk}|z_{N_j},\Theta)$, which depends on the segmentation $z_{N_j}$ of the neighborhood $N_j$. The exact structure of $f_{\text{MRF}}$ depends on the smoothness assumptions of $Z$.

The Hammersley-Clifford theorem \cite{39} states that for the function $f_{\text{MRF}}$ to be an MRF, it must be of the form,

$$f_{\text{MRF}}(Z|\Theta) = \frac{1}{M} \exp\{-U(Z|\Theta)\}, \quad (2.10)$$

where $U(\cdot|\Theta)$, called the Gibbs potential, is usually a sum of functions of the neighborhoods of each voxel and $M$ is a normalizing constant. The Potts model \cite{40} is a common example of the Gibbs Potential,

$$U(z_{jk}|\Theta) = \beta \sum_{i \in N_j} \delta(z_{jk}, z_{ik}), \quad (2.11)$$

where $\delta(\cdot, \cdot)$ is a Kronecker Delta, and $\beta$ is a positive weight, indicating maximum prior when $z_j = z_i \ \forall i \in N_j$. Variations of this model have been successfully used in previous brain tissue segmentation methods \cite{37,41,42}. One such variation enforces
segmentation smoothness by introducing the product of the segmentations at every voxel with its neighbors [6],

\[ U(z_j|\Theta) = \beta \sum_{k=1}^{K} \sum_{i \in \mathcal{N}_j} \sum_{\ell=1,\ell \neq k}^{K} z_{jk}z_{i\ell}, \tag{2.12} \]

where small values of \( U \) are obtained when the neighborhood of the \( j^{th} \) voxel has the same segmentation as the \( j^{th} \) voxel.

Usually the MRF models contains a few “interaction coefficients” (such as \( \beta \)) that are difficult to estimate. They are often estimated by heuristics or by a global search using simulated annealing [43] or Monte-Carlo methods. However, approximate solutions can be found using mean field theory [44], where instead of jointly maximizing the probabilities of \( z_j \) along with the probabilities of \( z_{N_j} \), the “mean values” of the neighborhood field \( z_{N_j} \), i.e., the posteriors \( w_{N_j} \), are used to estimate the posterior of \( z_j \). This idea has also been used in many segmentation algorithms [45, 46].

### 2.3.2 Fuzzy C-means

In this section, we describe the Fuzzy C-means (FCM) model for image segmentation. FCM [7] is a clustering algorithm that has been extensively used in medical image segmentation [8, 24]. FCM converges readily, is scale and shift invariant, and allows for the straightforward incorporation of multichannel data. Furthermore, FCM yields soft segmentations (in the form of “membership functions”, similar to posterior probabilities in GMM) that have been used as intermediate data structures for
CHAPTER 2. BACKGROUND

further analysis [23].

Mathematically, FCM can be expressed as the solution to the minimization of an energy function, given by,

$$E_{FCM} = \sum_{j \in \Omega} \sum_{k=1}^{K} u_{jk}^q (y_j - v_k)^2, \text{ subject to } \sum_{k=1}^{K} u_{jk} = 1, u_{jk} \in [0, 1].$$

(2.13)

$y_j$ is the image intensity value at the $j^{th}$ voxel, $u_{jk}$ is the desired membership function for the $j^{th}$ voxel for the $k^{th}$ class, $v_k$ is the cluster centroid of the $k^{th}$ cluster, $\Omega$ is the image domain, $K$ is the total number of classes. For our purpose, we always use $K = 3$ to delineate CSF, GM and WM. The parameter $q$ is a weighting exponent and is constrained by $q > 1$. If Eqn. 2.13 is minimized w.r.t $u_{jk}$ and $v_k$'s, the following iterative update equations are obtained,

$$u_{jk}^{(m+1)} = \frac{|y_j - v_k^{(m)}|^{-\frac{2}{q-1}}}{\sum_{\ell=1}^{K} |y_j - v_\ell^{(m)}|^{-\frac{2}{q-1}}} \sum_{j \in \Omega} u_{jk}^{(m+1)q} y_j$$

$$v_k^{(m+1)} = \frac{\sum_{j \in \Omega} u_{jk}^{(m+1)q} y_j}{\sum_{j \in \Omega} u_{jk}^{(m+1)q}}.$$  

(2.14)

If $q = 1$, FCM reduces to the hard K-means algorithms with the membership functions taking binary values. As the value of $q$ increases, the fuzziness of the membership functions also increases. A typical value of $q$ is 2, although different values have been suggested based on experimental observations [47]. Several modifications of FCM have been proposed to compensate for the partial volume effect, which occurs due to the mixing of two tissues at a voxel because of the limited resolution of the MR
imaging system. One of the modification is to model the cluster centroids as a mixture of multiple tissues instead of a single tissue \[41,48,49\]. Modifications have been proposed for a simultaneous correction of intensity inhomogeneities by introducing a smoothly varying multiplicative “gain-field” into the energy function \[24,50\].

We briefly describe a particular modification of FCM, called Fuzzy and Noise Tolerant Adaptive Segmentation Method (FANTASM) \[6\] that is proposed to include both segmentation smoothness as well as bias field correction. The energy function of FANTASM is given by,

\[
E_{\text{FANTASM}} = \sum_{j \in \Omega} \sum_{k=1}^{K} u_{jk}^q (y_j - g_j v_k)^2 + \lambda_1 \sum_{j \in \Omega} \sum_{r=1}^{2} (D_r \ast g_j)^2 \\
+ \lambda_2 \sum_{j \in \Omega} \sum_{r=1}^{2} \sum_{s=1}^{2} (D_r \ast D_s \ast g_j)^2 + \beta \sum_{j \in \Omega} \sum_{k=1}^{K} \sum_{l \in N_j} \sum_{m=1, m \neq k}^{K} u_{lm}^2.
\]

As before, \(N_j\) denotes the neighborhood of the \(j^{th}\) voxel. The first term denotes the same clustering error as Eqn. 2.13. The inhomogeneity is represented by a smooth multiplicative field \(g_j\). The smoothness of \(g_j\)’s is guaranteed by penalizing its 1\(^{st}\) and 2\(^{nd}\) derivatives in the 2\(^{nd}\) and 3\(^{rd}\) terms, respectively, where \(D_1\) and \(D_2\) denote the horizontal and vertical finite difference operators. The smoothness of the membership of the \(j^{th}\) voxel for the \(k^{th}\) class (\(u_{jk}\)’s) is emphasized by penalizing the memberships of the other classes of its neighbors. \(\lambda_1, \lambda_2\) and \(\beta\) are weights that are often determined either empirically or by cross-validation. Assuming \(\lambda_1 = \lambda_2 = 0\), the update equations for \(u_{jk}\) and \(v_k\) are similar to Eqn. 2.14 and are provided in \[6\].
2.4 Summary

In this chapter, we have given an overview of methods that are commonly used in MR image segmentation. Two models, one based on Gaussian mixtures and the other based on FCM are discussed as well as some of their modifications. We will build upon these models in the next Chapter.
Chapter 3

Methods for Consistent Image Segmentation

3.1 Introduction

Various automated segmentation techniques have been proposed to segment brain tissues—typically CSF, GM, and WM—from MR images. Accurate and reliable tissue segmentation is extremely important to the neuroscience community because it is a key step in nearly every image-based study of the brain in health and disease [1,13,14]. Manual segmentation by experts is still considered to be the gold standard in brain quantification, though automated or semi-automated segmentation is acceptable for large-scale studies in which the image acquisition parameters are identical and manual segmentation is impractical.
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

Fully automated brain tissue segmentation algorithms can be sensitive to noise, partial volume effects, acquisition protocols, scanner differences, and imaging artifacts such as intensity inhomogeneities, zippers (RF interference), and ringing. Techniques have been proposed to address all of these limitations and have been very successful in large part. Most algorithms incorporate spatial smoothness to reduce isolated misclassification due to noise and local artifacts (cf. [37]). Intensity inhomogeneities are either estimated in preprocessing (e.g., [35,51]) or incorporated within the classification algorithm itself (e.g., [6,24,52]). Incorporation of statistical atlases (cf. [53,54]) and control of topology [9] have been used to reduce misclassification error through prior knowledge. The partial volume effect is typically addressed by producing a soft classification, i.e., one that provides membership functions or posterior probabilities associated with each tissue class. The effect can also be addressed by super-resolution methods [55], probabilistic models, or topological methods [9,56,57].

Compensation for different acquisition protocols or scanner differences has been particularly problematic for tissue segmentation algorithms [18]. Approaches to normalize histograms to a common scale have been proposed [2,58,59], and most recent algorithms use some kind of explicit or implicit intensity normalization preprocessing in practice. Achieving true pulse sequence independence currently requires the use of special pulse sequences [60,61] that permit computation of the underlying tissue parameters to which a segmentation algorithm can be applied [61]. Though admirable in spirit, common practice precludes routine use of special pulse sequences,
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

and modern study designs have typically relied on the use of multiple scanners or
types of scanners or multiple structural acquisition protocols with fixed parameters
in order to yield images whose segmentations can be quantitatively compared within
a particular study.

Two classes of tissue classification methods have emerged as leading algorithms for
MR brain image segmentation: methods [8,24,62] based on fuzzy c-means (FCM) [7]
and methods based on a Bayesian framework using a finite Gaussian mixture model
assumption [37,53,63]. Both approaches have been augmented to account for spatial
smoothness, most commonly using an MRF. At this time, the performances of these
methods are very similar “across the board” and the algorithms are widely used in
large-scale studies. Yet experience shows that algorithm parameters must be tuned in
order to achieve satisfactory results when acquisition parameters change. We suggest
in this chapter that both classes of algorithms operate with a less accurate model of
image intensity and that improving the model can provide consistency and robustness
in segmentation to pulse sequence changes.

The chapter is organized as follows. First we describe our improvement on the
FCM algorithm in Chapter 3.2. Then we describe the Rician model based segment-
tation algorithm in Chapter 3.3 and show that it is a better model for the image
intensities compared to a Gaussian mixture model. We conclude with results showing
that both models improve consistency in segmentations while maintaining the
same level of accuracy as the existing ones.
3.2 Fuzzy C-means with Variable Compactness

In this section, we describe an FCM based fully automatic method for segmentation of $T_1$-w brain images. This method, called Fuzzy and Noise Tolerant Adaptive Segmentation Method with Variable Compactness (FANTASM-VC) is built upon a previous segmentation method called FANTASM [6]. FANTASM, a modification of the FCM algorithm itself [7], includes models for spatial smoothness as well as the inhomogeneity correction. However, the underlying FCM framework, although a discriminative clustering method, lacks a model for the variability of the clusters; thus, is not able to robustly distinguish between clusters with different variances. This effect is seen on FANTASM segmentations of SPGR and MPRAGE acquisitions of MR images, where clusters with different variation are erroneously segmented, resulting in inconsistent segmentation between two acquisitions of the same subject. An example is shown Fig. 3.1 top row, where the two imaging sequences produce quite different tissue contrasts, as seen in the corresponding histograms in the bottom row. While the SPGR sequence provides a good contrast between CSF and GM, MPRAGE provides good contrast between GM and WM. The inconsistency is visible in their FANTASM hard segmentations. The GM-WM boundary is poorly segmented in the left temporal lobe of the SPGR scan, while the CSF is grossly overestimated in the frontal lobe for MPRAGE, as shown in the red circles.
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

Figure 3.1: SPGR and MPRAGE acquisitions of a subject, their hard segmentations by FANTASXM, and the corresponding histograms.

To correct this inconsistency in the segmentations, we propose a novel variation of the FCM framework by including a “compactness parameter” to account for the variability in clusters.

3.2.1 Previous Works

Many variants of FCM have been proposed to account for different properties of clusters. The variations in clusters have been modeled using a variance term in the
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

\( \ell_2 \) norm [64], denoted by FCMV, given by

\[
\min E_{\text{FCMV}} = \sum_{j \in \Omega} \sum_{k=1}^{K} u_{jk}^q \left( \frac{y_j - v_k}{\sigma_k} \right)^2, \quad \text{subject to } \sum_{k=1}^{K} u_{jk} = 1, u_{jk} \in [0, 1],
\]

(3.1)

where \( \sigma_k \)'s are the variances of the clusters, usually determined empirically or by cross-validation, as shown in [65]. Kernel based FCM [66] has been introduced to account for the high-dimensionality of the data, usually given by the following form

\[
\min E_{\text{FCMK}} = \sum_{j \in \Omega} \sum_{k=1}^{K} u_{jk}^q (1 - K(y_j, v_k)), \quad \text{subject to } \sum_{k=1}^{K} u_{jk} = 1, u_{jk} \in [0, 1],
\]

(3.2)

where \( y_j \) is a high dimensional data vector and \( K \) is a kernel function, often Gaussian. Weighted FCM energy functions are proposed [67] to account for cluster variations

\[
\min E_{\text{FCMW}} = \sum_{j \in \Omega} \sum_{k=1}^{K} u_{jk}^q w_{jk}(y_j - v_k)^2, \quad \text{subject to } \sum_{k=1}^{K} u_{jk} = f_k, u_{jk} \geq 0.
\]

(3.3)

The \( f_k \)'s are non-negative constants that are also determined via iteration. Many variants of Eqn. 3.3 have also been proposed, often determining the weights \( w_{jk} \)'s based on outlier detection and cross-validation. Algorithms have been proposed to determine optimal values of \( q \) based on the convergence rate [68] as well as a cross-validation on the data [47].

3.2.2 Compactness Parameters

In this section, we investigate the effect of \( q \) on the resulting membership functions and propose a new modification of the FCM energy function. As discussed in
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

Section 2.3.2, the FCM energy function is given by

\[
E_{\text{FCM}} = \sum_{j \in \Omega} \sum_{k=1}^{K} u_{jk}^q (y_j - v_k)^2, \quad \text{subject to } \sum_{k=1}^{K} u_{jk} = 1, u_{jk} \in [0, 1],
\]  

(3.4)

and the corresponding update equations for \( v_k \) and \( u_{jk} \)'s are given by

\[
u_{jk} = \frac{|y_j - v_k|^{-\frac{2}{q-1}}}{\sum_{\ell=1}^{K} |y_j - v_\ell|^{-\frac{2}{q-1}}},
\]  

(3.5)

\[v_k = \frac{\sum_{j \in \Omega} u_{jk}^q y_j}{\sum_{j \in \Omega} u_{jk}^q}.
\]  

(3.6)

It has been shown that \( q = 2 \) or \( q = 3 \) is often a reasonable choice [47] that provides both robust segmentations as well as a fast convergence rate, primarily attributed to the fact that they lead to either the \( \ell_2 \) or \( \ell_1 \) norm on the error metric \( |y_j - v_k| \).

As we are always interested in a 3-class segmentations in neuroimaging, for the set of randomly chosen triplet \( \mathbf{v} \equiv [v_1, v_2, v_3] = [10, 50, 80] \), we plot the memberships \( u_{j1}, u_{j2}, \) and \( u_{j3} \) as functions of the intensities \( y_j \in [0, 100] \) in Fig. 3.2. As suggested from the update equations (Eqn. 3.5), the membership \( u_{jk} = 1 \) when \( y_j = v_k \), also evident from the figure. We make two important observations from Fig. 3.2;

1. \( u_{j1} = u_{j2} \) when \( y_j = \frac{v_1 + v_2}{2} \) and \( u_{j2} = u_{j3} \) when \( y_j = \frac{v_2 + v_3}{2} \), irrespective of \( q \).

   This can also be derived from Eqn. 3.5.

2. \( u_{j3} > 0 \) at \( y_j = 0 (\ll v_3) \), and \( u_{j1} > 0 \) at \( y_j = 1 (\gg v_1) \), also evident from Eqn. 3.5.
Clearly, both of these properties are undesirable. The first property implies the absence of any variance factor of the data by having equal membership at exactly half-way between the centroids, irrespective of $q$. The second property implies that if the data $y_j$ is far away from all the centroids, i.e., if $y_j$ is an outlier, it may have high membership for an unwanted class. E.g., assuming finite $v_k$’s, if $y_j \to \pm \infty$, $u_{j1} = u_{j2} = u_{j3} = 1/3$. This can also be derived from Eqn. 3.5. This phenomenon is often undesirable in medical imaging context, where a very bright WM outlier may get a high CSF and GM membership. A visualization of this effect is shown in Fig. 3.3. Define $\mathbf{u} = [u_{j1}, u_{j2}, u_{j3}]$ be a point in a 3D space, and $\mathbf{u} \equiv \mathbf{u}(y_j; v_k)$ is a function of $y_j$ for a particular set of $v_k$’s. Thus $\mathbf{u}$ lies on a simplex, because $u_{j1} + u_{j2} + u_{j3} = 1$ and $u_{j1}, u_{j2}, u_{j3} \in [0, 1]$, as shown in Fig. 3.3(a). If $\mathbf{u}(y_j; v_k)$ is plotted for $y_j \in (-\infty, +\infty)$ for $q = 2$ and $q = 3$, we get the curves in Fig. 3.3(b), showing a 2D projection of the
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

Figure 3.3: (a) The simplex of $u$ and (b) its trajectory for different values of $y_j$.

The simplex of $u$ and (b) its trajectory for different values of $y_j$. The three vertices in Fig. 3.3 starting from bottom left in clockwise direction are $v_1$, $v_2$, and $v_3$, respectively. As $y_j$ increases from $-\infty$ to $+\infty$, the clockwise direction of the trajectory is shown by green arrows in Fig. 3.3(b). As the “fuzziness coefficient” $q$ increases from 2 to 3, the membership trajectory moves closer to the centroid of the triangle, $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, indicating more fuzziness in the memberships. It can be shown that the trajectory of $u$ for $q = 1$ consists of only three vertices, indicating accumulation of all the intermediate points to their nearest vertices giving a K-means hard segmentation. Assuming $v_1 < v_2 < v_3$, $u(\infty; v_k) = u(-\infty; v_k) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, indicating that the asymptotic points on the trajectory of $u$ are the centroid of the triangle.

For fixed $v_k$’s, if $y_j \in [v_1, v_2]$, it is expected that $y_j$ is a mixture of CSF and GM,
with \( u_{j3} \) being small. Similarly \( u_{j1} \) is expected to be small for \( y_j \in [v_2, v_3] \), as it is a mixture of GM and WM. Considering the effect of \( q \) on the memberships, we observe that, if \( q \) increases, WM membership \( u_{j3} \) for \( y_j \in [v_1, v_2] \) increases, and so does the CSF membership \( u_{j1} \) for \( y_j \in [v_2, v_3] \), also shown in Fig. 3.3. This demonstrates that the parameter \( q \) captures information about the variability of all the clusters.

Based on this interpretation, we want to make sure that if the WM cluster has large variance, then \( u_{j3} \) for \( y_j \in [v_2, v_3] \) should be fuzzier independent of \( u_{j1} \) and \( u_{j2} \). And similarly, if CSF cluster has large variance, then \( u_{j1} \) for \( y_j \in [v_1, v_2] \) should be fuzzier independently of \( u_{j2} \) and \( u_{j3} \). This gives us the motivation for treating \( q \) as a parameter to capture variation in clusters.

Previous methods have used separate \( q \)'s for each cluster. However, it does not lead to any closed form solutions and heuristics are to be used to solve the system. However, in a similar manner as [47], we introduce compactness parameters \( p_k, k = 1, \ldots, K \), in the FCM framework as a minimization of the following energy function that leads to closed form solutions,

\[
E_{\text{FCMVC}} = \sum_{j \in \Omega} \sum_{k=1}^{K} \bar{u}_{jk}^2 |y_j - v_k|^{2p_k}, \text{ subject to } \sum_{k=1}^{K} u_{jk} = 1, u_{jk} \in [0, 1]. \tag{3.7}
\]
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

\[ p = [1.3, 1.2, 1.0] \quad p = [1.0, 1.2, 1.3] \]

Figure 3.4: Trajectories of \( u \) and \( \bar{u} \) for different values of \( p \).

Minimization of \( E_{FCMVC} \) gives the membership functions and centroids as,

\[
\bar{u}_{jk} = \frac{|y_j - v_k|^{-2p_k}}{K \sum_{l=1}^{K} |y_j - v_l|^{-2p_l}}
\]

\[
u_k = \frac{\sum_{j \in \Omega} \bar{u}_{jk}^2 |y_j - v_k|^{2p_k - 2} y_j}{\sum_{j \in \Omega} \bar{u}_{jk}^2 |y_j - v_k|^{2p_k - 2}}.
\]

We note that \( p = [p_1, \ldots, p_K] = [1, \ldots, 1] \) is exactly the same as FCM with \( q = 2 \). The plot of memberships \( \bar{u}_{jk} \) and \( u_{jk} \) for the same set of \( v_k \)’s are shown in Fig. 3.4 for two sets of \( p \). We observe that for \( \forall y_j \in [v_1, v_2] \), \( \bar{u}_{j1} \) with \( p_1 > p_2 > p_3 \) (Fig. 3.4(a)) has decreased from \( u_{j1} \) with \( p = [1.0, 1.0, 1.0] \). Similarly \( \forall y_j \in [v_2, v_3] \), \( \bar{u}_{j3} \) with \( p_1 < p_2 < p_3 \), has decreased from \( u_{j3} \) (Fig. 3.4(b)). It shows that with increase in \( p_k \), memberships decrease for a fixed \( |y_j - v_k| \). Thus we define \( p_k \) as the “compactness
parameter” (as opposed to variance) of the \( k^{th} \) cluster. The memberships \( \bar{u}_{jk} \)'s are shown for two different sets of \( p \) in Fig. 3.5 with the same \( v = [10, 50, 80] \), as before. Now using different compactness parameters for different classes with \( p_1 > p_2 > p_3 \) (Fig. 3.5(a)), the point of equal memberships for \( u_{j1} = u_{j2} \), has moved from \( \frac{v_1 + v_2}{2} \) and slightly towards \( v_1 \), indicating lower variance for the 1\(^{st} \) cluster. Also using a much smaller \( p_3 < p_2 = p_1 \), \( u_{j1} \approx 0 \) at \( y_j = 100 \), unlike Fig. 3.2(a), indicating outlier suppression. Thus using suitable \( p \), the variation in cluster size can be accommodated in the FCMVC framework.

Figure 3.5: FCMVC memberships for different values of \( p \).
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

3.2.3 Integration with FANTASM

Based on the proposed method FCMVC (Eqn. 3.7), we now modify the existing FANTASM energy function [6] to FANTASM-VC [10];

\[
E_{\text{FANTASMVC}} = \sum_{j \in \Omega} \sum_{k=1}^{C} u_{jk}^2 (y_j - g_j v_k)^2 p_k + \lambda_1 \sum_{j \in \Omega} \sum_{r=1}^{R} (D_r \ast g_j)^2 \\
+ \lambda_2 \sum_{j \in \Omega} \sum_{r=1}^{R} \sum_{s=1}^{R} (D_r \ast D_s \ast g_j)^2 + \beta \sum_{j \in \Omega} \sum_{k=1}^{C} u_{jk}^2 \sum_{l \in N_j} \sum_{m \neq k} u_{lm}^2 
\]  

(3.10)

\(y_j\) is the image intensity at \(j^{th}\) voxel, \(\Omega\) is the image domain, \(u_{jk}\)'s are the memberships, \(p_k\)'s are the compactness parameters, \(v_k\)'s are the centroids. Here, \(g_j\) is a scalar gain field to account for any image inhomogeneity, \(D_r\) denotes the first order finite difference operator, \(R = 1, 2\) denote the vertical and horizontal directions, \(\lambda_1\) and \(\lambda_2\) are the regularization parameters on the gain field, \(N_j\) is the set of first order neighbors of the \(j^{th}\) voxel and \(\beta\) is a smoothing coefficient. \(\lambda_1, \lambda_2,\) and \(\beta\) are determined by cross-validation as shown in [6].

Minimization of \(E_{\text{FANTASMVC}}\) can be performed using the following algorithm.

1. Choose suitable compactness parameters \(p_k\) (discussed in the following section).

2. Obtain initial estimate of \(v_k\). This is usually done by a 3-class Gaussian mixture model, as described in Chapter 2.3.1.

3. Initialize \(g_j = 1, \forall j \in \Omega\).
4. Compute membership functions,
\[ u_{jk} = \left( \frac{|y_j - g_j v_k|^2 p_k + \beta \sum_{l \in N_j} \sum_{m \neq k} u_{lm}^2}{\sum_{n=1}^{K} \left( |y_j - g_j v_n|^2 p_n + \beta \sum_{l \in N_j} \sum_{m \neq n} u_{lm}^2 \right)} \right)^{-1}. \] (3.11)

5. Update the centroids,
\[ v_k = \frac{\sum_{j \in \Omega} u_{jk}^2 g_j y_j |y_j - g_j v_k|^{2p_k-2}}{\sum_{j \in \Omega} u_{jk}^2 g_j^2 |y_j - g_j v_k|^{2p_k-2}}. \] (3.12)

6. Update gain field coefficients by solving the spatially varying equation,
\[ 2y_j \sum_{k=1}^{K} u_{jk}^2 p_k |y_j - g_j v_k|^{2p_k-2} v_k = 2g_j \sum_{k=1}^{K} u_{jk}^2 p_k |y_j - g_j v_k|^{2p_k-2} v_k^2 + \lambda_1 (H_1 \ast g)_j + \lambda_2 (H_2 \ast g)_j. \] (3.13)

7. If the algorithm converges, stop; otherwise go to step 4.

Proofs of the Eqn. 3.11–Eqn. 3.13 are given in the appendix.

### 3.2.4 Estimation of Compactness Parameters

We have established a fully automatic segmentation algorithm FANTASM-VC, whose energy function (Eqn. 3.10) is iteratively minimized using coordinate-descent type iterative update equations (Eqn. 3.11-Eqn. 3.13). In this section, we show that using different compactness parameters for SPGR and MPRAGE scans, their segmentations can be made more similar.
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

The compactness parameters $p$ are estimated based on a set of SPGR training data [23], for which we have manually selected landmarks on the cortical boundaries [69]. An exhaustive search is performed to estimate $p$, such that the segmentation isocontours generated by FANTASM-VC were close to the landmarks. The compactness parameters estimated on such training data are $p_S = [0.9, 1.00, 1.06]$. Next we estimated the compactness parameters for MPRAGE scans using a training set of SPGR and MPRAGE scans of the same subject. The parameters for MPRAGE are estimated based on an exhaustive search so that the isocontours of the MPRAGE segmentations line up with the isocontours of SPGR segmentations that are done using $p_S$. The estimated compactness for MPRAGE scans are $p_M = [1.25, 1.00, 0.95]$. We use these trained parameters for all the subsequent data.

As the primary objective of variable compactness $p$ is to introduce a variance-like effect in the FCM framework, we compare FANTASM-VC with the particular modification FCMV [64] given by Eqn. 3.1, where the variation of clusters is achieved by directly introducing variance terms $\sigma_k$’s. The corresponding iterative update equations are similar to Eqn. 3.11–Eqn. 3.13. The corresponding variances for SPGR as well as MPRAGE scans are estimated exactly in the same way described above. The estimated variances for the three classes were found to be $\sigma^2_S = [3.1, 1.44, 1]$ for SPGR and $\sigma^2_M = [1, 1.52, 2.55]$ for MPRAGE scans. We used these values of $\sigma_S$ and $\sigma_M$ on all the test images.
3.2.5 Segmentation Consistency

We conducted two experiments to show that FANTASM-VC can improve the consistency of the segmentation of the same subject under two different acquisition protocols. Our first experiment involves two synthetic phantom data sets which we used to simulate the same object imaged with MPRAGE and SPGR imaging parameters. We generated the phantoms from a fuzzy classification truth model using the statistical model outlined in [70]. The phantoms possessed 5% noise and did not have any inhomogeneity. The misclassification rate, defined as the ratio of the # of correctly classified voxels against the total # of non-background voxels, is reported in Table 3.1 for classifying the MPRAGE and SPGR phantoms using GMM (as described in Section 2.3.1), FANTASM, FCM with Fuzzy Covariance matrix (FCMV Eqn. 3.1) and FANTASM-VC. It is observed that FANTASM has nearly the same misclassification rate as FANTASM-VC on SPGR phantoms, but fails to do as well on MPRAGE phantoms. However, FANTASM-VC has the lowest misclassification rate for both SPGR and MPRAGE phantoms.

Our second experiment involved pairs of real SPGR and MPRAGE $T_1$-w images for five subjects. We computed the hard segmentations using each of the methods used in the first experiment and report the Jaccard coefficient between the hard segmentations, averaged over all three classes. We used smoothing and inhomogeneity corrections while using GMM and FCMV as well. The Jaccard coefficients are reported in Table 3.2, and Fig. 3.6 shows the segmentations for one of the data
Table 3.1: Percent misclassification rate for SPGR/MPRAGE with Ground Truth for two simulated phantoms

<table>
<thead>
<tr>
<th></th>
<th>GMM</th>
<th>FANTASM</th>
<th>FCMV</th>
<th>FANTASM-VC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPGR</td>
<td>5.64</td>
<td>3.62</td>
<td>5.66</td>
<td>3.55</td>
</tr>
<tr>
<td>MPRAGE</td>
<td>12.78</td>
<td>5.11</td>
<td>4.45</td>
<td>3.51</td>
</tr>
<tr>
<td>SPGR</td>
<td>7.51</td>
<td>6.32</td>
<td>8.43</td>
<td>4.72</td>
</tr>
<tr>
<td>MPRAGE</td>
<td>9.54</td>
<td>6.11</td>
<td>5.32</td>
<td>4.86*</td>
</tr>
</tbody>
</table>

* Significantly smaller than the other three MPRAGE misclassification rates

sets. We observe that FANTASM tends to over-estimate CSF in MPRAGE. The reason is attributed to the fact that the variance of the CSF class is usually smaller than that of GM class in MPRAGE contrasts. Thus using smaller variance for the CSF class \(p_M(\text{CSF}) = 1.25\) compared to GM \(p_M(\text{GM}) = 1.0\), FANTASM-VC can successfully segment CSF, while FANTASM, having a larger variance \(q = 2\) or \(p_M(\text{CSF}) = 1\), over-estimates CSF. It is also noted that GMM often over-estimates GM in SPGR scans because of the poor GM-WM contrast, indicated by the unimodal histogram of the SPGR image, cf. Fig. 3.2. FCMV and FANTASMVC performs similarly on three of the subjects, while FCMV has a gross failure on one of the MPRAGE subject (#5) indicated by a Jaccard coefficient < 0.5. FANTASM-VC performs best on all the five subjects.
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

Table 3.2: Jaccard Coefficients, averaged over all three classes, between SPGR and MPRAGE hard segmentations for five normal subjects

<table>
<thead>
<tr>
<th>Subject #</th>
<th>GMM</th>
<th>FANTASM</th>
<th>FCMV</th>
<th>FANTASM-VC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.7383</td>
<td>0.5611</td>
<td>0.7781</td>
<td>0.7952</td>
</tr>
<tr>
<td>2</td>
<td>0.7047</td>
<td>0.4764</td>
<td>0.7501</td>
<td>0.7538</td>
</tr>
<tr>
<td>3</td>
<td>0.8134</td>
<td>0.6670</td>
<td>0.8391</td>
<td>0.8559</td>
</tr>
<tr>
<td>4</td>
<td>0.7707</td>
<td>0.5209</td>
<td>0.6363</td>
<td>0.7921</td>
</tr>
<tr>
<td>5</td>
<td>0.7377</td>
<td>0.4205</td>
<td>0.4173</td>
<td>0.7588</td>
</tr>
</tbody>
</table>

3.2.6 Discussion

As the FCM framework intrinsically assumes uniform cluster size, it can sometimes produce biased classification, manifested in the inconsistency in the segmentations obtained from two different pulse sequences of the same subject, giving rise to different tissue contrasts. A new compactness parameter in the FCM framework was introduced to account for the variances in data clusters. We outlined an automatic fuzzy segmentation method based on this parameter. We have satisfied our primary goal of obtaining more similar segmentations from the same data acquired under differing protocols, SPGR and MPRAGE. The proposed approach was tested on single channel data, although it can be readily extended to a multichannel algorithm.
Figure 3.6: Comparison between hard segmentation of real MPRAGE and SPGR images.
3.3 Consistent Segmentation using a Rician Classifier

In the previous section, we described a segmentation method based on the Fuzzy C-means framework. The FCM method is not based on an underlying intensity model, though one can tease apart the variational formulation in order to assert its basic assumptions. In its conventional formulation, FCM is a clustering method that associates voxels to all classes in proportion to the value of its computed membership functions. The basic formulation is not Bayesian, and there is no formula relating the underlying tissue intensities to the observed intensities via an explicit noise model. Accommodations have been made to account for clusters that might not have the same size [10, 64, 65], but the added parameters must generally be known in advance and tuned to any given pulse sequence.

In this section, we develop a segmentation algorithm based on a Bayesian modeling of intensities. The most common Bayesian formulations are based on a finite GMM (Section 2.3.1), in which the conditional probability of the image intensity for a particular tissue type is Gaussian. The parameters of the underlying Gaussian conditional probabilities and the mixture coefficients are typically estimated using the EM algorithm. The model choice together with the estimation procedure automatically accommodates for clusters that might be of different sizes and relative proportions. It is logical to assume that the additional flexibility of this model together with the
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

Bayesian optimality would lead to a better result than FCM. However, there are numerous papers that support the contrary opinion.

This section is organized as follows. First we provide a motivation and rationale for a better modeling of the intensities. Then we provide the segmentation algorithm Rician Classification using EM (RiCE [20]) and show that better modeling of the tissue intensities provides more accurate as well as consistent segmentations across pulse sequences, which we expect because the pulse sequences do not change the noise in acquisition process; thus, the variation in pulse sequences can be taken into account by more accurate modeling of the image intensities.

3.3.1 Motivation

We are led to question the underlying assumption of Gaussian models [37, 71] of the MR intensities in the current Bayesian methods. In conventional MR imaging, the acquired raw data is the underlying signal in “real” (in-phase) and “imaginary” (quadrature phase) channels, each of which is corrupted by additive, zero-mean, i.i.d. Gaussian noise. The complex image intensities are obtained using the Fourier transform, which preserves the Gaussian nature of the noise in the real and imaginary components of the image intensities [72]. Since the observed image intensities are formed by taking the complex modulus of the real and imaginary parts of the complex image, each image voxel becomes a Rician random variable [26–28, 32].

The underlying signal values are generally different at each voxel because of bio-
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

logical variation. Therefore, the probability distribution that describes the collection of all voxels taken together is a Rician mixture model in which there is a different conditional Rician probability density function for each underlying signal value. By noting that within each tissue class the underlying signal intensities are close in value, this rich mixture model can be approximated by one that has only three conditional Rician probability densities, one for each tissue class. When the underlying signal values are large relative to the noise, it is known that a Rician distribution can be approximated by a Gaussian distribution [26]. But since this approximation becomes less accurate with smaller underlying signal values, we can expect the greatest impact of using this Rician mixture model versus a Gaussian mixture model to be in the tissue classes having the smallest underlying signal values.

To illustrate this point, in Fig. 3.7(a) we show the smoothed histogram of intensities in an inhomogeneity corrected [35] MPRAGE image together with two fitted distributions, one using a mixture of three Gaussians (blue) and one using a mixture of three Ricians (red). It is observed that the Rician fit is better, an observation that can be quantitatively verified by noting that the Kullback-Leibler (KL) distance [73] between the image histogram and the Gaussian fit is 0.0418 and between the image histogram and the Rician fit is 0.0097. In Fig. 3.7(b), the fits of the individual class conditional probabilities derived from the Gaussian (blue) and Rician (red) fitting process. It is observed that the CSF densities show the most difference, which is to be expected since these intensities are the lowest. The WM densities are most similar,
which makes sense since these tissues have the highest intensities in this $T_1$-weighted pulse sequence, and are likely to be well approximated by a Gaussian as a result.

Figure 3.7: (a) The histogram (solid black) of an inhomogeneity corrected MPRAGE image (shown inset), overlapped with a Gaussian (dotted blue) and Rician (dashed dot) fitting. (b) CSF, GM, WM distributions as obtained from the Rician (dot) and Gaussian (solid colored) fit.

This observation motivates us to propose a brain image tissue segmentation algorithm based on an underlying finite Rician mixture model. We primarily focus on the difference between Rician and Gaussian models of the tissue intensities. Consequently, we do not include any bias-field correction in our method, instead, we pre-process all the data using a non-parametric inhomogeneity correction method (N3) [35]. Although the inhomogeneities in different MR sequences can depend on the sequence itself, N3 has been shown to work well on different sequences [74, 75]. In order to include smoothness on the resulting segmentation, the algorithm includes an MRF model. This fully automatic algorithm does not require parameter choices, relying instead on the assumption that cluster intensity distributions will be Ri-
cian regardless of the pulse sequence. The main contribution of this work is to improve segmentation consistency between different pulse sequences having $T_1$-w contrast. We compare our method with a Gaussian intensity model approach, SPM (spm_segment function) [63, 76, 77], a Gaussian model approach on log-transformed intensities, FAST [53] and two FCM based approaches, Freesurfer [78] (mri_ms_EM function), and FANTASM [6].

### 3.3.2 Method

As described in Section 2.2, MR image intensities follow a Rician distribution,

$$f_R(y|v, \sigma) = \frac{y}{\sigma^2} e^{-\left(\frac{y^2 + v^2}{2\sigma^2}\right)} I_0\left(\frac{yv}{\sigma^2}\right), \quad y \geq 0, \sigma > 0.$$  \hspace{1cm} (3.14)

where $v = \sqrt{A_R^2 + A_I^2}$ and $I_p$ is the modified $p^{th}$ order Bessel function of the first kind.

As shown in Fig. 2.3(a), a Rician PDF is quite different from a Gaussian for low SNR, where SNR is defined as $v/\sigma$. For higher SNR ($>2$), it can be shown that the Rician distribution asymptotically approaches a Gaussian distribution with mean $\sqrt{v^2 + \sigma^2}$ and variance $\sigma^2$ [26], however Fig. 2.3(b) shows a Rician PDF with SNR = 2, with the corresponding asymptotic Gaussian mean $\sqrt{v^2 + \sigma^2}$ and variance $\sigma^2$. Clearly, the Gaussian PDF is biased for high SNR (=3) and any estimator based on a Gaussian assumption will also be biased. For example, the CSF having a low SNR follows the Rician more closely than a Gaussian (Fig. 3.7(a)). Thus a Gaussian approximation of the histogram will lead to a biased segmentation and a Rician estimation of the
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

histogram will be more appropriate.

As we want to classify a brain MR image into three major tissue classes, CSF, GM, and WM, we propose a 3-class finite mixture model using the Rician distribution. Given a voxel intensity $y_j, j \in \Omega$, $\Omega$ being the image domain, define $z_{jk}$ as the indicator function of the $j^{th}$ voxel belonging to the $k^{th}$ class, $k = 1, \ldots, K$, $K = 3$, for the three classes. Thus, $z_{jk}$ is equivalent to the hidden underlying true segmentation of the tissues. Also define the Rician parameters for the $k^{th}$ class to be $\{v_k, \sigma_k\}$. Let the unknown prior probabilities of observing $y_j$ from the $k^{th}$ class be $\pi_{jk}$. Now a finite mixture model representation of the likelihood of observing $y_j$ is given by

$$f(y_j, z_j|\Theta) = \prod_{k=1}^{K} [\pi_{jk} f_R(y_j|\Theta)]^{z_{jk}},$$

(3.15)

where $z_j = [z_{j1}, z_{j2}, \ldots, z_{jk}]$ is a $1 \times K$ vector of indicator functions and $f_R$ is defined in Eqn. 3.14. Define $Z = \{z_{jk} : j \in \Omega, k = 1 \ldots K\}$ be the true underlying segmentation of the whole image. The parameter collection $\Theta$ can be defined as,

$$\Theta = \bigcup_{j \in \Omega} \bigcup_{k=1}^{K} \{v_k, \sigma_k, \pi_{jk}\}.$$  

(3.16)

In the most general case, the $\pi_{jk}$’s can be treated as unknown parameters, but the number of such parameters will be large ($K \times |\Omega|$). Biologically, the underlying segmentation $Z$ should be locally smooth. The local smoothness is often captured by introducing an MRF on the segmentation [38, 71], which is essentially a smoothness criteria on the prior probabilities $\pi_{jk}$. We parametrize $\pi_{jk}$ using an MRF approach and redefine $\Theta$ so that the number of unknown parameters is smaller. The log-likelihood
of Eqn. 3.15 is extended to include segmentation smoothness as,

$$ f(y_j, z_j|\Theta) = \prod_{k=1}^{K} \left[ f_{\text{MRF}}(z_{jk}|z_{Nj}, \Theta) f_R(y_j|\Theta) \right]^{z_{jk}}$$  \hspace{1cm} (3.17) 

where $z_{Nj}$ as the underlying segmentation of a neighborhood $N_j$ of the $j^{th}$ voxel. Thus the unknown prior probabilities $\pi_{jk}$ in Eqn. 3.15 are replaced by a spatially varying function $f_{\text{MRF}}(z_{jk}|z_{Nj}, \Theta)$ following the smoothness model described in Section 2.3.1.

The exact structure of $f_{\text{MRF}}$ depends on the smoothness assumptions of $Z$. The Hammersley-Clifford theorem \[39\] states that for the function $f_{\text{MRF}}$ to be a Markov Random Field, it must be of the form,

$$ f_{\text{MRF}}(Z|\Theta) = \frac{1}{M} \exp \left\{ -U(Z|\Theta) \right\} ,$$  \hspace{1cm} (3.18) 

where $U(\cdot|\Theta)$, called the Gibbs potential, is usually a sum of functions of the neighborhoods of each voxel and $M$ is a normalizing constant. Unlike the Ising model and the Potts model \[40\], which require Monte-Carlo methods to solve the system, a computationally simpler enhancement to these models has been suggested in \[45,46\], where $U$ is taken as a sum of Gaussian functions. We follow this idea and define the MRF as,

$$ U(Z|\Theta) = \sum_{j \in \Omega} U(z_j|z_{Nj}, \Theta) = \sum_{j \in \Omega} \sum_{k=1}^{K} \ell_{jk} \sum_{i \in N_j} (z_{jk} - z_{ik})^2,$$  \hspace{1cm} (3.19) 

where $\ell_{jk}$ is a weighing function. From this Gibbs potential, a natural choice of $f_{\text{MRF}}(z_{jk}|z_{Nj}, \Theta)$ is

$$ f_{\text{MRF}}(z_{jk}|z_{Nj}, \Theta) = \frac{1}{\sqrt{2\pi \beta_k|N_j|L}} \exp \left\{ -\frac{\sum_{i \in N_j} (z_{jk} - z_{ik})^2}{2\beta_k^2} \right\} .$$  \hspace{1cm} (3.20)
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

Here, \( \mathcal{L} \) is a normalizing constant so as to make \( \sum_k f_{\text{MRF}}(z_{jk}|z_{Nj}, \Theta) = 1 \). The assumption behind such an MRF is that without any prior knowledge on the smoothness of the underlying segmentations \( \mathcal{Z} \), \( z_{jk} \) is assumed to be Gaussian distributed with mean \( \frac{1}{|N_j|} \sum_{i \in N_j} z_{ik} \) and variance \( \beta_k^2 \). This formulation assures that the spatial prior \( f_{\text{MRF}} \) is high if and only if segmentation of the \( j \)th voxel is the same as the segmentation of its neighborhood. It is also possible to estimate the variances \( \beta_k^2 \) by EM. Thus the parameter collection \( \Theta \) becomes

\[
\Theta = \bigcup_{k=1}^{K} \{v_k, \sigma_k, \beta_k\}. \tag{3.21}
\]

Now that we have defined \( \Theta \) and \( f_{\text{MRF}} \) in Eqn. 3.17, the maximum likelihood estimate of \( \Theta \) is described via the EM algorithm (Eqn. 2.6–Eqn. 2.7). The E-step requires computation of \( E(z_{jk}|y_j, \Theta) \). Using the fact that \( z_{jk} \) is a binary variable with \( z_{jk} \in \{0, 1\} \), it can be shown that \( P(z_{jk} = 1|y_j, \Theta) = E(z_{jk}|y_j, \Theta) \). Thus the conditional probability is also the conditional expectation. Define \( w_{jk}^{(m)} = E(z_{jk}|y_j, \Theta^{(m)}) \) as the conditional expectation at the \( m \)th iteration of the EM algorithm. Then the E-step gives

\[
\hat{f}_{\text{MRF}}(z_{jk}|z_{Nj}, \Theta^{(m)}) \approx \frac{1}{\sqrt{2\pi \beta_k^{(m)}|N_j|}} \mathcal{L}^{(m)} \exp \left\{ \frac{-\sum_{i \in N_j} (w_{ik}^{(m)} - w_{ik}^{(m)})^2}{2\beta_k^{(m)^2}} \right\}, \tag{3.22}
\]

\[
w_{jk}^{(m+1)} \approx \frac{\hat{f}_{\text{MRF}}(z_{jk}|z_{Nj}, \Theta^{(m)}) f_R(y_j|\Theta^{(m)})}{\sum_{k=1}^{K} \hat{f}_{\text{MRF}}(z_{jk}|z_{Nj}, \Theta^{(m)}) f_R(y_j|\Theta^{(m)})}, \tag{3.23}
\]

where \( z_{jk} \) is replaced by its current conditional expectation \( w_{jk}^{(m)} \) following the mean-field approach \([38, 44]\) to approximate the true conditional MRF by its current esti-
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

mantes.

The M step requires estimation of \( \Theta \) given the current segmentation \( w_{jk}^{(m)} \). The update equations are given by

\[
v_k^{(m+1)} = \frac{\sum_{j \in \Omega} w_{jk}^{(m)} y_j^{(m)} \gamma_{jk}^{(m)}}{\sum_{j \in \Omega} w_{jk}^{(m)}}, \tag{3.24}
\]

\[
\sigma_k^{(m+1)^2} = \frac{\sum_i w_{jk}^{(m)} \left( y_j^2 + v_k^{(m+1)^2} - 2 y_j v_k^{(m+1)} \gamma_{jk}^{(m)} \right)}{2 \sum_i w_{jk}^{(m)}}, \tag{3.25}
\]

\[
\beta_k^{(m+1)} = \sqrt{\frac{\sum_{j \in \Omega} \left( \sum_{i \in N_j} w_{jk}^{(m)} - w_{ik}^{(m)} \right)^2}{N}}. \tag{3.26}
\]

Here, \( N \) is the number of voxels in the image domain and

\[
\gamma_{jk}^{(m)} = \frac{I_1 \left( \zeta_{jk}^{(m)} \right)}{I_0 \left( \zeta_{jk}^{(m)} \right)} \quad \text{where} \quad \zeta_{jk}^{(m)} = \frac{y_j v_k^{(m)}}{\sigma_k^{(m)^2}}.
\]

We continue iterating through the EM algorithm until the increases in log-likelihood of successive iterations are below a threshold. The derivations of Eqns. 3.23–3.26 are provided in the Appendix.

The algorithm is executed in the following way. The parameters \( \{v_k, \sigma_k, \beta_k\} \) are first initialized by a k-means algorithm, then the estimates are fed to a 3-class GMM. The output of the GMM is used as the initialization of RiCE. Other clustering algorithms can also be used for initialization, but empirically, we have found that a good solution is reached quickly and the log-likelihood increases rapidly this way. This is
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

in accordance with previous findings [79], although the theoretical evidence, to the best of our knowledge, is still lacking [80].

We evaluate Eqns. 3.22 – 3.26 to get the parameters $\Theta^{(m)}$ and the posteriors $w^{(m)}_{jk}$. The final values of the $w^{(m)}_{jk}$’s are the expectations for the $j^{th}$ voxel to be included in the $k^{th}$ class, referred to as the “soft classification”. The hard segmentation for the $j^{th}$ voxel is given by $\max_k \{w_{jk}\}$, for each voxel $j$.

### 3.3.3 Validation

#### 3.3.3.1 Brainweb Phantom Validation

We first validated RiCE on the Brainweb phantom [81] and compared it with SPM [76], FAST [42], FANTASM [6] and a FCM based segmentation from Freesurfer [78], (mri_ms_EM function). SPM uses a Gaussian intensity model and it tries to recover the non-Gaussianity of the intensity distribution by modeling it with multiple Gaussians. FAST uses a Gaussian model on the log transformed intensities. Freesurfer and FANTASM use different variations of FCM. Thus, RiCE is directly comparable to SPM, while we compare it with the other methods to show the advantages of using a Rician model.

The phantom data comprises 15 phantoms, made up of five different noise levels (0-9%) and three different inhomogeneity levels (0, 20, 40%). Both the soft classification and the hard segmentation of the three tissues are shown in Fig. 3.8. The ground
Figure 3.8: Comparison of RiCE (2nd row) with ground truth (1st row) on true hard segmentation and fuzzy membership functions of a Brainweb phantom with 3% noise truth and the fuzzy memberships, from which the phantoms are generated, are also available and shown in the top row of Fig. 3.8. We use the true hard segmentation to find Dice coefficients of the three tissue classes for each of the methods.

Table 3.3 presents Dice coefficients for each of the noise levels averaged over three inhomogeneity levels. RiCE is comparable to the other methods, ranking in the top two in 16 out of 20 cases. As the phantoms are corrupted by Rician noise [81], RiCE gives better CSF segmentation than the Gaussian based method (SPM) on low noise levels, with a slightly reduced performance on high noise levels (7 – 9%), where it becomes comparable to both FAST and SPM. FAST, Freesurfer, SPM and RiCE do not perform as well as FANTASM on low noise data. We believe the reason for this is the small standard deviation of the PDF of the tissue classes, for which the EM
 iterations become unstable and may not converge to the true minima.

Table 3.3: Dice coefficients between the ground truth and hard segmentations of tissue classes, averaged over three inhomogeneity levels, are shown for different noise levels of Brainweb phantoms.

<table>
<thead>
<tr>
<th>Noise Level</th>
<th>0%</th>
<th>3%</th>
<th>5%</th>
<th>7%</th>
<th>9%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CSF</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAST</td>
<td>0.9312</td>
<td>0.9170</td>
<td>0.9295</td>
<td>0.9236</td>
<td>0.9255</td>
</tr>
<tr>
<td>Freesurfer</td>
<td>0.8560</td>
<td>0.8598</td>
<td>0.8561</td>
<td>0.8341</td>
<td>0.8014</td>
</tr>
<tr>
<td>FANTASM</td>
<td>0.9520</td>
<td>0.9456</td>
<td>0.9350</td>
<td>0.9176</td>
<td>0.8978</td>
</tr>
<tr>
<td>SPM</td>
<td><strong>0.9700</strong></td>
<td><strong>0.9547</strong></td>
<td>0.9400</td>
<td>0.9266</td>
<td>0.9010</td>
</tr>
<tr>
<td>RiCE</td>
<td>0.9561</td>
<td>0.9500</td>
<td><strong>0.9411</strong></td>
<td><strong>0.9301</strong></td>
<td><strong>0.9266</strong></td>
</tr>
<tr>
<td><strong>GM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAST</td>
<td>0.8394</td>
<td>0.9347</td>
<td>0.9337</td>
<td>0.9251</td>
<td><strong>0.9123</strong></td>
</tr>
<tr>
<td>Freesurfer</td>
<td>0.8496</td>
<td>0.8969</td>
<td>0.8611</td>
<td>0.8064</td>
<td>0.7474</td>
</tr>
<tr>
<td>FANTASM</td>
<td>0.9682</td>
<td>0.9582</td>
<td>0.9429</td>
<td>0.9179</td>
<td>0.8881</td>
</tr>
<tr>
<td>SPM</td>
<td>0.8997</td>
<td><strong>0.9590</strong></td>
<td>0.9426</td>
<td>0.9248</td>
<td>0.8952</td>
</tr>
<tr>
<td>RiCE</td>
<td><strong>0.9465</strong></td>
<td>0.9580</td>
<td><strong>0.9444</strong></td>
<td><strong>0.9250</strong></td>
<td>0.9100</td>
</tr>
<tr>
<td><strong>WM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAST</td>
<td>0.7448</td>
<td>0.9628</td>
<td>0.9545</td>
<td><strong>0.9416</strong></td>
<td>0.9292</td>
</tr>
<tr>
<td>Freesurfer</td>
<td>0.8691</td>
<td>0.9602</td>
<td>0.9230</td>
<td>0.8733</td>
<td>0.8231</td>
</tr>
<tr>
<td>FANTASM</td>
<td>0.9734</td>
<td>0.9647</td>
<td>0.9511</td>
<td>0.9304</td>
<td>0.9020</td>
</tr>
<tr>
<td>SPM</td>
<td>0.8483</td>
<td>0.9541</td>
<td>0.9575</td>
<td>0.9355</td>
<td>0.9014</td>
</tr>
<tr>
<td>RiCE</td>
<td><strong>0.9718</strong></td>
<td><strong>0.9710</strong></td>
<td><strong>0.9654</strong></td>
<td>0.9332</td>
<td><strong>0.9322</strong></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAST</td>
<td>0.8318</td>
<td>0.9405</td>
<td>0.9401</td>
<td><strong>0.9324</strong></td>
<td>0.9209</td>
</tr>
<tr>
<td>Freesurfer</td>
<td>0.8569</td>
<td>0.9105</td>
<td>0.8819</td>
<td>0.8362</td>
<td>0.7861</td>
</tr>
<tr>
<td>FANTASM</td>
<td><strong>0.9670</strong></td>
<td>0.9581</td>
<td>0.9443</td>
<td>0.9231</td>
<td>0.8949</td>
</tr>
<tr>
<td>SPM</td>
<td>0.8977</td>
<td>0.9518</td>
<td>0.9471</td>
<td>0.9308</td>
<td>0.9033</td>
</tr>
<tr>
<td>RiCE</td>
<td>0.9609</td>
<td><strong>0.9589</strong></td>
<td><strong>0.9548</strong></td>
<td>0.9262</td>
<td><strong>0.9294</strong></td>
</tr>
</tbody>
</table>
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

3.3.3.2 IBSR Validation

The next validation experiment was conducted on 18 normal healthy subjects from the Internet Brain Segmentation Repository (IBSR) [82]. The MR brain data sets and their manual segmentations were provided by the Center for Morphometric Analysis at Massachusetts General Hospital. The $T_1$-w coronal data is acquired on a 1.5T scanner. The manual whole head segmentations are used as a ground truth. Fig. 3.9 shows a slice of an image, with the manual and automatic segmentations from the five methods. As the manual segmentation does not include cortical CSF as a class, we combine CSF and GM as one class to compute Dice between the manual segmentation and the automatic segmentations. Table 3.4 shows the Dice coefficients of hard segmentation from each algorithm.

RiCE holds a higher score than FAST, SPM, and FANTASM for GM segmen-
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

tation, yielding a statistically significant improvement in these two cases (p-values of 0.012, 0.00002, 0.34, and 0.0004 for a pairwise t-test with FAST, FANTASM, Freesurfer, and SPM, respectively). For the WM segmentation, the performance of RiCE is not significantly different from the others. This experiment thus indicates that making the more rigorous Rician assumption does not deteriorate the performance of WM and GM segmentation and the segmentations from RiCE are comparable to those from the current available methods on WM and GM.

Table 3.4: Dice coefficients of GM, WM, and a volume weighted average (WA) between manual segmentations and hard segmentations obtained by FAST, FANTASM, Freesurfer FCM based segmentation (Freesurfer), SPM, and RiCE.

<table>
<thead>
<tr>
<th></th>
<th>FAST</th>
<th>FANTASM</th>
<th>Freesurfer</th>
<th>SPM</th>
<th>RiCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>GM</td>
<td>Mean</td>
<td>0.9271</td>
<td>0.9186</td>
<td>0.9340</td>
<td>0.9131</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.0109</td>
<td>0.0099</td>
<td>0.0076</td>
<td>0.0217</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0.8685</td>
<td>0.8685</td>
<td>0.8660</td>
<td>0.8558</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.0105</td>
<td>0.184</td>
<td>0.0151</td>
<td>0.0283</td>
</tr>
<tr>
<td>WA</td>
<td>Mean</td>
<td>0.9071</td>
<td>0.9100</td>
<td>0.9128</td>
<td>0.9030</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.0106</td>
<td>0.0132</td>
<td>0.0095</td>
<td>0.0251</td>
</tr>
</tbody>
</table>
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

Table 3.5: Dice comparison of Rician and Gaussian mixture models on BLSA data

<table>
<thead>
<tr>
<th></th>
<th>CSF Mean</th>
<th>SD</th>
<th>GM Mean</th>
<th>SD</th>
<th>WM Mean</th>
<th>SD</th>
<th>WA Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gauss</td>
<td>0.6872</td>
<td>0.0429</td>
<td>0.6700</td>
<td>0.0544</td>
<td>0.8205</td>
<td>0.0367</td>
<td>0.7376</td>
<td>0.0423</td>
</tr>
<tr>
<td>Rice</td>
<td>0.7589*</td>
<td>0.0386</td>
<td>0.7289</td>
<td>0.0384</td>
<td>0.8535*</td>
<td>0.0200</td>
<td>0.7924*</td>
<td>0.0249</td>
</tr>
</tbody>
</table>

*Statistically significantly larger than the Gaussian model (p-value < 0.05).

3.3.4 Comparison with a Gaussian model

In this sections, we will show the efficacy of using the Rician model over a comparable Gaussian one, by showing the improvement in segmentation consistency, both in terms of tissue classes as well as cortical surfaces.

3.3.4.1 Segmentation consistency

We carry out a consistency performance experiment on a set of 3T data from the Baltimore Longitudinal Study of Aging (BLSA) [1, 83], comprised of $T_1$-w axial MPRAGE and SPGR acquisitions (256 × 256 × 124 volumes having the resolution of 0.9375 × 0.9375 × 1.5 mm) of 14 normal subjects, ages in the range of 69 – 92. The SPGR acquisitions are registered to their corresponding MPRAGE acquisition using a rigid registration [84] and stripped using a hybrid registration based skull-stripping algorithm [33, 34]. Then each of the images is bias-corrected using N3.
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

Ideally, we expect to be able to generate identical segmentations of each subject from the different acquisitions. To compare with a Gaussian model, we modify Eqn. 3.17 keeping the smoothness $f_{MRF}$, while changing the Rician pdf $f_R(y|v, \sigma)$ from Eqn. 3.14 to a Gaussian one $f_G(y|v, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y-v)^2}{2\sigma^2}\right\}$, thereby modifying the Eqns. 3.22 – 3.26 accordingly.

Average Dice coefficients between the hard segmentations obtained from SPGR

![Figure 3.10: Rician and Gaussian fitting of histograms for an SPGR and an MPRAGE scan of the same subject](image)

Figure 3.10: Rician and Gaussian fitting of histograms for an SPGR and an MPRAGE scan of the same subject
and MPRAGE acquisitions of the same subject are reported in Table 3.5. Mean and standard deviations (Std) are calculated based on 14 normal subjects. The $p$-value for a null hypothesis, that the CSF Dice coefficient for the Rician model is smaller than that of the Gaussian model, is 0.0001. The $p$-values for a similar hypothesis on the GM, WM, and WA Dice coefficients are 0.022, 0.001, and 0.011, respectively. Thus, the consistency improves significantly on CSF segmentation, which is expected because the Rician distribution models the CSF intensity regime better than the Gaussian one, as seen by the fitting of the histograms of the SPGR and MPRAGE images (Fig. 3.10(a)–(b)), shown in Fig. 3.10(c)–(d). The KL distances between the actual histogram and the Rician and Gaussian fitting is 0.0129 and 0.0342, respectively, for MPRAGE, and 0.0876 and 0.1012 for SPGR. Thus, better fitting of the histograms provides more accurate delineation between the tissue classes. There is a large variability in the GM segmentation for both the Rician and the Gaussian models, which can be explained by the variability of the intensities of the sub-cortical structures, which is not explicitly modeled in this scenario.

### 3.3.4.2 Cortical surface consistency

Cortical thickness is an important measure for the neuroscience community [13]. As a consequence, robust, and accurate delineation of cortical surfaces are of importance. We study the Rician model on the consistent delineation of the cortical surfaces. We use a Cortical Reconstruction Using Implicit Surface Evolution
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

Figure 3.11: CRUISE outer surface comparison between the Rician model and a comparable Gaussian one.

(CRUISE) [23] to generate inner and outer surfaces from the soft classification. As the Rician model is visually more effective in modeling the CSF intensities of MPRAGE sequences (see Fig. 3.10(a)), we expect the CSF delineation of an MPRAGE scan (Fig. 3.11(a)) to be more accurate, which is shown in Fig. 3.11(d). A zoomed-in view is shown in Fig. 3.11(b). The CSF distribution in the image histogram is poorly fitted by a Gaussian in Fig. 3.10(d), which results in a under-estimation of the CSF-GM boundary, shown in Fig. 3.11(c), while a Rician model fits the histogram better and results in a more accurate estimate of the outer surface (Fig. 3.11(d)). An overlap of both the surfaces are shown in Fig. 3.11(e).

To show the improved consistency, we compare the CRUISE inner cortical surfaces
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

Figure 3.12: Comparison of the consistency in CRUISE inner cortical surface between the Rician model and a Gaussian one generated from the SPGR and MPRAGE acquisitions of the same subject. This is also shown in Fig. 3.12, where the CRUISE inner surfaces generated using the Rician model are shown to be closer in these two acquisitions. The Gaussian model does not lead to accurate estimation of the inner surface on the SPGR image due to the poor GM-WM contrast and the heavy partial volume effect (Fig. 3.12(e)), while a Rician model is better in this scenario (Fig. 3.12(f)). Quantitative distances between these surfaces are reported in Table 3.6. The surface distance is the mean of the distances between one surface and the other, while the distance from a point on the
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

Table 3.6: CRUISE surface Differences between Gaussian and Rician models.

<table>
<thead>
<tr>
<th></th>
<th>Inner Surface</th>
<th></th>
<th>Outer Surface</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
<td>Std</td>
</tr>
<tr>
<td>Gaussian</td>
<td>1.2276</td>
<td>0.1807</td>
<td>0.7869</td>
<td>0.1497</td>
</tr>
<tr>
<td>Rician</td>
<td>0.7022*</td>
<td>0.0987*</td>
<td>0.6001*</td>
<td>0.0901*</td>
</tr>
</tbody>
</table>

*Statistically significantly smaller than the Gaussian one (p-value < 0.05).

surface is the shortest distance to the other surface. The results are averaged on a pool of 14 normal subjects. A significantly large improvement in average inner surface difference is observed with the Rician model. Using a null hypothesis that the surface differences arising from RiCE are smaller than that of the corresponding Gaussian model, the p-values obtained from a t-test are 0.00001 and 0.022 for inner and outer surfaces, respectively.

3.3.5 Comparison with other methods

In this section, we compared the overall performance of our method with four other methods. Fig. 3.13 shows the comparison of the hard segmentations using the five algorithms. The Dice coefficients of the three classes and their volume weighted “average” Dice are shown in Table 3.7, averaged over 14 subjects. It can be visually seen that both the CSF and GM segmentation are more similar in the case of RiCE.
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

Figure 3.13: Hard segmentations of a pair of SPGR and MPRAGE acquisition of a BLSA subject, obtained using 5 different algorithms.

The \( p \)-values for a null hypothesis, that CSF Dice coefficient for RiCE is smaller than that of FAST/FANTASM/Freesurfer/SPM are 0.0016, 0.00001, 0.0002, and 0.0046 respectively. The \( p \)-values for a similar hypothesis on the GM, WM, and mean Dice coefficients (WA) are \([0.0128, 0.009, 0.003, 0.25], [0.1258, 0.00001, 0.0002, 0.24],\) and \([0.425, 0.0007, 0.0041, 0.07]\), respectively, thus showing significant improvement in CSF segmentation consistency over the other four methods. This experiment also
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

Table 3.7: Dice coefficients between segmentations of a $T_1$-w SPGR and $T_1$-w MPRAGE acquisitions for RiCE, FAST, FANTASM, Freesurfer, and SPM segmentations.

<table>
<thead>
<tr>
<th></th>
<th>FAST</th>
<th>FANTASM</th>
<th>Freesurfer</th>
<th>SPM</th>
<th>RiCE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CSF</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>0.6758</td>
<td>0.6317</td>
<td>0.6829</td>
<td>0.7207</td>
<td>0.7589 *</td>
</tr>
<tr>
<td><strong>Std</strong></td>
<td>0.0431</td>
<td>0.0504</td>
<td>0.0401</td>
<td>0.0264</td>
<td>0.0386</td>
</tr>
<tr>
<td><strong>GM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>0.6889</td>
<td>0.6697</td>
<td>0.6381</td>
<td>0.7201</td>
<td>0.7289 †</td>
</tr>
<tr>
<td><strong>Std</strong></td>
<td>0.0621</td>
<td>0.0549</td>
<td>0.0690</td>
<td>0.0280</td>
<td>0.0384</td>
</tr>
<tr>
<td><strong>WM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>0.8288</td>
<td>0.7960</td>
<td>0.8141</td>
<td>0.8589</td>
<td>0.8535</td>
</tr>
<tr>
<td><strong>Std</strong></td>
<td>0.0329</td>
<td>0.0274</td>
<td>0.0352</td>
<td>0.0195</td>
<td>0.0200</td>
</tr>
<tr>
<td><strong>WA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>0.7392</td>
<td>0.7151</td>
<td>0.7309</td>
<td>0.7785</td>
<td>0.7924</td>
</tr>
<tr>
<td><strong>Std</strong></td>
<td>0.0407</td>
<td>0.0358</td>
<td>0.0376</td>
<td>0.0220</td>
<td>0.0249</td>
</tr>
</tbody>
</table>

* Statistically significantly larger than all the other ones ($p$-value < 0.05).

† Statistically significantly larger than FAST, FANTASM, and Freesurfer ($p$-value < 0.05).

shows that the Rician model does not do worse than a Gaussian model on GM and WM segmentation (e.g., SPM), while improving on FCM based methods. Thus, the Rician model is significantly more consistent in a Gaussian model on low SNR regimes.

Fig. 3.14 shows a visual comparison between the difference of the surfaces generated using the soft classifications from FAST, FANTASM, Freesurfer, SPM, and RiCE using the CRUISE pipeline [23]. The mean surface difference in mm, aver-
Figure 3.14: Surface difference between the inner and outer CRUISE cortical surfaces generated from a pair of SPGR and MPRAGE scans of the same subject aged over 14 normal subjects, between surfaces generated from SPGR and MPRAGE images are reported in Table 3.8. Using a null hypothesis that the inner surface differences arising from FAST/FANTASM/Freesurfer/SPM are smaller than that of RiCE, the \( p \)-values obtained from a \( t \)-test are 0.0004, 0.000006, 0.0003, and 0.0421, respectively. A similar hypothesis on the outer surfaces give the following \( p \)-values 0.0032, 0.000001, 0.000001, and 0.00005 for FAST/FANTASM/Freesurfer/SPM, respectively, confirming that RiCE produces more consistent cortical surface delineations.

3.3.6 Discussion and Future Work

In this section, we have proposed a Rician pdf–based brain MR segmentation technique. We concentrated on consistent segmentation of three primary tissues, cerebrospinal fluid, gray matter, and white matter, from \( T_1 \)-w MR images acquired.
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

Table 3.8: Surface differences of the consistency experiment.

<table>
<thead>
<tr>
<th></th>
<th>Inner Surface</th>
<th></th>
<th>Outer Surface</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
<td>Std</td>
</tr>
<tr>
<td>FAST</td>
<td>0.8852</td>
<td>0.1996</td>
<td>0.7607</td>
<td>0.1376</td>
</tr>
<tr>
<td>FANTASM</td>
<td>1.2375</td>
<td>0.1997</td>
<td>0.9234</td>
<td>0.0891</td>
</tr>
<tr>
<td>Freesurfer</td>
<td>1.0356</td>
<td>0.1551</td>
<td>0.9446</td>
<td>0.2042</td>
</tr>
<tr>
<td>SPM</td>
<td>0.7829</td>
<td>0.0949</td>
<td>0.8213</td>
<td>0.0917</td>
</tr>
<tr>
<td>RiCE</td>
<td>0.7106 *</td>
<td>0.1017</td>
<td>0.6114 *</td>
<td>0.1001</td>
</tr>
</tbody>
</table>

* Statistically significantly smaller than all the other ones (p-value < 0.05).

with two different pulse sequences, MPRAGE and SPGR. The underlying acquisition parameters, like repetition time, inversion time, or flip angle, are usually different from one sequence to another, which gives rise to the variability of the tissue contrast. With exact knowledge of the acquisition parameters and the imaging sequences, consistent tissue segmentations can be obtained [60], but for most studies, either the parameters are not available or the imaging sequences are difficult to model accurately. Hence, most statistical segmentation algorithms rely on probabilistic modeling of the intensities only. It is difficult to remove inconsistencies in the segmentations between images from different pulse sequences without the exact knowledge of the acquisition
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

process, which is the primary source of the variability in the contrast.

Both SPGR and MPRAGE sequences are often used to obtain $T_1$-w MR images. They are gradient-echo sequences, but have widely variable tissue contrast due to the difference in acquisition processes and the imaging parameters. Nevertheless, the MR image intensity at each voxel follows Rician distribution for both these pulse sequences, although most of the current statistical model based segmentation techniques assume an underlying Gaussian distribution. Specifically, it can be seen that CSF and GM, having low SNR in $T_1$-w images, are not modeled correctly by Gaussians (Fig. 3.7 and Fig. 3.10). As a result, the segmentations of $T_1$-w images with different pulse sequences become inconsistent. We have shown that introducing a Rician mixture model produces more consistent segmentation between SPGR and MPRAGEs, both in terms of hard segmentation of tissues and delineation of cortical surfaces. The use of the Rician distribution to replace Gaussian distributions is shown to be promising. Modeling tissue classes in this mono-model manner ignores the true complexity of tissue structures and the local variation that is possible within a tissue. This topic, in light of this advancement in the correct tissue model, is a rich area for future work.

Our algorithm is fully automatic and no training data is required. We correct the image inhomogeneities by a non-parametric model and use Markov random field to introduce segmentation consistency. We have validated the algorithm on the Brainweb phantom and IBSR 20 normal subjects. The improvement in segmentation consis-
CHAPTER 3. IMPROVING SEGMENTATION CONSISTENCY

tency is demonstrated on 14 BLSA subjects having both SPGR and MPRAGE scans. The algorithm takes approximately 10 minutes on a 3GHz Intel processor on a Linux workstation. Future work will focus on incorporating \textit{a priori} information via statistical atlases.

3.4 Summary

In this chapter, we have described two methods, FANTASM-VC and RiCE, for the segmentation of MR images, both primarily developed to achieve consistency in segmentations between images acquired with different pulse sequences. While FANTASM-VC, based on fuzzy C-means, gives membership functions for the tissues, RiCE models the MR image intensities with a finite mixture model and provides posterior probabilities in an optimal way. Both the methods are shown to produce more consistent segmentations across two $T_1$-w acquisitions, SPGR and MPRAGE.

Although the methods are quite successful in segmenting images, they overlook a critical aspect of normalizing the intensities. While segmenting images from different subjects obtained in different sites and scanners, it is usually inherently assumed that their intensities are normalized. However, intensity distributions change with the slightest modification in the calibration of the scanners, pulse sequence parameters and even with the temperature of the scanners. Thus, instead of improving segmentation methods to achieve consistent segmentations, in the following section,
we explore the effect of intensity normalization, such that any segmentation method provides a certain level of consistency.
Chapter 4

A Sparse Prior Based Image Normalization and Synthesis

The tissue contrast of an MR neuroimaging dataset has an impact on the performance of image analysis tasks such as segmentation and registration [18]. Post-processing of MR brain images, particularly image segmentation, has been used for many scientific purposes such as furthering our understanding of normal aging [1,12], disease progression [13,14] and prognosis [15]. When large multi-site and multi-center studies are involved (e.g., [3,16,17]), two major difficulties in carrying out consistent image processing across all the available imaging data often emerge. First, due to scanner and pulse sequence differences, the image intensities in the most fundamental anatomical acquisitions – typically $T_1$-w SPGR or MPRAGE pulse sequences – do not share a common scale, resulting in inconsistencies among the processed results.
CHAPTER 4. SPARSE PRIOR BASED CONTRAST SYNTHESIS

Second, some additional tissue contrasts that might have been acquired at one site for a local study, such as $T_2$-w or Fluid Attenuated Inversion Recovery (FLAIR) acquisitions, might not have been acquired at another site, resulting in entirely missing tissue contrasts. This may prevent certain image processing steps from being carried out or their results applied to alternate data may be inconsistent with the rest of the study.

Unlike computed tomography, it is a fundamental problem in MRI that the image intensities do not have any specific numeric meaning, and their values differ with the pulse sequence parameters, the specific implementation of the pulse sequence, the scanner manufacturer, and the scanner’s calibration [21]. As a concrete example, the intensity histograms are evidently quite different for two common structural imaging pulse sequences of the same subject on the same scanner (Fig. 4.1, first two columns) and using the same pulse sequence on different scanners (Fig. 4.1, three rightmost columns). This lack of consistency in the image intensity scale causes inconsistencies in image processing tasks such as segmentation [11,18,20], which is of primary concern in this chapter.

In this chapter, we describe a technique that uses image patches from an atlas to synthesize contrasts not present or not intensity normalized in the original dataset. The proposed image synthesis technique is demonstrated using two applications. First, it is used to normalize the intensity of images acquired using the same pulse sequence but with different parameters or on different scanners. Second, it is
Figure 4.1: The top row shows four different acquisitions of the same subject under different scanners and pulse sequences, while the bottom row shows the corresponding intensity histograms. The scale on the horizontal axis is arbitrary.

used to synthesize images with a different tissue contrast than that which was acquired. Unlike previous synthesis methods, the proposed method does not require images to be acquired using a particular pulse sequence or set of pulse sequences or to carry out an image registration procedure as part of the process. The normalization method can be applied as a pre-processing step to a segmentation method and is shown to yield more consistent segmentations on large datasets and to work in the presence of mild pathologies.
CHAPTER 4. SPARSE PRIOR BASED CONTRAST SYNTHESIS

4.1 Background on Sparse Priors

We use the idea of sparse priors in our MIMECS approach. The idea behind sparse priors comes from the fact that most signals that we observe are usually sparse in some basis, thus it is better not to observe the full signal, but a part of it, and reconstruct the whole signal “exactly” from those small number of measurements. By observing a small number of measurements, the dimensionality of the data can also be reduced in a lossless fashion. Recently this idea of sparse representation has been successfully applied to many image processing algorithms, like denoising [85], image restoration [86, 87] and super-resolution [88], often improving on the state-of-the-art algorithms. We use the sparse representation of image patches to synthesize new image contrasts.

Suppose we want to reconstruct a signal \( x \in \mathbb{R}^N \) which is \( s \)-sparse, i.e., has at most \( s \) non-zero elements. We observe another signal \( b \in \mathbb{R}^d \), \( s < d < N \), such that each element of \( y \) can be obtained by an inner product of \( x \) and another vector from \( \mathbb{R}^N \). Then, \( x \in \mathbb{R}^N \) can be recovered exactly from \( b \in \mathbb{R}^d \), with \( b = Ax \), \( d < N \), \( x \) being \( s \)-sparse, \( A \in \mathbb{R}^{d\times N} \), under some restrictions on \( d \), \( s \), \( N \) and \( \Phi \). It is also a way to reduce the dimension of sparse data \( (x) \) from \( N \) to \( d \) in a lossless way.

Clearly, \( b = Ax \), \( d < N \), is an under-determined system, with \( N \) unknowns and \( d \) constraints. Hence, additional constraints on \( x \) are needed to solve the system. One
CHAPTER 4. SPARSE PRIOR BASED CONTRAST SYNTHESIS

assumes a sparsity prior on $\mathbf{x}$ by penalizing its support

$$\hat{\mathbf{x}} = \min ||\mathbf{x}||_0 \text{ such that } ||\mathbf{b} - A\mathbf{x}||_2^2 < \epsilon_1,$$  \hspace{1cm} (4.1)

where $\epsilon_1$ is the noise in the measurement and the $\ell_0$ norm $||\cdot||_0$ indicates the number of non-zero elements in the vector. Although this approach provides some simple conditions on $A$ to have a feasible solution, it is a combinatorial problem and NP-hard. It has been shown that under certain conditions on $||\mathbf{x}||_0$, the optimal solution of Eqn. 4.1 can also be found by solving an $\ell_1$ problem

$$\hat{\mathbf{x}} = \min ||\mathbf{x}||_1 \text{ such that } ||\mathbf{b} - A\mathbf{x}||_2^2 < \epsilon_2,$$  \hspace{1cm} (4.2)

where $||\cdot||_1$ is the $\ell_1$ norm. This is a convex programming problem and can be solved in polynomial time. If $\epsilon_2$ is unknown, Eqn. 4.2 can be written in the following form

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \{||\mathbf{b} - A\mathbf{x}||_2^2 + \lambda||\mathbf{x}||_1\},$$  \hspace{1cm} (4.3)

where $\lambda$ is a weighting factor. The sparsity on $\hat{\mathbf{x}}$ increases as $\lambda$ increases.

Eqn. 4.3 can also be thought of as a dimensionality reduction problem, where an $N \times 1$ vector $\mathbf{x}$ is projected onto a linear space $A$ to obtain a $d \times 1$ vector $\mathbf{b}$, and this process is invertible if it is possible to find an appropriate set of vectors ($A$) and $\mathbf{x}$’s are “sufficiently” sparse. Recent works [89,90] suggest constraints on $A$ as well as on $\mathbf{x}$ for unique and optimal solutions of Eqn. 4.2 as well as the equivalence of Eqn. 4.1 and Eqn. 4.2. In a simplest case, when $A \in \mathbb{R}^{d \times 2d}$ consists of two orthonormal bases, $\hat{\mathbf{x}}$, the solution of Eqn. 4.2, is also the unique sparsest optimal solution, i.e., the solution
of Eqn. 4.1 if $||\hat{x}||_0$ is bounded above. The bound is a function of the maximum of the dot-products between columns of $A$. Mathematically, let $A = [\Psi, \Xi]$, where $\Psi$ and $\Xi$ are $d \times d$ orthonormal matrices. Assume that $\psi_i$ and $\xi_i$ are the columns of $\Psi$ and $\Xi$. Then the uniqueness of Eqn. 4.2 is guaranteed for any $x$ with sparsity $||x||_0 < \frac{1}{M}$, where

$$M = \sup_{i,j} |\psi_i^T \xi_j|, 1 \leq i, j \leq d. \quad (4.4)$$

This result can be extended to any $A$ that satisfies a global restricted isometry property (RIP) [91]. However, to the best of our knowledge, there is no closed form solution, such as Eqn. 4.4, to find if any matrix $A$ satisfies RIP. As a thumb rule, a desirable property of a $A$ is that the columns should be as uncorrelated as possible, following the result in Eqn. 4.4. See [92,93] for more details on RIP.

### 4.1.1 Phase Transition Diagram

Phase transition diagrams [94] have been proposed to find asymptotic conditions on $A$, $x$ and $b$’s so as to satisfy the equivalence of Eqn. 4.1 and Eqn. 4.2. It has been shown that if the elements of $A$ are i.i.d. $\mathcal{N}(0, \frac{1}{d})$, then the $\ell_0$ and $\ell_1$ problems are equivalent with a high probability under certain conditions on the sparsity of $x$ [93,95], given by

$$s \leq \frac{Cd}{\log \frac{N}{d}}, \quad (4.5)$$
where $s$ is the sparsity of $\mathbf{x}$ and $A \in \mathbb{R}^{d \times N}$. The exact value of $C$ is given in [91] for Gaussian matrices. The upper bound on $s$ indicates that if $\mathbf{x}$ is sufficiently sparse, the $\ell_0$ and $\ell_1$ solutions are equivalent. Similar simple bounds can be obtained for Fourier matrices as well as other orthogonal matrix ensembles.

However, the upper bound is not generalizable to any arbitrary matrices. In those cases, phase transition diagrams try to find an asymptotic relation between $\rho = s/d$ and $\delta = d/N$ [94,96], similar to one in Eqn. 4.5, by plotting the probability that the solutions of the $\ell_0$ and $\ell_1$ problems, i.e., Eqn. 4.1 and Eqn. 4.2, are different. The probability is, in general, a function of both $\rho$ and $\delta$. $\rho$ denotes the undersampling ratio, i.e., how sparse the data $\mathbf{x}$ actually is, and $\delta$ denotes the oversampling ratio,
i.e., how much reduction in the dimensionality of the data (from $x$ to $b$) is tolerable for perfect reconstruction, ensuring $\rho, \delta \in [0, 1]$. Once the probability map is obtained via simulation for different values of $\rho$ and $\delta$ for a particular $A$, a 0.5 level curve will give an estimate of the maximum tolerable sparsity of $x$ to be perfectly reconstructed using that $A$. An example of the asymptotic 0.5 level curve of the phase transition diagram for a random matrix is shown in Fig. 4.2, where the elements of the matrix are taken from $\mathcal{N}(0, 1)$. The bottom half of the curve denotes a set of $(\rho, \delta)$ pairs for which the probability that $\ell_0$ and $\ell_1$ problems are different is less than 0.5. Thus, the operating point $(\rho, \delta)$ should always lie in the bottom half on the curve. The exact form of the curve is only available for Gaussian matrices [96]. For other matrices, only asymptotic level curves can be obtained via simulation. As an example, if $x \in \mathbb{R}^{100}$ and $b \in \mathbb{R}^{40}$, the operating point is $\delta = 0.40$. From the diagram, the maximum $\rho$ can be approximately 0.35, indicating, $s \leq 0.35d = 14$. Thus maximum 14-sparse vectors $x$ can be recovered exactly via this $A$ with a probability greater than 0.5. We use the phase transition diagram in to evaluate our choice of matrices that we use in sparse priors framework.

4.2 Previous Works

Some popular ways to do intensity normalization are histogram equalization or matching by non-rigid registration [97], intensity scaling based on landmarks [2, 4, 98],
intensity correction by a global piecewise linear [2], or a polynomial mapping [59]. A common problem with these methods is that the histogram matching is never perfect with discrete-valued images. Also landmark based methods are mostly initialized by manually chosen landmarks, which are time consuming to create and lack reproducibility. Another approach to normalization is the use of the peak intensities of the WM or GM in the histogram [99] or by matching the image histogram to a target histogram [100]. Histograms of the segmented sub-cortical structures can also be matched to the individual histograms of the sub-cortical structures of a registered atlas [58]. Although this method produces consistent sub-cortical segmentations over datasets acquired under different scanners, it requires a detailed segmentation of the images. If multiple images of the same subject are obtained with predefined, precisely calibrated acquisition parameters, then the images can be normalized by incorporating the acquisition physics [60]. Unfortunately, it is not feasible to obtain images with such parameters in routine clinical imaging.

In many neuroimaging studies and in routine clinical imaging, different pulse sequences are used to detect different features of the brain. For example, to quantify the condition of Alzheimer’s disease by the progression of white matter hyperintensities (WMH) [101, 102], most studies focus on acquiring $T_2$-w, $P_D$-w, and FLAIR sequences to provide better contrast for the WMHs compared to a conventional $T_1$-w image. A more classical example of the importance of pulse sequence selection for the purpose of cortical delineation is the choice of $T_1$-w SPGR compared to $T_1$-w
CHAPTER 4. SPARSE PRIOR BASED CONTRAST SYNTHESIS

MPRAGE. Clearly, CSF-GM contrast is better in SPGRs, while GM-WM contrast is better in MPRAGEs. If a tissue contrast is needed for an image processing task, but was not acquired in the study, it is useful to synthesize an image having the missing contrast. One approach to this synthesis task is to simulate the missing pulse sequences using imaging equations [103]. The underlying maps of tissue proton density $P_D$ and relaxation times $T_1$, $T_2$, and $T_{2}^*$ are required for this purpose. These can be found by acquiring multiple images of the same subject with predefined acquisition parameters like repetition time ($T_R$), echo time ($T_E$), flip angle, etc. [60,61]. This approach has two disadvantages: 1) the imaging equations are often not very accurate, and 2) for large studies, images cannot all be acquired with predefined acquisition parameters.

Another way to generate alternate contrasts is to register the subject to a multiple-contrast atlas and to transfer the desired contrast intensities from the atlas space to the subject space [104]. This approach fails if the registration is poor or if there are pathologies or unknown tissues in the subject that are not present or spatially distributed differently in the atlas. In the example shown in Fig. 4.3, a $T_1$-w SPGR atlas (Fig. 4.3(a)) was registered to a similar contrast subject image (Fig. 4.3(c)) using the (inverse consistent and diffeomorphic) SyN registration algorithm [5]. The atlas MPRAGE contrast intensities (Fig. 4.3(b)) were then mapped through the transformation to the subject space, yielding a synthetic subject MPRAGE image (Fig. 4.3(e)). By comparing this result to the subject’s true MPRAGE image
(Fig. 4.3(d)), it is apparent that this approach failed to adequately define the ventricles (Fig. 4.3(d), arrow adjacent to the ventricles) due to shortcomings in the registration, and it also failed to represent the white matter lesions that are present in the original image (Fig. 4.3(d), arrow denoting the lesion posterior to the ventricles) because such lesions are not present in the atlas.

The method proposed in this section, MR image example-based contrast synthesis (MIMECS), combines the idea of example-based image hallucination [55] with sparse priors [88, 92] to synthesize multiple MR contrasts using patches from a multiple-contrast atlas. The core idea of MIMECS is that a subject patch can be reconstructed from a linear combination of a few relevant patches from a dictionary. It is different from classical histogram matching in the sense that a patch can be thought of as a feature vector of the center voxel of the patch thus including local neighborhood information for that voxel. We use MIMECS in two applications: 1) image intensity normalization over scanners or pulse sequences with different parameters and 2) generating a missing tissue contrast in a dataset. Both uses are for the purpose of generating consistent segmentations from multi-site, multi-scanner neuroimaging studies. There are several advantages of MIMECS compared to previous synthesis methods. First, it is a completely automatic pre-processing step that can precede any image processing task. Second, there is no need to use multiple pulse sequences in order to estimate underlying tissue parameters. Third, even though MIMECS uses an atlas, there is no need to carry out subject-to-atlas registration, thus avoiding
Figure 4.3: (a) Atlas $T_1$-w SPGR and (b) its corresponding $T_1$-w MPRAGE. (c) A subject $T_1$-w SPGR scan and (d) its $T_1$-w MPRAGE image. The atlas SPGR is deformably registered (using SyN [5]) to the subject SPGR. This deformation is applied to the atlas MPRAGE to obtain (e) a synthetic subject MPRAGE.
potential errors due to mis-registration or missing tissues. Finally, it does not require any segmentation of the subject.

4.3 Method

As discussed in Sec. 4.1, sparse signal reconstruction assumes that because most signals are sparse in some way, it is not necessary to observe the full signal in order to accurately reconstruct it. Suppose we want to reconstruct a vector $x \in \mathbb{R}^N$ that is $s$-sparse, i.e., having at most $s$ non-zero elements, and our observations are $b = Ax$, where $b \in \mathbb{R}^d$, $s < d < N$, and $A \in \mathbb{R}^{d \times N}$. Since this is an under-determined system, we cannot directly invert the system to find $x$, and additional penalties or constraints must therefore be used. Since $x$ is known to be sparse, it makes sense to formulate an objective function that tries to find a sparse solution that is also consistent with the measurements, as follows

$$\hat{x} = \min \| x \|_0 \text{ such that } \| b - Ax \|_2^2 < \epsilon_1.$$  \hspace{1cm} (4.6)

Here, $\epsilon_1$ is the noise level in the measurements, $\| \cdot \|_p$ is the $\ell_p$ norm; in particular $\| x \|_0$ is the number of non-zero elements in $x$. There are relatively simple conditions on $A$ that guarantee a feasible solution to Eqn. 4.6, however the solution of this problem is combinatorial in nature, and therefore NP-hard. It has been shown that if $\| x \|_0$ is small [95, 105], the optimal solution to Eqn. 4.6 can also be found by solving

$$\hat{x} = \min \| x \|_1 \text{ such that } \| b - Ax \|_2^2 < \epsilon_2,$$
CHAPTER 4. SPARSE PRIOR BASED CONTRAST SYNTHESIS

This is a convex programming problem and can be solved in polynomial time. If $\epsilon_2$ is unknown, we can rewrite the above in the following form

$$\hat{x} = \arg \min_x \{ ||b - Ax||_2^2 + \lambda ||x||_1 \}, \quad (4.7)$$

where $\lambda$ is a weighting factor. The resulting sparsity in $\hat{x}$ decreases as $\lambda$ increases. The formulation in Eqn. 4.7 is now the standard way to frame a “sparse prior” estimation problem and is the core idea behind MIMECS.

The MIMECS method is based on analysis of image patches, which are $p \times q \times r$ 3D patches associated with each voxel of the image. Patches are typically small and centered on the voxel of interest—e.g., $p = q = r = 3$ or $p = q = r = 5$. It is convenient for writing out the mathematics to describe a patch as the 1D vector $b$ of size $d \times 1$, where $d = pqr$.

### 4.3.1 Atlas Description

We define an atlas as a pair of images, $\{a_1, a_2\}$, where $a_1$ and $a_2$ have tissue contrasts $C_1$ and $C_2$, respectively, and are co-registered and sampled on the same voxel locations in space. At each voxel, 3D patches can be defined on each image and stacked as 1D vectors of size $d \times 1$. The patches are then denoted by the $d \times 1$ vectors $a_1(i)$ and $a_2(i)$, where $i = 1, \ldots, N$, is an index over the voxels of $a_1$. Then we define the $C_1$ and $C_2$ contrast dictionaries $A_1$ and $A_2 \in \mathbb{R}^{d \times N}$, where the columns of $A_1$ and $A_2$ are patches $a_1(i)$ and $a_2(i)$, from the atlas. By construction, a column
of $A_1$ contains a patch whose corresponding patch in the second contrast image can be found in the same column of $A_2$.

4.3.2 Imaging Model

The mathematical relationship between two MR images $b_1$ and $b_2$ of the same person, acquired at the same time but with differing pulse sequences or contrasts (say $C_1$ and $C_2$), is nonlinear in general and is described by the underlying physics of nuclear magnetic resonance. When working with patches, however, it is reasonable to approximate the relationship between the image values in the $j$-th patch $b_c(j)$ having contrast $c$ (= 1 or 2) and the underlying tissue values $\Theta(j)$ with a linear approximation

$$b_c(j) = W_c \Theta(j) + \eta_c(j).$$

where $W_c$ is a linearization of the pulse sequence with contrast $c$ at patch $j$ and $\eta_c(j)$ is noise. The underlying tissue values $\Theta(j)$ often consists of the $T_1$, $T_2$ and $P_D$ values for the $j^{th}$ voxel, as seen in imaging equations, e.g., [103]. If $W_c$ were known and the noise were i.i.d. Gaussian, then an estimate of $\Theta(j)$ might be given by the normal equations, and we could estimate a pseudo-inverse of the transfer function $G$ relating the $b_c$'s as follows

$$b_2(j) = G b_1(j), \quad (4.8)$$
CHAPTER 4. SPARSE PRIOR BASED CONTRAST SYNTHESIS

where

\[ G = W_2(W_1^T W_1)^{-1} W_1^T. \] (4.9)

However, for real images, \( W_c \) is generally not known because it depends on precise knowledge of the imaging parameters, and therefore \( G \) is very difficult to estimate. We avoid estimation of \( G \), by using the implicit relationship between atlas patch pairs in a sparse priors algorithm [88], the details of which are given in Sec. 4.3.3, and the idea is as follows. Given a patch written as a linear combination of relevant patches from a contrast dictionary, an alternative contrast can be synthesized using the same linear combination and the alternative contrast dictionary. Thus, assuming for some patch \( b_1(j) \) we have identified coefficients \( x(j) \) that yield the patch as a linear combination of \( C_1 \) atlas patches, i.e., \( b_1(j) = A_1 x(j) \). Then an estimate \( \hat{b}_2(j) \) of the \( C_2 \) contrast patch is given by

\[
\hat{b}_2(j) = G b_1(j),
\]

\[
= G A_1 x(j),
\]

\[
= A_2 x(j). \] (4.10)

Thus, the alternative contrast patch is estimated using the examples from the dictionaries \( A_1 \) and \( A_2 \) without an explicit estimation of \( G \).
4.3.3 Contrast Synthesis Algorithm

We first decompose the subject (input) image $b_1$ into $d \times 1$ patches $b_1(j)$, where $j$ is an index over the voxels of the input image. We maintain that for a subject image patch $b_1(j)$, a small number of relevant and similar examples can always be found from a rich and overcomplete patch dictionary $A_1$ which is of the same contrast as $b_1$. It is unlikely that a single patch from $A_1$ will perfectly match $b_1(j)$, but it is quite likely that an optimal linear combination of a small number of patches will yield a very close approximation. The problem of finding a few patches to form the linear combination can be solved by assuming the weight vector $x(j)$ is sparse, leading to

$$b_1(j) \approx A_1 x(j), \text{ for some } x(j) \in \mathbb{R}^N, \text{ with } ||x(j)||_0 \ll N, \forall j, \quad (4.11)$$

where $A_1$ is a subset of patches taken from the atlas image $a_1$. A subset of patches is used to save computation time.

The sparsest representation $x(j)$ is just a single column of $A_1$ which is approximately equal to $b_1(j)$. In this case, the corresponding column of $A_2$ gives the $C_2$ contrast of $b_1(j)$. However, we seek a small number $s$ of patches from $A_1$ whose linear combination gives $b_1(j)$. The motivation behind choosing a sparse $x(j)$ to reconstruct $b_1(j)$ is twofold. First, the linear approximation $G$ of the imaging equations, as assumed in Eqn. 4.9, will hold for those atlas patches that are close in intensity, i.e., those that are of the same tissue classes, so it would be undesirable to pick a large number of patches that might tend to mix the tissue classes. Second, if too many
similar atlas patches are used to reconstruct $b_1(j)$, then the corresponding $C_2$ contrast patch will be overly smooth due to the cumulative effects of small mismatches in each patch. Empirically, we have found that non-sparse $x(j)$ tends to produce smooth images acting much like a denoising process, while highly sparse $x(j)$ produces noisy images. In Sec. 4.3.5, an empirical way of choosing the sparsity level is described.

Since the combinatorics of Eqn. 4.11 makes it infeasible to solve this problem directly, we use an $\ell_1$ minimization strategy and reformulate the problem as

$$\hat{x}(j) = \arg \min_{x \geq 0} \left[ ||b_1(j) - A_1 x||_2^2 + \lambda ||x||_1 \right], \forall j,$$

subject to $||a_1(i)||_2^2 = 1, \ i = 1, \ldots, N$. \hspace{1cm} (4.12)

As in Eqn. 4.7, $\lambda$ is a weighting factor that encourages a sparse solution when it is large. Given this optimal combination of the patches in $A_1$, which best approximates the subject patch $b_1(i)$, the $C_2$ contrast patch is estimated using Eqn. 4.10. The full contrast image $b_2$ is reconstructed by taking the union of all central voxels in the reconstructed patches.

There are two key differences in our solution Eqn. 4.12 as compared to Eqn. 4.7. The first difference is the positivity constraint on the coefficients $x$ in Eqn. 4.12. The reason we impose this condition is to encourage the selected patches—i.e., those used in the sparse reconstruction of the subject patch—to have the same anatomical meaning as the subject patch. For example, we do not want a boundary that is gray matter on the left and white matter on the right to be approximated with a negative coefficient multiplying a patch having exactly the opposite orientation.
CHAPTER 4. SPARSE PRIOR BASED CONTRAST SYNTHESIS

Although the image being approximated might “look” just the same, a patch that is synthesized with a new contrast would correspond to tissues having the wrong underlying physical parameters, and would therefore be likely to have the wrong appearance. This condition is therefore designed to encourage “gray matter patches” to be used to synthesize “gray matter patches” and so on. We note that computational aspects of this positivity constraint has been previously explored in Lasso [106], and that this constrained $\ell_1$ minimization procedure has been shown to be equivalent to an $\ell_0$ minimization procedure (like Eqn. 4.1 with a similar positivity constraint) if $x(j)$ is sufficiently sparse [107].

The second difference between Eqn. 4.12 and Eqn. 4.7 is the requirement that the $\ell_2$ norm of columns of $A_1$ (denoted $a_1(i), i = 1, \ldots, N$) be unity. This constraint is necessary in order to guarantee the uniqueness of the solution, and is a common feature of patch-based techniques in computer vision and image processing [88, 108]. However, unit normalization of patches taken from the atlas MR images, removes the relationship between overall patch intensity and the patch texture, and it also discourages the establishment of a correct anatomical meaning of the patches as described above. Therefore, in the next section, we propose a novel approach that normalizes patches in the $(d + 1)$-dimensional space rather than the $d$-dimensional space in order to preserve the desired intensity information.
4.3.4 Normalization of the Matrices

To guarantee a unique solution to Eqn. 4.12 (by removing the scaling ambiguity), the columns of $A_1$ are normalized such that $||a_1(i)||_2^2 = 1, \forall i = 1, \ldots, N$. However, if scale is directly removed from an MR image patch then a key feature in distinguishing tissue types is lost and patches used in the sparse reconstruction are less likely to be selected from the same tissue type. It is common in the sparse reconstruction literature to use prior training to learn how to select a subset of patches from which to synthesize a given subject patch. There are selection methods based on image classification in order to restrict the dictionary to patches that are likely to come from appropriate tissue classes [19]. However, an important goal of the method is to eliminate both the dependency on prior training and the requirement for a classification step prior to synthesis. We have accomplished this using the “trick” of normalizing the patches in a $(d + 1)$-dimensional space rather than the $d$-dimensional space, which we now describe.

All the atlas and subject patches are first globally normalized such that their maximum of their norms is unity. Define

$$m = \max_{i,j} \{ ||a_1(i)||^2, ||b_1(j)||^2 \}.$$ 

Then the patches are scaled as follows

$$\tilde{b}_1(j) = \frac{b_1(j)}{m}, \forall j,$$

$$\tilde{a}_1(i) = \frac{a_1(i)}{m}, \forall i.$$
CHAPTER 4. SPARSE PRIOR BASED CONTRAST SYNTHESIS

This global scaling guarantees that relative intensities are preserved and that all intensities of both subject and atlas fall in the range \([0, 1]\). Now, we project both the subject patch and all atlas patches to a unit sphere in \(\mathbb{R}^{d+1}\) as follows

\[
\begin{align*}
\mathbf{b}'_1(j) &= \left[ \frac{\tilde{\mathbf{b}}_1(j)}{\sqrt{1 - ||\tilde{\mathbf{b}}_1(j)||^2}} \right], \forall j, \\
\mathbf{a}'_1(i) &= \left[ \frac{\tilde{\mathbf{a}}_1(i)}{\sqrt{1 - ||\tilde{\mathbf{a}}_1(i)||^2}} \right], \forall i.
\end{align*}
\]

At this point, both the subject patch and the atlas patches are normalized to unity, satisfying \( ||\mathbf{a}'_1(i)||^2 = 1, \forall i, \) and \( ||\mathbf{b}'_1(j)||^2 = 1, \forall j. \) The normalization criterion in Eqn. 4.12 is therefore automatically satisfied if the \( \mathcal{C}_1 \) contrast dictionary \( A_1 \) is changed to \( A'_1 \in \mathbb{R}^{(d+1)\times N} \) such that the columns of \( A'_1 \) are \( \mathbf{a}'_1(i), i = 1, \ldots, N. \) We therefore rewrite Eqn. 4.12 as

\[
\hat{x}(j) = \arg \min_{\mathbf{x} \geq 0} \left[ ||\mathbf{b}'_1(j) - A'_1\mathbf{x}||^2 + \lambda ||\mathbf{x}||_1 \right], \forall j.
\]

The solution of Eqn. 4.13 yields a nonnegative combination of the columns of \( A'_1 \) that sparsely approximates \( \mathbf{b}'_1(j). \) The resulting reconstructed patch matches both the pattern and intensities within the target patch. Since we do not actually want to synthesize a patch in the higher dimension, we use the solution \( \mathbf{x}(j) \) to synthesize the \( \mathcal{C}_2 \) patch in \( \mathbb{R}^d \) using the first \( d \) components of \( \mathbf{x}(j) \) in Eqn. 4.13. This works correctly because we maintained the association of columns in \( A_1 \) and \( A'_1. \) Once \( \hat{\mathbf{b}}_2(j) \) is found, we use only its central value in the reconstructed image \( \mathbf{b}_2 \) to create the final estimated
4.3.5 Dictionary and Parameter Selection

For all the experiments reported in this and subsequent sections, we used patches of dimension $3 \times 3 \times 3$ voxels. We used three freely available $\ell_1$ solver packages, Sparselab [109], the large-scale $\ell_1$-regularized least-squares ($\ell_1$-ls) [110] and CVX [111], to optimize Eqn. 4.13. From our experience, Sparselab and $\ell_1$-ls are the faster of the three, typically having $\sim 1$ ms run-time per patch, while CVX on average takes $\sim 10$ ms. However, CVX is more robust, managing to producing reasonable results in the cases where both Sparselab and $\ell_1$-ls fail to converge. Our algorithm uses Sparselab for the majority of computations, and for those occasions when it fails to converge, we use $\ell_1$-ls, if it fails also, we use CVX.

A typical $256 \times 256 \times 198$ isotropic MR image, with voxels of size 1 mm$^3$, contains about one million non-background patches. It is computationally intractable to solve Eqn. 4.13 with all such patches included in the dictionary $A'_1$. Recall that for $3 \times 3 \times 3$ patches, our dictionary $A'_1 \in \mathbb{R}^{28 \times N}$, is potentially a $28 \times 1,000,000$ matrix, as $N$ was originally assumed to be indexing over all patches in the atlas image $a_1$. To avoid the computational bottleneck of large $N$ we construct a dictionary for each patch $b'_1(j)$, that is of size $N = 100$. This is done by selecting 100 patches from the global collection of one million patches, thus drastically reducing run time. In keeping with the literature [108], we construct our patch dictionary, $A'_1$, to consist of only those
atlas patches that are close to $b'_1(j)$ in an $\ell_2$ sense. To achieve this in our setting, we sort the atlas patches, $a'_1(i), i = 1, 2, \ldots, N$, by their $\ell_2$ distance to the current patch, $b'_1(j)$. The nearest 100 patches, $a'_1(i), i = 1, 2, \ldots, 100$, are then used to generate the dictionary, $A'_1$, for the current patch, then $\hat{x}(j)$ is found using Eqn. 4.13. To accelerate the search for the 100 nearest patches, we use a k-d tree [112]. We construct $A'_2$ as the corresponding $C_2$ contrast patches for the collection $a'_2(i), i = 1, 2, \ldots, 100$. Our selection of $N = 100$ is based on our experience, having observed that smaller $N$ does not always yield the best candidate patches and larger $N$ generally shows little improvement in the reconstruction result.

The sparsity regularization term, $\lambda$, is an important parameter in our algorithm as it is used for tuning the sparsity of the reconstruction. Through empirical experiments involving hundreds of thousands of patch reconstructions, we have found the algorithm itself to be very stable for $\lambda \in [0, 0.90]$; but the quality of reconstruction varies within this range. To get a quantitative sense of the synthesis, we first experimentally assess how the error in synthesis is affected by the choice of $\lambda$ and choose an optimal $\lambda$ based on experimental validation. Later we show that the $\lambda$ thus found, conforms to a theoretical justification based the equivalence of $\ell_0$ and $\ell_1$ problems using phase transition diagrams.

We define the synthesis error metric $||b_2 - \hat{b}_2||_2$ as the $\ell_2$ intensity difference, per voxel, of the images $b_2$ and $\hat{b}_2$ averaged over all non-zero voxels. In particular, we created an atlas pair $\{a_1, a_2\}$ using patches from one part of a subject’s SPGR ($C_1$) and
CHAPTER 4. SPARSE PRIOR BASED CONTRAST SYNTHESIS

Figure 4.4: Effect of sparsity of $x(j)$ on the $C_2$ contrast: The left most column shows a subject’s SPGR (top) and MPRAGE (bottom) acquisitions. The second, third, and fourth columns of the top row show synthetic MPRAGEs generated using another portion of the subject’s MPRAGE as the atlas. The synthetic MPRAGEs were generated using $\lambda$ values of 0.05, 0.80, and 0.95, respectively. The plot shows the MPRAGE synthesis error vs. the average sparsity of all $x(j)$’s, averaged over all non-zero voxels. The average sparsity scale is on the top of the plot while the sparsity regularization parameter, $\lambda$, is plotted on the bottom axis.

MPRAGE ($C_2$) image, shown in the left most column of Fig. 4.4, and then synthesized MPRAGE images using patches from a different part of the SPGR image. The MPRAGE synthesis error was evaluated over the synthetic image for a collection of
different $\lambda$’s, yielding the plot shown in Fig. 4.4. For this experiment, $\lambda \in [0.05, 0.90]$ yields images with very similar synthesis errors, while $\lambda > 0.90$ gives images with higher errors. Example synthesis results are also shown in Fig. 4.4 for three different values of $\lambda$. We conclude from this experiment that values of $\lambda$ in the range $[0.05, 0.90]$ are reasonable and produce similar error levels. However, we choose $\lambda = 0.80$ in all of our experimental results as it gives the lowest error, as seen from Fig. 4.4.

Figure 4.5: Phase transition diagram of a $A_1$, constructed using patches from a real SPGR image

We now show that this choice conforms with the theoretical consideration that we require a solution to Eqn. 4.6 by solving Eqn. 4.7. Following this notion, we should use a large $\lambda$ from this range because sparser solutions produced by $\ell_1$ minimization are
more likely to agree with the desired $\ell_0$ solution [94–96] as discussed in Sec. 4.1. The limits on the sparsity bound is shown using a phase transition diagram of an $A_1$ matrix that is constructed using a real SPGR image, shown in Fig. 4.4. The phase transition diagram, as shown in Fig. 4.5(a), is a plot of failure probability of an $\ell_1$ problem, such as Eqn. 4.7, to provide the correct solution as obtained from its corresponding $\ell_0$ problem, such as Eqn. 4.6. Asymptotic 0.5 level curves for both unconstrained (Eqn. 4.7) and non-negative constrained (Eqn. 4.12) problems are shown in the lower and upper black dotted lines, respectively, for a random $A$ matrix (for Eqn. 4.6 and Eqn. 4.7 ) whose elements are chosen from $N \sim (0,1)$. These curves are obtained from [96]. The 0.5 level curve for a $A_1$, composed of $N = 100$ patches of dimension $3 \times 3 \times 3$ voxels from a real SPGR image (Fig. 4.4), is shown as the solid black line. Here we note that although the dashed level curves are optimal for a Gaussian matrix, our level curve is generated with patches from real MR images, which violates the i.i.d. nature of the matrix elements. Unlike the method described in [94] to generate the diagram, we vary the number of patches $(N)$ in the dictionary instead of varying the patch size $d$. We choose our operating point on the diagram as $(\delta, \rho)$, $\delta = \frac{d}{N}$ and $\rho = \frac{s}{d}$, such that it is on the 0.5 level curve of $\Phi_1$, satisfying the equivalence of the $\ell_0$ and $\ell_1$ problems, thereby guaranteeing the uniqueness and optimality of the solutions. In this case, if $\lambda = 0.8$, average $s = 5.5$ (from Fig. 4.4), thus the operating point is $\delta = d/N = 0.27$, $\rho = s/d = 0.2$, which is a point on the 0.5 level curve. Thus using the choice of $\lambda$, we keep the average reconstruction error low as well as satisfy
the equivalence of $\ell_0$ and $\ell_1$ problem.

## 4.4 Results

All MR images were first skull-stripped [33, 34], corrected for intensity inhomogeneities using N3 [35], and then normalized using a global linear scaling such that their peak WM intensities (identified using [6]) were at unity.

### 4.4.1 Validation on T1-w Phantom

We first evaluate MIMECS on the Brainweb phantom data [81]. To show the effectiveness of the sparse reconstruction formulation Eqn. 4.13, we choose $a_1$ to be the T1-w phantom with 0% noise and set $a_1 = a_2$. The test image $b_1$ was chosen as the phantom with $n\%$ noise, where $n \in \{0, 1, 3, 5, 7, 9\}$. To evaluate the patch matching process, we defined $\hat{b}_2$ as the $C_2$ contrast of $b_1$, as obtained from Eqn. 4.13, in terms of our patch notation we have $\hat{b}_2(j) = A_2x(j)$. In this experiment, we expected $\hat{b}_2$ to be the same as $a_1$. Accordingly, we chose $||a_1 - \hat{b}_2||_2$ and $||a_1 - b_1||_2$ as the evaluation metrics. Where $||a_1 - b_1||_2$ is measuring the baseline noise present in $b_1$ relative to $a_1$, this noise is characteristic of the phantom itself. Whereas $||a_1 - \hat{b}_2||_2$ denotes the residual noise in $\hat{b}_2$ after reconstructing patches from a noise-free image $a_2 = a_1$.

Fig. 4.6 show the atlas and the test images for $n = 3, 5$ and the corresponding errors in the patch-matching process. Fig. 4.7, on the left, shows a plot of the two
CHAPTER 4. SPARSE PRIOR BASED CONTRAST SYNTHESIS

Atlas $C_1$ and $C_2$

Figure 4.6: The left column shows the Brainweb phantom with 0% noise, used as both $C_1$ and $C_2$ contrast atlas. The next four columns show the subject images, reconstructed images using the atlas, baseline noise and residual noise after reconstruction, respectively, for two noise levels in the two rows.

Error metrics, $||a_1 - \hat{b}_2||_2$ and $||a_1 - b_1||_2$. $||a_1 - \hat{b}_2||_2 = 0$ when $a_1 = b_1$, i.e., when $n = 0$. This is in accordance with the fact that if $A'_1$ contains $b'_1(j)$ as one of its
columns for any patch $b'_1(j)$, then Eqn. 4.13 is minimized with $||x(j)||_0 = 1$. Also, with increasing baseline noise, the residual noise $||a_1 - \hat{b}_2||_2$ is less than the baseline noise $||a_1 - b_1||_2$ because the patches in $\hat{b}_2$ are chosen from the noise free atlas $a_1$ itself. Ideally, when $a_1$ and $b_1$ are the noise-free and noisy versions of the same image, for any particular subject patch $b'_1(j)$, there is exactly one corresponding patch $a'_1(i) \in A'_1$ for some $i$. However, with increased noise in $b_1$, a combination of patches is usually found so that the noise in $b_1$ is reduced by a linear patch combination, i.e., $||b'_1(j) - A'_1x(j)||^2 < ||b'_1(j) - a'_1(i)||^2$. Thus $\hat{b}_2$ resembles $a_1$ more closely than $b_1$ with $||a_1 - \hat{b}_2||_2 < ||a_1 - b_1||_2$, as shown in the two right most columns of Fig. 4.6. Residual noise increases at a much slower rate than the baseline noise in Fig. 4.7, indicating the robustness of the patch matching procedure with respect to noise.

![Figure 4.7: Plots of the baseline and residual noise vs. the noise levels available in the Brainweb phantom. The left plot shows when $C_1$ and $C_2$ atlases are both the same $T_1$-w image. The plot on the right shows when $C_1$ and $C_2$ are $T_1$-w and $T_2$-w images respectively.](image)
CHAPTER 4. SPARSE PRIOR BASED CONTRAST SYNTHESIS

Figure 4.8: The top row shows the $T_1$-w and $T_2$-w phantoms from Brainweb, used as the $C_1$ and $C_2$ contrast atlases. The second and third rows show the corresponding subject images, with their synthetic MIMECS result, along with baseline and residual noise images for Brainweb noise levels of 3% and 5%, respectively.

4.4.2 Validation on T2-w phantom

To validate the contrast synthesis formulation Eqn. 4.10, we used $T_2$-w as the $C_2$ contrast, again using the Brainweb computational phantom. Assuming $a_1$ and $a_2$ to be the $T_1$-w and $T_2$-w phantoms with 0% noise, we synthesized the $T_2$ contrast for $b_1$, which were again chosen to be the $T_1$-w phantoms with $n\%$ noise, $n \in \{0, 1, 3, 5, 7, 9\}$. 
CHAPTER 4. SPARSE PRIOR BASED CONTRAST SYNTHESIS

Fig. 4.8 shows the synthetic T2 contrast images for two noise levels along with the true T2-w images. The first error metric $||a_2 - b_2||_2$ gives the baseline noise present in the actual T2-w test image $b_2$. The second error metric $||a_2 - \hat{b}_2||_2$ gives the difference between the atlas T2-w image and the synthetic image, i.e., the residual noise after reconstruction. Fig. 4.7, on the right, shows the plot of each of these two metrics. Clearly, with increasing input noise, the synthetic $T_2$-w contrast image $\hat{b}_2$ has less noise, which demonstrates a robustness of the alternate contrast reconstruction to noise. Also, when $a_1 = b_1$, i.e., with $n = 0$, the reconstruction error is zero and so is the residual noise, providing a sanity check for Eqn. 4.13. As can be seen from the last column of Fig. 4.8, erroneous estimation of $\mathbf{x}(j)$s occur primarily at the CSF-background boundary of the images. We believe this happens because of the presence of background values of zero in those boundary patches, for which the $\ell_1$ estimation is deficient.

We note that a registration based method may perform better than MIMECS on this test data. Ideally, $b_1$ for $n = 9$ can be perfectly registered to the atlas $a_1$ because they are the same phantom with different noise. If $b_1$ is registered to $a_1$ to obtain an estimate $\hat{b}_1$, then $\hat{b}_1 = b_1$. Thus the transformation, which is the identity in this scenario, can be applied to $a_2$ to get $\hat{b}_2 = b_2$. Then the residual noise $||a_2 - \hat{b}_2||_2 = 0$, although the baseline noise $||a_1 - b_1||_2 > 0$. In real data, however, this is not the case as a perfect registration is seldom achieved, as shown in Fig. 4.3. We will show more examples in Sec. 4.5 to demonstrate the superior performance of MIMECS for
contrast synthesis in comparison to a registration based method.

### 4.4.3 Test-Retest consistency

We used the freely available Kirby21 data [113] to test the reproducibility of our method. This dataset contains 21 normal subjects, each having two MPRAGE scans from a Philips scanner, taken 25 minutes apart. Each volume is of size $256 \times 256 \times 170$ with $1.0 \times 1.0 \times 1.2$ mm resolution. For each subject, we used one scan as both the atlas $a_1$ and $a_2$ images, and normalized the second scan intensities to the atlas scan. We repeated this process after switching the atlas and scan images. Thus, in this experiment we are testing normalization between real images with different noise levels, scanned in the same scanner with the same pulse sequence.

The top row of Fig. 4.9 shows the original scans, referred to as #1 and #2 respectively, and their difference image. The middle row shows the synthetic version of #1 using #2 as the atlas, denoted as #1(S). With #2(S) denoting the synthetic version of #2 using #1 as the atlas, which is shown on the bottom row. The difference images between the synthetic images and the original images, shown in both the middle and both row, closely resemble the difference between the original scans. The small differences between the synthetic and atlas images, indicate the quality of the normalization achieved by MIMECS. We note that non-zero differences occur for two reasons, the underlying noise is different in the two original images, and the convergence limit of the sparse reconstruction algorithm.
CHAPTER 4. SPARSE PRIOR BASED CONTRAST SYNTHESIS

Figure 4.9: The top row shows an example of a repeat scan from the Kirby21 dataset and the corresponding difference image. The middle row shows the synthetic version of scan #1 using scan #2 as both $C_1$ and $C_2$ atlas contrasts, as well as the difference image between the synthetic result and both of scans #1 and #2, respectively. The bottom row shows another synthetic result, this time synthesizing scan #2 with scan #1 as the $C_1$ and $C_2$ atlas contrasts. The scale is for the difference images.
Figure 4.10: The top row shows two example images of repeat scans from the Kirby21 dataset, as well as their corresponding synthetic images. The bottom row shows the hard segmentations of the original images and their synthetic counterpart.

To show the effect of normalization on image segmentation applied to the processed images, the Dice similarity coefficient [114] between the hard segmentations (produced in each case using the TOADS method described in [9]) of the original and synthetic images are reported in Table 4.1. The segmentations are shown in Fig. 4.10. The table shows that there is no significant difference between the Dice coefficients of the segmentations of the synthetic and original images. A paired $t$-test with a null hypothesis that the Dice between the synthetic scans is different from that between the original scans, gave a $p$-value < 0.05 thus the images are statistically indistinguishable. Therefore, assuming that one scan is a noisy version of the other or vice
versa, MIMECS does not change their segmentations significantly while normalizing.

Table 4.1: Dice coefficients (mean ± standard deviation) between the segmentations of the original and synthetic images

<table>
<thead>
<tr>
<th></th>
<th>#1 vs. #2</th>
<th>#1 vs. #2(S)</th>
<th>#2 vs. #1(S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSF</td>
<td>0.9271 ± 0.026</td>
<td>0.9301 ± 0.013</td>
<td>0.9242 ± 0.022</td>
</tr>
<tr>
<td>Ventricle</td>
<td>0.9344 ± 0.016</td>
<td>0.9387 ± 0.031</td>
<td>0.9290 ± 0.022</td>
</tr>
<tr>
<td>GM</td>
<td>0.9404 ± 0.022</td>
<td>0.9389 ± 0.029</td>
<td>0.9422 ± 0.020</td>
</tr>
<tr>
<td>Basal Ganglia</td>
<td>0.8900 ± 0.039</td>
<td>0.9000 ± 0.029</td>
<td>0.8939 ± 0.032</td>
</tr>
<tr>
<td>Thalamus</td>
<td>0.8956 ± 0.036</td>
<td>0.8899 ± 0.042</td>
<td>0.8912 ± 0.039</td>
</tr>
<tr>
<td>WM</td>
<td>0.9389 ± 0.019</td>
<td>0.9360 ± 0.017</td>
<td>0.9390 ± 0.020</td>
</tr>
</tbody>
</table>

Table 4.2: Mean KL distances between histograms of the original/synthetic images of BIRN data

<table>
<thead>
<tr>
<th></th>
<th>Philips 1.5T</th>
<th>GE 4T</th>
<th>Siemens 3T</th>
<th>Siemens 3T #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Philips 1.5T</td>
<td>—</td>
<td>0.28 / 0.04</td>
<td>0.47 / 0.07</td>
<td>0.67 / 0.09</td>
</tr>
<tr>
<td>GE 4T</td>
<td>0.72 / 0.04</td>
<td>—</td>
<td>1.22 / 0.02</td>
<td>0.11 / 0.06</td>
</tr>
<tr>
<td>Siemens 3T</td>
<td>0.17 / 0.04</td>
<td>0.28 / 0.03</td>
<td>—</td>
<td>0.88 / 0.06</td>
</tr>
<tr>
<td>Siemens 3T #2</td>
<td>0.41 / 0.05</td>
<td>0.26 / 0.06</td>
<td>0.16 / 0.03</td>
<td>—</td>
</tr>
</tbody>
</table>
CHAPTER 4. SPARSE PRIOR BASED CONTRAST SYNTHESIS

4.4.4 Intensity Normalization

MR intensity distributions vary widely across scanners, as shown in Fig. 4.11, where the same subject was scanned in four different scanners with different field strengths and different acquisition parameters. The histograms (Fig. 4.11 bottom row) of the inhomogeneity corrected images show substantial variations in the intensity distributions. As the observed MR intensities are not associated with any physical meaning, the difference in intensity distributions gives rise to inconsistencies in the segmentation, previously reported for both normal [18,20] and diseased subjects [115]. In this experiment, MIMECS was used to normalize images obtained from different scanners. We used two multi-scanner datasets: 1) the traveling subject data from the BIRN study [3] and 2) the morphometry BIRN (MBIRN) calibration study.

The traveling subject data contains five normal subjects, each of them having five SPGR scans (one each from a GE 1.5T, GE 4T, Philips 1.5T and two Siemens 3T scanners) and two MPRAGE scans (from a GE 3T and a Siemens 1.5T). Each scan was first skull-stripped using the mask generated by skull-stripping that subject’s Siemens 1.5T MPRAGE scan [33,34]. We used MIMECS to normalize all the SPGR scans of every subject to an MPRAGE contrast by synthesizing them using a pair of SPGR-MPRAGE images of that subject as the atlas. The normalization strategy is consistent with the goal of carrying out consistent segmentations, and from our experience better segmentations are usually achieved using the MPRAGE contrast.
CHAPTER 4. SPARSE PRIOR BASED CONTRAST SYNTHESIS

For every subject, the atlases $a_1$ and $a_2$ were chosen to be the GE 1.5T SPGR and the corresponding Siemens 1.5T MPRAGE scan, respectively.

We normalized each of the remaining four SPGR scans of every subject to the corresponding MPRAGE contrast using $a_1$ and $a_2$. The KL distances [73] between the histograms of the original scans and the normalized scans are reported in Table 4.2. The distances are averaged over five subjects and each subject has five SPGR scans from five different scanners. In all the cases, the KL distances between synthetic images are significantly less than the original images. The atlases ($a_1$ and $a_2$), the original scans, and the synthetic images of one subject along with the corresponding histograms are shown in Fig. 4.11. Both qualitatively and quantitatively, both the synthetic images and their histograms are closer to one another as compared to the original scans.

Next, MIMECS was used as a normalization technique on the MBIRN calibration study. The dataset contains five subjects, each of them having four SPGR scans from each of four scanners: two GE 1.5T scanners, one Siemens 1.5T scanner, and one Marconi 1.5T scanner. For each scanner, two scans were acquired with 20° flip angle and the other two with 30° flip angle. Each of the 16 scans of every subject was stripped with the same mask and corrected by N3 for intensity inhomogeneities. To normalize all the scans with 20° flip angles to those with 30° flip angles, we chose a subject specific atlas pair ($a_1$ and $a_2$) as those acquired from the Siemens 1.5T scanner with 20° and 30° flip angles, respectively. These are shown in the top
CHAPTER 4. SPARSE PRIOR BASED CONTRAST SYNTHESIS

Figure 4.11: Intensity normalization result on BIRN data
CHAPTER 4. SPARSE PRIOR BASED CONTRAST SYNTHESIS

$C_1$ CONTRAST $C_2$ CONTRAST

Figure 4.12: Intensity normalization on MBIRN data: The top row shows $C_1$ and $C_2$ atlas images with 20° and 30° flip angles, respectively. The second row shows four other subjects with 20°$^\circ$irc flip angles and the corresponding synthetic result with 30°$^\circ$irc flip angles, generated by MIMECS, is shown in the third row.

row of Fig. 4.12 for one subject. The repeated scans of the same subject with 20°$^\circ$irc flip angles from the two GE 1.5T scanners are shown in the second row of Fig. 4.12
and their corresponding synthetic images are shown in the bottom row of Fig. 4.12.

Average KL distances between the histograms of the original images and the synthetic ones are reported in Table 4.3. The distances are averaged over five subjects, each having eight SPGR acquisitions of $20^\circ$ flip angles. In all the cases, the distances between the synthetic images are significantly lower than that between the original images. The data reveals substantial improvement in intensity normalization. The small difference and the lower KL-distances from Table 4.3 indicate the good normalization capability of MIMECS with respect to variations of pulse sequence parameters.

Table 4.3: Mean KL distances between histograms of the original/synthetic images for MBIRN data

<table>
<thead>
<tr>
<th></th>
<th>GE 1.5T #1</th>
<th>GE 1.5T #2</th>
<th>Siemens 1.5T</th>
<th>Marconi 1.5T</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE 1.5T #1</td>
<td>—</td>
<td>0.111 / 0.011</td>
<td>0.166 / 0.030</td>
<td>0.194 / 0.021</td>
</tr>
<tr>
<td>GE 1.5T #2</td>
<td>0.080 / 0.017</td>
<td>—</td>
<td>0.202 / 0.018</td>
<td>0.210 / 0.034</td>
</tr>
<tr>
<td>Siemens 1.5T</td>
<td>0.172 / 0.011</td>
<td>0.128 / 0.005</td>
<td>—</td>
<td>0.108 / 0.024</td>
</tr>
<tr>
<td>Marconi 1.5T</td>
<td>0.156 / 0.017</td>
<td>0.212 / 0.011</td>
<td>0.143 / 0.033</td>
<td>—</td>
</tr>
</tbody>
</table>

4.4.5 Consistent Segmentation

In this section, we show that normalization leads to more consistent segmentation across scanners. Because MR image intensities do not possess any specific physical
Table 4.4: Dice coefficients (mean ± standard deviation) between segmentations of original scans, registered and transformed versions, and MIMECS synthesis for BIRN data.

<table>
<thead>
<tr>
<th></th>
<th>GE 4T</th>
<th>Philips 1.5T</th>
<th>Siemens 3T</th>
<th>Siemens 1.5T</th>
<th>GE 3T</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Original</strong></td>
<td>0.81 ± 0.04</td>
<td>0.75 ± 0.03</td>
<td>0.86 ± 0.03</td>
<td>0.82 ± 0.02</td>
<td>0.79 ± 0.03</td>
</tr>
<tr>
<td><strong>Registration</strong></td>
<td>0.78 ± 0.04</td>
<td>0.81 ± 0.03</td>
<td>0.87 ± 0.02</td>
<td>0.80 ± 0.01</td>
<td>0.78 ± 0.03</td>
</tr>
<tr>
<td><strong>MIMECS</strong></td>
<td>0.84 ± 0.03</td>
<td>0.83 ± 0.02 *</td>
<td>0.89 ± 0.02 *</td>
<td>0.85 ± 0.01 *</td>
<td>0.84 ± 0.01 *</td>
</tr>
<tr>
<td><strong>Philips 1.5T</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original</td>
<td>0.70 ± 0.04</td>
<td>0.79 ± 0.02</td>
<td>0.76 ± 0.02</td>
<td>0.73 ± 0.01</td>
<td></td>
</tr>
<tr>
<td>Registration</td>
<td>0.71 ± 0.03</td>
<td>0.77 ± 0.03</td>
<td>0.77 ± 0.04</td>
<td>0.71 ± 0.03</td>
<td></td>
</tr>
<tr>
<td><strong>MIMECS</strong></td>
<td>0.79 ± 0.03 *</td>
<td>0.84 ± 0.03 *</td>
<td>0.82 ± 0.01 *</td>
<td>0.78 ± 0.02 *</td>
<td></td>
</tr>
<tr>
<td><strong>Siemens 3T</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original</td>
<td>0.78 ± 0.02</td>
<td>0.74 ± 0.02</td>
<td>0.79 ± 0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Registration</td>
<td>0.83 ± 0.02</td>
<td>0.75 ± 0.03</td>
<td>0.80 ± 0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MIMECS</strong></td>
<td>0.84 ± 0.01</td>
<td>0.84 ± 0.02 *</td>
<td>0.83 ± 0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Siemens 1.5T</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original</td>
<td>0.79 ± 0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Registration</td>
<td>0.80 ± 0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MIMECS</strong></td>
<td>0.83 ± 0.01 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Statistically significantly larger than the original Dice coefficient (with a p-value < 0.05).
<table>
<thead>
<tr>
<th>ORIGINAL IMAGES</th>
<th>GE 1.5T</th>
<th>GE 4T</th>
<th>PHILIPS 1.5T</th>
<th>SIEMENS 1.5T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segmentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIMECS Result</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segmentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.13: Segmentation result on BIRN data: The top two rows show the original scans of one subject from four different scanners and their corresponding segmentations. The bottom two rows show the MIMECS synthetic versions and their segmentations.
meaning, the contrast of the tissues are highly variable with respect to the scanners and pulse sequences. Most image processing techniques are primarily dependent on image intensities, and therefore tend to produce inconsistent results between scanners and pulse sequences. We applied MIMECS on the BIRN traveling subjects in order to synthesize normalized MPRAGE images as the $C_2$ contrast. The SPGR and MPRAGE atlases, $a_1$ and $a_2$ respectively, and the normalization process are described in the previous section. We compared MIMECS with a registration based contrast synthesis method. Each SPGR scan of a subject was first deformably registered to the $C_1$ contrast atlas $a_1$ using a deformable registration algorithm SyN [5]. Then the deformation was applied to the MPRAGE atlas $a_2$ to obtain the MPRAGE contrast. For MPRAGE scans of a subject, $a_1$ and $a_2$ were both MPRAGE images. An example of such a registration is shown in Fig. 4.3. The Dice coefficients between the segmentations of the original scans, their registered and transformed contrasts, and their synthetic versions are reported in Table 4.4.

To quantify the improvement in segmentation consistency, we use the following null hypothesis, the Dice coefficient between the segmentations of the synthetic images obtained from two different scanners is less than that obtained from the segmentations of the original images. One-sided $t$-tests give $p$-values less than 0.05 in 11 out of 15 pairs of comparisons in Table 4.4, indicating significant improvement in those cases. A similar hypothesis comparing synthetic images with the registered and transformed images demonstrating significant improvements ($p$-values < 0.05), again in 11 out of
CHAPTER 4. SPARSE PRIOR BASED CONTRAST SYNTHESIS

15 cases. In the other 4 cases, MIMECS does not improve the segmentation consistency significantly in comparison to the original or the registration approach. This experiment indicates that synthesizing contrasts improves segmentation consistency. In some cases, registered and transformed images reduces the segmentation consistency compared to the original scans. In these cases, further examination showed that the deformable registration had performed poorly.

4.5 Robustness to Pathologies

From our experiments, we have found that MIMECS is robust to mild pathologies in a subject, even if the atlas does not possess any pathology. To show the robustness of MIMECS with respect to pathologies, we include a demonstration with two OASIS [116] subjects, one with very large ventricles (see Fig. 4.14 Subject #1) and one with severe lesions (see Fig. 4.14 Subject #2). If we choose a normal atlas—normal ventricles and no lesions—(Fig. 4.14 top row), a deformable registration does not perform well, as shown in the third row of Fig. 4.14. Even when an atlas may have lesions, registration will depend on the location of the lesions. However, with MIMECS, the lesion contrast and ventricle boundary are synthesized well, as shown in the bottom row of Fig. 4.14.
Figure 4.14: The top row shows an MPRAGE scan of a normal subject, used as both $C_1$ and $C_2$ contrast atlases in MIMECS and also used for the registration based synthesis. The second, third, and fourth rows show two OASIS subjects in each column, the atlas image registered and transferred to be in the subject space and the MIMECS result.
CHAPTER 4. SPARSE PRIOR BASED CONTRAST SYNTHESIS

4.6 Discussion

Histogram matching has been a popular approach for correcting intensity variations due to scanners [117] or pulse sequences [4]. Histogram matching tries to find the one-to-one correspondence between two image intensity values based on the cumulative distributions of their intensity distributions. Of course this one-to-one mapping can be erroneous if the histogram bins are not chosen carefully. Additionally, high intensity outliers can bias the matching process. For a robust piecewise transformation of histograms, landmarks can be carefully picked or selected in a semi-automated fashion [2]. MIMECS does not have these drawbacks, as there is no single one-to-one intensity mapping—which we believe is a major benefit of this approach. It is robust to outliers as well since the $\ell_1$ minimization process is carried out with unit-normalized patches, and it is fully automated with no need for landmark selection.

MIMECS is also shown to be robust to noise, see the Brainweb validation experiments (Sec. 4.4.1 and Sec. 4.4.2) and the test-retest consistency experiment (Sec. 4.4.3). This can be attributed to the fact that the sparsity of any of the $x(j)$ coefficients is greater than one, which has the consequence that several atlas patches are averaged together to reconstruct the alternate contrast patch. Additionally, for any patch $b_1(j)$, its $A_1$ (or $A'_1$) consists of similar looking patches, following a non-local means type criterion [108], which also provides for noise-robustness.

The success of patch-matching to a multi-contrast atlas is based on two assumptions. First, the contrast variation between the subject and the atlas $C_1$ contrast
images is not too large such that a good linear approximation can always be found from a rich dictionary. Second, the $C_2$ contrast image has a locally monotonic relationship with the $C_1$ contrast image, as inherently assumed in Eqn. 4.9. Comparison of a $T_1$-w SPGR and a fluid attenuated inversion recovery (FLAIR) image with lesions provide an example where this assumption is violated. The lesions have a GM-like contrast in $T_1$-w SPGR, while it is bright in FLAIR images. Thus a current limitation of MIMECS is that a local monotonic tissue contrast must be present in the atlas to synthesize it in the subject.
Chapter 5

A Generative Model for Image Normalization and Synthesis

In the previous chapter, we described a sparse prior approach that uses patches from a multi-contrast atlas to synthesize different contrasts for a subject. There are two drawbacks of the previous approach. First, Eqn. 4.10 implies that the $C_2$ contrast patches are combined using the same weight that are obtained from the matching of the $C_1$ contrast patches. Thus the relation between $C_1$ and $C_2$ contrast images is not utilized, although evidently they are not independent. Second, the assumption in Eqn. 4.11 does not have an underlying generative model; thus, it is not amenable to a probabilistic interpretation. In this chapter, we propose two generative frameworks for the normalization and the synthesis process, leading to methods that are quite different in nature to the method described in Chapter 4. The proposed framework
also uses patches from a multi-contrast atlas as before. While synthesizing, however, the new framework successfully utilizes informations from both the $C_1$ and $C_2$ contrast atlases in an optimal way to generate new patches.

This chapter is organized as follows. First, we provide a brief introduction to image normalization methods and describe necessary assumptions. Then we describe in Sec. 5.1 the generative model for the intensity normalization between images acquired in different scanners or with different pulse sequences parameters (e.g., different repetition times or flip angles) but with same imaging protocol (e.g., SPGR). We show that our normalization produces better consistency in segmentation between images than popular histogram or landmark based methods. Then we describe in Sec. 5.2 the model for image synthesis and show the improvement in segmentation consistency between images acquired with different scanners or with different pulse sequences (e.g., from SPGR vs. MPRAGE).

5.1 Bayesian Intensity Normalization

5.1.1 Previous Works

Several intensity standardization algorithms have been proposed to bring MR intensities to a common scale. Histogram matching algorithms try to match one subject histogram to a reference histogram by matching the intensities of manually or automatically chosen landmarks [2] or by minimizing some information-theoretic
CHAPTER 5. GENERATIVE MODEL FOR IMAGE NORMALIZATION AND SYNTHESIS

criteria [118,119] between the histograms. Another method uses intrinsic MR properties, such as longitudinal and transverse relaxation times ($T_1$ and $T_2$) or proton density ($P_D$) to normalize the subject intensities to a reference volume, acquired with a different pulse sequence [60]. A segmentation based technique has also been proposed to account for the scanner variations by matching intensity distributions of similar tissues between two images [58]. Landmark based methods usually suffer from the fact that reliable landmarks are sometimes difficult to find as they are either manually chosen or based on a segmentation itself. Using tissue properties can be difficult because one needs precise knowledge of the imaging equations pulse sequence parameters, which are often not known.

In Chapter 4, we have described a patch-based intensity normalization method to overcome these problems. In this section, a generative model based intensity normalization technique which has strong parallels with the idea of coherent point drift [29], is presented. We still use patches instead of intensities of a single voxel as features, because a patch around a voxel contains neighborhood information about the center voxel. We propose a patch matching technique that takes patches from the subject image and finds its best matching patches in a target image. We benefit from not having to choose landmarks and no pulse sequence parameters are required.
CHAPTER 5. GENERATIVE MODEL FOR IMAGE NORMALIZATION AND SYNTHESIS

5.1.2 Motivation

The subject and atlas patches represent a pattern of intensities that arise from same pulse sequences (e.g., SPGR) and are scaled to a similar intensity range. Therefore, an atlas patch that has a pattern of intensities that is similar to a given subject patch might arise from the same distribution of tissues. In that case, a patch in the atlas can be expected to represent an approximate distribution of the subject in that patch. This is the principle of patch-based contrast normalization. One could find a patch within the atlas that is close (or closest) to the subject patch and then use the atlas patch directly as a replacement for the subject patch.

We propose a generative method that specifically relates subject patches to atlas patches. We postulate that subject patches are random vectors whose probability densities are Gaussian with means given by an unknown atlas patch and with unknown covariance matrix. This framework captures the notion that an atlas patch can be used to describe a subject patch and that the subject patch might vary somewhat (Gaussian deviation) from the selected atlas patch.

5.1.3 Method

We want to normalize the intensities of a normal subject $S$ (i.e., no disease) to a normal atlas $A$. We assume that $S$ and $A$ are acquired with similar imaging sequences, e.g., either SPGR or MPRAGE, but differ in either or both the scanner or
CHAPTER 5. GENERATIVE MODEL FOR IMAGE NORMALIZATION AND SYNTHESIS

the imaging parameters (e.g., repetition time, echo time, or flip angle). At each voxel of an image, 3D patches are considered and stacked into 1D vectors of size $d \times 1$. There are $N$ subject patches $x_i$ and $M$ atlas patches $y_k$, where $N$ and $M$ are the number of non-zero image voxels in $S$ and $A$, respectively. Define $X = \{x_i\}$ and $Y = \{y_j\}$ to be the collection of subject and atlas patches. To match the subject intensities to the atlas, we match $X$ to $Y$. First, we make sure that the peak white matter (WM) intensities of both $S$ and $A$ are the same (which provides a rough normalization of the two data sets). We assume that each subject patch $x_i$ is a realization of a Gaussian random vector whose mean is one of the atlas patches–i.e., $x_i \sim \mathcal{N}(y_j, \Sigma_j)$ for some $j$. Then both the best matching atlas patch for each subject patch and the covariances of all atlas patches are defined by maximum likelihood and found using the expectation maximization algorithm (EM) [36], as described in detail below. The normalized image is produced by replacing the center pixel of each observed patch by the corresponding value of the matching atlas patch.

5.1.3.1 Algorithm

To find the correspondence between patches, let $z_{ij}$ be an indicator function having the value one when the $i^{th}$ subject patch $x_i$ originates from a Gaussian distribution having its mean as the $j^{th}$ atlas patch $y_j$ with covariance matrix $\Sigma_j$, and is zero otherwise. The probability of observing $x_i$ is written as

$$P(x_i|z_{ij} = 1, \Sigma_j, y_j) = \frac{1}{\sqrt{2\pi|\Sigma_j|^{1/2}}} \exp \left\{ -\frac{1}{2}(x_i - y_j)^T \Sigma_j (x_i - y_j) \right\}. \quad (5.1)$$
Here we make the assumption that $\Sigma_j$ is a diagonal matrix—i.e., $\Sigma_j = \sigma_j^2 I$, where $I$ is a $d \times d$ identity matrix. Although the voxels in a patch need not be of the same tissue class, we will model the covariance as $\sigma_j^2 I$, assuming that their perturbation around the different means (i.e., the atlas patches $y_j$) is the same. However, this strict assumption can be relaxed by assuming full-rank or a diagonal matrix for $\Sigma_j$, which will lead to estimation of more parameters. However, for computational simplicity, in all of our analysis, we use $\Sigma_j = \sigma_j^2 I$. Assuming the i.i.d. nature of the patches and a uniform prior probability $P(z_{ij} = 1|\Sigma_j, y_j) = \frac{1}{M}$, the joint probability distribution of all the subject patches is given by

$$P(X, Z|Y, \Theta) = C \prod_{j=1}^{M} \prod_{i=1}^{N} \left( \frac{1}{\sigma_j} \exp \left\{ -\frac{||x_i - y_j||^2}{2\sigma_j^2} \right\} \right)^{z_{ij}}$$ \hspace{1cm} (5.2)

where $Z = \{z_{ij}\}$, $\Theta = \{\sigma_j : j = 1, \ldots, M\}$, and $C$ is a normalizing constant. An estimate of $\sigma_j$'s and posterior probabilities $P(z_{ij} = 1|X, Y, \Theta)$ can be obtained by maximizing the joint probability using EM. See Eqn. 2.6 – Eqn. 2.7 for the details of EM algorithm.

Assuming $w_{ij} = E(z_{ij}|x_i, y_j, \sigma_j)$, the E-step gives

$$w_{ij} = P(z_{ij} = 1|x_i, y_j, \sigma_j),$$

$$= \frac{P(x_i|z_{ij} = 1, y_j, \sigma_j)P(z_{ij} = 1|y_j, \sigma_j)}{\sum_k P(x_i|z_{ik} = 1, y_k, \sigma_k)P(z_{ij}|y_j, \sigma_j)},$$

$$= \frac{1}{\sigma_j} \phi \left( \frac{x_i - y_j}{\sigma_j} \right), \quad \phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}||t||^2}. \hspace{1cm} (5.3)$$

The M-step gives the update equations for $\sigma_j$'s as
CHAPTER 5. GENERATIVE MODEL FOR IMAGE NORMALIZATION AND SYNTHESIS

\[ \sigma_j^2 = \frac{\sum_{i=1}^{N} w_{ij} ||x_i - y_j||^2}{\sum_{i=1}^{N} w_{ij}}. \] (5.4)

The algorithm is deemed to have converged when the maximum difference between the posteriors in successive iterations is below a threshold. Then the central voxel of the subject patch \(x_i\) is replaced with the central voxel of \(y_t\), where \(t = \arg \max_j (w_{ij})\).

Here we mention that instead of using uniform priors, it is also possible to use data priors, such that

\[ P(z_{ij}|\Sigma_j, y_j) = f_y, \]

where \(f_y\) is the fraction of voxels having intensity \(\in (y - \Delta, y + \Delta)\), where \(y\) is the intensity value of the middle voxel of the patch \(y_j\), and \(\Delta\) is some small number. \(f_y\) can be taken as a smooth version of the image histogram also. Eqn. 5.3 – Eqn. 5.4 can be rewritten to include \(f_y\) also.

For a 256 \(\times\) 256 \(\times\) 199 MR brain image of resolution 1mm\(^3\), \(M\) and \(N\) are typically \(\approx 10^7\). It is therefore memory and time intensive to compute \(w_{ij}\)’s for every \(x_i\)’s. We observed experimentally that \(w_{ij}\) is very close to zero if \(||x_i - y_j||\) is large. Following the assumption that \(\Theta_S(i)\) and \(\Theta_A(j)\) should be the same tissue, we assume \(w_{ij} = 0\) if \(||x_i - y_j|| > \delta\), which is similar to a non-local criteria [108]. However, in our computation, for each \(i\), we choose its closest \(D\) atlas patches to have non-zero \(w_{ij}\), their indices denoted by \(\Omega_i\), a set of cardinality \(D\). Thus if any atlas index \(j\) is not in \(\Omega_i\), we fix \(w_{ij} = 0\). Eqn. 5.4 is modified as follows

\[ \sigma_j^2 = \frac{\sum_{i,j \in \Omega_i} w_{ij} ||x_i - y_j||^2}{\sum_{i,j \in \Omega_i} w_{ij}}. \] (5.5)
CHAPTER 5. GENERATIVE MODEL FOR IMAGE NORMALIZATION AND SYNTHESIS

\[ \alpha_S = 30^\circ \]

\[ \alpha_S = 45^\circ \]

\[ \alpha_S = 60^\circ \]

\[ \alpha_S = 75^\circ \]

Figure 5.1: Mean squared errors (MSE) between atlas (\(\alpha_A = 90^\circ\)) and subject images (\(\alpha_S = 30^\circ, 45^\circ, 60^\circ, 75^\circ\)) are shown for various noise levels.
5.1.4 Results

5.1.4.1 Phantom validation

We first validated the algorithm on the Brainweb phantom data [81]. We used flip angle $\alpha$ as the varying imaging parameter of SPGR scans of a normal phantom. Phantoms with $\alpha_S = 30^\circ, 45^\circ, 60^\circ, 75^\circ$ with noise levels $n = 0, 1, 3, 5\%$ were used to normalize to a phantom with $\alpha_A = 90^\circ$ and 0\% noise. Our method is compared with histogram matching and a landmark based method [2] where the landmarks are found using a Gaussian mixture model algorithm. Fig. 5.1 shows the mean squared errors (MSE) between the atlas and the subjects before and after normalization with these three methods. Clearly the patch based method outperforms the other two for all values of $\alpha_S$ and $n = 1, 3, 5$. For 0\% noise, all three methods perform similarly, because the lack of any partial volume or noise makes the choice of landmarks accurate and histogram matching perfect. At higher noise levels, histogram matching becomes dependent on the number of bins and the estimation of landmarks becomes less robust. The atlas and a subject with $\alpha_S = 30^\circ$ at 5\% noise are shown in Fig. 5.2 (top row), and the normalized images and their differences from the atlas are shown in the middle and bottom row, respectively. Clearly, both histogram matching and landmark based normalization leaves a bias in intensities in large regions, e.g., gray matter (GM), while our patch-based method fails only in recovering sharp edges such as at GM to WM transitions.
CHAPTER 5. GENERATIVE MODEL FOR IMAGE NORMALIZATION AND SYNTHESIS

ATLAS ($\alpha_A = 90^\circ$)          SUBJECT ($\alpha_S = 30^\circ$)

Figure 5.2: Top row shows the atlas ($\alpha_A = 90^\circ$) and the subject ($\alpha_S = 30^\circ$). Middle row shows the histogram matched, landmark based normalized and patch based normalized images. Bottom row shows their differences with the atlas.
CHAPTER 5. GENERATIVE MODEL FOR IMAGE NORMALIZATION AND SYNTHESIS

Figure 5.3: Top row shows the subject ($\alpha = 20^\circ$), atlas ($\alpha = 30^\circ$) and the corresponding normalized image. Bottom row shows a plot of the histograms of the above three images.
5.1.4.2 Experiments on real data

We next used five healthy subjects from the morphometry BIRN [3] calibration study to show the effects of normalization. Each subject has 14 SPGR scans, 7 each with $\alpha = 20^\circ$ and $30^\circ$ from 4 different scanners. Each scan of a subject was first rigidly registered to its GE 1.5T $\alpha = 30^\circ$ scan and then skull-stripped [34] using the same mask generated from the GE 1.5T $\alpha = 30^\circ$ scan. Then each image was corrected for intensity inhomogeneities using N3 [35]. For every subject, we normalized every $\alpha = 20^\circ$ scan to the corresponding $\alpha = 30^\circ$ scan. Fig. 5.3 shows one set of $\alpha = 20^\circ$ and $\alpha = 30^\circ$ scans of a subject along with the normalized image in the top row.

Our method does not seek to match histograms, however it is useful to compare them after normalization. Fig. 5.3 bottom row, shows the histogram of the normalized image more strongly resembles the atlas than the original subject image.

Table 5.1: Dice coefficients of hard segmentations obtained of images before and after normalization, comparing our method with histogram matching and a landmark based matching.

<table>
<thead>
<tr>
<th></th>
<th>CSF</th>
<th>GM</th>
<th>WM</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>0.882</td>
<td>0.788</td>
<td>0.927</td>
<td>0.884</td>
</tr>
<tr>
<td>Hist. Match</td>
<td>0.881</td>
<td>0.828</td>
<td>0.917</td>
<td>0.891</td>
</tr>
<tr>
<td>Landmark based</td>
<td>0.876</td>
<td>0.776</td>
<td>0.926</td>
<td>0.882</td>
</tr>
<tr>
<td>Patch based</td>
<td>0.874</td>
<td>0.845*</td>
<td>0.943*</td>
<td>0.902</td>
</tr>
</tbody>
</table>

*Statistically significantly larger than the other three ($p$-value < 0.05).
CHAPTER 5. GENERATIVE MODEL FOR IMAGE NORMALIZATION AND SYNTHESIS

As described earlier, segmentations can become inconsistent with the variation of scanners and imaging parameters [18, 58, 60]. We now demonstrate that the patch based normalization achieves more consistent segmentation between images acquired with different scanners and pulse sequence parameters. We compare our method with histogram matching and a landmark based intensity transformation [2] where three landmarks on the subject and the atlas are automatically chosen based on a 3-class Gaussian mixture model segmentation. After normalization, we used fuzzy C-means [7] to generate segmentations. We report Dice similarity coefficients in Table 5.1 between segmentations of the three primary tissues, CSF, GM, and WM, and their weighted average. The values are averaged over five subjects, each with seven scans. Patch based normalization produces significantly higher Dice coefficients for GM and WM compared to the other two, while the CSF Dice segmentations are comparable. Histogram matching often decreases the similarity with the original, as it is heavily dependent on the number of bins. Landmark based matching can also deteriorate the segmentation consistency, when a suitable landmark is not found. This is often the case if the image histogram is not multi-modal (e.g., the blue line in Fig. 5.1). Both Figs. 5.2 and 5.3 reveal a smoothing effect on the reconstructed images. Though not unexpected in patch-based methods, this effect needs further investigation.
5.1.5 Discussion and Conclusion

We have described a patch based MR intensity normalization framework that can normalize scans having similar acquisition protocol (e.g., SPGR in the above experiments) but acquired on different scanners or with different acquisition parameters (e.g., different flip angles). We validated our algorithm on phantoms and real scans and showed it gives rise to more consistent segmentations after normalization.

5.2 Bayesian Image Synthesis

In this section, we describe a Bayesian framework for tissue contrast synthesis and compare it to the sparse prior based technique described in Chapter 4. Previous approaches to MR intensity standardization as well as synthesis have depended on the underlying tissue properties—proton density ($P_D$) and relaxation times ($T_1$, $T_2$, and $T_2^*$)—as it is possible to estimate these at every voxel and apply imaging equation of the unobserved pulse sequence to obtain the desired contrast [60]. However, these methods require multiple acquisitions as well as some knowledge of the imaging equations and other imaging parameters, which are often not known or difficult to estimate. Histogram based solutions try to match the histogram of a subject to an atlas histogram with the desired contrast by matching intensities of landmarks [2] or minimizing some information theoretic criteria [119] between histograms. Such approaches are never perfect on discrete valued images. Individual histograms of seg-
CHAPTER 5. GENERATIVE MODEL FOR IMAGE NORMALIZATION AND SYNTHESIS

mented sub-cortical structures can also be matched to the corresponding histograms of an atlas of different contrast [58], however this requires a detailed segmentation of the images which can itself be dependent on the contrast. The problem can also be tackled through the registration of a subject to a multi-contrast atlas and thereby applying the found deformation to the desired atlas contrast [104]. As registration is seldom perfect, this method is prone to failure.

To circumvent these issues, we proposed an atlas based patch matching approach MIMECS [19] in Chapter 4, that neither requires estimation of any tissue parameters nor any atlas-to-subject registration. To synthesize one contrast \( C_2 \) from another \( C_1 \), co-registered multi-contrast atlases are used—the atlas contains contrasts \( C_1 \) and \( C_2 \). The subject and atlas images are decomposed into patches, and the subject patches are approximated by a sparse linear combination of patches from the atlas for contrast \( C_1 \). The corresponding \( C_2 \) patches are linearly combined using the weights from the sparse linear combination to generate the synthetic \( C_2 \) contrast subject patch.

A drawback of this approach, is the failure to use any information from the matching atlas \( C_2 \) contrast patches, although it is evident that the \( C_1 \) and \( C_2 \) contrast intensities are not independent. In this section, we propose a generative model for this type of patch matching synthesis algorithm that takes into account information from both contrasts, by incorporating the idea of coherent point drift [29]. The subject and the atlas images are decomposed into patches, generating corresponding “patch-clouds”. The subject patch-cloud is matched to the atlas cloud using a number of Gaussian
mixture models. Then the synthetic $C_2$ contrast of the subject patches is found as a maximum a-posteriori estimate from the model. To show the effect of synthesis on segmentation consistency, we generate synthetic contrasts for two datasets [1, 3], that are acquired with different pulse sequences (both SPGR and MPRAGE), thus having different contrasts. We show that our approach produces more consistent segmentation than histogram matching and a landmark based approach [2].

The rest of this section is organized as follows. First, we describe the imaging model and the synthesis algorithm. Then a validation study of the segmentation consistency results on the Brainweb phantom [81] data are shown. Then our method is compared with a landmark based method and a histogram method to show improved segmentation consistency on two real datasets having acquisitions from multiple scanners and multiple pulse sequences.

5.2.1 Atlas description

We define the atlas as a pair of co-registered images $\{a_1, a_2\}$ having contrasts $C_1$ and $C_2$, respectively, and are of the same resolution as well. The subject, having the same resolution, is denoted by $b$ and is of contrast $C_1$. Both $b$ and $a_1$ are normalized such that their WM peak intensities are at unity. At each voxel of an image, 3D patches (of size $p \times q \times r$) are considered and stacked into 1D vectors of size $d \times 1$, where $d = pqr$. Atlas $C_1$ and $C_2$ contrast patches are denoted by $y_j$ and $v_j$, respectively, where $j = 1, \ldots, M$. Subject $b$ is comprised of $C_1$ contrast patches, denoted by $x_i, i =$
CHAPTER 5. GENERATIVE MODEL FOR IMAGE NORMALIZATION AND SYNTHESIS

1, \ldots, N. The unobserved $C_2$ contrast subject patches are denoted by $u_i$. $N$ and $M$ represent the number of non-zero voxels in the subject and the atlas, respectively. We combine the patch pairs as $2d \times 1$ vectors $p_i = [x_i; u_i]$ and $q_j = [y_j; v_j]$. Now the subject and the atlas patch clouds are defined as the collection of patch-pairs $P = \{p_i\}$ and $Q = \{q_j\}$.

5.2.2 Contrast Synthesis

As described in Sec. 5.1, the subject and atlas $C_1$ patches represent a pattern of intensities that arise from the same pulse sequence (tissue contrast) and are scaled to a similar intensity range. Therefore, an atlas patch that has a pattern of intensities that is similar to a given subject patch might arise from the same distribution of tissues. In that case, the $C_2$ patch in the atlas can be expected to represent an approximate $C_2$ contrast distribution of the subject in that patch. As proposed earlier, one could find a single patch within the atlas that is close to the subject patch and then use the corresponding $C_2$ atlas patch directly in synthesis. This method is too restrictive as the nearest atlas patch may be far from the subject patch. A slightly more complex way to use this is to find a sparse collection of atlas patches that can reconstruct the subject patch perfectly, then use the same combination of $C_2$ patches to reconstruct a synthetic image. This method, described in [19], may oversmooth the result by combining too many patches. Neither of these approaches uses the $C_2$ atlas patches in selecting the combination. In the present method, we want to retain the property
CHAPTER 5. GENERATIVE MODEL FOR IMAGE NORMALIZATION AND SYNTHESIS

of combining a small number of patches to avoid oversmoothing, but we also want to permit the subject patch to differ somewhat from the associated atlas patches and also take advantage of the $C_2$ patches in the atlas during the $C_1$ patch selection.

We propose a generative method that specifically relates subject $C_1$ and $C_2$ patches to atlas $C_1$ and $C_2$ patches. Since atlas patches may not be plentiful enough to closely resemble all subject patches we consider all convex combinations of pairs of atlas patches. We then postulate that subject patches are random vectors whose probability densities are Gaussian with means given by an unknown convex combinations of pairs of atlas patches and with unknown covariance matrices. This framework captures the notion that a small number of atlas patches (just two in this description) should be used to describe a subject patch and that the subject patch might vary somewhat (Gaussian deviation) from the selected combination of atlas patches. In order to tie the $C_1$ and $C_2$ contrasts together, we further assume that the subject’s unknown $C_2$ patch is a random vector whose mean is the (same) convex combination of the same two atlas patches associated with the $C_1$ contrast, with a covariance matrix that can be different, in principle. We note that the framework of using two atlas patches can be generalized to include multiple atlas patches at the cost of computation overhead, as described in Sec. 5.2.3.

The idea can be summarized succinctly, by considering a subject patch $p_i$ and two associated atlas patches $q_j$ and $q_k$. Then $p_i$ is assumed to arise from the Gaussian distribution,
\[ P_i \sim \mathcal{N}(\alpha_{it} \mathbf{q}_j + (1 - \alpha_{it}) \mathbf{q}_k, \Sigma_t), \quad t \equiv \{j, k\}, \quad (5.6) \]

where \( \Sigma_t \) is a covariance matrix corresponding to the \( j^{th} \) and \( k^{th} \) atlas patch pairs, and \( t \in \Psi, \quad \Psi = \{\{j, k\}; j, k = 1, \ldots, M, j > k\} \), \( \Psi \) being the set of all pairs of atlas patch indices, \( \alpha_{it} \) being a mixing coefficient for the \( i^{th} \) subject patch to the \( t^{th} \) atlas patch-pairs. In essence, each subject patch follows a \( \binom{M}{2} \)-class Gaussian mixture model (GMM). We assume i.i.d. nature of the patches and maximize the probability of observing the subject patches \( \mathbf{x}_i \) using expectation-maximization (EM) \cite{EM} to find the synthetic contrast patches \( \mathbf{u}_i \).

To find the correspondence between a subject patch and an atlas patch pair, define \( z_{it} \) as the indicator function that \( \mathbf{p}_i \) comes from a GMM of \( t = \{j, k\}^{th} \) atlas pair, \( \sum_{t \in \Psi} z_{it} = 1 \quad \forall i, \quad z_{it} \in \{0, 1\} \). Then the probability of observing \( \mathbf{p}_i \) can be written as

\[ P(\mathbf{p}_i | \mathbf{q}_j, \mathbf{q}_k, z_{it} = 1, \Sigma_t, \alpha_{it}) = \frac{1}{\sqrt{2\pi |\Sigma_t|}} \exp \left\{ -\frac{1}{2} \mathbf{h}_{it}^T \Sigma_t^{-1} \mathbf{h}_{it} \right\}, \]

where \( \mathbf{h}_{it} = \mathbf{p}_i - \alpha_{it} \mathbf{q}_j - (1 - \alpha_{it}) \mathbf{q}_k, \quad t \equiv \{j, k\}. \quad (5.7) \]

The prior probability of having \( \mathbf{p}_i \) originating from the distribution of the \( t^{th} \) pair is \( P(z_{it} = 1 | \mathbf{q}_j, \mathbf{q}_k, \Sigma_t, \alpha_{it}) \). Without any knowledge of \( \mathbf{x}_i \), this prior should ideally depend on a classification of the patch cloud \( \mathbf{Q} \). However, we avoid any classification of patches by assuming a uniform prior. Thus, the joint probability becomes

\[ P(\mathbf{p}_i, z_{it} = 1 | \mathbf{q}_j, \mathbf{q}_k, \Sigma_t, \alpha_{it}) = \frac{1}{|\Psi|} P(\mathbf{p}_i | \mathbf{q}_j, \mathbf{q}_k, z_{it} = 1, \Sigma_t, \alpha_{it}). \quad (5.8) \]
CHAPTER 5. GENERATIVE MODEL FOR IMAGE NORMALIZATION AND SYNTHESIS

Assuming $Z = \{z_{it}\}$ and using i.i.d. assumption on patches, the joint likelihood of all patches can be written as

$$P(P, Z|Q, \Theta) = D \prod_{t \in \Psi} \prod_{i=1}^{N} \left[ \frac{1}{|\Sigma_t|} \exp \left\{ -\frac{1}{2} h_{it}^T \Sigma_t^{-1} h_{it} \right\} \right]^{z_{it}}, \quad (5.9)$$

where $h_{it}$ is defined in Eqn. 5.7 and $\Theta = \{u_i, \Sigma_t, \alpha_{it}; i = 1, \ldots, N, t \in \Psi\}$ denotes the set of parameters in the model. $D$ is a normalization constant. Although it is possible to maximize Eqn. 5.9 by EM assuming full-rank $\Sigma_t$, we have experimentally found it to be less robust. Instead, we assume it to be separable and diagonal,

$$\Sigma_t = \begin{bmatrix} \sigma_{1t}^2 I & 0 \\ 0 & \sigma_{2t}^2 I \end{bmatrix},$$

simplifying Eqn. 5.9 to

$$P(P, Z|Q, \Theta) = D \prod_{t \in \Psi} \prod_{i=1}^{N} \left[ \frac{1}{\sigma_{1t}\sigma_{2t}} \exp \left\{ -\frac{||f_{it}||^2}{2\sigma_{1t}^2} \right\} \exp \left\{ -\frac{||g_{it}||^2}{2\sigma_{2t}^2} \right\} \right]^{z_{it}},$$

$$f_{it} = x_i - \alpha_{it}y_j - (1 - \alpha_{it})y_k, \quad g_{it} = u_i - \alpha_{it}v_j - (1 - \alpha_{it})v_k. \quad (5.10)$$

The new set of parameters is $\Theta = \{u_i, \sigma_{1t}, \sigma_{2t}, \alpha_{it}; i = 1, \ldots, N, t \in \Psi\}$. The MAP estimator of $\Theta$ are found by maximizing Eqn. 5.10 using EM [36]. Details of EM algorithm is provided in Sec. 2.3.1 (Eqn. 2.6 – Eqn. 2.7). Using EM, $z_{it}$’s based on a current estimate of the parameters $\Theta$, and then updates $\Theta$ based on the estimates of $z_{it}$. This is described as

1. **E-step:** to find new update $\Theta^{(m+1)}$ at the $m^{th}$ iteration, compute the expectation $Q(\Theta^{(m+1)}|\Theta^{(m)}) = E[\log P(Z|P, Q, \Theta^{(m+1)})|P, Q, \Theta^{(m)}]$. 

CHAPTER 5. GENERATIVE MODEL FOR IMAGE NORMALIZATION AND SYNTHESIS

2. M-step: find new estimates $\Theta^{(m+1)}$ based on the previous estimates using

$$\Theta^{(m+1)} = \arg \max_{\Theta^{(m+1)}} Q(\Theta^{(m+1)}|\Theta^{(m)}).$$

The E-step requires the computation of $E(z_{it}|P, Q, \Theta^{(m)})$. Given that $z_{it}$ is an indicator function, it can be shown that $E(z_{it}|P, Q, \Theta^{(m)}) = w_{it}^{(m)}$, where $w_{it}^{(m)}$ is the posterior probability of $p_i$ originating from the Gaussian distribution of the $t^{th}$ atlas patches $q_j$ and $q_k$, given by

$$w_{it}^{(m+1)} = \frac{\exp \left\{ \frac{-\|f_{it}^{(m)}\|^2}{2\sigma_{1t}^{(m)^2}} \right\} \exp \left\{ \frac{-\|g_{it}^{(m)}\|^2}{2\sigma_{2t}^{(m)^2}} \right\}}{\sum_{\ell \in \Psi} \frac{1}{\sigma_{1\ell}^{(m)} \sigma_{2\ell}^{(m)}} \exp \left\{ \frac{-\|f_{it}^{(m)}\|^2}{2\sigma_{1\ell}^{(m)^2}} \right\} \exp \left\{ \frac{-\|g_{it}^{(m)}\|^2}{2\sigma_{2\ell}^{(m)^2}} \right\}}, \quad (5.11)$$

where $f_{it}^{(m)}$ and $g_{it}^{(m)}$ are the expressions defined in Eqn. 5.10 but with $\alpha_{it}^{(m)}$. $f_{it\ell}^{(m)}$ and $g_{it\ell}^{(m)}$ denote the corresponding values with atlas patches belonging to the $\ell^{th}$ pair, $\ell \in \Psi$, with $\alpha_{it\ell}^{(m)}$. The M-step involves the maximization of the log of the expectation with respect to the parameters given the current $w_{it}^{(m)}$. The update equations are given by

$$u_{it}^{(m+1)} = \frac{\sum_{i \in \Psi} w_{it}^{(m)} \left( \alpha_{it}^{(m)} v_j + (1 - \alpha_{it}^{(m)})v_k \right)}{\sum_{i \in \Psi} w_{it}^{(m)}}, \quad (5.12)$$

$$\sigma_{it}^{(m+1)^2} = \frac{\sum_{i=1}^{N} w_{it}^{(m)} \|x_i - \alpha_{it}^{(m)}y_j - (1 - \alpha_{it}^{(m)})y_k\|^2}{\sum_{i=1}^{N} w_{it}^{(m)}}, \quad (5.13)$$

136
\[ \sigma_{2t}^{(m+1)^2} = \frac{\sum_{i=1}^{N} w_{it}^{(m)} ||u_i^{(m+1)} - \alpha_{it}^{(m)} v_j - (1 - \alpha_{it}^{(m)}) v_k||^2}{\sum_{i=1}^{N} w_{it}^{(m)}} \]  

\[ \alpha_{it}^{(m+1)}: F(\alpha_{it}^{(m+1)}) = 0, \text{ where } F(x) = Ax^2(1-x) - Bx(1-x) + 2x - 1 \]

and \[ A = \frac{||y_k - y_j||^2}{\sigma_{1t}^{(m+1)^2}} + \frac{||v_k - v_j||^2}{\sigma_{2t}^{(m+1)^2}}, \]

\[ B = \frac{(y_k - x_i)^T(y_k - y_j)}{\sigma_{1t}^{(m+1)^2}} + \frac{(v_k - u_i^{(m+1)})^T(v_k - v_j)}{\sigma_{2t}^{(m+1)^2}}, \alpha_{it}^{(m+1)} \in (0, 1). \]  

(5.14) 

It should be noted that from Eqn. 5.15, \( F(0) = -1, F(1) = 1, \forall A, B; \) thus there is always a feasible \( \alpha_{it}^{(m)} \in (0, 1). \) The EM algorithm is said to converge at iteration \( m, \) if \( ||u_i^{(m+1)} - u_i^{(m)}|| < \delta \) for some small \( \delta. \) Once EM converges, the final \( u_i^{(m)} \) is considered to be the synthetic \( C_2 \) contrast, and the middle voxel of \( u_i^{(m)} \) is used as the \( C_2 \) contrast replacement of the \( i^{th} \) voxel.

5.2.3 Results

As there are \( M \) atlas and \( N \) subject patches, the algorithm described above is \( O(NM^2). \) For a typical \( 256 \times 256 \times 198 \) MR image with \( 1\text{mm}^3 \) resolution, \( M, N \sim 10^7. \) Thus it is extremely computationally expensive to use all patches. As described in Sec. 5.2.2, the imaging model is valid for those atlas and subject patches that are close in intensities. Using a non-local type of criteria \([108] \), for every subject patch \( x_i, \) we choose a feasible set of \( L \) atlas patches such that they are the \( L \) nearest neighbors of \( x_i. \) Thus the \( i^{th} \) subject patch follows an \( (L^2) \)-class GMM and the algorithm becomes
CHAPTER 5. GENERATIVE MODEL FOR IMAGE NORMALIZATION AND SYNTHESIS

\(O(NL^2)\). In all our experiments, we choose \(3 \times 3 \times 3\) patches with \(L = 40\) and \(\delta = 0.001 \max(a_2)\).

![Brain images](image)

Figure 5.4: (a) Atlas \(C_1\) contrast (\(\alpha = 90^\circ\), \(n = 0\%\)) and (b) \(C_2\) contrast volume. (c) Subject \(C_1\) contrast (\(\alpha = 45^\circ\), \(n = 3\%\)), (d) original subject \(C_2\) contrast (\(n = 3\%\)) image and (e) synthetic subject \(C_2\) contrast image. (f) Plot of baseline and residual noise between atlas and synthetic \(C_2\) contrast images for \(\alpha = 45^\circ\) and \(\alpha = 60^\circ\).

5.2.3.1 Phantom validation

Our first experiment used Brainweb phantoms [81]. \(C_1\) and \(C_2\) were chosen as the \(T_1\)-w and \(T_2\)-w contrasts, respectively. We synthesized \(T_2\)-w phantoms from the corresponding \(T_1\)-w ones for a variety of imaging parameters and noise levels. Atlases \(a_1\) and \(a_2\) were chosen as the phantoms with flip angle (\(\alpha\)) \(90^\circ\) and \(0\%\) noise. The
repetition time ($T_R$) and echo time ($T_E$) for both the atlas $a_1$ and the subject $b$ were kept fixed at $T_R = 15$ms, $T_E = 2$ms. We used subject $b$ as the $T_1$-w phantoms with $\alpha = 45^\circ$ and $60^\circ$, which are typical values used in real imaging. The noise levels were varied from $n = 0, 1, 3, 5, 7\%$ for the subject and the corresponding $T_2$-w synthetic images $\hat{b}_2$ were generated. Ideally, since the atlas and the subject are the same, $\hat{b}_2$ should be close to $a_2$. Fig. 5.4(a)–(e) show the $T_1$-w and $T_2$-w atlases ($\alpha = 90^\circ$, $n = 0\%$) and a subject ($\alpha = 45^\circ$, $n = 3\%$) along with its synthetic $T_2$-w image.

To quantify the errors in the synthesis, we define $b_2$ as the original $T_2$-w images for different noise levels. The change in $\alpha$ does not change $b_2$. Two error metrics are defined, baseline noise ($||a_2 - b_2||^2$) and the residual noise ($||a_2 - \hat{b}_2||^2$). Baseline noise denotes the noise present in the original $T_2$-w phantoms (Fig. 5.4(d)) and residual noise denotes the synthesis errors. Fig. 5.4(f) shows the plot of these two metrics for two different flip angles over different noise levels. For $n = 0\%$, there is a non-zero residual noise, although the atlas and the subject are the same. This is attributed to the fact that we model every subject patch $x_i$ as a linear combination of two atlas patches $y_j$ and $y_k$. Even if there is a patch $y_\ell$ that exactly matches $x_i$ ($y_\ell = x_i$), the model chooses two patches ($y_\ell$ and $y_t$) to model $x_i$, $y_t$ being the second nearest neighbor of $x_i$, thus introducing systematic variations in intensities. We observe that this effect is less prominent as the noise level increases. Similar errors for $\alpha = 45^\circ$ and $\alpha = 60^\circ$ show that the synthesis is robust to the difference between the pulse sequence parameters of $a_1$ and $b$. This result motivates us to synthesize alternate contrasts using...
CHAPTER 5. GENERATIVE MODEL FOR IMAGE NORMALIZATION AND SYNTHESIS

Figure 5.5: Top row shows the original SPGR, MPRAGE, the patch based synthetic MPRAGE and the landmark based synthetic MPRAGE contrast of a BLSA subject [1], bottom row shows their [6] segmentation.

atlases of similar $C_1$ contrast but acquired with different pulse sequence parameters.

5.2.3.2 Experiments with real data

In this section, we describe the effect of synthesis on segmentation consistency on two real dataset. The Baltimore Longitudinal Study of Aging (BLSA) [1] contains subjects having both SPGR and MPRAGE acquisitions from a GE 3T scanner. Each volume is of size $256 \times 256 \times 199$, with $0.94\text{mm}^3$ resolution. All images were first skull-stripped [34], corrected for any intensity inhomogeneity [35], and scaled such that their WM peak intensities are at unity. From our experience, images with the
Table 5.2: Dice coefficients of hard segmentations between the original MPRAGE acquisition and the synthetic ones as well as the original SPGR acquisitions are shown for 20 BLSA subjects [1]. Our method is compared with histogram matching and a landmark based synthesis [2].

<table>
<thead>
<tr>
<th></th>
<th>CSF</th>
<th>GM</th>
<th>WM</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SPGR</strong></td>
<td>0.815 ± 0.030</td>
<td>0.703 ± 0.052</td>
<td>0.854 ± 0.023</td>
<td>0.793 ± 0.030</td>
</tr>
<tr>
<td><strong>Hist. Match</strong></td>
<td>0.784 ± 0.053</td>
<td>0.751 ± 0.031</td>
<td>0.859 ± 0.012</td>
<td>0.804 ± 0.026</td>
</tr>
<tr>
<td><strong>Landmark</strong></td>
<td>0.828 ± 0.027</td>
<td>0.768 ± 0.026</td>
<td>0.863 ± 0.013</td>
<td>0.818 ± 0.016</td>
</tr>
<tr>
<td><strong>Patch based</strong></td>
<td>0.815 ± 0.028</td>
<td>0.791 ± 0.029 *</td>
<td>0.872 ± 0.014 *</td>
<td>0.821 ± 0.019 *</td>
</tr>
</tbody>
</table>

* Statistically significantly larger than the other three ($p$-value < 0.05).

MPRAGE contrast usually produce more accurate cortical segmentations than those with SPGR; thus, we synthesized MPRAGE contrast from the SPGR acquisitions for 20 subjects. Atlases $a_1$ and $a_2$ are chosen as the SPGR and MPRAGE contrast of a randomly chosen subject in the dataset.

Our method is compared with histogram matching and a landmark based synthesis [2]. Histogram matching is usually sensitive to the choice of bin size and a landmark based method is sensitive to the accuracy of the landmarks. We chose three intensity landmarks, namely mean CSF, GM, and WM intensities, each on both $a_2$ and $b$ using a 3-class Gaussian mixture model. Once the landmarks were found, the intensities between them were linearly scaled. A fuzzy segmentation [6] method was used to segment the images into three tissue classes, CSF, GM, and WM.
Fig. 5.5 shows the two contrasts of a subject and the synthetic images, obtained by landmark matching and our method, along with their segmentations. The red circled region shows the similarity between original MPRAGE and the synthetic one, where the SPGR segmentation fails to detect some CSF and the landmark based synthesis generally overestimates GM. The segmentation consistency is quantitatively described in Table 5.2, where the Dice similarity coefficients between the MPRAGE segmentation and synthetic ones are shown for the three tissue classes. Weighted averages of these three Dice coefficients, weighted by the volume of those tissues, are also reported. The Dice coefficients are averaged over 20 subjects. Our patch based synthesis significantly improves GM and WM segmentation consistency from other methods. However, it performs equally to the landmark based method on CSF segmentation. This is attributed to the fact that CSF is a convoluted structure, for which Dice coefficient is not a good overlap metric, the problem is also influenced by the partial voluming of CSF with GM in these narrow structures. In general, a landmark based synthesis works well for WM, as a WM landmark can be detected robustly. However, for SPGR images, a GM landmark can be erroneous as the image histograms are often uni or bi-modal [2]. Nevertheless, smaller standard deviations in the GM and WM Dice coefficients indicate robustness with respect to pulse sequence changes.

In the next experiment, we generated synthetic images on BIRN [3] travelling subject dataset. The dataset contains 5 subjects, each having 5 SPGRs (one each
CHAPTER 5. GENERATIVE MODEL FOR IMAGE NORMALIZATION AND SYNTHESIS

Figure 5.6: Atlas and subject SPGRs along with the corresponding segmentations from a GE 1.5T, GE 4T, Philips 1.5T, two from a Siemens 3T scanners) and an MPRAGE acquisition (four subjects having GE 3T MPRAGE scans and one having a Siemens 1.5T MPRAGE scan). To show segmentation consistency, we synthesized MPRAGE contrasts of all the SPGR acquisitions of every subject, using \( a_1 \) and \( a_2 \) as
CHAPTER 5. GENERATIVE MODEL FOR IMAGE NORMALIZATION AND SYNTHESIS

Figure 5.7: Landmark based and patch based synthesis of the atlas and subject SPGRs, along with the corresponding segmentations.
Table 5.3: Average Dice coefficients of hard segmentations between the MPRAGE acquisition and the synthetic ones as well as the original SPGR acquisitions for 5 BIRN [3] subjects.

<table>
<thead>
<tr>
<th>Method</th>
<th>CSF</th>
<th>GM</th>
<th>WM</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPGR</td>
<td>0.577 ± 0.049</td>
<td>0.713 ± 0.053</td>
<td>0.820 ± 0.019</td>
<td>0.749 ± 0.028</td>
</tr>
<tr>
<td>Hist. Match</td>
<td>0.606 ± 0.052</td>
<td>0.801 ± 0.025</td>
<td>0.862 ± 0.018</td>
<td>0.805 ± 0.014</td>
</tr>
<tr>
<td>Landmark</td>
<td>0.481 ± 0.066</td>
<td>0.682 ± 0.067</td>
<td>0.816 ± 0.059</td>
<td>0.757 ± 0.078</td>
</tr>
<tr>
<td>Patch based</td>
<td>0.704 ± 0.049 *</td>
<td>0.833 ± 0.014 *</td>
<td>0.915 ± 0.007 *</td>
<td>0.863 ± 0.008 *</td>
</tr>
</tbody>
</table>

* Statistically significantly larger than the other three (p-value < 0.05).

the GE 1.5T SPGR and the GE 3T MPRAGE of the corresponding subject. Fig. 5.6 shows the atlas and four other SPGR acquisitions of a subject. The landmark based and patch based synthesis results along with their hard segmentations are shown in Fig. 5.7. Table 5.3 quantitatively shows that the Dice coefficients between the patch based synthesized images and the original MPRAGE segmentation is significantly higher than the other methods. Both landmark based method and histogram matching depend on the shape and the range of the histograms of the subject and the target. Since the SPGRs were acquired on a variety of scanners, the shape of the image histograms vary widely [19], resulting in higher standard deviation in the average Dice coefficients for both CSF and GM segmentations. However, our method has significantly low standard deviation, indicating robustness to the variation in scanners.
5.2.4 Discussion and conclusion

We have proposed a generative framework for patch based MR tissue contrast synthesis. Using a pair of registered atlases having two different contrasts $C_1$ and $C_2$, we can synthesize $C_2$ contrast for a $C_1$ contrast subject by matching subject patches to the multi-contrast atlas. Our method does not need any image sequence parameters nor does it require any atlas-to-subject registration. As it does not directly rely on any histogram information or landmarks based on histograms, it is more robust to scanner variations compared to histogram matching or a landmark based matching technique, as shown on a multi-scanner dataset [3]. In future, we will explore the effect of convex combination of more than two atlas patches to reconstruct a subject patch.

5.3 Summary

In this chapter, we have described a Bayesian framework for patch based MR intensity normalization and synthesis techniques. In contrast to the one proposed in Chapter 4, this method is based on the assumption that the underlying tissue MR properties (i.e., $T_1$, $T_2$, $P_D$, etc) of the subject and the atlas are of similar nature. The range and distributions of these MR properties for normal subjects are well-known in literature [120–122]. Based on this assumption, we propose a linear combination of atlas patches to match a subject patch, thereby matching the subject patch cloud to
CHAPTER 5. GENERATIVE MODEL FOR IMAGE NORMALIZATION AND SYNTHESIS

an atlas patch cloud in an optimal way. Experiments on both the phantom and real images show that the improvement in intensity matching as well as the synthesis of missing contrasts leads to improvement in segmentation consistency between scans acquired with various pulse sequences, scanners and pulse sequence parameters.
Chapter 6

Longitudinal Intensity Normalization

In previous chapters, we described two MR intensity normalization methods, one based on sparse priors (Chapter 4) and another based on Gaussian mixture models (Chapter 5). We showed that both methods can normalize images acquired with different scanners and imaging sequences (Sec. 4.4.4 and Sec. 5.1.4.2). However, the temporal aspect of the normalization was not addressed before. Thus, when normalizing longitudinal scans of a single subject, previous methods did not use longitudinal information, and this yields image sequences that may lack intensity smoothness over time. An example is shown in Fig. 6.1, where 5 time-points of a normal subject (having total 9 time points) are shown in the top row, along with the corresponding 3-class segmentations using FCM [7]. Each time point is registered to the first
one. The plots of the relative volumes of the three primary tissues, CSF, GM, and WM are shown in the bottom row. Without any longitudinal smoothness constraint in the segmentation process, we observe a large variation in the segmentation volumes, which is not expected in normal subjects or even those with gradual disease or aging [1, 14, 123–125].

This chapter presents a patch based method to normalize temporal intensities
CHAPTER 6. LONGITUDINAL INTENSITY NORMALIZATION

from longitudinal normal brain MR images. Longitudinal intensity normalization is relevant for subsequent processing, steps such as segmentation, so that rates of change of tissue volumes, cortical thickness, or shapes of brain structures can be estimated robustly. Instead of using intensities at each voxel, we use patches as image features. Once all the time-points of a longitudinal dataset are registered, the longitudinal intensity change at each patch is assumed to follow an auto-regressive (AR(1)) process. An estimate of the normalized intensities of a patch at every time-point are generated from a hidden Markov model, where the hidden states are the unobserved normalized patches and the outputs are the observed patches. A validation study on a phantom dataset shows good segmentation overlap with the truth, and an experiment with real data shows more stable rates of change for tissue volumes with the temporal normalization than without.

6.1 Background

Many image processing techniques, such as segmentation, are needed to understand normal aging [12] as well as the progression of diseases. Analysis of 4D temporal data–multiple 3D images at different time-points–is relevant in this scenario to estimate the rates of change of image statistics (i.e., tissue shape and volume, and cortical thickness). Image intensities need to be on a standardized scale to achieve temporally consistent results. However, 3D segmentation algorithms, performed independently
CHAPTER 6. LONGITUDINAL INTENSITY NORMALIZATION

on a longitudinal dataset, seldom achieve the desired longitudinal stability, giving rise to the need for a 4D normalization technique.

Figure 6.2: (a)-(c) shows the first, fourth, and sixth year of a normal subject. Original intensities of a voxel, the center voxel of the blue box in (a), are plotted in (d) as a blue line, while an AR(1) fit of the intensities is shown as a red line.

Several 3D intensity normalization techniques have been proposed in the literature to bring the varying MR intensity ranges to a common scale. Most of these methods are based on deforming the intensity histogram to match a template histogram from an atlas [19] or using spatial landmarks [4,4,98], or some information-theoretic criteria [97,100,119]. They work well in normalizing one 3D volume to another, however they
CHAPTER 6. LONGITUDINAL INTENSITY NORMALIZATION

do not produce the desired stability on a 4D dataset. A state-of-the-art 4D image segmentation technique, CLASSIC [126], was proposed to account for the desired smoothness of the longitudinal segmentation. Unfortunately the algorithm is closely tied to a particular segmentation algorithm and can not easily be extended to other segmentation methods. In this paper, we propose a 4D image intensity normalization method that can be used as a pre-processing step to any segmentation algorithm. Instead of using intensities at each voxel, our method uses patches as features, because a 3D patch around a voxel encodes the neighborhood information at that voxel. We assume that the longitudinal change of intensity of a patch follows an AR(1) process and then estimate the normalized intensity of that patch using a hidden Markov model (HMM) [127]. The hidden states of the HMM are the unobserved normalized patches, while the corresponding outputs are the observed image patches. The motivation for such a model is shown in Figs. 6.2(a)–(c) where three time-points are shown for the longitudinal BLSA dataset [124]. The images are scaled so that their white matter (WM) peaks are unity and Figs. 6.2(b) and (c) are registered to Fig. 6.2(a). The scaled intensities of a selected voxel, shown in the center of the blue square in Fig. 6.2(a), are plotted as a blue line in Fig. 6.2(d). An AR(1) fit of all the intensities is also shown in Fig. 6.2(d) as a red line, with $R^2 = 0.91$. In the following section, we describe the algorithm in detail.
CHAPTER 6. LONGITUDINAL INTENSITY NORMALIZATION

6.2 Method

We assume that there are $T$ time-points in our 4D dataset. The 3D volume of each time-point is denoted as $S_t, t = 1, \ldots, T$, and are all registered to $S_1$. The $S_t$’s have also been linearly scaled such that their WM peaks are at unity. Each 3D volume $S_t$ is made up of small $p \times q \times r$ 3D patches, which we stack as 1D vectors $y_i^{(t)} \in \mathbb{R}^d$, where $i$ denotes the spatial location of the center voxel of the patch and $d$ is the dimension, $d = pqr$. In our HMM model, shown in Fig. 6.3, each of the $y_i^{(t)}$’s has a hidden state, denoted as $x_i^{(t)}$, which is the corresponding unobserved normalized patch. Since the images are registered, we assume the hidden states $x_i^{(1\cdots T)} \equiv \{x_i^{(1)}, \ldots, x_i^{(T)}\}$ follow a Markov process and the transition of the normalized patches from $(t-1)^{th}$ time-point $x_i^{(t-1)}$ to the $t^{th}$ time-point $x_i^{(t)}$ is an AR(1) process,

$$x_i^{(t)} = M_i x_i^{(t-1)} + \epsilon_i^{(t)} \text{ for } t \geq 2, \text{ with } \epsilon_i^{(t)} \sim \mathcal{N}(0, \sigma_{\epsilon,i}^2 I),$$

(6.1)

where $M_i$ is a spatially varying matrix denoting the parameter of the AR(1) process. $\epsilon_i^{(t)}$ denotes the time-varying noise, and is assumed to be a zero mean uncorrelated Gaussian with diagonal covariance, $I$ is the identity matrix.

The collection of parameters of the AR(1) process, namely $M_i$, controls how much the previous state $x_i^{(t-1)}$ contributes to the next state $x_i^{(t)}$. For simplicity, we assume that $M_i$ is a diagonal matrix, $M_i = \text{diag}\{m_i^{(1)}, \ldots, m_i^{(d)}\}$. If $m_i^{(t)}$ is close to zero, then the process looks like white noise. As $m_i^{(t)}$ approaches 1, the $t^{th}$ state gets more contribution from the $(t-1)^{th}$ state, and the result is a smoothing effect of the
CHAPTER 6. LONGITUDINAL INTENSITY NORMALIZATION

Figure 6.3: A hidden Markov model with normalized patches $x_i^{(t)}$ as hidden states, while the observed patches $y_i^{(t)}$'s are the output. The transition from the $(t - 1)^{th}$ time-point to the $t^{th}$ time-point is an AR(1) process.

intensities, as usually observed in AR processes. As an example, Fig. 6.2(d) shows the 1D intensities with the best fitted parameter $m = 0.89$. Without any prior knowledge on the values of $m^{(\ell)}$, we assume a Gaussian distribution, $m_i^{(\ell)} \sim \mathcal{N}(1, \sigma_{m,i}^2)$. This is based on the assumption that most often intensities remain similar within an error range, e.g., inside deep WM, but occasionally, as shown in a voxel near ventricles (Fig. 6.2(a)), they decrease in intensity over time, indicating a change of tissue type at that location.

The observed patches are the outputs of the HMM and they are assumed to be obtained from the hidden states as, 

$$P(y_i^{(t)} | x_i^{(t)}) = \frac{1}{\sqrt{2\pi\sigma_i}} \exp \left\{ -\frac{1}{2} \left( \frac{y_i^{(t)} - x_i^{(t)}}{\sigma_i} \right)^2 \right\}.$$

Here, we have also assumed the errors in obtaining $y_i^{(t)}$ from $x_i^{(t)}$ to be uncorrelated zero mean Gaussian. Using the conditional independence of $x_i^{(1\cdots T)}$ and $y_i^{(1\cdots T)}$, we can write

$$P(x_i^{(t)} | x_i^{(1)}, \ldots, x_i^{(t-1)}) = P(x_i^{(t)} | x_i^{(t-1)}), P(y_i^{(t)} | x_i^{(1\cdots T)}) = P(y_i^{(t)} | x_i^{(t)}).$$

The collection of all unknown parameters are denoted as $\Theta = \{\sigma_i, \sigma_{m,i}, \sigma_{c,i}, m_i^{(\ell)}\}$. Now the parameter set $\Theta$ and the hidden states $x_i^{(1\cdots T)}$ are estimated by a maximum-a-
posteriori (MAP) estimation criteria by maximizing the posterior probability of $x_i^{(1:T)}$ as

$$\begin{align*}
(x_i^{(1:T)}, \Theta) &= \arg \max_{x_i^{(1:T)}, \Theta} P(x_i^{(1:T)}, \Theta | y_i^{(1:T)}) \\
&= \arg \max_{x_i^{(1:T)}, \Theta} P(y_i^{(1:T)} | x_i^{(1:T)}, \Theta) P(x_i^{(1:T)} | \Theta) P(\Theta).
\end{align*}$$

(6.2)

We simplify $P(y_i^{(1:T)} | x_i^{(1:T)}, \Theta)$ and $P(x_i^{(1:T)} | \Theta)$ and use the prior distribution of $m_i^{(\ell)}$ as $P(\Theta)$. Assuming i.i.d. nature of patches and discarding some normalization constants, the posterior probability becomes

$$P(x_i^{(1:T)}, \Theta | y_i^{(1:T)}) \propto \frac{1}{\sigma_i^T \sigma_{\epsilon,i}^{T-1} \sigma_{m,i}} \exp \left\{- \sum_{t=1}^{T} \frac{||y_i^{(t)} - x_i^{(t)}||^2}{2\sigma_i^2} - \sum_{t=2}^{T} \frac{||x_i^{(t)} - M_i x_i^{(t-1)}||^2}{2\sigma_{\epsilon,i}^2} - \sum_{\ell=1}^{d} \frac{(m_i^{(\ell)} - 1)^2}{2\sigma_{m,i}^2} \right\}. \quad (6.3)$$
Maximizing the posterior w.r.t. $\Theta$ yields the following update equations,

$$m_i^{(t)} = \frac{1}{\sigma_{e,i}^2} \sum_{t=2}^{T} x_i^{(t)}(\ell)x_i^{(t-1)}(\ell) + \frac{1}{\sigma_{m,i}^2},$$

$$\sigma_i^2 = \frac{1}{T} \sum_{t=1}^{T} ||y_i^{(t)} - x_i^{(t)}||^2,$$

$$\sigma_{m,i}^2 = \sum_{\ell=1}^{d} (m_i^{(\ell)} - 1)^2,$$

$$\sigma_{e,i}^2 = \frac{1}{T-1} \sum_{t=2}^{T} ||x_i^{(t)} - M_i x_i^{(t-1)}||^2,$$

$$x_i^{(t)} = \left( \left( \frac{1}{\sigma_i^2} + \frac{1}{\sigma_{e,i}^2} M_i^T M_i \right)^{-1} \left( \frac{y_i^{(t)}}{\sigma_i^2} + \frac{M_i x_i^{(t+1)}}{\sigma_{e,i}^2} + \frac{M_i x_i^{(t-1)}}{\sigma_{e,i}^2} \right) \right)^{-1} \left( \frac{y_i^{(t)}}{\sigma_i^2} + \frac{M_i x_i^{(t+1)}}{\sigma_{e,i}^2} + \frac{M_i x_i^{(t-1)}}{\sigma_{e,i}^2} \right),$$

$$t = 2, \ldots, T - 1,$$

$$x_i^{(1)} = \left( \frac{I}{\sigma_i^2} + \frac{1}{\sigma_{e,i}^2} M_i^T M_i \right)^{-1} \left( \frac{y_i^{(1)}}{\sigma_i^2} + \frac{M_i x_i^{(2)}}{\sigma_{e,i}^2} \right),$$

$$x_i^{(T)} = \left( \frac{1}{\sigma_i^2} + \frac{1}{\sigma_{e,i}^2} \right)^{-1} \left( \frac{y_i^{(T)}}{\sigma_i^2} + \frac{M_i x_i^{(T-1)}}{\sigma_{e,i}^2} \right) \quad (6.4)$$

We use a coordinate-descent type optimization to solve Eqn. 6.4 until the difference in the posterior, Eqn. 6.3, between successive iterations is sufficiently small. After Eqn. 6.4 reaches convergence, the center voxel of $y_i^{(t)}$ is replaced with the center voxel of $x_i^{(t)}$.

### 6.3 Results

In our first experiment, we used a tissue atrophy simulation method [128] on the Brainweb [81] T1-w phantom to simulate temporal data. Ventricles of the phantom were deformed using six different atrophy radii to simulate the normal aging at six dif-
Table 6.1: The table shows the (mean ± std) $R^2$ of linear fits of the tissue volumes obtained from hard segmentations of the original image and also the hard segmentations based on three normalization approaches: histogram matched; landmark based [4]; and our approach (Longitudinal). The $R^2$ values are based on an average of seven subjects.

<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Hist. Matched</th>
<th>Landmark Based</th>
<th>Longitudinal</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSF</td>
<td>0.362 ± 0.361</td>
<td>0.045 ± 0.027</td>
<td>0.710 ± 0.280</td>
<td>0.724 ± 0.380 †</td>
</tr>
<tr>
<td>GM</td>
<td>0.819 ± 0.067</td>
<td>0.068 ± 0.079</td>
<td>0.630 ± 0.183</td>
<td>0.930 ± 0.037 *</td>
</tr>
<tr>
<td>WM</td>
<td>0.578 ± 0.180</td>
<td>0.075 ± 0.117</td>
<td>0.379 ± 0.248</td>
<td>0.903 ± 0.019 *</td>
</tr>
</tbody>
</table>

* Statistically significantly larger than the other three ($p$-value < 0.05).
† Statistically significantly larger that original and histogram matching ($p$-value < 0.05).

Different time-points. Then the images were normalized using this method and the hard segmentations [7] were compared with the ground truth obtained from the simulation. We calculated Dice coefficients between the hard segmentations of un-normalized and normalized images with the truth. A paired $t$-test shows that the Dice coefficients are not statistically different ($p$-value = 0.014), indicating that normalization does not deteriorate the stability of the segmentation when the segmentation is very close to the truth.

We used seven subjects from the BLSA dataset [124] to show the segmentation stability. Each subject has 8–11 time-points. For one subject, three time points and the corresponding normalized images are shown in Fig. 6.3(a). We compared our method with histogram matching and a landmark based histogram transformation.
CHAPTER 6. LONGITUDINAL INTENSITY NORMALIZATION

method [4]. For these two algorithms, we used each subject’s first time-point as the reference histogram. The relative volumes of cerebro-spinal fluid (CSF), gray matter (GM), and WM were computed from the hard segmentations of the original, normalized, and matched volumes, plotted for one subject in Fig. 6.3(b). Clearly, normalized volumes are more stable, which is also reflected in Table 6.1, where average coefficient of determination $R^2$ from linear fitting from the volumes are shown. Larger $R^2$ for normalized volumes indicate smoother rates of change. $R^2$ values of the fits for longitudinal normalization are significantly larger than the original and two other matching algorithms, except for the CSF on the landmark based method. This can be attributed to the high variability of CSF segmentation in the SPGR images.

6.4 Conclusion

We have described a method to normalize 4D longitudinal data which can be used as a pre-processing step to segmentation algorithms. Although this method currently assumes that the time-points are equally spaced, future work would overcome this limitation by introducing age factors into the time-dependent noise variances $\sigma_{m,i}^2$. 
CHAPTER 6. LONGITUDINAL INTENSITY NORMALIZATION

Figure 6.4: (a) Three time-points (1\textsuperscript{st}, 4\textsuperscript{th} and 9\textsuperscript{th}) of a subject and the corresponding normalized images are shown. (b) After segmentation, CSF, GM, and WM relative volumes are plotted with respect to age.
Chapter 7

Appendix

A brief derivation of Eqn. 3.23 to Eqn. 3.26 are given in this appendix.

To do the E-step, we find the \( w_{jk}^{(m)} \) as

\[
u_{jk}^{(m+1)} = E(z_{jk}|y_j, \Theta^{(m)})
\]

\[
= 1.P(z_{jk} = 1|y_j, \Theta^{(m)}) + 0.P(z_{jk} = 0|y_j, \Theta^{(m)}),
\]

\[
= P(z_{jk} = 1|y_j, \Theta^{(m)}),
\]

\[
= \frac{P(y_j|z_{jk} = 1, \Theta^{(m)})P(z_{jk} = 1|\Theta^{(m)})}{\sum_{k=1}^{K} P(y_j, z_{jk} = 1|\Theta^{(m)})},
\]

\[
= \frac{f_{\text{MRF}}(z_{jk}|z_{N_j}, \Theta^{(m)}) f_R(y_j|\Theta^{(m)})}{\sum_{k=1}^{K} f_{\text{MRF}}(z_{jk}|z_{N_j}, \Theta^{(m)}) f_R(y_j|\Theta^{(m)})},
\]

where \( f_{\text{MRF}}(z_{jk}|z_{N_j}, \Theta^{(m)}) \) is given by Eqn. 3.20. Using mean-field approximation [44] to replace \( z_{jk} \) by the current estimate of its expectation \( w_{ij}^{(m)} \), we obtain Eqn. 3.23.
CHAPTER 7. APPENDIX

The M-step provides the estimation of $\Theta$. $Q(\Theta^{(m+1)}|\Theta^{(m)})$ becomes,

$$E \left[ \log f(Z|\Theta^{(m+1)})|y, \Theta^{(m)} \right]$$

$$= \sum_{j \in \Omega} \sum_{k=1}^{K} w_{jk}^{(m)} \log \left\{ f_{\text{MRF}}(w_{jk}^{(m)}|w_{Nj}^{(m)}, \Theta^{(m+1)}) f_{R}(y_j|\Theta^{(m+1)}) \right\},$$

$$= \sum_{j \in \Omega} \sum_{k=1}^{K} w_{jk}^{(m)} \log \left\{ f_{\text{MRF}}(w_{jk}^{(m)}|w_{Nj}^{(m)}, \Theta^{(m+1)}) \right\} + \sum_{j \in \Omega} \sum_{k=1}^{K} w_{jk}^{(m)} \log \left\{ f_{R}(y_j|\Theta^{(m+1)}) \right\}.$$

We note that the 1st term of $E \left[ \log f(Z|\Theta^{(m+1)})|y, \Theta^{(m)} \right]$ is explicitly independent of $v_{k}^{(m+1)}$ and $\sigma_{k}^{(m+1)}$ and the 2nd term is explicitly independent of $\beta_{k}^{(m+1)}$. So the expectation $E \left[ \log f(Z|\Theta^{(m+1)})|y, \Theta^{(m)} \right]$ is maximized w.r.t. $v_{k}^{(m+1)}$, by setting the partial derivative of the 2nd term w.r.t. $v_{k}^{(m+1)}$ to zero,

$$\frac{\partial}{\partial v_{k}^{(m+1)}} \sum_{j \in \Omega} \sum_{k=1}^{K} w_{jk}^{(m)} \left[ \log \left\{ f_{R}(y_j|\Theta^{(m+1)}) \right\} \right] = 0,$$

$$\Rightarrow \frac{\partial}{\partial v_{k}^{(m+1)}} \sum_{j \in \Omega} \sum_{k=1}^{K} w_{jk}^{(m)} \left[ \log \frac{y_j}{\sigma_{k}^{(m+1)}} - \frac{v_{k}^{(m+1)} y_j^2 + y_j^2}{2\sigma_{k}^{(m+1)}^2} + \log I_{0} \left( \frac{y_j v_{k}^{(m+1)}}{\sigma_{k}^{(m+1)}} \right) \right] = 0.$$

Simplifying this equation and also using the fact that $\frac{d}{dx} I_{0}(x) = I_{1}(x)$, a coordinate descent equation for $v_{k}^{(m+1)}$ is obtained in Eqn. 3.24.

Similarly, Eqn. 3.25 is obtained by setting the partial derivative of the 2nd term w.r.t. $\sigma_{k}^{(m+1)}$ to zero.

As the 2nd term of $E \left[ \log f(Z|\Theta^{(m+1)})|y, \Theta^{(m)} \right]$ is explicitly independent of $\beta_{k}^{(m+1)}$, we equate the partial derivative of $\sum_{j,k} w_{jk}^{(m)} \left[ \log \left\{ f_{\text{MRF}}(w_{jk}^{(m)}|w_{Nj}^{(m)}, \Theta^{(m+1)}) \right\} \right]$ w.r.t
CHAPTER 7. APPENDIX

\( \beta_k^{(m+1)} \) to zero,

\[
\frac{\partial}{\partial \beta_k^{(m+1)}} \sum_{j \in \Omega} \sum_{k=1}^K \left[ \log \left( |N_j| \beta_k^{(m+1)} \right) + \frac{\sum_{i \in N_j} \left( w_{jk}^{(m)} - w_{ik}^{(m)} \right)^2}{2 \beta_k^{(m+1)^2}} \right] = 0,
\]

\[
\Rightarrow \sum_{j \in \Omega} \left[ \frac{1}{\beta_k^{(m+1)}} - \frac{\sum_{i \in N_j} \left( w_{jk}^{(m)} - w_{ik}^{(m)} \right)^2}{\beta_k^{(m+1)^3}} \right] = 0,
\]

to get update equation Eqn. 3.26.
Bibliography


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


[51] U. Vovk, F. Pernu, and B. Likar, “Multi-feature Intensity Inhomogeneity Cor-
BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


[77] D. T. Chard, G. J. Parker, C. M. Griffin, A. J. Thompson, and D. H. Miller,
BIBLIOGRAPHY


BIBLIOGRAPHY


[95] D. L. Donoho, “For most large underdetermined systems of linear equations,
the minimal $\ell_1$-norm solution is also the sparsest solution,” *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, 2006.


[101] R. Barber, P. Scheltens, A. Ghokar, C. Ballard, I. McKeith, P. Ince, R. Perry, and J. O’Brien, “White matter lesions on magnetic resonance imaging in de-
mentia with Lewy bodies, Alzheimer’s disease, vascular dementia, and normal

tation of T2-FLAIR Multiple Sclerosis Lesions,” in Midas Journal, MICCAI
Workshop, 2008.

timization of 3-D MP-RAGE Sequences for Structural Brain Imaging,” Neuro-

textbook of deformable neuroanatomies,” Proc. Natl. Acad. Sci. USA, vol. 90,


[106] R. Tibshirani, “Regression Shrinkage and Selection via the Lasso,” Journal of

Linear Equations by Linear Programming,” Proc. of Natl. Acad. of Sciences,
vol. 102, no. 27, pp. 9446–9451, 2005.


[119] F. Jager and J. Hornegger, “Nonrigid Registration of Joint Histograms for In-


[125] K. M. Kennedy, K. I. Erickson, K. M. Rodrigue, M. W. Voss, S. J. Colcombe,
BIBLIOGRAPHY


Vita

Snehashis Roy was born in Kolkata, India on August 24th, 1983. He received the B.Tech. degree from Indian Institute of Technology, Kharagpur, India, in 2006 and the M.S.E. degree from the Johns Hopkins University, Baltimore, MD in 2010, both in Electrical and Computer Engineering. He entered the Ph.D. program in the Department of Electrical and Computer engineering at the Johns Hopkins University in 2006. His research interest is in the area of medical image analysis with primary focus on brain tissue segmentation, MRI intensity normalization, and MR contrast synthesis of brains. His work on MR image contrast synthesis achieved the best poster award in the 22nd International Conference on Information Processing in Medical Imaging, Kloster Irsee, Germany, 2011.