To My Parents, My Wife and Sons, My Brother and Friends Who Share the Happiness with Me

Acknowledgements

I would like to express my honest gratitude to my advisors, Professor Tianhu Lei and Professor Tülay Adalı, for being so generous with their energy, guidance, input, and encouragement, throughout the course of this research. Specifically, since this research covers both electrical engineering and biomedical science, their supervision provides a privilege for me to work on this subject with the best environment.

Deep appreciation is expressed to two major readers, Professor Tülay Adalı and Dr. Robert F. Wagner, for their careful reading and detailed comments for improving this dissertation. The deep thanks also forward to Professors Joel M. Morris and Wilfred Sewchand, for serving on my committee.

I have gained greatly from interactions with other researchers. In particular a number of discussions with Professor Shih-Chung Ben Lo, Professor Zsolt Szabo, and Dr. Keith A. Wear had a large influence. The work was also greatly affected by many conversations with Jason Xuan, Youhong Lu, and Darmadi Komo.

I would like to thank Professor Seong K. Mun, for providing the opportunity for my visiting research at Georgetown University Medical Center. This wonderful experience makes my research life more productive.

I would like to thank Dr. Sun Yat-Sen Scholarship Foundation for being so generous with the financial support to me in completing my Ph.D degree.

The research which led to this dissertation has been supported by both Department of Radiation Oncology and Department of Electrical Engineering.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Magnetic resonance (MR) imaging - one of the fastest growing areas in scientific world - is a non-invasive computed imaging technique that has been opening many medical frontiers by enabling physicians to study the internal activities of human body visually. MR imaging has attracted much attention in the research community primarily due to its rich biomedical information content and the fact that it has great clinical potential [9, 16, 32, 34, 35, 36, 37]. MR imaging is also a particularly exciting field in image processing research, because it presents new challenges for the development of theory and techniques to perform sophisticated visual information processing. It also requires a concrete and unified understanding of the underlying physics, statistics, signal processing, and information formation of the MR imaging process [1, 2, 16, 22, 30].

In an effort to improve MR imaging quality, the physical and mathematical problems have been intensively studied during the last two decades [10, 11, 14]. Unfortunately, most computed images do not give a perfect representation of the true scene context. Statistical image analysis can provide a better way for the improvement of MR imaging process [16, 22]. Furthermore, imaging process and image analysis are traditionally considered to belong to different research areas. Hence, little work has been done on the unification of these two procedures, although they can obviously benefit a great deal from each other. The main goal of this dissertation is to provide a new framework for stochastic modeling of MR images by incorporating imaging statistics and random field theory, and to develop an unsupervised learning procedure for model-based MR image analysis.

The motivation for conducting such a statistical approach also arises from the fact that there is great demand for MR image analysis in related clinical applications, such as; automatic image analysis, computer-aided diagnosis, radiation treatment planning, and medical informatics [9, 16, 18, 19, 35]. For example, in effective radiation treatment planning, accurate and automated segmentation of therapy-relevant anatomical structures and disease targets, is highly critical. Currently, this is achieved by manual delineation which is a major effort to be carried out for each patient and there are only several existing semi-automated systems with limited capabilities [35]. Hence, it is very important to develop more sophisticated computing techniques to automate this process as much as possible. The change of biomedical quantities and the appearance of disease patterns are the early signs of cancer, and only an earlier detection can increase the chance of survival for patients with cancer. However, an accurate diagnosis highly depends on the sensitivity of object quantities and the visibility of disease patterns. Because of the limitations of imaging hardware in the forms of noise and clutter, and of the human eye with low contrast resolution, these small parameter changes and early disease patterns can be hidden. Hence, the efficiency and effectiveness of tissue characterization can be highly increased by an image analysis system that is capable of finding, enhancing, and segmenting the disease patterns and changes in object quantities [19].

1.2 Background

A great deal of research has been done on the analysis of x-ray CT image [17], ultrasound image [33], and mammographic image [19]. MR image analysis has recently been a very active research field, primarily because of the considerable use of this modality and the fast advances in its clinical applications. From a practical point of view, however, it is very important to have a reliable and complete framework for MR image modeling and analysis as well as development of related techniques.

1.2.1 Image Statistics

The physical and mathematical problems of MR imaging have been intensively studied during the last two decades. A good summary can be found in [11]. However, although a lot of pioneer work is reported on the statistical analysis for x-ray CT [17], positron emission tomography [29], and ultrasound scan [23], there has been little work done on MR imaging statistics, such as random source identification, statistical analysis of pixel images, and statistical treatment of context images [1]. Edelstein *et al.* [30] discusses noise sources and effects in MR imaging concentrating mainly on thermal noise. The missing point is that object variability, an important random source in micro-scale, is ignored, and only results on first order statistics are presented. Wagner *et al.* [25] conducts an intensive research on thermal noise in MR imaging and concludes that object variability is not a significant random source in pixel scale. However, all results are explained in the raw data domain, not in the reconstruction domain, i.e., pixel image domain. Godtliebsen *et al.* [22] reports a pioneer work on the statistical properties of MR image in terms of a comparative real MR data analysis. Although their results are very useful in objectively validating MR image statistics, they do not provide theoretical justification. Fuderer [76] apply information theory to quality evaluation in MR imaging in which many assumptions were made in order to derive some conclusions. The problem is that some of the assumptions do not hold for most cases. Hence, an objective investigation for a complete statistical description of MR image, from the imaging equations to the image random field, is needed. This problem is solved by characterizing the statistical properties of MR images as the standard problems of statistics and justifying the heuristic assumptions that has appeared in the literature based on the underlying physics and mathematics.

1.2.2 Computational Image Modeling

The objective of computational modeling for MR image is to capture the intrinsic stochastic character of image in a structure with a few parameters so as to understand the nature of imaging process. Many statistical image models have been proposed and applied to model based MR image analysis. Liang [24] proposes statistical models of *a-priori* information for MR image by modifying Gaussian model with neighboring correlation constraints, however, the justification of the models is not stated clearly and some model parameters are chosen heuristically. Godtliebsen *et al.* [22] use a Markov random field model in a noise reduction technique for MR image. Since no justification is given, the method is unable to present significant progress in either theoretical or practical aspects. The COVIRA (Computer Vision in Radiology) project mentioned by Elliott *et al.* [35] is an advanced medical informatics to improve the diagnosis and planning of treatment for patients with brain tumor and other diseases. They use texture image models and conduct intensive research on texture descriptors and statistical knowledge base, but this research has been slowed down because the preliminary results lack good mathematical justification. Santago *et al.* [32] modify finite normal mixture (FNM) model for the quantification of MR brain images by developing a supervised partial volume modeling (PVM). The drawback is that it is not practically useful since in fact their PVM can not provide a flexible framework for real clinical cases. Since the previously proposed models lack objective justification, a number of fundamental issues remain unexplored. Hence, for MR image analysis, a better understanding of the

[1] In this dissertation, pixel image refers to the intensity or gray level, context image refers to the label or membership of image pixels.

problems is needed as well as efficient algorithms. Within a mathematical framework, the development of suitable models for both pixel images and context images by using MR image statistics is required.

1.2.3 Unsupervised Parameter Estimation

The key step in unsupervised stochastic model-based MR image analysis is parameter estimation which is a process to quantify model parameters such that the resulting model best fits the observed image [62]. For the computational efficiency, it is desirable to estimate model parameter values directly from the observed data before final pixel classification [61]. In general, the parameter set of the models can be classified into two groups: structural and regional parameters. The problem addressed here is the combined estimation of these parameters based on their statistical nature and the appropriate optimization criteria selected. Many previously proposed techniques exhibit a typical drawback: they are only partially unsupervised since some parameters are often heuristically or empirically specified. In the work reported recently by Liang *et al.* [18], a hidden Markov random field (MRF) model is used that provides promising solution to the problem. The disadvantage is that parameter estimation is performed in a partially unsupervised manner in which the Markov parameter value and the scale factor in determining image region numbers are specified empirically. Santago *et al.* [32] use simulated annealing method to determine model parameter values based on FNM and PVM in a supervised mode. It has difficulty in dealing with abnormal cases encountered in a real situation, and is a time-consuming procedure. In addition, none of them [18, 32] has addressed issues such as algorithm initialization and local minima. We construct a unified information theoretic criterion for this optimization procedure with a completely unsupervised model-fitting implementation.

1.2.4 Contextual Image Segmentation

After parameter estimation, image segmentation is formulated as a statistical classification problem. According to the statistical nature of pixel labels and the optimization criterion that is used, most detectors fall into one of the three classifiers: maximum likelihood (ML), Bayesian, and maximum a posterior probability (MAP). However, in spite of impressive results shown by many model-based segmentation schemes, the framework still lacks of a satisfactory solution. For example, most non-contextual segmentation algorithms are likely to perform poorly since there may not be sufficient local information to make a good decision, and the Markov random field (MRF) based contextual techniques are often computationally prohibitive. Brummer *et al.* [34] has done automatic detection of brain contours in MRI data sets by using a simple thresholding method and then improving it by morphological filtering. It may be a promising approach but ignores the possible noise effects. Joliot *et al.* [37] has conducted research on three-dimensional segmentation and interpolation of MR brain images by using an approach similar to Brummer's. Hall *et al.* [36] provides a comparison of neural network and fuzzy clustering techniques in MR brain image segmentation, only some preliminary result has been presented. They use both neural network and fuzzy clustering approaches which does not take the neighborhood information into consideration. Hence, more efficient implementation is required to translate richer context statement into the algorithm by naturally unifying edge and region information in a single unified framework.

In summary, we can conclude that based on our literature survey, currently no work has provided reasonable and suitable framework for the MR image analysis that fulfills all the requirements from clinical application, mathematical justification, and algorithm implementation.

1.3 Overview of Dissertation

Main topics discussed in this dissertation are: imaging formulation, image statistics, image modeling, parameter estimation, and image segmentation, for MR images. Also, many experimental simulations are conducted to verify the ideas that formulate the related problems, to examine the algorithms for solving these problems, and to demonstrate their potential practical applications. Since accuracy and efficiency are

two important practical issues, some related topics are also discussed in order to provide a balance between the usually opposing requirements of achieving high accuracy and efficiency at the same time.
The dissertation is organized as follows:

Chapter 2 gives a brief review on the physical and mathematical principles in MR imaging. Certain key issues in imaging physics and mathematics are introduced. In particular, a convenient mathematical definition of MR imaging quantity is given, and a more complete discussion on the point spread function (PSF) of MR imaging is presented. It is emphasized that an MR image is a discrete and interpolated visual representation of the spatial distribution of the linear magnetic resonance coefficients[2].

Chapter 3 presents a complete statistical description of MR imaging, covering the fundamental statistical properties of MR Image. Both thermal noise and object variability, as major distortion sources, are considered. The statistical properties of pixel images, such as Gaussianity, dependence, stationarity, and ergodicity, are characterized as the standard problems of statistics, and are justified to form a basis for establishing pixel image models. Specifically, main assumptions commonly used in current literature have been verified. Furthermore, the stochastic aspects of context images within the framework of Markov regularization is explored. It is found that the spatial continuity and inhomogeneity can be represented in terms of randomization rule and inhomogeneous Markov configuration.

In Chapter 4, we introduce stochastic approach for modeling MR image with different formulations. Based on MR image statistics, the correct use of conventional finite normal mixture (FNM) models for the specified tasks in MR image analysis is first justified with new definitions, observations, and theorems, which leads to a better mathematical understanding, and then the different versions of this framework is proposed to include the multiresolution and localized formulation by incorporating correlation among local pixels. The new formulations impose context regularities into the model structure and parameters in an intuitively appealing manner, which simplifies problem formulation. Furthermore, an inhomogeneous hidden MRF model is proposed to motivate a unifying theme for incorporating context regularities by local randomization and Gibbs measures. The natural unification of the boundary and region information in this framework is recognized in which Gaussian parameters represent region information while Markov parameters reflect boundary discontinuity.

In Chapter 5, we address the problem of unsupervised parameter estimation. A unified optimization procedure is proposed based on information theoretic criteria. Specifically, we use a model-fitting procedure for detecting the number of image regions, by using Akaike information criterion (AIC), minimum description length (MDL), and a newly proposed minimum conditional bias/variance (MCBV) formulation. For the estimation of regional parameters, a hybrid algorithm is developed based on the ML principle for achieving parameter estimates in three steps: adaptive Lloyd-Max histogram quantizer (ALMHQ) algorithm for parameter initialization, block-wise classification-maximization (BCM) algorithm for refining parameter initialization, and expectation-maximization (EM) algorithm for finalizing parameter estimation. The new parameter estimation procedure received efficient and accurate responses, in which the likelihood of local minima is reduced and the ML parameter estimates are asymptotically unbiased. Finally, Markov parameter values in the Gibbs measure is identified by using entropy rate information in which boundary defined Markov configuration (BDMC) algorithm is developed.

Chapter 6 focuses on the development of efficient algorithms for contextual MR image segmentation and the justification of optimization criterion by considering its behavior at local/global scale. Non-contextual pixel classification is performed to initialize image segmentation. Both ML and Bayesian classifiers are introduced, and the discussion centers on pixel label statistics and misclassification error distribution. Within the framework of relaxation labeling procedure, two algorithms are then derived for final contextual segmentation: contextual Bayesian relaxation labeling (CBRL) and modified iterated conditional mode (MICM) algorithms. Derivations show that the iterative local Bayesian detections in CBRL algorithm results in a consistent labeling solution based on localized SFNM formulation, and MICM algorithm approximates the MAP solution directly using conditional Gibbs measure. The originality of the proposed approach lies in the fact that CBRL algorithm uses a non-parametric mode to impose local spatial consistent constraint by

[2] The definition of linear magnetic resonance coefficients will be given in Section 2.3.1.

using compatibility measure in a contextual image segmentation, and MICM algorithm integrates boundary information into decision maker by an inhomogeneous Markov configuration.

In Chapter 7, we examine the efficacy of the ideas and algorithms that are developed in this dissertation, with the focus on the unsupervised stochastic model based MR image analysis. We show that by using the techniques developed in this thesis it is possible to achieve an accurate and efficient performance in many practical MR image analysis applications, with a better mathematical understanding, richer context representation, and less computational requirements. A summary of major conclusions and future work is given in Chapter 8.

Chapter 2

MR Imaging

2.1 Introduction

MR imaging is a technique to create a 2D or 3D picture of the interior of an object. The underlying physics of MR imaging is nuclear magnetic resonance (NMR) which is based on the interaction between an atomic nuclei and external magnetic fields. A typical NMR experiment is first to perturb the system from its equilibrium configuration, and then to observe how it relaxes back to the equilibrium. During this relaxation process a free induction decay (FID) signal is generated and collected. This signal mainly depends on the nuclear spin density ρ, the spin-lattice relaxation time T_1, and the spin-spin relaxation time T_2, and can be modulated and coded by the proper pulse sequences. Applying image reconstruction algorithm to this FID signal, a spatial distribution of magnetization weighted by ρ, T_1 and T_2 will be formed, which is known as MR image [9, 10, 16]. The basic nucleus considered here is the one of the hydrogen (water and lipids). Hence, the information about ρ and T_1 and T_2 provides a more precise picture of tissue functionality.

In order to analyze MR images, its physics, mathematics, and statistics must be fully studied. This consideration is strongly shared by many researchers for other imaging modalities. For example, in physical and mathematical aspects, [17] defines the imaging quantity in x-ray CT as relative linear attenuation coefficient; [23] conducts an intensive research on fundamental correlation length in ultrasound images and made an explicit connection between the point spread function (PSF) and the auto covariance function (ACF); in statistical aspect, [20] proposes a theoretical discussion on quantum noise and object variability for photon imaging system; [25] provides a unified SNR analysis of medical imaging systems. Some parallel research has been done for MR imaging. For example, [22] compares several statistical methods in MR imaging by analyzing real MR image data; and [30] and [76] discuss the noise process in MR imaging from different points of view, etc.

In this chapter, we present a brief review covering both the physics and signal processing of MR imaging in a unified way. From quantum mechanics of nuclear spins, Bloch equation is discussed in detail to specialize the knowledge about NMR. By introducing spatial coding and pulse sequence techniques, the measured FID signal is formulated as a modulated Fourier integration. After defining MR imaging quantity using new concept of linear magnetic resonance coefficient, we discuss the standard MR image formation algorithm in terms of 2D forward- and inverse- Fourier transform pairs. In particular, we derive a more complete formulation on the system point spread function of MR imaging and give a more clear formulation of MR imaging equations, in which the practical requirements for window functions and the discrete finite length effects are emphasized.

2.2 Imaging Physics

2.2.1 Nuclear Magnetic Resonance

Nuclei that contain an odd number of protons possess a nucleus spin and a magnetic momentum $\vec{\mu}(t)$. When an external magnetic field $\vec{H}_0$ is applied, the randomly oriented nuclei tend to align in either parallel or anti-parallel directions with this field. Some nuclei rotate like a spinning top. The rotation frequency is called Larmor frequency

$$\omega_0 = \gamma H_0 \tag{2.1}$$

where γ is the gyromagetic ratio. The spins due to alignment have two energy levels: $+\mu H_0$ and $-\mu H_0$. According to Gibbs law, a larger population of spins are at the low energy level. By transmitting a radio-frequency (RF) signal with energy $\Delta E = 2\mu H_0$ to nuclei in the form of an RF magnetic field $\vec{H_{RF}}$, the protons at the low energy level are to be excited to the high energy level. After removing RF field $\vec{H_{RF}}$, the excited protons tend to return to their low energy level by emitting an RF signal. The emitted RF signal is called free induction decay (FID) signal which can be detected by antenna coils.

Normally, we do not describe $\vec{\mu}(t)$ of a single nuclear. Instead, we will describe the behavior of N identical nuclei in a small volume element V_0 centered at (x, y, z)[1], contributing to the magnetization $\vec{M}(x, y, z, t)$, defined by

$$\vec{M}(x, y, z, t) = \sum_{(x', y', z') \epsilon V_0} \vec{\mu}(x', y', z', t) \tag{2.2}$$

where $\vec{M}(x, y, z, t)$ in the above equation is called the total nuclear magnetization in V_0. We assume that the number of nuclei in V_0 is sufficiently large to allow $\vec{M}(x, y, z, t)$ to be a continuous function in space and to be a constant in micro scale. It does not include the effects due to random thermal motion of nuclei and their coherent movements in the volume element V_0. Thus, the magnetization $\vec{M}(x, y, z, t)$ is governed by external field $\vec{H}(x, y, z, t)$

$$\frac{d}{dt}\vec{M}(x, y, z, t) = \gamma \vec{M}(x, y, z, t) \times \vec{H}(x, y, z, t) \tag{2.3}$$

In this equation, we ignore the interactions of the nuclei with the surrounding molecular environment (spin-lattice interaction) and those between the close nuclei (spin-spin interaction).

When $\vec{H}(x, y, z, t) = \vec{H}_0$, the solution of Eq. (2.3) is

$$M_x(t) = M_x(t_0) \cos \omega_0 (t - t_0) + M_y(t_0) \sin \omega_0 (t - t_0) \tag{2.4}$$

$$M_y(t) = M_y(t_0) \cos \omega_0 (t - t_0) - M_x(t_0) \sin \omega_0 (t - t_0) \tag{2.5}$$

$$M_z(t) = M_z(t_0) \tag{2.6}$$

It was found that two relaxation mechanisms are associated with excited nuclear spins, i.e. transverse (spin-spin) relaxation, and longitudinal (spin-lattice) relaxation. The spin-lattice interaction involves the transfer of energy from the spin system to the molecular environment. Since the energy of total spin system only depend on $M_z(x, y, z, t)$. The spin-lattice interaction will contribute to the changes of $M_z(x, y, z, t)$ which relaxes back to the equilibrium value $\vec{M}(x, y, z, \infty)$ exponentially with a time constant T_1 called the spin–lattice relaxation time. The spin–spin interaction does not involve energy transfer. But it will affect $\vec{M}_{xy}(x, y, z, t)$, influence the neighboring spins by changing their Larmor frequency, and will result phase de-coherence in the xy–plane. Also in this case, $\vec{M}_{xy}(x, y, z, t)$ would decrease to zero exponentially with a time constant T_2 called the spin–spin relaxation time in which a central resonance frequency is identical to the Larmor frequency.

[1] For convenience, to define some density functions, V_0 is also referred to as the unit volume element in Section 2.2.3.

2.2.2 Bloch Equation and FID Signal

Based on the above discussion, the whole NMR experiment can be described by Bloch equation [10, 11], given by

$$\frac{d}{dt}\vec{M}(x,y,z,t) = \gamma\vec{M}(x,y,z,t) \times \vec{H}$$
$$- \frac{\vec{M}_{xy}(x,y,z,t)}{T_2(x,y,z)} - \frac{M_z(x,y,z,t) - M_z(x,y,z,\infty)}{T_1(x,y,z)} \tag{2.7}$$

where $\vec{H} = \vec{H}_0(x,y,z,t) + \vec{H}_{RF}(x,y,z,t) + \vec{H}_G(x,y,z,t)$, $\vec{H}_0$ is the static magnetic field, $\vec{H}_{RF}$ is the RF-field, and $\vec{H}_G$ is the gradient field. Bloch equation provides us with a macroscopic model for $\vec{M}(x,y,z,t)$ in a deterministic sense.

In general, Bloch equation is applied to the relaxation process. The solution can be easily obtained [11], and is given by

$$M_x(x,y,z,t) = (M_x(x,y,z,t_0)\cos(\omega_0 t + \phi_0) - M_y(x,y,z,t_0)\sin(\omega_0 t + \phi_0))$$
$$\times \exp(-\frac{t}{T_2(x,y,z)}) \tag{2.8}$$
$$M_y(x,y,z,t) = (M_y(x,y,z,t_0)\cos(\omega_0 t + \phi_0) + M_x(x,y,z,t_0)\sin(\omega_0 t + \phi_0))$$
$$\times \exp(-\frac{t}{T_2(x,y,z)}) \tag{2.9}$$
$$M_z(x,y,z,t) = (M_z(x,y,z,t_0) - M_z(x,y,z,\infty))\exp(-\frac{t}{T_1(x,y,z)}) + M_z(x,y,z,\infty) \tag{2.10}$$

2.2.3 Spatial Encoding and Signal Acquisition

By placing a couple of conductive coils around the system, the relaxation and rotation of $\vec{M}_{xy}(x,y,z,t)$ will induce a complex electric signal in the coils, i.e., FID signal. The spatial encoding procedure for modulating FID signal in MR imaging includes selective excitation, frequency encoding, and phase encoding. Though MRI technique is intrinsically three dimensional, cross-sectional imaging is most routinely performed. The following discussion will focus on this two-dimensional case by assuming the coherent magnetization density is uniformly distributed along z-axis.

A selective excitation process is used to planar images in such a way that only the spins belonging to a selected thin slice will contribute to the signal. This is performed by a selective excitation sequence in which a linear gradient magnetic field along z-axis and an RF pulse lying in the xy-plane are applied such that they satisfy the resonance condition

$$\omega_{RF} = \omega_0 \tag{2.11}$$

where $G_{(\cdot)}$ is the linear gradient function. Thus, a thin slice has been selected and only the spins belonging to this slice will contribute to the FID signals. This procedure is called slice selection.

To obtain an MR image, a physical hardware is needed to encode the contributions of the volume-element magnetization in the slice and a mathematical tool is needed to decode the measured signals. After the slice selection by a 90^0 selective RF pulse, the signal acquisition is performed in the presence of a linear gradient magnetic field along x in the field. The effective external field will be

$$H_z(x) = H_0 + G_x x \tag{2.12}$$

so that the readout Larmor frequency, i.e., the relaxation central frequency, is proportional to the linear gradient magnetic field along x. Clearly, the readout frequency is constant on the straight lines $x = x_0$ and is a linear function of x

$$\omega(x) = \gamma(H_0 + G_x x) \tag{2.13}$$

This procedure is known as frequency encoding and can be generalized later in the so-called Fourier zeugmatography [11].

The Fourier zeugmatography consists of the application, after the slice selection but before the readout encoding, of a linear gradient along y in the field for a time interval t_y such that the phasing field is

$$H_z(y) = H_0 + G_y y \tag{2.14}$$

and the recorded phase follows

$$\phi(y) = \gamma(H_0 + G_y y)t_y \tag{2.15}$$

Similar to the frequency encoding, the recorded phase is constant on the straight lines $y = y_0$ and is a linear function of y. This procedure is also called phase encoding.

By introducing this spatial encoding technique into Eqs. (2.8-2.10), the contributions of the volume-element magnetization in the slice are encoded into a two-dimensional time domain complex FID signal. Let $\rho(x, y)$ denote the spin density at the volume element $V_0(x, y)$ referred to as the unit volume. Reference [11] show that at the equilibrium state, the relationship between the magnetization $M(x, y)$ and the spin density $\rho(x, y)$ is

$$M(x, y) = \rho(x, y)\frac{\gamma^2 h^2 I(I + 1)}{3KT_s}H_0 \tag{2.16}$$

where I is the spin quantum number, h is Planck's constant, K is Boltzmann's constant, and T_s is the absolute temperature. Hence, the received FID signal, through the detecting antenna coils, can be expressed in a form of two-dimensional Fourier integral, given by

$$s(t_x, t_y) = \Delta_1 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \rho(x, y)\exp(-\frac{t_x}{T_2(x, y)} + j\gamma(G_x x t_x + G_y y t_y))dxdy \tag{2.17}$$

where the base frequency determined by H_0 has been removed by hardware simply for increasing the readout resolution. At the equilibrium state, Δ_1 is a constant given by

$$\Delta_1 = \frac{\gamma^3 h^2 I(I + 1)}{3KT_s}QH_0^2 \tag{2.18}$$

where Q is the coil parameter. Note that Δ_1 is proportional to the Larmor frequency γH_0. Signal $s(t_x, t_y)$ is the only measurable quantity in MR imaging, known as the received/detected FID signal. In the following discussion, we simply call $s(t_x, t_y)$ as the FID signal. In practice, we are not directly measuring the FID signal, but its spin-echo signal instead. The technique for doing this, called spin-warp [11] and pulse sequence, will be described in next section.

2.2.4 Pulse Sequences and Signal Modulation

Consider the ensemble of signals $s(t_x, t_y)$, obtained by the experiments. The spin warp is carried out based on [11]: at time $T_E/2$ from the end of the selective pulse, a 180^0 non-selective pulse is performed for a negligible time, where T_E is the echo time. At $t = T_E$, a signal called spin-echo is produced, whose height at $t = T_E$ is proportional to

$$\rho(x, y)\exp(-T_E/T_2(x, y)) \tag{2.19}$$

For $t > T_E$, signal decays because of x gradient, spin-spin interaction, and field inhomogeneities. Shifting the time origin at $t = T_E$, we have

$$s(t_x, t_y) = \Delta_1 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \rho(x, y)\exp(-\frac{T_E}{T_2(x, y)})\exp(-\frac{t_x}{T_2(x, y)} + j\gamma(G_x x t_x + G_y y t_y))dxdy \tag{2.20}$$

It can be seen that the spin warp technique produces a modulated FID signal depending on $\rho(x, y)$ weighted by a function of the transversal relaxation time $T_2(x, y)$.

There are similar techniques, called pulse sequences, which produce a signal depending on $\rho(x,y)$ weighted by a function of $T_1(x,y)$ and/or $T_2(x,y)$. These sequences are very useful, because the modulated FID signal will provide us more information about the tissue characterization. We summarized three major schemes as the follows.

1. Partial Saturation (PS):
$$\rho(x,y)\exp(-\frac{T_E}{T_2(x,y)})(1-\exp(-\frac{T_R}{T_1(x,y)})) \tag{2.21}$$
where T_R is the repetition time.

2. Inversion Recovery (IR):
$$\rho(x,y)\exp(-\frac{T_E}{T_2(x,y)})\{1-(2-\exp(-\frac{T_R-T_I}{T_1(x,y)}))\exp(-\frac{T_I}{T_1(x,y)})\} \tag{2.22}$$
where T_I is the inversion time.

3. Spin Echo (SE):
$$\rho(x,y)\exp(-\frac{T_E}{T_2(x,y)})(1-2\exp(-\frac{T_R-T_I}{T_1(x,y)})+\exp(-\frac{T_R}{T_1(x,y)})) \tag{2.23}$$

As a reasonable approximation [11], since the actual data acquisition is performed only for a time interval that is much shorter than T_2, let $\rho_P(x,y)$ denote the modulated spin density by pulse sequence, we can write Eq. (2.20) in a general form as[2]
$$s(t_x,t_y)=\Delta_1\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\rho_P(x,y)\exp(j\gamma(G_xxt_x+G_yyt_y))dxdy \tag{2.24}$$
where the modulated spin density is determined by the specific scanning pulse sequence that is being used.

2.3 Imaging Mathematics

2.3.1 Imaging Quantity

Based on Eqs. (2.16) and (2.24), instead of using magnetization $\vec{M}(x,y)$ in Bloch equation, by including the effects of signal measurement and pulse sequence modulation, the imaging quantity in the imaging equations can be easily defined as follows:

Definition: *The linear magnetic resonance coefficient in MR imaging is defined by*
$$\beta(x,y)=\Delta_1\rho_P(x,y) \tag{2.25}$$

This definition is consistent with general measure theory and also with the underlying physical phenomena [21]. Note that linear magnetic resonance coefficient is simply a function of spin density, and has a linear relationship with magnetization $\vec{M}(x,y)$.

From Eqs. (2.24-2.25) the linear magnetic resonance coefficient $\beta(x,y)$ can be expressed in terms of inverse Fourier integral onto the FID signal $s(t_x,t_y)$, given by
$$\beta(x,y)=\frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}s(t_x,t_y)\exp(-j\gamma(G_xxt_x+G_yyt_y))d(\gamma G_xt_x)d(\gamma G_yt_y) \tag{2.26}$$

Note that Eq. (2.24) and Eq. (2.26) form a pair of continuous Fourier transform.

[2] When the exact formulation is used, the effect by Lorentzian line-spread function in x direction should be considered [14].

2.3.2 Standard Image Formation Algorithm

An interesting aspect of MR imaging is its diversity in image formation, data collection, and reconstruction. More than a dozen of different MRI methods based on the Fourier Projection Theorem and the Direct Inverse Fourier Transform have appeared recently, each of them has advantages and disadvantages, depending upon various factors such as; imaging time, available field gradient, pulse sequences, signal-to-noise ratio, and artifacts associated with the reconstructed image [10, 11, 14]. The most commonly used image formation algorithm is based on the two-dimensional inverse Fourier transform method [11]. Thus, MR image reconstruction is characterized as a decoding process in which the sequential FID signals are decoded to form an MR image.

Although the FID signals in MR imaging are continuous, the image reconstruction must be implemented in a discrete fashion by computer. We discuss the duel approximation problem involved in this procedure: based on $M \times N$ samples of the spatial-continuous signal $s(t_x, t_y)$, we approximate the spatial-continuous signal $\beta(x, y)$ by applying the rectangular rule of numerical integration.

Let

$$\omega_x = \gamma G_x t_x, \quad \omega_y = \gamma G_y t_y \tag{2.27}$$

We rewrite Eq. (2.24)

$$s(t_x, t_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \beta(x, y) \exp(j(\omega_x x + \omega_y y))dxdy \tag{2.28}$$

and Eq. (2.26)

$$\beta(x, y) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s(t_x, t_y) \exp(-j(\omega_x x + \omega_y y))d\omega_x d\omega_y \tag{2.29}$$

Thus

$$s(t_x, t_y) = \mathcal{F}_2\{\beta(x, y)\}, \quad \beta(x, y) = \mathcal{F}_2^{-1}\{s(t_x, t_y)\} \tag{2.30}$$

Using discrete Fourier transform (DFT), sampling intervals and signal ranges are defined by

$$m\Delta t_x = t_x, \quad n\Delta t_y = t_y \tag{2.31}$$

$$k\Delta x = x, \quad l\Delta y = y \tag{2.32}$$

and

$$1 \le m, k \le M, \quad 1 \le n, l \le N \tag{2.33}$$

we have

$$s(m, n) = \sum_{k=0}^{M-1}\sum_{l=0}^{N-1} \hat{\beta}(k, l) \exp(j(\frac{2\pi mk}{M} + \frac{2\pi nl}{N})) \tag{2.34}$$

and

$$\hat{\beta}(k, l) = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} s(m, n) \exp(-j(\frac{2\pi mk}{M} + \frac{2\pi nl}{N})) \tag{2.35}$$

When

$$\Delta\omega_x = \gamma G_x \Delta t_x, \quad \Delta\omega_y = \gamma G_y \Delta t_y \tag{2.36}$$

and

$$s(m, n) = s(m\Delta\omega_x, n\Delta\omega_y) = s(t_x, t_y)|_{t_x = m\Delta t_x, t_y = n\Delta t_y} \tag{2.37}$$

then, from Eq. (2.24) we can rewrite Eq. (2.34) as

$$\hat{\beta}(k, l) = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \beta(x, y) \exp(j(\Delta\omega_x x + \Delta\omega_y y))dxdy$$

$$\times \exp(-j(\frac{2\pi mk}{M} + \frac{2\pi nl}{N})) \tag{2.38}$$

On the other hand, it is assumed that the time-domain sampling rate satisfies the Nyquist condition in order to avoid the under-sampling aliasing [15]. When only $M \times N$ samples of the estimate $\hat{\beta}(k,l)$ are required at the equally spaced sampling

$$\Delta x = 2\pi/(M\Delta\omega_x) \tag{2.39}$$

$$\Delta y = 2\pi/(N\Delta\omega_y) \tag{2.40}$$

then replacing $\Delta\omega_y$ by $2\pi/(M\Delta x)$, $\Delta\omega_y$ by $2\pi/(N\Delta y)$, and letting $f(k,l) = \hat{\beta}(k,l)$, it is easy to show that Eq. (2.38) can be expressed as

$$f(k,l) = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} s(\frac{2\pi m}{M\Delta x}, \frac{2\pi n}{N\Delta y}) \exp(-j(\frac{2\pi mk}{M} + \frac{2\pi nl}{N})) \tag{2.41}$$

$$= \mathcal{F}^{-1}_{M,N}\{s(\frac{2\pi m}{M\Delta x}, \frac{2\pi n}{N\Delta y})\} \tag{2.42}$$

where $\mathcal{F}^{-1}_{M,N}$ is the two-dimensional inverse DFT that can be implemented by the efficient inverse FFT. Note that here we use an explicit expression for $s(m,n)$ by imposing the Nyquist constraint.

2.3.3 Imaging Equations

The MR imaging equations consist of a pair of equations. The Fourier integral that describes the FID signals is called the projection equation:

$$s(m,n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \beta(x,y) \exp(j\gamma(G_x x\Delta t_x m + G_y y\Delta t_y n))dxdy \tag{2.43}$$

In order to derive the reconstruction equation in terms of system point spread function (PSF), we need to know how $f(k,l)$ is related to $\beta(x,y)$. Simply following Eq. (2.41) by substituting discrete version of Eq. (2.24),we have

$$f(k,l) = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} s(\frac{2\pi m}{M\Delta x}, \frac{2\pi n}{N\Delta y}) \exp(-j(\frac{2\pi mk}{M} + \frac{2\pi nl}{N})) \tag{2.44}$$

$$= \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \beta(u,v)\exp(j(\frac{2\pi mu}{M\Delta x} + \frac{2\pi nv}{N\Delta y}))dudv \exp(-j(\frac{2\pi mk}{M} + \frac{2\pi nl}{N})) \tag{2.45}$$

$$= \frac{1}{MN}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \beta(u,v)dudv \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} \exp(-j(\frac{2\pi m(k-u/\Delta x)}{M} + \frac{2\pi n(l-v/\Delta y)}{N})) \tag{2.46}$$

$$= \frac{1}{MN}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \beta(u,v)\frac{1-\exp(-j(2\pi(k\Delta x-u)/\Delta x + 2\pi(l\Delta y-v)/\Delta y)}{1-\exp(-j(2\pi(k\Delta x-u)/(M\Delta x) + 2\pi(l\Delta y-v)/(N\Delta y))}dudv \tag{2.47}$$

$$= \frac{1}{MN}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \beta(u,v)\frac{\sin(\pi(k\Delta x-u)/\Delta x)\sin(\pi(l\Delta y-v)/\Delta y)}{\sin(\pi(k\Delta x-u)/(M\Delta x))\sin(\pi(l\Delta y-v)/(N\Delta y))}$$

$$\times \exp(-j(\frac{\pi(k\Delta x-u)(M-1)}{M\Delta x} + \frac{\pi(l\Delta y-v)(N-1)}{N\Delta y}))dudv \tag{2.48}$$

By substituting the original parameters into the equation, we can write the close form of the reconstruction equation as:

$$f(k,l) = \frac{1}{MN}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \beta(u,v)h(k-\frac{u}{\Delta x}, l-\frac{v}{\Delta y})dudv \tag{2.49}$$

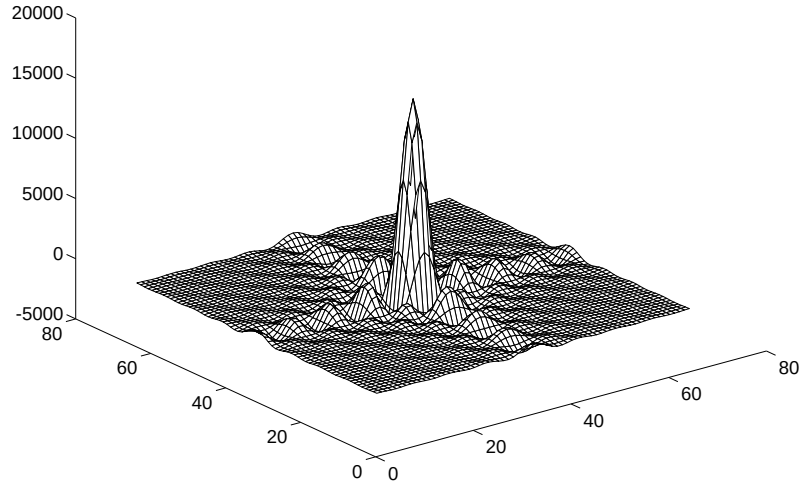

Figure 2.1: Point Spread Function (PSF) of MR Imaging System.

where $h(k-\frac{u}{\Delta x}, l-\frac{v}{\Delta y})$ is the system PSF given by:

$$h(u,v) = \frac{\sin(\frac{M}{2}\gamma G_x u \Delta t_x)}{\sin(\frac{1}{2}\gamma G_x u \Delta t_x)} \frac{\sin(\frac{N}{2}\gamma G_y v \Delta t_y)}{\sin(\frac{1}{2}\gamma G_y v \Delta t_y)} e^{-j\gamma(\frac{M-1}{2}G_x u \Delta t_x + \frac{N-1}{2}G_y v \Delta t_y)} \tag{2.50}$$

Figure 2.1 shows a 2D plot of $h(u,v)$. It can be seen that the estimated linear magnetic resonance coefficients $f(k,l)$ is an *interpolated* version of true function $\beta(u,v)$ in which the approximation is obtained by convolving the true function $\beta(u,v)$ with the system PSF [13]. In this theoretic derivation, a *discrete* rectangular *window function* $W(m,n)$ is introduced in the time domain

$$W(m,n) = \begin{cases} 1 & (m < M, n < N) \\ 0 & (m \geq M, n \geq N) \end{cases} \tag{2.51}$$

Since the inverse transform of this window function, i.e., system PSF (as an interpolating function), has a quite strong oscillating side-lobes, the resultant estimates $f(k,l)$ have significant overlap or aliasing [12]. This is called Gibbs phenomenon [15]. Therefore, in practical implementation, a different window function is chosen to reduce greatly the amplitude of these oscillations [13]. Reference [15] shows that the basic requirements for window function are

1. $W(m,n)$ is non-negative, real, and even function;
2. $W(0,0) = 1$ in order to assure a asymptotic unbiased estimation;
3. $h(u,v)$ is strictly positive definite such that $f(k,l) \geq 0$ being physically meaningful.

In MR image reconstruction, we assume these conditions are satisfied. A possible candidate of appropriate window function is proposed by [13] as

$$W(m,n) = \begin{cases} 1 - \epsilon\frac{|(2m-M)(2n-N)|}{MN} & (m < M, n < N) \\ 0 & (m \geq M, n \geq N) \end{cases} \tag{2.52}$$

where parameter ϵ takes values in the range [0, 1]. For simplicity, let $\epsilon = 1$ (*Bartlett Window*), we obtain a better system PSF [15]

$$h(u,v) = \frac{\sin^2(\frac{M}{2}\gamma G_x u \Delta t_x)}{\sin^2(\frac{1}{2}\gamma G_x u \Delta t_x)} \frac{\sin^2(\frac{N}{2}\gamma G_y v \Delta t_y)}{\sin^2(\frac{1}{2}\gamma G_y v \Delta t_y)} \tag{2.53}$$

Note that references [11, 14] have come to basically the similar formulations of system PSF using 2D Fourier reconstruction algorithm, called Fourier zeugmatography and Lorentzian line-spread function. However, their theoretic derivation is based on the assumption that an infinite length FID signal is available or a continuous Fourier transform can be implemented. Clearly, the actual situation does not satisfy these assumptions.

2.4 Summary

In this chapter, we discuss Bloch equation, spatial encoding methods, pulse sequences techniques, and image reconstruction algorithms. In our justifications and derivations, the physics and signal processing in MR imaging find their common point in a unified way. Specifically, MR imaging equations (Eq. (2.43) and Eq. (2.49)) provide us a complete description of the forward and inverse transformations in the imaging process, in a deterministic sense. Furthermore, based on more complete system PSF derived here, we show that the resultant MR image is a discrete and interpolated representation of the linear magnetic resonance coefficients. In fact, the properties of system PSF are not only satisfies the requirements of digital signal processing, but also related to image quality assessment and statistical justification of imaging process ergodicity. Based on the review of MR imaging in this chapter, a statistical investigation is conducted in the following chapters to obtain the knowledge of MR imaging process in a statistical sense.

Chapter 3

MR Image Statistics

3.1 Introduction

As discussed in Section 1.2.1, although some pioneer work has been reported on the statistical analysis for x-ray CT [17], positron emission tomography [29], and ultrasound scan [23], there has been little work on MR imaging statistics and model-based MR image analysis. For example, [20] discusses noise sources and their effects in MR imaging, [25] presents a unified SNR analysis of medical imaging systems, [76] apply information theory to the quality evaluation of MR imaging, and [22] compares several currently used statistical algorithms in MR imaging, but none of them provides a complete statistical analysis that can incorporate MR imaging statistics into MR image modeling by justifying many heuristic assumptions typically made in the literature [76]. In this chapter, we present a full investigation of the stochastic aspects of MR imaging and discuss its applications in MR image modeling. A complete statistical description of MR imaging, including the imaging equation and the random field theory, is provided. Both object variability and thermal noise are considered as degradations presented in the pixel images which are generated by Fourier transform reconstruction algorithm. Gaussianity, dependence, stationarity, and ergodicity of MR pixel images are characterized as the standard problems of statistics. Their justification are given in order to form the basis of the stochastic image modeling and analysis.

In the context of stochastic image modeling and regularization, the statistical properties of both pixel and context images are very important and have to be considered together. For pixel images in MR or other modalities, many investigations on image statistics have been conducted [24, 42, 61, 76]. Reference [20] has conducted an intensive research on the objective assessment of imaging statistics of x-ray and gamma ray, by considering the effects of both quantum noise and object variability. In [23], a pioneer work is reported on the speckle statistics in medical ultrasound images in which the first order and the second order statistics have received intensive treatment. References [25, 30] study the intrinsic signal-to-noise ratio and noise power spectrum for MR imaging, both object variability and thermal noise are considered and evaluated theoretically. Reference [22] conducts a comparative work for evaluating several statistical approaches in MR imaging, in which intensive real MR data analysis is included; reference [76] aides to measure the bivariant information content of MR images in which six assumptions are made and partially justified.

Furthermore, the statistics of context image is another challenging topic. Using randomization rule and stochastic regularization, many investigations have been conducted [40, 41, 42, 45]. In spite of some impressive results shown in literature, most of the techniques proposed previously exhibit at least one of the following drawbacks. For example, constraints on context representation are generally imposed mathematically without objective justification by the context image statistics [45, 61]; since the true context is unobservable in general, the assumed models always contain some empirical parameters that are given heuristically [24, 32, 45]; homogeneous assumption is applied to the context image even though it obviously is not true [18, 67].

In our MR statistical investigation, the following issues are addressed:

1. MR image statistics with the typical MR data samples through theoretical and experimental studies;
2. MR image statistics in the context image;
3. Relationship between the MR image statistics and the MR image models.

3.2 Noise Consideration

3.2.1 Object Variability

Governed by quantum thermodynamics and balanced metabolism, the nuclei system will approach a dynamical equilibrium state in macro-scale and only coherent magnetization contributes to the NMR signal, i.e., linear magnetic resonance coefficients. The actual micro-scene is that, even in a vary small volume, there are millions of proton spins randomly flipping, determined by the spin coherence [25], and randomly moving, controlled by the spin fluctuation.

Reference [25] shows that the statistical behavior of the populations of spins, e.g., spin density of Eq. (2.16) in V_0, may be calculated by considering the total sample of N spins to hold a binomial distribution with parameter κ, where κ is the probability that a spin is at low energy level. Then spin density at two energy levels will have the variance

$$\sigma_{SC}^2 = N\kappa(1-\kappa) \tag{3.1}$$

When N approaches infinity, spin density has an asymptotically Gaussian distribution [21]. Since spin density cannot be negative or infinity, the approximated Gaussian distribution of spin density must be limited in a positive finite interval determined by physical settings of the MR imaging system. We define this intrinsic randomness in spin density as *object variability*, i.e., the intrinsic noise associated with the origin of the NMR signal. It is believed that the object variability in MR imaging is mainly due to spin statistics [11, 16].

We will also investigate second-order statistics of the spin system, which mainly describes the microscopic correlation between two single spins as well as two spin densities. We first consider the case of two spin correlations. Reference [26] shows that spin coherent statistics has two important properties:

a) the correlation between two spins can be shown to decrease exponentially as a function of inter-spin distance

$$R(n) = \sin^2(2\phi)\exp[-n\ln(\lambda_+/\lambda_-)] \tag{3.2}$$

where $\lambda_{(.)}$ are the real eigenvalues of the Hermitian matrix in the spin system, λ_+ and λ_- denote the largest and the smallest eigenvalues, ϕ is the angle between two spin vectors, and n is the separation number of lattices with the units of the lattice spacing. Note that the order of lattice spacing is *much less* than the dimension of element volume V_0 [1].

b) two spins on the lattice chain are statistically independent if the inter-spin distance is relative large (e.g., greater or equal to the dimension of element volume) since the correlation is expected to decrease *rapidly* as the distance increases.

These important observations are violated only when two spins are close to a critical point [11, 26] which is not within the scope of our consideration. Now we extend these properties to the case of correlation between two spin densities. As discussed before, spin density $\rho(x,y)$ is defined on element volume and obtained by the summation over all spins within this element volume. Since only linear operation is involved, from properties a) and b), it is reasonable to have the extensions that:

c) Covariance between two spin densities, $\rho(x,y)$ and $\rho(u,v)$, can be shown to decrease exponentially as a function of the separation $||(x,y),(u,v)||$ between two volume elements

$$K_\rho((x,y),(u,v)) = \sigma_{\rho(x,y)}\sigma_{\rho(u,v)}\exp(-\frac{||(x,y),(u,v)||}{\Delta r}) \tag{3.3}$$

[1] In spatial dimension, Lattice spacing has an order of 10^{-10} m, and element volume has an order of 10^{-8} m.

where $\sigma_{\rho(.,.)}$ is the standard deviation of $\rho(.,.)$, and Δr is defined as the effective correlation distance that has an order of 10^{-8} m.

d) Two spin densities are statistically independent if their spatial separation is relatively large than the dimension of element volume in all directions.

In MR imaging, linear magnetic resonance coefficient $\beta(x,y)$ is a deterministic function of spin density $\rho(x,y)$ (Eq. (2.25)), by the above reasoning, we can well *approximate* the covariance function of $\beta(x,y)$ in the scale of volume element by an exponential function:

$$K_\beta((x,y),(u,v)) = E[(\beta(x,y)-\mu(x,y))(\beta(u,v)-\mu(u,v))]$$

$$= \sigma_{\beta(x,y)}\sigma_{\beta(u,v)} \exp(-\frac{||(x,y),(u,v)||}{\Delta r}) \tag{3.4}$$

where E denotes the operation of expectation, $\mu(.,.)$ is the mean of the linear magnetic resonance coefficient, and $\sigma_{\beta(.,.)}$ is the standard deviation of $\beta(.,.)$. Note that although the $\sigma_{\rho(.,.)}$ or $\sigma_{\beta(.,.)}$ may have different *finite* values for either auto- or cross- covariance the envelop function $\exp(.)$ is shift-invariant. This important extension is supported by real data analysis in [76].

In a summary, this section shows that spin density $\rho(x,y)$ is a truncated asymptotic Gaussian. From Eq. (2.25), it is straightforward to conclude that linear magnetic resonance coefficient $\beta(x,y)$ also has a truncated asymptotic Gaussian density. Furthermore, due to the uniqueness of tissue characterization [16, 17], within a homogeneous object, all linear magnetic resonance coefficients should contribute equally likely. In other words, the mean $E[\beta(x,y)]$ and the variance $E[(\beta(x,y)-E[\beta(x,y)])^2]$ must be the same for all (x,y). Thus, the covariance function of $\beta(x,y)$ within that region can be simplified as

$$K_\beta((x,y),(u,v)) = \sigma^2 \exp(-\frac{||(x,y),(u,v)||}{\Delta r}) \tag{3.5}$$

where σ^2 is the variance of the linear magnetic resonance coefficient[2].

3.2.2 Thermal Noise

During data acquisition, another dominant random source enters the NMR signal, this is the thermal noise. This random character is mainly due to the thermal current in the body and the detection system [25]. More precisely, the source of the noise is thermally generated, by randomly fluctuating noise currents in the body which are picked up by the receiving antenna. This noise is inevitable and can not be eliminated by any antenna design [30]. In addition, photon noise in a communication channel also contributes to the thermal noise. Compared to object variability, thermal noise enters NMR data in the time domain with a discrete fashion, and forms statistically independent random source from signal. Moreover, since in double-balanced mixer with phase detectors thermal noise takes on complex values, its real and imaginary parts are statistically independent.

Many previous work (e.g. [25, 30]) has shown that thermal noise is white with zero mean and a constant variance and can be well characterized by a Gaussian random process. According to the Johnson noise formula [10], the variance of thermal noise per measurement can be expressed as

$$\sigma_{TN}^2 = 4kTR_e\Delta f_t F \tag{3.6}$$

where k is Boltzmann's constant, T is the absolute temperature, R_e is the effective resistance of the receiving system, Δf_t is the bandwidth of the communication receiver in Hertz, and F is the noise figure of the front end amplifier. Therefore, from Eqs. (2.18) and (3.7), the associated signal-to-noise ratio in the time domain can be written as

$$\frac{\Delta_1}{\sigma_{TN}^2} = \frac{\gamma^3 h^2 I(I+1)QH_0^2}{12k^2T^2R_e\Delta f_t F} \tag{3.7}$$

[2] For simplicity, σ_β^2 is simply written as σ^2.

The property of thermal noise in MR imaging is quite different from other imaging modalities which use ionizing radiation, such as x-ray CT or emission tomography, where the noise in any volume element is proportional to the square root of the signal. In NMR signal the thermal noise is constant. Reference [25] shows that the randomness in MR imaging derives primarily from thermal noise while object variability is negligible in general cases, but we will consider both object variability and thermal noise since we are interested in the scale of volume element.

3.2.3 Pixel Image

This subsection investigates how these two randomness sources, object variability and thermal noise, contribute to the statistics of the final reconstructed pixel images.

Let $s_n(m,n)$ denote the thermal noise. Rewriting MR imaging equations in both time and spatial domain, i.e., raw data/projection domain and pixel images/reconstruction domain, by

$$\hat{s}(m,n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \beta(x,y)\exp(j\gamma(G_x x\Delta t_x m + G_y y\Delta t_y n))dxdy + s_n(m,n) \tag{3.8}$$

Eq. (2.49) becomes

$$\begin{aligned}\hat{f}(k,l) &= \frac{1}{MN}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\beta(u,v)h(k-\frac{u}{\Delta x},l-\frac{v}{\Delta y})dudv \\ &+ \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} s_n(m,n)\exp(-j(\frac{2\pi mk}{M}+\frac{2\pi nl}{N}))\end{aligned} \tag{3.9}$$

It is clear that each pixel image contains two parts, the first one is sampled from the convolution of the linear magnetic resonance coefficient $\beta(u,v)$ with the system PSF (first term of (3.9)), the second one consists of a series of linear orthogonal transformations (IDFT) applied onto the original thermal noises $s_n(m,n)$ (second term of (3.9)).

Define a new random field for thermal noise in the spatial domain by

$$f_n(k,l) = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} s_n(m,n)\exp(-j(\frac{2\pi mk}{M}+\frac{2\pi nl}{N})) \tag{3.10}$$

Then, since the thermal noise $s_n(m,n)$ is white with zero mean

$$E[s_n(m,n)s_n(m',n')] = \sigma_{s_n}^2\delta(m-m',n-n') \tag{3.11}$$

$$E[s_n(m,n)] = 0 \tag{3.12}$$

Thus, the mean and autocorrelation function of this new random field are:

$$E[f_n(k,l)] = 0 \tag{3.13}$$

$$R_{f_n}((k,l),(k',l')) = E[f_n(k,l)f_n^*(k',l')] \tag{3.14}$$

$$\begin{aligned}&= E[\frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} s_n(m,n)\exp(-j(\frac{2\pi mk}{M}+\frac{2\pi nl}{N})) \\ &\times \frac{1}{MN}\sum_{m'=0}^{M-1}\sum_{n'=0}^{N-1} s_n(m',n')\exp(j(\frac{2\pi m'k'}{M}+\frac{2\pi n'l'}{N}))]\end{aligned} \tag{3.15}$$

$$= \frac{1}{M^2N^2}\sum_{m=0}^{M-1}\sum_{m'=0}^{M-1}\sum_{n=0}^{N-1}\sum_{n'=0}^{N-1} E[s_n(m,n)s_n(m',n')]$$

$$\times \exp(j(\frac{2\pi(m'k'-mk)}{M}+\frac{2\pi(n'l'-nl)}{N})) \tag{3.16}$$

$$= \frac{1}{M^2N^2}\sum_{m=0}^{M-1}\sum_{m'=0}^{M-1}\sum_{n=0}^{N-1}\sum_{n'=0}^{N-1}\sigma_{s_n}^2\delta(m-m',n-n')$$

$$\times \exp(j(\frac{2\pi(m'k'-mk)}{M}+\frac{2\pi(n'l'-nl)}{N})) \tag{3.17}$$

$$= \frac{\sigma_{s_n}^2}{M^2N^2}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\exp(j(\frac{2\pi m(k'-k)}{M}+\frac{2\pi n(l'-l)}{N})) \tag{3.18}$$

$$= \frac{\sigma_{s_n}^2}{MN}\delta(k'-k,l'-l) \tag{3.19}$$

Therefore, a property of pixel image related to the thermal noise can be stated as

Property 3.2.3: *Because of the orthogonality of the inverse discrete Fourier transform used in MR image reconstruction, the thermal noise component in the final pixel images is white with zero mean and has a constant variance.*

3.3 Statistical Properties of Pixel Images

3.3.1 Gaussianity

From Eqs. (2.47) and (3.9), pixel image $\hat{f}(k,l)$ is a weighted sum of large number of $(M\times N)$ random variables (FID signal $s(m,n)$ and thermal noise $s_n(m,n)$). Both FID signal and thermal noise have finite means and variances, thus, according to generalized central limit theorem, pixel image $\hat{f}(k,l)$ has an asymptotic Gaussian distribution. Due to the physical limitation, this Gaussian is truncated. Therefore, Gaussianity of the final MR pixel images can be summarized by the following statement:

Property 3.3.1: *Given an MR image, each pixel is a random variable with a truncated asymptotic Gaussian distribution, and the whole image is a Gaussian random field.*

This conclusion is strongly supported by the real MRI data analysis. Reference [22] has shown that, based on a number of experiments in which a single set of measurement of a slice is compared with an average image made from eight sets of measurements of the same slice, the results from the real (or the imaginary) part of the reconstructed pixel image is influenced by the noise following a Gaussian distribution with zero mean. The histogram of the noise for the real part of the reconstructed data clearly show that, even in the tails, a Gaussian distribution gives a very good fit. Furthermore, the modulus image will have a Rayleigh distribution. Rayleigh density approaches to a truncated Gaussian when the SNR is relatively high [40]. Reference [22] also shows that noise at each pixel is approximately following a Gaussian distribution in the modulus images.

3.3.2 Dependence

According to second-order moment input/output relations of linear systems [21], the covariance function of the convolved linear magnetic resonance coefficients can be expressed as

$$K_f((k,l),(k',l')) = K_\beta((k,l),(k',l'))\odot h^*((k,l),(k',l'))\odot h((k,l),(k',l')) \tag{3.20}$$

where $\odot$ denotes continuous linear convolution, and $K_\beta(.,.)$ is the covariance function of the true linear magnetic resonance coefficients.

Similar to Eq. (3.3), when the distance between two pixel images is denoted by $\Delta d = ||(k,l),(k',l')||$, we have

$$K_f((k,l),(k',l')) = \sigma(k,l)\sigma(k',l')\exp(-\frac{\Delta d}{\Delta r})\odot h^*((k,l),(k',l'))\odot h((k,l),(k',l'))$$

$$= \sigma(k,l)\sigma(k',l') \exp(-\frac{\Delta d}{\Delta r}) \odot h^*(\Delta d) \odot h(-\Delta d) \tag{3.21}$$

Thus, it is easy to show that when Δd approaches to infinity, the covariance of $f(.,.)$ goes to zero since the first convolution function in Eq. (3.21) is narrow ranged. This property is called asymptotically independent or weakly dependent. In our case, it is

$$\lim_{||(k,l),(k',l')|| \to \infty} K_f((k,l),(k',l')) = 0. \tag{3.22}$$

Furthermore, the dependence between any two pixel images can also be evaluated by the covariance function $K_{\hat{f}}((k,l),(k',l'))$ given by

$$K_{\hat{f}}((k,l),(k',l')) = K_f((k,l),(k',l')) + K_{f_n}((k,l),(k',l'))$$

$$= K_f((k,l),(k',l')) + \frac{\sigma_{s_n}^2}{MN}\delta(k'-k,l'-l) \tag{3.23}$$

Thus, the dependence in MR pixel images can be stated as follows:

Property 3.3.2: *Any two pixel image rv's in an MR image are asymptotically independent, that is, weakly dependent. Their correlation is mainly governed by the system point spread function under the condition of narrow ranged microscopic correlation.*

The importance of (3.23) derives from the fact that correlation between any two pixels in a MR image is only determined by the correlation from object variability part in which both microscopic correlation and system point spread function make contributions. Property 3.2.2 is strongly confirmed by real MRI data analysis. References [22, 76] have reported that, based on their experiments, pixel images seem to be uncorrelated since the plot of the correlation function of MR image shows a *rapidly* decreasing with the inter-pixel distance increasing.

3.3.3 Stationarity

After investigating the first and second order statistics of MR pixel images, we are ready to explore the other two important properties: that is, stationarity and ergodicity. As indicated in Section 3.2.1, a homogeneous object in MR image has a unique linear magnetic resonance coefficient $\beta(x,y)$, that is, the unique mean and variance

$$E[\beta(x,y)] = \mu, \quad E[(\beta(x,y)-\mu)^2] = \sigma^2. \tag{3.24}$$

From Eqs. (3.9) and (3.10), it is easy to show that

$$E[\hat{f}(k,l)] = C_1, \quad E[(\hat{f}(k,l)-C_1)^2] = C_2 \tag{3.25}$$

where C_1 and C_2 are constant mean and variance of pixel image $\hat{f}(x,y)$. Note that for different homogeneous object in image, $C_{(.)}$ is determined uniquely by the mean and variance of underlying linear magnetic resonance coefficient $\beta(x,y)$ of that object. We define a meaningful image region (or simply say, region) mathematically as follows:

Definition: *A group of pixel images are said to form a region in an MR image if they have the same means and variances.*

Thus, based on above definition, an image region in the final reconstructed MR image will correspond to a homogeneous object.

Then within an image region, we can simply rewrite Eq. (3.21) as

$$K_f(\Delta d) = \sigma^2 \exp(-\Delta d) \odot h^*(\Delta d) \odot h(-\Delta d) \tag{3.26}$$

Therefore, the autocorrelation function of $\hat{f}(i,j)$ will be

$$R_{\hat{f}}((k,l),(k',l')) = E[\hat{f}(k,l)\hat{f}(k',l')]$$

$$= \begin{cases} \sigma^2 \exp(-\Delta d) \odot h^*(\Delta d) \odot h(-\Delta d) + C_1^2 & (k \neq k', l \neq l') \\ C_2 + C_1^2 & (k = k', l = l') \end{cases} \tag{3.27}$$

Since correlation only depends on the spatial index difference, according to reference [21] $\hat{f}(i,j)$ is a stationary field in the wide-sense, also, in the strict-sense. By assuming that an MR image contains several distinct regions, we state stationarity as follows:

Property 3.3.3: *Given an MR image, each region is stationary. The whole image is piecewise stationary.*

3.3.4 Ergodicity

For a given region in an MR image, we can rewrite Eq. (3.23) as

$$K_{\hat{f}}((k,l),(k',l')) = \begin{cases} \sigma^2 \exp(-\Delta d) \odot h^*(\Delta d) \odot h(-\Delta d) & (k \neq k', l \neq l') \\ C_2 & (k = k', l = l') \end{cases} \tag{3.28}$$

which indicates that $K_{\hat{f}}(0) < \infty$, and $K_{\hat{f}}((k,l),(k',l')) \to 0$ when $\Delta d \to \infty$. Therefore, $\hat{f}(k,l)$ has a mean ergodic theorem with limiting sample average C_1

$$\lim_{M,N\to\infty} \frac{1}{MN} \sum_{k=0}^{M} \sum_{l=0}^{N} \hat{f}(k,l) = C_1 \tag{3.29}$$

with probability one [21] [Theorem 7.1.]. Note that the variances of both thermal noise and object variability are assumed to be finite.

Then, we directly use Theorem 7.3. (Birkhoff-Khinchin) in reference [21] to show that $\hat{f}(k,l)$ has a variance ergodic theorem. Reference [21] shows that a stationary Gaussian random field is ergodic if it has strictly positive definite covariance function $K_{\hat{f}}$ such that $K_{\hat{f}}((k,l),(k',l'))$ is finite for all Δd and approaches 0 as $\Delta d \to \infty$ with probability one. Therefore, according to Theorem 7.3 in [21],

$$\lim_{M,N\to\infty} \frac{1}{MN} \sum_{k=0}^{M} \sum_{l=0}^{N} [\hat{f}(k,l) - C_1]^2 = C_2 \tag{3.30}$$

A detailed mathematical proof for independent case can be found in [17]. The above discussion can be summarized by the following property:

Property 3.3.4: *An MR image is a piecewise ergodic random field, and each region satisfies both mean and variance ergodic theorems.*

The importance of this property is that in unsupervised image analysis the quantification of mean and variance for each region is necessary. With ergodic theorems, the spatial averages of pixel images can be performed to estimate these quantities. Thus, only one image, i.e., only one realization of random field, is required.

As a summary, in Section 3.3, we discussed the effects of two major random sources involved in MR imaging, derived several statistical properties of pixel images, and defined some important concepts. The conclusions we have obtained are strongly supported by the real MR data analysis [22, 76]. Thus, based on the pixel image statistics, we justified main heuristic assumptions on MR image that have been made in the literature. This work provides a basis for further image modeling, and its application to pixel image modeling will be discussed in Section 4.2.

3.4 Statistical Properties of Context Image

3.4.1 Stochastic Regularization and Markovian Property

As briefly defined in Chapter 1, each pixel is described in terms of pixel image and context image, where the pixel image refers to its intensity (gray level), and the context image refers to its label (membership). In Section 3.3, we define an image region as a group of pixel images which have the same mean and variance and justify the stationarity property of image regions. This investigation is based on the moments of the intensities of pixels. We also show that an image region in MR image corresponds to a homogeneous object, and the whole image may contain several regions. Thus, there is a one-to-one correspondence between pixel label and image region, i.e., homogeneous object. In order to have a complete characterization of MR image statistics, we need to investigate the statistical properties of context image.

Another motivation for studying the statistics of context images in this research is to provide a basis for searching a mathematical tool which can translate statement about the context information, in both local and global scales, into model structures and parameter values, in a traceable format. As discussed in many references [42, 45], although true context is unobservable, a priori expectations, specific scene knowledge, and contextual clues can help us to eliminate possible ambiguities and recover missing information so as to perceive the scene "correctly". In MR imaging, nearby locations typically have similar tissue types; tones, although having locally slow variations, represent homogeneous regions; boundaries are usually smooth and persistent; and objects, such as tissue and organ, have preferred relations and orientations. These "regularities" are rarely deterministic, rather, they describe the correlations and likelihoods of possible outcomes in the real scene. These kind of knowledge can be captured mathematically and exploited in a stochastic framework. The procedure for formulating this framework is called stochastic regularization, and will be further explored in Section 4.3.

In MR image, regions are piecewise contiguous, thus, pixel label takes on discrete values and the labels of nearby pixels are strongly correlated. This correlation is primarily local and can be represented by Markovian property. Markovian property describes that the probability distribution of a label, given all other pixel labels, only depends on its neighboring pixel labels. Mathematically, the concept of Markovian property can be described by the following two local properties:

1. (Positivity) $P(L_i = l_i) > 0$;
2. (Markov Property) $P(L_i = l_i|\mathbf{l}) = P(L_i = l_i|\mathbf{l}_{\partial i})$.

where l_i is the label of pixel i, and ∂i denotes the neighborhood of pixel i. This natural constraint can be also justified based on the mechanism of biological development [16].

3.4.2 Spatial Continuity and Context Representation

Local context images can be described by Markovian property. Global context images are generally reflected by image regions and boundaries. Within one region, the context exhibits a strong spatial continuity, while on region boundaries, the context reflects the natural discontinuity. The level of spatial continuity in local context images will be measured by the dependence (or correlation) among pixel labels, and represented by the parameter values that can reflect the measure. This consideration will be further explained in Sections 5.5.

We propose a simple randomization rule to mathematically describe the statistical properties of the local context images.

For each pixel i, in a second order neighborhood system[3], after randomly reordering its eight neighboring pixel labels without specifying orientation, we can calculate the conditional histogram of local context images. Since this conditional histogram is a probability measure, we can define a meaningful probability space for

[3] Given a central pixel, second order neighborhood system refers to its eight neighboring pixels [45]. The mathematical description is presented in Section 4.3

the dependence among local context images based on this conditional probability measure that determines the probability of the context image of the central pixel when the context images of its neighboring pixels are given.

Similarly, by randomly reordering all pixels in the whole image without specifying orientation, we can calculate the histogram of global context images. Clearly, the global histogram is also a probability measure. We define a probability space based on this probability measure and refer it to as the global context information. This multinomial probability distribution determines the unconditional probability of any context image in the whole image, which will be further discussed in Section 4.2.3.

3.5 Summary

In this chapter, a full investigation on MR image statistics is presented. We study the statistical background of two major random sources in MR imaging: object variability and thermal noise, and verify their contributions to statistical properties of pixel images. We showed that the pixel image generated by two dimensional DFT (image reconstruction) algorithm can be characterized by an asymptotic Gaussian distribution. We justify the piece-wise stationarity and the asymptotic independence of pixel images, and show that each region satisfies mean and variance ergodic theorems. We also demonstrate that these important properties provide a basis for establishing the stochastic models for pixel images, and hence greatly helping the statistical model based MR image analysis.

By discussing the stochastic regularities exhibited by context images, we introduce Markovian property and provide its mathematical description. We demonstrate that the intrinsic dependence (or correlation) among context images can be characterized by Markovian property. We propose a simple randomization rule to mathematically realize the statistical treatment of context images, and show that the underlying probability space in global and local scales are well defined. Therefore, this study provides a good mathematical interpretation for the stochastic regularization of context images.

Within a statistical framework, ergodicity represents the relationship between one realization and its ensemble. A semi-formal definition of ergodicity is that, under some regularization conditions, every realization of one random field holds some common statistical properties of the ensemble. Conceptually, the ergodic theorems and the ergodicity are not the same, in practice, however, we are always interested in the ergodic theorems because they can be directly used in signal detection and estimation. The key concept we used in our justification for ergodic theorems is *strongly mixing* condition based on the property of asymptotic independence.

Chapter 4

Stochastic Models for MR Image

4.1 Introduction

Objective of stochastic modeling in image analysis is to capture the intrinsic character of images in a structure with a few parameters so as to understand the statistical nature of the imaging process. Stochastic image models are useful to quantitatively specify natural constraints and general assumptions in the sense of statistics about the physical world and the imaging process. They play a crucial role in unsupervised[1] MR image analysis, such as; parameter estimation (image quantification) and image segmentation (pixel classification).

Two different types of stochastic image models are required in practical applications [40]:

1. *Models for Pixel Images*: The models for pixel images are specified to describe the type of randomness involved in the observed pixel random variables. Many stochastic models have been proposed to analyze the tone or texture pixel images, such as; the conditional FNM (CFNM) [60], the standard FNM (SFNM) [31, 62, 61], the generalized Gaussian mixture (GGM) [61], the Gaussian random field (GRF) [24], the autoregressive (AR) model [44], and the Markov/Gibbs random field (MRF/GRF) [45].

2. *Models for Context Images*: The models for the underlying true context images are designed to serve as prior contextual constraints on unobserved pixel labels in terms of stochastic regularization [47]. The MRF model has been the most popular so far in this domain although it has a number of fundamental issues still unexplored and unanswered [45].

We can summarize important and unsolved problems in statistical image modeling as follows [41, 42, 45, 47, 60, 61, 63, 67, 70], the main unsolved issues are as follows:

1. General assumptions about image statistics are made without justification;

2. Connections between image statistics and stochastic image models are not well discussed;

3. Specification of stochastic models for different tasks in image analysis has not been addressed;

4. The mathematical and physical aspects of model formulations lack clear interpretation;

5. Unsupervised model-based image analysis is vastly inferior to natural biological solution [42].

Similar to a recent work [17], in Chapter 3, we discussed the statistical properties of MR image. Rather than making heuristic assumptions on image statistics, these statistical properties can be utilized to establish a proper framework of stochastic modeling for MR images. For convenience in the following discussions, the main results derived in Chapter 3 are quoted below. For an MR image:

[1] In this dissertation, *unsupervised* means that there is not any prior knowledge about one specific observed MR image.

1. Each pixel is a random variable with a truncated, asymptotic Gaussian distribution. The whole image is a Gaussian random field.
2. Any two pixel image rv's are asymptotically independent (i.e., weakly dependent). Their correlation is mainly governed by the system point spread function with a narrow ranged microscopic correlation [2].
3. Each region is stationary. The whole image is piecewise stationary.
4. Each region satisfies both mean and variance ergodic theorems. The whole image is an ergodic random field.
5. Context images have multinomial distributions, they are correlated and satisfy Markovian property.

In this chapter, based on these properties, we present a new framework for stochastic MR image modeling and provide a better mathematical understanding of the related issues. Since recently there has been considerable interest in using FNM (for pixel images) and MRF (for context images) models for medical image analysis, and impressive application results using these models have been presented [60, 61, 62], our discussions mainly focus on these two kinds of models and show that this framework results in efficient algorithms for image analysis. Figure 4.1 presents a block diagram that describes the main stochastic models proposed for pixel and context images and their applications to the development of related algorithms.

4.2 Finite Normal Mixture Modeling

4.2.1 Conditional Finite Normal Mixture

Based on the image statistics of computed medical imaging modalities, two kinds of FNM models have been used to describe the tone images which include x-ray CT and MRI [18, 61, 62]. Assume an MR image with N^2 pixels and K regions, then the probability density function (pdf) for each pixel image has a *conditional*[3] Gaussian density, *given its true pixel label*:

$$p(x_i|l_i^*) = \prod_{k=1}^{K} [\frac{1}{\sqrt{2\pi}\sigma_k} \exp(-\frac{(x_i-\mu_k)^2}{2\sigma_k^2})]^{I(l_i^*,k)} \tag{4.1}$$

where μ_k and σ_k^2 are the region-dependent mean and variance, l_i^* is the true pixel label, and $I(l_i^*,k)$ is the indicator function

$$I(l_i^*,k) = \begin{cases} 1 & (l_i^* = k) \\ 0 & (l_i^* \neq k) \end{cases}$$

The joint pdf of all pixel images is

$$P(\mathbf{x}|\mathbf{l}^*) = \prod_{i=1}^{N^2} \prod_{k=1}^{K} [\frac{1}{\sqrt{2\pi}\sigma_k} \exp(-\frac{(x_i-\mu_k)^2}{2\sigma_k^2})]^{I(l_i^*,k)} \tag{4.2}$$

where $\mathbf{l}^* = (l_1^*, ..., l_{N^2}^*)$ and all l_i^* are considered to be functionally independent. This is the CFNM model [60]. It is important to observe that both the regional parameters (μ_k, σ_k) and the contextual variables l_i^* are non-random unknown constants.

[2] Here microscopic correlation refers to the intrinsic correlation among spin densities as the input of imaging system.

[3] Here 'conditional' means that for each pixel image i, its true label is given with one to one correspondence.

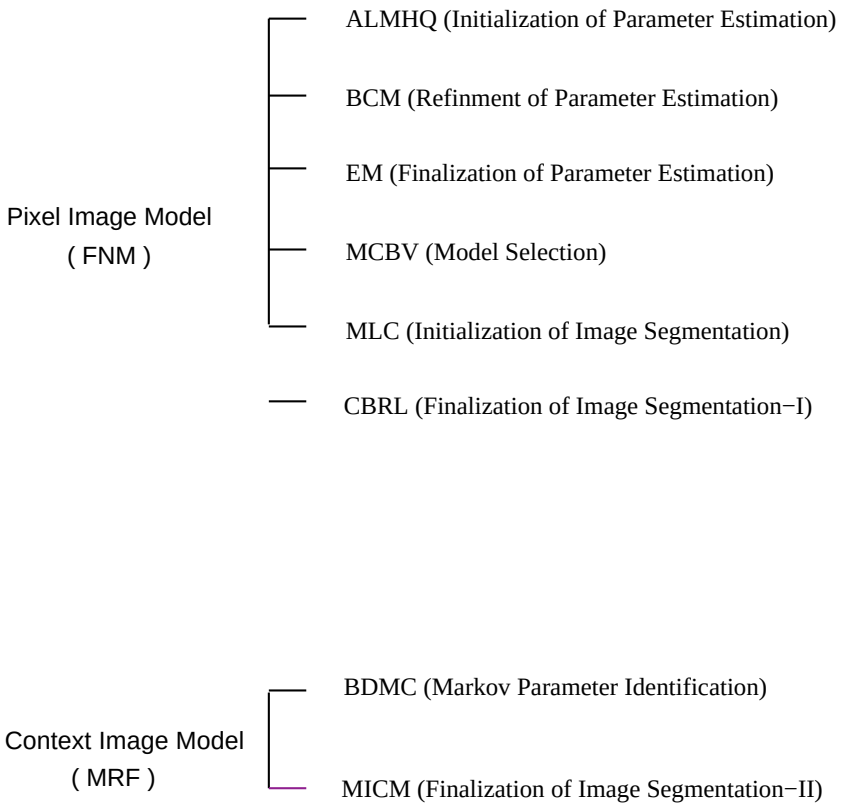

Figure 4.1: An overview of statistical modeling for MR images

4.2.2 Standard Finite Normal Mixture

By *randomly* reordering all pixels in the underlying probability space, i.e., ignoring information regarding the spatial ordering of pixels, we can treat pixel labels as random variables and introduce a new probability measure by using a multinomial distribution with unknown parameters π_k. This distribution reflects the distribution of the number of pixels in each region, i.e., the global context information [47, 67]. Thus, the relevant (sufficient) statistics are the tone statistics for each region and the *number* of pixels in each region [67]. The *new standard*[4] pdf of pixel image becomes

$$q(x_j) = \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{2\pi}\sigma_k} \exp(-\frac{(x_j - \mu_k)^2}{2\sigma_k^2}) \tag{4.3}$$

and the corresponding joint pdf for the whole image is

$$Q(\mathbf{x}) = \prod_{j=1}^{N^2} \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{2\pi}\sigma_k} \exp(-\frac{(x_j - \mu_k)^2}{2\sigma_k^2}) \tag{4.4}$$

which is recognizable as the distribution of N^2 independent and identically distributed rv's with a marginal distribution formed by a mixture of K Gaussian components [67]. This is the standard finite normal mixture (SFNM) model [31]. An important observation is that the marginal pdf $q(x)$ parameterizes the true distribution and can be estimated by the histogram of pixel images.

4.2.3 Comments on CFNM and SFNM Models

This section discusses the relationship between CFNM and SFNM models as well as their advantages and limitations, and the validity of the independence approximation of pixel images made in the formulation.

4.2.3.1 Relationship between CFNM and SFNM

As described in Section 4.2.2, SFNM model is obtained by assuming that the image pixels have been randomly reordered in the same sample space as the CFNM model [67]. Mathematically, we derive the formulation Eq. (4.3) as follows; Let both the pixel images x_j and the context images l_j be random variables, according to the total probability law [21], the marginal probability measure for any pixel image can be obtained by writing down the joint probability density of the x_j and l_j and then summing it over all possible outcomes of l_j [47]

$$q(x_j) = \sum_{k=1}^{K} p(x_j, l_j = k) \tag{4.5}$$

By using Bayesian formula and setting $p(l_j = k) = \pi_k$, we have

$$q(x_j) = \sum_{k=1}^{K} p(l_j = k) p(x_j | l_j = k) = \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{2\pi}\sigma_k} \exp(-\frac{(x_j - \mu_k)^2}{2\sigma_k^2}) \tag{4.6}$$

Clearly, the relationship between π_k and true l_j^*'s satisfies

$$\pi_k = \frac{1}{N^2} \sum_{j=1}^{N^2} I(l_j^*, k) \tag{4.7}$$

[4] Here 'standard' means that for each reordered pixel image j, its true label exists, but the correct one to one correspondence is not given.

where the true pixel labels l_j^* are interpreted as the *true* value (realization) of random variables l_j *associated* with the observed pixel images [65]. Thus, in SFNM modeling we *choose* l_j to be *random* without *realizing* its true value[5]. The reason for doing this is that when the one to one correspondence between pixel image and its true label is ignored, the uncertainty can be used to define or treat l_j as a random variable.

Histogram of an image is defined as follows:

Definition: *For a given image* $\mathbf{x}$ *consisting a set of pixel images* $\mathbf{x} = (x_1, ..., x_J)$*, its histogram is defined by*

$$q_{\mathbf{x}}(u_l) = \frac{1}{J} \sum_{i=1}^{J} I(x_i, u_l) \tag{4.8}$$

where $I(\cdot, \cdot)$ *is the indicator function,* u_l *is the gray level, and* J *is the total number of pixel images.*

From the definitions of CFNM, SFNM, and histogram, we have the following observation:

Observation 4.2.3.1: *SFNM model possesses the same sample space as the CFNM model, but with a new probability measure. The pixels in SFNM model are identically distributed. The histogram of the image (Eq. (4.8)) is a sample estimation of true SFNM distribution (Eq. (4.3)).*

This observation can be simply explained as follows: The data space in CFNM can be called *classified*, and the same data space in SFNM is *unclassified* [31]. Thus, SFNM and CFNM have the same sample space, and the uncertainty contained in the *unclassified* data set contributes to the new probability measure in SFNM.

4.2.3.2 Independence Approximation of Pixel Images

In Section 4.2.1 and 4.2.2, the formulation of both CFNM (Eq. 4.2) and SFNM (Eq. 4.4) models is based on the assumption that pixel images can be considered to be statistical independent. Thus, a natural question is the effectiveness of FNM modeling in the description of pixel images with specified structure and parameter set so as to capture the useful statistical character for image analysis [62]. In this work, we use independent structure (Eq. 4.2 and 4.4) and emphasize FNM distributions (Eq. 4.1 and 4.2). For the correct use of FNM models, one criterion used for evaluation of the models is the fitting between the FNM distribution and the true pdf based on the given pixel images[6]. More precisely, the discussion addressed here focuses on the impact of independence approximation (i.e., ignoring the dependence among pixel images) in FNM modeling on the quantitative model fitting procedure, i.e., the estimation of FNM parameter set based on the given pixel images $\mathbf{x}$.

The discussion is carried out in two steps. First, we reformulate the likelihood function of $\mathbf{x}$ with FNM models by a relative entropy representation. Therefore, the quantitative model fitting is equivalent to the least relative entropy estimation, and the representative of the given pixel image $\mathbf{x}$ for the true image distribution is provided by the histogram of pixel images. Then, we use the statistical properties of MR pixel images to show the convergence property of pixel image histogram [21, 75]. Hence we conclude that for MR pixel images, the independence approximation in FNM modeling will not essentially affect the quantitative model fitting procedure for extracting the information about the true image distribution.

Given the relationship between the likelihood function and the probability measure, we define likelihood functions of CFNM and SFNM based on Eqs. (4.2) and (4.4) as follows:

$$\mathcal{L}_{CFNM}(\mathbf{r}) = \prod_{i=1}^{N^2} \prod_{k=1}^{K} [\frac{1}{\sqrt{2\pi}\sigma_k} \exp(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2})]^{I(l_i^*, k)} = \prod_{i=1}^{N^2} p_{\mathbf{r}^{(k)}}(x_i) \tag{4.9}$$

$$\mathcal{L}_{SFNM}(\mathbf{r}) = \prod_{j=1}^{N^2} \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{2\pi}\sigma_k} \exp(-\frac{(x_j - \mu_k)^2}{2\sigma_k^2}) = \prod_{j=1}^{N^2} q_{\mathbf{r}}(x_j) \tag{4.10}$$

[5] The procedure for realizing the true value of l_j, given x_j, is called pixel classification, i.e., image segmentation. This will be further discussed in Chapter 6.

[6] By using Eq. 4.1 and 4.3, we emphasize first order statistics.

where $\mathbf{r}$ denotes the model parameter vector.

First, we will show that under the independence approximation, maximizing likelihood function is equivalent to minimizing the relative entropy between the histogram and the estimated pdf.

Theorem 4.2.3.1: *If $x_1, ..., x_{N^2}$ are identically distributed according to true distribution $q^*(x)$, then the likelihood of $\mathbf{r}$ in SFNM is determined only by the histogram and is given by*

$$\mathcal{L}_{SFNM}(\mathbf{r}) = 2^{-N^2[H(q_{\mathbf{X}})+D(q_{\mathbf{X}}||q_{\mathbf{r}})]} \tag{4.11}$$

where $q_{\mathbf{X}}$ is the histogram, $q_{\mathbf{r}}$ is an estimation of $q^*(x)$, $H(\cdot)$ denotes the entropy, and $D(\cdot||\cdot)$ denotes the relative entropy.

Proof of Theorem 4.2.3.1:

Since the multiplication over j in Eq. (4.10) is not affected by the pixel order, we regroup them in increasing order of the gray levels $u_l : u_1 < u_2, ..., < u_L$. Thus

$$\mathcal{L}_{SFNM}(\mathbf{r}) = \prod_{j=1}^{N^2} q_{\mathbf{r}}(x_j) = \prod_{l=1}^{L} \left(\prod_{x_j = u_l} q_{\mathbf{r}}(x_j) \right) \tag{4.12}$$

By the definition of histogram (4.8), the number of pixels with gray level u_l equals $N^2 q_{\mathbf{X}}(u_l)$, hence we have

$$\mathcal{L}_{SFNM}(\mathbf{r}) = \prod_{l=1}^{L} q_{\mathbf{r}}(u_l)^{N^2 q_{\mathbf{X}}(u_l)} \tag{4.13}$$

$$= \prod_{l=1}^{L} 2^{N^2 q_{\mathbf{X}}(u_l) \log_2 q_{\mathbf{r}}(u_l)} \tag{4.14}$$

$$= \prod_{l=1}^{L} 2^{N^2 [q_{\mathbf{X}}(u_l) \log_2 q_{\mathbf{r}}(u_l) - q_{\mathbf{X}}(u_l) \log_2 q_{\mathbf{X}}(u_l) + q_{\mathbf{X}}(u_l) \log_2 q_{\mathbf{X}}(u_l)]} \tag{4.15}$$

$$= 2^{-N^2 \sum_{l=1}^{L} [q_{\mathbf{X}}(u_l) \log_2 \frac{1}{q_{\mathbf{X}}(u_l)} + q_{\mathbf{X}}(u_l) \log_2 \frac{q_{\mathbf{X}}(u_l)}{q_{\mathbf{r}}(u_l)}]} \tag{4.16}$$

$$= 2^{-N^2[H(q_{\mathbf{X}})+D(q_{\mathbf{X}}||q_{\mathbf{r}})]} \quad \square \tag{4.17}$$

Similarly, by regrouping Eq. (4.2) based on l_i^* in an order according to region index $k = 1, ..., K$, it is straightforward to obtain the following corollary for the CFNM model.

Corollary: *If $x_1, ..., x_{N^2}$ are piece-wisely and identically (piece-wise stationary) distributed according to the true distributions $p^{*(1)}(x), ..., p^{*(K)}(x)$, then the likelihood of $\mathbf{r}$ in CFNM is determined only by the piecewise histograms and is given by*

$$\mathcal{L}_{CFNM}(\mathbf{r}) = 2^{-N^2 \sum_{k=1}^{K} \pi_k [H(p_{\mathbf{X}}^{(k)}) + D(p_{\mathbf{X}}^{(k)} || p_{\mathbf{r}^{(k)}})]} \tag{4.18}$$

where $p_{\mathbf{X}}^{(k)}$ is the k region histogram, and $p_{\mathbf{r}^{(k)}}$ is the parameterized estimation of $p^{*(k)}(x)$.

Since quantitative model fitting procedure is in general based on the likelihood function of the model parameter vector $\mathbf{r}$, Theorem 4.2.3.1 and the corollary show that for a given image with histogram $q_{\mathbf{X}}$ (or $p_{\mathbf{X}}^{(k)}$), if $q_{\mathbf{X}}$ (or $p_{\mathbf{X}}^{(k)}$) is used as a basis and the parameter vector $\mathbf{r}$ is chosen such that the relative entropy between the estimated pdf $q_{\mathbf{r}}$ (or $p_{\mathbf{r}}^{(k)}$) and the histogram $q_{\mathbf{X}}$ (or $p_{\mathbf{X}}^{(k)}$) is minimized, then this fitting procedure is the same as maximizing likelihood function under the independence approximation of pixel images. In other words, the quality of the quantitative model fitting procedure will be essentially determined by the asymptotic convergence property of $q_{\mathbf{X}}(x)$ to $q^*(x)$ (or $p_{\mathbf{X}}^{(k)}(x)$ to $p^{*(k)}(x)$), since even ignoring the correlation among pixel images, the information about the true pdf contained originally in $\mathbf{x}$ can be *sufficiently* represented by the histogram of pixel images for our interests.

Next, we show that when N^2 approaches infinity, the histogram converges to the true pdf with probability one. Which is given in the following theorem.

Theorem 4.2.3.2: *Given an MR image with true piece-wise distributions* $p^{*(1)}x), ..., p^{*(K)}(x)$ *with a mixture probability* $\{\pi_k\}$*, then for* $k = 1, ..., K$*, pixel images* $\mathbf{x}_k$ *of each region satisfies*

$$\lim_{N^2 \to \infty} Pr(p_{\mathbf{x}_k}(u_l) = p^{*(k)}(u_l)) = 1 \tag{4.19}$$

Proof of Theorem 4.2.3.2:

For each gray level u_l, we apply indicator function $I(x, u_l)$ to pixel images $\mathbf{x}_k$ of the kth region. By the definition of histogram, we have the relationship between the histogram of the kth region $p_{\mathbf{x}_k}(u_l)$ and the sample average of the indicator functions $I(x_j, u_l)$, given by [21]

$$p_{\mathbf{x}_k}(u_l) = \frac{1}{\pi_k N^2} \sum_{j=1}^{\pi_k N^2} I(x_j, u_l) \tag{4.20}$$

Since pixel images $\mathbf{x}_k$ are Gaussian, stationary, and ergodic[7], also, indicator function is a deterministic measurable function, Birkhoff-Khinchin theorem [21] shows that

$$\lim_{N^2 \to \infty} Pr\left(\frac{1}{\pi_k N^2} \sum_{j=1}^{\pi_k N^2} I(x_j, u_l) = E[I(x_j, u_l)] \right) = 1 \tag{4.21}$$

By the fundamental theorem of expectation [21], we have

$$E[I(x_j, u_l)] = \sum_u I(x_j = u, u_l) p^{*(k)}(u) = p^{*(k)}(u_l) \tag{4.22}$$

Thus, by substitute Eqs. (4.20) and (4.22) into Eq. (4.21) we get

$$\lim_{N^2 \to \infty} Pr(p_{\mathbf{x}_k}(u_l) = p^{*(k)}(u_l)) = 1. \quad \square$$

By applying indicator function to pixel images $\mathbf{x}_{(\cdot)}$ of each region, it is straightforward to obtain the same conclusion for $k = 1, ..., K$.

According to the relationship between the SFNM mixture and its components, we have

$$q_{\mathbf{X}}(u_l) = \sum_{k=1}^{K} \pi_k p_{\mathbf{x}_k}(u_l)$$

and

$$q^*(u_l) = \sum_{k=1}^{K} \pi_k p^{*(k)}(u_l)$$

Therefore, when Eq. (4.19) holds for all k components, the following corollary is true.

Corollary: *Given an MR image with the true distribution* $q^*(x)$*, then for pixel images* $\mathbf{x}$

$$\lim_{N^2 \to \infty} Pr(q_{\mathbf{X}}(u_l) = q^*(u_l)) = 1 \tag{4.23}$$

[7] Strictly positive definite covariance function for MR image is shown in Section 3.3.4.

Theorem 4.2.3.2 and and the corollary show that the histogram $q_{\mathbf{X}}(x)$ converges to the true distribution $q^*(x)$ with probability one as $N^2 \rightarrow \infty$.

In summary, Theorem 4.2.3.1 shows that minimizing the relative entropy between estimated pdf $q_{\mathbf{R}}$ and the histogram $q_{\mathbf{X}}$ is equivalent to maximizing likelihood function under independence approximation among pixel images $\mathbf{x}$. Theorem 4.2.3.2 shows that when the number of pixel images N^2 approaches infinity, histogram $q_{\mathbf{X}}$ approaches the true pdf q^* with probability one. Thus, when N^2 is sufficiently large, minimization of the relative entropy between $q_{\mathbf{R}}$ and q^* can be well approximated by the minimization the relative entropy between $q_{\mathbf{R}}$ and $q_{\mathbf{X}}$. This fitting procedure can be practically implemented by maximizing likelihood function under independence approximation of pixel images. Hence, regarding the independence approximation, the correct use of FNM models for our specified purposes and interests in MR image analysis has been justified.

However, some unsatisfactory results in applications raise the questions or objections for the FNM modeling. In order to answer these questions, we discuss the advantages and limitations of the FNM models, and also, provide an introduction for some FNM extensions we will present in the following sections.

4.2.3.3 Advantages and Limitations of CFNM and SFNM

For the CFNM model, both region-defined variables (μ_k, σ_k^2) and contextual variables l_i^* are model parameters. Therefore, in an unsupervised mode, image analysis addresses the problem of the combined estimation of these parameters. Since CFNM model does not have an identically distributed structure, according to reference [65], an unbiased solution to the likelihood equation of Eq. (5.1) based on Eq. (4.9) is not assured, that is, the biases may be introduced into the estimations [47]. This disadvantage will be seen more clearly in Section 5.2.2. Furthermore, since all contextual variables l_i^* are considered to be functionally independent in the CFNM model, one limitation of this model is that it can only reflect the local context information in pixel scale, i.e., the correlation among pixel labels is missing [40].

On the other hand, the SFNM model provides a new structure of likelihood function in which the original data are treated to be i.i.d.. Thus, for the same reason given by [65], an asymptotically unbiased maximum likelihood (ML) estimation can be obtained. However, because the context associated with the individual pixel is replaced by the new parameter π_k in Eq. (4.10), the SFNM model can only reflect the global context information in the image scale, i.e., the context images are hidden [64, 72].

Based on above discussions, we show that the biased ML estimation in using CFNM model is caused by the inappropriate model structure and the lack of contextual regularization. We also show that the price to be paid for i.i.d. structure in using SFNM model is the disability of representing local context information explicitly. CFNM model seems to have more evident meaning than SFNM model for image segmentation, while SFNM model has a better structure for parameter estimation [47, 60, 62].

Since context information has been proven to be important in unsupervised image analysis [40, 41], for some specified applications such as parameter estimation and image segmentation, it is desirable to have the suitable mathematical formulations of FNM model that can incorporate context images associated with pixel images and reflect the correlation among them with an appropriate structure. For example, in order to include global context into the CFNM model, an improvement of Eq. (4.9) has been proposed [47, 59]

$$\mathcal{L}_{JFNM}(\mathbf{r}) = \prod_{i=1}^{N^2} \prod_{k=1}^{K} [\frac{\pi_k}{\sqrt{2\pi}\sigma_k} \exp(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2})]^{I(l_i^*, k)} \tag{4.24}$$

This is the joint FNM (JFNM) model. The major problems with this model are that the marginal likelihood function does not correspond to a probability measure and the correlation among pixel labels is still missing. In the next two sections, we present our recent work on the multiresolution and localized formulations of FNM model to improve the performance of the conventional FNM model for specified purposes.

4.2.4 Multiple Resolution CFNM

By dividing original image into disjoint blocks with $b \times b$ pixels[8] and assuming that all pixels within one block belong to the same image region, a multiple resolution CFNM can be formed. The joint pdf of all pixel images in each block under CFNM modeling is

$$p(\mathbf{x}_r|l_r^*) = \prod_{k=1}^{K}[\prod_{i=1}^{b^2}\frac{1}{\sqrt{2\pi}\sigma_k}\exp(-\frac{(x_i-\mu_k)^2}{2\sigma_k^2})]^{I(l_r^*,k)} \tag{4.25}$$

where r is the index of the block, and l_r^* is the true label associated with the rth block. Then, assuming N^2/b^2 is an integer, the joint pdf of all blocks for the whole image becomes

$$P(\mathbf{x}|\mathbf{l}^*) = \prod_{r=1}^{N^2/b^2}\prod_{k=1}^{K}[\prod_{i=1}^{b^2}\frac{1}{\sqrt{2\pi}\sigma_k}\exp(-\frac{(x_i-\mu_k)^2}{2\sigma_k^2})]^{I(l_r^*,k)} \tag{4.26}$$

where $\mathbf{l}^* = (l_1^*, ..., l_{N^2/b^2}^*)$ and all l_r^* are assumed to be functionally independent since contextual information is primarily local. The new formulation of Eq. (4.26) incorporates local context images by using an idea similar to vector quantization (VQ) and considers b to be a parameter.

4.2.5 Localized SFNM

By assuming that context images are correlated random variables and exhibit local regularities, we propose a localized SFNM formulation for pixel images by incorporating these local regularities statistically. Within the framework of SFNM modeling, new formulation provides a form of local consistency constraints between context images in terms of a stochastic regularization by using Markovian property. Its utilization in contextual image segmentation will be further discussed in Section 6.3.1.

Using a strategy similar to the one in [70], for each pixel i, we define the spatial constraint as the set of all pairs (l_i, l_j) such that the consistency between l_i and l_j can be measured by the compatibility function $R_{ij}(l_i, l_j)$. By requiring that context images be consistent locally, constraints are defined only over neighboring pixels. That is, when there is no correlation between labels, or when i and j are not neighbors, the compatibility $R_{ij}(l_i, l_j)$ is zero [70]. For simplicity, we define the neighborhood of pixel i by opening a $b \times b$ window with pixel i being the central pixel, and let ∂i denote its neighborhood [9] and use indicator function represent the constraints $R_{ij}(l_i, l_j) = I(l_i, l_j)$. Note that pairs of labels are now either compatible or incompatible. Similar to reference [70], we can compute the frequency of neighbors of a pixel i which has labels compatible to a given label k at i when the labels of its neighbors $\mathbf{l}_{\partial i}$ are given

$$\pi_k^{(i)} = p(l_i = k|\mathbf{l}_{\partial i}) = \frac{1}{b^2-1}\sum_{\partial i} I(l_i = k, l_{j|\partial i}) \tag{4.27}$$

where $\mathbf{l}_{\partial i} = (l_{j|\partial i} : j = 1, ..., b^2, j \neq i)$. Note that $\pi_k^{(i)}$ can be interpreted as the *conditional* prior of regions (conditional probability of l_i) since its value depends on the uncertainty resulting from the given $\mathbf{l}_{\partial i}$, and it provides a form of non-parameterized formulation of Markovian regularization [42]. From the similarity to the SFNM modeling [62], by considering $\pi_k^{(i)}$ as the local conditional prior of regions, we formulate the localized SFNM distribution for x_i as

$$q(x_i|\mathbf{l}_{\partial i}) = \sum_{k=1}^{K}\pi_k^{(i)}\frac{1}{\sqrt{2\pi}\sigma_k}\exp(-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}) \tag{4.28}$$

[8] Generally, N^2/b^2 is not an integer. For the residual pixels on the boundary, we give a specified treatment which is described in Section 5.2.2.

[9] For neighborhood, when pixel is at the corner, a quarter window is used; when pixel is at the boundary, a semi-window is used.

Thus, by further requiring conditional independence of pixel images, the joint distribution of $\mathbf{x}$, given the context images $\mathbf{l}$, becomes

$$Q(\mathbf{x}|\mathbf{l}) = \prod_{i=1}^{N^2} \sum_{k=1}^{K} \pi_k^{(i)} \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}\right) \tag{4.29}$$

where $\mathbf{l} = (l_i : i = 1, ..., N^2)$. Clearly, for different $\mathbf{l}$, the joint distribution of $\mathbf{x}$ is different.

4.2.6 Comments on Multiple Resolution CFNM and Localized SFNM

By introducing multiresolution CFNM, the correlation among neighboring context images is taken into account in a simple way. Therefore, the parameter estimation using multiresolution approach is expected to be more accurate, that is, multiresolution CFNM formulation may provide a good approximation when an appropriate resolution can be selected according to a suitable criterion. The simulation results of regional parameter estimation (given in Section 7.2.1.) support this consideration. In addition, since this approach provides a multiresolution mode for algorithm implementation, the likelihood of being trapped in local minima/maxima can be reduced by using coarse resolution solutions to guide finer solutions [67]. However, multiresolution approach can only provide an approximation, and the resultant analysis may not be sufficiently accurate since partial volume effect[10] may be introduced by the boundary mismatch. In this work, multiresolution formulation of CFNM is mainly used to support the multiresolution implementation of parameter estimation in Chapter 5.

Localized SFNM formulation is specifically designated to incorporate local context regularities into the framework of SFNM modeling. By assuming context images be locally correlated random variables with Markovian property, new formulation provides a way of stochastic regularization on context images. The differences between localized SFNM and SFNM are that: in SFNM, the prior of regions π_k is determined by the *true* context images (functionally independent), so the formulation does not explicitly include the context images l_j (statistically independent). The uncertainty resulting from the underlying $\mathbf{l}^*$ provides a fixed global context regularization by Eq. (4.7) which is shift-invariant; in localized SFNM, *interim* realization of context images (statistically dependent) are given, and for each pixel image its neighboring context images is conditioned. Therefore, the uncertainty from the given $\mathbf{l}_{\partial i}$ provides a better local regularization $\pi_k^{(i)}$ in terms of Eq. (4.27) which is shift-variant. New formulation statistically reflects Markovian property of context images [42], presents a link to the local consistency constraints, and provides a possibility of using relaxation labeling procedure for contextual image segmentation [70], which will be further explained in Chapter 6.

4.3 Markov Random Field Modeling

4.3.1 Markovian Property and Global Gibbs Measure

In Section 3.4, we show that by defining a neighborhood system, Markovian Property describes a local dependence model in terms of conditional probability measure. In Section 4.2.5, we propose the localized SFNM to incorporate the Markovian Property into the pixel image modeling using a conditional probability framework. Another way to do this is to use Gibbs measure that will be introduced in this section.

From the equivalence between a Gibbs random field (GRF) and a Markov random field (MRF), many investigations have shown that GRF (or global Gibbs measure) model provides a joint probability measure for $\mathbf{L}$ in which local Markovian Property is preserved [42]. Theoretically, GRF is expressed in the following form

$$P(\mathbf{l}) = \frac{1}{Z_{\mathbf{l}}} \exp(-U(\mathbf{l})) \tag{4.30}$$

[10] Due to the limit of resolution, an MR image may consist of pixels made up of a mixture of tissue types. There mixed tissue pixels represent partial volume effect.

where $U(\mathbf{l})$ is the energy function, and the normalizing constant $Z_{\mathbf{l}}$ is the partition function. A neighborhood system can be established by specifying the clique function (or potential function) $V_c^{(i)}(\mathbf{l})$, and the relationship between energy function and clique function is [40]

$$U(\mathbf{l}) = \sum_{i=1}^{N^2} V_c^{(i)}(\mathbf{l}) \tag{4.31}$$

It has been shown that a unique GRF exists for every MRF and vice versa, as long as the GRF is defined in terms of cliques in a neighborhood system [42]. By translating global or local context information into the energy function of Gibbs measure through clique structure and Markov parameter, an appropriate MRF model can be established. The most typical configuration of MRF model is the pairwise interaction neighborhood system in which spatial neighbors occur only in pairs [41]. So the energy function can be written by

$$U(\mathbf{l}) = \sum_{i=1}^{N^2} \sum_{r=1}^{c} [\theta_i I(l_i, l_{i:\pm r})] \tag{4.32}$$

where $\{\theta_i\}$ is the Markov parameter[11], and c denotes clique with $c = 2$ for a first-order model and $c = 4$ for a second-order model. More precisely, a second-order neighborhood system is: $\{i : +r | r = 1, 2, 3, 4\} \cup \{i : -r | r = 1, 2, 3, 4\}$ [40]. In this research, we use a second-order pairwise interaction neighborhood system in the MRF model.

4.3.2 Hidden MRF Model and Conditional Gibbs Measure

Different from the localized SFNM formulation, in order to insert contextual information into pixel image modeling, a posterior conditional probability measure is used within the context of Bayesian formulation. References [40, 42] show that requiring the conditional independence of the observed pixel images, given the context images, is sufficient to ensure that the posterior distribution is an MRF. That is, a posteriori conditional probability function for the context images $\mathbf{l}$, given the observed pixel images $\mathbf{x}$ also has the form of a GRF. This is the hidden MRF model [44, 75]. Mathematically, we have

$$P(\mathbf{l}|\mathbf{x}) = \frac{1}{Z_{\mathbf{l}|\mathbf{x}}} \exp(-U(\mathbf{l}|\mathbf{x})) \tag{4.33}$$

where $Z_{\mathbf{l}|\mathbf{x}}$ is a normalizing constant. Based on the pairwise interaction neighborhood system, the corresponding energy function is

$$U(\mathbf{l}|\mathbf{x}) = \sum_{i=1}^{N^2} [\frac{1}{2} \ln(\sigma_{l_i}^2) + \frac{(x_i - \mu_{l_i})^2}{2\sigma_{l_i}^2} + \sum_{r=1}^{c} [\theta_i I(l_i, l_{i:\pm r})]] \tag{4.34}$$

In many applications, one is interested in the local property of a hidden MRF. Equivalence theorem in [42] shows that local Markovian property can be mathematically derived from Gibbs measure. For each pixel i, the conditional probability of l_i can be written as [45]

$$p(l_i|\mathbf{l}_{\partial i}, \mathbf{x}) = \frac{1}{Z_i} \exp(-U(l_i, \mathbf{l}_{\partial i}, \mathbf{x})) \tag{4.35}$$

where Z_i is a normalizing constant, and

$$U(l_i, \mathbf{l}_{\partial i}, \mathbf{x}) = \frac{1}{2} \ln(\sigma_{l_i}^2) + \frac{(x_i - \mu_{l_i})^2}{2\sigma_{l_i}^2} + \sum_{r=1}^{c} [\theta_i I(l_i, l_{i:\pm r})] \tag{4.36}$$

In a contrast to the region-defined parameters $(\pi_{(\cdot)}, \mu_{(\cdot)}, \sigma_{(\cdot)}^2)$, Markov parameter $\theta_{(\cdot)}$ is interpreted as boundary-defined parameter. This will be further explained in Section 5.5.

[11] In this work, Markov parameter $\{\theta_i\}$ is designed to reflect the dependence or correlation among context images.

4.4 Summary

In this chapter, we present a new framework for the stochastic modeling of MR images using random field theory. For pixel images, we introduce FNM models and show that SFNM model can be obtained mathematically from CFNM model using randomization, therefore SFNM and CFNM models have the same sample space but different event space and probability measures (Theorem 4.2.3.1). Based on the statistical properties of pixel images, we prove several theorems so as to justify the correct use of the independence approximation used in FNM models. We show that the quality of model fitting procedure based on FNM modeling only depends on the histogram of data and is not seriously affected by the independence approximation since the histogram of the image converges to the true distribution (Theorem 4.2.3.2). By incorporating the statistical properties of context images, we presented the multiresolution and localized formulations of the FNM model. Multiresolution CFNM provides a basis for multiple resolution algorithm implementation in parameter estimation, and localized SFNM uses statistical local consistency constraints (Markovian property) to provide a regularization tool for possible contextual image segmentation. In order to analyze the properties of FNM models, we present a discussion with respect to their structures and applications. Finally, we introduce the concept of MRF modeling for context images and propose an inhomogeneous[12] hidden MRF to model both pixel and context images using Bayesian formulation. We show that this configuration naturally integrates region and boundary information.

[12] Here "inhomogeneous" means that Markov parameter θ_i values are spatially varying.

Chapter 5

Parameter Estimation

5.1 Introduction

Unsupervised stochastic model-based image analysis addresses the problem of combined *estimation* of regional $(\pi_k, \mu_k, \sigma_k^2)$, structural K, and Markov θ_i parameters, and *detection* of the context variables (l_i^*, l_j), given the observed pixel images $\mathbf{x}$. Parameter estimation mainly focuses on determining model parameter values while image segmentation solves the problem of context variable detection. Figure 5.1 presents a block diagram that describes the main contents of parameter estimation and image segmentation.

Although parameter estimation and image segmentation are counterparts of each other in image analysis, it is desirable to estimate model parameters directly from the observed data before segmentation [62, 69]. In this chapter, we will discuss the principles and algorithms used for parameter estimation based on the SFNM model.

Detection of the structural parameter K, the number of image regions directly affects the quality of the resulting regional parameter estimation. One approach for detecting the number of image regions is to use a sequence of hypothesis tests [52]. The problem in this approach is the subjective judgement in the selection of the threshold for different tests. Some recent work [18, 61, 62, 67] applied two information theoretic criteria: Akaike information criterion (AIC) and minimum description length (MDL) for this problem. In this work, we use AIC and MDL, and propose a new approach based on minimum conditional bias/variance (MCBV) criterion, for detecting the number of image regions.

The counterpart for using information theoretic criterion in region number detection is to find ML estimate of the regional parameter vector $\mathbf{r}$. The ordinary way of calculating the ML estimator is to take the derivative of the likelihood function $\mathcal{L}(\mathbf{r})$ with respect to the parameter vector $\mathbf{r}$, set it equal zero, and solve the resulting ML equation [65]:

$$D\mathcal{L}(\mathbf{r}) = \frac{\partial \mathcal{L}(\mathbf{r})}{\partial \mathbf{r}} = 0. \tag{5.1}$$

Because of the complexity of the model structure, i.e., high dimensionality and non-linearity, the likelihood equation may not be solved explicitly. Based on the ML method, or equivalently, least relative entropy (LRE) method [31], we derive three iterative relaxation algorithms to obtain the approximated ML solutions, and provide the mathematical interpretation of these algorithms.

As indicated by many researchers (see e.g. [61, 62, 67]), in the unsupervised mode, performance of the algorithm used for parameter estimation may suffer from high computational complexity and local minima. For example, in order to get a sufficiently accurate ML solution, most of the iterative methods are both time- and memory-intensive, and only guarantee convergence to one of local minima [32, 62]. Therefore, the detection of the number of image regions may also result in some unsatisfactory solution since the accuracy of information theoretic criterion is very sensitive to the local minima of ML estimates of regional parameters [61]. In this work, we address the importance of parameter initialization, and overcome above two major problems using systematic histogram analysis and multiresolution classification.

For convenience in discussion, we summarize the main algorithms we developed in this chapter as follows.

- Adaptive Lloyd-Max Histogram Quantization (ALMHQ) algorithm: an adaptive iterative procedure for initializing regional parameter values based on Lloyd-Max optimal scalar quantization in which the histogram is modeled by the SFNM.
- Block-wise classification-maximization (BCM) Algorithm: a two-step iterative procedure for refining regional parameter initialization by using ML method and conducting multiple resolution CM algorithm.
- Expectation-Maximization (EM) Algorithm: a two-step iterative procedure for finalizing ML regional parameter estimation based on Bayesian probabilistic pixel classification.
- Boundary Defined Markov Configuration (BDMC) Algorithm: a look-up table mapping procedure for Markov parameter identification based on local randomization and entropy rate estimation.

Figure 5.1 presents a block diagram that describes the main contents of parameter estimation and image segmentation. Note that in our approach parameter estimation is performed based on the image data before final image segmentation, and both parameter estimation and image segmentation are achieved through the multistage algorithms.

5.2 Regional Parameter Initialization

5.2.1 Histogram Quantization with Minimum Distortion

In rate distortion theory, Lloyd-Max (LM) scalar quantization has been proven to be optimal in the sense that it results in minimum distortion [49, 75]. We extend the application of LM quantization to the initialization of regional parameter values. After obtaining the optimal multiple thresholds, the initial regional parameter values can be calculated simply by using the definitions of the parameters. Considering histogram as a probability measure, distortion is defined as the mean squared value of the quantization error. For a given number of ranges K, the global distortion D can be described by specifying the thresholds (t_k, t_{k+1}) and the mean of the kth region $(k = 1, ..., K)$

$$D = \sum_{k=1}^{K} \int_{t_k}^{t_{k+1}} (u - \mu_k)^2 q_{\mathbf{X}}(u) du \tag{5.2}$$

Clearly, the necessary conditions for minimizing D with fixed K can be derived by differentiating D with respect to the t_k's and μ_k's and setting the derivatives to zero:

$$\frac{\partial D}{\partial t_k} = 0 \implies \mu_k = 2t_k - \mu_{k-1}, \ k = 2, ..., K, \tag{5.3}$$

$$\frac{\partial D}{\partial \mu_k} = 0 \implies \int_{t_k}^{t_{k+1}} (u - \mu_k) q_{\mathbf{X}}(u) du = 0, \ k = 1, ..., K. \tag{5.4}$$

That is, μ_k is the centroid of the area of $q_{\mathbf{X}}(u)$ between t_k and t_{k+1}. This method is particularly suitable to compromising the profile and the details of the histogram, so is relatively noise insensitive.

Because of the complicated functional mixtures likely to be induced by $q_{\mathbf{X}}(u)$, Eqs. (5.3) and (5.4) are not a set of simultaneous equations we can solve with a closed-form. A method is to pick μ_1, calculate the succeeding t_k's and μ_k's, and if μ_K is (or close to in discrete cases) the true centroid of the last component, denoted by μ_K^*, then μ_1 is chosen correctly; otherwise, update μ_1 as a function of the distance between μ_K and μ_K^*. We develop an adaptive iterative algorithm to carry out this procedure numerically. In this procedure, we insert a parameter α to control the learning rate. Since only discrete version of $q_{\mathbf{X}}(u)$ is available, an

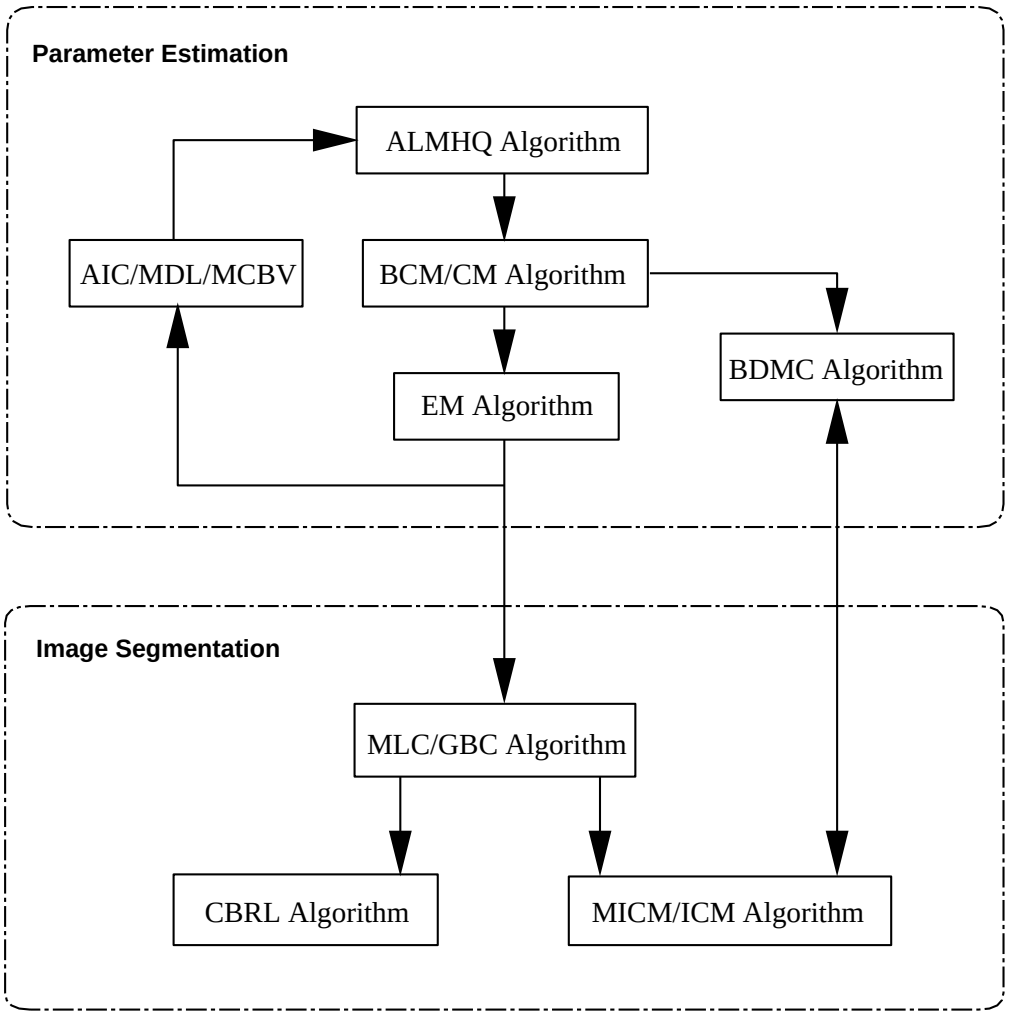

Figure 5.1: Flowchart of parameter estimation and image segmentation

appropriate setting of α is important for the stability of the algorithm in the numerical implementation. The algorithm is summarized as follows.

ALMHQ Algorithm:

1. Fix K and set α[1] and ϵ
2. Set $m = 0$ and select $\mu_1^{(m)}$
3. For $k = 1, ..., K-1$
   - set $t_1^{(m)} = u_{min}$,
   - compute $t_{k+1}^{(m)}$ by

$$\sum_{u=t_k^{(m)}}^{t_{k+1}^{(m)}} u q_{\mathbf{X}}(u) = \mu_k^{(m)} \sum_{u=t_k^{(m)}}^{t_{k+1}^{(m)}} q_{\mathbf{X}}(u) \tag{5.5}$$

   - compute $\mu_{k+1}^{(m)}$ by

$$\mu_{k+1}^{(m)} = 2t_{k+1}^{(m)} + \mu_k^{(m)} \tag{5.6}$$

   - set $t_{K+1}^{(m)} = u_{max}$ and compute $\mu_K^{(m)*}$ by Eq. (5.5)
4. if $|\mu_K^{(m)} - \mu_K^{(m)*}| < \epsilon$ then
   Go to step 5
   Otherwise

$$\mu_1^{(m+1)} = \mu_1^{(m)} + \alpha(\mu_K^{(m)*} - \mu_K^{(m)})$$
$$m = m + 1$$

   Go to step 3
5. Compute

$$\sigma_k^2 = \sum_{u=t_k^{(m)}}^{t_{k+1}^{(m)}} (u - \mu_k^{(m)})^2 q_{\mathbf{X}}(u)$$

$$\pi_k = \sum_{u=t_k^{(m)}}^{t_{k+1}^{(m)}} q_{\mathbf{X}}(u)$$

5.2.2 Block-Wise CM (BCM) Algorithm

BCM algorithm is developed to refine the regional parameter initialization using ML estimation in a multiresolution mode. Similar to CM algorithm [62], pixel classification (block-wise) is a necessary step in parameter estimation. The first step of the algorithm is to classify each block to each region (i.e., obtain $l_r^{(m)}$ using least relative entropy (LRE) criterion and the current estimates $\pi_k^{(m)}$, $\mu_k^{(m)}$, and $\sigma_k^{2(m)}$ (C-step). The second step of the algorithm is to update parameter estimates (i.e., obtain $\pi_k^{(m)}$, $\mu_k^{(m)}$, and $\sigma_k^{2(m)}$) by using ML criterion and the classification result $l_r^{(m)}$ (M-step). The iterations between these two steps continue until no block changes its label. By assuming that each block belongs to a single region, we define histogram of pixel images in the rth block as the feature vector of that block and approximate it by a normal distribution f_r, since it is a suitable treatment for tone images. Furthermore, the relative entropy between the feature vector f_r and the regional distribution $p_{\mathbf{r}^{(k)}}$ is used as the meaningful information distance between the feature and template, which is given by

$$D(f_r || p_{\mathbf{r}^{(k)}}) = \log \frac{\sigma_k}{\sigma_r} + \frac{\sigma_r^2 - \sigma_k^2 + (\mu_r - \mu_k)^2}{2\sigma_k^2} = d(r,k) \tag{5.7}$$

[1] $\alpha = 1/K^3$, is an empirical value found to be a suitable learning rate based on our intensive numerical experiments.

where μ_r and σ_r^2 are the sample mean and variance of the rth block, μ_k and σ_k^2 are the sample mean and variance of the kth region, and $d(r,k)$ is the distance between the rth block and the kth region.

Applying BCM algorithm to the images with different block sizes $b = b_{MAX}, ..., 1$, a multiple resolution processing procedure, from the coarser resolution to the finer resolution, is formed. The advantages of the multiresolution approach are: 1) computational complexity is reduced; 2) likelihood of being trapped in local minima/maxima is reduced. We use global relative entropy (GRE) between the histogram and the estimated SFNM distribution as a measure of the estimation bias, and select the parameter estimation corresponding to the resolution b_0 which satisfies

$$b_0 = \arg\left\{ \min_{1 \le b \le b_{MAX}} \sum_u q_{\mathbf{x}}(u) \log \frac{q_{\mathbf{x}}(u)}{q_{\hat{\mathbf{r}}}(u)} \right\} \tag{5.8}$$

where $\hat{\mathbf{r}}$ is the ML estimate of parameter vector $\mathbf{r}$ for a fixed block size b. BCM algorithm is summarized as follows:

BCM Algorithm:

1. Given $\mathbf{r}^{(0)}$ calculated by ALMHQ,
   $b = b_{MAX}$, $m = 0$
2. C-step: for $r = 1, ..., \frac{N^2}{b^2}$
   compute the relative entropy distance $d(r,k) = \log \frac{\sigma_k^{(m)}}{\sigma_r} + \frac{\sigma_r^2 - \sigma_k^{2(m)} + (\mu_r - \mu_k^{(m)})^2}{2\sigma_k^{2(m)}}$
   classify the rth block into the kth region by $l_r^{(m)} = \arg\{\min_{1 \le k \le K} d(r,k)\}$
3. M-step: for $k = 1, ..., K$, update regional parameters:

$$\pi_k^{(m+1)} = \frac{b^2 \sum_{r=1}^{N^2/b^2} I(l_r^{(m)}, k)}{N^2}$$

$$\mu_k^{(m+1)} = \frac{\sum_{r=1}^{N^2/b^2} I(l_r^{(m)}, k) \mu_r}{\sum_{r=1}^{N^2/b^2} I(l_r^{(m)}, k)}$$

$$\sigma_k^{2(m+1)} = \frac{\sum_{r=1}^{N^2/b^2} I(l_r^{(m)}, k)(\mu_r - \mu_k^{(m+1)})^2}{\sum_{r=1}^{N^2/b^2} I(l_r^{(m)}, k)}$$

4. If $(\mathbf{l}^{(m+1)} - \mathbf{l}^{(m)}) = 0$, go to step 5.
   Otherwise $m = m + 1$ and go to step 2.
5. If $b = 1$, go to step 6.
   Otherwise $b = b - 1$ and go to step 2.
6. Choose b_0 and the corresponding $\pi_k^{(m)}$, $\mu_k^{(m)}$, and $\sigma_k^{2(m)}$ according to Eq. (5.6).

Note that $\frac{N^2}{b^2}$ may not in general be an integer so corner or edge effect may occur. Since for most real MR images corner or edge does not contain meaningful information, the practical way to avoid this problem is to choose an integer which is closest but less than $\frac{N^2}{b^2}$ in the implementation of the algorithm. In addition, since for a given data structure, different block-wise partition may result in non-uniqueness of the "best" resolution b_0, and a full search of the unique b_0 is time consuming, we consider Eq. (5.8) as a systematic way to determine a sub-optimal b_0. Also, note that by the law of large numbers, the variation of b_0 will be small when the number of total observations is large; and our numerical experiments also show that the parameter estimates corresponding to b_0 are always better than the initial values computed by ALMHQ algorithm, i.e., they achieve lower GRE values. We apply Eq. (5.8) to the parameters computed by ALMHQ and BCM, and use GRE between the histogram and the estimated SFNM distribution as a measure of the goodness of parameter estimation.

5.3 Regional Parameter Finalization

5.3.1 Expectation-Maximization Algorithm

We use EM algorithm to finalize the ML estimates by further reducing the bias after the parameter initialization, i.e., ALMHQ followed by BCM. The reasons for using EM algorithm as the final step in parameter estimation are that: a) ML principle is directly applied to all pixels in the algorithm (i.e., pixel-based resolution); b) the algorithm has the capability of achieving flexible classifier shapes using probabilistic memberships; c) parameter estimates satisfy the ergodic theorems and therefore are asymptotically unbiased.

Based on the SFNM model, reference [62] shows that, EM algorithm can be theoretically derived by first estimating the likelihood of the complete data (both pixel and context images) using the incomplete data (only pixel images) and the current parameter estimates (E-step), and then maximizing the estimated likelihood to generate the updated parameter estimates (M-step). The algorithm cycles back and forth between these two steps to reach a stationary point of the likelihood [18, 46, 62]. We emphasize the importance of the parameter estimation performance and use relative entropy between the histogram and the estimated SFNM distribution as a stopping criterion. The stopping criterion can be expressed by

$$|\sum_u q_{\mathbf{X}}(u)\log\frac{q_{\mathbf{X}}(u)}{q_{\hat{\mathbf{r}}^{(m+1)}}(u)} - \sum_u q_{\mathbf{X}}(u)\log\frac{q_{\mathbf{X}}(u)}{q_{\hat{\mathbf{r}}^{(m)}}(u)}| \le \epsilon \tag{5.9}$$

where $\hat{\mathbf{r}}$ is the ML estimate of the parameter vector, and ϵ is a prespecified small number.

We summarize EM algorithm as follows.

EM Algorithm:

1. Given $\mathbf{r}^{(0)}$ calculated by BCM algorithm at $b = b_0$
   $m = 0$
2. E-step: for $j = 1, ..., N^2,\ k = 1, ..., K$, compute the probabilistic membership

$$z_{jk}^{(m)} = \frac{\pi_k^{(m)}\frac{1}{\sqrt{2\pi}\sigma_k^{(m)}}\exp(-\frac{(x_j-\mu_k^{(m)})^2}{2\sigma_k^{2(m)}})}{\sum_{k=1}^K \pi_k^{(m)}\frac{1}{\sqrt{2\pi}\sigma_k^{(m)}}\exp(-\frac{(x_j-\mu_k^{(m)})^2}{2\sigma_k^{2(m)}})} \tag{5.10}$$

3. M-step: for $k = 1, ..., K$, compute the updated parameter estimates

$$\begin{cases} \pi_k^{(m+1)} = \frac{1}{N^2}\sum_{j=1}^{N^2} z_{jk}^{(m)} \\ \mu_k^{(m+1)} = \frac{1}{N^2\pi_k^{(m+1)}}\sum_{j=1}^{N^2} z_{jk}^{(m)} x_j \\ \sigma_k^{2(m+1)} = \frac{1}{N^2\pi_k^{(m+1)}}\sum_{j=1}^{N^2} z_{jk}^{(m)}(x_j - \mu_k^{(m+1)})^2 \end{cases} \tag{5.11}$$

4. If Eq. (5.7) is satisfied, Stop.
   Otherwise, $m = m + 1$ and Go to Step 2.

5.3.2 Comments on EM Algorithm

EM algorithm is an example of deterministic relaxation method. References [31, 46, 62] show that

$$\mathcal{L}(\mathbf{r}^{(m+1)}) \ge \mathcal{L}(\mathbf{r}^{(m)}) \tag{5.12}$$

That is, the likelihood increases with the iterations although there is no guarantee of convergence to the global maximum. Our numerical experiments support this statement. Reference [65] shows that the parameter estimates from EM algorithm are asymptotically unbiased. In probabilistic classification, instead of realizing

the true value of pixel label l_j, Bayesian formulation is used to assign every pixel to each region with a posterior conditional probability measure, given by[2]

$$z_{jk} = p(l_j = k|x_j) = \frac{p(l_j = k)p(x_j|l_j = k)}{p(x_j)} \tag{5.13}$$

i.e., we emphasize probabilistic membership by keeping l_j *random*. Mathematical formulation of EM algorithm can be interpreted by the ergodic theorems since regional parameter estimates are obtained simply by using the *weighted* sample averages over classified regional pixels.

5.4 Model Selection

5.4.1 Detection of the Number of Image Regions

In unsupervised stochastic model-based image analysis, the number of image regions is usually detected as part of parameter estimation procedure before image segmentation. The procedure for detecting this number is called model selection. In this work, instead of directly determining the correct number K_0 of the distinctive image regions by the conventional hypothesis testing [31], a model-fitting procedure is utilized to select a model from several competing candidates such that the selected model best fits the given image data [62]. These candidates have the same functional form of pdf (or likelihood) Eq. (4.4) [3] but with different K ($K = K_{MIN}, ..., K_{MAX}$) where K_{MIN} and K_{MAX} are prespecified. Therefore, the number of image regions can be determined by this model fitting procedure. In order to evaluate the goodness of fitting, two well-known information criteria (AIC and MDL) are used and one newly derived formulation (MCBV) is proposed. The theoretical justification and interpretation, and numerical comparisons are also presented.

5.4.2 AIC and MDL Criteria

Akaike proposes to select the model that gives the minimum AIC, defined by

$$AIC(K_a) = -2\log(\mathcal{L}(\hat{\mathbf{r}}_{ML})) + 2K_a \tag{5.14}$$

where $\mathcal{L}(\hat{\mathbf{r}}_{ML})$ is the likelihood of $\hat{\mathbf{r}}_{ML}$, and K_a is the number of free adjustable parameters in the model [51, 62]. The first term represents a form of the information theoretic distance (*Kullback-Leibler mean information*) between the histogram and its SFNM model [75], and the second term, $2K_a$, is the penalty term reflecting both approximation and bias correction [51, 55]. The AIC tries to reformulate the problem explicitly as a problem of *approximation* of the true structure by the model. AIC criterion implies that the correct number of the distinctive image regions K_0 can be obtained by

$$K_0 = \arg\left\{\min_{1\le K\le K_{MAX}} AIC(K_a)\right\} \tag{5.15}$$

From a quite different point of view, Rissanen reformulates the problem explicitly as an information coding problem in which the best model fit is measured such that it assigns high probabilities to the observed data while at the same time the model itself is not too complex to describe [53, 55]. In other words, a shortest total code length is preferred. So the model is selected by minimizing the total description length defined by

$$MDL(K_a) = -\log(\mathcal{L}(\hat{\mathbf{r}}_{ML})) + 0.5K_a\log N^2. \tag{5.16}$$

The first term in MDL is identical to the corresponding one in AIC, and the second term is defined as the model complexity penalty. Note that, different from AIC, the penalty term in MDL takes into account the

[2] In CM algorithm, the corresponding variable to z_{jk} is $I(l_i^*, k)$.

[3] Since the derivations of information criteria require that the observations are i.i.d. in the model, we use SFNM formulation in this work.

number of observations. Similarly, the correct number of the distinctive image regions K_0 must satisfy

$$K_0 = \arg\left\{\min_{1\le K\le K_{MAX}} MDL(K_a)\right\} \tag{5.17}$$

5.4.3 Comments on AIC and MDL

In general, when ML estimation is used, the fitting by using a higher-order model will always be better than the lower order models since it includes the lower order cases, and the corresponding estimation will generally choose a large or infinite number of image regions to achieve the best or perfect fitting to the observed data, which clearly is not an interesting nor a useful result [53]. Therefore, these two criteria suggest to use approximation (AIC) or optimal coding (MDL) [51, 55].

However, the justifications for the optimality of these criteria and the assumptions made in the derivations are somewhat indirect and remain unsolved [53, 55]. For example, reference [52] shows that AIC cannot be asymptotically consistent and the heuristics of Akaike do not seem to lead to any proof of its optimality; reference [53, 55] show that prior distribution to assign to parameters in MDL remains never be settled to everyone's satisfaction and other distributions could be contemplated; references [18, 67] show that the modified AIC and MDL (by an empirical factor) result in better estimations; our recent work [4] found that even though a universal prior for parameters is acceptable, the optimal precision determined to describe parameters in MDL seems to have no connection with the precision for expressing the observed data.

We present an alternative approach to the problem by deriving a new formulation. Nevertheless, it was Akaike/Rissanen's work that was the inspirational source to this work, but some new interpretations are presented and justified with information theoretic means.

5.4.4 Minimum Conditional Bias and Variance Criterion

Our new formulation for model selection is based on minimum conditional bias and variance (MCBV) criterion. Similar to the idea in references [56, 57], our intuitively appealing criterion comes from a very fundamental and attractive principle: the number of image regions can not be arbitrary or infinite, because such a model might be said to have low 'bias' but the price to be paid is high 'variance'. Whether or not it turns out to be entirely satisfactory should be decided on experiments [57].

Using a similar procedure as in MDL [55], model fitting is optimized by achieving minimum description length for both data and model parameters, or equivalently, for the total estimation error. This criterion is introduced based on an entropy measure, which is aimed at supplying the missing structure dependent term [57]. The underlying principle is stated as: "*the parameters in a model which determine the value of the maximum entropy should be assigned values which minimize the maximum entropy*" (Jaynes' principle).

Invoked by Bayes's law, let model parameter vector be denoted by $\mathbf{r}$, the joint distribution of data $\mathbf{x}$ and model parameter estimates $\hat{\mathbf{r}}$ can be factored as follows:

$$P(\mathbf{x},\hat{\mathbf{r}}) = Q(\mathbf{x}|\hat{\mathbf{r}})f_{\hat{\mathbf{R}}}(\hat{\mathbf{r}}) \tag{5.18}$$

where $Q(\cdot|\hat{\mathbf{r}})$ is the estimated distribution of the true distribution $Q^*(\cdot)$ for the data, and $f_{\hat{\mathbf{R}}}(\cdot)$ is the suitable distribution for parameter estimate $\hat{\mathbf{r}}$. We assume that $Q(\cdot|\hat{\mathbf{r}})$ can be obtained by estimating the parameter vector $\mathbf{r}$ from observations $\mathbf{x}$ by some estimation method. According to reference [75], by using optimal coding scheme based on $P(\mathbf{x},\hat{\mathbf{r}})$, the average total description length $L(\mathbf{x},\hat{\mathbf{r}})$ required to describe both data and model parameters can be calculated in terms of the joint entropy, given by

$$L(\mathbf{x},\hat{\mathbf{r}}) = L(\mathbf{x}|\hat{\mathbf{r}}) + L(\hat{\mathbf{r}}) \tag{5.19}$$

where the connection between the entropy and description length is made [57] such that $L(\mathbf{x}|\hat{\mathbf{r}})$ $(= H(\mathbf{x}|\hat{\mathbf{r}}))$ is the description length for data, and $L(\hat{\mathbf{r}})$ $(= H(\hat{\mathbf{r}}))$ is the description length for model parameters.

Subjecting to the natural requirements that the parameterized distribution estimate equals to the sample estimate, a very neat interpretation states [57] that the maximum of the entropy $H(\mathbf{x}|\hat{\mathbf{r}})$ according to Jaynes'

principle, is precisely the negative of the logarithm of the likelihood function $\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}})$ corresponding to the entropy-maximizing distribution for $\mathbf{x}$, we have

$$\max_{q_{\hat{\mathbf{r}}}=q_{\mathbf{X}}} L(\mathbf{x}|\hat{\mathbf{r}}) = -\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}})) \tag{5.20}$$

where $q_{\hat{\mathbf{r}}}$ is the estimated SFNM distribution of the true distribution $q^*(\cdot)$ such that the distribution for each region is normal and the mixture prior distribution is determined by the uniform randomization which corresponds to the maximum uncertainty. We can also maximize $L(\hat{\mathbf{r}})$ based on Jaynes' principle so as to assure that the components of $\hat{\mathbf{r}}$ are normal and independent [75]. Suppose that model parameter vector have K_a independent components ($\mathbf{r} = (r_1, ..., r_{K_a})$), maximizing the entropy of the parameter estimates results in

$$\max_{\hat{\mathbf{r}}=\mathbf{r}_{\mathbf{X}}} L(\hat{\mathbf{r}}) = \sum_{k=1}^{K_a} H(\hat{r}_k) \tag{5.21}$$

where $\mathbf{r}_{\mathbf{X}}$ denotes the sample estimate of $\mathbf{r}$. Note that when variance of parameter estimate is determined by the corresponding sample estimate, normal distribution gives the maximum entropy by a theorem due to Shannon.

The joint maximum entropy of Eqs. (5.20) and (5.21) is a function of the model parameter K_a, and $\hat{\mathbf{r}}$. A minimization, then, leads to the following characterization of the optimum estimations according to Jaynes principle:

$$\min_{K_a,\hat{\mathbf{r}}} \max L(\mathbf{x}, \hat{\mathbf{r}}) = \min_{K_a,\hat{\mathbf{r}}} [-\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}})) + \sum_{k=1}^{K_a} H(\hat{r}_k)] \tag{5.22}$$

For the interpretation of the right side of Eq. (5.22), reference [57] shows that both $-\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}))$ and $\sum_{k=1}^{K_a} H(\hat{r}_k)$ represent natural estimation errors about their true models and should be treated on an equal basis. We decomposite the first term into the following two terms according to Theorem 4.2.3.1:

$$-\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}})) = N^2 H(q_{\mathbf{X}}) + N^2 D(q_{\mathbf{X}}||q_{\hat{\mathbf{r}}}) \tag{5.23}$$

Since true distribution is not available in general, we consider the image histogram $q_{\mathbf{X}}(u)$, the sample estimation of the true distribution as the relevant statistics, and when $N^2 \to \infty$, $q_{\mathbf{X}} \to q^*$. Note that for a given $\mathbf{x} = (x_1, ..., x_{N^2})$, $q_{\mathbf{X}}$ is fixed. Therefore, we can interpret $N^2 H(q_{\mathbf{X}})$ as the constant description length for the data based on the true model, and $N^2 D(q_{\mathbf{X}}||q_{\hat{\mathbf{r}}})$ as the extra description length for the data based on the estimated model [75]. We define $-\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}))$ as the conditional[4] model bias. Nevertheless, we need a formulation for second statistics formulation about $\hat{\mathbf{r}}$ in the second term of Eq. (5.22). By assuming that a suitable formulation of the distribution about $\hat{\mathbf{r}}$ is available and demoting its estimate by $\prod_{k=1}^{K_a} f_{\hat{r}_k}(\hat{r}_k)$, we define $\sum_{k=1}^{K_a} H(\hat{r}_k)$ as the conditional model variance since the corresponding entropy solely depends on the variances of the estimated parameters.

Therefore, if the cost of model variance is defined as the description length to represent its parameters [57], it can be seen that the cost of adding new parameters to the model must be balanced by the reduction they permit in the ideal code length for the reconstruction error [75]. Hence we justify the desired optimality in a most natural manner. By taking the advantage of the fact that parameter estimation is separable as regional and structural, for fixed K_a, regional parameters can be estimated by the ML principle. Hence, for the purpose of model selection, we define MCBV criterion as

$$MCBV(K) = -\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}_{ML})) + \sum_{k=1}^{K_a} H(\hat{r}_{kML}) \tag{5.24}$$

Based on this new formulation, MCBV is to select a model with K_0 distinctive image regions if

$$K_0 = \arg\left\{ \min_{1 \le K \le K_{MAX}} MCBV(K) \right\} \tag{5.25}$$

[4] Here 'conditional' means that only one set of observations $\mathbf{x}$ is given.

Thus, we can compute the value of $MCBV(K)$ for $K = 1, 2, ..., K_{max}$ and determine K_0 according to Eq. (5.25). Once K_0 is determined, $\hat{\mathbf{r}}_{ML}$ is taken as the final estimated parameter vector. New approach has a simple optimal appeal in that it selects a minimum conditional bias and variance model, i.e., if two models are about equally likely, MCBV selects the one whose parameters can be estimated with the smallest variance [4, 57].

In view of the fact that ML estimator is consistent and normal under certain conditions [57, 65], it is reasonable to use normal distribution to describe $\hat{\mathbf{r}}_{ML}$ in which the variance of the estimates asymptotically converges to its *Cramer-Rao Lower Bound* (CRLB). In the discussion of EM algorithm, we have shown that ML estimates satisfy the natural requirement that they converges to the sample estimates, which is shared by Rissanen in [57]. Similar to AIC and MDL, in practical applications, in order to obtain a simplified closed-form of $\sum_{k=1}^{K_a} H(\hat{r}_{kML})$ we discuss the formulation of variances of ML estimates and the approximations used in numerical computation.

References [65, 78] show that for SFNM model, the CRLBs of ML estimates are

$$\begin{aligned} Var(\hat{\pi}_{kML}) &= \frac{\pi_k(1-\pi_k)}{N^2} \\ Var(\hat{\mu}_{kML}) &= \frac{\sigma_k^2}{N^2\pi_k} \\ Var(\hat{\sigma}_{kML}^2) &= \frac{2\sigma_k^4(N^2\pi_k - 1)}{N^4\pi_k^2} \end{aligned} \tag{5.26}$$

Since the true values of model parameters are not available, we use their ML estimates in the calculation. This reasonable treatment is shared by Rissanen [53, 57] and supported by the consistency of ML estimates [65]. Moreover, reference [62] shows that the total number of independent model parameters in SFNM model is

$$K_a = 3K - 1 \tag{5.27}$$

Recall that for a given variance normal distribution maximizes the entropy of the data, we adopt that ML parameter estimates have asymptotically normal distributions denoted by $(f_{\hat{r}_k}(\hat{r}_k) = \{N(r_k, Var(r_{kML}))\}, k = 1, ..., K_a)$. Reference [75] shows that for the normal distribution, we have

$$H(\hat{r}_{kML}) = \frac{1}{2}\log 2\pi e Var(\hat{r}_{kML}) \tag{5.28}$$

Therefore, we have a practical MCBV formulation with code-length expression

$$MCBV(K) = -\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}_{ML})) + \sum_{k=1}^{K_a} \frac{1}{2}\log 2\pi e Var(\hat{r}_{kML}) \tag{5.29}$$

We organize the procedure for calculating MCBV values as follows:

MCBV Algorithm:

1. For $K = K_{MIN}, ..., K_{MAX}$, given $\hat{\mathbf{r}}_{ML}$ finalized by EM algorithm
2. Compute the optimal code-length for model bias

$$L_1 = -\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}_{ML}))$$

3. Compute the optimal code-length for model variance
   for $k = 1, ..., 3K - 1$
   - Calculate $Var(\hat{r}_{kML})$
   - Compute

$$L_2 = \sum_{k=1}^{3K-1} \frac{1}{2}\log(2\pi e Var(\hat{r}_{kML}))$$

4. Compute $MBVC(K) = L_1 + L_2$.
5. Select the K_0 that minimizes $MBVC(K)$.

5.4.5 Discussion

MCBV for model selection employs ideas similar to those of AIC and MDL, however the major differences of the approach and the resulting formulation can be summarized as follows:

1) By using Theorem 4.2.3.1, we decompose $-\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}))$ into two terms (Eq. (5.23)). We recognize that for a given image $\mathbf{x}$ with large N^2 in Eq. (5.23), the first term does not relate to the estimator, it is simply the joint entropy of data based on the true model. Hence the second term, joint relative entropy provides a natural measurement of the inefficiency of using the estimated distribution instead of true distribution, i.e., the distance between the estimated and the true distributions. Considering Eqs. (5.19) and (5.23) together, our bias/variance decomposition of ML estimation error provides a better interpretation for the problem of model selection, which is strongly shared by Gemen in [56]. AIC uses an approximation of relative entropy by taking the expectation which makes the physical interpretation difficult[52, 58]; MDL measures data complexity by self-information of $\mathbf{x}$ without further decomposition, and hence the link to the estimation error is indirect[5].

2) To describe the statistics of model parameter estimates, we use an asymptotically normal prior distribution based on Jaynes' principle and the properties of ML estimation[6], and let these normal prior distributions have different means and variances for different model parameters. MDL considers model parameter estimates as non-random truncated integers and uses the same universal prior distribution (non-informatic) for all model parameter estimates, based on the theoretic coding scheme. Our numerical experiments show that this assumption is not appropriate in some cases since model parameter estimates are quite random with a variance beyond the range of universal prior distribution.

3) The second term in MCBV highly depends on the data structure, and is a non-linear function of K. Assume that the means in data structure are less crucial since they may not affect the absolute description length, we use all other parameters of data structure (N^2, π_k, σ_k^2). The formulation in Rissanen's minimax approach [57] implies a similar strategy, that is, when some small regions with high variances are included in the model, they make large contributions to the penalty term, and if this can not be balanced against a reduction in the model bias, they will be eliminated first, that is, only meaningful distinct regions are selected [67]. By taking a complicated approximation and optimization procedure, the universal optimal precision make the second term in MDL to solely depend on K and N^2, so it is not clear why the models with the same K and N^2 and different combination of other parameters should have the same complexity.

4) Both AIC and MDL are originally derived for a random time series [51, 53, 55, 57, 58]. We derive MCBV based on SFNM model for an image random field in which the SFNM modeling is re-justified by Jaynes's principle. We also conduct extensive numerical simulations to show that the model selection based on MCBV criterion are consistent with those of the AIC and MDL, and further discuss the estimation properties for noisy data.

5.5 Markov Parameter Identification

5.5.1 Entropy Rate Estimation of Context Images

In order to perform contextual image segmentation based on inhomogeneous hidden MRF model, the identification of Markov parameter θ_i is a prerequisite. As discussed in Sections 3.4 and 4.3, Markov parameter is designed to reflect the dependence or correlation among context images, and the level of the dependence or correlation can be mathematically represented by the conditional histogram of context images based on the available context information. Clearly, this is inconvenient for the scale MRF parameter identification. Based on the initial context images from BCM algorithm or the interim context images from MICM algorithm, we propose to use entropy rate as a measure of the dependence and to identify Markov parameter values using look-up table. This approach has a clear physical interpretation since lower entropy rate means higher dependence.

[5] However, in Rissanen's minimax approach [57] the link to the prediction error is established for time series model.

[6] This is also suggested by Rissanen in references [55, 57].

Reference [75] shows that for a stationary random process $L_1, L_2, ..., L_m$, the entropy rate $H(\mathcal{L})$ is defined as the rate of the entropy growth

$$H(\mathcal{L}) = \lim_{m\to\infty} \frac{1}{m} H(L_1, L_2, ..., L_m) \tag{5.30}$$

and for a stationary Markov chain, the entropy rate can be easily reformulated as

$$H(\mathcal{L}) = \lim_{m\to\infty} H(L_m|L_{m-1}, ..., L_1) = \lim_{m\to\infty} H(L_m|L_{m-1}) = H(L_2|L_1) \tag{5.31}$$

Let inhomogeneous MRF be locally stationary, the local entropy rate $H(\mathcal{L}_i)$ is given by

$$H(\mathcal{L}_i) = H(L_i|\mathbf{L}_{\partial i}) = \sum_{\mathbf{l}_{\partial i}} P(\mathbf{l}_{\partial i}) H(L_i|\mathbf{l}_{\partial i}) \tag{5.32}$$

Since in general the true a prior probability $P(\mathbf{l}_{\partial i})$ is not available, we use first order stochastic approximation to estimate the entropy rate [21, 65, 75]

$$H(\mathcal{L}_i) \approx H(L_i|\mathbf{l}_{\partial i}) = -\sum_{k=1}^{K} p(l_i = k|\mathbf{l}_{\partial i}) \log p(l_i = k|\mathbf{l}_{\partial i}) \tag{5.33}$$

where

$$p(l_i = k|\mathbf{l}_{\partial i}) = \sum_{j=1}^{b^2-1} \frac{I(l_{j|\partial i}, k)}{b^2 - 1} \tag{5.34}$$

Finally, inhomogeneous MRF parameter is identified according to the following look-up table

$$\theta_i = \frac{\alpha}{H(\mathcal{L}_i) + \eta} \tag{5.35}$$

where α is the scale factor, η is the shifting offset. These parameters are determined empirically based on the experience and their selection is discussed in Section 7.2.4.

5.5.2 Boundary Defined Markov Configuration (BDMC) Algorithm

Based on available context images, the actual procedure for identifying Markov parameter values is described as follows.

BDMC Algorithm:

1. Given initial context images $\mathbf{l}^{(0)}$
2. For each pixel $i = 1, ..., N^2$
   - Determine conditional probabilities according to Eq. (5.34)
   - Calculate local entropy rate according to Eq. (5.33)
3. Assign Markov parameter values by look-up table Eq. (5.35).

5.6 Summary and Discussions

In this chapter, we discuss the methods and algorithms for unsupervised parameter estimation. The observed pixel images are modeled by SFNM and the goodness of model fitting is measured by the GRE criterion. For regional parameter estimation, we first derive ALMHQ algorithm to initialize the parameter estimates, in which a self-organizing procedure is used to search for a minimum distortion solution; then we develop BCM algorithm to refine the parameter initialization based on a multiresolution scheme. The procedure employs a coarse-to-fine resolution loop to search for minimum bias solution; finally, we introduce EM algorithm to finalize ML parameter estimates, in which the stopping criterion is chosen as GRE. We also show that the

ML estimates generated by these algorithms are consistent with the underlying mean and variance ergodic theorems. The numerical results of our experimental simulations show the significant improvement at each stage of the hybrid algorithm (ALMHQ, BCM, EM) for regional parameter estimation. The explanation is as follows. Based on the SFNM distribution, parameter initialization by ALMHQ is obtained by finding an optimal quantizer to the image histogram in which the parameter values are determined by the optimal thresholds. Since most real MR images have the unimodal histograms, biases are introduced by simple thresholding which do not even disappear asymptotically. BCM is based on the same principle as that in CM but with a multiresolution mode. By considering the fact that an MR image is piece-wise stationary, it is reasonable to assume that the number of pixels inside the regions are much large than the number of pixels on the boundaries. Therefore, based on the ergodic theorem, the M-step in BCM algorithm gives lower biased parameter estimates. Theoretically, the parameter estimates from EM are the ML solution of the SFNM distribution. Since the SFNM model has an i.i.d. structure, the ML estimates are asymptotically unbiased through the probabilistic assignment of each pixel to all regions. In addition, the determination of b_0 by GRE in BCM is a critical issue. But it is believed that it will not conflict with the main stream of the whole procedure. Practically, in case that the parameter refinements by BCM corresponding to b_0 has a higher GRE than those from ALMHQ, BCM step can be simply skipped.

For structural parameter detection, i.e., to determine the number of image regions, we use a model-fitting procedure to automatically select a model from several competing candidates, based on the information theoretic criteria, AIC, MDL and a newly proposed approach MCBV. The selected model achieves a good fitting to the observed data in terms of obtaining a distribution estimation with minimum bias and variance. Finally, based on interim pixel classification of BCM or MICM, we derive BDMC algorithm for quantifying Markov parameter values.

The whole procedure for unsupervised parameter estimation is summarized as follows:

1. Calculate the histogram of the whole image $q_{\mathbf{X}}$;
2. For $K = 1, ..., K_{MAX}$
   - Initialize regional parameters by ALMHQ algorithm (Eqs. (5.3-5.4));
   - Refine regional parameter initialization by BCM algorithm (Eqs. (5.5-5.6));
   - Finalize regional parameter estimation by EM algorithm (Eqs. (5.7-5.9));
3. Determine the number of image regions K_0 according to AIC, MDL, and MCBV (Eqs. (5.12, 5.14, 5.27));
4. Save the estimates of both regional and structural parameters with $K = K_0$;
5. Initialize Markov parameter values by BDMC algorithm (Eq. (5.33)).

New approach has several important advantages. In particular, since ML estimate for SFNM model is separable in its structural parameters K and regional parameters $(\pi_k, \mu_k, \sigma_k^2)$, the algorithms are adaptive and self-organizing. Furthermore, the application of ALMHQ and BCM algorithms lead to an accurate and efficient parameter initialization in which the likelihood of local minima in the final solution is reduced and the speed of convergence for ML estimation is increased. By using probabilistic classification, EM algorithm provides the accurate ML estimates of regional parameters. In addition, the model selection procedure based on information theoretic approach not only provides a basis for unsupervised image analysis but also presents a better information match in parameter estimation. Finally, inhomogeneous Markov parameter identification by BDMC Algorithm provides a systematic way to translate the context information into model parameter settings.

Chapter 6

Image Segmentation

6.1 Introduction

MR image segmentation addresses the detection of contextual variables (l_i^*, l_j), given observed pixel images $\mathbf{x}$. In this study, MR image segmentation is formulated as a standard detection and estimation problem, and is performed based on the results of parameter estimation. The methods and algorithms used for image segmentation depend on image models, the nature of context variables, and the optimization criteria utilized [18, 40, 41, 42, 44, 45].

Many approaches for image segmentation have been proposed [35, 36, 37, 45], and they can be classified into two main categories: non-contextual and contextual. In non-contextual approaches, when context images are considered to be non-random unknown constants, pixels are classified into different image regions by maximizing the joint likelihood with respect to pixel labels l_i^*. This is called the ML classifier (MLC) [59, 60]. On the other hand, when context images are considered to be statistically independent random variables, multiple likelihood ratio testing can be used to partition pixels into different image regions by minimizing the total misclassification error. This is called global Bayesian classifier (GBC) [62]. In this work, we use MLC for the initialization in image segmentation.

Most of non-contextual approaches are likely to perform poorly since locally there may not be sufficient information to make a good decision [67]. Therefore, many contextual approaches are proposed to improve the performance of the image segmentation scheme. For example, references [24, 64] conduct fuzzy classification by using spatial correlation among pixel images; reference [63] compares a spatial thresholding method with MRF model-based ICM approach; and references [72, 77] evaluate popular MRF model-based algorithms by intensive numerical simulations. Some researchers also extended these techniques to perform fast segmentation by using multiresolution strategy [67]. Detailed discussions can be found in [67, 69, 71].

In order to obtain a better solution by using contextual MR image segmentation, two deterministic relaxation algorithms are proposed. Based on the localized SFNM formulation, we derive contextual Bayesian relaxation labeling (CBRL) algorithm to improve pixel classifications by locally minimizing misclassification error. Based on the inhomogeneous hidden MRF model, we derive modified iterated conditional mode (MICM) algorithm to search for an MAP approximation by incorporating boundary defined Markov configuration. References [41, 45] show that no algorithmic way of translating statements about the true context in the image into the MRF parameters has been found and the procedures for actually doing this are purely heuristic. We propose to use entropy rate as a measure to identify MRF parameters and let the algorithm automatically update these values based on the interim context information (BDMC algorithm) during its iterations.

We summarize the algorithms for image segmentation as follows:

1. MLC Algorithm: a multiple thresholding procedure for initializing image segmentation based on ML principle.

2. GBC Algorithm: a multiple thresholding procedure for initializing image segmentation based on Bayesian decision rule.
3. CBRL Algorithm: a deterministic relaxation labeling procedure for contextual image segmentation based on localized Bayesian decision rule.
4. MICM Algorithm: a modified ICM procedure for contextual image segmentation based on inhomogeneous hidden MRF configuration and MAP decision rule.

6.2 Initialization of Image Segmentation

6.2.1 ML Classifier and Bayesian Classifier

When true pixel label l_i^*'s are considered to be functionally independent and non-random constants in the FNM formulation, MLC can be designed by directly maximizing individual likelihood function of pixel images for all pixels (Eq. (4.9)) [60]. That is, pixel i will be classified into kth region, if

$$l_i^* = \arg\left\{\max_k \frac{1}{\sqrt{2\pi}\sigma_k}\exp(-\frac{(x_i-\mu_k)^2}{2\sigma_k^2})\right\} \tag{6.1}$$

In fact, this is equivalent to the ML multiple thresholding approach. Thus, we have

$$l_i^* = \arg\left\{\min_k d_{ik}\right\} \tag{6.2}$$

where d_{ik} is the modified Mahalanobis distance given by [62]

$$d_{ik} = \log(\sigma_k^2) + (x_i-\mu_k)^2\sigma_k^{-2} \tag{6.3}$$

On the other hand, when pixel labels are considered to be random variables, pixel classification can be performed by realizing the true values of the pixel labels l_j given the observation x_j. When the global context information π_k's are taken as the prior probabilities in SFNM modeling, GBC is most commonly used for image segmentation. By minimizing the expectation of total Bayes classification error, pixel j will be classified into kth region if

$$l_j = \arg\left\{\max_k \frac{\pi_k}{\sqrt{2\pi}\sigma_k}\exp(-\frac{(x_j-\mu_k)^2}{2\sigma_k^2})\right\} \tag{6.4}$$

Similarly, by using distance measure, we can write

$$l_j = \arg\left\{\min_k d_{jk}\right\} \tag{6.5}$$

where

$$d_{ik} = \log(\sigma_k^2) - 2\log(\pi_k) + (x_j-\mu_k)^2\sigma_k^{-2} \tag{6.6}$$

6.2.2 Comments on Thresholding Methods

In MLC and GBC approaches, true (in MLC) or random (in GBC) pixel labels are estimated or detected independently (Eqs. (6.2) and (6.6)). Thus, they belong to non-contextual image segmentation. Due to the term, $-2\log(\pi_k)$, in Eq. (6.6), the differences between MLC and GBC are that:

1. The misclassification error of MLC is uniformly distributed over the whole image and the total classification error may be high.

2. By using global context information π_k in the decision maker, GBC minimizes the total misclassification error but some minor regions may have locally higher misclassification error.

Since the performance of non-contextual image segmentation may be poor when the SNR of observed image is low, and a uniformly distributed misclassification is more suitable for further contextual image segmentation. Hence, we use MLC to initialize image segmentation.

6.3 Contextual Image Segmentation

6.3.1 Contextual Bayesian Relaxation Labeling (CBRL) Algorithm

When pixel labels are considered to be correlated random variables, contextual image segmentation can be performed to improve initial segmentation by using neighborhood context regularities. In general, most approaches seek for either a consistent labeling solution or a MAP solution to the posteriori probability of pixel labels when given pixel images $P(\mathbf{l}|\mathbf{x})$. In consistent labeling approach, the criterion is to minimize global consistency measure by using a system of inequalities. In MAP approach, the criterion is to minimize the posterior cost of the total misclassification error [65].

Contextual Bayesian relaxation labeling (CBRL) algorithm is developed to obtain a consistent labeling solution based on localized SFNM formulation (Eq. (4.28)). The structure of relaxation labeling is motivated by two basic concerns: 1) the decomposition of a global computation into a network of simple local computations; 2) the suitable use of local context regularities in solving ambiguities. Therefore, we introduce a theory of consistency of discrete relaxation labeling and formalize its relationship to global optimization by using a similar strategy as that in [70].

In Section 4.2.5, we have defined the spatial consistency (smoothness) constraints by the binary compatibility measure functions $R_{ij}(l_i, l_j) = I(l_i, l_j)$, and formulated the local average constraints by the normalized summation of the compatibility of pixel i with its all pair-wise neighbors $\pi_k^{(i)}$ (see Eq. (4.27)). Invoked by Bayes formulation, we interpret $\pi_k^{(i)}$ as the local conditional prior of regions, and define the component in localized SFNM of Eq. (4.28) by the support function

$$S_i(k) = \pi_k^{(i)} \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}\right) \tag{6.7}$$

where $S_i(k)$ is a function of k. We can now define consistency for image segmentation in terms of the satisfaction of a system of inequalities [70].

Definition: *Let* $\mathbf{l}$ *be the labels which are assigned to pixel images* $\mathbf{x}$. *Then* $\mathbf{l}$ *is consistent if*

$$S_i(l_i) \geq S_i(k)$$

for all k *and for* $i = 1, ..., N^2$.

Note that when the support function is suitably defined from a local model for good context regularization, all labels in a consistent labeling have maximum support at each pixel simultaneously. Note that the consistency condition is not a pure maximization problem but simply a criterion for determining if a labeling is consistent [70]. Therefore, in order to formalize the consistent labeling to the global optimization, we use a similar justification as that in [70], i.e., the notion of consistency defined above suggests a measure for guiding the update of a nearly consistent labeling into a consistent one. Thus, we can define

$$A(\mathbf{l}) = \sum_{i=1}^{N^2} \sum_{k} I(l_i, k) S_i(k) \tag{6.8}$$

as the average local consistency measure where each of the terms in the sum represents the local consistency of $\mathbf{l}$. Since the N^2 values of $\sum_k I(l_i, k) S_i(k)$ and the value of $A(\mathbf{l})$ will be large, it would seem natural to

attempt to maximize it in order to satisfy the criterion expressed in the definition of consistency. However, since the N^2 quantities are not independent, we cannot find a single $\mathbf{l}$ which maximizes these N^2 values simultaneously, while our goal is to find a consistent labeling given an initial labeling.

The real problem corresponds to the statement that each of the N^2 values

$$\sum_k I(l_i, k) S_i(k), \qquad i = 1, ..., N^2. \tag{6.9}$$

should be independently maximized. Reference [70] (Theorem 5.1) shows that when the spatial compatibility measure is symmetric, i.e., $R_{ij}(l_i, l_j) = R_{ji}(l_j, l_i)$, if $A(\mathbf{l})$ attains a local maximum at $\mathbf{l}$, then $\mathbf{l}$ is a consistent labeling. Hence, a consistent labeling can be accomplished by locally maximizing $A(\mathbf{l})$. Ordinarily, stochastic discrete gradient descent procedure is used in which each pixel is randomly visited and its label is updated. This procedure can be implemented by relaxation labeling, and is equivalent to finding one of the local maxima of average local consistency, which, according to Theorem 5.1 in [70], are consistent labelings. Reference [79] (Theorem A.1 and Lemma 2.1) shows that under certain conditions the relaxation labeling based on stochastic discrete gradient descent principle converges to a stopping point that no pixel label needs to be updated by the local maximization rule Eq. (6.9), which corresponds to at least one local maximum of $A(\mathbf{l})$. According to Proposition 7.5 in [70], when labeling $\mathbf{l}$ is a stopping point, then if the compatibility measure is symmetric, $\mathbf{l}$ is consistent. Thus, we now have a method for finding consistent labelings, given an initial labeling assignment. However, since each image has different data structure, the performance surface of $A(\mathbf{l})$ may be very complicated. For the convergence of relaxation labeling, it has been shown that stochastic discrete gradient descent procedure has a stopping point at $\mathbf{l}$ if and only if $\mathbf{l}$ is a consistent labeling. Thus, if the stochastic discrete gradient descent procedure, which is approximated by the relaxation labeling algorithm, converges to a stopping point, then we have found a consistent labeling [70]. Fortunately, Theorem 9.1 in [70] shows that for each stopping point $\mathbf{l}_0$, there exists a neighborhood of $\mathbf{l}_0$ such that if $\mathbf{l}^{(m)}$ enters the neighborhood, then $\lim_{m\to\infty} \mathbf{l}^{(m)} = \mathbf{l}_0$.[1] That is, when a good initial assignment $\mathbf{l}^{(0)}$ is given, the likelihood of infinite cycle behavior is extremely low [70].

Based on localized SFNM formulation, we develop CBRL algorithm to conduct contextual image segmentation according to the updating rule of Eq. (6.9). That is, we update l_i^{old} by the value of l_i that maximizes $\sum_k I(l_i, k) S_i(k)$, i.e., classify pixel i into kth region if

$$l_i = \arg\left\{ \max_k \pi_k^{(i)} \frac{1}{\sqrt{2\pi}\sigma_k} \exp(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}) \right\} \tag{6.10}$$

Iterations are needed to search for a consistent labeling in which each pixel is randomly visited in turn and its label is updated according to Eq. (6.10). During this relaxation process, our numerical experiments show that classification error is decreased at every iteration and converges to a local minimum. The whole procedure can be summarized as follows:

CBRL Algorithm:

1. Given $\mathbf{l}^{(0)}$, m=0
2. Update pixel labels
   - Randomly visit each pixel for $i = 1, .., N^2$ (by random permutation of pixel ordering)
   - Update its label l_i according to

$$l_i^{(m)} = \arg\left\{ \max_k \pi_k^{(i)(m)} \frac{1}{\sqrt{2\pi}\sigma_k} \exp(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}) \right\}$$

3. When the percentage of label changing less that 1%, stop; otherwise, $m = m + 1$ and repeat (2).

[1] Because a complete consistent labeling may not be reached in a practical implementation, the relaxation labeling algorithm, as an approximation with finite iterations, can provide a quite reasonable and accurate consistent labeling [18].

In our simulations, the iteration times for a complete convergence takes about 10 to 15 whole frame scans, and experience has also shown that 5 or 6 whole frame scans of an image with a neighborhood system defined by a 5×5 window are sufficient for the convergence to a stationary point.

6.3.2 Modified Iterated Conditional Mode (MICM) Algorithm

Modified iterated conditional mode (MICM) algorithm is developed to obtain a MAP solution based on the hidden MRF modeling. Let ϵ denote the expected segmentation error, i.e., the posterior cost of misclassification given pixel images $\mathbf{x}$ is

$$\epsilon = \frac{1}{N^2} \sum_{i=1}^{N^2} [1 - I(l_i^*, l_i)] \tag{6.11}$$

Based on the following equivalence [67]

$$\arg\{\max_k P(l_i = k, \mathbf{l}_{S|i}|\mathbf{x})\} = \arg\{\max_k p(l_i = k|\mathbf{l}_{S|i}, \mathbf{x})\} \tag{6.12}$$

where $\mathbf{l}_{S|i}$ denotes pixel labels of all pixels except pixel i, reference [45] shows that choosing the labeling that minimizes ϵ is equivalent to maximizing the marginal posterior distribution so that label l_i is updated to be the one that satisfies

$$p(l_i^{new}|\mathbf{x}, \mathbf{l}_{S|i}) \geq p(l_i^{old}|\mathbf{x}, \mathbf{l}_{S|i}) \tag{6.13}$$

for all pixels. Furthermore, by imposing the Markovian constraint, we have the following proportionalities [41]

$$p(l_i|\mathbf{x}, \mathbf{l}_{S|i}) \propto p(x_i|l_i)p(l_i|\mathbf{l}_{\partial i}) \propto p(l_i = k|x_i, \mathbf{l}_{\partial i}) \tag{6.14}$$

That is,

$$\arg\{\max_k p(l_i|\mathbf{x}, \mathbf{l}_{S|i})\} = \arg\{\max_k p(x_i|l_i)p(l_i|\mathbf{l}_{\partial i})\} = \arg\{\max_k p(l_i|x_i, \mathbf{l}_{\partial i})\} \tag{6.15}$$

Based on the inhomogeneous hidden MRF formulation, MICM algorithm is proposed as a computationally feasible alternative to obtain an MAP solution for contextual image segmentation according to Eq. (6.15). From Eq. (4.35), we have

$$p(l_i = k|x_i, \mathbf{l}_{\partial i}) = \frac{1}{Z_i} \exp(-U(l_i, \mathbf{l}_{\partial i}, \mathbf{x})) \tag{6.16}$$

Therefore, maximizing the conditional probability in Eq. (6.16) is equivalent to minimizing the energy function $U(l_i, \mathbf{l}_{\partial i}, \mathbf{x})$. That is, pixel i will be classified into kth region, if

$$l_i = \arg\left\{\min_k U(l_i = k, \mathbf{l}_{\partial i}, \mathbf{x})\right\} \tag{6.17}$$

Combined with BDMC algorithm for updating MRF parameters, the modified ICM algorithm is described below.

MICM Algorithm:

1. Given $\mathbf{l}^{(0)}$ (from MLC) and initial θ_i's (from BCM/BDMC)
2. Randomly visit each pixel for i from 1 to N^2 (by random permutation of pixel ordering)
   - Calculate all possible values of the energy function (Eq. (4.36))
   - Update l_i to a new value according to

$$l_i = \arg\left\{\min_k U(l_i = k, \mathbf{l}_{\partial i}, \mathbf{x})\right\}$$

3. Update all θ_i's using the BDMC procedure (Eq. (5.33))
4. Repeat (2)-(3) until the percentage of label change is less than 1%.

6.3.3 Comments on CBRL and MICM Algorithms

Many variations of contextual image segmentation algorithms have been proposed based on MRF modeling in the literature, such as simulated annealing (SA), iterated conditional modes (ICM), and maximizer of the posterior marginals (MPM). Reference [77] presents a detailed comparative study of these approaches and states that the ICM algorithm performs consistently well, provides reasonable and a robust segmentation, and has clear computational advantages.

In this work, CBRL algorithm is developed based on the localized SFNM formulation and the optimality of consistent labeling, in which Markovian property is incorporated in the derivation. Since the classification rule of CBRL in each iteration is similar to GBC, the differences between CBRL and GBC need to be reiterated. GBC uses the global context information π_k in its decision maker. Since π_k represents an unconditional and shift-invariant prior regularization, and pixel labels are considered as independent random variables, neither initial labeling nor update iterations are required in the GBC approach, i.e., image segmentation can be performed by only visiting all pixels once independently. On the other hand, in CBRL approach, $\pi_k^{(i)}$ represents a conditional and shift-variant local prior regularization, and pixel labels are considered as dependent random variables, an appropriate initial labeling and finite updating iterations are required to search for a consistent labeling. We can view consistency as a "locking-in" property. That is, since the support function defined for a given pixel depends on the current labels at neighboring pixels, this neighborhood influences the update of the given pixel through the compatibility constraints. With the constraint propagation, the relaxation process iteratively updates the label assignments to find a more consistent labeling with the neighboring labels, ideally so that each pixel is designated a unique label [70]. Our experience shows that CBRL algorithm performs well and is efficient without any parameter setting, and the configuration of neighborhood system affects the quality of image segmentation.

Furthermore, by dropping global homogeneity assumption and using Gibbs formulation, MICM algorithm adaptively incorporates boundary information into the decision maker (through Markov parameters in local energy function). The advantages are that the likelihood of phase transition is reduced [45] and the probability of correct classification is increased. MICM algorithm uses a specified iterations through random pixel visit. At each single cycle in the iterations, for updating l_i, we have

$$P(\mathbf{l}|\mathbf{x}) = p(l_i|\mathbf{l}_{S|i}, \mathbf{x}) P(\mathbf{l}_{S|i}|\mathbf{x}) \tag{6.18}$$

which never decreases at any stage so the eventual convergence is assured [66].

6.4 Summary and Discussions

In this chapter, we formulate MR image segmentation as a standard estimation and detection problem based on the estimated model parameters and the given pixel images. Based on the proper FNM formulations and appropriate criteria, we introduce two non-contextual classifiers for initial image segmentation: MLC and GBC. Since the spatial distribution of misclassification error resulting from MLC is uniform, MLC is selected for image segmentation initialization.

In order to improve the initial image segmentation, we derive two contextual algorithms for finalizing MR image segmentation: CBRL and MICM. Within the framework of consistent labeling, CBRL algorithm is a non-parametric relaxation labeling approach in which the average local consistency measure is maximized by using local Bayesian detection everywhere based on the localized SFNM formulation. On the other hand, by using MRF modeling, MICM algorithm adopts the conventional ICM approach but by using an inhomogeneous hidden GRF configuration. As a counterpart of deterministic relaxation labeling in MICM algorithm, Markov parameter identification is formulated as a problem of boundary defined contextual parameter identification. We use BDMC algorithm for quantifying Markov parameter values and integrate it into MICM algorithm.

The whole procedure for MR image segmentation is summarized below:

1. Initialize image segmentation by MLC;

2. Initialize Markov parameters θ_i's by BDMC Algorithm;
3. Finalize image segmentation by either CBRL or MICM algorithm;
4. Save the result of pixel classification.

The major advantages of the new approaches are that: the connection between image modeling and image segmentation is clear; in a non-parametric mode, Markovian property is utilized as a spatial consistency constraint in contextual image segmentation by CBRL algorithm that leads to a satisfying segmentation result; boundary information is used by decision maker in MICM algorithm in terms of inhomogeneous MRF configuration so as to achieve a better contextual image segmentation.

Chapter 7

Experimental Verification

7.1 Introduction

In this chapter, we apply the algorithms developed in chapters 5 and 6 to unsupervised MR image segmentation, and show that the algorithms perform consistently well as defined by the proper criteria chosen. We first test the algorithms with simulated data to highlight specific characteristics of the algorithms by presenting a tone image with controlled parameter values. The methods used for the verification are:

1) Global relative entropy (GRE) between the histogram and the estimated SFNM distribution for regional parameter estimation;

2) Correct rate of information theoretic criteria (with SNR[1] as a parameter) for structural parameter detection;

3) Visual inspection and percentage misclassification evaluation for image segmentation.

The images used in this verification are:

1) Simulated image (Figure 7.1)

2) Real MR image (MRI #1) (Figure 7.2)

3) Real MR image (MRI #2) (Figure 7.3)

4) Real MR image (MRI #3) (Figure 7.4)

5) Test image (Figure 7.5)

The parameter settings of simulated and test images are given in Table 7.1 and 7.2.

[1] The signal-to-noise ratio (SNR) is defined by

$$SNR = 10 \log_{10} \frac{(\Delta\mu)^2}{\sigma^2}$$

where $\Delta\mu$ is the mean difference between regions, and σ^2 is the noise power.

Table 7.1: Parameter settings of Simulated Image (SNR=6 dB)

k	π	μ	σ^2
1	0.25	86	400
2	0.125	126	400
3	0.5	166	400
4	0.125	206	400

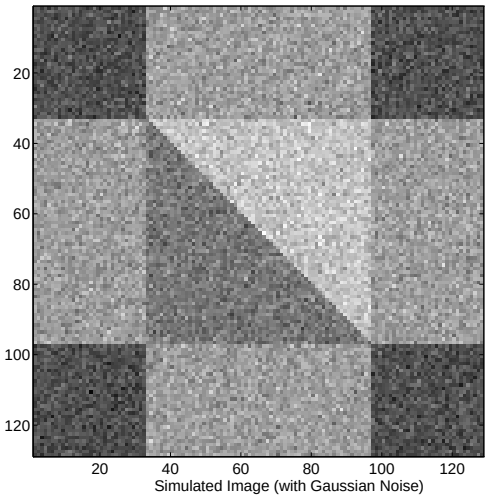

Figure 7.1: Simulated Image ($K_0 = 4$, SNR=6 dB)

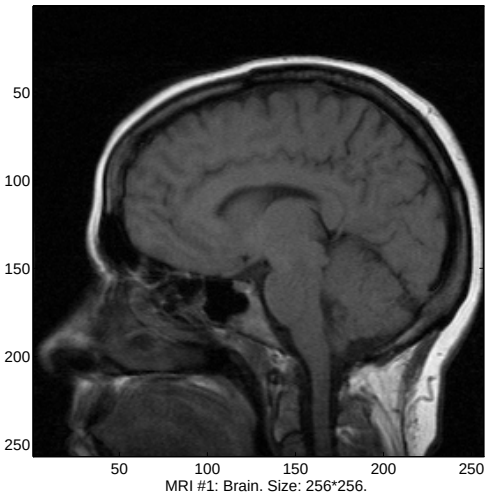

Figure 7.2: Real MR Image (MRI #1) ($K_0 = 8$)

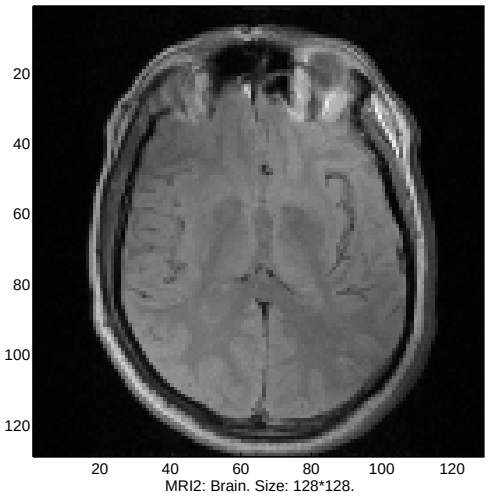

Figure 7.3: Real MR Image (MRI #2) ($K_0 = 10$)

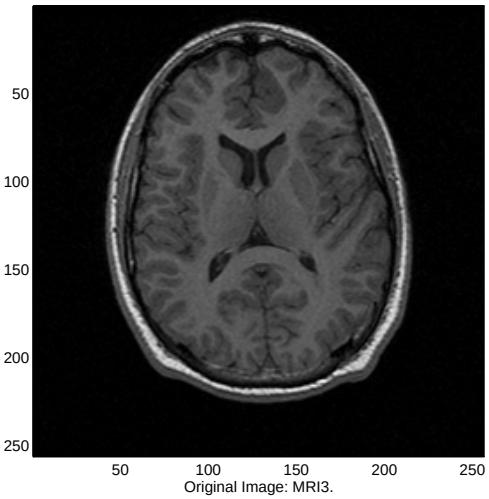

Figure 7.4: Real MR Image (MRI #3) ($K_0 = 8$)

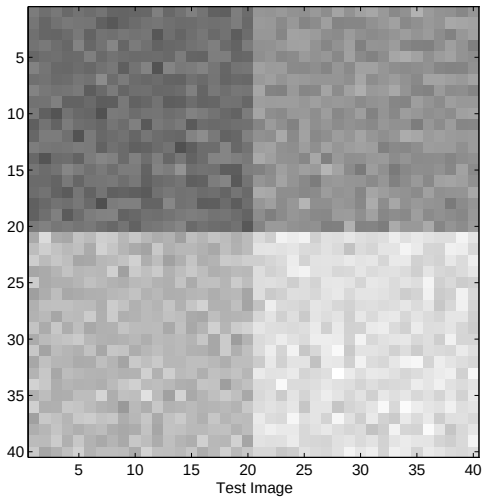

Figure 7.5: Test Image ($K_0 = 4$, SNR=10 dB)

Table 7.2: Parameter Setting for Test Image (SNR=10 dB).

k	π	μ	σ^2
(Case #1)			
1	0.25	30	9
2	0.25	40	9
3	0.25	50	9
4	0.25	60	9
(Case #2)			
1	0.25	300	900
2	0.25	400	900
3	0.25	500	900
4	0.25	600	900
(Case #3)			
1	0.25	3000	90000
2	0.25	4000	90000
3	0.25	5000	90000
4	0.25	6000	90000

Table 7.3: Result of ALMHQ for Simulated Image

k	π	μ	σ^2
1	0.234	81.45	235
2	0.234	130.54	158
3	0.364	167.45	157
4	0.185	204.54	177

This chapter is organized as follows: Section 7.2 mainly focuses on the verification of parameter estimation. In Section 7.3, five different approaches for image segmentation are tested and compared. A summary and discussion of results is presented in Section 7.4. Figure 7.6 shows the major experiments in the verification and their relationships to related algorithms.

7.2 Parameter Estimation

7.2.1 Initialization of Regional Parameters

The initialization of regional parameters $(\pi_k, \mu_k, \sigma_k^2)$ is conducted in two steps: First ALMHQ algorithm (Eqs. (5.3) - (5.4)) is used to initialize parameter estimates by quantizing the histogram of the observed image, and then BCM algorithm (Eqs. (5.7 - 5.8) is used to refine the initialization results.

Verification of ALMHQ algorithm is carried out by experiment 1 on the simulated image. The results are shown in Figure 7.7 and Table 7.3 where the noisy curve is the observed histogram and the smooth curve is the initialized SFNM distribution. In the figure, four peaks are clearly distinguished without being affected by the noise and parameter values are close to the true values given in Table 7.1.

Another method for parameter initialization is the CM algorithm used in [62]. The CM algorithm however needs an initialization itself and highly depends on the data structure. In our experimental studies, ALMHQ algorithm performs quite satisfactorily in most of cases, and provides a robust method for systematic parameter initialization.

ALMHQ algorithm is also applied to the real MR images: MRI #1, MRI #2, and MRI #3. Figure 7.8 shows the result of ALMHQ initialization for MRI #1, and Table 7.4 and Table 7.5 show the results of ALMHQ for MRI #2 and MRI #3 respectively. [2]

Using the results of ALMHQ algorithm as the input, BCM algorithm is performed to refine the regional parameter initialization. For simulated image results, a plot of global relative entropy with different resolutions is given in Figure 7.9, which shows that the suitable resolution is $b_0 = 3$ (for $b = 1, ..., 7$). At this resolution, the GRE value is 0.0081258 nats. The refined parameter initialization by BCM algorithm is shown in Figure 7.10. Based on our comparative numerical experiments, it is seen that BCM algorithm dramatically reduces the bias in the initialization and has the ability to reduce the likelihood of being trapped in a local minima because of its multi-resolution evolution. The refined regional parameter estimates of simulated image is given in Table 7.6.

For real MR image MRI #1, since the true number of image regions is not available, we use information criteria to detect this number and present the corresponding results. Model selection procedure by information theoretic criteria (AIC/MDL/MCBV) (see Section 5.4)) suggested that $K_0 = 8$ (over $K = 2 \sim 10$), and Eq. (5.6) suggested that the suitable resolution is $b_0 = 4$ (over $b_0 = 1 \sim 5$). The multi-resolution segmentation results are shown in Figure 7.11 and the refined regional parameter estimates by BCM algorithm are given in Table 7.7. A quite low GRE is obtained, it is about 0.056887 nats.

[2] Since we can not differentiate estimated distribution from histogram, we have used for MRI # 2 and MRI # 3.

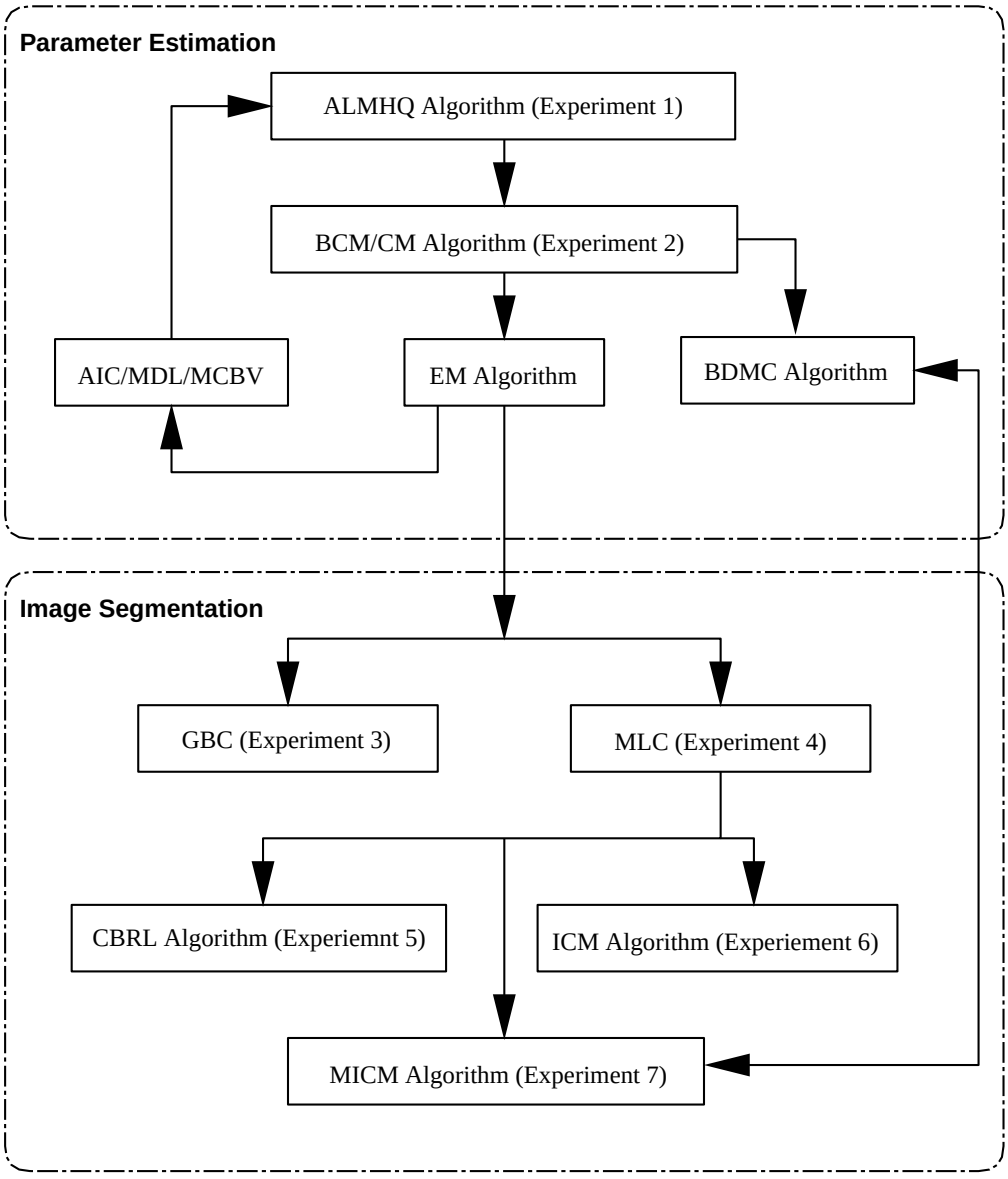

Figure 7.6: Flowchart of Experimental Verification

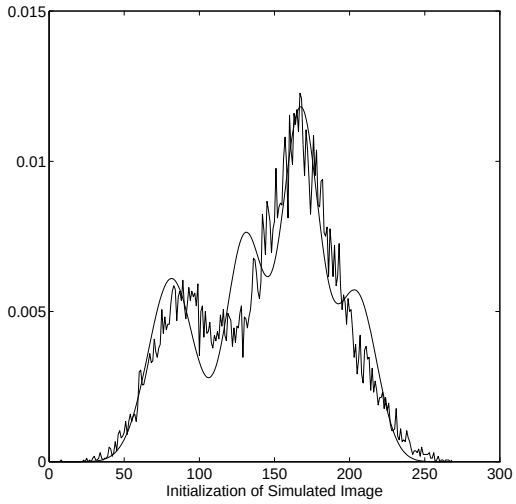

Figure 7.7: Result of ALMHQ for Simulated Image (GRE=0.0398998 nats)

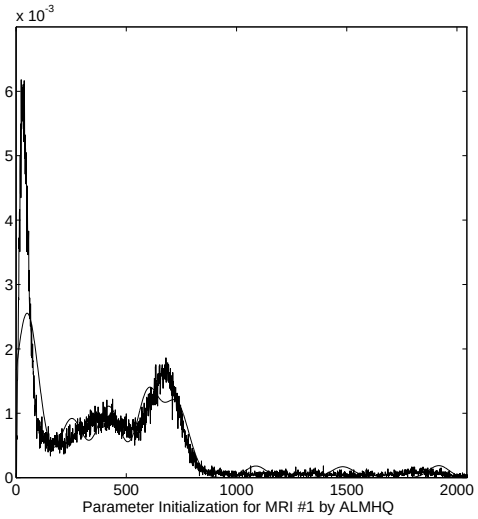

Figure 7.8: Result of ALMHQ for MRI #1 (GRE=0.25786 nats)

Table 7.4: Result of ALMHQ for MRI #2 (GRE=0.167294 nats)

k	π	$\mu(\times 10^3)$	$\sigma^2(\times 10^3)$
1	0.39531541	0.07028045	4.77365140
2	0.03607177	0.38971954	3.30403484
3	0.04278564	0.72428045	3.77127442
4	0.04510498	1.03571954	4.42601973
5	0.06683349	1.38428045	4.43038947
6	0.11804199	1.61371954	4.77525354
7	0.15515136	1.83628045	6.06341221
8	0.10571289	2.05771954	6.31367534
9	0.02386474	2.40428045	3.87563688
10	0.01111769	2.95171954	7.77271762

Table 7.5: Result of ALMHQ for MRI #3 (GRE=0.444136 nats)

k	π	μ	σ^2
1	0.5521	5.8421	63.8375
2	0.0428	42.1579	39.9845
3	0.0585	77.8421	41.9493
4	0.1356	104.1579	50.5766
5	0.1085	125.8421	40.8023
6	0.0957	148.1579	72.2267
7	0.0139	233.8421	73.1012
8	0.0110	318.1579	51.3025

Table 7.6: Refined regional parameter estimates for Simulated Image by BCM Algorithm ($b_0 = 3$) (GRE=0.0081258 nats)

k	π	μ	σ^2
1	0.2295	84.1069	351.9481
2	0.1272	120.9573	427.1215
3	0.4807	163.4669	410.2669
4	0.1626	200.1592	486.4619

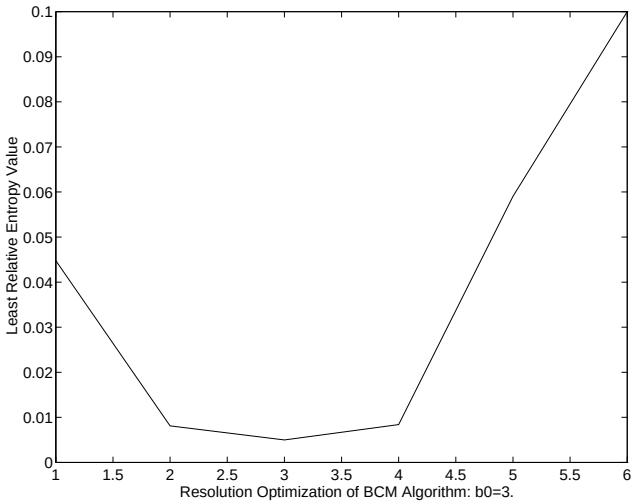

Figure 7.9: GRE Curve of b for Simulated Image.

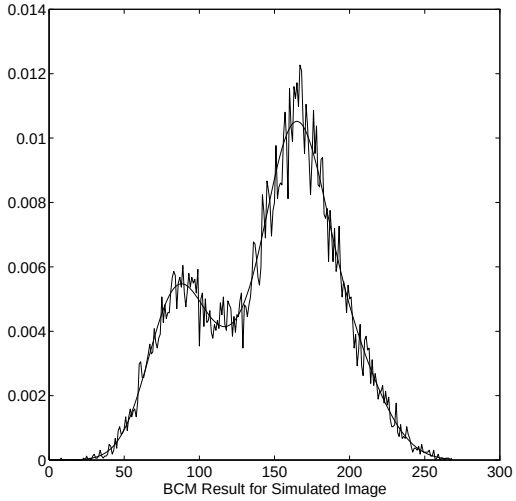

Figure 7.10: Result of BCM Algorithm: Simulated Image. ($b_0 = 3$, GRE=0.0081258 nats)

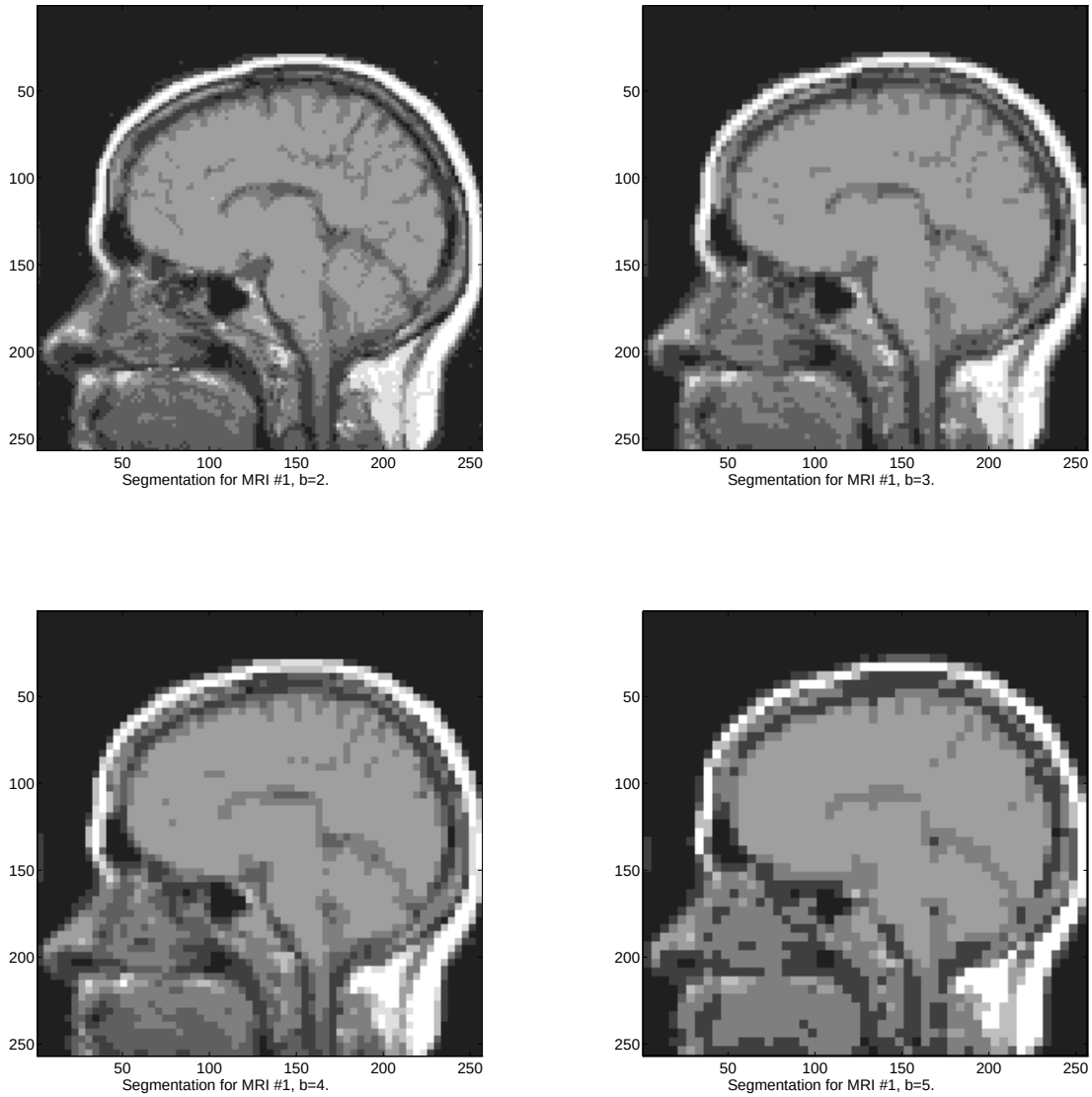

Figure 7.11: Multi-resolution Segmentation: MRI #1.

Table 7.7: Refined regional parameter estimates for MRI #1 by BCM algorithm ($b_0 = 4$).

k	π	$\mu(\times 10^3)$	$\sigma^2(\times 10^5)$
1	0.2999	0.0429	0.0058
2	0.1150	0.1926	0.0707
3	0.1607	0.3800	0.0614
4	0.1303	0.5544	0.0482
5	0.2085	0.6961	0.0317
6	0.0235	0.7038	1.1281
7	0.0321	1.2151	0.8111
8	0.0296	1.7532	0.3415

Table 7.8: Refined regional parameter estimates for MRI #2 by BCM algorithm ($b_0 = 3$)

k	π	$\mu(\times 10^3)$	$\sigma^2(\times 10^5)$
1	0.3218	0.0625	0.0053
2	0.0466	0.2577	0.5594
3	0.0769	0.7148	1.7575
4	0.0549	1.1685	4.0152
5	0.0576	1.4084	1.1489
6	0.1257	1.6082	0.1995
7	0.0076	1.7050	9.6741
8	0.1494	1.8475	0.1323
9	0.0763	2.0469	0.2087
10	0.0516	2.1470	4.2307

The same procedure is applied to MRI #2. Model selection procedure suggested that $K_0 = 10$ over $K = 2 \sim 12$ and Eq. (5.6) suggested that the suitable resolution is $b_0 = 3$ over $b = 1 \sim 4$. The multi-resolution segmentation results are shown in Figure 7.12. The refined regional parameter estimates by BCM algorithm are given in Table 7.8. The GRE is reduced from 0.167294 nats of ALMHQ result to about 0.165476 nats.

7.2.2 Finalization of Regional Parameter Estimates

BCM reduces the likelihood of being trapped in local minima and offers fast computation but since most often the suitable resolution determined by Eq. (5.6) is not 1, it only provides parameter estimates at a coarse resolution. Hence EM algorithm which is known to be asymptotically convergent is employed to finalize parameter estimates.

Figure 7.13 gives the result of EM algorithm for simulated image. The GRE value is about 0.0080381 nats. The final regional parameter estimates for simulated image are given in Table 7.9.

For real MR image MRI #1, Figure 7.14 gives the plots of the histogram and the final estimated SFNM distribution. The final GRE value is about 0.030688 nats. The regional parameter estimates are presented in Table 7.10.

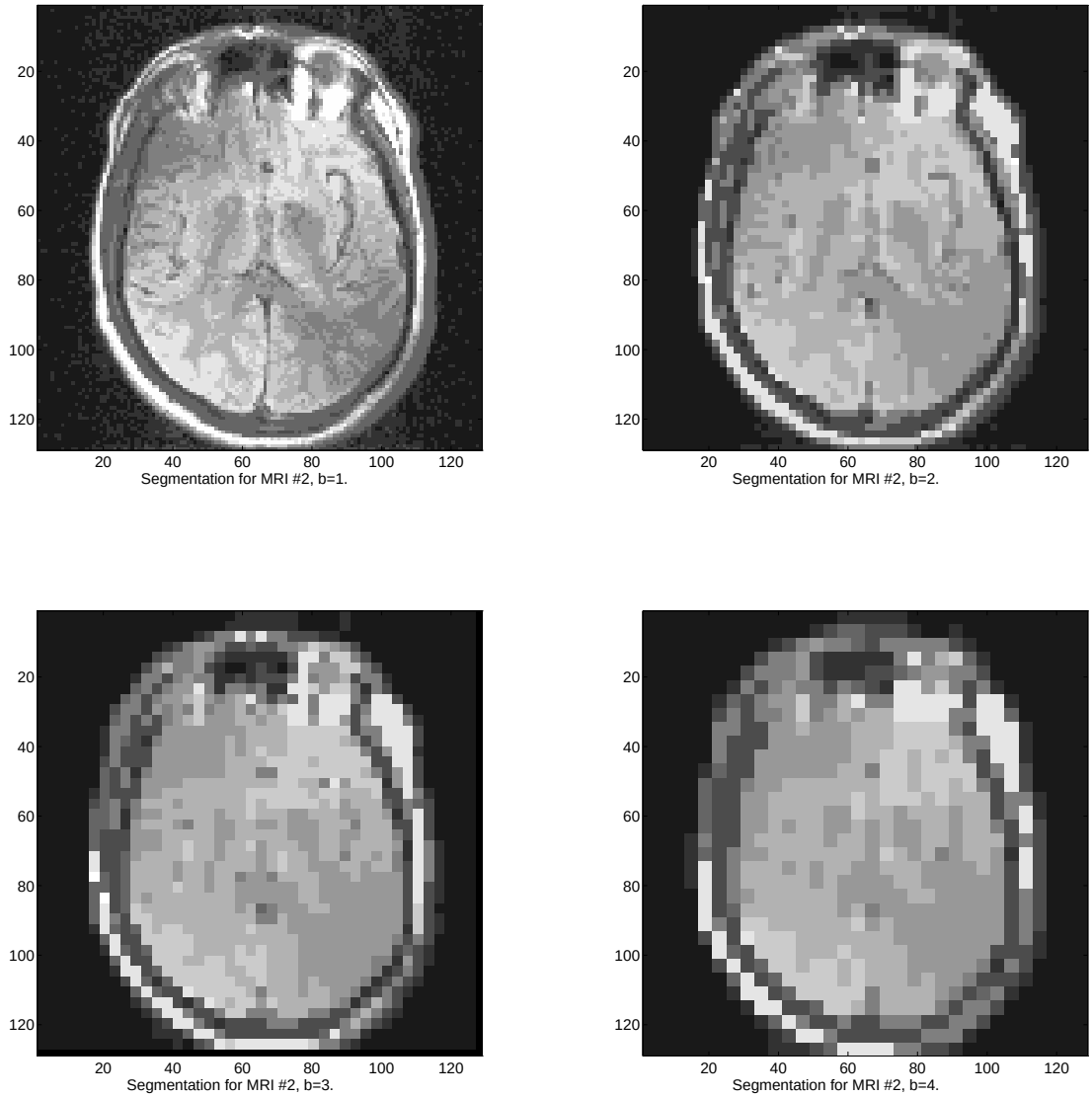

Figure 7.12: Multi-resolution Segmentation: MRI #2.

Table 7.9: Final Parameter Estimates for Simulated Image by EM Algorithm. (GRE=0.0080381)

k	π	μ	σ^2
1	0.2303	83.9360	353.7141
2	0.1352	121.3515	365.4675
3	0.4773	164.2710	373.2707
4	0.1572	201.4632	462.5701

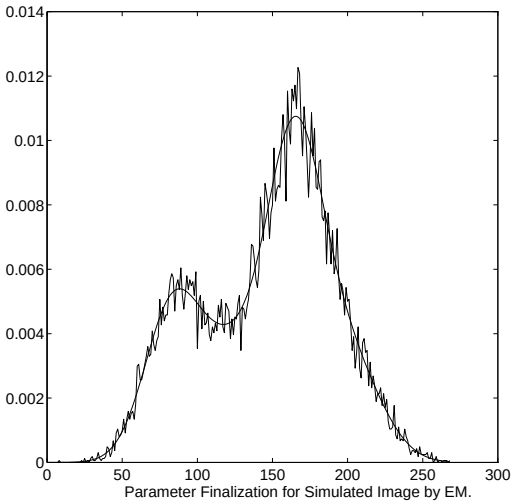

Figure 7.13: Result of EM Algorithm for Simulated Image. (GRE=0.0080381)

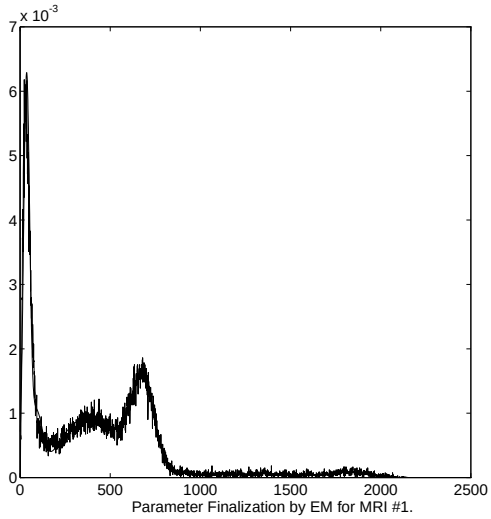

Figure 7.14: Result of EM Algorithm for MRI #1 .

Table 7.10: Final Parameter Estimates for MRI #1 by EM Algorithm. (GRE=0.030688)

k	π	$\mu(\times 10^3)$	$\sigma^2(\times 10^4)$
1	0.2013	0.0321	0.0197
2	0.1102	0.0679	0.0700
3	0.2450	0.3060	1.5502
4	0.1690	0.5510	1.4528
5	0.1872	0.6832	0.3494
6	0.0342	0.8174	4.5394
7	0.0323	1.3927	5.1709
8	0.0206	1.8456	1.1330

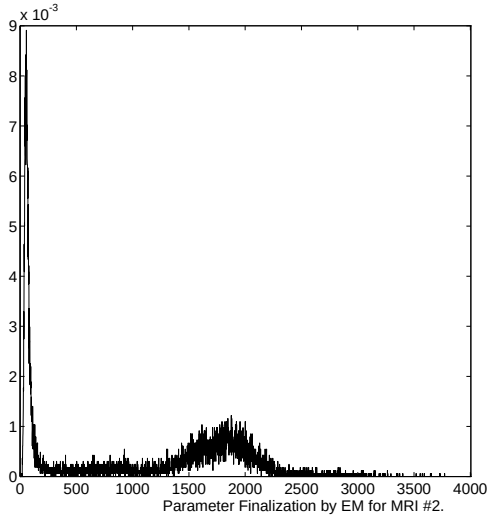

Figure 7.15: Result of EM algorithm for MRI #2.

For MRI #2, Figure 7.15 gives the histogram and the estimated SFNM distribution. Table 7.11 presents the final regional parameter estimates of MRI #2. The GRE value is around 0.165057 nats.

We also conduct an experiment with MRI #3 in which the initial parameter estimates from ALMHQ algorithm are finalized directly by the EM algorithm. The results are given in Table 7.12. It is observed that when applied without refinement by the BCM algorithm, EM algorithm requires more iterations to obtain a similar low GRE value as with BCM.

Finally, a brief summary on the performance of the algorithms for regional parameter estimation is given by Table 7.13 (given the correct number of image regions).

7.2.3 Model Selection

One of the major steps in unsupervised image analysis is the detection of the model structural parameter K, i.e., the number of image regions. This detection task is implemented by using a model fitting procedure based on information theoretic criteria. In this section, we apply AIC (Eq. (5.12)), MDL (Eq. (5.14)), and

Table 7.11: Final Parameter Estimates for MRI #2 by EM Algorithm. (GRE=0.165057)

k	π	$\mu(\times 10^3)$	$\sigma^2(\times 10^5)$
1	0.2950	0.0563	0.0023
2	0.0893	0.1049	0.0129
3	0.0822	0.5396	0.6882
4	0.0574	1.2512	1.4812
5	0.0610	1.3130	1.2668
6	0.1251	1.5948	0.1539
7	0.1419	1.8537	0.1010
8	0.0071	1.8865	3.4486
9	0.0811	2.0513	0.1324
10	0.0593	2.2352	2.7228

Table 7.12: Final Parameter Estimates for MRI #3 by EM Finalization (GRE=0.036097 nats)

k	π	μ	$\sigma^2(\times 10^3)$
1	0.3250	3.3325	0.0028
2	0.2104	7.6726	0.0091
3	0.0826	46.6317	0.5101
4	0.1688	103.2050	0.1962
5	0.0745	120.9055	0.1518
6	0.0998	142.7683	0.0745
7	0.0252	181.1221	2.9174
8	0.0138	296.2685	1.0604

Table 7.13: Summary of Regional Parameter Estimation (GRE Values)

Algorithm/Image	Simulated Image	MRI #1	MRI #2	MRI #3
ALMHQ	0.0398998	0.25786	0.167294	0.444136
BCM	0.0081258	0.056887	0.165476	-
EM	0.0080381	0.030688	0.165057	0.036097

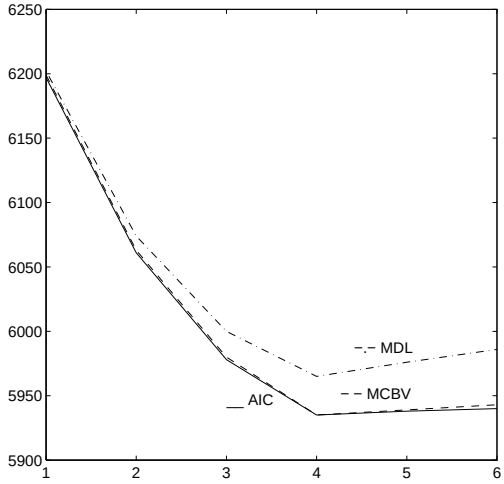

Figure 7.16: AIC/MDL/MCBV Curves of Test Image (Case #1)

Table 7.14: Number of image regions detected by information criteria

Test Image	Case #1	Case #2	Case #3
SNR (dB)	10	10	10
AIC	4	4	4
MDL	4	4	4
MCBV	4	4	4

MCBV (Eq. (5.22)) criteria to the test image (figure 7.5) for three cases where the data structures and the SNRs are the same, but the variances are different.

First, we conduct the model fitting procedure for the test image to evaluate the performances of the three information theoretic criteria. The AIC/MDL/MCBV curves, as a function of the number of image regions K, are plotted in Figures 7.16, 7.17, and 7.18.

Table 7.14 summarizes the number of image regions estimated by these criteria. Figures 7.16 7.17 7.18 and Table 7.14 show that for these three cases for the test image, all AIC, MDL, and MCBV criteria can reach their minima. They perform consistently well and correctly identify the number of image regions for these cases.

Applying AIC/MDL/MCBV criteria to MRI #1, #2, and #3, we obtain the results: $K_0 = 8$ for MRI #1, $K_0 = 10$ for MRI #2, and $K_0 = 8$ for MRI #3. For MRI #1 and #2, the curves are shown in Figure 7.19 and 7.20. For MRI #3, the results of all these three information criteria are given in Table 7.15. Our experiments show that all these three criteria provide consistent results.

7.2.4 Markov Parameter Configuration

Based on the interim image segmentation results from BCM or MLC or MICM, Markov parameter identification is performed by using entropy rate estimation and BDMC algorithm. By first calculating the map

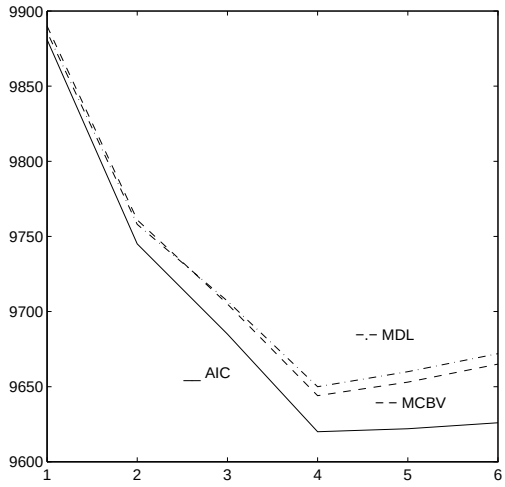

Figure 7.17: AIC/MDL/MCBV Curves of Test Image (Case #2)

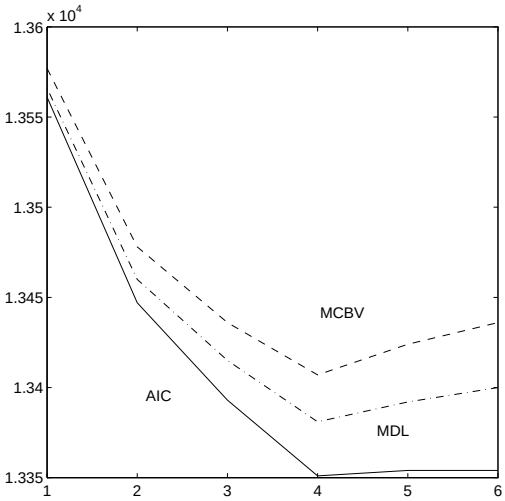

Figure 7.18: AIC/MDL/MCBV Curves of Test Image (Case #3)

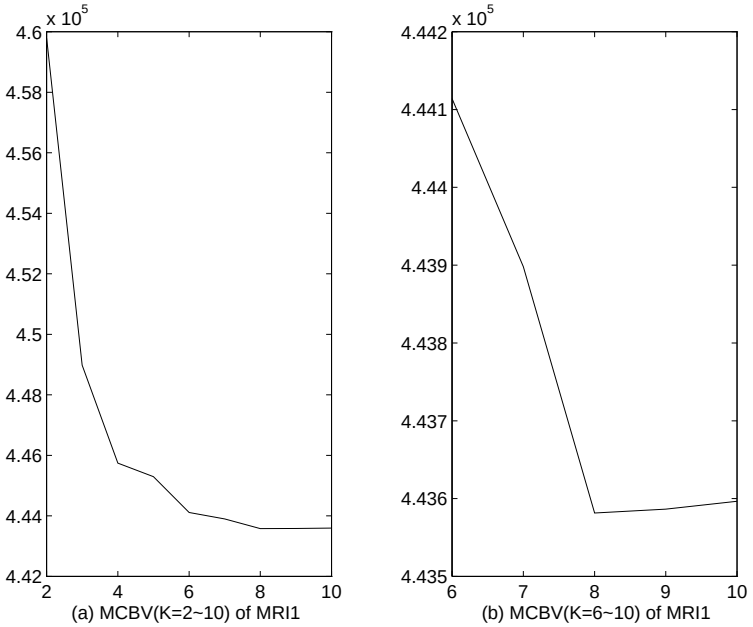

Figure 7.19: MCBV Curve of MRI #1

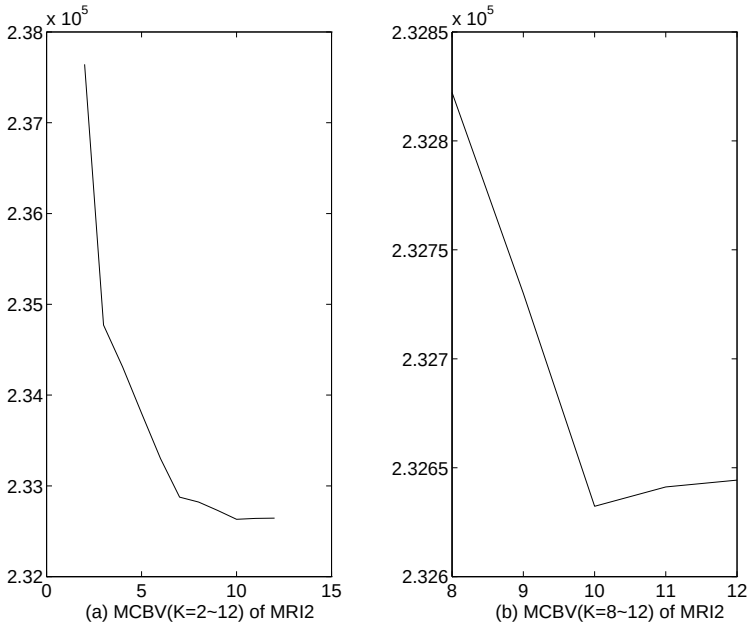

Figure 7.20: MCBV Curve of MRI #2

Table 7.15: The Values of AIC, MDL, and MBVC for MRI #3

K	AIC	MBVC	MDL
2	295944	295948	295967
3	291667	291674	291703
4	288845	288856	288895
5	288397	288410	288461
6	288380	288395	288458
7	288184	288205	288275
8	**285685**	**285706**	**285790**
9	287674	287693	287792
10	286317	286348	286449

of entropy rate according to Eq. (5.39), the inhomogeneous Markov parameter assignments are conducted based on a look-up table defined by Eq. (5.37). In [45], after extensive numerical experiments it is reported that the assignment of $\theta = 0.4$ along the boundaries and $\theta = 1.5 - 6.0$ within the smooth regions, resulted in good segmentation. By using these empirical knowledge as a guideline, we use $\alpha = 0.5199$ and $\eta = 0.0867$ in our experiments (Eq. (5.41)). Figure 7.21 gives the numerical Markov parameter assignments for the simulated image. It can be seen that the proposed method (BDMC algorithm) provides a satisfactory result in which all boundaries are clearly detected and the noise effect is highly suppressed. However, edges are blurred.

Similarly, BDMC algorithm is also applied to MRI #1, #2, and #3. Figure 7.22 shows one of the corresponding results for MRI #2 where the value of θ_i takes values between 0.4 to 6.0.

7.3 Image Segmentation

7.3.1 Initialization of Image Segmentation

In this section, we apply MLC (GBC), CBRL, and MICM algorithms to obtain image segmentation results for the simulated image and three MR images, in which the MLC (GBC) algorithms are used for the initialization of the segmentation, and CBRL and MICM algorithms are used for finalizing the segmentation results. For assessment of image segmentation quality, in addition to visual inspection, we also use the percentage of misclassification error as a quantitatively measure

In Experiment 3, GBC is tested with simulated image for image segmentation. The result is presented in Figure 7.23, where misclassification error is around 15%. We notice that even using global context information, GBC reproduces a poor recovery of true context when SNR is low. Also, the misclassification errors are much higher in minor regions than in major regions.

In Experiment 4, MLC is used to initialize image segmentation for the simulated image resulting in a misclassification error of about 30%. The result is shown in Figure 7.24 where we notice that MLC produces a quite noisy segmented image. However, since misclassification by MLC are uniformly distributed over the whole image, it is more suited for image segmentation initialization than the GBC approach. Hence, in this work, we use MLC for the initialization of image segmentation. Figure 7.25 and 7.26 show the corresponding initial segmentation results for MLC when applied to MRI #1 and MRI #2,

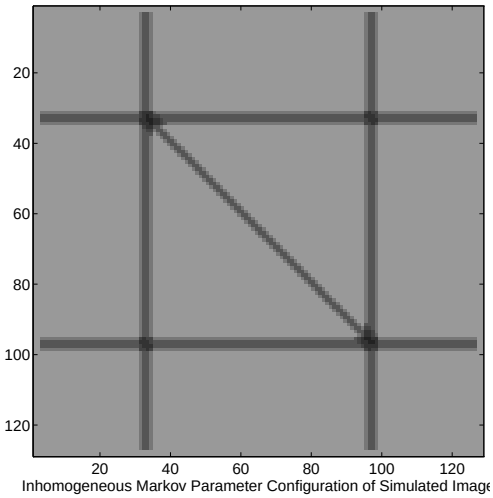

Figure 7.21: Inhomogeneous Markov Parameter Configuration by BDMC: Simulated Data.

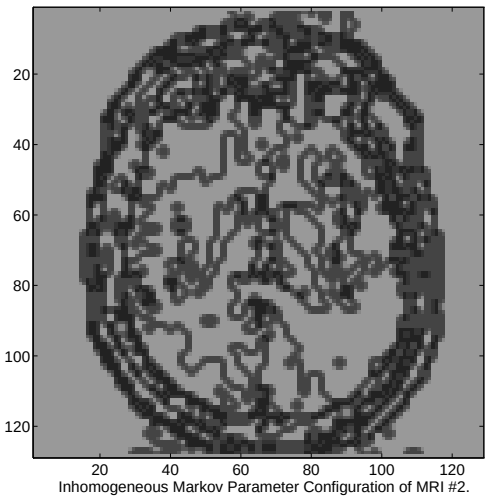

Figure 7.22: Markov Parameter Configuration by BDMC Algorithm: MRI2.

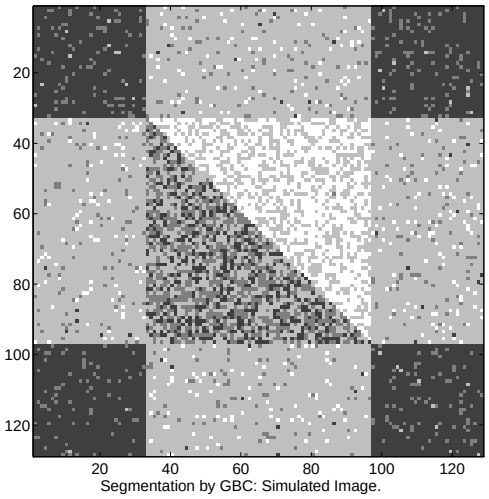

Figure 7.23: Result of GBC Algorithm for Simulated Image.

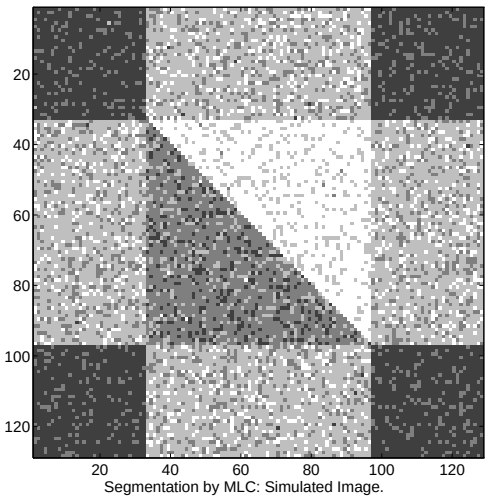

Figure 7.24: Result of MLC Algorithm: Simulated Image.

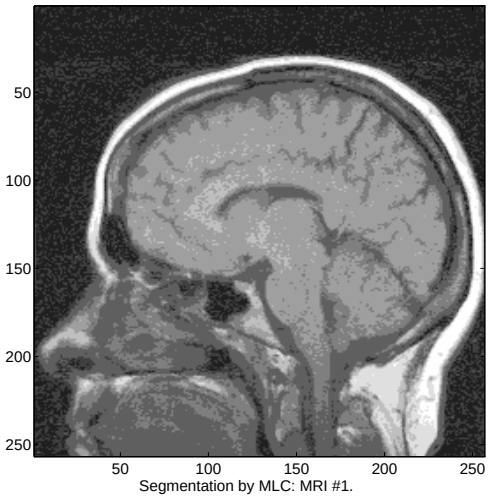

Figure 7.25: Result of MLC Algorithm for MRI #1.

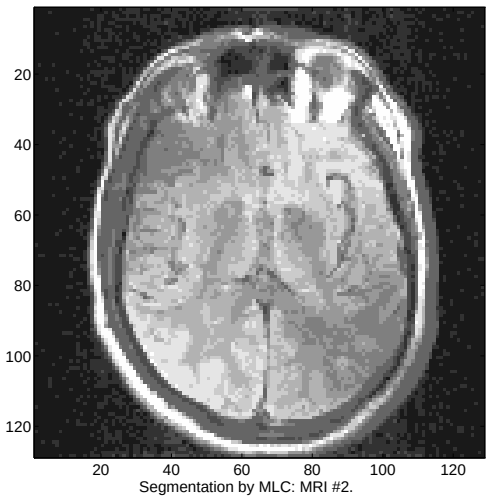

Figure 7.26: Result of MLC Algorithm for MRI #2.

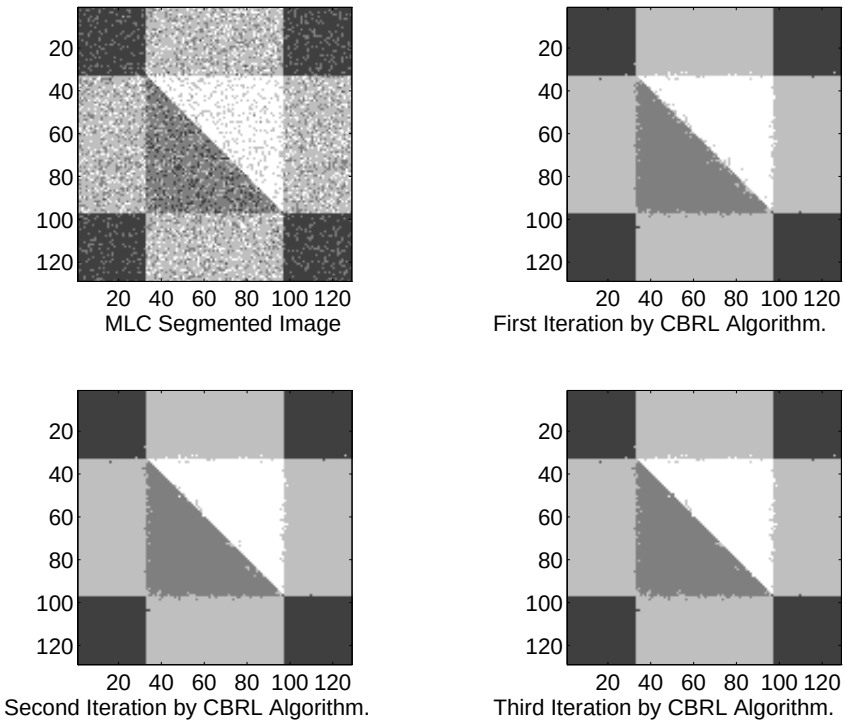

Figure 7.27: Image Segmentation by CBRL: Simulated Image.

7.3.2 Finalization of Image Segmentation

Using the MLC result as a starting point, contextual image segmentation is implemented first by the CBRL algorithm in Experiment 5. For simulated image, the result is shown in Figure 7.27 where images (a), (b), (c) are the results from the first three iterations of CBRL. Clearly, a dramatic improvement is obtained by using local context in terms of non-parametric Bayesian binding. Also, the convergence is fast since after the first iteration most of the misclassification are removed. The final percentage of misclassification errors is about 0.7935%.

For real MR images #1 and #2, the results (after three iterations) are given in Figure 7.28 (a) and (b). By visual inspection, the final segmentations of these images are of acceptable quality.

CBRL is also applied to MRI #3 to segment major ROIs (corresponding to major tissue types). The segmented major and minor ROIs are presented in Figure 7.29 and 7.30, respectively. Among these segmented images, region 1 mainly represents white matter, region 2 mainly includes gray matter and subcutaneous fat, region 3 includes CSF and internal occiptal protuberance and skin, region 4 represents partial volume effects and temporalis muscle, region 5-6 correspond to the skull, and region 7-8 represent air and background.

The contextual image segmentation is also implemented by MICM algorithm. In Experiment 6, we apply MLC and conventional ICM algorithms to the simulated image. In Experiment 7, we apply MLC, BDMC, and MICM algorithms to the same simulated image. Finally, we compare the results from MICM and ICM algorithms, as well as those from MICM and CBRL algorithms.

ICM algorithm uses a homogeneous MRF configuration. Two settings are chosen for the Markov parameters, they are: 1) $\theta = 0.4$ and 2) $\theta = 6$. The results from these two settings are shown in Figure 7.31 and 7.32, respectively. These two Figures show that: 1) when the Markov parameter θ is chosen too small, the misclassification due to the noise effect will not be removed out; 2) when the Markov parameter θ is chosen

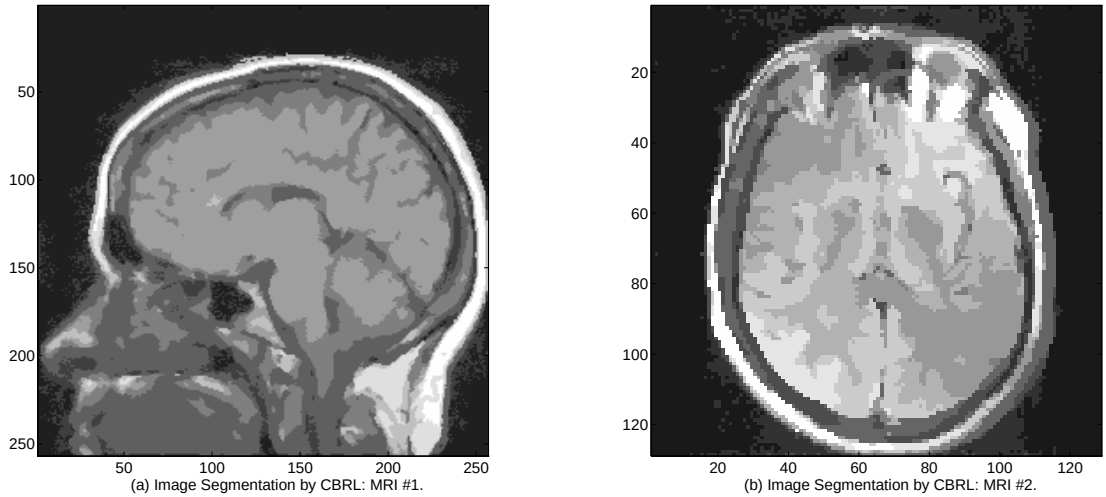

Figure 7.28: Result of CBRL Algorithm: (a) MRI #1, (b) MRI #2.

Table 7.16: Comparison of CBRL, ICM, and MICM Algorithms: Simulated Data.

Item	CBRL Result	ICM Result	MICM Result
Misclassification Error	0.7935%	0.7508%	0.3113%

too big, the recovered context is too "smooth" such that many local details are missing, for example, the sharp angles. Our experiments suggest $\theta = 1.5$ as best choice for the simulated image. The corresponding percentage of misclassification error for this choice of θ is about 0.7508%.

MICM algorithm uses an inhomogeneous Markov parameter configuration. Markov parameter settings are given by BDMC algorithm. The segmentation results are shown in Figure 7.33, which provides almost "perfect" true context information. It is clear that in Figure 7.33 the noise effect is almost completely removed and the image details are preserved. Most probably, the reason for this improvement is that for this case an inhomogeneous Markov configuration, i.e., a more appropriate parameter setting, is used. The percentage of misclassification error is about 0.3113%.

Comparison of results of Experiment #6 and #7 reveal that MICM outperforms the conventional ICM, and results of Experiment #5 and #7 reveal that MICM provides a better segmentation than CBRL algorithm. As for the computational speed, our studies show that both CBRL and MICM take about 3-4 iterations to achieve similar performances. Table 7.16 summarizes our numerical comparison of the segmentations from these experiments.

We also apply MICM algorithm to MRI #1, MRI #2, and MRI #3. Figure 7.34 (a) and (b) present the segmented images using MICM algorithm for MRI #1 and MRI #2, and Figure 7.35 and 7.36 present eight segmented regions in MRI #3.

By visual inspection, these results are satisfactory and meaningful, since major image regions which represent skin, bone, white matter, gray matter, cerebrospinal fluid, as well as their mixtures, are well separated. A simple explanation can be stated as follows. Basically, for the brain tissue, three major components are white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). By considering

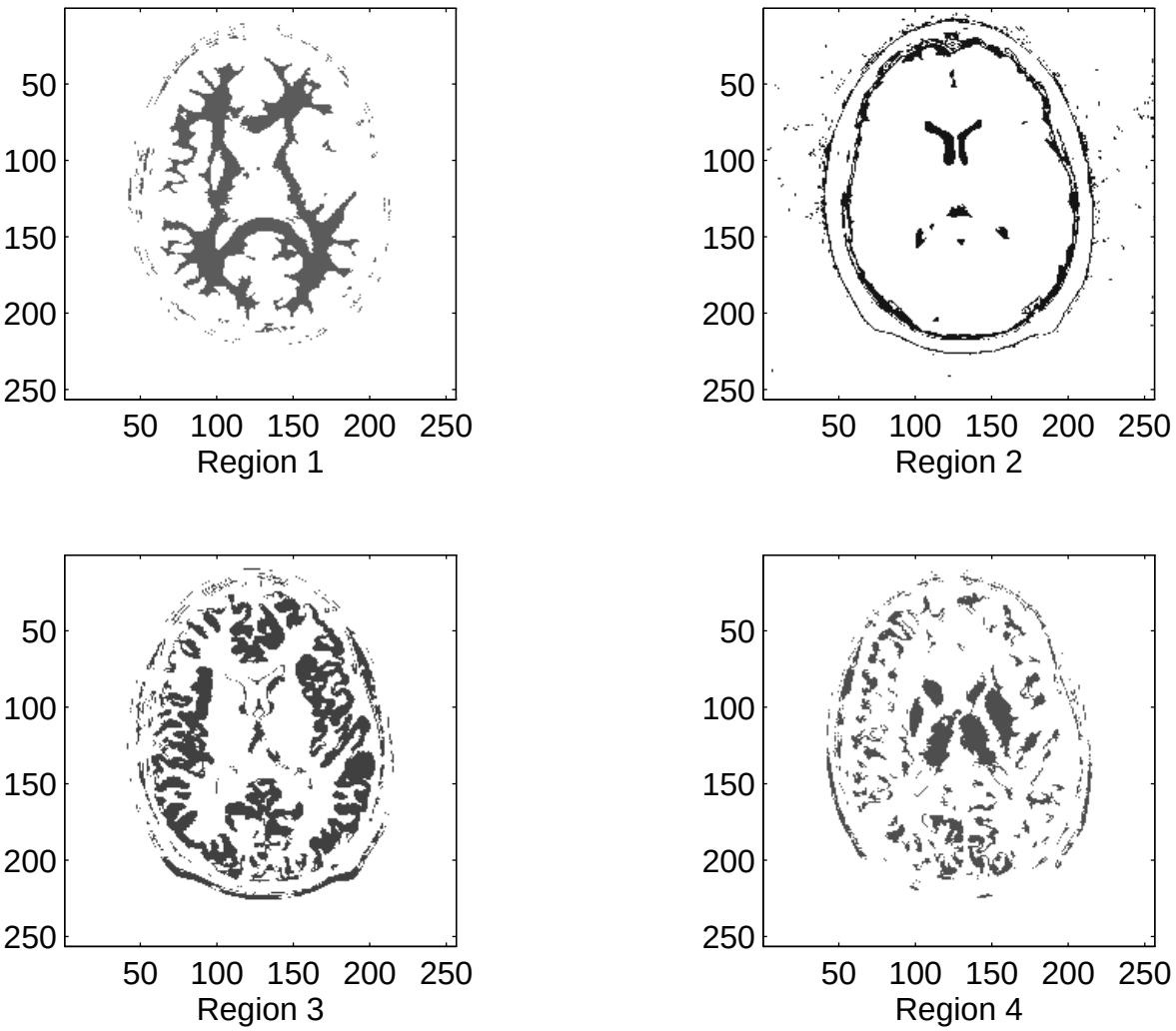

Figure 7.29: Major ROIs in MRI3 (CBRL Result).

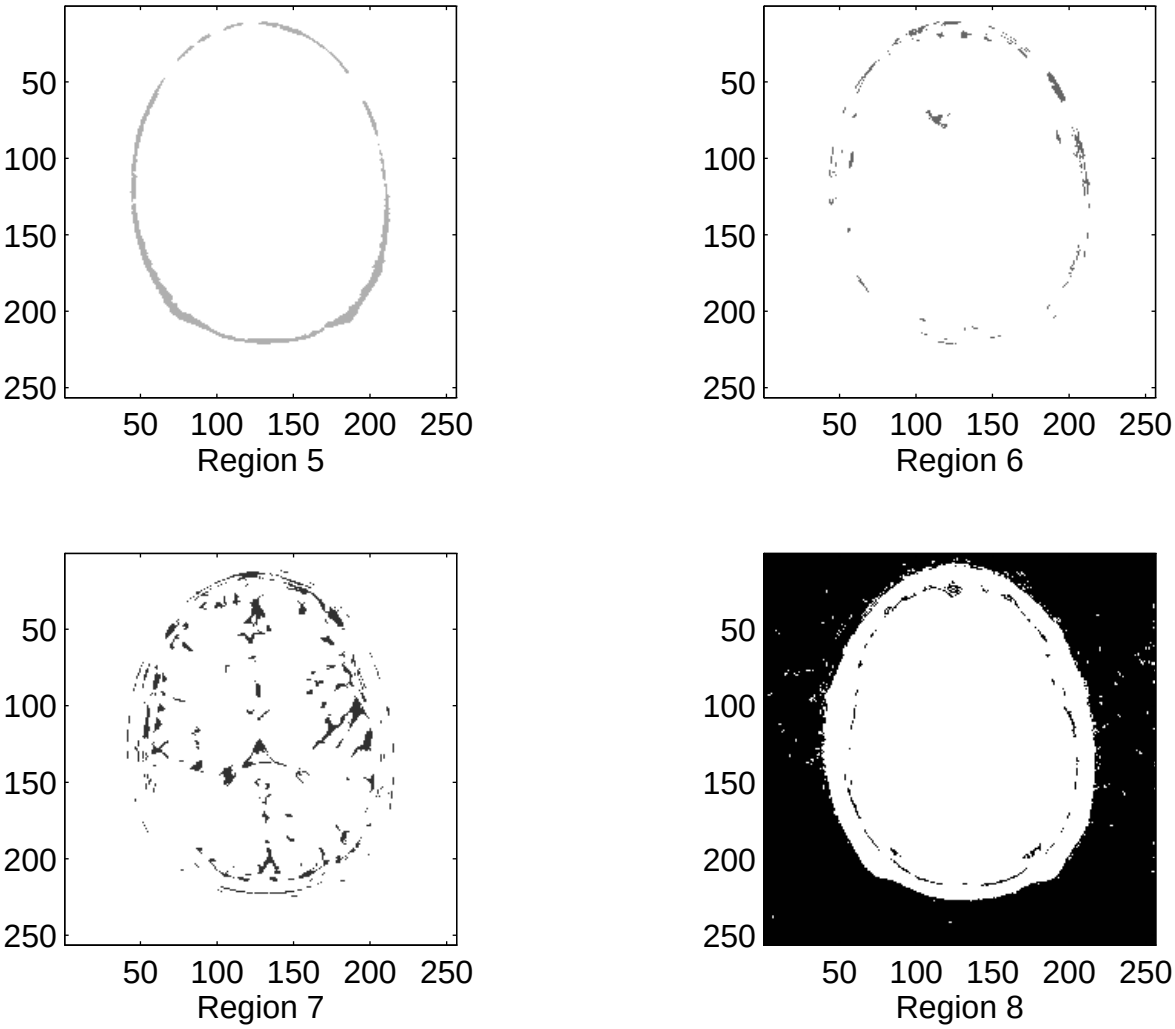

Figure 7.30: Minor Regions in MRI3 (CBRL Result).

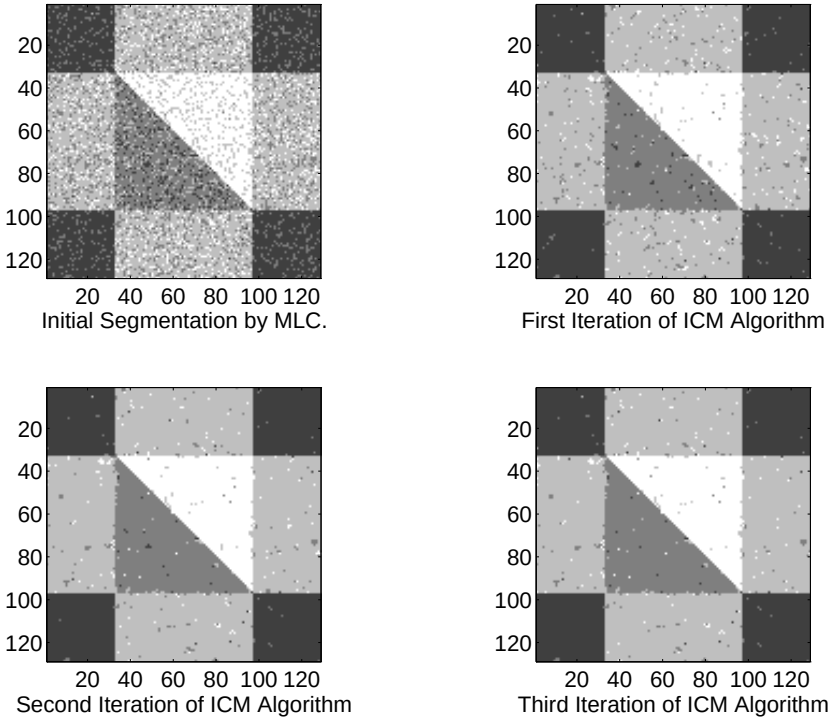

Figure 7.31: Image Segmentation by Conventional ICM: Simulated Image. ($\theta = 0.4$)

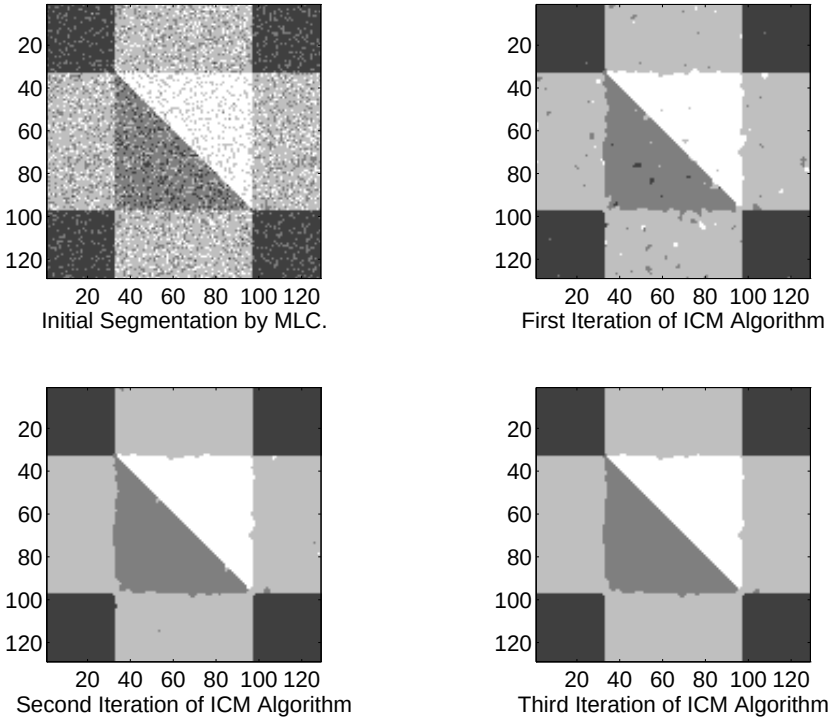

Figure 7.32: Image Segmentation by Conventional ICM: Simulated Image. ($\theta = 6$)

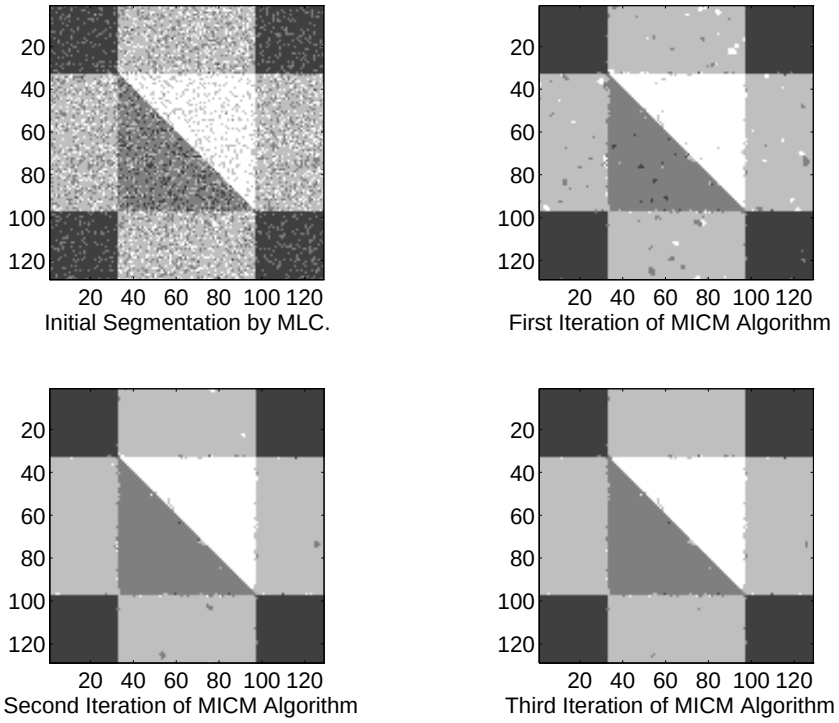

Figure 7.33: Image Segmentation by MICM: Simulated Image.

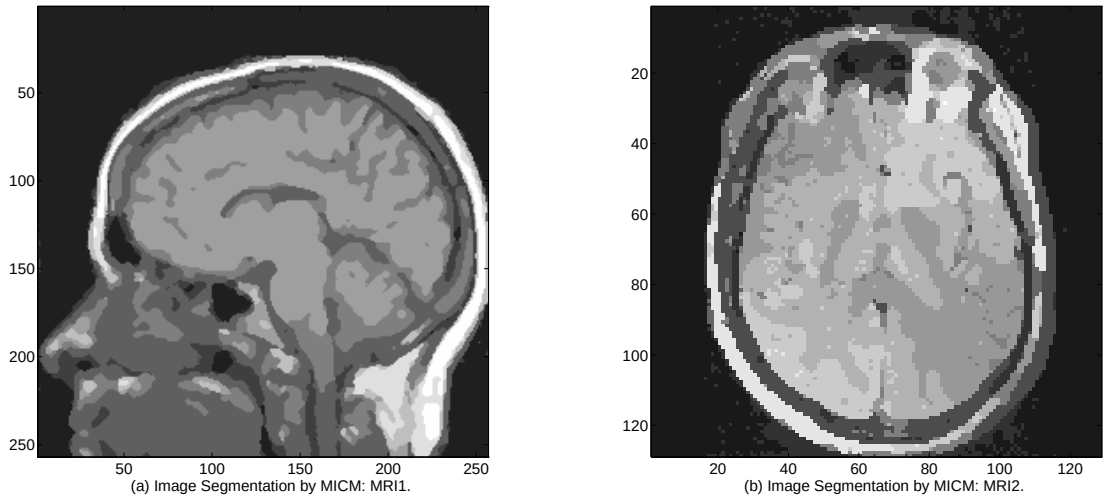

Figure 7.34: Image Segmentation by MICM: (a) MRI #1, (b) MRI #2.

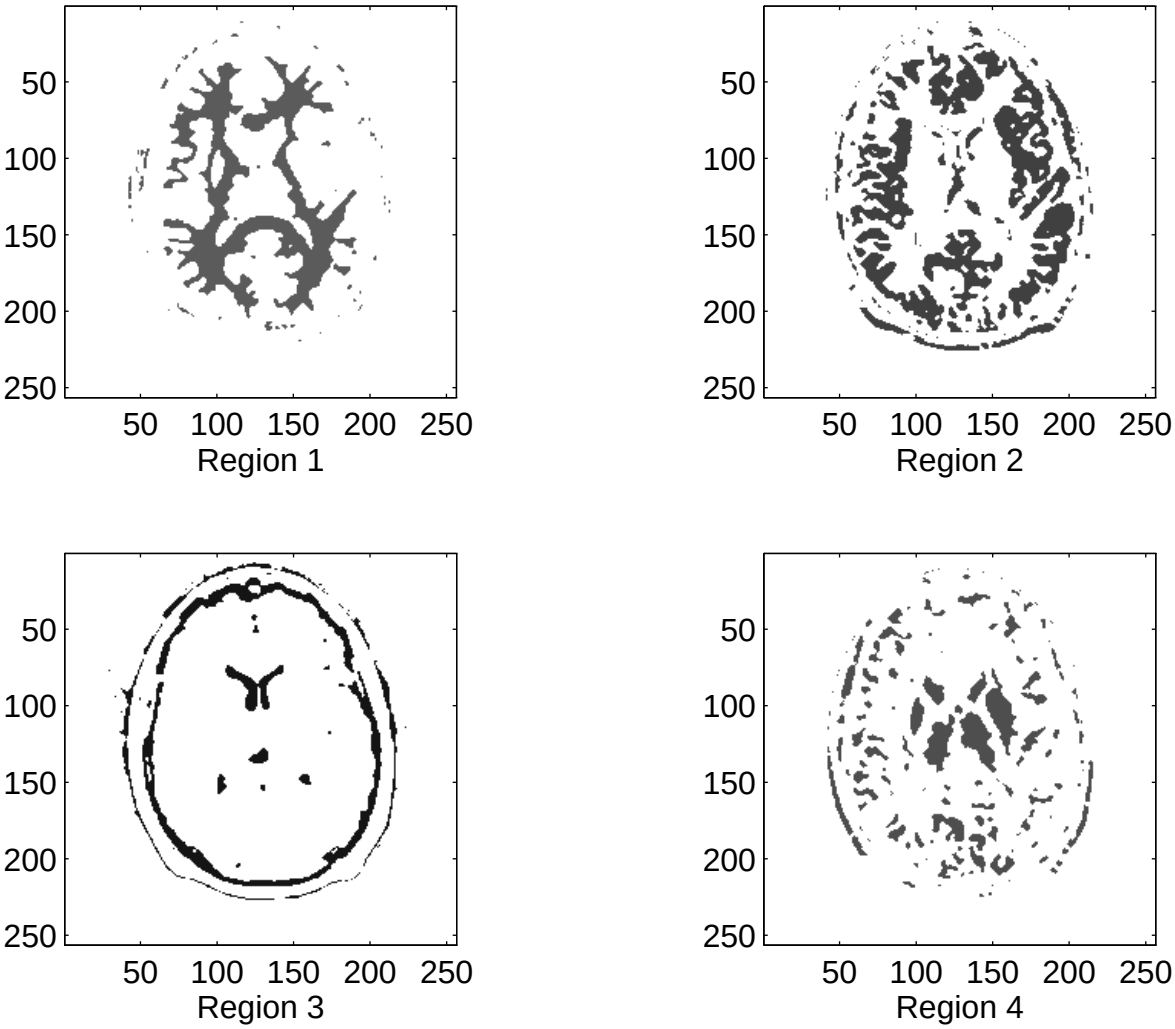

Figure 7.35: Major Regions in MRI #3 (MICM Result).

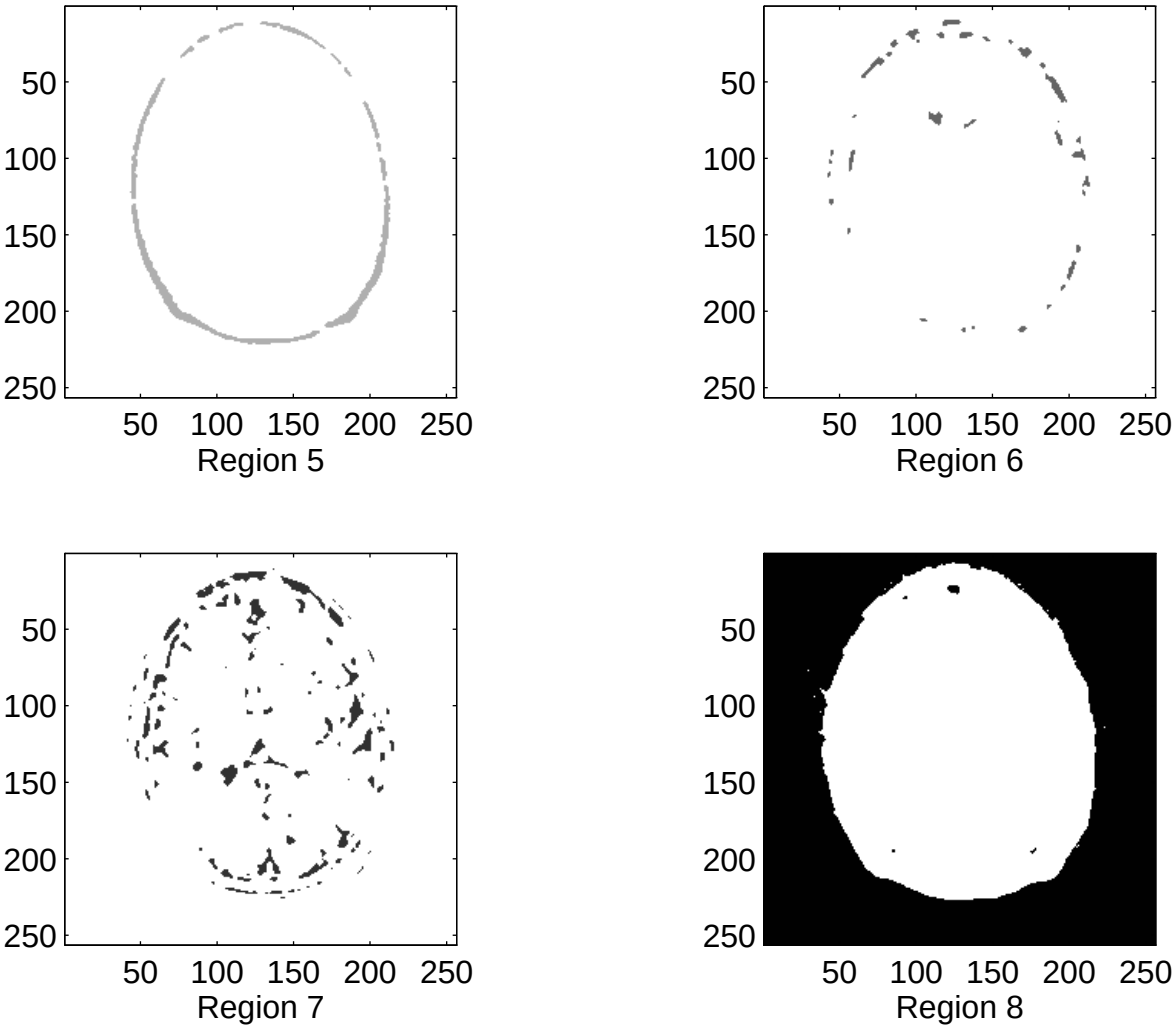

Figure 7.36: Minor Regions in MRI #3 (MICM Result).

partial volume effect, four possible mixture components will be CSF/GM (CG), GM/WM (GW), CSF/WM (CW), and CSF/GM/WM (CGW) [32]. The other uninteresting components we need to include may be background, skull, diploic space, temporalis muscle, internal occipital protuberance, and subcutaneous fat. Since each image is taken at a different position, these tissue components may only be partially captured by one image. Figure 7.37 (a) and (b) show the final segmented images of MRI #3 for further visual comparison. It can be seen that the results are consistent and close.

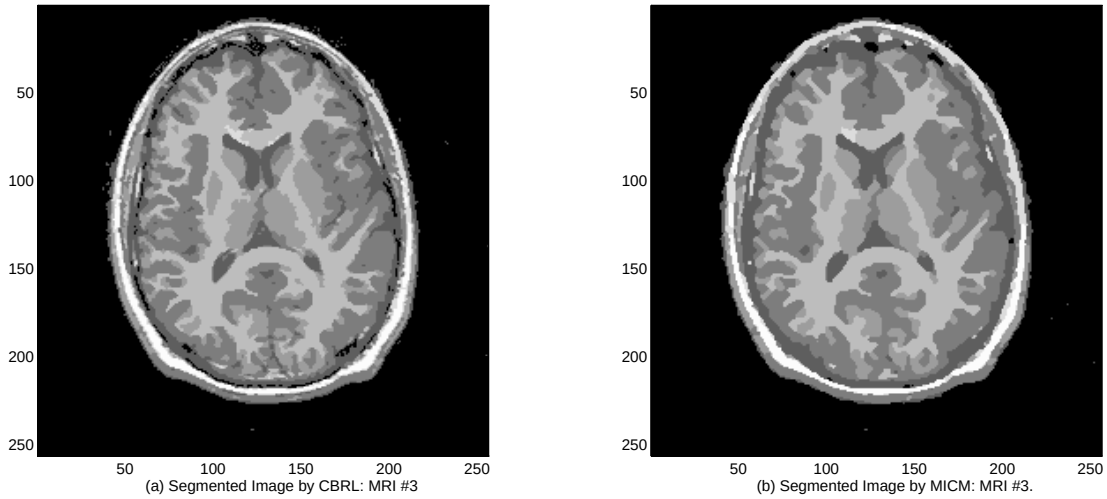

Figure 7.37: Final Segmented MRI #3: (a) Result of CBRL, (b) Result of MICM.

Table 7.17: Comparison of BCM and EM Algorithms with Simulated Data.

Item	BCM Result	EM Result	Improvement
GRE Value	0.0081258	0.0080381	0.0000877

7.4 Summary and Discussions

1) One problem in current unsupervised image analysis techniques is that there is not a systematic method for initializing model parameters. Our study shows that ALMHQ algorithm may be a satisfactory approach. CM algorithm which has been used for both parameter estimation and image segmentation [59, 60, 62] might also be a good alternative. Our study also shows that when image SNR is high, the segmentation by MLC is acceptable; on the other hand, when image SNR is low, MLC provides a good initialization for final image segmentation.
2) Another problem in current unsupervised image analysis techniques is that the EM solution may correspond to a local minimum. Our study shows that the likelihood of being trapped in a local minimum can be reduced by using BCM algorithm which also provides fast computation. Reference [62] shows that although CM algorithm needs much less iterations the than EM algorithm, its performance highly depends on the choice of a well-structured data; however, if EM algorithm is directly used, the computational demanding is high. BCM algorithm provides a solution for this conflicting demand. Also, BCM algorithm shows that it is more robust and insensitive to the initial input. Table 7.17 summarizes our numerical comparison of the performance of BCM and EM Algorithms for the simulated image. Clearly, EM algorithm leads to the most accurate parameter estimation. But the pay-off is the high computational demanding.
3) Based on our numerical simulations on both parameter estimation and segmentation, it is seen that the final results are not affected by the independence approximation of pixel images. This fact strongly supports the corresponding theoretic validation in Chapter 4. We also conjecture that some imperfect results from these algorithms are mainly due to the violation of the statistical properties only in some small regions of the image.

4) For the finalization of image segmentation, contextual image segmentation algorithms (CBRL and MICM) are performed based on the unsupervised parameter estimation and non-iterative initial segmentation. The connections and differences between contextual image segmentation and unsupervised parameter estimation as well as non-iterative initial segmentation need to be reiterated. Unsupervised parameter estimation is based on the observed pixel images and addresses the estimation of unknown constant parameters by using ML principle. When SFNM model is used, parameters $(\pi_k, \mu_k, \sigma_k^2)$ can be estimated by ML-EM procedure before final segmentation [18, 61]. Note that parameter estimation only provides partial and global information about the regions, since the correspondences between pixels and regions are missing in which the probabilistic memberships are estimated instead of realizing true pixel labels. As discussed before, true pixel labels are the mathematical indices which link pixels to regions, thus, pixel labels can represent context or region process in both local and global scale. Non-iterative initial segmentation is solely based on the observed pixel images and addresses the detection of pixel labels (independent) by using ML (MLC) or Bayesian (GBC) principle. Since the local correlations among pixel labels are ignored, i.e., no local context or region regularities are imposed and no iteration is needed, when realizing true pixel labels by applying Bayesian detection to the probabilistic memberships, the overlaps of regions in intensity domain result in noisy segmentations in spatial domain [18]. Therefore, contextual image segmentation is based on both observed pixel images and local context or region regularities and addresses the detection of correlated pixel label random variables by using MAP principle. Although region process is combined, pixel labels are correlated, thus, during relaxation labeling, local context changes and iterations are needed. As a conclusion, contextual image segmentation uses region process to impose the local context regularities into image segmentation combined with pixel image information modeled by FNM. Since the underlying true region or context is unique, the appropriate region process (by suitable consistency constraint formulation and Markov parameter identification) in context image segmentation is expected to provide a better segmentation result (close to unique true context) and reduce misclassification which is also consistent with unsupervised parameter estimation when the ergodic theorems hold.

5) As a final issue, the advantages of inhomogeneous MRF configuration and the corresponding MICM algorithm are worthy of being reiterated. Theoretically, Gibbs sampling and MRF model-based image segmentation form a pair of forward and backward problems. By assuming a homogeneous MRF, some numerical experiments have shown the inconsistency of these two procedures [45]. For example, a realization of an MRF with $\theta = 1.5$ tends to have very unbalanced numbers of labels because of phase transition, although the segmentation results under the same parameter setting are quite satisfactory. This idea is strongly shared by many researchers [41, 45]. It is expected that with our inhomogeneous model, MICM algorithm provides a new systematic way for this process.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

An investigation into MR imaging statistics and model-based MR image analysis is presented in this dissertation. The studies cover both the physics and signal processing aspects of MR imaging in a unified way.

For direct Fourier transform MR image reconstruction approach, we derived a more complete closed-form of point spread function (Eq. (2.53)), and shown that the resultant MR image is a discrete and interpolated representation of the linear magnetic resonance coefficients.

Two major random sources in MR imaging: object variability and thermal noise, have been studied. We show that the pixel image generated by direct Fourier transform MR image reconstruction can be characterized by an asymptotic Gaussian distribution; pixel images forms an asymptotic independent and piece-wise stationary random field; and each image region satisfies mean and variance ergodic theorems (Properties 3.3.1-3.3.4). By discussing the stochastic regularities exhibited by context images, we demonstrate that the intrinsic dependence (or correlation) among context images can be characterized by the Markovian property. We propose a simple randomization rule to mathematically realize the statistical treatment of context images, and therefore provide a good mathematical interpretation for the stochastic regularization of context images.

A new framework on the stochastic modeling of MR image is presented. For pixel images, we introduce FNM models and show that SFNM model can be obtained mathematically from CFNM model using randomization, and therefore SFNM model has the same sample space as CFNM but different event space and probability measures (Observation 4.2.3.1). We justify the correct use of FNM models for the specified purpose in MR image analysis in terms of new definitions, observations (Observations 4.3.2.1), and theorems (Theorems 4.2.3.1-4.2.3.2). Specifically, we show that parameter estimation will not be seriously affected by the independence approximation in SFNM modeling. By incorporating the statistical properties of context images, we extend FNM modeling to its multiresolution and localized formulations and shown that new formulations resulted in efficient algorithms for different purposes. A novel comparison of their structures and applications is also presented. Finally, we propose an inhomogeneous hidden MRF model for context images and incorporate pixel images by using Bayesian formulation. We show that this configuration naturally integrates region and boundary information into a framework.

Parameter estimation and segmentation can be solved in an unsupervised mode. We first derive ALMHQ algorithm to initialize the parameter estimates, and then develop BCM algorithm to refine the parameter initialization by using a multiresolution procedure. Finally, we introduce EM algorithm to finalize ML parameter estimates. We show that the ML estimates generated by EM algorithms are consistent with mean and variance ergodic theorems. For determining the number of image regions, a model fitting procedure is developed by using the information theoretic criteria, AIC, MDL, and a new derived approach (MCBV) based on the minimum bias and variance principle. We demonstrated that this model fitting approach

performs consistently well and provides a systematic basis for unsupervised image analysis.

By comparing CM and EM algorithm, we show that both CM and EM can produce asymptotically unbiased ML estimates, for some data structure under proper conditions. In order to improve the initial image segmentation by MLC, we developed two algorithms for finalizing MR image segmentation: CBRL and MICM (BDMC algorithm for quantifying Markov parameter values is integrated into MICM algorithm). Local consistency constraint defined by compatibility function reflects Markovian property and is utilized in the contextual image segmentation by CBRL algorithm. Boundary information is used by decision maker in MICM algorithm in terms of inhomogeneous MRF configuration which results in a better contextual image segmentation.

The proposed algorithms are numerically and empirically verified with simulated image and real MR images. Based on our experiments, ALMHQ algorithm has been proven to provide a very satisfied parameter initialization, and CM algorithm is a good alternative. BCM algorithm reduces the likelihood of solutions being trapped in local minima, and provides fast computation. When image SNR is high, the segmentation with simple MLC is acceptable, when image SNR is low, MLC provides a good initialization. The initial segmentation can be improved dramatically by using local context in terms of non-parametric Bayesian binding in CBRL algorithm. Both CBRL and MICM algorithms receive very fast convergence and outperform than the conventional ICM approach with a lower misclassification error.

8.2 Discussions

With regard to our image statistics, image modeling, and image analysis technique, there are still several interesting and important problems which need to be addressed. Our discussions mainly focus on the limitations of new framework, the relationship between the parameter estimation and image segmentation, and the potential application of this approach to other modalities.

We proposed our framework and developed new techniques for MR image analysis. Under certain conditions, we justified the statistical properties of MR images so as to use the model-based approach. The main assumptions we have made are that the imaging system is a linear system, object variability and thermal noise are two major random sources, and the possible deterministic artifacts can be ignored. In fact, because of the inhomogeneity of the magnetic field, the non-linearity of the RF receiver, and the artifacts from the reconstruction algorithm, the corresponding distortions may have significant impact on the image statistics. Therefore, possible unsatisfactory results may occur when applying our technique to the situations which do not satisfy the assumptions. In addition, in image segmentation, we use non-contextual classifier (MLC/GBC) for initializing the segmentation and use contextual classifiers (CBRL, MICM) for finalizing the segmentation. Since the intrinsic trade-off between reducing the noise effect and smoothing the sharp edges, if the parameter setting or structure selecting of the prior constraint is not appropriate, the shape and size of the isolated or small disease patterns may be eliminated or misclassified. However, for the major applications of this research, such as radiation treatment planing and partial volume correction using computed images, these limitations are not significant.

As we discussed before, in unsupervised image analysis, parameter estimation and image segmentation can be performed either simultaneously or sequentially. Since in general an accurate image segmentation is time-consuming even when the true model parameter values are given, in this work, in order to achieve an efficient performance, we first estimate SFNM model parameters and then segment image into the regions based on the framework of Markov formulation. Since the stochastic model and optimization criterion in parameter estimation are different from those in image segmentation, possible inconsistency may occur in some level when no special care is conducted. This problem can be explained based on the relationship between parameter estimation and image segmentation. In our case, the task in parameter estimation is to obtain ML estimates of $(\pi_k, \mu_k, \sigma_k^2)$. Since the SFNM model is formulated for *unclassified* data with an i.i.d. structure, asymptotically unbiased ML estimates can be obtained *before* image segmentation. By using EM algorithm, the parameter estimates are determined based on the estimated probabilistic memberships of all pixels *without* realizing the true pixel labels for all pixels. On the other hand, the task in image segmentation

is to obtain Bayesian or MAP solution of l_i for all pixels based on the observation x_i (a realization based on (μ_k, σ_k^2)), therefore, the i.i.d. structure is no longer meaningful. In fact, even the true model parameter values are given, image segmentation techniques, when using different models or criteria, do not give the perfect true scenes. According to ergodic theorems, the misclassification error from image segmentation with respect to the true context will result in the different parameter estimates. In particular, when the prior probabilities of regions are not appropriately determined, the misclassification error could be large. There are two possible ways to solve this inconsistency problem. One is to perform parameter estimation and image segmentation simultaneously based on one model. Another is to develop a modified FNM model which can reflect the true correlation among pixels and can be well estimated.

Finally, since this technique is model-based and mainly uses FNM and MRF models, it is suitable for the tone-dominant images when they satisfy the statistical requirements for these models. In general, it can be applied to the imaging modalities such as x-ray CT, x-ray, and CR, with little modification. For ultrasound scan or PET/SPECT, possible major modifications are required.

8.3 Future Work

Image analysis, particularly, the unsupervised approach, provides valuable, accurate, and quantitative information about the true scene represented by a visual picture. In medicine, it can improve the human vision in the diagnosis and is an important part of computer aided radiology. This dissertation addresses some topics in MR image analysis. Some future work is listed below.

1) With minor modifications it is possible to generalize FNM to include partial volume effect. This is particularly interesting, since it allows an exhaustive analysis of histogram without heuristic assumptions in the conventional partial volume modeling [32], and the possible use of self-organizing map, Gabor expansion, and wavelet techniques.

2) Conceptually, dependence among pixel images and context images are different, however, they are inherently connected. In order to assure the validity of the conditional independent assumption required for hidden MRF modeling, a possible way is to mathematically fuse these two kinds of correlation together into the optimization of inhomogeneous MRF parameter assignment. More theoretic work and numerical experiments are necessary.

3) MCBV approach described in Chapter 5 demonstrates promise for image region number detection. However, it has not been rigorously justified. While the problem of evaluating its performances with compared to AIC and MDL was preliminary treated in this dissertation, further work for justification or proof is required.

4) All of the unsupervised MR image analysis approaches suggested in this work were within the context of sequential parameter estimation and segmentation. Based on the multiple resolution segmentation algorithm such as BCM, simultaneously parameter estimation and image segmentation may achieve fast and accurate image analysis.

Bibliography

[1] Y. Wang and T. Lei, "Statistical Analysis of MR Imaging and Its Applications in Image Modeling," in *Proc. First IEEE Intl. Conf. Image Processing*, Austin, Texas, November 13-16, 1994.

[2] Y. Wang and T. Lei, "A New Stochastic Model-Based Image Segmentation Technique for MR Images," in *Proc. First IEEE Intl. Conf. Image Processing*, Austin, Texas, November 13-16, 1994.

[3] Y. Wang, T. Lei, and T. Adalı, "Hidden MRF Model-Based Algorithms for MR Image Analysis," *Proc. IEEE Medical Imaging Conference*, Norfolk, Virginia, November 3-5, 1994.

[4] Y. Wang, T. Lei, and J. M. Morris, "Detection of the Number of Image Regions by Minimum Bias/Variance Criterion," in *Proc. SPIE Conf. Visual Commu. Image Processing*, Chicago, Illinois, September 26-28, 1994.

[5] Y. Wang and T. Lei, "A New Look at Finite Mixture Models in Medical Image Analysis," *Proc. IEEE Intl. Symp. Speech, Image Processing & Neural Networks,* April 14-16, Hong Kong, 1994.

[6] Y. Wang and J. Zhang, "Engineering Aspects in the Design of Low Field MRI Systems," TR-JTU-E31, Center for Medical Imaging and Instrumentation, Shanghai 1991.

[7] Y. Wang, "A Note on Pulse Sequence Design for Low Field MRI Systems," TR-Weida-MRI14, Weida MRI, Inc., Guangdong, 1990.

[8] T. Adalı, and Y. Wang, "Maximum likelihood neural networks for medical image quantification," in *Proc. First IEEE Int. Conf. on Image Processing.*, Austin, Texas, November 13-16, 1994.

[9] R. P. Crease, "Biomedicine in the age of imaging," *Science*, Vol. 261 July 1993.

[10] W. Hinshaw and A. H. Lent, "An introduction to NMR imaging: from the Bloch equation to the imaging equation," *Proc. IEEE*, Vol. 71, 338-350, 1983.

[11] G. Sebastiani and P. Barone, "Mathematical principles of basic magnetic resonance imaging in medicine," *Signal Processing* 25, 227-250, 1991.

[12] A. Macovski, "Physical Problems of Computerized Tomography," *Imaging Technology*, IEEE Press, New York 1986

[13] R M. Lewitt, "Reconstruction Algorithms: Transform Methods," *Imaging Technology*, IEEE Press, New York 1986

[14] Z. H. Cho, H. S. Kim, H. B. Song, and J. Cumming, "Fourier Transform Nuclear Magnetic Resonance Tomographic Imaging," *Imaging Technology*, IEEE Press, New York 1986

[15] A. V. Oppenheim and R. W. Schafer, *discrete-Time Signal Processing*, Prentice-Hall, Inc. 1989.

[16] C. E. Swenberg and J. J. Conklin, *Imaging Techniques in Biology and Medicine*, Academic Press, Inc. 1988.

[17] T. Lei and W. Sewchand, "Statistical approach to X-ray CT imaging and its applications in image analysis—PartI," *IEEE Trans. Med. Imag.* Vol. 11, No. 1, 53-61, March 1992.

[18] Z. Liang, J. R. MacFall, and D. P. Harrington, "Parameter Estimation and Tissue Segmentation from Multispectral MR Images," *IEEE Trans. Med. Imag.* Vol. 13, No. 3, September 1994.

[19] J. Dengler, S. Behrens, and J. F. Desaga, "Segmentation of Microcalcifications in Mammograms," *IEEE Trans. Med. Imag.* Vol. 12, No. 4, December 1993.

[20] H. H. Barrett, "Objective assessment of image quality: effects of quantum noise and object variability" *J. Opt. Soc. Am. A* Vol. 7, No. 7, pp. 1266-1274, July 1990.

[21] R. Gray and L. Davisson, *Random Processes–A Mathematical Approach for Engineers*, Englewood Cliffs, NJ: Prentice-Hall, Inc. 1986.

[22] F. Godtliebsen and E. Spjotvoll, "Comparison of statistical methods in MR imaging," *Intl. J. Imag. Sys. Tech.* Vol. 3, pp. 33-39, 1991.

[23] R. F. Wagner, M. F. Insana, and S. W. Smith, "Fundamental correlation lengths of coherent speckle in medical ultrasonic images," *IEEE Trans. Ultrason. Ferroelect. Freq. Contr.*, Vol. UFFC-33, No. 3, pp257-264, May 1986.

[24] Z. P. Liang, "Statistical models of a priori information for image processing: neighboring correlation constraints," *J. Opt. Soc. Am. A*, Vol. 5, No. 12, pp. 2026-2031, December 1990.

[25] R. F. Wagner and D. G. Brown, "Unified SNR analysis of medical imaging systems," *Phys. Med. Biol.*, Vol. 30, No. 6, 489-518, 1985.

[26] C. F. Tejero, "A method for evaluating two-spin correlations of a one-dimensional Ising model," *Am. J. Phys.*, 56(2), 1988.

[27] I. Shenberg and A. Macovski, "Resolution and Noise Considerations in MRI Systems with Time-Varying Gradients," *IEEE Trans. Med. Imaging*, Vol. 4, No. 3, September 1985.

[28] G. Muller and R. E. Shrock, "Dynamical correlation functions for quantum spin systems," *J. Appl. Phys.*, 55(6) 1984.

[29] Y. Vardi, L. A. Shepp, and L. Kaufman, "A statistical model for positron emission tomography," *J. Ame. Stat. Assoc.*, Vol. 80, No. 389, pp. 8-37, March 1985.

[30] W. A. Edelstein, G. H. Glover, and R. W. Redington, "The intrinsic signal-to-noise ratio in NMR imaging," *Magn. Reson. Med.* 3, 604-618, 1986.

[31] D. M. Titterington, A. F. M. Smith, and U. E. Markov, *Statistical analysis of finite mixture distributions*, New York: John Wiley & Sons., 1985.

[32] P. Santago and H. D. Gage, "Quantification of MR brain images by mixture density and partial volume modeling," *IEEE Trans. Med. Imag.*, Vol. 12, No. 3, pp. 566-574, 1993.

[33] R. Momenan, R. F. Wagner, B. Garra, M. H. Loew, and M. F. Insana, "Image Staining and Differential Diagnosis of Ultrasound Scans Based on the Mahalanobis Distance," *IEEE Trans. Med. Imag.*, Vol. 13, No. 1, March 1994.

[34] M. E. Brummer, R. M. Merserrau, R. L. Eisner, and R. J. Lewine, "Automatic Detection of Brain Contours in MRI Data Sets," *IEEE Trans. Med. Imaging*, Vol. 12, No. 2, June 1993.

[35] P. J. Elliott, J. M. Knapman, and W. Schlegel, "Interactive image segmentation for radiation treatment planning," IBM Systems Journal, Vol. 31, No. 4, pp. 620-634, 1992.

[36] L. O. Hall, A. M. Bensaid, L.P. Clarke, R.P. Velthuizen, M.S. Silbiger, and J.C. Bezdek, "A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain," *IEEE Trans. Neural Networks*, Vol. 3, No. 5, September 1992.

[37] M. Joliot and B. M. Mazoyer, "Three-Dimensional Segmentation and Interpolation of Magnetic Resonance Brain Images," *IEEE Trans. Med. Imaging*, Vol. 12, No. 2, June 1993.

[38] S. C. B. Lo, E. L. Chen, S. K. Mun, and J. Chen, "A Method for Splitting Digital Value in Radiological Image Compression," *Med. Phys.*, Vol. 18, No. 5, pp. 939-946, September 1991.

[39] D. P. Casasent and J. S. Smokelin, "Neural Net Design of Macro Gabor Wavelet Filters for Distortion-Invariant Object Detection in Clutter," *Opt. Eng.*, Vol. 33, No. 7, pp. 2264-2271, July 1994.

[40] A. K. Jain, *Fundamentals of Digital Image Processing*, Pretice-Hall, Inc., 1989.

[41] J. Besag, "Spatial Interaction and Statistical Analysis of Lattice Systems," J. Roy. Stats. Sco. 1974.

[42] D. Geman, *Random Fields and Inverse Problems in Imaging*, Springer-Verlag, 1990.

[43] R. L. Kashyap and R. Chellappa, "Estimation and Choice of Neighbors in Spatial-Interaction Models of Images," *IEEE Trans. Information Theory*, Vol. 29, No. 1, January 1983.

[44] F. Young and K. S. Fu, *The Handbook of Pattern Recognition and Image Processing*, Academic Press, 1986.

[45] R. C. Dubes and A. K. Jain, "Random field models in image analysis," *J. Appl. Statist.,* Vol. 16, No. 2, 1989.

[46] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Review*, Vol. 26, No. 2 April 1984.

[47] D. M. Titterington, "Comments on 'Application of the conditional population mixture model to image segmentation'," *IEEE Trans. Pattern Recognition and Machine Intelligence,* Vol. 6, No. 5, September 1984.

[48] L. I. Perlovsky and M. M. McManus, "Maximum likelihood neural networks for sensor fusion and adaptive classification," *Neural Networks*, Vol. 4, 1991.

[49] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inform. Theory*, Vol. IT-6, pp.7-12, Mar. 1960.

[50] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice-Hall, Inc. 1991.

[51] H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, Vol. 19, No. 6, December 1974.

[52] G. Schwarz, "Estimating The Dimension of A Model," *The Annals of Statistics*, Vol. 6, No. 2, 1978.

[53] J. Rissanen, "A Universal Prior for Integers and Estimation by Minimum Description Length," *The Annals of Statistics,* Vol. 11, No. 2, 1983.

[54] M. Wax and T. Kailath, "Detection of Signals by Information Theoretic Criteria," *IEEE Trans. Acoust. Speech, Signal Processing,* Vol. 33, No. 2, April 1985.

[55] J. Rissanen, "Modeling by shortest data description," *Automatica*, Vol. 14, 1978.

[56] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural Computation*, 4, 1992.

[57] J. Rissanen, "Minmax entropy estimation for vector processes," *System Identification: Advances and Case Studies*, pp. 97-117, Academic Press, 1976

[58] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, World Scientific, 1989.

[59] T. Kurita, N. Otsu, and N. Abdelmalek, "Maximum likelihood thresholding based on population mixture models, " *Pattern Recognition*, Vol. 25, No. 10, 1992.

[60] S. L. Sclove, "Application of the Conditional Population-Mixture Model to Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 5, No. 4, pp428-433, 1983.

[61] J. Zhang and J.M. Modestino, "A Model-Fitting Approach to Cluster Validation with Application to Stochastic Model-Based Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 10, October 1990.

[62] T. Lei and W. Sewchand, "A New Stochastic Model-Based Image Segmentation Technique for X-ray CT Image," *IEEE Transactions on Medical Imaging,* Vol. 11, N0. 1, 53-61, March 1992.

[63] K. V. Mardia and T. J. Hainsworth, "A spatial thresholding method for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 6, November 1988.

[64] J. T. Kent and K. V. Mardia, "Spatial classification using fuzzy membership models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 5, September 1988.

[65] H.V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlay, 1988.

[66] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Statist. Soc. B*, Vol. 48, No. 3, 1986.

[67] C. Bouman and B. Liu, "Multiple Resolution Segmentation of Texture Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 2, February 1991.

[68] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, No. 6, November 1984.

[69] B. S. Manjunath and R. Chellappa, "Unsupervised Texture Segmentation Using Markov Random Field Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 5, May 1991.

[70] R. A. Hummel and S. W. Zucker, "On the foundations of relaxation labeling processes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 5, No. 3, May 1983.

[71] K. Sauer and C. Jones, "Bayesian block-wise segmentation of interframe differences in video sequences," *Graphical Models and Image Processing*, Vol. 55, No. 1, March 1993.

[72] H. H. Nguyen and P. Cohen, "Gibbs random fields, fuzzy clustering, and the unsupervised segmentation of textured images," *Graphical Models and Image Processing*, Vol. 55, No. 1, January 1993.

[73] D. Geman, S. Geman, C. Graffigen, and P. Dong, "Boundary Detection by Constrained Optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 7, July 1990.

[74] S. Lakshmanan and H. Derin, "Simultanneous parameter estimation and segmentation of Gibbs random fields using simulated annealing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 8, August 1989.

[75] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc. 1991.

[76] M. Fuderer, "The information content of MR images," *IEEE Trans. Med. Imag.* Vol. 7, No. 4, 368-380, December 1988.

[77] R. C. Dubes, A. K. Jain, S. G. Nadabar, and C. C. Chen, "MRF Model-Based Algorithms for Image Segmentation," *Proc. Pattern Recognition*, Vol. 1, pp. 808-814, 1990.

[78] L. I. Perlovsky, "Cramer-Rao Bounds for The Estimation of Normal Mixtures," *Pattern Recognition Letters*, Vol. 10, 1989.

[79] J. L. Marroquin and F. Girosi, "Some extensions of the K-means algorithm for image segmentation and pattern classification," Technical Report, MIT Artificial Intelligence Laboratory, January 1993.