To appear: Signal Processing for Magnetic Resonance Imaging and Spectroscopy, (Chapter 14), Marcel Dekker, 2000.

14
Stochastic Model Based Image Analysis

Yue Wang[1] and Tülay Adalı[2]

[1]Department of Electrical Engineering and Computer Science
Catholic University of America, Washington, DC 20064

[2]Department of Computer Science and Electrical Engineering
University of Maryland Baltimore County, Baltimore, MD 21250
adali@engr.umbc.edu

1 Introduction

Quantitative analysis of magnetic resonance (MR) images deals with the problem of estimating tissue quantities and segmenting the anatomy into contiguous regions of interest. The problem has recently received much attention largely due to the improved fidelity and resolution of MR imaging systems, and the effective clinical utility of image analysis and understanding in diagnosis, monitoring, and intervention (e.g. see references [1]– [4]). For example, pathological studies show that many neurological diseases are accompanied by subtle abnormal changes in brain tissue quantities and volumes as shown in reference [2]. Because of the virtual impossibility for clinicians to quantitatively analyze these pathological changes associated with a specific disease directly from MR images, considerable effort is required to develop accurate image analysis algorithms for identifying and quantifying these changes in-vivo as discussed in references [5]– [11].

The major tasks of MR image analysis involve tissue quantification and image segmentation. In the stochastic model-based approach, it is typically assumed that each pixel can be decomposed into a *pixel image* and a *context image* and each is considered separately. A pixel image is defined as the observed gray level associated with the pixel, whereas the context image is defined as the membership of the pixel associated with different regions.

In our treatment of the stochastic MR analysis, we first address the problem of MR image statistics and in the next section, provide the complete statistical description of MR imaging along with justification of the common assumptions made for MR image modeling. This section provides the

basis for the following section, section 3, that introduces the statistical models for pixel and context images. The finite mixture models, and in particular, the standard finite normal mixtures (SFNM), have been widely used for pixel images, and efficient algorithms are available for calculating the parameters of the model. On the other hand, by incorporating statistical properties of context images, Markov random fields (MRF) have proven to be the most popular way to impose local consistency constraints on context images in terms of a stochastic regularization scheme given in reference [12]. We introduce both of these models and show that by using an *adaptive* SFNM model where information theoretic criteria are used to determine the number of tissue types, we can incorporate the partial volume effect, tissue abnormality, and inhomogeneity distortion, into a single stochastic model [3, 10]. Also, an inhomogeneous MRF is introduced that uses the entropy rate distribution to determine probabilistic boundary sites in the image, and the spatial discontinuity levels to assign the model parameter values. This permits an adaptive construction of the patient specific site model for improving both tissue quantification and image segmentation as given in references [13, 14].

Section 4 describes the estimation of the model parameters, i.e., the quantification and segmentation stages. The final section provides a set of two and three dimensional image analysis results. We also address the problem of algorithm performance evaluation and define post-global relative entropy for the assessment of the final image segmentation performance, shown in reference [15]. Based on the observation that the parameter values of a particular tissue type in the quantified and segmented images should be very close, this criterion suggests an indirect but objective approach for the difficult problem of performance evaluation in image segmentation in reference [16].

2 MR Image Statistics

2.1 Introduction

As discussed in Section 1, although some pioneer work has been reported on the statistical analysis of x-ray CT [17], positron emission tomography [18], and ultrasound scan [19], there has been little work carried out on MR imaging statistics and model-based MR image analysis. For example, reference [20] discusses noise sources and their effects in MR imaging, reference [21] presents a unified SNR analysis of medical imaging systems, reference [22] applies information theory to the quality evaluation of MR imaging, and reference [24] compares several currently used statistical algorithms in MR imaging. However, none of these researchers provide a complete statistical

analysis that can incorporate MR imaging statistics into MR image modeling by justifying the many heuristic assumptions typically made in the literature (e.g., see reference [22]). In this section, 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 is 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 important and have to be considered together. Many investigations have been conducted on image statistics of pixel images in MR or other modalities (e.g. references [10, 22, 25, 29]). Reference [20] has conducted intensive research on the objective assessment of imaging statistics of x-ray and gamma rays, by considering the effects of both quantum noise and object variability. A pioneer work is reported in reference [19], on the speckle statistics in medical ultrasound images in which the first order and the second order statistics received intensive treatment. References [21, 26] 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 [24] conducts a comparative work for evaluating several statistical approaches in MR imaging, in which intensive real MR data analysis is included; reference [22] contributes to the measurement of bivariant information content of MR images in which six assumptions are made and partially justified.

Furthermore, the problem for researchers relates to the statistics of context image is another challenging topic. Using a randomization rule and stochastic regularization, many investigations have been conducted (see references [13, 25, 27, 28]) showing important properties. However, they suffer from some limitations, such as constraints on context representation, which are generally imposed mathematically without objective justification by the context image statistics [13, 29]. Since the true context is unobservable in general, the assumed models always contain some empirical parameters that are given heuristically [3, 10, 13]; homogeneous assumption is applied to the context image even when it is obviously not true [10, 30].

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

1. MR image statistics as applicable to typical MR data samples using theoretical and experimental studies;

2. MR image statistics in the context image;

3. The relationship between the MR image statistics and the MR image models.

2.2 Statistical Properties of Pixel Images

2.2.1 Gaussianity

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

Property 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 [24] has shown that, based on a number of experiments in which a single set of measurement of a slice is compared to 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 produces and excellent fit. Furthermore, the modulus image will have a Rayleigh distribution. Rayleigh density approaches a truncated Gaussian when the SNR is relatively high [27]. Reference [24] also shows that noise at each pixel is approximately following a Gaussian distribution in the modulus images.

2.2.2 Dependence

According to second-order moment input/output relations of linear systems [31], the covariance function of the convolved linear magnetic resonance coefficients can be expressed as

$$K_f((k,l),(k',l')) = K_\beta((k,l),(k',l')) \odot h^*((k,l),(k',l')) \odot h((k,l),(k',l')) \tag{1}$$

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

When the distance between two pixel images is denoted by $\Delta d = ||(k,l),(k',l')||$, we have

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

$$= \sigma(k,l)\sigma(k',l')\exp(-\frac{\Delta d}{\Delta r}) \odot h^*(\Delta d) \odot h(-\Delta d) \tag{2}$$

where Δr is defined as the effective correlation distance that has an order of 10^{-8} m. Thus, it is easy to show that when Δd approaches infinity, the covariance of $f(\cdot,\cdot)$ goes to zero since the first convolution function in Equation 2 is narrow ranged. This property is called asymptotically independent or weakly dependent. In our case, it is

$$\lim_{||(k,l),(k',l')||\to\infty} K_f((k,l),(k',l')) = 0. \tag{3}$$

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

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

$$= K_f((k,l),(k',l')) + \frac{\sigma_{s_n}^2}{MN}\delta(k'-k,l'-l) \tag{4}$$

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

Property 2: *Any two pixel image random variable (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 Equation 4 derives from the fact that the correlation between any two pixels in an MR image is only determined by the correlation of the object variability part in which both microscopic correlation and system point spread function make contributions. This property is strongly confirmed by real MRI data analysis. References [22, 24] have reported that, based on their experiments, pixel images seem to be uncorrelated since the plot of the correlation function of MR an image is shown to be *rapidly* decreasing with the inter-pixel distance increasing.

2.2.3 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. A homogeneous object in

MR image has a unique linear magnetic resonance coefficient $\beta(x, y)$, that is, the unique mean and variance is

$$E[\beta(x,y)] = \mu, \quad E[(\beta(x,y) - \mu)^2] = \sigma^2. \tag{5}$$

We can note that by reference [23]

$$E[\hat{f}(k,l)] = C_1, \quad E[(\hat{f}(k,l) - C_1)^2] = C_2 \tag{6}$$

where C_1 and C_2 are the constant mean and variance of a pixel image $\hat{f}(x, y)$. Note that for different homogeneous object in an image, $C_{(\cdot)}$ 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, we 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 the 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 Equation 3.21 as

$$K_f(\Delta d) = \sigma^2 \exp(-\Delta d) \odot h^*(\Delta d) \odot h(-\Delta d) \tag{7}$$

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

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

$$= \begin{cases} \sigma^2 \exp(-\Delta d) \odot h^*(\Delta d) \odot h(-\Delta d) + C_1^2 & (k \neq k', l \neq l') \\ C_2 + C_1^2 & (k = k', l = l') \end{cases} \tag{8}$$

Since correlation only depends on the spatial index difference, according to reference [31], $\hat{f}(i, j)$ is a stationary field in the wide-sense, and also in the strict-sense. By assuming that an MR image contains several distinct regions, we state stationarity as follows:

Property 3: *Given an MR image, each region is stationary. The whole image is piecewise stationary.*

2.2.4 Ergodicity

For a given region in an MR image, we can rewrite Equation 4 as

$$K_{\hat{f}}((k,l),(k',l')) = \begin{cases} \sigma^2 \exp(-\Delta d) \odot h^*(\Delta d) \odot h(-\Delta d) & (k \neq k', l \neq l') \\ C_2 & (k = k', l = l') \end{cases} \tag{9}$$

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

$$\lim_{M,N\to\infty} \frac{1}{MN} \sum_{k=0}^{M} \sum_{l=0}^{N} \hat{f}(k,l) = C_1 \tag{10}$$

with probability one [31]. Note that the variances of both thermal noise and object variability are assumed to be finite.

Then, we use the Birkhoff-Khinchin theorem [31] to show that $\hat{f}(k,l)$ has a variance ergodic theorem. Reference [31] 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 [31],

$$\lim_{M,N\to\infty} \frac{1}{MN} \sum_{k=0}^{M} \sum_{l=0}^{N} [\hat{f}(k,l) - C_1]^2 = C_2 \tag{11}$$

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

Property 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 a random field, is required.

2.3 Statistical Properties of Context Image

2.3.1 Stochastic Regularization and Markovian Property

As mentioned above, each pixel in an MR image 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 class label (membership). In the previous section, 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 an 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., a homogeneous object. In order to have a complete characterization of MR image statistics, we need to investigate the statistical properties of the 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 a statement about the context information, in both local and global scales, into model structures and parameter values, in a traceable format. As discussed in a number of references [13, 25], although true context is unobservable, a priori expectations, specific scene knowledge, and contextual clues can help 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 such as; tones, although these have locally slow variations representing homogeneous regions; boundaries which are usually smooth and persistent; and objects, such as tissue and organs, have preferred relations and orientations. These "regularities" are rarely deterministic, rather, they describe the correlations and likelihoods of possible outcomes in a real scene. This type of knowledge can be captured mathematically and exploited in a stochastic framework.

In MR images, regions are piecewise contiguous, thus, a 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 a Markovian property. The Markovian property implies that the probability distribution of a label, only depends on its neighboring pixel labels, given all other 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) $P(L_i = l_i | \mathbf{l_{S|i}}) = P(L_i = l_i | \mathbf{l}_{\partial i})$.

where l_i is the label of pixel i, ∂i denotes the neighborhood of pixel i, and $\mathbf{l_{S|i}}$ all pixel labels except pixel i. This natural constraint can be also justified based on the mechanism of biological development [32].

2.3.2 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.

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 [1] 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 its dependence on 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 the next section.

3 Stochastic Models for an MR Image

3.1 Introduction

The objective of stochastic modeling in image analysis is to capture the intrinsic character of images in a structure with few parameters so as to understand the statistical nature of the imaging

[1] Given a central pixel, a second order neighborhood system refers to its eight neighboring pixels [13].

process. Stochastic image models are useful to quantitatively specify natural constraints and general assumptions for the purpose of statistics about the physical world and the imaging process. They play a crucial role in unsupervised 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 [27]:

1. *Models for Pixel Images*: The models for pixel images are specifically designed to describe the type of randomness involved in the observable pixel random variables. Many stochastic models have been proposed to analyze the tone or texture pixel images, such as; the conditional FNM (CFNM) [33], the standard FNM (SFNM) [17, 29, 35], the generalized Gaussian mixture (GGM) [29], the Gaussian random field (GRF) [10], the autoregressive (AR) model [34], and the Markov/Gibbs random field (MRF/GRF) [37].
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 [35]. 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 [37].

In the previous section, we discussed the statistical properties of MR image. Rather than making heuristic assumptions on image statistics, these statistical properties can be utilized to establish an appropriate framework of stochastic modeling for MR images. A summary of the previous section is as follows.

For an 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.

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

5. Context images have multinomial distributions, they are correlated and satisfy the Markovian property.

In this section, based on these properties, we present a new framework for stochastic MR image modeling and provide a better mathematical understanding of the related issues. 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 [17, 29, 33]. Our discussion mainly focuses on these two types of models and show that this framework results in efficient algorithms for image analysis.

3.2 Image Modeling

3.2.1 Pixel Modeling

Given a digital image consisting of $N \equiv N_1 \times N_2$ pixels, assume that this image contains K regions and that each pixel is decomposed into a pixel image x and a context image l. By ignoring information regarding the spatial ordering of pixels, we can treat context images (i.e., pixel labels) as random variables and describe them using a multinomial distribution with unknown parameters π_k. Since this parameter reflects the distribution of the total number of pixels in each region, π_k can be interpreted as a prior probability of pixel labels determined by the global context information. Thus, the relevant (sufficient) statistics are the pixel image statistics for each component mixture and the number of pixels of each component. The marginal probability measure for any pixel image, i.e., the finite mixtures distribution, can be obtained by writing the joint probability density of x and l and then summing the joint density over all possible outcomes of l, i.e., by computing $p(x_i) = \sum_l p(x_i, l)$, resulting in a sum of the following general form:

$$p(x_i) = \sum_{k=1}^{K} \pi_k p_k(x_i), \quad i = 1, \cdots, N \tag{12}$$

where x_i is the gray-level of pixel i. $p_k(x_i)$s are conditional region probability density functions (pdfs) with the weighting factor π_k, satisfying $\pi_k > 0$, and $\sum_{k=1}^{K} \pi_k = 1$. The generalized Gaussian pdf given region k is defined by [29]

$$p_k(x_i) = \frac{\alpha \beta_k}{2\Gamma(1/\alpha)} \exp\left[- |\beta_k (x_i - \mu_k)|^{\alpha} \right], \quad \alpha > 0, \quad \beta_k = \frac{1}{\sigma_k} \left[\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)} \right]^{1/2}. \tag{13}$$

where μ_k is the mean, $\Gamma(\cdot)$ is the Gamma function, β_k is a parameter related to the variance σ_k by

$$\beta_k = \frac{1}{\sigma_k}\left[\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right]^{1/2}. \tag{14}$$

When $\alpha \gg 1$, the distribution tends to a uniform pdf; for $\alpha < 1$, the pdf becomes sharper, for $\alpha = 2.0$, one has the Gaussian pdf, and for $\alpha = 1.0$ the Laplacian pdf. Therefore, the generalized Gaussian model is a suitable model to fit the histogram distribution of those images whose statistical properties are unknown since the kernel shape can be controlled by selecting different α values. The finite Gaussian mixture model (FGGM) for $\alpha = 2$, is also commonly referred to as the standard finite normal mixture model and has been the most frequently used form. It can be written as

$$p_k(x_i) = \frac{1}{\sqrt{2\pi}\sigma_k}\exp(-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}) \qquad i = 1, 2, \cdots, N \tag{15}$$

where μ_k and σ_k^2 are the mean and variance of the kth Gaussian kernel and K is the number of Gaussian components.

The whole image can be closely approximated by an independent and identically distributed random field $\mathbf{X}$. The corresponding joint pdf is

$$P(\mathbf{x}) = \prod_{i=1}^{N}\sum_{k=1}^{K}\pi_k p_k(x_i) \tag{16}$$

where $\mathbf{x} = [x_1, x_2, \cdots, x_N]$, and $\mathbf{x} \in \mathbf{X}$. Based on the joint probability measure of pixel images, the likelihood function under finite mixture modeling can be expressed as

$$\mathcal{L}(\mathbf{r}) = \prod_{i=1}^{N} p_{\mathbf{r}}(x_i) \tag{17}$$

where $\mathbf{r} : \{K, \alpha, \pi_k, \mu_k, \sigma_k, k = 1, \cdots, K\}$ denotes the model parameter set.

3.2.2 Context Modeling

In MR images, regions are piecewise contiguous, thus, a pixel label takes on discrete values and the labels of nearby pixels are strongly correlated. Using the equivalence between a Gibbs distribution and an MRF, it has been shown that a Gibbs distribution provides a joint probability measure for $\mathbf{l}$ in the following form [2, 10, 13]:

$$P(\mathbf{l}) = \frac{1}{Z_{\mathbf{l}}}\exp(-U(\mathbf{l})) \tag{18}$$

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 $V_c^{(i)}(\mathbf{l})$, where $U(\mathbf{l}) = \sum_{i=1}^{N} V_c^{(i)}(\mathbf{l})$. The most typical configuration of an MRF model is the pairwise interaction neighborhood system in which spatial neighbors occur only in pairs, i.e., the second order model [13]. We define the neighborhood of pixel i, denoted by ∂i, by opening a 3×3 window with pixel i as the central pixel. The energy function can then be written as

$$U(\mathbf{l}) = \sum_{i=1}^{N} \sum_{j \in \partial i} [\theta_i I(l_i, l_j)] \tag{19}$$

where $\{\theta_i\}$ is the Markov parameter and $I(.,.)$ is the indicator function [4]. By translating local context information into the energy function of the Gibbs measure, by the use of a clique structure or a Markov parameter, an inhomogeneous MRF model can be established. Again it is important to note that, in our formulation, the Markov parameter θ_i in Equation 19 is considered to be shift-variant, not a fixed constant as in the most previous work [2, 10]. We believe that the correlation among context images is primarily local and should be reflected by the Markov parameter values. For example, within one region, the context exhibits strong spatial continuity, while on region boundaries, the context reflects the natural discontinuity.

By simply applying the Bayes law, we can construct a unified framework to integrate the pixel image model with the context image model. It has been shown that requiring the conditional independence of the observed pixel images, and given the context images, is sufficient to ensure that the posterior distribution is an MRF [13, 37]. That is, a posterior conditional probability function for the context images $\mathbf{l}$, given the observed pixel images $\mathbf{x}$, has the form of a Gibbs distribution given by

$$P(\mathbf{l}|\mathbf{x}) = \frac{1}{Z_{\mathbf{l}|\mathbf{x}}} \exp(-U(\mathbf{l}|\mathbf{x})) \tag{20}$$

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

$$U(\mathbf{l}|\mathbf{x}) = \sum_{i=1}^{N} \left\{ \prod_{k=1}^{K} [\frac{1}{2} \ln(\sigma_k^2) + \frac{(x_i - \mu_k)^2}{2\sigma_k^2}]^{I(l_i,k)} + \sum_{j \in \partial i} [\theta_i I(l_i, l_j)] \right\} \tag{21}$$

using Equations 15 and 19.

Equation 20 refers to a hidden MRF, where the local property can be derived as

$$p(l_i|\mathbf{l}_{\partial i}, \mathbf{x}) = \frac{1}{Z_i} \exp(-U(l_i|\mathbf{l}_{\partial i}, \mathbf{x})) \tag{22}$$

where Z_i is a normalizing constant, $\mathbf{l}_{\partial i}$ local context images, and

$$U(l_i|\mathbf{l}_{\partial i}, \mathbf{x}) = \prod_{k=1}^{K} [\frac{1}{2}\ln(\sigma_k^2) + \frac{(x_i - \mu_k)^2}{2\sigma_k^2}]^{I(l_i,k)} + \sum_{j\in\partial i}^{c} [\theta_i I(l_i, l_j)]\,. \tag{23}$$

The spatial statistical dependence among pixel images is one of the fundamental issues in the mathematical formulation for image analysis. In our approach, we tackle the problem as follows: For the purpose of tissue quantification, maximum likelihood (ML) estimation, based on the SFNM model given in Equation 17, is used. In [38] we prove a convergence theorem showing that, when the pixel images are asymptotically independent, the parameter estimates based on Equation 17 converge to their true values with probability one for a sufficiently large N. For the purpose of image segmentation, maximum posterior probability (MAP) approach is employed to update $\mathbf{l}$ according to Equation 20. Since in the formulation, we allow θ_i to be adjustable, the assignment of their values will incorporate the unified correlation among both the pixel and context images. Our tests with a wide class of simulated and real data demonstrate the plausibility of the approach.

3.3 Model Identification

Once the model is chosen, identification addresses the estimation of the local region parameters $(\pi_k, \mu_k, \sigma_k, k = 1, \cdots, K)$ and the structural parameters (K, α). In particular the estimation of the order parameter K, is referred to as model order selection.

3.3.1 Parameter Estimation

With an appropriate system likelihood function, the objective of model identification is to estimate the model parameters by maximizing the likelihood function, or equivalently, minimizing the relative entropy between the image histogram $p_{\mathbf{x}}(u)$ and the estimated pdf $p_{\mathbf{r}}(u)$, where u is the gray level [4, 39]. There are a number of approaches to perform the ML estimation of finite mixture distributions [47]. The most popular method is the expectation-maximization (EM) algorithm [40, 41]. EM algorithm first calculates the posterior Bayesian probabilities of the data based on the observations, and obtains the current parameter estimates (E-step). Then it updates parameter estimates using generalized mean ergodic theorems (M-step). The procedure moves back and forth between these two steps. The successive iterations increase the likelihood of the model parameters being estimated. A neural network interpretation of this procedure is given in [42].

We can use relative entropy (the Kullback-Leibler distance) [43] for parameter estimation, i.e., we can measure the information theoretic distance between the histogram of the pixel images, denoted by $p_{\mathbf{x}}$, and the estimated distribution $p_{\mathbf{r}}(u)$ which we define as the global relative entropy (GRE)

$$D(p_{\mathbf{x}}||p_{\mathbf{r}}) = \sum_u p_{\mathbf{x}}(u) \log \frac{p_{\mathbf{x}}(u)}{p_{\mathbf{r}}(u)}. \tag{24}$$

It can be shown that, when relative entropy is used as the distance measure, distance minimization is equivalent to the maximum likelihood (ML) estimation of the model parameters [39, 4].

For the case of the FGGM model, the EM algorithm can be applied to the joint estimation of the parameter vector and the structural parameter α as follows [40]:

EM Algorithm:

1. For $\alpha = \alpha_{min}, \cdots, \alpha_{max}$

   - $m = 0$, given initialized $\mathbf{r}^{(0)}$
   - E-step: for $i = 1, \cdots, N,\ k = 1, \cdots, K$, compute the probabilistic membership

$$z_{ik}^{(m)} = \frac{\pi_k^{(m)} p_k(x_i)}{\sum_{k=1}^K \pi_k^{(m)} p_k(x_i)} \tag{25}$$

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

$$\begin{cases} \pi_k^{(m+1)} = \frac{1}{N} \sum_{i=1}^{N_1 N_2} z_{ik}^{(m)} \\ \mu_k^{(m+1)} = \frac{1}{N\pi_k^{(m+1)}} \sum_{i=1}^{N} z_{ik}^{(m)} x_i \\ \sigma_k^{2(m+1)} = \frac{1}{N\pi_k^{(m+1)}} \sum_{i=1}^{N} z_{ik}^{(m)} (x_i - \mu_k^{(m+1)})^2 \end{cases} \tag{26}$$

   - When $|GRE^{(m)}(p_{\mathbf{x}}||p_{\mathbf{r}}) - GRE^{(m+1)}(p_{\mathbf{x}}||p_{\mathbf{r}})| \leq \epsilon$ is satisfied, go to Step 2

     Otherwise, $m = m + 1$ and go to E-Step.

2. Compute GRE, and go to Step 1

3. Choose the optimal $\hat{\mathbf{r}}$ which corresponds to the minimum GRE.

However, the EM algorithm, generally, has the reputation of being slow, since it has a first order convergence in which new information acquired in the expectation step is not used immediately

[44]. Recently, a number of on-line versions of the EM algorithm were proposed for large scale sequential learning (e.g. see [4, 45]– [48]). Such a procedure obviates the need to store all the incoming observations, and changes the parameters immediately after each data point allowing for high data rates. Titterington [47] developed a stochastic approximation procedure which is closely related to the probabilistic self organizing mixtures (PSOM) algorithm we are introducing here, and shows that the solution can be consistent. Other similar formulations are due to Marroquin *et al.* [45] and Weinstein *et al.* [48].

For the adaptive estimation of the SFNM model parameters, we can derive an incremental learning algorithm by simple stochastic gradient descent minimization of $D(p_{\mathbf{x}}||p_{\mathbf{r}})$ [4, 49] given in Equation 24 with $p_{\mathbf{r}}$ given by Equation 15:

$$\mu_k{}^{(t+1)} = \mu_k^{(t)} + a(t)(x_{t+1} - \mu_k^{(t)})z_{(t+1)k}^{(t)}, \tag{27}$$

$$\sigma_k^{2(t+1)} = \sigma_k^{2(t)} + b(t)[(x_{t+1} - \mu_k^{(t)})^2 - \sigma_k^{2(t)}]z_{(t+1)k}^{(t)},$$

$$k = 1, \cdots, K. \tag{28}$$

where $a(t)$ and $b(t)$ are introduced as the learning rates, two sequences converging to zero, and ensuring unbiased estimates after convergence. For details of the derivation and approximations, see [4, 9]. Based on a generalized mean ergodic theorem [50], updates can also be obtained for the constrained regularization parameters, π_k, in the SFNM model. For simplicity, given an asymptotically convergent sequence, the corresponding mean ergodic theorem, i.e., the recursive version of the sample mean calculation, should hold asymptotically. Thus, we define the interim estimate of π_k as in reference [51]:

$$\pi_k^{(t+1)} = \frac{t}{t+1}\pi_k^{(t)} + \frac{1}{t+1}z_{(t+1)k}^{(t)}. \tag{29}$$

Hence the updates given by Equations 27, 28, and 29 together with an evaluation of Equation 25 using Equation 15, provide the incremental procedure for computing the SFNM component parameters. Their practical use however, requires strongly mixing condition and a decaying annealing procedure (learning rate decay) [50, 52, 53]. In finite mixtures parameter estimation, the algorithm initialization must be chosen carefully and appropriately. In [51], an adaptive Lloyd-Max histogram quantization (ALMHQ) algorithm is introduced for threshold selection which is also well suited to initialization in an ML estimation. It can be used for initializing the network parameters: μ_k, σ_k^2, and π_k, $k, 1, 2, \cdots, K$.

3.3.2 Model Order Selection

The determination of the region parameter K directly affects the quality of the resulting model parameter estimation and in turn, affects the result of segmentation. In a statistical problem formulation such as the one introduced in the previous section, the use of information theoretic criteria for the problem of model determination arises as a natural choice. Two popular approaches are Akaike's information criterion (AIC) [55], and Rissanen's minimum description length (MDL) [56]. Akaike proposes the selection of the model that gives the minimum AIC, that is defined by

$$AIC(K_a) = -2\log(\mathcal{L}(\hat{\mathbf{r}}_{ML})) + 2K_a \tag{30}$$

where $\hat{\mathbf{r}}_{ML}$ is the maximum likelihood estimate of the model parameter set $\mathbf{r}$, and K_a is the number of free adjustable parameters in the model [17, 55] and is given by $3K - 1$ for the SFNM model. The AIC selects the correct number of the image regions K_0 such that

$$K_0 = \arg\left\{\min_{1\leq K\leq K_{MAX}} AIC(K_a)\right\}. \tag{31}$$

Rissanen addresses the problem from a quite different point of view. He reformulates the problem explicitly as an information coding problem in which the best model fit is measured such that high probabilities are assigned to the observed data, while at the same time the model itself is not too complex to be described [56]. The model is selected by minimizing the total description length defined by

$$MDL(K_a) = -\log(\mathcal{L}(\hat{\mathbf{r}}_{ML})) + 0.5K_a\log(N). \tag{32}$$

Similarly, the correct number of distinctive image regions K_0 can be estimated as

$$K_0 = \arg\left\{\min_{1\leq K\leq K_{MAX}} MDL(K_a)\right\}. \tag{33}$$

It is also worth noting that Schwartz [57] arrives at the same formulation as Rissanen by using Bayesian arguments.

A more recent formulation of an information theoretic criterion, the minimum conditional bias and variance criterion (MCBV) [4, 58] 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. The formulation is based on the fundamental argument that the value of the structural parameter can not be arbitrary or infinite, because such an estimate might be said to have a low 'bias' but the price to be paid is high 'variance' [25].

Since the joint maximum entropy is a function of K_a and $\hat{\mathbf{r}}$, by taking advantage of the fact that model estimation is separable into components and structure, we define the MCBV criterion as

$$\text{MCBV}(K) = -\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}_{ML})) + \sum_{k=1}^{K_a} H(\hat{r}_{kML}) \tag{34}$$

where $-\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}_{ML}))$ is the conditional bias (a form of information theoretic distance) [59, 50], and $\sum_{k=1}^{K_a} H(\hat{r}_{kML})$ is the conditional variance (a measure of model uncertainty) [59, 53], of the model. As both of these two terms represent natural estimation errors about their true models, they can be treated on an equal basis. A minimization of the expression in Equation 34 leads to the following characterization of the optimum estimation

$$K_0 = \arg\left\{\min_{1\leq K\leq K_{MAX}} \text{MCBV}(K)\right\}. \tag{35}$$

That is, if the cost of model variance is defined as the entropy of parameter estimates, 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. A practical MCBV formulation with code-length expression is further given by [50, 58]

$$\text{MCBV}(K) = -\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}_{ML})) + \sum_{k=1}^{K_a} \frac{1}{2}\log 2\pi e Var(\hat{\mathbf{r}}_{kML}) \tag{36}$$

where the calculation of $H(\hat{\mathbf{r}}_{kML})$ requires the estimation of the true ML model parameter values. It is shown that, for sufficiently large number of observations, the accuracy of the ML estimation tends quickly to the best possible as determined by the Cramer-Rao lower bounds (CRLBs) [53]. Thus, the CRLBs of the parameter estimates are used in the actual calculation to represent the "conditional" bias and variance [54]. We have found that, experimentally, the MCBV formulation for determining the value of K_0, exhibits very good performance consistent with both the AIC and the MDL criteria. It should be noted, however, that these are not the only plausible approaches to the problem of order selection; other approaches such as cross validation techniques may also be quite useful [60]– [64].

3.3.3 Segmentation

Image segmentation is a technique for partitioning the image into meaningful regions corresponding to different objects. It may be considered to be a clustering process where the pixels are classified

into attributed tissue types according to their gray-level values and spatial correlation. More precisely, image segmentation addresses the realization of context images $\mathbf{l}$, given the observed pixel images $\mathbf{x}$ [36]. Based on the inhomogeneous MRF model given by Equation 20, there are several approaches for performing the pixel classification. For example, we can use ML classification to directly maximize the individual likelihood function of the pixel images, i.e., the first term in Equation 20, by searching the optimum $\mathbf{l}$ where the true pixel labels $\mathbf{l}^*$ are considered to be functionally independent constants [36]. The major problem associated with this approach is that the classification error is high when the observed pixel images are noisy. However, it may well function as an initial solution since the classification error is spatially uniformly distributed. By considering both terms in Equation 20, we derive the modified iterated conditional modes (MICM) algorithm to search for a MAP image segmentation. The structure of this relaxation labeling procedure is based on two basic considerations: 1) decomposition of the global computation scheme into a network performing simple local computations; and 2) use of suitable local context regularities in resolving ambiguities [36].

Let ϵ denote the expected segmentation error, i.e., the posterior cost of misclassification given pixel images $\mathbf{x}$, based on the following equivalence

$$\arg\{\max_k P(l_i = k, \mathbf{l}_{S|i}|\mathbf{x})\} = \arg\{\max_k p(l_i = k|\mathbf{l}_{S|i}, \mathbf{x})\} \tag{37}$$

where $\mathbf{l}_{S|i}$ denotes pixel labels of all pixels except pixel i. Reference [13, 37] shows that choosing the labeling that minimizes ϵ is equivalent to maximizing the marginal posterior distribution such that label l_i is updated to satisfy

$$p(l_i^{(m+1)}|\mathbf{x}, \mathbf{l}_{S|i}) \geq p(l_i^{(m)}|\mathbf{x}, \mathbf{l}_{S|i}) \tag{38}$$

for all pixels. Furthermore, by imposing the Markovian constraint, we have the following relationship [13, 37]:

$$p(l_i|\mathbf{x}, \mathbf{l}_{S|i}) \propto p(x_i|l_i)p(l_i|\mathbf{l}_{\partial i}) \propto p(l_i = k|x_i, \mathbf{l}_{\partial i}). \tag{39}$$

That is,

$$\arg\{\max_k p(l_i|\mathbf{x}, \mathbf{l}_{S|i})\} = \arg\{\max_k p(x_i|l_i)p(l_i|\mathbf{l}_{\partial i})\} = \arg\{\max_k p(l_i|x_i, \mathbf{l}_{\partial i})\}. \tag{40}$$

Hence, based on the inhomogeneous hidden MRF formulation, the MICM algorithm is constructed as a computationally feasible alternative to obtain an MAP solution. From Equation 22, we can show that maximizing the conditional probability is equivalent to minimizing the energy function

$U(l_i|\mathbf{l}_{\partial i}, \mathbf{x})$. That is, pixel i will be classified into the kth region, if

$$l_i = \arg\left\{\min_k U(l_i = k|\mathbf{l}_{\partial i}, \mathbf{x})\right\}. \tag{41}$$

The MICM algorithm uses a specified number of iterations by randomly visiting all pixels. At each single cycle in the iterations, for the update of l_i, we have

$$P(\mathbf{l}|\mathbf{x}) = p(l_i|\mathbf{l}_{S|i}, \mathbf{x})P(\mathbf{l}_{S|i}|\mathbf{x}) \tag{42}$$

which never decreases, hence assuring eventual convergence [13, 37]. This approach demonstrates how a network of discrete units can be used to search an optimal solution in a problem where incorporation of context constraints is important.

To perform image segmentation, based on the inhomogeneous MRF model requires assignment of the Markov parameter θ_i [13]. An important area in image analysis that has recently been emphasized is the use of patient-specific site models to guide various medical image analysis tasks [14]. Hence, in the assignment of the Markov parameter, we first construct a patient-specific probabilistic atlas directly from the interim context images (e.g., after ML classification), where the entropy rate distribution is used as a measure of the pixel label dependence, and then, assign the Markov parameter values based on this model through a pre-defined look-up table. This approach has a clear physical interpretation, since a lower entropy rate means higher dependence and the probabilistic atlas coincides with the boundary allocation [31].

For a stationary random process $l_1, l_2, ..., l_N$, the entropy rate $H(\mathcal{L})$ is defined as the rate of entropy growth [50]. When dealing with a stationary Markov chain, the entropy rate can be calculated by

$$H(\mathcal{L}) = \lim_{N\to\infty} H(l_N|l_{N-1}, ..., l_1) = \lim_{N\to\infty} H(l_N|l_{N-1}) = H(l_2|l_1). \tag{43}$$

Assume that the inhomogeneous MRF is locally stationary, then the local entropy rate, i.e., the entropy rate distribution, is given by

$$H(\mathcal{L}_i) = \sum_{\mathbf{l}_{\partial i}} P(\mathbf{l}_{\partial i})H(l_i|\mathbf{l}_{\partial i}). \tag{44}$$

Since the probability distribution of $\mathbf{l}_{\partial i}$ is generally unavailable, we use first order stochastic approximation to estimate the entropy rate [31, 53]:

$$H(\mathcal{L}_i) \approx H(l_i|\mathbf{l}_{\partial i}) = -\sum_{k=1}^{K} p(l_i = k|\mathbf{l}_{\partial i}) \log p(l_i = k|\mathbf{l}_{\partial i}) \tag{45}$$

By opening a 3×3 neighborhood window, the conditional probability is given by

$$p(l_i = k | \mathbf{l}_{\partial i}) = \sum_{j \in \partial i} \frac{I(l_j, k)}{8}. \tag{46}$$

It can be shown that when $\mathbf{l}_{\partial i}$ represents a uniform block the corresponding entropy rate is zero. Finally, we use the following look-up table to assign the Markov parameter values

$$\theta_i = \frac{\alpha}{H(\mathcal{L}_i) + \eta} \tag{47}$$

where α is the scale factor and η is the shifting offset, and these coefficients are empirically determined based on our experience. We refer to the plot of θ_i as the *patient-specific probabilistic atlas* or the *site model*. Finally, we introduce, in addition to the iterated update of context images $\mathbf{l}$ using Equation 41 and subsequent modification of the Markov parameter θ according to the look-up table Table 47, the fine tuning of the tissue conditional likelihood densities through a "classification-based learning" scheme [38]. The classification-based learning method uses the *misclassified* pixels to adjust the tissue density functions, which are previously estimated using the EM algorithm, so that a minimum classification error can be achieved. Integrated into the MICM algorithm, the *reinforced* and *anti-reinforced* learning rules are used to update the tissue quantities μ_k and σ_k^2 in the first term of Equation 20 at each iteration:

$$\begin{array}{ll}
\text{Reinforced learning:} & \mu_{l_i^{(n+1)}}^{(n+1)} = \mu_{l_i^{(n+1)}}^{(n)} + \kappa(x_i - \mu_{l_i^{(n+1)}}^{(n)}) \\
& \sigma_{l_i^{(n+1)}}^{2(n+1)} = \sigma_{l_i^{(n+1)}}^{2(n)} + \kappa[(x_i - \mu_{l_i^{(n+1)}}^{(n)})^2 - \sigma_{l_i^{(n+1)}}^{2(n)}] \\
\text{Anti-reinforced learning:} & \mu_{l_i^{(n)}}^{(n+1)} = \mu_{l_i^{(n)}}^{(n)} - \kappa(x_i - \mu_{l_i^{(n)}}^{(n)}) \\
& \sigma_{l_i^{(n)}}^{2(n+1)} = \sigma_{l_i^{(n)}}^{2(n)} - \kappa[(x_i - \mu_{l_i^{(n)}}^{(n)})^2 - \sigma_{l_i^{(n)}}^{2(n)}]
\end{array} \tag{48}$$

where the *misclassified* pixels are those with $l_i^{(n+1)} \neq l_i^{(n)}$ and κ is the learning rate constant, which typically chosen to be small ($\kappa \ll 1$). It is important to note that, if the pixels are re-classified into $l_i^{(n+1)}$th tissue type, reinforced learning will be applied to pull the corresponding kernel closer to the host region, while anti-reinforced learning will be used to push the kernel of the $l_i^{(n)}$th tissue type away from the problematic region [38]. This *unsupervised* adaptive learning rule incorporates the partial volume effect and the inhomogeneity distortion into the relaxation labeling to improve image segmentation. The misclassified pixels associated with the spatial discontinuities are used to fine-tune the discriminant thresholds in the decision domain. We refer to those misclassified pixels as the boundary-associated pixels since significant partial volume effect and inhomogeneity distortion most often occurs along the boundaries [3].

4 Algorithm and Experiment

So far, we have described the theory on which image analysis scheme supported by the stochastic site model, with tissue quantification and segmentation is based. We now complete the description of our algorithm by considering the major steps in the implementation and illustration of its performance. Although tissue quantification and image segmentation may be simultaneously performed [3, 6, 5], a more accurate result can be achieved, if the two objectives are considered separately [4, 10, 13]. Guided by the two information theoretic criteria, our algorithm progressively proceeds by fitting a SFNM with model order selection to the histogram of pixel images, and then constructing a consistent relaxation labeling of the context images. A summary of the major steps is as follows:

1. For each value of K, perform an ML tissue quantification by applying the EM algorithm (Equations 25-26).
2. Scan the values of $K = K_{\min,}...K_{\max}$, by using AIC in Equation 30 and MDL Equation 32 to determine the suitable number of tissue types K_0.
3. Select the result of tissue quantification corresponding to the value of K_0 determined in step 2.
4. Initialize image segmentation using the ML classification method.
5. Construct the probabilistic atlas or site model according to the entropy rate distribution (Equations 45-46) and the look-up table Table 47.
6. Finalize image segmentation by applying the MICM algorithm by implementing relaxation labeling Equation 41 and local parameter tuning Equation 48.

The performance of the algorithm is evaluated in terms of the global relative entropy (GRE) value for tissue quantification, visual judgement, and post-GRE value for image segmentation, as well as computational complexity and robustness [4, 16, 37].

We first illustrate the application of our algorithm on a simple synthetic data set generated from a mixture of four Gaussian densities in a two-dimensional space, shown in Figure 1 (left). Each component represents one tissue type. The value for each of the components is set to a constant value and normally distributed noise is added to the simulation image resulting in a 6 dB SNR. Figure 1 (right) shows the corresponding histogram of the pixel images (noisy curve). It is clear

k	$\pi/\hat{\pi}$	$\mu/\hat{\mu}$	$\sigma^2/\hat{\sigma}^2$
1	0.25/0.24	86/84	400/354
2	0.125/0.13	126/121	400/365
3	0.5/0.48	166/164	400/373
4	0.125/0.15	206/201	400/463

Table 1: True and estimated parameter values for the simulated image in Figure 1.

that these four simulated tissue types overlap considerably. We then apply the two information theoretic criteria (AIC and MDL) to detect the number of the components. The curves of the AIC and MDL, as functions of the number of tissue types K, are plotted in the Figure 2, where the left and right plots correspond to two different noise levels. The minima of these curves indicate the correct number of the image components. It is clear that the number of tissue components suggested by both criteria is correct. The results of the distribution learning, using the EM algorithm, are shown in Figure 1 (right: smooth curve). With the correct K_0, the GRE between the estimated SFNM and the image histogram is about 0.008 nats. The numerical results are given in Table 1.

Having determined the parameters of the SFNM model, we conduct image segmentation in which the performance of three different algorithms is compared. The ML pixel classification is applied to initialize image segmentation, in which the pixel is classified into the kth tissue component based on the maximum likelihood, i.e., $l_i = \arg\{\min_k \prod_{k=1}^{K} [\frac{1}{2}\ln(\sigma_k^2) + \frac{(x_i-\mu_k)^2}{2\sigma_k^2}]^{I(l_i,k)}\}$. The result is shown in Figure 3 (left). The corresponding classification error is about 30% and is spatially and uniformly distributed over the whole image. We then apply the ICM algorithm to update the pixel labeling with two settings: $\theta = 0.4$ and $\theta = 6$. The results are shown in Figure 3 (middle, right). It can be concluded that, since the ICM algorithm uses a homogeneous MRF configuration, a smaller θ will lead to a noisy image segmentation with a higher classification error and a larger θ will lead to a smooth context recovery with missing details. For the given simulated image, we determine that $\theta = 1.5$ as the best choice, and with this setting, the corresponding classification error is down to 0.7508%.

Finally, we use the MICM algorithm together with the Markov parameter assignment procedure, to improve image segmentation. Based on Equations 45–47, our intensive numerical experiments suggest that the assignment of $\theta = 0.4$ along the boundaries, and $\theta = 1.5 - 6.0$ within the regions, results in good performance in a typical tone image [13, 37]. Using this empirical knowledge as a guideline, we select $\alpha = 0.5$ and $\eta = 0.1$ to construct the look-up table. Figure 4 (left) gives the distribution of the Markov parameter values for the simulated image, clearly showing the structure

of the boundaries. The result using the MICM algorithm is shown in Figure 4 (right), that provides an almost a "perfect" true context in which the image details are well preserved while most noise effects have been removed. The final classification error is further reduced to 0.3113%.

As an example of a more complex problem, we consider a set of real MR brain images used for identification of different tissue space domains (3D SPGR). A slice from such a data set is shown in Figure 5 (left). This is a T2-weighted image parallel to the AC-PC line. The data are acquired with a GE Sigma 1.5 Tesla system. The imaging parameters are TR 35, TE 5, flip angle 45^0, 1.5 mm effective slice thickness, 0 gap, 124 slices with in-plane 192×256 matrix, and a 24 cm field of view, and 1 nex. Since the skull, scalp, and fat in the original brain images do not contribute to the brain tissue, we edit the MR images to exclude non-brain structures prior to tissue quantification and segmentation. A close look at the whole three dimensional data set indicates that the histogram has considerably different characteristics from slice to slice and the tissue types are all greatly overlapping [4]. Our experiment assesses the feasibility of the algorithm on model selection and probabilistic atlas construction with real MR images. After the initial image segmentation using the ML pixel classification, the corresponding site model is given in Figure 5 (right). Applying both AIC and MDL to this image, we obtain $K_0 = 10$. The corresponding MDL curve is shown in Figure 6 (left) with a close up around the minimum on the right in Figure 6. Our experience has shown that the overall performance of these two information theoretic criteria is fairly consistent with real MR brain images.

As a final example, we consider tissue quantification and image segmentation of a normal T1-weighted MR brain image as shown in Figure 7 (left). As discussed in the literature [2, 3], the brain tissue is generally composed of three principal types, i.e., white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), and their pair-wise combinations. Santago [3] propose a six-tissue model to represent the primary tissue types and the partial volume effect, defined as CSF-White (CW), CSF-Gray (CG), and Gray-White (GW). In our implementation, we also consider the triple mixture tissue, defined as CSF-White-Gray (CWG). More importantly, since the MR images clearly show distinctive intensities at local brain areas, the functional areas within a tissue type need to be considered. The caudate nucleus and putamen are two particularly important local brain functional areas. We calculate AIC(K) and MDL(K) for $K = 2, ..., 10$. The results with these two criteria are given in Table 2, which suggest that this brain image contains 8 tissue types. The histogram of the pixel images is shown in Figure 7 (right: noisy curve). With $K_0 = 8$ the result of tissue quantification using our algorithm is given by Figure 7 (right: smooth curve) with a GRE value of 0.02-0.04 nats. The actual parameter quantities corresponding to the estimation agree with those of a physician's qualitative analysis results [4]. The MICM based tissue segmentation,

K	2	3	4	5	6	7	8	9	10
AIC	295944	291667	288845	288397	288380	288184	**285685**	287674	286317
MDL	295967	291703	288895	288461	288458	288275	**285790**	287792	286449

Table 2: Model selection using AIC and MDL values.

together with the local parameter tuning Equation 48, is performed and the pixel label updates are terminated after 5-10 iterations, since further iterations produced almost identical results. The segmentation results are given in Figure 8. Although the segmentation results contain some small isolated spots (less than 4-pixel size), the proposed approach is quite encouraging. It is seen that the boundaries of WM, GM, and CSF are delineated successfully. The regions with different gray levels are satisfactorily segmented, especially, the major brain tissues are clearly identified representing eight brain tissue types: (1) CSF, (2) CG, (3) CGW, (4) GW, (5) GM, (6) putamen area, (7) caudate area, and (8) WM. These segmented tissue types also agree with the results of radiologists' evaluation.

We have tested our algorithm on a total of 124 MR brain images. As illustrated by both simulated and real examples here, we have observed that the overall approach can be very effective at quantifying and segmenting different tissue types. It is important to emphasize that in unsupervised image analysis with real data sets, an objective evaluation of different techniques is a particularly difficult undertaking [6, 10, 16]. It is difficult to quantify the merit of a particular algorithm, and the dependability of evaluations, by the simple mathematical measures, is questionable [4]. Therefore, the quality of tissue quantification and image segmentation, usually, depends heavily on subjective and qualitative judgements. In this work, we use the evaluation performed by radiologists, and the GRE value to measure the quality of tissue quantification. We believe that this criterion is quite reasonable since the ML estimation is unbiased under the SFNM modeling. Also the image histogram, as the reference, is the only subjective quantity directly available from the image data [47, 53]. Furthermore, for the assessment of image segmentation, we use both the post-GRE value and the visual inspection, where the post-GRE value is generated by the sample averages based on $\mathbf{l}$

$$\begin{cases} \hat{\hat{\pi}}_k = \frac{1}{N}\sum_{i=1}^{N} I(l_i,k) \\ \hat{\hat{\mu}}_k = \frac{1}{N\pi_k}\sum_{i=1}^{N} I(l_i,k)x_i \\ \hat{\hat{\sigma}}_k^2 = \frac{1}{N\pi_k}\sum_{i=1}^{N} I(l_i,k)(x_i-\mu_k)^2 \end{cases} \tag{49}$$

and is given by

$$D(f_{\mathbf{z}}||f_{\mathbf{l}}) = \sum_{x\in\mathcal{X}} f_{\mathbf{z}}(x)\log\frac{f_{\mathbf{z}}(x)}{f_{\mathbf{l}}(x)} \tag{50}$$

Method	ML-EM	ICM	MICM	
	(*soft*)	(*hard*-homogeneous)	(*hard*-inhomogeneous)	
GRE value (nats)	0.0067	0.4406	0.1578	

Table 3: Comparison of segmentation error resulting from different clustering method.

as an indirect, but objective criterion, where $f_{\mathbf{l}}(x)$ denotes the SFNM distribution corresponding to the segmented image. Considering the probabilistic pixel label $\mathbf{z}$ and the realized pixel label $\mathbf{l}$, the pre-segmentation SFNM $f_{\mathbf{z}}$ and post-segmentation SFNM $f_{\mathbf{l}}$ are related through the image segmentation obtained by the MICM algorithm. The post-GRE segmentation criterion presented here states that the relative entropy between the ML estimate of the SFNM (via soft clustering) and the SFNM obtained from the segmented image (via hard clustering) is minimal, if the image components are correctly segmented. The parameter values of a particular tissue type in the estimated histogram are most likely to be equal to the parameter values of the corresponding tissue type in the segmented region if the pixel images are properly classified. This correspondence is lost in the case of misclassification. Our experiment with the MR brain image leads to a result of $D(f_{\mathbf{x}}||f_{\mathbf{l}}) = 0.1578$ nats and $D(f_{\mathbf{x}}||f_{\mathbf{z}}) = 0.0067$ nats, which coincides with that suggested by the information theoretic "Pythagorean" theorem (Theorem 12.6.1 in [50]), $D(f_{\mathbf{x}}||f_{\mathbf{l}}) \geq D(f_{\mathbf{x}}||f_{\mathbf{z}}) + D(f_{\mathbf{z}}||f_{\mathbf{l}})$. This result implies that the classification error in image segmentation contributes to the additional quantification error $D(f_{\mathbf{z}}||f_{\mathbf{l}})$. Table 3 gives the summary of the results.

5 Summary

We have presented a complete procedure for quantifying and segmenting major brain tissue types from MR images. This method, as illustrated by well-planned simulation and pilot applications in brain tissue analysis, is capable of revealing abnormalities within data and can be applied to clinical problems such as those encountered in tissue segmentation and quantitative diagnosis. It is important to emphasize that although the experiment results here are presented for two-dimensional images, their application to three-dimensional data sets is straightforward and has also been implemented recently [65]. We show two examples of 3D segmentation results in Figures 9 and 10.

The main limitations of the current approach are that: 1) there is no systematic way to initialize the model parameter values, and 2) the algorithm for both the tissue quantification and model

selection is frequently trapped in a local optima. Thus, future research should focus on parameter initialization supported by a generic site model and a multiresolution or hierarchical scheme for both model selection and tissue quantification. Additional work should include a theoretical investigation of the new formulation in model selection and the objective criterion for image segmentation evaluation.

Figure Captions:

Figure 1.
Simulated image with four Gaussian components (left) and the corresponding histogram of the pixel images (right)

Figure 2.
Model selection with simulated image using AIC, MDL, and MCBV.

Figure 3.
Results of image segmentation using the ML pixel classification (left), the ICM algorithm after the 3rd iteration with $\theta = 0.4$ (middle), and the ICM algorithm after the 3rd iterations with $\theta = 6$ (right)

Figure 4.
Contextual image segmentation using a probabalistic site model (left) supported approach (right)

Figure 5.
A real MR brain image and the corresponding probabilistic site model

Figure 6.
An MDL curve of the MR image in Fig. 5, the left picture is the global profile and the right picture shows the local detail, where the minimum of the MDL value corresponds to $K_0 = 10$

Figure 7.
A T1-weighted "normal" MR brain image and the corresponding image histogram

Figure 8.
Result of image segmentation using the MICM algorithm and local parameter fine tuning

Figure 9.

Figure 10.

References

[1] A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer, "Morphometric analysis of white matter lesions in MR images: method and validation," *IEEE Trans. Med. Imaging*, vol. 13, no. 4, pp. 716–724, December 1994.

[2] H. S. Choi, D. R. Haynor, and Y. Kim, "Partial volume tissue classification of multichannel magnetic resonance images–A mixture model," *IEEE Trans. Med. Imaging*, Vol. 10, pp. 395–407, September 1994.

[3] 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, September 1993.

[4] Y. Wang, T. Adalı, S.-Y. Kung, and Z. Szabo, "Quantification and segmentation of brain tissue from MR images: A probabilistic neural network approach," *IEEE Trans. Image Processing*, Special Issue on Applications of Neural Networks to Image Processing, vol. 7, no. 8, pp. 1165–1181, Aug. 1998.

[5] H. E. Cline, W. E. Lorensen, R. Kikinis, and R. Jolesz, "Three-dimensional segmentation of MR images of the head using probability and connectivity," *J. Comp. Assisted Tomography*, Vol. 14, pp. 1037–1045, 1990.

[6] 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 Nets.*, Vol. 3, pp. 672–682, 1992.

[7] 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.

[8] 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.

[9] Y. Wang, T. Adalı, C. Lau, and S.-Y. Kung, "Quantitative analysis of MR brain image sequences by adaptive self-organizing mixtures," *Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*, Special Issue on Neural Networks for Biomedical Image Processing, vol. 18, no. 3, pp. 219–239, Apr. 1998.

[10] 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, pp. 441–449, September 1994.

[11] J. Xuan and T. Adalı, "Segmentation of magnetic resonance brain image: integrating region growing and edge detection," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, Washington, DC, Nov. 1995, vol. 3, pp. 544–547.

[12] Y. Wang, T. Adalı, M. T. Freedman, and S. K. Mun, "MR brain image analysis by distribution learning and relaxation labeling," *Proc. 15th South. Biomed. Eng. Conf.*, pp. 133–136, Dayton, Ohio, March 1996.

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

[14] G. P. Robinson, A. C. Colchester, and L. D. Griffin, "Model-based recognition of anatomical objects from medical images," *Proc. Infor. Processing Med. Imaging*, Springer-Verlag, 1993.

[15] Y. Wang, T. Adalı, J. Xuan, Z. Szabo, "Magnetic resonance image analysis by information theory and stochastic site model," submitted to *IEEE Trans. Medical Imag.*

[16] Y. J. Zhang, "A Survey on Evaluation Methods for Image Segmentation," *Pattern Recognition,* Vol. 29, No. 8, pp. 1335–1346, 1996.

[17] T. Lei and W. Sewchand, "Statistical Approach to X-Ray CT Imaging and Its Application in Image Analysis–Part II: A New Stochastic Model-Based Image Segmentation Technique for X-Ray CT Image," *IEEE Trans. on Med. Imaging,* Vol. 11, No. 1, pp. 62–69, March 1992.

[18] 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.

[19] 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, pp. 257–264, May 1986.

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

[22] M. Fuderer, "The information content of MR images," *IEEE Trans. Med. Imaging*, Vol. 7, No. 4, pp. 368–380, 1988.

[23] Y. Wang, *MR Imaging Statistics and Model-Based MR Image Analysis*, Ph.D. Dissertation, University of Maryland, Baltimore County, May 1995.

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

[25] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural Computation*, 4, pp. 1–52, 1992.

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

[27] A. K. Jain, "Markov Random fields: Theory and Application," Boston, Academic Press, 1993.

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

[29] J. Zhang and J. W. Modestino, "A Model-Fitting Approach to Cluster Validation with Application to Stochastic Model-Based Image Segmentation," *IEEE Trans. on PAMI,* Vol. 12, No. 10, pp. 1009–1017, October 1990.

[30] C. Bouman and B. Liu, "Multiple Resolution Segmentation of Texture Images," *IEEE Trans on Pattern Anal. and Machine Intell.*, Vol. 13, No. 2, pp. 99–113, February 1991.

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

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

[33] 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, pp. 428–433, 1983.

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

[35] D. M. Titterington, "Comments on application of the conditional population-mixture model to image segmentation'," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 6, No. 5, pp. 656–658, September 1984.

[36] R. A. Hummel and S. W. Zucker, "On the foundations of relaxation labeling processes," *IEEE Trans. Pattern Ana. Machine Intell.*, Vol. 5, No. 3, May 1983.

[37] 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.

[38] Y. Wang, S-H Lin, H. Li, and S-Y Kung, "Data mapping by probabilistic modular network and information theoretic criteria," *IEEE Trans. Signal Processing*, pp. 3378–3397, Dec 1998.

[39] T. Adalı, X. Liu, and M. K. Sönmez, "Conditional distribution learning with neural networks and its application to channel equalization," *IEEE Trans. Signal Processing*, vol. 45, no. 4, pp. 1051–1064, Apr. 1997.

[40] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Roy. Statist. Soc. Ser. B*, Vol. 39, pp. 1–38, 1977.

[41] R. A. Redner and N. M. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Rev.* , Vol. 26, pp.195–239, 1984.

[42] L. Perlovsky and M. McManus, "Maximum likelihood neural networks for sensor fusion and adaptive classification," *Neural Networks*, Vol. 4, pp. 89–102, 1991.

[43] L. Kullback, and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, pp. 79–86, 1951.

[44] L. Xu and M. I. Jordan, "On convergence properties of the EM algorithm for Gaussian mixture," Technical Report, MIT Artificial Intelligence Laboratory, Jan. 1995.

[45] 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, Jan. 1993.

[46] R.M. Neal and G.E. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants", *Learning in Graphical Models*,M. I. Jordan (editor), pp. 355–368, Dordrecht: Kluwer Academic Publishers, 1998.

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

[48] E. Weinstein, M. Feder, and A. V. Oppenheim, "Sequential algorithms for parameter estimation based on the Kullback-Leibler information measure," *IEEE Trans. Acou. Speech, and Signal Processing*, Vol. 38, No. 9, pp. 1652-1654, 1990.

[49] Y. Wang and T. Adalı, "Efficient learning of finite normal mixtures for image quantification," in *Proc. IEEE Intl. Conf. Acoust., Speech, and Signal Processing*, Atlanta, Georgia, pp. 3422–3425, 1996.

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

[51] Y. Wang, T. Adali, and B. Lo, "Automatic threshold selection by histogram quantization," *SPIE J. Biomedical Optics*, Vol. 2, No. 2, pp. 211–217, April 1997.

[52] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New York: Macmillan College Publishing Company, 1999.

[53] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlay, 1996.

[54] L. I. Perlovsky, "Cramer-Rao Bounds for the estimation of normal mixtures," *Pattern Recognition Letters*, Vol. 10, pp. 141–148, 1989.

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

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

[57] G. Schwarz, "Estimating the dimension of a model", *Ann. Statist.*, vol. 6, No. 2, pp. 461-464, 1978.

[58] Y. Wang, "Image Quantification and The Minimum Conditional Bias/Variance Criterion," *Proc. 30th Conf. Info. Sci. Sys.*, pp. 1061–1064, Princeton, March 20-22, 1996.

[59] J. Rissanen, "Minimax entropy estimation of models for vector processes," *System Identification*, pp. 97–119, 1987.

[60] J. H. Friedman, "On bias, variance, 0/1 - loss, and the curse-of-dimensionality," Technical Report, Stanford University, 1996.

[61] L. Perlovsky, W. Schoendorf, B. Burdick, and D. M. Tye, "Model-based neural network for target detection in SAR images," *IEEE Trans. Image Processing*, Vol. 6, No. 1, pp. 203–216, January 1997.

[62] H. Li, S. C. Lo, Y. Wang, W. Hayes, M. T. Freedman, and S. K. Mun, "Detection of masses on mammograms using advanced segmentation techniques and an HMOE classifier," *Digital Mammography*, Elsevier Science B.V. 1996.

[63] J. L. Marroquin, "Measure fields for function approximation," *IEEE Trans. Neural Nets.*, Vol. 6, No. 5, pp. 1081–1090, 1995.

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

[65] K. Woods, Y. Wang, Z. Szabo, and M. T. Freedman , "Unsupervised Tissue Quantification and Segmentation from 3-D MR Brain Images," *Proc. Intl. Conf. Signal and Image Processing*, Las Vagas, 1998.