USC–SIPI REPORT #157

Unsupervised Hierarchical Segmentation of Textured Images Based on Homogeneity Testing

by

Zhenyu Wu and Richard Leahy

June 1990

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Room 400
Los Angeles, CA 90089-2564 U.S.A.

Contents

List of Figures

List of Tables

Abstract

A new unsupervised hierarchical segmentation scheme is presented and is applied to the problem of tissue classification of magnetic resonance (MR) images of the human brain. The images under study are modeled as a mosaic of 'homogeneous' subimages where each subimage is modeled as a first order Gauss-Markov random field (GMRF) with unknown parameters. The segmentation goal is to group the pixels into regions which, under a suitable hypothesis, are homogeneous GMRFs. An analysis of homogeneity testing for GMRF is presented and two tests are proposed : the hierarchical likelihood ratio test and the dispersion test. Useful analytic results are derived for the case of an uncorrelated random field. Based on the homogeneity tests a hierarchical segmentation approach is suggested. The image is represented by a quadtree, and a split and merge procedure is applied to find large homogeneous regions within the image. This is followed by a segmentation refinement step involving connected component labeling and region growing, in which most of the unclassified pixels will be attached to the neighboring regions to which they are most similar. Finally, a constrained maximum likelihood agglomerative clustering procedure is used to reduce the number of segmented regions. The idea here is to allow the algorithm to learn about the image by first clustering the 'easy' pixels while delaying decisions on the 'hard' pixels until more information has been gathered. One difficulty encountered in implementing the algorithm is evaluating the likelihood for irregularly shaped regions. A new highly accurate approximation for the determinant of the covariance matrix based on eigenanalysis is proposed to overcome this problem.

Index Terms- Segmentation, random fields, hypothesis testing, magnetic resonance imaging, quadtree.

1 Introduction

In this paper an unsupervised hierarchical segmentation scheme for textured images is described, in which the image is modeled as a mosaic of regions where the pixel intensities in each region are characterized by a homogeneous GMRF with unknown parameters.

Markov random field (MRF) models have been widely studied as a mathematical model for images [1], [2], [3], [4]. In addition to textured image segmentation, they have also been used to solve related problems such as image restoration and smoothing [5], [6], [7], [8] and texture synthesis and classification [9], [10], [11]. Among the numerous papers on textured image segmentation in the literature, there are two general approaches which are closely related to this work: Bayesian and "maximum likelihood".

Within the Bayesian framework, a doubly stochastic process model is assumed: a prior Gibbs distribution for the region formation, and a second MRF for the texture of each region [12], [13]. The segmentation may be achieved by maximizing the aposteriori probability (MAP) or the posterior marginal probability (MPM) [14]. MAP solutions have been found using either iterative relaxation [7], [15], [16], [17], dynamic programming applied iteratively to strips of the image [12], [18], and stochastic relaxation through simulated annealing [5], [13], [19].

Another approach is hierarchical 'maximum likelihood' segmentation as proposed by Cohen and Cooper [15], [16]. In this case the image is modeled as a mosaic of GMRFs with known parameters. The image is partitioned into small square windows (*e.g.* 16×16) and each window is assumed to contain pixels of at most two classes. Each window is represented by a quadtree and segmented independently using a hierarchical approach. Starting from the root node, each block is divided into four sub-blocks and the best grouping of four is then obtained by maximizing their joint likelihood. This procedure is repeated recursively for each of the sub-blocks so that the image can be segmented from coarse to fine scale. Except at the root level, the likelihood conditioned on the resulting classification at a coarser level of the surrounding blocks should be used to achieve a good segmentation. Due to the restriction of at most two textures in each window, this hierarchical segmentation method may perform well only for images without fine details.

The techniques mentioned above are mostly supervised. However unsupervised extensions have been reported [19] or may be achieved by incorporating some unsupervised learning algorithms suggested in [20], [21].

This paper is focused on developing an unsupervised segmentation scheme based on the GMRF model. The unsupervised approach is often preferable to a supervised method in situations such as tissue classification for MR images, where the parameters for a given type of tissue may vary from patient to patient and from machine to machine, and may even vary within the same image depending on its relative location. The images are modeled

as a mosaic of 'homogeneous' subimages, each characterized by a first order GMRF with unknown parameters. The choice of a first order neighborhood was motivated by the need for computational tractability, however the variations within a region of an MR image, corresponding to a single tissue type, generally exhibit very local correlation rather than large scale textual characteristics and consequently this model should be appropriate in most cases. A preliminary version of the algorithm was reported in [21]. The proposed algorithm has some parallels to the hierarchical ML methods in [15], [16], namely that the images are modeled as a mosaic of 'homogeneous' GMRFs; the joint likelihood plays an important role in the segmentation; and the images are segmented in a hierarchical fashion.

The basis of the segmentation scheme is homogeneity testing for GMRF's. There are two cases to consider: labeled data and unlabeled data. Labeled data case refers to the situation where the data are already grouped into homogeneous subsets. One such example is the problem of deciding whether two homogeneous regions merge together to form a larger homogeneous region. The likelihood ratio test [22], [23] is chosen here as the homogeneity criterion. Similar tests have been used by other researchers for the same purpose [20]. Unfortunately when segmenting an image, the data are unlabeled. In other words, the data may be mixtures of samples from different populations, and a decision needs to be made about the occurrence of population mixtures. Chen and Pavlidis [24] have addressed this problem for uncorrelated data. Generalizations for correlated data have also been reported [25], [20]. Their testing strategies are all similar: divide the data into a few groups and test the homogeneity based on statistics computed from those data groups. It has been found in our analysis that these tests usually have a low power of detecting inhomogeneity, regardless of the choice of the homogeneity criterion, because the implicit assumption that these data groups were homogeneous themselves is often invalid. Here a statistical analysis of this issue is presented and two new and more powerful tests are proposed : the dispersion test and the hierarchical likelihood ratio test. For the special case of white GMRF's, the power function for these tests can be computed and comparisons are made.

The segmentation algorithm seeks to group the image pixels into regions which, under a suitable hypothesis, are homogeneous GMRFs. Following an analysis of homogeneity testing, a three step unsupervised hierarchical segmentation algorithm is proposed. At first a familiar split-and-merge procedure [24] is used to find large homogeneous regions. Its purpose is to allow the algorithm to learn the statistics of the different classes based on these large regions. A segmentation refinement step is then introduced through connected component finding and region growing. Finally a step-wise constrained maximum likelihood agglomerative clustering procedure is used to merge the resulting regions to a smaller number. In all three steps homogeneity testing is involved, for either labeled or

unlabeled data.

As in most of the segmentation techniques which use the MRF model, the problem of computing the partition function of the MRF for regions of irregular shape is encountered. For the particular case of GMRFs, the challenge lies in evaluating the determinant of a large matrix. Computation of the partition function is limited to very small irregular regions and a few regular shaped regions such as a rectangular lattice where analytic expressions for eigenvalues of the matrix are available [26], [27]. Some researchers introduce additional independence assumptions in their formulation so that computing the partition function becomes unnecessary [18], although those assumptions are often difficult to justify. Others use the so called pseudo-likelihood [18], [20] to approximate the true likelihood. The pseudo-likelihood is defined as the product of the conditional likelihoods for all pixels within the region [28], [29], which can be interpreted as an approximation to the likelihood of the residual errors resulting from GMRF by assuming that the residuals are independent. However, the likelihoods of the residuals and the data are numerically different, although they contain equivalent statistical information. Here a new highly accurate approximation is proposed to compute the partition function for a first order GMRF. The approximation is based on eigenanalysis of the inverse of the covariance matrix. Its evaluation is carried out as a summation over the pixel sites of the region, where each term depends only on the pixel site and its neighborhood. This form allows efficient computation when merging regions. The approximation is also used for GMRF parameter estimation.

The paper is organized as follows. The image model is briefly described in section 2, and the partition function approximation and related computing are presented in section 3. The next section analyzes the problem of homogeneity testing for GMRF and evaluates different homogeneity tests. The proposed segmentation scheme is developed in section 5 and examples of segmentation experiments are presented in section 6.

2 Image Modeling

Let $\mathbf{X}$ denotes a real valued image defined on the image domain Z, usually a rectangular lattice. Let Z be composed by K exhaustive and mutually exclusive subsets $\{Z_k\}_{k=1}^{K}$, and X_k, the subimage defined on Z_k, be modeled as a Gauss-Markov random field (GMRF). The image $\mathbf{X}$ can then be described by its joint probability density function (pdf) given by :

$$f(\mathbf{X}) = \prod_{k=1}^{K} f(X_k),$$

provided the X_k and X_j are independent for $k \neq j$. Based on this composite image model, the segmentation becomes a problem of fitting GMRF parameters and estimating the partition $\{Z_k\}$.

The GMRF model for X_k on Z_k is defined by the following joint probability density function [2] :

$$f(X_k) = \frac{|B_k|^{\frac{1}{2}}}{(2\pi\sigma_k^2)^{\frac{N_k}{2}}} e^{-\frac{1}{2\sigma_k^2}(X_k-\mu_k\cdot\underline{1})^T B_k (X_k-\mu_k\cdot\underline{1})} \tag{1}$$

where N_k = number of pixels on Z_k, $\underline{1}$ denotes an $N_k \times 1$ vector consisting of all 1's, and for a first order neighborhood

$$B_k(i,j) = \begin{cases} 1 & \text{if } j = i \\ -\beta_{1k} & \text{if } j \in N_H(i) \\ -\beta_{2k} & \text{if } j \in N_V(i) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $N_H(i)$ and $N_V(i)$ denote the sets of horizontal and vertical nearest neighbors of pixel i, respectively. Here β_{1k} and β_{2k} should be properly chosen so that B_k is positive definite.

The GMRF X_k defined in (1) is introduced by Besag [2] as an approximation of a stationary spatial process observed on a finite window Z_k. Some alternative approximations are based on the spectral analysis of the underlying stationary process [30], [31]. However these are not easily tractable models for image segmentation. As an approximation X_k itself is only asymptotically stationary. Therefore when X_k is said to be "homogeneous" it should only be interpreted as a random field defined by the homogeneous parametric model given in (1) and (2).

X_k, as defined in (1), has a free boundary. In other words, if X_k were viewed as a portion of an image on an infinite lattice, then $f(X_k)$ is identical to the conditional density given that all the pixels outside Z_k have been set to the mean, μ_k. Such a boundary assumption is preferable to others such as toroidal wrapping due to the potential shape irregularity of the image partition $\{Z_k\}$.

Here only a first order neighborhood and constant mean are considered. However we note that extensions to parameterized mean and higher order neighborhood are possible.

3 Likelihood Related Computations

For notational convenience the subscript k used in (1) and (2) will be dropped. Given the GMRF image model, the following operations are frequently performed when segmenting an image :

- Given an observation X on Z, compute the likelihood $f(X)$ for fixed parameter set $\theta = (\mu, \sigma^2, \beta_1, \beta_2)$.
- Find the maximum likelihood estimate (MLE) of θ given X.
- Efficiently update $f(X)$ and/or the MLE of θ when modifying Z.

Note that the evaluation of $|B|$, or equivalently $\ln|B|$, is involved in all of those operations. Unfortunately computing $\ln|B|$ is a costly operation, which becomes prohibitive when Z is a large irregularly shaped region. Evaluating $\ln|B|$ is equivalent to computing the partition function for Gibbs distributions [2], [4] for the case where the random field is Gaussian.

Exact evaluation of $\ln|B|$ is limited to small or certain regularly shaped regions. Ord [26] gives an analytic formula to compute $\ln|B|$ for a rectangular lattice with the free boundary assumption. Similar expression for the toroidal boundary condition is given by Besag and Moran [27]. Numerical approximations of the likelihood function for GMRFs have also been proposed in the literature, most within the context of parameter estimation. Whittle [30] computes $\ln|B|$ from the spectral density of the underlying stationary spatial process, which is observed only through a window Z. Due to the use of an asymptotic argument, only the number of pixels on Z is directly reflected in the approximation, which makes it insensitive to the shape of Z. Similar arguments are used in [31], [32]. Besag [2] also compares the parameter estimation results using his coding scheme against Whittle's method. His coding scheme is later generalized to the pseudo-likelihood function [28], [29] for more efficient parameter estimation. Although the pseudo-likelihood is not originally intended for likelihood approximation, it has been widely used for that purpose. However the pseudo-likelihood is at best an approximation of the likelihood of residual errors, which is numerically different from the likelihood of the data even though both contain equivalent statistical information. Ripley [4] also mentions a simple approximation of $\ln|B|$ using Taylor expansion when discussing the fact that the likelihood function may not be a concave function of the GMRF parameters.

3.1 Approximate evaluation of $\ln|B|$

Grouping all pixels on Z into "$*$" and "$+$" classes, as shown in figure 1, and using the property that the "$*$" pixels are independent of each other conditioned on the "$+$" pixels, and *vice versa* [2], B can be decomposed after reordering X as

$$B = \begin{bmatrix} I_* & B_{*+} \\ B_{+*} & I_+ \end{bmatrix} = \begin{bmatrix} I_* & B_{*+} \\ 0 & I_+ \end{bmatrix} \begin{bmatrix} I_* - B_{*+}B_{+*} & 0 \\ B_{*+} & I_+ \end{bmatrix}$$

Therefore defining $Q = B_{*+}B_{+*}$ one can show :

$$\ln|B| = \ln|I_* - Q| = \sum_{k=1}^{N_*} \ln(1-\lambda_k) \tag{3}$$

where $0 \leq \lambda_k < 1$ are the eigenvalues of Q. It is impractical to compute the $\{\lambda\}$ for a large region Z, unless Z is a regular shape such as a rectangular lattice and the $\{\lambda_k\}$ can be evaluated analytically [26]. The proposed approximations for $\ln|B|$ are based on the sum and the square sum of the $\{\lambda_k\}$, whose computation is provided by the following theorem.

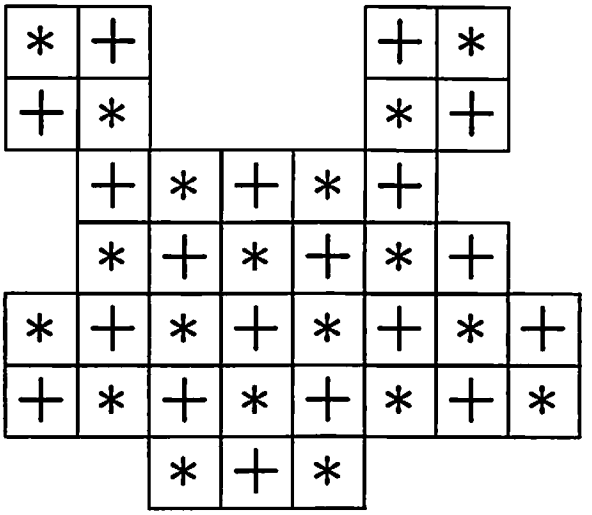


Figure 1: Indexing pattern for a first order GMRF on an irregular lattice

Theorem 1 *From the definitions of B and Q,*

$$\sum_{k=1}^{N_s} \lambda_k = tr(Q) = A_1\beta_1^2 + A_2\beta_2^2 \tag{4}$$

$$\sum_{k=1}^{N_s} \lambda_k^2 = tr(Q^2) = A_{11}\beta_1^4 + A_{22}\beta_2^4 + A_{12}\beta_1^2\beta_2^2 \tag{5}$$

Denote by $n_H(i)$, $n_V(i)$, $n_{H3}(i)$ and $n_{V3}(i)$ the number of first order horizontal, vertical and third order horizontal and vertical neighbors of site i in region Z, respectively; $N_2(i)$ the set of second order neighbors of i connected through Z and $n_C(i,j)$ the number of common (first order) neighbors between i and j. The constants in (4) and (5) are then computed as :

$$\begin{aligned}
A_1 &= \sum_{i\in Z_s} n_H(i) \\
A_2 &= \sum_{i\in Z_s} n_V(i) \\
A_{11} &= \sum_{i\in Z_s} \{n_H^2(i) + n_{H3}(i)\} \\
A_{22} &= \sum_{i\in Z_s} \{n_V^2(i) + n_{V3}(i)\} \\
A_{12} &= \sum_{i\in Z_s} \{2\cdot n_H(i)\cdot n_V(i) + [\sum_{j\in N_2(i)} n_C(i,j)]^2\}
\end{aligned} \tag{6}$$

Note that the constants (A_1, A_2, A_{11}, A_{22}, A_{12}) are shape descriptors of Z. The proof for the theorem is given in Appendix A. In particular when Z is square, which is often encountered in a quadtree based hierarchical segmentation, the theorem simplifies to:

Corollary 1 *If Z is a $n \times n$ square array, then*

$$\begin{aligned} A_1 &= A_2 = n(n-1) \\ A_{11} &= A_{22} = n(3n-5) \qquad (n>1) \\ A_{12} &= 12(n-1) \end{aligned}$$

The first approximation for $\ln|B|$ consists of finding a sequence $\{\hat{\lambda}_k\}$ "close" to $\{\lambda_k\}$ from which $\ln|B|$ is computed. Specifically, the following empirical model for the sequence $\{\hat{\lambda}_k\}$ was used :

$$\hat{\lambda}_k = m_\lambda + C_k \cdot S_\lambda \qquad k = 1, \cdots, N_* \tag{7}$$

where m_λ and S_λ^2 are the sample mean and variance of $\{\lambda_k\}$ computed from (4) and (5), and

$$C_k = C_1 + (k-1)^p D,$$

where the constants C_1 and D are determined using $\sum_{k=1}^{N_*} C_k = 0$ and $\sum_{k=1}^{N_*} C_k^2 = N_*$, so that $\{\hat{\lambda}_k\}$ will have the same sum and square sum as $\{\lambda_k\}$. In our segmentation experiments $p = 3$ was chosen. The sequence $\{\hat{\lambda}_k\}$ is based on a heuristic model, the advantage of this model is that $\{C_k\}$ does not depend on θ, hence the MLE of θ becomes simpler. In cases where any of the resulting $\hat{\lambda}_k$ fall outside the valid interval $[0,1)$, a reduced length sequence $\hat{\lambda}_k$ should be recomputed by assigning some of the $\hat{\lambda}$'s to 0.

Repeated use of the above approximation may still be too time consuming. In this case a much faster but less accurate approximation may be used instead. From the Taylor expansion of $\ln|B|$ as given below,

$$\ln|B| = -\sum_{n=1}^{\infty} [\frac{1}{n} \sum_{k=1}^{N_*} \lambda_k^n]$$

a simple approximation for $\ln|B|$ is proposed

$$\begin{aligned} \ln|B| &\approx -\sum_{k=1}^{N_*} \lambda_k - \frac{1}{2}\sum_{k=1}^{N_*} \lambda_k^2 - \sum_{n=3}^{\infty} [\frac{N_*}{n} \cdot \rho^n] \\ &= -\sum_{k=1}^{N_*} \lambda_k - \frac{1}{2}\sum_{k=1}^{N_*} \lambda_k^2 + N_* \cdot [\ln(1-\rho) + \rho + \frac{1}{2}\rho^2] \end{aligned} \tag{8}$$

where

$$\rho = \frac{1}{N_*}(\sum_{k=1}^{N_*} \lambda_k + \frac{1}{2}\sum_{k=1}^{N_*} \lambda_k^2).$$

Note that the approximation consists of replacing the $\{\lambda_k\}$ by a constant ρ from the third term on in the Taylor expansion of $\ln|B|$. The two approximations described above are compared with the exact value of $\ln|B|$ for several regions and combinations of β_1 and β_2 in Section 6.3.

3.2 Sufficient statistics of θ

In order to efficiently compute and update the likelihood $f(X)$ defined in (1), we keep track of a set of minimal sufficient statistics of θ. Expanding the exponent term of $f(X)$ gives

$$(X-\mu\cdot\underline{1})^T B(X-\mu\cdot\underline{1}) = X^TBX - 2\mu\cdot X^TB\underline{1} + \mu^2\cdot\underline{1}^TB\underline{1}.$$

Substituting B from (2) :

$$\begin{aligned} X^TBX &= \sum_{i\in Z}x_i^2 - (\sum_{i\in Z}\sum_{j\in N_H(i)}x_ix_j)\beta_1 - (\sum_{i\in Z}\sum_{j\in N_V(i)}x_ix_j)\beta_2 \\ X^TB\underline{1} &= \sum_{i\in Z}x_i - (\sum_{i\in Z}n_H(i)\cdot x_i)\beta_1 - (\sum_{i\in Z}n_V(i)\cdot x_i)\beta_2 \\ \underline{1}^TB\underline{1} &= N - (\sum_{i\in Z}n_H(i))\beta_1 - (\sum_{i\in Z}n_V(i))\beta_2 \\ &= N - 2A_1\beta_1 - 2A_2\beta_2 \end{aligned} \tag{9}$$

Hence the vector

$$(\sum_{i\in Z}x_i, \sum_{i\in Z}x_i^2, \sum_{i\in Z}\sum_{j\in N_H(i)}x_ix_j, \sum_{i\in Z}\sum_{j\in N_V(i)}x_ix_j, \sum_{i\in Z}n_H(i)\cdot x_i, \sum_{i\in Z}n_V(i)\cdot x_i)$$

is a sufficient statistics for θ. The first four terms are commonly expected statistics, the last two are closely related to the boundary shape of Z.

Observe that the sufficient statistics, as well as the shape descriptors defined earlier, are summations of local properties which depend on each individual pixel and its neighbors exclusively. Therefore their computation can be carried out in either recursive or parallel fashion, and consequently they are ideal for region split-and-merge or relaxation type of algorithms. This form of computing sufficient statistics should be more efficient than the hierarchical form proposed in [16].

3.3 MLE of parameter set θ

For any given β_1 and β_2, the MLE of μ and σ^2 are easily computed [4] :

$$\hat{\mu} = \frac{X^TB\underline{1}}{\underline{1}^TB\underline{1}} \tag{10}$$

$$\hat{\sigma}^2 = \frac{(X-\hat{\mu}\cdot\underline{1})^TB(X-\hat{\mu}\cdot\underline{1})}{N} \tag{11}$$

It can be shown that $\hat{\mu}$ and $\frac{N\hat{\sigma}^2}{\sigma^2}$ are independent random variables with distributions $\mathcal{N}(\mu, \frac{\sigma^2}{\underline{1}^TB\underline{1}})$ and χ^2 with $N-1$ degree of freedom, respectively. The above results can be easily extended when the subimage mean becomes a parametric function, *i.e.* $Ex(i) = \underline{\alpha}^T\cdot\underline{g}(i)$. Linear and quadric functions of lattice index are common examples for fitting subimage mean tendencies [20].

When β_1 and β_2 are unknown it is only practical to compute approximate MLEs of β_1 and β_2 for regions of irregular shape then compute $\hat{\mu}$ and $\hat{\sigma}^2$ according to (10) and (11). The coding method [2] and the pseudo-likelihood method [28], [29] are the popular estimation techniques in the literature. However these estimates often produce a non-positive definite matrix $\hat{B}$, where $\hat{B}$ is the resulting matrix B evaluated at the estimates of β_1 and β_2. Here the approximation of $\ln|B|$ described above is used in conjunction with the asymptotic domain constraint $|\beta_1| + |\beta_2| \leq 0.5$ [2] to find an approximate MLE. Without this constraint, the resulting $\hat{B}$ may still not be positive definite. In our experiment the barrier method [33] was used to solve this constrained optimization problem.

We have also observed that the likelihood surface defined on (β_1, β_2) is quite smooth. Consequently we may sample the (β_1, β_2) domain and perform the maximization over the resulting finite set of (β_1, β_2). Since the likelihood function may be evaluated from sufficient statistics and the shape descriptors, once they are computed it requires little extra effort to evaluate the likelihood functions associated with each (β_1, β_2) of the sample set. This later approach was preferred and implemented in the segmentation algorithm.

4 Testing Homogeneity for GMRFs

Let X be a subimage defined on an arbitrary region consisting of N pixels. In this section we develop tests for whether X is a realization of a homogeneous GMRF (the null hypothesis). Under the alternative hypothesis H_1, X is a mosaic of homogeneous GMRFs, *i.e.* $X = (X_1, X_2, \cdots, X_K)^T$ where K is the number of GMRFs in X and X_i denotes the N_i samples from the i-th GMRF parameterized by $\theta_i = (\mu_i, \sigma_i^2, \beta_{1i}, \beta_{2i})$. The null hypothesis H_0 may be more conveniently stated as $\theta_i = \theta$, $\forall i$ for X partitioned into K groups. Here only tests against the class of equal variance alternatives, *i.e.* $\sigma_i^2 = \sigma^2$, $\forall i$, are studied. This class excludes images in which the variance is the only distinguishable feature.

4.1 Homogeneity tests for labeled data

When the data are labeled, *i.e.* the partition $\{X_i\}$ of region X is given, testing homogeneity falls within the conventional framework of statistical hypotheses testing. The tests for labeled data described here will be used in the segmentation algorithm to decide whether two homogeneous regions should be merged together to form a larger one. For the limited class of alternative hypotheses stated above, the likelihood ratio test is used.

Let $\hat{\mu}$, $\hat{\sigma}^2$ and $\hat{B}$ denote the MLE of μ, σ^2 and B under H_0, and μ_i^*, σ_i^{*2} and B_i^* denote the MLE of μ_i, σ^2 and B_i under H_1 for the i-th subimage. The MLE of B refers to the matrix constructed using the MLE of β_1 and β_2. Then the likelihood ratio test [22], [20]

for rejecting H_0 can be simplified to :

$$\tilde{F}_K = \frac{N-K}{K-1} \cdot \frac{|\hat{B}|^{-\frac{1}{N}} \cdot \hat{\sigma}^2 - |B^*|^{-\frac{1}{N}} \cdot \sigma^{*2}}{|B^*|^{-\frac{1}{N}} \cdot \sigma^{*2}} > C \tag{12}$$

where K is the number of classes, N is the number of pixels in X, $B^* = \mathrm{diag}(B_1^*, B_2^*, \cdots, B_K^*)$, and

$$\sigma^{*2} = \frac{1}{N}\sum_{i=1}^{K} N_i \sigma_i^{*2} = \frac{1}{N}\sum_{i=1}^{K} (X_i - \mu_i^* \cdot \underline{1})^T B_i^* (X_i - \mu_i^* \cdot \underline{1})$$

In the case where the common variance σ^2 is known, the corresponding likelihood ratio test for rejecting the null hypothesis H_0 is

$$\tilde{W}_K = \frac{1}{\sigma^2}(N\hat{\sigma}^2 - \sum_{i=1}^{K} N_i \sigma_i^{*2}) + \ln|B^*| - \ln|\hat{B}| > C \tag{13}$$

In order to use the tests $\tilde{F}_K$ and $\tilde{W}_K$ defined in (12) and (13), we must set the threshold C for any given significance test level. Unfortunately both $\tilde{F}_K$ and $\tilde{W}_K$ are usually complicated functions of the data, hence approximations are often needed for setting the threshold C.

Let us rewrite $\hat{\sigma}^2$ as :

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{N}\sum_{i=1}^{K} (X_i - \hat{\mu} \cdot \underline{1})^T \hat{B}_i (X_i - \hat{\mu} \cdot \underline{1}) \\ &= \frac{1}{N}\sum_{i=1}^{K} [(X_i - \hat{\mu}_i \cdot \underline{1})^T \hat{B}_i (X_i - \hat{\mu}_i \cdot \underline{1}) + (\underline{1}^T \hat{B}_i \underline{1}) \cdot (\hat{\mu} - \hat{\mu}_i)^2] \end{aligned}$$

where $\hat{B}_i$ is the i-th diagonal block of $\hat{B}$, and

$$\hat{\mu}_i = \frac{X_i^T \hat{B}_i \underline{1}}{\underline{1}^T \hat{B}_i \underline{1}}.$$

Due to the consistency property of the MLE of β_1 and β_2 [26], if all subimages $\{X_i\}$ are sufficiently large, estimates of the β's, based on X or individual X_i, should be approximately equal under the null hypothesis H_0, *i.e.* $\hat{B} \approx B^*$. Hence under H_0,

$$\tilde{F}_K \approx \frac{N-K}{K-1} \cdot \frac{\sum_{i=1}^{K} (\underline{1}^T B_i^* \underline{1}) \cdot (\mu^* - \mu_i^*)^2}{\sum_{i=1}^{K} (X_i - \mu_i^* \cdot \underline{1})^T B_i^* (X_i - \mu_i^* \cdot \underline{1})}.$$

Note that the right hand side is approximately an $\mathcal{F}_{K-1,N-K}$ distributed random variable [22], because its numerator and denominator become asymptotically independent. Similarly,

$$\tilde{W}_K \approx \frac{1}{\sigma^2}\sum_{i=1}^{K} [(\underline{1}^T \hat{B}_i \underline{1}) \cdot (\hat{\mu} - \hat{\mu}_i)^2]$$

which is approximately χ^2_{K-1} distributed.

Another possible approximation for setting the threshold C is to use the property that the log of the likelihood ratio is asymptotically proportional to a χ^2, under the null hypothesis H_0 [22], [20].

Due to the complexity of the tests, evaluating their performance is very difficult. However there is a special sub-class of GMRFs for which the power functions for both tests can be computed. Assume all K GMRF's are white, *i.e.* $\beta_{1i} = \beta_{2i} = 0 \ \forall \ i$, then the homogeneity testing as formulated becomes the well studied problem of mean comparison for normal populations [22], [23]. In terms of image processing, our image model becomes uniform regions (constant mean for each region) with Gaussian additive noise. It can be shown that $\tilde{F}_K$ and $\tilde{W}_K$ are non-central $\mathcal{F}_{K-1,N-K}$ and non-central χ^2_{K-1} distributed, respectively. For both tests the non-centrality parameter vanishes under the null hypothesis H_0. Note that the approximations used earlier for $\tilde{F}_K$ and $\tilde{W}_K$ now become exact. For the white random field case :

$$\tilde{W}_K = \frac{1}{\sigma^2} \sum_{i=1}^{K} N_i \cdot (\bar{X} - \bar{X}_i)^2. \tag{14}$$

Here $\hat{\mu}$ and $\hat{\mu}_i$ are replaced by the sample means $\bar{X}$ and $\bar{X}_i$ respectively.

Given the probability distributions, the power of $\tilde{F}_K$ and $\tilde{W}_K$ can be computed for any fixed K and $\{N_i\}$. The power function of a test is defined as the probability of detecting inhomogeneity at a given significance level. In figure 2 the power function of $\tilde{W}_2$ (curve a) is plotted for $N_1 = 4$ and $N_2 = 12$, against the normalized mean difference ($= \frac{\mu_1 - \mu_2}{\sigma}$). Other curves in fig. 2 are described in the next subsection. The power function of $\tilde{F}_2$ was also computed but not plotted since the difference between the two are negligible. This suggests that the power of $\tilde{W}$ has a low sensitivity towards the true value of variance, even for a relatively small sample size. In this two population case both tests, $\tilde{F}_2$ and $\tilde{W}_2$, are uniformly most powerful (UMP) for their respective hypotheses [22], [23].

4.2 Homogeneity tests for unlabeled data

The tests described in the last section are valid only for labeled data X. However X is often unlabeled, which makes homogeneity testing more difficult. Several tests will be proposed and their performance will be analyzed here. In order to make the analysis simpler, we assume that the common variance σ^2 is known throughout this section.

Homogeneity tests for white GMRF

We begin the analysis focused on the limited sub-class of white GMRF for two reasons. Firstly homogeneity testing for white GMRFs is important in its own right. It would provide a solution to the problem of resolving population mixtures for a given data sample. Applications in image processing include the restoration of uniform images corrupted by

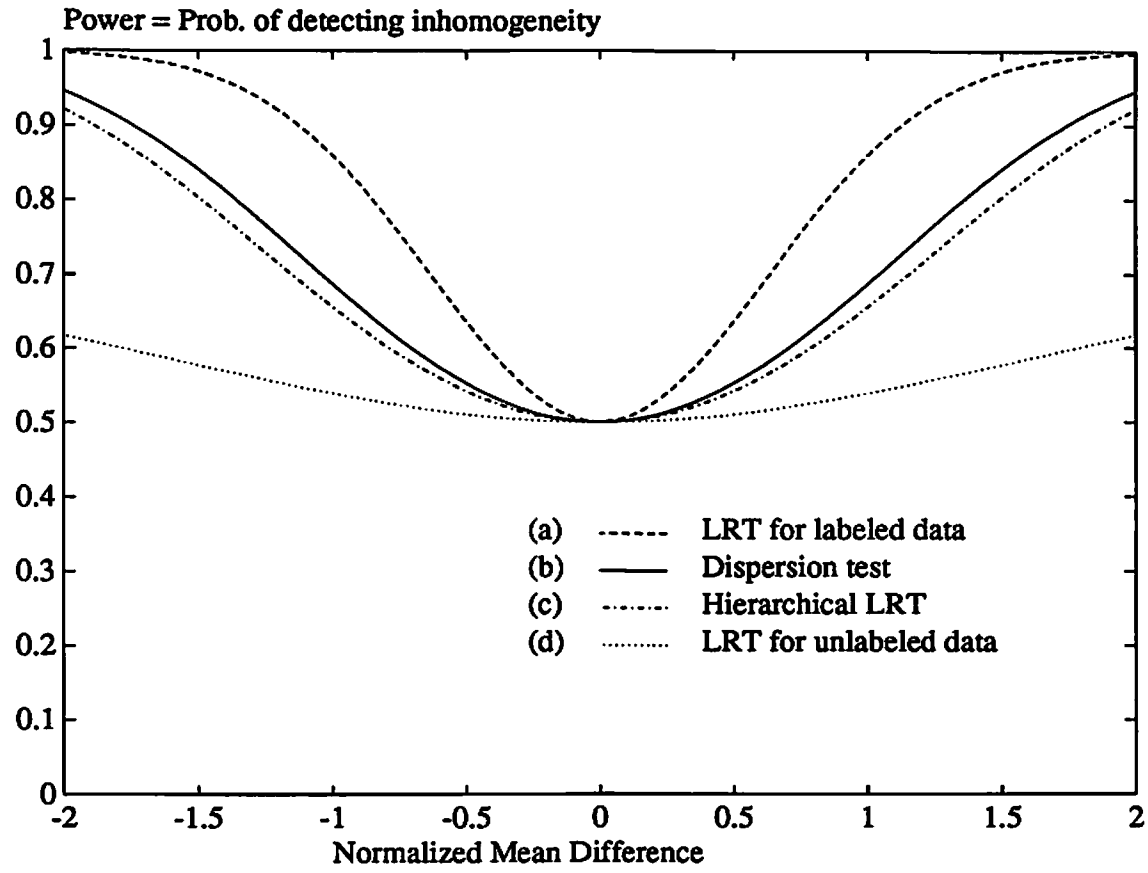


Figure 2: Example of power functions for (a) likelihood ratio test (LRT) for labeled data; (b) dispersion test; (c) hierarchical LRT; (d) LRT applied to unlabeled data with arbitrary partition

additive noise. Examples of this application are presented in Section 6. Also useful analytic results for homogeneity testing can be derived for white GMRF, which may prove useful for solving the more general problem.

One common solution for testing homogeneity of unlabeled data is the following [24], [20] : divide X into K' subimages $\{X_i'\}$, usually of equal size, then test the homogeneity of X based on the statistics of $\{X_i'\}$. Examples of homogeneity criteria include the $\tilde{W}_{K'}$ [20] and the maximum difference in the sample means of X and $\{X_i'\}$ [24]. This approach is simple but its power is usually poor. Note that there is always an implicit assumption that the $\{X_i'\}$ are homogeneous themselves, regardless of the choice of homogeneity criterion; such an assumption is often invalid.

Consider the case where $\tilde{W}_{K'}$ is used. For an arbitrary partition $\{X_i'\}$, $\tilde{W}_{K'}$ is a non-central $\chi^2_{K'-1}$ random variable, as for the case of labeled data, but often with a much reduced non-centrality parameter. The overall power of $\tilde{W}_{K'}$ is then computed as the average over all possible partitions $\{X_i'\}$ according to a prior distribution of $\{X_i'\}$. For

the same two class example ($N_1 = 4, N_2 = 12$) used in the previous subsection, the power is computed as follows : the 16 samples are arbitrarily split into two subsets of 8, and the least informative prior (binomial in this case) is used to compute the overall power of the $\tilde{W}_2$ test. As shown in figure 2 (curve d), the resulting power is disappointing.

One remedy is to make the subimages $\{X_i'\}$ smaller, so they have a higher probability of being homogeneous. Although more subsets must be tested, which contributes some power loss, the overall power is still improved. In the limit every subset has only one pixel, then the $\tilde{W}$ test becomes proportional to the sample variance of X. This special case is named here the dispersion test : reject H_0 if

$$\tilde{D}_N = \tilde{W}_N = \frac{N\hat{\sigma}^2}{\sigma^2} > C. \tag{15}$$

Despite its simple appearance, the $\tilde{D}_N$ test has a surprisingly good power. Figure 2 shows the power function of $\tilde{D}_N$ (curve b) when applied to the same two class problem.

A more elaborate approach is also proposed : apply the dispersion test if the size of X is small; otherwise, divide X into K' groups and apply recursively the same test to all groups, then apply the $\tilde{W}_{K'}$ test defined in (14) only if all groups test homogeneous. This procedure is named here the hierarchical likelihood ratio test. This test has a clear purpose : to avoid testing homogeneity among subimages unless subimages themselves have a high probability of being homogeneous. The test can be naturally implemented within a hierarchical data structure, in our case a quadtree.

In order to claim the proposed hierarchical test as a genuine hypothesis test, it is necessary to know how to set the involved thresholds so that the overall test has the desired significance level. The following theorem provides the solution to this problem.

Theorem 2 *Under either H_0 or H_1 all tests involved in the hierarchical likelihood ratio test scheme are statistically independent.*

This theorem is an extension to the well known fact that the sample mean and the sample variance are independent when data are drawn from a normal population [22]. The proof is also similar and is provided in Appendix B. Using this theorem, the overall significance level of the hierarchical test can be computed from the significance levels of the individual tests involved in the testing scheme. It is, however, still an open question as to how to distribute the overall significance level within the hierarchical test so that the resulting test power is optimized. An empirical method is used here to assign the individual significance levels.

The power of a two-level hierarchical test, applied to the same two class problem, is computed : the 16 samples are split equally into 2 groups, an equal significance level is assigned to all three individual tests and the binomial prior is used to compute the overall power. The resulting power is plotted in figure 2 (curve c). One can observe that the

hierarchical test is more powerful than the $\tilde{W}$ test applied to unlabeled data, but less powerful than the dispersion test. This result could be considered encouraging, since the power is computed for the least informative prior. In other words this computed power is the lower bound for the power function of the hierarchical test for any prior distribution [34]. In contrast, the dispersion test makes no use of a prior. Another advantage of using the hierarchical test is the possibility of detecting homogeneous portions even if the overall test fails.

Homogeneity tests for general unlabeled GMRF

When local interactions among neighboring pixels are also considered, the homogeneity issue becomes more complex. In addition to analytic complexity, there are important model defects of which we need to be aware. Consider two contiguous subimages, X_1 and X_2. According to our mosaic image model, they are uncorrelated when H_0 is rejected, but they become correlated if H_0 should be accepted. One way to avoid dealing with such an undesirable discontinuity is not to use inter-region statistics when testing homogeneity, as in the case of the hierarchical likelihood ratio test. Another defect is caused by the use of the GMRF model itself as an approximation to a stationary spatial process. If observations are added or removed the resulting model is not compatible with the original model in the sense that the same pixels will have different characterization within the two models.

Based on the result obtained for white GMRFs (Fig. 2) the use of tests whose decision are based on statistics computed from arbitrarily formed subregions should be avoided due to their low power of detecting inhomogeneity. Therefore only the dispersion test and the hierarchical likelihood ratio test will be used for testing homogeneity.

Due to the difficulty of obtaining analytical results, these tests are only presented as generalizations of their counterparts in the white GMRF case. The dispersion test is defined as a scaled MLE of the variance σ^2, *i.e.* reject H_0 if

$$\tilde{D}_N = \frac{N\hat{\sigma}^2}{\sigma^2} > C. \tag{16}$$

Here $\hat{\sigma}^2$ is computed according to (11). Under the null hypothesis H_0, $\tilde{D}$ is approximately χ^2. Consequently C can be approximately computed for a given significance test level.

The hierarchical likelihood ratio test is defined in a similar manner as for the white GMRF case, except the test $\tilde{W}$ in (13) should be used in place of (14. For the general case, the thresholds for the tests involved in the hierarchical scheme can only be set approximately for a given significance level.

5 Hierarchical Segmentation Algorithm

The hierarchical segmentation algorithm can be divided into three major components : split-and-merge learning, patch forming, and constrained ML agglomerative patch clustering.

5.1 Split-and-merge learning

The main goal of the split-and-merge procedure is to allow the algorithm to learn and to be trained through hierarchically finding large homogeneous regions within the image. This is the most crucial part of our unsupervised segmentation scheme.

Let the image be represented by a quadtree with the top of the tree corresponding to the entire image, the next level to the four quadrants of the image and so on [35]. Starting from the top of the tree, we test homogeneity for the region corresponding to each node using either the dispersion test or the hierarchical test. The node is then labeled accordingly: if the region is homogeneous a label is attached together with the estimated GMRF parameters to that node and the tree is terminated at this node; if the region is inhomogeneous the node remains unlabeled and the same node labeling strategy is applied to its four quadrants. Furthermore, each newly labeled region is tested for possible merging with previously formed homogeneous regions. This allows larger homogeneous regions to be formed gradually, so that the GMRF parameter estimates become more and more reliable. Here the likelihood ratio test $\tilde{W}_2$ should be used. As the node size becomes small, the parameters estimated from large homogeneous regions can provide supervision for deciding whether to merge the node to bigger regions. In this way the algorithm can be 'trained' by these early large homogeneous regions for later classifying the remaining 'hard' (small) regions. This split-and-merge mechanism is illustrated in figure 3. In completion, the split-and-merge step produces an image partition.

When testing homogeneity two kinds of errors may occur : a region is labeled inhomogeneous while H_0 is true (type I error) or a region is labeled homogeneous while H_0 is false (type II error). The type II error is related to how accurately the algorithm can be trained. Once committed, this error is unlikely to be corrected. On the other hand, the type I error controls the rate of training for the algorithm and is often self-correctable. In addition a step-wise ML agglomerative clustering can also be applied to reduce the type I error after the split-and-merge. A large number of type I errors may however inhibit the early formation of large homogeneous regions, resulting in a very noisy segmentation. Therefore a balance between the two types of error is needed. A general rule is to keep the probability of type II error small while testing large regions, then gradually increase this probability as the region size becomes small.

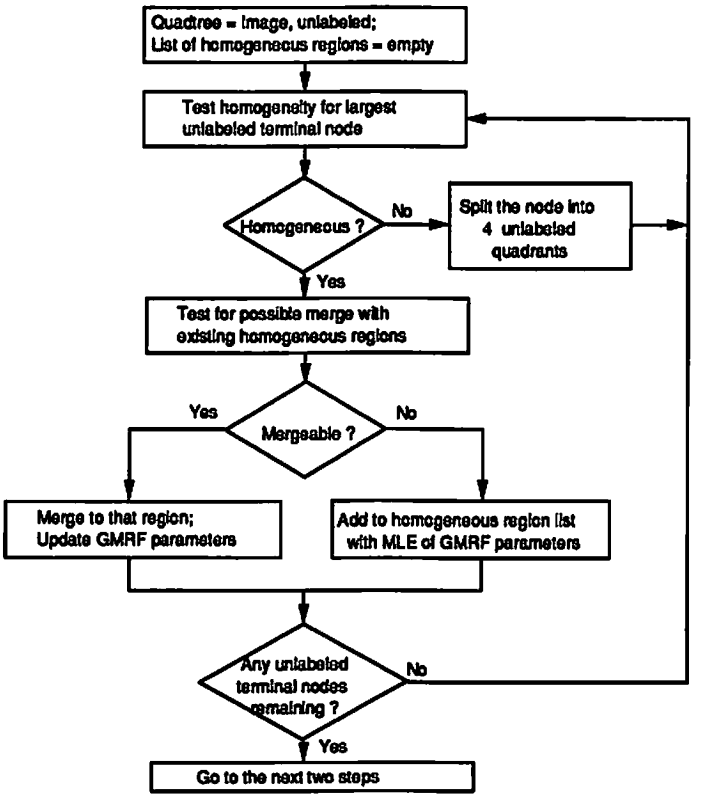


Figure 3: Block diagram for split-and-merge learning

5.2 Patch forming

The resulting segmentation from the split-and-merge is often noisy. This is not surprising since no smoothness constraint is imposed. Each segmented region (set of pixels assigned to the same label) may have many isolated subregions, most of which may contain very few pixels. Hence it is natural to think of breaking down each segmented region into connected subregions (patches) in order to get a smoother segmentation.

The patch forming algorithm consists of the following steps :

1. Patch labeling: For each resulting region from the split-and-merge, find all its spatially connected components (patches), and label them accordingly : a different label is assigned to each patch except those with very few pixels (≤ 4 pixels in our experiment) which are left unlabeled.

2. Patch growing: Among all pairs of unlabeled patches and their neighboring labeled patches, select the one with the smallest mean difference and expand the labeled patch by attaching the unlabeled one to it. Repeat this step until the smallest mean difference exceeds a given threshold or no unlabeled patch is left.

In the above algorithm, the patch labeling is carried out using a modified version of the algorithm in [36].

5.3 Constrained ML agglomerative patch clustering

The number of patches resulting from the previous step is usually too large. A constrained maximum likelihood agglomerative patch clustering step is therefore used to reduce the number of labels. The patch clustering is performed as follows :

> Among all pairs of labeled neighboring patches, select the one which when merged produces the smallest reduction in their joint likelihood while satisfying certain regularization constraints. A new patch is then formed by merging these two patches. The procedure is repeated until a stopping condition is met.

This clustering is a generalization of the clustering method described by Duda and Hart [37], in the sense that regularization constraints are imposed and the clustering is performed on sets of correlated pixels (patches).

When implemented the following specifications are used :

1. In addition to connectivity only one other regularizing constraint is used: the maximum gradient on the boundary between two candidate patches should not exceed a given threshold. Here the gradient is computed using only pixels that lie in the union of the candidate patches.

2. The stopping condition for clustering is that no pair of patches can be found which give a smaller likelihood reduction than a pre-set threshold while meeting the regularizing constraints.

3. A common variance is used for computing the likelihood for all patches. This variance is estimated as a weighted sum of ML variance estimates of the labeled patches.

When merging patches, the criterion of maximum likelihood seeks to put together patches with similar GMRF features, while the connectivity and edge constraints tend to limit the extension of resulting patches and to preserve the local contrast displayed between neighboring patches. The patch clustering is only step-wise optimal. After patch clustering, most of the resulting patches are large and distinct from their surrounding patches. Therefore when applied to the problem of tissue classification of MR images, the final tissue labeling of the patches can be easily performed through an interactive patch editing.

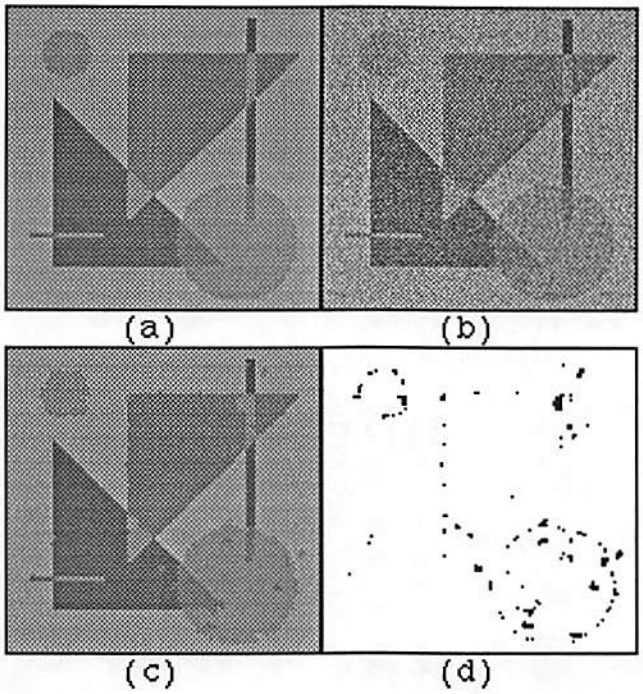


Figure 4: Segmentation of 4 region hand-drawn image : (a) noise-free image, (b) noisy image (SNR = 2), (c) segmentation, (d) error (1.3%)

6 Experimental Results

The segmentation scheme described above has been applied to the restoration of images of uniform regions corrupted by Gaussian white noise, and tissue classification of MR images of the human brain.

6.1 Restoration of uniform image with additive noise

Although the problem solved in this subsection is somewhat artificial, the experiment can help to evaluate the performance of our segmentation scheme. Figure 4.a shows a 4 region computer generated image of 128×128 pixels with 256 grey levels. This image is adapted from the example of [19], but it is not an exact replica due to the lack of specification. The intensities for the 4 uniform regions are 80, 112, 144 and 176, respectively. The image was then corrupted by Gaussian additive noise corresponding to a signal-to-noise ratio SNR = 2 (Fig. 4.b). The SNR is defined as the ratio between the minimum intensity difference between the mean of these uniform regions and the standard deviation of the noise. The pixel-wise maximum likelihood classifier (thresholds at 96, 128, 160) gives a 21.1% error percentage. In contrast, if the same ML classifier is applied to the patches resulting from the unsupervised hierarchical segmentation, the error drops to only 1.3% (Fig. 4.d). A second case is shown in Fig. 5. Fig. 5.b shows the corrupted image with

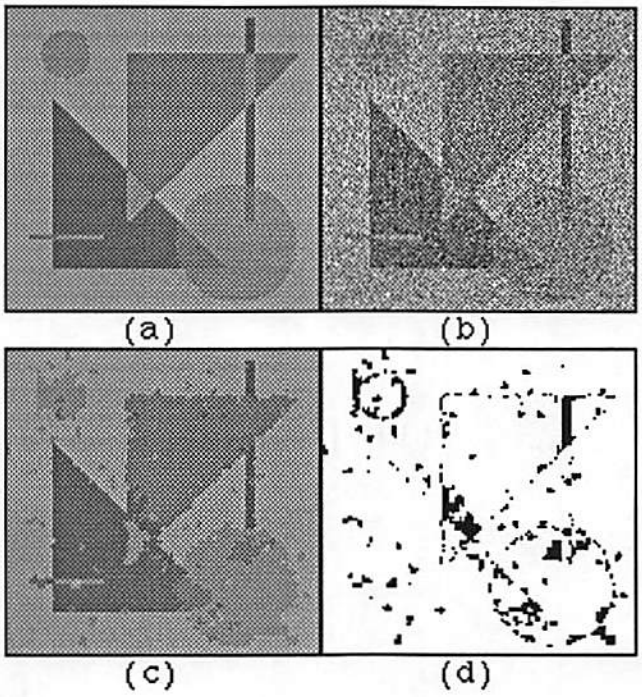


Figure 5: Segmentation of 4 region hand-drawn image : (a) noise-free image, (b) noisy image (SNR = 1), (c) segmentation, (d) error (7.3%)

SNR = 1. The segmentation error (Fig. 5.d) is 7.3% compared to an error of 40.3% using a ML classifier.

6.2 MR image segmentation

MR imaging is a powerful non-invasive diagnostic tool which produces anatomical images of the human body based on the nuclear magnetic resonance properties of hydrogen. One undesirable characteristic of MR images is that the parameters for a given type of tissue may change from patient to patient, from scanner to scanner and with variations with the imaging protocol. In addition they may also vary as a function of the tissue's location within the same image. The last is particularly disturbing, since this variation may results in undistinguishable features (GMRF parameters) for two different types of tissue if they are far apart from each other. This constitutes a strong motivation for segmenting the image into connected patches and using regularizing constraints to preserve local contrast between neighboring patches. This feature ambiguity problem is experienced in our example shown in figure 6.

Figure 6 shows a typical MR image of the human brain for a patient with a large lesion on the left side of the brain. The image size is 256 × 256 quantized to 12 bits per pixel. Before segmentation, the skull portion of the image was removed from the image by finding the contour of the boundary between the skull and the enclosed soft tissues. In order to

illustrate how the segmentation method works, we will discuss in some detail the partial results produced at each step of the algorithm when applied to the image in figure 6. In the example all GMRF parameters are estimated from the image and no prior information of region partition is assumed.

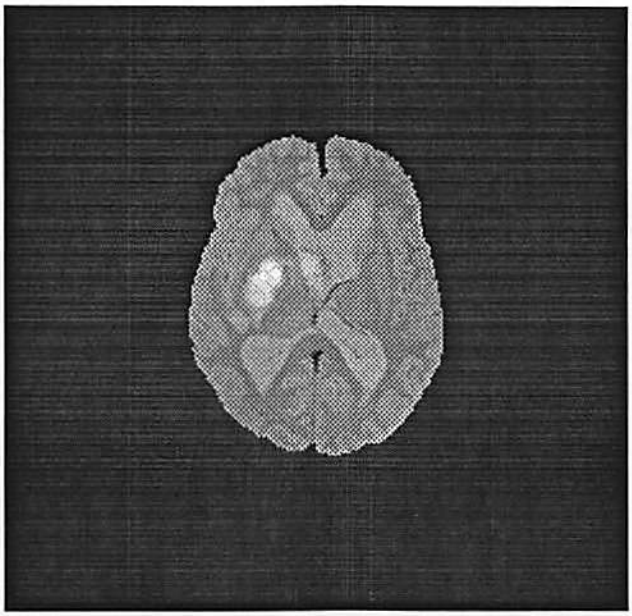


Figure 6: A cross-section MR image of the brain for a patient with brain tumor

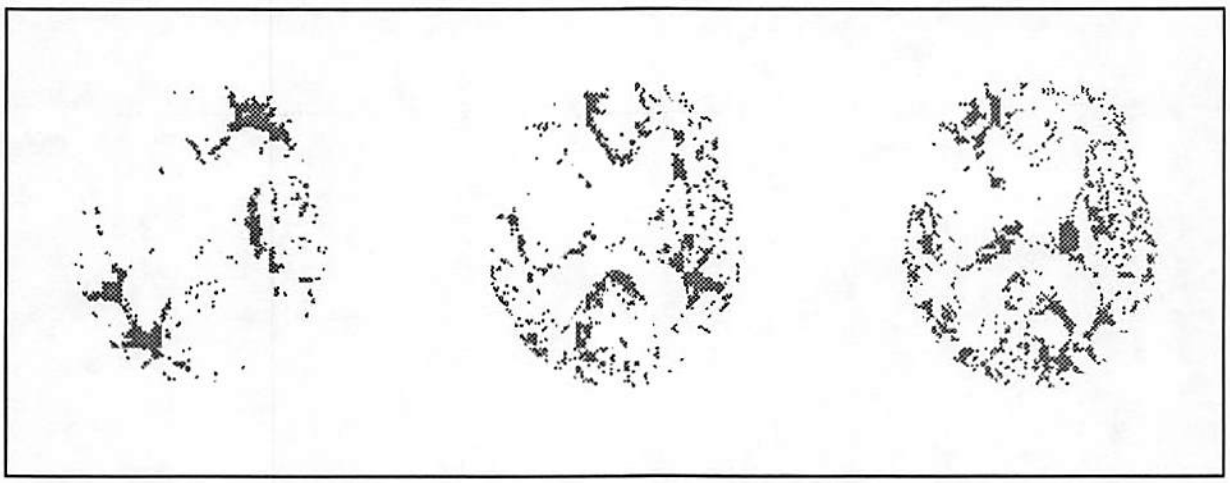
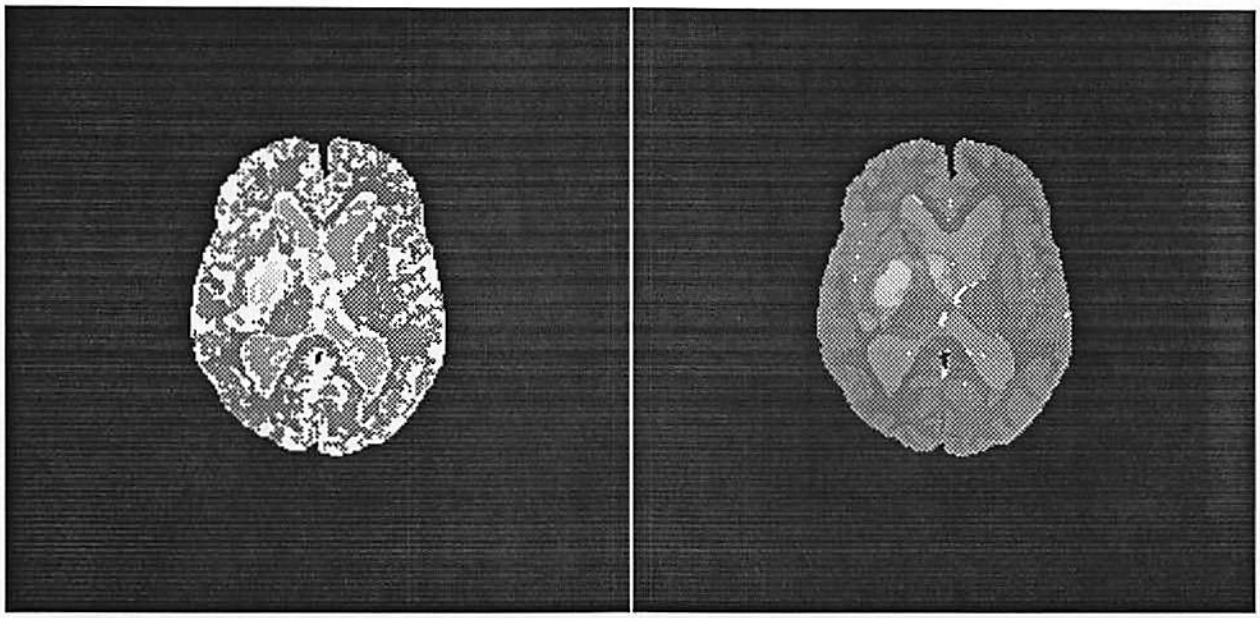


Figure 7: Partial results of split-and-merge

Results of the split-and-merge : The split-and-merge algorithm produces 16 regions. Figure 7 shows three of these regions whose pixels correspond mostly to the white matter in the brain. Note that they are rather noisy. However our main goal at this point, as



(a) (b)

Figure 8: Resulting patches before and after patch growing

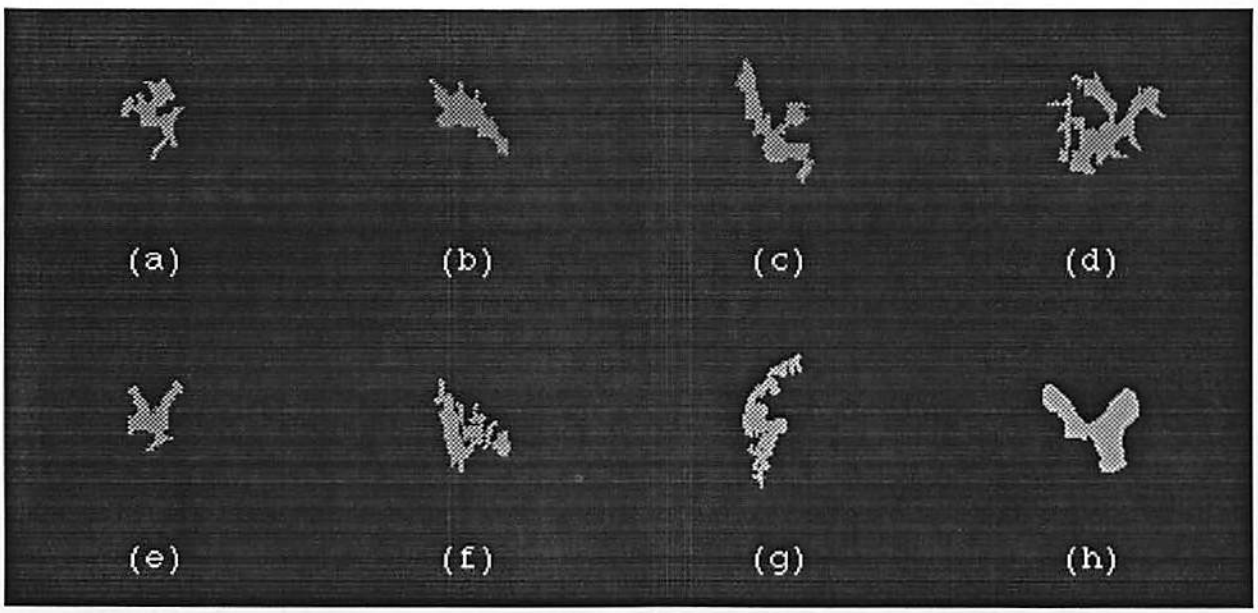
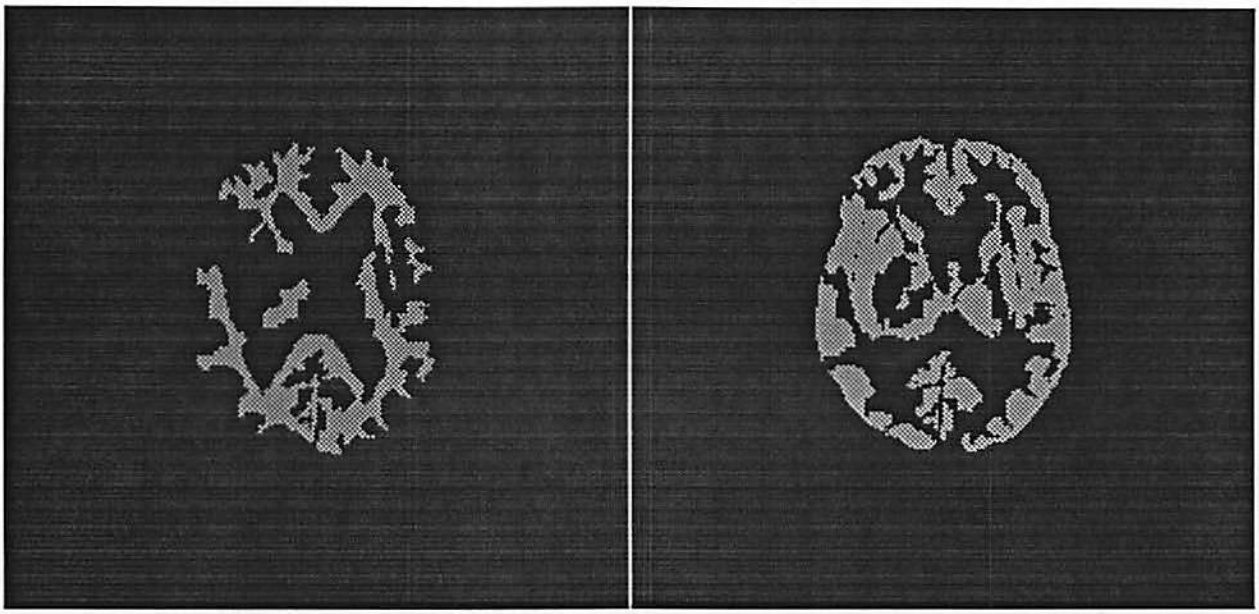
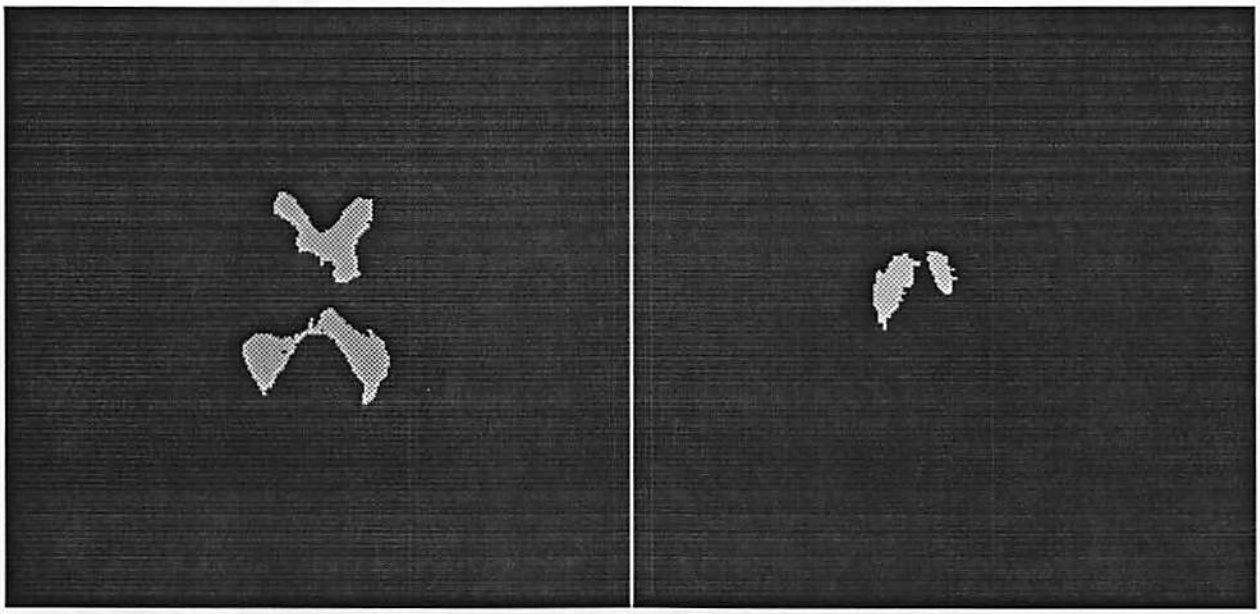


Figure 9: Examples of resulting patches after clustering



(a) White matter (b) Grey matter



(c) Ventricles (d) Tumor

Figure 10: Resulting segmented regions

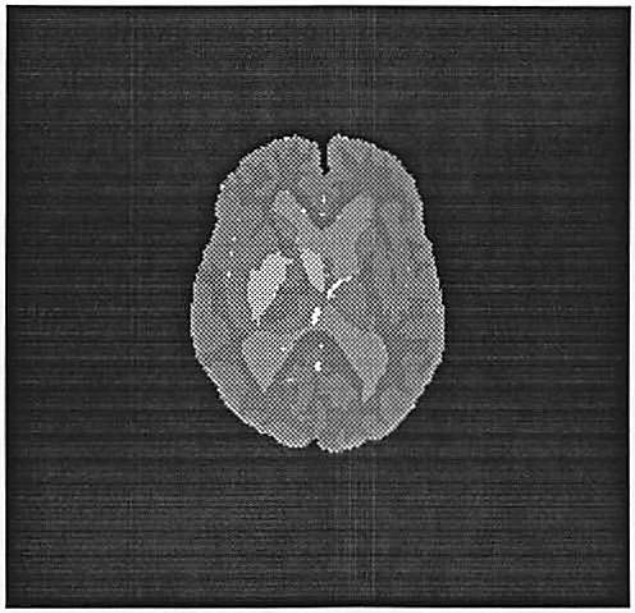


Figure 11: Combined segmentation map

stated before, is not to produce a smooth segmentation but to find large homogeneous regions.

Results of patch forming : The connected component labeling algorithm is applied to each of the regions obtained from the split-and-merge. Figure 8.a shows all labeled patches displayed with their mean grey level while the unlabeled patched are shown in white. Note that the unlabeled patches fall mainly into the boundaries between different tissues. The unlabeled patches are then assigned a label through patch growing, the resulting patch mean image is shown in figure 8.b. Note that the main features of the original image are preserved.

Results of patch clustering : Figure 9 shows a few resulting patches after constrained ML patch clustering. In figures 9.a-e the patches belong to white matter; in figure 9.f-g, grey matter; and in figure 9.h, ventricle. In this case we further classified the patches into one of four tissue types. Figures 10.a-d shows the resulting segmented regions corresponding respectively to (a) white matter; (b) grey matter; (c) ventricle; and (d) tumor. In each figure only pixels classified to that label are displayed while the rest are set to black. The combined segmentation map is shown in figure 11.

6.3 Accuracy study for the proposed approximations for $\ln|B|$

In Table 1 we show the values of the partition function (defined as $\frac{1}{2}\ln|B|$) for the eight patches displayed in fig. 9 for various (β_1, β_2). For each (β_1, β_2) the values in the three columns are computed respectively by the direct method, the accurate approximation

through $\{\hat{\lambda}_k\}$ and the fast approximation through Taylor expansion. For $\beta_1 = \beta_2 = 0.25$, which is near the boundary of the valid domain for $\ln|B|$, the error for the first approximation is small while the error for the second becomes large. As the (β_1, β_2) moves further from the boundary, the error of the second approximation reduces rapidly $(\beta_1 = 0.15, \beta_2 = 0.3)$ and eventually becomes negligible $(\beta_1 = 0.1, \beta_2 = 0.15)$.

	Size	$\beta_1 = 0.25, \beta_2 = 0.25$			$\beta_1 = 0.15, \beta_2 = 0.30$			$\beta_1 = 0.10, \beta_2 = 0.15$		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
a	298	-21.849	-21.550	-19.127	-17.754	-17.735	-16.625	-4.207	-4.213	-4.191
b	615	-52.601	-50.940	-43.606	-40.839	-40.704	-37.629	-9.440	-9.439	-9.384
c	559	-46.492	-45.448	-39.157	-36.707	-36.656	-33.914	-8.496	-8.502	-8.453
d	489	-37.034	-36.346	-32.053	-29.325	-29.216	-27.310	-6.951	-6.951	-6.915
e	569	-46.217	-45.395	-39.289	-36.485	-36.430	-33.802	-8.513	-8.514	-8.466
f	477	-35.987	-35.740	-31.410	-28.935	-29.059	-27.124	-6.803	-6.864	-6.828
g	414	-31.524	-31.384	-27.440	-25.679	-25.858	-24.085	-6.016	-6.045	-6.013
h	602	-55.461	-54.815	-45.653	-42.456	-42.904	-39.298	-9.698	-9.790	-9.728

Table 1: Values of partition function for patches shown in fig. 6 calculated using : (1) direct method, (2) accurate approximation, (3) fast approximation

.

7 Conclusion

An unsupervised hierarchical segmentation scheme has been developed for segmenting textured images. The images are modeled as a mosaic of "homogeneous" first order GMRFs, and the segmentation seeks to group the image pixels into homogeneous regions. Based on homogeneity testing a three step segmentation scheme have been developed : a split-and-merge procedure for finding large homogeneous regions, a parch forming for classifying boundary pixels and a final constrained ML agglomerative patch clustering to reduce the number of resulting patches.

An analysis on testing homogeneity for GMRFs has been presented. For the case of labeled data the likelihood ratio test is used; and for the unlabeled data case two new tests, the dispersion test and the hierarchical likelihood ration test, are proposed. For the case of white Gaussian random fields useful analytic results are derived for these tests, their power functions are computed, and their performance are compared against a commonly used testing strategy.

Accurate approximations have been proposed for computing the likelihood for data observed on irregularly shaped regions. This is equivalent to compute the partition function for Gibbs distributions for the case where the random field is Gaussian.

The segmentation scheme has been applied to tissue classification in MR images of the human brain. The preliminary results are very promising. Currently we are looking at the extension to the segmentation of 3D images, and to higher order neighborhoods. We also plan to incorporate anatomical information in order to automatically assign tissue label to the resulting patches.

Acknowledgment

This work is supported in part by the Z. A. Kaprielian Innovation Research Fund. The authors would like to thank Michael J. Apozzo,MD and Hervey D. Segall,MD for providing the MR images and for their assistance in the interpretation of these images.

A Appendix : Proof of Theorem 1

Let denote $N_H(i)$, $N_V(i)$, $N_2(i)$, N_{H3} and N_{V3} the (first order) horizontal and vertical neighbor sets, second order neighbor set, third order horizontal and vertical neighbor sets of a site i, respectively. Only neighboring pixels connected to i through Z are included in these neighbor sets. By definition, $B(i,k) = 0$ when i and k are not (first order) neighbors of each other. Hence

$$\begin{aligned} Q(i,j) &= \sum_{k \in Z_+} B(i,k) \cdot B(k,j) \\ &= \sum_{k \in N_H(i)} B(i,k) \cdot B(k,j) + \sum_{k \in N_V(i)} B(i,k) \cdot B(k,j) \\ &= -\beta_1 \sum_{k \in N_H(i)} B(k,j) - \beta_2 \sum_{k \in N_V(i)} B(k,j), \end{aligned}$$

By substituting $Q(i,j)$, we have

$$\begin{aligned} \sum_{i \in Z_*} \lambda_i &= \mathrm{tr}(Q) \\ &= \sum_{i \in Z_*} Q(i,i) \\ &= \sum_{i \in Z_*} \{\beta_1^2 \sum_{k \in N_H(i)} 1 + \beta_2^2 \sum_{k \in N_V(i)} 1\} \\ &= \beta_1^2 \sum_{i \in Z_*} n_H(i) + \beta_2^2 \sum_{i \in Z_*} n_V(i) \\ &= A_1 \beta_1^2 + A_2 \beta_2^2 \\ \sum_{i \in Z_*} \lambda_i^2 &= \mathrm{tr}(Q^2) \\ &= \sum_{i \in Z_*} \sum_{j \in Z_*} Q^2(i,j) \\ &= \sum_{i \in Z_*} \{Q^2(i,i) + \sum_{j \in N_{H3}(i)} Q^2(i,j) + \sum_{j \in N_{V3}(i)} Q^2(i,j) + \sum_{j \in N_2(i)} Q^2(i,j)\} \\ &= \sum_{i \in Z_*} \{[\beta_1^2 \, n_H(i) + \beta_2^2 \, n_V(i)]^2 + \beta_1^4 \, n_{H3}(i) + \beta_2^4 \, n_{V3}(i) \\ &\qquad + \beta_1^2 \beta_2^2 \cdot [\sum_{j \in N_2(i)} n_C(i,j)]^2\} \\ &= \sum_{i \in Z_*} \{\beta_1^4 \cdot [n_H^2(i) + n_{H3}(i)] + \beta_2^4 \cdot [n_V^2(i) + n_{V3}(i)] \\ &\qquad + \beta_1^2 \beta_2^2 \cdot \{2 \, n_H(i) n_V(i) + [\sum_{j \in N_2(i)} n_C(i,j)]^2\}\} \\ &= A_{11} \beta_1^4 + A_{22} \beta_2^4 + A_{12} \beta_1^2 \beta_2^2 \end{aligned}$$

B Appendix : Proof of Theorem 2

The following lemma is an extension to the well known result in statistics that the sample mean and sample variance of i.i.d. normal random variables are independent. The proof follows the proof of that result given in [22].

Lemma 1 : Let $X = (X_1, X_2, \cdots, X_K)^T$ be K independent random vectors, where $X_i = (x_{i1}, x_{i2}, \cdots, x_{iN_i})^T$ are N_i independent normal random variables with with mean μ_{ij} and variance σ^2, then $\bar{X}$ and $(\bar{X}_1 - \bar{X}, \bar{X}_2 - \bar{X}, \cdots, \bar{X}_K - \bar{X})$ are statistically independent, where the bar denotes the sample mean of the corresponding vector.

Proof : Write the joint characteristic function of the $K+1$ quantities involved:

$$\phi_{(\bar{X}, \bar{X}_1 - \bar{X}, \cdots, \bar{X}_K - \bar{X})}(t, t_1, \cdots, t_K) = \mathrm{E}\{\exp(j[t\bar{X} + \sum_{i=1}^{K} t_i(\bar{X}_i - \bar{X})]\}.$$

But observe that $\bar{X} = \sum \frac{N_i \bar{X}_i}{N}$ and define $\bar{t} = \frac{1}{N}\sum t_i$, hence

$$t\bar{X} + \sum_{i=1}^{K} t_i(\bar{X}_i - \bar{X}) = \sum_{i=1}^{K}[(\frac{t}{N} - \bar{t})N_i + t_i]\bar{X}_i = \sum_{i=1}^{K} a_i \bar{X}_i$$

which is a linear combination of independent, normal random variables $\{\bar{X}_i\}$ with coefficients

$$a_i = (\frac{t}{N} - \bar{t})N_i + t_i = \frac{N_i}{N}t + (t_i - N_i\bar{t})$$

having the properties

$$\sum_{i=1}^{K} \bar{\mu}_i a_i = \bar{\mu}t + \sum_{i=1}^{K} \bar{\mu}_i(t_i - N_i\bar{t})$$

and

$$\begin{aligned}
\sum_{i=1}^{K} \frac{a_i^2}{N_i} &= \sum_{i=1}^{K} \frac{1}{N_i}[\frac{t^2 N_i^2}{N^2} + \frac{2tN_i}{N}(t_i - N_i\bar{t}) + (t_i - N_i\bar{t})^2] \\
&= t^2 \sum_{i=1}^{K} \frac{N_i}{N^2} + 2t \sum_{i=1}^{K} \frac{t_i}{N} - 2t \cdot \bar{t} + \sum_{i=1}^{K} \frac{(t_i - N_i\bar{t})^2}{N_i} \\
&= \frac{t^2}{N} + \sum_{i=1}^{K} \frac{(t_i - N_i\bar{t})^2}{N_i}
\end{aligned}$$

The joint characteristic function of $\bar{X}$ and $(\bar{X}_1 - \bar{X}, \cdots, \bar{X}_K - \bar{X})$ can then be written as

$$\begin{aligned}
\mathrm{E}[\exp(\sum_{i=1}^{K} j a_i \bar{X}_i)] &= \prod_{i=1}^{K} \exp(j\bar{\mu}_i a_i - \frac{\sigma_i^2}{2}a_i^2) \\
&= \exp(j\sum_{i=1}^{K} \bar{\mu}_i a_i - \frac{\sigma^2}{2}\sum_{i=1}^{K} \frac{a_i^2}{N_i}) \\
&= \exp[j\bar{\mu}t - \frac{\sigma^2 t^2}{2N}] \cdot \exp[j\sum_{i=1}^{K} \bar{\mu}_i(t_i - N_i\bar{t}) - \frac{\sigma^2}{2}\sum_{i=1}^{K} \frac{(t_i - N_i\bar{t})^2}{N_i}].
\end{aligned}$$

The last form of the joint characteristic function is separable in t and $(t_1, \cdots, t_N)$, therefore $\bar{X}$ and $(x_1 - \bar{X}, x_2 - \bar{X}, \cdots, x_N - \bar{X})$ are independent.

Proof of theorem 2 : At an arbitrary point of the hierarchical testing scheme, the data X is divided into $(X_1, X_2, \cdots, X_K)$ and the test for homogeneity is

$$\tilde{W}_K = \frac{1}{\sigma^2} \sum_{i=1}^{K} N_i \cdot (\bar{X}_i - \bar{X})^2.$$

Since $\tilde{W}_K$ is exclusively a function of $\{\bar{X}_i - \bar{X}\}$, $\tilde{W}_K$ must be independent of $\bar{X}$ following lemma 1. Also note that the only statistic of X_i that $\tilde{W}_K$ uses is its sample mean $\bar{X}_i$. Similarly the test at one level above will only use the statistic $\bar{X}$ from the data set X. Therefore the test at one level above should be independent of $\tilde{W}_K$. By repeating the same argument, we show that all tests involved in the hierarchical scheme are independent.

References

[1] J. W. Woods, "Two-dimensional discrete Markovian fields," *IEEE Trans. Information Theory*, vol. 18, pp. 232–240, Mar. 1972.

[2] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Stat. Soc., Series B*, vol. 36, pp. 192–236, 1974.

[3] A. K. Jain, "Advances in mathematical models for image processing," *Proceedings of the IEEE*, vol. 69, pp. 502–528, May 1981.

[4] B. D. Ripley, *Statistical Inference for Spatial Processes.* Cambridge University Press, 1988.

[5] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, pp. 721–741, Nov. 1984.

[6] H. Derin, H. Elliott, R. Cristi, and D. Geman, "Bayes smoothing algorithms for segmentation of binary images modeled by Markov random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, pp. 707–720, Nov. 1984.

[7] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Stat. Soc., Series B*, vol. 48, pp. 259–302, 1986.

[8] R. Chellappa and R. L. Kashyap, "Digital image restoration using spatial interaction model," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 30, pp. 461–472, Jun. 1982.

[9] G. R. Cross and A. K. Jain, "Markov random field texture models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 5, pp. 25–39, Jan. 1983.

[10] R. Chellappa and R. L. Kashyap, "Texture synthesis using 2-D noncausal autoregressive models," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 33, pp. 194–203, Feb. 1985.

[11] R. Chellappa and S. Chatterjee, "Classification of textures using Gaussian Markov random field," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 33, pp. 959–963, Aug. 1985.

[12] H. Derin and W. S. Cole, "Segmentation of textured images using Gibbs random fields," *Comp. Vision Graphics Image Processing*, vol. 35, pp. 72–98, 1986.

[13] T. Simchony and R. Chellappa, "Stochastic and deterministic algorithms for MAP texture segmentation," *in Proc. IEEE ICASSP*, 1988.

[14] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *J. Amer. Stat. Soc.*, vol. 82, pp. 76–89, Mar 1987.

[15] F. S. Cohen and D. B. Cooper, "Real time textured-image segmentation based on noncausal Markovian random field models," *in Proc. of 3rd Int. Conf. on Robot Vision & Sensory Control*, Nov. 1983.

[16] F. S. Cohen and D. B. Cooper, "Simple parallel hierarchical and relaxation algorithm for segmenting noncausal Markovian random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, pp. 195–219, Mar. 1987.

[17] R. H. Lee and R. Leahy, "Multi-spectral tissue classification of mr images using sensor fusion approaches," *in Proc. SPIE conf. on Medical Imaging IV*, 1990.

[18] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images and Gibbs random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, pp. 39–55, Jan. 1987.

[19] R. Lankhmanan and H. Derin, "Simultaneous parameter estimation and segmentation of Gibbs random field using simulated annealing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 799–813, Aug. 1989.

[20] J. F. Silverman and D. B. Cooper, "Bayesian clustering for unsupervised estimation of surface and texture models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, pp. 482–495, July 1988.

[21] Z. Wu and R. Leahy, "A new unsupervised hierarchical segmentation algorithm for textured images," *in Proc. IEEE ICASSP*, 1990.

[22] B. W. Lindgren, *Statistical Theory.* McMillan Publishing Co., Inc., New York, 3rd ed., 1976.

[23] E. L. Lehamann, *Testing Statistical Hypotheses.* John Wiley & Sons, Inc., 2nd ed., 1986.

[24] P. C. Chen and T. Pavlidis, "Image segmentation as an estimation problem," *Comp. Graphics Image Processing*, vol. 12, pp. 153–172, 1980.

[25] P. C. Chen and T. Pavlidis, "Segmentation by texture using correlation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 5, pp. 64–69, 1983.

[26] K. Ord, "Estimation methods for models of spatial interaction," *J. Amer. Stat. Soc.*, vol. 70, pp. 120–126, Mar 1975.

[27] J. Besag and P. A. P. Moran, "On the estimation and testing of spatial interaction in Gaussian lattice processes," *Biometrika*, vol. 62, pp. 555–562, 1975.

[28] J. Besag, "Statistical analysis of non-lattice data," *The Statistician*, vol. 24, pp. 179–195, 1975.

[29] R. L. Kashyap and R. Chellappa, "Estimation and choice of neighbors in spatial-interactive models of images," *IEEE Trans. Information Theory*, vol. 29, pp. 60–72, Jan. 1983.

[30] P. Whittle, "On stationary processes in the plane," *Biometrika*, vol. 41, pp. 434–449, 1954.

[31] X. Guyon, "Parameter estimation for a stationary process on a d-dimensional lattice," *Biometrika*, vol. 69, pp. 95–105, 1982.

[32] K. V. Mardia and R. J. Marshall, "Maximum likelihood estimation of models for residual covariance in spatial regression," *Biometrika*, vol. 71, pp. 135–146, 1984.

[33] D. Luenberger, *Linear and Nonlinear Programming.* Addison-Wesley, 2nd ed., 1984.

[34] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, New York, 2nd ed., 1980.

[35] H. Samet, "The quadtree and related hierarchical data structures," *Computing Surveys*, vol. 16, pp. 187–260, 1984.

[36] H. Samet, "Connected component labeling using quadtree," *J. ACM*, vol. 28, pp. 487–501, 1981.

[37] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis.* Wiley, New York, 1973.