

Multiresolution Gauss Markov Random Field Models for Texture Segmentation *

Santhana Krishnamachari[‡] and Rama Chellappa[†]

[‡] Image Processing Department
Communication Technology Division
COMSAT Laboratories
Clarksburg, MD 20871

[†] Department of Electrical Engineering and
Center for Automation Research
University of Maryland
College Park, MD 20742

Abstract: This paper presents multiresolution models for Gauss Markov random fields (**GMRF**) with applications to texture segmentation. Coarser resolution sample fields are obtained by subsampling the sample field at the fine resolution. Although the Markov property is lost under such resolution transformation, coarse resolution non-Markov random fields can be effectively approximated by Markov fields. We present two techniques to estimate the GMRF parameters at coarser resolutions from the fine resolution parameters, one by minimizing the Kullback-Leibler distance and another based on local conditional distribution invariance. We also allude to the fact that different GMRF parameters at the fine resolution can result in the same probability measure after subsampling and present the results for first and second order cases.

We apply this multiresolution model to texture segmentation. Different texture regions in an image are modeled by GMRFs and the associated parameters are assumed to be known. Parameters at lower resolutions are estimated from the fine resolution parameters. The coarsest resolution data is first segmented and the segmentation results are propagated upwards to the finer resolution. We use the iterated conditional mode (ICM) minimization at all resolutions. A confidence measure is attached to the segmentation result at each pixel and passed on to the higher resolution. At each resolution, ICM is restricted only to pixels

*Supported in part by Grant #ASC 9318183 from National Science Foundation.

with low confidence measure. Our experiments with synthetic, Brodatz texture and real satellite images show that the multiresolution technique results in a better segmentation and requires lesser computation than the single resolution algorithm.

1 Introduction

There has been an increasing emphasis in using statistical techniques for modeling and processing images in the image analysis community. Most of the research has been restricted to Markov random field (**MRF**) models, rightly so, because of the local statistical dependence of images. Introduction of the Markov models in a Bayesian framework has resulted in a unified, coherent framework for processing images which enables posing many image processing problems as statistical inference problems. MRFs have been used to model various image features such as textures, edges, region labels. Since MRF models express global statistics in terms of the local neighborhood potentials, all computations are restricted to a local window. This spawned a lot of interest in developing algorithms that utilize local computations to achieve global optimization. But the main drawback of MRF algorithms is that the optimization schemes associated with the energy functions are iterative. Typical MRF algorithms visit all lattice sites in a specific order and perform a local computation at each site; this is repeated until some form of convergence is reached. In essence, even though MRF models ensure that the conditional statistics are local, still there are long range interactions that necessitate the use of iterative algorithms. Even though the individual iterations involve only simple local computations, the iterative nature of these algorithms contributes to the computational burden.

Energy functions associated with most non-trivial MRF problems are extremely non-convex, so conventional gradient based methods are to no avail. Best results are obtained by using simulated annealing [12], which is a stochastic relaxation technique, but it is computationally very taxing. There are two different approaches that have been used to reduce the computational burden. The first is to use non-optimal, deterministic methods that converge to a local optimal point, but still provide reasonably good results. Geiger and Giroso [11] and Zhang [35] use mean field approximations that lead to deterministic relaxation algorithms. Wu and Doerschuk [34] use a tree approximation that replaces the lattice on which an MRF is defined by an acyclic tree which allows replacing the iterative MRF computations by recursive computations.

The second approach is to use multiresolution techniques. Two important aspects of multiresolution approaches are: (1) divide and conquer and (2) action at a distance [29]. Multigrid methods proposed by Terzopoulos [31], substantially reduce the computation when applied to some computer vision problems. Chen and Pavlidis [5] use a hierarchical approach to texture segmentation, but do not directly use MRFs. Bouman and Liu [2] use a causal autoregressive

Gaussian model and a quad-tree structure to perform multiresolution segmentation. Gidas [14] shows some connection between the renormalization group studies in statistical mechanics [10] and the multiresolution MRF framework to process images. A hierarchical image analysis scheme based on renormalization group ideas is presented also by Matsuba in [24]. Perez and Heitz [27] present a multiscale MRF model, where optimization is performed sequentially over specific subsets of original configuration space. The result of optimization at each scale is used to initialize the optimization at the subsequent finer scale. Cohen and Cooper [6] present a hierarchical scheme, where the data at lower resolution is divided into blocks and the conditional probabilities of these blocks given the neighboring blocks, are obtained. They use a recursive scheme to calculate the within-block and between-block interaction values.

Jeng, in [16], discusses the effect of subsampling resolution transformation on MRFs and presents two results: first, the Markov property is not preserved for a general subsampling scheme and, second, it is preserved under some specific subsampling schemes depending on the size and shape of the neighborhood. However, these subsampling schemes under which the Markov property is preserved are not very interesting.

In [20] Lakshmanan and Derin present an excellent discussion on multiresolution GMRF models. It is shown that the GMRFs lose their Markov property under subsampling and expressions for the power spectral density functions at coarser resolution are obtained. It is also shown that for the special case of second order GMRFs with separable autocovariances, the Markov property is retained under subsampling. In addition, a *covariance invariance* approximation is presented to approximate the coarser resolution data by GMRFs. This GMRF approximation is such that the covariances associated with pairs of sites that are neighbors are equal to the corresponding entries in the covariance matrix associated with the exact non-Markov density. Many interesting properties of this estimator such as maximizing the entropy and minimizing the Kullback-Leibler (**KL**) distance can be found in [20].

Luetttgen, *et al.* [22] propose a multiscale model using a tree construction for bilateral (called reciprocal processes by the authors) Markov processes in 1-D by using a midpoint deflection construction. The tree is constructed in such a way that any node in the tree is conditionally independent of every other node given the parent node. Such a construction allows individual subtrees to be processed independent of other subtrees, thereby reducing the computational load. This representation is extended to 2-D by using a midline deflection construction and a quadtree

graph. However, the construction of this multiscale model is such that the set of lattice points that is retained at a particular scale is not distributed uniformly (spatially) on the set of lattice points at the finest scale. This might be an undesirable feature of this multiscale model for image analysis.

In this paper, we present two multiresolution GMRF models. Given that the data at the fine resolution is a GMRF, the goal is to obtain suitable models at coarser resolutions. Data at coarser resolutions are obtained by subsampling the fine resolution data. Under these resolution transformations, coarser resolution data are non-Markov. We present two schemes to estimate the parameters corresponding to the “best” GMRF approximation at lower resolutions from the parameters at the fine resolution, one by directly minimizing the KL distance (relative entropy) and the second by minimizing the KL distance between the conditional densities (conditional relative entropy). We also show that the computations for these two estimators turn out to be similar to the traditional maximum likelihood [25], [19], [30] and the pseudo likelihood estimators [1], except that the sample covariances are replaced by covariances calculated with respect to the exact non-Markov measure that is being approximated. We present results on the existence of different sets of GMRF parameters at fine resolution that result in statistically identical coarser resolution random fields. As an application, we consider the texture segmentation problem and performed segmentation over multiple resolutions using our multiresolution GMRF model. The coarsest resolution image is initially segmented and the results of segmentation along with a confidence measure are passed on to the immediate higher resolution. This is repeated until the fine resolution image is segmented. We show that the multiresolution technique performs better than the single resolution approach.

The rest of the paper is organized as follows. Section 2 introduces the GMRF and the basics of the resolution transformation. Section 3 presents the Markov approximation for non-Markov fields based on KL distance minimization and local conditional distribution invariance approximation. Section 4 carries the details of many-to-one nature of transformation of the GMRF parameters from the fine to coarse resolution. Section 5 presents various aspects of the multiresolution segmentation and Section 6 carries synthetic and real experiments. Section 7 concludes the paper.

2 GMRFs and Resolution Transformation

In this section we introduce basic notations used for GMRFs in the rest of the paper and also present results on loss of the Markov property under resolution transformation.

2.1 The GMRF Model

We use the following notation :

$t = (t_1, t_2), s = (s_1, s_2)$: coordinates of grid points on a 2-D lattice

$\Omega = \{s : 0 \leq s_1 \leq M - 1, 0 \leq s_2 \leq N - 1\}$: a two dimensional lattice

\underline{X} : a random field on Ω , represented as a vector by a row-wise scan ordering

X_s : the random variable at site s

η, ψ, ξ : neighborhood sets

The set of lattice points that are contained in the neighborhood of a site s is denoted by η_s . The elements that are included in the neighborhood of the site marked s for different neighborhood orders are shown in the Figure 2. Note that the set of sites in the neighborhood for a given order also includes all the sites that are present in all the lower orders.

For the first order case, $\eta = \{(1, 0), (0, 1), (-1, 0), (0, -1)\}$, and $\eta_s = \{s + r : r \in \eta\}$.

If \underline{X} is modeled by a GMRF with a symmetric neighborhood η , then \underline{X} can be written as [17]:

$$\begin{aligned} X_s &= \sum_{r \in \eta} \theta_r X_{s+r} + e_s \\ B_\theta \underline{X} &= \underline{e}. \end{aligned}$$

For a finite lattice, the neighborhood for the lattice sites at the boundaries of the lattice is not complete. This problem is circumvented by assuming that the lattice is folded into a toroid. The matrix B_θ is a $MN \times MN$ matrix, \underline{e} is a zero mean, Gaussian noise, with autocorrelation given by :

$$E[e_s e_{s+r}] = \begin{cases} \sigma^2 & \text{if } r = (0, 0) \\ -\theta_r \sigma^2 & \text{if } r \in \eta \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Hence the GMRF can be completely characterized by the set of parameters $\{\theta, \sigma^2\}$. The parameter set θ should satisfy the following conditions :

$$1. \quad \theta_r = \theta_{-r} \quad \forall r \in \eta$$

$$2. \quad 1 - \underline{\theta}^T \underline{\phi}_s > 0 \quad \forall s \in \Omega \quad (2)$$

where $\underline{\phi}_s$ is a vector whose length is equal to the number of elements in the neighbor set η . The individual elements of $\underline{\phi}_s$ are given by:

$$\cos \left(\left(\frac{2\pi s_1}{M} \frac{2\pi s_2}{N} \right) \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \right) \quad r \in \eta.$$

The first condition is necessary to ensure stationarity and the second to ensure that the covariance matrix of \underline{X} is positive definite.

The covariance matrix of \underline{e} is $\sigma^2 B_\theta$ and it can be shown that the covariance matrix of \underline{X} is $\Sigma = \sigma^2 B_\theta^{-1}$ [17]. The joint probability density function of \underline{X} can be written as follows:

$$p(\underline{X} = \underline{x}) = \frac{\sqrt{\det B_\theta}}{(2\pi\sigma^2)^{\frac{MN}{2}}} \exp\left\{-\frac{1}{2\sigma^2} \underline{x}^T B_\theta \underline{x}\right\}. \quad (3)$$

Also \underline{X} exhibits the Markov property [17],

$$\begin{aligned} p(x_s | x_t, \forall t \neq s, t \in \Omega) &= p(x_s | x_{s+r}, r \in \eta) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{[x_s - \sum_{r \in \eta} \theta_r x_{s+r}]^2}{2\sigma^2}\right\}. \end{aligned} \quad (4)$$

The power spectrum $S_x(\omega)$ of \underline{X} can be shown to be [17]:

$$S_x(\omega) = \frac{\sigma^2}{1 - \sum_{r \in \eta} \theta_r \cos\left[\frac{2\pi}{M} r_1 \omega_1 + \frac{2\pi}{N} r_2 \omega_2\right]} \quad (5)$$

where $\omega = \{\omega_1, \omega_2\}$, and $0 \leq \omega_1 \leq M - 1, 0 \leq \omega_2 \leq N - 1$.

2.2 GMRFs and Resolution Transformation

Let $\Omega^{(0)} = \Omega = \{(s_1, s_2) : 0 \leq s_1 \leq M - 1, 0 \leq s_2 \leq N - 1\}$ be a rectangular lattice and M and N are assumed to be powers of 2. The superscript stands for the level in the image pyramid, $\Omega^{(0)}$ being the lattice at the fine resolution, $\Omega^{(k)}$ represents the lattice which is obtained by subsampling $\Omega^{(0)}$, k times (Figure 1). Let $\underline{X}^{(k)}$ represent a random field, obtained by ordering the random variables on $\Omega^{(k)}$. The parameters of a GMRF defined on a lattice $\Omega^{(k)}$ are denoted by $\{\underline{\theta}^{(k)}, [\sigma^2]^{(k)}\}$ and the associated neighborhood is denoted by $\eta^{(k)}$. The covariance matrix and the power spectrum associated with $\underline{X}^{(k)}$ are denoted by $\Sigma^{(k)}$ and $S_x^{(k)}(\omega)$ respectively. The probability distributions defined on a lattice $\Omega^{(k)}$ are indexed by $p^{(k)}(\cdot)$.

We restrict our discussion to resolution transformation obtained by subsampling (see Figure 1). But the results can be easily extended to block-to-point type transformation, where the coarse resolution data is obtained by averaging the fine resolution data over a local window.

Let $\underline{X}^{(0)}$ be a GMRF defined on $\Omega^{(0)}$ with parameters $\{\underline{\theta}^{(0)}, [\sigma^2]^{(0)}\}$ and a neighborhood $\eta^{(0)}$. The power spectrum of $\underline{X}^{(0)}$ can be written as in Eq. (5) :

$$S_x^{(0)}(\omega) = \frac{[\sigma^2]^{(0)}}{1 - \sum_{r \in \eta^{(0)}} \theta_r^{(0)} \cos[\frac{2\pi}{M}r_1\omega_1 + \frac{2\pi}{N}r_2\omega_2]} \quad (6)$$

where $\omega = \{(\omega_1, \omega_2) : 0 \leq \omega_1 \leq M - 1, 0 \leq \omega_2 \leq N - 1\}$. The subsampling resolution transformation is defined as:

$$X_s^{(k)} = X_{2s}^{(k-1)}$$

defined for all $s \in \Omega^{(k)}$.

Equivalently,

$$\underline{X}^{(k)} = D_0^k \underline{X}^{(0)} \quad (7)$$

where the matrix D_0^k , has to be properly defined [20].

The resulting subsampled field $\underline{X}^{(k)}$ is Gaussian, with covariance $\Sigma^{(k)} = [D_0^k] \Sigma^{(0)} [D_0^k]^T$. The power spectrum of $\underline{X}^{(k)}$ can be shown to be [20]:

$$S_x^{(k)}(\omega) = \frac{1}{2^{2k}} \sum_{r \in C_k} S_x^{(0)}(\omega + r') \quad (8)$$

where $r' = (\frac{M}{2^k}r_1, \frac{N}{2^k}r_2)$ and $C_k = \{r : 0 \leq r_1 \leq 2^k - 1, 0 \leq r_2 \leq 2^k - 1\}$.

It can be observed that $S_x^{(k)}(\omega)$ cannot be written in the form of Eq. (5) with a finite neighborhood. Therefore, the subsampled fields $\underline{X}^{(k)}$ are non-Markov, except for the special case of second order separable correlation GMRFs [20].

3 Markov Approximations

As mentioned in the last section, GMRFs become non-Markov under subsampling and the probability density functions (pdf) of the subsampled random fields can be obtained. However, if the coarser resolution data are modeled by the exact non-Markov Gaussian measures, conventional optimization techniques based on Markov properties cannot be applied. In this section we show

that it is possible to obtain good Markov approximations for coarser resolution fields. Two different estimators to obtain the parameters of GMRFs at lower resolutions from the parameters at the fine resolution are presented. We also exemplify the connection between these two estimators and the estimators that are commonly used to estimate the GMRF parameters from a data sample, the maximum likelihood [25], [19], [30] and the pseudo likelihood estimators [1].

3.1 Kullback-Leibler (KL) Distance Minimization

In this section, we show that given any pdf p , it is possible to obtain a GMRF approximation of p by minimizing the KL distance $D(p \parallel q)$ [18], where q belongs to the family of GMRF pdfs. The KL distance measure is widely used to obtain approximate probability measures with desired properties. The estimator presented in this subsection is same as the estimator originally proposed by Lakshmanan and Derin [20]. The estimator presented in [20] is motivated by “covariance matching” and is also shown to minimize the KL distance. Our presentation is based on directly minimizing the KL distance, but the two estimators are the same, because of the uniqueness of this estimator.

The problem at hand can be stated as follows:

Given a stationary probability measure $p(\underline{x})$, find another probability measure $q^*(\underline{x})$ such that:

$$\begin{aligned} q^* &= \arg \min_q D(p \parallel q) \\ &= \arg \min_q E_p \left[\log \frac{p(\underline{X})}{q(\underline{X})} \right] \end{aligned} \quad (9)$$

where $E_p(\cdot)$ represents the expectation with respect to the p-measure and $q(\underline{x})$ belongs to the family of GMRF densities with a specific neighborhood η , indexed by $(\underline{\theta}, \sigma^2)$. As seen before, GMRFs can be completely characterized by $(\underline{\theta}, \sigma^2)$. From Eq. (3), $q(\underline{x})$ can be written as:

$$q(\underline{x}) = \frac{\sqrt{\det B_\theta}}{(2\pi\sigma^2)^{\frac{MN}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\underline{x}^T B_\theta \underline{x}\right\}.$$

The quadratic form $\underline{x}^T B_\theta \underline{x}$ in $q(\underline{x})$ can be simplified as [30]:

$$\underline{x}^T B_\theta \underline{x} = C(0) - \underline{\theta}^T \underline{C}$$

where \underline{C} is a vector whose length is equal to the number of elements in the neighborhood η and

is as follows:

$$C(0) = \sum_{s \in \Omega} x_s^2; \quad C(r) = \sum_{s \in \Omega} x_s x_{s+r} \quad \forall r \in \eta.$$

and the determinant of the B_θ matrix can be written as [17]:

$$\det B_\theta = \prod_{s \in \Omega} (1 - \theta^T \underline{\phi}_s).$$

Using the above equations, $q(\underline{x})$ can be written as:

$$q(\underline{x}) = \frac{\prod_s \sqrt{(1 - \theta^T \underline{\phi}_s)}}{(2\pi\sigma^2)^{\frac{MN}{2}}} \exp\left\{-\frac{1}{2\sigma^2}[C(0) - \theta^T C]\right\}.$$

Let $\{\underline{\theta}^*, [\sigma^2]^*\}$ be the parameters corresponding to $q^*(\underline{x})$. Now, rewriting Eq. (9) and performing the minimization in terms of the parameters:

$$\begin{aligned} (\underline{\theta}^*, [\sigma^2]^*) &= \arg \min_{(\underline{\theta}, \sigma^2)} E_p \left[\log \frac{p(\underline{X})}{q(\underline{X})} \right] \\ &= \arg \max_{(\underline{\theta}, \sigma^2)} E_p [\log q(\underline{X})] \\ &= \arg \max_{(\underline{\theta}, \sigma^2)} E_p \left[\frac{1}{2} \sum_s \log(1 - \theta^T \underline{\phi}_s) - \frac{MN}{2} \log \sigma^2 - \frac{(C(0) - \theta^T C)}{2\sigma^2} \right] \\ &= \arg \max_{(\underline{\theta}, \sigma^2)} \frac{1}{2} \sum_s \log(1 - \theta^T \underline{\phi}_s) - \frac{MN}{2} \log \sigma^2 \\ &\quad - \frac{1}{2\sigma^2} (E_p[C(0)] - \theta^T E_p[C]) \end{aligned} \tag{10}$$

where,

$$\begin{aligned} E_p[C(r)] &= E_p \left[\sum_{s \in \Omega} X_s X_{s+r} \right] \\ &= (MN) E_p[X_s X_{s+r}]. \end{aligned} \tag{11}$$

Remark: Observe that Eq. (10) is similar to the maximum likelihood expression [25], [19], [30], except that the $C(r)$ values obtained from the data are replaced by expectation values with respect to the p - measure. Hence, in terms of computation, the maximization is exactly similar to the maximum likelihood computation. Given the p - measure, we only need a few moment values $E_p[C(r)]$, followed by the maximization of Eq. (10) to obtain the Markov approximation.

Lakshmanan and Derin [20] also remark that their covariance invariance estimator, which minimizes the KL distance, also maximizes the likelihood of a sample observation with sample covariance equal to the covariance of the p - measure.

3.2 Local Conditional Distribution Invariance Approximation

We propose another method to estimate the best GMRF parameters of a non-Markov random field, based on a KL distance measure between local conditional distributions (conditional relative entropy). In MRF applications all optimizations are performed based on the local conditional distribution, so, we believe an estimator based on that should be well suited for image analysis applications.

The Markov approximation presented in this section is based on linear estimation. Before presenting the details, we will provide a known result regarding the linear estimation of a GMRF. Let Z be a GMRF defined by $(\underline{\theta}, \sigma^2)$ with a neighborhood ψ . Then the best estimate of Z_s based on the elements of ψ is given by [4]:

$$\hat{z}_s = \sum_{r \in \psi} \theta_r z_{s+r}$$

and the mean square error

$$E(Z_s - \hat{Z}_s)^2 = \sigma^2.$$

The conditional density $p(z_s|z_r, r \in \psi)$ is Gaussian with conditional mean $\sum_{r \in \psi} \theta_r z_{s+r}$ and conditional variance σ^2 .

Let \underline{X} be a random field with a stationary non-Markov probability measure $p(\underline{x})$ and let $q^*(\underline{x})$ be a GMRF approximation such that:

$$q^*(x_s|x_{s+r}, r \in \eta) = \arg \min_q D[p(x_s|x_{s+r}, r \in \eta) \parallel q(x_s|x_{s+r}, r \in \eta)], \quad (12)$$

where the minimization is performed over the entire family of GMRF pdfs with a chosen neighborhood η . In addition, under certain conditions (given at the end of the section), $q^*(x_s|x_{s+r}, r \in \eta)$ is exactly equal to $p(x_s|x_{s+r}, r \in \eta)$.

Since $q(\underline{x})$ belongs to the family of GMRF densities, $q(x_s|x_{s+r}, r \in \eta)$ will be of the form given in Eq. (4).

$$q(x_s|x_{s+r}, r \in \eta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{[x_s - \sum_{r \in \eta} \theta_r x_{s+r}]^2}{2\sigma^2}\right\}.$$

Let $(\underline{\theta}^*, [\sigma^2]^*)$ be the parameters corresponding to $q^*(\underline{x})$. To simplify the notation, let \underline{Y} be the vector containing the neighborhood random variables in a proper order. For a first order neighborhood,

$$\underline{Y}^T = \left(X_{s+(1,0)} \quad X_{s+(0,1)} \quad X_{s+(-1,0)} \quad X_{s+(0,-1)} \right).$$

Now,

$$\begin{aligned}
(\underline{\theta}^*, [\sigma^2]^*) &= \arg \min_{(\underline{\theta}, \sigma^2)} E_p \left[\log \frac{p(X_s | \underline{Y})}{q(X_s | \underline{Y})} \right] \\
&= \arg \max_{(\underline{\theta}, \sigma^2)} E_p [\log q(X_s | \underline{Y})] \\
&= \arg \max_{(\underline{\theta}, \sigma^2)} E_p \left[-\frac{1}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (X_s - \sum_{r \in \eta} \theta_r Y_r)^2 \right] \\
&= \arg \min_{(\underline{\theta}, \sigma^2)} \frac{1}{2} \log \sigma^2 + \frac{1}{2\sigma^2} (E_p[X_s - \sum_{r \in \eta} \theta_r Y_r])^2.
\end{aligned} \tag{13}$$

It can be seen that the $\underline{\theta}^*$ parameters corresponding to $q^*(\mathbf{x})$ are obtained by minimizing the second term in the Eq. (13)

$$\underline{\theta}^* = \arg \min_{\underline{\theta}} E_p [X_s - \sum_{r \in \eta} \theta_r Y_r]^2 \tag{14}$$

and using the $\underline{\theta}^*$ obtained, we can estimate the $[\sigma^2]^*$ that minimizes Eq. (13),

$$[\sigma^2]^* = E_p [X_s - \sum_{r \in \eta} \theta_r^* Y_r]^2. \tag{15}$$

then,

$$\begin{aligned}
\underline{\theta}^* &= \arg \min_{\underline{\theta}} E_p [X_s - \underline{\theta}^T \underline{Y}]^2 \\
\underline{\theta}^* &= [E_p(\underline{Y} \underline{Y}^T)]^{-1} E_p(X_s \underline{Y})
\end{aligned} \tag{16}$$

and,

$$\begin{aligned}
[\sigma^2]^* &= E_p(X_s^2) - E_p(X_s \underline{Y}^T) [E_p(\underline{Y} \underline{Y}^T)]^{-1} E_p(X_s \underline{Y}) \\
&= E_p(X_s^2) - [\underline{\theta}^*]^T E_p(X_s \underline{Y}).
\end{aligned} \tag{17}$$

In addition, the estimated $\underline{\theta}^*$ parameters should satisfy the positivity conditions in Eq. (2):

$$1 - [\underline{\theta}^*]^T \underline{\phi}_s > 0 \quad \forall s \in \Omega.$$

Now, returning back to multiresolution discussion, let $X^{(0)}$ be a GMRF defined by $(\underline{\theta}^{(0)}, [\sigma^2]^{(0)})$ and $X^{(k)}$ be the field obtained by subsampling $X^{(0)}$, k times. The non-Markov $X^{(k)}$ can be approximated by a GMRF by minimizing Eq. (12). The minimization requires the autocorrelation

values $E_{p^{(k)}}[X_s^{(k)}X_{s+r}^{(k)}]$ which can be computed, given the GMRF parameters for $X^{(0)}$ as shown below.

$$\begin{aligned} X_s^{(k)} &= X_{2^k s}^{(0)} \\ E_{p^{(k)}}[X_s^{(k)}X_{s+r}^{(k)}] &= E_{p^{(0)}}[X_{2^k s}^{(0)}X_{2^k(s+r)}^{(0)}]. \end{aligned}$$

For any two lattice sites u and v in $\Omega^{(0)}$ the correlation is given by [17]:

$$E_{p^{(0)}}[X_u^{(0)}X_v^{(0)}] = \frac{[\sigma^2]^{(0)}}{MN} \sum_{s \in \Omega^{(0)}} \frac{(\lambda_M^{s_1 u_1} \lambda_N^{s_2 u_2})(\lambda_M^{-s_1 v_1} \lambda_N^{-s_2 v_2})}{1 - [\theta^{(0)}]^T \underline{\phi}_s} \quad (18)$$

where $\lambda_n^i = \exp(\sqrt{-1} \frac{2\pi i}{n})$.

Under the assumption that the covariance matrix with respect to p - measure is positive definite, the function in Eq. (14) to be minimized is convex and is minimized over a convex set defined by $1 - [\theta^{(k)}]^T \underline{\phi}_s > 0$, for $\forall s \in \Omega$. If the solution lies inside the convex set, it can be obtained from Eq. (16). Otherwise, a gradient descent procedure can be used.

Remarks:

1. If the $\underline{\theta}^*$ obtained from Eq. (16) satisfies the positivity conditions and if p is Gaussian, then $p(x_s|x_{s+r}, r \in \eta) = q^*(x_s|x_{s+r}, r \in \eta)$. Since $p(\underline{x})$ is Gaussian, $p(x_s|x_{s+r}, r \in \eta)$ is also Gaussian with conditional mean $\sum_{r \in \eta} \theta_r^* x_{s+r}$ (which is the best linear estimate of X_s in terms of $X_{s+r}, r \in \eta$) and conditional variance $[\sigma^2]^*$ (which is the corresponding minimum mean square error of the estimator) [26]. $q^*(\underline{x})$ being a GMRF with parameters $(\underline{\theta}^*, [\sigma^2]^*)$, from the discussion at the beginning of this section, has the conditional distribution $q^*(x_s|x_{s+r}, r \in \eta)$ with the conditional mean $\sum_{r \in \eta} \theta_r^* x_{s+r}$ and conditional variance $[\sigma^2]^*$. However, the joint densities $p(\underline{x})$ and $q(\underline{x})$ on the whole lattice are not the same, $p(\underline{x})$ is a non-Markov density and $q(\underline{x})$ is a Markov density.

2. It is worth observing that Eq. (13) is similar to the pseudo likelihood estimate [4], [1] where the GMRF parameters are obtained by minimizing the products of local conditional densities over the entire lattice. The pseudo likelihood estimator uses the sample covariances obtained from the observed sample field, whereas our local conditional distribution invariance estimator uses the covariances calculated with respect to the p - measure.

3.3 PSD Comparisons

We show the validity of local conditional distribution invariance approximation to estimate the GMRF parameters at lower resolutions by comparing the exact power spectrum function of the

Level (k)	Order (m)	$D(k, m)$
1	3	0.72
	4	0.32
2	2	0.26
	3	0.21
3	2	0.023
	3	0.022

Table 1: Comparison of the power spectral functions of the exact non-Markov fields and the GMRF approximations.

subsampled fields and the power spectrum associated with the GMRF approximation.

Let $\underline{X}^{(0)}$ be a second order GMRF with $\underline{\theta}^{(0)} = \{ \theta_{(1,0)} = 0.2, \theta_{(0,1)} = -0.1, \theta_{(1,1)} = -0.25, \theta_{(-1,1)} = 0.15 \}$ and $[\sigma^2]^{(0)} = 6.0$. Let $S_x^{(k)}(\omega)$ be the exact power spectral density function at $\Omega^{(k)}$ and let $MS_x^{(k)}(\omega, m)$ be the power spectrum for the m-th order Markov approximation. We calculate a normalized absolute difference $D(k, m) = \frac{\sum_{(\omega \in \Omega^{(k)})} |S_x^{(k)}(\omega) - MS_x^{(k)}(\omega, m)|}{size(\Omega^{(k)})}$, where $size(\Omega^{(k)})$ is the number of lattice sites in $\Omega^{(k)}$. Table 1 shows the values of $D(k, m)$, for values of $k = 1, 2, 3$. Figure 3(a) shows the power spectrum at the fine resolution calculated by Eq. (5) and Figure 3(b) shows the exact power spectrum (non-Markov) at $\Omega^{(1)}$ calculated by Eq. (8). Figure 3(c) shows the power spectrum of the third order Markov approximation with the GMRF parameters obtained from the local conditional distribution invariance approximation and Figure 3(d) shows the same for the fourth order Markov approximation. Figure 4(a) shows the exact power spectrum (non-Markov) at $\Omega^{(2)}$. Figure 4(b) shows the second order Markov approximation and Figure 4(c) shows the third order Markov approximation. From these figures and Table 1, it is easy to see that the power spectrums of these Markov approximations are very similar to the power spectrums of the exact non-Markov fields.

Comments:

1. For any k, $D(k, m + 1) \leq D(k, m)$. This is true because the GMRF approximation of order $m + 1$ includes order m .
2. We have observed that $D(k + 1, m) \leq D(k, m)$, i.e., as the level of subsampling increases

the order of the GMRF approximation need not be increased.

3. If the order of the GMRF model at the fine resolution is m , then an m or $m + 1$ order GMRF approximation at $\Omega^{(1)}$ and an m order approximation at the subsequent coarser resolutions results in very small values of the normalized absolute difference $D(k, m)$. This can be used as a thumb rule to select the order of GMRF approximations at coarser resolutions.

4 Parameters Resulting in Identical PDFs at Coarser Resolutions

In the previous section, we presented methods to approximate subsampled random fields by GMRFs assuming that data at the fine resolution is modeled by a GMRF. It is also necessary to analyze if different GMRF parameters at the fine resolution can result in probabilistically identical coarser resolution random fields. Since we are dealing with Gaussian fields, it suffices to check the covariance matrices of the subsampled fields instead of the pdfs. However, the covariance elements are complicated functions of the parameters (see Eq. (18)). Therefore, we look at the power spectrum of the subsampled random fields which are simpler functions of the parameters. We show that there exists different sets of GMRF parameters, which on subsampling result in the same pdf at the lower resolution. Since the parameter $[\sigma^2]^{(0)}$ is a multiplicative factor in the power spectral function, we assume it be equal to one and investigate the existence of different sets of θ parameters that result in the identical coarser resolution random fields.

Case 1: First order GMRF on $\Omega^{(0)}$

The first order GMRF model is defined by the parameters $(\theta_{(1,0)}, \theta_{(0,1)}, 1)$ and the power spectral density function is given by:

$$S_x^{(0)}(\omega) = \frac{1}{1 - 2(\theta_{(1,0)} \cos \frac{2\pi\omega_1}{M} + \theta_{(0,1)} \cos \frac{2\pi\omega_2}{N})}. \quad (19)$$

From Eq.(8) the power spectral function on subsampling can be written as:

$$S_x^{(1)}(\omega) = \frac{1}{4}[S_x^{(0)}(\omega) + S_x^{(0)}(\omega + (\frac{M}{2}, 0)) + S_x^{(0)}(\omega + (0, \frac{N}{2})) + S_x^{(0)}(\omega + (\frac{M}{2}, \frac{N}{2}))]. \quad (20)$$

Using Eq. (19) in Eq. (20) and after some manipulations, we obtain:

$$S_x^{(1)}(\omega) =$$

$$\frac{1}{2} \left[\frac{1}{1 - 4(\theta_{(1,0)} \cos \frac{2\pi\omega_1}{M} + \theta_{(0,1)} \cos \frac{2\pi\omega_2}{N})^2} + \frac{1}{1 - 4(\theta_{(1,0)} \cos \frac{2\pi\omega_1}{M} - \theta_{(0,1)} \cos \frac{2\pi\omega_2}{N})^2} \right]. \quad (21)$$

Claim:

For a first order GMRF at the fine resolution, the only set of parameters that results in the same power spectrum at $\Omega^{(1)}$ is $(\theta_{(1,0)}, \theta_{(0,1)})$, $(-\theta_{(1,0)}, \theta_{(0,1)})$, $(\theta_{(1,0)}, -\theta_{(0,1)})$, $(-\theta_{(1,0)}, -\theta_{(0,1)})$.

Proof: From Eq. (21) it can be inferred that, the sets of parameters $(\theta_{(1,0)}, \theta_{(0,1)})$, $(-\theta_{(1,0)}, \theta_{(0,1)})$, $(\theta_{(1,0)}, -\theta_{(0,1)})$, $(-\theta_{(1,0)}, -\theta_{(0,1)})$ will result in the same $S_x^{(1)}(\omega)$. For these sets of parameters, the subsampled fields are statistically indistinguishable. We need to show that these are the only set of parameters that result in the same power spectrum at the lower resolution.

Since the power spectrum has to be the same for every $0 \leq \omega_1 \leq \frac{M}{2} - 1$ and $0 \leq \omega_2 \leq \frac{N}{2} - 1$, we can substitute specific values of (ω_1, ω_2) , to get the necessary conditions.

Let, $\omega_1 = 0$ and $\omega_2 = N/4$

$$S_x^{(1)}(0, \frac{N}{4}) = \frac{1}{1 - 4\theta_{(1,0)}^2}.$$

Hence, the only possible values of $\theta_{(1,0)}$ that will result in the same $S_x^{(1)}(0, \frac{N}{4})$ are $\theta_{(1,0)}$ and $-\theta_{(1,0)}$. The same can be proved for $\theta_{(0,1)}$ by taking $\omega_1 = N/4$ and $\omega_2 = 0$. This proves our claim.

Case 2: Second order GMRF on $\Omega^{(0)}$

The second order GMRF model is defined by the parameters $(\theta_{(1,0)}, \theta_{(0,1)}, \theta_{(1,1)}, \theta_{(-1,1)}, 1)$ and the power spectral density function is given by:

$$S_x^{(0)}(\omega) = \frac{1}{1 - 2[\theta_{(1,0)} \cos \frac{2\pi\omega_1}{M} + \theta_{(0,1)} \cos \frac{2\pi\omega_2}{N} + \theta_{(1,1)} \cos(\frac{2\pi\omega_1}{M} + \frac{2\pi\omega_2}{N}) + \theta_{(-1,1)} \cos(\frac{2\pi\omega_1}{M} - \frac{2\pi\omega_2}{N})]}.$$

The power spectral function on subsampling is given by:

$$S_x^{(1)}(\omega) = \frac{1}{4} \left[\frac{1}{1 - 2(\theta_{(1,0)} \cos \frac{2\pi\omega_1}{M} + \theta_{(0,1)} \cos \frac{2\pi\omega_2}{N} + \theta_{(1,1)} \cos(\frac{2\pi\omega_1}{M} + \frac{2\pi\omega_2}{N}) + \theta_{(-1,1)} \cos(\frac{2\pi\omega_1}{M} - \frac{2\pi\omega_2}{N}))} + \frac{1}{1 - 2(\theta_{(1,0)} \cos \frac{2\pi\omega_1}{M} - \theta_{(0,1)} \cos \frac{2\pi\omega_2}{N} - \theta_{(1,1)} \cos(\frac{2\pi\omega_1}{M} + \frac{2\pi\omega_2}{N}) - \theta_{(-1,1)} \cos(\frac{2\pi\omega_1}{M} - \frac{2\pi\omega_2}{N}))} + \frac{1}{1 - 2(-\theta_{(1,0)} \cos \frac{2\pi\omega_1}{M} + \theta_{(0,1)} \cos \frac{2\pi\omega_2}{N} - \theta_{(1,1)} \cos(\frac{2\pi\omega_1}{M} + \frac{2\pi\omega_2}{N}) - \theta_{(-1,1)} \cos(\frac{2\pi\omega_1}{M} - \frac{2\pi\omega_2}{N}))} \right]$$

$$+ \frac{1}{1 - 2(-\theta_{(1,0)} \cos \frac{2\pi\omega_1}{M} - \theta_{(0,1)} \cos \frac{2\pi\omega_2}{N} + \theta_{(1,1)} \cos(\frac{2\pi\omega_1}{M} + \frac{2\pi\omega_2}{N}) + \theta_{(-1,1)} \cos(\frac{2\pi\omega_1}{M} - \frac{2\pi\omega_2}{N}))}] . \quad (22)$$

Claim:

For a second order GMRF at the fine resolution, the only set of parameters that result in the same power spectrum at $\Omega^{(1)}$ is,

$$(\theta_{(1,0)}, \theta_{(0,1)}, \theta_{(1,1)}, \theta_{(-1,1)}), (-\theta_{(1,0)}, \theta_{(0,1)}, -\theta_{(1,1)}, -\theta_{(-1,1)}),$$

$$(\theta_{(1,0)}, -\theta_{(0,1)}, -\theta_{(1,1)}, -\theta_{(-1,1)}), (-\theta_{(1,0)}, -\theta_{(0,1)}, \theta_{(1,1)}, \theta_{(-1,1)}).$$

Proof:

The proof is presented in the Appendix.

Similar results can be obtained for higher order cases. As we have shown for the first and second order cases, it is easy to show for all higher order cases that there exists atleast four sets of parameters that will result in the same power spectrum at the coarser resolution. The proof to show that these are the only sets of parameters gets more complicated for higher order cases, however, it is our conjecture that this is true.

5 Texture Segmentation

Computer vision and image analysis algorithms use various visual cues to analyze and interpret an image of a complex scene. These visual cues include, among others, photometric and geometric cues. Photometric cues include shading, texture, etc., from which features such as edges and regions are obtained. Texture is one of the basic characteristics of a visible surface and provides useful information for scene segmentation and understanding. Texture is a very important property for the analysis of remote sensed satellite images, their segmentation into various vegetation classes.

Texture classification and segmentation problems have been addressed by several authors with different approaches that can be broadly classified into two, namely, structural and statistical approaches. Structural approaches are aimed at regular textures that exhibit a strong structural behavior. These approaches define a basic primitive and placement rules. Such techniques were used by Rosenfeld [28], Lu and Fu [21] and Tomita [32]. Structural approaches are well suited for macro textures, but are not useful for micro textures where it is hard to define a basic primitive

and placement rules. Statistical methods are better suited for such micro textures. Most of the early statistical methods use first and second order properties to discriminate between textures. Haralick [15] suggested various local statistical measures based on the gray level dependence, later generalized by Davis, et.al. [8].

More complex statistical models using MRFs [33] emerged later and have been used in various image processing problems. Cross and Jain [7] and Chellappa [4] have shown the applicability of GMRF models to synthesize textures. Chellappa and Chatterjee [3] have used GMRF models for maximum likelihood segmentation of textures. Geman and Graffigne [13] used a general MRF model for segmentation. They showed that this non-Gaussian MRF model was not very well suited for synthesis, but still performed well for segmentation. Derin and Elliot [9] used Gibbs distributions and presented a non-optimal method using dynamic programming. Manjunath et.al. [23] proposed a stochastic learning technique to improve the ICM results and presented a neural network implementation for texture segmentation.

Texture segmentation problem is the labeling of pixels in a lattice to one of V texture classes, based on a texture model and the observed intensity field. Each site in the lattice carries a class label (say, $L_s = v$, $v \in \{1, 2, \dots, V\}$) and this label field is modeled by an MRF. We do not directly observe the label field, but a function of the labels, the intensity field. The intensity field is modeled by a GMRF, whose parameters depend on the value of label field at that site. The goal is to estimate the unobserved label field from the observed intensities by optimizing a suitable error criterion.

We model the label field by an MRF with a neighborhood ψ :

$$p(L = l) = \frac{1}{Z} \exp \left[\beta \sum_{s \in \Omega} U(l_s) \right] \quad (23)$$

where $U(l_s)$ is the number of neighbors in ψ that have the same label as l_s . This model is also called a pairwise interaction model.

The local conditional probability of the label field is given by:

$$p(l_s | l_{s+r}, r \in \psi) = \frac{\exp [\beta U(l_s)]}{\sum_{l'_s=(1,2,\dots,V)} \exp [\beta U(l'_s)]}. \quad (24)$$

The GMRF parameters corresponding to a label v , are denoted by $(\theta(v), \sigma^2(v))$. The conditional density of the intensity field can be written as follows, from Eq. (4):

$$p(X_s = x_s | L_s = v, X_{s+r}, r \in \eta)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2(v)}} \exp\left\{-\frac{1}{2\sigma^2(v)}\left[x_s - \sum_{r \in \eta} \theta_r(v)x_{s+r}\right]^2\right\}. \quad (25)$$

Now given the intensities, the label field can be estimated by optimizing a suitable criterion. The maximum a posterior (MAP) error criterion solution can be obtained by:

$$\max_{\underline{L}} P(\underline{X}, \underline{L}) = \max_{\underline{L}} P(\underline{X}|\underline{L})P(\underline{L}). \quad (26)$$

This optimization requires stochastic relaxation methods and is computationally very expensive. So, we restrict ourselves to the iterated conditional mode method. The ICM solution is obtained by performing the following optimization at each lattice site [1]:

$$\max_{L_s} P(X_s|L_s, X_{s+r}, r \in \eta)P(L_s|L_{s+r}, r \in \psi).$$

This is equivalent to,

$$\min_{L_s} \frac{1}{2} \log(\sigma^2(v)) + \frac{1}{2\sigma^2(v)} \left[x_s - \sum_{r \in \eta} \theta_r(v)x_{s+r} \right]^2 - \beta U(L_s = v) \quad (27)$$

the minimization is performed by visiting the pixels in raster scan order for all $s \in \Omega$ and stopped when no further changes in the labels occur.

5.1 Multiresolution Segmentation

The segmentation algorithm presented above is a single resolution algorithm. As we have discussed before, data at lower resolutions can be approximated by a GMRF. Thus the same algorithm can be applied at lower resolutions too. Our multiresolution algorithm includes the following steps. First, given the number of classes and the associated parameters at the fine resolution, the GMRF parameters at lower resolutions are obtained by the local conditional distribution invariance approximation. Then, segmentation is performed at the coarsest resolution using Eq. (27) with the corresponding parameters and the results of segmentation are passed on to the immediate higher resolution. This is repeated until the fine resolution is reached. At each resolution a confidence measure is attached to the segmentation result at each pixel and propagated to the finer resolution. We address issues regarding confidence measures in this section.

5.1.1 Confidence Measures

After obtaining the segmentation result by ICM convergence at one resolution, the results have to be propagated to the immediate higher resolution. Since we obtain resolution transformation by subsampling, we have a quad tree type of graph. If $L^{(k)}$ is the segmentation result at the k th resolution, the labels in the $(k - 1)$ th level are initialized as:

$$L_s^{(k-1)} = L_{\lfloor s/2 \rfloor}^{(k)}. \quad (28)$$

In addition, at level k , after the ICM converges, we attach a confidence measure $C_s^{(k)}$ to the segmentation result obtained at site s .

At level k , after the convergence of ICM iterations, let \dot{v} and \ddot{v} be such that,

$$\begin{aligned} \dot{v}_s &= \arg \max_{v \in \{1, 2, \dots, V\}} P(X_s | L_s = v, X_{s+r}, r \in \eta) P(L_s = v | L_{s+r}, r \in \psi) \\ \ddot{v}_s &= \arg \max_{v \in \{1, 2, \dots, V\}, v \neq \dot{v}_s} P(X_s | L_s = v, X_{s+r}, r \in \eta) P(L_s = v | L_{s+r}, r \in \psi) \end{aligned}$$

and the confidence measure is defined as,

$$C_s^{(k)} = \frac{P(X_s | \dot{v}_s, X_{s+r}) P(\dot{v}_s | L_{s+r})}{P(X_s | \ddot{v}_s, X_{s+r}) P(\ddot{v}_s | L_{s+r})}. \quad (29)$$

These confidence measures at level k are propagated upwards to level $k - 1$ in the same manner as in Eq. (28). At level k , ICM is restricted to only those pixels with the confidence measure such that,

$$\frac{1}{C_s^{(k)}} \geq c^{(k)} \quad (30)$$

where $c^{(k)}$ is a confidence threshold at level k . Also from the definition,

$$0 < \frac{1}{C_s^{(k)}} \leq 1.0.$$

For the coarsest resolution $c^{(1)} = 0$, i.e., ICM is performed over all sites in the lattice.

The confidence measure defined in Eq. (29) has a hypothesis testing interpretation. The C_s expression is exactly the same as the hypothesis test ratio to compare the hypotheses that the label at site s is \dot{v} or \ddot{v} . This interpretation can be used in the following ways:

1. In an application, if misclassification between textures v_1 and v_2 results in a higher misclassification cost, say a ($a > 1$), compared to other types of misclassification which have a cost of 1, then, at the end of ICM iteration, if $\dot{v}_s = v_1$ and $\ddot{v}_s = v_2$ or vice versa, the confidence measure C_s at that site can be replaced by $\frac{C_s}{a}$ to reflect the higher misclassification cost.

2. Let the parameters corresponding to different textures at a particular resolution be such that the discrimination between some textures is low, while the rest of the textures can be easily discriminated. In such cases, if \check{v} and $\check{\check{v}}$ belong to the set of textures that have low discrimination, then the result of classification at that site is not reliable and hence the C_s can be set to zero.

5.2 MRF on Resolution Transformation

We have already addressed the issue of modeling GMRFs at lower resolutions. The label field is modeled by an MRF. This MRF also loses the Markov property under resolution transformation and can be approximated by a Markov field at lower resolutions. For the pairwise interaction model there is only one parameter β to be estimated in Eq. (23). However, it is hard to find approximations as we did in the case of Gaussian fields. Fortunately, segmentation results are not heavily dependent on this parameter. Therefore, we have chosen $\beta = \{0.5, 0.3, 0.15\}$ for different resolutions with the smaller values used at coarser resolutions. We have experimented with different sets of values of β and found that parameter the of the label MRF does not have a great bearing on the segmentation, hence a rigorous estimation may not be necessary.

6 Experiments

We present experimental results with simulated, Brodatz texture images and real satellite images to show that the multiresolution algorithms perform better than the single resolution both in terms of the classification accuracy and computational requirements. In all the experiments, the confidence threshold $c^{(k)} = \{0.6, 0.25, 0.0\}$, is used for the different levels with smaller values used at coarser resolutions. Multiresolution results presented in this section are obtained by performing the algorithm over three resolutions. In all cases, percentages of correct classification are reported.

6.1 Synthetic Image

We generated texture images using the technique given in [4]. Three third order GMRF textures are generated with parameters $\{ (\theta_{(1,0)} = 0.0934154, \theta_{(0,1)} = 0.520252, \theta_{(1,1)} = 0.0303413, \theta_{(-1,1)} = 0.0180476, \theta_{(2,0)} = -0.0216434, \theta_{(0,2)} = -0.148331), \sigma^2 = 0.9342 \}$, $\{\underline{\theta} = (0.308257, 0.468389, -$

0.0755398, -0.0755797, -0.0407557, -0.100678), $\sigma^2 = 1.8472$ }, $\{\theta = (0.406875, 0.423393, -0.178478, -0.188702, -0.0649544, -0.121439), \sigma^2 = 1.264811$ }. Figure 5(a) shows the composite image with these three textures. Figure 5(b) shows the single resolution segmentation result and Figure 5(c) shows the result for multiresolution segmentation.

6.2 Brodatz Images

We have tested our algorithm on textures from the Brodatz texture album. Figure 6(a) contains grass, calf leather, wool, and wood textures. The original GMRF parameters are estimated by maximum likelihood estimation. Figure 6(b) shows the single resolution segmentation and Figure 6(c) shows the multiresolution segmentation. We have another interesting plot of $\frac{1}{C_s^{(k)}}$ for the level $k = 1$ in Figure 6(d). The brighter points in this image correspond to points of low confidence measure. As expected, all the boundary regions between different textures have low confidence measures. In texture segmentation, classification near the texture boundaries is usually more ambiguous.

We also experimented with another more complex set of textures from the Brodatz album. Figure 7(a) shows a three class (pigskin, raffia, and water) texture image. Figure 7(b) shows the single resolution segmentation and Figure 7(c) shows the multiresolution segmentation. Figure 7(d) shows the corresponding confidence measure plot.

6.3 Real Images

Figure 8(a) shows a section of a single channel of a multispectral sensor (MSS) image over Africa. We chose three classes corresponding to river, forest, and deforestation. The GMRF parameters obtained from small sections of a different part of the image are used to classify the image shown. Unfortunately exact class maps are not available. Figure 8(b) shows the single resolution result and Figure 8(c) shows the multiresolution result. Clearly, we can see that the multiresolution algorithm has performed better, with lesser computation, than the single resolution algorithm.

Table 2 shows the comparison between single and multiresolution algorithms in terms of the classification accuracy and the number of computation units required. To compare the computational requirements between the single resolution and multiresolution approaches, we define a unit of computation to be the computation required to perform ICM at a single pixel site.

Image	Resolution			
	Single		Multiple	
	Classification %	Computation	Classification %	Computation
Synthetic	89.84	2686976	96.75	431031
Brodatz1	86.04	1114112	92.75	679444
Brodatz2	68.11	2490368	84.54	710980
Africa	–	2160000	–	426105

Table 2: Comparison of the performances of single and multiresolution algorithms.

Finally, we present results of multiresolution segmentation on a thematic mapper (TM) image consisting of four classes corresponding to river, forest, deforestation and regrowth. Figure 9(a) shows a section of thematic mapper (TM) data and Figure 9(b) shows the 4-class multiresolution segmentation result.

7 Summary

Multiresolution models and algorithms play an important role in image analysis. These algorithms not only help to reduce the computational time, but also help to analyze the given information at different spatial scales. We have presented techniques based on minimizing KL distances, to estimate the parameters of GMRFs at coarser resolutions and have used it for texture segmentation application. GMRFs are widely used in many image processing applications including restoration, segmentation, compression, etc., and the proposed models can be used for these applications.

Also, this can be extended to perform unsupervised texture segmentation. However, as mentioned in Section 4, GMRF parameters at a lower resolution can correspond to more than one set of parameters at fine resolution. Hence the problem of retrieving the GMRF parameters at fine resolution given the parameters at coarse resolution has to be addressed for unsupervised segmentation.

Appendix

By observing Eq. (22) it can be inferred that the above four sets of parameters will result in the same $S_x^{(1)}(\omega)$. The constraints we have obtained on $(\theta_{(1,0)}, \theta_{(0,1)})$ for the first order case are also valid here.

Case a: Let $(\theta_{(1,0)}, \theta_{(0,1)}, \theta_{(1,1)}, \theta_{(-1,1)})$, and $(\theta_{(1,0)}, \theta_{(0,1)}, \theta_{(1,1)}, \theta_{(-1,1)})$ result in the same power spectrum at $\Omega^{(1)}$. The goal is to find all possible values of $\theta_{(1,1)}$ and $\theta_{(-1,1)}$ that satisfy the above condition. Let (ω_1, ω_2) be such that,

$$\frac{2\pi\omega_1}{M} + \frac{2\pi\omega_2}{N} = \frac{\pi}{2} \text{ or } \frac{3\pi}{2}, \quad \omega_1 \neq 0, \omega_2 \neq 0. \quad (31)$$

Let $\cos(\frac{2\pi\omega_1}{M} - \frac{2\pi\omega_2}{N}) = \gamma$, $2(\theta_{(1,0)} \cos(\frac{2\pi\omega_1}{M}) + \theta_{(0,1)} \cos(\frac{2\pi\omega_2}{N})) = \alpha$ and $2(\theta_{(1,0)} \cos(\frac{2\pi\omega_1}{M}) - \theta_{(0,1)} \cos(\frac{2\pi\omega_2}{N})) = \beta$.

$$\begin{aligned} S_x^{(1)}(\cdot, \cdot) &= \frac{1}{4} \left[\frac{1}{1 - \alpha - 2\gamma\theta_{(-1,1)}} + \frac{1}{1 - \beta + 2\gamma\theta_{(-1,1)}} \right. \\ &\quad \left. + \frac{1}{1 + \beta + 2\gamma\theta_{(-1,1)}} + \frac{1}{1 + \alpha - 2\gamma\theta_{(-1,1)}} \right] \end{aligned} \quad (32)$$

$$= \frac{1}{2} \left[\frac{1 - 2\gamma\theta_{(-1,1)}}{(1 - 2\gamma\theta_{(-1,1)})^2 - \alpha^2} + \frac{1 + 2\gamma\theta_{(-1,1)}}{(1 + 2\gamma\theta_{(-1,1)})^2 - \beta^2} \right] \quad (33)$$

Observe that $(1 - 2\gamma\theta_{(-1,1)}) > 0$, $(1 + 2\gamma\theta_{(-1,1)}) > 0$, $(1 - 2\gamma\theta_{(-1,1)})^2 - \alpha^2 > 0$ and $(1 + 2\gamma\theta_{(-1,1)})^2 - \beta^2 > 0$, because all four terms in Eq. (32) are positive.

The first and second order derivative of Eq. (33) can be shown to be as follows:

$$\begin{aligned} \frac{dS_x(\cdot, \cdot)}{d\theta_{(-1,1)}} &= \frac{\alpha^2\gamma + \gamma(1 - 2\gamma\theta_{(-1,1)})^2}{[(1 - 2\gamma\theta_{(-1,1)})^2 - \alpha^2]^2} + \frac{-\beta^2\gamma - \gamma(1 + 2\gamma\theta_{(-1,1)})^2}{[(1 + 2\gamma\theta_{(-1,1)})^2 - \beta^2]^2} \\ \frac{d^2S_x(\cdot, \cdot)}{d\theta_{(-1,1)}^2} &= \frac{12\alpha^2\gamma^2(1 - 2\gamma\theta_{(-1,1)}) + 4\gamma^2(1 - 2\gamma\theta_{(-1,1)})^3}{[(1 - 2\gamma\theta_{(-1,1)})^2 - \alpha^2]^3} \\ &\quad + \frac{12\beta^2\gamma^2(1 + 2\gamma\theta_{(-1,1)}) + 4\gamma^2(1 + 2\gamma\theta_{(-1,1)})^3}{[(1 + 2\gamma\theta_{(-1,1)})^2 - \beta^2]^3} \end{aligned} \quad (34)$$

It can be seen that the second derivative is always positive. We have shown that $S_x^{(1)}(\cdot, \cdot)$ is a convex function of $\theta_{(-1,1)}$. For the set of (ω_1, ω_2) under consideration, the power spectrum is independent of $\theta_{(1,1)}$. Because of the strict convexity of the power spectrum, if $(\theta_{(1,0)}, \theta_{(0,1)}, \cdot, \theta_{(-1,1)})$, and $(\theta_{(1,0)}, \theta_{(0,1)}, \cdot, \theta_{(-1,1)})$ have the same power spectrum at a given ω_1 and ω_2 , then there can

be a maximum of only one value for $\theta_{(-1,1)}$. However, from Eq. (33) it can be seen that this value will not be independent of ω_1 and ω_2 . Therefore, there does not exist any other value of $\theta_{(-1,1)} \neq \theta_{(-1,1)}$, such that the power spectrums are equal at all (ω_1, ω_2) in Eq.(31). The same can be shown for $\theta_{(1,1)}$ by taking all (ω_1, ω_2) such that $\frac{2\pi\omega_1}{M} - \frac{2\pi\omega_2}{N} = \frac{\pi}{2}$.

Remark 1: If $(\theta_{(1,0)}, \theta_{(0,1)}, \theta_{(1,1)}, \theta_{(-1,1)})$ and $(\theta_{(1,0)}, \theta_{(0,1)}, \theta_{(1,1)}, \theta_{(-1,1)})$ result in the same power spectrum at $\Omega^{(1)}$ then, $\theta_{(1,1)} = \theta_{(1,1)}$ and $\theta_{(-1,1)} = \theta_{(-1,1)}$. (Note: If $\theta_{(1,0)} = 0$ and $\theta_{(0,1)} = 0$, then $\theta_{(1,1)} = -\theta_{(1,1)}$ and $\theta_{(-1,1)} = -\theta_{(-1,1)}$ can result in the same power spectrum at $\Omega^{(1)}$. However, here there is no distinction between Case a and Case b.)

Case b: Let $(\theta_{(1,0)}, \theta_{(0,1)}, \theta_{(1,1)}, \theta_{(-1,1)})$, and $(-\theta_{(1,0)}, \theta_{(0,1)}, \theta_{(1,1)}, \theta_{(-1,1)})$ result in the same power spectrum at $\Omega^{(1)}$. The goal is to find all possible values of $\theta_{(1,1)}$ and $\theta_{(-1,1)}$. For a given (ω_1, ω_2) that satisfy Eq. (31), the power spectrum with parameters $(-\theta_{(1,0)}, \theta_{(0,1)}, \theta_{(1,1)}, \theta_{(-1,1)})$ is a strict convex function of $\theta_{(-1,1)}$ (Eq. (34)). Therefore, there can be a maximum of two different values of $\theta_{(-1,1)}$. One of them is $-\theta_{(-1,1)}$ from our claim. It follows from the Remark 1, that there does not exist another value of $\theta_{(-1,1)}$, such that $(-\theta_{(1,0)}, \theta_{(0,1)}, \dots, -\theta_{(-1,1)})$ and $(-\theta_{(1,0)}, \theta_{(0,1)}, \dots, \theta_{(-1,1)})$ have the same power spectrum at $\Omega^{(1)}$. So, the only possible value for $\theta_{(-1,1)}$ is $-\theta_{(-1,1)}$. The same argument holds good for $\theta_{(1,1)}$, when we use the set of all (ω_1, ω_2) such that $\frac{2\pi\omega_1}{M} - \frac{2\pi\omega_2}{N} = \frac{\pi}{2}$. Therefore, if $(\theta_{(1,0)}, \theta_{(0,1)}, \theta_{(1,1)}, \theta_{(-1,1)})$ and $(-\theta_{(1,0)}, \theta_{(0,1)}, \theta_{(1,1)}, \theta_{(-1,1)})$ result in the same power spectrum at $\Omega^{(1)}$, then $\theta_{(1,1)} = -\theta_{(1,1)}$ and $\theta_{(-1,1)} = -\theta_{(-1,1)}$.

Similar argument can be presented for the other two sets of parameters.

References

- [1] J. Besag, "On the Statistical Analysis of Dirty Pictures," *Journal of the Royal Statistical Society*, Vol. 48, pp. 259–302, 1986.
- [2] C. Bouman and B. Liu, "Multiple Resolution Segmentation of Textured Images," *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 13, pp. 99–113, Feb. 1991.
- [3] S. Chatterjee and R. Chellappa, "Maximum Likelihood Texture Segmentation Using Gaussian Markov Random Field Models," *IEEE International Conf. on Acoustics, Speech, and Signal Processing*, June 1985.

- [4] R. Chellappa, “Two-dimensional Discrete Gaussian Markov Random Field Models for Image Processing,” in *Progress in Pattern Recognition* (L. N. Kanal and A. Rosenfeld, eds.), pp. 79–112, Elsevier, 1985.
- [5] P. C. Chen and T. Pavlidis, “Image Segmentation as an Estimation Problem,” *Computer Vision, Graphics, and Image Processing*, Vol. 12, pp. 153–173, 1980.
- [6] F. S. Cohen and D. B. Cooper, “Simple Parallel Hierarchical and Relaxation Algorithms for Segmenting Noncausal Markovian Random Fields,” *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 9, pp. 195–219, March 1987.
- [7] G. R. Cross and A. K. Jain, “Markov Random Field Texture Models,” *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 5, pp. 25–39, Jan 1983.
- [8] L. S. Davis, S. Johns, and J. K. Aggrawal, “Texture analysis using generalized co-occurrence matrices,” *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 1, pp. 251–259, July 1979.
- [9] H. Derin and H. Elliot, “Modeling and Segmentation of Noisy and Textured Images Using Gibbs Random Field,” *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 9, pp. 39–55, Jan. 1987.
- [10] C. Domb and M. Green, “Phase Transition and Critical Phenomena.” Academic Press, 1979.
- [11] D. Geiger and F. Girosi, “Parallel and Deterministic Algorithms for MRFs: Surface Reconstruction,” *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 13, pp. 401–413, May 1991.
- [12] S. Geman and D. Geman, “Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images,” *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 6, pp. 721–741, Nov 1984.
- [13] S. Geman and C. Graffigne, “Markov Random Field Image Models and Their Application to Computer Vision,” *Proc. International Congress of Mathematicians*, pp. 1496–1517, 1986.
- [14] B. Gidas, “A Renormalization Group Approach to Image Processing,” *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 11, No. 2, pp. 164–180, 1989.
- [15] R. Haralick, “Statistical and Structural Approaches to Texture,” *Proc. IEEE*, Vol. 67, pp. 610–621, May 1979.

- [16] F. C. Jeng, "Subsampling of Markov Random Fields," *Jour. of Visual Communication and Image Representation*, Vol. 3, pp. 225–229, Sep. 1992.
- [17] R. L. Kashyap, "Analysis and Synthesis of Image Patterns by Spatial Interaction Models," in *Progress in Pattern Recognition* (L. N. Kanal and A. Rosenfeld, eds.), North-Holland, Amsterdam, 1981.
- [18] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *Ann. Math. Stat.*, Vol. 22, pp. 79–86, 1951.
- [19] H. Kunsch, "Thermodynamics and Statistical Analysis of Gaussian Random Fields," *Z. Wahr. Ver. Geb.*, pp. 407–421, Nov. 1981.
- [20] S. Lakshmanan and H. Derin, "Gaussian Markov Random Fields at Multiple Resolutions," in *Markov Random Fields: Theory and Applications* (R. Chellappa, ed.), pp. 131–157, Academic Press, 1993.
- [21] S. Lu and S. Fu, "Stochastic Tree Grammar Inference for Texture Synthesis and Discrimination," *Computer Vision, Graphics, and Image Processing*, Vol. 9, pp. 234–245, 1979.
- [22] M. R. Luetzgen, W. C. Karl, A. S. Willsky, and R. R. Tenney, "Multiscale Representations of Markov Random Fields," *IEEE Trans. on Signal Processing*, Vol. 41, pp. 3377–3397, Dec. 1993.
- [23] B. S. Manjunath, T. Simchony, and R. Chellappa, "Stochastic and Deterministic Networks for Texture Segmentation," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 38, pp. 1039–1049, June 1990.
- [24] I. Matsuba, "Renormalization Group Approach to Hierarchical Image Analysis," *IEEE International Conf. on Acoustics, Speech, and Signal Processing*, pp. 1044–1047, 1988.
- [25] P. A. P. Moran and J. Besag, "On the Estimation and Testing of Spatial Interaction in Gaussian Lattice," *Biometrika*, Vol. 62, pp. 552–562, 1975.
- [26] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Book Company, 1965.

- [27] P. Perez and F. Heitz, "Multiscale Markov Random Fields and Constrained Relaxation in Low Level Image Analysis," *IEEE International Conf. on Acoustics, Speech, and Signal Processing*, Vol. 3, pp. 61–64, 1992.
- [28] A. Rosenfeld, "Visual Texture Analysis," Tech. Rep. 70-116, University of Maryland, June 1970.
- [29] A. Rosenfeld, "Some Useful Properties of Pyramids," in *Multiresolution Image Processing and Analysis* (A. Rosenfeld, ed.), Springer-Verlag, 1984.
- [30] G. Sharma and R. Chellappa, "Two-Dimensional Maximum Entropy Power Spectra," *IEEE Transaction on Information Theory*, Vol. IT-31, pp. 90–99, Jan 1985.
- [31] D. Terzopoulos, "Image Analysis using Multigrid Relaxation Methods," *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 8, pp. 129–139, March 1986.
- [32] F. Tomita, "Description of texture by a structural analysis," *Proc. IJCAI*, 1979.
- [33] J. W. Woods, "Two-dimensional Discrete Markovian Fields," *IEEE Transaction on Information Theory*, Vol. 18, pp. 232–240, 1972.
- [34] C. H. Wu and P. C. Doerschuk, "Tree Approximations to Markov Random Fields," *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 17, pp. 391–343, Apr. 1995.
- [35] J. Zhang, "The Mean Field Theory in EM Procedures for Markov Random Fields," *IEEE Trans. on Signal Processing*, Vol. 40, pp. 2570–2583, Oct. 1992.

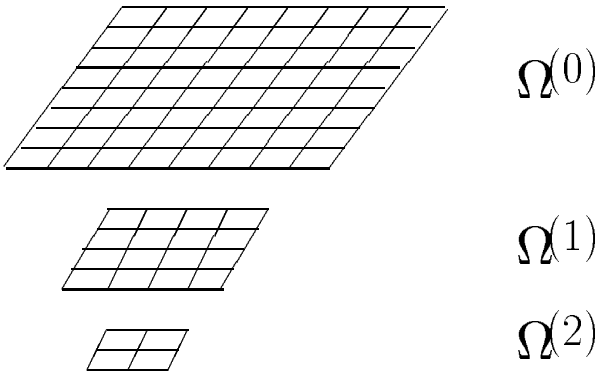
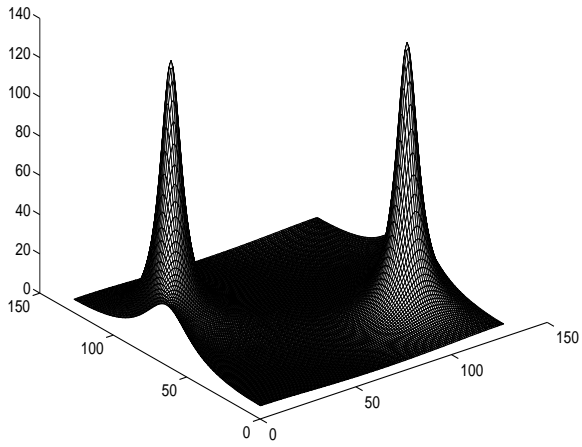


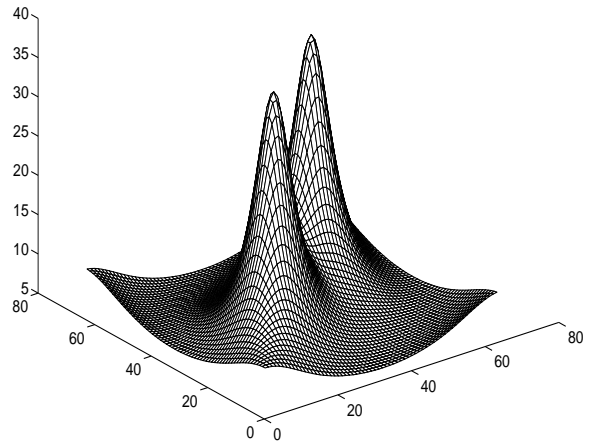
Figure 1: Resolution transformation.

5	4	3	4	5
4	2	1	2	4
3	1	s	1	3
4	2	1	2	4
5	4	3	4	5

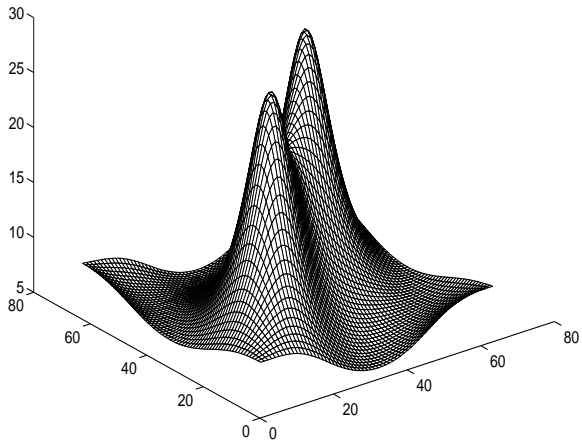
Figure 2: Neighbor sets for different MRF orders [7]



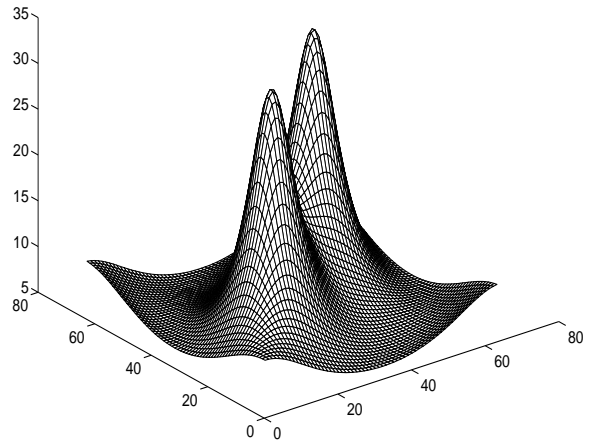
(a)



(b)

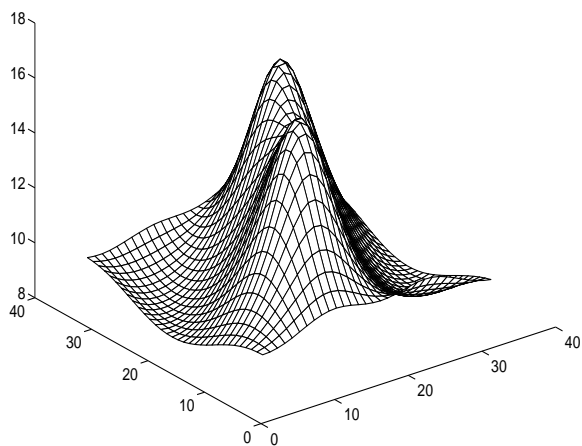


(c)

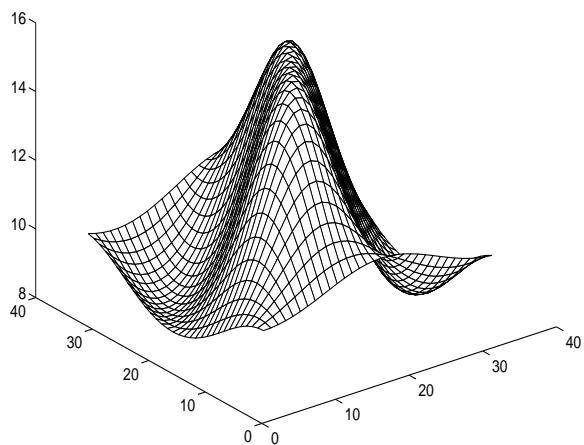


(d)

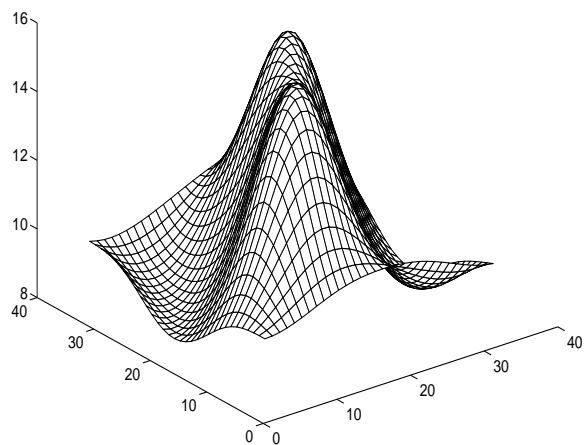
Figure 3: (a) $S_x^{(0)}(\omega)$ – Power spectrum of GMRF at $\Omega^{(0)}$, (b) $S_x^{(1)}(\omega)$ – Exact non-Markov power spectrum at $\Omega^{(1)}$, (c) $MS_x^{(1)}(\omega, 3)$ – Power spectrum of third order Markov approximation at $\Omega^{(1)}$, (d) $MS_x^{(1)}(\omega, 4)$ Power spectrum of fourth order Markov approximation at $\Omega^{(1)}$.



(a)

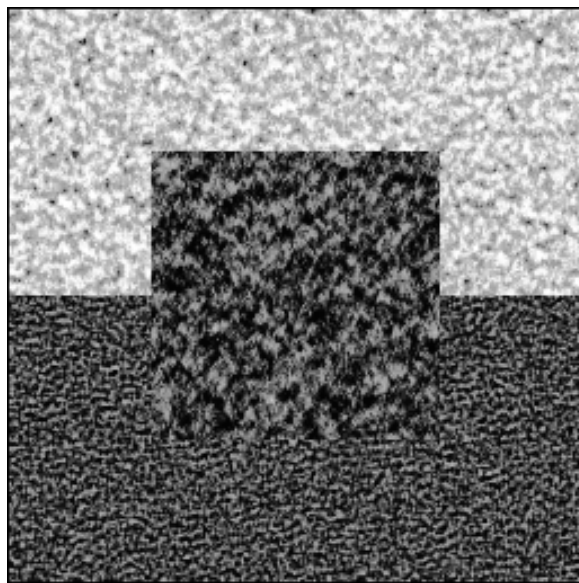


(b)

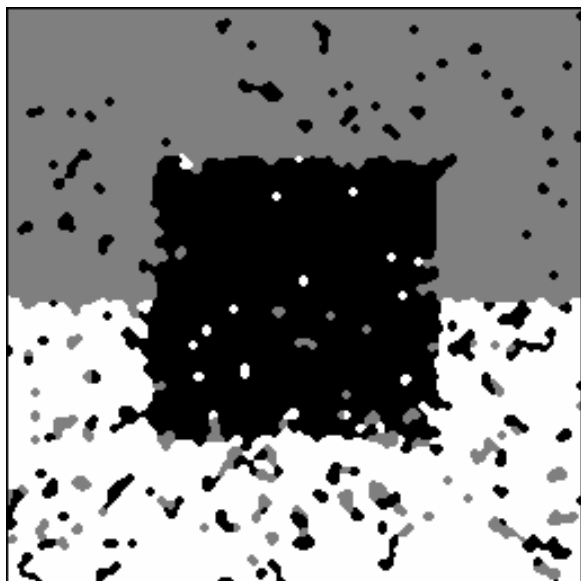


(c)

Figure 4: (a) $S_x^{(2)}(\omega)$ – Exact non-Markov power spectrum at $\Omega^{(2)}$, (b) $MS_x^{(2)}(\omega, 2)$ – Power spectrum of second order Markov approximation at $\Omega^{(2)}$, (c) $MS_x^{(2)}(\omega, 3)$ – Power spectrum of third order Markov approximation at $\Omega^{(2)}$.



(a)

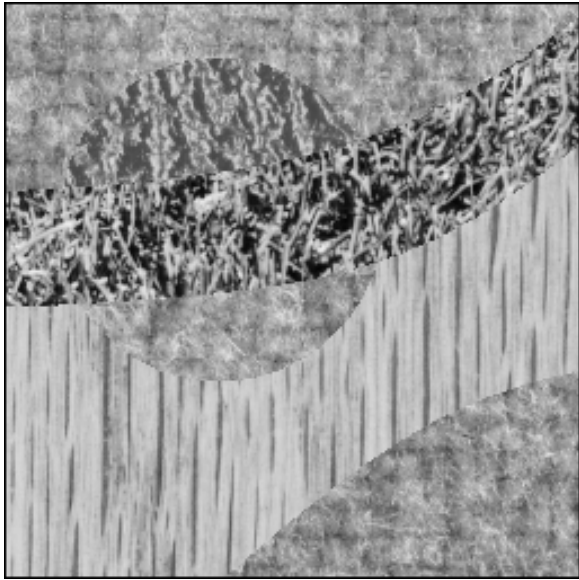


(b)

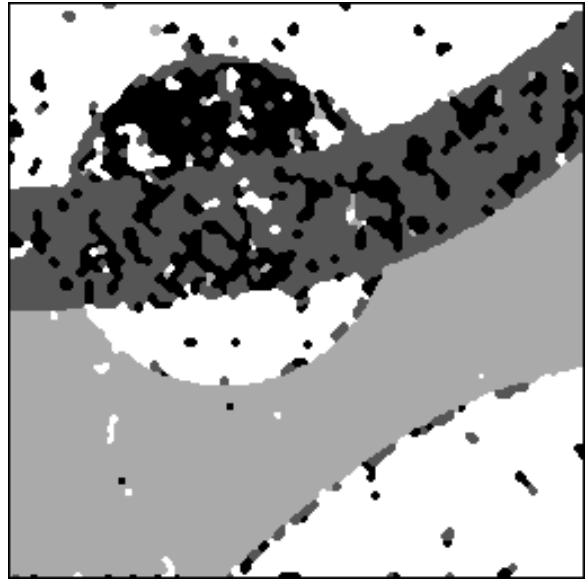


(c)

Figure 5: (a) Synthetic texture image, (b) Single resolution segmentation result, (c) Multiresolution segmentation result.



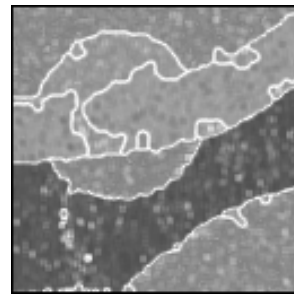
(a)



(b)

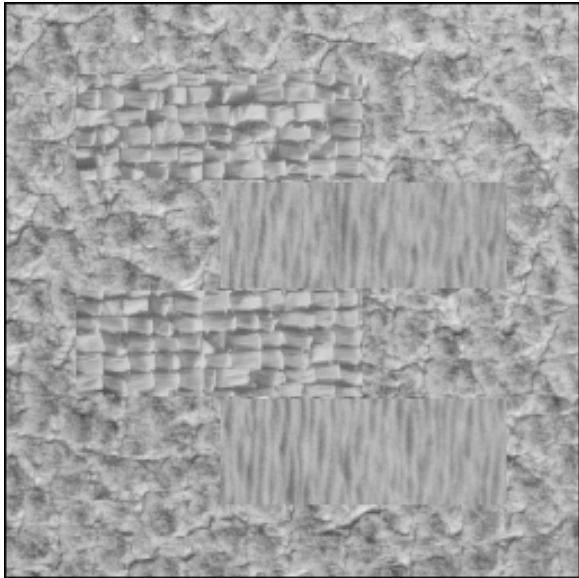


(c)

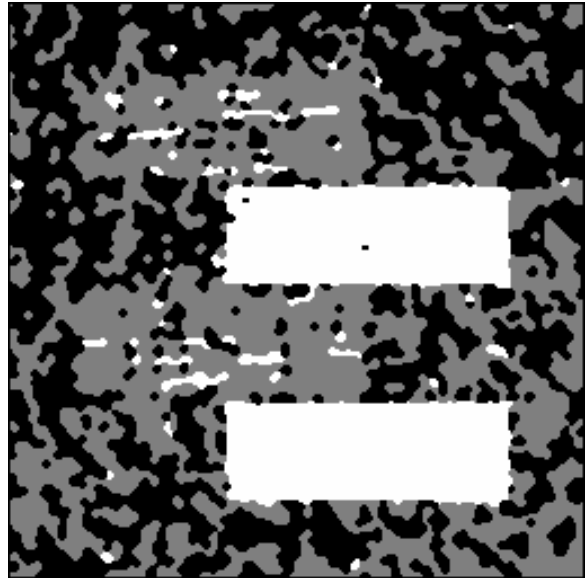


(d)

Figure 6: (a) Brodatz texture image, (b) Single resolution segmentation result, (c) Multiresolution segmentation result, (d) Confidence measures.



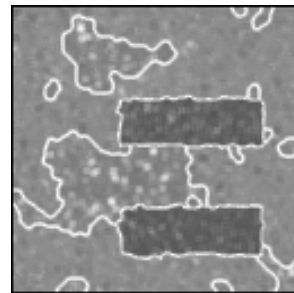
(a)



(b)



(c)

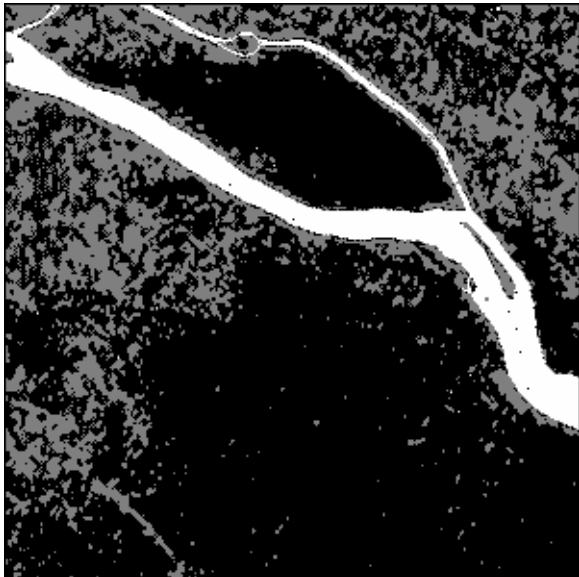


(d)

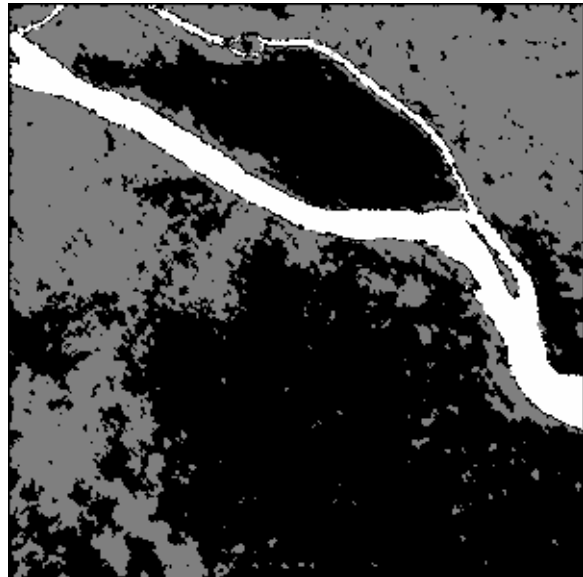
Figure 7: (a) Brodatz texture image, (b) Single resolution segmentation result, (c) Multiresolution segmentation result, (d) Confidence measures.



(a)

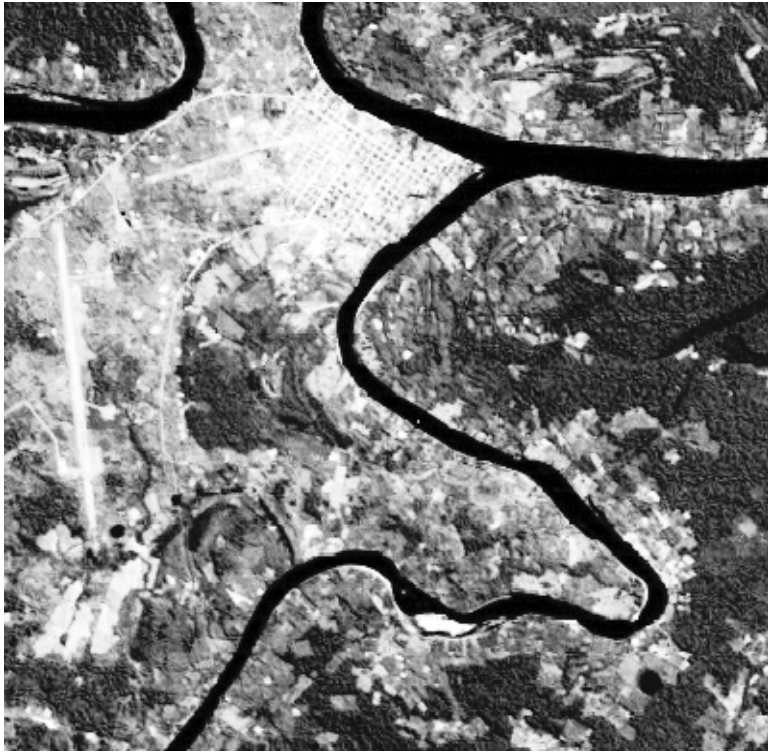


(b)

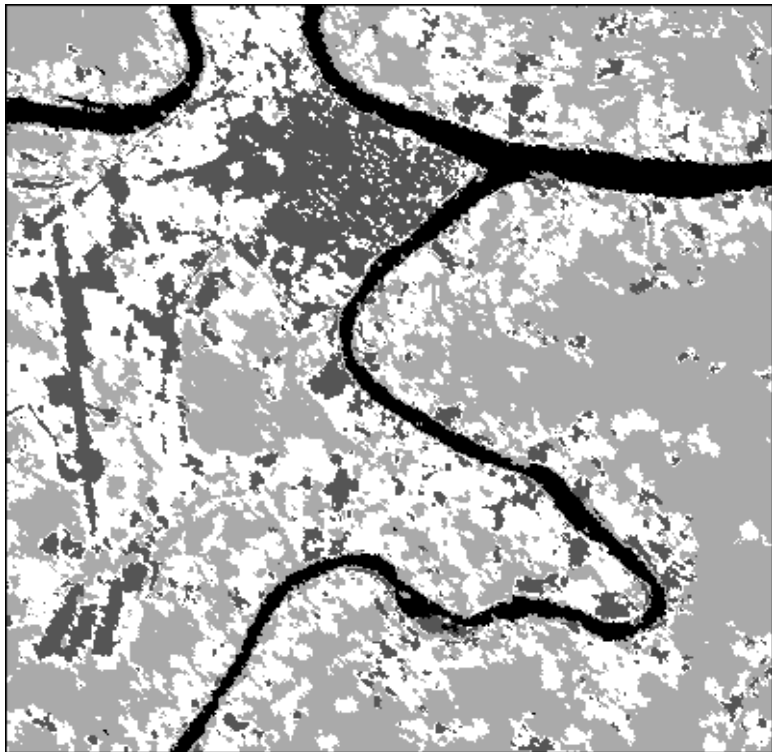


(c)

Figure 8: (a) Remotely sensed MSS image, (b) Single resolution segmentation result, (c) Multiresolution segmentation result.



(a)



(b)

Figure 9: (a) Remotely sensed TM image, (b) Multiresolution segmentation result.