1 Introduction

There has been an increasing emphasis in using statistical techniques for modeling and processing images in the image analysis community. Among the several possible 2-D models for images most of the research has been restricted to Markov random field **(MRF)** models, rightly so, because of the local statistical dependence of images. MRF models have been used to characterize prior beliefs about various image features such as textures, edges, region labels etc. Also, establishment of the connection between MRF models and the Gibbs distribution has resulted in the influx of many statistical mechanical techniques like simulated annealing [19], mean field methods [14] in image processing. Problems in both statistical mechanics and image processing involve a large number of micro states (pixels), with their local interactions deciding the properties of the macro state (image) and the complexity of the problem arises from the fact that any desired optimization has to be performed over a large number of states. 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. In the rest of this section, we provide a brief review of literature on texture segmentation and multiresolution techniques.

Texture classification and segmentation problems have been addressed by several authors with a wide range of 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 [27], Lu and Fu [23] and Tomita [31]. 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 [18] suggested various local statistical measures based on the gray level dependence, later generalized by Davis, et.al. [11].

More complex statistical models using MRFs [32] emerged later and have been used in various image processing problems. Cross and Jain [10] and Chellappa [6] have shown the applicability of GMRF models to synthesize textures. Chellappa and Chatterjee [5] have used GMRF models for maximum likelihood segmentation of textures. Geman and Graffigne [16] 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 [12] used Gibbs distributions and presented a non-optimal method using dynamic programming. Manjunath et.al. [24] proposed a stochastic learning technique to improve the ICM results and presented a neural network implementation for texture segmentation.

The main drawback of the MRF techniques is that the minimization schemes associated with the energy functions are iterative and are usually computationally expensive. Best results are obtained by using simulated annealing [15] which optimizes the MAP criterion, but is computationally very taxing. There are two different approaches that have been used to ease the computational burden. The first is to use non-optimal, deterministic methods that converge to a local minima, but still provide reasonably good results [1], [14]. 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 [28]. Multigrid methods proposed by Terzopoulos [30], applied to some computer vision problems substantially reduced the computation. Chen and Pavlidis [7] used a hierarchical approach to texture segmentation, but did not directly use MRFs. Bouman and Liu [3] used a Gaussian causal autoregressive model and a quad-tree structure to perform multiresolution segmentation. Gidas [17] investigated the multiresolution MRF framework to process images and showed some connection between these ideas and similar ones in the renormalization group studies in statistical mechanics [13]. A hierarchical image analysis scheme based on renormalization group ideas was presented by Matsuba in [25].

Cohen and Cooper [8] presented a hierarchical scheme, where the data at lower resolution is divided into blocks and the conditional probabilities of the blocks given the neighboring blocks were obtained. They used a recursive scheme to calculate the within-block and between-block interaction values. They did not try to model the data at lower resolution by GMRFs. Lakshmanan and Derin [22] used covariance invariance approximation to approximate the lower resolution data by GMRFs. This GMRF approximation has the property that in the associated covariance matrix the entries corresponding to pairs of sites that are neighbors are equal to the entries in the covariance matrix associated with the exact non-Markov density. Many interesting properties of this estimator such as maximizing entropy, minimizing Kullback-Leibler **(KL)** distance can be found in [22].

In this paper, we present two schemes to estimate the parameters of GMRFs 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 (conditional relative entropy) between the conditional densities. We also show that the computations for these two estimators turn out to be similar to traditional maximum likelihood [26], [21], [29] and pseudo likelihood estimators [2] for GMRF parameters except that the sample covariances are replaced by covariances calculated with respect to the exact non-Markov measure. We also present results on the existence of different sets of GMRF parameters at fine resolution which on subsampling results in statistically identical coarser resolution processes. As an application, we have chosen the texture segmentation problem and performed segmentation over multiple resolution using the multiresolution GMRF model. The coarsest resolution data is initially segmented and the results of segmentation along with a confidence measure is passed on the immediate higher resolution and so on, until the fine resolution data is segmented. We have shown that the multiresolution technique performs better than 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 discusses the 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 Markovianity under resolution transformation.

2.1 The GMRF Model

The GMRF models can be used on a finite, infinite lattice or on any general graph. Here we introduce the aspects GMRF models on a finite lattice. Let $\mathbf{\Omega^{(0)}} = \{(i,j) : 0 \leq i \leq M-1, 0 \leq j \leq N-1\}$ be a rectangular lattice. The superscript stands for the level in the image pyramid, $\mathbf{\Omega^{(0)}}$ being the lattice at the fine resolution, $\mathbf{\Omega^{(k)}}$ represents the lattice which is obtained by subsampling $\mathbf{\Omega^{(0)}}$, k times (Figure 1). Let $X^{(k)}$ represent a random vector, obtained by ordering the random variables on the two-dimensional lattice $\mathbf{\Omega^{(k)}}$, through a row-wise scan. The elements of $\mathbf{\Omega^{(k)}}$ are indexed by s and t, where $s = (s_1, s_2)$. If $X^{(0)}$ is modeled by a GMRF, then the joint probability density function of $X^{(0)}$ can be written as follows:

$$P^{(0)}(X^{(0)} = x) = \frac{1}{(2\pi)^{\frac{MN}{2}}(\det \Sigma^{(0)})^{\frac{1}{2}}} \exp\{-\frac{1}{2}x^T[\Sigma^{(0)}]^{-1}x^T\}$$

where $\Sigma^{(0)}$ is the covariance matrix of $X^{(0)}$.

Equivalently, the process $X^{(0)}$ can be written in terms of a non-causal interpolative process. For a site s, let $\eta^{(0)}$ be the symmetric neighborhood that contains the set of sites that are chosen to be the neighbors of $X_s^{(0)}$ and let $\theta_r^{(0)} = \theta_{-r}^{(0)}$. We always use r to index into the neighbor set in the rest of the paper.

$$X_s^{(0)} = \sum_{r \in \eta^{(0)}} \theta_r^{(0)} x_{s+r}^{(0)} + e_s^{(0)}$$

where $e_s^{(0)}$, is zero mean, Gaussian noise, with autocorrelation given by:

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

Hence the GMRF process can be completely characterized by the set of parameters $\{\underline{\theta}^{(0)}, [\sigma^{(0)}]^2\}$. Also $X_s^{(0)}$ exhibits the Markov property,

$$\begin{aligned} P^{(0)}(X_s^{(0)}/X_t^{(0)}, \forall t \neq s, t \in \Omega^{(0)}) &= P^{(0)}(X_s^{(0)}/X_{s+r}^{(0)}, r \in \eta^{(0)}) \\ &= \frac{1}{\sqrt{2\pi[\sigma^{(0)}]^2}} \exp\{-\frac{[x_s^{(0)} - \sum_{r\in\eta^{(0)}} \theta_r^{(0)} x_{s+r}^{(0)}]^2}{2[\sigma^{(0)}]^2}\} \end{aligned} \quad (1)$$

The power spectrum $S_\omega^{(0)}$ of the process $X^{(0)}$ can be shown to be:

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

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

2.2 Resolution Transformation

In this paper we restrict ourself to resolution transformation obtained by subsampling. 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 2x2 window. Other types of resolution transformation including Gaussian, Laplacian pyramids [4] have been used in image processing literature for various other problems.

The subsampling resolution transformation is defined as:

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

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

Equivalently,

$$X^{(k)} = D_0^k X^{(0)} \quad (3)$$

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

The resulting subsampled process $X^{(k)}$ is Gaussian, with covariance $\Sigma^{(k)} = [D_0^k]\Sigma^{(0)}[D_0^k]^T$. The power spectrum of $X^{(k)}$ can be shown to be [22]:

$$S_\omega^{(k)} = \frac{1}{2^{2k}} \sum_{r\in C_k} S_{\omega+r'}^{(0)} \quad (4)$$

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 this is similar to frequency domain aliasing due to sampling, in time series applications. It can be observed that $S_\omega^{(k)}$ cannot be written in the form of Eq. (2) with a finite neighborhood. Therefore, the subsampled process $X^{(k)}$ is non-Markov, except for the special case of second order separable correlation processes [22].

3 Markov Approximations

As mentioned in the last section, the GMRFs lose Markovianity under subsampling, the same is true for general MRFs. In the case of GMRFs, it is atleast possible to find the probability density function (pdf) of subsampled processes, whereas in the case of general MRFs it is not possible to obtain exact expressions for the pdf of subsampled process. However, if the lower 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 very good Markov approximations for coarser resolution processes. Cohen et.al [9] refer to the possibility of approximating rotated and scaled textures by GMRFs. In this section two different estimators to estimate 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, namely, the maximum likelihood and pseudo likelihood estimators.

3.1 Kullback-Leibler 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 $D(p \parallel q)$ [20], where q belongs to the family of GMRF pdfs. KL distance measure is widely used to obtain approximate probability measure with desired properties. The problem at hand, can be stated as follows:

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

$$q^* = \arg\min_q D(p \parallel q)$$

$$= \quad \arg\min_{q} \sum_{\underline{\mathrm{x}}} p(\underline{\mathrm{x}}) \log \frac{p(\underline{\mathrm{x}})}{q(\underline{\mathrm{x}})} \tag{5}$$

where,

$$q(\underline{\mathrm{x}}) = \frac{1}{(2\pi)^{\frac{MN}{2}} (\det \Sigma)^{\frac{1}{2}}} \exp\{-\frac{1}{2} \underline{\mathrm{x}}^T \Sigma^{-1} \underline{\mathrm{x}}\}$$

and the covariance matrix Σ is such that, the power spectrum is of the form in Eq. (2). As seen before, GMRFs can be completely characterized by $(\underline{\theta}, \sigma^2)$ and $q(\underline{\mathrm{x}})$ can be written in terms of these parameters.

The quadratic form $\underline{\mathrm{x}}^T \Sigma^{-1} \underline{\mathrm{x}}$ can be simplified as:

$$\underline{\mathrm{x}}^T \Sigma^{-1} \underline{\mathrm{x}} = C(0) - \underline{\theta}^T \underline{C}$$

where

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

and

$$\det \Sigma = \frac{\sigma^{MN}}{\prod_{s \in \Omega} (1 - \underline{\theta}^T \underline{\phi}_s)}, \quad where \ \ \underline{\phi}_s = cos \left((\frac{2\pi s_1}{M} \frac{2\pi s_2}{N}) \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \right) \forall s \in \Omega.$$

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

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

Rewriting Eq. (5):

$$\begin{aligned} (\underline{\theta}^*, [\sigma^2]^*) &= \arg\min_{(\underline{\theta}, \sigma^2)} \sum_{\underline{\mathrm{x}}} p(\underline{\mathrm{x}}) \log \frac{p(\underline{\mathrm{x}})}{q(\underline{\mathrm{x}})} \\ &= \arg\max_{(\underline{\theta}, \sigma^2)} \sum_{\underline{\mathrm{x}}} p(\underline{\mathrm{x}}) \log q(\underline{\mathrm{x}}) \\ &= \arg\max_{(\underline{\theta}, \sigma^2)} \sum_{\underline{\mathrm{x}}} p(\underline{\mathrm{x}}) [\sum_s \log(1 - \underline{\theta}^T \underline{\phi}_s) - \frac{MN}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (C(0) - \underline{\theta}^T \underline{C})] \\ &= \arg\max_{(\underline{\theta}, \sigma^2)} \sum_s \log(1 - \underline{\theta}^T \underline{\phi}_s) - \frac{MN}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (E_p[C(0)] - \underline{\theta}^T E_p[\underline{C}]) \end{aligned} \tag{6}$$

where $E_p(.)$ represents the expectation with respect to the p-measure.

$$\begin{aligned} E_p[C(r)] &= E_p[\sum_{s\in\Omega} X_s X_{s+r}] \\ &= (MN)E_p[X_s X_{s+r}]. \end{aligned} \tag{7}$$

Remark: Observe that Eq. (6) is very similar to the maximum likelihood expression, 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. Second, given the p - measure, we only need a few moment values $E_p[C(r)]$, followed by the maximization of Eq. (6) to obtain the Markov approximation. Lakshmanan and Derin [22] also remark that their covariance invariance estimator which also minimizes the KL distance 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 GMRF parameters of a non-Markov process based on a KL distance (conditional relative entropy) measure between local conditional distributions. In MRF applications all optimizations are performed based on the local conditional distribution, so, we believe an estimator based on that should be better 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 process. Let $\underline{Z}$ be a GMRF defined by $(\underline{\theta}, \sigma^2)$ with a neighborhood ψ. Then the linear estimate of Z_s based on the elements of ψ is given by:

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

and the mean square error

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

The conditional density $Pr(z_s/z_r, r \in \psi)$ is gaussian with conditional mean $\sum_{r\in\psi} \theta_r z_{s+r}$ and conditional variance $[\sigma]^2$.

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

$$q^*(x_s/x_{s+r}, r \in \eta) = \arg\min_q D[p(x_s/x_{s+r}, r \in \eta) \ \ || \ \ q(x_s/x_{s+r}, r \in \eta)], \tag{8}$$

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

Since $q(\underline{\mathrm{x}})$ belongs to the family of GMRF density, $q(x_s/x_{s+r})$ will be of the form in Eq. (1). The following results regarding the parameters $(\underline{\theta}^*, [\sigma^2]^*)$ corresponding to $q^*(\underline{\mathrm{x}})$ can be obtained from Eq. (8).

To simplify the notations, let $\underline{\mathrm{Y}}$ be the vector containing the neighborhood random variables in a proper order. For a first order neighborhood,

$$\underline{\mathrm{Y}}^T = \left(\begin{array}{cccc} X_{s+(1,0)} & X_{s-(1,0)} & X_{s+(0,1)} & X_{s-(0,1)}. \end{array} \right)$$

$$\begin{aligned} (\underline{\theta}^*, [\sigma^2]^*) &= \arg\min_{(\underline{\theta},\sigma^2)} \sum_{x_s,\underline{\mathrm{y}}} p(x_s, \underline{\mathrm{y}}) \log \frac{p(x_s/\underline{\mathrm{y}})}{q(x_s/\underline{\mathrm{y}})} \\ &= \arg\max_{(\underline{\theta},\sigma^2)} \sum_{x_s,\underline{\mathrm{y}}} p(x_s, \underline{\mathrm{y}}) \log q(x_s/\underline{\mathrm{y}}) \\ &= \arg\max_{(\underline{\theta},\sigma^2)} \sum_{x_s,\underline{\mathrm{y}}} p(x_s, \underline{\mathrm{y}})[-\frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(x_s - \sum_{r\in\eta}\theta_r y_r)^2] \\ &= \arg\min_{(\underline{\theta},\sigma^2)} \frac{1}{2}\log\sigma^2 + \frac{1}{2\sigma^2}(E_p[X_s - \sum_{r\in\eta}\theta_r Y_r]^2). \end{aligned} \tag{9}$$

It can be shown that the $\underline{\theta}^*$ parameters corresponding to $q^*(\underline{\mathrm{x}})$ are the coefficients of the best linear estimator and can be obtained as follows:

$$\underline{\theta}^* = \arg\min_{\underline{\alpha}} E_p[X_s - \sum_{r\in\eta}\alpha_r X_{s+r}]^2 \tag{10}$$

and using the $\underline{\theta}^*$ obtained, we can estimate the $[\sigma^2]^*$ as,

$$[\sigma^2]^* = E_p[X_s - \sum_{r\in\eta}\theta_r^* X_{s+r}]^2. \tag{11}$$

then,

$$\begin{aligned}\underline{\theta}^* &= \arg\min_{\underline{\alpha}} E_p[X_s - \underline{\alpha}^T \underline{\mathrm{Y}}]^2 \\ \underline{\theta}^{*T} &= E_p(X_s \underline{Y}^T)[E_p(\underline{Y}\underline{Y}^T)]^{-1} \end{aligned} \qquad (12)$$

and,

$$\begin{aligned}[\sigma^2]^* &= E_p(X_s^2) - E_p(X_s \underline{\mathrm{Y}}^T)[E_p(\underline{\mathrm{Y}}\underline{\mathrm{Y}}^T)]^{-1} E_p(X_s \underline{\mathrm{Y}}) \\ &= E_p(X_s^2) - [\underline{\theta}^*]^T E_p(X_s \underline{\mathrm{Y}}). \end{aligned} \qquad (13)$$

In addition, the estimated $\underline{\theta}^*$ parameters should satisfy the positivity conditions [6]:

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

Now, returning back to multiresolution discussion, let $X^{(0)}$ be a GMRF defined by $(\underline{\theta}^{(0)}, [\sigma^{(0)}]^2)$ and $X^{(k)}$ be the process obtained by subsampling $X^{(0)}$ k times. The non-Markov $X^{(k)}$ can be approximated by a GMRF by minimizing Eq. (8). The minimization requires the autocorrelation values $E(X_s^{(k)} X_{s+r}^{(k)})$ which can be computed, given the GMRF parameters for $X^{(0)}$ as shown below.

$$\begin{aligned} X_s^{(k)} &= X_{2^k s}^{(0)} \\ E(X_s^{(k)} X_{s+r}^{(k)}) &= E(X_{2^k s}^{(0)} X_{2^k(s+r)}^{(0)}) \end{aligned}$$

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

$$E(X_u^{(0)} X_v^{(0)}) = \frac{1}{MN} \sum_{s \in \Omega^{(0)}} \frac{(\lambda_M^{s_1 u_1} \lambda_N^{s_2 u_2})(\lambda_M^{s_1 v_1} \lambda_N^{s_2 v_2})}{1 - [\underline{\theta}^{(0)}]^T \underline{\phi}_s} \qquad (14)$$

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

Algorithm:

1. Compute the autocorrelations required in Eq. (12) using Eq. (14) and the inverse of the matrix $[E(\underline{Y}\underline{Y}^T)]^{-1}$.
2. Compute $\underline{\theta}^{(k)}$ using Eq. (12).

3. If the computed $\underline{\theta}^{(k)}$, satisfy the conditions $1 - [\underline{\theta}^{(k)}]^T \phi_s > 0$, for every s in $\Omega^{(k)}$, then compute $[\sigma^{(k)}]^2$ from Eq. (13) and stop.

4. If the conditions are not satisfied, then scale $\underline{\theta}^{(k)}$ to,

$$\underline{\theta}_n^{(k)} = \underline{\theta}^{(k)} \frac{0.99}{\max_{s \in \Omega^{(k)}} [\underline{\theta}^{(k)}]^T \phi_s}$$

5. Compute a direction $\underline{\vartheta}$ such that

$$[\nabla^2_{\underline{\theta}} f(\underline{\theta})]_{\underline{\theta}_n^{(k)}} \underline{\vartheta} = -[\nabla_{\underline{\theta}} f(\underline{\theta})]_{\underline{\theta}_n^{(k)}}$$

   where $f(\underline{\alpha}) = E_p[X_s^{(k)} - \sum_{r \in \eta^{(k)}} \alpha_r X_{s+r}^{(k)}]^2$

6. Find the largest m for a fixed β ($\beta < 1$), such that

$$\underline{\theta}_{n+1}^{(k)} = \underline{\theta}_n^{(k)} + \beta^m \underline{\vartheta}$$

   and $\underline{\theta}_{n+1}^{(k)}$ satisfy the positivity condition.

7. Go to step 5 if

$$\mid f(\underline{\theta}_n^{(k)}) - f(\underline{\theta}_{n+1}^{(k)}) \mid > \tau$$

Lemma: Under the assumption that the covariance matrix with respect to p - measure is positive definite, the function in Eq. (10) to be minimized is convex and is minimized over a convex set defined by $1 - [\underline{\theta}^{(k)}]^T \underline{\phi}_s > 0$, for every s in Ω. Hence the minimization can be performed by a gradient descent procedure.

Remarks:

1. If the $\underline{\theta}^*$ obtained from Eq. (12) satisfies the positivity conditions, ie., the algorithm terminates at step 3 and given that p is gaussian, then it can be shown that $p(x_s/x_{s+r}, r \in \eta) = q^*(x_s/x_{s+r}, r \in \eta)$, both conditional densities are gaussian with conditional mean $\sum_{r \in \eta} \theta_r^* x_{s+r}$ and conditional variance $[\sigma^2]^*$, thus the estimator *preserves the local conditional distribution.*

2. It is worth observing that solving Eqs. (10) and (11) is similar to the pseudo likelihood estimate [2] 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 covariance 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 subsampled process and the power spectrum associated with the GMRF approximation.

We show the results for $\underline{\theta}^{(0)} = \{0.2, -0.1, -0.25, 0.15\}$ and $[\sigma^{(0)}]^2 = 6.0$. Let $S_\omega^{(k)}$ be the exact power spectral density function at the k-th subsampled stage and let $MS_\omega^{(k)}(m)$ be the power spectrum for the m-th order Markov approximation. We calculate the normalized absolute difference $D(k,m) = \frac{\sum_{(\omega \in \Omega^{(k)})} |S_\omega^{(k)} - MS_\omega^{(k)}(m)|}{size(\Omega^{(k)})}$, 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 2 shows the power spectrum at the fine resolution calculated by Eq. (2) and Figure 3 shows the exact power spectrum (non-Markov)at the once subsampled resolution calculated by Eq. (4). Figure 4 shows the power spectrum for a third order Markov approximation with GMRF parameters obtained from the local conditional distribution invariance approximation and Figure 5 shows the same for a fourth order Markov approximation. Figure 6 shows the exact power spectrum (non-Markov) of twice subsampled process. Figure 7 shows the second order Markov approximation and Figure 8 shows the third order Markov approximation. From these figures and Table 1, it is easy to see that the power spectrum of the Markov approximations are very similar to the power spectrum of the exact non-Markov processes.

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

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 $D(k+1, m) \leq D(k, m)$, ie., 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 finest resolution is m, then an m or $m+1$ order GMRF approximation at the first level and an m order approximation at the subsequent levels results in very small values of the normalized absolute difference $D(k, m)$ and hence can be used as a good GMRF approximation.

4 Parameters Resulting in Identical PDFs at Lower Resolution

In the previous section, we suggested methods to approximate subsampled processes by GMRFs assuming that data at the fine resolution is modeled by a GMRF. In this process, it is necessary to analyze that if different GMRF parameters at the fine resolution can result in the same process (probabilistically) at the coarser resolutions. Since we are dealing with Gaussian processes, it suffices to check the covariance matrices of the subsampled processes instead of the pdf. However, the covariance elements are complicated functions of the parameters. Therefore, we look at the power spectrum of the subsampled processes 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 and hence the same estimated parameters of the GMRF approximation.

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

The first order GMRF model is defined by the parameters $(\theta_{(1,0)}, \theta_{(0,1)}, [\sigma^{(0)}]^2)$ and the power spectral density function is given by:

$$S_{\omega}^{(0)} = \frac{[\sigma^{(0)}]^2}{1 - 2(\theta_{(1,0)} \cos \frac{2\pi\omega_1}{M} + \theta_{(0,1)} \cos \frac{2\pi\omega_2}{N})}. \tag{15}$$

The power spectral function on subsampling is given by:

$$S_{\omega}^{(1)} = \frac{1}{4}[S_{\omega}^{(0)} + S^{(0)}_{\omega+(\frac{M}{2},0)} + S^{(0)}_{\omega+(0,\frac{N}{2})} + S^{(0)}_{\omega+(\frac{M}{2},\frac{N}{2})}]. \tag{16}$$

By using Eq. (15) in Eq. (16) and after some manipulations, we obtain:

$$\begin{aligned} S_{\omega}^{(1)} &= \\ 2 \quad [\sigma^{(0)}]^2 \quad & [\frac{1}{1 - 4(\theta_{(1,0)} \cos \frac{2\pi\omega_1}{M} + \theta_{(0,1)} \cos \frac{2\pi\omega_2}{N})^2} + \frac{1}{1 - 4(\theta_{(1,0)} \cos \frac{2\pi\omega_2}{M} - \theta_{(0,1)} \cos \frac{2\pi\omega_2}{N})^2}]. \end{aligned} \tag{17}$$

Claim:

For a first order GMRF at the fine resolution, the only set of parameters that result in the same power spectrum at the lower resolution 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. (17) it can be infered 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 $S_{\omega}^{(1)}$. For these sets of parameters the subsampled processes 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 $0 \leq \omega_1 \leq M-1$ and $0 \leq \omega_2 \leq N-1$, we can substitute specific values of (ω_1, ω_2), to get the necessary conditions.

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

$$S^{(1)}_{(0,\frac{N}{4})} = \frac{4[\sigma^{(0)}]^2}{1 - 4\theta^2_{(1,0)}}$$

Hence, the only possible values of $\theta_{(1,0)}$ that will result in the same $S^{(1)}_{(0,\frac{N}{4})}$ 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 $(\theta_{(1,0)}, \theta_{(0,1)}\theta_{(1,1)}, \theta_{(-1,1)}, [\sigma^{(0)}]^2)$ and the power spectral density function is given by:

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

The power spectral function on subsampling is given by:

$$\begin{aligned} S^{(1)}_{\omega} = & \\ & \frac{1}{4}[\sigma^{(0)}]^2[\frac{1}{1-2(\theta_{(1,0)} \cos\frac{2\pi\omega_1}{M}+\theta_{(0,1)} \cos\frac{2\pi\omega_2}{N}+\theta_{(1,1)} \cos(\frac{2\pi\omega_1}{M}+\frac{2\pi\omega_2}{N})+\theta_{(-1,1)} \cos(\frac{2\pi\omega_1}{M}-\frac{2\pi\omega_2}{N}))} \\ & +\frac{1}{1-2(\theta_{(1,0)} \cos\frac{2\pi\omega_1}{M}-\theta_{(0,1)} \cos\frac{2\pi\omega_2}{N}-\theta_{(1,1)} \cos(\frac{2\pi\omega_1}{M}+\frac{2\pi\omega_2}{N})-\theta_{(-1,1)} \cos(\frac{2\pi\omega_1}{M}-\frac{2\pi\omega_2}{N}))} \\ & +\frac{1}{1-2(-\theta_{(1,0)} \cos\frac{2\pi\omega_1}{M}+\theta_{(0,1)} \cos\frac{2\pi\omega_2}{N}-\theta_{(1,1)} \cos(\frac{2\pi\omega_1}{M}+\frac{2\pi\omega_2}{N})-\theta_{(-1,1)} \cos(\frac{2\pi\omega_1}{M}-\frac{2\pi\omega_2}{N}))} \\ & +\frac{1}{1-2(-\theta_{(1,0)} \cos\frac{2\pi\omega_1}{M}-\theta_{(0,1)} \cos\frac{2\pi\omega_2}{N}+\theta_{(1,1)} \cos(\frac{2\pi\omega_1}{M}+\frac{2\pi\omega_2}{N})+\theta_{(-1,1)} \cos(\frac{2\pi\omega_1}{M}-\frac{2\pi\omega_2}{N}))}] \quad . \end{aligned} \tag{18}$$

Claim:

For a second order GMRF at the fine resolution, the only set of parameters that result in the same power spectrum at the lower resolution 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:

By observing Eq. (18) it can be infered that the above four sets of parameters will result in the same $S^{(1)}_{w}$. The constraints we have obtained on $(\theta_{(1,0)}, \theta_{(0,1)})$ for the first order case also hold good here.

Let, $w_1 = M/4$ and $w_2 = N/4$,

$$S^{(1)}_{(\frac{M}{4},\frac{N}{4})} = \frac{4[\sigma^{(0)}]^2}{1 - 4(\theta_{(1,1)} - \theta_{(-1,1)})^2}$$

Therefore, a constraint on the set of $(\theta_{(1,1)} - \theta_{(-1,1)})$ that can result in the same $S^{(1)}_w$ is,

$$\mid (\theta_{(1,1)} - \theta_{(-1,1)}) \mid = constant. \tag{19}$$

Let, $w_2 = \frac{Nw_1}{M} + \frac{N}{4}$, so that $\cos(\frac{2\pi\omega_1}{M} - \frac{2\pi\omega_2}{N}) = 0$. For convenience, let, $\cos(\frac{2\pi\omega_1}{M}) = \alpha_1$, $\cos(\frac{2\pi\omega_2}{N}) = \alpha_2$ and $\cos(\frac{2\pi\omega_1}{M} + \frac{2\pi\omega_2}{N}) = \alpha_3$, Writing the expression for $S^{(1)}_w$:

$$S^{(1)}_{(\omega_1,\frac{N\omega_1}{M}+\frac{N}{4})} = \frac{2(1-2\alpha_1\theta_{(1,0)})}{(1-2\alpha_1\theta_{(1,0)})^2-4(\alpha_2\theta_{(0,1)}+\alpha_3\theta_{(1,1)})^2} + \frac{2(1+2\alpha_1\theta_{(1,0)})}{(1+2\alpha_1\theta_{(1,0)})^2-4(\alpha_2\theta_{(0,1)}-\alpha_3\theta_{(1,1)})^2}. \tag{20}$$

If $(-\theta_{(1,0)}, \theta_{(0,1)}, \theta'_{(1,1)}, \theta'_{(-1,1)})$ should result in the same spectrum as $(\theta_{(1,0)}, \theta_{(0,1)}, \theta_{(1,1)}, \theta_{(-1,1)})$ for all ω_1 in Eq. (20), then

$$\theta'_{(1,1)} = -\theta_{(1,1)}.$$

Combining this result with Eq. (19) implies that

$$\theta'_{(-1,1)} = -\theta_{(-1,1)}.$$

Similar results can be shown for the other sets of parameters and for higher order processes.

5 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 process. Each site in the lattice carries a class label from $(1, 2, \ldots, V)$ and this label process is modeled by an MRF. We do not directly observe the label process, but a function of the labels, the intensity process. The intensity process is modeled by a GMRF, the parameters $(\underline{\theta}, \sigma^2)$ of which depend on the label process at that site. The goal is to estimate the unobserved label process from the observed intensity process by optimizing a suitable error criterion.

The following conventions are used. The GMRF parameters corresponding to label v, $(v \in \{1, 2, \dots V\})$ are written as $(\underline{\theta}(v), \sigma^2(v))$. The symmetric neighborhood for the GMRF is denoted by η and for the MRF label process by ψ and the index r is used to index through the neighborhood of the both label and intensity processes.

The intensity process is defined as follows:

$$P(X_s = x_s / L_s = v, X_r, L_r, r \in \eta) = \frac{1}{\sqrt{2\pi\sigma^2(v)}} \exp\{-\frac{1}{2\sigma^2(v)}[x_s - \sum_{r\in\eta} \theta_r(v) x_{s+r})]^2\} \quad (21)$$

The above equation is exact if the label field is homogeneous in the η neighborhood. If not, the intensity values corresponding to the site where $L_r \neq v$ can be replaced by the mean value.

The label process is modeled by an MRF. The following model is also called a pairwise interaction model.

$$P(\underline{L} = \underline{l}) = \frac{1}{Z} \exp(\beta \sum_{s\in\Omega} U(l_s))$$

where, $U(l_s)$ is the number of neighbors in ψ that have the same label as l_s.

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

$$P(L_s / L_r, r \in \psi) = \frac{\exp(\beta U(l_s))}{\sum_{l'_s = (1,2,\dots V)} \exp(\beta U(l'_s))} \quad (22)$$

Now given the intensity process, the label process can be estimated by minimizing a suitable criterion. The *maximum a posterior* (MAP) error criterion solution can be obtained by:

$$max_{\underline{L}} P(\underline{X}, \underline{L}) = max_{\underline{L}} P(\underline{L}/\underline{X}) P(\underline{X}). \quad (23)$$

This optimization requires stochastic relaxation methods and is computationally very expensive. So, we restrict ourselves to iterated conditional mode (ICM) method, a greedy algorithm that converges to a local maxima. ICM solution is obtained by:

$$\max_{L_s} \qquad P(L_s / L_r, X_s, X_r)$$

$$\begin{aligned} \max_{L_s} \quad & P(X_s/X_r, L_s, L_r)P(L_s/L_r, X_r) \\ \max_{L_s} \quad & P(X_s/X_r, L_s, L_r)P(L_s/L_r) \end{aligned} \tag{24}$$

This is equivalent to,

$$\min_{L_s} \quad \log(\sigma(L_s = v)) + \frac{1}{2\sigma^2(v)}[x_s - \sum_{r \in \eta} \theta_r(L_s = v)x_{s+r}]^2 - \beta U(L_s = v) \tag{25}$$

the minimization is performed by visiting the pixels in raster scan 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 singe resolution algorithm. As we have discussed in the previous sections, data at lower resolutions can be approximated by a GMRF process. Thus the same algorithm can be applied at lower resolutions too. Our multiresolution algorithms 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 as discussed before. Then segmentation is performed at the coarsest resolution by minimizing Eq. (25) with the corresponding parameters and the results of segmentation are passed on to the immediate higher resolution and this process 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 the 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 propagated upwards to the immediate higher resolution. Since we obtain resolution transformation by subsampling, we have a quad tree type of graph. If $\underline{L}^{(k)}$ is the segmentation result at the kth resolution, the labels in the $(k-1)$th level are initialized as:

$$L_s^{(k-1)} = L_{\lfloor s/2 \rfloor}^{(k)}. \tag{26}$$

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

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

$$\begin{aligned} \dot{v}_s &= \arg \max_{v \in \{1,2,\ldots,V\}} P(L_s = v / L_r, X_s, X_r) \\ \ddot{v}_s &= \arg \max_{v \in \{1,2,\ldots,V\} \setminus \dot{v}_s} P(L_s = v / L_r, X_s, X_r) \end{aligned}$$

and the confidence measure is defined as,

$$C_s^{(k)} = \frac{P(\dot{v}_s / L_r, X_s, X_r)}{P(\ddot{v}_s / L_r, X_s, X_r)} \tag{27}$$

These confidence measure at level k are propagated upwards to level $k-1$ in the same manner as in Eq. (26). At level k, ICM is restricted to only those pixels with confidence measure such that,

$$\frac{1}{C_s^{(k)}} \geq c^{(k)} \tag{28}$$

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

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

For the coarsest resolution $c^{(\cdot)} = 0$, ie., ICM is performed over all sites in the lattice.

The confidence measure defined in Eq. (27) has a hypothesis testing interpretation. The C_s expression is exactly 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. If at a particular resolution the parameters corresponding to different textures are such that discrimination between some textures are low, while the rest of the textures can be easily discriminated. In such cases, if $\dot{v}$ and $\ddot{v}$ 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 process in the two-tier model is defined by an MRF. This process also loses Markovianity under resolution transformation and can be approximated by a Markov process at lower resolutions. For the pairwise interaction model there is only one parameter β to be estimated in Eq. (22). 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 $\beta = \{0.5, 0.3, 0.15, 0.1, 0.05\}$ 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 of the label MRF process does not have a great bearing on the segmentation, hence a rigorous estimation may not necessary.

6 Experiments

We present experimental results with simulated, Brodatz texture images and real satellite images and show that the multiresolution algorithms performs better both in terms of the classification accuracy and computational requirement. In all the experiments, the confidence threshold $c^{(k)} = (0.6, 0.25, 0.15, 0.1, 0.05)$, is used for the different levels. In all cases the misclassification percentage error is reported. Maximum likelihood initialization is used for the ICM in the single resolution and at the lowest level for multiresolution. Multiresolution results presented in this section are obtained by performing the algorithm over three resolutions.

6.1 Synthetic Image

We synthesised texture images using the technique given in [6]. Three third order GMRF textures are generated with parameters $\{\underline{\theta}$= (0.0934154, 0.520252, 0.0303413, 0.0180476, -0.0216434,-0.148331), $\sigma^2 = 0.9342$ $\}$, $\{\underline{\theta}$= (0.308257, 0.468389, -0.0755398, -0.0755797, -0.0407557, -0.100678), $\sigma^2 = 1.8472$ $\}$, $\{\underline{\theta}$= (0.406875, 0.423393, -0.178478, -0.188702, -0.0649544, -0.121439), $\sigma^2 = 1.264811$ $\}$. Figure 9 shows the composite image with these three textures (after histogram equalization for visual clarity). Figure 10 shows the single resolution segmentation result and Figure 11 shows the result for multiresolution segmentation.

6.2 Brodatz Images

We have tested our algorithm on textures from Brodatz texture album. Figure 12 contains grass, calf leather, wool and wood textures. The original GMRF parameters are estimated by maximum likelihood estimation. Figure 13 shows the single resolution segmentation and Figure 14 shows the multiresolution segmentation. We have another interesting plot of the $\frac{1}{C_s^{(k)}}$ for the level $k = 1$ in Figure 15. The brighter points in the image correspond to points of low confidence measure. As expected all the boundary regions between different textures have a low confidence measure. In texture segmentation classification near the texture boundaries is usually more ambiguous.

Table 2 shows the comparison between single and multiresolution algorithms in terms of the misclassification error percentage and 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	
	Error %	Computation	Error%	Computation
Synthetic	10.16	2686976	3.25	431031
Brodatz1	13.96	1114112	7.25	679444
Brodatz2	31.89	2490368	15.46	710980
Africa	–	2160000	–	426105

Table 2

We also experimented with another more complex set of textures from Brodatz album. Figure 16 shows a three class (pigskin, raffia and water) texture image. Figure 17 shows the single resolution segmentation and Figure 18 shows the multiresolution segmentation. Figure 19 shows the corresponding confidence measure plot.

6.3 Real Image

Figure 20 shows a section of a single channel of a multispectral sensor (MSS) image over Africa. (The image has been displayed after histogram equalization, the classes are not so disparate in the actual image). Unfortunately exact class maps are not available. However, we chose three classes corresponding to river, forest, deforestation and the GMRF parameters are obtained from small sections from a different part of the image and then used to classify the image shown. Figure 21 shows the single resolution result and Figure 22 shows the multiresolution result. Obviously, we can see that the multiresolution algorithm has performed better, with lesser computation, than the single resolution algorithm.

We also present the results of multireslution segmentation for two other satellite images. We have not presented the performance comparison in these cases, since exact class maps are not available. Figure 23 shows another section of MSS image and Figure 24 shows the corresponding 4-class multiresolution segmentation result. Figure 25 shows a section of thematic mapper (TM) data and Figure 26 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 in different levels of specificity. We have provided two schemes based on minimizing KL distances, to estimate the parameters of GMRF at lower resolutions and have successfully used it for texture segmentation application. Also, this can be extended to perform unsupervised texture segmentation. However, as mentioned in Section 4, GMRF parameters at 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. GMRFs are widely used in many image processing applications including restoration, segmentation and compression. The proposed scheme can be used for multiscale restoration, segmentation and image compression techniques used in progressive image transmission.

References

[1] J. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society*, Vol. 36, pp. 192–326, 1974.

[2] J. Besag, "On the Statistical Analysis of Dirty Pictures," *Journal of the Royal Statistical Society*, Vol. 48, pp. 259–302, 1986.

[3] C. Bouman and B. Liu, "Multiple Resolution Segmentation of Textured Images," *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 13, pp. 99–113, Feb 1991.

[4] P. Burt and E. Adelson, "The Laplacian Pyramid as a Compact Image Code," *IEEE Transactions on Communications*, Vol. COM-31, No. 4, pp. 532–540, 1983.

[5] S. Chatterjee and R. Chellappa, "Maximum Likelihood Texture Segmentation Using Gaussian Markov Random Field Models," *IEEE International Conf. on Acoustics, Speech, and Signal Procssing*, June 1985.

[6] 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, Elsavier, 1985.

[7] P. C. Chen and T. Pavlidis, "Image Segmentation as an Estimation Problem," *Computer Vision,Graphics and Image Processing*, Vol. 12, pp. 153–173, 1980.

[8] 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.

[9] F. S. Cohen, Z. Fan, and M. A. Patel, "Classification of Rotated and Scaled Textured Images Using Gaussian Markov Random Field Models," *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 13, pp. 192–202, February 1991.

[10] G. R. Cross and A. K. Jain, "Markov Random Field Texture Models," *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 5, pp. 25–39, 1983.

[11] L. S. Davis, S. Johns, and J. K. Aggrawal, "Texture analysis using generalized co-occurence matrices," *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 1, pp. 251–259, July 1979.

[12] 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, January 1987.

[13] C. Domb and M. Green, "Phase Transition and Critical Phenomena." Academic Press, 1979.

[14] D. Geiger and F. Girosi, "Parallel and Deterministic Algorithms for MRF's:Surface Reconstruction," *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 13, pp. 401–413, May 1991.

[15] 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, 1984.

[16] S. Geman and C. Graffigne, "Markov Random Field Image Models and Their Application to Computer Vision," *Proc. Internatioinal Congress of Mathematicians*, pp. 1496–1517, 1986.

[17] B. Gidas, "A Renormalization Group Approach to Image Processing," *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 11, No. 2, pp. 164–180, 1989.

[18] R. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE*, Vol. 67, No. 5, pp. 610–621, 1979.

[19] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by Simulated Annealing," *Science*, pp. 671–680, 1983.

[20] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *Ann. Math. Stat.*, Vol. 22, pp. 79–86, 1951.

[21] H. Kunsch, "Thermodynamics and Statistical Analysis of Gaussian Random Fields," *Z. Wahr. Ver. Geb.*, pp. 407–421, Nov 1981.

[22] 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, Acadmic Press, 1993.

[23] 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.

[24] 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.

[25] I. Matsuba, "Renormalization Group Approach to Hierarchical Image Analysis," *IEEE International Conf. on Acoustics, Speech, and Signal Procssing*, pp. 1044–47, 1988.

[26] 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.

[27] A. Rosenfeld, "Visual Texture Analysis," Tech. Rep. 70-116, University of Maryland, June 1970.

[28] A. Rosenfeld, "Some Useful Properties of Pyramids," in *Multiresolution Image Processing and Analysis* (A. Rosenfeld, ed.), Springer-Verlag, 1984.

[29] G. Sharma and R. Chellappa, "Two-Dimensional Maximum Entropy Power Spectra," *IEEE Transaction on Information Theory*, Vol. IT-31, pp. 90–99, Jan 1985.

[30] D. Terzopoulos, “Image Analysis using Multigrid Relaxation Methods,” *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 8, pp. 129–139, March 1986.

[31] F. Tomita, “Description of texture by a structural analysis,” *Proc. IJCAI*, 1979.

[32] M J. W. Woods, “Two-dimensional Discrete Markovian Fields,” *IEEE Transaction on Information Theory*, Vol. 18, pp. 232–240, 1972.

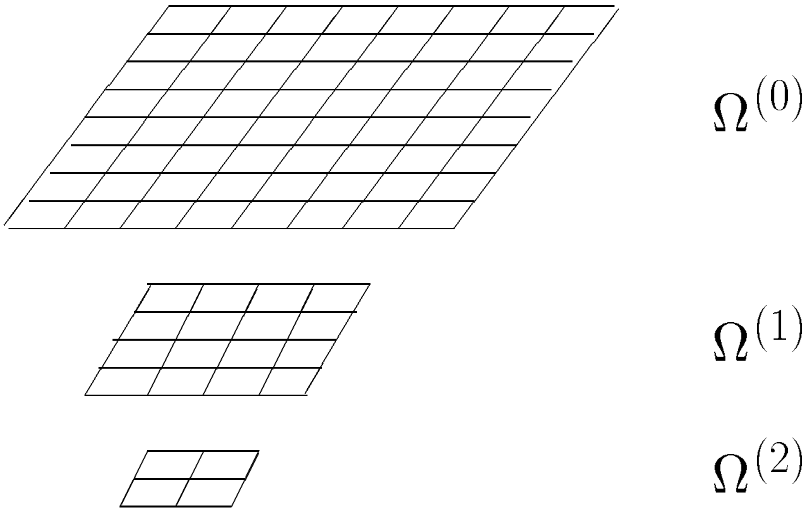

Figure 1: Resolution Transformation

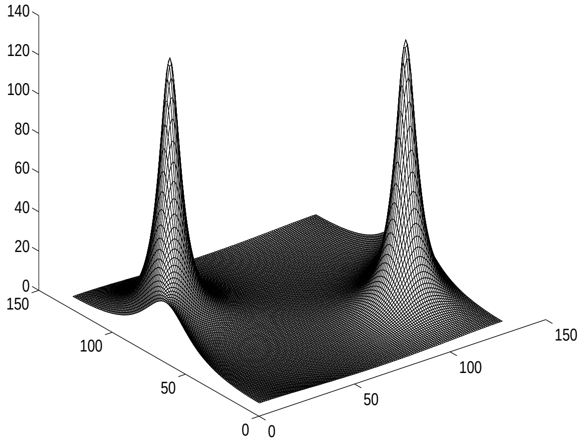

Figure 2: $S_w^{(0)}$

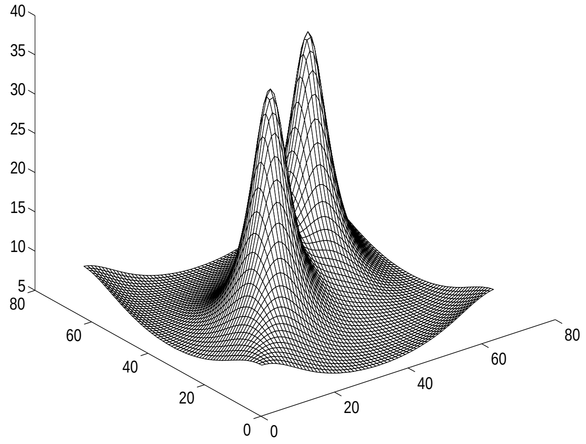

Figure 3: $S_w^{(1)}$

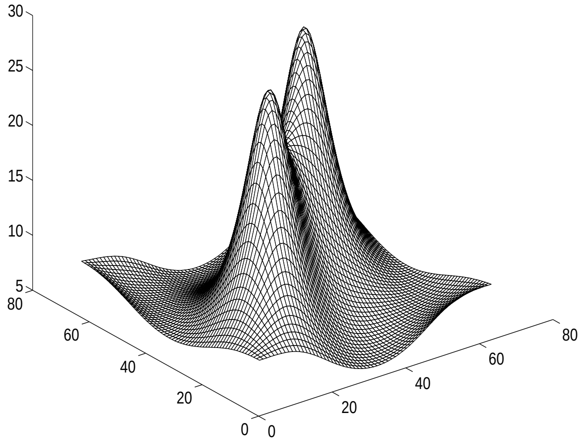

Figure 4: GMRF $MS_w^{(1)}(3)$

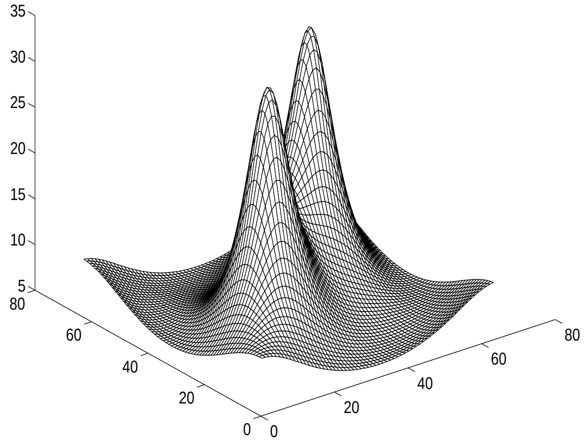

Figure 5: GMRF $MS_w^{(1)}(4)$

Figure 2: Power spectrum of GMRF at $\Omega^{(0)}$, Figure 3: Exact non-Markov power spectrum at $\Omega^{(1)}$, Figure 4: Power spectrum of third order Markov approximation at $\Omega^{(1)}$ and Figure 5: Power spectrum of fourth order Markov approximation at $\Omega^{(1)}$.

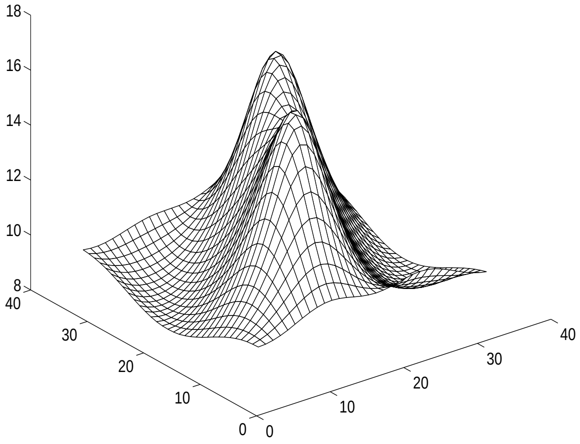

Figure 6: $S_w^{(2)}$

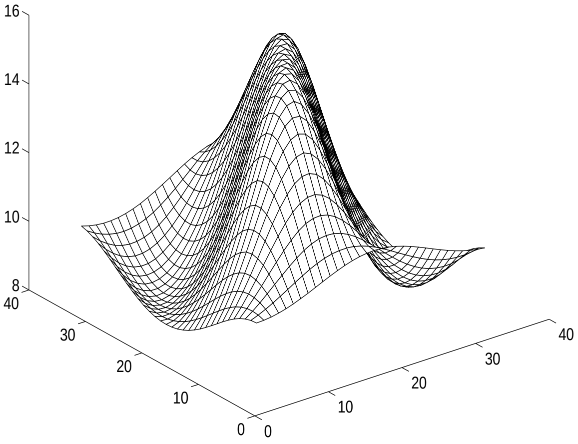

Figure 7: GMRF $MS_w^{(2)}(2)$

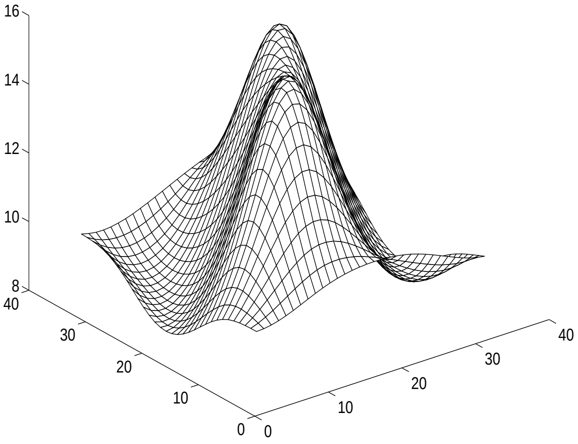

Figure 8: GMRF $MS_w^{(2)}(3)$

Figure 6: Exact non-Markov power spectrum at $\Omega^{(2)}$, Figure 7: Power spectrum of second order Markov approximation at $\Omega^{(2)}$ and Figure 8: Power spectrum of third order Markov approximation at $\Omega^{(2)}$.