4.3

Random Field Models

P. Fieguth
University of Waterloo, Ontario

J. Zhang
University of Wisconsin, Milwaukee

1 Introduction

Random fluctuations in intensity, color, texture, object boundary or shape can be seen in most real-world images, as illustrated in Fig. 1. The causes for these fluctuations are diverse and complex, often due to factors such as non-uniform lighting, random fluctuations in object surface orientation and texture, complex scene geometry, and noise.[1] Consequently the processing of such images becomes a problem of statistical inference [1], which requires the definition of a statistical model corresponding to the image pixels.[2]

While simple image models can be obtained from statistics such as the mean, variance, histogram, and correlation [2, 3], a more general approach is to use *random fields*. Indeed, as a two-dimensional extension of the one-dimensional random process, a random field model provides a complete statistical characterization for a given class of images, from which all statistical properties of the images can, in principle, be derived. Combined with frameworks for statistical inference, such as maximum-likelihood (ML) and Bayesian estimation, random field models have led to significant advances in many statistical image processing applications, including image restoration, enhancement, classification, segmentation, compression and synthesis.

Early studies of random fields can be traced to the 1970s, with many of the early results summarized in [4]. Among the wide variety of proposed models, the most used is perhaps the AR (autoregressive) model and its various extensions (e.g., [3]). A landmark paper by Geman and Geman [5] in 1984 addressed Markov random field (MRF) models, attracted great attention, and invigorated research in image modeling. Indeed the MRF, coupled with the Bayesian framework, has been the focus of many studies [6–8]. Section 2 will introduce notation and provide an overview of random field models, emphasizing the autoregressive and Markov fields.

With the advent of multiresolution processing techniques,[3] such as the pyramid [9] and wavelets [10], much of the current research in random field models focuses on multiscale models [11–25]. This interest has been motivated by the significant advantages they may have in *computational power* and *representational power* over the single-resolution/single-scale models. Specifically, multiresolution/multiscale processing can provide drastic computation reduction and represent a highly complicated model by a set of simpler models.

Multiresolution/multiscale models that aim at computation-reduction include various multiresolution/multiscale MRFs [17–19] and multiscale tree models [12–16]. Through their connection to multigrid methods [27], these models often can improve convergence in iterative procedures. The multiscale tree model is described in more detail in Section 3.

Multiresolution/multiscale models that aim at representing highly complicated random fields (e.g., those with nonlinear and long range interactions) include various nonlinear hierarchical/multiresolution/multiscale texture models

[1]Chapter 4.5 of this book is devoted to noise and noise models.

[2]Also see Chapter 4.7, which considers statistical models for photographic images.

[3]Also see Chapter 4.2 of this book.

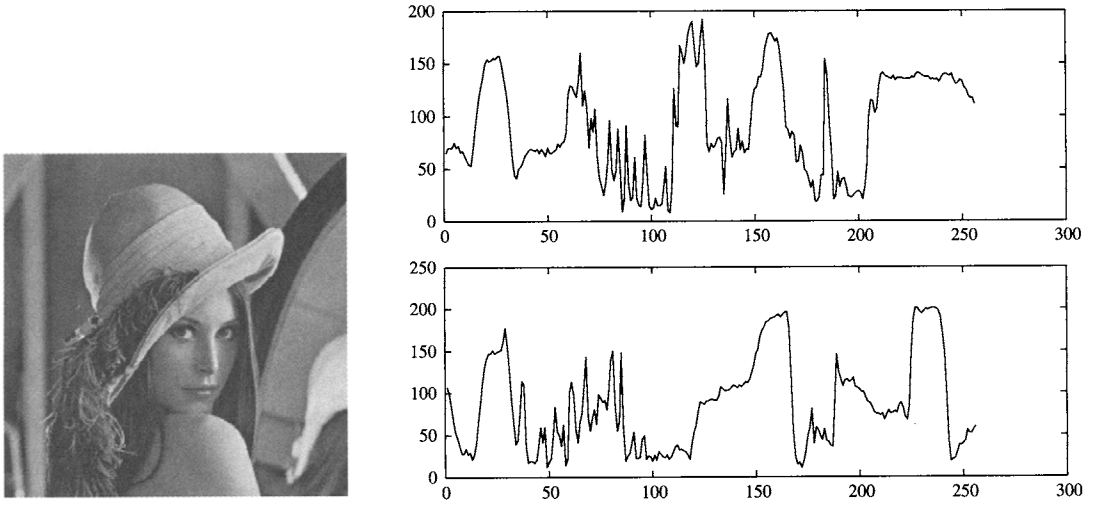

FIGURE 1 A typical image (left), two rows of which are plotted (right); the fine-scale details appear nearly "random" in nature.

[21–26]. Section 4 describes such a model based on the Gaussian mixture.

2 Random Fields — Overview

A random field $\boldsymbol{x}$ is a collection of random variables arranged on a lattice L:

$$\boldsymbol{x} = \{x_{\boldsymbol{i}}, \boldsymbol{i} \in L\}. \tag{1}$$

In principle the lattice can be any (possibly irregular) collection of discrete points; however, it is most convenient and intuitive to visualize the lattice as a rectangular, regular array of sites:

$$L = \{(i,j) | 1 \le i \le N, 1 \le j \le M\} \tag{2}$$

in which case a random field is just a set of random pixels

$$\boldsymbol{x} = \{x_{i,j}, (i,j) \in L\}. \tag{3}$$

As with random variables or random vectors, any random field can, in principle, be completely characterized by its associated probability measure $p(\boldsymbol{x})$. The detailed form of $p(\cdot)$ will depend on whether the elements $x_{i,j}$ are discrete, in which case $p(\boldsymbol{x})$ denotes a probability distribution, or continuous, in which case $p(\boldsymbol{x})$ denotes a probability density function.

Suppose we take an image of modest size, say $N = M = 256$; this implies that $p(\cdot)$ must explicitly characterize the joint statistics of 65536 elements. However, often $p(\cdot)$ is a cumbersome and computationally inefficient means of defining the statistics of the random field. Indeed, a great part of the research into random fields involves the discovery or definition of *implicit* statistical forms which lead to effective or faithful representations of the true statistics, while admitting computationally efficient algorithms.

Broadly speaking there are four typical problems associated with random fields:

1. Representation: how is the random field represented and parameterized? In general the probability distribution $p(\cdot)$ is not computable for large fields, except in pathological cases, for example, in which all of the elements are independent:

$$p(\boldsymbol{x}) = \prod_{(i,j) \in L} p(x_{i,j}). \tag{4}$$

2. Parameter Estimation: given a parameterized statistical model, for example in the form $p(\boldsymbol{x}|\boldsymbol{\theta})$, and a sample image $\boldsymbol{y}$, estimate the parameters $\boldsymbol{\theta}$. Typically we are interested in the *maximum likelihood* (ML) estimates

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} \{p(\boldsymbol{y}|\boldsymbol{\theta})\} \tag{5}$$

   Such an approach can be used to estimate any parameters on which the field statistics depend, such as correlation length, temperature, or ambient color.
3. Estimation: given a prior statistical model $p(\boldsymbol{x}|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is known or has been estimated (above), and given observations $\boldsymbol{y}$, estimate the image $\boldsymbol{x}$. From the observation model $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$ we can infer the posterior distribution $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$, leading to the *maximum a posteriori* (MAP) estimate

$$\hat{\boldsymbol{x}}_{MAP} = \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})\}. \tag{6}$$

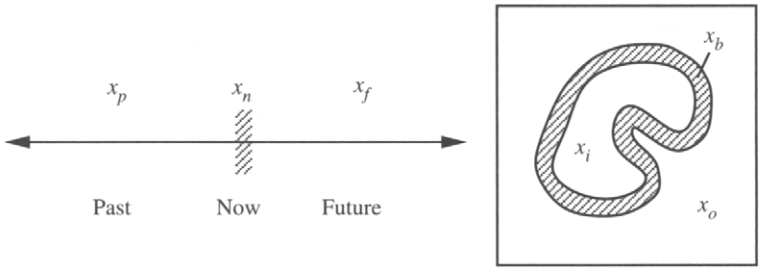

FIGURE 2 The Markov property: given a boundary, the two separated portions of the field are conditionally independent.

In general the MAP estimate may be quite complicated. In many imaging settings we prefer a simplified model,

$$\boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{v} \tag{7}$$

where f is linear and $\boldsymbol{v}$ is a noise signal with known statistics, and choose to find the least-squares estimate

$$\hat{\boldsymbol{x}} = \arg\min_{\hat{\boldsymbol{x}}} E\left[|(\boldsymbol{x} - \hat{\boldsymbol{x}})|^2 \middle| \boldsymbol{y} \right]. \tag{8}$$

Least-squares estimates are of interest in reconstruction or inference problems; for example, denoising images or interpolation.

4. Sampling: whereas MAP estimation produced the *best* estimate $\boldsymbol{x}$, maximizing some distribution over $\boldsymbol{x}$, in many contexts we are interested in the inherent variability present in $\boldsymbol{x}$.
   Prior sampling involves generating random samples of $\boldsymbol{x}$ from distribution $p(\boldsymbol{x}|\boldsymbol{\theta})$; whereas posterior sampling involves finding random samples from the posterior distribution $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$, such that the random $\boldsymbol{x}$ are subject to constraints imposed by the observations $\boldsymbol{y}$. To generate fields $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ statistically sampled from $p(\boldsymbol{x}|\ldots)$ implies that the probability of generating $\boldsymbol{x}$ is $P(\boldsymbol{x}|\ldots)$.[4]
   Such random field realizations are used in stochastic image compression, random texture synthesis, lattice-physics simulations.

2.1 Markov Random Fields

The fundamental notion associated with Markovianity is one of conditional independence: a one-dimensional process x_n is Markovian (Fig. 2) if the knowledge of the process at some point x_n decouples the "past" $\boldsymbol{x}_p$ and the "future" $\boldsymbol{x}_f$:

$$p(\boldsymbol{x}_f|x_n, \boldsymbol{x}_p) = p(\boldsymbol{x}_f|x_n), \qquad p(\boldsymbol{x}_p|x_n, \boldsymbol{x}_f) = p(\boldsymbol{x}_p|x_n). \tag{9}$$

[4]Assuming that $\boldsymbol{x}$ is discrete. Sampling is straightforward for continuous-state $\boldsymbol{x}$, however, notions of probability and sampling much be formulated more carefully.

The decoupling extends perfectly into two dimensions, except that the natural concepts of "past" and "future" are lost, since there is no natural ordering of the elements in a grid; instead, a random field $\boldsymbol{x}$ is Markov (Fig. 2) if the knowledge of the process on a boundary set b decouples the inside and outside of the set:

$$p(\boldsymbol{x}_i|\boldsymbol{x}_b, \boldsymbol{x}_o) = p(\boldsymbol{x}_i|\boldsymbol{x}_b), \qquad p(\boldsymbol{x}_o|\boldsymbol{x}_b, \boldsymbol{x}_i) = p(\boldsymbol{x}_o|\boldsymbol{x}_b). \tag{10}$$

This boundary concept, although elegantly intuitive, is lacking in details (e.g., how "thick" does the boundary need to be?). It is often simpler, and more explicit, to talk about separating a single element $x_{i,j}$ from the entire field $\boldsymbol{x}$ conditioned on a local neighborhood $\mathcal{N}_{i,j}$ [5, 6]:

$$p(x_{i,j}|\{x_{k,l}, (k,l) \in L\backslash(i,j)\}) = p(x_{i,j}|\{x_{k,l}, (k,l) \in \mathcal{N}_{i,j}\}). \tag{11}$$

The shape and extent of $\mathcal{N}_{i,j}$ is one aspect which characterizes the nature of the random field:

- Causal/Noncausal: A neighborhood structure is causal (Fig. 3) if all elements of the neighborhood live in one half of the plane; e.g.,

$$(k,l) \in \mathcal{N}_{i,j} \Longrightarrow l < j, \text{ or } l = j, k < i. \tag{12}$$

  That is, if the field can be reordered into a one-dimensional random vector which is Markov, satisfying (9). Otherwise, more typically, the neighborhood is noncausal (Fig. 3).

- Order: The order of a neighborhood reflects its extent. A first-order neighborhood is shown in Fig. 4, which also illustrates the pattern followed by higher order neighborhoods.

The above discussion is entirely formulated in terms of the probability density $p(\boldsymbol{x})$, which is often impractical for large random fields; the following sections summarize the two,

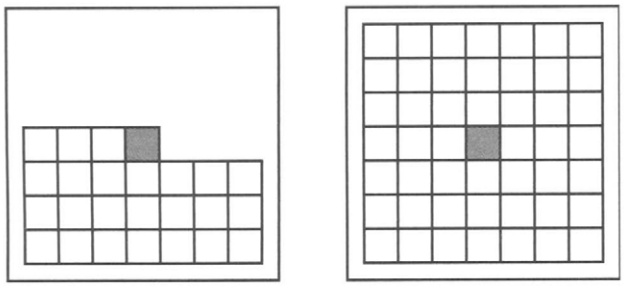

FIGURE 3 Regions of support for causal (left) and noncausal (right) neighborhoods of the shaded element.

broad alternatives which have been developed:

1. Gauss-Markov Random Fields: $\boldsymbol{x}$ is Gaussian, in which case the field can be characterized explicitly in terms of expectations rather than probability densities.
2. Gibbs Random Fields: an energy $E(\boldsymbol{x})$ is associated with each possible field $\boldsymbol{x}$; a probability density is then constructed implicitly from $E(\boldsymbol{x})$ in such a way that $p(\boldsymbol{x})$ satisfies (10),(11).

2.2 Gauss-Markov Random Fields

When the random field $\boldsymbol{x}$ is Gaussian [6], then conditional independence is equivalent to conditional uncorrelatedness, so instead of (11) we write

$$\hat{x}_{i,j} = E[x_{i,j}|\{x_{k,l}, (k,l) \in L\backslash(i,j)\}] = E[x_{i,j}|\{x_{k,l}, (k,l) \in \mathcal{N}_{i,j}\}] \tag{13}$$

where $\hat{x}_{i,j}$ is the estimated value of $x_{i,j}$. However if $\boldsymbol{x}$ is Gaussian the expectation is known to be linear, so the right side of (13) can be rewritten as

$$\hat{x}_{i,j} = \sum_{(k,l)\in\mathcal{N}_{i,j}} a_{i,j,k,l} x_{k,l}. \tag{14}$$

Alternatively, we can describe the field elements directly, instead of their estimates:

$$x_{i,j} = \hat{x}_{i,j} + w_{i,j} = \sum_{(k,l)\in\mathcal{N}_{i,j}} a_{i,j,k,l} x_{k,l} + w_{i,j} \tag{15}$$

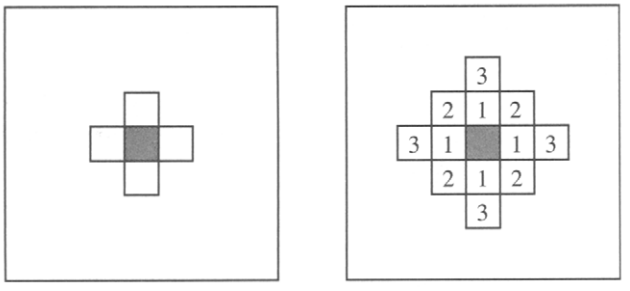

FIGURE 4 Left: The region of support of a first-order neighborhood. Right: Neighborhood size as a function of model order.

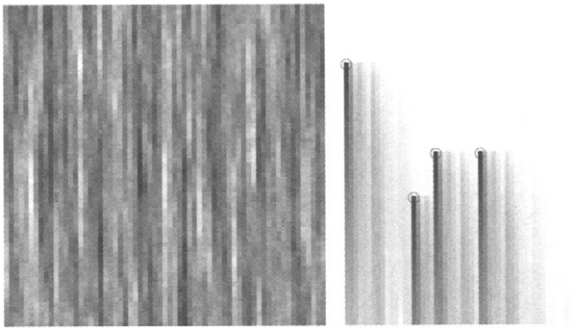

FIGURE 5 A causal model produces a reasonable sample (left), but shows obvious limitations when computing estimates (right) from sparse observations (circled).

where $w_{i,j}$ is the estimation error process. If the random field is stationary then the coefficients simplify as

$$a_{i,j,k,l} = a_{i-k,j-l}. \tag{16}$$

2.2.1 Causal GMRFs

If a random field $\boldsymbol{x}$ is causal, each neighborhood $\mathcal{N}_{i,j}$ must limit its support to one half of the plane, as sketched in Fig. 3. These are known as non-symmetric half-plane (NSHP) models, and lead to very simple, $\mathcal{O}(NM)$ autoregressive equations for sampling and estimation.

Specifically, there must exist an ordering of the field elements into a vector $\vec{x}$ such that each element depends only on the values of elements lying earlier in the ordering, in which case (15) is the autoregressive equation to generate prior samples and the Kalman filter can be used for estimation.

These models have limited applicability, since most random fields are not well represented causally. The limitations of the causal model are most obvious when computing estimates from sparse observations, as shown in Fig. 5, since the arrangement of the estimates is obviously asymmetric with respect to the observations.

2.2.2 Toroidally Stationary GMRFs

A second special case is that of toroidally-stationary fields; that is, rectangular fields in which the left and right edges are considered adjacent, as are the top and bottom.[5] In other words, the field is periodic. The correlation structure of a toroidally-stationary $N \times M$ field therefore takes the form

$$E[x_{i,j}x_{i+\Delta i,j+\Delta j}] = E[x_{0,0}x_{\Delta i \bmod N, \Delta j \bmod M}]. \tag{17}$$

[5] That is, topologically, the wrapping of a rectangular sheet onto a torus or doughnut.

The significance of *any* such stationarity is that the resulting covariance matrix is diagonalized by the two-dimensional FFT, leading to extremely fast algorithms. Specifically, let Λ be the correlation structure of the field,

$$\Lambda_{i,j} = E[x_{0,0}x_{i,j}] \quad 1 \le i \le N, 1 \le j \le M \tag{18}$$

then prior samples may be computed as

$$\boldsymbol{s} = \text{ifft}_2\{\text{sqrt}(\text{fft}_2(\boldsymbol{\Lambda})) \cdot \text{fft}_2(\boldsymbol{q})\} \tag{19}$$

where sqrt() and $\cdot$ are element-by-element operations, and $\boldsymbol{q}$ is an array of unit-variance, independent Gaussian samples. Similarly, given a set of observations with Gaussian error,

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{v}, \quad \text{cov}(\boldsymbol{v}) = \boldsymbol{\sigma}^2 I \tag{20}$$

the least-squares estimates may be computed as

$$\hat{\boldsymbol{x}} = \text{ifft}_2\{\text{fft}_2(\boldsymbol{\Lambda}) \cdot \text{fft}_2(\boldsymbol{y})/(\text{fft}_2(\boldsymbol{\Lambda}) + \boldsymbol{\sigma}^2)\} \tag{21}$$

where again $\cdot$ and / are performed element-by-element.

In general the circumstances (17), (20) are restrictive: the field must be toroidally-stationary, regularly sampled, and densely-measured with constant error variance; however the FFT approach is fast, $\mathcal{O}(NM \log(NM))$, when these circumstances are satisfied (for example texture synthesis or image denoising). The FFT was used to generate the "wood" texture of Fig. 6.

2.2.3 Noncausal GMRFs

In general, we can rewrite (15), stacking the random field into a vector $\vec{x}$ (by rows or by columns):

$$G\vec{x} = \vec{w}. \tag{22}$$

	.009	.014	.006	
.001	.116	.250	.141	.009
.052	.551		.551	.052
.009	.141	.250	.116	.001
	.006	.014	.009	

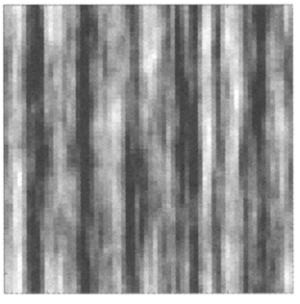

FIGURE 6 A fourth-order stationary MRF, synthesized using the FFT method, based on the coefficients[6] $a_{i,j}$.

[6]The coefficients in the figure are rounded; the exact values may be found in [16].

If the field is small (a few thousand elements) and G is invertible we can, in principle, solve for the prior covariance P of the entire field:

$$G\vec{x} = \vec{w} \Longrightarrow GPG^T = \text{cov}(\vec{w}) = W \Longrightarrow P = G^{-1}WG^{-T}. \tag{23}$$

Prior samples can be computed using the Cholesky decomposition[7] of P, least-squares estimates computed by inverting P, and posterior samples from the Cholesky decomposition of the estimation error covariance $\tilde{P}$, however none of these operations are practical for large fields.

Instead, methods of *domain decomposition* (e.g., nested-dissection or Marching methods) [12, 27, 28, 29] are often used. The goal is to somehow break a large field into smaller pieces.

Suppose we have a first-order GMRF; although the individual field elements cannot be ordered causally, if we divide the field into columns

$$\boldsymbol{x} = [\vec{x}_1\, \vec{x}_2 \ldots \vec{x}_N] \tag{24}$$

then the sequence of columns are a one-dimensional first-order Markov process; that is,

$$E[\vec{x}_{i+1}|\vec{x}_1, \ldots, \vec{x}_i] = E[\vec{x}_{i+1}|\vec{x}_i]. \tag{25}$$

Since the field is Gaussian, we can express $\vec{x}_i$ via a linear model

$$\vec{x}_{i+1} = A_i\vec{x}_i + B_i w_i. \tag{26}$$

The estimates $\hat{\vec{x}}_i$ still need to be computed acausally, however this can be accomplished efficiently, $\mathcal{O}(\min(NM)^3)$ using the RTS smoother [30]. A tremendous number of variations exist: different decomposition schemes, approximate or partial matrix inversion, reduced update etc., of which one multiscale approach is described in Section 3.

2.3 Gibbs Random Fields

Gibbs random fields (GRFs) are random fields characterized by neighboring-site interactions. These were originally used in statistical physics [31, 32] to study the thermodynamic properties of interacting particle systems, such as lattice gases, and their use in image processing was popularized by papers of Besag [33, 34] and Geman and Geman [5]. The neighboring interactions in GRFs lead to effective and intuitive image models — for example, to assert the piecewise-continuity of image intensity. Hence, the GRF is

[7]A common matrix operation, available in standard mathematics packages such as MATLAB.

often used as a prior model in a Bayesian formulation to enforce image constraints.

Mathematically, a GRF $\boldsymbol{x}$ is described/defined by a Gibbs distribution:

$$p(\boldsymbol{x}) = \frac{1}{Z} e^{-\beta E(\boldsymbol{x})}. \tag{27}$$

Although it would appear that we are focusing once again on probability densities, $p(\boldsymbol{x})$ is, in fact, never evaluated. All inferences of $\boldsymbol{x}$ will take place implicitly, strictly through evaluations of the *energy function* $E(\boldsymbol{x})$. The energy function is normally written as a sum of local interaction terms, called *clique potentials*:

$$E(\boldsymbol{x}) = \sum_{c \in C} V_c(\boldsymbol{x}), \tag{28}$$

where c is a *clique*, i.e., either a single site or a set of sites that are all neighbors of each other; $V_c(\cdot)$ is the clique potential, a function of the random variables associated with c; and C is the set of all possible cliques. Finally, $\beta > 0$ is a constant, also known as the temperature parameter, and

$$Z = \sum_{\boldsymbol{x}'} e^{-\beta E(\boldsymbol{x}')} \tag{29}$$

is a normalization constant, known as the *partition function*. The enormity of the sum prevents Z from being evaluated for all but the tiniest problems.

The attractiveness of the GRF approach lies in allowing a simple, intuitive energy function E to describe enormously complicated probability functions p. For example an $E(\boldsymbol{x})$, involving only elements of $\boldsymbol{x}$, such as in the standard binary Ising model [31]

$$E(\boldsymbol{x}) = +b_1 \sum_{i,j} x_{i,j} x_{i,j-1} + b_2 \sum_{i,j} x_{i,j} x_{i-1,j} \tag{30}$$

implicitly describes a *prior* model on $\boldsymbol{x}$, whereas the introduction of measurements or constraints

$$E(\boldsymbol{x}|\boldsymbol{y}) = a \sum_{i,j} y_{i,j} x_{i,j} + b_1 \sum_{i,j} x_{i,j} x_{i,j-1} + b_2 \sum_{i,j} x_{i,j} x_{i-1,j} \tag{31}$$

just as simply implicitly describes a *posterior* model. In both examples, the cliques are either a single site $\{(i,j)\}$, or two neighboring sites $\{(i,j),(i,j-1)\}$, $\{(i,j),(i-1,j)\}$. Typical realizations of the prior model are shown in Fig. 7.

Given the above brief introduction to GRFs, it is natural to ask how they relate to the MRFs and how to address the basic random field problems (Section 2).

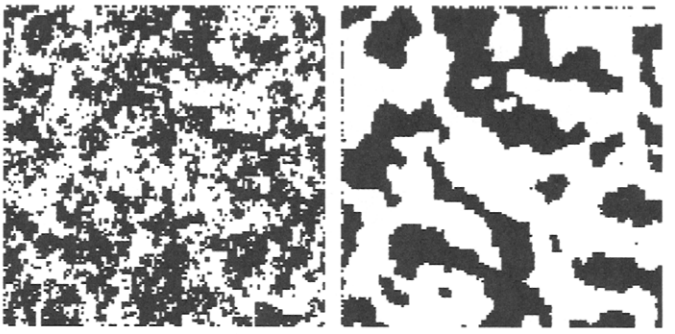

FIGURE 7 Typical samples of the Ising model (30). Left: $b_1 = b_2 = -0.4$. Right: $b_1 = b_2 = -1.0$.

First, according to the Hammersley-Clifford theorem [33], the GRF and MRF are equivalent. As a result, the Ising model described above is an MRF and the Gauss-Markov models described in Section 2.2 have associated energy functions and clique potentials. For example, the energy function for the a first-order noncausal GMRF model is [35]

$$E(\boldsymbol{x}) = \frac{1}{2}\sigma^2 \sum_{i,j} [x_{i,j} - \sum_{(k,l) \in \mathcal{N}_{i,j}} a_{i-k,j-l} x_{k,l}]^2 \tag{32}$$

from which the clique potentials can be identified.

Second, in considering basic problems such as estimation or parameter estimation, the typical ML or MAP approach

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{x}|\boldsymbol{\theta}), \qquad \hat{\boldsymbol{x}}_{MAP} = \arg\max_{\boldsymbol{x}} p(\boldsymbol{x}|\boldsymbol{y}),$$

is impractical for GRFs because actually evaluating $p(\boldsymbol{x})$ requires the calculation of Z in (29), which is impossible. Instead, local approaches, based in evaluations of evergy, are required.

Parameter Estimation

Besag proposed to maximize the *pseudo-likelihood* [33]

$$q(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{i,j} p(x_{i,j} | \{x_{k,l}, (k,l) \in \mathcal{N}_{i,j}\}, \boldsymbol{\theta}), \tag{33}$$

which is made up by a set of conditional probabilities, with respect to $\boldsymbol{\theta}$. These conditional probabilities, also called the *local characteristics*, are easily calculated since the partition function no longer appears and each term in (33) can be evaluated locally:

$$p(x_{i,j} | \{x_{k,l}, (k,l) \in \mathcal{N}_{i,j}\}, \boldsymbol{\theta}) = \frac{e^{-\beta \sum_{c \ni (i,j)} V_c(\boldsymbol{x})}}{\sum_{x_{i,j}} e^{-\beta \sum_{c \ni (i,j)} V_c(\boldsymbol{x})}} \tag{34}$$

Sampling

We will see that finding the estimate $\hat{\boldsymbol{x}}$ which maximizes the posterior $p(\boldsymbol{x}|\boldsymbol{y})$ is a special case of randomly sampling from $p(\boldsymbol{x}|\boldsymbol{y})$, thus sampling is the more fundamental problem, discussed first.

A number of techniques have been developed for sampling, such as the Metropolis algorithm [36] or the Gibbs sampler [5]. The Gibbs sampler is applicable to both continuous-state [35] and discrete-state problems, the discrete case as follows:

Step 0: Initialize with a sample from a i.i.d. random field.

Step 1: Scan the image from left to right, top to bottom. At each site (i, j), sample $x_{i,j}$ from the conditional distribution $p(x_{i,j}|\{x_{k,l}, (k, l) \in \mathcal{N}_{i,j}\})$. Whereas the *joint* distribution $p(\boldsymbol{x})$ cannot be evaluated, the *conditional* is straightforward, because the partition function (29) is a simple sum over the possible discrete values of $x_{i,j}$.

Step 2: Repeat Step 1 many times; after many iterations $\boldsymbol{x}$ is a statistical sample of the random field [5].

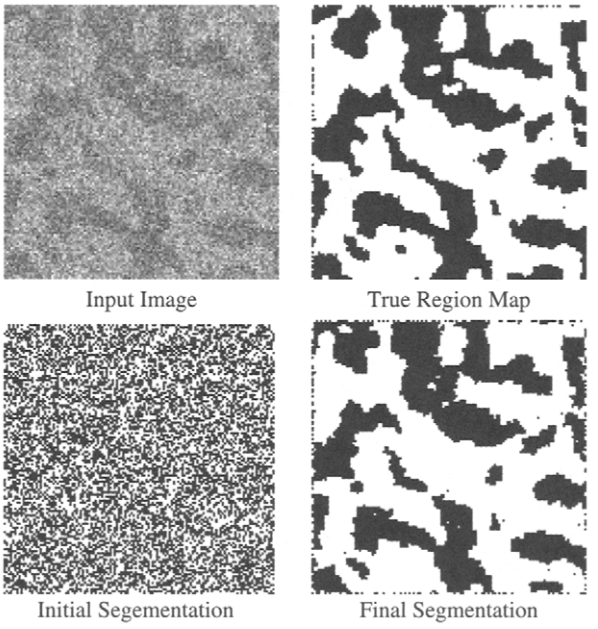

FIGURE 8 Segmentation by simulated annealing.

Estimation

If we consider a GRF (27), as we reduce the temperature $T = 1/\beta$ random samples of $p(\boldsymbol{x})$ become increasingly biased towards *likely* $\boldsymbol{x}$. In the limit as $T \longrightarrow 0$, random samples of $p(\boldsymbol{x})$ must approach the globally optimum sample, which is the MAP estimate $\hat{\boldsymbol{x}}_{MAP}$.

The difficulty, however, is that the Gibbs sampler would never actually converge to this optimum if $T = 0$ is set; instead, the optimum needs to be found through a gradual annealing process, in which a temperature parameter is slowly decreased:

Step 0: Start with an i.i.d. sample image and an initial temperature $T = T_0$.

Step 1: Perform one iteration of the Gibbs sampler, finding $\boldsymbol{x}_k$.

Step 2: Select a lower temperature T_{k+1} and repeat Step 1 till convergence.

Theoretically, this produces a global optimum when $k \longrightarrow \infty$ and $T \longrightarrow 0$ [5], where k is the number of iterations. In practice, however, the temperature has to be lowered very slowly, e.g., according to $T(k) = C/\log(1 + k)$, which is computation-intensive. Hence suboptimal techniques are often used, among them are Besag's iterative conditional mode (ICM) [34] and the mean field theory [37, 38].

Finally, Fig. 8 provides an example on how the GRF can be used in a Bayesian formulation. Specifically, consider the problem of segmenting an image into two types of image regions, labeled -1 and $+1$. Suppose the true labels are described by binary field $\boldsymbol{x}$, and we are given corrupted measurements $\boldsymbol{r} = m\boldsymbol{x} + \boldsymbol{v}$, where $\boldsymbol{v}$ is an additive zero-mean white Gaussian noise. The problem of segmentation, then, is to label each pixel of $\boldsymbol{r}$ to either -1 or $+1$; that is, we seek to estimate $\boldsymbol{x}$. In a Bayesian formulation, we find $\boldsymbol{x}$ as

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \log p(\boldsymbol{x}|\boldsymbol{r}) \propto \arg\max_{\boldsymbol{x}} \{\log p(\boldsymbol{r}|\boldsymbol{x}) + \log p(\boldsymbol{x})\}. \tag{35}$$

If an Ising model is adopted for $p(\boldsymbol{x})$ to enforce the region continuity constraint (i.e., neighboring pixels are likely to be in the same region), it can be shown easily that $p(\boldsymbol{x}|\boldsymbol{r})$ is also an Ising model (with an external field) and the segmentation can be obtained by simulated annealing or ICM or the MFT. For more details and more realistic examples, see [37].

3 Multiscale Random Fields

There are elegant, natural fits between Markov models and hierarchical structures. Consider the generalization of the boundary in Fig. 2 to the boundary $\boldsymbol{x}_b$ shown in Fig. 9. From (10) we see that each of the quadrant fields $\boldsymbol{x}_i$ satisfies

$$p(\boldsymbol{x}_i|\boldsymbol{x}_b, \boldsymbol{x}_j, j \neq i) = p(\boldsymbol{x}_i|\boldsymbol{x}_b). \tag{36}$$

That is, given some properly-chosen boundary state $\boldsymbol{x}_b$, the closed regions separated by the boundary become conditionally independent, and can therefore be processed separately. Constructing further boundaries within each of these regions allows the domain decomposition to continue recursively, leading to a hierarchical structure.

Similarly, given a Gibbs distribution (27) and local energy function (28), this fine-scale model implicitly defines the energy function corresponding to coarse-scale regions, each containing some number of fine-scale pixels. Constructing a

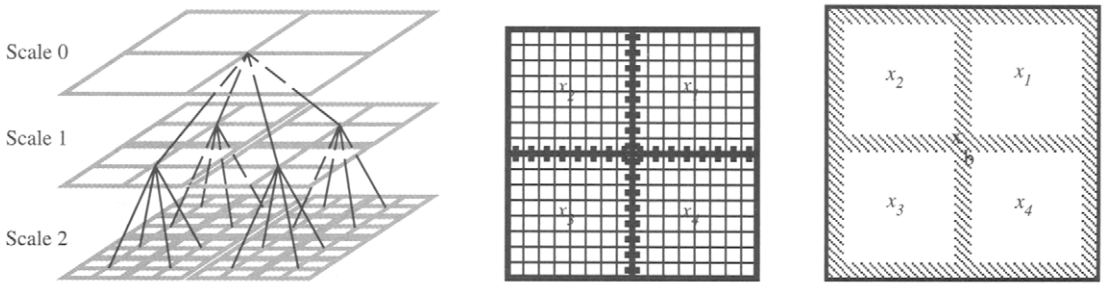

FIGURE 9 Hierarchical modeling of random fields on a quad-tree (left). Boundary models for MRFs: the state x_b kept at each node (middle) can be chosen to decorrelate the four quadrants represented by that node, allowing a decomposition. Multiresolution models for GRFs: knowing the Gibbs model at some finest scale implicitly defines a model for coarse-scale pixels (right) as the aggregation of constraints between the contained finest-scale pixels.

nested set of increasingly coarse regions once again leads to a hierarchy.

The existence of a hierarchical structure does not alone motivate its use. However, the sampling and estimation of large, multidimensional random fields is computationally very complex, and hierarchical structures have shown considerable promise in reducing this complexity. This section will discuss two hierarchical approaches, first a brief discussion of the discrete-state case, and secondly a more detailed look at Gauss-Markov random fields on trees.

3.1 Discrete-State Models

We seek hierarchical approaches to modeling and computation for two reasons: first, for those models which are inherently multiscale (concrete, for example, contains cracks ranging in size from sub-micron to millimetre) and, secondly, to more effectively model those non-local relationships on the finest scale, but which become progressively more local (and simple) on coarser levels.

A wide variety of hierarchical [39–42] and region-based [43] methods exist. The most common discrete-state problems in image processing relate to image classification and segmentation[8] and are designed for estimation [39, 41, 42], thus guided and driven by measurements, as opposed to random sampling, which is purely model based. If the image noise is modest, then in many cases the coarse scales serve mostly as a weak regularizer for a densely-measured, well-conditioned fine-scale estimation problem. On the other hand, problems involving sparse or no measurements, although less studied, possess a finest scale which is poorly conditioned, and where the coarse scales have a great deal to offer.

Consider, for example, the Ising model of (30), where $b_1 = b_2 = b$.

For small β the sampled field is essentially random, thus all samplers, whether flat or hierarchical, converge quickly. However, as β increases the structures present in x grow larger, and the longer-range relationships become increasingly difficult to deduce from a local model, thus hierarchical methods begin to outperform (Fig. 10). The hierarchical samplers proceed from coarse to fine, in a single pass, sampling for some iterations at each scale.

The key challenge relates to the definition of coarse-scale models. The coarsification of (30) is not obvious [40], and is most often circumvented by defining coarse-scale models implicitly in terms of the finest-scale by projection [41]. In the Ising model this implies the widely-used model $\beta_s = 2^s\beta$ at s scales above the finest.

The problem is that this is wrong. Figure 10(a) makes it clear that for small β the coupling should *decrease* with scale. Using $\beta_s = 2^s\beta$ leads to stiff, large-scale structures sampled on coarse scales which need to be undone at finer scales. If a properly-renormalized model is used with the correct value of β at each scale, then the sampler needs only to insert the details unresolvable at the coarser scale, a *much* easier task than undoing incorrect structure, and thus converging *far* faster, as seen in Fig. 10(b).

3.2 Continuous-State Models

The multiscale statistical modeling of continuous random fields based on a particular multiscale tree structure has been the focus of research for several years [11, 12, 15, 16]. The motivations driving this research are broad, including $1/f$-processes, stochastic realization theory, and a variety of application areas (computer vision, synthetic aperture radar, groundwater hydrology, ocean altimetry and hydrography). Although the method is applicable more broadly, we will focus on the GMRF case.

A statistical characterization of a random field in multiscale form (detailed below) possesses the following attributes:

- Efficient prior and posterior sampling
- Efficient least-squares estimation [11, 15] (i.e., computing $\hat{\boldsymbol{x}}$ in (8))
- Efficient likelihood calculation [14] (i.e., computing $p(\boldsymbol{y}|\boldsymbol{\theta})$ in (5))
- The ability to accommodate non-local or distributed measurements

[8] See Chapters 4.9–4.13 of this book.

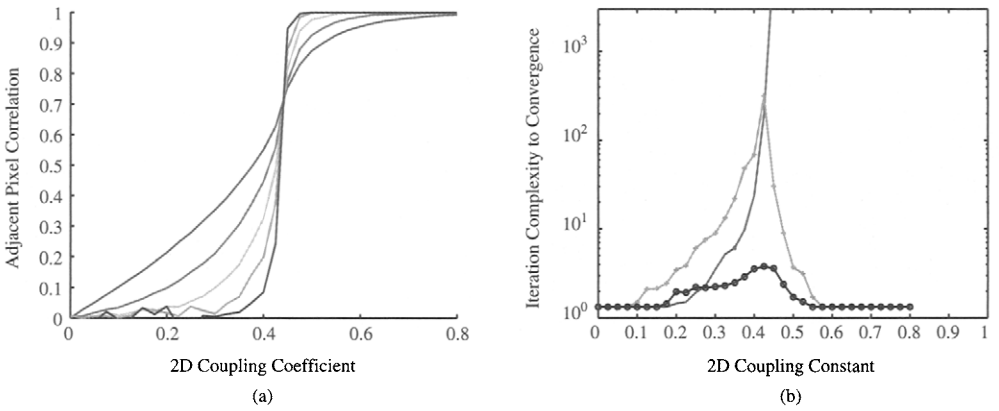

FIGURE 10 Hierarchical sampling for the Ising model (30). (a) The correlation between adjacent pixels as a function of the coupling β and scale. Coarser scales (steep) are progressively random or uniform, depending on the value of β relative to criticality (at $\beta \approx 0.44$). (b) The convergence of a flat (divergent), standard hierarchy (intermediate), and renormalized hierarchical model (bottom) as a function of coupling, for three different scales. Around criticality, where structures are most complex, the renormalized model converges more than 100 times faster.

- Computational complexity is unaffected by nonstationarities in the random field or in the measurements (compare with the FFT, Section 2.2).

From (36), the Gaussianity of the field allows us to write

$$E\left[\vec{x}_i|\vec{x}_b, \vec{x}_j, j \neq i\right] = E\left[\vec{x}_i|\vec{x}_b\right] = \mathcal{A}_i\vec{x}_b \tag{37}$$

in other words,

$$\vec{x}_i = \mathcal{A}_i\vec{x}_b + \vec{\omega}_i \tag{38}$$

where $\vec{\omega}_i$ is uncorrelated with $\vec{\omega}_j$ for $i \neq j$. There is, however, no reason to content ourselves with limiting the decomposition of the field into four quadrants. We can proceed further, creating a boundary $\boldsymbol{x}_{b_i}$ within quadrant i; from (38) it follows that we can write the boundary as

$$\vec{x}_{b_i} = A_i\vec{x}_b + \vec{w}_i. \tag{39}$$

We can continue the successive subdivision of the field into smaller pieces; Fig. 9 shows such a set of boundaries organized onto a tree structure. Now *every* vector on our tree is a boundary; that is, our random field is described in terms of a set of hierarchical boundaries, terminating with an individual pixel at each tree element at the finest level of the tree.

Let $\vec{x}_{s,i}$ be the ith boundary at scale s on the tree; then

$$\vec{x}_{s,i} = A_{s,i}\vec{x}_{s-1,p(i)} + B_{s,i}\vec{w}_{s,i} \tag{40}$$

where $p(i)$ represents the "parent" of the ith boundary, and $\vec{w}$ is a white-noise process.

Having developed a multiscale model (40) for a random field, we can also introduce measurements

$$\vec{y}_{s,i} = C_{s,i}\vec{x}_{s,i} + D_{s,i}\vec{v}_{s,i} \tag{41}$$

where $\vec{v}$ is a white-noise process. Local point measurements are normally associated with the individual pixels at the finest level of the tree, however with the appropriate definition of $\vec{x}_{s,i}$ at coarser levels of the tree, non-local measurements can also be accommodated.

It is (40), (41) which form the basis for the multiscale environment; *any* random field (whether Markov or not) which can be written in this form leads to efficient algorithms for sample paths, estimation and likelihood calculation, as mentioned earlier.

It should be noted that (40) is essentially a distributed marching algorithm, here marching over boundaries on a tree rather than across space (Section 2.2). Indeed, the marching principle applies to determining $A_{s,i}$, $B_{s,i}$: if we write each boundary as $\vec{x}_{s,i} = L_{s,i}\vec{x}$ where $\vec{x}$ is the original random field with covariance P, then

$$A_{s,i} = \left\{L_{s,i}PL_{s-1,p(i)}\right\}\left\{L_{s-1,p(i)}PL_{s-1,p(i)}\right\}^{-1} \tag{42}$$

$$B_{s,i} = \left\{L_{s,i}PL_{s,i}\right\} - A_{s,i}\left\{L_{s-1,p(i)}PL_{s-1,p(i)}\right\}^{-1}A_{s,i}^T. \tag{43}$$

With the determination of the parameters $A_{s,i}$, $B_{s,i}$, the definition of the multiscale random field is complete. There are four issues to address to put such a model into practice:

- How does the structure (40), (41) lead to an fast prior/posterior sampler?
  Prior sampling follows trivially from (40): the scale-to-scale dynamics are simulated, with random white noise synthesized for $\vec{w}_{s,i}$.
  Posterior sampling is possible because the multiscale estimation errors *also* obey a multiscale model [13].

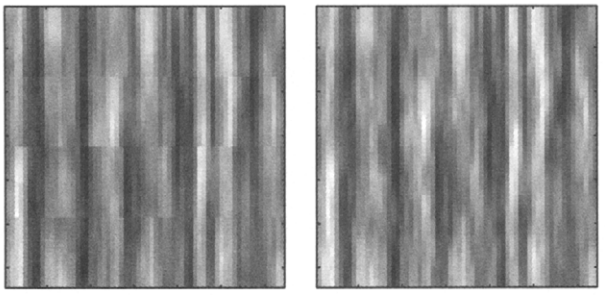

FIGURE 11 Multiscale estimation of random-field textures. Left: Reduced-order model with texture artifacts. Right: Overlapped multiscale model.

Adding a sample from this error model to the estimates yields a posterior sample.

- How does the structure (40), (41) lead to an efficient estimator?
  The Kalman filter and Rauch-Tung-Striebel smoother [30] can be applied to (40) to compute estimates; however since every state $x_{s,i}$ is a boundary — only a small subset of the random field — the matrices operations at each tree node are modest. Algorithms have been published and are available on the Internet [12, 15].
- How are the boundaries $L_{s,i}$ determined?
  For Gauss-Markov random fields $\vec{x}$, Section 2.2 discusses the criteria for a boundary to conditionally decorrelate subsets of the field. For non-Markov fields the "optimal" choice of $L_{s,i}$ is much more complicated; further discussion may be found in [16].
- Can further computational efficiency be gained via approximations?
  For large random fields (10^6 pixels), an "exact" solution may require orders of magnitude more memory and computational effort than an approximate method yielding essentially the same estimates. This is typically accomplished by selecting $L_{s,i}$ to very nearly, although not perfectly, decorrelate subsets of the field. Such approximations, and methods of dealing with possible resulting artifacts, are discussed at length in [16].

3.3 Examples

We will briefly survey three examples. Our first example [12, 16] continues with the MRF texture of Fig. 6. Figure 11 shows two estimates of this texture based on noisy measurements: using a reduced-order multiscale model, which illustrates the artifacts which may be present with poorly-approximated models, and using an overlapped multiscale model, having the same computational complexity, but free of artifacts.

Figure 12 shows a related example, which illustrates the ability to estimate posterior samples, given a sparse set of measurements from a random field. The nonstationarity of the measurement structure makes the FFT inapplicable, but poses no problems for the multiscale approach.

Finally Fig. 13 highlights two random fields of significant interest in oceanography [15] and remote-sensing: the estimation of ocean height (altimetry) and temperature (hydrography) fields from sparse, nonstationary, noisy measurements.

4 A Nonlinear/Non-Gaussian Model: the Gaussian Mixture

Many real-world images, especially those with texture, contain nonlinear and possibly long range spatial interactions that can not be easily captured by linear (i.e., GMRF type) models. For example, an image containing sharp gray level transitions (e.g., a checkerboard) cannot be described by a stationary autoregressive model (a type of GMRF). In such cases, nonlinear (i.e., non-Gaussian) and possibly high-order models are what is needed [20].

The Gaussian mixture has attracted a lot of attention as a versatile model for non-Gaussian random variables [44, 45]. Indeed, under relatively mild conditions, the probability density function (PDF) of a non-Gaussian random variable can be approximated arbitrarily closely by a Gaussian mixture [46]. In this section, we briefly describe a Gaussian mixture based image model and illustrate its efficacy through

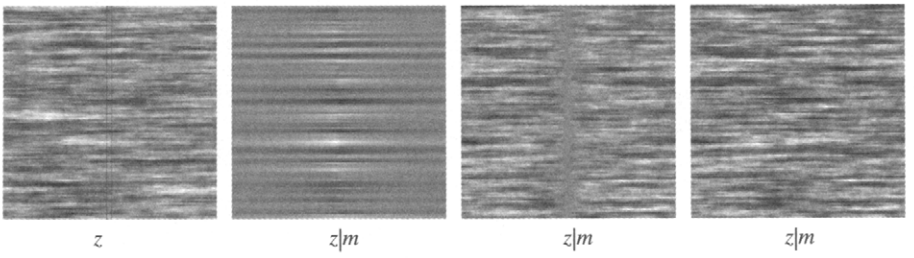

FIGURE 12 The process of posterior sampling. The leftmost two panels show a sample from an anisotropic prior model, and estimates based on the central measured columns. The third panel shows the sampled estimation error, where a low-variance zero-mean band can be seen where the estimation uncertainties are small, around the measurements. The final panel shows the sampled posterior, consistent with both the measurements and the prior statistics.

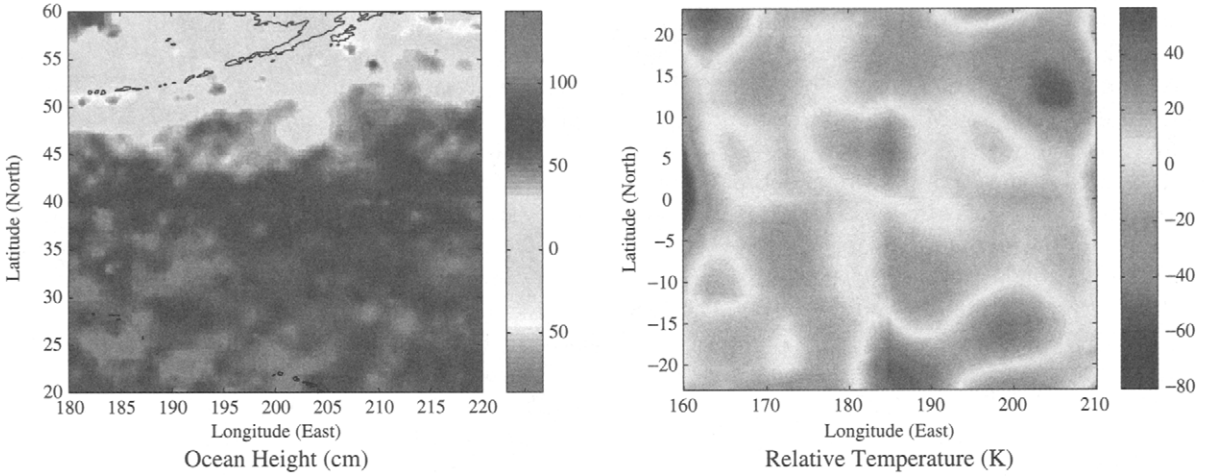

FIGURE 13 Multiscale estimation of remotely-sensed fields. Left: North-Pacific altimetry based on Topex/Poseidon data. Right: Equatorial-Pacific temperature estimates based on in-situ ship data (see color insert).

applications in anomaly detection and texture modelling. More details of this model and related work can be found in [25, 47] and the references therein.

4.1 The Gaussian Mixture Model

We first review the concept of a multivariate Gaussian mixture and then describe how it can be used to model images. Suppose $\boldsymbol{u} = [u_1, u_2, \ldots, u_n]^T$ is a column random vector whose PDF has the following form:

$$p(\boldsymbol{u}) = \sum_{k=1}^{K} \pi_k p_k(\boldsymbol{u}) \tag{44}$$

where $p_k(\cdot)$'s are multivariate Gaussians with mean $\boldsymbol{m}_k$ and covariance matrix C_k, and

$$\pi_k > 0, \qquad \text{for } k = 1, 2, \ldots, K, \text{ and} \qquad \sum_{k=1}^{K} \pi_k = 1. \tag{45}$$

Then $p(\cdot)$ is called a multivariate Gaussian mixture. A related result is the conditional PDF of a Gaussian mixture [47], [54]:

$$p(u_n|u_{n-1}, u_{n-2}, \ldots, u_1) = \sum_{k=1}^{K} \pi_k' p_k(u_n|u_{n-1}, u_{n-2}, \ldots, u_1) \tag{46}$$

where $p_k(u_n|u_{n-1}, u_{n-2}, \ldots, u_1)$ are the conditional PDFs obtained from $p_k(\boldsymbol{u})$ and

$$\pi_k' = \pi_k \frac{\int p_k(\boldsymbol{u}) du_n}{\int p(\boldsymbol{u}) du_n} \tag{47}$$

are functions of $u_{n-1}, u_{n-2}, \ldots, u_1$.

Multivariate Gaussian mixtures have many applications in statistics [44], especially in clustering [45], and in image modeling. Specifically, suppose $\boldsymbol{x} = \{x_{i,j}\}$ is an image with a given neighborhood (Fig. 4). Suppose for each site (i, j) we define a pixel block (Fig. 14)

$$\boldsymbol{x}_{i,j} = \{x_{i,j} \text{ and } x_{k,l}, (k, l) \in \mathcal{N}_{i,j}\} \tag{48}$$

and organize it into a one-dimensional vector according to raster ordering. Then, a simple multivariate Gaussian mixture image model can be obtained by adopting a multivariate Gaussian mixture for the vector $\boldsymbol{x}_{i,j}$ [20]. Notice that in order for such a model to be useful in practice, some stationarity conditions are usually imposed. For example, it is commonly assumed that the multivariate Gaussian mixture $p(\boldsymbol{x}_{i,j})$ is invariant with respect to i, j, i.e.,

$$p(\boldsymbol{x}_{i,j}) = p(\boldsymbol{x}_{i',j'}) \tag{49}$$

for any i, j and i', j'.

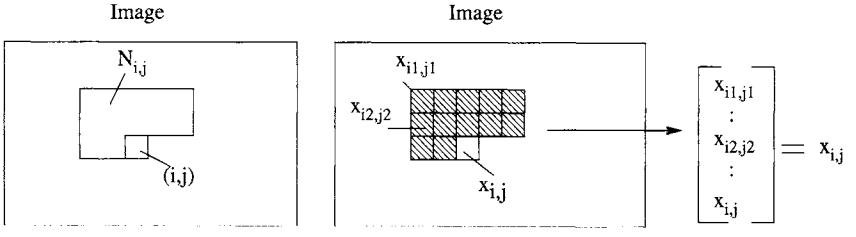

FIGURE 14 Illustration of a pixel block.

More elaborate stationary multivariate Gaussian mixture models have been considered for textures with more complicated structures, for example, see the multiresolution model of [20] and the wavelet-based multiresolution model of [25] (also described briefly below).

Finally, compared to the random field models described in previous sections, these block based Gaussian mixture model may seem "incomplete" — providing only the marginal PDFs but not the joint PDF of the entire field, $p(\boldsymbol{x})$. This issue can be resolved in two steps: first, we can find the conditional PDF

$$p(x_{i,j}|x_{k,l}, (k,l) \in \mathcal{N}_{i,j}) = \frac{p(\boldsymbol{x}_{i,j})}{\int_{x_{i,j}} p(\boldsymbol{x}_{i,j}) dx_{i,j}} \tag{50}$$

which, under a somewhat larger neighborhood system, induces an MRF and consequently a joint PDF $p(\boldsymbol{x})$ (a Gibbs distribution) for the entire random field [33].

4.2 Some Applications

Here, we briefly describe two applications of the block-based Gaussian mixture image model: anomaly detection and multiresolution texture modeling.

Anomaly Detection. Many practical image processing applications amount to the detection of small objects (anomalies) that are different from the dominating background texture and clutter. For example, in aerial photo interpretation, one may want to detect small man-made objects from fields and vegetation. Similarly, in manufacturing quality inspection, one may want to detect small defects in fabric and paper.

An effective technique for anomaly detection is to compute the image prediction error [48], [49], [50], by computing the prediction $\hat{x}_{i,j}$ of each image pixel using its neighboring pixels:

$$e_{i,j} = x_{i,j} - \hat{x}_{i,j}. \tag{51}$$

If the magnitude of the prediction error is large, the corresponding pixel is then considered different from its surroundings and can be considered to be part of an anomaly. The remaining question is how to compute the prediction $\hat{x}_{i,j}$.

Suppose we model the image by a random field with neighborhood $\mathcal{N}_{i,j}$. Then, an optimal predictor in the minimum mean square error sense is

$$\hat{x}_{i,j} = E[x_{i,j}|x_{k,l}, (k,l) \in \mathcal{N}_{i,j}]. \tag{52}$$

When the image model is Gaussian (e.g., Gauss-Markov), the optimal predictor is a linear combination of the neighboring pixels and the prediction calculation is easy. However, as mentioned previously, many real-world images are dominated by non-linear interactions which cannot be captured well by Gaussian models. In such cases, a linear predictor based on a Gaussian model tends to generate large prediction error magnitudes even when anomaly is not present, thereby rendering the prediction error technique ineffective.

Clearly one solution to this problem is to use a non-Gaussian model, in which case the optimal predictor (52) is non-linear and, in general, its analytic expression is difficult, if not impossible, to obtain. Fortunately, because of (46) and (47), the Gaussian mixture image model described in Section 4.1 is one of the few cases where an analytic expression of the optimal nonlinear predictor can be found [47]. Furthermore, because a non-Gaussian model can be well approximated by an appropriately chosen Gaussian mixture model, a wide range of non-Gaussian image texture/clutter can be modeled by such mixtures.

Figure 15 illustrates one example of anomaly detection, in which the object (anomaly) of interest is a helicopter in the middle of the image in the presence of background texture and clutter. The detection results are presented as a prediction error image (anomaly) and a spatial average of the squared prediction error (smoothed anomaly). Bright spots in the latter image correspond to likely anomalies. For comparison purposes detection results are shown for both Gaussian mixture and Gauss-Markov models. The results illustrate the efficacy of the Gaussian mixture model: it produces better detection and fewer false alarms.

Multiresolution Texture Modeling. Many real world textured images contain nonlinear and long range interactions. Gaussian mixtures are effective in capturing nonlinear interactions, but only on a relatively short range, to avoid excessive model complexity. To capture long range nonlinear interactions, however, we can extend the Gaussian mixture into a multiresolution approach, for example through the use of wavelets [25], where the desired long range interactions may be realized through short range interactions at a coarse

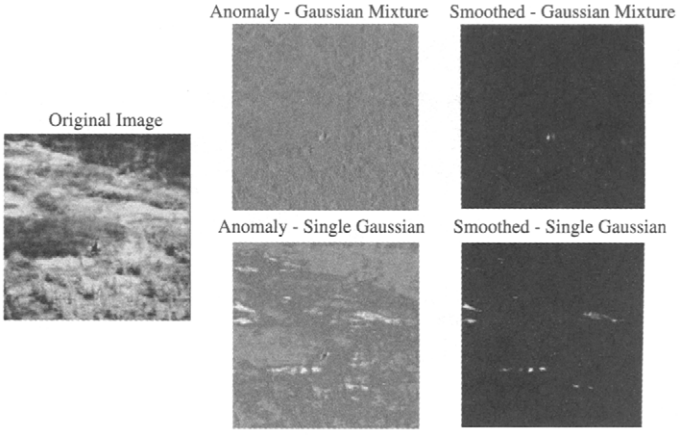

FIGURE 15 An example of anomaly detection.

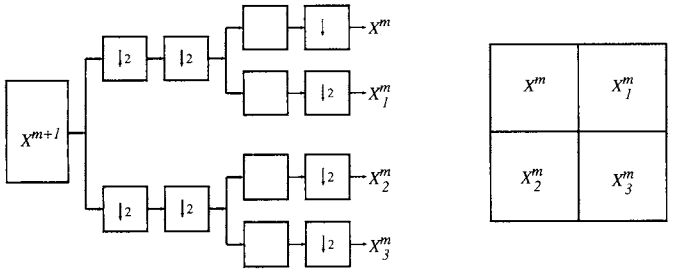

FIGURE 16 Wavelet decomposition.

resolution. In other words, a high-order fine resolution model may be achieved through a set of low-order coarse resolution models.

Let L be a square lattice and $\boldsymbol{x} = \{x_{i,j}, (i,j) \in L\}$ be a random field used to model a class of images. Suppose L represents the finest resolution and denote $\boldsymbol{x}$ by $\boldsymbol{x}^0$. Suppose that for some positive integer M, $\boldsymbol{x}^0$ has a wavelet expansion

$$\boldsymbol{x}^0 \sim \{\boldsymbol{w}^{-1}, \boldsymbol{w}^{-2}, \ldots, \boldsymbol{w}^{-M}, \boldsymbol{x}^{-M}\} \tag{53}$$

where the $\boldsymbol{w}^m$, $m = -1, -2, \ldots, -M$, are the wavelet coefficients at various levels, and where each $\boldsymbol{w}^m$ contains three subsets $\boldsymbol{w}_1^m, \boldsymbol{w}_2^m, \boldsymbol{w}_3^m$ corresponding to the LH, HL, and HH orientations, as shown in Fig. 16. The sign $\sim$ denotes equivalence in the sense that the wavelet coefficients can be used to reconstruct $\boldsymbol{x}^0$.

Since $\boldsymbol{x}$ is completely determined from its wavelet coefficients, we can model $\boldsymbol{x}$ by modeling the wavelet coefficients. In other words, $\boldsymbol{x}$ can be modeled by specifying the joint probability density:

$$\begin{aligned} p(\boldsymbol{x}^0) &\sim p(\boldsymbol{w}^{-1}, \boldsymbol{w}^{-2}, \ldots, \boldsymbol{w}^{-M}, \boldsymbol{x}^{-M}) \\ &= p(\boldsymbol{w}^{-1} | \boldsymbol{w}^{-2}, \ldots, \boldsymbol{w}^{-M}, \boldsymbol{x}^{-M}) \\ &\quad \times p(\boldsymbol{w}^{-2} | \boldsymbol{w}^{-3}, \cdots, \boldsymbol{w}^{-M}, \boldsymbol{x}^{-M}) \ldots p(\boldsymbol{x}^{-M}) \\ &= p(\boldsymbol{w}^{-1} | \boldsymbol{x}^{-1}) p(\boldsymbol{w}^{-2} | \boldsymbol{x}^{-2}) \ldots p(\boldsymbol{x}^{-M}) \end{aligned} \tag{54}$$

where we have used the fact that $(\boldsymbol{w}^m, \ldots, \boldsymbol{w}^{-M}, \boldsymbol{x}^{-M}) \sim \boldsymbol{x}^{(m+1)}$ for $-M < m < -1$. Now, the problem of specifying $p(\boldsymbol{x}^0)$ becomes that of specifying densities $p(\boldsymbol{x}^{-M})$ and $p(\boldsymbol{w}^m | \boldsymbol{x}^m)$, where the complexity (model-order) for these latter models can be considerably lower than that for the fine resolution model.

Since it is well known that the wavelet transform reduces correlation within and between resolutions, two assumptions can be made to simplify the modeling of $p(\boldsymbol{w}^m | \boldsymbol{x}^m)$:

- *Assumption 1* (Inter-band Conditional Independence): $\boldsymbol{w}_n^m$, $n = 1, 2, 3$ are independent given $\boldsymbol{x}^m$.
- *Assumption 2* (Inter-band Independence): $\boldsymbol{w}_n^m$, $n = 1, 2, 3$ are independent of each other as well as independent of $\boldsymbol{x}^m$.

Assumption 2 is stronger than Assumption 1 and seems to work well for textures that are characterized by random micro-structures. However, when the texture contains long range correlation and large structural elements, the weaker and more realistic Assumption 1 is needed [25].

Now, we can model the random field at each resolution by using the block-based Gaussian mixture model as described in Section 4.1. For example, $\boldsymbol{x}^{-M}$ can be modeled through

$$p(x_{i,j}^{-M}, x_{k,l}^{-M}, (k,l) \in \mathcal{N}_{i,j}) = \sum_{\kappa=1}^{K} \pi_\kappa p_\kappa(x_{i,j}^{-M}, x_{k,l}^{-M}, (k,l) \in \mathcal{N}_{i,j}) \tag{55}$$

where $\mathcal{N}_{i,j}$ is a causal or non-causal neighborhood and $p_\kappa(\cdot)$ are Gaussian density functions. Similar block Gaussian mixture PDFs can also be defined for each wavelet subband $\boldsymbol{w}_n^m$ (conditioned on $\boldsymbol{x}^m$), where $m = 1, 2, \ldots, M$ and $n = 1, 2, 3$. More details can be found in [25].

The efficacy of the multiresolution Gaussian mixture model can be illustrated through an example of texture synthesis. Specifically, in Fig. 17 we have shown an original texture image (from the Brodatz texture set [51]), synthesis from a multiresolution Gaussian model (i.e., multiresolution AR model), and synthesis from multiresolution Gaussian mixture models (one with a causal and one with a non-causal neighborhood system). The multiresolution Gaussian mixture models provide significantly better results than those from the multiresolution Gaussian model. Furthermore, the non-causal neighborhood system provides better results than those

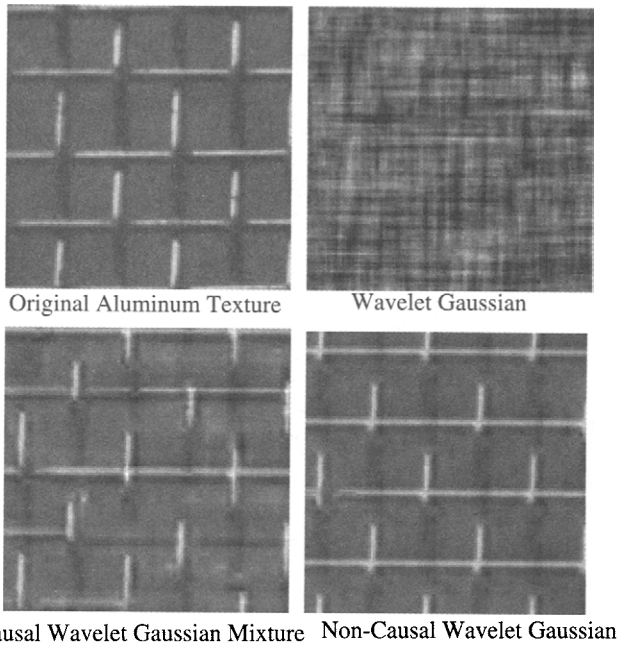

FIGURE 17 Examples of texture synthesis.

of the causal neighborhood system. More examples and discussion can be found in [25].

Acknowledgment

The authors would like Dongyan Wang and Dehong Ma for producing some of the figures.

References

[1] G. Casella and R. L. Berger, *Statistical Inference*, Wadswoth & Brooks, Pacific Grove, California, 1990.

[2] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, 2 Ed., Academic Press, 1982.

[3] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, 1989.

[4] A. Rosenfeld, *Image Modeling*, Academic Press, 1981.

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

[6] H. Derin and P. A. Kelly, "Discrete-index Markov-type random processes," *IEEE Proc.*, 77, 1485–1511, Oct., 1989.

[7] R. Chellappa and A. Jain, eds., *Markov Random Fields —Theory and Applications*, Academic Press, 1993.

[8] G. Winkler, *Image analysis, random fields, and dynamic Monte Carlo methods: a mathematical introduction*, Springer-Verlag, 1995.

[9] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Comm.*, 31, 532–540, 1983.

[10] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. PAMI*, 11, 674–693, Nov., 1989.

[11] K. Chou, A. Willsky, A. Benveniste, "Multiscale recursive estimation, data fusion, and regularization," *IEEE Trans. on Automatic Control* (39) 3, 464–478, 1994.

[12] M. Luettgen, W. Karl, A. Willsky, and R. Tenney, "Multiscale representations of Markov random fields," *IEEE Trans. Signal Processing*, 41, 3377–3396, Dec., 1993.

[13] M. Luettgen, A. Willsky, "Multiscale smoothing error models," *IEEE Trans. on Automatic Control* (40) 1, 1995.

[14] M. Luettgen, A. S. Willsky, "Likelihood calculation for a class of multiscale stochastic models, with application to texture discrimination.," *IEEE Trans. Image Processing* (4) 2, 194–207, 1995.

[15] P. Fieguth, W. Karl, A. Willsky, C. Wunsch, "Multiresolution optimal interpolation and statistical analysis of TOPEX/ POSEIDON satellite altimetry," *IEEE Trans. Geoscience and Remote Sensing* (33) 2, 280–292, 1995.

[16] W. Irving, P. Fieguth, A. Willsky, "An overlapping tree approach to multiscale stochastic modeling and estimation," *IEEE Trans. on Image Processing* (6) 11, 1517–1529, 1997.

[17] C. Bouman and B. Liu, "Multiple resolution segmentation of textured images," *IEEE Trans. Pattern Anal. Machine Intel.*, 13, 99–113, Feb., 1991.

[18] C. Bouman and M. Shapiro, "A multiscale random field model for Bayesian image segmentation," *IEEE Trans. Image Processing*, 3, 162–177, 1994.

[19] F. Heitz, P. Perez, and P. Bouthemy, "Parallel visual motion analysis using multiscale Markov random fields," *Proc. IEEE Workshop on Visual Motion*, Princeton, NJ, Oct., 1991.

[20] K. Popat and R. Picard, "Novel cluster-based probability model for texture synthesis," *SPIE*, 2094, 756–768, 1993.

[21] J. M., Francos, A. Z. Meiri, and B. Porat, "A unified texture model based on a 2-D Wold-like decomposition," *IEEE Trans. Signal Processing*, 2665–2678, Aug., 1993.

[22] S. C. Zhu, Y. Wu, D. Mumford, "Frame: Filters, random fields and maximum entropy: towards a unified theory for texture modeling" *International Journal of Computer Vision*, 27, 1–20, March 1998.

[23] J. De Bonet and P. Viola, "A non-parametric multiscale statistical model for natural images," in *Adv in Neural Infor Processing*, 9, MIT Press, Dec. 1997.

[24] J. Zhang, D. Wang, and Q. Tran, "Wavelet-based stochastic image modeling," *SPIE Imaging 97*, Proc. SPIE 3026, 293–304.

[25] J. Zhang, D. Wang, and Q. Tran, "Wavelet-based multiresolution stochastic models," *IEEE Trans. Image Processing*, 7, 1621–1626, Nov. 1998.

[26] E. P. Simoncelli and J. Portilla, "Texture characterization via joint statistics of wavelet coefficient magnitudes," *Proc. ICIP 98*, Chicago, Illnois, Oct. 1998.

[27] W. Hackbusch, *Multi-grid methods and applications*, Springer-Verlag, 1985.

[28] A. George, *Computer solution of large sparse positive definite systems*, Prentice-Hall, 1981.

[29] P. Roache, *Elliptic marching methods and domain decomposition*, CRC Press, 1995.

[30] H. Rauch, F. Tung, C. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA Journal*, (3) 8, 1965.

[31] D. Chandler, *Introduction to Modern Statistical Mechanics*, Oxford University Press, 1987.

[32] K. Wilson, "Problems in physics with many scales of length," *Scientific American* (241), 158–179, 1979.

[33] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Statist. Soc.*, Series B, 36, 192–226, 1974.

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

[35] F.-C. Jeng and J. W. Woods, "Simulated annealing in compound Gauss-Markov random fields," *IEEE Trans. Infor. Theory*, 36, 94–107, Jan., 1990.

[36] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, "Equations of state calculations by fast computing machines," *J. Chem. Phys.*, 21, 1087–1091, 1953.

[37] J. Zhang, "The mean field theory in EM procedures for Markov random fields," *IEEE Trans. ASSP*, 40, 2570–2583, October, 1992.

[38] G. L. Bilbro and W.E. Snyder, "Applying of mean field annealing to image noise removal," *J. Neural Network Computing*, 5–17, Fall, 1990.

[39] C. Bouman and M. Shapiro, "A multiscale random field model for Bayesian image segmentation," *IEEE Image Processing* **3** (1994), no. 2, 162–177.

[40] Basilis Gidas, "A renormalization group approach to image processing problems," *IEEE Trans. PAMI* **11** (1989), no. 2, 164–180.

[41] Z. Kato, M. Berthod, and J. Zerubia, "A hierarchical Markov random field model . . .," GMIP **58** (1996).

[42] Jan Puzicha and Joachim M. Buhmann, Multiscale annealing for grouping and unsupervised texture segmentation, CVIU **76** (1999), no. 3, 213–230.

[43] R. H. Swendson and J. S. Wang, "Nonuniversal critical dynamics in Monte Carlo simulations," Physical Review Letters **58** (1987), 86–88.

[44] D. M. Titterington, A. F. M. Smith and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*, John Wiley & Sons, New York, 1985.

[45] G. J. McLachlan and D. Peel, *Finite Mixture Models*, New York: Wiley, 2000.

[46] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, 1979.

[47] J. Zhang and D. Ma, "Nonlinear prediction for Gaussian mixture image models," To appear in *IEEE Trans. Image Processing*.

[48] C. W. Therrien, T. F. Quatieri and D. E. Dudgeon, "Statistical model-based algorithms for image analysis," *Proc. IEEE*, 74, April, 1986.

[49] X. Yu, et al., "Automatic target detection and recognition in multiband imagery: a unified ML detecting and estimation approach," *IEEE Trans. Image Processing*, 6, 143–156, Jan. 1997.

[50] D. M. Silva et al., "Optimal detection of small targets in a cluttered background," *Optical Engineering*, 37, No. 1, 83–92, 1998.

[51] P. Brodatz, *Textures; A Photographic Album for Artists and Designers*, Dover Publications, 1966.

[52] B. D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, 1996.

[53] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Soc. Statist., Series B.*, No. 1, 1–38, 1977.

[54] M. A. T. Figueiredo, "On Gaussian radial basis function approximations: interpretation, extensions, and learning strategies," in *Proc. ICPR 2000*, 2, 618–621.

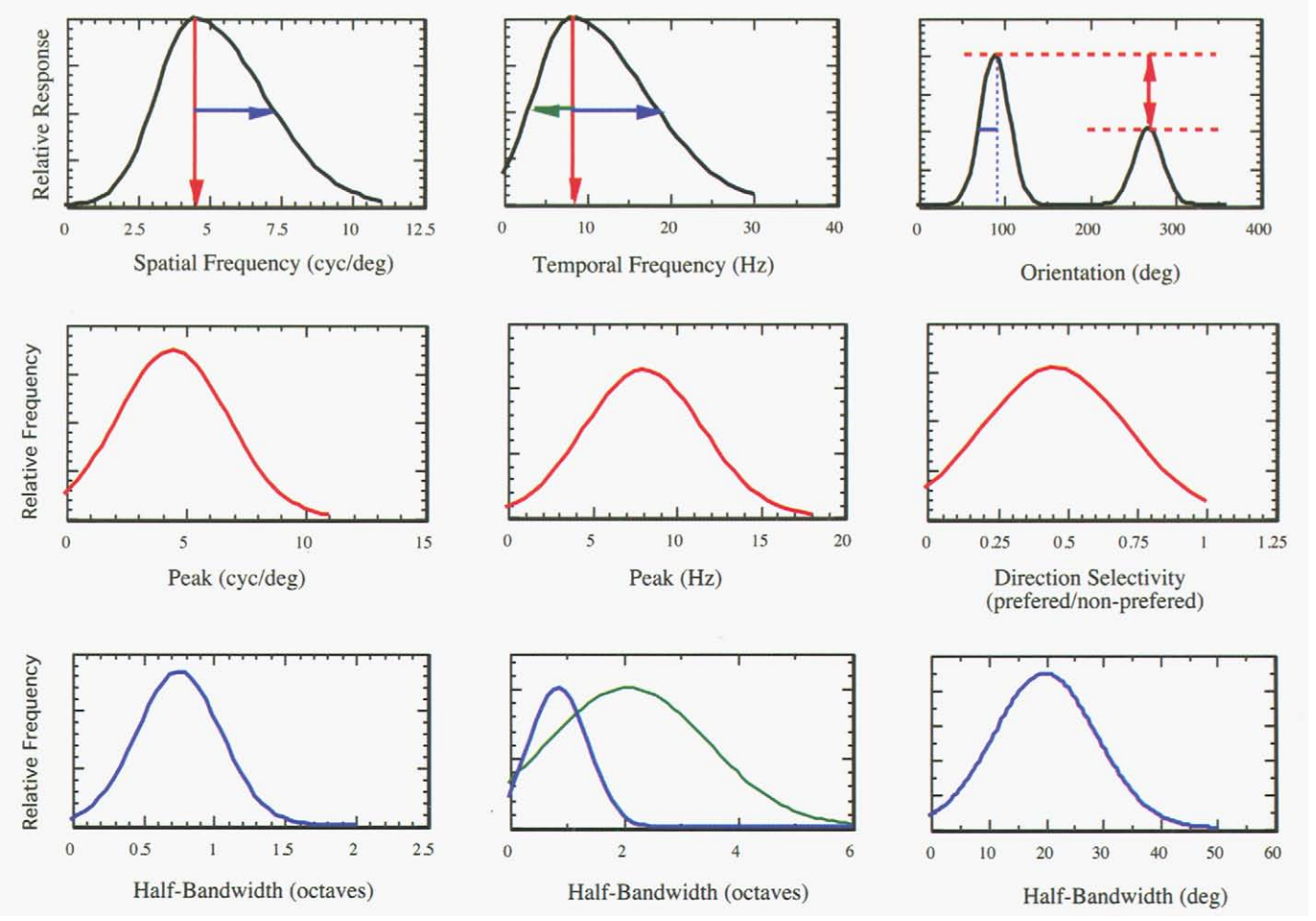

FIGURE 4.1.6 Left column: the upper panel shows a spatial frequency tuning profile typical of cell such as shown in Fig. 5. The middle and lower panels show distribution estimates of the two parameters of peak sensitivity (middle) and half-bandwidth in octaves (lower) for cells in macaque visual cortex. Middle column: same as the left column, but showing the temporal frequency response. As the response is asymmetric in octave bandwidth, the lower figure shows separate distributions for the upper and lower half-bandwidths (blue and green, respectively). Right column: the upper panel shows the response of a typical cortical cell to the orientation of a drifting sinusoidal grating. The estimate of half-bandwidth for macaque cortical cells is shown in the middle panel. The ratio of responses between the optimal direction and its reciprocal is taken as an index of directional selectivity; the estimated distribution of this ratio is plotted in the lower panel (the index cannot exceed unity by definition).

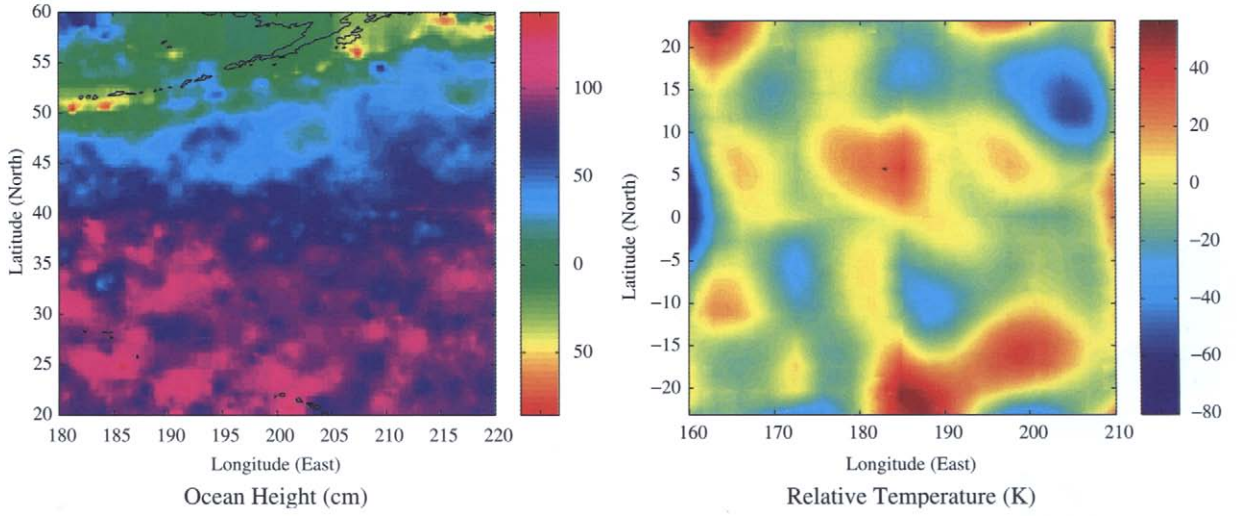

FIGURE 4.3.13 Multiscale estimation of remotely sensed fields. Left: North-Pacific altimetry based on Topex/Poseidon data. Right: Equatorial-Pacific temperature estimates based on in-situ ship data.