Progress in Pattern Recognition 2
L.N. Kanal and A. Rosenfeld (Editors)



# TWO-DIMENSIONAL DISCRETE GAUSSIAN MARKOV RANDOM FIELD MODELS FOR IMAGE PROCESSING

**R. CHELLAPPA**
**Department of EE-Systems and**
**Image Processing Institute**
**University of Southern California**
**Los Angeles, California**

**Abstract**

This paper is concerned with a systematic exposition of the usefulness of two-dimensional (2-D) discrete Gaussian Markov random field (GMRF) models for image processing applications. Specifically, we discuss the following topics: Notion of Markovianity on a plane, statistical inference in GMRF models, and their applications in several image related problems such as image synthesis and compression, image classification and image restoration.

## I. INTRODUCTION

In the analysis and processing of 2-D images one encounters a large amount of data. In a typical image restoration or compression problem, it is not uncommon to work with images defined on grids of dimensions 256 x 256 or 512 x 512. To enable efficient processing of this data, it would be preferable to have an underlying model that explains the dominant statistical characteristics of the given data. Subsequent processing of the images can be efficiently done using the models fitted to the images. Although it is very difficult to identify the underlying physical mechanisms that could have possibly generated the observed data, any analytical expression that explains the nature and extent of dependency of a pixel intensity on intensities of its neighbors can be said to be a model.

A typical image is represented by the gray level variations defined over a rectangular or square lattice. One of the important characteristics of image data is the special nature of the statistical dependence of the gray level at a lattice point on those of its neighbors. The different classes of image models suggested in the literature attempt to characterize this dependence among the neighboring pixels. One way of characterizing the statistical dependence among the neighboring pixels is to represent y(s) as a linear weighted combination of $\{y(s+r), r \varepsilon N\}$ and additive noise where N is known as the neighbor set and does not contain (0,0). Specific restrictions on the members of the neighbor set N yield representations familiar in image processing literature For example, the popular "causal models" are obtained when N is defined as a subset of the set $\{(i,j): i \leq 0, j \leq 0, (i,j) \neq (0,0)\}$. One of the features of using the causal model is that the resulting image processing algorithms are recursive. One can generalize the causal models to obtain the class of "unilateral" models by including more neighbors, but still preserving the recursive structure of the image processing algorithms. The unilateral models result when N is a subset of the non symmetric half plane $S^+$ defined recursively as in [Goodman & Ekstrom, 1980]. Unlike in 1-D discrete stochastic process, where the existence of a preferred direction is inherently assumed, no such preferred ordering is appropriate for a 2-D discrete lattice. Thus, it is possible that an observation y(s) is dependent on neighboring observations in all directions leading to noncausal or bilateral representation. The simplest non-causal model is obtained when y(s) is dependent on its east, west, north and south neighbors. One can consider, more general representation by including the diagonal neighbors and so on.

In this paper we are concerned with a particular class of 2-D noncausal models known as the Gaussian Markov random field (GMRF) models. Let $\{y(s),\ s\in\Omega\}$, $\Omega=\{s=(i,j);\ 0\leq i,\ j\leq M-1\}$ be the observations from an image. It is postulated that this data is generated by an appropriate 2-D GMRF model. The 2-D GMRF models characterize the statistical dependency among pixels by requiring that

$$p(y(s)|\ \text{all } y(r),\ r\neq s) = p(y(s)|\ \text{all } y(s+r),\ r\in N)$$

where N is the appropriate symmetric neighbor set. For instance $N=\{(0,1),(0,-1),(-1,0),(1,0)\}$ corresponds to the simplest GMRF model and by including more neighbors we can construct higher order GMRF models. Since GMRF models are defined only for symmetric neighbor sets, often N is equivalently characterized using an asymmetrical neighbor set $N_s$; i.e., if $r\in N_s$ then $-r\notin N_s$ and $N=(r:r\in N_s)\cup(-r:r\in N_s)$. Since the introduction of GMRF models in the literature [Rosanov, 1967; Woods, 1972] there has been considerable interest in using GMRF models for image restoration, image classification, synthesis and coding.

Prior to any practical application of GMRF models two major problems have to be tackled, viz. the estimation of parameters in GMRF models and the choice of appropriate N for the given image. We discuss several estimation methods for GMRF models. Of particular interest are the statistical properties of these estimates like asymptotic consistency and efficiency. We also discuss decision rules for choosing the appropriate N for the given image. To illustrate the usefulness of GMRF models for image processing applications, algorithms for texture synthesis and coding, texture classification and image restoration are given alongwith pictorial examples.

The organization of the paper is as follows: the representation of GMRF models in 1-D and 2-D is discussed in Section 2. Procedures for the synthesis of an image pattern obeying a known GMRF model are given in Section 3. Methods for estimation of parameters in GMRF models are discussed in Section 4 along with the statistical properties of the estimates. Decision rules for choosing the appropriate structure of the model are given in Section 5. The usefulness of the GMRF models for texture synthesis, texture classification and image restoration is illustrated in Sections 6,7, and 8.

## 2. MODEL REPRESENTATION

We first give a brief review of the relevant theory of 1-D models and subsequently discuss the representation of 2-D models.

### 2.1 One-Dimensional GMRF Models

Definition 2.1: Suppose we have a set of discrete Gaussian observations $\{r(t),\ t=1,2,\ldots,N\}$. Then, $\{r(.)\}$ is said to be a unilateral $m^{th}$ order Markov process in a strict sense [Doob, 1953] if,

$$p(r(t)|\ \text{all } r(j),\ j<t) = p(r(t)\ |\text{all } r(t-j),\ 1\leq j\leq m), \tag{2.1}$$

The sequence $\{r(.)\}$ obeying (2.1) has the representation

$$r(t) = \sum_{j=1}^{m} \theta_j\, r(t-j) + \sqrt{\beta}\ \omega(t),\ t=1,2,\ldots,N, \tag{2.2}$$

with associated initial conditions $\{r(0),\ r(-1),\ldots,r(1-m)\}$. In (2.2), $\{\omega(.)\}$ is an independent and identically distributed (IID) Gaussian noise sequence with mean zero and variance unity.

Definition 2.2. The set $\{r(.)\}$ is said to be a unilateral $m^{th}$ order Markov process in a wide sense [Doob, 1953] if,

$$E(r(t) \mid \text{all } r(j),\ j < t) = E(r(t) \mid \text{all } r(t-j),\ 1 \leq j \leq m)$$

and

$$\text{Var}(r(t) \mid \text{all } r(j),\ j < t) = \text{Var}(r(t) \mid \text{all } r(t-j),\ 1 \leq j \leq m), \tag{2.3}$$

For any arbitrary continuous distribution of $\{\omega(.)\}$, the sequence $\{r(.)\}$ in (2.2) is a wide sense unilateral $m^{th}$ order Markov process. Whether Gaussian or not, strict sense Markovianity implies Markovianity in wide sense, but the converse is true only for Gaussian variables. The notion of unilateral Markov process can be extended to include dependence on neighbors on either side leading to bilateral Markov process.

Definition 2.3: The sequence $\{r(.)\}$ is said to obey a $m^{th}$ order bilateral Markov process in a strict sense if

$$p(r(t) \mid \text{all } r(i),\ i \neq t) = p(r(t) \mid r(t-1),\ldots,\ r(t-m),\ r(t+1),\ldots,r(t+m)) \tag{2.4}$$

The sequence $\{r(.)\}$ obeying (2.4) has the representation [Woods, 1972; Bartlett, 1975]

$$r(t) = \sum_{i=-m}^{m} \theta_i\, r(t-i) + e(t),$$

where

$$\theta_k = \theta_{-k} \text{ and } \{e(\bullet)\}$$

is a correlated Gaussian noise sequence with the conditional structure

$$p(e(s) \mid \text{all } r(t),\ t \neq s) = p(e(s) \mid \text{all } r(s+t),\ -m \leq t \leq m,\ t \neq 0), \tag{2.5}$$

Equation (2.5) implies a weaker condition involving second order properties, as in

$$E(e(s) \mid \text{all } r(t),\ t \neq s) = 0,$$

and

$$\text{Var}(e(s) \mid \text{all } r(t),\ t \neq s) = \nu, \tag{2.6}$$

which in turn implies that the correlated noise sequence $\{e(.)\}$ has the following correlation structure,

$$\begin{aligned} E(e(t)\ e(s)) &= \nu,\ t=s \\ &= -\theta_{t-s}\nu,\ |t-s| \leq m \\ &= 0, \text{ otherwise} \end{aligned} \tag{2.7}$$

Equation (2.7) can be derived as a special case of the derivation for 2-D GMRF models given later. The bilateral Markov Model is driven by a correlated noise sequence. A consequence of (2.5) is

$$E(r(s) \mid \text{all } r(t),\ t \neq s) = \sum_{i=-m}^{m} \theta_i\, r(s-i)$$

A wide sense bilateral Markov model can be defined similar to a wide sense unilateral process. As in the case of unilateral models whether Gaussian or not, strict sense Markovianity implies wide sense Markovianity, but the converse is true only for Gaussian variables.

If $\{r(.)\}$ obeys a unilateral strict sense model of order m then it also obeys a bilateral strict sense model of order m [Woods, 1972]. Further, if 1-D spectral factorization theorem applies, a 1-D bilateral strict sense Markov process (2.5) of order m always possesses an equivalent unilateral wide sense Markov Process (2.3) of order m with identical spectral density function. Thus, under Gaussian assumption, when spectral factorization applies

$$(2.1) \leftrightarrow (2.4) \tag{2.8}$$

For non-Gaussian variables, (2.8) is valid in the wide sense.

Example 2.1 [Bartlett, 1975]: Consider the first-order unilateral Markov model $r(i) = \theta r(i-1) + \sqrt{\beta}\,\omega(i)$ where $\{\omega(.)\}$ is an IID Gaussian sequence. The spectral density function of $\{r(.)\}$ is

$$S_r(\lambda) = \frac{\beta}{(1+\theta^2 - 2\theta\cos\lambda)}$$

The equivalent first-order bilateral model has the representation

$$r(i) = \frac{\theta}{1+\theta^2}\,(r(i-1) + r(i+1)) + e(i)$$

where $\{e(i)\}$ is a correlated noise sequence with the correlation structure

$$E(e)i)\ e(j)) = \frac{\beta}{1+\theta^2},\ i=j$$

$$= \frac{-\beta\theta}{(1+\theta^2)^2},\ |i-j| = 1$$

$$= 0,\ \text{Otherwise}$$

2.2 Unilateral 2-D GMRF Models

Assume that the observations $\{y(s), s\varepsilon\Omega\}$, obey the difference equation;

$$y(s) = \sum_{r\varepsilon N} \theta_r\, y(s+r) + \sqrt{\beta}\,\omega(s),\ \forall\ s, \tag{2.9}$$

The set N associated with (2.9) is a subset of the set $\{(i,j)\colon i\leq 0,\ j\leq 0,\ (i,j)\neq(0,0)\}$ and the noise sequence $\{\omega(s)\}$ is IID. Typically, (2.9) has a set of initial conditions associated with it. For $N=\{(-1,0),\ (0,-1),\ (-1,-1)\}$, (2.9) is characterized by a set of initial conditions made of the observations along the first row and first column. The familiar "causal" models are obtained by restricting N as a subset of quarter plane. The causal models can be generalized to include as many previously scanned points as possible by using the notion of non-symmetric half plane $S^+$ [Goodman & Ekstrom, 1980] recursively defined as

1. $s\varepsilon S^+,\ r\varepsilon S^+ \rightarrow r+s\varepsilon S^+$

2. $s\varepsilon S^+ \rightarrow -s \notin S^+$

3. $(0,0) \notin S^+$

The definition of $S^+$ is not unique. Without loss of generality, one can consider those half planes generated by vectors collinear with cartesian axes. With N being a subset of $S^+$, the corresponding $\{y(s)\}$ is said to be strict sense unilateral Markov with respect to a unilateral neighbor set N if

$$p(y(s) \mid \text{all } y(r),\ r \in \Omega_{s,N}) = p(y(s) \mid \text{all } y(s+r),\ r \in N) \tag{2.10}$$

where $\Omega_{s,N}$ is defined in terms of s and N as follows:

1. $s \notin \Omega_{s,N}$
2. $s+r \in \Omega_{s,N}$ for all $r \in N$
3. $r \in \Omega_{s,N} \rightarrow (r+t) \in \Omega_{s,N}$ for all $t \in N$ provided $r+t \neq s$

Some examples of possible structures of $\Omega_{s,N}$ are shown in Fig. 2.1 [Kashyap, 1981b].

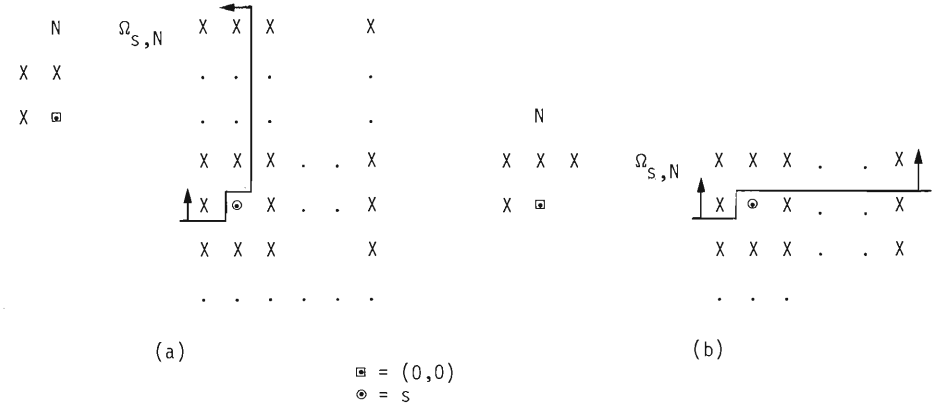


**Fig. 2.1 Examples of N and $\Omega_{s,N}$ for unilateral GMRF Models [Kashyap, 1981b].**

2.3 Non-Causal or Bilateral 2-D GMRF Models

Unlike 1-D discrete time series, where the existence of a preferred direction is inherently assumed, no such preferred ordering of the discrete lattice is appropriate. In other words, the notion of "past" and "future" as understood in unilateral 1-D Markov processes is restrictive in 2-D as it implies a particular ordering of the lattice. It is quite possible that an observation at s, may be dependent on neighboring observations in all directions. The simplest non causal model is obtained when the dependence is on the nearest north, south, east and west neighbors. One can think of more general dependence on nearest diagonal neighbors and neighbors farther. Such a noncausal representation in 2-D is of interest not only due to its generality, but is significant due to lack of spectral factorization in two-dimensions. Thus, if the given observation set $\{y(s)\}$ indeed obeys a non causal model whose spectrum does not factorize, any finite order unilateral representation for this data will only be approximate. In our subsequent discussion, the noncausal or bilateral GMRF models will be simply refered to as GMRF models.

Definition 2.4: The observation set $\{y(s)\}$ is said to be strictly Markov with respect to a symmetric neighbors set N if,

$$p(y(s) \mid \text{all } y(r),\ r \neq s) = p(y(s) \mid \text{all } y(s+r),\ r \in N), \tag{2.11}$$

A hierarchy of GMRF models can be defined as follows: when $N_S = \{(0,1), (1,0)\}$, we obtain a first-order GMRF model and with $N_S = \{(0,1),(1,0),(-1,1),(1,1)\}$, a second-order model and so on. Figure 2.2 illustrates this hierarchy up to seventh-order model. A wide sense GMRF model can be defined as follows:

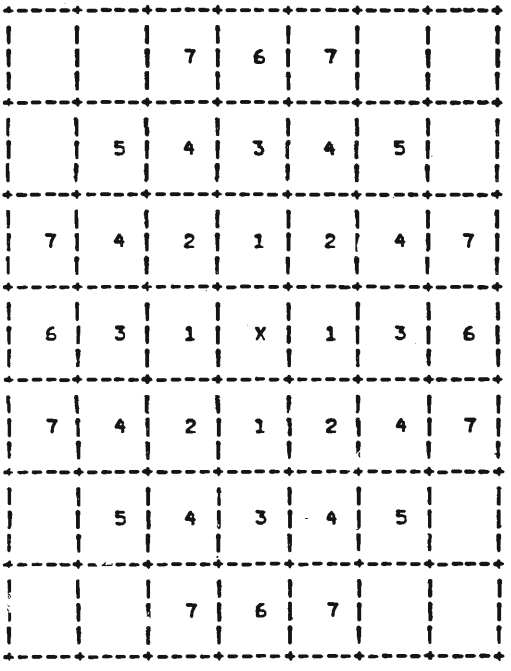


**Fig. 2.2 Hierarchy of GMRF models. The numbers indicate the order of the model relative to x. [G.R. Cross, Ph.D., Thesis, Michigan State University, 1980].**

Definition 2.5: The observation set $\{y(s)\}$ is said to be wide sense Markov with respect to a symmetric neighbor set N if,

$$E(y(s) \mid \text{all } y(r),\ r \neq s) = E(y(s) \mid \text{all } y(s+r),\ r \in N) \tag{2.12}$$

and

$$\text{Var}(y(s) \mid \text{all } y(r),\ r \neq s) = \text{Var}(y(s) \mid \text{all } y(s+r),\ r \in N) > 0 \tag{2.13}$$

In our discussion, the second of the two conditions is restricted to be

$$\text{Var}(y(s) \mid \text{all } y(r),\ r \neq s) = \nu > 0. \tag{2.14}$$

The definition (2.4) of strict Markovianity is often referred to as local Markov property compared to the global Markovian definition used in [Rosanov, 1967; Woods, 1972]. The global definition of Markovianity is as follows [Woods, 1972]. Let a band of minimum width P (a band of minimum width P is a set of contiguous lattice points with a well-defined inside and outside $\Omega^+$ and $\Omega^-$, such that all points in $\Omega^+$ are at least distance P from every point in $\Omega^-$) denoted as $\partial\Omega$ (the "present") separate the discrete lattice $\Omega$ into two regions $\Omega^+$ and $\Omega^-$ (the "future" and "past"). The set $y(s)$, $s \in \Omega^+$ is said to be Markov of order P if,

$$p(y(s),\ s \in \Omega^+ \mid y(s),\ s \in \Omega^-,\ s \in \partial\Omega) = p(y(s),\ s \in \Omega^+ \mid y(s),\ s \in \partial\Omega)$$

This global definition includes the local Markov definition 2.4. A proof is given in [Rosanov, 1967] to show that for GMRF models local Markovianity implies global Markovianity.

Given a finite image we can analyze the image as a finite slice of an underlying infinite lattice image. This approach leads to the class of GMRF models known as infinite lattice GMRF models. In general the infinite lattice models, do not give computationally attractive algorithms for image processing. Another class of models, known as the finite lattice models, obtained by assuming special boundary conditions, yields computationally efficient processing algorithms using fast transforms like discrete Fourier, discrete sine and cosine. We first give the representations of infinite lattice GMRF models, discuss some computational problems associated with this representation and proceed to the representation of finite lattice GMRF models.

2.3.1 Infinite Lattice GMRF Models [Rosanov, 1967; Woods, 1972]

Assume that the observations $\{y(s), s\varepsilon\Omega\}$ have zero mean, Gaussian distribution and obey the following difference equation

$$y(s) = \sum_{r\varepsilon N_s} \theta_r(y(s+r) + y(s-r)) + e(s) \tag{2.15}$$

where the stationary Gaussian noise sequence e(s) has the following properties.

$$\begin{aligned} E(e(s)\, e(r)) &= -\theta_{s-r}\nu,\ (s-r)\varepsilon N \\ &= \nu,\ s=r \\ &= 0,\ \text{otherwise.} \end{aligned} \tag{2.16}$$

The set $N_s$ is the asymmetric neighbor set mentioned in Section 1. Using (2.15) and (2.16), one can prove that

$$\begin{aligned} E(e(s)\, y(r)) &= 0,\ r \neq s \\ &= \nu,\ r = s \end{aligned} \tag{2.17}$$

In view of the Gaussian assumption (2.17) implies

$$E(e(s)\ |\text{all } y(r),\ r \neq s) = 0$$

which in turn implies

$$p(y(s)\ |\text{all } y(r),\ r \neq s) = p(y(s)\ |\text{all } y(s+r),\ r\varepsilon\ N) \tag{2.18}$$

where N is related to $N_s$ as mentioned in Section 1.

Another equivalent way of deriving the representation is to begin with the definition of Markovianity as in (2.11) and show that $\{y(s)\}$ and $\{e(s)\}$, satisfy (2.15) and (2.16) respectively. This is done as follows [Rosanov, 1967; Woods, 1972]: since (2.11) is true,

$$E(y(s)\ |\text{all } y(r),\ r \neq s) = \sum_{r\varepsilon N} \theta_r\, y(s+r) \tag{2.19}$$

Equation (2.19) means that the difference

$$e(s) = y(s) - \sum_{r\varepsilon N} \theta_r\, y(s+r)$$

is orthogonal to all the variables $y(r)$, $r \neq s$. Thus,

$$E(e(r)\, y(s)) = 0, \quad r \neq s$$
$$= \nu \text{ (for some } \nu), \quad r=s \tag{2.20}$$

Since $y(s)$ is Gaussian and zero mean, (2.20) implies

$$E(e(r) \,|y(s),\ r \neq s) = 0 \tag{2.21}$$

Equations (2.19) – (2.21) imply that $y(s)$ should satisfy (2.15). Equation (2.15), together with (2.21), helps to establish (2.16).

Since it is required that

$$E(e(s)e(t)) = -\theta_r \nu, \quad r = s-t$$
$$= E(e(t)e(s)) = \theta_{-r} \nu$$

it follows that GMRF model representation is defined only for symmetric neighbor sets.

A fundamental difference between 1-D and 2-D Markov models is that in 2-D case, there is no simple connection between the noncausal and unilateral GMRF models. This is mainly due to lack of spectral factorization in two-dimensions. Consequently, given that a 2-D sequence $\{y(s)\}$ is strict or wide sense Markov with respect to a neighbor set N, there may not exist any neighbor set N', so that $\{y(s)\}$ is unilateral strict or wide sense Markov with respect to N'. However, any 2-D sequence $\{y(.)\}$ which is unilateral strict Markov with respect to a neighbor set N is also bilateral strict Markov with respect to a symmetric neighbor set N". An example of N and N" is given below and more examples are in [Kashyap, 1981b; Abend, et al., 1965].

Example 2.2 [Jain, 1976]: Consider the 2-D causal Gaussian white noise driven model,

$$y(s) = p\, y(s+(0,-1)) + p\, y(s+(-1,0))$$
$$- p^2 y(s+(-1,-1)) + \omega(s)$$
$$\text{with } E(\omega(s)) = 0 \text{ and } E(\omega^2(s)) = (1-p^2)$$

The equivalent noncausal GMRF model is given by

$$y(s) = \alpha \sum_{r \in N_1} y(s+r) + \alpha^2 \sum_{r \in N_2} y(s+r) + e(s)$$

where

$$N_1 = \{(-1,0),(1,0),(0,-1),(0,1)\},$$
$$N_2 = \{(-1,1),(1,-1),(-1,-1),(1,1)\}$$

with the correlation structure of the Gaussian noise sequence $\{e(s)\}$ being specified as

$$E(e(s)\, e(r)) = \beta^2, \quad s = r$$
$$= -\alpha\beta^2, \quad r-s \in N_1$$
$$= \alpha^2\beta^2, \quad r-s \in N_2$$
$$= 0, \text{ otherwise}$$

where

$$\alpha = \frac{p}{1+p^2} \text{ and } \beta = \frac{1-p^2}{1+p^2}$$

2.3.2. Finite Lattice GMRF Models

Effectively, finite lattice models are obtained by making specific assumptions regarding the neighbors of the boundary pixels. The infinite lattice GMRF models considered in the previous section, although devoid of these assumptions, are difficult to analyze. For example, computation of even some of the basic quantities like covariance matrices is very expensive. The synthesis of images obeying infinite lattice GMRF models can be done only by iterative methods [Cross & Jain, 1983] which are time consuming. On the other hand, the finite lattice GMRF models are easy to analyze and lead to computationally elegant solutions in Wiener filtering, synthesis and compression of images. We discuss a specific finite lattice GMRF model representation that assumes doubly periodic boundary conditions. This model is represented as,

$$y(s) = \sum_{r \in N} \theta_r \, r(s \oplus r) + e(s) \tag{2.22}$$

where $\oplus$ indicates sum modulo M along both the co-ordinate axes and $\{e(s)\}$ is a stationary noise sequence with the correlation structure in (2.16). By defining

$$\underline{y} = \text{Col}\cdot [y(0,0),y(0,1),y(0,M-1),y(1,0)\ldots$$

$$y(1,M-1),\ldots,y(1,M-1),\ldots,y(M-1,M-1)],$$

an $M^2$-vector,

and

$$\underline{e} = \text{Col.}[e(0,0),\ldots,e(0,M-1),\ldots,e(M-1,0),\ldots,e(M-1,M-1)], \text{ an } M^2\text{-vector}$$

Equation (2.22) can now be equivalently written as,

$$B(\underline{\theta})\, \underline{y} = \underline{e}, \tag{2.23}$$

where $B(\underline{\theta})$ is a block-circulant symmetric matrix

$$B(\underline{\theta}) = \begin{bmatrix} B_{0,0} & B_{0,1} \cdots & B_{0,M-1} \\ B_{0,M-1} & B_{0,0} & \ldots B_{0,M-2} \\ B_{0,1} & \cdots & \ldots B_{0,0} \end{bmatrix} \tag{2.24}$$

where each $B_{0,i}$ is M x M matrix and

$$B_{0,i} = B_{0,M-i}$$

A necesssary and sufficient condition to ensure bounded input bounded output stability is

$$\mu_s \triangleq (1 - 2\underline{\theta}^T \underline{\phi}_s) > 0, \; \forall \, s \in \Omega, \tag{2.25}$$

where

$$\underline{\phi}_s = \text{Col.}[\cos \frac{2\pi}{M} s^t r, \; r \in N_s], \tag{2.26}$$

The specific representation given in (2.22) can be arrived at by assuming that the given image is represented on a toroidal lattice. Restrictive as it may appear, the toroidal structure has been widely used in several investigations. It was shown in a pioneering work [Onsager, 1944] that in 2-D, one can solve problems relating to finite lattices analytically using periodic boundary conditions and then obtain the results corresponding to infinite lattices by limiting arguments. In [Moran, 1973; 1973], it was shown how an infinite GMRF model can be constructed from a toroidal lattice GMRF model by letting the lattice size tend to infinity. Toroidal lattice assumptions have been made in [Moran & Besag, 1975] to obtain Gaussian maximum likelihood estimates of parameters of GMRF model. One of the important conclusions of these investigations is that, the final results obtained with and without periodic boundary conditions are not significantly different from each other. Simulation results supporting the above may be found in [Chellappa, 1981].

Several other boundary conditions have been considered in the literature [Kashyap, 1981a; Jain, 1976; 1981] yielding different finite lattice GMRF models. For instance, a useful finite lattice GMRF model is obtained when $\{y(s)\}$ is reflected along the boundaries and N is $\{(-1,0), (1,0),(0,-1),(0,1)\}$. When toroidal representation is used the efficient computational algorithms are obtained for any arbitrary symmetric N; however, the computational advantages are obtained using the boundary conditions mentioned above only for the particular set $N=\{(0,1),(0,-1),(-1,0),(1,0)\}$. Hence, the toroidal representation is more useful.

At this point we would like to point out the differences between the noncausal GMRF models discussed in this section and the class of noncausal models known as spatial autoregressive (SAR) models introduced in a pioneering paper [Whittle, 1954]. The representation of $\{y(s)\}$ obeying a SAR model is

$$y(s) = \sum_{r \in N} \theta_r \, y(s+r) + \sqrt{\beta} \, \omega(s), \tag{2.27}$$

In (2.27), $(\theta_r, r \in N)$ and $\beta$ are unknown parameters, and $\omega(.)$ is an IID noise sequence with zero mean and unit variance. The neighbor set N does not include (0,0) and need not be symmetric. However, if N has symmetric neighbors the corresponding coefficients should be equal; i.e., if $r \in N$ and $-r \in N$, $\theta_r = \theta_{-r}$, otherwise the parameters are not identifiable [Besag, 1974]. If N is symmetric, a necessary and sufficient condition for $\{y(.)\}$ to be stationary is

$$\{1 - \sum_{(i,j) \in N} \theta_{i,j} \, z_1^{\,i} \, z_2^{\,j}\} \neq 0,$$

for all $z_1$ and $z_2$ such that $|z_1| = |z_2| = 1$. Stability conditions for other cases may be found in [Bose, 1982; Jain, 1981].

The main difference between the SAR model and the Gaussian GMRF model is that the $\{y(s)\}$ obeying (2.27) is not Markov with respect to the noncausal neighbor set N, i.e.,

$$p(y(s) \mid \text{all } y(r),\ s \neq r) \neq p(y(s) \mid \text{all } y(s+r),\ r \in N)$$

However, when $\{y(s)\}$ is Gaussian, it is possible to construct a set $N_1$ which is a superset of N such that $y(.)$ obeying a Gaussian SAR model (2.27) is non causal Markov with respect to $N_1$. For instance, if $N = \{(0,1),(0,-1),(-1,0), (1,0)\}$ then $N_1 = [s: s \in N_s \text{ or } -s \in N_s]$ where $N_s = [(1,0), (0,1),(1,1),(-1,1),(0,2),(2,0)]$. The converse is not always true. Given a GMRF model with neighbor set N, there may not exist a finite parameter SAR model with the same second-order properties. A simple example is the GMRF model with $N=\{(0,1),(0,-1),(-1,0),(1,0)\}$. The SAR models are useful for image restoration [Chellappa & Kashyap, 1982a], texture synthesis [Chellappa & Kashyap, 1982b] texture classification [Kashyap et al., 1982] and 2-D spectral estimation [Chellappa & Sharma, 1983]. Discussion of estimation methods and the decision rules for choice of appropriate N may be found in [Kashyap & Chellappa, 1983].

## 3. SYNTHESIS OF IMAGES USING GMRF MODELS

Prior to application of GMRF models for image synthesis and coding, it is useful to see the kind of image patterns that are generated by GMRF models. In this experiment we assume certain model structures and parameters characterizing the models, use a pseudo random number generator to construct {e(s)} and generate {y(s)} obeying (2.15). It should be pointed that since the models are noncausal, in general recursive generation schemes will not be useful; for the case, when the spectral density function of the GMRF model factorizes, recursive generation schemes can be used.

The synthesis procedure begins with the representation

$$B(\underline{\theta})\ \underline{y} = \underline{e} \tag{3.1}$$

where $B(\underline{\theta})$ is a symmetric $M^2 \times M^2$ matrix. Assuming that $\underline{\theta}$ takes values so that $B^{-1}(\underline{\theta})$ exists, the image vector $\underline{y}$ can be written as,

$$\underline{y} = B^{-1}(\underline{\theta})\ \underline{e}, \tag{3.2}$$

Direct implementation of (3.2) is not very attractive due to the computational burden of inverting an $M^2 \times M^2$ symmetric matrix where M could be 32, 64 or 128. However, if we assume that y(s) is represented on a toroidal lattice as in (2.22), then $B^{-1}(\underline{\theta})$ is a block-circulant matrix with eigenvalues $1/\mu_s$ and eigenvectors

$$\underline{f}_s = \text{Col.}[1, \lambda_i, \lambda_i^2\underline{t}_j, \ldots, \lambda_i^{M-1}\underline{t}_j],\ s = (i,j),\ \text{an } M^2\text{-vector}$$

$$\underline{t}_j = \text{Col.}[1, \lambda_j, \lambda_j^2, \ldots, \lambda_j^{M-1}],\ \text{an M-vector} \quad \text{and} \quad \lambda_i = \exp(\sqrt{-1}\ \frac{2\pi i}{M})$$

Thus, $\underline{y}$ can be computed as [Kashyap, 1981a; 1981b]

$$\underline{y} = \frac{1}{M^2} \sum_{s\in\Omega} \frac{\underline{f}_s\, x_s}{\mu_s}$$

where

$$x_s = \underline{f}_s^{*T}\ \underline{e}, \tag{3.3}$$

Since $\underline{e}$ is Gaussian with correlation matrix $B(\underline{\theta})$, one can generate $\underline{e}$ as follows: let $\{\omega(s), s\in\Omega\}$ be an array of zero mean unit variance white Gaussian array. Generate an array of random variables, $\{\varepsilon(s), s\in\Omega\}$ as,

$$\varepsilon(s) = \omega(s)(\mu_s)^{1/2}$$

and let

$$\xi(s) = \frac{1}{M} \sum_{r\in\Omega} \varepsilon(r) \exp[-\sqrt{-1}\ \frac{2\pi}{M}\ (s^T r)]$$

It can be shown [Woods, 1972] that the correlation matrix of $\xi\{(s)\}$ is identical to that of $\{e(s)\}$. For the case of toroidal lattice representation one can save some computations by using the fact that the square root of $B^{-1}(\underline{\theta})$ is also block-circulant with eigenvalues $1/\sqrt{\mu_s}$. Consider the representation [Kashyap, 1981a]

$$\sqrt{B}(\underline{\theta})\ \underline{y} = \sqrt{\nu}\ \underline{n} \tag{3.4}$$

where $\underline{n}$ is $M^2$ vector of zero mean and unit variance Gaussian random numbers. It can be easily checked that (3.4) is an equivalent representation of (2.23). The synthesis of $\underline{y}$ obeying (3.4) can be done as

$$\underline{y} = \frac{1}{M^2} \sum_{s \in \Omega} \frac{\underline{f}_s x_s}{\sqrt{\mu_s}} \tag{3.5}$$

where

$$x_s = \underline{f}_s^{*T} \underline{n} \tag{3.6}$$

We have given in Fig. 3.1, sixteen 64 x 64 image patterns generated using (3.5). The details of the models are in Table 3.1. To make the display of synthetic textures pleasing to the eyes, a constant $\alpha$ = 30.0 (corresponding to the mean of the image) has been added. The gray scale values of the images lie in the range 0-63. It can be seen from an inspection of Fig. 3.1, that the generated patterns are quite varied and some of them look similar to natural textures. Diagonal neighbor sets seem to induce diagonal patterns, as in the windows in positions (1,2),(1,3),(2,2), and (2,3). The role played by adding nearest neighbors can be illustrated using patterns (4,2),(4,3), and (2,1). The window (4,2) corresponds to $N_s$ = {(1,0),(0,1),(2,0),(0,2)} and produces horizontally oriented, macro structured strip patterns. By adding the symmetric neighbor (1,1) a diffused version in window (4,3) is produced. When an extra symmetric neighbor (1,-1) is added we obtain a more micro-structured pattern resembling water. By adding similar neighbors to the model in (4,1), vertically oriented patterns can be produced. To illustrate the role played by the coefficients in generating the patterns, twelve, 64 x 64 patterns corresponding to the GMRF model $N_s$ = {(-1,1),(1,1)}, $\alpha$=30.0, $\nu$ =1.1111, $\theta_{-1,1}$ = .28 and $\theta_{1,1}$ = -.14 were generated. It was found [Chellappa, 1981] that the basic pattern is still retained in all images but the "busyness" of the pattern varied. All the patterns considered, thus far, were generated using the same pseudo random number generator. As illustrated in [Chellappa, 1981], different pseudo random sequences produce very small perturbations in the patterns.

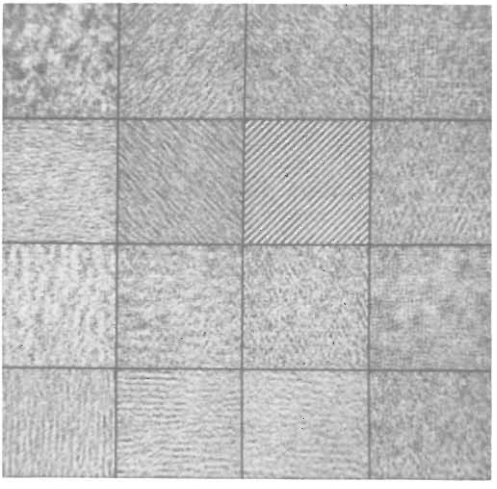

**Fig. 3.1 Synthetic patterns generated by GMRF models in Table 3.1.**

## 4. ESTIMATION IN GMRF MODELS

### 4.1 Introduction

In any practical application of GMRF models, two problems have to be tackled, namely, given the structure of the model, methods for estimating the parameters of the model and

**TABLE 3.1 DETAILS OF GMRF MODELS CORRESPONDING TO SYNTHETIC IMAGES IN FIG. 3.1. FOR ALL OF THEM $\alpha = 30.0$, $\nu = 1.1111$**

| Model Number (row, column) | Neighbor Set $N_S$ and Coefficients |
|---|---|
| (1,1) | (1,0), (0,1)<br>.2794 .1825 |
| (1,2) | (−1,1), (1,1)<br>.28 −.14 |
| (1,3) | (−1,1), (1,1)<br>−.14 .28 |
| (1,4) | (1,0), (1,−1), (0,1), (1,1)<br>.3357 −.25 .3246 −.2126 |
| (2,1) | (1,0), (1,−1), (0,1), (1,1), (2,0), (0,2)<br>−.2061 .0536 −.2061 .0536 −.0123 −.0580 |
| (2,2) | (1,−1), (1,1), (2,0), (0,2), (2,−2), (2,2)<br>−.2341 .4682 .0655 .0655 −.0655 −.0655 |
| (2,3) | (−1,1), (1,1)<br>.28 −.22 |
| (2,4) | (1,0), (0,1), (2,0), (0,2), (3,0),(0,3),(4,0),(0,4)<br>.12 −.10 .08 −.09 −.11 .11 −07 .09 |
| (3,1) | (1,0), (0,1), (3,0), (0,3)<br>.16 .10 .12 −.14 |
| (3,2) | (1,0), (0,1), (3,0), (0,3)<br>.10 .16 −.14 .12 |
| (3,3) | (1,0), (0,1), (1,1), (−1,1), (3,0), (0,3)<br>.12 −.10 .08 −.09 −.11 .11 |
| (3,4) | (0,2), (2,0)<br>.2794 .1825 |
| (4,1) | (1,0), (0,1), (2,0), (0,2)<br>.12 −.24 .16 −.18 |
| (4,2) | (1,0), (0,1), (2,0), (0,2)<br>−.24 .12 −.18 .16 |
| (4,3) | (1,0), (0,1), (1,1), (2,0), (0,2)<br>−.24 .12 .11 −.18 .16 |
| (4,4) | (0,1), (1,0)<br>−.12 .18 |

secondly, estimating the structure of the appropriate model itself. By first assuming that the neighbor set N of the GMRF model is known, we discuss several estimation schemes for GMRF models. Particularly, we stress the asymptotic statistical properties like consistency and efficiency of these estimates. The problem of choosing appropriate GMRF model for the given data is considered in Section 5.

Since in noncausal GMRF models, the expectation of $y(s)$ conditioned on $y(r)$, $r \neq s$ is a function of $y(s+r)$, $r \in N$,

$$p(y(s),\ s \in \Omega) \neq \prod_{s \in \Omega} p(y(s) |\ \text{all } y(s+r),\ r \in N),$$

Hence, it is extremely difficult to write the likelihood function $p(y(s), s \in \Omega)$ except in Gaussian situations. In Gaussian case, one can easily write an expression for log $p(y(s),\ s \in \Omega\ |\underline{\theta}, \nu))$ by evaluating the mean and covariances of $\{y(s)\}$. In general, due to the fact that the Jacobian of the transformation matrix $B(\underline{\theta})$ in (2.24) is not unity, but a complicated function of $\underline{\theta}$ for

GMRF models, log $p(y(s), s\varepsilon\Omega|\underline{\theta},\nu)$ is computationally involved for infinite lattice GMRF models. However, for the special case of finite lattice GMRF models, discussed in Section 2.3.2, one can evaluate easily the Jacobian of the transformation matrix and hence the likelihood function. The resulting likelihood function is a non-quadratic function of the parameters necessitating the use of numerical optimization algorithms.

To avoid the difficulty associated with obtaining the likelihood function in a general case and the associated computations, an ingenious estimation scheme known as the coding method was developed in [Besag, 1972; 1974]. This method is useful for both Gaussian and non-Gaussian models. However, the coding estimate is not as efficient as the ML estimate as it only uses part of the given data. An estimate whose efficiency lies between that of the coding and ML estimates for Gaussian GMRF models was analyzed in [Kashyap & Chellappa, 1983]; this estimate was recommended in an earlier paper [Woods, 1972].

In this section, we discuss several estimation schemes mentioned above and give the results of the LS method applied to synthetic data generated by known GMRF model. A comparison of efficiencies of the coding, LS and ML estimates is given for a simple isotropic GMRF model with $N=\{(0,-1), (0,1), (-1,0), (1,0)\}$.

## 4.2 Coding Method

Assume that the observations $\{y(s)\}$ obey the GMRF model

$$y(s) = \sum_{r\varepsilon N_s} \theta_r[y(s+r) + y(s-r)] + e(s), \forall s,$$

where $\{e(s)\}$ is a zero mean correlated noise sequence with variance $\nu$ and correlation structure given in (2.16). Consider the case of GMRF model with $N_s = \{(0,1), (1,0)\}$, characterized by $\underline{\theta} = \text{Col.}\ [\theta_r, r\varepsilon N_s]$. The conditional distribution $p(y(s)|y(s+r'),$ all $r'\varepsilon N)$ is of the form,

$$p(y(s)\ |y(s+r),\ r\varepsilon N) = \frac{1}{(2\pi\nu)^{1/2}} \exp[-\frac{1}{2\nu}(y(s)- \sum_{r\varepsilon N} \theta_r\ y(s+r)^2]$$

and the coding estimate $\underline{\theta}'_c$ is given by

$$\underline{\theta}'_c = [\sum_{\Omega_0} \underline{q}(s)\underline{q}^T(s)]^{-1}\ (\sum_{\Omega_0} \underline{q}(s)y(s))$$

where

$$\underline{q}(s) = \text{Col.}[y(s+r') + y(s-r'),\ r'\varepsilon N_s]$$

and $\Omega_0$ is a subset of $\Omega$ (consisting of sites marked X in Fig. 4.1). The coding scheme yields another estimate similar to $\underline{\theta}'_c$, say $\underline{\theta}''_c$, with the sum being evaluated over $\Omega - \Omega_0$. One of the main disadvantages of this method is that the estimates thus obtained are not efficient [Moran & Besag, 1975] due to the partial utilization (50%) of the data. Also, for a particular GMRF model, more than one coding scheme can be realized, yielding several estimates likely to be highly dependent. For instance, in the Mercer-Hall wheat data and a GMRF model with $N_s = \{(0,1), (1,0), (1,1), (-1,1)\}$ the estimates of $\theta_{0,1}$ obtained by four possible coding schemes are •043, •085, •243, •236 [Besag, 1974] and a simple averaging of these highly dependent estimates is not satisfactory.

## 4.3 A Consistent LS Estimation Scheme

Consider the estimates

$$\underline{\theta}^* = [\sum_{\Omega_1} \underline{q}(s)\ \underline{q}^T(s)]^{-1}\ (\sum_{\Omega_1} \underline{q}(s)y(s)) \tag{4.1}$$

$$\nu^* = \frac{1}{M^2} \sum_{\Omega_I} (y(s) - \underline{\theta}^{*T} \underline{q}(s))^2, \tag{4.2}$$

where $\Omega_I$ is the region of $\Omega$ containing the interior pixels. The estimate $\underline{\theta}^*$ obtained is an improvement over $\underline{\theta}'_c$ or $\theta''_c$. The statistical properties of this estimate derived in [Kashyap & Chellappa, 1983] are summarized below in a theorem.

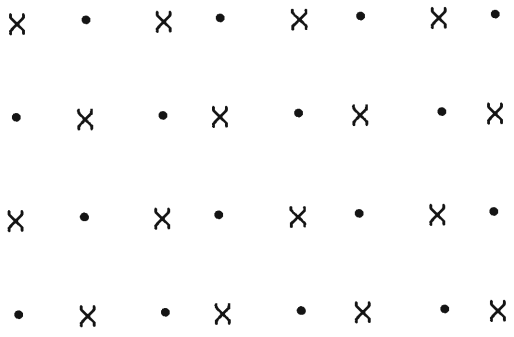

**Fig. 4.1 Coding pattern for a first-order scheme [Besag, 1974].**

Theorem 4.1: Let $y(s)$, $s \in \Omega$ be the set of observations obeying the GMRF model (2.15). Then

1. The estimate $\underline{\theta}^*$ is asymptotically consistent.

2. The asymptotic covariance matrix of $\underline{\theta}^*$, is

$$E(\underline{\theta}-\underline{\theta}^*)(\underline{\theta}-\underline{\theta}^*)^T = \frac{1}{M^2}\left[\nu\, \underline{Q}^{-1} + 2\nu^2(\underline{Q}^T\underline{Q})^{-1} - \frac{\nu}{M^2}\underline{Q}^{-1}\sum_s \sum_{\substack{r \\ (s-r)\in N}} \theta_{(s-r)} \underline{W}_{r,s}(\underline{Q}^T)^{-1}\right]$$

where

$$\underline{Q} = E[\underline{q}(s)\underline{q}^T(s)]$$

and

$$\underline{W}_{r,s} = E[\underline{q}(r)\underline{q}^T(s)]$$

For the isotropic conditional model with $N_s = \{(0,1), (1,0)\}$, the asymptotic expected mean square error is

$$E(\theta-\theta^*)^2 = \frac{2\theta^2(1-4\theta\rho_{1,0})^2}{4M^2\rho^2_{1,0}} \quad \text{where} \quad \rho_{1,0} = \frac{\text{cov}(y(s),y(s+(1,0)))}{\text{cov}(y^2(s))} \tag{4.3}$$

The elements of matrices $\underline{Q}$ and $\underline{W}_{r,s}$ are functions of normalized autocorrelation coefficients $\rho_{k,1}$.

We give the results of applying the LS estimation scheme to some synthetic data generated from a known GMRF model. The synthetic generation scheme in Section 3 was used. The results of the estimation scheme are given below.

Experiment 4.1: (Gaussian Data and a GMRF model with $N_s = \{(-1,1), (1,1)\}$.

The true model is defined as follows: $\nu = 1.1111$, $\theta_{-1,1} = -.14$, and $\theta_{1,1} = .28$. Using this model, synthetic data was generated. For estimation of parameters, consistent estimation scheme in (4.1), (4.2) was used. The actual values of the estimates of the parameters are $\theta^*_{-1,1} = -.1410$, $\theta^*_{1,1} = 0.27875$ and $\nu^* = 1.1033$ which are numerically close to the true parameters. As a consequence, the pattern generated by the GMRF model with estimates of parameters is close to the true pattern.

4.4 ML Estimation of Parameters

The ML estimates are obtained by minimizing the function, [Kunsch, 1981]

$$(-2/M^2)\log p(\underline{y}|\underline{\theta},\nu) = \frac{1}{(2\pi)^2}\int[\log S(\lambda,\underline{\theta})]\, d\lambda + (1/\nu)[C_0 - 2\sum_{r\in N_s}\theta_r C_r] + \log 2\pi \tag{4.4}$$

where

$$S(\lambda,\underline{\theta}) = \frac{\nu}{(1-2\sum_{r\in N_s}\theta_r \cos \lambda^T \cdot r)}, \tag{4.5}$$

and $C_r$ are the sample correlations

$$C_r = \frac{1}{M^2}\sum_{s\in\Omega} y(s)y(s+r)$$

The estimates obtained by minimizing (4.4) should satisfy

$$(1-2\sum_{r\in N_s}\theta_r \cos \lambda^T \cdot r) > 0$$

for the model to be stable. The ML estimate $\hat{\underline{\theta}}$ thus obtained also satisfy

$$\frac{1}{4\pi^2}\int \cos(\lambda^T \cdot r) S(\lambda,\hat{\underline{\theta}})\, d\lambda = C_r, \tag{4.6}$$

The ML estimate are computationally involved due to the evaluation of the integral. Preliminary results indicate [Sharma & Chellappa, 1983] that convergence can be obtained for first- and second-order MRF models in about 30-40 iterations. The IMSL routine ZXMIN was used in this study for parameter estimation.

An usefulness of (4.6) is in developing a GMRF model based approach for 2-D maximum entropy power spectrum (MEPS) estimation. Using the fact that the 2-D MEPS has a structure similar to that of the spectrum (4.5) of GMRF models and (4.6), a model based approach has been developed [Sharma & Chellappa, 1983] recently for 2-D MEPS estimation. The ML estimation scheme may not be practical for image and signal processing applications involving large values of M. By using special boundary conditions, computational load can be reduced as discussed in the next section.

4.5 Estimation of Parameters in Finite Lattice Models

The coding method and LS estimation scheme in Section 4.3 can be directly extended to finite toroidal lattices by summing over $\Omega$. Numerical simulations indicate that the differences in the values of $\underline{\theta}^*$ for infinite and finite toroidal lattice representations are

negligible. The ML estimate for general GMRF models can be obtained by assuming the toroidal lattice representation for {y(s)} and Gaussian structure for {e(s)} and writing down the log-likelihood function. For the toroidal representation corresponding to (2.22) the log-likelihood function log $p(\underline{y}|\underline{\theta},\nu)$ can be written as [Moran & Besag, 1975],

$$\log p(\underline{y}|\underline{\theta},\nu) = \sum_{s\in\Omega} \log (1-2\underline{\theta}^T\underline{\Phi}_s) - (M^2/2) \log 2\pi\nu - \frac{1}{2\nu}\underline{y}^T B(\underline{\theta})\underline{y} \tag{4.7}$$

Note that the contribution of the exponent term of the probability density function is linear in $\underline{\theta}$ unlike the case of SAR models, where this term is quadratic [Whittle, 1954; Kashyap & Chellappa, 1983]. Numerical optimization procedures like Newton-Raphson can be used to obtain the ML estimates. Since the summation term can be computed using 2-D FFT, (4.7) is easier to evaluate compared to (4.5). An asymptotic technique for 2-D MEPS estimation using toroidal representation is in [Chellappa, Hu & Kung, 1983].

Similarly likelihood functions can be written down for GMRF models represented on other types of finite lattices. For instance, [Kashyap, 1981a] suppose we divide the finite lattice $\Omega$ into two mutually exclusive and totally inclusive subsets $\Omega_I$, the interior set and $\Omega_B$, the boundary set.

$\Omega_B = \{s = (i,j);\ s\in\Omega$ and $(s+r) \notin \Omega$ for any member $r\in N\}$

$\Omega_I = \Omega - \Omega_B$

Let the GMRF model be characterized by a neighbor set $N = \{s_1=(0,1), \bar{s}_1=(0,-1), s_2=(1,0), \bar{s}_2=(-1,0)\}$. Consider

$$y(s) = \sum_{i=1}^{2} \theta_i(y(s+s_i) + y(s+\bar{s}_i)) + e(s),\ s\in\Omega_I$$

and

$$y(s) = \sum_{i=1}^{2} \theta_i(y_1(s+s_i) + y_1(s+\bar{s}_i)) + e(s),\ s\in\Omega_B$$

where

$$y_1(s+\bar{s}_i) + y_1(s+s_i) = 2\ y(s+s_i) \text{ if } (s+s_i)\in\Omega,\ (s+\bar{s}_i)\notin\Omega$$

$$= 2\ y(s+\bar{s}_i) \text{ if } (s+s_i)\notin\Omega,\ (s+\bar{s}_i)\in\Omega$$

$$= y(s+s_i) + y(s+\bar{s}_i), \text{ otherwise}$$

The above equations can be written in the form of

$$B(\underline{\theta})\ \underline{y} = \underline{e}$$

where $\underline{y}$ and $\underline{e}$ are $M^2 \times 1$ vectors of arrays {y(s)} and {e(s)}. One can show [Kashyap, 1981a] that the eigenvalues $\mu_{i,j}$, $(i,j)\in\Omega$ of $B(\underline{\theta})$ are

$$\mu_{ij} = 1 + 2\theta_1\phi_1 + 2\theta_2\phi_j$$

where

$$\phi_j = \cos(j\pi/M-1)$$

with the corresponding eigenvector

$$\underline{n}_{ij} = \text{Col.}\ [\eta_0(\phi_j)\ \eta^i,\ \eta_1(\phi_j)\eta^i,\ldots,\eta_{M-1}(\phi_j)\eta^i],\ 0 \le i, j \le M-1$$

$$\eta^j = \text{Col.}\ [\eta_0(\phi_j),\ \eta_1(\phi_j),\ldots,\eta_{M-1}(\phi_j)],\ \text{an M-vector and}$$

$$\eta_i(\phi_j) = \cos(j\ i\ \pi/M-1),\ i,j = 0,1,\ldots,M-1.$$

Hence, the exact likelihood function for this finite model is

$$\log p(\underline{y}|\underline{\theta},\nu) = \frac{1}{2}\sum_{(i,j)\in\Omega}\log \mu_{ij} - (M^2/2)\log 2\pi\nu \quad -\frac{1}{2\nu}\sum_{(i,j)\in\Omega}\|z(\lambda_{ij})\|^2\mu_{ij}, \tag{4.8}$$

where $z(\lambda_{i,j})$ are the 2-D discrete cosine transform of $\{y(s)\}$, defined by the eigenvectors of $B(\underline{\theta})$ given in (2.24). However, it may be noted that the likelihood function in (4.8) is exact only for the particular neighbor set $N = \{(0,1), (0,-1), (-1,0), (1,0)\}$. Similar likelihood function using discrete sine transforms may be obtained for the finite lattice model

$$y(s) = \sum_{i=1}^{2}\theta_i(y(s+s_i) + y(s+\bar{s}_i)) + e(s),\ s\in\Omega_I$$

and

$$y(s) = \sum_{i=1}^{2}\theta_i(y_1(s+s_i) + y_1(s+\bar{s}_i)) + e(s),\ s\in\Omega_B$$

where

$$y_1(s+s_i) + y_1(s+\bar{s}_i) = y(s+s_i) \text{ if } (s+\bar{s}_i) \notin \Omega$$

$$= y(s+\bar{s}_i) \text{ if } (s+s_i) \notin \Omega$$

The families of finite lattice GMRF models having sine or cosine eigen-vectors is not as rich as that of finite lattice GMRF models having discrete Fourier eigenvectors, in that the latter representation is valid for arbitrary symmetric N while the former two are valid only for symmetric N involving nearest neighbors. Since in any given application, models with neighbor sets more general than the nearest neighbors may be required, one may obtain ML estimates for these models using toroidal assumption.

4.6. Comparison of Estimates

We compare the asymptotic variance of the estimate (4.1), with the asymptotic variances of the coding estimate and ML estimate for the isotropic conditional model with $N = \{(0,1), (1,0)\}$. From [Moran & Besag, 1975], the asymptotic variance of the coding estimate is

$$M^2 \text{ Var } (\theta'_c) = \frac{(1-4\rho_{1,0})}{2\rho_{1,0}},$$

Also from [Moran & Besag, 1975], the variance of the ML estimate, $\hat{\theta}_{ML}$ is

$$\text{Var } (\hat{\theta}_{ML}) = \frac{0.5}{M^2(I(\theta)-4V^2_{10}(\theta))},$$

where

$$I(\theta) = \frac{1}{4\pi^2}\int_0^{2\pi}\int \frac{(\cos x+\cos y)^2\, dxdy}{(1-2\theta(\cos x+\cos y))^2}, \tag{4.9}$$

and

$$V_{st}(\theta) = \frac{1}{4\pi^2}\int_0^{2\pi}\int \frac{\cos(sx+ty)\, dxdy}{(1-2\theta(\cos x+\cos y))}, \tag{4.10}$$

Tabulated values of $V_{10}(\theta)$, $\alpha_{1,0}$ and $I(\theta)$ are available in [Moran & Besag, 1975] for different values of $\theta$. Using these values, and (4.3), (4.9), and (4.10), the columns 2-4 of Table 4.1 are

**TABLE 4.1 COMPUTATION OF ASYMPTOTIC VARIANCES AND EFFICIENCIES OF DIFFERENT ESTIMATES IN ISOTROPIC CONDITIONAL MODEL WITH $N_S = \{(0,1),(1,0)\}$**

| $4\theta$ | $M^2 var(\hat{\theta}_{ML})$ | $M^2 var(\theta_C)$ | $M^2 var(\theta^*)$ | $eff(\theta_C)$ | $eff(\theta^*)$ |
|---|---|---|---|---|---|
| .1 | .4928 | .497 | .494 | .991 | .9975 |
| .2 | .472 | .489 | .478 | .965 | .987 |
| .3 | .437 | .474 | .450 | .921 | .971 |
| .4 | .390 | .454 | .412 | .859 | .946 |
| .5 | .333 | .427 | .365 | .779 | .912 |
| .6 | .267 | .393 | .309 | .681 | .864 |
| .7 | .197 | .349 | .244 | .564 | .807 |
| .8 | .1243 | .296 | .1753 | .419 | .709 |
| .9 | .0556 | .224 | .1004 | .248 | .553 |

*Columns 2, 3, and 5 are from [Moran & Besag, 1975].

computed. The asymptotic efficiencies, in columns 5 and 6, are defined as below:

$$eff(\theta'_c) = Var(\hat{\theta}_{ML})/Var(\theta'_c)$$

and

$$eff\ (\theta^*) = Var\ (\hat{\theta}_{ML})/Var(\theta^*)$$

It is evident that the estimate $\theta^*$ computed using (4.1) is more efficient than the coding estimate but is not as good as ML estimate.

## 5. DECISION RULES FOR THE CHOICE OF APPROPRIATE GMRF MODEL

### 5.1 Motivation and Possible Approaches

One of the problems to be tackled in fitting a model to the given image data is the selection of the order of the model to be used. Since the structure of the GMRF model is defined by the underlying neighbor set N of the model, the problem is to decide the appropriate N to be used.

From 1-D time series analysis [Box & Jenkins, 1976], it is known that a model of appropriate order should be fitted to obtain good results in applications like forecasting and control. A similar situation is true in the case of 2-D GMRF models. In fact, the problem becomes more difficult due to the rich variety of model structures. For instance, within the class of GMRF models, different neighbor sets account for different image patterns as shown in Section 3 and the quality of the reconstructed image varies considerably depending on how similar the underlying model is to the true model and hence the use of appropriate neighbor set is important.

The possible approaches are using pairwise hypothesis testing [Anderson, 1962], Akaike's information criterion (AIC), [Akaike, 1974], Bayes approach [Kashyap, 1977; Schwarz, 1978], and Hannan's procedure [Hannan & Quinn, 1979]. The pairwise hypothesis testing method has been used for GMRF models in the modeling of field data [Besag, 1974] and textures [Cross & Jain, 1983]. The main criticisms of this approach are that the resulting decision rules are not always transitive, i.e., if a model $C_1$ is preferred to $C_2$ and $C_2$ is preferred to $C_3$ then it does not follow always that $C_1$ is preferred to $C_3$ [Kashyap, 1977]. Further, the decision rules are not consistent, i.e., the probability of choosing an incorrect model does not go to zero even as the number of observations goes to infinity.

The model selection problem comes under the category of multiple decision problem. A method that is well suited for this problem is to compute a test statistic for different models and choose the one corresponding to the minimum. The AIC criterion and the Bayes method are two such procedures. The AIC test statistics can be computed for GMRF models from the expression of the log-likelihood function given (4.5) or (4.7); the best model is the one which minimizes the AIC statistic. The AIC method, in general gives transitive decision rules but is not consistent even for one-dimensional autoregressive models [Kashyap, 1980].

In the Bayes approach of fitting models to the image, various possible models are postulated as mutually exclusive hypotheses $C_i$, $1 \leq i \leq k$. The hypothesis that maximizes the posterior probability density $P(C_i|y(s),s\epsilon\Omega)$ is chosen as the appropriate model with minimum probability of error. This approach involves obtaining an expression for the likelihood of the observations and integrating it over the parameters using an appropriate prior probability density function. One can also develop a decision rule for assigning $\{y(s)\}$ to one of the hypotheses $C_i$, i=1,...,k so as to minimize a suitable criterion function and also determine the probability of error associated with the decision. The loss function can be chosen to reflect the particular needs of the problem, such as forecasting or estimation of spectral density [Kashyap, 1977].

In this section, we give a decision rule obtained using Bayesian method for finding an appropriate GMRF model. We assume that the given observation set could have possibly been generated by one of k mutually exclusive models or hypotheses denoted by $C_i$, $1 \leq i \leq k$. These hypotheses are GMRF models of different neighbor sets. Under this assumption one can write the expression for joint density of observations, as described in Section 4 and integrate this probability density function over the parameter space using a regular prior probability density function and asymptotic integration [Lindley, 1961]. Using this expression and the prior probabilities $P(C_i)$, i=1,...,k, of the hypotheses, a decision rule for choosing a model with minimum probability of error can be designed. If the prior probability densities are given, one can explicitly compute the posterior densities and take appropriate decisions. For situations where prior densities are not known explicitly, we suggest using that portion of the posterior density function, which does not involve the prior densities. Though this rule does not have the minimum error rate property, it is asymptotically weakly consistent.

5.2 Decision Rules

Currently, known methods for choosing an appropriate GMRF model from a class of such models use pairwise hypothesis testing procedures employing coding estimates. In addition to being non-transitive and inconsistent, this method has several drawbacks due to the usage of coding estimates. Since more than one coding estimate results for a GMRF model, it is quite possible that a GMRF model gets accepted or rejected when different coding estimates are used. Further, when we are choosing between a first-order GMRF model and a second-order GMRF model, to ensure that the likelihoods are comparable, the coding schemes corresponding to the higher order model should be used leading to the use of highly inefficient estimates of the lower order model. Decisions rules that are transitive, asymptotically consistent and which do not possess the above mentioned drawbacks can be derived by using the Bayes procedure. Instead of getting into details, we simply state the problem, give the decision rule and simulation results.

Suppose we have k sets $N_{s1},\ldots N_{sk}$ of neighbors containing $m_1, m_2,\ldots,m_k$ members, respectively. Corresponding to each $N_{si}$, we write the GMRF model as

$$y(s) = \sum_{r\epsilon N_{si}} \theta_{ir} \, (y(s+r) \oplus y(s-r)) + \sqrt{\nu_i}\, e(s), \; s\epsilon\Omega$$

$\oplus$ denotes modulo M along both the co-ordinate axes, $\theta_{ir} \neq 0$, $r\epsilon N_{si}$, $\nu_i > 0$, i=1,...,k and e(s) is Gaussian. Then, the decision rule for the choice of appropriate neighbors is: [Kashyap & Chellappa, 1983] choose the neighbor set $N_{si}^*$ if,

$$i^* = \text{Argument} \min_n \{g_n\}$$

where

$$g_n = -2 \sum_{s \in \Omega} \log(1 - \underline{\theta}_n^{*T} \underline{\phi}_{ns}) + M^2 \log \nu_n^* + m_n \log (M^2), \tag{5.1}$$

and

$$\underline{\theta}_n^{*T} = \text{Col.}\,[\theta_{rn}^*,\ r \in N_{sn}]$$

$$\underline{\phi}_s = \text{Col.}\,[\cos \frac{2\pi}{M} (s^T r),\ r \in N_{sn}]$$

The model selection procedure consists of computing $g_n$ for different models and choosing the one corresponding to the lowest $g_n$. One can write similar test statistics for general GMRF models using (4.4).

To illustrate the usefulness of the decision rule in (5.1), we consider the synthetic data generated by the GMRF model with $N_s = \{(-1,1), (1,1)\}$ considered in Section 4.3. The test statistics $g_n$ in (5.1) were computed for each of the fitted models in Table 5.1. The decision rule correctly picks up the true model. To illustrate the usefulness of the decision rule for choosing GMRF models in image modeling, synthetic patterns corresponding to the models in Table 5.1 were generated in [Chellappa, 1981]. It was found that the patterns corresponding to the inappropriate models 1,2, and 4 are not similar to the original pattern. The patterns

**TABLE 5.1 DETAILS OF GMRF MODELS FITTED TO THE GAUSSIAN DATA GENERATED BY $N_S = \{(-1,1),(1,1)\}$, $\nu = 1.1111$, $\theta_{-1,1} = -.14$, $\theta_{1,1} = .28$**

| Number | Neighbor Set $N_S$ | $\hat{\nu}$ | Estimate of Coefficients | Test Statistics $g_n$ |
|---|---|---|---|---|
| 1 | (1,1) | 1.1638 | $\theta_{1,1} = .3116$ | 1575.3 |
| 2 | (−1,1) | 1.3520 | $\theta_{-1,1} = .2093$ | 1628.9 |
| 3 (True Model) | (−1,1), (1,1) | 1.1033 | $\theta_{-1,1} = -.1410$, $\theta_{1,1} = .27875$ | 1464.30 |
| 4 | (0,1), (1,0) | 1.4934 | $\theta_{0,1} = -.0052$, $\theta_{1,0} = -.0020$ | 1659.7 |
| 5 | (−1,1), (1,1), (0,1) | 1.1033 | $\theta_{-1,1} = -.14101$, $\theta_{1,1} = .27877$, $\theta_{1,0} = -.0051$ | 1472.7 |
| 6 | (−1,1), (1,1), (0,1) | 1.1033 | $\theta_{-1,1} = -.1410$, $\theta_{1,1} = .27873$, $\theta_{0,1} = -.001717$ | 1472.2 |
| 7 | (−1,1), (1,1) (0,1), (1,0) | 1.1033 | $\theta_{-1,1} = -.14101$, $\theta_{1,1} = .27876$, $\theta_{1,0} = .0049$, $\theta_{0,1} = -.0009$ | 1481.0 |

corresponding to models 5,6, and 7 are similar to the original pattern. In general, the patterns corresponding to a specific model A and another model, which includes all the neighbors in model A and some extra neighbors appear very similar, making visusal judgement subjective. However, the quantiative decision rule correctly rejects these overparametrized models.

Another test statistic for the choice of GMRF models can be given by extending Hannan's results [Hannan & Quinn, 1979] for GMRF models. The decision rule is to choose the neighbor set $N_{si}^*$ if,

$$i^* = \text{Argument min } \{g_n\}$$

where

$$g_n = -2 \sum_{s \in \Omega} \log (1 - \underline{\theta}_n^{*T} \underline{\phi}_{n,s}) + M^2 \log \nu_n^* + 2m_n \, c \log \log M^2, \; c > 1 \tag{5.2}$$

Since log log $M^2$ increases with $M^2$slower than log M, the probability of choosing a higher order model is less when (5.2) is used.

## 6. TEXTURE SYNTHESIS AND CODING

In Section 3, it was shown that the class of GMRF models is capable of generating a wide variety of image patterns which exhibit local replication attribute, an essential ingredient of texture. In this section we give experimental results of synthesis of some 128 x 128 textures from Brodatz album. The synthetic textures generated using 12 parameter GMRF models retain most of the characteristics of the original textures. Before giving the experimental results, we justify the use of GMRF models for texture synthesis by showing that the theoretical variograms of many GMRF model possess an oscillatory behavior, a characteristic of variograms of many naturally occurring textures.

### 6.1 Theoretical Variograms of GMRF Models

The second-order properties of y(.) can be conveniently described in terms of the variogram V(r) at displacement r defined as

$$V(r) = E[y(s) - y(s+r)]^2$$

For stationary case,

$$V(r) = 2 R_0 (1 - \rho_r) \tag{6.1}$$

where $\rho_r$ is the normalized autocorrelation function and

$$R_0 = E(y^2(s))$$

Since for the infinite lattice GMRF model in (2.15)

$$\rho_r = \frac{\int_{-0.5}^{0.5} \int \exp(\sqrt{-1}\,(2\pi/M)\,\lambda^T r) S(\lambda) d\lambda}{\int_{-0.5}^{0.5} \int S(\lambda) d\lambda}, \tag{6.2}$$

where

$$S(\lambda) = \frac{\nu}{(1 - 2\underline{\theta}^T \underline{\phi}_\lambda)}$$

and

$$\underline{\phi}_\lambda = \text{Col. } [\cos (2\pi/M)\,\lambda^T r, \; r \in N_s]$$

the variogram at an arbitrary displacement may be obtained using (6.2). Simple expressions (not involving numerical integration) for the variograms may be obtained for GMRF model in (2.22); specifically, the normalized autocorrelation function $\rho_r$ at lag r can be written as

$$\rho_r = \frac{\sum_{s \in \Omega} \exp(\sqrt{-1}\ (2\pi/M)\ s^T r)/\mu_s}{\sum_{s \in \Omega} 1/\mu_s}$$

where $\mu_s$ is in (2.24). The variogram V(r) related to $\rho_r$ as in (6.1) are plotted in Fig. 6.1 for several GMRF models. The details of the models are given [Chellappa, 1981]. For each model we have plotted variograms along the four directions r = (o,i), (i,o), (i,i) and (i,-i) in a discrete lattice. The structure of the possible variograms of 2-D GMRF models is quite varied and many of them possess an oscillatory behavior, a characteristic of the variograms of natural textures [Schacter, et al., 1978].

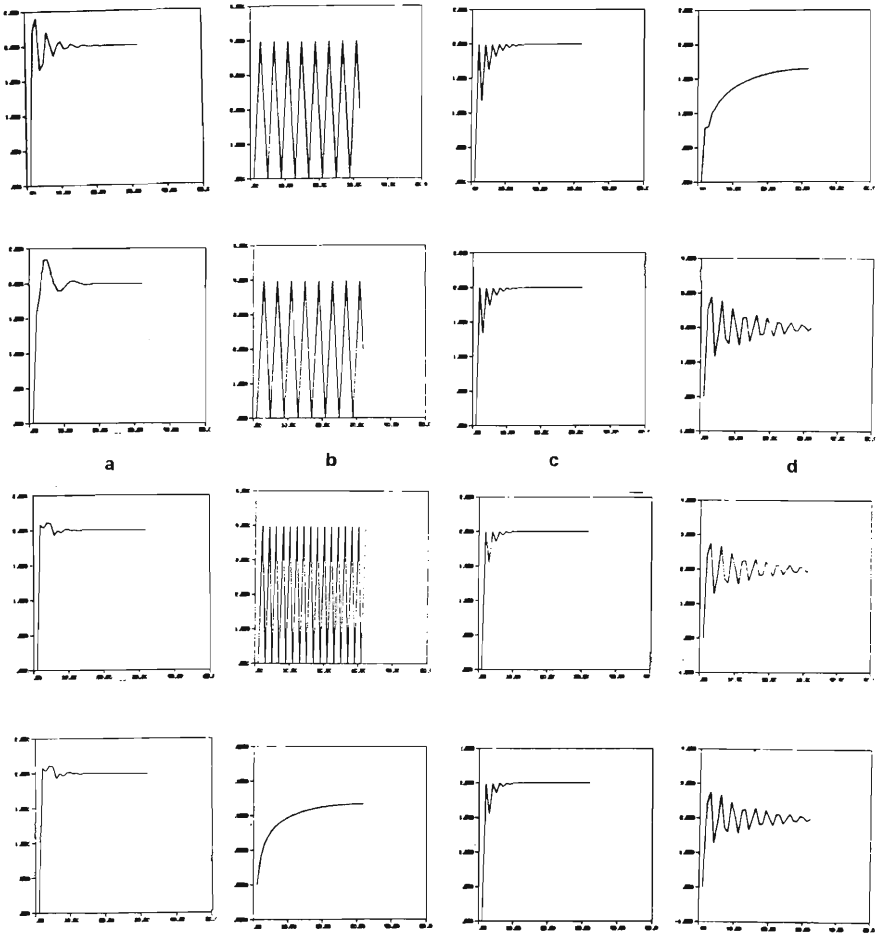


**Fig. 6.1 The theoretical variograms of some GMRF models. V(o,k) versus k is plotted in the first row, V(k,o) versus k in the second row, V(k,k) versus k in the third row, and V(k,-k) versus k in the fourth row.**

6.2 Real Texture Synthesis and Coding

Our discussions in Sections 3 and 6.1 have been concerned with synthetic image patterns and variograms of simulated GMRF models. The next step is to investigate the appropriateness of MRF models for four 128 x 128 natural textures wood, grass, tree bark, and plastic bubble from Brodatz album. The gray levels of the texture are in the range 0–255. We give the results of fitting models for textures as a sequence of experiments. In the first experiment a fourth-order GMRF model with $N_s$ = {(0,1), (1,0), (−1,1), (1,1), (2,0), (0,2), (−2,1), (1,2), (1,−2), (2,1)} was fitted to wood, grass, tree bark, and an eighth-order model was fitted to plastic bubble. To synthesize natural textures pseudo random Gaussian numbers were used instead of $\underline{n}$ in (3.4). The original and synthesized textures are given in the top rows of Figs. 6.2–6.5. Note that we have to store only 12 (10 dimensional vector $\theta^*$, $\alpha^*$ and $v^*$ where $\alpha^*$ is the sample mean) parameters for the fourth-order and 24 parameters for the eighth-order model. The generated wood texture retains the dominant vertical streakline patterns. The synthesized grass and plastic buble are not very close to the original texture.

When a GMRF model is fitted to the given texture, the information contained in the orginal texture is split between the estimates of parameters of the model and the residuals defined as

$$\hat{\underline{n}} = \sqrt{H(\underline{\theta}^*)}\underline{y},$$

In the previous experiment only the parameters of the fitted GMRF model were used to generate the texture. To the extent that the GMRF model or toroidal lattice is appropriate, the components of $\underline{n}$ are uncorrelated and hence can be represented by much fewer number of bits compared to the original texture. In Figs. 6.2–6.5 (bottom left), we have given the synthetic textures generated using the fourth- and eighth-order models and residuals quantized to one bit. For quantization, the Max quantizer [Max, 1960] for Gaussian residuals was used. Since there are 128 x 128 residuals each represented by 1 bit and 12 parameters of the fourth-order model each represented by 32 bits, the compression factor is around 7.816 for wood, grass, and tree bark; for plastic bubble the comparison factor is 7.641 as an eight-order model is required. However, by only retaining half of the total number of quantized residuals (only every other quantized residual is used) and using a pseudo random number to generate the other of the residuals one can almost double the compressions factors. The synthetic textures generated using this procedure are given in the bottom right windows of Figs. 6.2–6.5.

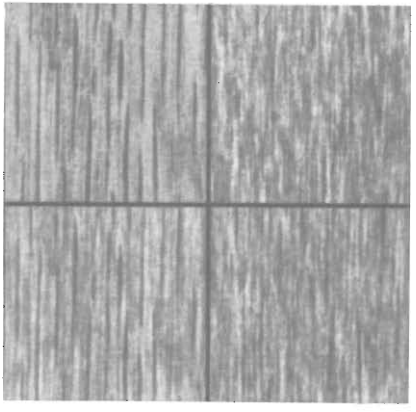

**Fig. 6.2 Synthesis of wood texture using a fourth-order GMRF model (12 parameters). Top left: original, top right: synthesized using the model parameters and a pseudo random number array. Bottom left: synthesized using the model parameters and residuals quantized to 1 bit each. Bottom right: only half the residuals are used.**

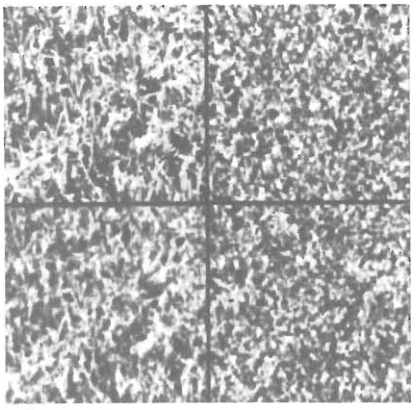

Fig. 6.3 Same as Fig. 6.2 for the grass texture

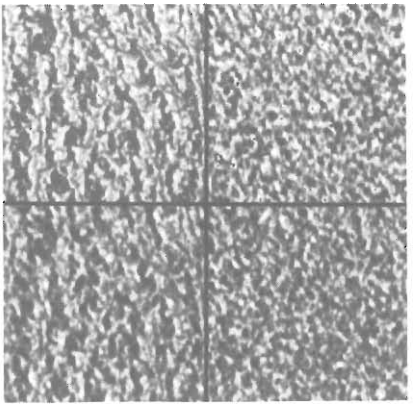

Fig. 6.4 Same as in Fig. 6.2 for the tree bark texture.

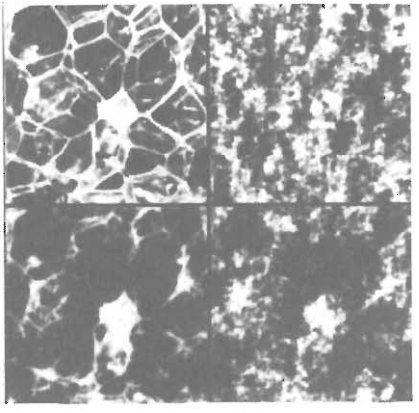

Fig. 6.5 Same as in Fig. 6.2 for the plastic bubble texture with an eighth-order model.

## 7. CLASSIFICATION OF TEXTURES

### 7.1 Preliminaries

Textural and contextual features are two important pattern elements in human interpretation of data. There are many applications like identification of large scale geological information, land use patterns and interpretation of aerial images where texture classification is an important element. A good review of literature on this subject can be found in [Haralick, 1979]. The texture classification problem can be stated as follows: there are a finite number of classes $C_i$, i= 1,2,...,r. A number of training textures belonging to each class are available. Based on the information extracted from these sets a rule is designed which classifies the given test image of unknown class to one of r classes. The key step in a classification problem is the choice of features which reduces the dimension of data to a computationally reasonable amount while preserving much of the information present in the actual data. Currently used methods utilize features from among the categories given below.

1. Features derived form second-order gray level statistics including the gray level co-occurrence matrices [Haralick, et al., 1973]

2. Fourier Power Spectrum [Weszka, et al., 1976]

3. Gray level difference statistics [Weszka, et al., 1976]

4. Gray level run length statistics [Weszka, et al., 1976]

5. Decorrelation methods [Faugeras, et al., 1980]

There are several drawbacks associated with the feature sets mentioned above. The justification of the use of any feature set is its ability to preserve all the relevant information contained in the $M^2$ pixels so that an image close to the original can be generated from the feature set. Such features are called as information preserving features. Since it is impossible to generate an image even close to the original image using any of the above mentioned features, these features are not information preserving. The success of information preserving features has been amply demonstrated in speech recognition using linear predictive coefficients and one can hope similar approaches might be successful in 2-D texture classification.

The decorrelation method feature extraction [Faugeras, et al., 1980] using Laplacian window has the underlying modeling interpretation: suppose the texture is represented by a 2-D separable, causal, Markov model

$$y(s) = \theta_1 y(s+(-1,0)) + \theta_2 y(s+(0,-1)) - \theta_1\theta_2 y(s+(-1,-1)) + \sqrt{\beta}\, w(s)$$

Then as $\theta_1$, $\theta_2 \to 1$ decorrelated residuals generated by the Laplacian operator are similar to those obtained by

$$\hat{\omega}(s) = y(s) - \theta_1^* y(s+(-1,0)) - \theta_2^* y(s+(0,-1)) + \theta_1^*\theta_2^* y(s+(-1,-1))$$

where $\theta_1^*, \theta_2^*$ and $\beta^*$ are LS estimates. Thus, from the point of view of image modeling, in the limit, the decorrelation features are extracted from this moments of the residuals. The loss of information during the process of dimensionality reduction may be significant since residuals themselves contain only partial information in the original data. One should also add $\theta^*_1, \theta^*_2$ as feature components. In the following section we show how GMRF models can be used for feature selection and classification of textures.

7.2.1: Fourier Features:

The Fourier features can be extracted by fitting an GMRF model of appropriate neighbor set to the given texture. The power spectrum given by

$$S(\lambda = (\frac{2\pi i}{M}, \frac{2\pi j}{M})) = \frac{\nu^*}{(1-2\underline{\theta}^{T^*}\underline{\Phi}_{ij}} \qquad (7.1)$$

can be used to extract ring and wedge features as in [Weszka, et al., 1976]. In (7.11), $\nu^*$ and $\underline{\theta}^*$ are the LS estimates (4.1) and (4.2) and $\underline{\Phi}_{ij}$ is

$$\underline{\Phi}_{ij} = \text{Col.}\,[\cos 2\pi/M\,(ik + j1),\,(k,1)\,\varepsilon N_s]$$

Well documented results in 1-D spectral estimation [Childers, 1975] show that the spectral estimates obtained by fitting a model possess better statistical properties compared to the

periodogram estimates. Thus, it may be expected that textural features extracted from S(.) in (7.1) perform better than the periodogram features.

7.2.2: Feature Extraction Using Model Parameters

To extract decorrelation features we simply fit an GMRF model (2.22) to the given texture and obtain the residuals

$$\hat{\underline{n}} = \sqrt{H(\underline{\theta}^*)}\,\underline{y}, \tag{7.2}$$

where $\underline{y}$ denotes the $M^2$ vector constructed from the texture. The decorrelation features can be extracted from the moments of $\hat{\underline{\nu}}$. Since the underlying modeling assumption is more general than the one corresponding to the Laplacian operator, the features derived from $\hat{\underline{n}}$ may perform better. One can also consider an expanded feature set using the estimates $\underline{\theta}^*$ as feature components. Since we can construct textures close to the original using $\underline{\theta}^*$ and $\nu^*$ alone (as discussed in Section 6.2), the set $(\underline{\theta}^*, \nu^*)$ is information preserving and hence should prove to be a good feature vector. A recent work using Bhattacharya distance between two textures obeying Gaussian GMRF models for texture classification is in [Kaneko & Yodogawa, 1982].

The transformation from the given texture to the feature vector is many-to-one and usually involves some loss of information. One can however derive a lossless feature vector for texture classification using GMRF models by using the notion of sufficient statistics [Kashyap, 1978]. Suppose that the vector $\underline{y}$ from the texture is represented by a Gaussian GMRF model (2.22) characterized by $\underline{\theta}$ and $\nu$. Then the probability density of $\underline{y}$ has the form

$$p(\underline{y}|\underline{\theta},\nu) = \frac{(\det\,(B(\underline{\theta})))^{1/2}}{(2\pi\nu)^{M^2/2}} \exp\,\{-(1/2\nu)\,\underline{y}^T B(\underline{\theta})\,\underline{y}\}, \tag{7.3}$$

where $B(\underline{\theta})$ is specified in (2.24). Now define a sample correlation function

$$C_d(r) = M^{-2} \sum_{s\in\Omega} y(s)\,y(s+r).$$

The quadric form $\underline{y}^T B(\underline{\theta})\,\underline{y}$ in (7.3) can be simplified as

$$\underline{y}^T B(\underline{\theta})\,\underline{y} = M^2[C_d(0) - \sum_{r\in N_s} \theta_r C_d(r)] = M^2(C_d(0) - \underline{\theta}^T \underline{C}_d)$$

where

$$\underline{C}_d = \text{Col.}[C_d(r),\ r\in N_s]$$

Thus, (7.3) can be written as

$$p(\underline{y}|\underline{\theta},\ \nu) = \frac{(\det\,(B(\underline{\theta})))^{1/2}}{(2\pi\nu)^{M^2/2}} \exp\,\{\frac{M^2}{2\nu}(C_d(0) - \underline{\theta}^T \underline{C}_d)\} \tag{7.4}$$

Using the factorization theorem [Duda & Hart, 1973],

$$\underline{\xi} = (C_d(0),\ \underline{C}_d)$$

is a sufficient statistic for $(\underline{\theta},\ \nu)$ and hence is a lossless feature vector. One can now design Bayes rule or other standard classification rules like nearest neighbor and so on [Duda & Hart, 1973] using $\underline{\xi}$ as the feature vector. Note that the dimensionality of the feature vector is (m+1) where m is the number of independent $\theta_r$'s. Experiments are currently being done to determine the effectiveness of $\underline{\xi}$ as the feature vector.

## 8. IMAGE RESTORATION

The restoration of degraded images has many fields of application, including space and biomedical imagery. The literature on image restoration is too enormous to be listed in detail and there are many different methodologies like minimum mean-square error (MMSE) restoration, maximum a posteriori probability restoration, and maximum entropy restoration. In this paper, we are concerned with MMSE methods of restoration.

Suppose $\underline{y}$ and $\underline{x}$ represent lexicographic ordered arrays of the original and degraded image, related as in (8.1).

$$\underline{x} = \underline{H}\,\underline{y} + \underline{\eta} \tag{8.1}$$

In (8.1), $\underline{H}$ is the block-circulant matrix corresponding to the blur caused by a nonseparable, space-invariant point-spread function (PSF) and $\underline{\eta}$ is a signal independent additive white noise of variance $\gamma$. Then, it is well known [Andrews & Hunt, 1977] that $\hat{\underline{y}}$, the MMSE estimate of $\underline{y}$, can be written as

$$\hat{\underline{y}} = \underline{Q}_y\,\underline{H}^T(\underline{H}\underline{Q}_y\,\underline{H}^T + \gamma\underline{I})^{-1}\,\underline{x} \tag{8.2}$$

where $\underline{Q}_y$ is the covariance matrix of $\underline{y}$. There are two problems to be considered in the evaluation of $\hat{\underline{y}}$, namely, the determination of the covariance matrix $\underline{Q}_y$ and the inversion of the matrix in (8.1). For an image of size M x M, $\underline{Q}_y$ is of dimension $M^2 \times M^2$, and it is not uncommon to have M = 128 in typical restoration applications. Thus, some assumptions have to be made to reduce the computational load. In [Andrews & Hunt, 1977, ch. 7], the block-Toeplitz covariance matrix $\underline{Q}_y$ is approximated by a block-circulant matrix. Since block-circulant matrices possess an eigenfunction expansion in terms of Fourier vectors, fast Fourier transform (FFT) computations are used for the implementation of (8.2). In the stochastic representation of images by finite difference approximations of partial differential equations [Jain & Jain, 1978], appropriate assumptions are made regarding the model representation for $\underline{y}$ so that $\underline{Q}_y$ has a symmetric, tridiagonal Toeplitz form. Since symmetric tridiagonal matrices are diagonalized by sine transforms, fast algorithms for the implementation of (8.2) have been developed. Another approach [Woods & Radewan, 1977; Woods, 1978; Murphy & Silverman, 1978] is to assume an underlying causal model for the image $\underline{y}$ displayed as a state space model. Then $\hat{\underline{y}}$ could be computed by Kalman filter recursive algorithms.

The other problem in evaluating (8.2) is the determination of $\underline{Q}_y$. Most of the MMSE algorithms assume the availability of a prototype of the original image. Then the covariance matrix $\underline{Q}_y$ is evaluated from the prototype by making appropriate assumptions about the correlation structure of the prototype image. For instance, considerable attention has been paid to the exponential separable autocorrelation function, whose parameters are estimated from the prototype. In [Andrews & Hunt, 1977], $\underline{y}$ of (8.2) is written in terms of spectral density function (SDF) of the prototype image, which is often estimated from the periodogram of the prototype. In [Pratt & Davarian, 1977], the correlation function is assumed to be separable, corresponding to an underlying causal separable model. The method of estimating $\underline{Q}_y$ or equivalently, the parameters of the corresponding models, is arbitrary. We are interested in the use of GMRF models in developing nonrecursive restoration schemes. We represent the given degraded image, $\underline{x}$ in (8.1), by appropriate GMRF models and formulate the MMSE restoration problem. The representation on a toroidal lattice leads to covariance matrices having a block-circulant structure leading to fast implementation of filters using FFT algorithms.

### 8.2. Restoration Using GMRF Model for $\underline{x}$

Assume that the degradation is due to noise only, i.e.,

$$\underline{x} = \underline{y} + \underline{n} \tag{8.3}$$

Assume that $\underline{x}$ obeys GMRF model in

$$x(s) = \sum_{r \in N} \theta'_r \, x(s \oplus r) + e(s) \tag{8.4}$$

where $\{e(s)\}$ is the correlated noise sequence with correlation structure specified in (2.16). The covariance matrix $\underline{Q}_x$ of $\{x(.)\}$ is

$$\underline{Q}_x = \underline{Q}_y + \gamma \underline{T} \tag{8.5}$$

Substitution of $\underline{Q}_y$ form (8.5) into (8.2) with $\underline{H} = \underline{I}$ yields $\underline{y}$, an estimate of the original image

$$\hat{\underline{y}} = (\underline{Q}_y - \gamma \underline{I})^{-1} \, \underline{Q}_x^{-1} \underline{x} \tag{8.6}$$

$\underline{y}$ in (8.6) can be computed using Fourier computations as,

$$\hat{\underline{y}} = \sum_{s \in \Omega} \underline{f}_s \frac{(\nu' - \gamma \mu'_s)}{\nu'} \underline{f}_s^{*T} \underline{x}, \tag{8.7}$$

where

$$\mu'_s = (1 - 2\underline{\theta}'^T \underline{\Phi}_s)$$

are defined in (2.25) and (2.26). The parameters $\underline{\theta}'$, and $\nu'$ can be estimated from the noisy image. The estimates $\underline{\theta}'^*$ and $\nu'^*$ are obtained as,

$$\underline{\theta}'^* = [\sum_{s \in \Omega} \underline{q}(s)\underline{q}^T(s)]^{-1} \, (\sum_{s \in \Omega} \underline{q}(s)x(s))$$

and

$$\nu'^* = (1/M^2) \sum_{s \in \Omega} (x(s) - \underline{\theta}'^{*T} q(s))^2$$

where

$$\underline{q}(s) = \text{Col.}[x(s+r) + x(s-r), \; r \in N_s]$$

One can use the steady-state component of the spectral density of the degraded image as an estimate of $\gamma$ [Andrews & Hunt, 1977]. The experimental results using this scheme are given below.

The original girl's image in Fig. 8.1 is corrupted by additive white noise of SNR = 7dB. GMRF models of different neighbor sets $N_s$ were fitted to the noisy image and (8.8) was used. The noisy image together with the filtered images are given in Fig. 8.2 for SNR = 7 dB. The filtered images of the 7 dB noisy image corresponding to neighbor sets $N_{s3}$ and $N_{s4}$ are good. Table (8.1) gives the details of the GMRF models fitted to the noisy image. Using the model with the neighbor set $N_{s4}$, the best estimate of $\gamma$ for 7 dB noisy image was obtained as 378.74 (true value being 393.01). The MSE corresponding to the best filter in Fig. 8.2 was 91.08 using the neighbor set $N_{s4}$ with the error between the original and noisy being 378.53, roughly a reduction in the MSE by a factor of 4. The numerical values of the MSE between $\underline{y}$ and $\hat{\underline{y}}$ for the models in Table 8.1 as well as the restoration results for 0dB additive noise case may be found in [Chellappa & Kashyap, 1982a].

**Fig. 8.1 Original uncorrupted image.**

**TABLE 8.1 DETAILS OF GMRF MODELS CORRESPONDING TO THE RESTORED IMAGES IN FIG. 8.2**

| Number | Neighbor set $N_S$ |
|---|---|
| $N_{S1}$ | {(0,1)(1,0)} |
| $N_{S2}$ | {(0,1),(1,0),(1,1)} |
| $N_{S3}$ | {(1,0),(1,−1),(0,1),(1,1)} |
| $N_{S4}$ | {(1,0),(0,1),(1,1),(1,−1),(0,2),(2,0)} |

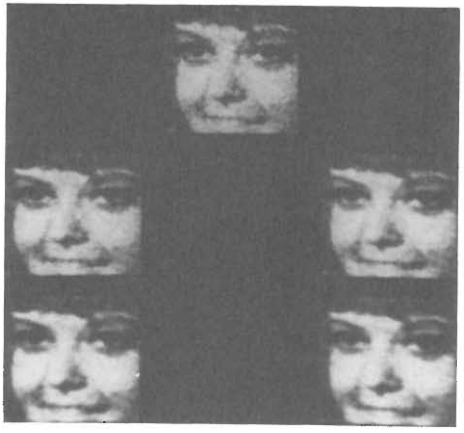

**Fig. 8.2 Filtering of noisy images using GMRF models when the prototype of the original is not available. Top: noisy image of SNR = 5. Second row: filtered images using models $N_{S1}$ and $N_{S2}$. Third row: filtered images using models $N_{S3}$ and $N_{S4}$.**

8.3 Errors in Variable Formulation of the Image Restoration Problem

Consider the image restoration problem

$$\underline{x} = \underline{y} + \underline{\eta}$$

where $\underline{x}$ and $\underline{y}$ are $M^2 \times 1$ vectors corresponding to the noisy and original noise-free images $\underline{\eta}$ is the zero mean additive signal independent white noise vector. Suppose that $\underline{y}$ is of zero mean and obeys an GMRF model of known structure and unknown parameters $\underline{\theta}$ and $\nu$ and that $\underline{\eta}$ is Gaussian of unknown variance $\gamma$. Since the MMSE filter involves, $\underline{\theta}$, $\eta$ and $\gamma$ and in practical applications one does not usually have $\underline{y}$ or a prototype but only $\underline{x}$, it is desirable to develop estimates of $\underline{\theta}$, $\nu$, and $\gamma$ from $\underline{x}$. For the sake of simplicity, let $\gamma = k\nu$, $k>0$. Then $\underline{x}$ has mean $\underline{\nu}$ and covariance matrix $\nu\underline{Q}^{-1}$ where

$$\underline{Q}^{-1} = K\underline{T} + B^{-1}(\underline{\theta})$$

where $B(\underline{\theta})$ is the transformation matrix from $\underline{e}$ to $\underline{y}$ in GMRF model. The log-likelihood function corresponding to a realization $\underline{x}$ is given by [Besag, 1977]

$$\log p(\underline{x}|\underline{Q},\nu) = -(1/2)\ M^2\log(2\pi\nu) + (1/2)\ \log\det\underline{Q} - \frac{1}{2\nu}(\underline{x}^T\underline{Q}\underline{x})$$

The estimate $\nu$ is

$$\hat{\nu} = \frac{1}{M^2}\ \underline{x}^T\underline{Q}\underline{x}$$

The determination of estimates for the remaining parameters consequently reduces to the problem of minimizing

$$-\frac{1}{M^2}\log\det\underline{Q} + \log\{\frac{1}{M^2}\ \underline{x}^T\underline{Q}\underline{x}\}$$

Suppose $\underline{x}$ is represented on a toroidal lattice then

$$\log\det\underline{Q} = -\sum_{s\in\Omega}\log h_s$$

where

$$h_s = k + (1-2\underline{\theta}^T\underline{\phi}_s)^{-1}$$

Also, using the block circulant property of $\underline{Q}$ it can be shown that

$$\underline{x}^T\underline{Q}\underline{x} = \sum_{s\in\Omega}|z_s|^2/h_s,\ \{z_s\} = FFT\{x(s)\}$$

which can be computed once for all. Once $\underline{\theta}$, $\nu$, and $\gamma$ are determined as discussed above, the MMSE filter can be implemented.

## 9. CONCLUSIONS

A systematic exposition of the usefulness of a class of 2-D noncausal models known as GMRF models for developing algorithms for image synthessis, clasification, and restoration has been given. Not covered in our paper are the representation of non Gaussian Markov random field (NGMRF) models, estimation of parameters and usefulness of NGMRF models in image processing. A scholarly paper on the representation and estimation by NGMRF models is that of Besag [Besag, 1974]. The NGMRF models have also been used for texture synthesis [Cross & Jain, 1983; Hassner & Sklansky, 1980].

## 10. ACKNOWLEDGEMENTS

The partial support of National Science Foundation under the grant ECS-82-04181 and USC Faculty Research Innovation Fund is gratefully acknowledged. The author is thankful to Professor R.L. Kashyap for introducing him to the area of stochastic modeling of images. The author is also thankful to Mr. Govind Sharma, Mr. Shankar Chatterjee, Mrs. Jerene Carey and Ms. Hilda Marti for their assistance in preparing this paper.

## REFERENCES

[1] Abend, K., et al., Classification of Binary Random Patterns, IEEE Trans. on Inform. Theory, Vol. IT-11 (October 1965) pp. 538-544.

[2] Akaike, H., A New Look at the Statistical Model Identification, IEEE Trans. on Automat. Contr., Vol. AC-19 (December 1974) pp. 716-722.

[3] Anderson, T.W., Determination of the Order of Dependence in Normally Distributed Time Series, in Proc. of the Symposium on Time Series Analysis, M. Rosenblatt, (ed.) (Wiley, New York, 1962).

[4] Andrews, H.C. and Hunt, B.R., Digital Image Restoration. (Prentice-Hall, New Jersey, 1977).

[5] Bartlett, M.S., The Statistical Analysis of Spatial Pattern, (Chapman and Hall, London, 1975).

[6] Besag, J.E., Nearest Neighbor Systems and the Auto-Logistic Model for Binary Data, J. Royal Stat. Soc. Ser. B, Vol. 13-34 (1972) pp. 75-83.

[7] Besag, J.E., Spatial Interaction and Statistical Analysis of Lattice Systems, J. Royal Stat. Soc. Ser. B, Vol. B-35 (1974) pp. 192-236.

[8] Besag, J.E., Errors-in-Variable Estimation for Gaussian Lattice Scheme, J. Royal. Stat. Soc., Ser.B (1977) pp. 73-78.

[9] Bose, N.K., Applied Multidimensional Systems Theory, (Van Nostrand Reinhold, New York, 1982).

[10] Box, G.E.P. and Jenkins, G.M., Time Series Analysis-Forecasting and Control, (Holden-Day, San Francisco, California, 1976).

[11] Chellappa, R., Stochastic Models for Image Analysis and Processing, Ph.D Thesis, Purdue Unviersity, W. Lafayette, Indiana, (August 1981).

[12] Chellappa, R. and Kashyap, R.L., Digital Image Restoration Using Spatial Interaction Models, IEEE Trans. on Acoustics, Speech and Signal Processing, Vol.ASSP-30 (June 1982a) pp. 461-472.

[13] Chellappa, R. and Kashyap, R.L., Texture Synthesis Using Spatial Interaction Models, Proc. of IEEE Computer Society Conf. on Pattern Recognition and Image Processing, Las Vegas, (June 1982b) pp. 226-230.

[14] Chellappa, R. and Sharma, G., Two-Dimensional Spectral Estimation Using Spatial Autoregressive Models, Proc. Intl. Conf. on Acoustics, Speech and Signal Processing, Boston, (April 1983).

[15] Chellappa, R., Hu, Y.H., and Kung. S.Y., On Two Dimensional Markov Spectrum Estimation, IEEE Trans. on Acoust., Speech and Signal Proc., Vol. ASSP-31 (August 1983).

[16] Childers, D.G., Modern Spectral Estimation, (IEEE Press, New York, 1975).

[17] Cross, G.R. and Jain, A.K., Markov Random Field Texture Models, IEEE Trans. on Patt. Anal. and Mach. Intel., Vol. PAMI-5 (January 1983) pp. 25-40.

[18] Doob, J.L., Stochastic Process, (Wiley, New York, 1953).

[19] Duda, R.H. and Hart, P.E., Pattern Classification and Scene Analysis, (John Wiley, New York, 1973).

[20] Faugeras, O.D., et. al., Decorrelation Methods of Texture Feature Extraction, IEEE Trans. on Patt. Anal. and Mach Intel., Vol. PAMI-2 (July 1980) pp. 323-332.

[21] Goodman, D.M. and Ekstrom, M.P., Multidimensional Spectral Factorization and Univariate AR Models, IEEE Trans. on Automat. Control, Vol. AC-25 (April 1980) pp. 258-262.

[22] Hannan, E.J. and Quinn, B.G., The Determination of the Order of an Autoregression, J. of Royal Statistical Society, Ser. B, Vol. B-41 (1979) pp. 190-195.

[23] Haralick, R.M., Statistical and Structural Approaches to Texture, Proc. IEEE, Vol. 67 (May 1979) pp. 786-804.

[24] Haralick, R.M., Shanmugam, K. and Dinstein, I., Textural Features for Classification, IEEE Trans. on Syst., Man and Cybern., Vol. SMC-3 (November 1973) pp. 610-621.

[25] Hassner, M. and Sklansky, J., Markov Random Fields as Models of Digitized Image Texture, Computer Graph. and Image Proc., Vol. CG1P-12 (April 1980) pp. 376-406.

[26] Jain, A.K., A Fast Karhunen-Loeve Transform for a Class of Random Processes, IEEE Trans. on Commun., Com-24 (September 1976) pp. 1023-1029.

[27] Jain, A.K., A Fast Karhunen-Loeve Transform for Digital Restoration of Images Degraded by White and Colored Noise, IEEE Trans. on Computers, Vol. C-26 (June 1977) pp. 560-571.

[28] Jain, A.K. and Jain, J.R., Partial Difference Equations and Finite Differences in Image Processing, Part II: Image Restoration, IEEE Trans. on Automat. Control, Vol. AC-23 (October 1978) pp. 817-833.

[29] Jain, A.K., Advances in Mathematical Models for Image Processing, Proc. of IEEE, Vol. 69 (May, 1981) pp. 512-528.

[30] Kashyap, R.L., A Bayesian Comparison of Different Classes of Dynamic Models Using Empirical Data, IEEE Trans. on Automat. Contr., Vol. AC-22 (October 1977) pp. 715-727.

[31] Kashyap, R.L., Optimal Feature Selection and Decision Rules in Classification Problems with Time Series, IEEE Trans. on Inform. Theory, Vol. 17 (May 1978) pp. 281-288.

[32] Kashyap, R.L., Inconsistency of the AIC Rule for Estimating the Order of Autoregressive Models, IEEE Trans. on Automat. Contr., Vol. AC-25 (October 1980) pp. 996-998.

[33] Kashyap, R.L., Random Field Models on Finite Lattices for Finite Images, Proc. of the Conf. on Information Sciences and Systems, Johns Hopkins University, (March 1981a) pp. 215-220.

[34] Kashyap, R.L., Analysis and Synthesis of Image Patterns by Spatial Interaction Models, in Progress in Pattern Recognition, Vol. 1, L.N. Kanal and A. Rosenfeld, (eds.), (North-Holland Publishing Company, 1981b).

[35] Kashyap, R.L., Chellappa, R. and Khotanzad, A., Texture Classification, Using Features Derived from Random Field Models, Patt. Recn. Letters, Vol. 1 (October 1982) pp. 43-50.

[36] Kashyap, R.L. and Chellappa, R., Estimation and Choice of Neighbors in Spatial Interaction Models of Images, IEEE Trans. on Inform. Theory, Vol. IT-29 (January 1983) pp. 60-72.

[37] Kunsch, H., Thermodynamics and Statistical Analysis of Gaussian Random Fields, Z.W. Ver Gebiete (ed.), Vol. 58 (November 1981) pp 407-421.

[38] Lindley, D.V., The Use of Prior Probability Distributions in Statistical Inference and Decisions, in Proc. 4th Berkeley Symposium on Math. Statistics and Probability, Vol. 1 (1961) pp. 453-468.

[39] Max, J., Quantizing for Minimum Distortion, IEEE Trans. on Information Theory, Vol. IT-6 (March 1960) pp. 7-12.

[40] Moran, P.A.P., A Gaussian Markovian Process on a Square Lattice, J. Appl. Prob., Vol. 10 (1973) pp. 54-62.

[41] Moran, P.A.P., Necessary Conditions for Markovian Processes on a Lattice, J. Appl. Prob., Vol. 10 (1973) pp. 605-612.

[42] Moran, P.A.P. and Besag, J.E., On the Estimation and Testing of Spatial Interaction in Gaussian Lattice, Biometrika, Vol. 62 (1975) pp. 555-562.

[43] Murphy, M.S. and Silverman, L.M., Image Model Representation and Line by Line Recursive Restoration, IEEE Trans. on Automat. Control, Vol. AC-23 (1978) pp. 808-816.

[44] Onsager, L., Crystal Statistics, I: A Two-Dimensional Model with an Order-Disorder Transition, Phys. Ref., Vol. 65 (1944) pp. 117-149.

[45] Pratt. W.K. and Davarian, F., Fast Computational Techniques for Pseudo Inverse and Wiener Image Restoration, IEEE Trans. on Computers, Vol. C-26 (June 1977) pp. 571-580.

[46] Rosanov, Y.A., On Gaussian Fields with Given Conditional Distributions, Theory of Prob. and its Applications, Vol. II (1967) pp. 381-391.

[47] Schacter, B., Rosenfeld, A. and Davis, L.S., Random Mosaic Models for Textures, IEEE Trans. on Syst., Man and Cybern., Vol. SMC-8 (September 1978) pp. 694-702.

[48] Schwarz, G., Estimating the Dimension of a Model, Ann. Statist., Vol. 6 (March 1978) pp. 461-464.

[49] Sharma, G. and Chellappa, R., A Model Based Approach for Two-Dimensional Maximum Entropy Power Spectrum Estimation, Submitted for publication.

[50] Weszka, J., Dyer, C.R. and Rosenfeld, A., A Comparative Study of Texture Measures for Terrain Classification, IEEE Trans. on Syst., Man., and Cybern., Vol. SMC-6 (April 1976) pp. 269-285.

[51] Whittle, R., On Stationary Processes in the Plane, Biometrika, Vol. 41 (1954) pp. 434-449.

[52] Woods, J.W., Two-Dimensional Discrete Markovian Random Fields, IEEE Trans. On Information Theory, Vol. 18 (March 1972) pp. 232-240.

[53] Woods, J.W., Markov Image Modeling, IEEE Trans. on Automat. Contr. Vol. AC-23 (October 1978) pp. 846-850.

[54] Woods, J.W. and Radewan, C.H., Kalman Filtering in Two-Dimensions, IEEE Trans. on Information Theory, Vol. IT-23 (July 1977) pp. 473-482.