COMPUTER GRAPHICS AND IMAGE PROCESSING 15, 301–318 (1981)

Decision Rules for Choice of Neighbors in Random Field Models of Images

R. L. KASHYAP*

School of Electrical Engineering, Purdue University, West Lafayette, Indiana 47907

R. CHELLAPPA*, †

School of Electrical Engineering, Purdue University, West Lafayette, Indiana 47907; and Computer Science Center, University of Maryland, College Park, Maryland 20742

AND

N. AHUJA

Coordinated Science Laboratory, University of Illinois, Urbana, Illinois 61801

Received October 23, 1979

Random field models have many applications in image processing and analysis. The main concern of this paper is to design a decision rule for fitting an appropriate random field model to a given image. We assume that the given image is a particular realization of a homogenous Gaussian discrete random field. We represent the underlying random field by a set of parametric models representing the spatial dependence. Using spectral representations of the random field and standard Bayesian methods, we develop a decision rule for choosing an appropriate model from a class of such models. We discuss the relevance of the theory developed in this paper for applications in image modeling and texture characterization.

1. INTRODUCTION

Random field models have many applications in image processing and analysis; for example, they can be used for design of image enchancement or restoration algorithms [1–3], for image coding [4–6], segmentation of images [7–8], and characterization of textures [9–10]. Typically, an image is represented by a two-dimensional scalar array, the gray level variations defined over a square grid. One of the important characteristics of this data is the statistical dependence of the gray levels within a neighborhood. For example, $y(s_1, s_2)$, the scalar gray level at position (s_1, s_2), might be statistically dependent on the values of gray levels over a neighborhood that includes $\{(s_1 - 1, s_2), (s_1 + 1, s_2), (s_1, s_2 - 1), (s_1, s_2 + 1)\}$. This is in contrast to the familiar time series models where the dependence is strictly on the past observations. Therefore any image model should adequately reflect the statistical dependence of a pixel on its neighbors in all directions.

The statistical dependence of gray levels in a neighborhood can be represented in several ways. Suppose we concatenate the observations from successive rows and obtain a one-dimensional series of real numbers. This series can be treated as a realization of a discrete one-dimensional stochastic process and can be represented by one-dimensional time series models. Seasonal autoregressive models and their

*Supported in part by the National Science Foundation under Grant ECS 8009041.
† Supported in part by the Air Force Office of Scientific Research under Grant AFOSR-77-3271.

0146-664X/81/040301-18$02.00/0

variants have been used to represent such time series data [4, 9]. Another method would be to retain the two-dimensional character of the data and represent the statistical dependence of the pixel $y(s_1, s_2)$ on the neighbors on the left and top of (s_1, s_2) as shown in Fig. 1. A special case of this representation with $a = b = 1$ has been used in the image processing literature quite extensively [1, 3, 10]. The above-mentioned approaches are inadequate for an image since they do not include the neighborhood dependence in all directions. As mentioned earlier, an image represents a statistical phenomenon on a plane and hence the notion of past and future as understood in classical time series analysis is not relevant. Hence any appropriate image model should include dependence in all directions.

Image models which include dependence in all directions (referred to as neighborhood models in the sequel) have been considered recently [3, 7, 8, 11]. The neighborhood dependence might include the four nearest pixels (east, west, north, and south pixels), eight nearest pixels [3, 11], or all the pixels inside a square window surrounding the pixel at (s_1, s_2) [7, 8] as shown in Fig. 2. In these models, the observation $y(s_1, s_2)$ is written as a linear weighted sum of observations over the corresponding neighborhood and an independent noise sequence. These models are characterized by a set of unknown weights or coefficients and the variance of the noise driving the model.

Prior to the use of these models two problems have to be tackled, namely, the estimation of the unknown parameters and the choice of an appropriate neighborhood of pixel dependency for the given image. The parameter estimation is usually handled by the *m*aximum *l*ikelihood (m.l.) method. The problem considered here is the choice of appropriate neighborhood. There is no need for any special restrictions on the image such as isotropy. Neighborhoods having arbitrary number of members can be considered. It is well known that even in the case of one-dimensional time series models, a model of appropriate order should be fitted before it can be used for applications. Hence the importance of the choice of an appropriate neighborhood for an image, before using it for further applications, need not be overemphasized.

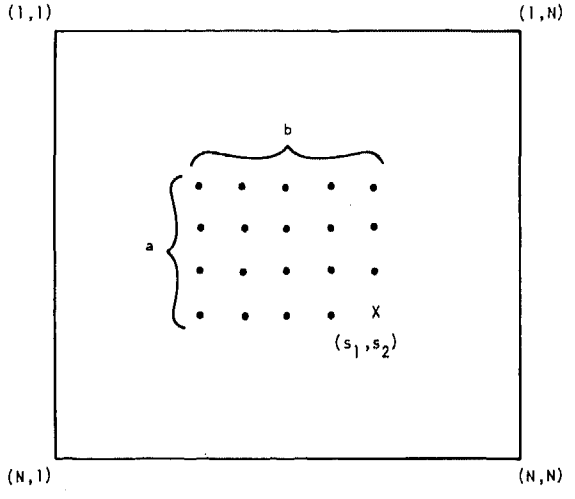

FIG. 1. Two-dimensional unilateral dependence.

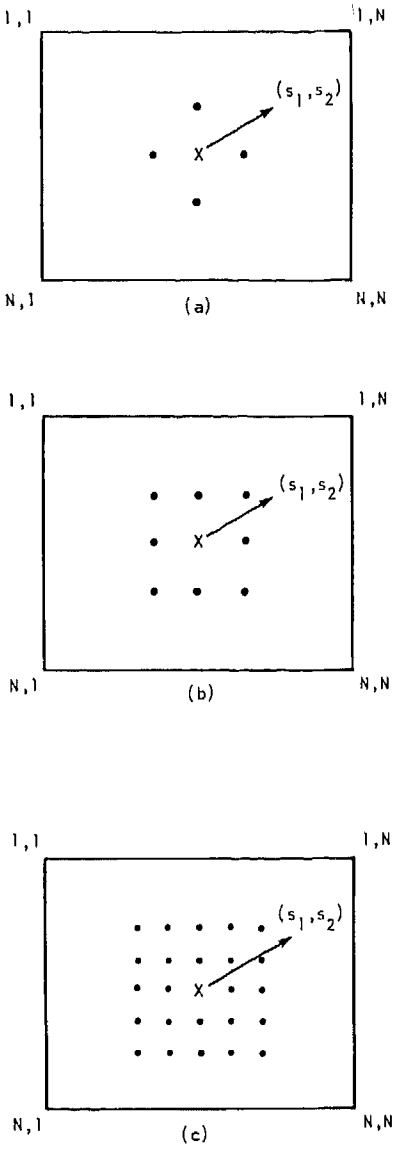

FIG. 2. Neighborhood dependence. (a) Four-neighborhood [3], (b) eight-neighborhood [3], and (c) the pixels within a surrounding window [7–8].

The main approaches to choosing neighbors in model building are the following:

(1) Pairwise hypothesis testing,

(2) *A*kaike's *i*nformation *c*riterion (AIC),

(3) The Bayesian approach.

In the pairwise hypothesis testing approach [12], a so-called null hypothesis is set up, (say) that the given two-dimensional data could be represented by an isotropic four-neighborhood model (e.g., one that includes east, west, north, and south

neighbors), and a decision rule is designed so that the probability of rejecting the null hypothesis when it is indeed true has a prespecified upper bound. The main criticism of this approach is that the resulting 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. Also, the decision rules may not always be transitive, i.e., given three hypotheses C_1, C_2, and C_3, if C_1 is preferred to C_2 and C_2 is preferred to C_3, then it does not follow that C_1 is preferred to C_3 [13].

The AIC method [14] computes the so-called AIC statistic of the given observations for each model. The best model is the one which minimizes the AIC statistic. This method gives a transitive but not consistent decision rule [15].

In the Bayesian approach [13] of fitting models to data, various possible models are postulated as mutually exclusive hypotheses $C_i, 1 \leq i \leq r$. The hypothesis that maximizes the posterior probability density $P(C_i | y(\mathbf{s}), \mathbf{s} \in \Omega_s)$ is chosen as the correct 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.

In this paper we propose a Bayesian method for finding a neighborhood model for a random field, using the spectral representation for the random field. Specifically, using the asymptotic Gaussian properties of the finite Fourier transforms $\{z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\}$, where Ω_λ is a set of discrete frequencies $\lambda_{ij} = (\lambda_0 i, \lambda_0 j)$, $\lambda_0 = 2\pi/N$, an explicit expression is given for $p(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda | \boldsymbol{\theta}_k)$, $\boldsymbol{\theta}_k$ being the vector of parameters in the model.

We integrate this probability density function over the variable $\boldsymbol{\theta}_k$ using the corresponding regular prior probability density function of $\boldsymbol{\theta}_k$, viz., $p(\boldsymbol{\theta}_k | C_k)$ to obtain $p(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda | C_k)$. Using this expression and the prior probabilities $P(C_k)$, $k = 1, 2, \ldots, r$, of the hypotheses, a decision rule for choosing a model with minimum probability of error is designed. This decision rule involves arbitrary quantities like prior densities of the parameters. A decision rule suppressing the terms involving the prior densities is also given. Though this rule does not have the minimum error rate property, it is asymptotically weakly consistent.

We drop the boldface notation in s and λ_{ij} throughout and use exp and e interchangeably to indicate the exponential function.

The transform domain approach used in this paper can be easily extended to include moving average and autoregressive and moving average models.

The organization of the paper is as follows: In Section 2, an explicit expression is derived for the probability density of the transforms of the observations given the neighborhood model obeyed by the observations. In Section 3, the problem of finding the appropriate neighborhood is posed as a class selection problem and decision rules are designed for choosing this neighborhood. Section 4 discusses the properties of the decision rule. The decision rule for the class of unilateral models is derived in Section 5. A brief discussion of the relevance of the theory developed here to image processing is given in Section 6.

2. EXPRESSION FOR THE PROBABILITY DENSITY $(p(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda | \phi, \rho)$ OF TRANSFORMS OF OBSERVATIONS OBEYING A RANDOM FIELD MODEL

In this section we derive an explicit expression for the probability density of the finite Fourier transforms of the observations $\{y(s), s \in \Omega_s\}$, where Ω_s is a square grid of side N, $s = (s_1, s_2)$ is a specific location on the grid and $1 \leq s_i \leq N$, $i = 1, 2$.

The given observations $y(s)$ are assumed to obey the random field model

$$E(\phi,\rho):\quad y(s)+\sum_{\mathbf{q}_k\in Q}\phi_k y(s+\mathbf{q}_k)=\sqrt{\rho}\,u(s), \tag{2.1}$$

$$\mathbf{s}\in\Omega_s,\qquad Q=\{\mathbf{q}_k=(q_{k,1},q_{k,2}),\quad k=1,2,\ldots m,$$

$$\mathbf{q}_k\neq(0,0),\quad q_{k,i}\text{ are integers}\}$$

Here $\{u(s), s\in\Omega_S\}$ is Gaussian i.i.d. noise with zero mean and unit variance. For simplicity, in what follows, we denote $\mathbf{q}_k$ by q_k.

The random field is homogeneous and Gaussian but not necessarily isotropic. A random field is said to be homogeneous if the following condition is satisfied:

$$R(s,t)=E\{(y(s)-E(y(s)))(y(t)-E(y(t)))\}$$

$$=R(s-t)$$

—i.e., the covariance function is translation invariant. In addition, if the covariance function $R(s,t)$ is also invariant to rotation, the random field is called isotropic. In general, images are *not* isotropic and hence the random field models of interest to us are only homogeneous and not necessarily isotropic.

Equation (2.1) is characterized by an unknown $(m+1)$-dimensional vector $\boldsymbol{\theta}^T=(\boldsymbol{\phi}^T,\rho)$ such that $\phi_j\neq 0$, $j=1,2,\ldots,m$ and $\rho>0$. Equation (2.1) represents the dependence of an observation at location (s_1,s_2) on its neighbors in the direction specified by Q. When $q_{k,1}$ and $q_{k,2}$ take only nonpositive values we obtain models where the observation at location (s_1,s_2) is a linear combination of the observations in a one-sided neighborhood. We assume that the coefficients in (2.1) satisfy the following condition to ensure homogenity [16]:

$$\Big|\sum_{q_k\in Q}\phi_k t_1^{q_{k,1}}t_2^{q_{k,2}}\Big|<1\qquad\text{whenever } |t_1|=|t_2|=1.$$

The two-dimensional transfer function of the system in (2.1) is given by

$$H\Big[\exp\big(\sqrt{-1}\,\lambda_{ij}\big),\phi\Big]=\Big[1+\sum_{q_k\in Q}\phi_k\exp\big(\sqrt{-1}\,\lambda_0(iq_{k,1}+jq_{k,2})\big)\Big]^{-1}. \tag{2.2}$$

In (2.2), $\exp(\sqrt{-1}\,\lambda_{ij})$ stands for the vector $[\exp(\sqrt{-1}\,\lambda_0 i),\exp(\sqrt{-1}\,\lambda_0 j)]$.The corresponding spectral density function evaluated at frequency $\{\lambda_{ij}=(\lambda_0 i,\lambda_0 j),(i,j)\in\Omega_s\}$ is given by

$$S_y(\lambda_{ij};\phi,\rho)=\rho\Big\|H\big(\exp\big(\sqrt{-1}\,\lambda_{ij}\big),\phi\big)\Big\|^2, \tag{2.3}$$

where $\|\alpha\|$ denotes the modulus of the complex variable α. To make our notation clear, we consider a few examples.

EXAMPLE 1. East, West, North, and South model.
Let the neighborhood be

$$Q = \{(1,0),(0,1),(-1,0),(0,-1)\}.$$

The equation for $y(\cdot)$ is

$$y(s_1,s_2) + \phi_1 y(s_1+1,s_2) + \phi_2 y(s_1,s_2+1) + \phi_3 y(s_1-1,s_2) + \phi_4 y(s_1,s_2-1) = \sqrt{\rho}\, u(s_1,s_2).$$

The transfer function is given by

$$H\left(\exp\left(\sqrt{-1}\,\lambda_{ij}\right),\phi\right) = \left[1 + \phi_1 e^{\sqrt{-1}\,\lambda_0 i} + \phi_2 e^{\sqrt{-1}\,\lambda_0 j} + \phi_3 e^{-\sqrt{-1}\,\lambda_0 i} + \phi_4 e^{-\sqrt{1}\,\lambda_0 j}\right]^{-1}.$$

The corresponding spectral density function evaluated at $\lambda_{ij} = (\lambda_0 i, \lambda_0 j)$ is

$$S_y(\lambda_{ij};\phi,\rho) = \rho\left(1 + \phi_1 e^{\sqrt{-1}\,\lambda_0 i} + \phi_2 e^{\sqrt{-1}\,\lambda_0 j} + \phi_3 e^{-\sqrt{-1}\,\lambda_0 i} + \phi_4 e^{-\sqrt{-1}\,\lambda_0 j}\right)^{-1} \times \left(1 + \phi_1 e^{-\sqrt{-1}\,\lambda_0 i} + \phi_2 e^{-\sqrt{-1}\,\lambda_0 j} + \phi_3 e^{\sqrt{-1}\,\lambda_0 i} + \phi_4 e^{\sqrt{-1}\,\lambda_0 j}\right)^{-1}.$$

For the isotropic case, $\phi_1 = \phi_2 = \phi_3 = \phi_4 = \phi$, we have

$$S_y(\lambda_{ij};\phi,\rho) = \rho(1 + 2\phi\cos\lambda_0 i + 2\phi\cos\lambda_0 j)^{-2}.$$

EXAMPLE 2. One-sided models of images.
Let the neighborhood be

$$Q = \{(0,-1),(-1,-1),(-1,0)\},$$

The corresponding equation being

$$y(s_1,s_2) + \phi_1 y(s_1,s_2-1) + \phi_2 y(s_1-1,s_2-1) + \phi_3 y(s_1-1,s_2) = \sqrt{\rho}\, u(s_1,s_2).$$

The transfer function for this system is given by

$$H\left(\exp\left(\sqrt{-1}\,\lambda_{ij}\right),\phi\right) = \left[1 + \phi_1 e^{-\sqrt{-1}\,\lambda_0 i} + \phi_2 e^{-\sqrt{-1}\,\lambda_0(i+j)} + \phi_3 e^{-\sqrt{-1}\,\lambda_0 j}\right]^{-1}.$$

The corresponding spectral density function is given by

$$S_y(\lambda_{ij};\phi,\rho) = \rho\left(1 + \phi_1 e^{-\sqrt{-1}\,\lambda_0 i} + \phi_2 e^{-\sqrt{-1}\,\lambda_0(i+j)} + \phi_3 e^{-\sqrt{-1}\,\lambda_0 j}\right)^{-1} \times \left(1 + \phi_1 e^{\sqrt{-1}\,\lambda_0 i} + \phi_2^{\sqrt{-1}\,\lambda_0(i+j)} + \phi_3 e^{\sqrt{-1}\,\lambda_0 j}\right)^{-1}.$$

For the special case when $\phi_2 = \phi_1\phi_3$ we obtain the familiar separable model [1–2].

Expression for the Probability Density of Transforms of Observations

In this section we derive an explicit expression for the probability density of the transform $\{z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\}$ of the observations $(y(s), s \in \Omega_s)$, given that the observations obey the model in (2.1).

To this end we first obtain an expression for $p(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda | \phi, \rho)$ and then integrate over (ϕ, ρ) by using an arbitrary but otherwise regular prior probability density function $p(\phi, \rho)$. An expression for $p(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda | \phi, \rho)$ is obtained by using the stochastic properties of finite Fourier transforms [17–18].

Consider the finite Fourier transforms $\{z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\}$ of the observations $(y(s), s \in \Omega_s)$ over a square grid $N \times N$, defined below:

$$z(\lambda_{ij}) = (N)^{-1} \sum_{s \in \Omega_s} \exp\left(-\sqrt{-1}\,\lambda_{ij}^T s\right) y(s), \qquad \lambda_{ij} \in \Omega_\lambda. \tag{2.4}$$

We now state a theorem regarding the distribution of $\{z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\}$.

THEOREM 1. *For an* infinite observation field, *the finite Fourier transforms* $\{z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\}$ *are* (*a*) *independent and* (*b*) *have a complex normal distribution with zero mean and variance*

$$E\left(z(\lambda_{ij}) z^*(\lambda_{ij})\right) = S_y(\lambda_{ij};\phi,\rho), \qquad \lambda_{ij} \in \Omega_\lambda, \tag{2.5}$$

where $S_y(\lambda_{ij};\phi,\rho)$ *is given by Eq.* (2.3).

Comments. (1) The distribution theory given in Theorem 1 is exact only when the observation field is infinite. For a finite field of observations the Fourier transforms are only approximately uncorrelated. Similarly, for finite observation fields, Eq. (2.5) is only approximately true. The smoother $S_y(\lambda_{ij};\phi,\rho)$ is in the vicinity of λ_{ij}, the better is this approximation. If $S_y(\lambda_{ij};\phi,\rho)$ is constant over $(-\pi,\pi)$, (2.5) is exact.

It is interesting to compare the expression obtained here for the probability density of the transforms of observations and the expressions obtained in Whittle

[12]. Whittle starts with an exact expression for the likelihood of the noisy variates $\{u(s), s \in \Omega_s\}$. Since for a general neighborhood model the Jacobian of the transformation from the noisy variates $u(\cdot)$ to the observations $y(\cdot)$ is difficult to evaluate, an approximate expression is used for the determinant. However, the expression obtained here is not an approximation to the likelihood function of the observations as in [12]. The density function considered here is the joint density of the finite Fourier transforms which is a one-to-one transformation with Jacobian unity (though a general proof can be given to establish this, a simple derivation is given in Appendix II for a 4×4 field). Consequently, the estimates of ϕ, ρ obtained by maximizing $p(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda | \phi, \rho)$ are only approximately the maximum likelihood estimates.

We now obtain an expression for the estimates of the unknown parameters ϕ, ρ. The expressions for the maximum likelihood estimates $\bar{\phi}$, $\bar{\rho}$ of (ϕ, ρ) are given below:

$$\bar{\phi} = \arg \min_{\phi \in R^m} \left\{ -\frac{1}{N} \sum_{\lambda_{ij} \in \Omega_\lambda} \ln \left\| H\left(\exp\left(\sqrt{-1}\,\lambda_{ij}\right), \phi\right) \right\|^2 \right.$$

$$\left. + \ln \sum_{\lambda_{ij} \in \Omega_\lambda} \|z(\lambda_{ij})\|^2 \Big/ \left\| H\left(\exp\left(\sqrt{-1}\,\lambda_{ij}\right), \phi\right) \right\|^2 \right\} \tag{2.6}$$

and

$$\bar{\rho} = \frac{1}{N^2} \sum_{\lambda_{ij} \in \Omega_\lambda} \|z(\lambda_{ij})\|^2 \Big/ \left\| H\left(\exp\left(\sqrt{-1}\,\lambda_{ij}, \bar{\phi}\right)\right) \right\|^2. \tag{2.7}$$

Equations (2.6) and (2.7) can be obtained by solving the two simultaneous equations in ϕ and ρ that result by equating the first derivatives of $\ln p(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda | \phi, \rho)$ w.r.t. ϕ and ρ to zero. We assume the existence of first and second derivatives of likelihood function w.r.t. ϕ and ρ.

To obtain $p(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda)$ we integrate $p(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda | \phi, \rho)$ over (ϕ, ρ) using an appropriate prior probability density function. We do not make any specific assumption about the structure of the prior probability density functions. They need be regular but are otherwise arbitrary. Letting $\theta^T = (\phi^T, \rho)$ and performing the integration

$$p\left(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\right) = \int \int p\left(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda | \phi, \rho\right) p(\phi, \rho)\, d\phi\, d\rho \tag{2.8}$$

we arrive at

THEOREM 2. *As the rectangle of observations becomes large in all dimensions, the probability density* $(p(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda)$ *is given by*

$$\ln p\left(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\right) = -\frac{N^2}{2} \ln \bar{\rho} - \tfrac{1}{2} \sum_{\lambda_{ij} \in \Omega_\lambda} \ln \left\| H\left(\exp\left(\sqrt{-1}\,\lambda_{ij}\right), \bar{\phi}\right) \right\|^2$$

$$- 2m \ln N + \ln p\left(\bar{\phi}, \bar{\rho}\right) - a(N) + b(m)$$

$$+ O(1) + O\left(\frac{1}{N}\right) \tag{2.9}$$

where $\bar{\phi}$ *and* $\bar{\rho}$ *are given in* (2.6) *and* (2.7), *respectively, and*

$$a(N) = 0.5(N^2 + N^2 \ln 2\pi) + \ln N, \tag{2.10}$$

$$b(m) = 0.5(m+1)\ln 2\pi, \tag{2.11}$$

and $O(1/N)$ *denotes a deterministic constant term behaving like* k_1/N *for large* N *where* k_1 *is independent of* N.

Comments. (1) Theorem 2 gives an explicit expression for the probability density of the transforms of observations from a random field characterized by the spectral density function

$$S_y(\lambda_{ij}; \phi, \rho).$$

(2) For large values of N, the contribution due to the $\ln p(\bar{\phi}, \bar{\rho})$ term is negligible compared to the first three terms in Eq. (2.9).

(3) The numerical values of the estimates $\bar{\phi}$ and $\bar{\rho}$ can be obtained by a gradient-type algorithm.

3. DECISION RULES FOR THE CHOICE OF NEIGHBORHOODS

We formulate the choice of neighborhood as a classical Bayesian decision theoretic problem. For an image it is reasonable to assume that an observation at (s_1, s_2) will not significantly depend on distant pixels. Hence we restrict our allowable neighborhood set to a maximum of eight neighbors, east, west, north, and south and the four diagonal neighbors. Thus our problem is to find an appropriate set of neighbors among the possible 2^8 neighborhood sets for the given image. We first define the notion of a class. A class is defined as a set of models having the same neighborhood set Q but differing in the parameters ϕ and ρ. The class C_i consists of all models of the form

$$C_i: \quad y(s) + \sum_{q_k \in Q_i} \phi_{i,k} y(s + q_k) = \sqrt{\rho_i}\, u(s), \qquad q_k \in Q_i, q_k \neq (0,0), \tag{3.1}$$

where $\phi_{i,k} \neq 0, k \in Q_i, i = 1, \ldots, r, \rho_i > 0, i = 1, 2, \ldots, r$, where r denotes the number of neighborhoods under consideration. ϕ_{ik}, and ρ_i, $i = 1, 2, \ldots, r$, are all unknown. Thus a class consists of an infinite number of models with the same neighborhood. The given set of observations $\{y(s), s \in \Omega_s\}$ is said to obey class C_i if $\{y(s), s \in \Omega_s\}$ obeys only one model in C_i. Two classes C_i and C_j are said to be mutually exclusive if the neighborhoods they represent differ in at least one neighbor. In our formulation of the problem of choosing an appropriate neighborhood, the unknown parameters to be estimated from the given data are the neighborhood set Q_i and the estimates of the unknown parameters ϕ, ρ. Let m_k denote the number of elements in the set Q_k.

Given r mutually exclusive neighborhoods $C_i, i = 1, 2, \ldots, r$, and a set of observations $\{y(s), s \in \Omega_s\}$, our aim is to find the most appropriate class for $\{y(s), s \in \Omega_s\}$. A decision rule for assigning $\{y(s), s \in \Omega_s\}$ to one of the r classes is designed to minimize a suitable criterion function chosen to reflect the particular needs of the problem.

Expression for $P(C_k|z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda)$

We first compute an expression for $P(C|z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda)$, the posterior probability density of the transforms of the given data having been generated by some model C_k, for every k, $k = 1, 2, \ldots, r$. Subsequently, we derive optimal decision rules to minimize the probability of error and discuss simplifications of the decision rule.

Let $\boldsymbol{\theta}_k^T = (\boldsymbol{\phi}_k^T, \rho_k), \boldsymbol{\phi}_k^T = (\phi_{k1}, \ldots, \phi_{km_k})^T$ and $p(\boldsymbol{\theta}_k|C_k)$ be the prior probability density function of the parameters under class C_k. an expression for $P(C_k|z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda)$ is given by

THEOREM 3. *Let the observations obey the class* C_k. *Then the posterior probability transform of the observations is given by*

$$\begin{aligned}
&\ln P\big(C_k|z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\big) \\
&= -\frac{N^2}{2} \ln \bar{\rho}_k - 2m_k \ln N - \tfrac{1}{2} \sum_{\lambda_{ij} \in \Omega_\lambda} \ln \left\| H_k\Big(\exp\big(\sqrt{-1}\,\lambda_{ij}\big), \bar{\boldsymbol{\phi}}_k\Big)\right\|^2 \\
&\quad + \ln p\big(\bar{\boldsymbol{\phi}}_k, \bar{\rho}_k|C_k\big) + \ln P(C_k) - \ln p\big(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\big) \\
&\quad - a(N) + b(m_k) + O(1),
\end{aligned} \tag{3.2}$$

where

$$a(N) = 0.5(N^2 + N^2 \ln 2\pi) + \ln N, \tag{3.3}$$

$$b(m_k) = (m_k + 1) \ln 2\pi. \tag{3.4}$$

$P(C_k), k = 1, 2, \ldots r$ are the prior probabilities of the classes, and $\bar{\boldsymbol{\phi}}_k$ and $\bar{\rho}_k$ have been defined in (2.6) and (2.7).

Proof. This follows from Theorem 2 and Bayes' rule,

$$P\big(C_k|z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\big) = \frac{p\big(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda|C_k\big)P(C_k)}{\sum_{n=1}^{r} p\big(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda|C_n\big)P(C_n)}. \tag{3.5}$$

Decision Rules

Consider a 0–1 loss function L which assigns unit cost to a wrong assignment of classes and zero cost otherwise, i.e.,

$$\begin{aligned}
L\big(C_i, d(y(s), s \in \Omega_s) = C_j\big) &= 0 \quad \text{if } C_i = C_j, \\
&= 1 \quad \text{if } C_i \neq C_j.
\end{aligned} \tag{3.6}$$

Since the finite Fourier transformation is one to one, the cost of wrong assignment of the observation set $\{y(s), s \in \Omega_s\}$ is the same as the cost of wrong assignment of the set $\{z(\lambda_{ij}\}, \lambda_{ij} \in \Omega_\lambda)$. Our intention is to choose the decision rule

to minimize the risk function

$$J(d) = \sum_{k=1}^{r} P(C_k) \int L\big(C_k, d(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda)\big) \times p\big(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda | C_i\big)\, d\,|z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda|. \tag{3.7}$$

Substituting the loss function in (3.6), we have

$$J\big(d(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda) = C_n\big) = \int \sum_{\substack{k=1 \\ k \neq n}}^{r} P\big(C_k | z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\big) \times p\big(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\big)\, d\,|z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda| \tag{3.8}$$

and the optimal decision rule is

$$d^*\big(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\big) = \arg \max_n \big\{P\big(C_n | z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\big)\big\}. \tag{3.9}$$

Substituting the posterior probability function from Theorem 3 in (3.9) and simplifying we get the following rule:

Choose class C_{k^*} if

$$k^* = \arg\Big\{ \min_n \big\{h_n\big(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\big\}\Big\}, \tag{3.10}$$

where

$$h_n\big(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\big) = \frac{N^2}{2} \ln \bar{\rho}_n + m_n \ln N^2 + \frac{1}{2} \sum_{\lambda_{ij} \in \Omega_\lambda} \ln \Big\| H_n\big(\exp\big(\sqrt{-1}\,\lambda_{ij}\big), \bar{\phi}_n\big)\Big\|^2 + m_n \ln 2\pi - \ln p\big(\bar{\theta}_n | C_n\big) - \ln P(C_n) + b(m_n), \tag{3.11}$$

where $\bar{\theta}_n = (\bar{\phi}_n^T, \bar{\rho}_n)^T$ is given by (2.6) and (2.7), and $b(m_n)$ is given by (3.4).

Comments. (1) The decision statistics in Eq. (3.11) involve prior probabilities of the hypotheses, $P(C_i)$, $i = 1, 2, \ldots, r$ and the probability density functions, $p(\bar{\theta}_n | C_n)$. For simplicity, we assume $P(C_n) = 1/r$, $n = 1, 2, \ldots, r$.

(2) The quantities $\bar{\rho}_n$ and $\bar{\phi}_n$ are computed using Eqs. (2.6) and (2.7), $p(\bar{\theta}_n | C_n)$ is computable for any known regular structure of the density function and $b(m_n)$ is computable using (3.4).

(3) For large values of N, the contribution due to $b(m_n)$ can be ignored.

Simplified Decision Rules

The decision rule given in (3.12) involves arbitrary quantities such as prior probability densities. Hence we suggest a decision rule in which the prior densities

are suppressed. The decision rule no longer minimizes the average probability of error but can be shown to be asymptotically consistent using a proof similar to one found in [19].

The simplified decision rule is:

Choose class k^* if

$$k^* = \arg\Big\{ \min_n g_n\big(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\big)\Big\}, \tag{3.12}$$

where

$$g_n\big(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\big) = N^2 \ln \bar{\rho}_n + 2m_n \ln N + \sum_{\lambda_{ij} \in \Omega_\lambda} \ln \Big\| H\Big(\exp\big(\sqrt{-1}\,\lambda_{ij}\big), \bar{\phi}_n\Big)\Big\|^2. \tag{3.13}$$

Equation (3.11) can be written as

$$g_n\big(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\big) = N^2 \ln \rho_n^* + 2m_n \ln N, \tag{3.14}$$

where

$$\ln \rho_n^* = \ln \bar{\rho}_n + \frac{1}{N^2} \sum_{\lambda_{ij} \in \Omega_\lambda} \ln \Big\| H_n\Big(\exp\big(\sqrt{-1}\,\lambda_{ij}\big), \bar{\phi}_n\Big)\Big\|^2. \tag{3.15}$$

The structure of the decision rule in (3.14) is characteristic of the Bayesian approach [15, 19–20].

4. PROPERTIES OF THE DECISION RULE

Asymptotic consistency. One of the important properties of a decision rule is the consistency property. A decision rule is said to be asymptotically consistent if the probability of choosing an incorrect model given the correct model goes to zero as the number of observations goes to infinity. We do not given an explicit proof for the consistency of the decision rule suggested in the previous section. A proof similar to that in [19] can be given to establish asymptotic consistency of the decision rule.

Generality. The theory developed here is valid for autoregressive spatial models. The theory can be easily extended to include moving average and autoregressive moving average models.

Parsimony. The expression for the decision function $g_k(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda)$ brings out the disadvantage of having too large a value for m_k. If we increase m_k, then ρ_i^* decreases, causing a *decrease* in $\ln \rho_i^*$. Thus $N^2 \ln \rho_i^*$ and $2m_k \ln N$ balance each other. This illustrates the desirability of keeping the unknown parameters to a minimum.

Transitivity. The decisions regarding pairwise comparison of the classes are transitive. This is because the decision function $g_k(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda)$ does not depend on any parameter outside class C_k.

5. SPECIAL CASES (ONE-SIDED RANDOM FIELD MODELS)

In this section we apply the general theory developed in the previous section to the random fields represented by one-sided models. For a weakly stationary process we have [18]

$$\ln \rho = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \ln S_y(\lambda; \phi, \rho)\, d\lambda. \tag{5.1}$$

Using the relation connecting the transfer functions and the spectral density we have

$$\ln \rho = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \ln \rho \left\| H\left(\exp(\sqrt{-1}\,\lambda), \phi\right)\right\|^2 d\lambda \tag{5.2}$$

$$= \ln \rho - \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \ln \left\| H\left(\exp(\sqrt{-1}\,, \lambda), \phi\right)\right\|^2 d\lambda. \tag{5.3}$$

Hence we have

$$\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \ln \left\| H\left(\exp(\sqrt{-1}\,\lambda), \phi\right)\right\|^2 d\lambda = 0. \tag{5.4}$$

Approximating the double integration by double summation, (5.4) reduces to

$$\sum_{\lambda_{ij} \in \Omega_\lambda} \ln \| H\left(\exp\left(\sqrt{-1}\,\lambda_{ij}\right), \phi\right)\|^2 = 0. \tag{5.5}$$

Using (5.5), the decision rule for these models is:
Choose k^* if

$$k^* = \arg\left\{\min\left\{g_n\left(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\right)\right\}\right\}, \tag{5.6}$$

where

$$g_n\left(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\right) = N^2 \ln \bar{\rho}_n + 2m_n \ln N. \tag{5.7}$$

Thus the main difference between the test statistic for the one-sided and the neighborhood models is due to the term

$$\sum_{\lambda_{ij} \in \Omega_\lambda} \ln \left\| H_n\left(\exp\left(\sqrt{-1}\,\lambda_{ij}\right), \bar{\phi}_n\right)\right\|^2.$$

6. DISCUSSION

We discuss the relevance of the theory developed here for applications in image modeling and texture characterization.

One of the main goals of any model building scheme is to develop appropriate representations for the given data. It is known that even for the comparatively less complicated autoregressive time series models, an appropriate model should be

used for good results in prediction and forecasting. Hence for the more complicated case of a stationary random field, the importance of choosing an appropriate neighborhood need not be overemphasized. Initially one often appeals to the physics of the process to choose the structure of the model. In applications like image modeling, where one deals with two-dimensional data, it is difficult to visualize the physics of the underlying random field.

Several approaches have been suggested in the literature for choosing the appropriate neighborhood for stationary random fields. The classical hypothesis testing approach has been suggested by Whittle [12] to choose between different neighborhoods. By approximating the Jacobian of the transformations Whittle obtains an approximate expression for the likelihood of observations of a random field. Likelihood ratio tests and significance levels have been used to identify a neighborhood. Whittle's procedure becomes very complicated when models other than autoregression are considered. Even for autoregressive models, the evaluation of the Jacobian is a nontrivial task. Also, the decision rules using pairwise hypothesis tests are not consistent and transitive. Larimore [14] has reconsidered the problem of inference of random fields using a spectral representation and the AIC criterion. However, even in the case of one-dimensional autoregressive models, the AIC criterion does not yield consistent decision rules [15]. Hence, it is not desirable to use the AIC criterion for the choice of neighborhood in a random field model.

The application of random field models in image processing and texture characterization has been considered only recently. Tou *et al.* [10] have used two-dimensional unilateral models for texture characterization. The neighborhood has been chosen using empirical autocorrelation functions. Neighborhood models have been used for image coding and restoration problems in [3, 5, 6]. The models considered are generally isotropic and no precise model identification schemes have been suggested. Akaike's FPE criterion has been used in [6–7] for the choice of appropriate neighborhood. Instead of using a truly two-dimensional procedure, the dependence along rows and columns has been determined independently. Also, least-square estimates, which are not consistent estimates of the parameters, have been used. We believe that in the absence of any consistent neighborhood selection rule for a stationary random field, the theory developed in this paper is relevant for image modeling, texture characterization, and the much broader problem of statistical inference of a stationary random field.

APPENDIX I

We prove Theorem 2.

Consider Eq. (2.9) repeated below:

$$\ln p\big(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda \mid \phi, \rho\big) = -\frac{N^2}{2} \ln 2\pi\rho - \frac{1}{2} \sum_{\lambda_{ij} \in \Omega_\lambda} \left[\ln \left\| H\Big(\exp\big(\sqrt{-1}\,\lambda_{ij}\big), \phi\Big) \right\|^2 + (1/\rho) \|z(\lambda_{ij})\|^2 \Big/ \left\| H\Big(\exp\big(\sqrt{-1}\,\lambda_{ij}\big), \phi\Big) \right\|^2 \right] \tag{1}$$

or letting $\theta^T = (\phi^T, \rho)$

$$p\big(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda | \theta\big) = (1/2\pi) \exp\left[\frac{N^2}{2} G(\phi, \rho)\right], \tag{2}$$

where

$$G(\phi, \rho) = -\left[\ln \rho + \frac{1}{N^2} \sum_{\lambda_{ij} \in \Omega_\lambda} \ln \left\| H\big(\exp\big(\sqrt{-1}\,\lambda_{ij}\big), \phi\big)\right\|^2 \right.$$

$$\left. + \frac{1}{N^2 \rho} \sum_{\lambda_{ij} \in \Omega_\lambda} \|z(\lambda_{ij})\|^2 \Big/ \left\| H\big(\exp\big(\sqrt{-1}\,\lambda_{ij}\big), \phi\big)\right\|^2 \right]. \tag{3}$$

To compute $p(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda)$ we integrate $p(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda | \theta)$ over θ by using a prior probability density $\rho(\theta)$:

$$p\big(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\big) = \int p\big(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda | \theta\big) p(\theta)\, d\theta. \tag{4}$$

Substituting (2) in (4) we have

$$p\big(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda\big) = (1/2\pi)^{N^2/2} \int \exp\left[\frac{N^2}{2} G(\theta)\right] p(\theta)\, d\theta. \tag{5}$$

Expanding $G(\theta)$ in Taylor's series in θ about $\theta = \bar{\theta}$, where

$$\bar{\theta} = \max_{\theta} G(\theta),$$

we have

$$\text{LHS of (5)} = \left(\frac{1}{2\pi}\right)^{N^2/2} \int \exp\left(\frac{N^2}{2}\right) \Big[G(\bar{\theta}) + [\nabla_\theta G(\theta)]^T_{\theta = \bar{\theta}} (\theta - \bar{\theta})$$

$$+ (\theta - \bar{\theta})^T \nabla^2_{\theta_i \theta_j} G(\theta)|_{\theta = \bar{\theta}} (\theta - \bar{\theta}) + \cdots \Big] p(\theta)\, d\theta. \tag{6}$$

Since

$$G(\theta) = \ln p\big(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda | \theta\big) + \text{const (independent of } \theta).$$

$\bar{\theta}$ is the maximum likelihood estimate of θ defined in Eq. (2.6) and (2.7).

By using the definition of $\bar{\theta}$ the linear term in the exponent vanishes. Thus the LHS of (6) reduces to

$$(1/2\pi)^{N^2/2} \exp\left[\left(\frac{N^2}{2}\right) G(\bar{\theta})\right]$$

$$\int \exp\left[\left(\frac{N^2}{2}\right) (\theta - \bar{\theta})^T \nabla^2_{\theta_i \theta_j} G(\theta)\Big]_{\theta = \bar{\theta}} (\theta - \bar{\theta}) + \cdots \right] p(\theta)\, d\theta \tag{7}$$

$$= \left(\frac{1}{2\pi}\right)^{N^2/2} \exp\left[\left(\frac{N^2}{2}\right) G(\bar{\theta})\right] p(\bar{\theta}) \left(\frac{2\pi}{N^2}\right)^{(m+1)/2}$$

$$\times \frac{1}{\left\{\det\left[\nabla^2_{\theta_i \theta_j} (G(\theta))\right]_{\theta = \bar{\theta}}\right\}^{1/2}} + O(1/N). \tag{8}$$

Observing that

$$G(\bar{\theta}) = -\left[\ln \bar{\rho} + (1/N^2) \sum_{\lambda_{ij} \in \Omega_\lambda} \ln \| H\left(\exp \sqrt{-1}\, \lambda_0(i+j), \bar{\phi}\right)\|^2 + 1\right]$$

and

$$\left\{\det\left[\nabla^2_{\theta_i,\theta_j}(G(\theta))\right]_{\theta = \bar{\theta}}\right\}^{1/2} = O(1)$$

and defining

$$a(N) = 0.5(N^2 + N^2 \ln 2\pi) + \ln N,$$
$$b(m) = (m+1) \ln 2\pi$$

we arrive at Theorem 2.

APPENDIX II

LEMMA. *The Jacobian of the transformation from the observation set* $(y(s), s \in \Omega_s)$ *to the finite Fourier transform* $(z(\lambda_{ij}), \lambda_{ij} \in \Omega_\lambda)$ *is unity.*

Proof. For simplicity, we consider a 4×4 case. We have

$$z(\lambda_{ij}) = (N)^{-1} \sum_{s \in \Omega_s} \exp\left(-\sqrt{-1}\, \lambda_0(is_1 + js_2)\right) y(s_1, s_2).$$

For a 4×4 case, $N = 4$,

$$\Omega_s = \{i,j\} \qquad 1 \le i,j \le 4$$

and

$$\Omega_\lambda = \left\{\frac{2\pi i}{4}, \frac{2\pi j}{4}\right\}, \qquad 1 \le i,j \le 4$$

In matrix notation,

$$\mathbf{z} = \mathbf{JY},$$

where $\mathbf{z}$ and $\mathbf{Y}$ are vectors of finite Fourier transforms and observations arranged in lexicographic order. The matrix $\mathbf{J}$ (16×16) can be written as

$$\mathbf{J} = (1/4)\begin{bmatrix} \mathbf{A} & e^{-\sqrt{-1}(\pi/2)}\mathbf{A} & e^{-\sqrt{-1}(2\pi/2)}\mathbf{A} & e^{-\sqrt{-1}(3\pi/2)}\mathbf{A} \\ e^{-\sqrt{-1}(\pi/2)}\mathbf{A} & e^{-\sqrt{-1}(3\pi/2)}\mathbf{A} & e^{-\sqrt{-1}(\pi/2)}\mathbf{A} & e^{-\sqrt{-1}(3\pi/2)}\mathbf{A} \\ e^{-\sqrt{-1}(2\pi/2)}\mathbf{A} & e^{-\sqrt{-1}(\pi/2)}\mathbf{A} & e^{-\sqrt{-1}2\pi}\mathbf{A} & e^{-\sqrt{-1}(3\pi/2)}\mathbf{A} \\ e^{-\sqrt{-1}(3\pi/2)}\mathbf{A} & e^{-\sqrt{-1}(3\pi/2)}\mathbf{A} & e^{-\sqrt{-1}(3\pi/2)}\mathbf{A} & e^{-\sqrt{-1}(3\pi/2)}\mathbf{A} \end{bmatrix},$$

where

$$\mathbf{A} = \begin{bmatrix} e^{-\sqrt{-1}\,\pi} & e^{-\sqrt{-1}\,(3\pi/2)} & e^{-\sqrt{-1}\,2\pi} & e^{-\sqrt{-1}\,(\pi/2)} \\ e^{-\sqrt{-1}\,(3\pi/2)} & e^{-\sqrt{-1}\,(\pi/2)} & e^{-\sqrt{-1}\,(3\pi/2)} & e^{-\sqrt{-1}\,(\pi/2)} \\ e^{-\sqrt{-1}\,2\pi} & e^{-\sqrt{-1}\,(3\pi/2)} & e^{-\sqrt{-1}\,\pi} & e^{-\sqrt{-1}\,(\pi/2)} \\ e^{-\sqrt{-1}\,(\pi/2)} & e^{-\sqrt{-1}\,(\pi/2)} & e^{-\sqrt{-1}\,(\pi/2)} & e^{-\sqrt{-1}\,(\pi/2)} \end{bmatrix}$$

Using Kronecker product notation,

$$\mathbf{J} = \mathbf{B} \times \mathbf{A}$$

where

$$\mathbf{B} = e^{-\sqrt{-1}\,(2\pi/2)}\mathbf{A}.$$

Hence

$$\det \mathbf{B} = \det \mathbf{A}.$$

From a theorem regarding the characteristic roots of Kronecker products [21] the characteristic roots of $\mathbf{A} \times \mathbf{B}$ are $a_i b_j$, where a_i are the characteristic roots of $\mathbf{A}$ and b_j are the characteristic roots of $\mathbf{B}$. Hence,

$$\det \mathbf{J} = (1/4)^{16} \prod_{1 \le i,j \le 4} a_i b_j = (1/4)^{16} (\det A)^4 (\det B)^4$$
$$= 1$$

by direct evaluation of $\det \mathbf{A}$.

ACKNOWLEDGMENTS

The continued interest and encouragement of Professor Azriel Rozenfeld is sincerely acknowledged.

REFERENCES

1. A. Habibi, Two-dimensional Bayesian estimate of images, *Proc. IEEE* **60**, July 1972, 878–883.
2. D. P. Panda and A. C. Kak, Recursive least squares smoothing of noise in images, *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-25**, Dec. 1977, 520–524.
3. A. K. Jain and J. R. Jain, Partial difference equations and finite difference methods in image processing, 2, Image restoration, *IEEE Trans. Automat. Control* **AC-23**, Oct. 1978, 817–833.
4. E. J. Delp, R. L. Kashyap, and O. R. Mitchell, Image data compression using autoregressive time series models, *Pattern Recognition* **11**, Dec. 1979, 313–323.
5. A. K. Jain and E. Angel, Image restoration, modeling and reduction of dimensionality, *IEEE Trans. Computers* **C-23**, May 1974, 470–476.
6. K. Deguchi and I. Morishita, Texture characterization and texture-based image partitioning using two-dimensional linear estimation techniques, *IEEE Trans. Computers* **C-27**, August 1978, 739–745.
7. K. Deguchi and I. Morishita, Image coding and reconstruction by two-dimensional optimal linear estimation, Proc. 4th IJCPR, pp. 530–532, Nov. 1978.
8. R. L. Kashyap, Univariate and multivariate random field models for images, Workshop on Image Modeling, Chicago, August 1979.
9. B. H. McCormick and S. N. Jayaramamurthy, Time series model for texture synthesis, *Internat. J. Comput. Inform. Sci.* **3**, 1974, 329–343.

10. J. T. Tou *et al.*, Pictorial texture analysis and synthesis, Proc. 3rd IJCPR, Nov. 1976.
11. J. W. Woods, Two-dimensional discrete Markovian fields, *IEEE Trans. Inform. Theory*, **IT-28**, March 1972, 232–240.
12. P. Whittle, On stationary processes in the plane, *Biometrika* **41**, 1954, 434–449.
13. R. L. Kashyap, A Bayesian comparison of different classes of dynamic models using empirical data, *IEEE Trans. Automat. Control* **AC-22**, Oct. 1977, 715–727.
14. W. E. Larimore, Statistical inference on stationary random fields, *Proc. IEEE* **65**, June 1977, 961–970.
15. R. L. Kashyap, Inconsistency of the AIC rule for estimating the order of autoregressive models, *IEEE Trans. Aut. Control*, **AC-25**, October 1980, 996–997.
16. M. S. Bartlett, *The Statistical Analysis of Spatial Patterns*, Chapman & Hall, London, 1975.
17. D. R. Brillinger, Fourier analysis of stationary processes, *Proc. IEEE* **62**, Dec. 1974, 1628–1642.
18. L. H. Koopmans, *The Spectral Analysis of Time Series*, Academic Press, New York, 1974.
19. R. L. Kashyap, Optimal choice of AR and MA parts in autoregressive and moving average models, unpublished notes.
20. G. Schwarz, Estimating the dimensions of a model, *Ann. Statist.* **6**, March 1978, 461–464.
21. R. Bellman, *Introduction to Matrix Analysis*, McGraw–Hill, New York, 1960.