Image models are very useful for image coding, compression, segmentation, interpretation as well as image enhancement and restoration. For many images in practical applications, statistical information is most important. This report deals with the fundamental statistical theory of image models including the topics of contextual analysis, stochastic random field, the local and global properties of the random field, ARMA systems, and the applications of the statistical image models.

The support of the Office of Naval Research Statistics and Probability Program is gratefully acknowledged.
I. INTRODUCTION

The way a digital image is processed depends largely on how it is modeled. Generally speaking, image models are useful in image coding, compression, interpretation, classification, texture characterization as well as image enhancement and restoration. Both statistical and structural models of images have been considered as the images contain both statistical and structural information. Purely structural models are too regular to be interesting. In most practical applications especially in the defense area, the statistical information is most important. With these applications in mind, this report deals with the fundamental theoretical topics in statistical image modeling. An extensive list of references is provided to cover many publications in this important area.

II. THE CONTEXTURAL ANALYSIS

Consider a digital image with M x N picture elements, i.e., M rows and N columns. A simple approximation of contextual dependence for the two-dimensional patterns is called Markov mesh [1], which is considered as a two-dimensional Markov chain. Assume that the image is partitioned into m x n subimages. Then this two-dimensional Markov chain is characterized by a transition probability matrix $P_{m,n}$ defined as

$$P_{m,n} = \begin{pmatrix}
P_{1,1} & P_{1,2} & \cdots & P_{1,b-1} & P_{1,b} & \cdots & P_{1,n} \\
P_{2,1} & P_{2,2} & \cdots & P_{2,b-1} & P_{2,b} & \cdots & P_{2,n} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
P_{a-1,1} & P_{a-1,2} & \cdots & P_{a-1,b-1} & P_{a-1,b} & \cdots & P_{a-1,n} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
P_{m,1} & P_{m,2} & \cdots & P_{m,b-1} & P_{m,b} & \cdots & P_{m,n}
\end{pmatrix}$$
where \( P_{ij} = P(x_j|x_i) \) is the transition probability. Here \( x_i, x_j \) are the vector measurements of subimages. It can be shown that for binary random patterns,

\[
P(p_{a,b} | p_{m,n}) = P(p_{a,b})
\]

That the transition probability \( P_{a,b} \) depends only on the transition probabilities of the neighboring subimages is a very important result in contextual analysis. The above result may be generalized to grayscale pictures. In general, the dependence on the eight neighboring subimages is most important in contextual analysis.

The use of neighbor dependence approximation for the spatial patterns was first studied by Chow [2]. The tree dependence he considered [3] can also be generalized to spatial patterns. In this case each subimage will depend on certain surrounding subimages in addition to the eight neighboring subimages. For image interpretation or classification, the compound decision theory provides a theoretical framework for decision making using the contextual information. However, the practical implementation of the compound decision rule has been limited to Markov dependence or neighbor dependence. Assume that a subimage depends only on its four adjacent subimages in the east-west and north-south directions, then the decision rule is to choose the class that maximizes the likelihood function [4].

\[
P(\theta)p(x_i | \theta) \prod_{i=1}^{4} P(n_i | \theta)
\]

where \( x \) and \( x_i \) correspond to vector measurements of the subimage under consideration and its neighboring subimage respectively and \( \theta = 1, 2, \ldots, n \) with \( n \) being the number of classes. Here
\[ P(x_i | \theta) = \sum_{j=1}^{m} P(n_{ij}, \theta_j)P(\theta_j | \theta) \]   \hspace{1cm} (4)

where \( \theta_j = 1, 2, \ldots, m \) and \( P(\theta_j | \theta) \) is the transition probability which is usually determined experimentally. If the four subimages in the four corners are also considered, then the likelihood function for the eight-neighbor dependence case is

\[ P(\theta) P(x | \theta) \prod_{i=1}^{8} p(x_i | \theta) \]   \hspace{1cm} (5)

where \( p(x_i | \theta) \) is also given by Eq. (4). The simple result given by Eqs. (3) and (5) is due to the assumption of conditional independence between the subimage and the neighboring subimages and the assumption that the contribution due to the adjacent subimages is independent of that due to subimages in four corners. Both assumptions are reasonable in theory though quite restrictive in practice. For example, the occurrence of one class at one subimage will affect the occurrences of its neighboring classes. Obviously, the results are not valid for dependence on arbitrary set of subimages in the neighborhood. Another problem with the compound decision rule is that the conditional probability densities are usually not available and the a priori information may not be accurate. Experimental results based on Gaussian assumption of probability densities have consistently demonstrated the performance improvement with the use of contextual information [4], [5]. Empirically it is possible to determine these densities from the histogram of gray levels of each subimage, which corresponds to an unconditional probability density [6].

As the statistical contextual analysis based on the neighborhood dependence model described above is quite restrictive in practice, further development in image modeling is much needed for contextual analysis.
III. STOCHASTIC RANDOM FIELD

The Markov random field is the most typical assumption in statistical models. Wong [7] considered the properties of a two-dimensional random field having finite first and second moments. He found that there is no continuous Gaussian random field of two dimensions which is both homogeneous and Markov of degree 1. A homogeneous random field has a covariance function that is invariant under translation as well as rotation. Woods [8] considered a more general definition of Markov mesh than that discussed in the last section. Hassner and Sklansky [9] also discussed a Markov random field model for images. They presented an algorithm that generates a texture from an initial random configuration and a set of independent parameters that specify a consistent collection of nearest neighbor conditional probabilities which characterize the Markov random field.

For many practical applications such as in military area, the homogeneous random field assumption is not valid because of the object boundaries. Nahi and Jahanshashi [10] suggested modeling the image as a background statistical process combined with a set of foreground statistical processes, each replacing the background in the regions occupied by the objects being considered. Let

\[ b(m,n) = \text{gray level at the mth row and nth column}, \]
\[ \gamma(m,n) = \text{a binary function carrying the boundary information}, \]
\[ b_b = \text{a sample gray level from the background process}, \]
\[ b_o = \text{a sample gray level from the object process}, \]
\[ v = \text{a sample gray level from the noise process}, \]

then the model can be written as

\[ b(m,n) = \gamma(m,n)b_o(m,n) + (1 - \gamma(m,n))b_b(m,n) + v(m,n) \]  

(6)

where \( \gamma \) incorporates the assumption of first order Markov process on the object boundaries.
Nuanq modeled image scan lines as a Markov jump process [11]. This model led to non-linear noise reduction and image segmentation algorithms that are superior to linear techniques. The recursive calculation of a conditional probability involving the boundary component of the scan lines was the key to the non-linear algorithms. Modestino [12] modeled the image as a marked point process evolving according to a spatial parameter. In another approach the image is considered as a spatially variant linear system superimposed by non-linear elements corresponding to object boundaries. Ingle and Woods [13] considered the use of a bank of Kalman filters corresponding to various correlation directions and demonstrated a considerable improvement in the visual quality compared with linear constant coefficient Kalman filtering. Chen [14] has employed an adaptive Kalman filtering that operates a generalized likelihood ratio test in parallel with the Kalman filter. An object boundary corresponds to a state jump that is detected and used to update the Kalman filter. It appears that both textural and temporal variations can be properly taken into account in the image enhancement.

IV. LOCAL AND GLOBAL MODELS

Local statistical image models emphasize on the use of local statistics while global models attempt a description of the random field by using the information from the entire field. In the absence of any knowledge or as assumption about the global process underlying a given image, one may attempt to describe the joint probability density of the gray level or other properties of the picture elements. To do this for the entire image involves extremely high dimensional space which is unrealistic. It would be easier to consider a small neighborhood. However, even for a 3 x 3 neighborhood, a nonparametric representation in a 9 dimensional space along with the asso-
Associated sample size and storage problems still present serious difficulty.

Thus even for local models, it would be desirable to "compress" the local properties to a low dimensional space. Local description of co-occurrence statistics for textures (e.g. [15]) uses only $2 \times 1$ neighborhoods. Different features can then be derived from the co-occurrence matrix for texture classification.

Most of the local models, however, use conditional probabilities of picture elements within a window, instead of their joint probability distributions described above. The Markov dependence assumption will make a picture element depend upon its neighbors. Let $r$ and $s$ be the row number and the column number associated with a picture element $x$. A conditional nearest-neighbor model is

$$P \{ x_{rs} | \text{all other values} \} = P \{ x_{rs} | x_{r-1,s}, x_{r+1,s}, x_{r,s-1}, x_{r,s+1} \}$$

and is also known as a Markov field. An efficient procedure to take into account the local dependence is the statistical theory of nearest-neighbor systems on a lattice [16]. If we consider the four neighbors in the east-west and north-south directions, then we have a non-causal model given by

$$x_{rs} = \beta_1 (x_{r-1,s} + x_{r+1,s}) + \beta_2 (x_{r,s+1}) + x_{r,s-1} + y_{rs}$$

where $r = 1, 2, \ldots, M$, $j = 1, 2, \ldots, N$; $\beta_1$ and $\beta_2$ are the coefficients to be determined; and $(y_{rs})$ is an uncorrelated Gaussian noise process with

$E[y_{rs}] = 0$, $\text{var} \{ y_{rs} \} = \sigma_i^2$, $i = 1, 2, \ldots, m$ where $m$ is the number of pattern classes under consideration. The coefficients $\beta_1$ and $\beta_2$ may also differ among various classes. A special case of Eq. (8) is the causal model given by

$$x_{rs} - y_i = \beta_1 (x_{r-1,s} - y_i) + (x_{r,s-1} - y_i) + y_{rs}$$
Addendum

Some new image processing result based on the adaptive Kalman filtering [14] is shown on this page. Figure A is the original scene. The object area with two airplanes and with additive noise is shown on Figure B. Figure C is the result of adaptive filtering. All pictures are in binary (two-level) form.
Each class $\omega_i$ corresponds to a set of parameters $(\mu_i, \beta_i, \sigma_i)$. The parameter values are given or estimated for each class. Classification consists in deciding which of the given sets of parameter values best describes, in terms of probabilities, the image to be classified. It is noted that the model given by Eq. (9) permits discrimination between classes $\omega_i$ and $\omega_j$ even when $(\mu_i, \sigma_i) = (\mu_j, \sigma_j)$ so long as $\beta_i \neq \beta_j$. That is the model performs classification by using the information about the inter-pixel correlation $\rho_s$ as well as the mean and standard deviation of gray levels. Both Eqs. (8) and (9) represent first order autoregressive model where the autoregression parameters $\beta_i$ describe the spatial correlation. An alternate expression for the model described by Eq. (9) is

$$x_{rs} = a_i + \beta_i [x_{r-1,s} + x_{r,s-1}] + \gamma_{rs}$$

$$= a_i + \beta_i z_{r,s} + \gamma_{rs}$$

where $a_i = (1-2\beta_i)\mu_i$, and $z_{r,s} = x_{r-1,s} + x_{r,s-1}$. Here $a_i$ and $\beta_i$ can be estimated from ordinary simple linear regression by the method of least squares. The maximum likelihood decision rule can be used to classify $x_{rs}$ which follows the Gaussian probability with all parameters represented by their least squares estimates.

It is noted that one difficulty with the model given by Eq. (8) is that, even if the $Y_{rs}$ and hence the $x_{rs}$ are Gaussian, the estimation of $\beta_1$ and $\beta_2$ from data is not a simple least-squares problem, because the Jacobian of the transformation from the noise variable $Y_{rs}$ to the observation $x_{rs}$ is difficult to evaluate. If the density function of the finite Fourier transform of $x_{rs}$ is considered, then the Jacobian of transformation will be unity [17]. Eqs. (8) and (9) can be generalized to higher order models. For the second order model, the picture elements identified as 1's and 2's in the following figure should be included in the linear model.
For the nth order model, the picture elements up to a distance of n should be included in the model. The expression is given by

\[ x_{rs} = \sum_{i=1}^{n} \beta_i (x_{r-l,s} + x_{r+l,s}) + \sum_{i=1}^{n} \beta_i' (x_{r,s-i} + x_{r,s+i}) + y_{rs} \] (11)

where the parameters \( \beta_i \), \( \beta_i' \) can be estimated by using the maximum likelihood principle [18] if only casual terms are used in Eq. (11).

For the global models considered, the Gaussian model is an oversimplification [19] even though it is mathematically tractable. The stationary Gaussian assumption requires that the mean vector of gray levels be a vector of identical components. Hunt [20] suggested to use a nonstationary Gaussian model which allows the mean vector to have unequal components.

It would be desirable to include some structural property in the global model. Matheron [21] used the term "regionalized variables" to emphasize the particular features of the picture elements whose complex mutual correlation reflects the structure of the underlying phenomenon. He assumed weak stationarity of the increments in the gray levels between picture elements. The second moment of the increments in the gray levels between picture elements at an arbitrary distance, called the variogram, is used to reflect the structure of the field. Knowledge of the variogram
is useful for the estimates of many global and local properties of the
field. A characterization similar to the variogram is given by the auto-
correlation function. However for real imagery good functional forms of
variogram and autocorrelation are seldom available. A reasonable approxi-
mation must be sought between the functional form of the image model and
the real data considered.

V. THE ARMA SYSTEMS

Eq. (11) is the autoregressive model in its general form. A more
general parametric model is called the autoregressive moving average (ARMA)
system which replaces \( Y_{rs} \) in Eq. (11) by a finite number of previous \( Y_{rs} \)
values. Of course only the causal terms are taken in Eq. (11). The
resulting ARMA or mixed model can provide a good representation of the
image with properly chosen coefficients. However an autoregressive model
of sufficient order should be adequate.

To determine the order of the model, a maximum likelihood decision
rule can be used for choices of neighbors [17]. A simpler procedure is
to use the Akaike Information Criterion (AIC) in each of the two dimensions
and thus the window size can be determined [22] by the final prediction error.
We consider two lines parallel to the horizontal and vertical axes passing
a point \((i, j)\), and employ one-dimensional estimators.

\[
\hat{x}_{i, j} = \sum_{p=1}^{\Pi} \beta^{(\Pi)}_{i} x_{i-p, j}
\]

\[
\hat{x}_{i, j} = \sum_{q=1}^{N} \beta^{(N)}_{i} x_{i, j-q}
\] (12)

where \( \Pi \) and \( N \) are the window sizes in the horizontal and vertical directions
respectively. Let \( S^{(\Pi)}_{\min} \) and \( S^{(N)}_{\min} \) be the minimum estimation errors. We deter-
mine the optimum values of $M$ and $N$ so that the two criterion functions

\[
FPE(M) = \frac{I + M + 1}{I - M - 1} S_{\min}^{(i)}
\]

\[
FPE(N) = \frac{J + N + 1}{J - M - 1} S_{\min}^{(N)}
\]

are minimized. Here the original image size is assumed as $I \times J$. A more accurate procedure would consider the two dimensions jointly and may lead to a smaller window size because of the correlation among the adjacent picture elements. Although, the AIC may lead to inconsistent result, it is by far the simplest criterion for determining the window size.

The coefficients in the autoregressive model should be determined from the autocorrelation function. However it would be desirable to develop an efficient two-dimensional Levinson recursion to compute the coefficients. The frequency domain analysis and the maximum entropy method are alternative procedures for estimation of coefficients.

The study of ARMA systems for image analysis is still at its infancy. Further development of ARMA systems is much needed for texture characterization and object boundary extraction.

VI. APPLICATIONS OF STATISTICAL IMAGE MODELS

In the ARMA systems, images are described by a few coefficients or parameters which may be coded for image transmission instead of transmitting the whole picture. This represents an important approach to image compression. Although a limited amount of effort has been made so far in this direction [23], development in two-dimensional ARMA models will find great applicability in image transmission.
For the image interpretation and classification, image models must emphasize on discrimination information such as the local statistics as the object and background must be statistically different. Different objects must have different model parameters. Although most image models are designed to represent or characterize an image rather than to discriminate among different objects or classes, useful discrimination information is contained in the image models.

Because of a large variety of images, many different image models are available. Statistical models are most effective for images which are rich in texture. The image processing techniques employed are often determined by image modeling. For example, Kalman filtering is particularly suitable for images modeled as a spatially variant linear system with additive noise. For images modeled with a multiplicative disturbance, a different image processing technique is required. On the other hand, preprocessed images usually make image modeling easier. Image models also provide useful knowledge about the image for image segmentation and restoration.

In summary, image modeling in general and statistical image modeling in particular provides an abstraction of rich information of various nature in the images and should be considered as an integral part of image analysis and synthesis.
REFERENCES


ADDITIONAL REFERENCES


