Advances in Mathematical Models for Image Processing

ANIL K. JAIN, MEMBER, IEEE

Invited Paper

Abstract—Several state-of-the-art mathematical models useful in image processing are considered. These models include the traditional fast unitary transforms, autoregressive and state variable models as well as two-dimensional linear prediction models. These models introduced earlier [51], [52] as low-order finite difference approximations of partial differential equations are generalized and extended to higher order in the framework of linear prediction theory. Applications in several image processing problems, including image restoration, smoothing, enhancement, data compression, spectral estimation, and filter design, are discussed and examples given.

I. Introduction

MATHEMATICAL models are becoming increasingly important because of their role in the development of useful algorithms for image processing. Virtually all applications of image processing utilize some sort of mathematical models. The continuing advances in high-speed digital processors, digital memories, and very-large-scale integration (VLSI) have led to successful algorithms for many difficult problems. Table I gives a description of some of the typical problems in image processing and their associated modeling requirements. A typical algorithm requires quantification of the processing criterion and a model such as bandwidth, power spectrum, etc., of the data (input to the algorithm). While most of the problems listed in Table I also occur in one-dimensional signal processing, special care is needed in the development of two- (and higher) dimensional algorithms. The major difference besides the higher dimensionality is that of causality. A large number of one-dimensional signal processing methods are based on the fact that the observed data is the output of a causal system. For two-dimensional images the data coordinates are spatial and any causality associated with an image is purely due to its scanning or acquisition technique. Therefore, it is not surprising that a large number of image processing algorithms for edge extraction, enhancement, restoration, data compression, etc., are noncausal.

The computational efficiency of algorithms is often measured by their memory and operation count requirements. The most efficient algorithms would be such that the required number of operations per pel would be independent of the size of the image. Unfortunately, a large number of algorithms require an operation count which is proportional to $\log N$, N, or higher for $N \times N$ images. Table II lists some of the desirable properties of two-dimensional models and algorithms which tend to minimize their computational complexity. In this paper, we will consider mathematical models which are useful for solving image processing problems such as listed in Table I and their tradeoffs with respect to the desirable characteristics of Table II.

We start by series expansion models in Section II, followed by one-dimensional stochastic models in Section III. In Sections IV, V, and VI two-dimensional causal, semicausal, and noncausal models, respectively, are discussed. In Section VII, we consider applications of these models in several image processing problems.

Notation and Definitions

1) We will denote two-dimensional sequences defined on a rectangular grid by $u_{i,j}$, $x_{i,j}$, etc. Upper case letters such as U, X, etc. will denote matrices. For example

$$U \triangleq \{u_{i,j}; 1 \leqslant i \leqslant N, 1 \leqslant j \leqslant M\}$$

denotes an $N \times M$ matrix whose elements are $u_{i,j}$. Whenever necessary we will also use $[U]_{i,j}$ to denote the (i,j)th element of a matrix U. The jth column of U will be denoted as u_j.

2) If A is a matrix then A^T is its transpose and A^* is its complex conjugate.

3) Let $\{u_{i,j}\}$ be an $N \times M$ sequence. We define an $NM \times 1$ lexicographic ordered vector u_r obtained from this sequence as

$$u_r = \mathcal{O}_r(u_{i,j}) \triangleq [u_{1,1}u_{1,2} \cdots u_{1,M}, u_{2,1} \cdots u_{2,M} \cdots u_{N,1} \cdots u_{N,M}]^T$$

where $\mathcal{O}_r$ signifies the row by row ordering operation. Similarly we define $\mathcal{O}_c$ as a column by column ordering operator and write

$$u_c = \mathcal{O}_c(u_{i,j}) \triangleq [u_{1,1}u_{2,1} \cdots u_{N,1}, u_{1,2} \cdots u_{N,2}, \cdots u_{1,M} \cdots u_{N,M}]^T.$$

4) When each $u_{i,j}$ is a random variable, we will call $\{u_{i,j}\}$, a discrete *random field*. A given image could be considered as a sample function of this random field.

5) The mean and covariances of a random field $\{u_{i,j}\}$ are defined as

$$Eu_{i,j} = \mu_{i,j}$$

$$\text{Cov}[u_{i,j}, u_{m,n}] \triangleq E(u_{i,j} - \mu_{i,j})(u_{m,n} - \mu_{m,n}) = r_u(i,j;m,n) \quad (1)$$

Manuscript received May 28, 1980; revised October 30, 1980. This work was supported by the U.S. Army Research Office, Durham, NC, under Grant DAAG29-78-G-0206.

The author is with the Signal and Image Processing Laboratory, Department of Electrical and Computer Engineering, University of California, Davis, CA 95616.

0018-9219/81/0500-0502$00.75 © 1981 IEEE

TABLE I
TYPICAL PROBLEMS IN IMAGE PROCESSING

	Problem	Description	Models
1.	SMOOTHING:	Given noisy image data filter it to smooth out the noise variations.	Noise & image, Power Spectra
2.	ENHANCEMENT	Bring out or enhance certain features of the image e.g., edge enhancement, contrast stretching, etc.	Features
3.	RESTORATION & FILTERING	Restore an image with known (or unknown) degradation as close to its original form as possible, e.g., image deblurring, image reconstruction, image registration, geometric correction etc.	Degradations, Criterion of "closeness".
4.	DATA COMPRESSION	Minimize the Number of Bits required to store/transmit an image for a given level of Distortion.	Distortion Criterion, Image as an Information source
5.	FEATURE EXTRACTION	Extract certain features from an image, e.g., edges.	Features, detection criterion
6.	DETECTION AND IDENTIFICATION	Detect and identify the presence of an object from a scene e.g., matched filter, pattern recognition and image segmentation, texture analysis etc.	Detection criterion, object and scene
7.	INTERPOLATION/ AND EXTRAPOLATION	Given image data at certain points in a region, estimate the image values of all other points inside this region (interpolation) and also at points outside this region (extrapolation).	Estimation Criterion, and Degree of smoothness of the data
8.	SPECTRAL ESTIMATION	Given image data in a region, estimate its power spectrum.	Criterion of Estimation, A-priori model for data
9.	SPECTRAL FACTORIZATION	Given the magnitude of the frequency response of a filter-design a realizable filter e.g., a stable 'causal' filter.	Criterion of realizability;
10.	SYNTHESIS	Given a description or some features of an image, design a system which reproduces a replica of that image; e.g., texture synthesis.	Features, Criterion of reproduction.

TABLE II
DESIRABLE PROPERTIES OF IMAGE PROCESSING ALGORITHMS

Property	Description
Linearity	Linear operations on data
Separability	Independent Row and Column operations
Shift Invariance	Operations leading to Toeplitz and Circulant Matrix Manipulations
Markovian or Finite Memory	Only local and/or sparse operation required in each pixel e.g., FIR Filters.

Often we will consider the special case when

$$\mu_{i,j} = \mu = \text{constant}$$

$$r_u(i, j; m, n) = r_u(i - m; j - n). \tag{2}$$

For notational simplicity, whenever there is no confusion, we will drop the subscript u. A random field satisfying (2) is also called *translational (or spatial) invariant, homogeneous, or wide-sense stationary*. For random fields with Gaussian statistics this also implies strict sense stationarity. Unless otherwise mentioned, the term "*stationary*" refers to wide-sense stationarity.

6) A random field $\{x_{i,j}\}$ will be called a *white noise field* whenever the random variables $\{x_{i,j}\}$ are mutually uncorrelated, i.e., its covariance function is of the form

$$r_x(i, j; m, n) = \sigma_x^2(i, j)\, \delta_{i,m}\delta_{j,n}$$

where $\delta_{i,m}$ is the Kronecker delta function and $\sigma_x^2(i, j)$ is the variance of $x_{i,j}$.

II. Series Expansion Models and Image Transforms

A. Unitary Transforms

A classical way of analyzing a function is by its series expansion in terms of a set of complete orthonormal functions. In the context of image processing a general orthogonal series expansion for an $N \times N$ image[1] $\{u_{i,j}\}$ is a pair of *unitary transformations* of the form

$$u_{i,j} = \sum_{k=1}^{N}\sum_{l=1}^{N} v_{k,l}\, a^*(i, j; k, l) \tag{3}$$

$$v_{k,l} = \sum_{i=1}^{N}\sum_{j=1}^{N} u_{i,j}\, a(i, j; k, l) \tag{4}$$

where $\{a(i, j, k, l)\}$, usually called the *image transform*, is a

[1] Much of the subsequent analysis is easily generalized to rectangular, $N \times M$ images. But for simplicity we shall only consider square images here.

TABLE III
TYPICAL FAST UNITARY TRANSFORM USED IN IMAGE PROCESSING

Transform	Formula
Discrete Fourier (DFT)	$a_{m,n} = \frac{1}{\sqrt{N}} \exp\left\{-j\frac{2\pi(m-1)(n-1)}{N}\right\}, 1 \le m,n \le N$
Discrete Cosine (DCT)	$a_{m,n} = \begin{cases} \frac{1}{\sqrt{N}}, & m=1,\ 1 \le n \le N \\ \sqrt{\frac{2}{N}} \cos\frac{(m-1)(2n-1)\pi}{2N}, & 2 \le m \le N,\ 1 \le n \le N \end{cases}$
Discrete Sine (DST)	$a_{m,n} = \sqrt{\frac{2}{N+1}} \sin\frac{mn\pi}{N+1}, \quad 1 \le m, n \le N$
Walsh-Hadamard (WHT)	$a_{m,n} = (-1)^{\sum_{i=0}^{p-1} m_i n_i}, \quad 1 \le m,n \le N=2^p$ m_i, n_i = ith binary digit (0 or 1) in the binary expansion of (m-1) and (n-1) respectively.

set of complete orthonormal basis functions satisfying the properties

$$\sum_{k=1}^{N}\sum_{l=1}^{N} a(i,j;k,l)\, a^*(i',j';k,l) = \delta_{i,i'}\delta_{j,j'} \quad \text{(Orthonormality)}$$

$$\sum_{i=1}^{N}\sum_{j=1}^{N} a(i,j;k,l)\, a^*(i,j;k',l') = \delta_{k,k'}\delta_{l,l'} \quad \text{(Completeness).} \tag{5}$$

The elements $v_{k,l}$ are called the *transform coefficients* and $\{v_{k,l}\}$ the *transformed image.* Equation (3) is a deterministic representation of an image considered as an $N^2 \times 1$ vector. Alternatively, the image $u_{i,j}$ is the output of a linear *spatially variant* (SV) *finite impulse response* (FIR) system whose *impulse response* (IR) or *point spread function* (PSF) is

$$h(i,j;k,l) = a^*(i,j;k,l), \qquad 1 \leqslant i,j,k,l \leqslant N \tag{6}$$

to a unit impulse $\delta_{i,k}\delta_{j,l}$. For each (i,j) the array $\{a^*(i,j;k,l)\}$ is also called a *basis image.* It is readily seen from (3) and (4) that the general unitary transformation would require $O(N^4)$ operations, one operation being a multiplication and a summation. For typical size ($N = 256$) images this means over a billion operations would be needed to compute the transform coefficients. To reduce dimensionality, the unitary transformations in (3) and (4) are restricted to the *product-separable* class, satisfying the condition

$$a(i,j;k,l) = a_{i,k}b_{j,l} \tag{7}$$

where $A = \{a_{i,j}\}$ and $B = \{b_{i,j}\}$ are unitary matrices (i.e., $A^{-1} = A^{*T}$). Often in image processing one chooses $B = A$ so that (3) and (4) yield

$$V = AUA^T \tag{8}$$

$$U = A^{*T}VA^*. \tag{9}$$

Now the transformations require column operations followed by row operations on the result, reducing the computations to $O(N^3)$ operations. Even this reduction is insufficient and the choice of image transforms is further restricted to *fast transforms.* Typically, these transform matrices have structural properties which lead to fast Fourier transform (FFT) type algorithms. Hence a transformation of the type $y = Ax$, for an $N \times 1$ vector x could be performed in $O(N \log N)$ operations so that for images the operation count is $O(N^2 \log N)$ or $\log N$ per pel. Examples of common fast unitary transforms are the discrete Fourier (DFT), cosine (DCT), sine (DST), Walsh-Hadamard (WHT) transforms [1]-[8] (see Table III), etc. Fig. 1 shows the basis images of the DCT. Other fast transforms include the Haar, Slant [1]-[3], and a family of sinusoidal transforms [9], [10]. A useful property of all unitary transforms is their energy conservation property

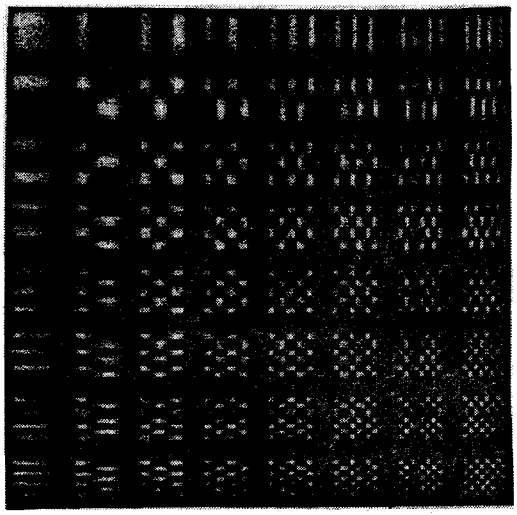


Fig. 1. 8 × 8 basis images of the DCT. Each 8 × 8 basis image if $B_{i,j}$ orthogonal to the rest and represents a spatially varying FIR of a system whose input is an impulse at (i,j).

$$\sum_{i=1}^{N}\sum_{j=1}^{N} |u_{i,j}|^2 = \sum_{k=1}^{N}\sum_{l=1}^{N} |v_{k,l}|^2 \tag{10}$$

known as Parseval's relation. This follows from the fact that a unitary transformation is simply a rotation of the image viewed as a vector in an N^2 dimensional vector space so that the length of the vector remains unchanged.

B. The Karhunen-Loeve Transform

Of particular significance among unitary transforms is the so-called Karhunen-Loeve transform (KLT) for random fields. Without loss of generality we are assuming zero-mean random fields. More generally, one could consider the autocorrelation function instead. It is the complete orthonormal set of basis images $\phi(i,j;k,l)$ determined from the eigenvalue equation

$$\sum_{m=1}^{N}\sum_{n=1}^{N} r(k,l;m,n)\,\phi(i,j;m,n) = \lambda_{i,j}\phi(i,j;k,l) \tag{11}$$

where $r(\cdot)$ is the image covariance function. For separable covariance functions, the KLT is also separable. Two significant properties which make the KLT very desirable are as follows [1], [11]-[13].

1) It completely decorrelates the transform coefficients, i.e.,

$$\text{Cov}\,[v_{k,l}(\mathcal{A}), v_{m,n}(\mathcal{A})] = \sigma^2_{k,l}(\mathcal{A})\,\delta_{k,m}\delta_{l,n}, \qquad \text{for } \mathcal{A} = \Phi \tag{12}$$

where $\mathcal{A}$ denotes an arbitrary $N^2 \times N^2$ unitary transform Φ

TABLE IV
TYPICAL COVARIANCE FUNCTION MODELS USED IN IMAGE PROCESSING

Model Description	Covariance Function $r(k,\ell) \triangleq \mathrm{Cov}[u_{i,j}, u_{i+k,j+\ell}]$	Comments
Separable	$\sigma^2 \rho_1^{\lvert k \rvert} \rho_2^{\lvert \ell \rvert}$	Typically $\rho_1 = \rho_2 = 0.95$ ρ_1, ρ_2 are one step correlation parameters
Nonseparable Exponential or Isotropic	$\sigma^2 \exp\{-\sqrt{\alpha_1 k^2 + \alpha_2 \ell^2}\}$	For $\alpha_1 = \alpha_2$, this is called the Isotropic model. $\rho_1 = \exp(-\alpha_1)$, $\rho_2 = \exp(-\alpha_2)$.

is the KLT and $\sigma^2_{k,l}(\mathfrak{A})$ are the variances of the $\mathfrak{A}$-transform coefficients $v_{k,l}(\mathfrak{A})$.

2) Compared to all other unitary transforms, the KLT packs the maximum expected energy in a given number of samples M, i.e.,

$$\sum_{k,l \in \mathfrak{I}(\Phi)} \sum \sigma^2_{k,l}(\Phi) > \sum_{k,l \in \mathfrak{I}(\mathfrak{A})} \sum \sigma^2_{k,l}(\mathfrak{A}), \quad \forall 1 \leq M \leq N^2 \tag{13}$$

where $\mathfrak{I}(\mathfrak{A})$ is the set containing M index pairs (k,l) corresponding to the largest M variances in the $\mathfrak{A}$-transform domain. This property serves as a basis for transform data compression techniques.

Table IV shows the two commonly used stationary covariance models used in image processing. For the separable model, the KLT is given by

$$\phi(i,j;k,l) = \psi_1(i,k)\,\psi_2(j,l) \tag{14}$$

where $\psi_n(i,j)$ is the KLT corresponding to the one dimensional covariance function $\rho_n^{|i-j|}$, $n = 1, 2$. Unfortunately, Ψ_n is not a fast transform. Depending on the value of ρ_n, it has been shown that a suitable fast sinusoidal transform [9], [10] could be found as a good approximation to the KLT. For example for $-0.5 \leq \rho_n \leq 0.5$, the sine transform (see Table III) and for $0.5 \leq \rho_n \leq 1$, the cosine transform [10], [14] are good substitutes for the KLT. For common monochrome images, the correlation parameters ρ_1 and ρ_2 are close to unity so that the DCT is the preferred fast transform. It has also been shown [10], [15] that the sinusoidal transforms have equivalent performance as $N \to \infty$. In image processing N can be quite large, and one often processes smaller blocks (typically 16×16) of an image at a time. The performance differences between the various transforms are significant enough to warrant the use of the KLT or a reasonable substitute of it. Recently it has been shown [16] that the separable DCT is a good substitute for the nonseparable KLT of other stationary random fields also, including those modeled by the nonseparable exponential covariance function shown in Table IV. Fig. 2 shows the data compression efficiency of the various unitary transforms for random fields modeled by the nonseparable exponential covariance function of Table IV. Here we are plotting the residual expected energy (also called the basis restriction error [10])

$$\sigma_e^2(\mathfrak{A}, M) = 1 - \left(\sum_{k,l \in \mathfrak{I}} \sum \sigma^2_{k,l}(\mathfrak{A}) \Big/ \sum_{k=1}^{N} \sum_{l=1}^{N} \sigma^2_{k,l}(\mathfrak{A}) \right) \tag{15}$$

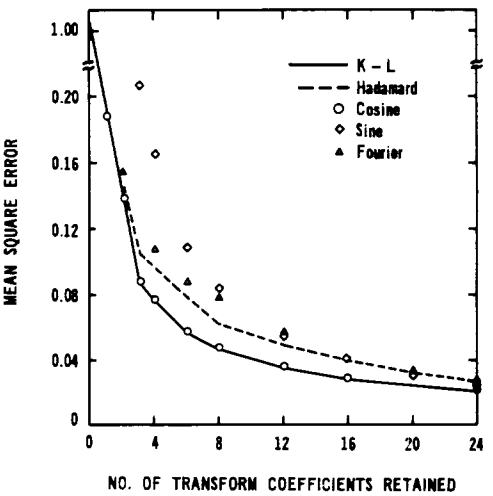


Fig. 2. Data compression efficiency of various transforms for 8×8 random fields with covariance function $r(k,l) = (0.95)^{\sqrt{k^2+l^2}}$.

remaining in the $N^2 - M$ samples of the $\mathfrak{A}$-transform. Clearly we see that the DCT has the best performance for the chosen values of ρ_1 and ρ_2. In the later sections we will see that the DCT and DST are useful in other statistical representations of images also. Image transforms have been applied extensively in data compression, noise smoothing, and restoration of images [1]-[3], [7], [17]-[19].

C. The Singular-Value Decomposition (SVD) Representation

Considering an $N \times N$ image as a matrix U of real numbers, it is possible to express it as [20]-[22]

$$U = \sum_{m=1}^{K} \lambda_m^{1/2}\, \Psi_m\, \Phi_m^T \tag{16}$$

where K is the rank of U and Ψ_m and Φ_m are the orthonormal eigenvector solutions of

$$U^T U \Phi_m = \lambda_m \Phi_m$$
$$U U^T \Psi_m = \lambda_m \Psi_m. \tag{17}$$

The quantities λ_m, $m = 1, \cdots, K$ are positive and (16) is called the singular value or the outer product expansion of U and can also be written as

$$U = \Psi \Lambda^{1/2} \Phi^T \tag{18}$$

where Ψ and Φ are $N \times K$ matrices whose column vectors are Ψ_m and Φ_m, respectively, and $\Lambda^{1/2}$ is the diagonal matrix of

elements $\{\lambda_m^{1/2}\}$. If the singular values λ_m are arranged in decreasing order then the partial sum

$$U_{\dot{M}} \triangleq \sum_{m=1}^{M} \lambda_m^{1/2} \Psi_m \Phi_m^T \tag{19}$$

is the best least squares approximation of U for any $M \geqslant 0$, i.e., the error

$$e_M = \sum_{i=1}^{N} \sum_{j=1}^{N} (u_{i,j} - u_{i,j}^{\cdot})^2 \tag{20}$$

is minimum for any fixed M. This means the energy packed in the M coefficients $\{\lambda_m, 1 \leqslant m \leqslant M\}$ is maximized by the SVD transform. We note that the KLT maximizes the *average energy* packed in M samples for an ensemble of images. Thus, for any given image, the SVD transform would be more efficient in terms of its data compression (and other least squares processing) ability. However, in view of (17) this transform would be different for each different image. Hence, unlike the KLT, a good fast transform substitute, independent of the image, cannot be found for the SVD. Although SVD transform has potential applications in image data compression and restoration problems [23]–[26], the computational burden introduced by it overwhelms its desirable least squares property. Fortunately, some of the iterative least squares algorithms [95] could be employed to achieve practically the same results. The SVD has other applications [27], e.g., in approximation of a two-dimensional power spectrum by a separable product of one-dimensional power spectra (see Section V), in the design of digital filters [22], [28] and also in texture analysis of images [29].

III. Image Representation by One-Dimensional Stochastic Processes

Often it is desired to design image processing algorithms for an ensemble of images. For practical reasons this ensemble is generally characterized by the mean and covariance functions. These functions could be specified by a mathematical formula (e.g., as in Table IV) or via the SDF, or simply as arrays of numerical values. An alternative is to consider the image ensemble as being generated by a linear system forced by white noise or a random sequence of known SDF. The impulse response of this linear system is often specified by a difference equation. Computational complexity as well as performance of various processing algorithms can be studied in terms of this difference equation.

A simple way to characterize an image is to consider it as a collection of one-dimensional signals, e.g., as an output of a raster scanner, or as a sequence of rows (or columns) ignoring the interrow (or column) dependencies. For such cases, one-dimensional representations of stochastic processes are useful. One dimensional stochastic models have been applied in line by line processing of images for DPCM coding, hybrid coding, recursive filtering and restoration, etc. [17], [38]–[42], [46], [52], [58].

A. Autoregressive Representations

If $\{u_k\}$ is a zero-mean stationary Gaussian random sequence, then a causal representation of the type

$$u_k, \forall k = \sum_{n=1}^{p} a_n u_{k-n} + \epsilon_k \qquad E\epsilon_k = 0 \qquad E\epsilon_k \epsilon_l = \beta^2 \delta_{k,l} \tag{21}$$

is called a (one-sided) autoregressive (AR) representation. The sequence $\{\epsilon_k\}$ is a zero-mean white-noise random process independent of the past outputs. AR models have the following important properties.

1) The quantity

$$\bar{u}_k = \sum_{n=1}^{p} a_n u_{k-n} = E[u_k | u_n, \forall n \leqslant k-1] \tag{22}$$

is the best mean-square predictor of u_k based on all of its past and depends only on the past p samples. Thus (21) becomes

$$u_k = \bar{u}_k + \epsilon_k \tag{23}$$

which says the sample at k is the sum of its minimum variance causal prediction estimate plus the prediction error. This is also called the *innovations representation.* The sequence $\{u_k\}$ defined by (21) is called a pth-order Markov process.

2) The AR process is stationary and causally stable (in the usual bounded-input–bounded-output (BIBO) sense) if and only if the roots of the polynomial

$$A_p(z) = 1 - \sum_{n=1}^{p} a_n z^{-n} \tag{24}$$

lie inside the unit circle. If stationary, then its spectral density function (SDF) is given by

$$S_u(z) = \beta^2 / [A_p(z) A_p(z^{-1})], \qquad z = e^{j\omega}, \quad -\pi < \omega < \pi. \tag{25}$$

From (21), it is seen that the transfer function of an AR representation is $1/A_p(z)$ which is an all pole model.

3) *Theorem 1:* Given an arbitrary set of positive definite real covariances $\{r_k\}$ on a window $W = \{-p \leqslant k \leqslant p\}$, there exists a unique AR model (21) whose parameters are identified by solving the linear Toeplitz system of equations

$$r_k - \sum_{n=1}^{p} a_n r_{k-n} = \beta^2 \delta_{k,0}, \qquad 0 \leqslant k \leqslant p. \tag{26}$$

Moreover, this model is causally stable and stationary.

The significance of this theorem is that the covariances generated by the model (obtained by the Fourier inverse of (25)) would match exactly the given covariances on the window W. The model covariances outside this window provide a positive definite *extrapolation* of the given covariance sequence.

An important consequence of this result is that any given SDF can be approximated arbitrarily closely by a finite-order AR spectrum. In other words, if $\{r_k\}$ is a covariance sequence corresponding to a given positive and analytic SDF $S(z)$, and if $\{a_n^p, n = 1, \cdots, p\}$ and β_p^2 denote the solution of (26), then

$$\lim_{p \to \infty} \beta_p^2 / [A_p(z) A_p(z^{-1})] = S(z), \qquad z = \exp(j\omega)$$

where $A_p(z)$ is given by (24) with a_n replaced by a_n^p. This relation states that the solution of (26), as $p \to \infty$, provides the spectral factorization of $S(z)$. If $S(z)$ happens to be a rational all pole spectrum, the a_n^p will be zero when $n > p$, for some $p < \infty$. In general (26) can be used to find a rational, all

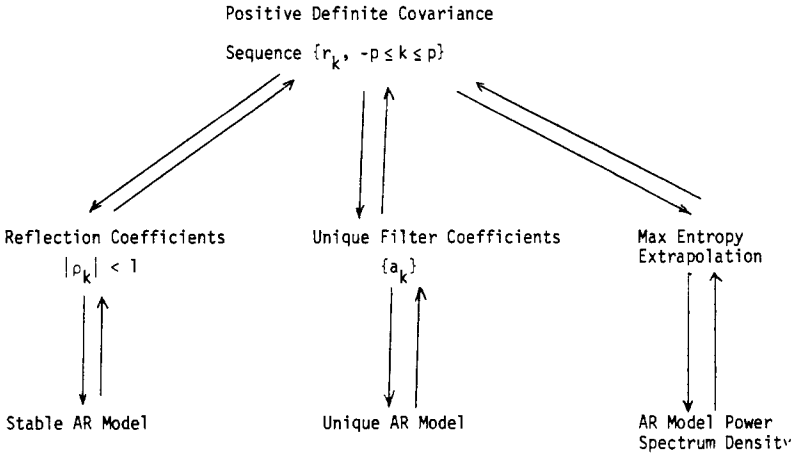


Fig. 3. Properties of the AR models.

pole spectral approximation of an arbitrary positive SDF. For these and other related ideas see Whittle [30] and Astrom [31].

4) *Levinson Algorithm [32], [33]:* The Toeplitz system of equations (26), also called the *normal equations*, can be solved in $O(p^2)$ operations by the recursions

$$a_{n+1,k} = \begin{cases} a_{n,k} - \rho_{n+1} a_{n,n+1-k}, & a_{n,0} = 1, \quad 1 \le k \le n \\ \rho_{n+1}, & k = n+1 \end{cases}$$

$$\beta_{n+1}^2 = \beta_n^2 (1 - \rho_{n+1}^2), \qquad \beta_0^2 = r_0$$

$$\rho_{n+1} = \frac{1}{\beta_n^2}\left[r_{n+1} - \sum_{k=1}^{n} a_{n,k} r_{n+1-k}\right], \qquad \rho_1 = r_1/r_0 \tag{27}$$

where $\beta^2 = \beta_p^2$, $a_n = a_{p,n}$ are the coefficients required in (26). The elements $\{\rho_n, 1 \le n \le p\}$ are called the *reflection coefficients*. If the covariance matrix $\{r_{k-n}, 0 \le k, n \le p\}$ is positive definite then it could be shown that $|\rho_n| < 1$. Moreover the sequences $\{r_k\}$, $\{a_k\}$ and $\{\rho_k\}$ are unique.

5) A necessary and sufficient condition for the stability of the AR model is that $|\rho_n| < 1$, $\forall n$, or equivalently that the matrix $\{r_{k-n}, 0 \le k, n \le p\}$ be positive definite.

6) *Theorem 2 (Maximum Entropy Extrapolation):* The given positive definite covariance sequence $\{r_k, -p \le k \le p\}$ has a unique maximum entropy extrapolation and is given by the covariances generated by the AR model of (21).

Although the proof of this result is available at various places in the literature [34], [36], the following simple proof is offered. We define as entropy of a covariance sequence $\{r_k, -N \le k \le N\}$ the average entropy of $(N+1)$ Gaussian random variables which is given by (within an additive constant)

$$H_N = \frac{1}{N+1} \log |R_N| \tag{28}$$

where $|R_N|$ is the determinant of the Toeplitz covariance matrix associated with the covariance sequence. From the theory of Toeplitz matrices it is easy to show that the determinant of R_N follows the simple recursion [35]

$$|R_N| = |R_{N-1}| \beta_N^2, \qquad \beta_N^2 \le \beta_{N-1}^2$$

From this, H_N is monotonically nonincreasing function of N. For $N \ge p$, to maximize H_N, we therefore simply need to have $\beta_k^2 = \beta_p^2$ for $N \ge k \ge p$. This means the optimum ρ_k are zero for $k > p$ and one obtains

$$a_{n+p,k} = \begin{cases} a_{p,k}, & \text{for } n \ge 0, \quad 1 \le k \le p \\ 0, & k \ge p+1. \end{cases} \tag{29}$$

Thus the pth-order AR model maximizes the entropy. By setting $\rho_{n+p} = 0$ in (27) the extrapolated covariances are obtained as

$$r_{n+p} = \sum_{k=1}^{p} a_k r_{n+p-k}, \qquad n \ge 1. \tag{30}$$

The SDF given by (25), determined via the coefficients $\{a_k\}$ is called the *maximum entropy spectrum* of the given covariance sequence. It is interesting to note that (29) provides the necessary conditions for extrapolation of $\{r_k\}$ from $k = p+1$ to $k = N$, for any $N > p$.

7) The reflection coefficients are such that at recursion step n in (27), the sum of the forward and backward predictive errors is minimized, i.e., if we define

$$\sigma_e^2 \triangleq E(\epsilon_{m,k}^+)^2 + E(\epsilon_{n,k}^-)^2 \tag{31}$$

$$\epsilon_{n,k}^+ \triangleq x_k - \sum_{m=1}^{n} a_{n,m} x_{k-m} = \epsilon_{n-1,k}^+ - \rho_n \epsilon_{n-1,k}^- \tag{32}$$

$$\epsilon_{n,k}^- \triangleq x_{k-n} - \sum_{m=1}^{n} a_{n,m} x_{k-n+m} = \epsilon_{n-1,k-1}^- - \rho_n \epsilon_{n-1,k}^+ \tag{33}$$

then ρ_n is such that σ_e^2 is minimized. This relation is useful in finding ρ_n directly when observed samples of the random process rather than their covariances are available [34].

Fig. 3 summarizes the properties of the AR models. In image processing, AR models have been found useful in modeling images line by line, as illustrated by the following examples.

Example 1: Consider zero-mean monochrome images whose covariance function models are listed in Table IV. If the image is processed one column at a time then the covariance function of pels on any column is of the form

$$r_k = E u_{i,j} u_{i+k,j} = \sigma^2 \rho_1^{|k|} \tag{34}$$

for both the models of Table IV. The corresponding AR representation is

$$x_i = \rho_1 x_{i-1} + \epsilon_i \qquad \beta^2 = E\epsilon_i^2 = \sigma^2(1 - \rho_1^2). \tag{35}$$

Example 2: In certain image processing applications, each column (or row) of an image is first unitarily transformed, i.e., for the jth column

$$v_j = Au_j, \quad j = 1, \cdots, N \tag{36}$$

and each element $v_j(i)$ is modeled as a first-order AR process, i.e.,

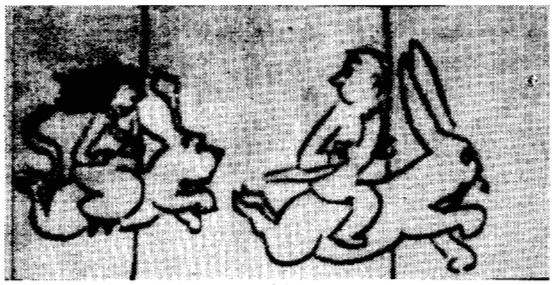
$$v_j(i) = a_1(i)\, v_{j-1}(i) + e_j(i), \quad i = 1, \cdots, N. \tag{37}$$

If A is the KLT, i.e., the elements $v_j(i)$ and $v_j(k)$ are uncorrelated (independent under Gaussian assumptions) for $i \neq k$, then (37) represents a set of N decoupled AR models. Such algorithms have been called *hybrid* (i.e., a combination of nonrecursive and recursive procedures) and have been used in image restoration and coding [17], [52], [58]. For a tutorial review of AR models see [37]. Practical examples are discussed in section VI.

Example 3: Common images represent a nonnegative luminance function. In many applications from the physics of image formation it is possible to model the image as a power spectrum. For example, in high-resolution radar imaging a target can be considered as a set of distributed scatterers. The overall radar cross section (RCS) of the target measured as a function of aspect angle has the properties of a power spectrum. Let $S(x)$, $-\frac{1}{2} < x < \frac{1}{2}$ represent a line of the image. Assume $S(x) > 0$ and that the Fourier series of $S(x)$ is uniformly convergent. Then its inverse Fourier transform is a valid covariance sequence. For a sampled image, if s_k, $\{0 \leqslant k \leqslant N-1\}$ represents a set of positive and bounded real numbers, it can be shown [38] that its DCT coefficients given by

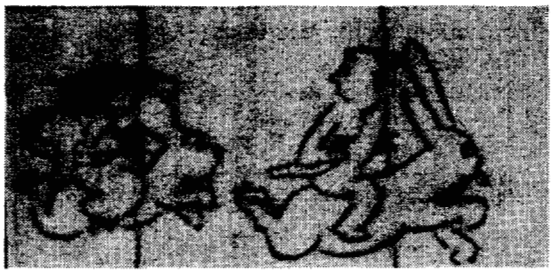
$$r_{-k} \triangleq r_k = \frac{1}{N} \sum_{m=0}^{N-1} s_m \cos \frac{\pi k}{N} \left(m + \frac{1}{2} \right), \quad 0 \leqslant k \leqslant N-1 \tag{38}$$

is a covariance sequence. Hence it is possible to find a pth-order AR model whose output covariances would match the $\{r_k\}$ over $[-p, p]$. In other words, one could synthesize each line of the image arbitrarily closely by a suitable AR model. Note that the image itself is being modeled by a deterministic function even though the AR model may characterize a random process whose covariances are related to the image. Fig. 4 shows an original and a synthesized image line by line by segments of $N = 32$ pels, generated by an eighth-order AR model. This figure also shows the image obtained by cosine transform coding when the same number of transform coefficients are retained. The higher resolution provided by the AR model approach is the result of the covariance extrapolation performed by that model. For other examples and details see [38].



(a)

(b)



(c)

Fig. 4. Line by line synthesis of an image by eighth-order AR models. Each line has been synthesized by line segments of 32 pixels. (a) Original. (b) Synthesis using Cosine transform, 4 to 1 sample reduction. (c) Synthesis using AR model, 4 to 1 sample reduction.

B. State Variable Models

State variable models have been used to represent two-dimensional images considered as the output of a raster scanner. The scanner output is modeled as a one-dimensional random process and is characterized by a set of state variable equations of the form

$$\frac{dx(t)}{dt} = A(t)x(t) + B(t)\varepsilon(t)$$
$$y(t) = C(t)x(t) \tag{39}$$

where $y(t)$ is the scanner output at time t, and $x(t)$ is an $n \times 1$ vector, $\varepsilon(t)$ is a $p \times 1$ zero-mean white-noise vector such that

$$E\varepsilon(t)\varepsilon^T(t') = K\delta(t - t'). \tag{40}$$

A, B, C, K are appropriate matrices which are determined such that $y(t)$ satisfies (approximately, if not exactly) the statistics of the scanner output.

The first attempt to model images by state variable techniques was made by Nahi and Asseffi [39], [40]. Although their final model has limitations because of several approximations, their modeling procedure does expose several difficulties in representing two-dimensional random field by one-dimensional models.

Consider an image being raster scanned from left to right and top to bottom with instantaneous repositioning so that the scanner output is continuous. Let $u(i, x)$ denote the brightness at a point x of the ith scanned line of the image, where $i = 0, \pm 1, \pm 2, \cdots, \pm\infty$, and x lies in the continuous interval $[0, M]$. The scanner output at any instant

$$t = jM + x, \quad 0 \leqslant x \leqslant M \tag{41}$$

is $s(t) = u(j, x)$, where the scanner speed is assumed to be unity. Thus the scanner acts like a stacking operator which stacks one row after another. It is easy to show that the scanner output $s(t)$ is a nonstationary process even though $u(i, x)$ is a two-dimensional stationary random field. Defining

$$\tau = iM + y, \quad 0 \leqslant y \leqslant M, \quad i = 0, 1, 2, \cdots \tag{42}$$

the covariance of the scanner output is given by [39]

$$r_s(x, y) = Es(t)s(t+\tau) = \begin{cases} r(i, y), & x + y \leqslant M. \\ r(i+1, M-y), & x + y > M. \end{cases} \tag{43}$$

Since r_s depends on both x and y, it is a nonstationary covariance function. It could be *approximated* as a stationary covariance by averaging it over $[0, M]$ to yield

$$\tilde{r}(\tau) = \frac{M-y}{M} r(i, y) + \frac{y}{M} r(i+1, M-y). \tag{44}$$

Now $\tilde{r}$ depends only in i and y (the parameters of τ). Thus it is a function of τ only and is therefore stationary. Now a rational approximation of $\tilde{r}(\tau)$ is made from which a state variable model can be determined using conventional spectral factorization techniques in one dimension. For an example see [39].

C. State Variable Model for the Nonstationary Scanning Process s(t) [41], [42]

Instead of approximating the covariance function $r_s(x, y)$ by a stationary covariance model, a time varying state variable model can be determined for the scanning process. Suppose the image has N scan lines and its covariance function is separable, i.e.,

$$r(n, \tau) = r_1(n) r_2(\tau) \tag{45}$$

where r_1, r_2 are stationary covariance functions. Suppose $r_2(\tau)$ has a state variable realization (A, B, C, K) and the $N \times N$ covariance matrix $R_1 = \{r_1(i-j),\ 1 \leqslant i, j \leqslant N\}$ has a lower-upper triangular (LU) factorization $R_1 = HH^T$. Then $r(n, \tau)$ has the realization

$$\begin{aligned} \dot{x} &= Ax + B\epsilon \\ x(kM) &= x_k \\ y &= Cx \\ E\epsilon(t)\epsilon^T(\tau) &= K\delta(t-\tau) \end{aligned}$$

$$s(t) = s_k(t) = \sum_{j=0}^{k} h_{k,j} y(t - (k-j)M),$$

$$kM < t \leqslant (k+1)M. \tag{46}$$

These equations give a nonstationary or the so-called *cyclostationary* representation of $s(t)$. This is because (46) is to be reinitialized after every M time units. Also, because of the delays involved, this representation is non-Markovian.

The derivation of (46) is quite simple, especially in comparison to the elaborate procedure of the earlier development. However, it is time varying, and being non-Markovian its usefulness as such is limited.

D. A Vector Scanning Model

Suppose the image considered above is scanned by a column of N raster scanners and define

$$\hat{s}(t) = [s_0(t) s_1(t) \cdots s_{N-1}(t)]^T$$

$$\hat{x}(t) = [x^T(t), \quad x^T(t-(N-2)M), \cdots, x^T(t-(N-1)M)]^T,$$

$$(N-1)M \leqslant t \leqslant NM.$$

Then the vector scanner $s(t)$ has a state variable *Markovian representation*

$$\begin{aligned} \frac{d\hat{x}}{dt} &= \mathfrak{A}\hat{x} + \mathfrak{B}\hat{\epsilon}(t) \\ E\hat{x}_0^T \hat{x}_0 &\triangleq P_0 = R_1 \\ E\hat{\epsilon}(t)\hat{\epsilon}^T(\tau) &= (I \otimes K)\delta(t-\tau), \quad (N-1)M \leqslant t, \quad \tau \leqslant NM \\ \hat{s}(t) &= (H \otimes C)\hat{x} \end{aligned} \tag{47}$$

where $\mathfrak{A} = I \otimes A$, $\mathfrak{B} = I \otimes B$, and $\otimes$ denotes the Kronecker product. Now we have a Markov model, but its dimensional has increased.

State variable models have been found useful in restoration of images degraded by spatially varying PSF's where Fourier techniques are not applicable and particularly when the PSF can be modeled as a finite impulse response and/or the degradation is a causal process (e.g., motion blur). For details see [45].

E. Noncausal Models [18], [46]

Earlier we saw that a causal AR representation is of the type

$$u_k = \bar{u}_k + \epsilon_k, \quad E\epsilon_k = 0$$

where $\bar{u}_k$ is the best linear mean square predictor of u_k based on the past values $\{u_l, l < k\}$ and $\{\epsilon_k\}$ is a white-noise sequence. Thus $\bar{u}_k$ is a minimum variance causal predictor of u_k. In an analogous fashion, we can define a minimum variance noncausal predictor $\bar{u}_k$ which depends on the past as well as the future values of u_k. Let $\{u_k\}$ be any zero-mean Gaussian random sequence and let $\bar{u}_k$ denote the best linear mean-square estimate of u_k based on all $\{u_l, l \neq k\}$. Writing

$$\bar{u}_k \triangleq \sum_{l \neq k} a_{k,l} u_l \tag{48}$$

we determine coefficients $a_{k,l}$ by minimizing the mean-square error $E[(u_k - \bar{u}_k)^2]$. This minimization gives the result as

$$r_{k,l} - \sum_{j \neq k} a_{k,j} r_{j,l} = \beta_k^2 \delta_{k,l}, \quad \beta_k^2 = \min E(u_k - \bar{u}_k)^2. \tag{49}$$

In matrix form after defining $a_{k,k} = -1$, this becomes

$$-AR = B \quad \text{or} \quad -A = BR^{-1} \tag{50}$$

where B is a diagonal matrix of elements β_k^2 and R is assumed to be positive definite. The noncausal *minimum variance representation* (MVR) of the (scalar) nonstationary random process $\{u_k\}$ is now defined as

$$u_k - \sum_{l \neq k} a_{k,l} u_l = \nu_k \quad \text{or} \quad -Au = \mathbf{v} \tag{51}$$

where u and $\mathbf{v}$ are vectors of elements $\{u_k\}$ and $\{\nu_k\}$ respectively and $\mathbf{v}$ is a random process that represents the noncausal prediction error. Using (50) and (51) we get

$$R^{-1}u = B^{-1}\mathbf{v}. \tag{52}$$

It is interesting that the minimum variance noncausal representation of (52) does not require any spectral factorization, but only needs the inversion of the covariance matrix R. Note that the random process $\{\nu_k\}$ is not a white-noise sequence. Instead, it is a "colored"-noise sequence whose co-

variance is obtained via (52) to give

$$R_v \triangleq E\mathbf{v}\mathbf{v}^T = BR^{-1}B. \tag{53}$$

The elements of the diagonal matrix B are determined by noting that $B \equiv \text{diag}\,[R_v]$ and are given by

$$\beta_k^2 = \frac{1}{[R^{-1}]_{k,k}}. \tag{54}$$

Hence, the noncausal representation (51) is complete by specifying R^{-1}. For an infinite stationary process with analytic SDF $S(\omega)$, the noncausal MVR becomes

$$r_0^+ u_k + \sum_{\substack{l=-\infty \\ l\neq 0}}^{\infty} r_l^+ u_{k-l} = \beta^{-2} v_k \tag{55}$$

where

$$r_k^+ = \frac{1}{2\pi}\int_{-\pi}^{\pi} S^{-1}(\omega)\exp(jk\omega)\,d\omega$$

is the covariance sequence corresponding to $1/S(\omega)$ and

$$\beta^2 = 1/r_0^+. \tag{56}$$

Example 4: A special case of interest is to find the noncausal MVR for causal AR processes. From (55), it follows that for the pth-order AR model of (21), the noncausal MVR is given by

$$u_k - \sum_{n=1}^{p} h_n(u_{k-n} + u_{k+n}) = v_k, \qquad \forall k$$

$$h_n = \frac{-\sum_{k=0}^{p} a_k a_{n+k}}{\sum_{k=0}^{p} a_k^2}, \qquad a_0 \triangleq -1$$

$$r_v(k) \triangleq Ev_j v_{j+k} = \beta_v^2\left[\delta_{k,0} - \sum_{n=1}^{p} h_n(\delta_{k,n} + \delta_{k,-n})\right]$$

$$\beta_v^2 = \frac{\beta^2}{\sum_{k=1}^{p} a_k^2}. \tag{57}$$

This is a $2p$th-order stochastic difference equation forced by a pth-order moving average process $\{v_k\}$.

If we define $N \times 1$ vectors $u = \{u_k\}$, $\mathbf{v} = \{v_k\}$, etc., (57) can be written as

$$H\mathbf{u} = \mathbf{v} + \mathbf{b} \qquad E\mathbf{v}\mathbf{b}^T = 0$$

$$R_v \triangleq E\mathbf{v}\mathbf{v}^T = \beta_v^2 H \tag{58}$$

where H is the $N \times N$ symmetric banded Toeplitz matrix whose entries along the mth subdiagonal are $-h_m$, and the $N \times 1$ vector $\mathbf{b}$ contains $2p$ nonzero terms involving the boundary variables $\{u_{1-k}, u_{N+k}, 1 \leq k \leq p\}$. Defining

$$\mathbf{u}^0 = H^{-1}\mathbf{v} \qquad \mathbf{u}^b = H^{-1}\mathbf{b} \tag{59}$$

we obtain an orthogonal decomposition of $\mathbf{u}$

$$\mathbf{u} = \mathbf{u}^0 + \mathbf{u}^b \qquad E[\mathbf{u}^0(\mathbf{u}^b)^T] = O \tag{60}$$

which is such that the covariance matrix of $\mathbf{u}^0$ is

$$R^0 \triangleq E\mathbf{u}^0\mathbf{u}^{0T} = \beta_v^2 H^{-1} \tag{61}$$

and the process $\mathbf{u}^b$ is completely determined by the $2p$ boundary variables $\{u_{1-k}, u_{N+k}, 1 \leq k \leq p\}$. This representation provides the following result which is useful in developing the so-called fast KLT algorithms [18].

Theorem 3: Let $\{u_k\}$ be a stationary, pth-order AR sequence. For any $N > 1$ if the $2p$ boundary values $\{u_{1-k}, u_{N+k}, 1 \leq k \leq p\}$ are given then the KLT of the sequence $\{u_k, 1 \leq k \leq N\}$ conditioned on these boundary values is the orthonormal set of eigenvectors of the $N \times N$ banded Toeplitz matrix H.

This theorem could be proven following [18] where the special case $p = 1$ was considered. It has been shown [10] that for a given H, a fast sinusoidal transform could be found as a good approximation to its KLT. For the case $p = 1$, the KLT is the fast sine transform (see Table III). Noncausal representations are useful in data compression and restoration of images using image transforms [46]–[53]. The boundary variables are processed first (e.g., filtered in image estimation problems) and $\tilde{u}^b$, an estimate of u^b is determined. The residual process $\tilde{u}^0 = u - u^b$ is then processed by the KLT of u^0. For example it has been shown for $p = 1$ that the noncausal representation can be used for block by block coding of a sequence in such a way that interblock redundancy could be exploited (via the boundary values) leading to an algorithm more efficient than the conventional block by block KLT coding where one does not consider the interblock effects. For details see [9], [17], [19]. Algorithms in other applications [93] will require manipulation of the banded Toeplitz matrix and certain circulant [54] and other [10], [49] decompositions have been found useful. In the sequel we will find the above theory useful in developing semicausal and noncausal representations for two-dimensional images.

IV. Linear Prediction Models in Two Dimensions

One important property of many one-dimensional systems is that of causality. For two-dimensional images, causality is not inherent in the data. Moreover, the data could be such that a causal realization by a finite-order linear system is not possible, even if the SDF is a rational function. This is because it is generally not possible to factorize a two-dimensional polynomial as a product of lower order polynomials. In general, one can think of causal, semicausal, and noncausal representations for two-dimensional images. These representations are the discrete equivalent of the classical categories, viz., initial-value (or hyperbolic), initial-boundary value (or parabolic) and boundary value (or elliptic), of two-dimensional linear systems characterized by partial differential equations. In this section we define such models in the framework of linear prediction. In the next section we will consider their realization from a given SDF.

Linear prediction models in two dimensions are useful in image data transmission and storage via DPCM coding and hybrid coding, design of recursive, semirecursive and nonrecursive filters for image estimation, restoration and filtering and in image analysis. Examples are considered in Section VI.

Let $\{u_{i,j}\}$ be an arbitrary zero-mean Gaussian random field and let $\bar{u}_{i,j}$ denote a prediction estimate of the random variable $u_{i,j}$.

A. Causal Prediction

Suppose the samples of the random field $\{u_{i,j}\}$ are arranged in any desired, one-dimensional ordered sequence $\{u(k)\}$. Then $\bar{u}_{i,j}$ is defined as a *causal prediction* of $u_{i,j}$ if it depends only

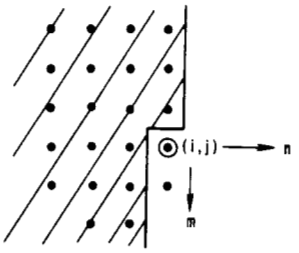


Fig. 5. Prediction region for causal models.

on the elements that occur before the element $u_{i,j}$. A common example occurs when an image is raster scanned, say, column by column, and $\bar{u}_{i,j}$ is a linear estimate based on all the elements scanned before arriving at (i, j), i.e.,

$$\bar{u}_{i,j} = \sum_{m,n \in \delta} a(i, j; m, n)\, u_{m,n},$$

$$\delta = \{m, n: n < j, \forall m\} \bigcup \{m, n: n = j, m < i\}. \quad (62)$$

Fig. 5 shows the set δ for causal prediction at (i, j). This definition of causality includes, as a special case, single quadrant causal predictors of the type

$$\bar{u}_{i,j} = \sum_{\substack{m=0 \\ (m,n) \neq (0,0)}}^{\infty} \sum_{n=0}^{\infty} a(i, j; m, n)\, u_{i-m, j-n}. \quad (63)$$

If u_j denotes the jth column vector of the image, (62) gives

$$\bar{u}_j = A^0_{j,j} u_j + \sum_{n<j} A_{j,n} u_n \quad (64)$$

where $A^0_{j,j}$ is a lower triangular matrix whose diagonal entries are zero.

B. Semicausal Prediction[2]

If the estimate $\bar{u}_{i,j}$ is causal in one of the coordinates and noncausal in the other, it is called a *semicausal predictor*. For example, a linear semicausal predictor which is causal in "j" and noncausal in "i" would be of the form

$$\bar{u}_{i,j} = \sum_{m,n \in \delta} a(i, j; m, n)\, u_{m,n},$$

$$\delta = \{m, n: n < j, \forall m\} \bigcup \{m, n: n = j, \forall m \neq i\} \quad (65)$$

where δ is shown in Fig. 6. In vector notation this becomes

$$\bar{u}_j = A^0_{j,j} u_j + \sum_{n<j} A_{j,n} u_n \quad (66)$$

where $\{A_{j,n}\}$ are matrices of elements $\{a(i, j; m, n)\ \forall i, m\}$ and $A^0_{j,j}$ has zeros at its diagonal elements.

C. Noncausal Prediction

The quantity $\bar{u}_{i,j}$ is defined as a noncausal prediction of $u_{i,j}$ if it can be written as a linear combination of possibly all the variables in the random field, except $u_{i,j}$ itself. For example, a *linear noncausal predictor* would be of the type

$$\bar{u}_{i,j} = \sum_{m,n \in \delta} a(i, j; m, n)\, u_{m,n},$$

$$\delta = \{m, n; (m, n) \neq (i, j)\} \quad (67)$$

and is shown in Fig. 7. Note that $\bar{u}_{i,j}$ contains terms from all the four quadrants about the point (i, j).

Example 4: The following are examples of causal, semicausal, and noncausal predictors.

Causal:

$$\bar{u}_{i,j} = a_1(i, j)\, u_{i-1,j} + a_2(i, j)\, u_{i,j-1} + a_3(i, j)\, u_{i-1,j-1}.$$

Semicausal:

$$\bar{u}_{i,j} = a_1(i, j)\, u_{i-1,j} + a_2(i, j)\, u_{i+1,j} + a_3(i, j)\, u_{i,j-1}.$$

Noncausal:

$$\bar{u}_{i,j} = a_1(i, j)\, u_{i-1,j} + a_2(i, j)\, u_{i+1,j} + a_3(i, j)\, u_{i,j-1} + a_4(i, j)\, u_{i,j+1}.$$

D. Minimum Variance Prediction

A minimum variance prediction estimate is one that minimizes the mean-square error

$$e_{i,j} = E(u_{i,j} - \bar{u}_{i,j})^2 \quad (68)$$

at each i, j. For a Gaussian random field, the minimum variance predictors would be linear, whose coefficients $a(\cdot\,\cdot;\cdot\,\cdot)$ could be determined from its covariance function. Minimization of (68) and use of (62), (65), or (67) yields the orthogonality relation

$$E\left[u_{i,j} - \sum_{m,n \in \delta} a(i, j; m, n)\, u_{m,n}\right] u_{p,q} = \beta^2_{i,j} \delta_{i,p}\, \delta_{j,q},$$

$$p, q \in \delta_0 \qquad \delta_0 \triangleq \delta \bigcup [i, j] \quad (69)$$

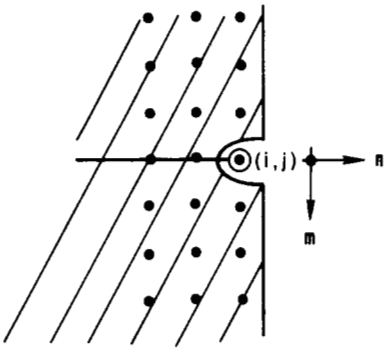


Fig. 6. Prediction region for semicausal models.

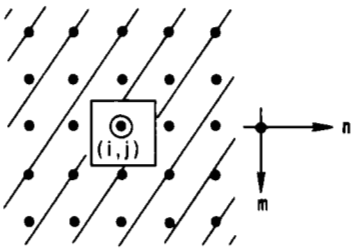


Fig. 7. Prediction region for noncausal models.

[2]We note that in the two-dimensional literature the term 'causal' is often used for single quadrant models only and the model of (62) is sometimes called 'semicausal' or nonsymmetric half-plane (NSHP) model. Our definition of semicausality includes all of the half-plane as indicated in (65).

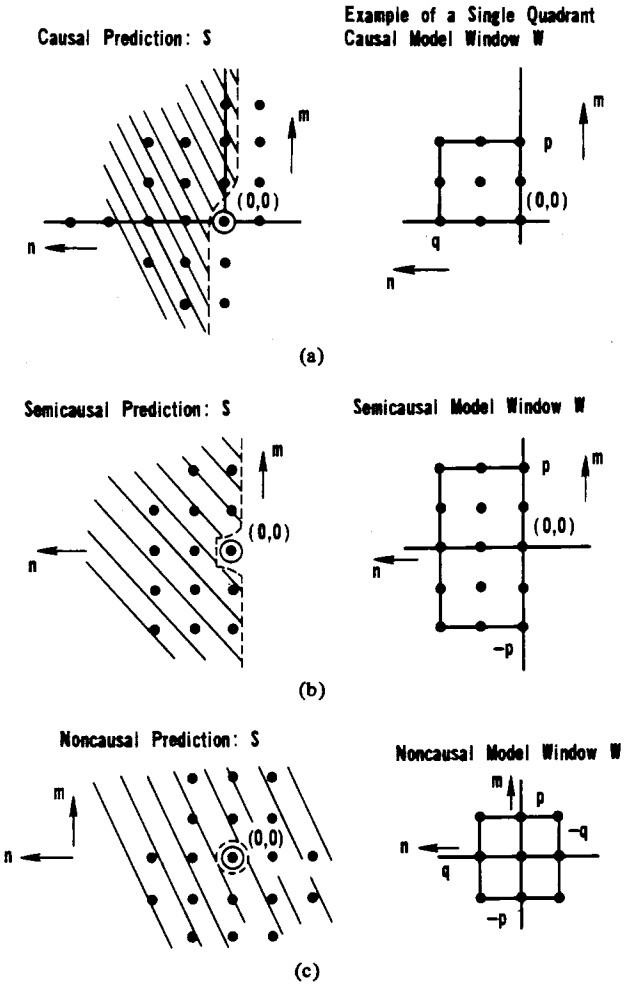


Fig. 8. Prediction regions and windows for stationary models.

Written in terms of covariances, we obtain

$$r(i,p;j,q) - \sum_{m,n\in\delta} a(i,m;j,n)\, r(m,p;n,q) = \beta_{i,j}^2 \delta_{i,p}\delta_{j,q}, \qquad p,q \in \delta_0 \quad (70)$$

where δ depends on whether $\bar{u}_{i,j}$ is causal, semicausal, or noncausal, and $\beta_{i,j}^2$ is the minimized value of $e_{i,j}$. A solution of the simultaneous equations (70) gives the unknowns $a(\cdot\,\cdot;\cdot\,\cdot)$ and $\beta_{i,j}^2$.

E. Stochastic Representation of Gaussian Fields

Let $\bar{u}_{i,j}$ be an arbitrary prediction of $u_{i,j}$. Then we define a stochastic representation of the random field $\{u_{i,j}\}$ as

$$u_{i,j} = \bar{u}_{i,j} + \epsilon_{i,j} \quad (71)$$

where $\{\epsilon_{i,j}\}$ is another random field such that the given covariance properties of $\{u_{i,j}\}$ are satisfied. There are three types of representations that we would be interested in considering here. These are as follows:

(i) minimum variance representations (MVR)
(ii) white-noise-driven representations (WNDR)
(iii) autoregressive moving average (ARMA) representations.

For minimum variance representations, $\bar{u}_{i,j}$ is chosen to be a minimum variance predictor. These representations could be causal, semicausal, or noncausal in structure. One basic difference among these representations is their spatial structure which leads to different types of processing algorithms. For WNDR $\{\epsilon_{i,j}\}$ is chosen to be a white-noise field resulting in several types of models including the Karhunen–Loeve representations. In ARMA representations, $\{\epsilon_{i,j}\}$ is a colored noise field with a truncated covariance function, i.e.,

$$E\epsilon_{i,j}\epsilon_{m,n} = 0, \qquad \forall |i-m| > K, \qquad |j-n| > L \quad (72a)$$

for some fixed integers $K > 0$, $L > 0$. For example the function

$$r_\epsilon(i,j;m,n) = \begin{cases} 1, & i=m,\ j=n \\ \alpha, & |i-m|=1,\ |j-n|=1 \\ 0, & \text{otherwise} \end{cases} \quad (72b)$$

represents a stationary moving average field.

F. Stationary Models

For stationary random fields, the covariances become a function of two variables, i.e.,

$$r(i,j;m,n) = r(i-m, j-n)$$

and the various predictors become spatially invariant (or shift invariant) yielding the representation

$$u_{i,j} = \sum_{m,n\in\hat{\delta}} \sum a_{m,n} u_{i-m,j-n} + \epsilon_{i,j} \quad (73)$$

where $\hat{\delta}$ is a window of index pairs (m, n) which is now independent of i, j and is defined as

$$\hat{\delta} = \begin{cases} \{n \geq 1, \forall m\} \bigcup \{n = 0, m \geq 1\}, & \text{for causal models} \\ \{n \geq 1, \forall m\} \bigcup \{n = 0, \forall m \neq 0\}, & \text{for semicausal models} \\ \{\forall (m,n) \neq (0,0)\}, & \text{for noncausal models.} \end{cases} \quad (74)$$

Note that $\hat{\delta}$ does not contain the origin (0, 0). Fig. 8 shows $\hat{\delta}$ for the various cases above. Often one is only interested in representations where $a_{m,n}$ are nonzero only over a finite window W, (see Fig. 8 for examples) called the *prediction window*, which is a subset of $\hat{\delta}$, so that

$$u_{i,j} = \sum_{m,n\in W} \sum a_{m,n} u_{i-m,j-n} + \epsilon_{i,j}. \quad (75)$$

In that event (75) becomes a constant coefficient stochastic difference equation with a rational transfer function

$$H(z_1^{-1}, z_2^{-1}) = \frac{1}{A(z_1^{-1}, z_2^{-1})} \triangleq \left[1 - \sum_{m,n\in W} \sum a_{m,n} z_1^{-m} z_2^{-n}\right]^{-1}. \quad (76)$$

The SDF of random fields represented by these stationary

models becomes

$$S_u(z_1, z_2) = \frac{S_\epsilon(z_1, z_2)}{\left[1 - \sum\sum_{m,n \in W} a_{m,n} z_1^{-m} z_2^{-n}\right]\left[1 - \sum\sum_{m,n \in W} a_{m,n} z_1^{m} z_2^{n}\right]} \quad (77)$$

where $S_\epsilon(z_1, z_2)$ is the SDF of $\{\epsilon_{i,j}\}$. In general, $\{\epsilon_{i,j}\}$ would be a moving average field so that (77) is a rational function expressed as a ratio of two-dimensional polynomials in z_1 and z_2.

Minimum Variance Models: If (75) is a minimum variance representation of a stationary field, then the orthogonality condition (69) becomes

$$r(k, l) - \sum\sum_{m,n \in W} a_{m,n} r(k - m, l - n) = \beta^2 \delta_{k,0} \delta_{l,0}, \qquad k, l \in \hat{\mathcal{S}}_0 \quad \hat{\mathcal{S}}_0 = \hat{\mathcal{S}} \bigcup (0, 0). \quad (78)$$

Defining

$$a_{0,0} \triangleq -1 \qquad W_0 = W \bigcup (0, 0) \quad (79)$$

(78) reduces to

$$-\sum\sum_{m,n \in W_0} a_{m,n} r(k - m, l - n) = \beta^2 \delta_{k,0} \delta_{l,0}, \qquad (k, l) \in \hat{\mathcal{S}}_0. \quad (80)$$

Mapping the arrays $\{a_{m,n}\}$, $\{u_{m,n}\}$, $\{\delta_{k,0}\delta_{l,0}\}$ with support on W_0 into vectors $\mathbf{a}$, $\mathbf{u}$, $\mathbf{l}$, respectively, by column ordering, we obtain

$$\mathcal{R}\mathbf{a} = -\beta^2 \mathbf{l} \quad (81)$$

where $\mathcal{R}$ is the covariance matrix of the vector $\mathbf{u}$. For example, if W_0 is the noncausal rectangular window $[-p, p] \times [-q, q]$ then $\mathbf{a}$ is of size $(2p + 1)(2q + 1)$ and $\mathcal{R}$ is a $(2p + 1) \times (2q + 1)$ doubly block Toeplitz matrix with basic dimension $(2p + 1) \times (2p + 1)$. The unit vector $\mathbf{l}$ takes the value 1 at the location i_0, which corresponds to the (0, 0) location in the window W_0. Equations (79) and (81) can be solved to give

$$\begin{aligned} \mathbf{a} &= -\beta^2 \mathcal{R}^{-1} \mathbf{l} \\ &\triangleq -\beta^2 \mathcal{B}_{i_0} \\ \beta^2 &= 1/\mathcal{B}_{i_0}(i_0) \\ &= 1/[\mathcal{R}^{-1}]_{i_0, i_0}. \end{aligned} \quad (82)$$

These equations are much easier to solve than (70) because the dimensionality of $\mathcal{R}$ equals the square of the number of points in the window W_0 rather than the square of the number of points in the image. For a given positive definite $\mathcal{R}$, a unique solution of (81) is obtained. However, for a given admissible set of predictor coefficients $a_{m,n}$, there is not a unique $\mathcal{R}$ which will yield these coefficients as the solution of (81). Moreover, for an arbitrary positive definite $\mathcal{R}$, the solution of (81) does not assure prediction coefficients which would yield stable models. In spite of these shortcomings, equation (81) is useful because i) it is a linear (Toeplitz block) system of equations and ii) it could be used in finding approximate causal, semicausal and noncausal MVR's which are stable.

G. White-Noise-Driven Representations (WNDR)

Consider an arbitrary matrix operator $\mathcal{A}$ and the linear representation

$$\mathcal{A}\mathbf{u} = \mathbf{e} \quad (83)$$

where $\mathbf{u}$ and $\mathbf{e}$ are vectors corresponding to the fields $\{u_{i,j}\}$ and $\{\epsilon_{i,j}\}$, respectively. When $\{\epsilon_{i,j}\}$ is required to be a white-noise field, we must have

$$\mathcal{A}\mathcal{R}\mathcal{A}^T = \mathcal{B} = \text{Diagonal} \qquad \mathcal{R} = E\mathbf{u}\mathbf{u}^T, \quad \mathcal{B} = E\mathbf{e}\mathbf{e}^T. \quad (84)$$

Thus, any white noise driven linear representation must satisfy the factorization equation (84). However, it does not have a unique solution $\mathcal{A}$ and therefore many different representations are possible. If $\mathcal{A}$ is a lower triangular matrix, then (83) is a causal representation. As we shall see, the causal MV representations are of this type. When $\mathcal{A}$ is a block lower triangular matrix, then (83) becomes a semicausal WNDR (but not necessarily MVR). Otherwise, it would be a noncausal WNDR. Since $\mathcal{R}$ is a covariance matrix, another solution of (84) is obtained when $\mathcal{A}$ is a unitary matrix containing the eigenvectors of $\mathcal{R}$ and $\mathcal{B}$ is the diagonal matrix of its eigenvalues. This would yield the KL representation of the random field. The choice of the operator $\mathcal{A}$ for WNDR's depends on the algorithmic considerations or on a given physical situation.

V. Realization of the Two-Dimensional Prediction Models

Now we consider the problem of identifying the foregoing three types of representations given the covariance function, or equivalently, (for the stationary case) the spectral density function of the image. When the desired representation is required to be causal and stable, the above problem is also called the *spectral factorization problem*. Let $S(z_1, z_2)$ represent the two-dimensional z-transform of a covariance sequence $r(m, n)$. When $z_1 = \exp(j\omega_1)$, $z_2 = \exp(j\omega_2)$, this becomes the SDF and will also be written as $S(\omega_1, \omega_2)$. The main result of this section is that it is possible to find finite order, stable causal, semicausal, and noncausal MVR's which would match a given well behaved SDF arbitrarily closely. Details are given in the Appendix.

A. Separable Models

When the covariance function is separable, e.g., in the stationary case, if $r(k - m, l - n) = r_1(k - m)\, r_2(l - n)$, the solution of (80) reduces to two independent one-dimensional models. For example, if $r_1(m)$ and $r_2(n)$ have the one-dimensional realizations

$$\sum_{m \in W_1} a_1(m) x(k - m) = e_1(k)$$

$$\sum_{m \in W_2} a_2(n) y(l - n) = e_2(l).$$

Then $r_1(m)\, r_2(n)$ has the realization

$$\sum_{m \in W_1, n \in W_2} \sum a_1(m)\, a_2(n)\, u_{k-m, l-n} = \epsilon_{k,l}$$

$$r_\epsilon(m, n) = r_{e_1}(m)\, r_{e_2}(n). \quad (85)$$

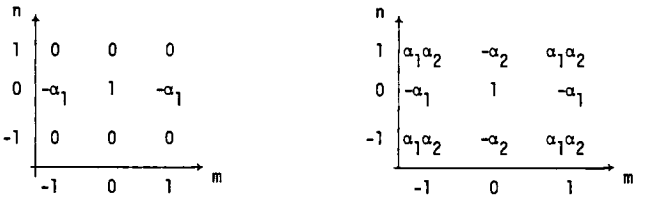


Fig. 9. Correlation arrays $\rho_\epsilon(m, n)$ for semicausal and noncausal model prediction errors.

Example 5: Consider the separable covariance model shown in Table IV. Using the above result the minimum variance causal, semicausal, and noncausal models are obtained as follows [51], [52].

Causal model (C1):

$$u_{i,j} = \rho_1 u_{i-1,j} + \rho_2 u_{i,j-1} - \rho_1 \rho_2 u_{i-1,j-1} + \epsilon_{i,j}$$
$$r_\epsilon(m, n) = \sigma^2(1 - \rho_1^2)(1 - \rho_2^2)\,\delta_{m,0}\,\delta_{n,0}. \tag{86}$$

Semicausal MVR (SC2):

$$u_{i,j} = \alpha_1(u_{i-1,j} + u_{i+1,j}) + \rho_2 u_{i,j-1} - \rho_2 \alpha_1(u_{i-1,j-1} + u_{i+1,j+1}) + \epsilon_{i,j}$$
$$r_\epsilon(m, n) = \sigma^2 \frac{(1 - \rho_1^2)(1 - \rho_2^2)}{(1 + \rho_2^2)} \rho_\epsilon(m, n)$$
$$\alpha_1 \triangleq \frac{\rho_1}{(1 + \rho_1^2)}. \tag{87}$$

Noncausal MVR (NC3):

$$u_{i,j} = \alpha_1(u_{i-1,j} + u_{i+1,j}) + \alpha_2(u_{i,j-1} + u_{i,j+1}) - \alpha_1\alpha_2(u_{i-1,j-1} + u_{i+1,j-1} + u_{i-1,j+1} + u_{i+1,j+1}) + \epsilon_{i,j}$$
$$r_\epsilon(m, n) = \sigma^2 \frac{(1 - \rho_1^2)(1 - \rho_2^2)}{(1 + \rho_1^2)(1 + \rho_2^2)} \cdot \rho_\epsilon(m, n). \tag{88}$$

Fig. 9 shows the correlation arrays $\rho_\epsilon(m, n)$ for the above two models.

Remarks:

1) The prediction window is necessarily rectangular for separable models. Thus causal MVR's are necessarily quarter-plane models for separable covariance functions although the converse is not true.

2) All the three models in the above example represent the same stationary random field. The spatial structural differences yield different types of algorithms. For example, it has been shown [9], [17], that the causal, semicausal and the noncausal models yield naturally the predictive, hybrid and transform coding algorithms for data compression of images. For other applications see [1], [52], [53], [55].

3) While the causal MVR's are white noise driven, the semicausal and noncausal MVR's are not.

4) If the given SDF $S(\omega_1, \omega_2)$ is not separable, then its best least squares separable approximation can be found via its singular value decomposition as

$$S(\omega_1, \omega_2) \simeq \hat{S}(\omega_1, \omega_2) \triangleq \lambda S_1(\omega_1) S_2(\omega_2)$$

where $S_1(\omega_1)$ and $S_2(\omega_2)$ are the solutions corresponding to the largest eigenvalue λ of the following eigenvalue equations

$$\int_{-\pi}^{\pi} K_1(\omega_1, y) S_1(y)\, dy = \lambda S_1(\omega_1), \qquad \int_{-\pi}^{\pi} S_1^2(\omega)\, d\omega = 1$$
$$\int_{-\pi}^{\pi} K_2(x, \omega_2) S_2(x)\, dx = \lambda S_2(\omega_2), \qquad \int_{-\pi}^{\pi} S_2^2(\omega)\, d\omega = 1$$

where

$$K_1(x, y) = \int_{-\pi}^{\pi} S(x, y') S(y, y')\, dy'$$
$$K_2(x, y) = \int_{-\pi}^{\pi} S(x', x) S(x', y)\, dx'.$$

If $S(\omega_1, \omega_2) > 0$, then it is easy to show that $\lambda > 0$, $S_1(\omega_1) > 0$, $S_2(\omega_2) > 0$ and S_1, S_2 would be valid one-dimensional SDF's if S is a valid two-dimensional SDF. Now $\hat{S}$ is a separable SDF and therefore an appropriate causal, semicausal or noncausal realization of $\hat{S}$ could be found. A similar method of approximating a SDF (or equivalently, the magnitude of the frequency response of a filter) by a positive separable function has been used in [27] for the design of separable two-dimensional filters when the given SDF is discrete. In that case, the above equations reduce to the matrix SVD relations of Section II-C.

B. Causal Minimum Variance Representations

It can be shown that the causal MVR's are also white-noise driven representations. In general identification of a causal MVR requires a two-stage factorization. For example if the SDF is a rational function of the form

$$S(z_1, z_2) = \frac{C_0}{\left[\sum_{n=-q}^{q} \sum_{m=-p_1}^{p_1} \alpha_{m,n} z_1^{-m} z_2^{-n}\right]} \tag{89}$$

where C_0 is a constant, then, the first-stage factorization is achieved by solving a set of normal equations (see (A12) in Appendix) parametric in z_1 and the result is

$$S(z_1, z_2) = \frac{\sum_{n=0}^{q} \hat{a}_n^q(z_1)\, r_n(z_1)}{\left[1 + \sum_{n=1}^{q} \hat{a}_n^q(z_1) z_2^{-n}\right]\left[1 + \sum_{n=1}^{q} \hat{a}_n^q(z_1^{-1}) z_2^{n}\right]}. \tag{90}$$

Then a second stage factorization of S (equation (A13)) gives

$$S(z_1, z_2) = \frac{\beta^2}{\left[\sum_{n=0}^{q} a_n^q(z_1) z_2^{-n}\right]\left[\sum_{n=0}^{q} a_n^q(z_1^{-1}) z_2^{n}\right]}. \tag{91}$$

In general $a_n^q(z_1)$ will be irrational functions expressed by their infinite Laurent series. Under certain conditions of positivity and analyticity of the SDF the causal model so determined would be stable and causally invertible (i.e., "minimum-phase"). However, one is confronted with the problem of solving an infinity of sets of q simultaneous equations ((A12)

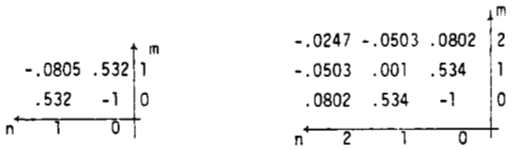


Fig. 10. Single quadrant causal MVR coefficients.

and (A13)) for the rational SDF of (89). An *approximate* rational factorization of S could be obtained by finding a rational approximation for $a_n^q(z_1)$. For example, let

$$a_n^q(z_1) \simeq \sum_{m=-p}^{p} a_{m,n} z_1^{-m}, \qquad n \geqslant 1$$

$$a_0^q(z_1) \simeq \sum_{m=0}^{p} a_{m,0} z_1^{-m}.$$

Then with the analyticity constraint on S, it is possible to find a suitable integer $p < \infty$, such that $\tilde{S}$, the rational factorized form obtained by replacing $a_n^q(z_1)$ by their approximations above would be of the form given in (77) and would be arbitrarily close to S. The realization of $\tilde{S}$ would now be of the form of (75). Thus it is possible to find finite-order causal MVR's which would come arbitrarily close to the infinite order causal MVR's of a given SDF. It can also be shown that the solution of (82) for a causal window W_0 and for suitably large p and q should yield a causal MVR which would realize the given SDF arbitrarily closely. The proof of this and related considerations such as stability will be considered elsewhere. Another approach of rationalizing (91) and retaining stability is via a reflection coefficient design method proposed by Marzetta [56].

Example 6: Consider the nonseparable exponential covariance model in Table IV with $\alpha_1 = \alpha_2 = 0.05$. Single quadrant causal MVR coefficients $a_{m,n}$ over $[0, 1] \times [0, 1]$ and $[0, 2] \times [0, 2]$ regions are obtained as shown in Fig. 10.

Fig. 11 compares the given covariance array with the one generated by the model on the $[4 \times 4]$ grid for the $[2 \times 2]$ model. Although both the models above turn out to be stable, the given and the model covariances do not match exactly. However, as the model order is increased, e.g., from $[1 \times 1]$ to $[2 \times 2]$ the covariance match has been found to improve.

$$r(k,\ell) = \exp\left\{-.05\sqrt{k^2+\ell^2}\right\}$$

.868	.894	.905	.894	.868
.894	.932	.951	.932	.894
.905	.951	1.000	.951	.905
.894	.932	.951	.932	.894
.868	.894	.905	.894	.868

ACTUAL COVARIANCES ON W_c

.696	.730	.697	.579	.479
.730	.819	.840	.701	.579
.697	.840	1.000	.840	.697
.579	.701	.840	.819	.730
.479	.579	.697	.730	.696

ESTIMATED COVARIANCES BY p=q=2 CAUSAL MODEL

.821	.853	.866	.853	.821
.858	.907	.932	.907	.858
.874	.934	1.000	.934	.874
.858	.907	.932	.907	.858
.821	.853	.866	.853	.821

ESTIMATED COVARIANCES BY p=q=2 SEMICAUSAL MODEL

Fig. 11. Covariances estimated by the causal and semicausal models.

C. Semicausal Minimum Variance Representations

For these models only a single stage factorization of the SDF is required. Unlike causal MVR's, semicausal MVR's are not white-noise-driven models. However, for a given semicausal MVR, a causal MVR can always be found under some mild restrictions [see Appendix]. But the causal realization may have a higher (even infinite) order and/or may become spatially varying as shown by the following example.

Example 7: Consider a spatially invariant, semicausal MVR which is noncausal in the "i" variable and causal in the "j" variable

$$u_{i,j} = a(u_{i-1,j} + u_{i+1,j}) + bu_{i,j-1} + \epsilon_{i,j}.$$

Let $\{u_{i,j}\}$ be a random field with $1 \leqslant i, j \leqslant N$, $u_{0,j} = u_{N+1,j} = 0$, $0 < a < \frac{1}{2}$, $|2a + b| < 1$. In vector notation this becomes

$$Qu_j = bu_{j-1} + \epsilon_j \tag{92}$$

where Q is a symmetric tridiagonal Toeplitz matrix whose diagonal elements are unity, and the subdiagonal elements equal $-a$. For this to be a minimum variance representation, the orthogonality conditions require (see (A14))

$$E\epsilon_j \epsilon_k^T = \beta^2 Q \delta_{j,k}$$

where β^2, the variance of $\epsilon_{i,j}$ is assumed to be constant. The matrix Q has an lower-upper factorization $Q = L^T \Gamma^{-1} L$, where

$$L = \begin{bmatrix} 1 & & & & \\ -r_2 & 1 & & & \\ & -r_3 & 1 & & \\ & & \ddots & \ddots & \\ & & & & -r_N \quad 1 \end{bmatrix}$$

$$\Gamma^{-1} = \text{Diag}\left\{\frac{a}{r_1}, \frac{a}{r_2}, \cdots, \frac{a}{r_N}\right\}$$

and $\{r_k\}$ are given by the backward recursion

$$r_k = \frac{a}{1 - ar_{k+1}}, \qquad r_{N+1} = 0, \qquad 1 \leqslant k \leqslant N.$$

Following Appendix A.2, we get a causal MVR as

$$u_{i,j} = r_i u_{i-1,j} + \frac{br_i}{a} \sum_{k=0}^{N-i+1} \alpha_{i,k} u_{i+k,j-1} + v_{i,j}$$

where $\alpha_{i,k}$ is the element in the kth upper diagonal and the ith row of $(L^T)^{-1}$. Now the sparse structure of the semicausal representation has been lost by this causal representation. Also, it is no longer a constant coefficient model. In the steady state, i.e., as i and $N \to \infty$, the asymptotic values of r and $\alpha_{i,k}$ are given by

$$r \triangleq r_\infty = \frac{1 - \sqrt{1 - 4a^2}}{2a}, \qquad \alpha_{i,k} = r^k, \qquad 0 < r < 1$$

and we obtain a constant coefficient, infinite order, causal MVR

$$u_{i,j} = r u_{i-1,j} + \frac{b}{a} \sum_{k=0}^{\infty} r^{k-1} u_{i+k,j-1} + v_{i,j}$$

$$E v_{i,j}^2 = \beta^2 (1 + r^2).$$

In a practical situation, one would truncate the summation to a finite number of terms to obtain an approximate, finite-order causal model. However, the number of terms that may have to be retained to achieve stability could well be quite large resulting in a high-order model.

For a rational SDF as in (89), one needs to solve (A25) for the given value of q from which a_n^q are determined. This gives the first stage of factorization in the z_2 variable and we obtain

$$S = \frac{-\beta^2 a_0^q(z_1)}{\left[\sum_{n=0}^{q} a_n^q(z_1) z_2^{-n}\right]\left[\sum_{n=0}^{q} a_n^q(z_1^{-1}) z_2^{n}\right]}. \tag{93}$$

No second stage factorization is required for semicausal MVR's. As in the case of causal models $\{a_n^q(z_1)\}$ will be generally irrational but analytic functions in the neighborhood of $|z_1| = 1$. When approximated by suitable rational functions, e.g., letting

$$a_n^q(z_1) \simeq \sum_{m=-p}^{p} a_{m,n} z_1^{-m}, \qquad a_{0,0} = -1, \qquad n > 0 \tag{94}$$

we obtain a finite-order semicausal MVR

$$u_{i,j} = \sum_{\substack{m=-p \\ m \neq 0}}^{p} a_{m,0} u_{i-m,j} + \sum_{n=1}^{q} \sum_{m=-p}^{p} a_{m,n} u_{i-m,j-n} + \epsilon_{i,j} \tag{95}$$

where $\{\epsilon_{i,j}\}$ is a moving average with SDF

$$S_\epsilon(z_1, z_2) = \beta^2 \left[1 - \sum_{m=-p}^{p} a_{m,0} z_1^{-m}\right]. \tag{96}$$

The SDF of this semicausal MVR is

$$S_u = \frac{S_\epsilon}{\left[\sum_{n=0}^{q} \sum_{m=-p}^{p} a_{m,n} z_1^{-m} z_2^{-n}\right]\left[\sum_{n=0}^{q} \sum_{m=-p}^{p} a_{m,n} z_1^{m} z_2^{n}\right]}. \tag{97}$$

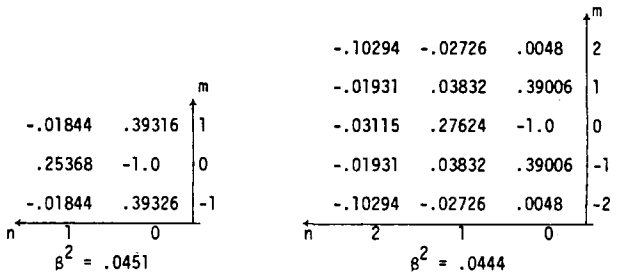


Fig. 12. Semicausal MVR coefficients.

Thus it is possible to find finite-order semicausal MVR's which would realize a given positive and analytic SDF arbitrarily closely. It can also be shown that for a semicausal prediction window with suitably large but finite values of p and q, the finite-order semicausal MVR obtained via the solution of (82) would realize a given positive and analytic SDF arbitrarily closely. Thus, even though the solution of (82) corresponding to a fixed size window W need not ensure an admissible (i.e., stable) representation, solving it successively (or recursively) for increasing size windows should eventually lead to an admissible as well as a reasonably accurate representation. The advantage, of course, is that one would be solving only finite-order equations, whereas the approach via equations (A24), (A25), in general requires solving an infinite set of equations (i.e., for every $|z_1| = 1$). In many examples (see below and in the next section), the acceptable values of p and q have been found not to be very large.

Remarks

1) The finite-order semicausal MVR's realize SDF's which contain both numerator and denominator polynomials (see (97)). However, the numerator polynomial is one dimensional and is in the noncausal dimension.

2) The above remark implies semicausal MVR's are driven by colored noise. In fact $\{\epsilon_{i,j}\}$ is a moving average process in the "i" variable and is white in the "j" variable.

3) Semicausal MVR's were introduced in [9] and have been found useful in developing semirecursive or the so-called hybrid algorithms which are recursive in one direction and are transform based in the other directions. See examples in Section VI and [51], [52], [55], [57]-[59].

4) The semicausal models are recursive (or causal) only in the j variable. A re-indexing in the i-variable to attempt to represent it as a two-dimensional causal model would yield an unstable model. Thus the model in Example 7 written as

$$u_{i,j} = \frac{1}{a} u_{i-1,j} - u_{i-2,j-1} - \frac{b}{a} u_{i-1,j-1} - \frac{1}{a} \epsilon_{i-1,j}$$

would be unstable if solved recursively in i and j.

Example 8: We return to the nonseparable exponential covariance function considered in Example 5. Fig. 12 shows the semicausal MVR coefficients for $p = q = 1$ and $p = q = 2$.

Fig. 11 shows the covariance match achieved by the semicausal MVR model. Compared to single quadrant causal model for $p = q = 2$, we find a better covariance fit is provided by the semicausal model (see Fig. 13).

D. Noncausal Representations

The results for two-dimensional noncausal MVR's are analogous to the one dimensional results discussed in Section III-E. All the relevant equations for the two-dimensional case are given in the Appendix. The following conclusions are made.

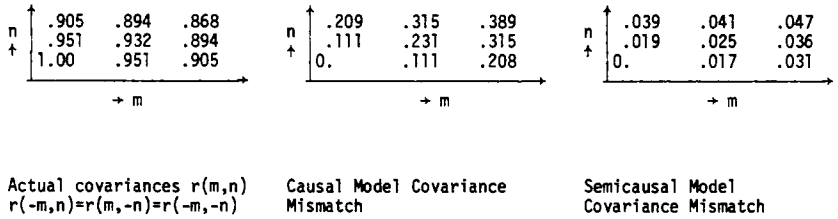


Fig. 13. Covariance mismatch of the causal and semicausal models.

1) An arbitrary positive SDF can be realized arbitrarily closely by a stationary finite-order noncausal MVR provided S^{-1} has a uniformly convergent Fourier series. Conversely, if the given SDF S is a rational function with a zeroth-order numerator polynomial, then there exists a unique finite-order noncausal MVR realization of S.

2) As in the case of causal and semicausal models, an admissible finite-order noncausal MVR with a specified spectral mismatch error could be identified via the finite-order block Toeplitz equation (82).

Example 9: Consider the SDF

$$S(z_1, z_2) = [1 - \alpha(z_1 + z_1^{-1} + z_2 + z_2^{-1})]^{-1}, \quad 0 < \alpha < \tfrac{1}{4}. \tag{98}$$

Since S^{-1} is a two-dimensional polynomial, the *noncausal MVR* is

$$u_{i,j} = \alpha(u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1}) + \epsilon_{i,j} \tag{99}$$

$$r_\epsilon(k, l) = \begin{cases} 1, & (k, l) = (0, 0) \\ -\alpha, & (k, l) = (\pm 1, 0), (0, \pm 1) \\ 0, & \text{otherwise.} \end{cases}$$

Now let us consider finding a semicausal MVR. Comparing S with (89) we see $q = 1$. The equation for $r_n(z)$ is obtained via (A11) and (98) as

$$[1 - \alpha(z_1 + z_1^{-1})]\, r_n(z_1) - \alpha[r_{n-1}(z_1) + r_{n+1}(z_1)] = \delta_{n,0}$$

which solves to give

$$r_n(z_1) = r_n(z_1^{-1}) = AC^{|n|}$$
$$A = 1/\sqrt{\alpha_0^2 - 4\alpha^2}$$
$$\alpha_0 = 1 - \alpha(z_1 + z_1^{-1})$$
$$C = [\alpha_0 - \sqrt{\alpha_0^2 - 4\alpha^2}\,]/2\alpha.$$

Using this in (A25) and (A26) for $q = 1$ and simplifying we obtain

$$\hat{a}_1 = \hat{a}_1^1 = -C(z_1^{-1})$$
$$h(z_1) = A^{-1}[1 - C(z_1)\, C(z_1^{-1})]^{-1} = \tfrac{1}{2}\{\alpha_0 + \sqrt{\alpha_0^2 - 4\alpha^2}\}$$
$$a_0 = a_0^1 = -\beta^2 h(z_1)$$
$$a_1 = \beta^2 h(z_1)\, C(z_1^{-1}) = \alpha\beta^2.$$

Note $h(z_1)$ is an irrational function of z_1. Assuming $0 < \alpha \ll 1$, we obtain the rational approximation

$$h(z_1) = 1 - \alpha(z_1 + z_1^{-1}) + 0(\alpha^2).$$

Ignoring $0(\alpha^2)$ terms we obtain

$$\beta^2 \simeq 1,\ a_0(z_1) = -1 + \alpha(z_1 + z_1^{-1}),\ a_1(z_1) = \alpha,\ S_\epsilon = -a_0(z_1)$$

Fig. 14. Homomorphic transform method for two-dimensional spectral factorization.

and the finite-order semicausal MVR

$$u_{i,j} = \alpha(u_{i-1,j} + u_{i+1,j}) + \alpha u_{i,j-1} + \epsilon_{i,j} \tag{100}$$

$$r_\epsilon(k, l) = \delta_{l,0} \begin{cases} 1, & k = 0 \\ -\alpha, & |k| = 1 \\ 0, & \text{otherwise.} \end{cases}$$

For $0 < \alpha < \frac{1}{3}$, equation (100) can be shown to be a stable semicausal model. The SDF corresponding to this model is

$$S_u = \frac{[1 - \alpha(z_1 + z_1^{-1})]}{[1 - \alpha(z_1 + z_1^{-1}) - \alpha z_2^{-1}]\,[1 - \alpha(z_1^{-1} + z_1) - \alpha z_2]}.$$

For $a = b = \alpha$ in Example 7, we see (100) and the model considered there are identical. A causal MVR approximation of (100) is obtained as discussed in that example.

E. Causal and Semicausal Model Realizations Via Homomorphic Transformation

The factorization required in the causal and semicausal models could be achieved via the so-called homomorphic transformation[3] shown in Fig. 14. One starts with the Fourier series of log S (the homomorphic transform of S) as

$$\tilde{S}(z_1, z_2) \triangleq \log S(z_1, z_2) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} c_{m,n} z_1^{-m} z_2^{-n},$$
$$|z_1| = 1,\ |z_2| = 1 \tag{101}$$

where it is assumed the sequence $\{c_{m,n}\}$, also called the *Cepstrum*, is absolutely summable, i.e.,

$$\sum_m \sum_n |c_{m,n}| < \infty. \tag{102}$$

Now, if $\tilde{S}$ is decomposed as a sum of, say, three components as

$$\tilde{S} \triangleq \tilde{S}_\epsilon + \tilde{H}^+ + \tilde{H}^- \tag{103}$$

then

$$S = \frac{e^{\tilde{S}_\epsilon}}{[e^{-\tilde{H}^+} e^{-\tilde{H}^-}]} \triangleq S_\epsilon H^+ H^- \tag{104}$$

is a product of three factors. If the decomposition is such that

[3] Also called Wiener–Doob factorization [61], [62] or Hilbert transformation [56].

$H^+(z_1, z_2) = H^-(z_1^{-1}, z_2^{-1})$ and $S_\epsilon(z_1, z_2) = S_\epsilon(z_1^{-1}, z_2^{-1})$ then there exists a stable two-dimensional linear system with transfer function $H^+(z_1, z_2)$ such that the SDF of the output is S if the SDF of the input is S_ϵ.

This algorithm is due to Ekstrom and Woods [64] and is a direct extension in two dimensions of a method by Wiener [61]. The causal and the semicausal models can be realized by finding their respective decompositions in such a way that the impulse response $h^+_{m,n} = \mathcal{F}^{-1}\{H^+(\omega_1, \omega_2)\}$ has causal and semicausal regions of support $\hat{\delta}_0 = \hat{\delta} \cup (0, 0)$ where $\hat{\delta}$ is defined in (74). We give the specific decompositions for causal MVR, semicausal WNDR, and semicausal MVR models.

Causal MVR:

$$\tilde{H}^+ = \sum_{m=1}^{\infty} c_{m,0} z_1^{-m} + \sum_{m=-\infty}^{\infty} \sum_{n=1}^{\infty} c_{m,n} z_1^{-m} z_2^{-n}$$

$$\tilde{H}^- = \sum_{m=-\infty}^{-1} c_{m,0} z_1^{-m} + \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{-1} c_{m,n} z_1^{-m} z_2^{-n}$$

$$\tilde{S}_\epsilon = c_{0,0}. \tag{105}$$

In view of (102), $\tilde{H}^+(z_1, z_2)$ is analytic in the region $\{|z_1| = 1, |z_2| \geq 1\} \cup \{|z_1| \geq 1, z_2 = \infty\}$. Hence $H^+(z_1, z_2)$ will be analytic in the same region. Therefore, the impulse response $h^+_{m,n}$ will be zero for $[-\infty \leq m \leq -1, n = 0] \cup [n < 0, \forall m]$, i.e., $\{h^+_{m,n}\}$ is causal.

Semicausal WNDR:

$$\tilde{H}^+ = \tfrac{1}{2} \sum_{m=-\infty}^{\infty} c_{m,0} z_1^{-m} + \sum_{m=-\infty}^{\infty} \sum_{n=1}^{\infty} c_{m,n} z_1^{-m} z_2^{-n}$$

$$\tilde{H}^- = \tfrac{1}{2} \sum_{m=-\infty}^{\infty} c_{m,0} z_1^{-m} + \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{-1} c_{m,n} z_1^{-m} z_2^{-n}$$

$$\tilde{S}_\epsilon = 0. \tag{106}$$

Now $\tilde{H}^+$ and hence H^+ will be analytic in the region $\{|z_1| = 1, |z_2| \geq 1\}$ and impulse response $h^+_{m,n}$ will be zero for $\{\forall m, n < 0\}$, i.e., $h^+_{m,n}$ is semicausal.

Semicausal MVR: A simple examination of (95)–(97) reveals that semicausal MVR's require that within a constant multiplier S_ϵ should equal $1/H^+(z_1, \infty)$. Using this condition, we obtain

$$\tilde{H}^+ = \sum_{m=-\infty}^{\infty} c_{m,0} z_1^{-m} + \sum_{m=-\infty}^{\infty} \sum_{n=1}^{\infty} c_{m,n} z_1^{-m} z_2^{-n}$$

$$\tilde{H}^- = \sum_{m=-\infty}^{\infty} c_{m,0} z_1^{-m} + \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{-1} c_{m,n} z_1^{-m} z_2^{-n}$$

$$\tilde{S}_\epsilon = -\sum_{m=-\infty}^{\infty} c_{m,0} z_1^{-m}. \tag{107}$$

The region of analyticity of $\tilde{H}^+$ and H^+ is $\{|z_1| = 1, |z_2| \geq 1\}$ and that of $\tilde{S}_\epsilon$ and S_ϵ is $\{|z_1| = 1, \forall z_2\}$. Hence $h^+_{m,n}$ is semicausal and S_ϵ is a SDF.

Example 10: Consider the SDF of Example 9. Assuming $0 < \alpha \ll 1$, we obtain

$$\tilde{S} \simeq \alpha(z_1 + z_1^{-1} + z_2 + z_2^{-1}) + 0(\alpha^2).$$

Ignoring $0(\alpha^2)$ terms we obtain

$$H^+ = \begin{cases} \alpha(z_1^{-1} + z_2^{-1}), & \text{causal MVR} \\ \frac{1}{2}\alpha(z_1 + z_1^{-1}) + \alpha z_2^{-1}, & \text{semicausal WNDR} \\ \alpha(z_1 + z_1^{-1}) + \alpha z_2^{-1}, & \text{semicausal MVR} \end{cases}$$

$$\tilde{S}_\epsilon = \begin{cases} 0, & \text{for causal MVR and semicausal WNDR} \\ -\alpha(z_1 + z^{-1}), & \text{for semicausal MVR.} \end{cases}$$

This gives

$$H^+ = 1/e^{-H^+} \simeq \begin{cases} 1/[1 - \alpha z_1^{-1} - \alpha z_2^{-1}], & \text{causal MVR} \\ 1/[1 - \frac{\alpha}{2}(z_1 + z_1^{-1}) - \alpha z_2^{-1}], & \text{semicausal WNDR} \\ 1/[1 - \alpha(z_1 + z_1^{-1}) - \alpha z_2^{-1}], & \text{semicausal MVR} \end{cases}$$

$$S_\epsilon \simeq \begin{cases} 1, & \text{causal MVR and semicausal WNDR} \\ 1 - (z_1 + z_1^{-1}), & \text{semicausal MVR.} \end{cases}$$

Remarks

1) We note that for noncausal models, one need not go through the above procedure. A suitable truncation of the Fourier series of S^{-1} yields the desired model.

2) In general, the Fourier series of $\tilde{S}$ and H^+ will contain an infinite number of terms even if S and $\tilde{H}^+$ (or its approximation) respectively were (finite order) polynomials. Therefore, a practical algorithm requires two stages of numerical approximations. First, in the evaluation of $\tilde{S}$ and then in the evaluation of H^+. An algorithm utilizing the DFT to obtain such an approximate spectral factorization and its error analysis has been considered by Rino [63] for one-dimensional problems. Extension of the algorithm in two dimensions is given in [64].

3) From the foregoing discussion the BIBO stability condition for finite-order causal and the semicausal models can be stated quite simply as follows.[4]

Theorem 4: A causal MVR whose transfer function H is given by

$$H(z_1, z_2) = 1/A(z_1, z_2)$$

$$A(z_1, z_2) = 1 - \sum_{m=1}^{p} a_{m,0} z_1^{-m} - \sum_{m=-p}^{p} \sum_{n=1}^{q} a_{m,n} z_1^{-m} z_2^{-n}$$

is stable if and only if i) $A(z_1, z_2) \neq 0$, $|z_1| \geq 1$, $z_2 = \infty$, ii) $A(z_1, z_2) \neq 0$, $|z_1| = 1$, $|z_2| \geq 1$.

Theorem 5: A semicausal WNDR or MVR whose transfer function H is given by

$$H = 1/A$$

$$A(z_1, z_2) = 1 - \sum_{\substack{m=-p \\ m \neq 0}}^{p} a_{m,0} z_1^{-m} - \sum_{m=-p}^{p} \sum_{n=1}^{q} a_{m,n} z_1^{-m} z_2^{-n}$$

[4] It is assumed that the causal, semicausal, and the noncausal model equations are solved, respectively, as initial, initial-boundary, and boundary value problems.

is stable if and only if $A(z_1, z_2) \neq 0$, $|z_1| = 1$, $|z_2| \geq 1$.

Theorem 6: A noncausal MVR or WNDR whose transfer function is given by

$$H = 1/A,\ A(z_1, z_2) = 1 - \sum_{m=-p}^{p} \sum_{n=-q}^{q} a_{m,n} z_1^{-m} z_2^{-n}$$

is stable if and only if $A(z_1, z_2) \neq 0$, $|z_1| = 1$, $|z_2| = 1$.

The conditions stated in these theorems assure H to be analytic in the appropriate regions in the z_1, z_2 hyperplanes so that its stability conditions are satisfied. Two-dimensional stability has been discussed almost invariably for causal models and occasionally for noncausal models. For details see [65]-[68].

F. Other Methods

Since exact factorization of a two-dimensional SDF $S(\omega_1, \omega_2)$ (rational or not) as the magnitude square of a rational function $H(\omega_1, \omega_2)$ is not possible in general, any number of other methods which could give reasonable rational approximations should be possible. The advantage of model identification via the prediction model orthogonality equations (Sections V-A to V-D) is that the equations to be solved are linear. The advantage of the homomorphic transform approach is that a suitable FFT (even if large in size) could be used to obtain the approximate models. In both procedures, however, situations exist where the model order may get very large in order to be stable.

In image processing applications it is generally desirable that the model order not be very large. This is because the image data is often processed block by block (often 16×16) and if the model order approaches the size of a column (or row), then nonrecursive or transform based methods become more efficient (even computationally). Therefore, identification techniques which give reasonable (low-) order models with assurance of stability would be useful in many applications such as restoration and data compression. High-order models, on the other hand, could be useful in image synthesis, object identification, spectral estimation, and digital filter design applications.

Among the various other methods for identifying image models, perhaps the most direct is the parameter identification method via the maximum-likelihood approach [69]. Sugimoto *et al.* [70], [71] have applied this method for identification of finite order semicausal models. Similar approaches are possible for causal and noncausal models although the algorithmic details would differ considerably.

Another method suggested in [51], [52] is based on the fact that the causal, semicausal, and noncausal models can also be considered as finite difference approximations of hyperbolic, parabolic, and elliptic partial differential equations, respectively. Hence, low-order stable models can be determined by identifying the parameters of the finite difference approximations subject to their stability constraints which are known from the theory of their parent partial differential equations.

Since the semicausal MVR and WNDR require only factorization in only one variable, one-dimensional system identification techniques [72] could potentially be applied for identifying these models. Once the semicausal MVR is known, the causal realization (or an approximation of it), if desired, could be found via the method of Section V-C.

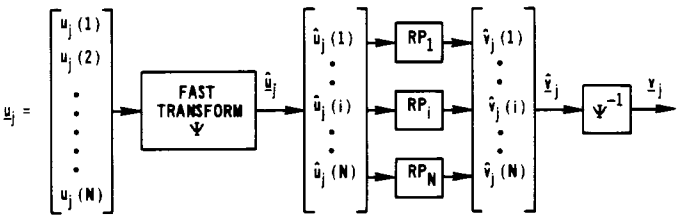


Fig. 15. Parallel structure of semicausal algorithms. RP_i is the ith channel recursive processor such as a recursive filter for image restoration or a DPCM encoder for data compression.

VI. Applications in Image Processing

The three types of models discussed in the foregoing have applications in most of the image processing problems listed in Table I. As mentioned before the choice of a particular model depends not only on the accuracy of the model but also on the associated algorithmic architecture. The three models yield three quite different processing techniques which will be described briefly.

A. Algorithmic Structures

The causal models can be easily viewed to represent images as the output of a line scanner and hence yield, quite naturally, algorithms which are recursive in nature. Causal MVR's have been found most useful in data compression of images [17] for real time transmission where the successive transmitted samples represent the prediction error between the new image-pel and its predicted value based on the past transmission. Causal MVR's have usefulness in the design of recursive filters for noise smoothing and restoration of blurred images (especially when the blurring process is also causal, e.g., in motion blur). It should be mentioned, however, that two-dimensional images generally do not have any causality (or dynamics) associated with them and the usual solutions of smoothing and most restoration problems are boundary valued (or noncausal). The recursive methods therefore have to sweep back and forth between the boundaries. This, together with the algorithmic complexities of two-dimensional recursions, have kept the users favor with more direct Fourier transform based algorithms [6], [7], [73], [75]. However, potential for these algorithms seems to be substantial for space variant restoration problems where Fourier techniques do not apply.

The semicausal models yield hybrid algorithms which are transform based (nonrecursive) in one dimension and recursive in the other. The finite-order semicausal MVR's have a structure where the model coefficients (e.g., matrices in (66)) be represented by banded Toeplitz matrices. It is possible to find a fast unitary sinusoidal transform [10] which would reduce these matrices to their diagonal forms (or nearly so). For example, the semicausal model of (92) reduces to a set of N decoupled first-order Markov processes

$$\lambda_k \hat{u}_j(k) = b\hat{u}_{j-1}(k) + \hat{e}_j(k), \qquad k = 1, \cdots, N \qquad (108)$$

under the transformation $\hat{u}_j = \Psi u_j$, Ψ = DST (see Table III), where λ_k are the eigenvalues of Q. Now, several image processing problems, e.g., noise smoothing, data compression, restoration, system identification etc. can be formulated as N one-dimensional problems, which could be solved in parallel. Fig. 15 shows the parallel structure of semicausal algorithms. Each line u_j of the image is first unitarily transformed and each transformed element $\hat{u}_j(k)$ is processed recursively in the

'j' variable. The output vector $\hat{v}_j$ is inverse transformed to give the processed vector v_j.

Semicausal models have been used in real-time data compression of images for transmission from remotely piloted vehicles (RPV) [17], [58], [74], in restoration and noise smoothing problems [52], [55], [70], [71] spectral estimation [59], etc. The associated algorithms combine the advantages of low memory and hardware complexity of recursive methods and the relatively high performance of transform processing methods.

Noncausal models have been used most widely either as moving average models or via the transform methods. A large number of commonly used image processing operators, e.g., spatial smoothing, gradient and compass edge detectors, bilinear interpolation, unsharp masks, etc., are moving average operators. It has been shown that noncausal MVR's naturally lead to transform domain algorithms and have been found useful in most of the problems listed in Table I.

B. Smoothing, Enhancement, and Restoration of Images

The common smoothing problem is to find the best linear mean-square estimate of the image $u_{i,j}$ given the noisy observations

$$y_{i,j} = u_{i,j} + n_{i,j}$$

where $\{n_{i,j}\}$ is a white-noise field independent of $\{u_{i,j}\}$. Causal models have been used by several authors [52], [76]–[79] to develop recursive filter implementations. Semicausal and vector scanning models have been considered in [52], [55], [70], [71], [80] to develop semirecursive or line to line recursive algorithms. Noncausal models have been shown to yield fast transform based nonrecursive algorithms [52], [73] and have also been found useful in developing moving average FIR filters.

The following two examples show how some of the commonly used operators in image processing are related to the linear prediction models. Use of appropriate prediction models (depending on the images) can lead to similar but better operators.

Example 11: Consider an image represented by the noncausal MVR of (99). The optimum filter transfer function is given by another noncausal MVR as

$$G(z_1, z_2) = \beta^2 / \{\beta^2 + \sigma_n^2[1 - \alpha(z_1 + z_1^{-1} + z_2 + z_2^{-1})]\}$$

where σ_n^2 is the variance of the noise. At signal-to-noise ratios (SNR) of approximately 5, G can be approximated by the FIR filter [52]

$$G \simeq T(z_1, z_2) = \tfrac{1}{2}\,[1 + \tfrac{1}{4}\,(z_1 + z_1^{-1} + z_2 + z_2^{-1})]$$

which is simply a commonly used spatial averaging filter. At a different signal to noise ratios another suitable moving average approximation can be found whose region of support would increase with the reciprocal of SNR. In fact, moving average filters can be adapted (via the noncausal models) to local changes in SNR. Fig. 16 shows that a noisy (SNR = 5) image optimally filtered via an appropriate noncausal model or its approximate spatial averaging filter is better than using the optimal (Wiener) filter based on the usual separable covariance model [52].

Example 12 (Enhancement): The prediction operator (causal, semicausal, or noncausal) denoted by $A(z_1, z_2)$ generally performs some sort of differentiation. When applied to a real-world image, the prediction error generally contains a nonstationary component which generally represents high spatial frequencies, e.g., edges. Thus the operator

$$C(z_1, z_2) = 1 + \lambda A(z_1, z_2)$$

would add to the image a quantity proportional to high spatial frequencies (or gradient). For example, in the case of the noncausal MVR considered above, with $\alpha \simeq \frac{1}{4}$

$$C = (1 + \lambda) - \frac{\lambda}{4}(z_1 + z_1^{-1} + z_2 + z_2^{-1}).$$

Now the operator A is simply a discrete Laplacian and C is called the "*Unsharp Mask*" used often for edge enhancement of images.

Image models have also been used in restoration of images blurred due to motion, atmospheric turbulence and other shift invariant PSF. State variable, line recursive as well as frame-recursive models have been used to develop recursive algorithms for deblurring of images [60]. Noncausal models have been used to develop transform-domain deblurring algorithms. Other restoration techniques such as iterative gradient methods, singular value decomposition have mostly been used [7], [73], [75], [95] for solving deterministic (e.g., least squares interpolation) problems although image models could be used (especially to improve their performance in the presence of noise).

C. Image Data Compression

This is one application where image models have probably had the most significant impact. Causal prediction models have been used most widely in the design of intraframe and interframe predictive or the so-called DPCM coders. Semicausal models have been employed in transform/DPCM coding for RPV and teleconferencing applications. Noncausal models give rise to transform coding algorithms which have been found to give high performance. The prediction model parameters determine the coder design details such as quantizer design, bit allocation, optimum transform, etc. These details and other related considerations for image data compression may be found in the recent survey articles [17], [81] and in [16], [57].

Example 13: Consider the semicausal model of (92) when it is white noise driven. It reduces to (108) after sine transformation of the vectors u_j. Fig. 17 shows images encoded via this model at an average rate of 1 bit/pel. This model has also been used effectively to simultaneously filter and encode noisy images. For details see [17], [57], [58].

D. Edge Extraction

Example 14 (Edge Extraction): An image can be considered as being composed of two components, i.e.,

$$u_{i,j} = u_{i,j}^s + u_{i,j}^b$$

where $u_{i,j}^s$ and $u_{i,j}^b$ represent the stationary and nonstationary (boundaries and edges) components respectively. Such two source models for images have been considered by Schreiber, Yan and Sakrison, Jain and Wang, and others [82]–[84], [57]. Using a WNDR for the stationary component, the image can be expressed as

$$\epsilon_{i,j} = \epsilon_{i,j}^s + \epsilon_{i,j}^b$$

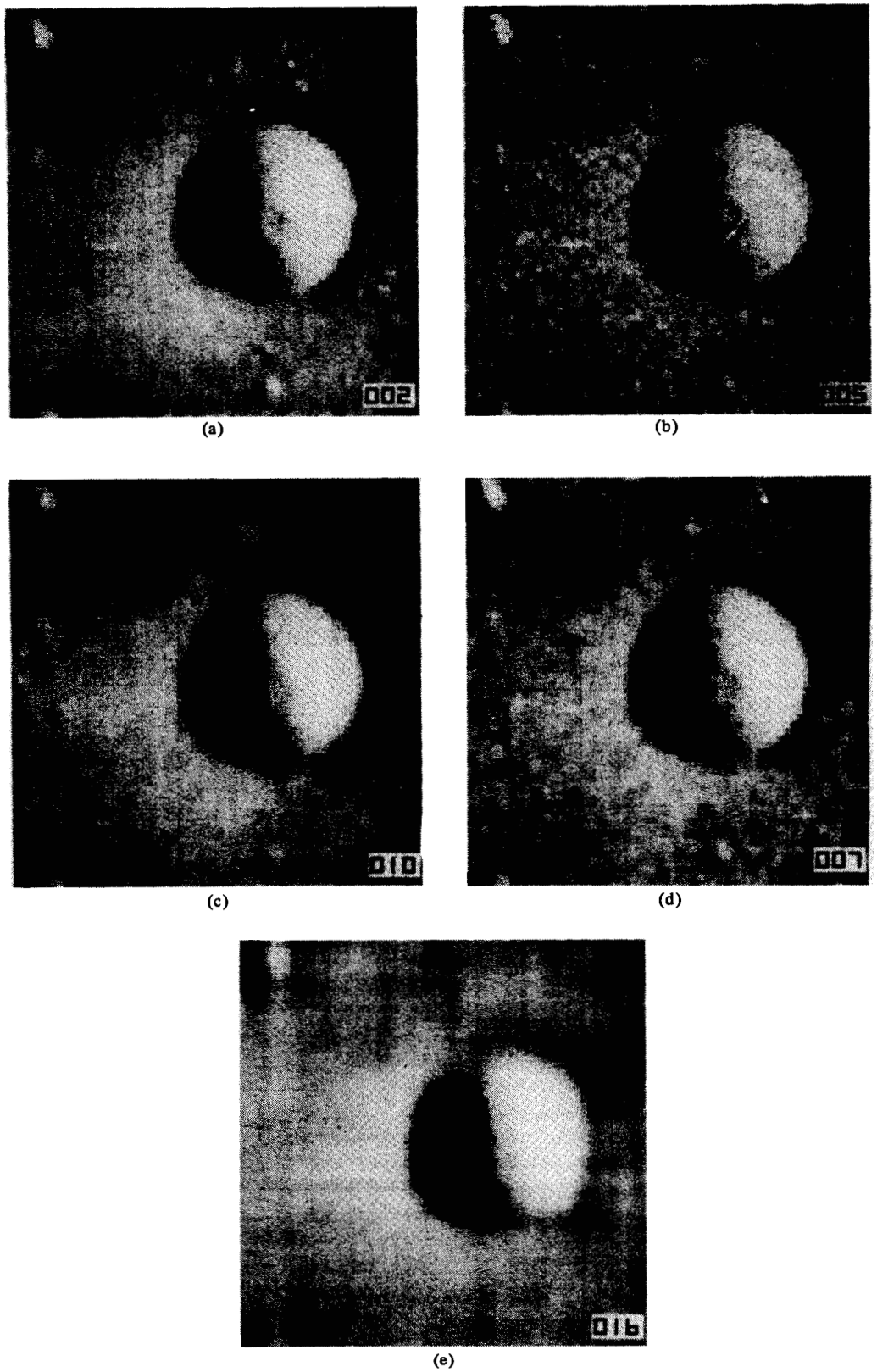


Fig. 16. Smoothing of a noisy image. (a) Original. (b) Noisy. (c) Optimum filtered using a noncausal model. (d) Spatial averaged. (e) Optimum filtered based on the separable covariance model.

where $\epsilon = \mathfrak{L}[u]$, and $\mathfrak{L}$ is the whitening prediction operator (causal, semicausal or noncausal). Now, $\epsilon_{i,j}$ can be viewed as a sum of signal $\epsilon^{p}_{i,j}$ (edges) and white noise $\epsilon^{s}_{i,j}$. The edge detection problem is therefore reduced to the detection of a signal in the presence of noise. Fig. 18 shows the edges detected using a noncausal model. Such two source models are useful in designing "edge preserving" image-processing algorithms.

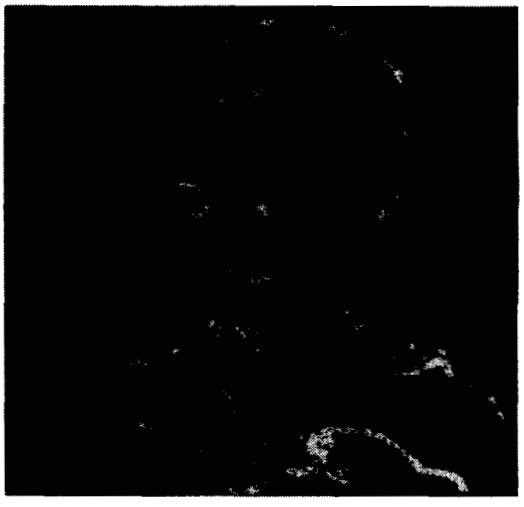
(a)

(b)

(c)

Fig. 17. Data compression of images via a semicausal model. (a) Original. (b) Encoded. 1 bit/pel. (c) Adaptive codings, 1 bit/pel. In the adaptive method, the model parameters are updated after every 16 X 16 block of pels. Mean-square error improvement due to adaptation is in excess of 3 dB.

(a)



(b)

Fig. 18. Edges extracted using a noncausal model.

E. Two-Dimensional Spectral Estimation

Let $r(m, n)$ be a two-dimensional positive definite covariance sequence given on a rectangular window $W_c = [-p_c, p_c] \times [-q_c, q_c]$. The problem is to find an estimate of the SDF associated with this covariance sequence. In two dimensions, the maximum entropy spectral estimation problem is to *find a positive definite extension of* $r(m, n)$ *outside* W_c *such that the entropy*

$$H = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \log S(\omega_1, \omega_2)\, d\omega_1\, d\omega_2$$

is maximized. Note that it is required that the Fourier coefficients of the estimated spectrum S match the given covariances exactly on W_c. The solution, if it exists, requires that the Fourier series of $1/S$ should be truncated outside the window W_c, i.e., S should be the SDF of a noncausal MVR, i.e., of the form

$$S(z_1, z_2) = \beta^2 \left[\sum_{m,n} \sum_{\in W_c} \alpha_{m,n} z_1^{-m} z_2^{-n} \right]^{-1}, \qquad |z_1| = 1, \quad |z_2| = 1 \quad (109)$$

where the coefficients $\alpha_{m,n}$, β^2 are determined by matching the given covariances to the Fourier coefficients of S on W_c. Unfortunately, this maximum entropy solution need not exist because, in two (and higher) dimensions, a positive definite sequence on a rectangular window need not have a positive

definite extension [85], [86]. However, if $r(m, n)$ has at least one positive definite extension then the maximum entropy solution (109) exists and is unique [94]. When the solution exists, the problem of finding $\alpha_{m,n}$ is nonlinear and cannot be, in general, reduced to a linear problem as in the one-dimensional case. This is because of the spectral factorization difficulty mentioned earlier.

However, it is possible to obtain positive SDF estimates if we relax the condition of exact covariance match on W_c. This is motivated by the fact that often the available covariances are estimated from data and are noisy. The foregoing finite-order causal, semicausal, and noncausal MVR's could be used to estimate the SDF. The algorithm simply requires solving (80) for $(k, l) \in W_0$ or equivalently (81) and (82) where W_0 is a subset of $\hat{\hat{\delta}}_0$ and could be a causal, semicausal, or noncausal prediction filter window. Note that the W_0 and W_c are not of equal size. For example, for a causal MVR, if $W_0 = [0, p] \times [0, q]$ then the covariance sequence is needed over $W_c = [-p, p] \times [-q, q]$. After having solved (80), the estimated SDF (if admissible) is given as follows.

Causal MVR:

$$S(z_1, z_2) = \beta^2 \Big/ \left| 1 - \sum_{m,n \in W} \sum a_{m,n} z_1^{-m} z_2^{-n} \right|^2, \qquad |z_1| = 1, \quad |z_2| = 1. \quad (110)$$

Semicausal MVR:

$$W_0 = [-p, p] \times [0, q], \qquad W_c = [-2p, 2p] \times [-q, q], \qquad W = W_0 - (0, 0)$$

$$S(z_1, z_2) = S_\epsilon(z_1, z_2) \Big/ \left| 1 - \sum_{m,n \in W} \sum a_{m,n} z_1^{-m} z_2^{-n} \right|^2, \qquad |z_1| = 1, \quad |z_2| = 1$$

$$S_\epsilon(z_1, z_2) = \beta^2 \left[1 - \sum_{m=1}^{p} a_{m,0} (z_1^{-m} + z_1^{m}) \right]. \quad (111)$$

Noncausal MVR:

$$W_0 = [-p, p] \times [-q, q],$$

$$W_c = [-2p, 2p] \times [-2q \times 2q], \qquad W = W_0 - (0, 0)$$

$$S(z_1, z_2) = \beta^2 \left[1 - \sum_{m,n \in W} \sum a_{m,n} z_1^{-m} z_2^{-n} \right]^{-1}, \qquad |z_1| = 1, \quad |z_2| = 1. \quad (112)$$

It should be noted that the solution of (81) and (82), while it guarantees admissible predictor coefficients, the estimated spectrum may not always be positive (i.e., admissible) for any arbitrary positive definite $\mathcal{R}$. This is because the solution yields coefficients which satisfy (80) only partially, i.e., for $k, l \in W_0$ (and not necessarily for all $k, l \in \hat{\delta}_0$). However, for a large enough W_0 and covariances that come from an analytic SDF, an admissible SDF can be found. In many practical examples, the window size W_0 has been found to be not large and in many cases it is actually quite small for admissible SDF.

Example 15: Consider the covariance sequence

$$r(m, n) = \frac{\cos 2\pi(m + n)}{8} + \frac{\cos 2\pi(m + n)}{12} + 0.05\delta_{m,0}\delta_{n,0}$$

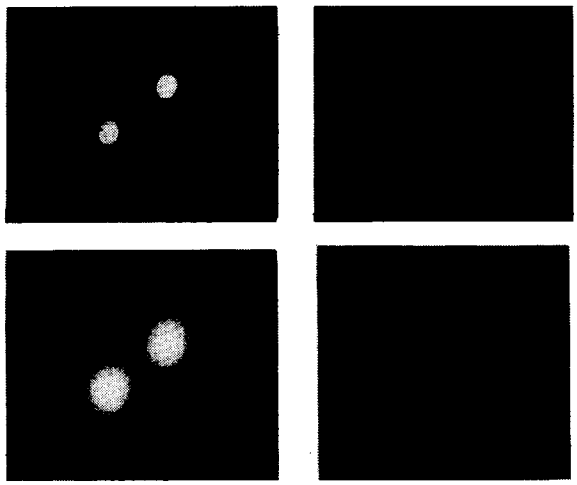


Fig. 19. Spectral estimation comparison of windowed DFT versus semicausal MVR methods. Top row; from left windowed DFT spectrum amplitude versus semicausal MVR spectrum amplitude. Bottom row; same quantities shown on dB scale.

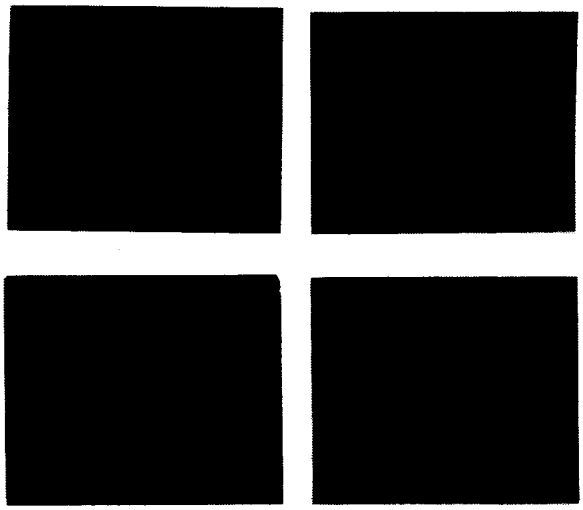


Fig. 20. Spectral estimation comparison of causal versus semicausal models. Top row from left; model spectrum amplitude versus semicausal model spectrum amplitude. Bottom row; same quantities shown on dB scale.

whose SDF has line spectra at $\pm(\frac{1}{8}, \frac{1}{8})$ and $\pm(\frac{1}{12}, \frac{1}{12})$ in the frequency plane. The covariances are available over $[-4, 4] \times [-2, 2]$. Figs. 19 and 20 shows the spectra estimated by a) Bartlett windowing and taking DFT, b) single quadrant causal MVR for $p = q = 5$, c) semicausal MVR, $p = 3, q = 4$. Note that objects not resolved by the DFT method are resolved by the causal and semicausal estimates. Spurious peaks appear in the causal model and the covariance match over the window W_c is not very good. The semicausal estimate is much more accurate both in terms of resolution, peak location and its covariance match. For other examples and more discussion see [59].

Remarks:

1) The causal and noncausal MVR spectra (110) and (112) are of the same form as the maximum entropy spectrum (109). However, they would not be equal in general. If the given covariance sequence is separable, then the maximum entropy and the causal, the semicausal and noncausal spectra would be identical.

2) The semicausal MVR spectra is ARMA type and is therefore not a maximum entropy spectra.

3) All the equations for calculating the prediction model coefficients are linear.

4) As expected from the theory of these two-dimensional prediction models, experimentally, it has been observed that the covariance match on the window W_c improves as p, q are increased.

F. Extension to Adaptive Models and Other Applications

Other applications of image models are in adaptive processing where the model parameters are updated with the changes in image properties–Fig. 17, for example, shows the result of adapting the semicausal model of (108) block by block. In general adaptations could be made continuously and recursively from pel to pel, as is done in many DPCM techniques for image coding. However, it has been observed that models which are adapted at every pel tend to be sensitive to noise and rarely show any significant improvement over more practical block by block adaptive models.

Applications also exist in identification, recognition and synthesis of images of texture. Image models also provide a priori information concerning performance bounds such as achievable signal to noise ratio, compression versus distortion, etc., [16], [52].

VII. Summary and Conclusions

We have presented several mathematical models which have been used and are of potential use in image processing. Included in our discussion were several traditional models (e.g., series expansion, one dimensional AR, state variable etc.) and some new models (viz. the causal, semicausal and noncausal prediction models). Not included in our discussion were considerations of the Markovian property, stability tests, controllability, observability and other such notions. Also, not considered were models for vector random fields which have potential applications in multispectral and color image processing. Theoretical results for these and related topics can be found in [66], [87]–[92]. It was shown however, that the models considered do have a large number of applications in image (and two-dimensional signal) processing. Among the prediction models considered, the semicausal and noncausal models seem to offer several algorithmic and performance tradeoffs against the purely causal (or recursive) methods.

Appendix

A.1 Causal Minimum Variance Representations

First we consider the general nonstationary case for arbitrary random fields. Let us define a block matrix $\mathcal{R}$ whose elements are the matrices $R_{j,k}$ which are the cross covariances $Eu_j u_k^T$. Then $\mathcal{R}$ is the covariance matrix of the column ordered array vector u. Let $\{u(k)\}$, $\{\bar{u}(k)\}$ and $\{e(k)\}$ be the column ordered sequences corresponding to the arrays $\{u_{i,j}\}$, $\{\bar{u}_{i,j}\}$ and $\{\epsilon_{i,j}\}$, respectively. From our definition of causality, we can write

$$u(k) = \bar{u}(k) + e(k).$$

The orthogonality condition (69) requires that $e(k)$ be orthogonal to $u(l)$ for $\forall l < k$. This means $\{e(k)\}$ must be a white-noise sequence because

$$Ee(l)\,e(k) = E[u(l) - \bar{u}(l)]\,e(k) = 0, \qquad \forall l < k.$$

Since $Ee(l)e(k) = Ee(k)e(l)$, the above is true also for $k < l$. Thus *causal MVR's are also white-noise driven representations*. If $\beta_{i,j}^2$ denotes the variance of $\epsilon_{i,j}$, we have

$$E\epsilon_{i,j}\epsilon_{i+m,j+n} = \beta_{i,j,n}^2 \delta_{m,0}\delta_{n,0}. \tag{A1}$$

In matrix notation, equations (64) and (71) become

$$-A_{j,j}u_j = \sum_{n<j} A_{j,n}u_n + \epsilon_j, \qquad \forall j$$

where

$$-A_{j,j} \triangleq I - A_{j,j}^0 \tag{A2}$$

is a unit lower triangular matrix (i.e., the diagonal entries are 1). This equation can also be written in terms of a *block lower triangular* matrix $\mathcal{A}$ whose block elements are the matrices $A_{j,n}$, as

$$-\mathcal{A}u = e \qquad \mathcal{A} = \{A_{j,n}\}. \tag{A3}$$

Since the diagonal blocks of $\mathcal{A}$ are $\{A_{j,j}\}$, which are lower triangular matrices themselves, the matrix $\mathcal{A}$ is a strictly lower triangular matrix. Equation (A3) implies

$$\mathcal{A}[Euu^T]\mathcal{A}^T = Eee^T \quad \text{or} \quad \mathcal{R} = \mathcal{A}^{-1}\mathcal{R}_e(\mathcal{A}^{-1})^T$$

$$\text{or} \quad \mathcal{R}^{-1} = \mathcal{A}^T\mathcal{R}_e^{-1}\mathcal{A} \tag{A4}$$

where $\mathcal{R}_e$ is a diagonal matrix of the variances of $e(k)$. This is a *lower-upper* factorization of $\mathcal{R}$. Thus, the causal MVR requires a factorization of the block covariance matrix $\mathcal{R}$ by a strictly lower triangular matrix. For a positive definite $\mathcal{R}$, such a factorization always exists and can be obtained in two stages of factorization. The first stage requires a block-lower triangular factorization of $\mathcal{R}$. The second stage requires lower triangular factorization of a sequence of matrix blocks. For example, for $N \times N$ images, define

$$\hat{A}_{j,n} \triangleq A_{j,j}^{-1}A_{j,n} \qquad R_{n,k} \triangleq Eu_n u_k^T. \tag{A5}$$

Then the desired solution of (A4) is obtained by first solving for $\hat{A}_{j,n}$ the so called normal equation

$$\sum_{n=1}^{j-1} \hat{A}_{j,n}R_{n,k} = -R_{j,k}, \qquad 1 \leqslant k \leqslant j-1, \qquad j = 2, \cdots, N \tag{A6}$$

which is a set of $(j-1) \times (j-1)$ block matrix equations.

Given $\{\hat{A}_{j,n}, n \leqslant j\}$, the unit lower triangular matrix $A_{j,j}$ is obtained by the lower-upper factorization

$$\left[\sum_{n=1}^{j} \hat{A}_{j,n}R_{n,j}\right]^{-1} = A_{j,j}^T(R_j^e)^{-1}A_{j,j} \tag{A7}$$

where R_j^e is the covariance matrix of the prediction error vector e_j and is diagonal. Once $A_{j,j}$ is known, $A_{j,n}$ is obtained via (A5). In practice the above algorithm will be quite tedious because the size of $\mathcal{R}$ is $N^2 \times N^2$ and the number of operations would be $0(N^6)$.

In the case of stationary models defined on the infinite plane if one starts with (73), a parallel result corresponding to (A1), (A5)–(A7), is obtained as

$$-A_0 u_j = \sum_{n=1}^{\infty} A_n u_{j-n} + e_j \qquad Ee_j e_k^T = \beta^2 I\delta_{j,k} \tag{A8}$$

and

$$-\sum_{n=1}^{q} \hat{A}_n^q R_{n-j} = R_{-j}, \quad 1 \leqslant j \leqslant q, \quad q = 1, 2, \cdots$$

$$\left[\sum_{n=1}^{q} \hat{A}_n^q R_n\right]^{-1} = \frac{1}{\beta^2}[A_0^q]^T[A_0^q], \quad R_n \triangleq E u_j u_{j+n}^T \tag{A9}$$

$$A_n^q \triangleq A_0 \hat{A}_n^q, \quad A_n = \lim_{q \to \infty} A_n^q, \quad \forall n \geqslant 0 \tag{A10}$$

where A_0 is a unit lower triangular matrix and $\hat{A}_n^q$, A_n^q, R_n, etc., are all doubly infinite Toeplitz matrices. If the given SDF $S(\omega_1, \omega_2)$ is positive and analytic, then following [56], it could be shown that there exists a unique[5] causal realization such as (A8). Moreover, the coefficients A_n could be obtained by solving the sequence of normal equations of (A9) and taking the limit $q \to \infty$ as in (A10). In terms of Z_1-transforms

$$r_n(z_1) \triangleq \sum_{m=-\infty}^{\infty} r(m, n) z_1^{-m}$$

$$a_n(z_1) \triangleq \sum_{m=-\infty}^{\infty} a_{m,n} z_1^{-m}, \quad n \geqslant 1$$

$$a_0(z_1) \triangleq \sum_{m=0}^{\infty} a_{m,0} z_1^{-m}, \text{ etc.} \tag{A11}$$

Equations (A9), (A10) reduce to scalar qth-order AR model equations, parametric in z_1

$$-\sum_{n=1}^{q} \hat{a}_n^q(z_1) r_{n-j}(z_1) = r_j(z_1^{-1}), \quad j = 1, \cdots, q \tag{A12}$$

where

$$-\sum_{n=0}^{q} \hat{a}_n^q(z_1) r_n(z) = \beta^2 / [a_0^q(z_1^{-1}) a_0^q(z_1)]$$

$$a_n^q(z_1) = a_0^q(z_1) \hat{a}_n^q(z_1), \quad \hat{a}_0^q(z_1) = 1, \quad n = 1, \cdots q \tag{A13}$$

where $a_0^q(z_1)$ is such that all its roots are inside the unit circle $|z_1| = 1$. If the SDF is a rational function of the form of (89), these equations can be used to obtain the causal MVR for the given value of q. In general this representation will be of infinite order. A finite-order approximation may be obtained as explained in Section V-B.

A.2 Semicausal Minimum Variance Representation

For semicausal models the orthogonality conditions require that for $\forall k < j$, the prediction error vector $\mathbf{e}_j$ be orthogonal to u_k, and at $k = j$, the elements $\epsilon_j(i)$ be orthogonal to $u_j(m)$, $\forall m \neq i$. These conditions yield the representation equation

$$-A_{j,j} u_j = \sum_{n<j} A_{j,n} u_n + \mathbf{e}_j, \quad A_{j,j} \triangleq -I + A_{j,j}^0, \quad \forall j$$

$$E u_k \mathbf{e}_j^T = B_j \delta_{j,k}, \quad k \leqslant j$$

$$E \mathbf{e}_k \mathbf{e}_j^T = E \mathbf{e}_j \mathbf{e}_k^T = -A_{j,j} B_j \delta_{j,k}, \quad \forall j, k \tag{A14}$$

where

$$B_j = \text{Diag}\,\{E\epsilon_{i,j}^2\} \triangleq \text{Diag}\,\{\beta_{i,j}^2\}.$$

From above $-A_{j,j}B_j$ is a covariance matrix and must therefore be symmetric. Since B_j is diagonal, $A_{j,j}$ could not be lower or upper triangular.[6] Also, $\{\mathbf{e}_j\}$ is a white-noise vector sequence. However, $\{\epsilon_{i,j}\}$ need not be a white-noise field because the elements of $\mathbf{e}_j$ could be mutually correlated. Recall that in the case of causal MVR's, $\{\epsilon_{i,j}\}$ was a white-noise field. Now (A14) can be written as a block lower triangular matrix equation

$$-\mathcal{A} u = \mathbf{e} \qquad \mathcal{A} = \{A_{j,n}\} \tag{A15}$$

where $A_{j,j}$ is no longer lower triangular (compare with (A3)). Once again the $\mathcal{A}$ satisfies an equation similar to (A4), the difference being $\mathcal{R}_{\mathbf{e}}$ would now be block diagonal and $\mathcal{A}$ is block lower triangular (rather than also being lower triangular). Now, the matrices $A_{j,n}$ could be found by a single stage lower-upper factorization by solving

$$-\sum_{n=1}^{j} A_{j,n} R_{n,k} = B_j \delta_{j,k}, \quad k \leqslant j. \tag{A16}$$

The solution is given by

$$A_{j,n} = A_{j,j} \hat{A}_{j,n}$$

$$A_{j,j} = -B_j H_j \qquad H_j \triangleq \left[\sum_{n=1}^{j} \hat{A}_{j,n} R_{n,j}\right]^{-1}$$

$$B_j(i,i) \triangleq \beta_{i,j}^2 = 1/H_j(i,i) \tag{A17}$$

where $\hat{A}_{j,n}$ is the solution of (A6).

It is now worth noting that a causal MVR could always be obtained from a semicausal MVR whenever the covariance matrix $R_j^{\epsilon} = E\mathbf{e}_j \mathbf{e}_j^T = -A_{j,j}B_j$ is positive definite so that it has a factorization

$$R_j^{\epsilon} = L_j^T \mathbf{\Gamma}_j^{-1} L_j \tag{A18}$$

where $\mathbf{\Gamma}_j = \{\gamma_j(k)\}$, is a diagonal matrix and L_j is a unit-lower triangular matrix. Then the first equation in (A14) can be written as

$$-\tilde{A}_{j,j} u_j = \sum_{n<j} \tilde{A}_{j,n} u_n + \nu_j \tag{A19}$$

where

$$\tilde{A}_{j,j} \triangleq B_j L_j B_j^{-1} \qquad \tilde{A}_{j,n} \triangleq B_j \mathbf{\Gamma}_j (L_j^T)^{-1} A_{j,n}$$

$$\nu_j \triangleq B_j \mathbf{\Gamma}_j (L_j^T)^{-1} \mathbf{e}_j. \tag{A20}$$

From these it follows that $\tilde{A}_{j,j}$ is also a unit-lower triangular matrix and $\{\nu_{i,j}\}$ is a white-noise random field since

$$R_j^{\nu} = E\nu_j \nu_j^T = B_j \mathbf{\Gamma}_j B_j \tag{A21}$$

is a diagonal matrix. Comparison of (A19) with (A2) shows that it is now a causal MVR. Given a semicausal MVR, a causal MVR is guaranteed provided the lower-upper factorization of (A18) exists, i.e., if R_j^{ϵ} is positive definite. However, it may not always be desirable to get to a causal representation because the spatial structure of this representation may be-

[5] This means the infinite predictor coefficients $a_{m,n}$ and the covariance sequence $r(m, n)$ are unique with respect to one another.

[6] Except when $A_{j,j}$ is diagonal. In that case however, equation (A14) will reduce to a causal model and would not be admissible here.

come more cumbersome than the semicausal case as shown by Example 7 in Section V-C.

In the case of stationary semicausal MVR's defined on the infinite plane the equations corresponding to (A16) and (A17) become

$$-A_0 u_j = \sum_{n=1}^{\infty} A_n u_{j-n} + \varepsilon_j \qquad E\varepsilon_j \varepsilon_k^T = -\beta^2 A_0 \delta_{j,k} \quad \text{(A22)}$$

$$-\sum_{n=1}^{q} \hat{A}_n^q R_{n-j} = R_{-j}, \qquad 1 \leqslant j \leqslant q, \qquad q = 1, 2, \cdots$$

$$-A_0^q = \beta^2 \left[R_0 + \sum_{n=1}^{q} \hat{A}_n^q R_n \right]^{-1}$$

$$A_n^q = A_0^q \hat{A}_n^q, \qquad \lim_{q \to \infty} A_n^q = A_n \quad \text{(A23)}$$

where $-A_0$ now is a symmetric, positive definite Toeplitz matrix. Written in terms of Z_1-transform variables defined in (A11) with

$$a_0(z_1) = \sum_{m=-\infty}^{\infty} a_{m,0} z_1^{-m} \quad \text{(A24)}$$

we obtain, again, the scalar qth-order AR model equations

$$-\sum_{n=1}^{q} \hat{a}_n^q(z_1) r_{n-j}(z_1) = r_j(z_1^{-1}), \qquad j = 1, \cdots, q \quad \text{(A25)}$$

and

$$-a_0^q(z_1) = \beta^2 h(z_1), \quad h(z_1) \triangleq \left[r_0(z_1) + \sum_{n=1}^{q} \hat{a}_n^q(z_1) r_n(z_1) \right]^{-1}$$

$$\beta^2 = 1 \Big/ \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} h(z_1)\, d\omega \right), \qquad z_1 = e^{j\omega},$$

$$S_\epsilon(z_1, z_2) = -\beta^2 a_0^q(z_1), \qquad a_n^q(z_1) = a_0^q(z_1) \hat{a}_n^q(z_1),$$

$$a_n(z_1) = \lim_{q \to \infty} a_n^q(z_1). \quad \text{(A26)}$$

For a positive and analytic SDF, a unique semicausal MVR such as (A22) can be shown to exist and its coefficients could be identified via (A23) or equivalently via (A25) and (A26).

A.3 Noncausal Representations

For noncausal MVR's, the orthogonality condition gives

$$r(i,j;p,q) - \sum_{(m,n) \neq (i,j)} \sum a(i,j;m,n) r(m,n;p,q) = \beta_{i,j}^2 \delta_{i,p} \delta_{j,q}.$$

Define

$$a(i,j;i,j) = -1, A_{j,n} = \{a(i,j;m,n), \forall i,m\}, \; B_j = \text{Diag}\,\{\beta_{i,j}^2, \forall i\}.$$

Using these we obtain the matrix equations similar to (50)

$$-\sum_n A_{j,n} R_{n,q} = B_j \delta_{j,q}, \quad \text{or} \quad -\mathcal{A}\mathcal{R} = \mathcal{B},$$

$$\text{or} \quad -\mathcal{A} = \mathcal{B}\mathcal{R}^{-1} \quad \text{(A27)}$$

where $\mathcal{A}$, $\mathcal{R}$, and $\mathcal{B}$ are block matrices defined as

$$\mathcal{A} = \{A_{j,n}, \forall j, n\}, \qquad \mathcal{R} = \{R_{j,n}, \forall j, n\}, \qquad \mathcal{B} = \text{Diag}\,\{B_j, \forall j\}.$$

From this, we can obtain, as in (54)

$$\beta_{i,j}^2 = 1/[(\mathcal{R}^{-1})_{i,i}]_{j,j}. \quad \text{(A28)}$$

Thus $\beta_{i,j}^2$ is obtained by finding the jth diagonal block of $\mathcal{R}^{-1}$. Then the ith diagonal term of this matrix block equals $\beta_{i,j}^{-2}$. $\mathcal{A}$ is now obtained directly from (A27).

The orthogonality condition also requires

$$Eu_{i,j}\epsilon_{p,q} = \beta_{i,j}^2 \delta_{i,p} \delta_{j,q}, \qquad \forall i, j$$

which gives

$$E\epsilon_{i,j}\epsilon_{p,q} = r_\epsilon(i,p;j,q) = \begin{cases} \beta_{i,j}^2, & (i,j) = (p,q) \\ -a(i,j;p,q)\beta_{i,j}^2, & (i,j) \neq (p,q). \end{cases} \quad \text{(A29)}$$

Hence, the noncausal MVR of an arbitrary Gaussian field $\{u_{i,j}\}$ is given by

$$u_{i,j} = \sum_{(m,n) \neq (i,j)} \sum a(i,j;m,n) u_{m,n} + \epsilon_{i,j} \quad \text{(A30)}$$

where $\{\epsilon_{i,j}\}$ is a zero-mean *nonwhite* Gaussian random field whose covariances are given by (A29). For stationary random fields defined on the infinite plane, the two-dimensional noncausal MVR is obtained by a direct extension of the one-dimensional result (see (55) and (56)) as

$$r_{0,0}^+ u_{i,j} + \sum_{\substack{k=-\infty \\ (k,l) \neq (0,0)}}^{\infty} \sum_{l=-\infty}^{\infty} r_{k,l}^+ u_{i-k,j-l} = r_{0,0}^+ \epsilon_{i,j} \quad \text{(A31)}$$

$$r_{k,l}^+ = \frac{1}{2\pi} \int_{-\pi}^{\pi} S^{-1}(\omega_1, \omega_2) \exp\{j(k\omega_1 + l\omega_2)\}\, d\omega_1\, d\omega_2, \qquad \forall k, l \quad \text{(A32)}$$

$$E\epsilon_{i,j}\epsilon_{i+k,j+l} = r_\epsilon(k,l) = \begin{cases} 1/r_{0,0}^+, & (k,l) = (0,0) \\ r_{k,l}^+/(r_{0,0}^+)^2, & (k,l) \neq (0,0). \end{cases} \quad \text{(A33)}$$

A rational approximation of $S^{-1}(z_1, z_2)$ will yield a finite-order (approximate) realization. Clearly factorization of S is not required in the case of noncausal MVR's. Thus if S^{-1} is a two-dimensional polynomial

$$S^{-1}(z_1, z_2) = \sum_{k=-p}^{p} \sum_{l=-q}^{q} r_{k,l}^+ z_1^{-p} z_2^{-q} \quad \text{(A34)}$$

then the noncausal MVR is

$$\sum_{k=-p}^{p} \sum_{l=-q}^{q} r_{k,l}^+ u_{i-k,j-l} = r_{0,0}^+ \epsilon_{i,j} \quad \text{(A35)}$$

where $\{\epsilon_{i,j}\}$ is a zero mean moving average field with SDF

$$S_\epsilon(z_1, z_2) = \left(\sum_{k=-p}^{p} \sum_{l=-q}^{q} r_{k,l}^+ z_1^{-p} z_2^{-q} \right) \Big/ (r_{0,0}^+)^2. \quad \text{(A36)}$$

REFERENCES

[1] A. K. Jain, *Multidimensional Techniques in Digital Image Processing.* To be published.

[2] W. K. Pratt, *Digital Image Processing.* New York: Wiley, 1978.

[3] N. Ahmed and K. R. Rao, *Orthogonal Transforms for Digital Signal Processing.* New York: Springer Verlag, 1975.

[4] E. O. Brigham, *The Fast Fourier Transform.* Englewood Cliffs, NJ: Prentice-Hall, 1974.

[5] H. F. Harmuth, *Transmission of Information by Orthogonal Signals.* New York: Springer Verlag, 1970.

[6] H. C. Andrews, *Computer Techniques in Image Processing.* New York: Academic Press, 1970.

[7] H. C. Andrews and B. R. Hunt, *Digital Image Restoration.* Englewood Cliffs, NJ: Prentice-Hall, 1977.

[8] A. Rosenfeld and A. Kak, *Digital Picture Processing.* New York: Academic Press, 1976.

[9] A. K. Jain, "Some new techniques in image processing," in *Proc. Symp. Current Math Problems in Image Science* (Naval Post Graduate School, Monterey, CA). North Hollywood, CA: Western Periodicals Co., Nov. 1976.

[10] A. K. Jain, "A sinusoidal family of unitary transforms," *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol. PAMI-1, pp. 356–365, Oct. 1979.

[11] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, pp. 417–441, and 498–520, 1933.

[12] S. Watanabe, "Karhunen-Loeve expansion and factor analysis, theoretical remarks and applications," *Trans. 4th Prague Conf. Inform. Theory, Statist. Decision Functions, and Random Processes* (Prague, Czechoslovakia) pp. 635–660, 1965.

[13] H. P. Kramer and M. V. Mathews, "A linear coding for transmitting a set of correlated signals," *IRE Trans. Inform. Theory*, vol. IT-2, pp. 41–46, Sept. 1956.

[14] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Trans. Computers*, vol. C-23, pp. 90–93, Jan. 1974.

[15] M. Hamidi and J. Pearl, "Comparison of the cosine and Fourier transforms of Markov-1 signals," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-24, pp. 428–429, Oct. 1976.

[16] J. R. Jain and A. K. Jain, "Interframe adaptive data compression techniques for images," Tech. Rep. SIPL-79-2, Signal and Image Processing Lab., Dep. Elec., Comput. Eng., Univ. California, Davis, Aug. 1979.

[17] A. K. Jain, "Image data compression—A review," *Proc. IEEE*, vol. 69, pp. 349–389, Mar. 1981.

[18] ——, "A fast Karhunen Loeve transform for a class of random processes," *IEEE Trans. Commun* , vol. COM–24, pp. 1023–1029, Sept. 1976.

[19] A. K. Jain, S. H. Wang, and Y. Z. Liao, "Fast Karhunen Loeve transform data compression studies," presented at Nat. Telecommun. Conf., Dallas, TX, Nov.–Dec. 1976.

[20] G. E. Forsythe and P. Henrici, "The cyclic Jacobi method for computing the principal values of a complex matrix," *Proc. Amer. Math. Soc.*, vol. 94, pp. 1–23, 1960.

[21] G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," *Numer. Math.*, vol. 14, pp. 403–420, 1970.

[22] S. Treitel and J. L. Shanks, "The design of multistage separable planar filters," *IEEE Trans. Geosci. Electron.*, vol. GE-9, pp. 10–27, Jan. 1971.

[23] T. S. Huang, W. F. Schreiber, and O. J. Tretiak, "Image processing," *Proc. IEEE*, vol. 59, pp. 1586–1609, Nov. 1971.

[24] M. M. Sondhi, "Image restoration: The removal of spatially invariant degradations," *Proc. IEEE*, vol. 60, pp. 842–853, July 1972.

[25] H. C. Andrews, "Two dimensional transforms," in *Picture Processing and Digital Filtering*, T. S. Huang, Ed. (Topics in Applied Physics Series, vol. 6), Berlin, Germany: Springer-Verlag, 1975.

[26] H. C. Andrews and C. L. Patterson, "Singular value decomposition (SVD) image coding," *IEEE Trans. Commun.*, vol. COM-24, pp. 425–432, Apr. 1976.

[27] R. E. Twogood and S. K. Mitra, "Computer aided design of separable two-dimensional digital filters," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-25, pp. 165–169, Apr. 1977.

[28] V. C. Klema and A. J. Laub, "The singular value decomposition: Its computation and some applications," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 164–176, Apr. 1980.

[29] B. Ashjari and W. K. Pratt, "Supervised classification with singular value decomposition texture," USCIPI Rep. 860, Image Proc. Inst., USC, Los Angeles, CA. Mar. 1979.

[30] P. Whittle, *Prediction and Regulation by Linear Least-Squares Methods.* London, England: English Univ. Press, 1954.

[31] K. Astrom, *Introduction to Stochastic Control Theory.* New York: Academic Press, 1970.

[32] N. Levinson, "The Wiener RMS error criterion in filter design and prediction," *J. Math. Phys.*, vol. 25, pp. 261–278, Jan. 1947.

[33] J. Durbin, "The filtering of time series models," *Rev. Int. Inst. Stat.*, vol. 28, pp. 233–244, 1960.

[34] J. Burg, "Maximum entropy spectral analysis," Ph.D. dissertation, Dep. Geophysics, Stanford Univ., Stanford, CA, 1975.

[35] S. Zohar, "Toeplitz matrix inversion: The algorithm of W. F. Trench," *J. Assoc. Comput. Mach.*, vol. 16, pp. 592–601, Oct. 1969.

[36] T. J. Ulrych and T. N. Bishop, "Maximum entropy spectral analysis and autoregressive decomposition," *Rev. Geophys. Space Phys.*, vol. 13, pp. 183–200, Feb. 1975.

[37] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, pp. 561–580, Apr. 1975.

[38] A. K. Jain and S. Ranganath, "Image coding by autoregressive synthesis," *Proc. IEEE ICASSP '80* (Denver, CO), pp. 770–773, Apr. 1980.

[39] N. E. Nahi and T. Assefi, "Bayesian recursive image estimation," *IEEE Trans. Comput.* (Short Notes), vol. C-21, pp. 734–738, July 1972.

[40] T. Assefi, "Two dimensional signal processing with application to image restoration," Tech. Rep. 32-1596, JPL, Caltech, Pasadena, CA, Sept. 1, 1974.

[41] N. E. Nahi and C. A. Franco, "Application of Kalman filtering to image enhancement," in *Proc. IEEE Conf. Decision and Control* (New Orleans, LA), pp. 63–65, Dec. 1972.

[42] S. R. Powell and L. M. Silverman, "Modeling of two dimensional covariance functions with application to image restoration," *IEEE Trans. Automat. Cont.*, vol. AC-19, pp. 8–12, Feb. 1974.

[43] N. E. Nahi, *Estimation Theory and Applications.* New York: Wiley, 1969.

[44] A. H. Jazwinsky, *Stochastic Processes and Filtering Theory.* New York: Academic Press, pp. 70–92, 1970.

[45] A. O. Aboutalib and L. M. Silverman, "Restoration of motion degraded images," *IEEE Trans. Circuits Syst.*, vol. CAS-22, pp. 278–286, Mar. 1975.

[46] A. K. Jain, "Noncausal representations for finite discrete signals," in *Proc. IEEE Conf. Decision and Control* (Tucson, AZ), 1974.

[47] A. K. Jain and E. Angel, "Image restoration, modeling and reduction of dimensionality," *IEEE Trans. Comput.*, vol. C-23, pp. 470–476, May 1974.

[48] A. K. Jain, "A fast Karhunen-Loeve transform for recursive filtering of images corrupted by white and colored noise," *IEEE Trans. Comput.*, vol. C-26, pp. 560–571, June 1977.

[49] ——, "An operator factorization method for restoration of blurred images," *IEEE Trans. Comput.*, vol. C-25, pp. 1061–1071, Nov. 1977.

[50] ——, "Image coding via a nearest neighbors image model," *IEEE Trans. Commun.*, vol. COM-23, pp. 318–331, Mar. 1975.

[51] ——, "Partial differential equations and finite difference methods in image processing, Part I–Image representation," *J. Optimization Theory Appl.*, vol. 23, no. 1, pp. 65–91, Sept. 1977.

[52] A. K. Jain and J. R. Jain, "Partial differential equations and finite difference methods in image processing, Part II: Image restoration," *IEEE Trans. Automat. Contr.*, vol. AC-23, pp. 817–834, Oct. 1978.

[53] A. Zvi Meiri, "The pinned Karhunen Loeve transform of a two dimensional Gauss-Markov field," in *Proc. 1976 SPIE Meeting* (San Diego, CA), 1976.

[54] A. K. Jain, "Fast inversion of banded Toeplitz matrices circular decompositions," *IEEE Trans. ASSP*, vol. ASSP-26, pp. 121–126, Oct. 1978.

[55] ——, "A semicausal model for recursive filtering of two dimensional images," *IEEE Trans. Comput.*, vol. C-26, pp. 343–350, Apr. 1977.

[56] T. L. Marzetta, "A linear prediction approach to two dimensional spectral factorization and spectral estimation," Ph.D. dissertation, Dep. Elec. Eng. Comput. Sci., MIT, Cambridge, MA, Feb. 1978.

[57] S. H. Wang, "Applications of stochastic models for image data compression," Ph.D. dissertation, Dep. Elec. Eng., SUNY Buffalo, NY, Sept. 1979; Also see S. H. Wang and A. K. Jain, Tech. Rep. SIPL-79-6, Signal and Image Processing Lab., Dep. Elec. Comput. Eng., Univ. California Davis, Sept. 1979.

[58] A. K. Jain and S. H. Wang, "Stochastic image models and hybrid coding," Final Rep., NOSC Contract N00953-77-C-003MJE, Dep. Elec. Eng., SUNY Buffalo, NY, Oct. 1977.

[59] A. K. Jain, "Spectral estimation and signal extrapolation in one and two dimensions," in *Proc. RADC Spectrum Estimation Workshop* (Rome, NY), pp. 195–214, Oct. 1979.

[60] E. Angel and A. K. Jain, "Frame to frame restoration of diffusion images," *IEEE Trans. Automat. Contr.*, vol. AC-23, pp. 850–855, Oct. 1978.

[61] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*. New York: Wiley, 1949.
[62] J. L. Doob, *Stochastic Processes*. New York: Wiley, 1953.
[63] C. L. Rino, "Factorization of spectra by discrete Fourier transforms," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 484-485, July 1970.
[64] M. P. Ekstrom and J. W. Woods, "Two dimensional spectral factorization with application in recursive digital filtering," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-24, pp. 115-128, Apr. 1976.
[65] T. S. Huang, "Stability of two dimensional recursive filters," *IEEE Trans. Audio Electroacoust.*, vol. AU-20, pp. 158-163, June 1972.
[66] *Proceedings IEEE*, vol. 65, Special Issue on Multidimensional Systems (Guest Editor N. K. Bose), June 1977.
[67] D. Goodman, "Some stability properties of two dimensional linear shift invariant filters," *IEEE Trans. Circuits Syst.*, vol. CAS-24, pp. 201-208, Apr. 1977.
[68] J. W. Woods, "Two dimensional discrete Markov fields," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 232-240, Mar. 1972.
[69] W. E. Larimore, "Statistical inference on stationary random fields," *Proc. IEEE*, vol. 65, pp. 961-970, June 1977.
[70] S. Sugimoto, H. Mizutani and T. Mizokawa, "Causality and recursive estimation of two dimensional random image fields," in *Proc. 8th SICE Symp. Control Theory*, pp. 145-150, Hachigi, Tokyo, 1979.
[71] H. Mizutani and S. Sugimoto, "Semicausal models and smoothing for 2-D random image fields," in *Proc. 11th JAACE Symp. Stoch. Syst.* (Tokyo, Japan), pp. 161-164, Nov. 1979.
[72] T. Kailath, D. O. Mayne, and R. K. Mehra, Eds., *IEEE Trans. Aut. Contr.*, Special Issue on System Identification and Time Series Analysis, vol. AC-19, Dec. 1974.
[73] B. R. Hunt, "The application of constrained least squares estimation to image restoration by digital computer," *IEEE Trans. Comput.*, vol. C-22, pp. 805-812, Sept. 1973.
[74] R. W. Means, E. H. Wrench, and H. J. Whitehouse, "Image transmission via spread spectrum techniques," ARPA Quart. Tech. Rep. ARPA-QR6, Nav. Ocean Syst. Cent., San Diego, CA, Jan.-Dec. 1975; Also see ARPA-QR8, Annu. Rep., Jan.-Dec. 1975.
[75] B. R. Hunt, "Digital image processing," *Proc. IEEE*, vol. 63, pp. 693-708, Apr. 1975.
[76] A. Habibi, "Two dimensional Bayesian estimate of images," *Proc. IEEE*, vol. 60, pp. 878-883, July 1972.
[77] M. Strintzis, "Comments on two dimensional Bayesian estimate of images," *Proc. IEEE*, vol. 64, pp. 1255-1257, Aug. 1976.
[78] N. E. Nahi and A. Habibi, "Decision directed recursive image enhancement," *IEEE Trans. Circuits Syst.*, vol. CAS-22, pp. 286-293, Mar. 1975.
[79] J. W. Woods and C. H. Radewan, "Kalman filtering in two dimensions," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 473-482, July 1977.
[80] M. S. Murphy and L. M. Silverman, "Image model representation and line by line recursive restoration," in *Proc. Conf. Decision Contr.* (Clearwater Beach, FL), pp. 601-606, Dec. 1976.
[81] A. Netravali and J. Limb, "Picture coding: A survey," *Proc. IEEE*, vol. 68, pp. 366-406, Mar. 1980.
[82] W. F. Schreiber, C. F. Knapp, and N. D. Kay, "Synthetic highs: An experimental TV bandwidth reduction system," *J. Soc. Motion Pic. Television Engrs.*, vol. 68, pp. 525-537, Aug. 1959.
[83] D. N. Graham, "Image transmission by two dimensional contour coding," *Proc. IEEE*, vol. 55, pp. 336-346, Mar. 1967.
[84] J. K. Yan and D. J. Sakrison, "Encoding of images based on a two component source model," *IEEE Trans. Commun.*, vol. COM-25, pp. 1315-1322, Nov. 1977.
[85] W. Rudin, "The extension problem for positive definite functions," *Ill. J. Math.*, vol. 7, pp. 532-539, 1963.
[86] B. W. Dickinson, "Two dimensional Markov spectrum estimates need not exist," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 120-121, Jan. 1980.
[87] R. P. Roesser, "A discrete state space model for linear image processing," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 1-10, Feb. 1975.
[88] S. Attasi, "Modelling and recursive estimation for double indexed sequences," IRIA Rep. IA/129, Domaine de Volvceau, Racquencourt, 78150 Le Chesnay, B.P. 5, France, July 1975.
[89] E. Fornasini and G. Marchesini, "State space realization theory of two dimensional filters," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 484-492, Aug. 1976.
[90] E. Wong, "Recursive causal linear filtering for two dimensional random fields," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 50-59, Jan. 1978.
[91] M. Morf, B. Levy, and S. Y. Kung, "New results in 2-D systems theory, Part I: 2-D polynomial matrices, factorization and coprimeness," *Proc. IEEE*, vol. 65, pp. 861-872, June 1977.
[92] S. Y. Kung, B. C. Levy, M. Morf, and T. Kailath, "New results in 2-D systems theory, Part II: Realization and the notions of controllability, observability and minimality," *Proc. IEEE*, vol. 65, pp. 945-960, June 1977.
[93] A. K. Jain and K. W. Au, "On linear estimation via fast Fourier transform," in *Proc. Conf. Decision and Control* (Ft. Lauderdale, FL), Dec. 1979.
[94] J. W. Woods, "Two dimensional Markov spectral estimation," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 552-559, Sept. 1976; also see vol. IT-26, pp. 129-130, Jan. 1980.
[95] E. S. Angel and A. K. Jain, "Restoration of images degraded by spatially varying point spread functions by a conjugate gradient method," *Appl. Opt.*, vol. 17, pp. 2186-2190, July 1980.

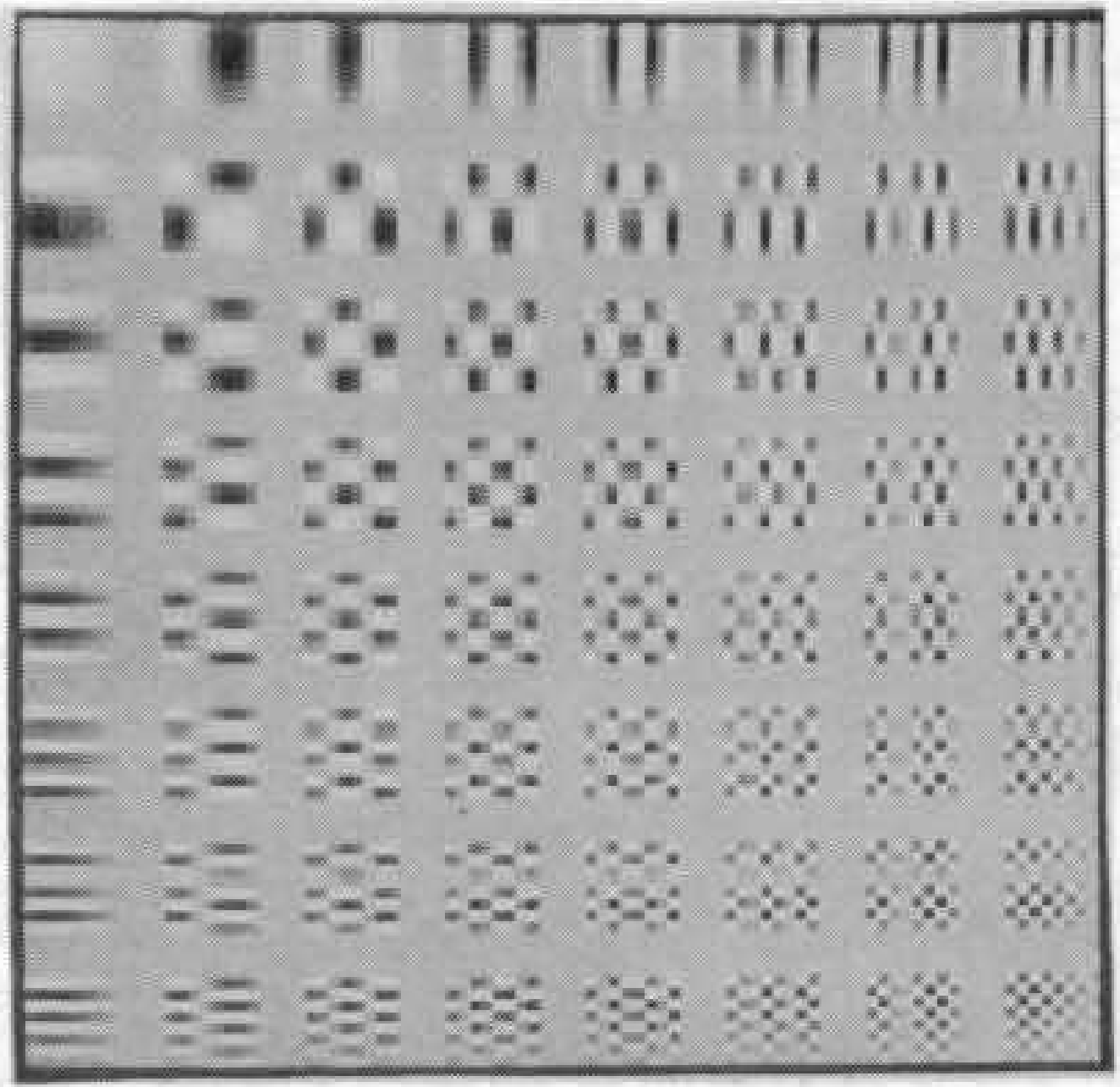
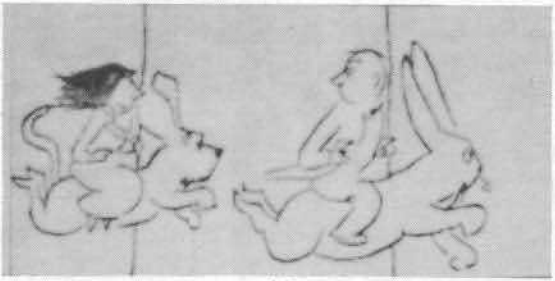
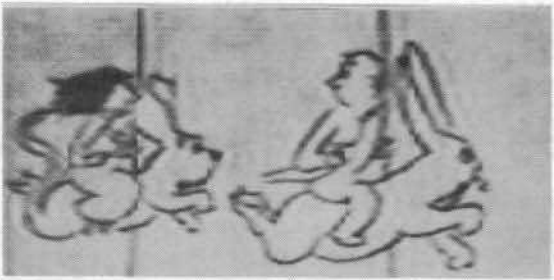


Fig. 1. 8 × 8 basis images of the DCT. Each 8 × 8 basis image if $B_{i,j}$ orthogonal to the rest and represents a spatially varying FIR of a system whose input is an impulse at (i, j).



(a)



(b)

(c)

Fig. 4. Line by line synthesis of an image by eighth-order AR models. Each line has been synthesized by line segments of 32 pixels. (a) Original. (b) Synthesis using Cosine transform, 4 to 1 sample reduction. (c) Synthesis using AR model, 4 to 1 sample reduction.

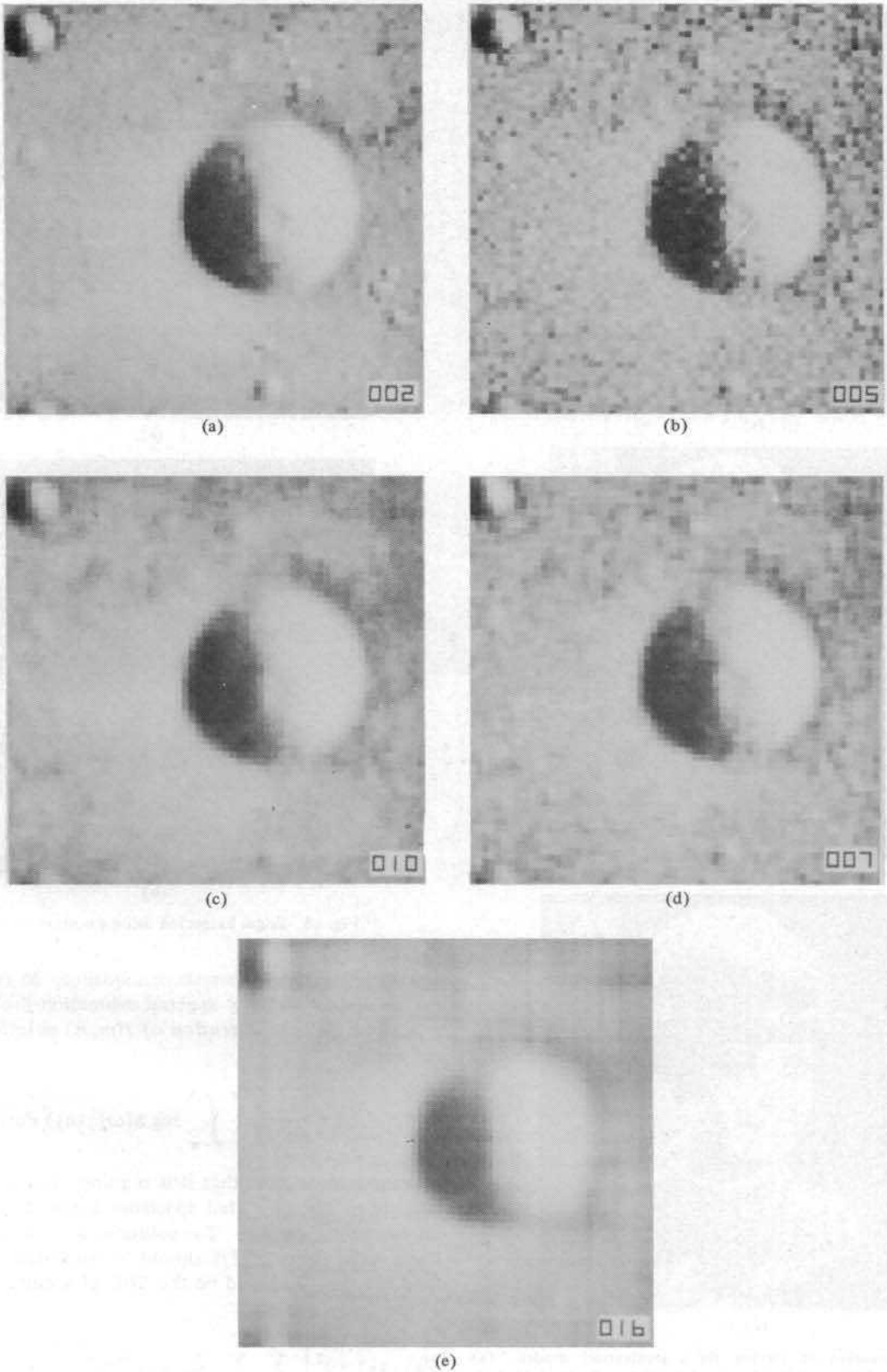


Fig. 16. Smoothing of a noisy image. (a) Original. (b) Noisy. (c) (d) Optimum filtered based on the separable covariance model.

(a)

(b)

(c)

Fig. 17. Data compression of images via a semicausal model. (a) Original. (b) Encoded. 1 bit/pel. (c) Adaptive codings, 1 bit/pel. In the adaptive method, the model parameters are updated after every 16×16 block of pels. Mean-square error improvement due to adaptation is in excess of 3 dB.

(a)

(b)

Fig. 18. Edges extracted using a noncausal model.

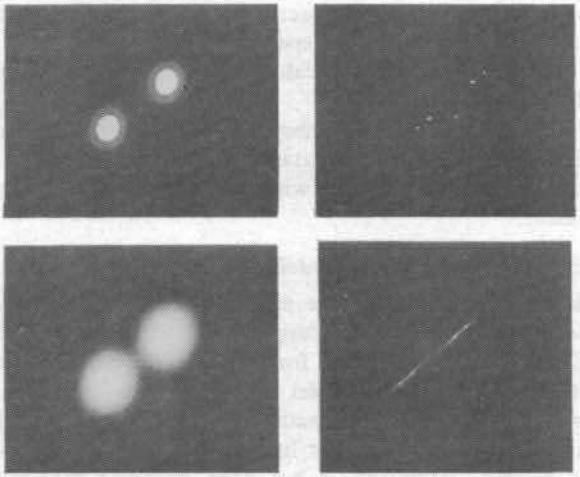


Fig. 19. Spectral estimation comparison of windowed DFT versus semicausal MVR methods. Top row; from left windowed DFT spectrum amplitude versus semicausal MVR spectrum amplitude. Bottom row; same quantities shown on dB scale.

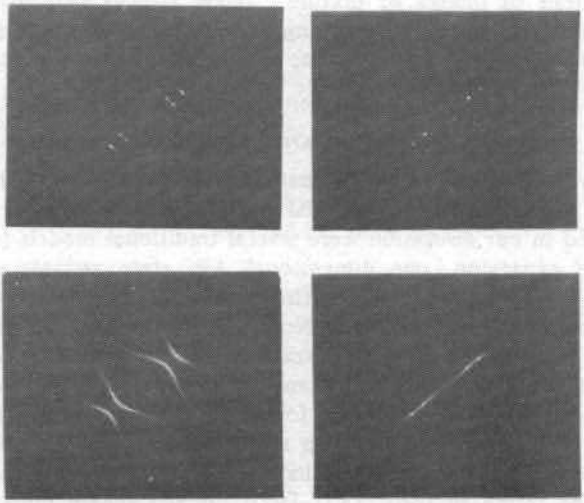


Fig. 20. Spectral estimation comparison of causal versus semicausal models. Top row from left; model spectrum amplitude versus semicausal model spectrum amplitude. Bottom row; same quantities shown on dB scale.