

Compound Gauss-Markov Random Fields for Image Estimation

Fure-Ching Jeng, *Member, IEEE*, and John W. Woods, *Fellow, IEEE*

Abstract—This paper is concerned with algorithms for obtaining approximations to statistically optimal estimates for images modeled as compound Gauss-Markov random fields. We consider both the maximum a posteriori probability (MAP) estimate and the minimum mean-squared error (MMSE) estimate for both image estimation and image restoration. Compound image models consist of several submodels having different characteristics along with an underlying structure model which governs transitions between these image submodels. Compound Gauss-Markov field models can be attractive for image estimation because the resulting estimates do not suffer the oversmoothing of edges that usually occurs with Gaussian image models.

Two different compound random field models are employed in this paper, the doubly stochastic Gaussian (DSG) random field and a newly defined compound Gauss-Markov (CGM) random field. We present MAP estimators for DSG and CGM random fields using simulated annealing, a powerful optimization method best suited to massively parallel processors. A fast converging algorithm called deterministic relaxation, which however converges to only a locally optimal MAP estimate, is also presented as an alternative for reducing computational loading on sequential machines. For comparison purposes, we also include results on the fixed-lag smoothing MMSE estimator for the DSG field and its suboptimal M -algorithm approximation.

The incorporation of causal and noncausal modeling together with causal and noncausal estimates on the same data sets allows meaningful visual comparisons to be made. We also include Wiener and reduced update Kalman filter (RUKF) estimates to allow visual comparison of the near optimal estimates based on compound Gauss-Markov models to those based on simple Gaussian image models.

I. INTRODUCTION

AS is well known, a linear-shift invariant (LSI) model is not well suited to image estimation, i.e., noise smoothing. The major problem is that the resulting image estimate suffers from oversmoothing of edges which are quite important to the human visual system (HVS). Partly as a result, space-variant filters have been proposed [1], [8], [12], [14] wherein each pixel (or block of pixels) is filtered by a switched linear filter with the switching governed by a visibility function. The visibility function can be based on either a local variance estimate [8] or an average estimate of the isotropic gradient [1]. A more complicated visibility function which depends upon the mix of the directional components and a high frequency component was used in [12]. One of the problems with all these approaches is that the filter switching is rather ad hoc and cannot be justified on a theoretical basis. It would be better if the filter transitions could be related to an image model.

Manuscript received February 9, 1988; revised April 23, 1990. This work was supported by the National Science Foundation under Grant ECS-8313880.

F.-C. Jeng is with Bellcore, Morristown, NJ 07960, and was visiting at the Center for Automation Research, University of Maryland at College Park, College Park, MD 20742, while this work was being done.

J. W. Woods is with the ECSE Department, Rensselaer Polytechnic Institute, Troy, NY 12180-3590.
IEEE Log Number 9041598.

One way to embed a mathematical approach into the above adaptive methods is to have the models' switching controlled by a hidden random field. Toward this end, a class of models called compound Gauss-Markov models has been proposed [5], [15]. A compound random field consists of two levels, upper and lower. The upper level random field is the observed image, which is composed of several LSI submodels representing a variety of local characteristics of images, e.g., edge orientation or texture. The lower or hidden level is a random field with a finite range whose purpose is to govern transitions between the observed LSI submodels.

The doubly stochastic Gaussian (DSG) field [15] possesses a conditional Gaussian, autoregressive (AR) upper or observed level, whose model coefficients have a causal support. They are switched by a hidden (hence lower level) 2-D causal Markov chain to generate the required local edge structure. The DSG random field is a Markov random field but the observed upper level component, by itself, is not, i.e., the image data component of the DSG is not Markov. In [15] a 2-D fixed-delay causal estimator called the M algorithm is presented for approximation to the minimum mean-squared error (MMSE) optimal estimate.

Geman and Geman, on the other hand, used a noncausal neighborhood system in their image model [5]. In their nomenclature, our lower level random field becomes their line process which was defined on an interpixel grid system. Their upper level random field is a noncausal conditional Markov random field also with a finite range, i.e., a noncausal 2-D Markov chain. Since they modeled images on a finite range space, they thereby excluded the Gaussian models, which are very widely used in image estimation and restoration. Generalizing their model, we have introduced an extension of the Gauss-Markov random field [10], [17]. Our compound Gauss-Markov (CGM) model [6] has continuous grey levels as an upper level observations model but retains the Geman's lower level line process (field) as a structural model.

As mentioned above, both DSG and CGM models will be used in this paper, both for estimation and restoration. Their MAP estimators will be developed by means of simulated annealing (stochastic relaxation) which was first introduced by Geman and Geman [5] where they presented a convergence proof for this iterative method. Their proof was confined to images that are modeled by a finite range space which, unfortunately, excludes the case of CGM models. When applying simulated annealing to our models, a proof for the convergence of simulated annealing to the MAP estimate was thus lacking. We have obtained such a proof for a broad subclass of CGM models [7]. This paper will concentrate on the algorithm and experimental results of applying simulated annealing with CGM models to image estimation and restoration. Those interested in the more theoretical aspects of this work are referred to [7].

It is well known that simulated annealing is a computation-

ally demanding method. Alternatively a quickly converging algorithm called *deterministic relaxation* can be used to get a local MAP estimate. This deterministic relaxation method has been called iterated conditional mode (ICM) in [2]. We also present results for this locally convergent algorithm.

We start by defining the two compound random fields. Then estimation problems are formulated and the various estimators are briefly described. Next, we present simulation results that provide comparisons between the various methods, i.e., causal versus noncausal model, compound Gaussian versus simple Gaussian, noncausal MAP versus causal *M*-algorithm, and simulated annealing versus deterministic relaxation. The intent here will be to see which algorithms result in subjectively good images. We then draw conclusions about which mathematical models best match this human visual system (HVS) viewpoint.

II. COMPOUND MARKOV RANDOM FIELDS

As described in the Introduction, a simple shift-invariant Gaussian AR model can often lead to oversmoothing of image data. In this section, we will present two compound Markov models, the noncausal CGM and the causal DSG, that address this problem.

A. Compound Gauss-Markov Random Fields

We can generalize the Gemans' compound model by incorporating a conditional Gauss-Markov observational model [10], [17] with their line field, as indicated below. A simple Gauss-Markov (GM) model can be described as follows:

$$s(m, n) = \sum_{kl \in \mathcal{R}} c_{kl} s(m - k, n - l) + w(m, n) \quad (1)$$

where $w(m, n)$ is a Gaussian random field satisfying the following covariance constraint:

$$E[w(m, n)w(k, l)] = \begin{cases} \sigma_w^2 & \text{if } (m, n) = (k, l) \\ -c_{m-k, n-l} \sigma_w^2 & \text{if } (m - k, n - l) \in \mathcal{R} \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

and \mathcal{R} is shown in Fig. 1. A random field described by the above equation is a Markov random field with a neighborhood support \mathcal{R} [10], [17]. By the equivalence of Markov random fields and Gibbs distributions, we can write the joint probability density function (pdf) $p(S)$ as the Gibbs pdf

$$p(S) = \frac{1}{Z_1} e^{-U_s(S)} \quad (3)$$

and

$$U_s(S) \triangleq \sum_{c_s \in C_s} V_{c_s}(S)$$

where c_s is a clique, C_s denotes the clique system for the given Markov neighborhood system, and Z_1 is a normalizing constant which is functionally independent of the matrix S , with elements $s_{ij} \triangleq s(i, j)$. Each $V_{c_s}(S)$ is a function on the sample space with the property that $V_{c_s}(S)$ involves only those $s(m,$

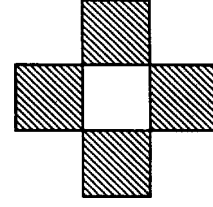


Fig. 1. Coefficient support region \mathcal{R} of the first-order model.

$n)$ of S for which $(m, n) \in c_s$. In the above GM model, we have

$$V_{c_s}(S) = \frac{s^2(m, n)}{2\sigma_w^2} \quad \text{or} \quad -\frac{c_{m-k, n-l} s(m, n) s(k, l)}{\sigma_w^2} \quad \text{if } (m - k, n - l) \in \mathcal{R}.$$

A CGM model consists of several conditionally Gauss-Markov submodels with an underlying structure or line field. Here

$$s(m, n) = \sum_{kl \in \mathcal{R}} c_{kl}^{l(m, n)} s(m - k, n - l) + w^{l(m, n)}(m, n) \quad (4)$$

where $w^{l(m, n)}(m, n)$ is a conditionally Gaussian noise whose variance controlled by the $l(m, n)$ and $l(m, n)$ is a vector which consists of four nearest neighbors of the line field surrounding the pixel $s(m, n)$, with $l(m, n)$ denoting the Gemans' line field. This line field takes on two values indicating whether a bond is broken or not. If a bond is broken between adjacent pixels, then there is little covariance between them, otherwise, a strong covariance exists.

The joint mixed probability density function (mpdf) of S, L is the Gibbs distribution

$$p(S, L) = \frac{1}{Z_2} e^{-[U_s(S|L) + U_l(L)]/T} \quad (5)$$

where the constant Z_2 is independent of S, L . Also T is a temperature parameter, U_s is defined in (3), and U_l is of the form

$$U_l(L) = \sum_{c_l \in C_l} V_{c_l}(L)$$

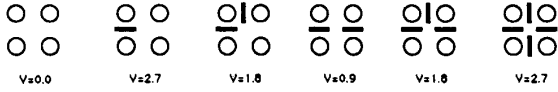
where the $N' \times N'$ matrix L has elements $l_{ij} \triangleq l(i, j)$, and where N'^2 is the total number of points in the line field over the finite observation region.

The clique system c_l used in this paper is the same system used in [5] as shown in Fig. 2. For a CGM model to have a valid conditional Markov random field given a realization of the line field L , there are covariance constraints on $w^{l(m, n)}(m, n)$:

$$E[w^{l(m, n)}(m, n)w^{l(k, l)}(k, l)] = \begin{cases} \sigma_{w^{l(m, n)}(m, n)}^2 & \text{if } (m, n) = (k, l) \\ -c_{m-k, n-l}^{l(k, l)} \sigma_{w^{l(m, n)}(m, n)}^2 & \text{if } (m - k, n - l) \in \mathcal{R} \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

By commutativity of covariance of two random variables, we have the following constraint on the model coefficients:

$$c_{m-k, n-l}^{l(k, l)} \sigma_{w^{l(m, n)}(m, n)}^2 = c_{k-m, l-n}^{l(m, n)} \sigma_{w^{l(k, l)}(k, l)}^2 \quad \text{if } (m - k, n - l) \in \mathcal{R}. \quad (7)$$

Fig. 2. The clique system c_l .

This constraint reduces the total number of free parameters that can be used in parameter identification. Consequently, classical unconstrained least squares parameter identification is not appropriate.

Since the CGM model has a Gibbs probability distribution, it is a Markov random field. Depending on the functional $U_s(S|L)$, the hidden random field L , by itself, may or may not be a Markov random field. This can be seen from the following expression for the probability function $P(L)$:

$$\begin{aligned} P(L) &= \frac{1}{Z_2} \int_S e^{-[U_s(S|L) + U_l(L)]/T} dS, \\ &= \frac{e^{-U_l(L)/T}}{Z_2} \int_S e^{-U_s(S|L)/T} dS. \end{aligned}$$

If $\int_S e^{-U_s(S|L)/T} dS$ has the form of Gibbs distribution for L , then L is a Markov random field, otherwise it is not. In general, $\int_S e^{-U_s(S|L)/T} dS$ may not be a Gibbs distribution for L .

B. Doubly Stochastic Gaussian Random Fields

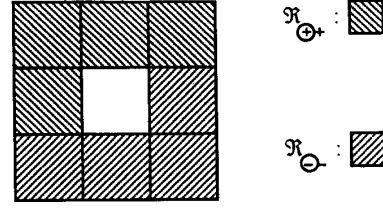
In the following, we will give a brief description of the DSG model. Before we describe this model, we need to introduce the 2-D causal Markov chain and a corresponding AR random field [15].

Definition: The 2-D Markov chain $l(m, n)$ is called a causal Markov chain if the following condition is satisfied:

$$\begin{aligned} P[l(m, n) | l(i, j), (i, j) \in (i, j) < (m, n)] \\ = P[l(m, n) | l(i, j) : (i, j) \in \mathcal{R}_{\oplus+}(m, n)] \end{aligned} \quad (8)$$

where $\mathcal{R}_{\oplus+}(m, n)$ is the local state support region shown in Fig. 3 and “ $(i, j) < (m, n)$ ” means “ $i < m$ or $i = m$ and $j < n$ ” which is consistent with progressive image scanning. In this paper, a 1×1 order Markov chain is used exclusively, i.e., $\mathcal{R}_{\oplus+}(m, n) = \{(m-1, n), (m+1, n-1), (m, n-1), (m-1, n-1)\}$. Thus we can write

$$\begin{aligned} P[l(m, n) | l(m-1, n)] \\ \triangleq P[l(m, n) | l(i, j) : (i, j) \in \mathcal{R}_{\oplus+}(m, n)]. \end{aligned}$$

Fig. 3. Coefficient support regions $\mathcal{R}_{\oplus+}(m, n)$ and $\mathcal{R}_{\ominus-}(m, n)$.

Definition: The random field $s(m, n)$ is called a causal AR random field if it can be described by

$$s(m, n) = \sum_{(k, l) \in \mathcal{R}_{\oplus+}} c_{kl} s(m-k, n-l) + w(m, n) \quad (9)$$

where $w(m, n)$ is a white noise field and $\mathcal{R}_{\oplus+} \triangleq \{(k, l) : (-k, -l) \in \mathcal{R}_{\oplus+}(0, 0)\}$.

We will use 1×1 order AR models exclusively in this paper.

Definition: A random field $s(m, n)$ is called a DSG random field [15] if it can be described as

$$s(m, n) = \sum_{(k, l) \in \mathcal{R}_{\oplus+}} c_{kl}^{l(m, n)} s(m-k, n-l) + w^{l(m, n)}(m, n) \quad (10)$$

where $l(m, n)$ is a causal Markov chain and the submodel coefficients at (m, n) in general depend upon $l(m, n)$.

The joint mpdf can be expressed as follows:

$$p(S, L) = p(S|L)P(L).$$

Note that the formulation of the joint mpdf of the DSG model is different from the CGM model. It is easily seen that, by construction, the line field of the DSG model itself is a Markov random field.

III. ESTIMATOR PROBLEM FORMULATIONS

We wish to estimate the original image $s(m, n)$ from the degraded observations $r(m, n)$ on an $N \times N$ region:

$$r(m, n) = \sum_{k, l \in \mathcal{R}_h} h(k, l) s(m-k, n-l) + v(m, n) \quad (11)$$

where $h(k, l)$ is a known point spread function (psf) with finite support region \mathcal{R}_h and $v(m, n)$ is zero-mean, white noise independent of $s(m, n)$ and having variance σ_v^2 . In the following context, two criteria for estimation will be discussed. We briefly specify the minimum mean-square error (MMSE) criteria as used in the fixed-delay causal estimator for the DSG model from [15]. Then we specify the maximum *a posteriori* probability (MAP) criteria as used in this work. Ultimately in Section VI, we will see some indication as to which estimation criteria is subjectively more relevant, i.e., which matches the HVS better.

A. Fixed-Lag Minimum Mean-Square Error Estimate

The object is to find the fixed-lag minimum mean-squared error (MMSE) estimate of $s(m-k, n-l)$ given the scanned set of all observations up to pixel (m, n) denoted

$$\begin{aligned} R(m, n) \triangleq \{ & r(1, 1), & r(1, 2) & \cdots, & r(1, N); \\ & r(2, 1), & r(2, 2), & \cdots, & r(2, N); \\ & \vdots, & \vdots, & \vdots, & \vdots; \\ & r(m-1, 1), & r(m-1, 2), & \cdots, & r(m-1, N); \\ & r(m, 1), & \cdots, & & r(m, n) \} \end{aligned}$$

where k, l represent fixed lag parameters (i.e., $k \geq 0, l \leq 0$). Due to the generalized causal nature of the DSG random field, this fixed-lag criterion is quite suitable. It has the advantage of enabling the processing of scanned image data as it arrives, i.e., without a frame store.

B. Maximum the A Posteriori Joint Density Estimate

The *a posteriori* joint mpdf for S and L given R can be decomposed as

$$p(S, L|R) = \frac{p(R|S, L)p(S, L)}{p(R)} = \frac{p(R|S)p(S|L)P(L)}{p(R)} \quad (12)$$

where the matrices S, L have been previously defined and R is defined as the $N \times N$ matrix with elements $r_{ij} \triangleq r(i, j)$.

Our object is to find the joint matrix estimate \hat{S} and \hat{L} such that

$$p(\hat{S}, \hat{L}|R) = \max_{S, L} p(S, L|R). \quad (13)$$

It can be seen that the MAP criteria leads to the need for a frame store to hold the entire image during processing. It also indirectly leads to iterative algorithms requiring massive parallelism for their efficient solution.

Interestingly, neither error criteria is clearly more relevant to image processing. For example, the MMSE estimate should yield an estimate that is closest to the observed data on average. And the MAP criteria, while currently more popular than MMSE, may conceivably yield an estimated image that is most probable given the model, but still very unlikely given the data.

IV. ESTIMATOR SOLUTIONS

In this section, we review a previously derived causal MMSE estimator for the DSG model and also present new noncausal MAP estimators [7] for both the CGM and DSG models. Then we present the suboptimal deterministic relaxation for the MAP estimate of both models.

A. Optimal MMSE Estimates

An optimal MMSE solution for the DSG model was derived in [15]. Using their notation, let $L_j(m, n)$ be the j th submodel sequence and $l_j(m, n)$ be the value of this sequence at the pixel (m, n) . Then

$$L_j(m, n) = \left\{ \begin{array}{cccc} l_j(1, 1), & l_j(1, 2), & \cdots, & l_j(1, N); \\ l_j(2, 1), & l_j(2, 2), & \cdots, & l_j(2, N); \\ \vdots, & \vdots, & \vdots, & \vdots; \\ l_j(m-1, 1), & l_j(m-1, 2), & \cdots, & l_j(m-1, N); \\ l_j(m, 1), & \cdots, & & l_j(m, n) \end{array} \right\}.$$

For the j th sequence, they wrote

$$\begin{aligned} \hat{s}_j[(m-k, n-l)|(m, n)] \\ \triangleq E[s(m-k, n-l)|R(m, n), L_j(m, n)]. \end{aligned}$$

Then the MMSE estimate was written as

$$\begin{aligned} \hat{s}[(m-k, n-l)|(m, n)] \\ = \sum_{j=1}^{L(n-1)N+m} \hat{s}_j[(m-k, n-l)|(m, n)] \\ \cdot P[L_j(m, n)|R(m, n)] \end{aligned} \quad (14)$$

where the *a posteriori* probability $P[L_j(m, n)|R(m, n)]$ was given as

$$\begin{aligned} P[L_j(m, n)|R(m, n)] \\ = kP[l_j(m, n)|l_j(m-1, n)] \\ \cdot P[r(m, n)|R(m-1, n), L_j(m, n)] \\ \cdot P[L_j(m-1, n)|R(m-1, n)] \end{aligned}$$

with k a normalizing constant and l_j a local state for the hidden Markov chain $l(m, n)$.

Notice that the number of terms in the summation in (14) grows exponentially. To reduce computation, we can keep a fixed number, say M , of paths at each pixel and discard the rest to control computations and storage space.

1) *Suboptimal Solution: M-Algorithm:* Instead of carrying forward the *a posteriori* probabilities $P[L_j(m, n)|R(m, n)]$ for all j , we now keep only the M most probable of the LM possible extensions and discard the rest. The *M*-algorithm can be formulated as follows:

$$\begin{aligned} \hat{s}[(m-k, n-l)|(m, n)] = \sum_{j=1}^M \hat{s}_j[(m-k, n-l)|(m, n)] \\ \cdot P[L_j(m, n)|R(m, n)] \end{aligned}$$

where the *a posteriori* probabilities are given by

$$\begin{aligned} P[L_j(m, n)|R(m, n)] \\ = k(m, n)P[l_j(m, n)|l_j(m-1, n)] \\ \cdot p[r(m, n)|R(m-1, n), L_j(m, n)] \\ \cdot P[L_j(m-1, n)|R(m-1, n)] \end{aligned}$$

with $j = 1, 2, \dots, M$ and $k(m, n)$ a normalizing constant.

B. Joint MAP Estimator—Simulated Annealing Approach

Since the criterion function $p(S, L|R)$ is nonlinear, it is extremely difficult to find the optimum solutions, \hat{S}, \hat{L} by any conventional method. In [5], a relaxation technique, called sto-

chastic relaxation or simulated annealing, is developed to search for MAP estimates from degraded observations. In the following, we will present a variation of simulated annealing appropriate for both CGM and DSG models.

Assume that $s(m, n)$ is either a CGM or a DSG random field and the observations $r(m, n)$ satisfy (11). Some notation is first

defined for the sake of conciseness:

$$S_{(m,n)} \triangleq \{s(k, l): \forall (k, l) \neq (m, n)\}$$

$$L_{(m,n)} \triangleq \{l(k, l): \forall (k, l) \neq (m, n)\}.$$

A) CGM Case: The conditional *a posteriori* density function $\pi^s(m, n)$ for $s(m, n)$ given $S_{(m,n)}$, L and R is

$$\begin{aligned} \pi^s(m, n) &\triangleq p[s(m, n)|S_{(m,n)}, L, R] = \frac{p(S, L, R)}{\int p(S, L, R) ds(m, n)} \\ &= \frac{p(R|S)p(S|L)P(L)}{\int p(R|S)p(S|L)P(L) ds(m, n)} \\ &= \frac{1}{Z_3} \exp \left(- \frac{\left(s(m, n) - \sum_{kl} c_{kl}^{l(m,n)} s(m-k, n-l) \right)^2}{2T\sigma_{w_{l(m,n)}}^2} \right. \\ &\quad \left. - \frac{\sum_{ij \in \mathcal{R}_h} \left(r(m+i, n+j) - \sum_{kl \in \mathcal{R}_h} h_{kl} s(m+i-k, n+j-l) \right)^2}{2T\sigma_v^2} \right) \end{aligned} \quad (15)$$

where Z_3 is a normalizing constant. The conditional *a posteriori* probability function $\pi^l(m, n)$ for $l(m, n)$ given $L_{(m,n)}$, S and R is

$$\begin{aligned} \pi^l(m, n) &\triangleq P[l(m, n)|L_{(m,n)}, S, R] = \frac{p(S, L, R)}{\sum_{l(m,n)} p(S, L, R)} \\ &= \frac{p(S, L)p(R|S)}{\sum_{l(m,n)} p(S, L)p(R|S)} \\ &= \frac{p(S, L)}{\sum_{l(m,n)} p(S, L)} \\ &= \frac{p(s(i, j), s(i', j')|L)p(S_{(i,j),(i',j')}|s(i, j), s(i', j'), L)P(L)}{\sum_{l(m,n)} p(s(i, j), s(i', j')|L)p(S_{(i,j),(i',j')}|s(i, j), s(i', j'), L)P(L)} \\ &= \frac{p(s(i, j), s(i', j')|L)P(L)}{\sum_{l(m,n)} p(s(i, j), s(i', j')|L)P(L)} \end{aligned} \quad (16)$$

where $l(m, n)$ is located between the two neighboring pixels $s(i, j)$, $s(i', j')$ and

$$S_{(i,j),(i',j')} \triangleq \{s(k, l): \forall (k, l) \neq (i, j) \text{ or } (i', j')\}.$$

The dependence of $p(s(i, j), s(i', j')|L)P(L)$ on $l(m, n)$ can be determined from (5). In fact, $\pi^l(m, n)$ can be explicitly written as follows:

$$\pi^l(m, n) = \frac{1}{Z_4} \exp \left(- \frac{\frac{s^2(i, j)}{2\sigma_{w_{l(i,j)}}^2} - \frac{c_{i-i', j-j'}^{l(i,j)}}{\sigma_w^2} s(i, j)s(i', j') + \frac{s^2(i', j')}{2\sigma_{w_{l(i',j')}}^2} + \sum_{(m,n) \in \mathcal{C}_l} V_{cl}(L)}{T} \right)$$

where Z_4 is a normalizing constant.

B) *DSG Case*: The conditional *a posteriori* density function $\pi^s(m, n)$ for $s(m, n)$ given $S_{(m,n)}$, L and R is

$$\begin{aligned} \pi^s(m, n) &= p[s(m, n) | S_{(m,n)}, L, R] \\ &= \frac{1}{Z_s} \exp \left[- \sum_{(i,j) \in \mathcal{R}_{o-}(m,n)} \left(\frac{s(i, j) - \sum_{kl} c_{kl}^{(i,j)} s(i-k, j-l)}{2T\sigma_{w(i,j)}^2} \right)^2 \right. \\ &\quad \left. - \sum_{ij \in \mathcal{R}_h} \left(\frac{r(m+i, n+j) - \sum_{kl \in \mathcal{R}_h} h_{kl} s(m+i-k, n+j-l)}{2T\sigma_v^2} \right)^2 \right] \end{aligned}$$

where Z_s is a normalizing constant and the support region $\mathcal{R}_{o-}(m, n)$ is specified in Fig. 3. The conditional *a posteriori* probability function $\pi^l(m, n)$ for $l(m, n)$ given $L_{(m,n)}$, S and R is

$$\begin{aligned} \pi^l(m, n) &= P[l(m, n) | L_{(m,n)}, S, R] \\ &= \frac{p(S, L, R)}{\sum_{l(m,n)} p(S, L, R)} = \frac{p(S, L)p(S|R)}{\sum_{l(m,n)} p(S, L)p(S|R)} \\ &= \frac{p(L)P(L)}{\sum_{l(m,n)} p(S|L)P(L)}. \end{aligned} \quad (17)$$

Recall that

$$\begin{aligned} p(S|L) &= \prod_{(i,j)} p(s(i, j) | \{s(k, l): (k, l) \\ &\quad \in \mathcal{R}_{o+}(i, j)\}, l(i, j)) \end{aligned}$$

and

$$P(L) = \prod_{(i,j)} P(l(i, j) | \{l(k, l): (k, l) \in \mathcal{R}_{o+}(i, j)\}).$$

Inserting these two equations into (17), we have

$$\begin{aligned} \pi^l(m, n) &= \frac{1}{Z_6} p(s(m, n) | \{s(k, l): (k, l) \in \mathcal{R}_{o+}(m, n)\}, \\ &\quad l(m, n)) \cdot P(l(m, n) | \{l(k, l): (k, l) \\ &\quad \in \mathcal{R}_{o+}(m, n)\}) \end{aligned} \quad (18)$$

where Z_6 is a normalizing constant. The remaining term in (18), $p(s(m, n) | \{s(k, l): (k, l) \in \mathcal{R}_{o+}(m, n)\}, l(m, n))$ can be expressed as

$$\begin{aligned} &p(s(m, n) | \{s(k, l): (k, l) \in \mathcal{R}_{o+}(m, n)\}, l(m, n)) \\ &= \frac{1}{\sqrt{2\pi T\sigma_{w(m,n)}^2}} \\ &\quad \cdot \exp \left[- \frac{\left(s(m, n) - \sum_{kl} c_{kl}^{(m,n)} s(m-k, n-l) \right)^2}{2T\sigma_{w(m,n)}^2} \right]. \end{aligned}$$

The simulated annealing method can be implemented in a sequential or parallel manner. We first describe sequential simulated annealing, which is appropriate for raster scanning of the image. We keep the temperature constant for each sweep of the image and reduce the temperature only after the complete sweep

(or iteration) of the image. Since $\pi^s(m, n)$ and $\pi^l(m, n)$ both depend on T , we henceforth denote them by $\pi_t^s(m, n)$ and $\pi_t^l(m, n)$, respectively.

1) *Sequential Simulated Annealing Algorithm*: The sequential simulated annealing algorithm can be described as follows: Let (m_t, n_t) , $t = 1, 2, \dots$, be the sequence in which the sites are visited for updating.

- 1) Set $t = 0$ and assign an initial configuration denoted as S_{-1} , L_{-1} and an initial temperature $T(0) = 1$.
- 2) The evolution $L_{t-1} \rightarrow L_t$ of the hidden system can be obtained by sampling the next point of the hidden system from the raster-scanning scheme based on the conditional probability mass function $\pi_t^l(m_t, n_t)$ and keeping the rest of L_{t-1} unchanged.
- 3) Set $t = t + 1$. Go back to step 2 until a complete sweep of the field L is finished.
- 4) The evolution $S_{t-1} \rightarrow S_t$ of the observed system can be obtained by sampling the new value of the observations $s(m_t, n_t)$ based on the conditional pdf $\pi_t^s(m_t, n_t)$ and keeping the rest of S_{t-1} unchanged.
- 5) Set $t = t + 1$. Go back to step 4 until a complete sweep of the whole image is achieved.
- 6) Go to step 2 until $t > t_f$, where t_f is a specified integer.

The following theorem from [7] guarantees that simulated annealing converges to the MAP estimate if the stated conditions are satisfied.

Theorem 1: If the following conditions are satisfied:

- a) $\sum_{kl} |c_{kl}^l| = \rho < 1 \quad \forall l$
- b) Let $T(t) \rightarrow 0$ as $t \rightarrow \infty$, and
- c) $T(t) \geq C/\log(1 + k(t))$,

then for any starting configuration S_{-1} , L_{-1} , we have

$$p(S_t, L_t | S_{-1}, L_{-1}, R) \rightarrow \pi_0(S, L), \quad \text{as } t \rightarrow \infty \quad (19)$$

where $\pi_0(\cdot, \cdot)$ is the uniform distribution over the MAP solutions and $k(t)$ is the sweep iteration number at time t .

2) *Parallel Simulated Annealing Algorithm*: As pointed out in [5], [11], stochastic relaxation converges very slowly to the MAP estimate because of the logarithmic temperature cooling schedule. Now, in [5], it was suggested that simulated annealing can be implemented in an asynchronous parallel machine to speed up the processing. However, an asynchronous machine may be rather difficult to implement and use compared to a single-instruction multiple-data (SIMD) machine. Here we present a coding scheme for implementation of simulated annealing on an SIMD machine as has been suggested by Murray *et al.* in [13], [4]. The total speed-up depends on the size of the neigh-

neighborhood of the Markov random field. In theory, we can have $2N^2/k$ speed-up if we have enough processors where N^2 is the total number of pixels in the image and k is the size of the neighborhood of the Markov random field. We describe this maximally parallel "coding scheme" next.

Instead of updating one pixel at each iteration, we can partition the entire image into disjoint regions (called coding regions) such that pixels which belong to the same region are conditionally independent given the data of all the other regions. The total number of coding regions depends on the size of the neighborhood support of the image model, hence the speed-up factor given above. Since the pixels in the same coding region are conditionally independent, we can update them simultaneously at each iteration by using the previous iteration result in the sequential simulated annealing procedure. A parallel version of our simulated annealing procedure can thus be described as follows.

- 1) Set $t = 0$ and assign an initial configuration denoted as S_{-1}, L_{-1} , an initial temperature $T(0) = 1$ and the disjoint coding regions for S denoted as $\Gamma_1^s, \Gamma_2^s, \dots, \Gamma_{\alpha_1}^s$ and the disjoint coding regions for L denoted as $\Gamma_1^l, \Gamma_2^l, \dots, \Gamma_{\alpha_2}^l$. Here, we assume that α_1 and α_2 are the total number of coding regions for S and L , respectively.
- 2) The evolution $L_{t-1} \rightarrow L_t$ of the line field can be obtained by sampling the new value of the coding region $\Gamma_{i_t}^l$ of the structure image L_t based on the conditional probability mass function $\pi^l(m, n) \forall (m, n) \in \Gamma_{i_t}^l$, where $\Gamma_{i_t}^l$ is the coding region visited at time t .
- 3) Set $t = t + 1$. Go back to step 2 until the whole line field is finished.
- 4) The evolution $S_{t-1} \rightarrow S_t$ of the image can be obtained by sampling the new value of the coding region $\Gamma_{i_t}^s$ of the image S_t based on the conditional density function $\pi^s(m, n) \forall (m, n) \in \Gamma_{i_t}^s$, where $\Gamma_{i_t}^s$ is the coding region visited at time t .
- 5) Set $t = t + 1$ and assign a new value to $T(t + 1)$ after each complete sweep of the image and go forward to the next step, otherwise keep the temperature unchanged and go back to the previous step.
- 6) Goto step 2 until $t > t_f$, where t_f is a specified integer.

The following theorem from [9] guarantees that this parallel processing version of simulated annealing will converge to the MAP estimate under the stated conditions.

Theorem 2: If the following conditions are satisfied:

- a) $\sum_{kl} |c_{kl}| = \rho < 1 \quad \forall l$
- b) $T(t) \rightarrow 0$ as $t \rightarrow \infty$, and
- c) $T(t) \geq C/\log(1 + k(t))$,

then for any starting configuration S_{-1}, L_{-1} , we have

$$p(S_t, L_t | S_{-1}, L_{-1}, R) \rightarrow \pi_0(S, L), \text{ as } t \rightarrow \infty. \quad (20)$$

C. Deterministic Search for the MAP Estimate

Instead of using a stochastic approach, we can use a deterministic method to search for a local optimal point instead of a global one. This deterministic algorithm is an extension of the iterated conditional mode of [2] to the continuous valued, compound Gauss-Markov random fields. An advantage of deterministic relaxation is that the convergence is much faster than that of simulated annealing on a sequential machine. The disadvantage is the local nature of the optimum achieved. The al-

gorithm can be described as follows: Let (m_t, n_t) , $t = 1, 2, \dots$ be the sequence in which the sites are visited for updating.

- 1) Start at any initial point S_0 and L_0 and set the temperature T to 1 throughout the whole algorithm and set $t = 1$. Set $\epsilon > 0$.
- 2) The evolution $L_{t-1} \rightarrow L_t$ of the hidden system can be obtained by maximizing the functional $\pi^l(m_t, n_t)$ and keeping the rest of L_{t-1} unchanged.
- 3) Set $t = t + 1$. Go back to step 2 until a complete sweep of the field L is finished.
- 4) The evolution $S_{t-1} \rightarrow S_t$ of the observed system can be obtained by maximizing the functional $\pi^s(m_t, n_t)$ and keeping the rest of S_{t-1} unchanged.
- 5) Set $t = t + 1$. Go back to step 4 until a complete sweep of the whole image is achieved.
- 6) Go to step 2 until

$$|p(S_t, L_t | R) - p(S_{t-1}, L_{t-1} | R)| < \epsilon. \quad (21)$$

- 7) Finally, set $\hat{S} = S_t$, $\hat{L} = L_t$.

Among the above five steps, only steps 2 and 3 need further explanation. The new value $l(m_t, n_t)$ in step 2 can be obtained by exhaustive search since $l(m_t, n_t)$ can take on only a small number of values. In distinction, since $\pi^s(m_t, n_t)$ is a quadratic form of $s(m_t, n_t)$, the new value $s(m_t, n_t)$ in step 3 can be easily obtained by taking a derivative of $\pi^s(m_t, n_t)$ and setting it to zero.

It is clearly seen that such a new choice of values for S and L at (m_t, n_t) always increases the value of the likelihood function, i.e.

$$p(S_t, L_t | R) \geq p(S_{t-1}, L_{t-1} | R). \quad (22)$$

In general, the deterministic search converges faster than simulated annealing, but not necessarily to the global optimal MAP estimate. As in Section IV-B2, a parallel version of deterministic search is possible.

V. PARAMETER ESTIMATION

Various parameters, such as model coefficients and transition probabilities, are needed before we can implement these filtering algorithms. Since these parameters are not often available, we must estimate them from the data. Here we take the viewpoint of estimating these parameters from training data that may be prototypical of the expected noisy images. Using four orientation directional edge operators scanning through the training image (the original image is used here), we partition this image into six disjoint regions. Each region corresponds to a different submodel which will be used to model that part of the image. After obtaining a model configuration, i.e., the hidden structure of the whole image, we can use maximum likelihood (ML) identification techniques to estimate the various submodels' coefficients.

For the causal DSG case, we can obtain ML identification by using the classic least squared error parameter identification method within each disjoint structural region. For the noncausal CGM case, the situation is more complicated and the difficulty is twofold. First, the ML identification for a noncausal Gauss-Markov model is not the classic least squares problem [10]. A method called the "coding scheme" has been widely used. However, this method is not efficient in that only 50% of the data is used. Also, the result is not unique since many different coding schemes exist. In [10], a consistent estimation method



Fig. 4. (a) Original lady's face image; (b) noisy image with $\text{SNR} = 10$ dB; (c) RUKF result for the NSHP AR model; (d) M -algorithm result with $M = 5$ for the DSG model.

was proposed for the noncausal Gauss-Markov model. This scheme is quite similar to least square error identification. We adopt this method for parameter identification in the CGM case. Second, due to the covariance constraints on the submodel coefficients (7), the set of coefficients obtained through the above identification procedures will be not consistent in general. There are two ways around this difficulty. One of them is to modify the identification results from the least squares method to make the solutions consistent. Another way is to heuristically assign a set of consistent model coefficients that seems reasonable, much as was done in [5]. Here we use this latter ad hoc approach to obtain the consistent set of model coefficients needed for our simulations.

The transition probabilities for the lower level Markov chain of the DSG random field can be obtained by using the ML identification technique given the model configuration of the image.

It turns out that histogram estimation is the optimal solution. For the line field in the CGM model, we use the same Gibbs distribution as in [5].

In the implementation of the M -algorithm, some approximations have been used for reducing computation [15]. For example, an approximate reduced updated Kalman filter (RUKF) [16] and steady-state Kalman filter gains were employed to approximate the exact 2-D Kalman filter.

VI. SIMULATION RESULTS

In this section, we present some simulation results and compare them with linear filtering results as in [16] and [3]. All simulations use first-order image models. For image estimation (no blur), white Gaussian noise was added to our intensity data at the signal-to-noise ratio (SNR) = 10 dB. For image resto-



Fig. 5. (a) Noisy image with SNR = 10 dB; (b) Wiener filter result for the NSHP model with SNR = 10 dB; (c) simulated annealing result for the DSG model at 200 iterations; (d) deterministic relaxation result for the DSG model at 25 iterations.

ration (blur plus noise), a 5×5 uniform blur has been used and blurred-signal-to-noise ratios (BSNR) = 20 and 40 dB were employed.

Fig. 4(a) shows an original lady's face image, (b) shows the 10 dB noisy image, (c) shows the RUKF filtering result, and (d) shows the M -algorithm result with $M = 5$. The remaining Figs. 5-10 are arranged as follows: (a) shows the input image (degraded image), (b) shows the linear filtering (Wiener filter) result, (c) shows the simulated annealing result with 200 iterations, and (d) shows the deterministic search result with 25 iterations.¹ Fig. 5 shows the estimation results for the DSG

¹Both iterative algorithms are, respectively, very well converged at these numbers of iterations.

model. The estimation results for the CGM model are shown in Fig. 6.

It is clearly seen that the estimate results from the compound Gauss-Markov models provide better visual quality, i.e., sharper edges and cleaner flat regions, than the simple Gaussian models. Although the deterministic relaxation results have some degree of degradation compared with those of simulated annealing, they are still superior to those of the simple Gaussian models, with the DSG model getting into more difficulties, perhaps because of its larger number of transition probabilities. We also note that the M -algorithm estimate with $M = 5$ of Fig. 4(d), while not as good as the simulated annealing result in Fig. 5(c), is still remarkably good and of much lower computational complexity.



Fig. 6. (a) Noisy image with SNR = 10 dB; (b) Wiener filter result for Gauss-Markov model with SNR = 10 dB; (c) simulated annealing result for the CGM model at 200 iterations; (d) deterministic relaxation result for the CGM model at 25 iterations.

Comparing the simulated annealing results for the DSG model (Fig. 5(c)) and the CGM model (Fig. 6(c)), the results are visually quite similar, with the DSG-based estimate showing a bit more detail and the CGM-based estimate apparently a bit smoother. However, this visual difference could easily be accounted for by differences in the model coefficients, transition probabilities, and line-field parameters.

Simulations for image restoration, i.e., deblurring, were also performed. Fig. 7 shows the estimate results for the DSG model restored from a blurred and noisy image with BSNR = 40 dB. The restoration results based on the CGM model from the same

blurred and noisy image with BSNR = 40 dB are shown in Fig. 8.

The restoration results in the BSNR = 40 dB case are quite good for both the LSI and the compound models although significant improvements have been obtained from the compound models in terms of artifact reduction. In particular the compound model-based result (Fig. 7(c)) shows much less ringing artifacts than the Gaussian model-based estimate (Fig. 7(b)). The results based on the causal DSG and noncausal CGM models are of very similar quality.

A more challenging task is the restoration for BSNR = 20



Fig. 7. (a) Blurred and noisy image with $\text{BSNR} = 40$ dB; (b) Wiener filter result for NSHP model with $\text{BSNR} = 40$ dB; (c) simulated annealing result for the DSG model at 200 iterations; (d) deterministic relaxation result for the DSG model at 25 iterations.

dB. Fig. 9 shows the results for the DSG model from a blurred and noisy image with $\text{BSNR} = 20$ dB. The restoration results based on the CGM model are shown in Fig. 10. Here, the differences among the restoration results for the various models become much stronger. The compound models provide sharper images without the noise amplification that results in the LSI Gauss-Markov based estimates. Comparing the two compound models, we find that the DSG model-based results appear somewhat better than the first-order CGM based estimates. Both compound models allow the restoration to proceed with strong noise reduction in fairly flat regions of the image, yet still retain

sharp edge restoration. The Wiener filter results clearly suffer from an inability to deal with this problem.

VII. CONCLUSIONS

Images are modeled as compound random fields in this paper using causal DSG models and new noncausal CGM models for both image estimation and restoration. A recently developed simulated annealing algorithm of the authors [7] has been applied to obtain MAP estimates for these models. To reduce the computation on sequential machines and for comparison of es-



Fig. 8. (a) Blurred and noisy image with BSNR = 40 dB; (b) Wiener filter result for the Gauss-Markov model with SNR = 40 dB; (c) simulated annealing result for the CGM model at 200 iterations; (d) deterministic relaxation result for the CGM model at 25 iterations.

timation quality, a deterministic relaxation method was also presented. We believe that the simulation results show that these two compound models can yield higher visual quality than can simple Gaussian models for both image estimation and restoration.

Although each of the compound models is more powerful than the LSI model, the increasing complexity and computation is a problem. Also, the difficult problem of parameter identification so as to avoid inconsistent sets of model coefficients for the CGM model, remains an open problem.

Finally we find that the simulations point to the fact that, while compound versus simple model is important, causal versus noncausal model did not seem to much affect the results. Also, but less strongly, the causal M -algorithm estimate with only five paths, i.e., $M = 5$, and an MMSE error criteria, seemed to perform quite comparably to the much more computationally demanding simulated annealing algorithm based on the MAP criteria in the noise-only case. This should be investigated more fully in the context of coarse-grained parallel algorithms for image restoration.



Fig. 9. (a) Blurred and noisy image with $\text{BSNR} = 20$ dB; (b) Wiener filter result for the NSHP model with $\text{SNR} = 20$ dB; (c) simulated annealing result for the DSG model at 200 iterations; (d) deterministic relaxation result for the DSG model at 25 iterations.



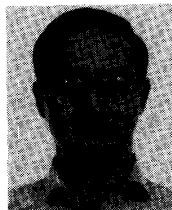
Fig. 10. (a) Blurred and noisy image with BSNR = 20 dB; (b) Wiener filter result for the Gauss-Markov model with SNR = 10 dB; (c) simulated annealing result for the CGM model at 200 iterations; (d) deterministic relaxation result for the CGM model at 25 iterations.

ACKNOWLEDGMENT

The authors wish to thank Prof. A. Rosenfeld for providing access to the computer facilities while F.-C. Jeng was visiting at the Center for Automation Research, University of Maryland.

REFERENCES

- [1] J. F. Abramatic and L. M. Silverman, "Nonlinear restoration of noisy images," *IEEE Trans. Pattern Analysis Mach. Intell.*, vol. PAMI-4, pp. 141-149, Mar. 1982.
- [2] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Stat. Soc. B*, vol. 48, pp. 259-302, 1986.
- [3] R. Chellappa and R. L. Kashyap, "Digital image restoration using spatial interaction models," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 461-472, June 1982.
- [4] H. Derin and C. S. Won, "A parallel image segmentation algorithm using relaxation with varying neighborhoods and its mapping to array processors," *Comput. Vision, Graph., Image Processing*, vol. CVGIP-40, pp. 54-78, 1987.
- [5] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Analysis Mach. Intell.*, vol. PAMI-6, pp. 721-741, Nov. 1984.
- [6] F. C. Jeng and J. W. Woods, "Image estimation by stochastic relaxation in the compound Gaussian case," in *Proc. ICASSP 1988* (New York, NY), Apr. 1988, pp. 1016-1019.
- [7] F. C. Jeng and J. W. Woods, "Simulated annealing in compound Gauss-Markov random fields," *IEEE Trans. Inform. Theory*, vol. 36, pp. 94-107, Jan. 1990.
- [8] F. C. Jeng and J. W. Woods, "Inhomogeneous Gaussian image models for image estimation and restoration," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, pp. 1305-1312, Aug. 1988.
- [9] F. C. Jeng, "Compound Gauss-Markov random fields for image estimation and restoration," Ph.D. dissertation, Rensselaer Polytechnic Institute, Troy, NY, June 1988.
- [10] R. L. Kashyap and R. Chellappa, "Estimation and choice of neighbors in spatial-interaction models of images," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 60-72, Jan. 1983.
- [11] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671-680, May 1983.
- [12] H. E. Knutsson, R. Wilson, and G. H. Granlund, "Anisotropic nonstationary image estimation and its applications: Part I-Restoration of noisy images," *IEEE Trans. Commun.*, vol. COM-31, pp. 388-397, Mar. 1983.
- [13] D. W. Murray, A. Kashko, and H. Buxton, "A parallel approach to the picture restoration algorithm of Geman and Geman on an SIMD machine," *Image Vision Comput.*, vol. 4, pp. 133-142, Aug. 1986.
- [14] S. A. Rajala and R. J. P. DeFigueiredo, "Adaptive nonlinear image restoration by a modified Kalman filtering approach," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 1033-1042, Oct. 1981.
- [15] J. W. Woods, S. Dravida, and R. Mediavilla, "Image estimation using doubly stochastic Gaussian random field models," *IEEE Trans. Pattern Analysis Mach. Intell.*, vol. PAMI-9, pp. 245-253, Mar. 1987.
- [16] J. W. Woods and C. H. Radewan, "Kalman filtering in two dimensions," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 473-482, July 1977.
- [17] J. W. Woods, "Two-dimensional discrete Markovian fields," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 232-240, Mar. 1972.



Fure-Ching Jeng (S'84-M'89) received the B.S.E.E. degree from the National Taiwan University, Taipei, Republic of China, in 1980, and the M.S. and Ph.D. degrees in electrical engineering from Rensselaer Polytechnic Institute, Troy, NY, in 1985 and 1988, respectively.

From 1984 to 1988 he was with the ECSE Department at Rensselaer Polytechnic Institute, as a Research Assistant. He is currently with Bell Communications Research in Morristown, NJ. His research interests include image segmentation, data compression, pattern recognition, and digital communications.



John W. Woods (S'67-M'70-SM'83-F'88) received the B.S., M.S., E.E., and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1965, 1967, and 1970, respectively.

Since 1976 he has been with the ECSE Department at Rensselaer Polytechnic Institute, Troy, NY, where he is currently a Professor. He has taught courses in digital signal processing, probability and stochastic processes, digital image processing, and communication systems. His research interests include estimation/restoration, detection, recursive digital filtering and data compression of images and other multidimensional data. He has authored or coauthored over 50 papers in these fields. He has coauthored one text in the area of probability, random processes, and estimation. During the academic year 1985/1986, he was Visiting Professor in the Information Theory Group at Delft University of Technology, The Netherlands. He directed the Circuits and Signal Processing Program at the National Science Foundation, Washington, DC, in 1987 and 1988.

Dr. Woods was corecipient of the 1976 and 1987 Senior Paper Awards of the IEEE Acoustics, Speech, and Signal Processing (ASSP) Society. He is a former member of the Digital Signal Processing Committee. He is a former Associate Editor for Signal Processing of the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. He is a former Chairman of the Society Technical Committee on Multidimensional Signal Processing. He is a former member of the Administrative Committee of the ASSP Society. Currently he is Chairman of the Signal Processing Society Education Committee. He received the Signal Processing Society Meritorious Service Award in 1990. He is a member of Sigma Xi, Tau Beta Pi, Eta Kappa Nu, and the AAAS.