Pyramid Implementation of Optimal-Step Conjugate-Search Algorithms for Some Low-Level Vision Problems

TAL SIMCHONY, MEMBER, IEEE, RAMALINGAM CHELLAPPA, SENIOR MEMBER, IEEE, AND ZE'EV LICHTENSTEIN

Abstract—**Optimization of a cost function arises in several low-level vision problems. The cost functions in these problems are usually derived from discretization of functionals obtained from regularization principles or stochastic estimation techniques using Markov random field models. A parallel pyramid implementation of the line search conjugate gradient algorithm for minimizing the cost functions is presented. By viewing the global cost function as a Gibbs energy function, we efficiently compute the gradients, inner products and the optimal-step size using the pyramid. The global Gibbs energy of a given configuration is broken into its basic energy terms associated with different cliques. The authors let each low-level processor sum the energies associated with its cliques. The local energy terms are added by the intermediate levels of the pyramid to the top, where the step size is determined by an efficient univariate search. Such implementation allows us to calculate the global energy in $O(\log n)$ operations, where n is the number of grid points in each direction. Implementation of this algorithm to shape from shading results in a multiresolution conjugate gradient algorithm. Robustness and efficiency of the algorithm are demonstrated for edge detection using the graduated nonconvexity (GNC) algorithm. The authors also present the usefulness of this formulation to image estimation based on Markov models. A compound model for the original image is defined consisting of a two-dimensional (2-D) noncausal Gauss–Markov random field (GMRF) to represent the homogeneous regions, and a line process to represent the discontinuities. A new deterministic algorithm based on the GNC formulation is derived to obtain a near optimal Maximum aposteriori probability (MAP) estimate of images corrupted by additive Gaussian noise. Since the algorithm depends on the model parameters, a new estimation technique for obtaining the compound GMRF model parameters is derived based on the expectation maximization (EM) algorithm. Experimental results are provided to illustrate the usefulness of this method.**

I. Introduction

MANY physical phenomena that govern the processes that are of interest to researchers in computer vision, can be described by their local characteristics. They are formulated mathematically using differential or integral equations in the deterministic approach, or Markov fields in the stochastic formulation. Examples for the deterministic formulation are the image irradiance equation, the optical flow equation [1], etc. An example for the stochastic formulation is image restoration using Markov random fields models [2]. Computer vision research is usually focused on the inverse problem [3], i.e., given an observation (usually an image, or a sequence of images), find a dense depth map, the velocity field, etc. These problems are in many cases ill posed in the sense of Hadamard [4]. In order to overcome the ill posedness, regularization methods have been suggested [3]. In this approach the original problems are reformulated as optimization problems. Instead of solving the original differential equation exactly, an integral consisting of two elements is constructed. The first element minimizes the mean square error between the observation and the corresponding calculated quantity (from the differential equation). The second term is a penalty term that favors smooth solutions. The desired solution is the minimum of this functional (integral).

There are several ways to obtain the minimum of the functional. One can use calculus of variations to convert the optimization problem into a set of differential equations [5]. This variational approach was extensively explored for the shape from shading (SFS) problem [6], which deals with the reconstruction of the orientation map of a surface from a single image using the irradiance equation. This formulation leads to nonlinear differential equations, which are not easy to solve unless the penalty term in the functional is sufficiently large (the solution is far from the original problem). Alternatively one can use direct optimization [5]. Direct methods discretize the integral in the functional and replace the partial derivatives by finite difference approximations to obtain cost functions of the different attributes' values at all grid points (which we call configurations). In the SFS problem one obtains a cost function on the orientation of the surface at all the grid points. The minimum of this cost function is the desired solution to the SFS problem [7]. Using standard calculus [7], the configuration that sets the gradient of the cost function with respect to the orientation at each grid point to zero is searched for. The same methodology is

Manuscript received September 16, 1988; revised March 30, 1989. This work was partially supported by the NSF Grant No. MIP-84-51010 and matching funds from IBM and AT&T and Hughes Aircraft Company. T. Simchony was supported by ECI Telecom. This work was partially presented at the Second International Conference on Computer Vision, Tampa, FL, December 1988.

The authors are with the University of Southern California, Signal and Image Processing Institute, Dept. of EE-Systems, University Park/MC-0272, Los Angeles, CA 90089.

IEEE Log Number 8930348.

0018-9472/89/1100-1408$01.00 ©1989 IEEE

used in [8], leading to a system of sparse nonlinear equations. The solution is obtained via a relaxation method in [7] or a direct method in [8]. The algorithms converge only when the penalty term is large, which yields an oversmoothed solution.

Cost functions also yield a suitable formulation for problems involving discontinuities in the observation or in the reconstructed attribute. An example for this is the weak membrane formulation [9] for the surface interpolation problem. A weak membrane (weak membrane is a membrane that satisfies weak continuity constraints) is fitted to noisy data, and used for edge detection by marking the points where discontinuities in the membrane height occur. The cost function derived using this model is not convex. To find the global minimum of the cost function F, it is suggested that the graduated nonconvexity algorithm (GNC) may be used. The GNC algorithm starts with a convex approximation F^* to the cost function F. Then a whole sequence of cost functions $F^{(P)}$, $0 \le P \le 1$ is constructed, so that $F^{(0)} = F$, the original nonconvex function, and $F^{(1)} = F^*$. In between, $F^{(P)}$ changes continuously, between $F^{(1)}$ and $F^{(0)}$. The GNC algorithm optimizes a whole sequence of $F^{(P)}$, for example $\{F^{(1)}, F^{(1/2)}, F^{(1/4)}, F^{(1/8)}, F^{(1/16)}\}$, one after the other, using the optimal solution of previous optimization step as the starting point for the next. The optimization step is then performed using the sequential over relaxation (SOR) method. The authors report a big degradation in the convergence rate of the SOR method when the processed image is corrupted with excessive noise.

The third cost function formulation is derived from stochastic modeling. When one models a process by a Markov field, and exploits the Gibbs–Markov equivalence theorem, a Gibbs energy function is obtained for the different configurations. As an example consider the problem of image estimation. We assume that the original image is represented by a compound model consisting of a two-dimensional (2-D) noncausal Gauss–Markov random field (GMRF) to represent the homogeneous regions and a line process model to represent the discontinuities (note that GMRF models can be viewed as extended regularization models). This model can also be used for modeling piecewise continuous functions required for surface interpolation and edge detection. The noise is assumed to be additive Gaussian and spatially independent. The objective is the maximum *aposteriori* distribution (MAP) solution, or the configuration that corresponds to the minimum of the Gibbs energy function. This problem illustrates the equivalence between a Gibbs energy function, and the cost functions obtained by direct optimization. Since the Gibbs energy function corresponding to the posterior distribution is not convex it can be minimized using simulated annealing [2], a computationally extensive proposition.

In this paper we propose a unified and practical method for minimizing the cost functions resulting from the deterministic and stochastic formulations described previously. The method exploits the equivalence between direct optimization of a deterministic problem, and the MAP solution using the Markov random fields, through the use of Gibbs energy function, and treats the two problems with the same algorithm. We chose an optimization algorithm [10] that is amenable to parallel computation to cope up with the computational complexity induced by the high-dimensional cost functions. The special structure of Gibbs energy functions is utilized to map the algorithm to a highly parallel implementation. Since in general the cost function is not quadratic, we use an extended version of the original conjugate gradient method—the line search extension proposed by Polak and Ribiere [10]. The algorithm is very efficient in solving high-dimensional unconstrained optimization problems. The complete algorithm is as follows.

- Step 1) Given X_0 compute $G_0 = \nabla f(X_0)$ and set $D_0 = -G_0$.
- Step 2) For $k = 0, 1, \cdots, n-1$:
  - a) Set $X_{k+1} = X_k + \alpha_k D_k$ where α_k minimizes $f(X_k + \alpha D_k)$;
  - b) Compute $G_{k+1} = \nabla f(X_{k+1})$;
  - c) Unless $k = n-1$, set $D_{k+1} = -G_{k+1} + \beta_k D_k$ where

$$\beta_k = \frac{(G_{k+1} - G_k)^T G_{k+1}}{G_k^T G_k};$$

  - d) Check termination criteria.
- Step 3) Replace X_0 by X_n and go back to Step 1).

The use of the algorithm requires computations of inner products in Step 2c), and finding the optimal step in Step 2a). The value of the optimal step depends on the global cost function. We use a pyramid to efficiently calculate the global cost function and the inner products required for computing β_k. We view the global cost function as a Gibbs energy function corresponding to a given configuration. We break the energy function to its basic energy terms associated with different cliques. We let each low-level processor sum the energies associated with its cliques, making sure that each clique's energy is summed only once. The local energy terms are added by the intermediate levels of the pyramid to the top, where the step size is determined by an efficient univariate search. Such an implementation allows us to calculate the global energy in $O(\log n)$ operations, where n is the number of grid points in each direction. This search is bound to converge to a maximal point of the cost function. The additional component of the algorithm is the computation of the gradient in Step 2b). This computation can be implemented in parallel at the bottom of the pyramid, because the gradients of the cost functions (for this class of problems) depend only on the values of the function in a small neighborhood of the pixel.

An additional improvement can be obtained for the SFS problem if one searches for the minimum of the discrete cost function using a multigrid conjugate gradient algorithm. The rate of convergence of the multigrid search is an order of magnitude faster than the single resolution

version. This improvement is due to better propagation of low frequency information on the coarse grid. Determination of the higher frequency modes occurs as the grid gets finer and finer. This multigrid search utilizes the same properties that make multigrid relaxation method so successful [11]. In our work we experimented with V-shape multigrid schedules, and we believe that other schedules such as W-shape will perform equally well.

We have also studied the use of the extended conjugate gradient search to replace the SOR step in Blake and Zisserman's [9] graduated nonconvexity (GNC) algorithm. Once again we use the Gibbs formulation in order to obtain an efficient and parallel implementation. The results we obtain using the conjugate gradient algorithm require a smaller number of iterations than the SOR method [9]. The convergence rate of the algorithm is much less sensitive to noise. This example demonstrates the robustness of the algorithm in comparison to other methods.

A deterministic relaxation algorithm is presented for obtaining the MAP estimate of gray-level images modeled by a compound GMRF, and degraded by additive Gaussian noise. This algorithm, which is an extension of the GNC algorithm, is able to find the near optimal MAP estimate corresponding to the nonconvex posterior Gibbs energy function. As a by-product, the line process configuration determined by the MAP estimate produces an accurate edge map without any additional cost. Unlike the simulated annealing method, the deterministic algorithm converges in a small number of iterations. Each of the optimization steps required by the GNC algorithm is performed using the extended conjugate gradient search implemented on a pyramid.

Due to the modeling assumption the restoration algorithm depends on the compound GMRF model parameters. We obtain estimates of the compound GMRF model parameters from the original image using a new expectation maximization (EM) estimation technique. The EM algorithm enables estimation of the GMRF model parameters without being affected by the edges present in the image.

Hierarchical pyramid based algorithms have been proposed for many computer vision tasks. High-level applications include multiresolution and top-down/bottom-up image analysis tasks, such as feature extraction for object recognition [12]. The pyramid structure is very efficient for a variety of low-level processes, e.g., averaging, histograming, edge detection, median filtering, and image segmentation [13]–[19]. Algorithms implemented on the pyramid exploit the pyramid's massively parallel and shallowly serial hierarchical computing ability. In this paper we use the pyramid structure to obtain a fast parallel implementation of the line search conjugate gradient algorithm, which we implement in both single-resolution, and multigrid fashion.

The organization of the paper is as follows. In Section II, the SFS problem is briefly discussed, and the multiresolution conjugate gradient algorithm for reconstructing the orientation of a surface is presented. A synthetic sphere image is experimented with both single- and multiresolution versions of the algorithm to show the reduction in computation time obtained with the multiresolution implementation. In Section III, we present the GNC algorithm with a modified optimization step, based on the conjugate gradient search. Experimental results for edge detection are presented. The experiments indicate that the algorithm is more robust than SOR. Section IV discusses estimation of gray-level images corrupted by additive noise. We present the compound GMRF model, and derive an extended GNC algorithm for obtaining the MAP estimate. A new parameter estimation technique for obtaining the compound model parameters is derived. We present experiments to show the performance of the new algorithm on real images, and obtain both image estimates and edge maps.

II. Shape From Shading

A. Discrete Cost Function for SFS and Gibbs Representation

The SFS problem is the problem of reconstructing the surface orientation from the observed image intensity. Let $E(x, y)$ be the observed image of intensity related to a surface $Z(x, y)$ as

$$E(x, y) = R(p, q, \beta, l, \rho) \tag{1}$$

where β is the illumination direction vector, l is the vector from the surface to the camera, ρ is the albedo term, $p = Z_x$ and $q = Z_y$ are the surface slopes and R is the reflectance map. In the case of a Lambertian surface, one can further write (1) as

$$E(x, y) = \frac{\rho\beta \cdot (-p, -q, 1)}{(1 + p^2 + q^2)^{1/2}}. \tag{2}$$

The discrete set up for the SFS problem was first suggested by Strat [7], using the following cost function:

$$\epsilon^2 \sum_{i=1}^{n} \sum_{j=1}^{m} \left(E_{ij} - R(p_{ij}, q_{ij})\right)^2 + \frac{\lambda}{\epsilon^2} \sum_{i=1}^{n} \sum_{j=1}^{m} e_{ij}^2 \tag{3}$$

where the first term corresponds to the irradiance error. The second term e_{ij} is the integrability penalty term that corresponds to an estimate for the integral around an elementary square path in the counter-clockwise direction, with the picture cell (i, j) in the lower left corner, i.e.,

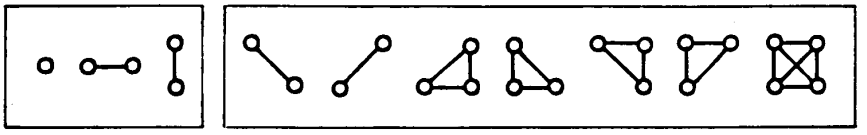
$$e_{ij} = \frac{\epsilon}{2}[p_{i,j} + p_{i+1,j} + q_{i+1,j} + q_{i+1,j+1} - p_{i+1,j+1} - p_{i,j+1} - q_{i,j+1} - q_{i,j}]. \tag{4}$$

In our work we use Strat's integrability term. A different discrete cost function was suggested by Lee [8]:

$$\mu = \epsilon^2 \sum_{i=1}^{n} \sum_{j=1}^{m} \left(E_{ij} - R(f_{ij}, g_{ij})\right)^2 + \kappa \sum_{i=1}^{n} \sum_{j=1}^{m} r_{ij} \tag{5}$$

where f and g are the surface orientation in stereographic coordinates and r_{ij} is a smoothing penalty term

$$r_{ij} = \left[(f_{i+1,j} - f_{i,j})^2 + (f_{i,j+1} - f_{i,j})^2 + (g_{i+1,j} - g_{i,j})^2 + (g_{i,j+1} - g_{i,j})^2\right] / h^2.$$



(a) (b)

Fig. 1. Gibbs energy cliques. (a) First-order MRF cliques ($c=1$). (b) Second-order MRF cliques ($c=2$).

In our formulation we work with p and q and use two different cost functions. The first is a straight forward modification of (5) to p and q coordinates, and the second adds the integrability constraint described in (3) and (4) to obtain:

$$\epsilon^2 \sum_{i=1}^{n}\sum_{j=1}^{m}\left(E_{ij}-R(p_{ij},q_{ij})\right)^2 + \frac{\lambda}{\epsilon^2}\sum_{i=1}^{n}\sum_{j=1}^{m} e_{ij}^2 + \kappa \sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij}. \quad (6)$$

We now explain the Gibbs formulation for the cost function.

Let $S=\{s_1, s_2, \cdots, s_N\}$ be the set of sites (grid points), $\mathcal{G}=\{\mathcal{G}_s, s\in S\}$ be a neighborhood system for S, meaning any collection of subsets of S for which 1) $s\notin \mathcal{G}_s$ and 2) $s\in\mathcal{G}_r \Leftrightarrow r\in\mathcal{G}_s$. $\mathcal{G}_s$ is the set of neighbors of s and $\{S,\mathcal{G}\}$ is a graph in the usual way. A subset $C\subset S$ is a clique if every pair of distinct sites in C are neighbors. $\mathcal{C}$ denotes the set of cliques [2]. In our problem S is the set of pixel sites (grid points) of the image intensity, $N=n^2$ where n is the number of grid points in each direction. It is easier to view s_m as a two-dimensional vector $s_m=\{(i,j)|i,j\in[1\cdots n]\}$. One can look into a homogeneous neighborhood system of the form

$$\mathcal{G}=\mathcal{F}_c=\left\{\mathcal{F}_{i,j}^c,(i,j)\in S\right\}$$

$$\mathcal{F}_{i,j}^c=\left\{(k,l)\in S|0<(k-i)^2+(l-j)^2<c\right\}.$$

Notice that the sites near the boundary have fewer neighbors than the interior ones. The cliques for $c=1$ are all of the form $\{(i,j)\}$, $\{(i,j),(i,j+1)\}$, $\{(i,j),(i+1,j)\}$ as shown in Fig. 1(a), and for $c=2$ we have the previously mentioned and the additional cliques described in Fig. 1(b). The cost function:

$$U(p,q)=\sum_{i,j}\epsilon^2\left(E_{ij}-R(p_{ij},q_{ij})\right)^2 + \kappa\left[(p_{i,j+1}-p_{i,j})^2+(p_{i+1,j}-p_{i,j})^2 + (q_{i,j+1}-q_{i,j})^2+(q_{i+1,j}-q_{i,j})^2\right] \quad (7)$$

can be broken into a sum of energy terms corresponding to a nearest neighbor system. The only cliques we need to consider correspond to $c=1$. Thus we can write

$$U(p,q)=\sum_{c\in\mathcal{C}} V_C(p,q).$$

Where $\mathcal{C}$ denote the set of cliques for $\mathcal{G}=\mathcal{F}_1$. In general the stencil we use for finite difference approximation to differential operators determines the spatial dependence of the cost function. It is reflected in the type of cliques we have to use when we break the cost function into its primal components. Thus for (6) the cliques consist of one or two elements corresponding to $c=2$. In (7) the spatial dependence corresponds to $c=1$ and the additional cliques consist of at most two elements. Note that to each pixel corresponds a two-element vector of unknowns: $(p_{i,j},q_{i,j})$. The unknowns are coupled together in the Gibbs energy function. This energy function is very similar to energy functions of coupled Markov Processes [2]. Minimizing our cost function is totally equivalent to solving for the configuration with maximum probability. A deterministic algorithm for solving this problem was suggested in [20], in which Besag minimizes the cost function for a grid point (i,j) given the values of its neighbor set. The algorithm uses only local computation, and can be implemented in parallel. The problem with using the algorithm when the unknowns can take values in a continuous set, is that the computations involved in minimizing the energy of the conditional distribution are not trivial. Thus this algorithm is not practical for our problem. All the deterministic algorithms have one limitation in common—they all guarantee convergence only to a local minimum. To avoid the difficulty we assumed that the minimum is unique. This is not always true, as discussed in [21]. We can use our formulation to deal with the multiple minima case. An algorithm for finding the global minimum of a Gibbs energy function was suggested by Geman and Hwang in [22], using an algorithm based on a simulated diffusion process. The algorithm performs a gradient search combined with a Brownian motion component, which is gradually attenuated to zero. This search can be implemented efficiently, on the architecture suggested in this paper. The only modification required, is adding a white noise component to each element of the computed gradient.

B. The Multiresolution Conjugate-Gradient Algorithm Implemented on a Pyramid

We first calculate the gradient corresponding to (5):

$$\frac{\partial\mu}{\partial p_{i,j}}=2\epsilon^2\left(E_{i,j}-R_{i,j}\right)\frac{\partial R_{i,j}}{\partial p_{i,j}}+\kappa\left(4p_{i,j}-\hat{p}_{i,j}\right) \quad (8)$$

where

$$\hat{p}_{i,j}=p_{i,j+1}+p_{i,j-1}+p_{i+1,j}+p_{i-1,j}$$

and

$$\frac{\partial \mu}{\partial q_{i,j}} = 2\epsilon^2\left(E_{i,j} - R_{i,j}\right)\frac{\partial R_{i,j}}{\partial q_{i,j}} + \kappa\left(4q_{i,j} - \hat{q}_{i,j}\right) \quad (9)$$

where

$$\hat{q}_{i,j} = q_{i,j+1} + q_{i,j-1} + q_{i+1,j} + q_{i-1,j}$$

for the Lambertian reflectance map, and light source located at the camera, (8) and (9) become:

$$\frac{\partial \mu}{\partial p_{i,j}} = 2\epsilon^2\left(E_{i,j} - \frac{1}{\left(1 + p_{i,j}^2 + q_{i,j}^2\right)^{\frac{1}{2}}}\right)\frac{p_{i,j}}{\left(1 + p_{i,j}^2 + q_{i,j}^2\right)^{\frac{3}{2}}} + \kappa\left(4p_{i,j} - \hat{p}_{i,j}\right)$$

$$\frac{\partial \mu}{\partial q_{i,j}} = 2\epsilon^2\left(E_{i,j} - \frac{1}{\left(1 + p_{i,j}^2 + q_{i,j}^2\right)^{\frac{1}{2}}}\right)\frac{q_{i,j}}{\left(1 + p_{i,j}^2 + q_{i,j}^2\right)^{\frac{3}{2}}} + \kappa\left(4q_{i,j} - \hat{q}_{i,j}\right). \quad (10)$$

Note that the spatial communication required is only with the nearest neighbors. Next we note that if the boundary and initial conditions are trivial, the gradient is zero—we are at a local maximum of the energy function. We can now use the previous example to illustrate the importance of enforcing integrability in the gradient search. One can see that if we start with an initial condition that satisfies ($p = q$) and the condition holds on the boundary, the search is then limited to the subspace ($p = q$), because the gradient is the same for p and q in every iteration. This problem is solved when the integrability term is added to obtain:

$$\frac{\partial \mu}{\partial p_{i,j}} = 2\epsilon^2\left[E_{i,j} - \frac{1}{\left(1 + p_{i,j}^2 + q_{i,j}^2\right)^{\frac{1}{2}}}\right]\frac{p_{i,j}}{\left(1 + p_{i,j}^2 + q_{i,j}^2\right)^{\frac{3}{2}}} + \kappa\left(4p_{i,j} - \hat{p}_{i,j}\right) + \frac{\lambda}{\epsilon^2}\left(e_{i,j} + e_{i,j-1} - e_{i-1,j-1} - e_{i-1,j}\right)$$

$$\frac{\partial \mu}{\partial q_{i,j}} = 2\epsilon^2\left[E_{i,j} - \frac{1}{\left(1 + p_{i,j}^2 + q_{i,j}^2\right)^{\frac{1}{2}}}\right]\frac{q_{i,j}}{\left(1 + p_{i,j}^2 + q_{i,j}^2\right)^{\frac{3}{2}}} + \kappa\left(4q_{i,j} - \hat{q}_{i,j}\right) + \frac{\lambda}{\epsilon^2}\left(e_{i,j-1} + e_{i-1,j-1} - e_{i,j} - e_{i-1,j}\right)$$

where $e_{i,j}$ was defined in (4). We now give the algorithm as follows.

1) Set the initial resolution to $l = 2^k$ and choose an initial solution on the coarse grid.
2) Given the initial configuration p^0, q^0 compute
$$G_0 = \nabla\mu\left(p^0, q^0\right) \text{ and set } D_0 = -G_0.$$
3) Calculate the optimal step for the given D_k using the global energy function (at the top of the pyramid) as described in the next paragraph.
4) Perform the descent step
$$p^{k+1} = p^k + \alpha_k D_k^p$$
$$q^{k+1} = q^k + \alpha_k D_k^q.$$
5) Calculate the gradient vector G_{k+1} for the given resolution (at the bottom of the pyramid).
6) Calculate a new conjugate direction
$$D_{k+1} = -G_{k+1} + \beta_k D_k$$
where
$$\beta_k = \frac{\left(G_{k+1} - G_k\right)^T G_{k+1}}{G_k^T G_k}.$$
7) Repeat Steps 3)–6) until either the reduction in the global energy is below a threshold or a fixed number of iterations have been completed.
8) Stop if you have reached the desired resolution.
9) Refine the resolution ($l = 2^{k+1}$), perform a local coarse to fine extension of the solution to obtain an initial configuration for the finer level.
10) Go back to Step 2).

Step 3) of the algorithm is implemented in the following manner. The top of the pyramid performs univariate minimization using an iterative algorithm, such as golden section search [23], to determine the optimal step. The algorithm uses the values of the cost function corresponding to different step sizes. In each iteration, a new value for the step size is propagated down the pyramid to the bottom layer, so that the processors on the bottom can calculate the updated values of p and q and then sum the energy terms associated with the cliques that correspond to their location. In order to avoid repeated terms in the sum, each energy term is weighted by a reciprocal of the number of elements that construct the corresponding clique. Then, the intermediate levels of the pyramid sum the energy terms corresponding to their "sons," in order to obtain the global energy for the configuration at the top. When the optimal step is determined at the top, it is propagated down to the bottom layer, and Step 4) is performed.

C. Experimental Results

The experiments were performed on a synthetic sphere image illuminated from the previously mentioned. The two gradients corresponding to the cost functions (6), (7) were used to obtain a reconstruction of the surface, assuming a Lambertian reflectance map, and constant albedo. The parameters λ and κ were chosen experimentally. We found that the rate of convergence is highly affected by the choice of λ and κ. Setting the smoothing part to small values slowed down the convergence rate significantly. As expected if smoothing dominated, the results started to differ from the original surface, because we have solved a different problem. The results for (7) were obtained using ten or less iterations at each resolution, starting with a 4×4 grid and finishing at 64×64 and 128×128 grids. The L_2 error is $O(10^{-2})$. When we decreased the smoothing

TABLE I
COMPARISON OF SINGLE AND MULTIRESOLUTION ALGORITHMS TO SHAPE FROM SHADING

Grid Size	Resolution	Iteration	Energy	L_2 Error$\times 10^{-2}$
64	multi	40	2.960476	2.12969
	single	104	2.888647	2.11380
128	multi	50	2.876788	1.15690
	single	252	2.851924	2.09900

part, better results were obtained and the accuracy improved to $O(10^{-3})$.

We repeated the first experiment with a single-resolution conjugate gradient search. The results are summarized in Table I. The experiment supports the belief that multiple resolutions increase the rate of convergence significantly.

III. Edge Detection Using the Graduated Nonconvexity Algorithm

Edge detection can be viewed as the problem of fitting a weak membrane (that is, an elastic membrane under weak continuity constraints) to a surface [9]. The location of discontinuities in the weak membrane correspond to discontinuities in the intensity (step edges). The problem has the following mathematical form. Find the configuration that minimizes the following functional corresponding to a weak membrane with a simple line process

$$F = \int \left\{ (u-d)^2 + \lambda^2 (\nabla u)^2 \right\} dA + \alpha \int dl \qquad (11)$$

where u is the estimated membrane height and d is the noisy data. The first integral is evaluated over the area in which the data d is defined, and the second along the length of all discontinuities. α is a penalty per unit length of discontinuity. λ is a characteristic length for smoothing the continuous portions of the data and is also a characteristic distance for interaction between discontinuity. Since the data is only available on grid points, and the problem does not have a closed form solution, we need to discretize the problem. Blake [9] suggests minimizing the cost function as

$$F = D + \sum_{ij} h_{\alpha,\lambda}(u_{i,j} - u_{i-1,j}, l_{ij}) + \sum_{ij} h_{\alpha,\lambda}(u_{i,j} - u_{i,j+1}, m_{ij}) \qquad (12)$$

where

$$D = \sum_{ij} (u_{i,j} - d_{i,j})^2 \qquad (13)$$

l_{ij} activates the line process in the northernly direction and

$$h_{\alpha,\lambda}(t,l) = \lambda^2 (t)^2 (1-l) + \alpha l$$

m_{ij} activates the line process in the easternly direction. The problem is thus reduced to the following optimization problem

$$\min_{\{u_{ij}\}} \left(D + \min_{\{l_{ij}\}} \left(\sum_{ij} h_{\alpha,\lambda}(u_{i,j} - u_{i-1,j}, l_{ij}) \right) + \min_{\{m_{ij}\}} \left(\sum_{ij} h_{\alpha,\lambda}(u_{i,j} - u_{i,j+1}, m_{ij}) \right) \right).$$

As D does not involve l_{ij}, m_{ij}, minimization over l_{ij}, m_{ij} can be performed and one is then left with minimization with respect to u_{ij}:

$$\min_{\{u_{ij}\}} F = \min_{\{u_{ij}\}} \left(D + \sum_{ij} g_{\alpha,\lambda}(u_{i,j} - u_{i-1,j}) + \sum_{ij} g_{\alpha,\lambda}(u_{i,j} - u_{i,j+1}) \right)$$

where

$$g_{\alpha,\lambda}(t) = \min_{l \in \{0,1\}} h_{\alpha,\lambda}(t,l) = \min\left(\lambda^2 (t)^2, \alpha\right).$$

By applying a convex approximation to F Blake obtains:

$$F^* = D + \sum_{ij} g^*_{\alpha,\lambda}(u_{i,j} - u_{i-1,j}) + \sum_{ij} g^*_{\alpha,\lambda}(u_{i,j} - u_{i,j+1}) \qquad (14)$$

where

$$g^*_{\alpha,\lambda}(t) = \begin{cases} \lambda^2 (t)^2, & |t| < q \\ \alpha - c^*(|t| - r)^2/2, & q \le |t| < r \\ \alpha, & |t| \ge r \end{cases}$$

where

$$r^2 = \alpha\left(\frac{2}{c^*} + \frac{1}{\lambda^2}\right), \quad q = \frac{\alpha}{\lambda^2 r}, \qquad \text{and } c^* = \frac{1}{4} \text{ for membrane.}$$

A one parameter family of cost functions $F^{(P)}$ is then defined by replacing g^* in (14) by $g^{(P)}$. $g^{(P)}$ is similar to g^* except that c^* is replaced by a variable c^{**}, that varies with P as

$$F^{(P)} = D + \sum_{ij} g^{(P)}_{\alpha,\lambda}(u_{i,j} - u_{i-1,j}) + \sum_{ij} g^{(P)}_{\alpha,\lambda}(u_{i,j} - u_{i,j+1}) \qquad (15)$$

with

$$g^{(P)}_{\alpha,\lambda}(t) = \begin{cases} \lambda^2 (t)^2, & |t| < q \\ \alpha - c^{**}(|t| - r)^2/2, & q \le |t| < r \\ \alpha, & |t| \ge r \end{cases}$$

where

$$c = \frac{c^*}{P}, \quad r^2 = \alpha\left(\frac{2}{c^{**}} + \frac{1}{\lambda^2}\right), \quad \text{and} \quad q = \frac{\alpha}{\lambda^2 r}.$$

The GNC algorithm begins by minimizing $F^{(P=1)} = F^*$. Then P is decreased from 1 to 0, which makes $g^{(P)}$ change steadily from g^* to g. For every value of P we minimize $F^{(P)}$ starting with the last configuration corresponding to the previous P (local minimum of $F^{(2P)}$). We suggest that minimization of $F^{(P)}$ can be performed efficiently using

the optimal step conjugate gradient algorithm described in Section II. The conjugate gradient algorithm takes fewer number of iterations than SOR, but each iteration requires substeps to determine the optimal step. The conjugate gradient search is less sensitive to the noise present in the image. The number of iterations increases only by 40 percent in the presence of 0 dB noise, while the SOR algorithm requires double the number of iterations. We view $F^{(P)}$ as a Gibbs energy function for the global configuration. The energy function is then broken into terms that correspond to cliques of a nearest neighbor system. The zero order terms are of the form

$$(u_{ij} - d_{ij})^2.$$

The energy terms corresponding to cliques of two elements e.g., $(i, j)(i, j+1)$ is of the form

$$g_{\alpha,\lambda}^{(P)}(t) = \begin{cases} \lambda^2(t)^2, & |t| < q \\ \alpha - c^{**}(|t| - r)^2/2, & q \le |t| < r \\ \alpha, & |t| \ge r \end{cases}$$

where $t = (u_{i,j} - u_{i-1,j})$ in the x direction and $t = (u_{i,j} - u_{i,j+1})$ in the y direction. Note that the energy term is a nonlinear function of the cliques' elements. We proceed by calculating the gradient of $F^{(P)}$ for points in the interior

$$\frac{\partial F^{(P)}}{\partial u_{ij}} = 2(u_{i,j} - d_{i,j}) + g_{\alpha,\lambda}^{(P)'}(u_{i,j} - u_{i,j-1}) + g_{\alpha,\lambda}^{(P)'}(u_{i,j} - u_{i-1,j}) + g_{\alpha,\lambda}^{(P)'}(u_{i,j} - u_{i,j+1}) + g_{\alpha,\lambda}^{(P)'}(u_{i,j} - u_{i+1,j}) \quad (16)$$

where

$$g_{\alpha,\lambda}^{(P)'}(t) = \begin{cases} 2\lambda^2(t), & |t| < q \\ -c^{**}(|t| - r)\sin(t), & q \le |t| < r. \\ 0, & |t| \ge r \end{cases}$$

On the boundary we need a modification, for the corner $i = 0,\ j = 0$:

$$\frac{\partial F^{(P)}}{\partial u_{00}} = 2(u_{0,0} - d_{0,0}) + g_{\alpha,\lambda}^{(P)'}(u_{0,0} - u_{0,1}) + g_{\alpha,\lambda}^{(P)'}(u_{0,0} - u_{1,0}) \quad (17)$$

and similarly at the other corners. For the boundary side $i = 0,\ j = 1, \cdots, N-1$:

$$\frac{\partial F^{(P)}}{\partial u_{ij}} = 2(u_{0,j} - d_{0,j}) + g_{\alpha,\lambda}^{(P)'}(u_{0,j} - u_{0,j-1}) + g_{\alpha,\lambda}^{(P)'}(u_{0,j} - u_{0,j+1}) + g_{\alpha,\lambda}^{(P)'}(u_{0,j} - u_{1,j}) \quad (18)$$

and similarly for the other sides.

We now have all the building blocks required for our algorithm, presented next.

1) Choose λ, h_0 (scale and sensitivity).
2) Set $\alpha = h_0^2\lambda/2$.
3) Set $P = 1.0$.
4) Given an initial configuration u^0 compute $G_0 = \nabla F^P(u^0)$ and set $D_0 = -G_0$.
5) Calculate (using a univariate search) the optimal step for the given D_k using the global energy function (at the top of the pyramid).
6) Perform the descent step
$$u^{(k+1)} = u^{(k)} + \alpha_k D_k$$
7) Calculate the gradient vector $G_{k+1} = \nabla F^P(u^{(k+1)})$ (at the bottom of the pyramid).
8) Calculate a new conjugate direction
$$D_{k+1} = -G_{k+1} + \beta_k D_k$$
where
$$\beta_k = \frac{(G_{k+1} - G_k)^T G_{k+1}}{G_k^T G_k}.$$
9) Repeat Steps 5)–8) until $\max_{i,j}|u_{i,j}^{(k+1)} - u_{i,j}^{(k)}| < \epsilon$, or the energy reduction is below some level.
10) If $P > c^*/\lambda$ then $P = P/2u^0 = u^{(k+1)}$ go back to Step 4).
11) Calculate the edge location using the following rules:

$$l_{i,j} = \begin{cases} 1, & \text{if } |u_{i,j} - u_{i-1,j}| > r \\ 0, & \text{if } |u_{i,j} - u_{i-1,j}| < q \\ \text{ambiguous}, & \text{otherwise} \end{cases}$$

$$m_{i,j} = \begin{cases} 1, & \text{if } |u_{i,j} - u_{i,j+1}| > r \\ 0, & \text{if } |u_{i,j} - u_{i,j+1}| < q \\ \text{ambiguous}, & \text{otherwise.} \end{cases}$$

The $l_{i,j}$ and $m_{i,j}$ are the edge indicator functions in the northernly and easternly directions respectively.

A. Experimental Results

1) Edge Detection using Conjugate Gradient: The experiments in this section were performed on a 128×128 airport image. We used the Polak–Ribiere conjugate gradient search. Edge detection was performed on the original image and the original image corrupted by additive Gaussian noise. The conjugate gradient method proved to be robust to noise in the sense that the convergence properties were not affected considerably by the amount of noise in the processed image. In each step of the GNC algorithm the termination criterion used was: stop if $f(x_k) - f(x_{k+1}) \le 100$. We preferred this test over the one: stop when $\|u_{k+1} - u_k\|_\infty \le a$ (where a is some constant), because it requires a smaller number of iterations without degradation in the computed edges. The GNC steps were performed till $p = c^*/\lambda$.

Experimental results are summarized in Fig. 2. Figs. 2(a) and 2(b) are the original airport image and the result obtained with $h = 10$, $\lambda = 4$ after 157 iterations. Figs. 2(c) and 2(d) are the image with 5-dB noise ($\sigma_n = 12$), and the result obtained with $h = 16$, $\lambda = 4$, where the number of

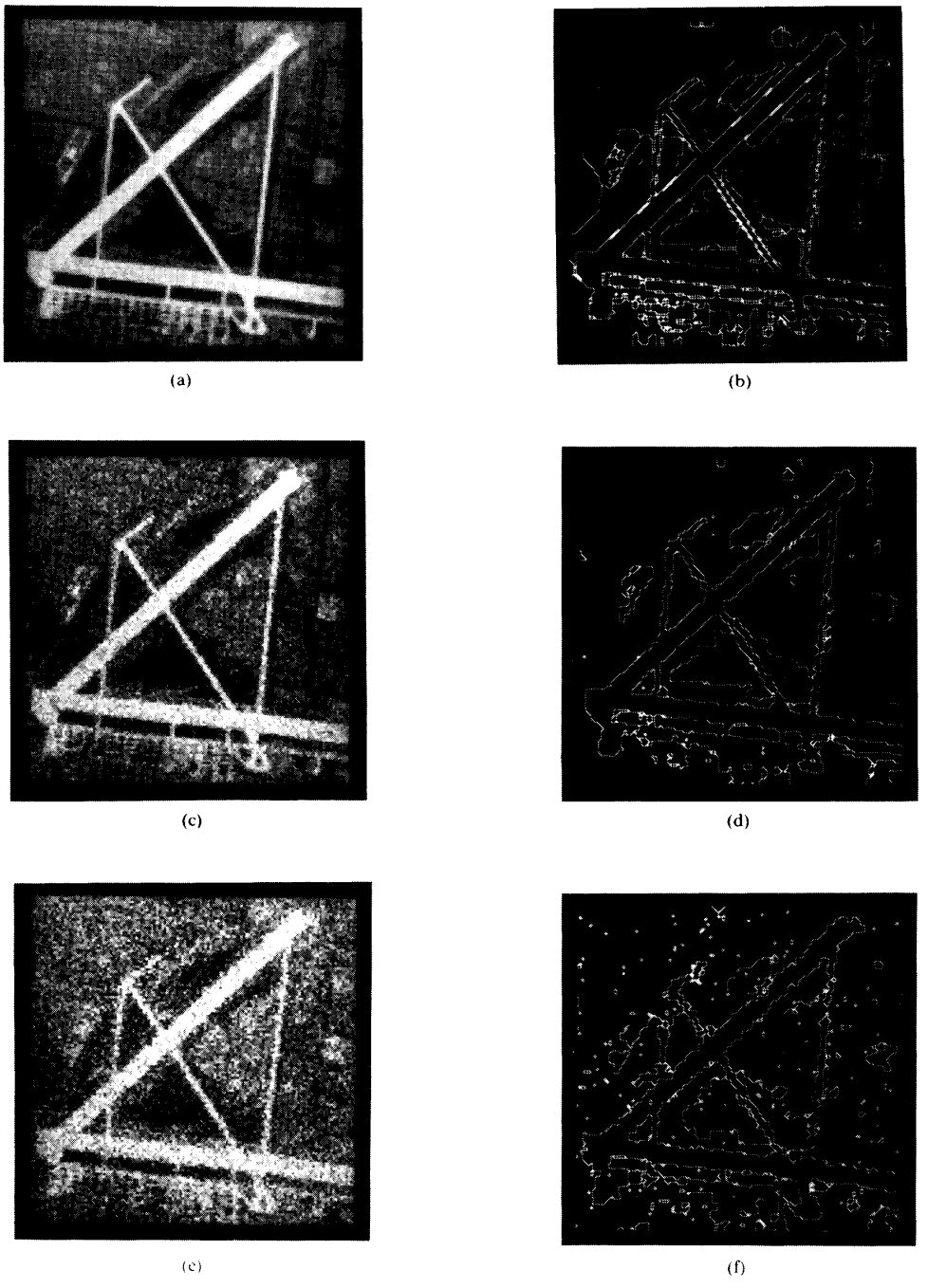


Fig. 2. Edge detection using conjugate gradient-based GNC algorithm. (a) Original airport image. (b) Edge estimate on original airport image. (c) Airport image corrupted by 5-dB additive noise. (d) Edge estimate on 5-dB noisy airport image. (e) Airport image corrupted by 0-dB additive noise. (f) Edge estimate on 0-dB noisy airport image.

iteration increased to 198. Figs. 2(e) and 2(f) are the image with 0-dB noise ($\sigma_n = 22$), and the result obtained with $h = 20$, $\lambda = 4$ with the algorithm terminating after 224 iterations.

2) Comparison between Conjugate Gradient and SOR: We first compared the sequential SOR and the black and red (parallel) SOR. The number of iterations and the final results were the same for these two algorithms. In order to compare the SOR and the conjugate gradient we used several images with different noise levels. The number of iterations in the SOR increases considerably when the noise level is increased, while the number of iterations in the conjugate gradient method is not so sensitive to the noise. The number of iterations in the SOR is greater than

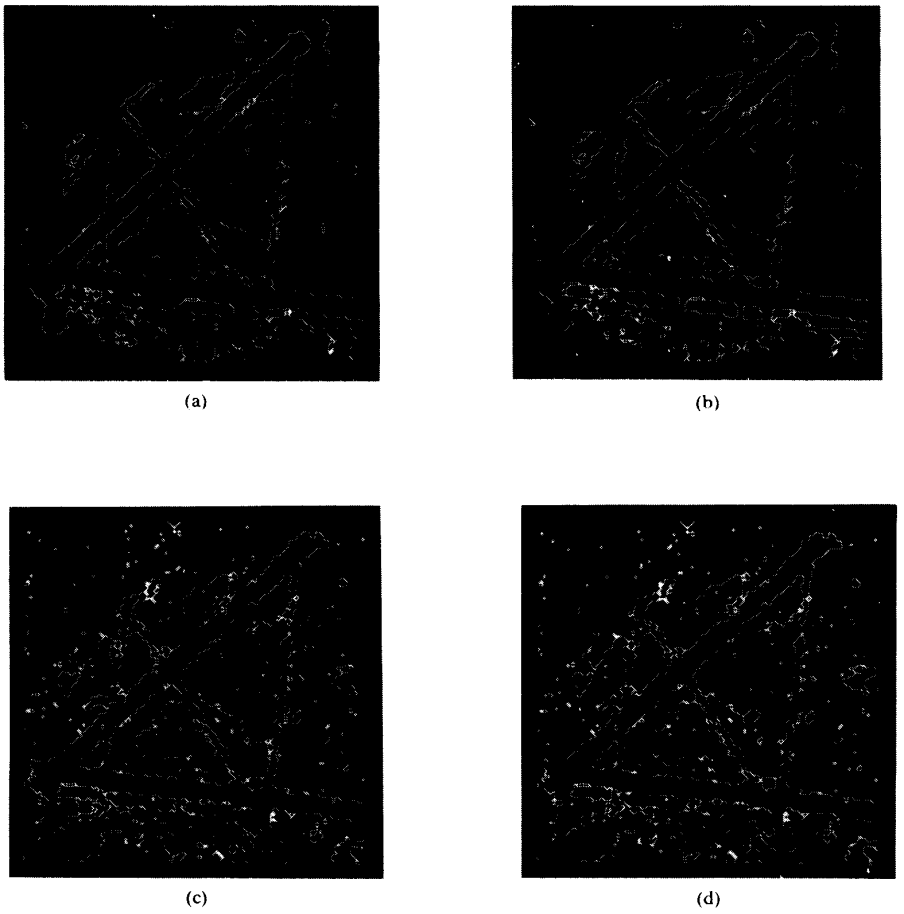


Fig. 3. Comparison between SOR and conjugate-gradient GNC results. (a) SOR edge estimate on 5-dB image in Fig. 2(c). (b) CG edge estimate on 5-dB image in Fig. 2(c). (c) SOR edge estimate on 0-dB image in Fig. 2(e). (d) CG edge estimate on 0-dB image in Fig. 2(e).

that of the conjugate gradient, but in each conjugate gradient step we perform more computations. The final results for these two methods are the same for all the noise levels. In Table II we summarize the results.

In Fig. 3 we present the comparison results for the two algorithms. Figs. 3(a) and 3(c) are the results from the 5-dB and 0-dB images using SOR algorithm, where Figs. 3(b) and 3(d) were obtained by using the conjugate gradient algorithm.

TABLE II
COMPARISON BETWEEN CONJUGATE GRADIENT AND SOR ALGORITHMS

Image	Algorithm	Iterations	Final Energy $\times 10^6$	Undefined $l_{i,j}, m_{i,j}$
Airport	SOR	328	1.3101	336
	CG	157	1.3088	336
Airport	SOR	477	4.1830	279
with 5-dB noise	CG	198	4.1814	253
Airport	SOR	617	8.9788	244
with 0-dB noise	CG	224	8.9722	227

IV. Graduated Nonconvexity Algorithm for Image Estimation Using Compound GMRF Models

Image estimation of gray-level images requires the ability of the estimate to preserve the sharpness of edges, which play an important role in human interpretation of images. Linear filters [24], [25], as well as MAP estimation using the homogeneous Gauss–Markov random field (GMRF) model [26] for the intensity tend to smear the edges. To avoid this problem a class of models called compound GMRF was suggested in [27]. The model suggested is a GMRF coupled with a line process. In this model the structure of the line process determines the value of the GMRF model coefficients. The GMRF parameters are required to satisfy some symmetry constraint, and the choice of the parameters for a given line configuration remains an open problem. Our simplified compound GMRF model is a member of the class of models proposed in [27]. In our formulation the line process breaks the correlation between neighboring pixels when the gray-level jump is above a threshold. The model can be viewed as an extension of the weak membrane model [9] used for sur-

face interpolation and edge detection. The GMRF breaks when edges occur, which in turn creates homogeneous GMRF patches separated by the line process. Qualitative and quantitative values correspond to all the new model parameters, and they can be estimated from the image. As an example the regularization term in the weak membrane model is replaced by the ratio of the noise variance to the GMRF model variance.

Given the distribution of the corrupting noise we write the conditional density of the degraded image conditioned on the original image. The degradation we consider in this work is additive white Gaussian noise. Thus the posterior density is written in terms of the GMRF model parameter estimates.

To obtain the global MAP estimate for an image obeying a compound GMRF model and corrupted by additive Gaussian noise, Jeng and Woods [27] use simulated annealing. This stochastic algorithm is slow, and can not in practice obey the theoretical requirement on the initial temperature. We use a modified version of the GNC algorithm [9]. The algorithm is deterministic and good results are obtained in less than 100 iterations. We have used the algorithm on a real airport image and obtained good restoration results. The configuration of the line process is determined by the MAP estimate. It provides an accurate estimate of the image edges without any additional cost.

In this section we also address the problem of estimating the GMRF model parameters from a noise-free realization of the original image. In order to avoid the edges one can limit the estimation domain only to homogeneous regions, which will make maximum likelihood (ML) estimation impractical. Instead we have developed the following procedure. We detect the edges on the image using the GNC algorithm [9] and discard the intensity data in a strip of four pixels centered at each edge. We develop an EM algorithm [28] that calculates the conditional expectation of the intensity given the parameters of the model and the neighboring intensity levels that is used to replace the discarded data, and then calculate the new parameters estimate from the smoothed image using least squares (LS) [25] or ML techniques. The process is repeated till it converges. As initial parameters we used the LS estimates of the image including the edges. The current version includes only results based on LS, and we are implementing the ML version. The parameters we obtain have much lower model variance then the parameters estimated from the image with the edges, although the image variance reduced only slightly. In the future we would like to extend this process to estimation from the noisy image itself.

A. Image and Noise Models

The GMRF models have been successfully used for modeling homogeneous intensity regions, such as natural texture (sand, grass, calf) [29], [30]. The GMRF model is not suitable for composite real-world scenes because it is unable to capture sharp transitions (edges). These problems can be partially overcome by introducing composite models [2], [27]. We develop a composite model able to describe homogeneous regions as well as sharp transitions from one region to the other. The model is an extension of the GMRF models. We first review the GMRF model and then extend it to the compound GMRF.

1) GMRF Model: Suppose that the original image $\{y(s)\}$ is defined on a finite lattice $\Omega = \{s = (i, j): 1 \le i, j \le M\}$. Consider a stationary Gauss–Markov model with mean y_m, parameters $\Theta_{s,r} = \Theta_{\{s-r\}}$ and toroidal boundaries for which

$$p(y(s)|y(r), r \in \Omega) = p(y(s)|y(r), r \in N_s)$$
$$= \frac{1}{\sqrt{2\pi\nu}} \exp\left\{ -\frac{1}{2\nu}\left[y(s) - y_m - \sum_{r \in N_s} \Theta_{s,r}(y(r) - y_m)\right]^2\right\} \tag{19}$$

where N_s characterizes the neighborhood dependence. This leads to the joint density function [31]:

$$p(y) = \frac{1}{\left\{(2\pi)^{M^2} \det\left(\nu B(\Theta)^{-1}\right)\right\}^{\frac{1}{2}}} \cdot \exp\left\{ -\frac{1}{2\nu}(y - y_m)' B(\Theta)(y - y_m)\right\} \tag{20}$$

where $B(\Theta)$ is a $M^2 \times M^2$ block-Toeplitz matrix, and y_m is the $M^2 \times 1$ vector of the constant mean y_m.

2) The Compound GMRF Model: In this section we extend the weak membrane (12) and GMRF (19) models to a compound GMRF and line process model. Since the parameters corresponding to the compound GMRF model can be estimated from the noisy image using bias compensated least square (BCLS) techniques [32], this extension enables us to get better reconstruction of weakly continuous surfaces. It also gives a qualitative meaning to λ in the weak membrane algorithm and allows us to estimate its value from the noise and model variances.

We define the following conditional distribution for the compound GMRF model as

$$p(y(s)|y(s+\tau), y(s-\tau), l(s,\tau), l(s-\tau,\tau), \tau \in N^*) = \frac{e^{-u(y(s)|y(s+\tau), y(s-\tau), l(s,\tau), l(s-\tau,\tau), \tau \in N^*)}}{Z}$$

where N^* is the set of shift vectors corresponding to the neighborhood of the GMRF model.

The line process notation is illustrated in Fig. 4. For a second order GMRF model $N^* = \{(0,1), (1,0), (1,1), (-1,1)\}$:

$$U(y(s)|y(s+\tau), y(s-\tau), l(s,\tau), l(s-\tau,\tau), \tau \in N^*) = \frac{1}{2\nu}\left[\sum_{\tau \in N^*} \Theta_\tau \left[(y(s) - y(s+\tau))^2 (1 - l(s,\tau)) + (y(s) - y(s-\tau))^2 (1 - l(s-\tau,\tau))\right] + \left(1 - \sum_{\tau \in N^*} 2\Theta_\tau\right) y(s)^2 \right].$$

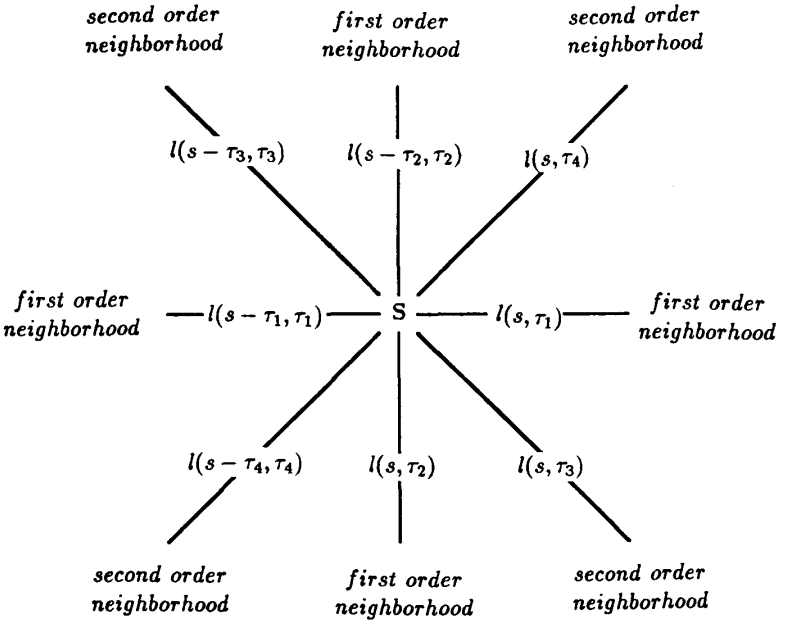


Fig. 4. Line process notation. $\tau_1=(0,1)$, $\tau_2=(1,0)$, $\tau_3=(1,1)$, $\tau_4=(-1,1)$.

The line process is now defined on the edges that connect the nodes (the grid points) that are neighbors of a given pixel.

The prior probability of the line process is

$$p\left(l(s,\tau_i)|l(r,\tau), r\in\Omega, \tau\in N^*, (r,\tau)\neq(s,\tau_i)\right) = \frac{e^{-U(l(s,\tau_i)|l(r,\tau), r\in\Omega,\tau\in N^*,(r,\tau)\neq(s,\tau_i))}}{Z}$$

where

$$U\left(l(s,\tau_i)|l(r,\tau), r\in\Omega, \tau\in N^*, (r,\tau)\neq(s,\tau_i)\right) = \beta l(s,\tau_i).$$

Recall from (19) that for a zero-mean GMRF model

$$U\left(y(s)|y(s+\tau), y(s-\tau), \tau\in N^*\right) = \frac{1}{2\nu}\left(y(s)-\sum_{\tau\in N^*}\Theta_\tau[y(s+\tau)+y(s-\tau)]\right)^2.$$

It is obvious that the conditional distribution of the compound GMRF model is equal to the GMRF model when $l(s,\tau)=0$, $l(s-\tau,\tau)=0$, for all $\tau\in N^*$. We shall assume that all $\Theta_\tau>0$, which is almost always the case for the first order neighborhoods.

The corresponding joint distribution of the compound GMRF is

$$p(y,l)=\frac{e^{-U(y,l)}}{Z}$$

where

$$U(y,l) = \sum_{s\in\Omega}\left\{\sum_{\tau\in N^*}\Theta_\tau\left[\frac{1}{2\nu}(y(s)-y(s+\tau))^2(1-l(s,\tau)) + \beta l(s,\tau)\right] + \frac{1}{2\nu}\left(1-\sum_{\tau\in N^*}2\Theta_\tau\right)y(s)^2\right\}.$$

One can see that if $(y(s)-y(s+\tau))$ is bigger than a threshold, it is cheaper to break the connection and introduce an edge in the same way it is done in the weak membrane formulation.

The Noise Model and the MAP Equation: We consider the following model for image intensity degraded by additive Gaussian noise [33]

$$x(s)=y(s)+n(s). \tag{21}$$

Since $n(s)$ is a spatially independent Gaussian noise variable, we can express the conditional density function $p(\boldsymbol{x}|\boldsymbol{y})$ as

$$p(\boldsymbol{x}|\boldsymbol{y})=\frac{1}{\left(2\pi\sigma_n^2\right)^{M^2/2}}\exp\left\{-\frac{1}{2\sigma_n^2}\sum_{s\in\Omega}(x(s)-y(s))^2\right\}. \tag{22}$$

Knowing the statistical properties of the original image $\boldsymbol{y}$ and the degradations, one can construct the *aposteriori* density function $p(\boldsymbol{y}|\boldsymbol{x})$ from the observation $\boldsymbol{x}$. Thus Bayes law leads to the *aposteriori* density

$$p(\boldsymbol{y}|\boldsymbol{x})=\frac{p(\boldsymbol{x}|\boldsymbol{y})p(\boldsymbol{y})}{p(\boldsymbol{x})} \tag{23}$$

where $\boldsymbol{y}$ is the original image we wish to estimate. The

MAP estimate is obtained by maximizing the *aposteriori* density function $p(\boldsymbol{y}|\boldsymbol{x})$ to find the most probable value of $\boldsymbol{y}$. Since the observation $\boldsymbol{x}$ is given, $p(\boldsymbol{x})$ can be treated as a constant, and we have

$$p(\boldsymbol{y}|\boldsymbol{x}) = \boldsymbol{K}p(\boldsymbol{x}|\boldsymbol{y})\,p(\boldsymbol{y}) \tag{24}$$

where constant $\boldsymbol{K}$ is equal to $1/p(\boldsymbol{x})$. Using the Gibbs formulation the global energy corresponding to the compound GMRF model with additive noise is

$$\begin{aligned} &U(\boldsymbol{y},\boldsymbol{l}|\boldsymbol{x}) \\ &= \sum_{s\in\Omega}\left\{\frac{(x(s)-y(s))^2}{2\sigma^2}+\frac{\left(1-\sum_{\tau\in N^*}2\Theta_\tau\right)y(s)^2}{2\nu}\right. \\ &\quad+\sum_{\tau\in N^*}\Theta_\tau\left[\frac{1}{2\nu}(y(s)-y(s+\tau))^2(1-l(s,\tau))\right. \\ &\quad\left.\left.+\beta l(s,\tau)\right]\right\} \\ &= \frac{1}{2\sigma^2}\sum_{s\in\Omega}\left\{(x(s)-y(s))^2\right. \\ &\quad+\lambda^2\left(1-\sum_{\tau\in N^*}2\Theta_\tau\right)y(s)^2 \\ &\quad+\sum_{\tau\in N^*}\Theta_\tau\left[\lambda^2(y(s)-y(s+\tau))^2(1-l(s,\tau))\right. \\ &\quad\left.\left.+\alpha l(s,\tau)\right]\right\} \end{aligned} \tag{25}$$

where $\lambda^2=\sigma^2/\nu$, $\alpha=2\beta\sigma^2$. The regularization parameter λ that did not have a quantitative value in the weak membrane model, has both qualitative meaning and quantitative value in the new model. It reflects the confidence we have in the data, as it is the ratio of the measurement noise variance to the GMRF model variance.

B. The GNC–GMRF Algorithm

One can see the similarity between the global energy function corresponding to the posterior density in (25) and the global energy Blake suggested to minimize in the weak membrane formulation (12). The weak membrane model is a special case of the compound GMRF model with a first order neighborhood system and isotropic parameters $\Theta_x=\Theta_y=0.25$ and with appropriate scaling of λ and α. Blake has developed the GNC algorithm to find the global minimum of (12). We modify the GNC algorithm to obtain an algorithm that is able to find a near optimal MAP solution for the compound GMRF model. For simplicity in the formulation we restrict our attention to a first order compound GMRF model although there are no difficulties in deriving the algorithm for higher order neighborhood systems. For the first order compound GMRF model we can write

$$\begin{aligned} U(\boldsymbol{y},\boldsymbol{l},\boldsymbol{m}|\boldsymbol{x}) = \sum_{i,j\in\Omega}\frac{1}{2\sigma^2}\Big\{&(x_{i,j}-y_{i,j})^2 \\ &+\lambda^2\left(1-2(\Theta_x+\Theta_y)\right)y_{i,j}^2 \\ &+\Theta_x\left[\lambda^2(y_{i,j}-y_{i-1,j})^2(1-l_{i,j})+\alpha l_{i,j}\right] \\ &+\Theta_y\left[\lambda^2(y_{i,j}-y_{i,j+1})^2(1-m_{i,j})+\alpha m_{i,j}\right]\Big\} \end{aligned}$$

where $l_{i,j}, m_{i,j}$ activate the line process in the x and y directions respectively. We can write the posterior energy function in the form

$$U=\frac{1}{2\sigma^2}\left\{D+\sum_{ij}\Theta_x h_{\alpha,\lambda}(y_{i,j}-y_{i-1,j},l_{ij})+\sum_{ij}\Theta_y h_{\alpha,\lambda}(y_{i,j}-y_{i,j+1},m_{ij})\right\} \tag{26}$$

where

$$D=\sum_{ij}(y_{i,j}-x_{i,j})^2+\lambda^2\left(1-2(\Theta_x+\Theta_y)\right)y_{i,j}^2 \tag{27}$$

where

$$h_{\alpha,\lambda}(t,l)=\lambda^2(t)^2(1-l)+\alpha l.$$

Note that $(1-2(\Theta_x+\Theta_y))>0$ because of the positivity requirement of the spectral density of the GMRF model. The problem is thus reduced to the following optimization problem:

$$\min_{\{y_{ij}\}}\left(\frac{1}{2\sigma^2}\left[D+\min_{\{l_{ij}\}}\left(\sum_{ij}\Theta_x h_{\alpha,\lambda}(y_{i,j}-y_{i-1,j},l_{ij})\right)+\min_{\{m_{ij}\}}\left(\sum_{ij}\Theta_y h_{\alpha,\lambda}(y_{i,j}-y_{i,j+1},m_{ij})\right)\right]\right).$$

As D does not involve l_{ij}, m_{ij}, minimization over l_{ij}, m_{ij} can be performed and one is then left with minimization with respect to y_{ij}:

$$\min_{\{y_{ij},l_{ij},m_{ij}\}}U=\min_{\{y_{ij}\}}\left\{\frac{1}{2\sigma^2}\left[D+\sum_{ij}\Theta_x g_{\alpha,\lambda}(y_{i,j}-y_{i-1,j})+\sum_{ij}\Theta_y g_{\alpha,\lambda}(y_{i,j}-y_{i,j+1})\right]\right\}$$

where

$$g_{\alpha,\lambda(t)}=\min_{l\in\{0,1\}}h_{\alpha,\lambda}(t,l)=\min\left(\lambda^2(t)^2,\alpha\right).$$

In the following, we assume that y is presented by a one-dimensional (1-D) array with a lexicographic order. Following Blake we look for a convex approximation U^*. The convexity of U^* is guaranteed by requiring that it has a positive definite Hessian matrix $H=\partial^2U^*/\partial y_{ij}\partial y_{k,l}$. Suppose g^* is designed to satisfy

$$\forall\, t\ \ g^{*\prime\prime}(t)\geq -c^*$$

where $c^*>0$. Then the "worst case" of H occurs when $\forall\, i,j\ \ g^{*\prime\prime}(y_{i,j}-y_{i,j+1})=-c^*$ and $g^{*\prime\prime}(y_{i,j}-y_{i-1,j})=-c^*$

so that

$$H = \left[2+2\lambda^2\left(1-2\left(\Theta_x+\Theta_y\right)\right)\right] I - c^* R.$$

The matrix R is a symmetric tri-diagonal block Toeplitz matrix:

$$R = \begin{bmatrix} Q & -\Theta_x I & & \\ -\Theta_x I & Q & -\Theta_x I & \\ & -\Theta_x I & Q & -\Theta_x I \\ & & -\Theta_x I & Q \end{bmatrix}$$

$$Q = \begin{bmatrix} 2\Theta_x+2\Theta_y & -\Theta_y & & \\ -\Theta_y & 2\Theta_x+2\Theta_y & -\Theta_y & \\ & -\Theta_y & 2\Theta_x+2\Theta_y & -\Theta_y \\ & & -\Theta_y & 2\Theta_x+2\Theta_y \end{bmatrix}.$$

To prove that H is positive definite it is sufficient to show that the largest eigenvalue $\mathscr{E}_{\max}$ of R satisfies $\mathscr{E}_{\max} \le 2 + 2\lambda^2(1-2(\Theta_x+\Theta_y))/c^*$.

The eigenvalues $\mathscr{E}_{i,j}$, $i, j \in (0 \cdots M-1)$ of R can be found using the sine transform [34].

$$\mathscr{E}_{i,j} = \Theta_x\left(2-2\cos\frac{2\pi i}{M}\right) + \Theta_y\left(2-2\cos\frac{2\pi j}{M}\right)$$

So that $\max \mathscr{E}_{i,j} = 4(\Theta_x+\Theta_y)$, thus to guarantee convexity c^* must satisfy

$$c^* \le \frac{1+\lambda^2\left(1-2\left(\Theta_x+\Theta_y\right)\right)}{2\left(\Theta_x+\Theta_y\right)}.$$

Following Blake we construct the best quadratic approximation g^* with a given bound $-c^*$ on its second derivative, satisfying the extra condition: for all t $g^*(t) \le g(t)$

$$g_{\alpha,\lambda}{}^*(t) = \begin{cases} \lambda^2(t)^2, & |t| < q \\ \alpha - c^*(|t|-r)^2/2, & q \le |t| < r \\ \alpha, & |t| \ge r \end{cases}$$

where

$$r^2 = \alpha\left(\frac{2}{c^*}+\frac{1}{\lambda^2}\right), \qquad q = \frac{\alpha}{\lambda^2 r}.$$

Thus we obtain the convex approximation for U:

$$U^* = \frac{1}{2\sigma^2}\left[D + \sum_{ij}\Theta_x g_{\alpha,\lambda}{}^*(y_{i,j}-y_{i-1,j}) + \sum_{ij}\Theta_y g_{\alpha,\lambda}{}^*(y_{i,j}-y_{i,j+1})\right]. \quad (28)$$

A one parameter family of cost functions $U^{(P)}$ is then defined by replacing g^* in (28) by $g^{(P)}$. $g^{(P)}$ is similar to g^* except that c^* is replaced by a variable c^{**}, that varies with P

$$U^{(P)} = \frac{1}{2\sigma^2}\left[D + \sum_{ij}\Theta_x g^{(P)}_{\alpha,\lambda}(y_{i,j}-y_{i-1,j}) + \sum_{ij}\Theta_y g^{(P)}_{\alpha,\lambda}(y_{i,j}-y_{i,j+1})\right] \quad (29)$$

with

$$g^{(P)}_{\alpha,\lambda}(t) = \begin{cases} \lambda^2(t)^2, & |t| < q \\ \alpha - c^{**}(|t|-r)^2/2, & q \le |t| < r \\ \alpha, & |t| \ge r \end{cases}$$

where

$$c^{**} = \frac{c^*}{P}, \qquad r^2 = \alpha\left(\frac{2}{c^{**}}+\frac{1}{\lambda^2}\right), \qquad \text{and } q = \frac{\alpha}{\lambda^2 r}.$$

The GNC–GMRF algorithm is derived similar to the GNC algorithm in Section III. In the GNC–GMRF algorithm we no longer choose the value of λ, but compute it from the equation $\lambda = \sqrt{\sigma^2/\nu}$. The GNC–GMRF algorithm begins by minimizing $U^{(P=1)} = U^*$. Then P is decreased from 1 to 0, which makes $g^{(P)}$ change steadily from g^* to g. For every value of P we minimize $U^{(P)}$ starting with the last configuration corresponding to the previous P (local minimum of $U^{(2P)}$). The minimization of $U^{(P)}$ is performed efficiently using the optimal step conjugate gradient algorithm.

C. Parameter Estimation

The compound GMRF model presented in Section IV-A-2 presents a new problem in parameter estimation. We would like to estimate the GMRF parameters only in homogeneous regions that do not include edges. The problem is that even if the location of the edges is known, removing the adjacent pixels from the rectangular domain we started with, leaves a highly irregular domain, which is cumbersome to work with. Furthermore in the case of ML estimation, the irregularity of the domain makes it impossible to write a close form expression for the likelihood function even if toroidal assumption are made. Instead we suggest a new algorithm based on the EM algorithm [28]. For the time being, we present the algorithm for estimating the parameters from the original image. We believe that the method can be extended to estimation techniques such as BCLS [32] that are based on the noisy image.

The idea behind the new algorithm is that we ignore the intensity data in strips centered at the edges and replace this data by the conditional mean of the GMRF model given the model parameters and the neighboring pixels intensity data. The conditional mean is calculated using a relaxation method, i.e., for the first order neighborhood we repeatedly set

$$y^{k+1}(i,j) = \Theta_x\left(y^k(i+1,j)+y^k(i-1,j)\right) + \Theta_y\left(y^k(i,j+1)+y^k(i,j-1)\right). \quad (30)$$

Note that we modify the intensity only in the strips centered at the edges. The iterations are bound to converge, because the stability requirement of the GMRF model parameters ensure that the eigenvalues of the iteration matrix are smaller than 1. Once we computed the condi-

TABLE III
PARAMETER ESTIMATION RESULTS FOR AIRPORT IMAGE USING THE EM ALGORITHM

Iteration	Θ_x	$\hat{\Theta}_x$	Θ_y	$\hat{\Theta}_y$	ν	$\hat{\nu}$	Image Variance
0	0.282700	0.266150	0.247330	0.232850	25.6184	51.97603	469.4704
1	0.300411	0.291597	0.213673	0.207403	6.55542	16.98208	359.0647
5	0.310806	0.304306	0.198853	0.194694	5.99739	12.96653	336.8897
14	0.312426	0.306504	0.196215	0.192496	5.98000	12.55508	350.1088

tional mean, we perform an estimation step based on LS or ML on the smoothed image. We use the estimated parameters to calculate a new conditional mean for the pixels in the edge vicinity, which in turn are used for a new parameter estimation step. In this paper we present results for the EM algorithm based on LS. We are currently working on the ML version. We use Blakes edge detector [9] to find the location of the edges in the image. In the future we intend to combine the image and parameters estimation. The results we obtain with the new algorithm differ significantly from the results obtained directly from the original image. For example the estimated model variance is much smaller, although the image variance is reduced only slightly. The results show the importance of suppressing the edges effect in the parameter estimation. The new algorithm is summarized as follows.

1) Get an initial estimate for the model parameters, using the LS technique on the original image.
2) Find the image edges location using the GNC algorithm [9].
3) By using the current estimate of the model parameters find the conditional mean of the intensity in a strip four pixels wide centered at the edge using the relaxation equation (30).
4) Calculate a new parameter estimate on the smoothed image calculated in Step 3).
5) Check if the parameters are stable. If not scale them by multiplying the vector Θ by

$$\psi = \frac{0.499}{\max_{s \in \Omega} \sum_{\tau_i \in N^*} \Theta_{\tau_i} \Phi_{s,\tau_i}}$$

where

$$\Phi_{s,\tau_i} = \cos\left(\frac{2\pi s^T \tau_i}{M}\right),$$

and

$$N^* = \{\tau_1, \tau_2\} = \{(0,1),(1,0)\}$$

are the shift vectors corresponding to the first order GMRF model, to ensure that the scaled parameters $\hat{\Theta}_{\tau_i} \tau_i \in N^*$ satisfy

$$\left(1 - 2 \sum_{\tau_i \in N^*} \hat{\Theta}_{\tau_i} \Phi_{s,\tau_i}\right) > 0 \; \forall \, s \in \Omega. \qquad (31)$$

6) If the change in the parameters is small enough stop, else go back to step 3).

Steps 4) and 5) of the algorithm can be replaced by constraint maximization of the likelihood function. We are currently working on the ML version of the algorithm. We experimented with the new algorithm on a real airport image and obtained the results summarized in Table III. Note that the parameter values do not change much after ten iterations.

D. Restoration Results

The original airport image corrupted by 5-dB and 10-dB additive white Gaussian noise was reconstructed using the modified GNC algorithm. As an initial condition for the algorithm we used the noisy image. Parameters were estimated from the original image using the new EM algorithm described in Section III. The results were obtained using at most 25 iterations for each value of P. For some values of P the algorithm converged in less than 20 iterations.

In the first experiment we restored an image corrupted by 10-dB additive white Gaussian noise. The edge threshold was chosen to be $h = 20$ and the noise variance was assumed to be known. The results after 100 iterations are presented in Fig. 5. We then repeated the experiment for the 5-dB additive white Gaussian noise case. In this experiment h was set to 25. The results are presented in Fig. 6. The results obtained for 10-dB noise case are very good. The estimates include all the information in the original image and in addition we obtain a precise edge map. We compare the results obtained assuming a GMRF model for the image intensity and show that the image estimates are substantially blurred. In the 5-dB case the results are good also, but a few noisy points were isolated by the line process. This is an inherent problem owing to the simple line process used in this model. It can partially be overcome by increasing λ. One can also modify the line process to include an energy term that will inhibit the formation of such edges as in [2], but the GNC algorithm can no longer be used for the more sophisticated line process.

V. CONCLUSION

In this paper, we used the stochastic formulation for low level vision problems, and demonstrated the equivalence between direct optimization of functionals and MAP estimates obtained from stochastic modeling. We used the Gibbs formulation to present the Markov field model, and the Gibbs energy functions to link the deterministic and stochastic formulations. Using the Gibbs formulation, we derived a pyramid implementation for the line-search con-

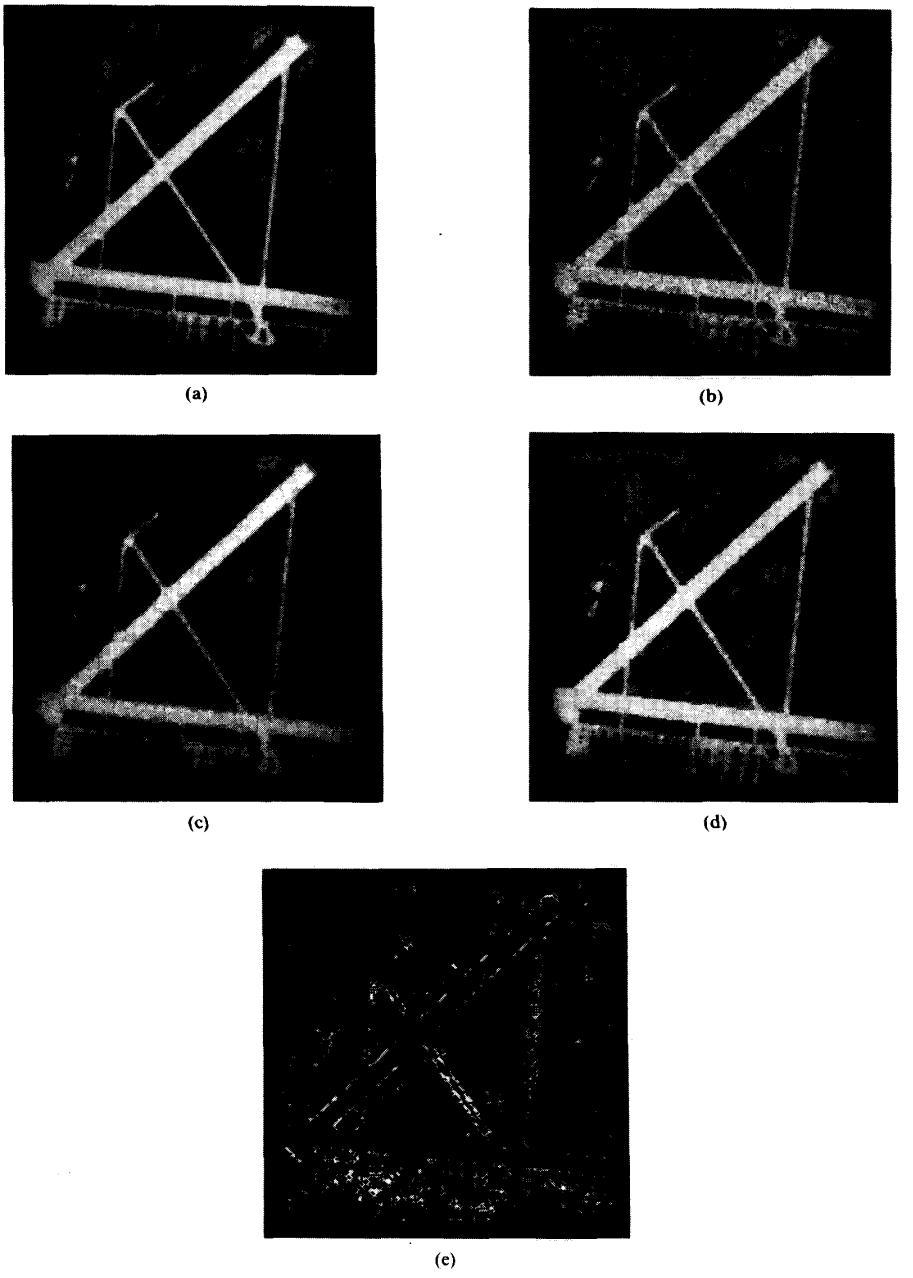


Fig. 5. Restoration and edge detection on airport image with 10-dB noise. (a) Original 128×128 airport image. (b) Image corrupted by 10-dB noise. (c) GMRF estimate after 100 iterations. (d) Compound GMRF estimate using GNC after 100 iterations. (e) Edge map obtained from GNC algorithm.

jugate gradient algorithm for early vision problems. The line-search conjugate algorithm was successfully used for nonlinear optimization problems, arising in SFS, image estimation and edge detection. We have also shown that the multigrid methods can successfully be incorporated in our algorithms to improve their convergence rate.

Coupled Markov fields models (such as compound GMRF) suggest a more realistic description of real life scenes. This observation is supported by the large reduction in the model variance, due to the ability to cope with the edges in the scene by using line processes. The new models require new parameter estimation techniques. An algorithm for obtaining the GMRF model parameters from the original image via an EM algorithm, was derived in this paper. The estimated parameters emphasize the importance of suppressing the influence of edges, when estimat-

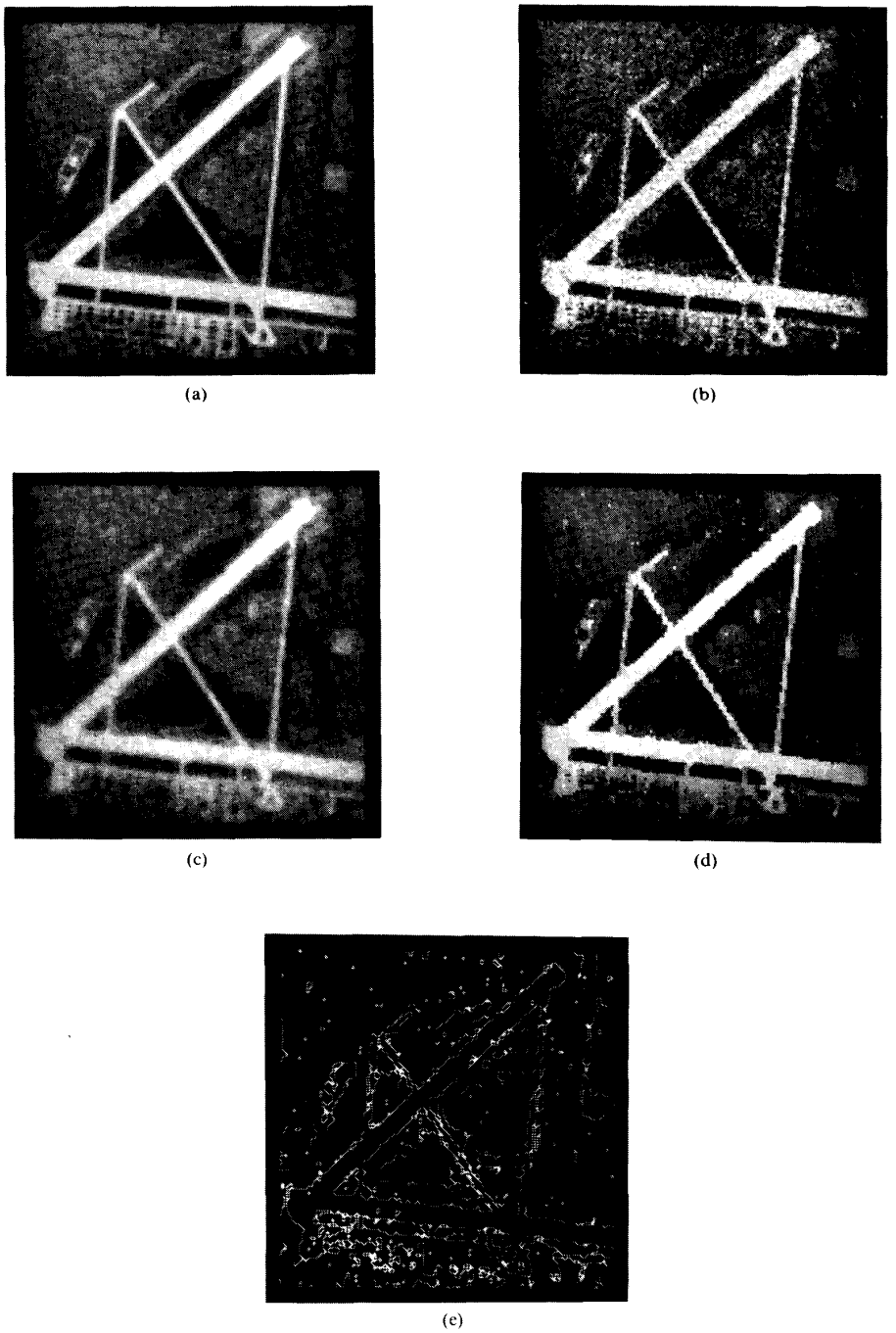


Fig. 6. Restoration and edge detection on airport image with 5-dB noise. (a) Original 128×128 airport image. (b) Image corrupted by 5-dB noise. (c) GMRF estimate after 100 iterations. (d) Compound GMRF estimate using GNC after 400 iterations. (e) Edge map obtained from GNC algorithm.

ing the parameters of the homogeneous regions. The work in the paper needs to be extended to a parameter estimation method from noisy images. We feel that it can be done using an EM algorithm

Existing stochastic relaxation techniques for optimizing the nonconvex cost functions are computationally unattractive. In some restricted cases one can use the GNC formulation to find the global minimum of these nonconvex cost functions. We have exploited the equivalence between the stochastic and deterministic formulations through the use of Gibbs energy function, to extend the GNC algorithm to image estimation in compound GMRF

for the additive noise case. The results are very encouraging, and the GNC–GMRF algorithm is fast and robust. Recently we have developed suboptimal relaxation methods in order to optimize nonconvex cost functions obtained from more sophisticated line processes than the one handled by the GNC formulation [35]. We have used a constrained minimization approach and let the line process variables evolve to the set $\{0,1\}$ gradually. Simulations results are encouraging.

REFERENCES

[1] B. K. P. Horn, *Robot Vision*. Cambridge, MA: The MIT Press, 1986.

[2] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, Nov. 1984.

[3] T. Poggio and V. Torre, "Ill-posed problems and regularization analysis in early vision," Artificial Intelligence Lab, M.I.T., A.I.M 773, Apr. 1984.

[4] J. Hadamard, *Lectures on the Cauchy Problem in Linear Partial Differential Equations*. New Haven, CT: Yale University Press, 1923.

[5] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, vol. 1. New York: Interscience, 1953.

[6] M. J. Brooks and B. K. P. Horn, "Shape and source from shading," in *Proc. Int. Joint Conf. Artificial Intell.*, Los Angeles, CA, Aug. 1985, pp. 932-936.

[7] M. Strat, "A numerical method for shape from shading from a single image," MS Thesis, M.I.T., Dept. of Elect. Engrg. and Comp. Sci., 1979.

[8] D. Lee, "A provably convergent algorithm for shape from shading," in *Proc. DARPA Image Understanding Workshop*, Miami Beach, FL, Dec. 1985, pp. 489-496.

[9] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.

[10] D. Luenberger, *Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley, 1984.

[11] D. Terzopoulos, "Image analysis using multigrid relaxation methods," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 129–139, Mar. 1986.

[12] Z. Li and L. Uhr, "Pyramid vision using key features to integrate image-driven bottom-up and model-driven top-down processes," *IEEE Trans. Syst. Man Cybern.* vol. SMC-17, pp. 250–263, Mar. 1987.

[13] P. J. Burt, T. H. Hong, and A. Rosenfeld, "Segmentation and estimation of image region properties through cooperative hierarchical computation," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-11, pp. 801–809, 1981.

[14] P. J. Burt and G. S. van der Wal, "Iconic image analysis with the pyramid vision machine," *Comput. Architecture Pattern Anal. Machine Intell.*, pp. 137–144, 1987.

[15] R. Miller, "Pyramid computer algorithms," Ph.D. Thesis, Dept. of Math., State Univ. of New York, Binghamton, 1984.

[16] A. Rosenfeld, *Multiresolution Image Processing and Analysis*. New York: Springer-Verlag, 1984.

[17] ____, "A report on the DARPA image understanding architectures workshop," in *Proc. of DARPA Image Understanding Workshop*, Los Angeles, CA, Feb. 1987, pp. 298-302.

[18] S. L. Tanimoto, "Paradigm for pyramid machine algorithms," *Parallel Structures for Comput. Vision*, M. T. B. Duff, Ed. New York: Academic, pp. 173–194, 1987.

[19] M. Shneier, "Using pyramid to define local threshols for blob detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-5, pp. 345–349, May 1983.

[20] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. B*, vol. 48, pp. 259–302, 1986.

[21] T. Simchony and R. Chellappa, "Direct analytical methods for solving Poisson equations in computer vision problems," in *IEEE Comput. Society Workshop Comput. Vision Problems*, Miami Beach, FL, Nov. 1987.

[22] S. Geman and C. Hwang, "Diffusions for global optimization," *SIAM J. Contr. Optimization*, vol. 24, pp. 1031–1043, Sept. 1986.

[23] L. E. Scales, *Introduction to Non-Linear Optimization*. New York: Spinger-Verlag, 1985.

[24] A. K. Jain and J. R. Jain, "Partial difference equations and finite differences in image processing, part II: Image restoration," *IEEE Trans. Automatic Contr.*, vol. AC-23, pp. 817–833, Oct. 1978.

[25] R. Chellappa and R. L. Kashyap, "Digital image restoration using spatial interaction models," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-30, pp. 461–472, June 1982.

[26] H. Jinchi, T. Simchony, and R. Chellappa, "Maximum *aposteriori* restoration of images corrupted by multiplicative noise," in *Proc. ICASSP 87, IEEE Conf. Acoust. Speech Signal Process.*, Dallas, TX, Apr. 1987.

[27] F. C. Jeng and J. W. Woods, "Image estimation by stochastic relaxation in the compound Gaussian case," in *Proc. ICASSP* 88, *IEEE Conf. Acoust. Speech Signal Process.*, New York, N.Y., Apr. 1988.

[28] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Roy. Statis. Soc. B*, no. 39, 1977.

[29] R. Chellappa and R. L. Kashyap, "Texture synthesis using two-dimensional noncausal autoregressive models," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-33, pp. 194–203, Feb. 1985.

[30] R. Chellappa, S. Chatterjee, and R. Bagdazian, "Texture synthesis and coding using Gaussian Markov random field models," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-15, pp. 298–303, Mar. 1985.

[31] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Statist. Soc. B*, vol. 36, pp. 192–236, 1974.

[32] H. Kaufman *et al.*, "Estimation and identification of two-dimensional images," *IEEE Trans. Automatic Contr.*, vol. AC-28, no. 7, pp. 745–755, July 1983.

[33] B. R. Hunt, "Digital image processing," *Proc. IEEE*, vol. 63, Apr. 1975, pp. 693–708.

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

[35] A. Rangarajan, R. Chellappa, and T. Simchony "Networks for image estimation using penalty mehtod," *Proc. Int. Joint Conf. Neural Networks*, Washington DC, June, 1989.

Tal Simchony (S'86–M'88) received the B.Sc. degree in mathematics and computer science and the M.Sc. degree in applied mathematics from Tel Aviv University in 1982 and 1985, respectively, and the Ph.D. degree in electrical engineering from University of Southern California, Los Angeles, CA in 1988.

In 1982–1985 he worked in the software industry, for ECI Telecom Ltd., Tel Aviv, Israel. He is currently a research associate in the Signal and Image Processing Institute, University of Southern California. His current research interests include image processing, computer vision and neural networks. He joined ECI Telecom in 1989.

Ramalingam Chellappa (S'78–M'79–SM'83) was born in Madras, India. He received the B.S. degree (honors) in electronics and communications engineering from the University of Madras in 1975, and the M.S. degree (with distinction) in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 1977, and the M.S. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, in 1978 and 1981, respectively.

During 1979–1981, he was a Faculty Research Assistant at the Computer Vision Laboratory, University of Maryland, College Park, MD. Since 1986 he has been an Associate Professor in the Electrical Engineering Systems, and as of September 1988, he became the Director of the Signal and Image Institute at the University of Southern California, Los Angeles. His current research interest are in signal and image processing, computer vision, and pattern recognition.

Dr. Chellappa is a member of the Tau Beta Pi and Eta Kappas Nu. He is a coeditor of two volumes of selected papers on image analysis and processing, published in Fall 1985, and an Associate Editor for IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING. He was a recipient of a National Scholarship from the Government of India during 1969–1975. He received the 1975 Jawaharlal Nehru Memorial Award from the Department of Education, Government of India, the 1985 Presidential Young Investigator Award, and the 1985 IBM Faculty Development Award. He was the General Chairman of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, June, 1989, IEEE Computer Society workshop on Artificial Intelligence for Computer Vision, and Program Co-Chairman of the NSF-sponsored workshop on Markov random fields.

Ze'ev Lichtenstein received the B.S.E.E. degree from Tel-Aviv University, Tel-Aviv, Israel in 1986.

In 1983–1987, he was with Scitex Corp., Israel, where he was involved in a design of a new scanner. Currently he is a research assistant in the Signal and Image Processing Institute at the University of Southern California, Los Angeles. His current research interests include signal processing, and optimization techniques for computer vision.