Efficient Multiresolution Counterparts to Variational Methods for Surface Reconstruction*

Paul W. Fieguth,† William C. Karl,‡ and Alan S. Willsky

Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, 35-437, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139
E-mail: pfieguth@uwaterloo.ca, wckarl@bu.edu, willsky@mit.edu

Variational methods have been employed with considerable success in computer vision, particularly for surface reconstruction problems. Formulations of this type require the solution of computationally complex Euler–Lagrange partial differential equations (PDEs) to obtain the desired reconstructions. Further, the calculation of reconstruction error covariances for such approaches are usually neglected.

In this paper we describe a computationally efficient multiscale approach to surface reconstruction which differs fundamentally from other multiresolution methods that are used to solve the Euler–Lagrange PDEs. Instead, we interpret the variational problem as a statistical estimation problem in order to define a nearby, but slightly *different*, multiscale estimation problem that admits efficient solutions for both surface reconstruction *and* the calculation of error statistics. In particular, the membrane and thin-plate variational models for surfaces are interpreted as $1/f^2$ prior statistical models for the surface and its gradients, respectively. Such $1/f^2$ behavior is then achieved using a recently introduced class of multiresolution models that admits algorithms with constant per-pixel computational complexity. © 1998 Academic Press

1. INTRODUCTION

The problem of surface reconstruction [1, 12, 15, 16, 36, 40] has been a topic of considerable interest in the field of computer vision for some time, involving the estimation of an unknown surface based on a set of noisy measurements of some function of the surface and its derivatives and based on a prior model for the surface (generally necessary to regularize the problem). Variational methods [5, 40] have enjoyed considerable success in dealing with surface reconstruction problems, both as an an-

alytical means of formulating the problem and as a means of determining a solution. Formulations of this type lead directly to Euler–Lagrange [5] partial differential equations (PDEs) to be solved in order to obtain the desired reconstructions. Except in those specific cases where the surface model and the measurement statistics are homogeneous, permitting FFT techniques to be applied, the solution of these equations can be a significant computational task. Moreover, the calculation of reconstruction error covariances [20, 31, 32, 37, 38] for such approaches are, for all practical purposes, completely infeasible, as their computation corresponds in essence to the calculation of the full inverse of the partial differential operator arising from the variational problem.

The research described in this paper is motivated by a class of problems for which FFT techniques are inapplicable and for which the solution of the Euler–Lagrange equations via variational approaches is impractical: we are interested in *large* estimation problems having spatially varying models, possibly sparse measurements or measurements of varying quality and resolution, and for which estimation error statistics are required. In this paper we describe an alternative multiscale approach to surface reconstruction that overcomes these computational difficulties. In our approach we take advantage of the dual interpretation of variational problems as statistical estimation problems; in particular, a variational problem with quadratic costs (i.e., a least squares problem) may be interpreted as a Gaussian statistical estimation problem. This point separates the approach of this paper from those of the multigrid [14, 27, 35, 39] and preconditioning [2, 3, 7, 8, 33, 43] literatures:

• Preconditioner approaches solve the linear system $Ax = b$ by effecting a transformation (typically of the form $x = Sy$) such that the condition number of $(S^T A S)$ in the transformed linear system is much lower than that of A, leading to a more rapid convergence for iterative algorithms. Similarly the multigrid algorithm transforms the problem into a coupled hierarchy of linear systems, which accelerates convergence. In both cases, a solution to $Ax = b$ is ultimately obtained.

• Our multiscale approach, on the other hand, is fundamentally different, because we end up solving a *different* problem $\bar{A}\bar{x} = b$. The choice of the prior model $\bar{A}$ is made such that it is

similar to A, but also such that $\bar{A}$ satisfies very particular properties that lead to an efficient, noniterative algorithm for both estimates *and* estimation error statistics.

The problem of finding an appropriate $\bar{A}$ (i.e., of finding an appropriate multiscale model) for those A typically used in variational surface reconstruction is the fundamental problem addressed by our paper. As has been noted by others [31, 32], the membrane and thin-plate variational models commonly used in surface reconstruction allow interpretations as $1/f^2$ prior statistical models, for which a recently introduced [4, 9, 23] class of multiresolution models is available. The work in this paper builds on these earlier multiscale efforts, in particular addressing two fundamental limitations encountered in [9] and [23]: an inability to estimate gradients and compute gradient error statistics and an inability to obtain smooth estimates. This paper develops a new multiresolution surface reconstruction model that is the counterpart of standard variational models, but which admits algorithms with constant per-pixel computational complexity for both surface reconstruction *and* the calculation of error statistics. Moreover, the flexibility of this multiresolution framework allows us to define a richer class of surface reconstruction models and algorithms corresponding to different prior models which have either more complicated or *no* variational counterparts, but that admit the same efficient solutions.

This paper is organized as follows. Section 2 reviews variational model development for surface reconstruction; Section 3 reviews the multiscale framework into which the surface reconstruction problem is to be cast. In Sections 4 and 5 we detail the construction of multiscale models given the variational counterpart. Finally in Section 6 several experimental results are presented.

2. BACKGROUND

2.1. Notation

The general surface reconstruction problem [36] involves estimating the shape of a surface given a discrete (and possibly sparse) set of noisy observations of some function of the surface and/or its gradients. The surface of interest is a two-dimensional function $z(x, y)$ presumed twice differentiable everywhere; denote by

$$p(x, y) = z_x(x, y) = \frac{\partial z(x, y)}{\partial x} \quad q(x, y) = z_y(x, y) = \frac{\partial z(x, y)}{\partial y}$$
(1)

the gradients of the surface at each point. Normally we shall refer simply to z, p, q with an implicit dependence on x and y. We are interested in least-squares problems of the type

$$\hat{z}(D) = \min_z \{(D - \mathcal{C}(z))^T \mathcal{R}^{-1}(D - \mathcal{C}(z)) + \mathcal{Z}(z)\}$$
(2)

$$D = \mathcal{C}(z) + V,$$
(3)

where D is a discrete vector of observations, corrupted by white noise V, and where V has covariance $\mathcal{R}$. $\mathcal{C}$ is the measurement function, and $\mathcal{Z}$ represents a prior model for the surface z. The specific nature of the measurement function $\mathcal{C}$ will typically depend on the specific surface reconstruction problem of interest; for example we may directly observe a sparse subset of the surface heights,

$$\mathcal{C}(z) = [z(x_1, y_1) \quad z(x_2, y_2) \quad \cdots]^T,$$
(4)

or we may just observe a subset of the surface gradients,

$$\mathcal{C}(z) = [p(x_1, y_1) \quad q(x_1, y_1) \quad \cdots]^T.$$
(5)

2.2. Surface Prior Models

A least-squares solution for the surface z given the measurements alone is typically ill-posed [1]; i.e., there is not a single optimum solution for the surface. In order to guarantee a unique solution the problem is regularized by asserting a prior model $\mathcal{Z}$ for the unknown surface, typically reflecting our prior knowledge about the surface to be reconstructed or, equivalently, asserting certain desired smoothness properties for the reconstructed surface.

A simple and common surface prior model is to assert a smoothness constraint [16]:

$$\mathcal{Z}_p(z) = \alpha \iint \{p_x^2 + p_y^2 + q_x^2 + q_y^2\} \, dx \, dy.$$
(6)

This constraint is also referred to as a "thin plate" term, in that (6) represents the potential energy in an isotropic thin plate [5]. An alternative function, representing the potential energy contained in a stretched membrane [5], punishes variations from $p = 0, q = 0$:

$$\mathcal{Z}_m(z) = \beta \iint \{p^2 + q^2\} \, dx \, dy.$$
(7)

Combining (6), (7) yields a variational formulation familiar to computer vision researchers [16]:

$$\hat{z}(D) = \min_z \left\{ (D - \mathcal{C}(z))^T \mathcal{R}^{-1}(D - \mathcal{C}(z)) \right.$$
$$\left. + \iint \alpha\{p_x^2 + p_y^2 + q_x^2 + q_y^2\} + \beta\{p^2 + q^2\} \, dx \, dy \right\}.$$
(8)

There are two common interpretations of this formulation:

- The regularization term of (8) is a special case of the class of two-dimensional generalized spline functionals [36, 38].
- The regularization term of (8) represents a deformable sheet [40] or a stiff surface, being acted upon by forces (i.e., the observations), where the resulting deformation is a function of the

specific stiffness properties of the sheet. In particular, a membrane term $\mathcal{Z}_m$ minimizes the surface "area" (like a rubber sheet), whereas the thin plate term $\mathcal{Z}_p$ minimizes the surface curvature (like a steel plate).

2.3. Explicit Estimation of Surface Gradients

One final addition to the variational formulation is appropriate. In many surface reconstruction applications, the gradients of the surface play a central role (the most notable example being the shape-from-shading problem [16, 17]). It is frequently of interest to estimate the gradients explicitly, rather than to infer them implicitly as a function of the estimated surface.

For example, consider a problem in which the measurements are functions of the surface gradients only, not the surface heights themselves. In such problems, we may be motivated to use the common variational equation [16]

$$\min_{p,q} \left\{ (D_g - \mathcal{C}_g(p,q))^T \mathcal{R}_g^{-1}(D_g - \mathcal{C}_g(p,q)) \right.$$
$$\left. + \iint_{R^2} \alpha \left(p_x^2 + p_y^2 + q_x^2 + q_y^2 \right) + \beta (p^2 + q^2) \, dx \, dy \right\},$$
(9)

where D_g represents gradient-dependent observations and where the measurement function $\mathcal{C}_g(p,q)$ explicitly depends upon gradient terms only. In general, the estimates of p, q resulting from such a variational equation will not correspond to the gradients of *any* surface z—this is the well-known integrability problem [12, 21]. Our multiscale surface reconstruction model (to be outlined in Section 4) will similarly be estimating the surface gradients explicitly, hence the relevance of the following discussion.

In order of p, q to be gradients of a surface, the consistency constraint

$$\oint_{\mathcal{L}} (p \, dx + q \, dy) = 0$$
(10)

must hold over all closed paths $\mathcal{L}$ in the plane [15]. In other words, $p_y = q_x$ must hold at all points in the plane, leading to the revised variational problem [15]

$$\min_{p,q} \left\{ (D_g - \mathcal{C}_g(p,q))^T \mathcal{R}_g^{-1}(D_g - \mathcal{C}_g(p,q)) \right.$$
$$+ \iint_{R^2} \alpha \left(p_x^2 + p_y^2 + q_x^2 + q_y^2 \right) + \beta (p^2 + q^2)$$
$$\left. + \gamma (p_y - q_x)^2 \, dx \, dy \right\}.$$
(11)

The inclusion of the additional penalty term in the above expression does not guarantee that (10) is exactly satisfied, although using a large value for γ will in general result in nearly consistent (p, q) fields. Even if this consistency relationship is exactly satisfied, however, we still have the nontrivial problem

of computing z from the (p, q) fields. On the other hand, this problem *and* the consistency problem may be avoided by explicitly reconstructing z as well as (p, q) through a simple surface-gradient consistency penalty

$$\iint_{R^2} (z_x - p)^2 + (z_y - q)^2 \, dx \, dy.$$
(12)

This leads to the following variational problem where, for generality, we allow both direct measurements $D_s = \mathcal{C}_s(z) + V_s$, as well as gradient measurements D_g:

$$\min_{z,p,q} \left\{ (D_s - \mathcal{C}_s(z))^T \mathcal{R}_s^{-1}(D_s - \mathcal{C}_s(z)) \right.$$
$$+ (D_g - \mathcal{C}_g(p,q))^T \mathcal{R}_g^{-1}(D_g - \mathcal{C}_g(p,q))$$
$$+ \iint_{R^2} \alpha \left(p_x^2 + p_y^2 + q_x^2 + q_y^2 \right) + \beta (p^2 + q^2)$$
$$\left. + \gamma ((z_x - p)^2 + (z_y - q)^2) \, dx \, dy \right\}.$$
(13)

This variational problem, similar to the one introduced in [15], forms the point of departure for our analysis and for the development of efficient multiscale counterparts to problems of this type. We will restrict our attention to *linear* measurements of the surface and its gradients, as in [35, 36].

2.4. Euler–Lagrange Equations

After a specific variational expression has been selected, the solution for the optimal estimated surface $\hat{z}(D)$ can be characterized in a straightforward manner. For example, the Euler–Lagrange equation [5, 36] corresponding to (8) is

$$\frac{\partial}{\partial z}(D - \mathcal{C}(z))\mathcal{R}^{-1}(D - \mathcal{C}(z))^T$$
$$- \alpha \left\{ \frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} \right\} + \beta \left\{ \frac{\partial^4 z}{\partial x^4} + 2\frac{\partial^4 z}{\partial x^2 y^2} + \frac{\partial^4 z}{\partial y^4} \right\} = 0. \quad (14)$$

This is an elliptic PDE which, after specifying appropriate boundary conditions, may be solved numerically by discretizing the PDE and applying numerical solution techniques. In special cases extremely efficient FFT techniques may be applied. Typically, however, FFT techniques are *not* applicable, leading researchers to propose a variety of other techniques [16, 33, 35, 40]. These techniques successfully estimate the surface by solving the PDE in a comparatively efficient manner; however, they are unable to produce estimation error statistics for reasons that are outlined next.

2.5. Variational Problems as Estimation Problems

In this section we briefly describe the dual interpretation of variational problems as estimation problems, a topic that has

been discussed in more detail in previous works [9, 23, 31]. Specifically, consider the variation of (13),

$$\min_{z, p, q} \left\{ (D - \mathcal{C}(z, p, q))^T \mathcal{R}^{-1} (D - \mathcal{C}(z, p, q)) \right.$$
$$+ \int \int_{R^2} [\mathcal{L}_1(z, p, q)^T \mathcal{L}_1(z, p, q)$$
$$\left. + \mathcal{L}_2(z, p, q)^T \mathcal{L}_2(z, p, q)] \, dx \, dy \right\}, \tag{15}$$

where $\mathcal{L}_1$, $\mathcal{L}_2$ are column vectors of linear functionals of z, p, q, and where D may contain both height and gradient measurements (e.g., the D_s, D_g of (13)) from the linear measurement functional $\mathcal{C}$. Although such variational expressions are elegantly represented in continuous space, the goal of implementing a practical estimator on a computer motivates the shift to discrete space. For reasons of consistency, similar notation is used for both continuous and discrete space expressions.

Let $Z^T = [\ldots, z(x_i, y_j), \ldots]^T$ represent a vector of samples of $z(x, y)$ on a 2-D grid; similarly define P, Q. Then (15) may be discretized as

$$\min_{Z, P, Q} \left\{ \left(D - C \begin{bmatrix} Z \\ P \\ Q \end{bmatrix} \right)^T \mathcal{R}^{-1} \left(D - C \begin{bmatrix} Z \\ P \\ Q \end{bmatrix} \right) \right.$$
$$+ \left\{ L_1 \begin{bmatrix} Z \\ P \\ Q \end{bmatrix} \right\}^T \left\{ L_1 \begin{bmatrix} Z \\ P \\ Q \end{bmatrix} \right\}$$
$$\left. + \left\{ L_2 \begin{bmatrix} Z \\ P \\ Q \end{bmatrix} \right\}^T \left\{ L_2 \begin{bmatrix} Z \\ P \\ Q \end{bmatrix} \right\} \right\}, \tag{16}$$

where L_1, L_2 are matrices representing discrete approximations of the linear functionals $\mathcal{L}_1$, $\mathcal{L}_2$ over the discrete grid and where C is a matrix which describes the measurements of Z, P, Q.

Each discrete optimization problem of the form of (16) possesses an estimation counterpart. Specifically, the optimization of (16) corresponds exactly to the problem of estimating Z, P, Q given the measurement model

$$D = \mathcal{C} \begin{bmatrix} Z \\ P \\ Q \end{bmatrix} + v \quad v \sim \mathcal{N}(0, R) \tag{17}$$

and a prior model

$$\begin{bmatrix} \mathcal{L}_1 \\ \mathcal{L}_2 \end{bmatrix} \begin{bmatrix} Z \\ P \\ Q \end{bmatrix} = w \quad w \sim \mathcal{N}(0, I), \tag{18}$$

where $\mathcal{N}(\mu, R)$ represents a Gaussian random vector with mean μ and variance R. Alternatively, we may interpret portions of

the "prior" model as "measurements" (we shall find this to be convenient later); for example, the above estimation problem is the same as the following problem having measurement model

$$\begin{bmatrix} D \\ 0 \end{bmatrix} = \begin{bmatrix} \mathcal{C} \\ \mathcal{L}_1 \end{bmatrix} \begin{bmatrix} Z \\ P \\ Q \end{bmatrix} + \bar{v} \quad \bar{v} \sim \mathcal{N} \left(0, \begin{bmatrix} R & 0 \\ 0 & I \end{bmatrix} \right) \tag{19}$$

and a corresponding prior model

$$\mathcal{L}_2 \begin{bmatrix} Z \\ P \\ Q \end{bmatrix} = \bar{w} \quad \bar{w} \sim \mathcal{N}(0, I). \tag{20}$$

The solution to both of these estimation problems is given by the Euler–Lagrange equations

$$\left(\begin{bmatrix} \mathcal{L}_1 \\ \mathcal{L}_2 \end{bmatrix}^T \begin{bmatrix} \mathcal{L}_1 \\ \mathcal{L}_2 \end{bmatrix} + \mathcal{C}^T \mathcal{R}^{-1} \mathcal{C} \right) \begin{bmatrix} \hat{Z} \\ \hat{P} \\ \hat{Q} \end{bmatrix} = \mathcal{C}^T \mathcal{R}^{-1} D \tag{21}$$

and for which the estimation error covariance is

$$\tilde{P} = \left(\begin{bmatrix} \mathcal{L}_1 \\ \mathcal{L}_2 \end{bmatrix}^T \begin{bmatrix} \mathcal{L}_1 \\ \mathcal{L}_2 \end{bmatrix} + \mathcal{C}^T \mathcal{R}^{-1} \mathcal{C} \right)^{-1}. \tag{22}$$

The computation of (21) corresponds to the solution of a PDE, a computationally difficult task. However it is (22) that is orders of magnitude more complex, as it corresponds to the *complete inversion* of a higher-order PDE operator. What we propose to do is to replace the prior (20) by something similar, such that computing (21) and any element of (22), in particular the diagonal elements (the estimation error variances), is easy.

3. MULTISCALE MODELING

The multiscale models of interest in this paper and originally introduced in [4, 23, 24] are scale recursive models defined on index sets that are organized as multilevel trees. A simple example of such a tree (a quadtree) for a 2-D random field is illustrated in Fig. 1. Here each level of the tree corresponds to a different scale of resolution in the representation of the random field, with coarser scales toward the top of the tree. This modeling framework is more flexible than the figure might suggest, however, because it is applicable to higher dimensional, asymmetric, and unusually shaped trees; for the purposes of this paper the quadtree structure of Fig. 1 will suffice.

Let s denotes any node on the tree and $s\bar{\gamma}$ its parent; then the state vector at each node is governed by the following coarse-to-fine recursion

$$x(s) = A(s)x(s\bar{\gamma}) + B(s)w(s), \tag{23}$$
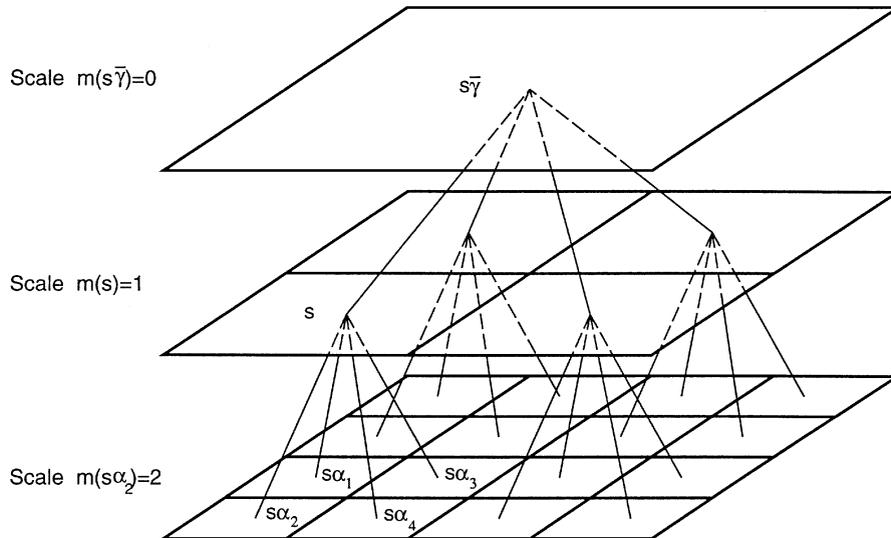
FIG. 1. A basic quadtree multiscale structure.

where $w(s)$ is a white Gaussian noise process with identity covariance. Moreover, the general measurement model associated with this framework allows measurement data at all scales,

$$d(s) = C(s)x(s) + v(s), \qquad (24)$$

where $v(s)$ is white, with covariance $R(s)$. The prior covariance $P_x(s)$ at each node s is determined by the recursive relationship in (23) and by the prior covariance at the root node,

$$P_x(0) = E[(x(0) - E[x(0)])(x(0) - E[x(0)])^T]. \qquad (25)$$

Superficially, this estimation problem would appear more difficult than the surface estimation problem of (17), (18), since we are now estimating quantities of interest on *all* resolutions. However for those multiscale estimation problems which can be written in the form of (23), (24) an extremely efficient realization of the optimal estimator is available [4, 23].

There are really two algorithms involved in proceeding from a variational description of an estimation problem, such as in (17), (18), to a set of multiscale estimates and error statistics:

1. The algorithm which takes us from the variational problem to a multiscale model of the form (23), (24). It is this algorithm which is novel to this paper and which will be the focus of the remaining sections.

2. The algorithm which takes us from a multiscale model plus data and computes estimates and error statistics.

The latter algorithm has been described in detail in [4, 23, 24]; however, we summarize the key points here.

Observe that *any* linear reconstruction, interpolation, or regularization problem has statistical interpretation in terms of specifying a prior statistical model and noisy measurements, which can be reduced in principle to a set of linear equations $Ax = b$ (e.g., as in (21)), where the solution yields the estimates and where the diagonal elements of A^{-1} are the error variances in the estimates of each component. Thus there are two related, but usually separate, issues: how do we solve $Ax = b$, and how do we get the diagonal elements of A^{-1}? The methods that have been used to compute error statistics, namely using Monte Carlo methods [31], computing A^{-1} iteratively row by row [31], or computing A^{-1} directly, require a great deal of computational effort *beyond* the effort needed to compute x. In particular, the computational load to exactly compute *all* of the diagonal elements of A^{-1} *dwarfs* the load required to solve $Ax = b$.

Our multiscale approach is fundamentally different. In particular, by choosing a prior statistical model with particular structure (i.e., by choosing a particular structure $\bar{A} \sim A$, consistent with (23), (24)), not only is the solution of the resulting $\bar{A}\bar{x} = b$ equation made much simpler and noniterative, but also in the process of solving this equation we directly compute the diagonal elements of $\bar{A}^{-1}$. As a comparison, consider the Kalman filter which computes time-recursive estimates of a dynamic process given a sequence of measurements. A key property of the Kalman filter is that the estimation error statistics are not computed separately from the estimates, rather the error statistics appear as a by-product of computing the estimates; in a manner of speaking, the error statistics are available "for free." Our multiscale approach is similar to that of the Kalman filter: we do not have a separate algorithm that computes estimation error statistics, rather these statistics are available "for free" as a by-product of our hierarchical Kalman filter-like algorithm. Just as a time-recursive stochastic model gives rise to the time-recursive Kalman filter, a scale-recursive model (23) on a tree gives rise to an efficient scale-recursive algorithm.

The challenge, then, is *not* one of calculating error statistics, rather one of finding a multiscale model which solves the

problem of interest. This challenge is addressed in the following sections.

4. ELEMENTARY MULTISCALE MODEL DEVELOPMENT

In this section we describe the basic elements required to map a variational problem such as (13), or equivalently its corresponding estimation problem (17), (18), into a closely related multiscale estimation problem of the form of (23), (24).

Recall that the estimation problem of (17), (18) consists of two components:

1. A measurement equation.
2. A prior model, parts of which may be interpreted as measurements (see Section 2.5).

The following two subsections discuss each of these two components in turn, followed by the development of a multiscale model appropriate for surface estimation.

4.1. Measurement Model

In the context of our multiscale tree, it is at the finest scale where we have measurements and wish to perform surface reconstruction. Thus the measurements (24) in our multiscale model are defined only at the finest scale. For example, in Section 6 we focus primarily on direct surface measurements. In such cases, if $z(s)$ is the first component of $x(s)$, then our measurement will take the form of (24) with $C(s) = [C_1(s), 0, \ldots]$ where $C_1(s) = 1$ at those finest scale nodes at which we have measurements.

4.2. Quadratic Penalties on State Derivatives, e.g., z_x^2

In one dimension the following variational cost

$$\int z_x^2 \, dx \qquad (26)$$

is equivalent to a Brownian motion prior model [23], i.e., a process with a $1/f^2$ spectrum (intuitive arguments why (26) may be considered a "fractal" prior are discussed in [4, 23, 32]). Such a process has the property that the variance of its increments (or more generally, its wavelet coefficients [41]) varies linearly with interval length (or, on a tree, varies geometrically with scale). As shown in [4, 23], we can capture such a geometric variation using a multiscale model such as

$$z(s) = 1 \cdot z(s\bar{\gamma}) + B_o 2^{-m(s)/2} w(s), \qquad (27)$$

where $m(s)$ denotes the scale of node s. The $1/f^2$-like nature of such a process is derived in [41] and illustrated empirically in [9]. As discussed in [23], a 2-D interpretation of the above also holds for variational penalties on derivatives of $z(x, y)$, so (6), (7) can be represented by quadtree models of (27).

4.3. Quadratic Penalties on Linear Combinations of State Variables

Other components of the prior involve quadratic costs of linear functions of variables, e.g.,

$$\iint ((z_x - p)^2 + (z_y - q)^2) \, dx \, dy \qquad (28)$$

from (12), which asserts the relationship between the surface and its gradients. As discussed in Section 2.5, we can (and do) choose to interpret such terms as part of the measurement model. For example, the penalty $\int \int (z_x - p)^2$ has the statistical interpretation as a measurement of the form

$$0 = z_x(x, y) - p(x, y) + v(x, y), \qquad (29)$$

where $v(x, y)$ is a spatially white "measurement noise" with a variance inversely proportional to the weight placed on the $(z_x - p)$ penalty in the variational problem.

"Measurements" of the form (29) are not in the most convenient form for multiscale implementation. However observe that the integrand in (28) can be expressed as

$$(z_x - p)^2 + (z_y - q)^2$$
$$= \frac{1}{2}(z_x + z_y - p - q)^2 + \frac{1}{2}(z_x - z_y - p + q)^2 \qquad (30)$$

which leads to the following types of measurements:

$$0 = z_x(x, y) + z_y(x, y) - p(x, y) - q(x, y) + v_1(x, y) \qquad (31)$$
$$0 = z_x(x, y) - z_y(x, y) - p(x, y) + q(x, y) + v_2(x, y). \qquad (32)$$

As we describe next, these measurements are readily captured in our quadtree structure.

In order to incorporate a discretized version of (31) or (32) into our multiscale framework we must define appropriate approximations to the derivatives $z_x(x, y)$ and $z_y(x, y)$. Since the concept of the gradient of a $1/f^2$-like surface is ill defined at best, we have some flexibility in how we choose to do this, and we have taken advantage of this flexibility to specify an approximation that leads to a very simple model. Consider Fig. 2, in which we have portrayed a parent pixel $s\bar{\gamma}$ and its four descendants s_1, s_2, s_3, s_4. We consider each node on the tree to represent a particular point on the surface, chosen to be the point in the center of the region aggregated by each node: the points are marked as a filled circle ● for each of the descendent nodes and as a cross × for node $s\bar{\gamma}$. Measured in units of finest scale pixels, the point ● in each of the child nodes is separated from × by a distance $\sqrt{2} \cdot 2^{M-m(s_1)-2}$, where M equals the number of scales on the tree. The quantities $(z_x + z_y)$ and $(z_x - z_y)$ represent directional derivatives of the field z along the two diagonal directions, i.e., along the directions from × to each ●. This suggests a natural approximation to either $z_x + z_y$ or $z_x - z_y$ in each

quadrant, depending on the direction of the line from $\times$ to each corresponding $\bullet$. For example, for nodes s_3 and s_4 we have

$$z_x(s_3) + z_y(s_3) \simeq 2^{2+m(s_3)-M}(-z(s\bar\gamma) + z(s_3)) \quad (33)$$

$$z_x(s_4) - z_y(s_4) \simeq 2^{2+m(s_4)-M}(-z(s\bar\gamma) + z(s_4)) \quad (34)$$

with analogous definitions at nodes s_1, s_2.

Consequently, if we wish to view (31) as a "measurement" we are led to discrete measurements of the form

$$0 = 2^{2+m(s_3)-M}(-z(s\bar\gamma) + z(s_3)) - p(s_3) - q(s_3) + v(s_3) \quad (35)$$

$$0 = 2^{2+m(s_2)-M}(+z(s\bar\gamma) - z(s_2)) - p(s_2) - q(s_2) + v(s_2). \quad (36)$$

In a similar fashion we can define measurements corresponding to (32):

$$0 = 2^{2+m(s_1)-M}(+z(s\bar\gamma) - z(s_1)) - p(s_1) + q(s_1) + v(s_1) \quad (37)$$

$$0 = 2^{2+m(s_4)-M}(-z(s\bar\gamma) + z(s_4)) - p(s_4) + q(s_4) + v(s_4). \quad (38)$$

For these "measurements" to be in the form of (24) all of the variables in (35)–(38) (other than the noise terms $v(s_i)$) must be in the respective state vectors $x(s_i)$. Thus each state $x(s_i)$ must include the parent value $z(s\bar\gamma)$; this is accomplished easily through state augmentation as described in the next subsection.

4.4. Elementary Multiscale Model Synthesis

Combining the preceding subsections leads to an elementary multiscale model corresponding to (13). First, the multiscale state dynamics:

$$\begin{bmatrix} z \\ p \\ q \\ zp \end{bmatrix}(s) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} z \\ p \\ q \\ zp \end{bmatrix}(s\bar\gamma)$$

$$+ \begin{bmatrix} B_s 2^{-m(s)/2} & 0 & 0 \\ 0 & B_g 2^{-m(s)/2} & 0 \\ 0 & 0 & B_g 2^{-m(s)/2} \\ 0 & 0 & 0 \end{bmatrix} w(s). \quad (39)$$

Here the first component of the dynamics captures the thin membrane term (7) and the next two the thin plate penalty (6). The last component accomplishes the state augmentation required for the penalty (12) as discussed in the preceding section: $zp(s)$ is simply $z(s\bar\gamma)$. The measurement equation accompanying this model includes both the actual measurements (at the finest scale) as well as measurements such as (35)–(38) which assert the surface-gradient consistency. For example, if we have only direct surface height measurements, the measurement equation at each of the four nodes s_i, $1 \le i \le 4$ in Fig. 2 would take the form
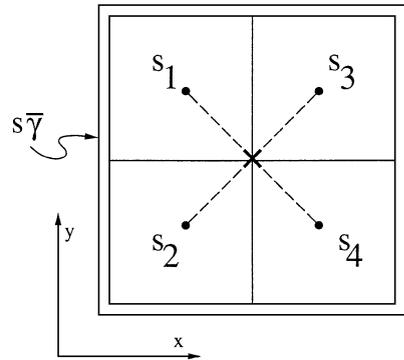
FIG. 2. A set of five node labels for gradient discussion purposes. s_1, s_2, s_3, s_4 are each children of coarser note $s\bar\gamma$. The $\times$ indicates the point represented by node $s\bar\gamma$; the solid circles $\bullet$ specify the points represented by nodes $s_1, \ldots, s_4$.

$$\begin{bmatrix} d(s_i) \\ 0 \end{bmatrix} = \begin{bmatrix} C_1(s_i) & 0 & 0 & 0 \\ -a_i 2^{2+m(s_i)-M} & -1 & b_i & a_i 2^{2+m(s_i)-M} \end{bmatrix}$$

$$\times \begin{bmatrix} z \\ p \\ q \\ zp \end{bmatrix}(s_i) + v(s_i), \quad (40)$$

where $C_1(s_i)$ (from Section 4.1) equals one at those nodes at which we have measurements (zero otherwise), and a_i, b_i are given by

$$\begin{array}{cccc} a_1 = 1 & a_2 = 1 & a_3 = -1 & a_4 = -1, \\ b_1 = 1 & b_2 = -1 & b_3 = -1 & b_4 = 1 \end{array} \quad (41)$$

where the a_i and b_i, which follow from (35)–(38), correspond to the measurement of either $z_x + z_y$ (for s_2 and s_3) or $z_x - z_y$ (for s_1 and s_4).

There are several final points that we should make about this multiscale model. First, there is the issue of specifying the unknown parameters, e.g., the B_s, B_g of (39). This is the same type of problem as the the selection of appropriate weights α, β, γ in the original variational formulation (13). However, now that we have a statistical interpretation of these terms, we can use that to our advantage in determining these quantities. In particular B_s, B_g have an explicit physical meaning, measured in real physical units, that represent the prior statistical knowledge of the surface; e.g., if the surface varies over a range of ± 15 cm, then $B_s = 10$ cm would make a reasonable choice. In addition, based on the multiscale likelihood and parameter identification methods that our models admit [11, 25], one can estimate the optimal values of these parameters from the data and perform model validation.

5. ADVANCED MULTISCALE MODEL DEVELOPMENT

The straightforward application of the elementary multiscale model (39), (40) derived in Section 4 would be to implement it
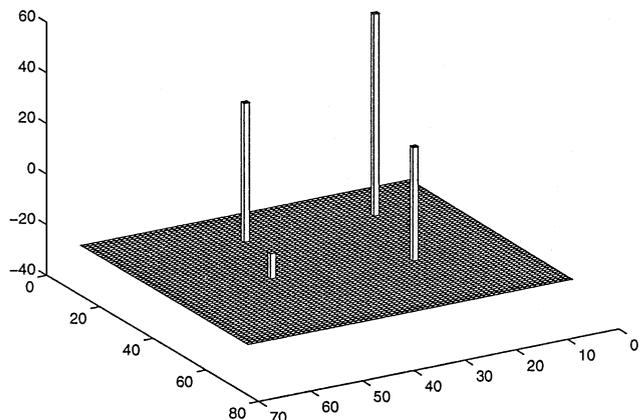
FIG. 3. Four sparse measurements used in testing the elementary multiscale model.

FIG. 5. Two nodes, s_1 and s_2, neighbors in physical space, but distantly separated in tree space.

directly on a quad-tree. Based on the four sparse measurements from Fig. 3, Fig. 4 demonstrates the result of such an approach. The computation of the estimates *and* estimation error variances (not shown) required about 3 s of SPARC-10 computer time. The most striking feature of these estimates is their blockiness, or lack of smoothness, characteristic of many quadtree algorithms; this section will address this blockiness issue.

From Fig. 1 we see that the upper-left and upper-right quadrants of the surface are correlated only via the state elements at the root node of the tree, despite the fact that these two quadrants share a relatively long common boundary. As a consequence, pixels such as s_1 and s_2 of Fig. 5 may be only weakly correlated because of their significant separation on the tree, despite their adjacency in physical space. As argued in [23], in some situations this blockiness is not a serious issue. In particular, in some problems the quality and quantity of information available for reconstruction may be sufficiently low that *no* statistically significant fine scale estimates can be computed. Nevertheless,

in many applications there are compelling reasons for producing smooth reconstructions, something that iterative variational solutions generally do well.

In this paper, we make use of a recently developed class of models to produce smooth estimates, based on the concept of "overlapping" trees. We refer the reader to two references [10, 19] which motivate and develop this method in greater detail, and limit ourselves here to a terse (but complete) description. As Fig. 1 suggests, in straightforward quadtree models the notes at a given level on the tree correspond to disjoint regions of the image plane, leading to the loss of correlation between pixels such as s_1 and s_2 in Fig. 5. The idea behind overlapping trees is to interpret nodes at a given level of the tree as corresponding to overlapping regions, as illustrated for a 1-D signal in Fig. 6, in which the bracket $\vdash\dashv$ at each node represents the region of physical space to which the state vector at that node corresponds. Although our overlapping regions are similar in spirit to the overlapping basis functions of multilevel preconditioners [33, 42, 43], this is where the similarity ends. It is the

FIG. 4. The surface estimates resulting from a straightforward implementation of the elementary multiscale model.
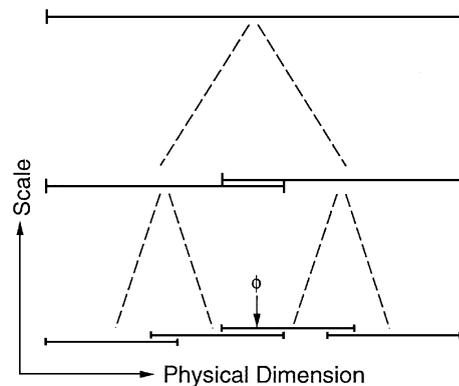
FIG. 6. Overlapping multiscale trees: adjacent nodes on the tree may represent overlapping regions of physical space.

desire to represent the process on an overlapped multiscale tree, but still consistent with (23), (24), that places very particular constraints on the overlapping framework, and the functions we end up using are unusual and bear no resemblance to the hierarchical functions of Szeliski [33] or Yaou and Chang [42].

Observe that the physical point ϕ, indicated in Fig. 6, is present in two of the intervals on the bottom level of the figure: clearly we have a redundant representation at the finest scale of the tree; i.e., several finest scale nodes may correspond to the same point in physical space, which raises the question of how tree states are mapped onto the image plane and how measurements in the image plane are mapped onto the tree. To see how this is done, let x_{dense} denote the fine scale field of interest (stacked into a column vector), with associated measurements

$$d_{\text{dense}} = C x_{\text{dense}} + v_{\text{dense}} \quad (42)$$

and let R_{dense} be the diagonal covariance of the measurement error v_{dense}. Next let $x_{\text{overlap}}, d_{\text{overlap}}, R_{\text{overlap}}$ represent the associated quantities on the finest scale of the tree. There are two requirements that we place on the mapping between x_{dense} and x_{overlap}:

- Each physical pixel in the image plane is mapped to one or more finest scale tree nodes.
- Each finest scale tree node corresponds to only one pixel in the image domain.

These requirements then lead to corresponding constraints on the "lifting" operation

$$x_{\text{overlap}} = G_x x_{\text{dense}} \quad (43)$$

namely that the elements of G_x are all zero or one and that each row has exactly one nonzero entry. A similar operator G_d "lifts" measurements d_{dense} in 2-D space to measurements on the finest level of the overlapped tree:

$$d_{\text{overlap}} = G_d d_{\text{dense}}. \quad (44)$$

G_d just contains those columns of G_x that correspond to the measurement points in 2-D space. Finally the diagonal measurement error covariance R_{overlap} is set to reflect the fact that measurements may be copied to more than one finest scale tree node,

$$G_d^T R_{\text{overlap}}^{-1} G_d = R_{\text{dense}}^{-1}; \quad (45)$$

i.e., the variance of each overlapped measurement is multiplied by the number of its redundant partners. It is also necessary to define a projection operator H:

$$x_{\text{dense}} = H x_{\text{overlap}}. \quad (46)$$

We first require that projection be the inverse of lifting, namely that $HG_x = 1$. Because G_x is one-to-many this constraint still allows considerable flexibility in choosing H, and we can use

that flexibility to achieve our smoothness objectives. We also require that H be a local operator, i.e., that each element of x_{dense} in (46) be a function only of the elements in x_{overlap} that correspond to the same physical location as the element of x_{dense}. In other words, if g_k denotes the kth column of G_x, and h_j^T the jth row of H, then

$$h_j^T g_k = \delta_{jk} (= 1 \text{ if } j = k \text{ and } 0 \text{ otherwise}). \quad (47)$$

Even with this constraint we still have additional freedom in selecting H, in that the only constraint on the nonzero elements of h_k is that they sum to 1. These nonzero values represent the "participation" of each of the redundant components of x_{overlap} in each 2-D pixel. By tapering these weights, sketched in Fig. 7a, so that they vary smoothly across overlapping regions we can achieve our smoothness objective. The procedure for specifying these weights, which is fully described in Appendix A, is recursive in scale. A simple illustration of the idea for a 1-D signal of length 10 is given in Fig. 7b. The weights shown represent the level of participation of each of the two tree nodes in each of the 10 signal points. For example, the signal point marked with a star, $\star$, has a participation factor of 0.4 from the left node and 0.6 from the right node. Each of these total participation factors (0.4, 0.6) is then further subdivided at finer levels by recursively applying the same tapering operation on all overlapped regions.

By adjusting the amount of overlap and the tapering of the weights we can adjust the amount of smoothness and remove artificial boundaries that lead to blockiness. However it is important to emphasize that the weighted averaging implied by
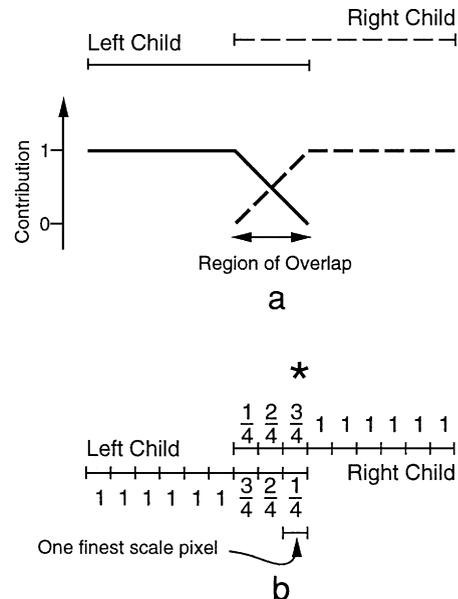
FIG. 7. Two overlapping nodes: the set of relative contributions to each finest scale pixel must sum to one. The contributions are tapered linearly over the region of overlap. (a) Tapering pictorially; (b) a specific example for two nodes which overlap by three pixels.

(46) does *not* introduce smoothing where it is not wanted; in particular, it does *not* correspond to lowpass filtering in space. Specifically, the fine scale nodes being averaged in (46) correspond to the same physical point. Thus, if the random field being modeled is known to have a statistical discontinuity at some location, that property will be reflected in *all* of the fine scale nodes to which that point corresponds.

Once we have chosen H and G_x and constructed a multiscale tree model for x_{overlap}, our estimation procedure proceeds by lifting the measurements from 2-D space to the tree according to (44), (45), performing multiscale estimation on the tree, and projecting back into the 2-D domain to produce the optimal estimates and estimation error covariances:

$$\hat{x}_{\text{dense}} = H\hat{x}_{\text{overlap}} \tag{48}$$

$$\tilde{P}_{\text{dense}} = H\tilde{P}_{\text{overlap}}H^T. \tag{49}$$

As discussed in Section 3, the estimation algorithm directly produced the error variances for the estimates at individual tree nodes—the diagonal elements of $\tilde{P}_{\text{overlap}}$. Each off-diagonal element of $\tilde{P}_{\text{overlap}}$ corresponds to the cross-covariance between estimation errors at pairs of tree nodes. From [24], each of these elements can be calculated with a number of 4×4 matrix operations proportional to the number of levels on the tree. Moreover, since $H^T H$ is quite sparse, only a comparatively small percentage of the off-diagonal elements of $\tilde{P}_{\text{overlap}}$ are required to compute the diagonal elements of $\tilde{P}_{\text{dense}}$ in (49). While the total load of these computations does not increase the total computation prohibitively, we have taken a far simpler approach by approximating the diagonal elements of $\tilde{P}_{\text{dense}}$ as

$$\text{diag}(\tilde{P}_{\text{dense}}) = \left\{ H \, \text{diag}(\tilde{P}_{\text{overlap}})^{\frac{1}{2}} \right\}^2. \tag{50}$$

This approximation corresponds to the worst-case assumption that the errors in the estimates at tree nodes corresponding to the same pixel are perfectly correlated. Thus the error variances in (50) not only represent an upper bound on the actual error variances, but also provide us with an accurate picture of the *relative* sizes of estimation error variances over the entire image domain.

Since the overlapping procedure we have just described involves a change in the multiscale tree structure from that used in the preceding section, it is also necessary to make a corresponding change in the model parameters on the tree. Because of the interpretation of thin-plate and membrane models as fractal priors, we can readily develop such a model motivated by results on hierarchical fractal surface synthesis in [28]. A key quantity in this construction is the scale to scale ratio of the dimensions of the multiscale regions. Specifically, each node of the tree represents a square region of image pixels; let w_m be the length of each such square region, measured in units of pixels, represented by nodes on scale m of the tree; then the key quantity is

$$r = \frac{w_m + 1}{w_m} \tag{51}$$

which is related to the *lacunarity* [26, 28], or texture, of the synthesized surface. Then in constructing a $1/f^2$ model on such a pyramidal structure the variance of the detail added in going from the mth to the $(m + 1)$st scale is proportional to $(1 - r)r^m$. For overlapped trees, $r > \frac{1}{2}$, thus the noise gains $2^{-m(s)/2}$ in (39) are replaced by $2(1 - r)r^{m(s)/2}$. Furthermore, in going from a nonoverlapped to an overlapped tree, the number of scales must be increased by a factor of $1/\log_2(1/r)$. The computational complexity increases as the number of tree levels is increased, so we have a tradeoff between the amount of overlap and computational load.

6. RECONSTRUCTION EXAMPLES

A meaningful comparison between our multiscale approach and established iterative methods is a challenging undertaking. The challenge stems from the fact that our multiscale approach is really addressing different primary objectives and is solving a different problem from that of iterative methods. Most iterative surface reconstruction methods, including those of Szeliski [33] and Yaou and Chang [42], are trying to solve the *same* estimation problem as efficiently as possible. That is, each method tries to solve $Ax = b$ for the same choice of A (either based on membrane or thin-plate types of models), *without* inverting A. Each of these methods produces a sequence of estimates $x_0, x_1, \dots$, and each of these sequences converges to the *same* limit x_∞—the desired surface. With such a concrete, common goal, the comparison of competing methods is obvious: investigate the rapidity of convergence of $\|x_i - x_\infty\|$ for each method.

Our multiscale approach does not fit into the picture of the previous paragraph, because we are solving a *different* problem $\bar{A}\bar{x} = b$. Consequently,

- the solution to our problem $\bar{A}\bar{x} = b$ will *never* be the same as that of $Ax = b$, regardless of the number of iterations used to solve the latter problem;
- we obtain error statistics (essentially the diagonal elements of $\bar{A}^{-1}$) for our estimation problem. These statistics are not computed for $Ax = b$ since they require the computation of A^{-1}, a problem much more difficult than the estimation of x. Thus not only are we solving a different estimation problem, but the nature of our results is fundamentally different.

The primary objective of our work is to present a method for surface reconstruction based on a modified prior model which permits the efficient computation of estimates and estimation error statistics. With this objective of our paper clearly stated, there are certain fundamental questions which are fair to ask in order to assess whether the objective has been accomplished:

1. Is the computational effort of our method on the same order as that of other surface reconstruction methods?
2. Is the modified prior a reasonable one?
3. Does the modified prior lead to issues of shift-invariance or estimation artifacts?

4. Is the modified prior model flexible enough to accommodate a range of surface reconstruction problems (e.g., discontinuities, nonstationarities)?

In terms of the computational effort, if the number of scales added to the tree to allow overlap is limited to two (we have never done otherwise), then our multiscale approach has a constant computational complexity per image pixel. As a point of comparison we used a multigrid implementation [14, 27, 35, 36, 39] based on W cycle (i.e., $\gamma = 2$) on a tree with a coarsest scale of 4×4 pixels and used ten Gauss–Seidel sweeps at each scale. If the number of iterations required to converge to an acceptable solution is independent of image size, then the computational complexity of this multigrid method, or that of preconditioned surface reconstruction methods such as that of Yaou and Chang [42] or Szeliski [33], similarly have constant computational complexity per pixel.

The second question concerns the reasonableness of our proposed multiscale prior model. We believe that the reasonableness is most effectively investigated through simple examples which, although visually less impressive, are easier to interpret and convey the advantages and disadvantages of a method more effectively than more complicated tests. In estimation problems, as the number of measurements increases and as the measurements noise variance decreases the prior model exerts less influence on the resulting estimates. Consequently in order to scrutinize the properties and qualities of our changed prior we could look at problems involving very few measurements or with very large measurement noise variances; problems having very sparse measurements are simpler to interpret, so we have chosen these as the vehicle by which to assess our changed prior in the first two subsections (6.1, 6.2).

The two remaining fundamental questions are examined in the experimental tests of the two following subsections (6.3, 6.4). In addition to these fundamental questions there are a number of *derivative* questions which might be asked, but which are of secondary interest for our paper; for example

1. What happens if we are not interested in estimation error statistics? What is the fastest method available to solve for the estimates alone?

2. Can the multiscale estimates $\bar{x}$ be used as effective initial conditions for iterative solutions to the original problem $Ax = b$?

3. Can we use the estimation residuals and error statistics to estimate the locations of discontinuities in the surface?

Although these questions are touched upon in our experiments, they are *not* the primary questions of our paper and we make no claims to have answered them conclusively.

6.1. Densely Sampled Measurements

Figure 8 summarizes our experimental results for surface reconstruction of a densely sampled surface. A smooth surface of size 64×64 pixels is shown in Fig. 8a and the corresponding dense measurements with added Gaussian noise (variance 25) are shown in Fig. 8b. The particular surface used does not correspond exactly to either the variational or multiscale prior and thus comparison between variational and multiscale reconstruction is not biased in favor of either formulation.

Figure 8c shows the reconstructed surface using an iterative multigrid algorithm [35] to solve the variational problem based on a thin-plate prior model (i.e., $\alpha = 1$, $\beta = 0$ in (13)). In this (and all future) comparison involving our multiscale approach and an iterative method (such as multigrid), the number of iterations for the iterative methods is selected such that the total computational effort of each method is the same. Figure 8d shows the reconstructed surface using our proposed multiscale algorithm computed using an overlapping tree with eight scales (a *non*overlapped tree for this surface would require seven scales) and a thin-plate-like prior model (i.e., the prior model is dominated by gradient constraints by choosing $B_s \gg 16B_g$). As indicated in the figure caption, $\mathcal{O} = (12, 6, 4, 3, 0, 0, 0)$, so the regions associated with nodes on the first level below the root node overlap by 12 pixels; at the next level the overlap is six pixels, etc.

The multigrid and multiscale reconstructions of Fig. 8 are arguably equally good—both give reasonable estimates of the original surface, and our multiscale approach shows no signs of blockiness. A more precise comparison is provided by Fig. 9. Here we consider a set of reconstruction problems in each of which only a randomly sampled subset of the noisy surface measurements is used. What Fig. 9 depicts is the Monte Carlo RMS reconstruction error for both the multiscale and multigrid methods as a function of the *fraction* of noisy surface measurements used. Observe that for the same computational effort, the multiscale algorithm performs comparably *or better* (particularly on sparse data sets) than the multigrid approach in an RMS sense. Thus the multiscale model which we have developed should *not* simply be viewed as an approximation to a variational equation, rather we are motivated by a certain variational form to develop a surface prior model—indeed, a conspicuously effective one—that leads to competitive reconstructions. Moreover it must be emphasized that the multiscale algorithm is computing surface estimates *and* estimation error variances in the time that multigrid computes surface estimates only.

6.2. Sparsely Sampled Measurements

The preceding example demonstrated that our algorithm provides competitive solutions for problems in which we have either dense or (randomly sampled) sparse data; in this section we use the sparse data of Fig. 10 to illustrate several additional issues.

The first issue concerns the easily controlled tradeoff (using our overlap method) between smoothness of reconstruction and computational complexity. Such a tradeoff is of interest because the difference between a rough reconstruction (as in Fig. 4) and a much smoother one may not be statistically significant (something which our method can determine). Thus, for purposes such as so-called line-of-sight or visibility calculations
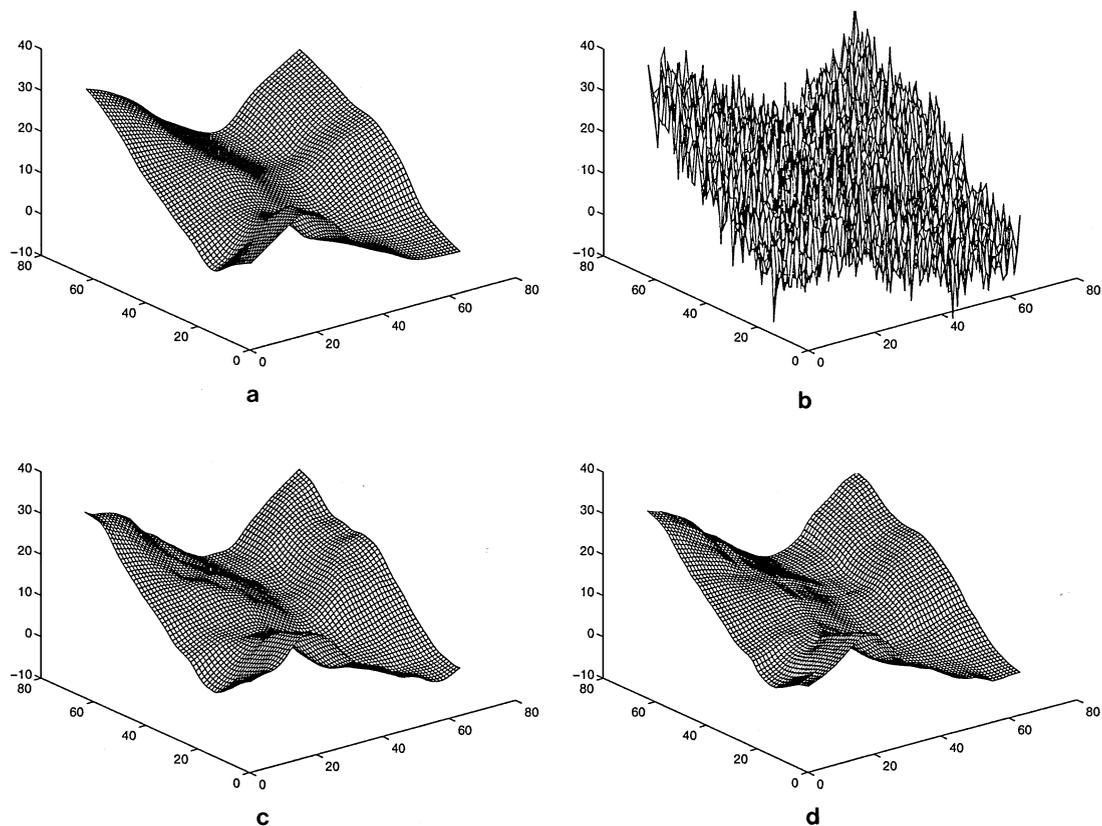
FIG. 8. Dense measurement reconstruction example: a surface is reconstructed based on dense measurements with 5 cm Gaussian noise. The multiscale and multigrid approaches involve the same number of computations. $B_s = 80$, $B_g = 0.4$, $\mathcal{O} = (12, 6, 4, 3, 0, 0, 0)$. (a) Original surface to be estimated; (b) original surface plus noise; (c) multigrid construction; (d) multiscale reconstruction.

for topographic maps, the use of a less smooth reconstruction can produce results that are just as accurate statistically, with much less computational effort, as those produced if a smoother reconstruction were used.

Figures 11a–11d show the surfaces reconstructed from the measurements of Fig. 10 and a thin-plate prior by means of four different approaches: Gauss–Seidel [6], conjugate–gradient [6], multigrid [35], and multiscale. The Gauss–Seidel and conjugate–gradient approaches are generally not practical algorithms for the surface reconstruction problems of interest; however they are well understood and many researchers have a sense for the

FIG. 9. RMS error in surface estimation using multiscale and multigrid methods as a function of the proportion of noisy measurements retained.
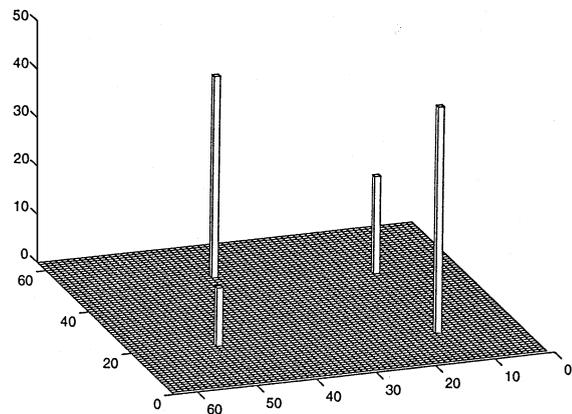
FIG. 10. Four surface samples which form the basis of the sparse-data reconstruction examples.
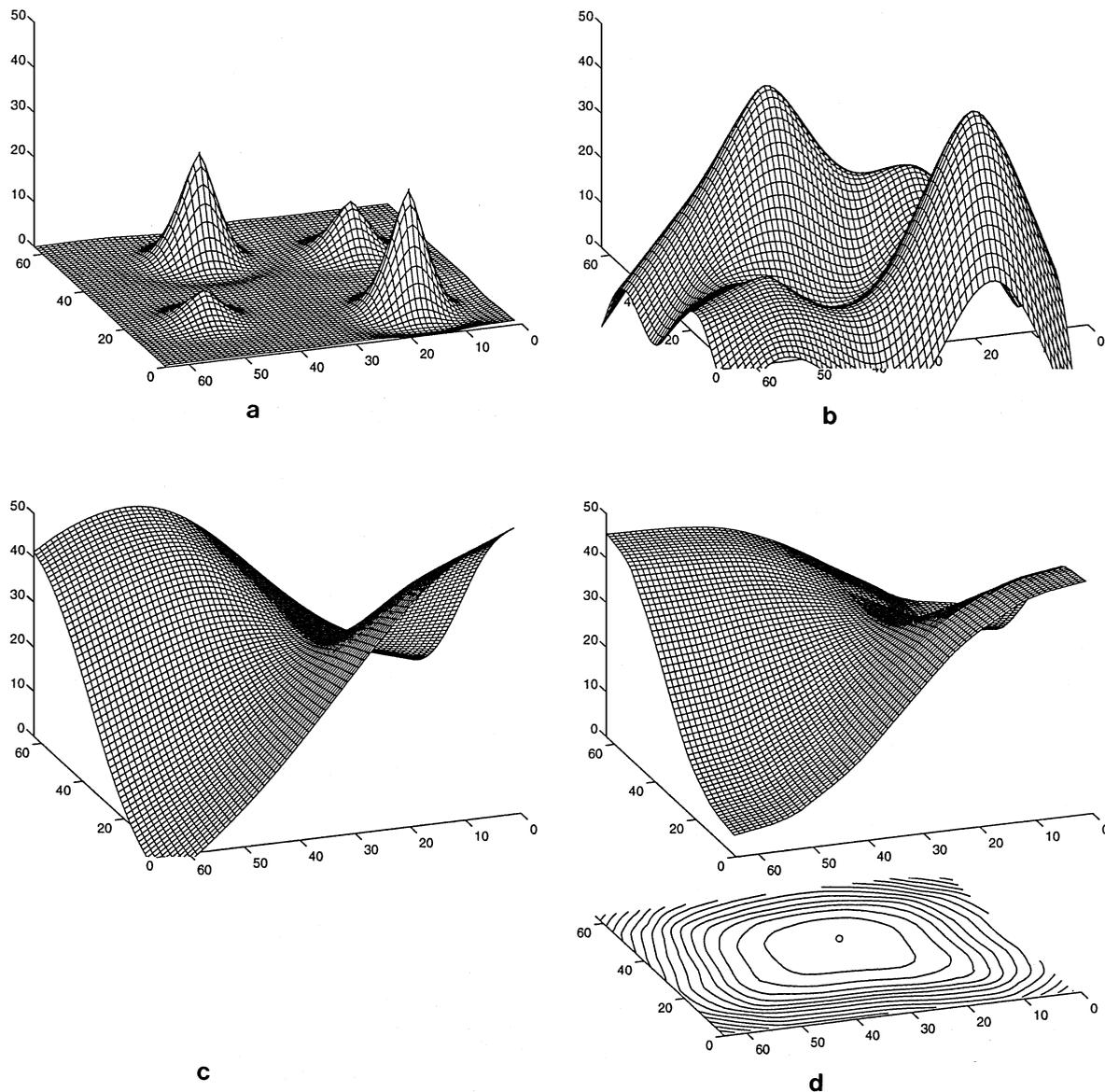
FIG. 11. Sparse measurement reconstruction examples: each reconstruction is based on the measurements of Fig. 10. Each of the methods uses the same number of total computations, except that the multiscale approach provides *both* surface estimates and error statistics. The contours of the error variance surface are shown in (d); the minimum of the error surface is marked with an "o." $B_s = 80$, $B_g = 0.05$, $\mathcal{O} = (38, 29, 24, 0, 0, 0, 0, 0)$. (a) Gauss–Seidel reconstruction; (b) conjugate–gradient reconstruction; (c) multigrid reconstruction; (d) multiscale reconstruction.

performance of these algorithms. Meaningful comparisons are achieved between these different algorithms by examining the surface estimated after a common amount of computational effort (about 35 s on a Sun SPARC-10). From Fig. 11 we can immediately see the problems associated with the Gauss–Seidel and conjugate–gradient methods, as neither is near to convergence. In contrast, both the multigrid and multiscale algorithms yield smooth estimates of arguably equal quality. In addition, the multiscale algorithm computes estimation error variances, shown in Fig. 11d, which provide useful information regarding the accuracy that can be expected from the reconstruction. In comparison, Fig. 12 shows a set of four surface reconstructions

paralleling the approaches of Fig. 11, but with one-fourth the computational effort. The multiscale-reconstructed surface of Fig. 12d is based on a tree having less overlap than its counterpart in Fig. 11d and possesses discontinuities in its estimated gradients, however, as discussed earlier, such artifacts will not be of concern in certain applications. Furthermore, Fig. 12d captures certain aspects of the true surface better than its multigrid competitor of Fig. 12c (compare in particular the upper-left portion of the two estimated surface).

A second, less fundamental, issue concerns the possibility of using our multiscale approach to accelerate iterative surface reconstruction solvers, in particular to using the multiscale
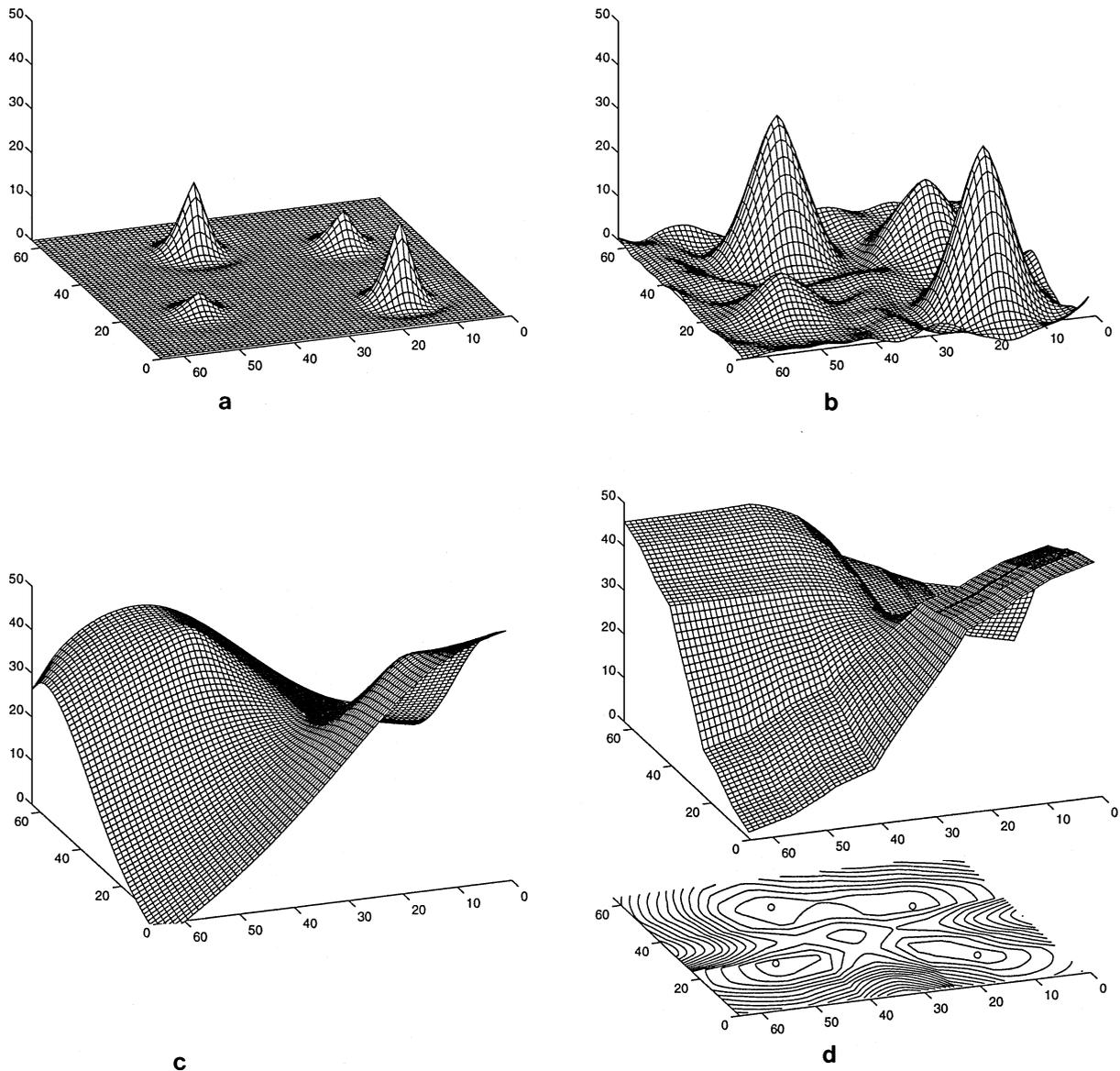
FIG. 12. Sparse measurement reconstructions using the same four methods of Fig. 11, but using one fourth of the computational effort in each case. Each minimum of the error surface in (d) is marked by an "o." $B_s = 80$, $B_g = 0.05$, $\mathcal{O} = (16, 10, 7, 0, 0, 0, 0)$. (a) Gauss–Seidel reconstruction; (b) conjugate–gradient reconstruction; (c) multigrid; (d) multiscale.

estimates as the initial guess for an iterative solver. For example, the reconstructions in Fig. 13 are computed by the application of one multigrid iteration to each of the two surfaces in Figs. 12c and 12d.

Figure 14 presents a more quantitative assessment: let $\text{rms}_{\text{MG}}(i)$ represent the RMS difference between the exact thin-plate solution and the reconstruction achieved after i multigrid iterations; let $\text{rms}_{\text{MS}}(i)$ represent the RMS error of the surface using multigrid, given as the initial guess the multiscale estimates from a seven-level tree, such that the total computational effort is the same as that of i multigrid cycles. Then Fig. 14 plots

$$\left(1 - \frac{\text{rms}_{\text{MS}}(i)}{\text{rms}_{\text{MG}}(i)}\right) \cdot 100\%; \tag{52}$$

that is, the percentage RMS improvement brought about preceding the iterative solution with our multiscale approach. Note that this percentage is significant, averaging more than 20%.

6.3. Shift-Variability of Multiscale Estimates

The previous experiments examined the estimates resulting from a fixed set of measurements; we now examine the
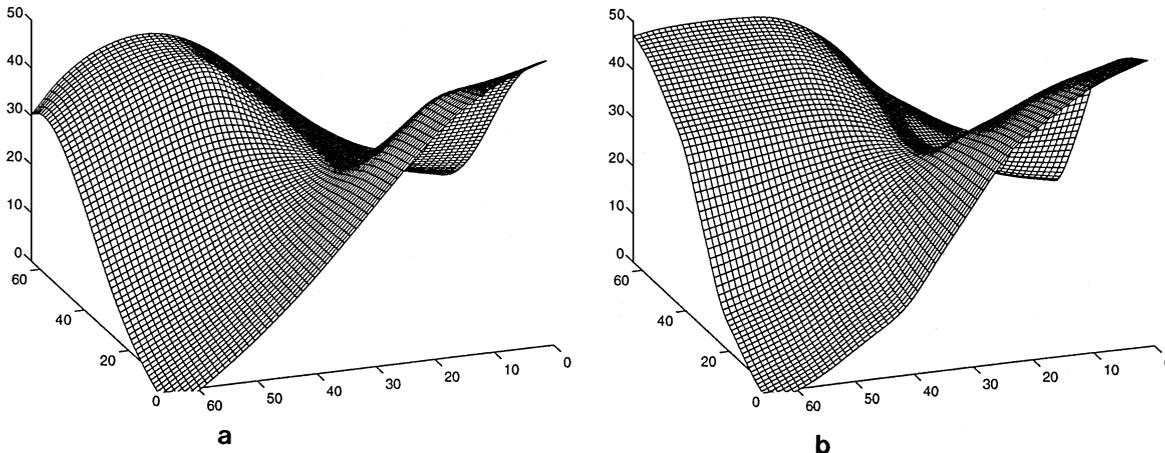
FIG. 13. Surface reconstructions computed by applying one multigrid iteration to the two surfaces Figs. 12c and 12d. $B_s = 80$, $B_g = 0.05$, $\mathcal{O} = (16, 10, 7, 0, 0, 0, 0)$. (a) Multigrid; (b) multiscale + multigrid.

variability of the estimates as the measurements are shifted in space.

Our method is not shift-invariant, due to the nonstationary structure of the multiscale tree; however, it is important to keep in mind that *any* sampled algorithm will fail to be shift-invariant. For example, standard discretization methods using regular grids do *not* produce strictly shift-invariant solutions either (e.g., if the measurements do not happen to fall on the regular grid points). Similarly, the iterative algorithms of Szeliski [33] and Yaou and Chang [42], which project the problem onto a new basis, will not produce shift-invariant reconstructions unless the algorithm is allowed to totally converge. To be sure, the issue of shift-invariant behavior is well understood for standard surface reconstruction algorithms, and it is fair to say that it is not nearly

as clear how shift-variations manifest themselves in our multiresolution framework.

Figure 15 illustrates the relationship between the proportional RMS surface variability (calculated as the ratio of the RMS variability of the estimates to the RMS value of the true surface) and the number of surface measurements; the values were determined empirically using a Monte Carlo approach. As expected, additional levels of overlap significantly attenuate the shift-variability of the estimator, such that with two extra levels of overlap the RMS variability is around one to two percent of the RMS surface value.

6.4. Surfaces with Discontinuities

As an example of estimating nonstationary fields, in this example we consider the estimation of surfaces possessing
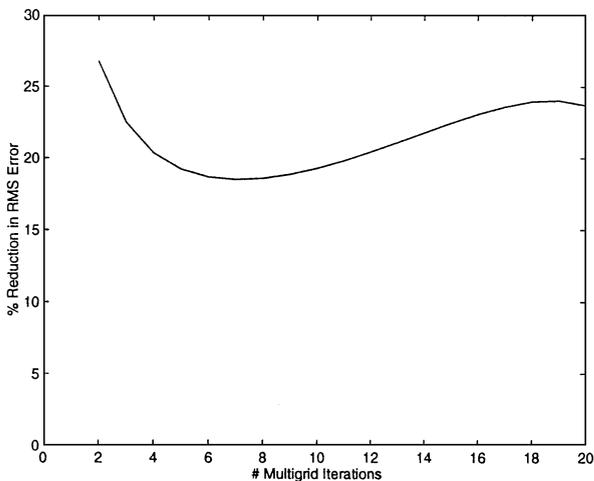
FIG. 14. Percent reduction in RMS error from just using i iterations of multigrid to computing estimates using the multiscale approach and using these estimates as the initial values for additional by multigrid iterations such that the total computation is equal to that of i multigrid iterations. The horizontal axis measures i.
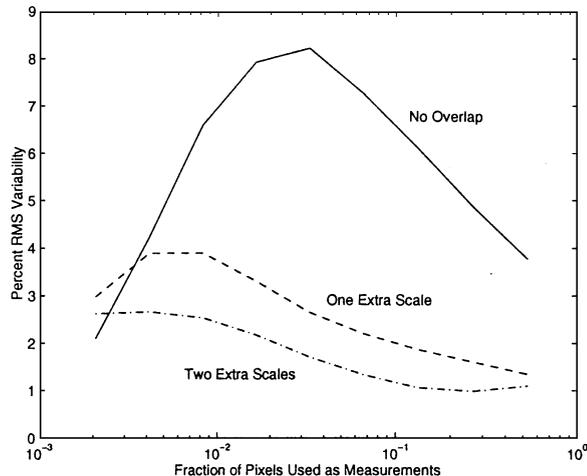
FIG. 15. Shift-variability of the multiscale estimates over random choices of the tree origin, as a function of the fraction of surface measurements. The vertical axis measures the RMS variability of the surface, normalized by the RMS value of the "truth" surface.
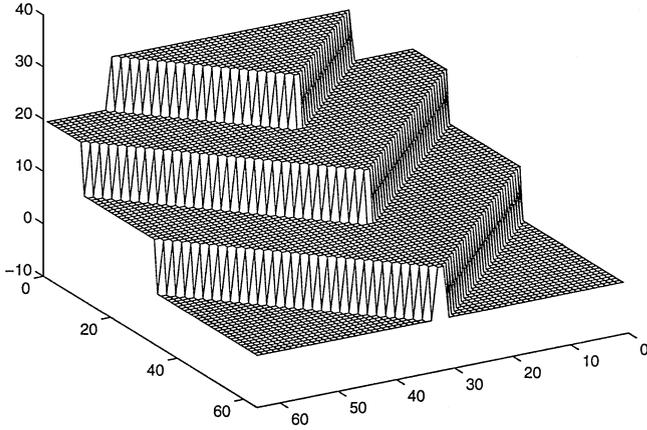
FIG. 16. Example of a discontinuous surface.

discontinuities. The overlapping multiscale tree model possesses a number of attributes which make it appropriate for such tasks:

• Unlike FFT-accelerated PDE methods, which require space-invariant surface models, the performance of the multiscale approach is unaffected by space-varying models (e.g., a piecewise thin-plate model broken by discontinuities).

• The smooth projection operator H_x of the overlapping tree does *not* blur the surface estimates spatially. As a result, the overlapping model is quite capable of capturing abrupt changes such as discontinuities.

• Because the multiscale estimator takes as input a statistical model for the unknown surface, not only the position of the discontinuities, but also the *statistics* of the discontinuity height may be specified.

The example surface which we use in this section is shown in Fig. 16. The 64×64 surface has four step discontinuities of height 10. The step edges are oriented diagonally so as to avoid a convenient alignment with the multiscale tree boundaries. The surface was measured by randomly sampling 30% of the surface elements and adding unit-variance Gaussian noise.

Suppose the locations of the discontinuities are known. We do not model each discontinuity line as a step of constant height, rather we assume that we have much less prior information and simply model each point along the discontinuity with independent zero-mean random variables: every time a branch of the multiscale tree crosses a discontinuity we will add $\eta = 100$ to the process noise variance for $z(s)$ of that branch; Fig. 17 illustrates this procedure for the example of Fig. 16:

• Let $(c_x(s), c_y(s))$ be the coordinate of the center of the region represented by multiscale node s; Fig. 17 shows these coordinates for the coarsest three tree scales.

• Let $l(s, s\bar{\gamma})$ be the line segment from $(c_x(s), c_y(s))$ to $(c_x(s\bar{\gamma}), c_y(s\bar{\gamma}))$.

• Let $\eta(s, s\bar{\gamma})$ be the sum of the variances of the discontinuities crossed by $l(s, s\bar{\gamma})$. That is $\eta(s, s\bar{\gamma})$ represents that part of the variance of $(z(s) - z(s\bar{\gamma}))$ that can be attributed to the presence of the modeled discontinuities.

Note that it is certainly possible for a given point of discontinuity to be crossed more than once in traversing the multiscale tree, i.e., in following a path from coarse to finer nodes. While one can certainly imagine adding differing amounts of uncertainty at each of these stages, we have used the simple procedure here of adding a variance of 100 at each such crossing; for example in
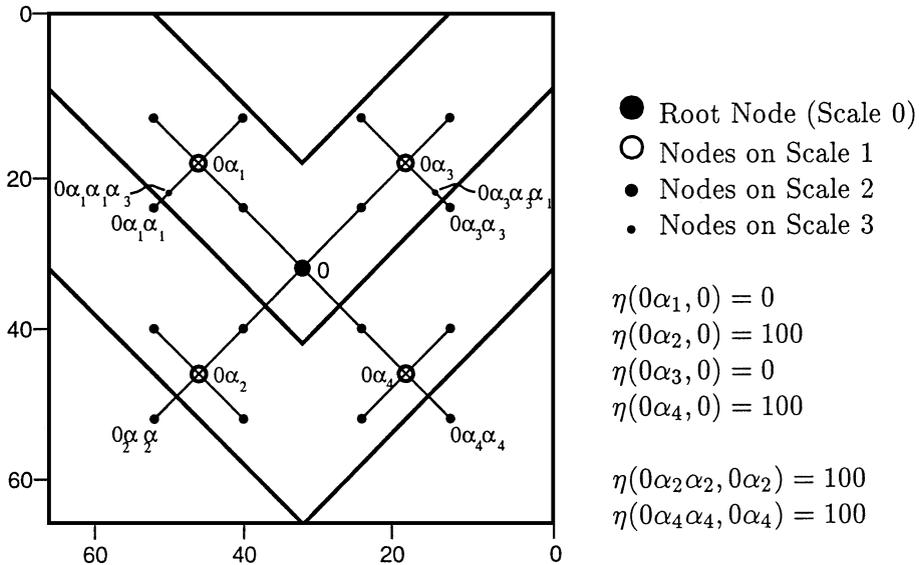
FIG. 17. This figure demonstrates the behavior of the function $\eta(s, t)$, which measures the increase in the variance of the surface due to discontinuity crossings. The positions of the circles ●, ○ represent the coordinates $(c_x(s), c_y(s))$ associated with the labeled multiscale nodes. The thick lines represent the locations of discontinuities, consistent with Fig. 16.

Fig. 17 we have $\eta(0\alpha_3\alpha_3, 0\alpha_3) = 100$ and $\eta(0\alpha_3\alpha_3\alpha_1, 0\alpha_3\alpha_3) = 100$.

We require a modified model at each node s for which $\eta(s, s\bar{\gamma}) > 0$; that is, for those nodes s where a discontinuity lies between s and its parent $s\bar{\gamma}$. When $\eta(s, s\bar{\gamma}) > 0$, (39), (40) must be modified to reflect the increased variance of the surface and the loss (or, more precisely, the irrelevance) of gradient information across the discontinuity

$$
\begin{bmatrix} z \\ p \\ q \\ zp \end{bmatrix}(s) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} z \\ p \\ q \\ zp \end{bmatrix}(s\bar{\gamma})
$$

$$
+ \begin{bmatrix} \left(B_s^2 2^{-m(s)} + \eta(s, s\bar{\gamma})\right)^{\frac{1}{2}} & 0 & 0 \\ 0 & (P_{og})^{\frac{1}{2}} & 0 \\ 0 & 0 & (P_{og})^{\frac{1}{2}} \\ 0 & 0 & 0 \end{bmatrix} w(s)
$$

$$
\tag{53}
$$

$$
d(s) = [C_1(s_i) \quad 0 \quad 0 \quad 0][z \quad p \quad q \quad zp]^T(s) + v(s), \tag{54}
$$

where P_{og} represents the gradient prior variance at the root node. Figure 18 shows the estimated surface and estimation error variances, based on this model, applied to an overlapping tree having nine scales. The surface estimates show the clear and well-preserved presence of the discontinuities. Similarly the error
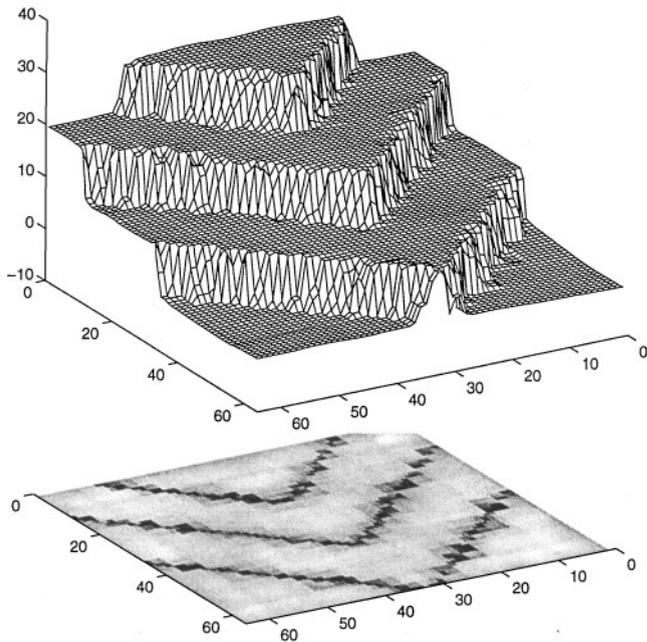
FIG. 18. Reconstruction of the surface of Fig. 16 and associated estimation error variances based on a knowledge of discontinuity locations and a sampling, at random, of one-third of the surface pixels having unit variance Gaussian noise added. The lower half of the figure plots the estimation error variances; darker regions represent greater uncertainty. $B_s = 20$, $B_g = 0.2$, $\mathcal{O} = (26, 19, 14, 9, 0, 0, 0, 0)$.
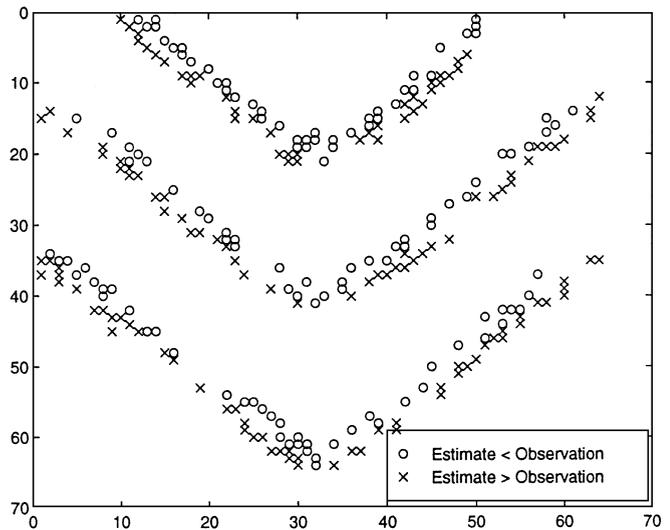
FIG. 19. Distribution of those measurement residuals, in excess of 3σ, from the estimates based on *no* knowledge of the discontinuity locations, with the same measurements as in Fig. 18. $B_s = 20$, $B_g = 0.2$, $\mathcal{O} = (26, 19, 14, 9, 0, 0, 0, 0)$.

variances are increased in the vicinity of the discontinuities, reflecting the fact that fewer measurements contribute to the surface estimates at those points.

As an aside, if the discontinuity locations are *unknown* the problem is much more complicated; however, we believe that our method represents a promising first step because of its key ability to compute posterior error statistics. For example, let us continue to use the same measurements, but now supply them to a multiscale model possessing no discontinuity information (as in (39)). We can examine the statistical significance of the measurement residual (= measurements − estimates), comparing them with their predicted variances, to identify those regions in which the multiscale thin-plate model is inconsistent with the measurements; these regions are plotted in Fig. 19. Initial progress in using this approach to automatically detect discontinuities is reported in [29].

6.5. Nonvariational Priors

While the multiscale surface reconstruction models used to this point are motivated by a certain variational thin-plate/membrane model (13), the results in the preceding sections demonstrate that our multiscale prior stands on its own as an equally valid prior model to those used in variational formulations. There are many surface statistically models which do not correspond to a thin-plate/membrane prior model and for which a variational optimization expression may be difficult to write and much more difficult to solve. Due to the flexibility of our framework, some of these surface models may readily be realized in a multiscale setting.

One such example is the class of $1/f^\mu$ prior models where μ is no longer constrained to equal 2. Although such prior models can be written in a variational form, solving the associated PDE
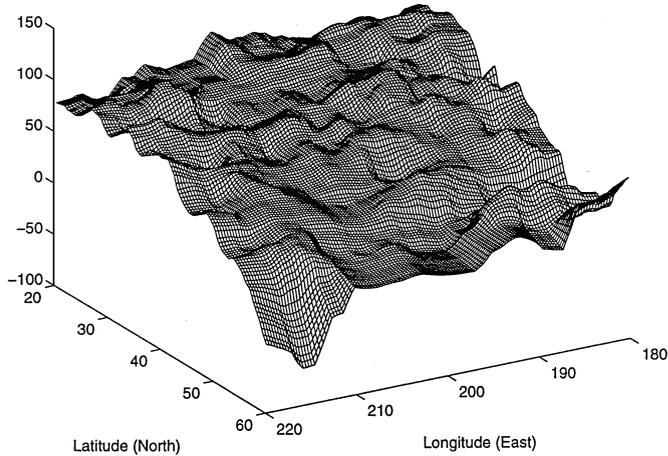
FIG. 20. A plot of the estimated ocean height, viewed from the north-east. $B_s = 35$cm, $B_g = 0.5$, $\mathcal{O} = (16, 10, 5, 3, 1, 1, 0, 0)$.

is very difficult and the inversion of the PDE operator, required to determine estimation error statistics, is much harder still.

Consider the following problem in ocean remote sensing [9]. A satellite orbiting the earth produces accurate but irregularly sampled measurements of the height of the ocean surface. It is desired to produce *smooth* maps of the ocean surface, its gradients, and corresponding error statistics on a regular, dense grid having upward of 100,000 points.

In [9], since multiscale models incorporating integrability constraints and estimating gradients had not yet been developed, we had to make do with a *much* simpler model

$$x(s) = x(s\bar{\gamma}) + 35 \text{ cm} \cdot 2^{m(s)(1-\mu)/2} \tag{55}$$

than the one considered in our current paper. Similarly, since the
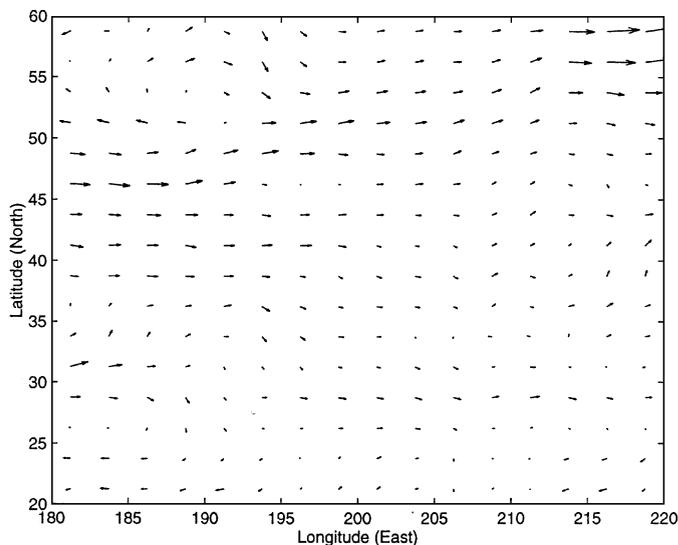
FIG. 21. The circulation field implied by the surface-gradient estimates of Fig. 20.

overlapping framework had not been developed, an ad hoc shifting approach was used (which improved smoothness somewhat, but in no way as well as the overlapped approach). Indeed, the techniques developed in this paper successfully address both the gradient and smoothness inadequacies of [9]. By setting $B_s = 35$ cm and by selecting a reasonable scale-to-scale variance for the gradient, e.g., $B_g = 0.5$, and by modifying the noise variances in (39) to $2^{m(s)(1-\mu)/2}$ we can jointly estimated the surface height and gradients from altimetric data. Figure 20 shows the resulting ocean height estimates for the north-east Pacific; the estimated oceanic circulation pattern, inferred from the estimated gradient field, is shown in Figure 21. The results confirm our physical expectations for the north Pacific ocean.

7. CONCLUSIONS

In this paper we have described and illustrated a multiresolution methodology for surface reconstruction. By using the dual interpretation of variational formulations as estimation problems and the relationship between standard variational penalties such as thin-plate and membrane models and fractal priors, we were able to define multiresolution estimation problems that possess similar interpretations and yield reconstructions of comparable quality. Furthermore, with the same computational effort, our multiscale method also produces estimation error variances, a task extremely difficult to accomplish within a variational setting. In addition, we have also shown that our multiscale estimator provides effective initial estimates for iterative solutions to variational formulations of the surface reconstruction problem.

Furthermore, while algorithms based on pyramidal quadtree representations often lead to reconstructions with blocky artifacts, a new overlapping scheme for multiresolution modeling greatly attenuates such artifacts. The appearance of artifacts may be of greater or lesser concern in different applications; by selecting the amount of overlap one has control over the computational complexity/artifact tradeoff. Furthermore, the variational formulation which motivated the multiscale model developed in Section 4 does not represent "truth," rather it is a convenient form of mathematical expression. Similarly the multiscale formulation offers not only a computationally attractive alternative, but also a flexible setting in which to construct surface prior models directly, including meaningful ones that have no simple variational counterparts.

APPENDIX A

The appendix will summarize the recursive procedure for defining G_x and H. We first consider the case of dyadic trees for 1-D signals, from which the extension to quadtrees is straightforward.

Let o_m represent the width of the overlap, measured in signal points, between the regions aggregated by sibling multiscale nodes on scale m; $\mathcal{O} = (o_1, o_2, \ldots, o_{M-1})$ represents a parameterization of the overlapping tree structure (for example, in Fig. 7, the two nodes overlap by three points). Let w_m

represent the width, measured in signal points, of the entire region associated with each multiscale node on scale m. The finest scale tree nodes must correspond to individual signal points, thus $w_{M-1} = 1$, $o_{M-1} = 0$, where M represents the number of scales on the tree. Furthermore,

$$w_{m-1} = 2w_m - o_m. \tag{56}$$

Consider any node s on the finest scale; s may be written as

$$s = 0\alpha_{i_1}\alpha_{i_2}\cdots\alpha_{i_{M-1}} \quad i_m \in \{0, 1\}, \tag{57}$$

where 0 is the root node, and α_0, α_1 correspond to downward left and right shifts, respectively. Let

$$\bar{n}(s) = \sum_{m=1}^{M-1} i_m 2^{M-m-1} \tag{58}$$

$$n(s) = \sum_{m=1}^{M-1} i_m(w_m - o_m) \tag{59}$$

then the projection matrix G_x is given by

$$G_x(a, b) = \begin{cases} 1 & \text{if } \exists s \ni \bar{n}(s) = a, n(s) = b \\ 0 & \text{otherwise.} \end{cases} \tag{60}$$

Furthermore, let

$$ofs_{M-1} = 0 \quad ofs_{m-1} = i_m(w_m - o_m) + ofs_m \tag{61}$$

$$h_m(s) = \begin{cases} 1 & i_m = 0; ofs_m < w_m - o_m \\ \frac{w - ofs_m}{o_m + 1} & i_m = 0; ofs_m \geq w_m - o_m \\ 1 & i_m = 1; ofs_m \geq o_m \\ \frac{ofs_m + 1}{o_m + 1} & i_m = 1; ofs_m < o_m \end{cases} \tag{62}$$

then the smoothing matrix H is given by

$$H(b, a) = \begin{cases} \prod_{m=1}^{M-1} h_m(s) & \text{if } \exists s \ni \bar{n}(s) = a, n(s) = b \\ 0 & \text{otherwise.} \end{cases} \tag{63}$$

The extension of these results to the quadtree case is straightforward if the quadtree is viewed as a product of two dyadic trees. Each node s on the finest scale may be represented as

$$s = 0\alpha_{i_1 j_1}\alpha_{i_2 j_2}\cdots\alpha_{i_{M-1} j_{M-1}} \quad i_m, j_m \in \{0, 1\}, \tag{64}$$

where α_{ij} represents the ith descendant in the x direction and the jth descendant along y. Consider the dyadic tree nodes

$$s_i = 0\alpha_{i_1}\alpha_{i_2}\cdots\alpha_{i_{M-1}} \tag{65}$$

$$s_j = 0\alpha_{j_1}\alpha_{j_2}\cdots\alpha_{j_{M-1}} \tag{66}$$

then the operators G_x, H, now viewed as projections between two-dimensional processes, are defined as

$$G_x(a, b) = \begin{cases} 1 & \text{if } \exists s \ni \{\bar{n}(s_i), \bar{n}(s_j)\} = a, \{n(s_i), n(s_j)\} = b \\ 0 & \text{otherwise.} \end{cases} \tag{67}$$

$$H(b, a) = \begin{cases} \prod_{i=1}^{M-1} h_m(s_i)h_m(s_j) & \text{if } \exists s \ni \{\bar{n}(s_i), \bar{n}(s_j)\} = a, \\ & \qquad \{\bar{n}(s_i), \bar{n}(s_j)\} = b \\ 0 & \qquad\qquad \text{otherwise.} \end{cases} \tag{68}$$

APPENDIX B

The implementation of the multiscale estimation algorithm [4, 23] is a rather complicated undertaking. In the interest of promoting the use of this algorithm and enabling interested researchers to apply it to problems of their own, we are making this code publicly available.

The code is written in MATLAB scripts and in C; the front-end visible to the user is written in MATLAB, and the multiscale computational engine is written in C. No programming experience is needed to try the software, although an understanding of MATLAB scripts would be required to customize our program for a different application.

Anyone interested in compiling and running our code will require MATLAB software and an ANSI-compatible C compiler (precompiled versions of the code, not requiring any compilation, are available for Sun-SPARC platforms). The programs are available via anonymous FTP from ocho.uwaterloo.ca (IP Address 129.97.172.37) in pub/Software/Multiscale Surface. The file README describes the purpose of each program and how to get started.

REFERENCES

1. M. Bertero, T. Poggio, and V. Torre, Ill-posed problems in early vision, *Proc. IEEE* **76**(8), 1988, 869–889.

2. J. Bramble, J. Pasciak, and A. Schatz, The construction of preconditioners for elliptic problems by substructuring III, *Math. Comp.* **51**(184), 1988, 415–430.

3. J. Bramble, R. Ewing, R. Paraskkevov, and J. Pasciak, Domain decomposition methods for problems with partial refinement, *SIAM J. Scie. Statist. Comput.* **13**(1), 1992, 397–410.

4. K. Chou, A. Willsky, and A. Benveniste, Multiscale recursive estimation, data fusion, and regularization, *IEEE Trans. Automat. Control* **39**(3), 1994, 464–478.

5. R. Courant and D. Hilbert, *Methods of Mathematical Physics,* Vol. 1, Interscience, New York, 1953.

6. B. Dahlquist, *Numerical Methods*, Prentice-Hall, Englewood Cliffs, NJ, 1974.

7. W. Dahmen and A. Kunoth, Multilevel preconditioning, *Numer. Math.* **63**(3), 1992, 315–344.

8. D. Evans (Ed.), *Preconditioning Methods: Analysis and Application*, Gordon & Breach, New York, 1983.

9. P. Fieguth, W. Karl, A. Willsky, and C. Wunsch, Multiresolution optimal interpolation and statistical analysis of Topex/Poseidon satellite altimetry, *IEEE J. Geosci. Remote Sensing* 1995, 280–292.

10. P. Fieguth, *Application of Multiscale Estimation to Large Scale Multidimensional Imaging and Remote Sensing Problems*, Ph.D. Thesis, Massachusetts Institute of Technology, 1995.

11. P. W. Fieguth and A. S. Willsky, Fractal estimation using models on multiscale trees, *IEEE Trans. Signal Process.* **44**(5), 1996, 1297–1300.

12. R. Franktot and R. Chellappa, A method for enforcing integrability in shape from shading algorithms, *IEEE Trans. Pattern Anal. Mach. Intell.* **10**(4), 1989, 439–451.

13. S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.* **6**(6), 1984, 721–741.

14. W. Hackbusch, *Multi-Grid Methods and Applications*, Springer-Verlag, Berlin, 1985.

15. J. Harris, *The Coupled Depth/Slope Approach to Surface Reconstruction*, MIT S.M. Thesis, 1986.

16. B. Horn, *Robot Vision*, MIT Press, Cambridge, MA, 1986.

17. B. Horn, Height and gradient from shading, *Int. J. Comput. Vision* **5**(1), 1990, 37–46.

18. Ikeuchi and B. Horn, Parallelizable shape from shading, *Artificial Intell.* **17**(1–3), 1981, 141–186.

19. W. Irving, P. Fieguth, and A. Willsky, An overlapping tree approach to multiscale stochastic modeling and estimation, *IEEE Trans. Image Process.* **6**(11), 1997, 1517–1529.

20. D. Keren and M. Werman, Probabilistic analysis of regularization, *IEEE Trans. Pattern anal. Mach. Intell.* **15**(10), 1993, 982–995.

21. K. Lee and C. Kuo, Shape from shading with a linear triangular element surface model, *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(8), 1993, 815–822.

22. D. Lee and J. Shiau, Thin plate splines with discontinuities and fast algorithms for their computation, *SIAM J. Scie. Comput.* **15**(6), 1994, 1311–1330.

23. M. Luettgen, W. Karl, and A. Willsky, Efficient multiscale regularization with applications to the computation of optical flow, *IEEE Trans. Image Process.* **3**(1), 1994, 41–64.

24. M. Luettgen and A. Willsky, Multiscale smoothing error models, *IEEE Trans. Automat. Control* **40**(1), 1995.

25. M. Luettgen and A. S. Willsky, Likelihood calculation for a class of multiscale stochastic models, with application to texture discrimination, *IEEE Trans. Image Process.* **4**(2), 1995, 194–207.

26. B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco, 1983.

27. S. McCormick, *Multilevel Adaptive Methods for Partial Differential Equations*, SIAM, Philadelphia, 1989.

28. D. Saupe, Algorithms for random fractals, in *The Science of Fractal Images* (Peitgen and Saupe, Eds.), Springer-Verlag, Berlin/New York, 1988.

29. M. Schneider, P. Fieguth, W. Karl, and A. Willsky, Multiscale methods for the segmentation of images, *IEEE ICASSP'96*, 1996.

30. S. Sinha and B. Schunck, A two stage algorithm for discontinuity detection, *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(1), 1992, 36–55.

31. R. Szeliski, *Bayesian Modeling of Uncertainty in Low-level Vision*, Kluwer Academic, Dordrecht/Norwell, MA, 1989.

32. R. Szeliski and D. Terzopoulos, From splines to fractals, *Comput. Graphics* **23**(3), 1989, 51–60.

33. R. Szeliski, fast surface interpolation using hierarchical basis functions, *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(6), 1990, 513–528.

34. R. Szeliski and D. Tannesen, Surface modeling with oriented particle systems, *Comput. Graphics* **26**(2), 1992, 185–194.

35. D. Terzopoulos, Multilevel computer processes for visual surface reconstruction, *Comput. Vision Graphics Image Process.* **24**, 1983, 52–96.

36. D. Terzopoulos, Regularization of inverse visual problems involving discontinuities, *IEEE Trans. Pattern Anal. Mach. Intell.* **8**(4), 1986, 413–424.

37. G. Wahba, Bayesian "confidence intervals" for the cross-validated smoothing spline, *J. Roy. Statist. Soc. B* **45**, 1983, 133–150.

38. G. Wahba, *Spline Models for Observational Data*, Series in Applied Mathematics, Vol. 59, SIAM, Philadelphia, 1990.

39. P. Wesseling, *An Introduction to Multigrid Methods*, Wiley, London, 1991.

40. G. Whitten, Scale space tracking and deformable sheet models for computational vision, *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(7), 1993.

41. G. Wornell, Wavelet-Based Representation for the $1/f$ Family of Fractal Processes, *Proc. IEEE*, Sept. 1993.

42. M. Yaou and W. Chang, Fast surface interpolation using multiresolution wavelet transform, *IEEE Trans. Pattern Anal. Mach. Intell.* **16**(7), 1994, 673–688.

43. H. Yserentant, Two preconditioners based on the multilevel splitting of finite element spaces, *Numer. Math.* **58**(2), 1990, 163–184.