Expectation-maximization algorithms for image processing using multiscale models and mean- field theory, with applications to laser radar range profiling and segmentation

Andy Tsai
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, Massachusetts 02139
E-mail: atsai@mit.edu

Jun Zhang
Department of Electrical Engineering and Computer Science
University of Wisconsin–Milwaukee
Milwaukee, Wisconsin 53201
E-mail: junzhang@csd.uwm.edu

Alan S. Willsky
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, Massachusetts 02139
E-mail: willsky@mit.edu

Abstract. We describe a new class of computationally efficient algorithms designed to solve incomplete-data problems frequently encountered in image processing and computer vision. The basis of this framework is the marriage of the expectation-maximization (EM) procedure with two powerful methodologies. In particular, we have incorporated optimal multiscale estimators into the EM procedure to compute estimates and error statistics efficiently. In addition, mean-field theory (MFT) from statistical mechanics is incorporated into the EM procedure to help solve the computational problems that arise from our use of Markov random-field (MRF) modeling of the hidden data in the EM formulation. We have applied this algorithmic framework and shown that it is effective in solving a wide variety of image-processing and computer-vision problems. We demonstrate the application of our algorithmic framework to solve the problem of simultaneous anomaly detection, segmentation, and object profile estimation for noisy and speckled laser radar range images.
[DOI: 10.1117/1.1385168]

Subject terms: EM algorithm; mean-field theory; multiscale models; laser radar; image processing.

Paper 200230 received June 8, 2000; accepted for publication Feb. 19, 2001.

1 Introduction

Estimation problems for 2-D random fields arise in contexts ranging from image processing to remote sensing. In some special cases—most notably spatially stationary statistics over a rectangular grid with regularly spaced measurements—very efficient algorithms based on the fast Fourier transform (FFT) can be used to obtain optimal estimates and error statistics. However, when one strays from these cases, the complexity of optimal solutions for many popular classes of stochastic models can become prohibitive with computational loads that do not scale well with problem size.

One approach that overcomes many of these problems involves the use of multiscale stochastic models[1,2] and associated multiscale estimation algorithms. These algorithms do not require stationarity, regularly shaped domains, or regularly spaced measurements. Moreover, these models have been shown to capture a rich class of random-field statistical behavior, making them useful for a number of important applications.

In this paper we extend the applicability of this multiscale framework by marrying it with two other computationally powerful techniques, namely the expectation-maximization (EM) algorithm and mean-field theory (MFT). The result is a methodology that can be applied to a greatly expanded range of applications, including texture segmentation, simultaneous classification and gain correction of magnetic resonance imaging (MRI) brain scans,[3] and a relaxed version of the Mumford-Shah variational approach to image segmentation. A number of these applications are described in some detail in Ref. 4, and Fig. 1 depicts an example of the application to a texture segmentation problem.

In particular, as described in Ref. 5, each of the individual textures shown in Fig. 1 can be well modeled with multiscale models, allowing us to apply efficient multiscale estimation procedures to problems such as restoring noise-corrupted versions of any one of these textures. However, the multiresolution algorithm by itself cannot solve the problem of restoring the image shown in Fig. 1, as it consists of several *different* textures, nor can it directly solve the problem of segmentation of such an image. However, by employing the EM concept of *hidden data* (in this case, the segmentation map indicating which of the two wood textures is visible in each pixel), we can indeed employ multiscale estimation as a core component of an algorithm that produced the results (restoration, segmentation, and associated error variances) shown in Fig. 1.

Figure 2 depicts the conceptual structure of our methodology. Viewing the EM algorithm as the central structure of an iterative procedure, we use multiscale estimation to solve the so-called M step of the procedure and MFT to provide an efficient approximation to the E step. In the next section we briefly describe the background topics that form the components of the triad in Fig. 2. In Secs. 3 and 4 we then develop our algorithm in the context of an important

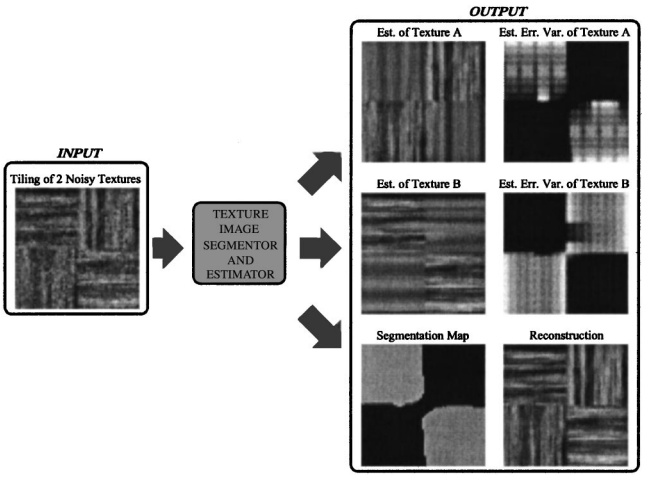


Fig. 1 Motivational example: texture image segmentation and estimation.

application, namely laser radar (ladar) image processing, target segmentation, and range profiling. We have chosen this application for several reasons. First, as we will see, the employment of our framework in this context involves more than just segmentation, and, as a result, using it as a vehicle illustrates both the breadth of problems to which it can be applied and how it is applied. Secondly, this particular application is of significant current interest, and, indeed, our results represent a substantial extension of previous work on ladar anomaly rejection and background range plane profiling. Following the development of our approach, we present results in Sec. 5 demonstrating its efficacy in simultaneously rejecting anomalies, profiling nonplanar backgrounds, detecting and segmenting targets, and providing range profiles of those targets. The paper then closes with a brief discussion and conclusion.

2 Background

As we have indicated, our framework involves the synthesis of three distinct methodologies for statistical inference, and in this section we provide concise summaries of each of these. More complete description of these topics can be found in some of the references (e.g., Refs. 1,2, and 5–8).

2.1 *The EM Procedure*

The EM procedure is a powerful iterative technique suited for calculating the maximum-likelihood estimates (MLEs) in problems where the observation can be viewed as incomplete data. The MLE of X, denoted as $\hat{X}_{\mathrm{ML}}$, based on the incomplete observed data Y, is defined as

$$\hat{X}_{\mathrm{ML}} = \arg\max_{X}\{\log p(Y|X)\}, \tag{1}$$

where $\log p(Y|X)$ is the log likelihood of Y given X. In many applications, calculating the MLE is difficult because the log-likelihood function is highly nonlinear and not easily maximized. To overcome these difficulties, the EM algorithm introduces an auxiliary function Q (along with some auxiliary random variables) that has the same behavior as the log-likelihood function (in that when the log-likelihood function increases, so does the auxiliary function) but is much easier to maximize.

Central to the EM method is the judicious introduction of an auxiliary random quantity W with log likelihood $\log p(W|X)$. The data W is referred to as the *complete data* because it is more informative than Y. The complete data W is not observed directly, but indirectly through Y via the relationship $Y=\mathcal{G}(W)$, where $\mathcal{G}$ is a many-to-one mapping function. Those unobserved variables are referred to in the EM formulation as the *hidden data* and denoted by H. The EM approach calculates $\hat{X}_{\mathrm{ML}}$ through an iterative procedure in which the next iteration's estimate of X is chosen to maximize the *expectation* of $\log p(W|X)$ given the incomplete data Y and the current iteration's estimate of X. The iteration consists of two steps:

- The E step computes the auxiliary function

$$Q(X|X^{[k]}) = \langle \log p(W|X) | Y, X^{[k]} \rangle, \tag{2}$$

  where $\langle \cdot \rangle$ represents expectation, and $X^{[k]}$ is the estimate of X from the k'th iteration. Often, this step reduces to calculating the expected value of H given Y and $X^{[k]}$.
- The M step finds $X^{[k+1]}$ such that

$$X^{[k+1]} = \arg\max_{X}\{Q(X|X^{[k]})\}.$$

  The maximization here is with respect to X, the first argument of the function Q. Intuitively, the M step is designed to use the expected value of the hidden data H found in the E step as if it were measured data in order to obtain the ML estimate of X. The EM algorithm can easily be extended to a maximum *a posteriori* (MAP) estimator by imposing a prior model, $p(X)$, on the estimated quantity X during this step. With this modification, the M step finds the $X^{[k+1]}$ such that

$$X^{[k+1]} = \arg\max_{X}\{Q(X|X^{[k]}) + \log p(X)\}. \tag{3}$$

Since the EM algorithm is iterative, initial conditions must be given to start it. As it is guaranteed to converge only to a local maximum of the likelihood function, choos-

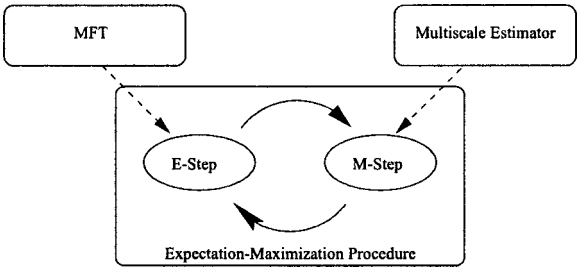


Fig. 2 Conceptual representation of the algorithmic framework.

ing initial conditions requires some care. The E and the M steps are evaluated repeatedly until convergence, typically specified as when $\|X^{[k+1]}-X^{[k]}\|<\delta$ for some small $\delta \geqslant 0$.

2.2 *Mean-Field Theory*

MFT, which provides a computationally efficient procedure for approximate calculation of expectations of large Markov random fields (MRFs), has its origins in statistical mechanics and is concerned with finding the mean field energy of an MRF. In particular, let $\mathcal{L}$ denote a two-dimensional rectangular lattice. Let $u=\{u_i, i\in\mathcal{L}\}$ be a MRF taking on the well-known Gibbs distribution (see, e.g., Ref. 8) given by

$$p(u)=\frac{\exp[-\zeta U(u)]}{\Gamma}.$$

The constant $\zeta \geqslant 0$ is used to adjust the strength of the energy function $U(u)$ defined as

$$U(u)=\sum_{\text{clq}} V_{\text{clq}}(u)$$

where the $V_{\text{clq}}(u)$ are the clique potentials used to capture the interactions between pixels. The normalization constant Γ is defined as

$$\Gamma=\sum_{\tilde{u}} \exp[-\zeta U(\tilde{u})].$$

For simplicity of exposition, pixel interactions are assumed to be at most pairwise. The energy function can now be written as

$$U(u)=\sum_{i\in\mathcal{L}}\left[V_{\text{clq}}(u_i)+\frac{1}{2}\sum_{l\in\mathcal{N}_i} V_{\text{clq}}(u_i,u_l)\right], \tag{4}$$

where the singleton clique $V_{\text{clq}}(\cdot)$ and the doubleton clique $V_{\text{clq}}(\cdot,\cdot)$ are the clique potentials for a single site and a pair of neighboring sites (horizontal or vertical), respectively. Finally, as a result of our simplifying assumption, $\mathcal{N}_i$ denotes the set of first-order neighbors of pixel i.

MFT is used to compute an approximation to the mean of the field u, i.e., to find

$$\begin{aligned}\langle u_i\rangle &= \sum_{u} u_i p(u)=\sum_{u_i} u_i \sum_{\mathcal{L}\setminus i} p(u)\\ &= \sum_{u_i} u_i \frac{\Sigma_{\mathcal{L}\setminus i}\exp[-\zeta U(u)]}{\Gamma}\end{aligned} \tag{5}$$

for each $i\in\mathcal{L}$ (where $\mathcal{L}\setminus i$ denotes the set of all pixels excluding i). It is clear from (5) that a large number of configurations associated with the energy function need to be evaluated in order to calculate this expectation, thus making its precise calculation both complex and computationally impractical. To overcome this, MFT approximates the weight on each possible value of u_i in the rightmost expression of Eq. (5). In particular, for each site i and each possible value of u_i, we approximate this weight by replacing the values of each of the u_j, $j\neq i$, by its mean value. That is, we define the mean field energy at site i as

$$U_i^{\text{mf}}(u_i)=U(u)|_{u_j=\langle u_j\rangle,\ j\neq i}=U_i^{\text{mf}'}(u_i)+R_i^{\text{mf}'}(\langle u_{\mathcal{L}\setminus i}\rangle), \tag{6}$$

where $U_i^{\text{mf}'}(u_i)$ includes all the clique potentials involving the value at site i and is given by

$$U_i^{\text{mf}'}(u_i)=V_{\text{clq}}(u_i)+\sum_{l\in\mathcal{N}_i} V_{\text{clq}}(u_i,\langle u_l\rangle), \tag{7}$$

and $R_i^{\text{mf}'}(\langle u_{\mathcal{L}\setminus i}\rangle)$ consists of terms involving only means at sites other than i. Further we use the mean-field approximation for the mean at each site:

$$\begin{aligned}\langle u_j\rangle &\approx \sum_{u_j} u_j \frac{\exp[-\zeta U_j^{\text{mf}'}(u_j)]}{\Gamma_j^{\text{mf}'}},\\ \Gamma_j^{\text{mf}'} &= \sum_{\tilde{u}_j} \exp[-\zeta U_j^{\text{mf}'}(\tilde{u}_j)].\end{aligned} \tag{8}$$

The MFT formulation supposes that for a particular pixel $i\in\mathcal{L}$, the influence of the energy field from other pixels can be approximated by that of their statistical means. Random fluctuations from the other pixels are neglected. Note that since $U_i^{\text{mf}}(u_i)$ depends on the means $\langle u_j\rangle$, $j\neq i$, the approximations to $\langle u_i\rangle$ in Eq. (8) depends on the means at other pixels. Thus the set of equations corresponding to Eq. (8) for all $i\in\mathcal{L}$ represents a set of simultaneous algebraic equations for all of the $\langle u_i\rangle$. Iterative algorithms can then be used to find the solution to this set of equations.

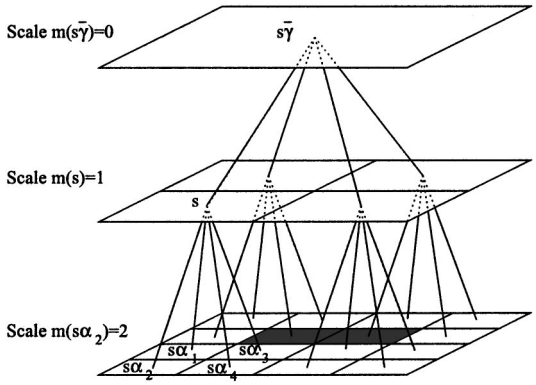


Fig. 3 An example of a multiscale quadtree.

2.3 *Multiscale Stochastic Models and Optimal Estimation*

In this subsection, we briefly review the multiscale statistical framework for modeling and estimating 2-D random fields.[1,2,5] Multiscale stochastic models for image processing are defined on a pyramidal data structure such as the quadtree shown in Fig. 3. Each level of the tree corresponds

to a different scale of resolution in the representation of the random field of interest. Define the index s to specify the nodes on the tree, $s\overline{\gamma}$ to represent the parent of node s, and $s\alpha_i$ to denote the q offspring of node s with $i=1,\ldots,q$ ($q=4$ for a quadtree). The scale of node s is denoted as $m(s)$ and is numbered sequentially from coarse to fine resolution. A multiscale model with state $\mathcal{X}(s)\in\mathcal{R}^n$ is specified by a recursion of the form

$$\mathcal{X}(s)=\mathcal{A}(s)\mathcal{X}(s\overline{\gamma})+\mathcal{B}(s)\mathcal{W}(s) \tag{9}$$

with $\mathcal{X}(0)\sim N(0,P_o)$, $\mathcal{W}(s)\sim N(0,I)$, and $\mathcal{W}(s)\in\mathcal{R}^m$. [The notation $N(\mu,\Lambda)$ represents a Gaussian random vector x with mean μ and variance Λ.] The matrices $\mathcal{A}(s)$ and $\mathcal{B}(s)$ are the parameters of the tree model, appropriately chosen to represent the random field of interest. The root node of the tree $\mathcal{X}(0)$ with prior covariance P_o provides the initial condition for starting the coarse-to-fine recursion. The driving white noise $\mathcal{W}(s)$ is independent of the initial condition $\mathcal{X}(0)$. The states at a given scale can be thought of as information about the process at that level of the tree. As $\mathcal{X}(s)$ evolves from coarse to fine scale, the term $\mathcal{A}(s)\mathcal{X}(s\overline{\gamma})$ predicts the states of the process for the next-finer scale while the term $\mathcal{B}(s)\mathcal{W}(s)$ adds new details to these states.

The measurement model associated with this multiscale stochastic model is

$$\mathcal{Y}(s)=\mathcal{C}(s)\mathcal{X}(s)+\mathcal{V}(s) \tag{10}$$

with $\mathcal{V}(s)\sim N(0,\mathcal{R}(s))$, $\mathcal{R}(s)$ the covariance of $\mathcal{V}(s)$, and $\mathcal{C}(s)$ the measurement matrix used to describe the nature of the measurement process. The general multiscale estimation framework allows easy and direct fusion of measurements at all scales (see, for example, Ref. 9). For the application on which we focus here, all measurements will be scalar observations at the finest scale, representing point measurements of the random field of interest.

The algorithm for optimal multiscale estimation consists of two steps: a fine-to-coarse sweep (which resembles the time-series Kalman filtering algorithm) followed by a coarse-to-fine sweep (which is analogous to the Rauch-Tung-Striebel smoothing algorithm). Among the most important properties of this algorithm is that it produces both optimal estimates and the covariance of the errors in these estimates with total complexity that is $\mathcal{O}(k^3n)$, where n is the number of fine-scale pixels and k is the dimension of the state $\mathcal{X}(s)$. Thus the computation load for this algorithm grows linearly with problem size. This should be contrasted with the complexity of other estimation approaches (e.g., those based on MRF models), which have complexity that is $\mathcal{O}(n^\alpha)$ for values of $\alpha>1$ for calculation of the estimates, and typically even greater complexity (often prohibitively high) for the computation of error covariances. Of course, the utility of the multiscale approach also depends on having models with modest state dimension k. Demonstration that such models do exist for a wide variety of random fields can be found in Refs. 5,10,11. Of specific relevance to this paper are the multiscale models developed in Ref. 10 corresponding to so-called thin-plate and thin-membrane random surfaces. A brief summary of these models is presented in Sec. 7.1.

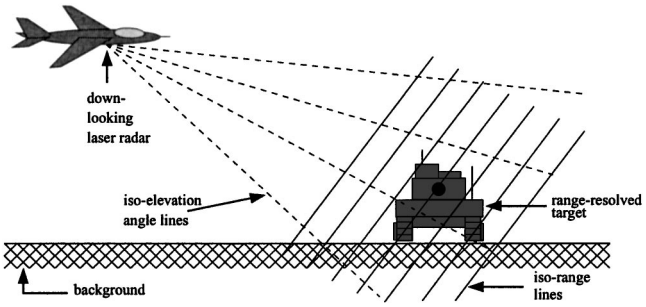


Fig. 4 Down-looking geometry for collecting 3-D images of ground-based scenes.

3 Probabilistic Models for Laser Radar Range Profiling and Segmentation

As illustrated in Fig. 4, and as described in detail in Refs. 12,13, raster-scanned and range-resolved ladar imaging provides the capability for producing 3-D images of ground-based objects from airborne platforms using a down-looking geometry. Utilization of such images [e.g., for automatic target recognition (ATR)] involves several basic functions. First, of course, is dealing with the uncertainties and sources of error in the ladar measurements. As described in Ref. 12, these errors consist of relatively small perturbations due to noise and occasional large anomalies due to deep speckle fades. Since the frequency of anomalies can be significant, taking account of them is a serious issue. Second, accurate localization and profiling of any target present in a ladar range image requires estimation of the unknown and spatially varying range of the ground. Detection and profiling of the target then requires segmenting out the region over which the range measurements stand out from the background, indicating the presence of the target.

As a first-generation preprocessor for an ATR system, Green and Shapiro[12] developed an algorithm that estimates the background range profile for a ladar range image corrupted by range anomalies. Only the problem of estimating the background range profile is addressed in their algorithm. The idea is that pixels that do not belong to the background are grouped together as either range anomalies or target range measurements. Their algorithm is based on an idealized single-pixel statistical model for the ladar range measurements together with the assumption that the background range profile is a plane parameterized by the elevation and azimuth range slopes and by the range intercept. The EM procedure is employed in Green and Shapiro's algorithm. The E step localizes and suppresses the range anomalies; the M step finds the ML estimate of the three parameters that characterize the planar background range profile. Simulation studies have confirmed that Green and Shapiro's technique is computationally simple with good range anomaly suppression capabilities.

The application of our EM–multiscale-MFT estimation framework to the ladar range imaging problem extends Green and Shapiro's work in several fundamental and prac-

tically significant ways. First, rather than modeling the background surface as planar, we allow smooth spatial variation of the background, using a multiresolution surface model analogous to the so-called thin-membrane surface model (see Sec. 7.1). Employing this multiresolution model allows us to use the very fast estimation procedure described in the previous section. In addition to estimating the background range profile, the approach we describe simultaneously performs two other significant functions, namely target segmentation and target range profile estimation. To accomplish this, we use a second multiscale model, in this case one corresponding to a so-called thin-plate prior, to model the target range profile within the image, together with several important sets of hidden variables. The first two of these sets, namely, the sets of pixels that correspond to anomalous background and target range measurements, are analogous to those used in Ref. 12. The other hidden variable, namely, the set of pixels segmenting the target and background regions and explicitly identifying the line of pixels corresponding to ground attachment (i.e., where the target meets the ground), allows us to perform target segmentation and profiling.

A uniformly spaced raster scan is used to collect a range image consisting of n pixels. We denote the 2-D lattice consisting of these n pixels as $\mathcal{L}$. Let $y=\{y_i, i\in\mathcal{L}\}$ denote the measured ladar range values, whose background range profile is denoted as $x_{\mathrm{bg}}=\{x_{\mathrm{bg}_i}, i\in\mathcal{L}\}$ and whose target range profile is denoted as $x_{\mathrm{tg}}=\{x_{\mathrm{tg}_i}, i\in\mathcal{L}\}$. Note that it is convenient (with no approximation) to think of both the background and the target as being defined over the entire image domain. The background obviously is, although it is occluded by the target over the (unknown) region in which the target resides. Thus, while the algorithm we describe produces estimation of background and target fields over the entire image domain, the target field estimate is only meaningful over the estimated target region, while over that same region the background estimate can be thought of as an optimal interpolation of the occluded portion of the background. For notational simplicity, let y, x_{bg}, and x_{tg} denote lexicographically ordered vectors of the measurement, background, and target fields, respectively. We next present the several elements of our models for these quantities, including the specification of the hidden variables employed in our EM-based procedure.

3.1 *Prior Models for x_{bg} and x_{tg}*

The random fields x_{tg} and x_{bg} are modeled as independent, zero-mean Gaussian random fields to represent the smoothness constraints on the random fields. Specifically, x_{bg} is modeled as the finest-scale field in a multiscale thin-membrane model, and x_{tg} as the finest-scale field in a multiscale thin-plate model (see Sec. 7.1 and Ref. 10).

In principle, then, the prior model for x_{bg} has the form $p(x_{\mathrm{bg}})=G(x_{\mathrm{bg}};0,\Lambda_{\mathrm{bg}})$, where

$$G(x;m,\Lambda)=(2\pi)^{-n/2}|\Lambda|^{-1/2}\times\exp[-\tfrac{1}{2}(x-m)^T\Lambda^{-1}(x-m)],$$

and where $\Lambda_{x_{\mathrm{bg}}}$ is the $n\times n$ covariance matrix for the background range profile x_{bg}. Likewise, the prior model for the target range profile x_{tg} is $p(x_{\mathrm{tg}})=G(x_{\mathrm{tg}};0,\Lambda_{\mathrm{tg}})$, where $\Lambda_{x_{\mathrm{tg}}}$ is the $n\times n$ covariance matrix for the target range profile x_{tg}. While we never need to explicitly calculate $\Lambda_{x_{\mathrm{bg}}}$ and $\Lambda_{x_{\mathrm{tg}}}$ or their inverses, it is convenient for our discussion to introduce these explicit models. Also for ease of notation, the two range profiles are stacked together to form the range profile vector X, i.e., $X=[x_{\mathrm{bg}}\ \ x_{\mathrm{tg}}]^T$. We consider the two range profiles to be statistically independent of one another, so that the prior model for X becomes

$$p(X)=G(X;0,\Lambda_X),\quad \Lambda_X=\begin{bmatrix}\Lambda_{x_{\mathrm{bg}}} & 0\\ 0 & \Lambda_{x_{\mathrm{tg}}}\end{bmatrix}. \tag{11}$$

3.2 *Definition of the Hidden Variables*

The hidden variables we introduce partition the observed image domain $\mathcal{L}$ into five regions. A convenient way in which to express this partitioning is through the introduction of five sets of indicator functions (taking only values 0 and 1) over $\mathcal{L}$, one for each region type. The simplest of these are the two sets of variables $h_{a\mathrm{tg}}=\{h_{a\mathrm{tg}_i}, i\in\mathcal{L}\}$ and $h_{a\mathrm{bg}}=\{h_{a\mathrm{bg}_i}, i\in\mathcal{L}\}$, indicating the presence of range anomalies over the target and background portions of the image, respectively. The reason for separating these anomalies into two fields is simple: we want to interpolate the target (background) range profile for anomalous pixels within the target (background) region.

The next two sets of hidden variables, $h_{\mathrm{bg}}=\{h_{\mathrm{bg}_i}, i\in\mathcal{L}\}$ and $h_{\mathrm{tg}}=\{h_{\mathrm{tg}_i}, i\in\mathcal{L}\}$, designate those portions of the observed range measurements that are (unambiguously) measurements of the background or target, respectively. The final set of hidden variables indicates the line of pixels along which the target and background profiles intersect—i.e., the ground attachment line of the target, specified by $h_{\mathrm{ga}}=\{h_{\mathrm{ga}_i}, i\in\mathcal{L}\}$. Together, the five sets of hidden variables indicate mutually exclusive sets of pixels partitioning $\mathcal{L}$. That is, if we define

$$H_i=[h_{\mathrm{bg}_i}\ \ h_{\mathrm{tg}_i}\ \ h_{\mathrm{ga}_i}\ \ h_{a\mathrm{bg}_i}\ \ h_{a\mathrm{tg}_i}]^T\quad \forall i\in\mathcal{L}, \tag{12}$$

then each H_i takes on only the values e_k, $k=1,\ldots,5$, where e_k is a binary column vector with 1 in the k'th component and 0 everywhere else. As an example, $H_i=e_2$ indicates pixel i as belonging to image class 2, i.e., the set of pixels corresponding unambiguously to target range measurements.

3.3 *Measurement and Pseudomeasurement Models*

The observed ladar image data are noise- and anomaly-corrupted measurements of either the target or background. The explicit dependence of these measurements on X and H is given by the following measurement model:

$$y_i=\begin{bmatrix}h_{\mathrm{bg}_i}+\dfrac{h_{\mathrm{ga}_i}}{2} & h_{\mathrm{tg}_i}+\dfrac{h_{\mathrm{ga}_i}}{2}\end{bmatrix}\begin{bmatrix}x_{\mathrm{bg}_i}\\ x_{\mathrm{tg}_i}\end{bmatrix}+[h_{\mathrm{tg}_i}+h_{\mathrm{bg}_i}+h_{\mathrm{ga}_i}\ \ h_{a\mathrm{bg}_i}+h_{a\mathrm{tg}_i}]\begin{bmatrix}v_{\delta R_i}\\ v_{\Delta R_i}\end{bmatrix}, \tag{13}$$

where we have adopted some of the notation from Refs. 12,13 to describe the noise and anomalous measurement distributions. In particular, $v_{\delta R} \sim N(0, \delta R^2 I)$ represents the noise in a nonanomalous range measurement, with δR corresponding to the range resolution of the ladar system, while $v_{\Delta R} \sim \mathcal{U}(R_{\min} + \Delta R/2, [(\Delta R)^2/12] I)$ corresponds to the uncertainty in an anomalous measurement, modeled as uniformly distributed over the full unambiguous range interval of the ladar, i.e., $\Delta R = R_{\max} - R_{\min}$. [The notation $\mathcal{U}(\mu, \Lambda)$ represents a uniform random vector with mean μ and variance Λ.] For the results here, we use the same ladar system parameters and statistics as in Ref. 12, namely, $\delta R = 2$ range bins, $R_{\max} = 1524$ range bins, and $R_{\min} = 0$ range bins, so that $\Delta R = 1524$ range bins. Each range bin translates to a physical distance of 1.1 m, with lower bin numbers indicating closer range.

Intuitively, the first term on the right of Eq. (13) determines the type of each range measurement y_i, while the second term indicates the type of noise associated with each range measurement. It is instructive to examine several specific examples to understand how the hidden variables influence the model of the observed data. For example, suppose that pixel i corresponds to a background measurement (so that $h_{\mathrm{bg}_i} = 1$ and all other hidden variables are zero). Then

$$y_i = [1 \quad 0] \begin{bmatrix} x_{\mathrm{bg}_i} \\ x_{\mathrm{tg}_i} \end{bmatrix} + [1 \quad 0] \begin{bmatrix} v_{\delta R_i} \\ v_{\Delta R_i} \end{bmatrix} = x_{\mathrm{bg}_i} + v_{\delta R_i}.$$

This equation says that y_i is equal to the background range profile x_{bg_i} plus some measurement noise, which we model as additive Gaussian noise. Suppose, as another example, pixel i corresponds to a range anomaly lying within the target region ($h_{a\mathrm{tg}_i} = 1$). Equation (13) then gives

$$y_i = [0 \quad 0] \begin{bmatrix} x_{\mathrm{bg}_i} \\ x_{\mathrm{tg}_i} \end{bmatrix} + [0 \quad 1] \begin{bmatrix} v_{\delta R_i} \\ v_{\Delta R_i} \end{bmatrix} = v_{\Delta R_i},$$

which is consistent with our modeling of a range anomaly. As a final example, suppose pixel i corresponds to a ground attachment measurement ($h_{\mathrm{ga}_i} = 1$). This means that x_{bg_i} and x_{tg_i} are implicitly equal, and from Eq. (13) we have

$$y_i = \frac{x_{\mathrm{bg}_i} + x_{\mathrm{tg}_i}}{2} + v_{\delta R_i},$$

which binds the ground attachment measurement to both x_{bg_i} and x_{tg_i}.

Note that when $h_{\mathrm{ga}_i} = 1$, we would like to ensure that x_{bg_i} and x_{tg_i} are approximately equal (since the target attaches to the ground at these points). There are two equivalent ways in which to accomplish this. The first is as part of the prior model for x_{bg} and x_{tg}, providing a nonlinear coupling between these two fields. The second approach, which leads to a far simpler estimation algorithm, involves incorporating this constraint as an additional set of pseudomeasurements: For those pixels for which $h_{\mathrm{ga}_i} = 1$, we observe a measurement value of 0 for the difference $x_{\mathrm{bg}_i} - x_{\mathrm{tg}_i}$, where the additive noise on this measurement has very small variance ϵ^2. That is, at each pixel we have the additional measurement

$$0 = [h_{\mathrm{ga}_i} \quad -h_{\mathrm{ga}_i}] \begin{bmatrix} x_{\mathrm{bg}_i} \\ x_{\mathrm{tg}_i} \end{bmatrix} + h_{\mathrm{ga}_i} v_{\epsilon_i} \quad \forall i \in \mathcal{L}, \tag{14}$$

where $v_\epsilon \sim N(0, \epsilon^2 I)$. In this setup, the parameter ϵ controls the tightness of the coupling between the two MRFs. Note that if $h_{\mathrm{ga}_i} = 0$ this equation correctly indicates that we are not specifying any additional constraint on $x_{\mathrm{bg}_i} - x_{\mathrm{tg}_i}$. However, if $h_{\mathrm{ga}_i} = 1$, this pseudomeasurement introduces a strong coupling between the *estimates* of x_{bg_i} and x_{tg_i}.

Putting this all together, at each pixel i we now have a 2-vector of measurements

$$Y_i = \begin{bmatrix} y_i \\ 0 \end{bmatrix} = C(H_i) X_i + F(H_i) V_i \tag{15}$$

with $V_i^T = [v_{\delta R_i} \quad v_{\epsilon_i} \quad v_{\Delta R_i}]$, and

$$C(H_i) = \begin{bmatrix} h_{\mathrm{bg}_i} + \dfrac{h_{\mathrm{ga}_i}}{2} & h_{\mathrm{tg}_i} + \dfrac{h_{\mathrm{ga}_i}}{2} \\ h_{\mathrm{ga}_i} & -h_{\mathrm{ga}_i} \end{bmatrix},$$

$$F(H_i) = \begin{bmatrix} h_{\mathrm{bg}_i} + h_{\mathrm{tg}_i} + h_{\mathrm{ga}_i} & 0 & h_{a\mathrm{bg}_i} + h_{a\mathrm{tg}_i} \\ 0 & h_{\mathrm{ga}_i} & 0 \end{bmatrix}.$$

Equation (15), then, gives us the second needed part of our probabilistic model, namely,

$$p(Y|X,H) = \prod_{i \in \mathcal{L}} p(Y_i|X_i, H_i) = \prod_{i \in \mathcal{L}} H_i^T \begin{bmatrix} G(y_i; x_{\mathrm{bg}_i}, \delta R^2) \\ G(y_i; x_{\mathrm{tg}_i}, \delta R^2) \\ G\left(y_i; \dfrac{x_{\mathrm{bg}_i} + x_{\mathrm{tg}_i}}{2}, \delta R^2\right) G(x_{\mathrm{bg}_i}; x_{\mathrm{tg}_i}, \epsilon^2) \\ 1/\Delta R \\ 1/\Delta R \end{bmatrix}. \tag{16}$$

(We ignore the digitization that is found in typical ladar systems, and treat Y as a continuous random variable.) The role of H_i is simply to select one of the five densities indicated on the right-hand side of Eq. (16).

3.4 *Probabilistic Model for the Hidden Variables*

The remaining part of the probabilistic model is that for the hidden variables, H. In particular, as we see in the next section, the quantity that is needed for the EM procedure is

$p(H|X)$. In our model we make the simplifying assumption that H and X are statistically independent, so

$$p(H|X)=p(H). \tag{17}$$

Note that for the variables $h_{a\mathrm{tg}}$ and $h_{a\mathrm{bg}}$ modeling the presence of ladar anomalies, this independence assumption is completely justified. However, for the others it represents a clear simplification. For example the condition $h_{\mathrm{ga}_i}=1$ (ground attachment) implies $x_{\mathrm{bg}_i}\approx x_{\mathrm{tg}_i}$, indicating a very strong dependence between H and X. However, as we have already indicated, we can capture this dependence through the pseudomeasurement (14) rather than through the prior. Note that there also is a dependence between $[h_{\mathrm{bg}_i},h_{\mathrm{tg}_i}]$ (indicating observation of background or target in pixel i) and $[x_{\mathrm{bg}_i},x_{\mathrm{tg}_i}]$: In particular, if the target is present in pixel i ($h_{\mathrm{tg}_i}=1$), then obviously the target occludes the background, implying that $x_{\mathrm{tg}_i}<x_{\mathrm{bg}_i}$. While we could introduce another pseudomeasurement to capture this relationship as well, for simplicity we have not done so, and thus our model indeed neglects this dependence. In principle this implies that the resulting estimates we produce could have target and background estimates that are inconsistent ($\hat{h}_{\mathrm{tg}_i}=1$ but $\hat{x}_{\mathrm{tg}_i}>\hat{x}_{\mathrm{bg}_i}$), but, as we illustrate using real ladar data, employing this simplification does not in fact introduce such inconsistencies.

We use a Gibbs prior to capture spatial dependences in the hidden variables:

$$p(H)=\Gamma^{-1}\exp[-\zeta U(H)],$$

where Γ is the normalization constant. Defining the energy function $U(H)$ requires specifying the neighborhood structure on the 2-D lattice $\mathcal{L}$. For our purposes it is sufficient to use first-order neighborhoods,[14] implying that cliques consist only of individual pixels and pairs of pixels that are either horizontal or vertical nearest neighbors. This implies that $U(H)$ can be written as

$$U(H)=\sum_{i\in\mathcal{L}}\left[V_{\mathrm{clq}}(H_i)+\frac{1}{2}\sum_{l\in\mathcal{N}_i}H_i^T\mathcal{W}_{il}H_l\right], \tag{18}$$

where $\mathcal{W}_{il}$ is a 5×5 matrix whose (s,t)th component is $V_{\mathrm{clq}}(H_i=e_s,H_l=e_t)$. The elements within each matrix $\mathcal{W}_{il}$ are chosen to encourage (discourage) certain types of interactions among the hidden variables by decreasing (increasing) the energy function $U(H)$. The specific choices for these matrices that we have used are as follows:

$$\mathcal{W}_{il}=\begin{bmatrix}-0.5 & 0.5 & 0 & 0 & 0\\ 0.5 & -0.5 & 0.5 & 0 & 0\\ 0 & 0 & 0.5 & 0 & 0\\ 0.5 & -0.5 & 0 & 0 & 0\\ -0.5 & 0.5 & 0 & 0 & 0\end{bmatrix}, \tag{19a}$$

when pixel l is above pixel i,

$$\mathcal{W}_{il}=\begin{bmatrix}-0.5 & 0.5 & 0 & 0 & 0\\ 0.5 & -0.5 & 0 & 0 & 0\\ 0 & 0.5 & 0.5 & 0 & 0\\ 0.5 & -0.5 & 0 & 0 & 0\\ -0.5 & 0.5 & 0 & 0 & 0\end{bmatrix}, \tag{19b}$$

when pixel l is below pixel i, and

$$\mathcal{W}_{il}=\begin{bmatrix}-0.5 & 0.5 & 0 & 0 & 0\\ 0.5 & -0.5 & 0 & 0 & 0\\ 0 & 0 & -0.5 & 0 & 0\\ 0.5 & -0.5 & 0 & 0 & 0\\ -0.5 & 0.5 & 0 & 0 & 0\end{bmatrix}, \tag{19c}$$

when pixel l is to the left or right of pixel i. In each of these matrices, the entries in the first row are chosen to encourage continuity of the background region (i.e., if a majority of the neighbors of pixel i are labeled as background, then pixel i is more likely to be background as well). The entries in the second row are chosen to encourage continuity of the target region and, in the case of Eq. (19a), also to discourage target pixels from appearing below the ground attachment pixels. The entries in the third row are chosen to encourage the ground attachment pixels to form a thin line below the target pixels. The entries in the fourth and fifth rows are chosen to help differentiate between an anomalous pixel located in the background region and an anomalous pixel located in the target region.

In order to incorporate $p(H_i)$, the single-pixel stationary prior probability density functions (pdfs) for H_i, into our Gibbs modeling of $p(H)$, we assign the singleton clique potential $V_{\mathrm{clq}}(\cdot)$ to be

$$p(H_i)=Z_i^{-1}\exp[-\zeta V_{\mathrm{clq}}(H_i)],$$
$$Z_i=\sum_{\tilde{H}_i}\exp[-\zeta V_{\mathrm{clq}}(\tilde{H}_i)]. \tag{20}$$

Based on the work in Ref. 12, the probability of the presence of range measurement anomalies is 0.05, and we distribute this prior probability equally between $p(H_i=e_4)$ and $p(H_i=e_5)$, the probabilities that a pixel measurement is an anomaly situated inside the background region and that it is an anomaly situated inside the target region, respectively. The prior probability that a particular pixel in the image is a ground attachment pixel is chosen to be 0.05. The remaining prior probability for H_i (0.9) is distributed equally between $p(H_i=e_1)$ and $p(H_i=e_2)$, the probabilities that a pixel corresponds to the background region and to the target region, respectively.* To summarize:

*Note that the values we have chosen for $p(e_1)$, $p(e_2)$, and $p(e_3)$ reflect relative values of pixel categories for ladar images in which we expect targets to fill roughly half of the image area. If targets take up significantly less image area, one can capture this by biasing the prior in favor of e_1. However, as the examples in Sec. 5 indicate, the results are not particularly sensitive to these values.

$p(H_i)=[0.45\ \ 0.45\ \ 0.05\ \ 0.025\ \ 0.025]H_i$.

Our full prior model for H then is

$$p(H)=\Gamma^{-1}\prod_{i\in\mathcal{L}}\left[p(H_i)Z_i\exp\left(-\frac{\zeta}{2}\sum_{l\in\mathcal{N}_i}H_i^T\mathcal{W}_{il}H_l\right)\right]. \quad (21)$$

4 The Profile Estimation and Segmentation Algorithm

The problem to be solved is the estimation of X given Y based on the prior model $p(X)$ given by Eq. (11), the measurement model $p(Y|X,H)$ given by Eq. (16), and the prior model for the hidden variables $p(H|X)=p(H)$ given by Eq. (21). In this section we describe the two steps in the EM procedure required to compute a local maximum of $\log p(X|Y)$, which we take as our estimate. This procedure also produces an estimate of H that is of direct use in segmenting out the target, rejecting and compensating for anomalies, etc.

4.1 *The E Step*

The quantity to be calculated in the E step is

$$Q(X|X^{[k]})=\langle\log p(Y,H|X)|Y,X^{[k]}\rangle,$$

which is a conditional expectation over H given Y and the current estimate $X^{[k]}$ of the range profile. Using Bayes's rule, Q can be rewritten as

$$Q(X|X^{[k]})=\langle\log p(Y|X,H)|Y,X^{[k]}\rangle+\langle\log p(H)|Y,X^{[k]}\rangle. \quad (22)$$

Since the M step is a maximization of $Q(X|X^{[k]})$ over X, we can discard the second term in Eq. (22), since it does not depend on X. Using the same notation for the result, we have that

$$Q(X|X^{[k]})=\sum_{i\in\mathcal{L}}\langle H_i|Y,X^{[k]}\rangle^T \times\begin{bmatrix}\log G(y_i;x_{\mathrm{bg}_i},\delta R^2)\\ \log G(y_i;x_{\mathrm{tg}_i},\delta R^2)\\ \log G\left(y_i;\frac{x_{\mathrm{bg}_i}+x_{\mathrm{tg}_i}}{2},\delta R^2\right)+\log G(x_{\mathrm{bg}_i};x_{\mathrm{tg}_i},\epsilon^2)\\ -\log\Delta R\\ -\log\Delta R\end{bmatrix}. \quad (23)$$

[Here we have also used the fact that each H_i is an indicator vector in order to take the log of $p(Y|X,H)$ in Eq. (16).] Thus the core of the E step is the computation of $\langle H_i|Y,X^{[k]}\rangle$. Using the formula for expectations, Bayes's rule, and Eqs. (16), (17), and (21), we find that

$$\langle H_i|Y,X^{[k]}\rangle=\Upsilon^{-1}\sum_H H_i\prod_{i\in\mathcal{L}}\left[p(Y_i|X_i^{[k]},H_i)p(H_i)\times\exp\left(-\frac{\zeta}{2}\sum_{l\in\mathcal{N}_i}H_i^T\mathcal{W}_{il}H_l\right)\right], \quad (24)$$

where

$$\Upsilon=\sum_{\tilde{H}}\prod_{i\in\mathcal{L}}\left[p(Y_i|X_i^{[k]},\tilde{H}_i)p(\tilde{H}_i)\times\exp\left(-\frac{\zeta}{2}\sum_{l\in\mathcal{N}_i}\tilde{H}_i^T\mathcal{W}_{il}\tilde{H}_l\right)\right].$$

From this, it is obvious that a large number of configurations of H need to be evaluated in order to compute Eq. (24). This makes its precise calculation both complex and computationally impractical. By applying MFT, this difficulty is overcome, and an approximation to $\langle H_i|Y,X^{[k]}\rangle$ is obtained by evaluating the following:

$$\langle H_i|Y,X^{[k]}\rangle\approx\Upsilon_i^{\mathrm{mf}\,-1}\sum_{H_i}\left[H_i p(Y_i|X_i^{[k]},H_i)p(H_i)\times\exp\left(-\zeta\sum_{l\in\mathcal{N}_i}H_i^T\mathcal{W}_{il}\langle H_l|Y,X^{[k]}\rangle\right)\right], \quad (25)$$

where

$$\Upsilon_i^{\mathrm{mf}}=\sum_{\tilde{H}_i}\left[p(Y_i|X_I^{[k]},\tilde{c}_i)p(\tilde{H}_i)\times\exp\left(-\zeta\sum_{l\in\mathcal{N}_i}\tilde{H}_i^T\mathcal{W}_{il}\langle H_l|Y,X^{[k]}\rangle\right)\right].$$

As Eq. (25) indicates, to find the mean field $\langle H|Y,X^{[k]}\rangle$ at i, one needs the mean field $\langle H|Y,X^{[k]}\rangle$ at the neighbors of i. Thus, Eq. (25) for all $i\in\mathcal{L}$ represents a set of simultaneous algebraic equations, which can be solved by embedding it within the E step of the overall EM procedure. In principle we can perform multiple Jacobi iterations to solve these equations within each E step. However, obtaining an accurate estimate of $\langle H|Y,X^{[k]}\rangle$ in each E step may not be necessary, since $\langle H|Y,X^{[k]}\rangle$ is merely used to guide the estimate of X in each M step. Empirically, we have found that it is sufficient and, moreover, computationally efficient to approximate Eq. (25) by performing only *one* Jacobi iteration in each E step. The $\langle H|Y,X^{[k-1]}\rangle$ calculated from the previous EM cycle serves as the initial guess for that single Jacobi's iteration. This approximation scheme has the natural notion of refining the estimate of $\langle H|Y,X^{[k]}\rangle$ as the number of EM cycles increases (because successive estimates of $\langle H|Y,X^{[k]}\rangle$ used for the Jacobi iterate in the E step are more accurate). Also, since the calculation of $\langle H|Y,X^{[k]}\rangle$ can be decomposed into local computations, its solution can be implemented in a parallel fashion to yield a fast and simple approximation.

4.2 *The M Step*

The M step calculates the $X^{[k+1]}$ in Eq. (3). It is straightforward to check that the quantity to be maximized in Eq. (3) is quadratic in X and that the maximum is given by

$$(\Lambda_X^{-1}+\mathcal{M})\,X^{[k+1]}=\mathcal{P}y, \quad (26)$$

where Λ_X is the block-diagonal prior covariance matrix in Eq. (11), and $\mathcal{M}$ and $\mathcal{P}$ are given by

$$\mathcal{M}=\begin{bmatrix} D_1 & D_3 \\ D_3 & D_2 \end{bmatrix}, \quad \mathcal{P}=\begin{bmatrix} D_4 \\ D_5 \end{bmatrix}, \tag{27}$$

where $D_1, \ldots, D_5$ are diagonal matrices given by

$$(D_1)_{ii}=\frac{\langle h_{\mathrm{bg}_i}|Y,X^{[k]}\rangle+\langle h_{\mathrm{ga}_i}|Y,X^{[k]}\rangle/4}{\delta R^2} + \frac{\langle h_{\mathrm{ga}_i}|Y,X^{[k]}\rangle}{\epsilon^2} \quad \forall i\in\mathcal{L}, \tag{28a}$$

$$(D_2)_{ii}=\frac{\langle h_{\mathrm{tg}_i}|Y,X^{[k]}\rangle+\langle h_{\mathrm{ga}_i}|Y,X^{[k]}\rangle/4}{\delta R^2} + \frac{\langle h_{\mathrm{ga}_i}|Y,X^{[k]}\rangle}{\epsilon^2} \quad \forall i\in\mathcal{L}, \tag{28b}$$

$$(D_3)_{ii}=\langle h_{\mathrm{ga}_i}|Y,X^{[k]}\rangle\left(\frac{1}{4\,\delta R^2}-\frac{1}{\epsilon^2}\right) \quad \forall i\in\mathcal{L}, \tag{28c}$$

$$(D_4)_{ii}=\frac{\langle h_{\mathrm{bg}_i}|Y,X^{[k]}\rangle+\langle h_{\mathrm{ga}_i}|Y,X^{[k]}\rangle/2}{\delta R^2} \quad \forall i\in\mathcal{L}, \tag{28d}$$

$$(D_5)_{ii}=\frac{\langle h_{\mathrm{tg}_i}|Y,X^{[k]}\rangle+\langle h_{\mathrm{ga}_i}|Y,X^{[k]}\rangle/2}{\delta R^2} \quad \forall i\in\mathcal{L}. \tag{28e}$$

The quantity $X^{[k+1]}$ can be viewed as the MAP estimate of X based on the prior model specified by $p(X)$ and a *measurement model* $p(\text{data}|X)$ specified implicitly by $Q(X|X^{[k]})$. [More precisely, $\log p(\text{data}|X)=Q(X|X^{[k]})$ + terms not depending on X.] The prior on X is reflected in the block-diagonal matrix Λ_X^{-1} in Eq. (26), while $\mathcal{M}$ and $\mathcal{P}$ reflect the measurement model. Specifically, by collecting measurements y in Eq. (13) and pseudomeasurements (all with value 0) in Eq. (14) into vectors, we can write

$$\begin{pmatrix} y \\ 0 \end{pmatrix}=C(H)X+F(H)V, \tag{29}$$

where V is a vector containing all of the noise sources $v_{\delta R_i}$, $v_{\Delta R_i}$, v_{ϵ_i} at all pixels with covariance matrix R. What the E step in essence does is to replace $C(H)$ and $F(H)$ by expected values (or approximations of these using MFT), yielding an effective measurement model of the form

$$\begin{pmatrix} y \\ 0 \end{pmatrix}=C_{[k]}X+F_{[k]}V. \tag{30}$$

Using this notation,

$$\mathcal{M}=C_{[k]}^T[F_{[k]}RF_{[k]}^T]^{-1}C_{[k]}, \tag{31a}$$

$$\mathcal{P}y=C_{[k]}^T[F_{[k]}RF_{[k]}^T]^{-1}\begin{pmatrix} y \\ 0 \end{pmatrix}, \tag{31b}$$

so that Eq. (26) is precisely the set of linear equations for the optimal estimate of X based on the measurement model (30). Moreover, the corresponding estimation error covariance is given by

$$\Lambda_e=(\Lambda_X^{-1}+\mathcal{M})^{-1}. \tag{32}$$

While the prior models for target and background are independent (as reflected in the block-diagonal structure of Λ_X), their estimates are coupled. For example, the E step replaces the various h's in Eq. (13) by their expectations. Thus while each pixel measurement is either target, background, or anomaly, the model (30) will in general characterize each measurement as a weighted average of these possibilities, using the expected values of the various h's as calculated in the E step. However, this coupling occurs in a pixel-by-pixel fashion, as reflected in the fact that $D_1, \ldots, D_5$ are diagonal. Indeed, by reordering the elements of X and y, we can view our problem as that of estimating a vector random field (consisting of the 2-D vector of target and background at each pixel) based on pixelwise measurements that are independent from pixel to pixel.

Finally, it is important to realize that despite all of this structure, Eq. (26) represents a huge set of equations to be solved ($2n^2$ for an $n\times n$ image), and the error covariance (32) is of the same very large size. However, by using the multiresolution thin-membrane and -plate priors mentioned previously and summarized in Sec. 7.1, the solution of Eq. (26) *and* the computation of the diagonal elements of Λ_e (i.e., the pixelwise error covariances) can all be accomplished with total per-pixel complexity that is independent of image domain size. This makes the EM procedure not only feasible but, in fact, quite efficient.

5 Experimental Results

In this section, we present two examples to illustrate the performance of the algorithm we have just described. The ladar range images we use are provided by MIT Lincoln Laboratory. These range images are formed by an airborne coherent CO_2-ladar system operating in pulsed-imager mode. The images are 43×126-pixel gray-scale range images.

In our first example, the range image is of a C50 tank positioned against a sloping featureless terrain as shown in Fig. 5(a). This range image was taken under conditions of high carrier-to-noise ratio (CNR), so that it has relatively little noise, with almost no anomalies present. We use this image as the ground truth to test the performance our algorithm. To simulate realistic range images that resemble the ones obtained under battlefield conditions (i.e., low-CNR conditions), a local-range-accuracy Gaussian noise with a standard deviation of $\delta R=2$ bins is artificially added to each pixel in the image. In addition, a Bernoulli process is used to randomly select 5% of the pixels in the image as anomalous pixels. For each such anomalous pixel, a random range value, chosen from a uniform distribution over

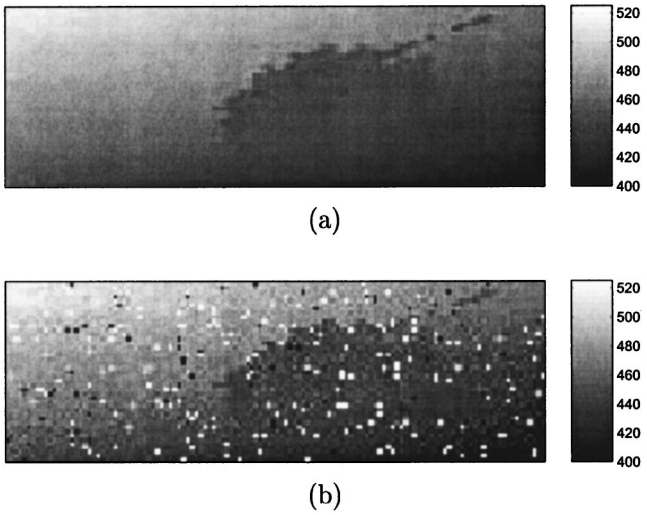


Fig. 5 Ladar range image of tank: (a) ground truth image, (b) ''observed'' image.

the range uncertainty interval $\Delta R = 1524$ bins, is assigned. The resulting ''observed'' ladar range image, shown in Fig. 5(b), very much resembles the ones obtained on the battlefield. The parameters used to obtain this simulated ladar image came from Ref. 12.

One of the goals of our algorithm is to estimate the background and the target range profiles, i.e., the range profiles of the sloping featureless terrain and the tank silhouette, based on the ''observed'' image. Figure 6 shows our estimate of the range profiles and their associated error statistics. The error statistics provide us with a useful measure of the uncertainty of our estimate of the range profiles. From top to bottom, this figure shows the estimate of the background range profile, its estimation error variances, the estimate of the target range profile, and its estimation error variances. The estimate of the background range profile in Fig. 6(a) demonstrates good anomaly suppression capabilities. The estimated profile is smooth and free of the obvious blocky artifacts sometimes associated with the use of a multiscale estimator.[5,10] This is directly attributable to the use of the overlapping tree surface models developed in Ref. 10 and summarized in Sec. 7.1. More interestingly, the estimate of the background is obtained over the *entire* field of view, including parts of the background range profile that are obscured from the field of view by the tank and the range anomalies. Even without background range measurements at these obscured locations, the multiscale estimator still does capture the coarse features of the smooth background profile there. The lack of background range measurements at the locations of the tank is consistent with the information conveyed in Fig. 6(b), which shows high estimation uncertainties around the region obscured by the tank.[†] Figure 6(c) shows the estimates of the tank's range profile. Note that the estimate of the target range profile is

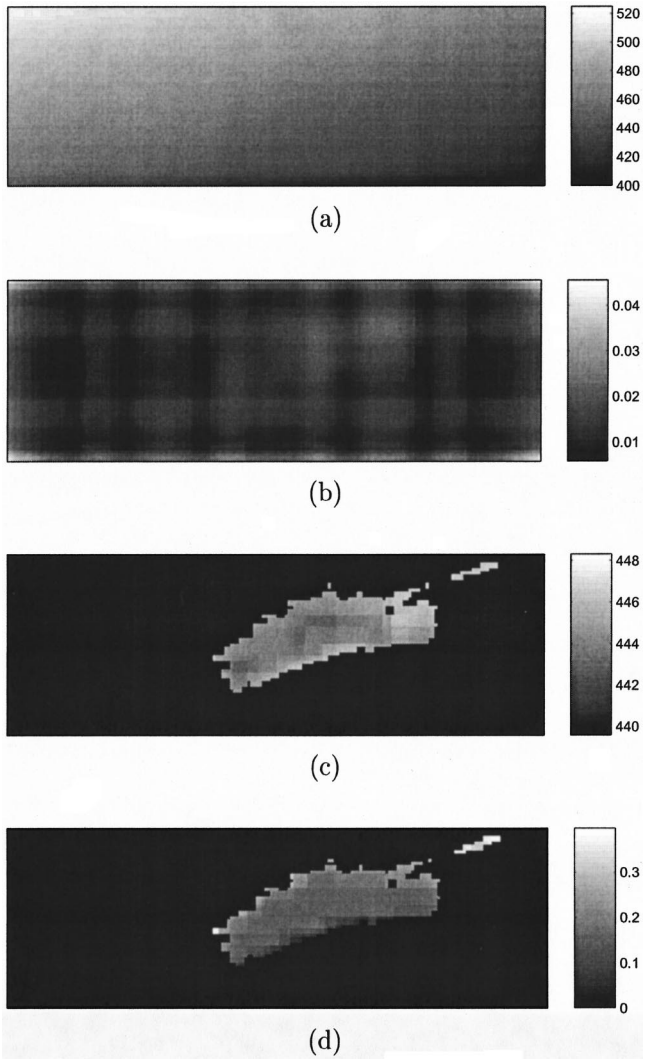


Fig. 6 Multiscale tank ladar range profile estimates: (a) estimate of background range profile, (b) estimation error variances of background range profile, (c) estimate of target range profile, (d) estimation error variances of target range profile.

relatively flat, which is consistent with our thin-plate modeling. Figure 6(d) shows the estimation error variances of the target range profile at locations where the tank appears.

The other goal of our algorithm is to provide a segmentation map of the ladar range image. By thresholding H, we obtain the segmentation map shown in Fig. 7. Figure 7(a) shows the locations of the target and the ground attachment pixels, while Fig. 7(b) shows the locations of the target and background range anomaly pixels. It is interesting to note that as designed, the ground attachment class constitutes a thin line corresponding to the line along which the target attaches to the background. Taken together, the target pixels and the target range anomaly pixels form the location map of the target of interest, which in this case is the tank. More importantly, the location map of the tank with its accompanying range profile can yield 3-D information about the target and thus plays an important role in ATR.

Using the estimate of the two range profiles and the information provided by the segmentation map, we can reconstruct an anomaly-free and denoised image from the

[†]Note that in contrast to the estimates in Fig. 6(a), the error variances in Fig. 6(b) do display some of the blocky artifacts associated with multiscale estimates, although even these are greatly reduced by using overlapped models. In fact, as discussed in Ref. 10, they can be reduced further by increasing the so-called degree of overlap, which however also increases the computational complexity.

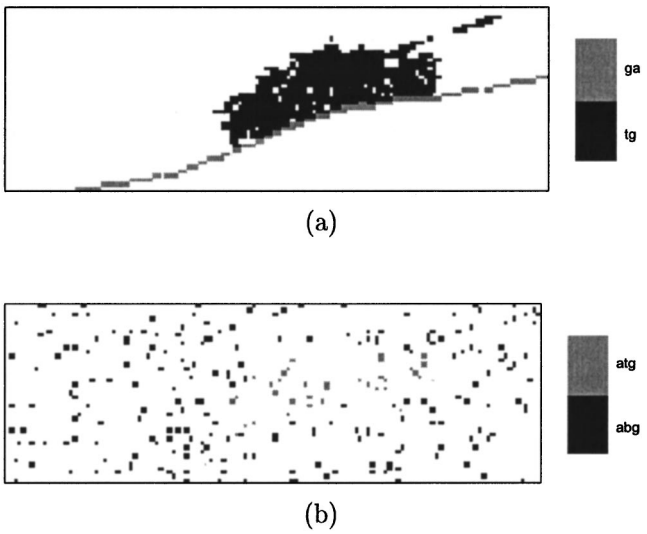


Fig. 7 Segmentation results of tank ladar range image: (a) target (tg) and ground attachment (ga) locations, (b) range anomaly locations inside (*a*tg) and outside (*a*bg) the target region.

raw noisy data in Fig. 5(b). This reconstructed ladar range image of the tank is shown in Fig. 8. The reconstructed image is remarkably similar to the ground-truth ladar range image in Fig. 5(a).

The results shown in Figs. 6–8 were obtained with $\zeta = 1$ and $\epsilon = 0.1$, after running our algorithm for 14 EM iterations. We chose 14 iterations as the stopping point of our iterative scheme based on the results shown in Fig. 9. The graphs in the figure show the one-norm distance measure of the estimates x_{bg} and x_{tg} between two successive iterations. After 14 iterations, the estimates appear to have stabilized, and it is at this point that we decided that the algorithm had converged. Note, however, that the knees in both of these curves occur after only two or three iterations, suggesting that we can obtain nearly as good reconstructions much more quickly if total computational load is a concern.

As another demonstration of the performance of our algorithm, we use another ladar range image, where the target of interest is a truck. Figures 10–14 convey the same set of information as their counterparts in Figs. 5–9.

6 Conclusion

The ladar target segmentation and range profiling algorithm that has been described in this paper has two components: estimation of the segmentation map and range profile estimation. Our contribution has been to combine them in an

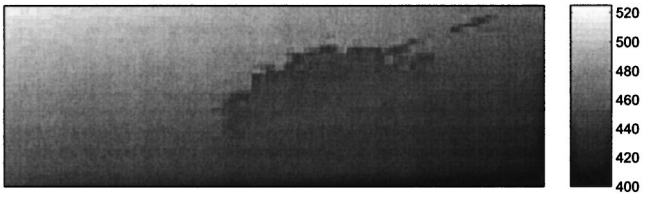


Fig. 8 Anomaly-free and denoised reconstruction of the tank ladar range image.

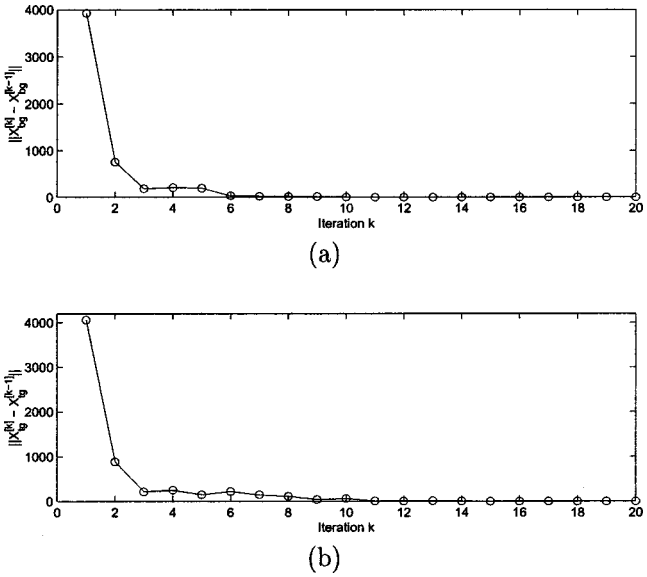


Fig. 9 Convergence results of tank ladar range profile estimates: (a) change in the estimate of background range profile with each iteration, (b) change in the estimate of target range profile with each iteration.

iterative fashion that yields a powerful new estimation procedure for ladar range image enhancement. Building on the previous work done by Green and Shapiro on planar range profiling, we derived a general algorithm for background range profile estimation that is able to deal with more complicated background. This extension uses a multiscale thin-membrane model for the background range profile; it is much less restrictive than the surface planar model used by Green and Shapiro and is more applicable for estimating the range profiles of natural scenes. As an added feature, we also developed our methodology using a thin-plate prior for target segmentation and estimation and hence provide 3-D fine range information about the target required, for example, in shape-based ATR. The multiscale methodology provides a computationally efficient and statistically opti-

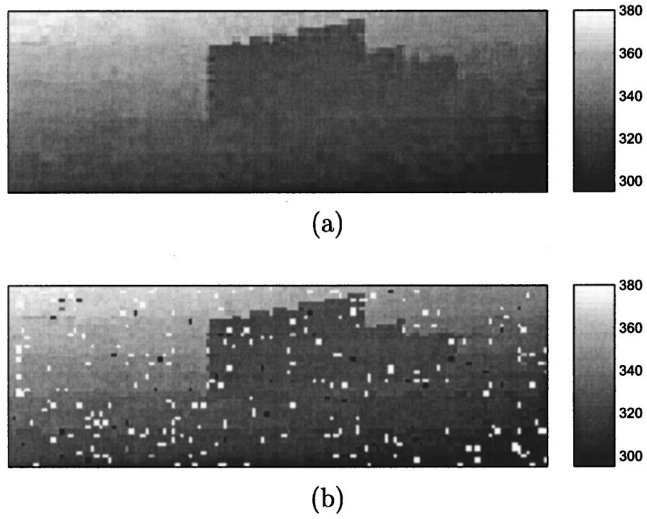


Fig. 10 Ladar range image of a truck: (a) ground truth image, (b) ''observed'' image.

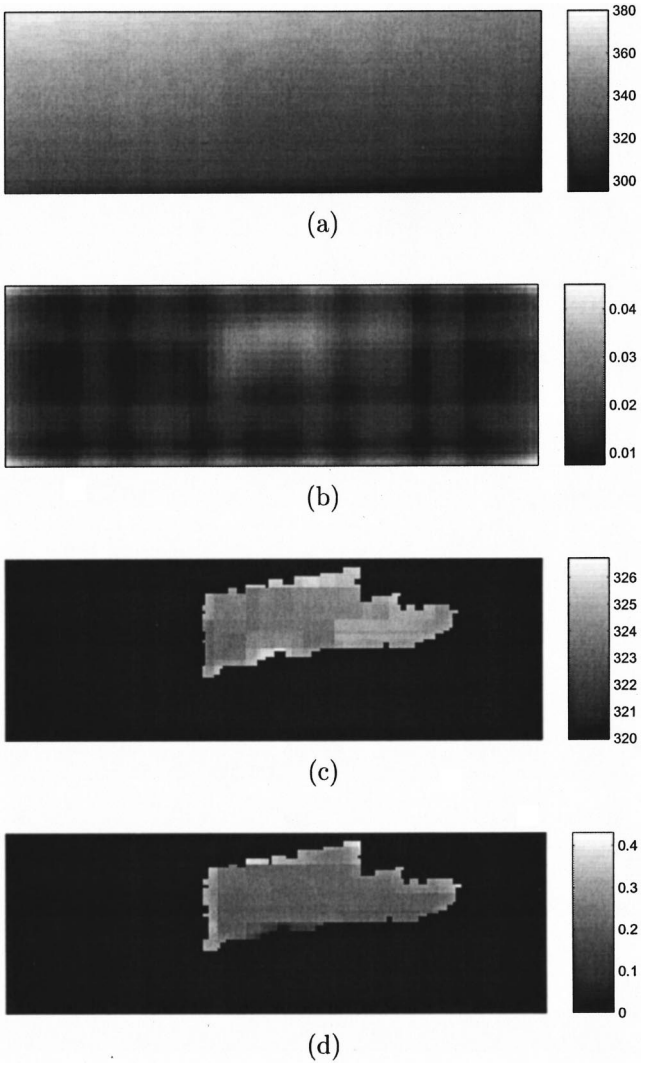


Fig. 11 Multiscale truck ladar range profile estimates: (a) estimate of background range profile, (b) estimation error variances of background range profile, (c) estimate of target range profile, (d) estimation error variances of target range profile.

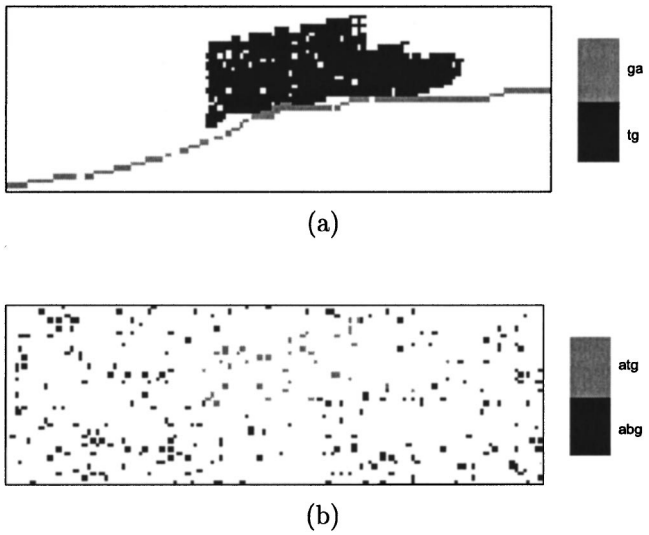


Fig. 12 Segmentation results of truck ladar range image: (a) target (tg) and ground attachment (ga) locations, (b) range anomaly locations inside (*a*tg) and outside (*a*bg) the target region.

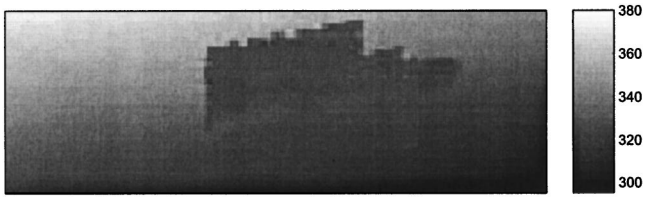


Fig. 13 Anomaly-free reconstruction of the truck ladar range image.

mal estimate of the two range profiles. The availability of accompanying estimation error variances provided by the multiscale framework is an attractive feature that can be used to assess the goodness of our estimates. In addition to profile estimation, our algorithm, aided by an approximation scheme derived from MFT, simultaneously provides an estimate of the segmentation map which can also be used for ATR.

More broadly, the major contributions of this paper lie in the formulation of a mathematically sound methodology based on the EM procedure that can be used to harness the powers of both multiscale methods and MFT in solving incomplete-data problems, extending well beyond the ladar range imaging application described here. For examples ranging from texture segmentation to simultaneous tissue classification and gain correction in MRI, we refer the reader to Refs. 3 and 4.

7 Appendix. Multiscale Thin-Plate and Thin-Membrane Models in the Overlapping Multiscale Tree Framework

Here we provide a brief sketch of the multiscale thin-plate and -membrane models introduced in Ref. 10. As discussed in Sec. 2.3, multiscale models are defined on trees, with a state $\mathcal{X}(s)$ at each node on the tree and with measurements

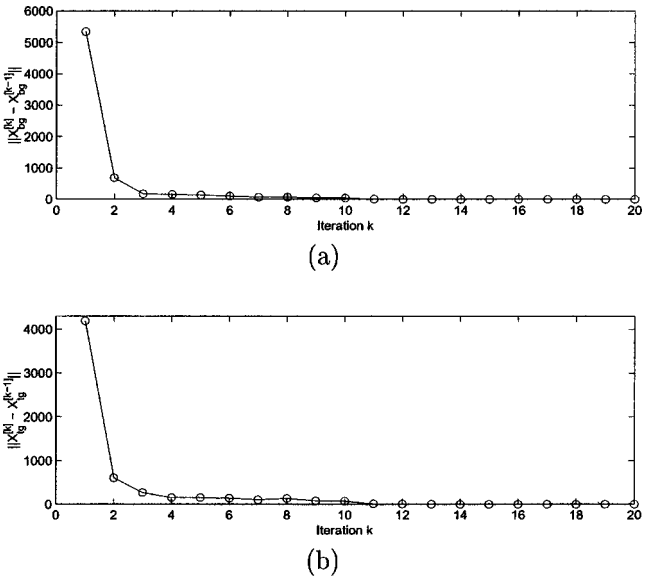


Fig. 14 Convergence results on truck radar range profile estimates: (a) change in the estimate of background range profile with each iteration, (b) change in the estimate of target range profile with each iteration.

$\mathcal{Y}(s)$ potentially defined at every node. In modeling target and background fields as surfaces, the representation of either surface resides at the finest scale of the tree, as do the real ladar range measurements and ground attachment pseudomeasurements described in Sec. 3. Thus the role of the ''hidden'' states and some additional pseudomeasurements at coarser scales in the tree is to shape the fine-scale surface statistics to approximate those corresponding to thin-plate and thin-membrane statistical models.

7.1 *Multiscale Thin-Plate and Thin-Membrane Models*

Thin-plate and -membrane models, which are frequently used in computer vision to assert smoothness constraints, correspond to particular variational problems. As discussed in Refs. 5, 10, and 15, these models can also be interpreted statistically, and they correspond to prior models for surfaces with self-similar statistics and, in particular with power spectral densities of the form $1/f^{\alpha}$ with α depending on whether a thin plate ($\alpha=4$) or membrane ($\alpha=2$) model is used. As discussed first in Ref. 5, such self-similar statistics can also be captured in multiscale models, albeit with slight differences in the prior, but with all of the computational advantages of multiscale models.

The basic thin-plate or -membrane multiscale model introduced in Ref. 10 is given by the following:

$$\begin{bmatrix} z \\ p \\ q \\ zp \end{bmatrix}(s)=\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} z \\ p \\ q \\ zp \end{bmatrix}(s\bar{\gamma}) + \begin{bmatrix} B_s 2^{-m(s)/2} & 0 & 0 \\ 0 & B_g 2^{-m(s)/2} & 0 \\ 0 & 0 & B_g 2^{-m(s)/2} \\ 0 & 0 & 0 \end{bmatrix} \times \mathcal{W}(s), \quad (33)$$

where the state at each node s consists of the height z of the surface at the scale and resolution corresponding to s, the x and y components (p and q) of the gradient of the surface (at the scale and location of nodes), and the surface height at the parent to node s at the next coarser scale [i.e., $zp(s)=z(s\bar{\gamma})$].

The thin-membrane model corresponds to Eq. (33) with $B_s \gg B_g$ (to diminish the effects of the states p and q and decrease the contribution of the thin-plate penalty), to capture the fact that this $1/f^2$-like model for z corresponds to what is in essence a self-similar random walk in scale. The multiscale counterpart of the thin-plate prior corresponds to $B_g \gg B_s$ in Eq. (33). However, as is well known, the components of the surface gradient must satisfy the so-called integrability constraints, whose strict enforcement increases the complexity of surface estimation considerably. An alternative is to relax this hard constraint and replace it by a pseudomeasurement at every node on the tree. Referring to Fig. 15, the pseudomeasurements at the

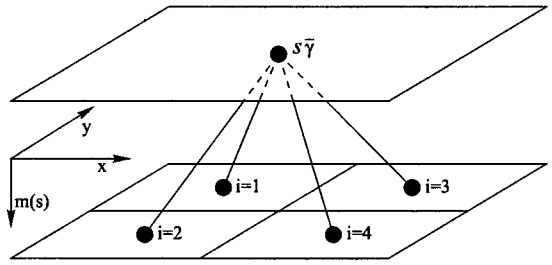


Fig. 15 Numbering scheme of nodes in the thin-plate and thin-membrane models.

four nodes s_i for $1 \leq i \leq 4$ with the same parent $s\bar{\gamma}$ are given by

$$0=\begin{bmatrix} -2^{m(s_i)-M} & a_i & b_i & 2^{m(s_i)-M} \end{bmatrix}\begin{bmatrix} z \\ p \\ q \\ zp \end{bmatrix}(s_i)+v(s_i), \quad (34)$$

where the values of a_i and b_i for $i \in \{1,2,3,4\}$ are given by

$$a_1=-1, \quad a_2=-1, \quad a_3=1, \quad a_4=1,$$

$$b_1=1, \quad b_2=-1, \quad b_3=1, \quad b_4=-1.$$

Here $2^{m(s_i)-M}[z(s_i)-zp(s_i)]$ represents a first-difference approximation to directional derivative of the surface in a direction (NW, SW, NE, SE) that depends on the particular value of i, while $a_i p+b_i q$ represents the approximation to the same directional derivative implied by the values of p and q at the resolution and location corresponding to node s_i. Equation (34) then simply states that the difference $v(s_i)$ in these two approximations to the directional derivative is a zero-mean white noise with variance 0.1.

The general model (33) allows both B_s and B_g to be nonzero, allowing us freedom in adjusting the relative importance of the thin-membrane and -plate smoothness penalties. In particular, to model the background field in our ladar application, we used values $B_s=30$ and $B_g=300$, while for the target field we set $B_s=4$ and $B_g=0.1$.

7.2 *Overlapping Multiscale Trees*

The application of this multiscale model based on the conventional multiscale quadtree and the associated estimation algorithm[1,2] results in a visually distracting blockiness in the estimates. The characteristic blocky artifact is one of the limitations often associated with the multiscale approach. In random fields with significant levels of regularity or smoothness, high correlations exist between spatially close neighbors. When using a standard multiscale quadtree representation such as that suggested by Fig. 3, there exist nodes on the tree (e.g., the two shaded squares in Fig. 3) that are spatially quite close but are much more distant as measured by the distance along the tree from one of these nodes to the other. As a result, prohibitively high-dimensional states are required to capture the high correlation between such nodes, implying that lower-dimensional

multiresolution models (and the estimates such models generate) may not adequately capture the correlation and smoothness of the field across such boundaries.

In Ref. 11 a method to overcome this problem was developed using the concept of overlapping trees. Instead of adhering to the convention of the multiscale quadtree framework, where distinct nodes at a given scale of the tree are assigned disjoint portions of the image domain, the overlapping multiscale quadtree allows distinct nodes at a given scale of the tree to correspond to *overlapping* portions of the image domain. As a result, any pixel that would fall near a major tree boundary in a standard quadtree representation actually has several replicas in an overlapped model, essentially placing versions of that pixel on both sides of the tree boundary in order to capture spatial correlations more easily.

There are several implications of this idea. The first is that at the finest scale of the tree we do indeed have several tree nodes corresponding to a single image pixel. As discussed in Ref. 11, there is then a need to define two operators. The first, namely *estimate projection*, maps the estimates of the states at this set of tree nodes into the estimate of the single pixel to which they all correspond, by taking a (weighted) average. The second, namely *measurement projection*, maps each measurement of an image pixel into a set of independent measurements of each of the tree nodes to which it corresponds. This is accomplished simply by replicating the observed measurement value at each of the tree nodes but modeling each of these tree measurements as being proportionately noisier, so that the total information content of this set of tree measurements is the same as the actual pixel measurement. For example if a pixel x corresponds to two tree nodes $x(s_1)$ and $x(s_2)$, then the measurement $y=x+v$, where v has variance σ^2, is mapped into two measurements $y_1=x(s_1)+v_1$ and $y_2=x(s_2)+v_2$, where v_1, v_2 are modeled as independent with variance $2\sigma^2$ and where the measurement values used are $y_1=y_2=y$.

The second implication of overlapping is that the size of regions corresponding to tree nodes decrease by *less* than a factor of 2 as we go from scale to scale. One consequence of this is that overlapping trees require more scales in order to reach the same finest partitioning of the image. This obviously increases the computational burden, although that is offset somewhat by the fact that lower-dimensional states can be used to capture the derived correlation structure. A second consequence is that the geometric scaling of the process noise $\mathcal{W}(s)$ in Eq. (33) must change in order to reflect the correct self-similar scaling laws implied by the thin-plate membrane model. In particular, if the linear dimension of the region represented by a node on an overlapped tree decreases by a factor of d rather than a factor of 2, then the noise gain $2^{-m(s)/2}$ on the right-hand side of Eq. (33) and in Eq. (34) as well should be replaced by $2(1-1/d)d^{-m(s)/2}$. In the results presented here, we used models with $d=2^{7/8}$.

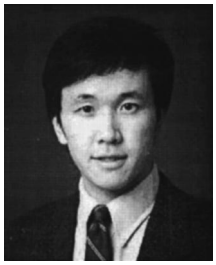
Acknowledgments

The authors would like to thank Dr. P.W. Fieguth, Dr. M. Schneider, Dr. W. M. Wells III, and Dr. W. C. Karl for their contributions to this paper. This work was supported by AFOSR grant F49620-98-1-0349 and by subcontract GC123919NGD from Boston University under the AFOSR Multidisciplinary Research Program.

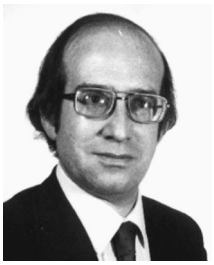
References

1. M. Basseville, A. Benveniste, K. Chou, S. Golden, R. Nikoukhah, and A. Willsky, ''Modeling and estimation of multiresolution stochastic processes,'' *IEEE Trans. Inf. Theory* **38**(2), 766–784 (1992).
2. K. Chou, A. Willsky, and A. Benveniste, ''Multiscale recursive estimation, data fusion, and regularization,'' *IEEE Trans. Autom. Control* **39**(3), 464–478 (1994).
3. A. Tsai, J. Zhang, and A. Willsky, ''Multiscale methods and mean field theory in EM procedures for image processing,'' presented at Eighth IEEE DSP Workshop 1998.
4. A. Tsai, ''Curve evolution and estimation-theoretic techniques for image processing,'' segmentation, PhD Thesis, MIT, Aug. 2000.
5. M. Luettgen, W. Karl, and A. Willsky, ''Efficient multiscale regularization with applications to the computation of optical flow,'' *IEEE Trans. Image Process.* **3**, 41–64 (1994).
6. D. Chandler, *Introduction to Modern Statistical Mechanics*, Oxford Univ. Press (1987).
7. A. Dempster, N. Laird, and D. Rubin, ''Maximum likelihood from incomplete data via the EM algorithm,'' *J. Roy. Statist. Soc. B* **39**, 1–38 (1977).
8. J. Zhang, ''The mean field theory in EM procedures for Markov random fields,'' *IEEE Trans. Signal Process.* **40**(10), 2570–2583 (1992).
9. M. Daniel and A. Willsky, ''A multiresolution methodology for signal-level fusion and data assimilation with applications to remote sensing,'' *Proc. IEEE* **85**, 164–180 (1997).
10. P. Fieguth, W. Karl, and A. Willsky, ''Efficient multiresolution counterparts to variational methods for surface reconstruction,'' *Comput. Vis. Image Underst.* **70**, 157–176 (1998).
11. W. Irving, P. Fieguth, and A. Willsky, ''An overlapping tree approach to multiscale stochastic modeling and estimation,'' *IEEE Trans. Image Process.* **6**, 1517–1529 (1997).
12. T. Green and J. Shapiro, ''Maximum-likelihood laser radar range profiling with the EM algorithm,'' *Opt. Eng.* **31**(11), 2343–2354 (1992).
13. D. Greer, I. Fung, and J. Shapiro, ''Maximum-likelihood multiresolution laser radar range imaging,'' *IEEE Trans. Image Process.* **6**, 36–46 (1997).
14. S. Geman and D. Geman, ''Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,'' *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 721–741 (1984).
15. R. Szeliski, *Bayesian Modeling of Uncertainty in Low-level Vision*, Kluwer Academic (1989).

Andy Tsai received his BS in electrical engineering from the University of California at San Diego in 1993, and his SM, EE, and PhD in electrical engineering from Massachusetts Institute of Technology in 1995, 1999, and 2000, respectively. He then continued his research as a postdoctoral research associate at the Laboratory for Information and Decision Systems at the Massachusetts Institute of Technology. Currently, he is a third-year medical student at Harvard Medical School. His research interests fall within the fields of computer vision and image processing.



Jun Zhang received his BS in electrical and computer engineering from Harbin Shipbuilding Engineering Institute, Harbin, China, in 1982 and was admitted to the graduate program of the Radio Electronic Department of Tsinghua University. After a brief stay at Tsinghua, he came to the USA for graduate study on a scholarship from the Li Foundation, Glen Cove, New York. He received his MS and PhD, both in electrical engineering, from Rensselaer Polytechnic Institute in 1985 and 1988, respectively. He joined the faculty of the Department of Electrical Engineering and Computer Science and currently is a professor. His research interests include image processing and computer vision, signal processing, and digital communications. He has been an associate editor of *IEEE Trans. Image Process.*



Alan S. Willsky received the SB degree and the PhD degree from Massachusetts Institute of Technology in 1969 and 1973, respectively. He joined the MIT faculty in 1973, and his present position is as the Edwin S. Webster Professor of Electrical Engineering. From 1974 to 1981 Dr. Willsky served as assistant director of the MIT Laboratory for Information and Decision Systems. He is also a founder and member of the board of directors of Alphatech, Inc., and a member of the U.S. Air Force Scientific Advisory Board. In 1975 he received the Donald P. Eckman Award from the American Automatic Control Council. Dr. Willsky has held visiting positions at Imperial College, London, L'Université de Paris-Sud; and the Institut de Recherche en Informatique et Systèmes Aléatoires in Rennes, France. He was program chairman for the 17th IEEE Conference on Decision and Control; has been an associate editor of several journals, including *IEEE Transactions on Automatic Control*; has served as a member of the Board of Governors and Vice President for Technical Affairs of the IEEE Control Systems Society; was program chairman for the 1981 Bilateral Seminar on Control Systems held in the People's Republic of China; was special guest editor of the 1992 special issue of the *IEEE Transactions on Information Theory* on wavelet transforms and multiresolution signal analysis; and served as co-chair of the 1998 ONR Working Group on the Role of Probability and Statistics in Command and Control. In 1988 he was made a Distinguished Member of the IEEE Control Systems Society. Dr. Willsky has given several plenary lectures at major scientific meetings, including the 20th IEEE Conference on Decision and Control; the 1991 IEEE International Conference on Systems Engineering; the SIAM Conference on Applied Linear Algebra, 1991; the 1992 Inaugural Workshop for the National Centre for Robust and Adaptive Systems, Canberra, Australia; the 1993 IEEE Symposium on Image and Multidimensional Signal Processing, Cannes, France; and the 1997 SPIE Wavelets and Applications Symposium, San Diego. Dr. Willsky is the author of the research monograph *Digital Signal Processing and Control and Estimation Theory* and is coauthor of the undergraduate text *Signals and Systems*. He was awarded the 1979 Alfred Noble Prize by the ASCE and the 1980 Browder J. Thompson Memorial Prize Award by the IEEE for a paper excerpted from his monograph. Dr. Willsky's present research interests are in problems involving multidimensional and multiresolution estimation and imaging, statistical image and signal processing, data fusion and estimation of complex systems, image reconstruction, and computer vision.