Regularization uses Fractal Priors

Richard Szeliski
Computer Science Department
Carnegie-Mellon University
Pittsburgh, PA 15213

Abstract

Many of the processing tasks arising in early vision involve the solution of ill-posed inverse problems. Two techniques that are often used to solve these inverse problems are regularization and Bayesian modeling. Regularization is used to find a solution that both fits the data and is also sufficiently smooth. Bayesian modeling uses a statistical prior model of the field being estimated to determine an optimal solution. One convenient way of specifying the prior model is to associate an energy function with each possible solution, and to use a Boltzmann distribution to relate the solution energy to its probability. This paper shows that regularization is an example of Bayesian modeling, and that using the regularization energy function for the surface interpolation problem results in a prior model that is fractal (self-affine over a range of scales). We derive an algorithm for generating typical (fractal) estimates from the posterior distribution. We also show how this algorithm can be used to estimate the uncertainty associated with a regularized solution, and how this uncertainty can be used at later stages of processing.

I. Introduction

Much of the processing that occurs in the early stages of vision deals with the solution of *inverse problems* [Horn, 1977]. The physics of image formation confounds many different phenomena such as lighting, surface reflectivity, surface geometry and projective geometry. Early visual processing attempts to recover some or all of these features from the sampled image array by making assumptions about the world being seen. For example, when solving the surface interpolation problem, i.e. the determination of a dense depth map from a sparse set of depth points (such as those provided by stereo matching), the assumption is made that surfaces vary smoothly in depth (except at object or part boundaries).

The inverse problems arising in early vision are generally *ill-posed* [Poggio and Torre, 1984], i.e. the data insufficiently constrains the desired solution. One approach to this problem, called *regularization*, imposes additional constraints in the form of smoothness assumptions. Another approach, *Bayesian modeling* [Geman and Geman, 1984], assumes a prior statistical distribution on the data being estimated, and models the image formation and sensing phenomena as stochastic (noisy) processes. Regularization can be viewed as a type of Bayesian modeling where the prior model is a Boltzmann distribution using the same energy function as the regularization. This paper shows that the average or most likely (optimal) estimate from the resulting posterior distribution is the same as the regularized solution. However, a *typical* sample from the posterior distribution is *fractal*, i.e. it exhibits self-similarity (and roughness) over a large range of scales [Pentland, 1984].

The fractal nature of the posterior distribution can be used to generate "realistic" fractal scenes with local control over elevation, discontinuities (either in depth or orientation) and fractal statistics. This paper presents an new algorithm for generating a sample from this distribution. This algorithm is a multigrid version of the Gibbs Sampler that is normally used for solving optimization problems whose energy function has many local minima [Szeliski, 1986]. We show that by using this algorithm we can also estimate the uncertainty associated with the regularized solution, for example by calculating the covariance matrix of the posterior distribution. The resulting error model can be used at later stages of processing along with the optimal estimate.

The remainder of this paper is structured as follows. Section II. reviews regularization techniques and shows an example of their application to the surface interpolation problem. Section III. discusses the application of Bayesian modeling to the solution of ill-posed problems, and shows that models that are Markov Random Fields can be specified by the choice of energy functions. Section IV. analyses the effects of regularization in the frequency domain, and derives the spectral characteristics of the Markov Random Fields that use the same energy functions. Section V. introduces fractal processes, and shows that the Markov Random Fields previously introduced are actually fractal. Section VI. gives a new algorithm for generating these fractals using multi-grid stochastic relaxation. Section VII. shows how this algorithm can be used to estimate the uncertainty inherent in regularized solutions. Section VIII. concludes with a discussion of possible applications of the results presented in this paper.

II. Regularization

Regularization is a mathematical technique used to solve ill-posed problems that imposes smoothness constraints on possible solutions[Tikhonov and Arsenin, 1977]. Given a set of data $\mathbf{d}$ from which we wish to recover the solution $\mathbf{u}$, we define an energy function $E_f(\mathbf{u}, \mathbf{d})$ which measures the compatibility between the solution and the sampled data. We then add a *stabilizing* function $E_s(\mathbf{u})$ which embodies the desired smoothness constraint and find the solution $\mathbf{u}^*$ that minimizes

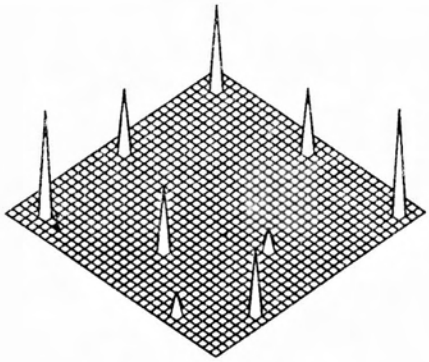

Figure 1: Sample data points

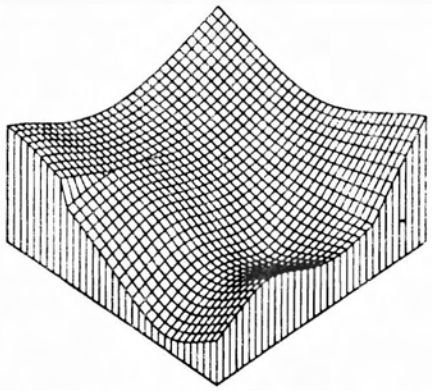

Figure 2: Regularized (thin plate) solution

the total energy

$$E(\mathbf{u}) = E_d(\mathbf{u}, \mathbf{d}) + \lambda E_p(\mathbf{u}) \tag{1}$$

The regularization parameter λ controls the amount of smoothing performed. In general, the data term $\mathbf{d}$ and solution $\mathbf{u}$ can be vectors, fields (two-dimensional arrays of data such as images or depth maps), or analytic functions (in which case the energy is a functional).

For the surface interpolation problem, the data is usually a sparse set of points $\{d_i\}$, and the desired solution is a two-dimensional function $u(x, y)$. The data compatibility term can be written as a weighted sum of squares

$$E_d(\mathbf{u}, \mathbf{d}) = \frac{1}{2} \sum_i w_i [u(x_i, y_i) - d_i]^2 \tag{2}$$

Two examples of possible smoothness functionals are the *membrane* model [Terzopoulos, 1984]

$$E_p(\mathbf{u}) = \frac{1}{2} \int \int \left(u_x^2 + u_y^2\right) \, dx \, dy \tag{3}$$

which is a small deflection approximation of the surface area, and the *thin plate* model

$$E_p(\mathbf{u}) = \frac{1}{2} \int \int \left(u_{xx}^2 + 2u_{xy}^2 + u_{yy}^2\right) \, dx \, dy \tag{4}$$

which is a small deflection approximation of the surface curvature (note that here the subscripts indicate partial derivatives). These two models can be combined into a single functional by using additional "rigidity" and "tension" functions, in order to introduce depth or orientation discontinuities [Terzopoulos, 1986].

As an example of a controlled-continuity regularizer, consider the nine data points shown in Figure 1. The regularized solution using a thin plate model is shown in Figure 2. Note that a depth discontinuity has been introduced along the left edge, an orientation discontinuity along the right, and that the regularized solution is very smooth away from these discontinuities.

The above stabilizer $E_p(\mathbf{u})$ is an example of the more general controlled-continuity constraint

$$E_p(\mathbf{u}) = \frac{1}{2} \sum_{m=0}^{n} \int w_m(\mathbf{x}) \sum_{j_1 + \cdots + j_d = m} \frac{m!}{j_1! \cdots j_d!} \left| \frac{\partial^m u(\mathbf{x})}{\partial x_1^{j_1} \cdots \partial x_d^{j_d}} \right|^2 d\mathbf{x} \tag{5}$$

where $\mathbf{x}$ is the (multi-dimensional) domain of the function u. This general formulation will be used in Section IV. to derive the spectral (frequency domain) characteristics of the stabilizer.

III. Bayesian modeling

The Bayesian modeling approach uses an *a priori* distribution $p(\mathbf{u})$ on the data being estimated, and a stochastic process $p(\mathbf{d}|\mathbf{u})$ relating the sampled data (input image) to the original data. According to Bayes' Rule, we have

$$p(\mathbf{u}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{u}) p(\mathbf{u})}{p(\mathbf{d})} \tag{6}$$

In its usual application [Geman and Geman, 1984], Bayesian modeling is used to find the *Maximum A Posteriori* (MAP) estimate, i.e. the value of $\mathbf{u}$ which maximizes the conditional probability $p(\mathbf{u}|\mathbf{d})$. In the more general case, the optimal estimator $\mathbf{u}^*$ is the value that minimizes the expected value of a loss function $L(\mathbf{u}, \mathbf{u}^*)$ with respect to this conditional probability.

Recently, Bayesian models that use Markov Random Fields have been used to solve ill-posed problems such as image restoration [Geman and Geman, 1984] and stereo matching [Szeliski, 1986]. A Markov Random Field (MRF) is a distribution where the probability of any one variable u_i is dependent on only a few neighbors,

$$p(u_i|\mathbf{u}) = p(u_i|\{u_j\}), \quad j \in N_i \tag{7}$$

In this case, the joint probability distribution $p(\mathbf{u})$ can be written as a Boltzmann (or Gibbs) distribution

$$p(\mathbf{u}) \propto \exp\left[-E_p(\mathbf{u})/T\right] \tag{8}$$

where T is called the "temperature". The "energy function" $E_p(\mathbf{u})$ can be written as a sum of local clique energies

$$E_p(\mathbf{u}) = \sum_{c \in C} E_c(\mathbf{u}) \tag{9}$$

where each clique energy $E_c(\mathbf{u})$ depends only on a few neighbors. Typically, the clique energy characterizes the local violation of the prior model or smoothness constraint.

The random vector $\mathbf{u}$ is sampled by a sensor which produces a data vector $\mathbf{d}$. We will model the measurement process

as having additive (multivariate) Gaussian noise

$$p(\mathbf{d}|\mathbf{u}) \propto \exp\left[-\frac{1}{2}(\mathbf{u}-\mathbf{d})^T\mathbf{A}(\mathbf{u}-\mathbf{d})\right] \equiv \exp[-E_d(\mathbf{u},\mathbf{d})] \tag{10}$$

From Bayes rule, we have

$$p(\mathbf{u}|\mathbf{d}) = \frac{p(\mathbf{u})p(\mathbf{d}|\mathbf{u})}{p(\mathbf{d})} \propto \exp[-E(\mathbf{u})] \tag{11}$$

where

$$E(\mathbf{u}) = E_p(\mathbf{u})/T + E_d(\mathbf{u},\mathbf{d}) \tag{12}$$

so that the posterior distribution is itself a Markov Random Field. Thus MAP estimation is equivalent to finding the minimum energy state. This shows that regularization is an example of the more general MRF approach to optimal estimation. The smoothing term (stabilizer) $E_p(\mathbf{u})$ corresponds to the *a priori* distribution, and the data compatibility term $E_d(\mathbf{u},\mathbf{d})$ corresponds to the measurement process.

While Bayesian modeling has previously been used in computer vision to find an optimal estimate, it has not been used to generate an error model. We propose to estimate additional (second order) statistics using this model, and to use these additional statistics at later stages of processing. For example, we can use these statistics when matching for object recognition or pose detection, or to optimally integrate new knowledge or measurements (by using Kalman filtering [Smith and Cheeseman, 1985]). We present a method for calculating these statistics in Section VII..

IV. Fourier analysis

By taking a Fourier transform of the function $u(\mathbf{x})$ and expressing the energy equations in the frequency domain, we can analyse the filtering behaviour of regularization and the spectral characteristics of the prior model. To simplify the analysis, we will set the weighting function $w_m(\mathbf{x})$ used in Equation 5 to a constant. While this analysis does not strictly apply to the general case, it provides an approximation to the local behaviour of the regularized system away from boundaries and discontinuities.

The Fourier transform [Bracewell, 1978] of a multidimensional signal $h(\mathbf{x})$ is defined by

$$\mathcal{F}\{h\} \equiv \int h(\mathbf{x})\exp(2\pi i\,\mathbf{f}\cdot\mathbf{x})\,d\mathbf{x} = H(\mathbf{f}) \tag{13}$$

and the transform of its partial derivative is given by

$$\mathcal{F}\{\frac{\partial h(\mathbf{x})}{\partial x_j}\} = (2\pi i f_j)H(\mathbf{f}) \tag{14}$$

By using Parseval's theorem

$$\int |h(\mathbf{x})|^2 d\mathbf{x} = \int |H(\mathbf{f})|^2 d\mathbf{f} \tag{15}$$

we can derive the smoothness functional E_p in terms of the Fourier transform $U(\mathbf{f}) = \mathcal{F}\{u\}$. The notation $E_p(U)$ denotes the energy associated with a signal U, which is derived from the original definition of $E_p(u)$ (in this case by using a Fourier transform). Applying the Equations 14 and 15 to Equation 5 we obtain

$$E_p(U) = \frac{1}{2}\int |G(\mathbf{f})|^2|U(\mathbf{f})|^2 d\mathbf{f} \tag{16}$$

where

$$|G(\mathbf{f})|^2 = \sum_{m=0}^{p} w_m|2\pi\mathbf{f}|^{2m} \tag{17}$$

For example, the membrane interpolator has $|G(\mathbf{f})|^2 \propto |2\pi\mathbf{f}|^2$ and the thin plate model has $|G(\mathbf{f})|^2 \propto |2\pi\mathbf{f}|^4$.

Since the Fourier transform is a linear operation, if $u(\mathbf{x})$ is Boltzmann distributed with energy $E_p(u)$, then $U(\mathbf{f})$ is also Boltzmann distributed with energy $E_p(U)$. Thus we have

$$p(U) \propto \exp\left[-\frac{1}{2}\int |G(\mathbf{f})|^2|U(\mathbf{f})|^2 d\mathbf{f}\right] \tag{18}$$

from which we see that the probability distribution at any frequency $\mathbf{f}$ is

$$p(U(\mathbf{f})) \propto \exp\left[-\frac{1}{2}|G(\mathbf{f})|^2|U(\mathbf{f})|^2\right] \tag{19}$$

Thus, $U(\mathbf{f})$ is a random Gaussian variable with variance $|G(\mathbf{f})|^{-2}$, and the signal $u(\mathbf{x})$ is correlated Gaussian noise with a spectral distribution

$$S_u(\mathbf{f}) = |G(\mathbf{f})|^{-2} \tag{20}$$

We can also use the same Fourier analysis techniques to determine the frequency response of regularization viewed as linear filtering. The result of this analysis (see [Szeliski, 1987] for details) is that the effective smoothing filter has a frequency response

$$H(\mathbf{f}) = \frac{1}{1+\sigma^2|G(\mathbf{f})|^2} \tag{21}$$

where σ is the standard deviation of the sensor noise (with uniform dense sensing). For the case of the membrane model and the thin plate model, the shape of the frequency response is qualitatively similar to that of Gaussian filtering. The overall posterior distribution (when the data confidence and prior model are spatially uniform) is the superposition of the regularized (smooth) solution and some correlated Gaussian noise. Fourier analysis can also be used to examine the convergence properties of the iterative algorithms discussed in Section VI. [Szeliski, 1987].

V. Fractals

Fractals are objects (geometric designs, coastlines, mountain surfaces) that exhibit self-similarity over a range of scales [Mandelbrot, 1982]. Fractals have been used to generate "realistic" images of terrain or surfaces that exhibit roughness, and to analyse certain types of structured noise. Brownian fractals are random processes or random fields that exhibit similar statistics over a range of scales. One common way to characterize such a fractal is to say that it follows a power law in its spectral density

$$S_v(f) \propto 1/f^\beta \tag{22}$$

This spectral density characterizes a fractal Brownian function $v_H(\mathbf{x})$ with $2H = \beta - E$, whose fractal dimension is $D = E + 1 - H$.

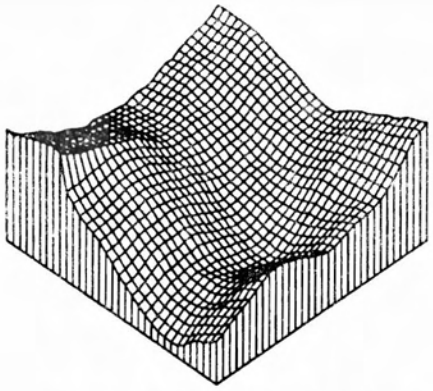

Figure 3: Fractal (random) solution

H (where E is the dimension of the Euclidean space) [Voss, 1985].

The spectral density of the regularization based prior models examined in the previous section is $|G(\mathbf{f})|^{-2}$. For a membrane interpolator, we have

$$S_{membrane}(\mathbf{f}) \propto |2\pi\mathbf{f}|^{-2} \tag{23}$$

while for a thin plate interpolator, we have

$$S_{thin-plate}(\mathbf{f}) \propto |2\pi\mathbf{f}|^{-4} \tag{24}$$

Thus, the prior models for a membrane and a thin plate are indeed fractal, since the spectral density is a power of the frequency.

The significance of this connection between regularization methods, Bayesian models and fractal models is two-fold. First, it shows that the smoothness assumptions embedded in regularization methods are equivalent to assuming that the underlying processes is fractal. When regularization techniques are used, it is usual to find the minimum energy solution (Figure 2), which also corresponds to the mean value solution for those cases where the energy functions are quadratic. Thus, the fractal nature of the process is not evident. A far more *representative* solution can be generated if a random (fractal) sample is taken from this distribution. Figure 3 shows such a random sample, generated by the algorithm that will be explained in section VI.. The amount of noise (and hence "bumpiness") that is desirable or appropriate can be derived from the data [Szeliski, 1987].

Second, the connection between Bayesian models and fractal models gives us a powerful new technique (described in Section VI.) for generating fractal surfaces for computer graphics applications. Previous techniques for generating fractals use either recursive subdivision algorithms [Fournier *et al.*, 1982] or the addition of correlated (pink) noise to some initial data [Pentland, 1984]. While the latter algorithm is equivalent to Bayesian modeling with uniform data and prior models, the Bayesian modeling approach can be extended to non-uniform data and the full controlled-continuity constraint. Thus, it is possible to constrain the desired fractal by placing control points at selected locations (using the discrete data formulation), or to introduce discontinuities such as cliffs or ridges. For example, the fractal in Figure 3 has been required to pass through the points in Figure 1, and has a depth discontinuity along the left edge and an orientation discontinuity along the right. The introduction of data points affects the local noise characteristics of the fractal *without* affecting the prior statistics. It thus generates a representative random sample that is true both to the fractal statistics being used and to the sampled (or desired) data points. This approach can also be used for doing interpolation of digital terrain models. Interpolators that have a smoothing behaviour between that of a membrane and a thin plate are better able to model the correct smoothness (fractal dimension) of natural terrain.

VI. Finite element solution

To simulate the Markov Random Field (or equivalently, to find the minimum energy solution) on a digital or analog computer, it is necessary to discretize the domain of the solution $u(\mathbf{x})$ by using a finite number of *nodal variables*. The usual and most flexible approach is to use finite element analysis [Terzopoulos, 1984]. We will restrict our attention to rectangular domains on which a rectangular mesh has been applied. As well, the input data points will be constrained to lie on this mesh.

As an example, let us examine the finite element approximation for the surface interpolation problem. Using a triangular conforming element for the membrane, and a non-conforming rectangular element for the thin plate (as in [Terzopoulos, 1984]), we can derive the energy equations

$$E_{membrane}(\mathbf{u}) = \frac{1}{2}\sum_{(x,y)}[(u_{x+1,y} - u_{x,y})^2 + (u_{x,y+1} - u_{x,y})^2] \tag{25}$$

for the membrane and

$$\begin{aligned} E_{thin-plate}(u) \quad = \quad & \frac{|\Delta x|^{-2}}{2}\sum_{(x,y)}[(u_{x+1,y} - 2u_{x,y} + u_{x-1,y})^2 + \\ & 2(u_{x+1,y+1} - u_{x,y+1} - u_{x+1,y} + u_{x,y})^2 + \\ & (u_{x,y+1} - 2u_{x,y} + u_{x,y-1})^2] \end{aligned} \tag{26}$$

for the thin plate, where $|\Delta x|$ is the size of the mesh (isotropic in x and y). These equations hold at the interior of the surface, i.e. away from the border points and discontinuities. Near border points or discontinuities some of the energy terms are dropped or replaced by lower continuity terms (see [Szeliski, 1987] for details). The equation for the data compatibility term is simply

$$E_d(\mathbf{u}, \mathbf{d}) = \frac{1}{2}\sum_{(x,y)} w_{x,y}(u_{x,y} - d_{x,y})^2 \tag{27}$$

with $w_{x,y} = 0$ at points where there is no input data.

If we concatenate all the nodal variables $\{u_{x,y}\}$ into one vector $\mathbf{u}$, we can write the prior energy model as one quadratic form

$$E_p(\mathbf{u}) = \frac{1}{2}\mathbf{u}^T\mathbf{A}_p\mathbf{u} \tag{28}$$

This quadratic form is valid for any controlled continuity stabilizer, though the coefficients differ. Similarly, for the data compatibility model we can write

$$E_d(\mathbf{u}, \mathbf{d}) = \frac{1}{2}(\mathbf{u} - \mathbf{d})^T\mathbf{A}_d(\mathbf{u} - \mathbf{d}) \tag{29}$$

where $\mathbf{A}_d$ is usually diagonal (for uncorrelated sensor noise). The resulting overall energy function $E(\mathbf{u})$ is quadratic in $\mathbf{u}$

$$E(\mathbf{u}) = \frac{1}{2}\mathbf{u}^T\mathbf{A}\mathbf{u} - \mathbf{u}^T\mathbf{b} + c \tag{30}$$

where

$$\mathbf{A} = \mathbf{A}_p + \mathbf{A}_d \quad \text{and} \quad \mathbf{b} = \mathbf{A}_d\mathbf{d} \tag{31}$$

and has a minimum at $\mathbf{u}^*$

$$\mathbf{u}^* = \mathbf{A}^{-1}\mathbf{b} \tag{32}$$

Once the parameters of the energy function have been determined, we can calculate the minimum energy solution $\mathbf{u}^*$ by using relaxation. For faster convergence on a serial machine, we use Gauss-Seidel relaxation where nodes are updated one at a time. At each step the selected node is set to the value that (locally) minimizes the energy function. The energy function for node u_i (with all other nodes fixed) is

$$E(u_i) = \frac{1}{2}a_{ii}u_i^2 + (\sum_{j \in N_i} a_{ij}u_j - b_i)u_i + k \tag{33}$$

and so the new node variable value is

$$u_i^+ = \frac{b_i - \sum_{j \in N_i} a_{ij}u_j}{a_{ii}} \tag{34}$$

The result of executing this iterative algorithm on the nine data points in Figure 1 is shown in Figure 2. Note that it is possible to use a parallel version of Gauss-Seidel relaxation so long as nodes that are dependent (have a non-zero a_{ij} entry) are not updated simultaneously. This parallel version can be implemented on a mesh of processors for greater computational speed.

The stochastic version of Gauss-Seidel relaxation is known as the "Gibbs Sampler" [Geman and Geman, 1984] or Boltzmann Machine [Hinton *et al.*, 1984]. Nodes are updated sequentially (or *asynchronously*), with the new nodal value selected from the local Boltzmann distribution

$$p(u_i) \propto \exp\left[-E(u_i)/T\right] \tag{35}$$

Since the local energy is quadratic

$$E(u_i) = a_{ii}(u_i - u_i^+)^2 + k \tag{36}$$

this distribution is a Gaussian with a mean equal to the deterministic update value u_i^+ and a variance equal to T/a_{ii}. Thus, the Gibbs Sampler is equivalent to the usual relaxation algorithm with the addition of some locally controlled Gaussian noise at each step. The resulting surface exhibits the rough (wrinkled) look of fractals (Figure 3). The amount of roughness can be controlled by the "temperature" parameter T. The "best" value for T can be determined by using parameter estimation techniques [Szeliski, 1987].

While the above iterative algorithms will eventually converge to the correct estimate, their performance in practice is unacceptably slow. To overcome this, multigrid techniques [Terzopoulos, 1984] can be used. The problem is first solved on a coarser mesh, then this solution is used as a starting point for the next finer level (thus this is a coarse-to-fine algorithm). In previous work [Terzopoulos, 1984] a more complex inter-level coordination strategy was used, but in this instance it has not been found to be necessary. The application of multigrid techniques to stochastic algorithms requires some care, since the energy equations must be preserved when mapping from a fine to a coarse level [Szeliski, 1987].

The application of a multigrid Gibbs Sampler to the generation of samples from a Markov Random Field with fractal priors results in a new algorithm for fractal generation. Like other commonly used techniques (random midpoint displacement, successive random additions [Voss, 1985]), it is a coarse-to-fine algorithm. It uses the interpolated coarse level solution as a starting point for the next finer level, just like successive random additions. However, the noise that is added at each stage is highly correlated. Since control points and discontinuities can be imposed at arbitrary locations, it gives more control over the fractal generation process.

VII. Estimating uncertainty

The preceding section has discussed how to obtain representative samples from the estimated posterior distribution. While this ability is useful in computer graphics, it is less relevant to the problems associated with computer vision. What is of interest is the optimal (or average) estimate, and also the uncertainty associated with this estimate. These uncertainty estimates can be used to integrate new data, guide search (set disparity limits in stereo matching), or dictate where more sensing is required.

For the Markov Random Field with a quadratic energy function (Equation 30), the probability distribution is a multivariate Gaussian with mean $\mathbf{u}^*$ and covariance $\mathbf{A}^{-1}$. Thus, to obtain the covariance matrix, we need only invert the $\mathbf{A}$ matrix. One way of doing this is to use the multigrid algorithm presented in the previous section to calculate the covariance matrix one row at a time [Szeliski, 1987]. However, this approach is time consuming, and storing all the covariance fields is impractical because of their large size (for a 512×512 image, the covariance matrix has 6.8×10^{10} entries).

An alternative to this deterministic algorithm is to run the multigrid Gibbs Sampler at a non-zero temperature, and to estimate the desired statistics (this is a Monte Carlo approach). For example, we can estimate the variance at each point (the diagonal of the covariance matrix) simply by keeping a running total of the depth values and their squares. Figure 4 shows the variance estimate corresponding to the regularized solution of Figure 2 (note how the variance increases near the edges and discontinuities). These variance values are an estimate of the confidence associated with each point in the regularized solution. Alternatively, they can be viewed as the amount of fluctuation at a point in the Markov Random Field (the "wobble" in the thin plate). Note that this error model is *dense*, since a measure of uncertainty is available at every point in the image. Error modeling in computer vision has not previously been applied to systems with such a large number of parameters.

The straightforward application of the Gibbs Sampler results in estimates that are biased or take extremely long to

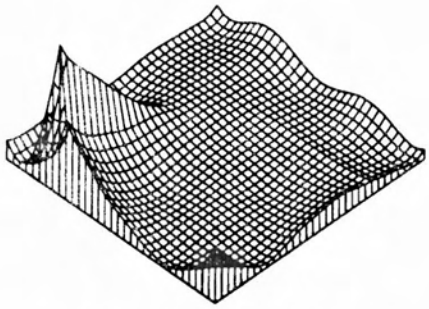

Figure 4: Variance estimate

converge. This is because the Gibbs Sampler is a multi-dimensional version of the Markov random walk, so that successive samples are highly correlated, and time averages are ergodic only over a very large time scale. To help decorrelate the signal, we can use successive coarse-to-fine iterations, and only gather a few statistics at the fine level each time [Szeliski, 1987].

The stochastic estimation technique can also be used with systems that have non-quadratic (and non-convex) energy functions. In this case, the mean and covariance are not sufficient to completely characterise the distribution, but they can still be estimated. For the example of stereo matching, once the best match has been found (by using simulated annealing), it may still be useful to estimate the variance in the depth values. Alternatively, stochastic estimation may be used to provide a whole distribution of possible solutions, perhaps to be disambiguated by a higher level process.

VIII. Conclusions

This paper has shown that regularization can be viewed as a special case of Bayesian modeling, and that such an interpretation results in prior models that are fractal. We have shown how this can be used to generate typical solutions to inverse problems, and also to generate constrained fractals with local control over continuity and fractal dimension. We have devised and implemented a multigrid stochastic algorithm that allows for the efficient simulation of the posterior distribution (which is a Markov Random Field).

The same approach has been extended to estimate the uncertainty associated with a regularized solution in order to build an error model. This information can be used at later stages of processing for sensor integration, search guidance, and on-line estimation. Work is currently under way [Szeliski, 1987] in studying related issues, such as the estimation of the model parameters, analysis of algorithm convergence rates, on-line estimation of depth and motion using Kalman filtering, and the integration of the multiple resolution levels into a single representation.

References

[Bracewell, 1978] R. N. Bracewell. *The Fourier transform and its applications*. McGraw Hill, New York, 2nd edition, 1978.

[Fournier *et al.*, 1982] A. Fournier, D. Fussel, and L. Carpenter. Computer rendering of stochastic models. *Commun. ACM*, 25(6):371–384, 1982.

[Geman and Geman, 1984] S. Geman and D. Geman. Stochastic relaxation, gibbs distribution, and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-6(6):721–741, November 1984.

[Hinton *et al.*, 1984] G. E. Hinton, T. J. Sejnowski, and D. H. Ackley. *Boltzmann Machines: Constraint Satisfaction Networks that Learn*. Technical Report CMU-CS-84-119, Carnegie-Mellon University, May 1984.

[Horn, 1977] B. K. P. Horn. Understanding image intensities. *Artificial Intelligence*, 8:201–231, April 1977.

[Mandelbrot, 1982] B. B. Mandelbrot. *The Fractal Geometry of Nature*. W. H. Freeman, San Francisco, 1982.

[Pentland, 1984] A. P. Pentland. Fractal-based description of natural scenes. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-6(6):661–674, November 1984.

[Poggio and Torre, 1984] T. Poggio and V. Torre. Ill-posed problems and regularization analysis in early vision. In *IUS Workshop*, pages 257–263, ARPA, October 1984.

[Smith and Cheeseman, 1985] R. C. Smith and P. Cheeseman. *On the representation and estimation of spatial uncertainty*. Technical Report (draft), SRI International, 1985.

[Szeliski, 1986] R. Szeliski. *Cooperative algorithms for solving random-dot stereograms*. Technical Report CMU-CS-86-133, Department of Computer Science, Carnegie-Mellon University, June 1986.

[Szeliski, 1987] R. Szeliski. *Uncertainty in Low Level Representations*. PhD thesis, Carnegie Mellon University, (in preparation) 1987.

[Terzopoulos, 1984] D. Terzopoulos. *Multiresolution Computation of Visible-Surface Representations*. PhD thesis, Massachusetts Institute of Technology, January 1984.

[Terzopoulos, 1986] D. Terzopoulos. Regularization of inverse visual problems involving discontinuities. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, PAMI-8(4):413–424, July 1986.

[Tikhonov and Arsenin, 1977] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems*. V. H. Winston and Sons, Washington, D. C., 1977.

[Voss, 1985] R. F. Voss. Random fractal forgeries. In R. A. Earnshaw, editor, *Fundamental Algorithms for Computer Graphics*, Springer-Verlag, Berlin, 1985.

Acknowledgements

I would like to thank Geoff Hinton and Takeo Kanade for their guidance in this research and their helpful comments on the paper. This work is supported in part by NSF grant IST 8520359, and by ARPA Order No. 4976, monitored by the Air Force Avionics Laboratory under contract F33615-84-K-1520.

AAAI 87

July 13-17, 1987

Sixth National Conference on Artificial Intelligence

Seattle, Washington

Volume 2

Proceedings