HIERARCHICAL POSTERIOR SAMPLING FOR IMAGES AND RANDOM FIELDS

Paul Fieguth

Systems Design Engineering
University of Waterloo
Waterloo, Ontario, Canada
pfieguth@uwaterloo.ca

ABSTRACT

The estimation of images and random fields from sparse and/or noisy data is highly-developed, to the point where methods such as least-squares estimation, simulated annealing, and wavelet shrinkage are quite standardized. The key problem, however, is that the estimates are *not* a realistic version of the random field, and do *not* represent a typical or representative sample of the system being studied.

Instead, what is often desired is that we find a random sample from the posterior distribution, a much more subtle and difficult problem than estimation. Typically this is solved using Markov-Chain Monte-Carlo / Simulated Annealing approaches, however these may be computationally challenging and slow to converge. In this paper we use hierarchical models to formulate a novel, fast posterior sampler.

1. INTRODUCTION

For the majority of image analysis problems, characterized by densely-measured images (e.g., from a camera), standard low-level estimation algorithms are more than adequate, and the challenge lies, instead, in the high-level modeling and interpretation of such images, not the subject of this paper.

However there does exist an important class of image processing problems, mostly scientific, in which the "image" is governed by some sort of known behaviour (a prior model), and the measurements are sparse (constrained by physics, time, and/or cost). One key problem in such sparse contexts is that the estimates are *not* a realistic version of the random field, as shown in Figure 1; typical or representative sample of the system being studied may be desired for purposes of visualization, further analysis, Monte Carlo studies etc.

Instead, what is required is that we find a random sample from the posterior distribution, a much more subtle and difficult problem than estimation, and, crucially, one which cannot be formulated as an optimization problem. Because the brute-force approach to sampling is computationally inconceivable, the Markov-Chain Monte-Carlo / Simulated Annealing approach is nearly universally used, especially for finite-state problems, however this approach can still be computationally challenging. In this paper we use a little-known property of multiscale statistical models to formulate a posterior sampler, exact in the case of Markov random fields, and approximate for other distributions.

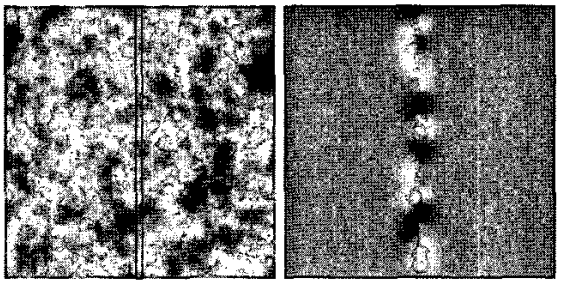


Fig. 1. Limitations of estimation in the context of sparsely-measured random fields: The left panel shows a random sample, of which the middle column is measured, from which estimates (right) are produced. The statistics of the estimates are clearly *very* different from the original, true field; in particular, it would be extremely unlikely for a random texture to be perfectly flat for 80% of the image, as these estimates are.

Section 2 describes the mathematical background behind sampling, followed by a development of the multiscale approach in Section 3 and experimental results in Section 4.

2. BACKGROUND

First, suppose the random field has the known Normal statistical distribution

$$\underline{z} \sim \mathcal{N}(\underline{\mu}, P) \tag{1}$$

Finding the eigendecomposition of $P = V\Lambda V^T$, the matrix $L = V\Lambda^{1/2}$ is referred to as the matrix square-root of P.

[0]The support of the Natural Science & Engineering Research Council of Canada is acknowledged.

0-7803-7750-8/03/$17.00 ©2003 IEEE

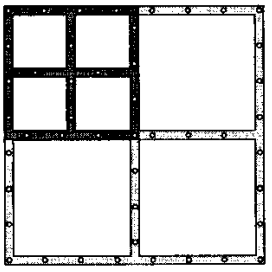


Fig. 2. A typical state arrangement at the coarsest level of a 2D multiscale quad-tree (light grey) and the top-left state at the next scale (dark grey). The spacing between pixels will be a function of the prior statistics, accuracy requirements, and computational needs.

Then a random *prior* sample of $\underline{z}$ may be found as

$$\underline{z} = \underline{\mu} + L\underline{w}, \ \underline{w} \sim \mathcal{N}(0, I). \quad (2)$$

Next, given linear measurements

$$\underline{m} = C\underline{z} + \underline{v}, \ \underline{v} \sim \mathcal{N}(0, R) \quad (3)$$

then the linear least-squares estimates $\hat{\underline{z}}$ and estimation error covariance $\tilde{P}$ may be computed, as usual, where $\tilde{\underline{z}} = \hat{\underline{z}} - \underline{z}$ is the estimation error. Of these two, the former computation of $\hat{\underline{z}}$ is relatively straightforward; the latter computation of $\tilde{P}$ is much more difficult, and is omitted in many contexts.

Our goal is to find a "typical" sample from random vector $\underline{z}$, constrained by the measurements $\underline{m}$; that is, we wish to find random samples from the distribution $p(\underline{z}|\underline{m})$. From the estimation, we have

$$E[\underline{z}|\underline{m}] = \hat{\underline{z}}, \ \ \mathrm{cov}(\underline{z}|\underline{m}) = \tilde{\mathbf{P}} \quad (4)$$

If we let $\tilde{L} = \tilde{V}\tilde{\Lambda}^{1/2}$ be the matrix square root of $\tilde{P} = \tilde{V}\tilde{\Lambda}\tilde{V}^T$, then the desired posterior samples, consistent with the measurements and with the prior model, are given by

$$(\underline{z}|\underline{m}) = \hat{\underline{z}} + \tilde{L}\underline{\tilde{w}}, \ \underline{\tilde{w}} \sim \mathcal{N}(0, I) \quad (5)$$

Finally, *if* pieces (e.g., the columns) of z obey a dynamic relationship, then sampling may be simplified. For example, if the columns $\underline{z}_i$ of image z obey a standard Gauss-Markov model

$$\underline{z}_{i+1} = A_i\underline{z}_i + \underline{q}_i, \ \ \underline{z}_o \sim \mathcal{N}(\underline{\mu}_o, P_o), \ \ \underline{q}_i \sim \mathcal{N}(0, Q_i), \quad (6)$$

then we would need to find square roots $P_o = L_o L_o^T, Q_i = L_i L_i^T$; clearly the key to efficiency is to keep the dimensions of the $\underline{q}_i$ as small as possible. Of course, to do *posterior* sampling, it is really a dynamic relationship for the estimation *errors* which we require:

$$\tilde{\underline{z}}_{i+1} = \tilde{A}_i\tilde{\underline{z}}_i + \tilde{\underline{q}}_i, \ \ \tilde{\underline{z}}_o \sim \mathcal{N}(\hat{\underline{z}}_o, \tilde{P}_o), \ \ \tilde{\underline{q}}_i \sim \mathcal{N}(0, \tilde{Q}_i). \quad (7)$$

Although it is not obvious that such a form is obeyed, it turns out that the error process of the Kalman smoother *does* obey such a dynamic process[1].

0	0	1	0	0
0	2	-8	2	0
1	-8	20.01	-8	1
0	2	-8	2	0
0	0	1	0	0

Table 1. Third-order "Thin-plate" coefficients.

0	-0.0091	0.0517	0.0008	0
0.0058	0.1405	-0.5508	0.1164	0.0085
-0.0139	-0.2498	1.0000	-0.2498	-0.0139
0.0085	0.1164	-0.5508	0.1405	0.0058
0	0.0008	0.0517	-0.0091	0

Table 2. Fourth-order "Wood-grain" coefficients.

3. MULTISCALE

The key challenge of Section 2, is how to compute the error covariance $\tilde{P}$ and the huge matrix square-roots, both extremely difficult! A variety of approaches have been proposed to decompose large estimation problems into smaller pieces; one in particular, the hierarchical multiscale one[7], has seen considerable effort and model development.

The multiscale model asserts a scale-to-scale process:

$$\underline{z}(s) = A(s)\underline{z}(ps) + \underline{w}(s), \ \ \underline{w}(s) \sim Q(s), \ \ \underline{z}(0) \sim P_0, \quad (8)$$

where s is an index on the nodes of a tree, 0 is the node corresponding to the tree root, and p is an operator, returning the parent ps of node s.

The usual arrangement, for two-dimensional random fields, is that $\underline{z}(s)$ live on a quad-tree, where the leaf-nodes of the tree represent the pixels of the random field of interest. This approach can model a variety of random fields of interest: certainly two-dimensional Markov random fields can be represented exactly [7], there has been success in modeling fractal and $1/f$-like processes, and results for more general stochastic realization [2, 5].

The estimator [7] which follows from (8) greatly resembles the Kalman smoother. As with the Kalman smoother, the algorithm consists of only two passes; it is not iterative. Furthermore the approach is exact, in the sense that the computed estimates are exact; any approximation lies in the choice of model itself.

The *key* benefit of the multiscale framework, for the context of this paper, is a mostly-overlooked result [6] which shows that the multiscale error process $\tilde{\underline{z}}(s)$ *also* obeys a multiscale model:

$$\tilde{\underline{z}}(s) = \tilde{A}(s)\tilde{\underline{z}}(ps) + \underline{\tilde{w}}(s), \ \ \underline{\tilde{w}}(s) \sim \tilde{Q}(s), \ \ \tilde{\underline{z}}(0) \sim \tilde{P}_0, \quad (9)$$

where $\tilde{A}(s), \tilde{Q}(s)$ are expressed [6] in terms of the model parameters $A(s), B(s)$, prior covariances $P(s)$, and estimation error covariances $\tilde{P}_u(s)$. Although the original intent of

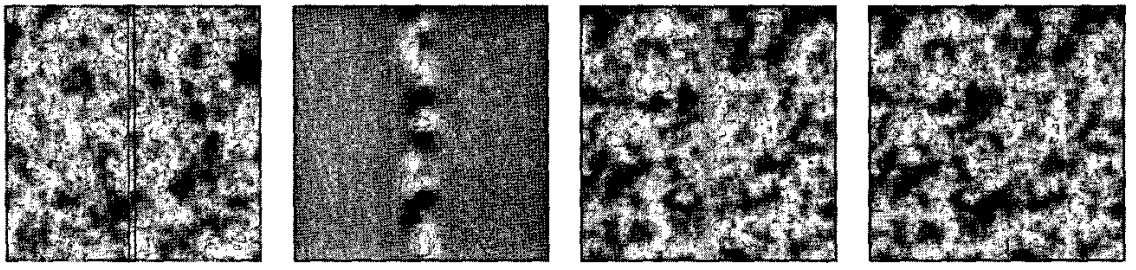


Fig. 3. The process of posterior sampling. The leftmost two panels repeat Figure 1, showing a sample from the prior model, and estimates based on the central measured column. The third panel shows the sampled estimation error, where a low-variance zero-mean band can be seen where the estimates are good. The final panel shows the sampled posterior, consistent with both the measurements and the prior statistics.

deriving the smoothing-error model was predominantly for other purposes of data fusion and estimation error covariance P calculation, clearly the availability of a multiscale smoothing error model leads to an approach to hierarchical posterior sampling, following from the dynamic approach discussed in the previous section:

1. The multiscale estimator computes $\underline{\hat{z}}(s)$, $\tilde{P}(s)$.
2. We then compute the smoothing error model $\tilde{A}(s)$, $\tilde{Q}(s)$.
3. Compute the matrix square-roots $\tilde{L}(0)$, $\tilde{L}(s)$.
4. Sample the smoothing error process as

$$(\underline{\tilde{z}}|\underline{m})(0) = \tilde{L}(0)\underline{\tilde{w}}(0), \quad \underline{\tilde{w}}(s) \sim \mathcal{N}(0, I) \qquad (10)$$

$$(\underline{\tilde{z}}|\underline{m})(s) = \tilde{A}(\underline{\tilde{z}}|\underline{m})(ps) + \tilde{L}(s)\underline{\tilde{w}}(s). \qquad (11)$$

5. Finally, add in the mean (the estimates)

$$(\underline{z}|\underline{m})(s) = \underline{\hat{z}}(s) + (\underline{\tilde{z}}|\underline{m})(s) \qquad (12)$$

With the principle of the method in place, the following section demonstrates the approach, its versatility, and computational issues.

4. RESULTS

Two prior models were used to illustrate and assess the proposed approach: a thin-plate Markov model and a Markov wood-grain sort of texture, the coefficients shown in Tables 1 and 2, respectively.

Figure 3 illustrates the posterior sampling process, building on the motivation from Figure 1. Because of stationarity and periodicity, the prior samples can be generated very efficiently using standard FFT methods, however the aperiodic structure of the measurements implies that the FFT cannot be used for posterior sampling[1]. The estimates $\underline{\hat{z}}|\underline{m}$ clearly reveal the measurement structure, and look nothing at all like the underlying texture. The sampled estimation error $\underline{\tilde{z}}|\underline{m}$ shows the random behaviour not constrained by the measurements; indeed, the constrained domain is clearly visible down the middle of the image. Finally the posterior sample $\underline{z}|\underline{m} = \underline{\hat{z}}|\underline{m} + \underline{\tilde{z}}|\underline{m}$ is a new random texture, consistent with the original where measured, and showing no evidence of the measurement structure[2].

Figure 3 has a fairly regular structure, however the multiscale model makes no particular assumptions regarding stationarity, so it is just as applicable in cases such as in Figure 4, which shows a decoupled process; the decoupling is clearly visible in the final, sampled image.

Finally, Figure 5 looks at issues of computational complexity. Comparisons with other approaches are difficult because of the limited approaches available: MCMC methods are iterative, and continuous-state annealing is difficult and slow; brute-force approaches are easily implemented, but horribly slow (about 28 hours on a 2GHz workstation); and FFT and Kalman filter methods apply only to stationary problems which are not of realistic interest.

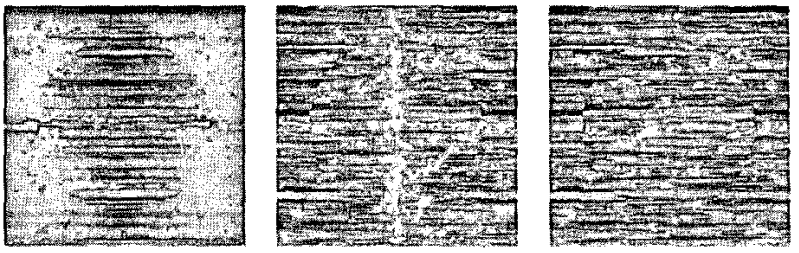
The panels of Figure 5 proceed from dense to relaxed state definition, reducing the computational complexity and ultimately introducing artifacts. An overlapping approach (e) may allow a reduced-state model to produce good results, as is the case here. The largest part of the computational effort is in setting up the model and computing the matrix square-roots. Once these square-roots have been computed and saved, any number of posterior samples can be produced extremely quickly (less than one second).

5. CONCLUSIONS

We have developed a multiscale approach for posterior sampling - the random sampling from the posterior distribution of a prior model plus constraints due to measurements.

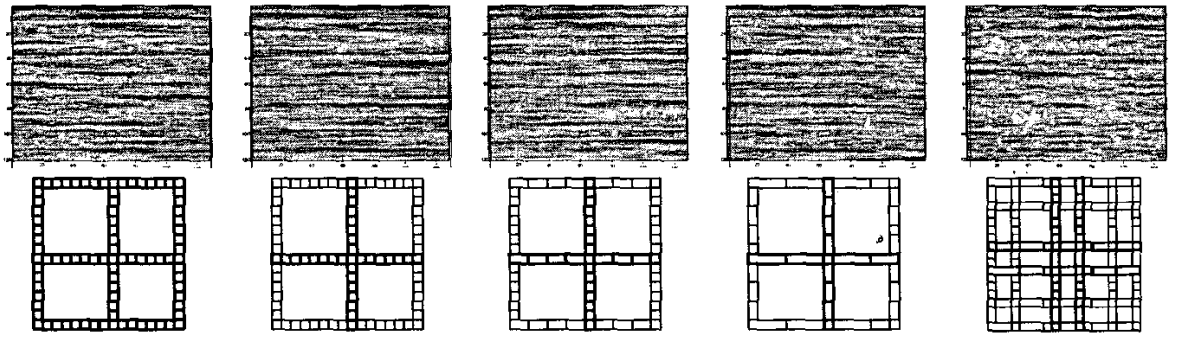
[1] In the artificial, special case that both prior and measurements are stationary / periodic, then indeed the FFT can be used to compute estimates, error variances, and posterior samples.

[2] The structure would, of course, be clearly visible if the measurement values grossly violated the prior model.



$\hat{z}|\underline{m}$ $\tilde{z}|\underline{m}$ $z|\underline{m}$

Fig. 4. An example of multiscale nonstationary estimation and sampling. We have two decoupled domains - one inside the circle, and the other outside. Measurements are taken in the central column and row.



a) Dense State 819 sec. b) Aperiodic 149 sec. c) Sparse Horz. 39 sec. d) Very Sparse 21 sec. e) Sparse, Overlapped 31 sec.

Fig. 5. An illustration of computational complexity and tradeoffs, illustrated for the wood-grain model, each panel showing a multiscale posterior sample, based on measurements down the middle column. The panels differ in the nature of the multiscale state: from left, a densely-sampled state, then a reduction in coarse state sampling, a reduction in horizontal sampling, a further reduction in density (so that artifacts appear), and a reduced state on an overlapped tree. In the state sketches, the bold pixels appear at all scales, the unbolded pixels appear only at finer scales. The running times, in seconds, are for the entire process - model building, estimation, estimation error variances, and posterior sampling. For comparison, brute-force posterior sampling on a 2GHz workstation would require approximately 28 hours.

The methods of this paper do not, as yet, address all of the challenges raised by the 3D modeling and imaging of porous media. In particular, the two needed extensions to this work are the development of more meaningful prior models, and the extension of the multiscale software to three-dimensional random fields.

6. REFERENCES

[1] M. Bello, A. Willsky, B. Levy, "Construction and applications of discrete-time smoothing error models," *Int. J. of Control* (50), pp.203-223, 1989

[2] M. Daniel, A. Willsky, "The modeling and estimation of statistically self-similar processes in a multiresolution framework," *IEEE Trans. Information Theory* (45) #3 , pp.955–970, 1999

[3] A. Gelfand, S. Hills, A. Racine-Poon, A. Smith, "Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling," *J. American Statistical Association* (85), pp.972–985, 1990

[4] K. Hanson, G. Cunningham, "Posterior Sampling with Improved Efficiency," *SPIE* (3338), pp.371-382, 1998

[5] W. Irving, A. Willsky, "A Canonical Correlations Approach to Multiscale Stochastic Realization," IEEE T. Aut. Cntrl. (46) #10, pp.1514-1528, 2001

[6] M. Luettgen, A. Willsky, "Multiscale Smoothing Error Models", *IEEE T. Aut. Cntrl.* (40) #1, 1995

[7] M. Luettgen, W. Karl, A. Willsky, R. Tenney, "Multiscale Representations of Markov Random Fields", *IEEE T. S.P.* (41) #12, pp.3377–3396, 1993

[8] M. Tanner, W. Wong, "The Calculation of Posterior Distributions by Data Augmentation," *J. American Statistical Association* (82), pp.548–550, 1987