EXPERIMENTS IN BAYESIAN IMAGE ANALYSIS

Stuart Geman [1]
Division of Applied Mathematics
Brown University
Providence, RI 02912/USA

ABSTRACT. We propose a statistical framework for modelling and analyzing pictures. The approach is Bayesian. It is given here in brief outline, with references to various reports and papers for a full discussion and for the results of experiments with a variety of applications.

1. INTRODUCTION

Computational image analysis encompasses a variety of applications involving a sensing device, a computer, and software for restoring and possibly interpreting the sensed data. Most commonly, visible light is sensed by a video camera and converted to an array of measured light intensities, each element corresponding to a small patch in the scene (a picture element, or pixel). The image is thereby digitized, and this format is suitable for computer analysis. In some applications, the sensing mechanism responds to other forms of light, such as in infrared imaging where the camera is tuned to the invisible part of the spectrum neighboring the color red. Infrared light is emitted in proportion to temperature, and thus infrared imaging is suitable for detecting and analyzing the temperature profile of a scene. Applications include automated inspection in industrial settings, medical diagnosis, and targeting and tracking of military objects. In single photon emission tomography, as a diagnostic tool, individual photons, emitted from a radiopharmaceutical (isotope combined with a suitable pharmaceutical) are detected. The objective is to reconstruct the distribution of isotope density inside the body from the externally-collected counts. Depending on the pharmaceutical, the isotope density may correspond to local blood flow (perfusion) or local metabolic activity. Other applications of computer vision include satellite imaging for weather and crop yield prediction, radar imaging in military applications, ultrasonic

[1] Research partially supported by Army Research Office contract DAAL03-86-K-0171 to the Center for Intelligent Control Systems, National Science Foundation grant number DMS-8352087, and the General Motors Research Laboratories.

In: *Proceedings of the Third Valencia International Meeting on BAYESIAN STATISTICS,* 1988 (to appear).

imaging for industrial inspection and a host of medical applications, and there is a growing role for video imaging in robotics.

The variety of applications has yielded an equal variety of algorithms for restoration and interpretation. Unfortunately, few general principals have emerged and no common foundation has been layed. Algorithms are by and large *ad hoc*; they are typically dedicated to a single application, and often critically tuned to the particulars of the environment (lighting, weather conditions, magnification, and so-on) in which they are implemented. It is likely that a coherent theoretical framework would support more robust and more powerful algorithms. A well-studied candidate is regularization theory (see [37],[40], and the similar "variational" approach in [4],[5],[39],[41]), which has been successfully applied to a variety of vision tasks. We have been exploring a related approach based upon probabilistic image models, well-defined principals of inference, and a Monte Carlo computation theory. Exploiting this framework, we have recently obtained encouraging results in several areas of application, including tomography, texture analysis, and scene segmentation.

In the following paragraphs, we lay out, briefly, our paradigm in its general formulation. We refer the reader to various manuscripts for more complete discussions of the methodology, and for applications to texture segmentation and classification ([14],[15],[17], [26]), boundary detection ([8],[11],[15],[16],[36],[37],), single photon emission tomography ([19]), and complex shape modelling and recognition ([28],[33], [35]).

ACKNOWLEDGEMENT. This is a summary of collaborative work, principally with Donald Geman, Basilis Gidas, Ulf Grenander, and Donald E. McClure.

2. BAYESIAN PARADIGM

In real scenes, neighboring pixels typically have similar intensities; boundaries are usually smooth and often straight; textures, although sometimes random locally, define spatially homogeneous regions; and objects, such as grass, tree trunks, branches and leaves, have preferred relations and orientations. Our approach to picture processing is to articulate such regularities mathematically, and then to exploit them in a statistical framework to make inferences. The regularities are rarely deterministic; instead, they describe correlations and likelihoods. This leads us to the Bayesian formulation, in which prior expectations are formally represented by a probability distribution. Thus we design a distribution (a prior) on relevant scene attributes to capture the tendencies and constraints that characterize the scenes of interest. Picture processing is then guided by this prior distribution, which, if properly conceived, enormously limits the plausible restorations and interpretations.

The approach involves five steps, which we shall briefly review here.

IMAGE MODEL. This is a probability distribution on relevant image attributes. Both for reasons of mathematical and computational convenience, we use *Markov random fields* (MRFs) as prior probability distributions. Let us suppose that we index all of the relevant attributes by the index set S. The set S is application specific. It typically includes indices for each of the pixels (about 512×512 in the usual video digitization) and may have other indices for such attributes as boundary elements, texture labels, object labels and so-on. Associated with each site $s \in S$ is a real-valued random variable X_s, representing the state of the corresponding attribute. Thus X_s may be the measured intensity at pixel s (typically, $X_s \in \{0, ...255\}$), or simply 1 or 0 as a boundary element at location s is present or absent.

The kind of knowledge we represent by the prior distribution is usually local, which is to say that we articulate regularities in terms of small local collections of variables. In the end, this leads to a distribution on $X = \{X_s\}_{s \in S}$ with a more or less local neighborhood structure. Specifically, our priors are Markov random fields: there exists a (symmetric) *neighborhood relation* $G = \{G_s\}_{s \in S}$, wherein $G_s \subseteq S$ is the set of neighbors of s, such that

$$\Pi(X_s = x_s | X_r = x_r, r \in S, r \neq s) = \Pi(X_s = x_s | X_r = x_r, r \in G_s)$$

$\Pi(a|b)$ is conditional probability, and, by convention, $s \notin G_s$. G symmetric means $s \in G_r \Leftrightarrow r \in G_s$. (Here, we assume that the range of the random vector X is discrete; there are obvious modifications for the continuous or mixed case.)

It is well known, and very convenient, that a distribution Π defines a MRF on S with neighborhood relation G if and only if it is Gibbs with respect to the same graph, (S, G).

The latter means that Π has the representation

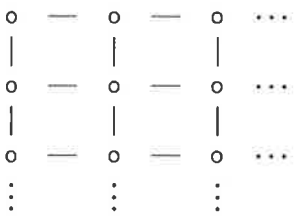
$$\Pi(x) = \frac{1}{z}\exp\{-U(x)\} \tag{2.1}$$

where

$$U(x) = \sum_{c \in C} V_c(x) \tag{2.2}$$

C is the collection of all cliques in (S, G) (a clique is a collection c of sites such that every two sites in c are neighbors), and $V_c(x)$ is a function depending only on $\{x_s\}_{s \in c}$. U is known as the energy, and has the intuitive property that the low energy states are the more likely states under Π. The normalizing constant, z, is known as the partition function. The Gibbs distribution arises in statistical mechanics as the equilibrium distribution of a system with energy function U.

As a simple example (too simple to be of much use for real pictures) suppose the pixel intensities are known, a priori, to be one of two levels, minus one (black) or plus one (white). Let S be the $N \times N$ square lattice, and let G be the neighborhood system that corresponds to nearest horizontal and vertical neighbors:



For picture processing, think of N as typically 512. Suppose that the only relevant regularity is that neighboring pixels tend to have the same intensities. An energy consistent with this regularity is the "Ising potential":

$$U(x) = -\beta \sum_{[s,t]} x_s x_t, \qquad \beta > 0$$

where $\sum_{[s,t]}$ means summation over all neighboring pairs $s, t \in S$. The minimum of U is achieved when $x_s = x_t, \quad \forall s, t \in S$. Under (2.1), the likely pictures are therefore the ones that respect our prior expectations; they segment into regions of constant intensities. This is called the Ising model. It models the equilibrium distribution of the spin states of the atoms in a ferromagnet. Aligned spins cooperate to produce a measureable magnetic field.

Obviously, β is an important parameter. The larger $\beta > 0$, the larger the typical region of constant intensity. Parameters such as β, that determine the detailed quantitative

behavior of the prior, are inevitably introduced in constructing image models. Whenever possible, we *estimate* these parameters from data. This raises a host of interesting computational and theoretical issues; we refer the reader to [2],[10],[12],[17],[19],[22],[24],[30], and [42] for some experiments with, discussions of, and partial solutions to the parameter estimation problem.

One very good reason for using MRF priors is their Gibbs representations. Gibbs distributions are characterized by their energy functions, and these are more convenient and intuitive for modelling than working directly with probabilities. Again, we refer the reader to the references for many more examples, and for applications.

DEGRADATION MODEL. The image model is a distribution $\Pi(\cdot)$ on the vector of image attributes $X = \{X_s\}_{s \in S}$. *By design*, the components of this vector contain all of the relevant information for the image processing task at hand. Hence, the goal is to estimate X. This estimation will be based upon partial or corrupted observations, and based upon the image model, i.e., the prior distribution. In emission tomography, X represents the spatial distribution of isotope in a target region of the body. What is actually observed is a collection of photon counts whose probability law is Poisson, with a mean function that is an attenuated Radon transform of X. In the texture labelling problem, X is the pixel intensity array combined with a corresponding array of texture labels. Each label gives the texture type of the associated pixel. The observation is only partial: we observe the pixels, which are just the digitized picture, but not the labels. The purpose is then to estimate the labels from the picture.

The observations are related to the image process X by a *degradation model.* This models the relation between X and the *observation process*, say $Y = \{Y_s\}_{s \in T}$. For texture analysis, we define $X = (X^P, X^L)$, where X^P is the usual grey-level pixel intensity process, and X^L is an associated array of texture labels. The observed picture is just X^P, and hence $Y = X^P$: the degradation is a projection. More typically, the degradation involves a random component, as in the tomography setting where the observations are Poisson variables whose means are related to the image process X. A simpler, and widely studied (if unrealistic), example is additive white noise. Let $X = \{X_s\}_{s \in S}$ be just the basic pixel process. In this case $T = S$, and for each $s \in S$ we observe

$$Y_s = X_s + \eta_s$$

where, for example, $\{\eta_s\}_{s \in S}$ is Gaussian with independent components, having means 0 and variances σ^2, and $\{\eta_s\}$ is independent of the X-process.

Formally, the degradation model is a conditional probability distribution, or density, for Y given X: $\Pi(y|x)$. If the degradation is just additive white noise, as in the above example, then

$$\Pi(y|x) = \left(\frac{1}{2\pi\sigma^2}\right)^{|S|/2} \exp\{-\frac{1}{2\sigma^2}\sum_{s\in S}(y_s - x_s)^2\}$$

For labelling textures, the degradation is deterministic; $\Pi(y|x)$ is concentrated on $y = x^P$, where $x = (x^P, x^L)$ has both pixel and label components.

POSTERIOR DISTRIBUTION. This is the conditional distribution on the image process X given the observation process Y. This posterior or *a posteriori* distribution contains the information relevant to the image restoration or image analysis task. Given an observation $Y = y$, and assuming the image model ($\Pi(x)$) and degradation model ($\Pi(y|x)$), the posterior distribution reveals the likely and unlikely states of the "true" (unobserved) image X. Having constructed X to contain all relevant image attributes, such as locations of boundaries, labels of objects or textures, and so on, the posterior distribution comes to play the fundamental role in our approach to image processing.

The posterior distribution is easily derived from Bayes' rule

$$\Pi(x|y) = \frac{\Pi(y|x)\Pi(x)}{\Pi(y)}$$

The denominator, $\Pi(y)$, is difficult to evaluate. It derives from the prior and degradation models by integration: $\Pi(y) = \int \Pi(y|x)\Pi(dx)$, but the formula is computationally intractable. Happily, our analysis of the posterior distribution will require only *ratios*, not absolute probabilities. Since y is fixed by observation, $1/\Pi(y)$ is a constant that can be ignored (see paragraph below on Computing).

As an example we consider the simple Ising model prior, with observations corrupted by additive white noise. Then

$$\Pi(x) = \frac{1}{z}\exp\{\beta\sum_{[s,t]} x_s x_t\}$$

and

$$\Pi(y|x) = \left(\frac{1}{2\pi\sigma^2}\right)^{|S|/2} \exp\{-\frac{1}{2\sigma^2}\sum_{s\in S}(y_s - x_s)^2\}$$

The posterior distribution is then

$$\Pi(x|y) = \frac{1}{z_p}\exp\{\beta\sum_{[s,t]} x_s x_t - \frac{1}{2\sigma^2}\sum_{s\in S}(y_s - x_s)^2\}$$

We denote by z_p the normalizing constant for the posterior distribution. Of course, it depends on y, but y is fixed. Notice that the posterior distribution is again a MRF. In the case of additive white noise, the neighborhood system of the posterior distribution is that of the prior, and hence local.

For a wide class of useful degradation models, including combinations of blur, additive or multiplicative colored noise, and a variety of nonlinear transformations, the posterior distribution is a MRF with a more or less local graph structure. This is convenient for our computational schemes, as we shall see shortly. We should note, however, that exceptions occur. Indeed, nonlocal graph structures that incorporate long-range interactions are useful and are completely consistent with the Bayesian paradigm. In tomography, for example, the posterior distribution is associated with a highly non-local graph. This particular situation incurs a high computational cost (see [19]) as a consequence of each site in the graph having a high degree.

ESTIMATING THE IMAGE. In our framework, image processing amounts to choosing a particular image x, given an observation $Y = y$. One choice is the maximum a posteriori, or MAP estimate:

$$\text{choose } x \text{ to maximize } \Pi(x|y)$$

The MAP estimate chooses the most likely x, given the observation. In many applications, our goal is to identify the MAP estimate, or a suitable approximation. Often, though, other estimators are more appropriate (see Besag [3] for an insightful discussion). We have found, for example, that the posterior mean ($\int x\Pi(dx|y)$) is more effective for tomography, at least in our experiments (see [19]).

The computational issues for posterior mean and MAP estimators are similar. For illustration we concentrate here on MAP estimation. In most applications we cannot hope to identify the true maximum a posteriori image vector x. To appreciate the computational difficulty, consider again the Ising model with added white noise:

$$\Pi(x|y) = \frac{1}{z_p} \exp\{\beta \sum_{[s,t]} x_s x_t - \frac{1}{2\sigma^2} \sum_{s \in S} (y_s - x_s)^2\} \tag{2.3}$$

This is to be maximized over all possible vectors $x = \{x_s\}_{s \in S} \in \{-1, 1\}^S$. With $|S| \sim 10^5$, brute force approaches are intractable; instead, we will employ a Monte Carlo algorithm which gives adequate approximations. (Remarkably, this particular example can be solved exactly, using techniques from network flow theory, see [27].)

Maximizing (2.3) amounts to minimizing

$$U_p(x) = -\beta \sum_{[s,t]} x_s x_t + \frac{1}{2\sigma^2} \sum_{s \in S} (y_s - x_s)^2$$

which might be thought of as the posterior energy. (As with z_p, the fixed observation y is suppressed in the notation $U_p(x)$.) More generally, we write the posterior distribution as

$$\frac{1}{z_p} \exp\{-U_p(x)\} \tag{2.4}$$

and characterize the MAP estimator as the solution to the problem

$$\text{choose x to minimize } U_p(x)$$

The utility of this point of view is that it suggests a further analogy to statistical mechanics, and a computation scheme for approximating the MAP estimate, which we shall now describe.

COMPUTING. When exploring a specific application, it is our consistent experience that computation-intensive algorithms for image analysis can be made fast by suitable compromises and exploitation of special structure. Still, as a research tool, it has been invaluable to have available a general computational framework, which we now describe in the context of MAP estimation.

Pretend that (2.4) is the equilibrium Gibbs distribution of a real system. Recall that MAP estimation amounts to finding a minimal energy state. For many physical systems the low-energy states are the most ordered, and these often have desirable properties. The state of silicon suitable for wafer manufacturing, for example, is a low-energy state. Physical chemists achieve low-energy states by heating and then slowly cooling a substance. This procedure is called *annealing*. Cerný [7] and Kirkpatrick [34] suggest searching for good minimizers of $U(\cdot)$ by *simulating* the dynamics of annealing, with U playing the role of energy for an (imagined) physical system. In our image processing experiments, we often use simulated annealing to find an approximation to the MAP estimator. (See Arts and van Laarhoven [1] for a good review of simulated annealing.)

Dynamics are simulated by producing a Markov chain, $X(1), X(2), \ldots$ with transition probabilities chosen so that the equilibrium distribution is the posterior (Gibbs) distribution (2.4). One way to do this is with the Metropolis algorithm [38]. More convenient for image processing is a variation we call *stochastic relaxation*. The full story can be found in [13],[16], and [28]. Briefly, in stochastic relaxation we choose a sequence of sites $s(1), s(2), \ldots \in S$ such

that each site in S is visited infinitely often. If $X(t) = x$, say, then $X_r(t+1) = x_r$, $\forall r \neq s(t), r \in S$, and $X_{s(t)}(t+1)$ is a sample from

$$\Pi(X_{s(t)} = \cdot | X_r = x_r, r \neq s(t)),$$

the conditional distribution on $X_{s(t)}$, given $X_r = x_r$, $\forall r \neq s(t)$. By the Markov property,

$$\Pi(X_{s(t)} = \cdot | X_r = x_r, r \neq s(t)) = \Pi(X_{s(t)} = \cdot | X_r = x_r, r \in G^p_{s(t)})$$

where $\{G^p_s\}_{s \in S}$ is the *posterior* neighborhood system, determined by the posterior energy $U_p(\cdot)$. The prior distributions that we have experimented with have mostly had local neighborhood systems, and usually the posterior neighborhood system is also more or less local as well. This means that $|G^p_{s(t)}|$ is small, and this makes it relatively easy to generate, Monte Carlo, $X(t+1)$ from $X(t)$. In fact, if Ω is the range of $X_{s(t)}$, then

$$\Pi(X_{s(t)} = \alpha | X_r = x_r, r \in G^p_{s(t)}) = \frac{\Pi(\alpha, {}_{s(t)}x)}{\sum_{\hat{\alpha} \in \Omega} \Pi(\hat{\alpha}, {}_{s(t)}x)} \tag{2.5}$$

where

$$(\alpha, {}_{s(t)}x)_r = \begin{cases} \alpha & r = s(t) \\ x_r & r \neq s(t) \end{cases}$$

Notice that (fortunately!) there is no need to compute the posterior partition function z_p. Also, the expression on the right-hand side of (2.5) involves only those potential terms associated with cliques containing $s(t)$, since all other terms are the same in the numerator and the denominator.

To simulate annealing, we introduce an artificial temperature into the posterior distribution:

$$\Pi_T(x) = \frac{\exp\{-U_p(x)/T\}}{Z_p(T)}$$

As $T \to 0, \Pi_T(\cdot)$ concentrates on low energy states of U_p. To actually find these states, we run the stochastic relaxation algorithm while slowly lowering the temperature. Thus $T = T(t)$, and $T(t) \downarrow 0$. $\Pi_{T(t)}(\cdot)$ replaces $\Pi(\cdot)$ in computing the transition $X(t) \to X(t+1)$. In [16] we showed that, under suitable hypotheses on the sequence of site visits, $s(1), s(2), \ldots$:

If $T(t) > c/(1 + \log(1+t))$, and $T(t) \downarrow 0$, then for all c sufficiently large $X(t)$ converges weakly to the distribution concentrating uniformly on $\{x : U(x) = min_y U(y)\}$.

More recently, our theorem has been improved upon by many authors. In particular, the smallest constant c which guarantees convergence of the annealing algorithm to a global minimum can be specified in terms of the energy function U_p (see [9],[20],[29], and

[31]). Also, see Brandt [6] and Gidas [23],[25] for some ideas about faster annealing via multiresolution methods, Geman & Hwang [18], Hwang & Sheu [32], and Gidas [21] for some extensions to continuous-time continuous-state annealing, and see Geman & Geman [13] for an annealing algorithm designed for *constrained* optimization.

REFERENCES

1. E. Arts and P. van Laarhoven, "Simulated annealing: a pedestrian review of the theory and some applications," in *NATO Advanced Study Institute on Pattern Recognition: Theory and Applications,* Spa, Belgium, 1986.

2. J. Besag, "Spatial interaction and the statistical analysis of lattice systems," J. of Royal Statistical Society, 36, pp. 192-236, 1974.

3. J. Besag, "On the statistical analysis of dirty pictures," (with discussion), *J. Royal Statist. Soc.,* series B, 48, pp. 259-302, 1986.

4. A. Blake, "The least disturbance principle and weak constraints," *Pattern Recognition Letters,* 1, pp. 393-399, 1983.

5. A. Blake and A. Zisserman, "Weak continuity constraints in computer vision," Report CSR-197-86, Dept. of Comp. Sci., Edinburgh Univ., 1986.

6. A. Brandt, "Multi-level approaches to large scale problems," in *Proceedings of the International Congress of Mathematicians 1986.* Ed. A. M. Gleason, American Mathematical Society, Providence, 1987.

7. V. Cerný, "A thermodynamical approach to the travelling salesman problem: an efficient simulation algorithm," Inst. Phys. and Biophysics, Comenius Univ., Bratislava, 1982.

8. B. Chalmond, "Image restoration using an estimated Markov model," Dept. of Mathematics, University of Paris, Orsay, 1987.

9. T.-S. Chiang and Y. Chow, "On eigenvalues and optimal annealing rate," Institute of Mathematics, Academia Sinica, Taipei, Taiwan, preprint 1987.

10. P.A. Devijver and M.M. Dekesel, "Learning the parameters of a hidden Markov random field image model: a simple example," in *Pattern Recognition Theory and Applications,* P.A. Devijver and J. Kittler, Eds., Heidelberg: Springer-Verlag, pp. 141-163, 1987.

11. D. Geman, "A stochastic model for boundary detection," *Image and Vision Computing,* May, 1987.

12. D. Geman, "Parameter estimation for Markov random fields with hidden variables and experiments with the EM algorithm," *Complex Systems Technical Report No. 21,* Div. of Applied Mathematics, Brown University, 1984.

13. D. Geman and S. Geman, "Relaxation and annealing with constraints," *Complex Systems Technical Report No. 35,* Div. of Applied Mathematics, Brown University, 1987.

14. D. Geman, S. Geman, and C. Graffigne, "Locating texture and object boundaries," in *Pattern Recognition Theory and Applications,* P.A. Devijver and J. Kittler, Eds., Heidelberg: Springer-Verlag, 1987.

15. D. Geman, S. Geman, C. Graffigne, and P. Dong, "Boundary detection by constrained optimization," preprint, Division of Applied Mathematics, Brown University, Providence, RI, 1988.

16. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern anal. Machine Intell.,* 6, pp. 721-741, 1984.

17. S. Geman and C. Graffigne, "Markov random field image models and their applications to computer vision," in: *Proceedings of the International Congress of Mathematicians, 1986.* Ed. A.M. Gleason, American Mathematical Society, Providence, 1987.

18. S. Geman and C.-R. Hwang, "Diffusions for global optimization," SIAM J. Control and Optimization, 24, pp. 1031-1043, 1986.

19. S. Geman and D.E. McClure, "Statistical methods for tomographic image reconstruction," in: *Proceedings of the 46th Session of the International Statistical Institute,* Bulletin of the ISI, Vol. 52, 1987.

20. B.Gidas, "Nonstationary Markov chains and convergence of the annealing algorithm," *J. Stat. Phys.,* 39, pp. 73-131, 1985.

21. B. Gidas, "Global minimization via the Langevin equation," Proceedings of 24th conference on Decision and Control, Ft. Lauderdale, Florida, pp. 774-778, December 1985.

22. B. Gidas, "Parameter estimation for Gibbs distributions," preprint, Division of Applied Mathematics, Brown University, Providence, RI, 1987.

23. B. Gidas, "A multilevel-multiresolution technique for computer vision via renormalization group ideas," Proceedings of SPIE-International Society for Optical Engineering, Vol. 880, High Speed Computing, pp. 214-218, 1988.

24. B. Gidas, "Consistency of maximum likelihood and pseudo-likelihood estimators for Gibbs distributions," Proceedings of the Workshop on Stochastic Differential Systems with Applications in Electrical/Computer Engineering, Control Theory, and Operations Research, Institute for Mathematics and its Applications, University of Minnesota, Springer,

1988.

25. B. Gidas, "A renormalization group approach to image processing problems," *IEEE Trans. Pattern Anal. Machine Intell.,* (to appear, 1989).

26. C. Graffigne, "Experiments in texture analysis and segmentation," Ph.D. Dissertation, Div. of Applied Mathematics, Brown University, 1987.

27. D.M. Greig, B.T. Porteous, and A.H. Seheult, Discussion: "On the statistical analysis of dirty pictures" by Julian Besag, *J. of the Royal Stat. Society,* B-48, pp. 259-302, 1986.

28. U. Grenander, "Tutorial in Pattern Theory," Lecture Notes Volume, Div. of Applied Mathematics, Brown University, 1984.

29. B. Hajek, "A Tutorial Survey of Theory and Applications of Simulated Annealing," *Proceedings of the 24th IEEE Conference on Decision and Control,* pp. 755-760, 1985.

30. G.E. Hinton and T.J. Sejnowski, "Optimal perceptual inference," *Proc. IEEE Conf. Comput. Vision Pattern Recognition,* 1983.

31. R. Holley and D. Stroock, "Simulated annealing via Sobolev inequalities," preprint, 1987.

32. C.-R. Hwang and S.-J. Sheu, "Large time behaviors of perturbed diffusion Markov processes with applications," I, II, and III. Institute of Mathematics, Academia Sinica, Taipei, Taiwan, preprint 1987.

33. D.M. Keenan and U. Grenander, "On the shape of plane images," Reports in Pattern Analysis No. 145, Division of Applied Mathematics, Brown University, Providence, RI, 1986.

34. S. Kirkpatrick, C.D. Gellatt, Jr. and M.P. Vecchi, "Optimization by simulated annealing," *Science,* 220, pp. 671-680, 1983.

35. A. Knoerr, "Modelling biological variation: an application to leaf recognition," Ph.D. Dissertation, Division of Applied Mathematics, Brown Unviersity, Providence, RI, 1988.

36. J.L. Marroquin, "Surface reconstruction preserving discontinuities," *Artificial Intell. Lab. Memo* 792, M.I.T., 1984.

37. J.L. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *J. Am. Stat. Assoc.,* Vol. 82, pp. 76-89, 1987.

38. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *J. Chem. Phys.,* Vol. 21, pp. 1087-1091, 1953.

39. D. Mumford and J. Shah, "Boundary detection by minimizing functionals, I," preprint, 1986.

40. T. Poggio, V. Torre and C. Koch, "Computational vision and regularization theory," *Nature,* 317, pp. 314-319, 1985.

41. D. Terzopoulos, "Regularization of inverse visual problems involving discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.,* 8, pp. 413-424, 1986.

42. L. Younes, "Estimation and annealing for Gibbsian fields," Tech. Rept., University of Paris, Dept. of Mathematics, Orsay, 1987.