



Probabilistic Solution of Ill-Posed Problems in Computational Vision

J. Marroquin; S. Mitter; T. Poggio

Journal of the American Statistical Association, Vol. 82, No. 397. (Mar., 1987), pp. 76-89.

Stable URL:

<http://links.jstor.org/sici?sici=0162-1459%28198703%2982%3A397%3C76%3APSOIPI%3E2.0.CO%3B2-D>

Journal of the American Statistical Association is currently published by American Statistical Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/astata.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

Probabilistic Solution of Ill-Posed Problems in Computational Vision

J. MARROQUIN, S. MITTER, and T. POGGIO*

Computational vision is a set of inverse problems. We review standard regularization theory, discuss its limitations, and present new stochastic (in particular, Bayesian) methods for their solution. We derive efficient algorithms and describe parallel implementations on digital parallel SIMD architectures, as well as a new class of parallel hybrid computers.

KEY WORDS: Computer vision; Markov random fields; Parallel processing.

1. INTRODUCTION

1.1 Computational Vision

Computational vision denotes a new field in artificial intelligence that has developed in the last 15 years. Its two main goals are to develop image understanding systems that automatically could provide scene descriptions from real images and to understand biological vision. Its main focus is on theoretical studies of vision, considered as an information processing task.

Since at least the work of David Marr (Marr 1982; see also Marr and Poggio 1977), it has been customary to consider vision as an information processing system that could be divided into several modules at different theoretical levels, at least as a first approximation. In particular, Marr suggested that the goal of the first step of vision is to obtain descriptions of physical properties of three-dimensional surfaces around the viewer, such as distance, orientation, texture, and reflectance. This first step of vision, up to what has been called $2\frac{1}{2}$ -D sketch or *intrinsic images*, is mainly bottom-up relying on general knowledge but no special high-level information about the scene to be analyzed.

The first part of vision—from images to surfaces—has been termed *early vision*. Although this point of view has been embraced widely (see a set of recent reviews, e.g., Barrow and Tennenbaum 1981; Brady 1982; Brown 1984; Poggio 1984), it is important to observe that its correctness is still to be proven. In particular, it is still unclear what the nature of the $2\frac{1}{2}$ -D sketch representation is, how different visual modules interact, how their output is fused, and what the role of high-level knowledge on early visual processes is. The critical problem of the organization of

vision and of the control of the flow of information from the different modules and how high-level knowledge is used is still very much an open problem.

In this article, we do not consider this larger issue. Our point of view is that a rigorous analysis of individual modules of vision is bound to play an important role in any full theory of vision.

1.2 Early Vision

Early vision consists of a set of processes that recover physical properties of visible three-dimensional surfaces from the two-dimensional images. Computational, biological, and epistemological arguments (see Marr and Poggio 1976) suggest that early vision processes are generic ones that correspond to conceptually independent modules that can be studied, at least to a very first approximation, in isolation. Some examples of early vision modules are edge detection, spatio-temporal interpolation and approximation, computation of optical flow, computation of lightness and albedo, shape from contours, shape from texture, shape from shading, binocular stereo, structure from motion, structure from stereo, surface reconstruction, and computation of surface color.

The standard definition of computational vision is that it is inverse optics. The direct problem—the problem of classical optics or of computer graphics—is to determine the images of three-dimensional objects. Computational vision is confronted with inverse problems of recovering surfaces from images. Much information is lost during the imaging process that projects a three-dimensional world into two-dimensional arrays (images). As a consequence, vision must rely on natural constraints, that is, general assumptions about the physical world, to derive an unambiguous output. This is typical of many inverse problems in mathematics and physics.

In fact, the common characteristics of most early vision problems, in a sense their deep structure, can be formalized: *early vision problems are ill posed in the sense defined by Hadamard* (1923). A problem is well posed when its solution (a) exists, (b) is unique, and (c) depends continuously on the initial data. Ill-posed problems fail to satisfy one or more of these criteria.

Bertero, Poggio, and Torre (1986) show precisely the mathematically ill-posed structure of several problems listed in Table 1 (see also Poggio and Torre 1984). The recognition that early vision problems are ill posed suggests immediately the use of regularization methods developed in mathematics and mathematical physics for solving the ill-posed problems of early vision (Poggio and Torre 1984).

* J. Marroquin is Professor, Centro de Investigación en Matemáticas, Apdo. Postal 402, Guanajuato, Gto., México. S. Mitter is Professor, Department of Electrical Engineering and Computer Science and Laboratory for Information and Decision Systems, and T. Poggio is Professor, Artificial Intelligence Laboratory, both at Massachusetts Institute of Technology, Cambridge, MA 02139. This article is one of several that were organized through the special editorial assistance of J. Michael Steele, Colin Goodall, and Douglas M. Bates. Support for J. Marroquin and S. Mitter was supplied by the Army Research Office under Contract ARO-DAAG29-84-K-0005 and by the Air Force Office of Scientific Research under Contract AFOSR 85-0027. This research was supported in part by Office of Naval Research Contract NR SRO-202 and by a gift from the Artificial Intelligence Center of the Hughes Aircraft Corporation to T. Poggio.

Table 1. Regularization in Early Vision

Problem	Regularization principle
Edge detection	$\int [(Sf - i)^2 + \lambda(f_{xx}^2)] dx$
Optical flow (area based)	$\int [(i_x u + i_y v + i_t)^2 + \lambda(u_x^2 + u_y^2 + v_x^2 + v_y^2)] dx dy$
Optical flow (contour based)	$\int \left[(V \times N - v^N)^2 + \lambda \left(\frac{\partial}{\partial s} V \right)^2 \right] ds$
Surface reconstruction	$\int [(S \times f - d)^2 + \lambda(f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2)] dx dy$
Spatiotemporal approximation	$\int [(Sf - i)^2 + \lambda(\nabla f \times V + f_t)^2] dx dy dt$
Color	$\ y - Az\ ^2 + \lambda\ Pz\ ^2$
Shape from shading	$\int [(E - R(f, g))^2 + \lambda(f_x^2 + f_y^2 + g_x^2 + g_y^2)] dx dy$
Stereo	$\int \{[\nabla^2 G * (L(x, y) - R(x + d(x, y), y))]^2 + \lambda(\nabla d)^2\} dx dy$

1.3 Standard Regularization in Early Vision

The main idea for “solving” ill-posed problems is to restrict the class of admissible solutions by introducing suitable a priori knowledge. In standard regularization methods, due mainly to Tikhonov (Tikhonov and Arsenin 1977), the regularization of the ill-posed problem of finding z from the data y , $Az = y$, requires the choice of norms $\|\cdot\|$ and of a stabilizing functional $\|Pz\|$. In standard regularization theory, A is a linear operator, the norms are quadratic, and P is linear. A method that can be applied follows.

Find z that minimizes

$$\|Az - y\|^2 + \lambda\|Pz\|^2, \quad (1)$$

where λ is a so-called regularization parameter.

In this method, λ controls the compromise between the degree of regularization of a solution and its closeness to the data [the first term in Eq. (1)]. P embeds the physical constraints of the problem. It can be shown for quadratic variational principles that under mild conditions the solution space is convex and a unique solution exists.

Poggio, Torre, and Koch (1985) showed that several problems in early vision can be “solved” by standard regularization techniques. Surface reconstruction, optical flow at each point in the image, optical flow along contours, color, and stereo can be computed by using standard regularization techniques. Variational principles that are not exactly quadratic but have the same form as Equation (1) can be used for other problems in early vision. The main results of Tikhonov can, in fact, be extended to some cases in which the operators A and P are nonlinear, provided that they satisfy certain conditions (Morozov 1984).

Standard regularization methods can be implemented efficiently by parallel architectures of the fine-grain type, such as the Connection Machine (Hillis 1985). Analog networks, either electrical or chemical, can also be a natural way of solving the variational principles dictated by standard regularization theory (Poggio and Koch 1984;

Poggio et al. 1985). A list of the problems that can be regularized by standard regularization theory or slightly nonlinear versions of it is shown in Table 1, together with the associated regularization principle.

1.4 Limitations of Standard Regularization Theory

This new theoretical framework for early vision shows clearly not only the attractions but also the limitations that are intrinsic to the standard Tikhonov form of regularization theory. Standard regularization methods lead to satisfactory solutions of early vision problems but cannot deal effectively and directly with a few general problems, such as *discontinuities* and *fusion of information from multiple modules*.

Standard regularization theory with linear A and P is equivalent to restricting the space of solution to generalized splines, whose order depends on the order of the stabilizer P . This means that in some cases the solution is too smooth and cannot be faithful in locations where discontinuities are present. In optical flow, surface reconstruction, and stereo, discontinuities are in fact not only present but are also the most critical locations for subsequent visual information processing. Standard regularization cannot deal well with another critical problem of vision, the problem of fusing information from different early vision modules. Since the regularizing principles of the standard theory are quadratic, they lead to linear Euler-Lagrange equations. The output of different modules can, therefore, be combined only in a linear way. Terzopoulos (1984; see also Poggio et al. 1985) showed how standard regularization techniques can be used in the presence of discontinuities in the case of surface interpolation. After standard regularization, locations where the solution f originates a large error in the second term of Equation (1) are identified (this needs setting a threshold for the error in smoothness). A second regularization step is then performed by using the location of discontinuities as boundary conditions.

A similar method could be used for fusing information from multiple sources: a regularizing step could be performed, and locations where terms of the type of the first term of Equation (1) give large errors would be identified. A decision step would then follow by setting appropriately various controlling parameters in those locations, thereby weighting in an appropriate way (for instance, vetoing some of) the various contributing processes.

In any case, one would like a more comprehensive and coherent theory capable of dealing directly with the problem of discontinuities and the problem of fusing information. So the challenge for a regularization theory of early vision is to extend it beyond standard regularization methods and their most obvious nonlinear versions.

1.5 Stochastic Route to Regularizing Early Vision

In this article, we will outline a rigorous approach to overcome part of the ill-posedness of vision problems, based on Bayes estimation and Markov random field (MRF) models, that effectively deals with the problems faced by the standard regularization approach. In this approach, the a priori knowledge is represented in terms of an appropriate probability distribution, whereas in standard regularization a priori knowledge leads to restrictions on the solution space. This distribution, together with a probabilistic description of the noise that corrupts the observations, allows one to use Bayes theory to compute the posterior distribution $P_{f|g}$, which represents the likelihood of a solution f given the observations g . In this way, we can solve the reconstruction problem by finding the estimate \hat{f} that either maximizes this a posteriori probability distribution [the so-called maximum a posteriori (MAP) estimate] or minimizes the expected value (with respect to $P_{f|g}$) of an appropriate error function. The class of solutions that can be obtained in this way is much larger than in standard regularization. In particular, we will show under which conditions this new method leads to solutions that are of the standard regularization type (see Sec. 3).

The price to be paid for this increased flexibility is computational complexity. New parallel architectures and possibly hybrid computers of the digital-analog type promise, however, to deal effectively with the computational requirements of the methods proposed here. We will discuss at the end of the article in some detail these new parallel architectures.

2. PROBABILISTIC MODELS

The key to success in the use of this approach is our ability to find a class of stochastic models (i.e., random fields) that have the following characteristics:

1. The probabilistic dependencies between the elements of the field should be local. This condition is necessary if the field is to be used to model surfaces that are only piecewise smooth; besides, if it is satisfied, the reconstruction algorithms are likely to be distributed and thus efficiently implementable in parallel hardware.

2. The class should be rich enough for a wide variety of qualitatively different behaviors to be modeled.

3. The relation between the parameters of the models and the characteristics of the corresponding sample fields should be relatively transparent, so that the models are easy to specify.

4. It should be possible to represent the prior probability distribution P_f explicitly, so that Bayes theory can be applied.

5. It should be possible to specify efficient Monte Carlo procedures, both for generating sample fields from the distribution so that the capability of the model to represent our prior knowledge can be verified and for computing the optimal estimators.

A class of random fields that satisfies these requirements is the class of MRF's on finite lattices (see Wong 1968; Woods 1972). An MRF has the property that the probability distribution of the configurations of the field can always be expressed in the form of a Gibbs distribution,

$$P_f(f) = 1/Z(e^{-1/T_0 U(f)}),$$

where Z is a normalizing constant, T_0 is a parameter (known as the "natural temperature" of the field), and the "energy function" $U(f)$ is of the form

$$U(f) = \sum_C V_C(f),$$

where C ranges over the "cliques" associated with the neighborhood system of the field and the potentials $V_C(f)$ are functions supported on them (a clique is either a single site or a set of sites such that any two sites belonging to it are neighbors of each other).

As an example, the behavior of piecewise constant functions can be modeled using first-order MRF models on a finite lattice L with generalized Ising potentials (Geman and Geman 1984):

$$\begin{aligned} V_C(f_i, f_j) &= -1, & \text{if } |i - j| = 1 \text{ and } f_i = f_j, \\ &= 1, & \text{if } |i - j| = 1 \text{ and } f_i \neq f_j, \\ &= 0, & \text{otherwise;} \end{aligned}$$

$$f_i \in Q_i = \{q_1, \dots, q_M\} \quad \text{for all } i \in L.$$

We will use a free boundary model, so that the neighborhood size for a given site will be: 4, if it is in the interior of the lattice; 3, if it lies at a boundary, but not at a corner; and 2 for the corners.

The Gibbs distribution,

$$P_f(f) = \frac{1}{Z} \exp \left[-\frac{1}{T_0} U_0(f) \right],$$

$$U_0(f) = \sum_{i,j} V(f_i, f_j), \quad (2)$$

defines a one-parameter family of models (indexed by T_0) describing piecewise constant patterns with varying degrees of granularity.

We will assume that the available observations g are obtained from a typical realization f of the field by a degrading operation (such as sampling) followed by corrup-

tion with iid noise (the form of whose distribution is known), so the conditional distribution can be written as

$$P_{g|f}(g; f) = \exp \left[-\alpha \sum_{i \in S} \Phi_i(f, g_i) \right], \quad (3)$$

where $\{\Phi_i\}$ are some known functions and α is a parameter.

The posterior distribution is obtained from Bayes's rule, as follows:

$$P_{f|g}(f; g) = \frac{1}{Z_P} \exp[-U_P(f; g)] \quad (4)$$

with

$$U_P(f; g) = \frac{1}{T_0} U_0(f) + \sum_{i \in S} \Phi(f, g_i). \quad (5)$$

For example, in the case of binary fields ($M = 2$) with the observations taken as the output of a binary symmetric channel (BSC) with error rate ε (Gallager 1968), we have

$$\begin{aligned} P(g_i | f_i) &= (1 - \varepsilon), & \text{for } g_i = f_i, \\ &= \varepsilon, & \text{for } g_i \neq f_i. \end{aligned}$$

The posterior energy reduces to

$$U_P(f; g) = \frac{1}{T_0} \sum_{i,j} V(f_i, f_j) + \alpha \sum_i (1 - \delta(f_i - g_i)), \quad (6)$$

where $f_i \in \{q_1, q_2\}$;

$$\begin{aligned} \delta(a) &= 1, & \text{if } a = 0, \\ &= 0, & \text{otherwise;} \end{aligned} \quad (7)$$

and

$$\alpha = \ln((1 - \varepsilon)/\varepsilon). \quad (8)$$

3. COST FUNCTIONALS

The Bayesian approach to the solution of reconstruction problems has been adopted by several researchers. In most cases, the criterion for selecting the optimal estimate has been the maximization of the posterior probability (the MAP estimate). It has been used, for example, by Geman and Geman (1984) for the restoration of piecewise constant images, by Grenander (1984) for pattern reconstruction, and by Elliot, Derin, Christi, and Geman (1983) and Hansen and Elliot (1982) for the segmentation of textured images [a similar criterion—the maximization of a suitably defined likelihood function—has been used by Cohen and Cooper (1984) for the same purposes].

In some other cases, a performance criterion, such as the minimization of the mean squared error, has been implicitly used for the estimation of particular classes of fields. For example, for continuous-valued fields with exponential autocorrelation functions, corrupted by additive white Gaussian noise, Nahi and Assefi (1972) and Habibi (1972) have used causal linear models and optimal (Kalman) linear filters for solving the reconstruction problem.

The minimization of the expected value of error functionals, however, has not been used as an explicit criterion

for designing optimal estimators in the general case. We will show that this design criterion is in fact more appropriate in our case, for the following reasons:

1. It permits one to adapt the estimator to each particular problem.
2. It is in closer agreement with one's intuitive assessment of the performance of an estimator.
3. It leads to attractive computational schemes.

As an example, we will now propose design criteria for two particular problems: image segmentation and surface reconstruction.

Consider a field f with N elements, each of which can belong to one of a finite set Q_i of classes. Let f_i denote the class to which the i th element belongs. The segmentation problem is to estimate f from a set of observations $\{g_1, \dots, g_p\}$. Note that f_i does not necessarily correspond to the image intensity. It may represent, for example, the texture class for a region in the image (as in Elliot et al. 1983).

A reasonable criterion for the performance of an estimate \hat{f} is the number of elements that are not classified correctly. Therefore, we define the segmentation error e_s as

$$e_s(f, \hat{f}) = \sum_{i=1}^N (1 - \delta(f_i - \hat{f}_i)), \quad f_i, \hat{f}_i \in Q_i. \quad (9)$$

In the case of the reconstruction problem, an estimate \hat{f} should be considered "good" if it is close to f in the ordinary sense, so the total squared error,

$$e_r(f, \hat{f}) = \sum_{i=1}^N (f_i - \hat{f}_i)^2, \quad (10)$$

will be a reasonable measure for its performance.

To derive the optimal estimators with respect to the criteria stated previously, we first present the general result (which can be found, e.g., in Abend 1968), which states that if the posterior marginal distributions for every element of the field are known, the optimal Bayesian estimator with respect to any additive, positive-definite cost functional C may be found by independently minimizing the marginal expected cost for each element.

In more precise terms, we will consider cost functionals $C(f, \hat{f})$ of the form

$$C(f, \hat{f}) = \sum_{i \in L} C_i(f_i, \hat{f}_i)$$

with

$$\begin{aligned} C_i(a, b) &= 0, & \text{if } a = b, \\ &> 0, & \text{if } a \neq b, \quad \text{for all } i. \end{aligned}$$

We will assume that the value of each element f_i of the field f is constrained to belong to some finite set Q_i (the generalization to the case of compact sets is straightforward). The optimal Bayesian estimator \hat{f}^* with respect to the cost functional C is defined as the global minimizer of the expected value of C over all possible f and g . One can

prove that this estimate can be found by minimizing independently the marginal expected cost for each element, that is,

$$\hat{f}_i^* = q : \sum_{r \in Q_i} C_i(r, q) P_i(r | g) \leq \sum_{r \in Q_i} C_i(r, s) P_i(r | g)$$

for all $s \neq q$ and for all $i \in L$,

where $P_i(r | g)$ is the posterior marginal distribution of the element i , as follows:

$$P_i(r | g) = \sum_{f: f_i=r} P_{f|g}(f; g).$$

The optimal estimators for the error criteria defined previously can be easily derived from this result: In the case of the segmentation problem, we get that

$$\hat{f}_i^* = q \in Q_i : P_i(q | g) \geq P_i(s | g) \text{ for all } s \neq q. \quad (11)$$

We will call this estimate the “maximizer of the posterior marginals” (\hat{f}_{MPM}).

For the reconstruction problem, the optimal estimate is

$$\hat{f}_i^* = q \in Q_i : (\bar{f}_i - q)^2 \leq (\bar{f}_i - s)^2 \text{ for all } s \neq q. \quad (12)$$

We will call this estimate the “thresholded posterior mean” (\hat{f}_{TPM}).

The main obstacle for the practical application of these results lies in the formidable computational cost associated with the exact computation of the marginals and the mean of the posterior distribution given by (5), even for lattices of moderate size. In the next section we will present a general distributed procedure that will permit us to approximate these quantities as precisely as we may want.

4. ALGORITHMS

The algorithms that we will propose are based on the use of the Metropolis (Metropolis et al. 1953) or Gibbs sampler (Geman and Geman 1984) schemes to simulate the equilibrium behavior of the coupled MRF described by Equation (5). We recall that the Markov chain generated by these algorithms is regular and their invariant measure is the posterior distribution $P_{f|g}$. The law of large numbers for regular chains (see, e.g., Kemeny and Snell 1960) establishes that the fraction of time that the chain

will spend on a given state f will tend to $P_{f|g}(f; g)$ as the number of steps gets large, independently of the initial state. This means that we can approximate the posterior marginals by

$$P_i(q | g) \approx \frac{1}{n - k} \sum_{t=k}^n \delta(f_t^{(i)} - q) \quad (13)$$

and \bar{f} by

$$\bar{f}_i \approx \frac{1}{n - k} \sum_{t=k}^n f_t^{(i)}, \quad (14)$$

where $f^{(i)}$ is the configuration generated by the Metropolis algorithm at time t and k is the time required for the system to be in thermal equilibrium. From these values, \hat{f}_{MPM} and \hat{f}_{TPM} can be easily computed using (11) and (12).

This procedure is related to the use of simulated annealing (Kirkpatrick, Gelatt, and Vecchi 1983) for finding the global minimum of U_p (i.e., the MAP estimate; see Geman and Geman 1984). In our case, however, we are interested in gathering statistics about the equilibrium behavior of the coupled field at a *fixed temperature* $T = 1$, rather than in finding the ground state of the system. This fact gives our procedure some distinct advantages, as follows:

1. It is difficult to determine in general the descent rate of the temperature (annealing schedule) that will guarantee the convergence of the annealing process in a reasonable time (it usually involves a trial-and-error procedure). Since we are running the Metropolis algorithm at a fixed temperature, this issue becomes irrelevant.
2. Since in our case we are using a Monte Carlo procedure to approximate the values of some integrals, we should expect a nice convergence behavior, in the sense that coarse approximations can be computed very rapidly and then refined to an arbitrary precision [in fact, it can be proved (see Feller 1950) that the expected value of the squared error of the estimates (13) and (14) is inversely proportional to n].

The main disadvantage of this procedure is that in the case of the segmentation problem, a large amount of memory might be required if the number of classes per element m is large [we need to store the $N(m - 1)$ numbers that define the posterior marginals].

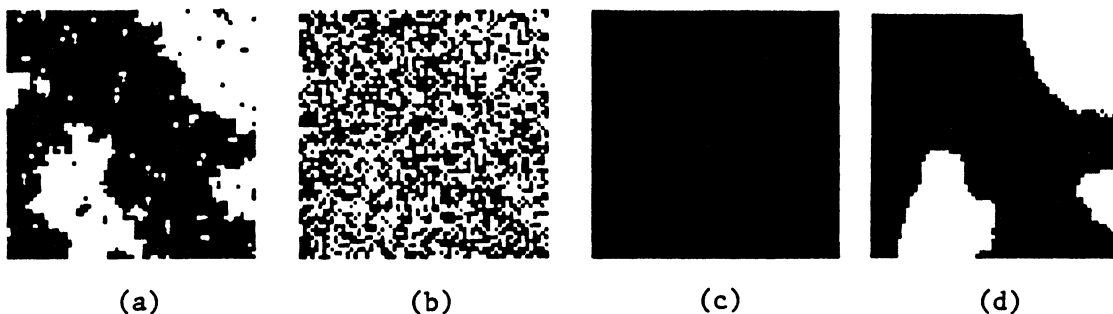


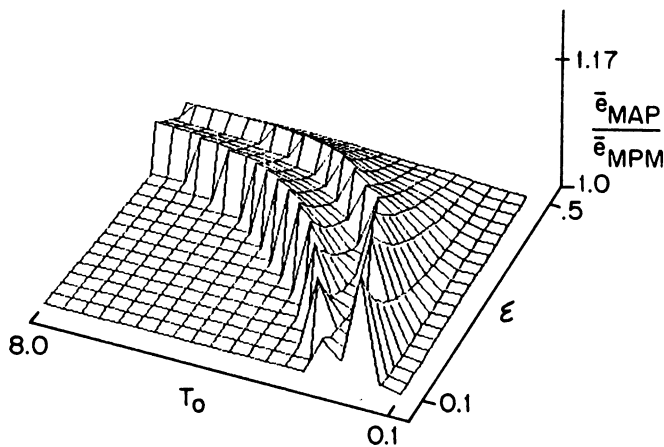
Figure 1. (a) Sample Function of a Binary MRF. (b) Output of a BSC (error rate: .4). (c) MAP Estimate. (d) Monte Carlo Approximation to the MPM estimate.

Table 2. Comparison of the Performance of the MAP and MPM Estimators

	f	g	\hat{f}_{MAP}	$\hat{f}_{MPM(M.C.)}$	$\hat{f}_{MPM(det.)}$
Energy	-5594.8	-226.0	-6660.9	-6460.0	-6427.0
Segmentation error	—	.4	.33	.128	.124

With respect to the relative performance, we point out that in many cases, particularly for high signal-to-noise ratios, the MAP estimate is usually close to the optimal one. If the noise level is high, however, the difference in the performances of the two estimators may be dramatic. This is illustrated in the example portrayed in Figure 1. Panel (a) represents a typical realization of a 64×64 binary Ising net with free boundaries, using a value of $T_0 = 1.74$ (.75 times the critical temperature of the lattice); panel (b) represents the output of a binary symmetric channel with error rate $\varepsilon = .4$; panel (c) represents the MAP estimate; and panel (d) represents an approximation to the MPM estimate [which we will label “MPM (M.C.)”] obtained by using the Metropolis algorithm and Equation (10) to estimate the posterior density. The corresponding values of the posterior energy U_P [Eq. (13)] and the relative segmentation error ($e_s/64^2$) are shown in Table 2. It is clear that the approximation to the MPM estimates shown in panel (d) is better than the MAP from almost any viewpoint.

An intuitive explanation for this behavior comes from the fact that the MAP estimator is implicitly minimizing the expected value of a cost functional $C_{MAP}(f, \hat{f})$, which is equal to zero only if $f_i = \hat{f}_i$ for all i , and is equal to, say, M otherwise. If the signal-to-noise ratio is sufficiently high, the expected value of the optimal segmentation error will be very close to zero, so \hat{f}_{MPM} and \hat{f}_{MAP} will coincide. In a high noise situation, however, the MAP estimator will tend to be too conservative, since from its viewpoint it is equally costly to make one or one thousand mistakes. The MPM estimator, in contrast, can make a better (although more risky) guess, since making a few mistakes has only a marginal effect on the expected cost.

Figure 2. Ratio of the Average Errors of the MAP and MPM Estimators for a 2×2 Ising Net.

A quantitative comparison of the performances of the MAP and MPM estimators, with respect to the segmentation error, can be obtained using the following ratio:

$$r = \frac{\bar{e}_{MAP}}{\bar{e}_{TPM}} = \frac{\sum_{f,g} \exp[-U_P(f;g)] e_s(f, \hat{f}_{MAP}(g))}{\sum_{f,g} \exp[-U_P(f;g)] e_s(f, \hat{f}_{TPM}(g))}.$$

In Figure 2 we show a plot of the ratio r for a 2×2 lattice, for different values of the error rate ε and the natural temperature T_0 . As expected, r is never less than 1. In the worst case (for $\varepsilon = .1$ and $T_0 = .2$) the error of the MAP estimate is 1.17 times that of the MPM estimate; if T_0 is not too small and ε is not too large, both estimates coincide, and as ε approaches .5 (low signal-to-noise ratio), the MPM estimate is consistently better than the MAP. An experimental analysis of larger lattices reveals a similar qualitative behavior, but the values of r are much larger in this case (see Table 2).

5. EXAMPLES OF APPLICATIONS IN VISION

5.1 Reconstruction of Piecewise Constant Functions

The efficient solution of this problem is relevant for several reasons: binary images (or images consisting of only a few gray levels) are directly useful in many interesting applications [e.g., object recognition and manipulation in restricted (industrial) environments]; besides, several perceptual problems, such as the segmentation of textured images (Cohen and Cooper 1984; Cross and Jain 1983; Elliot et al. 1983; Hansen and Elliot 1982) or the formation of perceptual clusters (Marroquin 1985), can be reduced to the problem of reconstructing a piecewise constant surface.

The prior model for this kind of function is given by Equations (1) and (2), and the posterior distribution is given by Equation (4). If the parameters that characterize the system (namely, the “natural temperature” T_0 and the noise parameter α) are known, the MPM estimator produces excellent results, such as the one illustrated in Figure 1.

In most practical cases, however, we are only given the noisy observations g and general qualitative information about the structure of the field and the noise, so f, α (which stands, e.g., for the error rate ε when the noise corruption corresponds to a BSC or for the variance σ^2 in the case of additive Gaussian noise), and T_0 have to be simultaneously estimated.

In principle, one could use again a Bayesian approach and, assuming prior independent uniform distributions for α and T_0 (in the ranges $[\alpha^0, \alpha^1]$ and $[T_0^0, T_0^1]$, respectively),

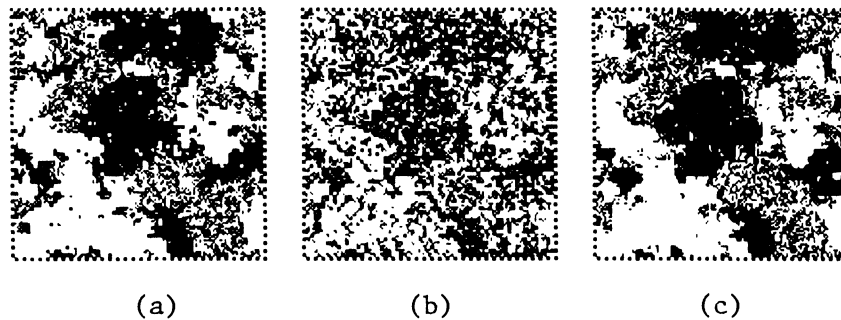


Figure 3. (a) Original Ternary MRF. (b) Noisy Observations (additive Gaussian noise). (c) Optimal (maximum likelihood) Estimate.

find those $\hat{\alpha}$, \hat{T}_0 , and \hat{f} that jointly maximize the posterior distribution:

$$P(f, \alpha, T_0 | g) = \frac{\exp[-U_P(\alpha, T_0, f)]}{(\alpha^1 - \alpha^0)(T_0^1 - T_0^0)Z(T_0)P_g(g)}.$$

The main difficulty here is the extraordinary computational complexity of the partition function

$$Z(T_0) = \sum_f \exp\left[-\frac{1}{T_0} U_0(f)\right],$$

which makes this approach impractical, except for very small lattices.

Another approach, with which we have obtained very good results, consists of defining a merit function for an estimate (obtained by using a particular value for the parameters) that is related to the degree of uniformity in the spatial distribution of the corresponding residuals. We have used, for example, a likelihood function \mathcal{L} , which we obtain by covering the lattice with a set of m nonoverlapping squares (say, 8 pixels wide), computing the relative variance of the noise parameter, estimated over each square, and adding all of these terms together, as follows:

$$\mathcal{L}(\hat{f}) = -\sum_{j=1}^m \left(\frac{\hat{\alpha} - \hat{\alpha}_j}{\hat{\alpha}} \right)^2,$$

where $\hat{\alpha}$ and $\hat{\alpha}_j$ denote the conditional (on \hat{f}) maximum likelihood estimates (MLE's) of the noise parameter, obtained by using the residuals over the whole lattice and over the j th square, respectively. The optimal estimate for f is then obtained as the global maximizer of \mathcal{L} over the appropriate region of the parameter space. An example of the performance of this scheme is presented in Figure

3, which shows the restoration of a ternary pattern corrupted by additive, white Gaussian noise.

Note that this estimation algorithm allows us to reconstruct a pattern f from the noisy observations g without having to adjust any free parameters. The only prior assumptions correspond to the qualitative structure of the field f (first-order, isotropic MRF) and to the nature of the noise process. In practice, this means that we can apply it to restore any piecewise uniform image with uniform granularity, even if it has not been generated by a Markov random process. In the particular case of a binary field sent through a BSC, we have developed a very efficient procedure for approximating the MPM estimator, which also permits us to find the optimal (maximum likelihood) estimate using only a one-dimensional search [see Marroquin (1985) for details]. We have used this algorithm to reconstruct a variety of binary images, with excellent results. In Figure 4 we show such a restoration. The observations (b) were generated from the synthetic image (a) with an actual error rate of .35 (assumed unknown). The MLE for f is shown in (c).

5.2 Reconstruction of Piecewise Continuous Functions

In this section we will illustrate the application of the local spatial interaction models and estimation techniques that we have described to the reconstruction of piecewise continuous functions from noisy observations taken at sparse locations.

In this reconstruction, it will be important not only to interpolate smooth patches over uniform regions, but to locate and preserve the discontinuities that bound these regions, since very often they are the most important parts

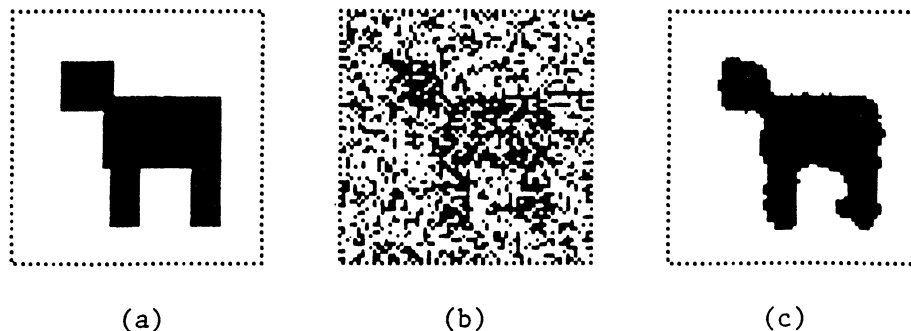


Figure 4. (a) Synthetic Image. (b) Noisy Observations. (c) Maximum Likelihood Estimate.

of the function. They may represent object boundaries in vision problems (such as image segmentation, depth from stereo, shape from shading, structure from motion, etc.), geological faults in geophysical information processing, and so on.

As we mentioned in Section 1.4, an approach to this problem (see Terzopoulos 1984; Grimson 1982) consists of, first, interpolating an everywhere smooth function over the whole domain, then applying some kind of discontinuity detector (followed by a thresholding operation) to try to find the significant boundaries, and, finally, reinterpolating smooth patches over the continuous subregions.

The results that have been obtained with this technique, however, are not completely satisfactory. The main problem is that the task of the discontinuity detector is hindered by the previous smooth interpolation operation. This becomes critical when the observations are sparsely located, since in this case the discontinuities may be smeared in the interpolation phase to such a degree that it may become impossible to recover them in the detection phase.

In contrast, in the Bayesian approach, the boundary detection and interpolation tasks are performed *at the same time*. In applying the general reconstruction algorithms developed previously to this problem, the main issue is the representation, in a meaningful way, of the concept of "piecewise continuity" in the form of a prior Gibbs distribution.

A flexible construction involves the use of two coupled MRF models: one to represent the function (the "surface") itself, and another to model the curves where the field is discontinuous. A coupled model of this kind was first used by Geman and Geman (1984) in the context of the restoration of piecewise constant images. Terzopoulos (in press) recently attempted to translate this idea in the continuous and deterministic framework of standard regularization.

This model can be adapted to our problem by modifying the choice of the potentials and the neighborhood structure of the coupled MRF's. Specifically, the following modifications are needed:

1. Since in our case the observations are sparse, it becomes necessary to expand the size of the neighborhoods of the line field, to prevent the formation of "thick" boundaries between the smooth patches (i.e., adjacent, parallel segments of active lines in these regions). In particular, we propose that the dual lattice be 8-connected, with nonzero potentials for the cliques of the form illustrated in Figure 5(a) and 5(b). The inclusion of the cliques of Figure 5(b) has the additional advantage of penalizing

the occurrence of sharp turns, permitting us to model the formation of piecewise smooth boundaries by using a binary line process instead of the 4-valued process proposed by Geman and Geman (1984). The potentials for these cliques are computed in the following way: Let V_a , V_b denote the potentials associated with the cliques C_a , C_b of Figure 5(a) and 5(b), respectively, and let S_k ($k \in \{a, b\}$) denote the number of line elements belonging to C_k that are "on" at a given time, that is,

$$S_k = \sum_{i \in C_k} l_i, \quad k = a, b.$$

The potentials V_k are given by

$$V_k = \beta \phi_k(S_k), \quad k = a, b,$$

where β is a constant, and the functions ϕ_k are defined by the following:

$$\begin{array}{ccccc} S_a & 0 & 1 & 2 & 3 & 4 \\ \phi_a & 0 & .4 & .25 & 1.2 & 2.0 \\ S_b & 0 & 1 & 2 \\ \phi_b & 0 & 0 & 10 \end{array}$$

It is not difficult to see that this choice of potentials will effectively discourage both the formation of thick boundaries ($S_b = 2$) and the presence of sharp turns ($S_a = 3$ and/or $S_b = 2$).

2. The potentials of the depth process, which is now continuous-valued, have to be modified to express the more relaxed condition of piecewise continuity (instead of piecewise constancy). Specifically, we propose

$$\begin{aligned} V(f_i, f_j, l_{ij}) &= (f_i - f_j)^2(1 - l_{ij}), & \text{for } |i - j| = 1, \\ &= 0, & \text{otherwise} \end{aligned}$$

(note that $l_{ij} \in \{0, 1\}$).

3. Unlike the case of piecewise constant surfaces, we now have to worry about the maximum absolute difference in the values of two adjacent depth sites that we are willing to consider as a "smooth" gradient (and not a discontinuity). This value, which in general is problem dependent, determines the magnitude of the constant β in Equation (2), which can be interpreted as the coupling strength between the two processes.

Assuming that the observations are corrupted by iid Gaussian noise, we get the following expression for the posterior energy:

$$\begin{aligned} U_P(f, l; g) &= \frac{1}{T_0} \sum_{i,j} (f_i - f_j)^2(1 - l_{ij}) \\ &+ \frac{1}{2\sigma^2} \sum_{i \in S} (f_i - g_i)^2 + \sum_{C_a} V_a(l) + \sum_{C_b} V_b(l), \end{aligned}$$

where S is the set of sites where an observation is present. As a performance criterion we will use a mixed cost functional of the form

$$e_m(f, l, \hat{f}, \hat{l}) = \sum_{i \in L_f} (f_i - \hat{f}_i)^2 + \sum_{j \in L_l} (1 - \delta(l_j - \hat{l}_j)),$$



Figure 5. (a) and (b) show different clique types for the line process.

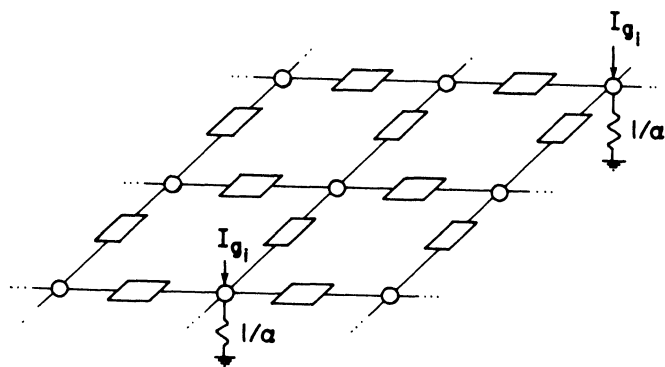


Figure 6. Hybrid Network Implementing the Surface Reconstruction Algorithm of Section 4. The voltage at every node represents the height of the surface. Inside every rectangular box there is a resistance of unit magnitude and a switch whose state is controlled by the corresponding line element.

where L_f , L_l denote the depth and line lattices, respectively. This error criterion means that the reconstructed surface should be as close as possible to the true (unknown) surface and that we should commit as few errors as possible in the assertions about the presence or absence of discontinuities.

Applying the results of Section 3, we find that the optimal estimators will be the *posterior mean* for f and the *maximizer of the posterior marginals* for l .

There is one serious difficulty that prevents us from applying directly the general Monte Carlo procedure that was derived previously to the computation of these optimal estimates: since the depth variables are continuous-valued, if we discretize them finely enough to guarantee sufficient precision of the results, the computational complexity of either the Metropolis or Gibbs sampler algorithms will be very large. One way around this difficulty is to note that for any fixed configuration of the line field, the posterior energy becomes a nonnegative definite quadratic form

$$U(f | l, g) = \sum_{i,j:l_{ij}=0} (f_i - f_j)^2 + \alpha \sum_{j \in S} (f_j - g_j)^2 + K, \quad (15)$$

where α and K are constants (note that the first sum is taken only over those pairs of sites whose connecting line element is "off" and the second one is taken over the set S). This means that the posterior distribution of the depth field is conditionally Gaussian, so for any fixed l we can find the optimal conditional estimator f_l^* as the minimizer of (15).

Let us define the set F^* as

$$F^* = \{(f, l) : f = f_l^*\}.$$

It is clear that, if \hat{f} , \hat{l} are the optimal estimates for our problem, we have that $(\hat{f}, \hat{l}) \in F^*$, which suggests that we can constrain the search for the optimal estimators to this set. This can be done, in principle, by replacing the posterior energy with the function $U^*(l) = U(f_l^*, l)$ (which depends only on l) and using the standard Monte Carlo procedures to find the optimal estimator \hat{l} . To illustrate this idea, let us consider a physical model in the next section.

Hybrid Parallel Computers. It is well known that the steady state of an electrical network that contains only (current or voltage) sources and linear resistors will be the global minimizer of a quadratic functional that corresponds to the total power dissipated as heat (Oster, Perelson, and Katchalsky 1971). It is, therefore, possible to construct an analog network that will find the equilibrium state of the depth field for a given, fixed configuration of the line process, that is, that will minimize the conditional energy (8) (see Poggio and Koch 1984; see also Poggio et al. 1985). This suggests a hybrid computational scheme in which the line field (whose state is updated digitally, using, say, the Metropolis or Gibbs sampler algorithms) acts as a set of switches on the connections between the nodes of the analog network whose voltages represent the depth process. In particular, if f_i represents the voltage at node i , the hybrid network can be represented as a 4-connected lattice of nodes (see Fig. 6) for which the following hold:

1. A resistance (of unit magnitude) and a switch (controlled by the line element l_{ij}) is present in every link between pairs i, j of adjacent nodes.

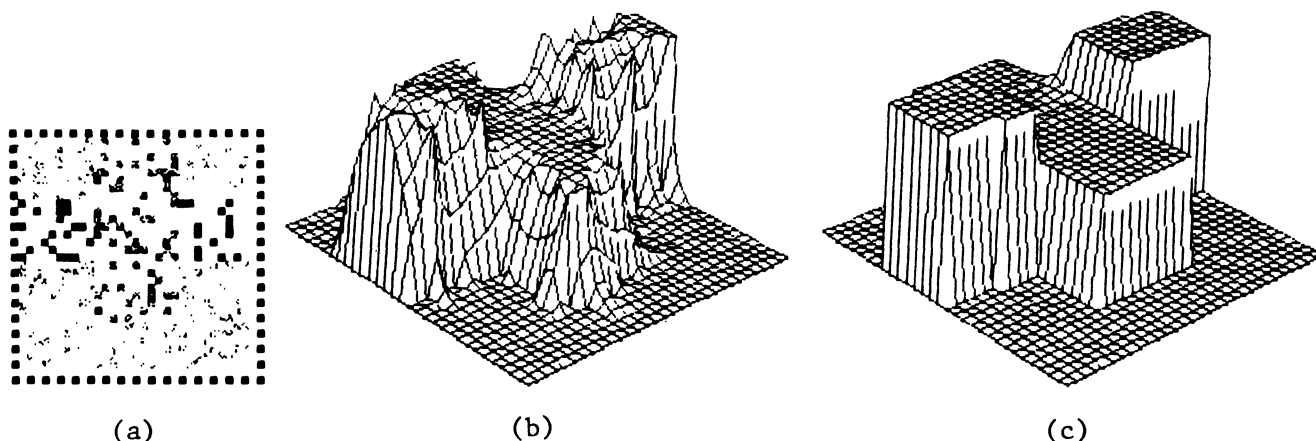


Figure 7. (a) Observations of Three Rectangles at Heights 2.0, 3.0, and 2.0 Over a Background at Height 1.0 (height coded by gray level; a white pixel means that the observation is absent at that point). (b) Equilibrium State of the Network With All Lines Turned "off." (c) Optimal Estimate.

2. If an observation g_i is present at site i , a current of magnitude equal to αg_i is injected to the corresponding node, which must also be connected to a common ground via a resistance of magnitude $1/\alpha$ [see Eq. (8)].

A direct application of Kirchoff current law shows that at each node i of this network we will have

$$\sum_{j \in N_i} (f_i - f_j)(1 - l_{ij}) + \alpha q_i f_i = \alpha q_i g_i,$$

which corresponds to the condition $\text{grad } U(f | l) = 0$, so the equilibrium configuration coincides with f_i^* .

This scheme can be used, in principle, to construct a special-purpose hybrid computer for the fast solution of problems of this type. In a digital machine, the exact implementation of this strategy will be, in general, very expensive computationally, since f_i^* must be computed every time a line site is updated. It is possible, however, to develop approximations that have an excellent experimental performance and lead to efficient implementations (Marroquin 1985). The performance of this method is illustrated in Figure 7, in which we show the observations (with height coded by gray level) (a), the initial state of the network (with all the lines turned "off") (b), and the final reconstructed surface (c), for a square at height 2.0 over a background at constant height = 1.0.

6. SIGNAL MATCHING

In all of the estimation problems we have studied so far, the posterior energy function had the form

$$U_P(f; g) = U_0(f) + \sum_i \Phi_i(f_i, g_i),$$

where $U_0(f)$ corresponded to the MRF model for the field f . The functions Φ_i , whose precise form depended on the particular noise model, were nondecreasing functions of the distance between f_i and g_i .

There are some cases, however, in which the conditional probability distribution of the observations $P_{g|f}(g; f)$ is multimodal (as a function of f), which causes the functions Φ_i to be nonmonotonic, so the solution to the problem remains ambiguous, even if the observations are dense and the signal-to-noise ratio is arbitrarily high. To illustrate this situation, we will study an important instance of it—the "signal-matching" problem, whose one-dimensional version is as follows.

Consider two one-dimensional, real-valued sequences h_L, h_R , where h_L is obtained from h_R by shifting some subintervals according to the "disparity sequence" d , as follows:

$$h_L(i) = h_R(i + d_i)$$

with

$$d_i \in Q = \{-m, -m + 1, \dots, -1, 0, 1, \dots, m\}.$$

The signal-matching problem is to find d given h_L, h_R . (In a more realistic situation, we do not observe h_L, h_R directly, but rather some noise-corrupted versions g_L, g_R).

Some interesting instances of this problem are the matching of stereoscopic images along epipolar lines (Marr and Poggio 1976), the computation of the dip angle of geological structures from electrical resistivity measurements taken along a bore hole, and the matching of DNA sequences.

To make the discussion more specific, we will consider a simple example in which the sequences h_L, h_R are binary Bernoulli sequences; we will assume that the noise corruption process can be modeled as a BSC with known error rate and that d is known to be a piecewise constant function. A well-known instance of this problem is the matching of a row of a random dot stereogram with density ρ (Julesz 1960) when the components of the stereo pair are corrupted by noise.

The stochastic model for the observations is then constructed by assuming that the right image is a sample function of a Bernoulli process A with parameter ρ : $g_R(i) = A(i)$. The left image is assumed to be formed from the right one by shifting it by a variable amount given by the disparity function d , except at some points where an error is committed with probability ε . Note that some regions that appear in the right image will be occluded in the left one (see Fig. 8). The "occlusion indicator" ϕ_d can be computed deterministically from d in the following way:

$$\begin{aligned} \phi_d(i) &= 1, & \text{if } d_{i-k} \geq d_i + k, \text{ for some integer} \\ & & k \in (0, m], \\ &= 0, & \text{otherwise.} \end{aligned} \quad (16)$$

The occluded areas are assumed to be "filled in" by an independent Bernoulli process B . The final model is then

$$\begin{aligned} g_L(i) &= g_R(i + d_i), & \text{with } \Pr(1 - \varepsilon), \text{ if } \phi_d(i) = 0, \\ &= 1 - g_R(i + d_i), & \text{with } \Pr(\varepsilon), \text{ if } \phi_d(i) = 0, \\ &= B_\rho(i), & \text{with } \Pr(1), \text{ if } \phi_d(i) = 1, \end{aligned} \quad (17)$$

Note that in the two-dimensional case, the index i denotes a site of a lattice and, therefore, it can be represented as a two-vector (i_1, i_2) whose components denote the column and row of the site, respectively. To simplify the notation, we will adopt the following convention throughout this

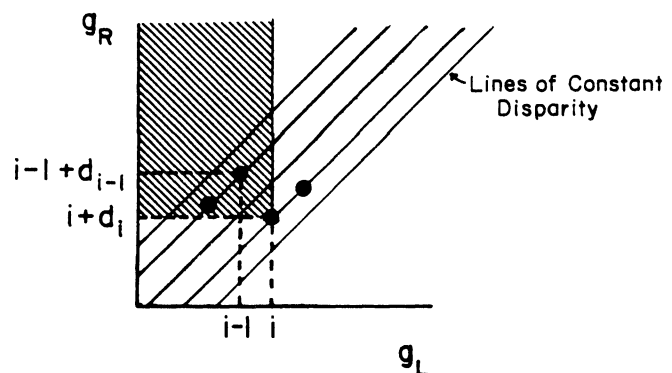


Figure 8. Occluded Regions. The horizontal and vertical axis represent points in one row of the left and right images, respectively. Matching points are represented by black circles. Any match in the shaded region will occlude the point i .

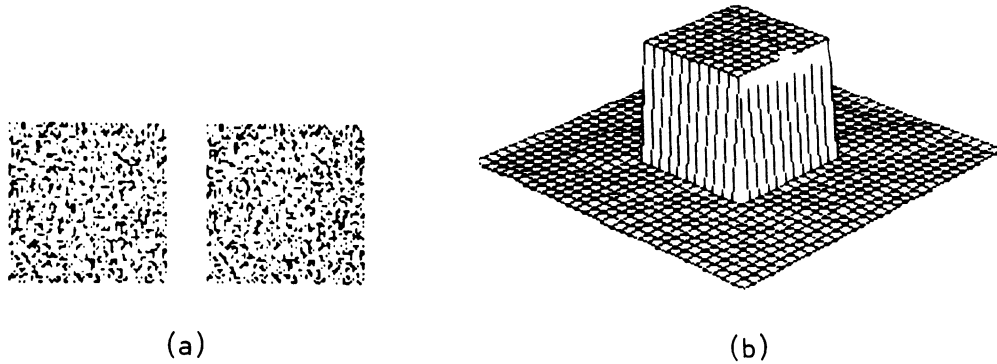


Figure 9. (a) Random Dot Stereogram. (b) Reconstructed Surface.

section: when a scalar is added to this vector index [as in $g_R(i + d_i)$ and d_{i+k}], it will be implicitly assumed that it is multiplied by the vector $(1, 0)$ [so the above expressions should be understood as $g_R(i + (d_i, 0))$ and $d_{i+(k,0)}$, respectively]. Using this convention, the observation model of Equation (17) can be applied either to the one-dimensional or to the two-dimensional cases.

Notice that even if the observations are noise free ($\varepsilon = 0$) the solution of the problem remains ambiguous, and it cannot be uniquely determined unless some prior knowledge about d (e.g., in the form of an MRF model) is introduced. The use of an MRF model in the stereo matching case corresponds to a quantification of the assumption of the existence of “dense solutions” [this term was introduced by Julesz (1960) and essentially corresponds to the assumption that the disparity d varies smoothly in most parts of the image; see also Marr and Poggio (1976)], and the use of the occlusion indicator corresponds to the “ordering constraint” [i.e., the requirement that if $i > j$, then $i + d_i > j + d_j$; we put $\phi_d = 1$ whenever this constraint is violated].

To formulate the estimation problem, we will consider the sequence g_L as “observations,” and g_R will play the role of a set of parameters. Thus, from (17), we have (assuming, for simplicity, that $\rho = \frac{1}{2}$)

$$\begin{aligned}
 P(g_L(i) = k \mid d, g_R) &= P_{g|d}(k) \\
 &= 1 - \varepsilon, & \text{if } \phi_d(i) = 0 \text{ and } g_R(i + d_i) = k, \\
 &= \varepsilon, & \text{if } \phi_d(i) = 0 \text{ and } g_R(i + d_i) \neq k, \\
 &= \frac{1}{2}, & \text{if } \phi_d(i) = 1.
 \end{aligned}$$

As a prior model for the disparity field, we may use a first-order MRF with generalized Ising potentials, such as the one presented in Section 5.1. Other models may also be used, including the coupled depth and line fields that we discussed in the previous section. For the present, let us assume that the simpler Ising model is adequate. Note that even when the matching problem is one-dimensional (we are assuming that there is no vertical disparity between the images, so the matching can be done on a row-by-row basis), the two-dimensional nature of the prior MRF model for the disparity introduces a coupling between matches

at adjacent rows. The posterior energy is

$$\begin{aligned}
 U_P(d; g) &= \frac{1}{T_0} \sum_{i,j} V(d_i, d_j) + \frac{1}{2} \sum_i \phi_d(i) \ln 2 \\
 &\quad + \frac{\alpha}{2} \sum_i (1 - \phi_d(i)) \delta(g_L(i) - g_R(i + d_i)),
 \end{aligned} \tag{18}$$

where $\alpha = \ln(\varepsilon/(1 - \varepsilon))$.

It is possible to apply the general Monte Carlo algorithms presented previously to approximate the optimal estimate \hat{d} with respect to a given performance measure (such as the mean squared error). Their use in this case, however, is complicated by the introduction of the occlusion function ϕ_d in the posterior energy: the size of the support for this function equals the total number of allowed values for the disparity [see Eq. (16)]. If this number is large, the computation of the increment in energy or of the conditional distributions (if the Gibbs sampler is used) may be quite expensive. In many cases, however, the size of the regions of constant disparity is relatively large compared with the size of the occluded areas. In these cases, one can approximate the posterior energy by

$$U_P(d) = \frac{1}{T_0} \sum_{i,j} V(d_i, d_j) + \frac{\alpha}{2} \sum_i \delta(g_L(i) - g_R(i + d_i))$$

and increase significantly the computational efficiency. It is also possible, particularly for the high signal-to-noise ratio case, to design deterministic, highly distributed algorithms for the efficient computation of the optimal estimator. The details of these designs can be found in Marroquin (1985).

To illustrate the performance of this approach, we present in Figure 9 a random dot stereogram portraying a square floating over a uniform background [panel (a)] and the reconstructed surface [panel (b)].

7. PARALLEL IMPLEMENTATIONS

7.1 Connection Machine Architectures

The general Monte Carlo procedure that we have presented for the approximation of the optimal Bayesian estimators of MRF's can be greatly accelerated if it is im-

plemented in a parallel architecture. A necessary condition for the convergence of the probability measures of the Markov chains defined by the Metropolis or Gibbs sampler algorithms to the posterior Gibbs distribution [and, therefore, for the convergence of the approximations given by Eqs. (13) and (14) to the desired estimates] is that if two sites belong to the same clique, they are never updated at the same time. It is important to note, however, that this condition is also sufficient only for the case of the Gibbs sampler: if one updates simultaneously the states of all nonneighboring sites, the reversibility of the resulting chain will be destroyed, so it will no longer be possible to guarantee the convergence of the Metropolis algorithm to the desired result (see Marroquin 1985).

If one implements the Gibbs sampler in a parallel architecture in which a processor is assigned to each site, the total execution time will be reduced by a factor of N/K , where K is the so-called "chromatic number" of the graph that describes the neighborhood structure and it is equal to the minimum number of colors needed to color the sites of the lattice in such a way that no two neighbors are the same.

An example of such a massively parallel architecture is the "Connection Machine" (Hillis 1985), constructed by Thinking Machines Corporation, Cambridge, Massachusetts. This machine was originally designed for the parallel processing of structured symbolic expressions, such as frames and semantic networks. It is a "Single Instruction Multiple Data" (SIMD) array processor consisting of 256,000 processing units (each with a single-bit arithmetic/logical unit and about 4K bits of storage) organized in a 4-connected lattice that is 512 elements square. In addition to this nearest-neighbor connectivity, it will be possible (although computationally more expensive) to connect any two processors in the array using a "Cross Omega" router network.

At each cycle of the machine, for which we will assume a duration of 1 microsecond, an instruction is executed by each processor and a single bit is transmitted to its neighbors. This means that the updating scheme can be implemented most efficiently if the field is first-order Markov, but higher-order processes can also be implemented without using the router by successively propagating the transmitted state (the execution time, therefore, will grow linearly with the order of the field).

To make this discussion more concrete, consider, as an example, the problem of finding the optimal estimate for a first-order MRF, where the random variable at each lattice site takes values on a set of cardinality M , with Ising potentials (i.e., the segmentation of a piecewise constant image) from noisy observations. Let us assume that the estimator is to be implemented in the "Connection Machine" and suppose that by the use of appropriate scaling factors all of the numbers can be represented as 16-bit integers. We will use the following conservative assumptions: 16 cycles of a single 1-bit processor are needed to perform 16-bit addition, subtraction, or comparison; 16^2 cycles are needed to perform multiplication or division;

2×16^2 cycles are needed for generating a pseudorandom number with uniform distribution on a given interval; 16 cycles are needed for memory transfer operations; and 6×16^2 cycles are needed for computing an exponential.

Assuming that we run 250 iterations of the system and ignoring the overhead time, we get that

$$\text{execution time} \approx 1.4(M - 1) \text{ seconds.}$$

For the particular case of binary images, we have developed a deterministic scheme for which this execution time can be reduced by an order of magnitude (see Marroquin 1985).

In the case of the reconstruction of piecewise smooth functions from sparse data, the optimal estimator can also be implemented in this machine. To study this implementation, we first note that the chromatic numbers of the graphs associated with the line and depth neighborhood systems are 4 and 2, respectively, which means that the coupled process has a chromatic number of 6. In Figure 10(a) we illustrate one possible "coloring."

The colors of the line process are represented by the numbers 1, 2, 3, 4, and those of the depth process are represented by white and black circles. The updating process can be implemented in a 4-connected architecture such as the "Connection Machine" by assigning one processor to each depth site and its four adjacent line elements. We will thus have two different populations of processors, whose configurations are shown in Figures 10(b) and 10(c), respectively.

Each complete iteration consists of six major cycles: in the first two, the state of the white and black depth variables is updated, respectively, and in the next four, the new states of the binary line variables stored in (say) the white processors are successively computed and transmitted to the corresponding memory locations of the neighboring black processors. Note that in this scheme we have some redundancy in the use of memory (each binary variable is stored twice), but the state of all of the elements needed for each updating operation is always available from adjacent processors. Considering that the Monte Carlo

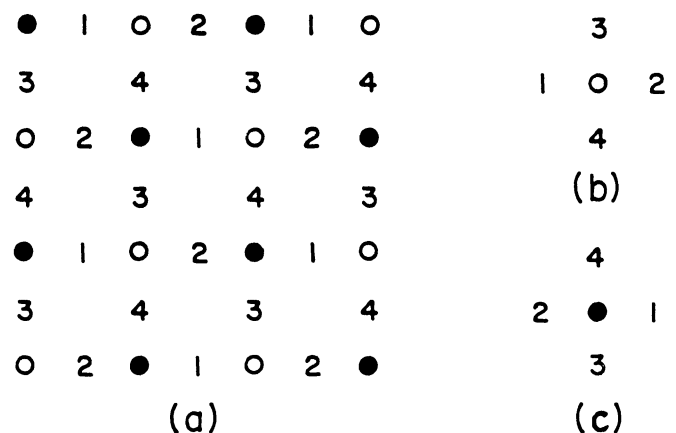


Figure 10. (a) Coloring of the Coupled Line-Depth Lattice. (b), (c) Elements Whose State Is Stored in Each of the Two Types of Processors of a 4-Connected Parallel Architecture.

algorithm requires about 200 iterations to converge, we estimate in this case an execution time of approximately 2.5 seconds, independent of the lattice size. As before, we have also developed in this case a deterministic scheme with very good experimental performance, for which the execution time can be reduced by at least an order of magnitude.

7.2 Hybrid Analog-Digital Computers and Hopfield Networks

As we mentioned in Section 5.2, the reconstruction of piecewise continuous functions can be achieved by coupling two MRF's, one corresponding to the continuous field and the other to the discontinuities. From this scheme we have suggested a special-purpose parallel computer consisting of an analog network of resistances (corresponding to the continuous intensity field) and a digital network (corresponding to the line process), coupled via D-A and A-D converters. The idea suggested by computer experiments (Marroquin 1985) is that the two processes can run on different time scales, a slow one for the digital part and a fast one for the analog network. In this way the two processes are effectively decoupled and the continuous field finds its equilibrium effective instantaneously after each update of the line process. Koch, Marroquin, and Yuille (1985) discussed implementations of this idea. This idea can be extended to *multilayered hybrid networks*, each layer corresponding to an MRF and being digital or analog depending on the continuous or binary nature of the field. Hybrid multilayered architectures of this type are especially attractive for implementing the fusion of several vision processes.

Finally, we mention that Koch et al. (1985) have been experimenting successfully with a special type of analog networks—Hopfield networks—whose equilibrium states correspond to approximations of the optimal estimators.

8. CONCLUSION

In this article we have presented a probabilistic approach to the solution of a class of perceptual problems. We showed that these problems can be reduced to the reconstruction of a function on a finite lattice from a set of degraded observations, and we derived the Bayesian estimators that provide an optimal solution. We also developed efficient distributed algorithms for the computation of these estimates and discussed their implementation in different kinds of hardware. To demonstrate the generality and practical value of this approach, we studied in detail several applications: the segmentation of noise-corrupted images, the reconstruction of piecewise smooth surfaces from sparse data, and the reconstruction of depth from stereoscopic measurements.

8.1 Connection With Standard Regularization

The MAP estimate of an MRF is obviously similar to a variational principle of the general form of Equation (3), since the use of this criterion defines the optimal estimator as the global minimizer of the posterior energy U_P [Eq.

(6)]: the first term measures the discrepancy between the data and the solution, and the second term is now an arbitrary “potential” function of the solution (defined on a discrete lattice). It is then natural to ask for the connection between standard regularization principles and the MRF approach. It turns out that an MAP estimate leads to the minimization of a functional U_P —in general not quadratic—that reduces to a quadratic functional, of the standard regularization type, when the MRF is continuous-valued, the noise is additive and Gaussian [the term $\sum \Phi_i(f, g_i)$ will be quadratic], and first-order differences of the field are zero-mean, independent, Gaussian random variables [thus the a priori probability distribution is a Gibbs distribution with quadratic potentials, so the term $U_0(f)$ is quadratic].

8.2 The Fusion Problem

This approach also permits, in principle, the incorporation of more than one modality of observations into a single estimation process, as well as the simultaneous estimation of several related functions from the same data set. This makes one hope that this framework could be useful in the solution of difficult problems that require such an integrated approach.

For instance, the stereo-matching problem in real situations has not been solved yet in a completely satisfactory way. The same can be said of other related perceptual problems, such as edge detection, image segmentation, the recovery of the shape of an object from a single two-dimensional image (the “shape from shading” problem), and the segmentation of a scene into distinct objects, as well as the recovery of their three-dimensional structure from the analysis of images formed at successive instants of time (the “structure from motion” problem). All of these problems are obviously related, and it is intuitively clear that the individual solutions that can be obtained should improve if the mutual constraints that the solution of each individual problem imposes on the others are taken into account. Thus the presence of a brightness edge should increase the likelihood of a depth edge and vice versa, the depth estimated from stereo should be compatible with the shape derived from shading, points belonging to the same region in an image should move together, and so on. We believe that these constraints can be incorporated in the potential functions of the corresponding MRF models, so the combined optimal estimation process represents, in fact, an integrated cooperative solution to these problems, with, it is hoped, a significantly improved performance.

[Received October 1985.]

REFERENCES

- Abend, K. (1968), “Compound Decision Procedures for Unknown Distributions and for Dependent States of Nature,” in *Pattern Recognition*, ed. L. Kanal, Washington, DC: Thompson Book Co., pp. 207–249.
- Barrow, H. G., and Tennenbaum, J. M. (1981), “Interpreting Line Drawings as Three Dimensional Surfaces,” *Artificial Intelligence*, 17, 75–117.

- Bertero, M., Poggio, T., and Torre, V. (1986), "Ill-Posed Problems in Early Vision," *Artificial Intelligence Laboratory Memo*, No. 924, Cambridge, MA: MIT.
- Brady, J. M. (1982), "Computational Approaches to Image Understanding," *Computing Surveys*, 14, 3-71.
- Brown, C. M. (1984), "Computer Vision and Natural Constraints," *Science*, 224, 1299-1305.
- Cohen, F. S., and Cooper, D. B. (1984), "Simple Parallel Hierarchical and Relaxation Algorithms for Segmenting Noncausal Markovian Random Fields," Technical Report LEMS-7, Brown University Laboratory for Engineering Man/Machine Systems.
- Cross, G. C., and Jain, A. K. (1983), "Markov Random Field Texture Models," *IEEE Transactions on PAMI*, 5, 25-39.
- Elliot, H., Derin, R., Christi, R., and Geman, D. (1983), "Application of the Gibbs Distribution to Image Segmentation," technical report, University of Massachusetts.
- Feller, W. (1950), *An Introduction to Probability Theory and Its Applications* (Vol. 1), New York: John Wiley.
- Gallager, R. G. (1968), "Information Theory and Reliable Communication," New York: John Wiley.
- Geman, S., and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.
- Grenander, U. (1984), "Tutorial in Pattern Theory," unpublished manuscript, Brown University, Division of Applied Math.
- Grimson W. E. L. (1982), "A Computational Theory of Visual Surface Interpolation," *Philosophical Transactions of the Royal Society of London*, Ser. B, 298, 395-427.
- Habibi, A. (1972), "Two Dimensional Bayesian Estimation of Images," *Proceedings of the Institute of Electrical and Electronics Engineers*, 60, 878-883.
- Hadarnard, J. (1923), *Lectures on the Cauchy Problem in Linear Partial Differential Equations*, New Haven, CT: Yale University Press.
- Hansen, A. R., and Elliot, H. (1982), "Image Segmentation Using Simple Markov Field Models," *Computer Graphics and Image Processing*, 20, 101-132.
- Hillis, D. (1985), *The Connection Machine*, Cambridge, MA: MIT Press.
- Julesz, B. (1960), "Binocular Depth Perception of Computer Generated Patterns," *The Bell System Technical Journal*, 39, 1125-1162.
- Kemeny, J. G., and Snell, J. L. (1960), *Finite Markov Chains*, New York: Van Nostrand.
- Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983), "Optimization by Simulated Annealing," *Science*, 220, 671-680.
- Koch, C., Marroquin, J., and Yuille, A. (1985), "Analog 'Neuronal' Networks in Early Vision," *Artificial Intelligence Laboratory Memo*, No. 751, Cambridge, MA: MIT.
- Marr, D. (1982), *Vision, A Computational Investigation Into the Human Representation & Processing of Visual Information*, San Francisco: W. H. Freeman.
- Marr, D., and Poggio, T. (1976), "Cooperative Computation of Stereo Disparity," *Science*, 194, 283-287.
- Marroquin, J. (1985), "Probabilistic Solution of Inverse Problems," Technical Report 860, MIT, Artificial Intelligence Laboratory.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953), "Equation of State Calculations by Fast Computing Machines," *Journal of Physical Chemistry*, 21, 1087.
- Morozov, V. A. (1984), *Methods for Solving Incorrectly Posed Problems*, New York: Springer-Verlag.
- Nahi, N. E., and Assefi, T. (1972), "Bayesian Recursive Image Estimation," *IEEE Transactions on Computers*, 21, 734-738.
- Oster, G. F., Perelson, A., and Katchalsky, A. (1971), "Network Thermodynamics," *Nature*, 234, 393-399.
- Poggio, T. (1984), "Vision by Man and Machine," *Artificial Intelligence Laboratory Memo*, No. 776, Cambridge, MA: MIT.
- Poggio, T., and Koch, C. (1984), "Analog Networks: A New Approach to Neural Computation," *Artificial Intelligence Laboratory Memo*, No. 783, Cambridge, MA: MIT.
- Poggio, T., and Torre, V. (1984), "Ill-Posed Problems and Regularization Analysis in Early Vision," *Artificial Intelligence Laboratory Memo*, No. 773, Cambridge, MA: MIT.
- Poggio, T., Torre, V., and Koch, C. (1985), "Computational Vision and Regularization Theory," *Nature*, 317, 314-319.
- Terzopoulos, D. (1984), "Multiresolution Computation of Visible-Surface Representations," unpublished Ph.D. thesis, MIT, Dept. of Electrical Engineering and Computational Science.
- Terzopoulos, D. (in press), "Integrating Visual Information for Multiple Sources for the Cooperative Computation of Surface Shape," in *From Pixels to Predicates: Recent Advances in Computational and Robotic Vision*, ed. A. Pentland, Norwood, NJ: Ablex.
- Tikhonov, A. N., and Arsenin, V. Y. (1977), "Solutions of Ill-Posed Problems," Washington, DC: Winston & Sons.
- Wong, E. (1968), "Two-Dimensional Random Fields and the Representation of Images," *SIAM Journal on Applied Mathematics*, 16, 756-770.
- Woods, J. W. (1972), "Two-Dimensional Discrete Markovian Fields," *IEEE Transactions on Information Theory*, 18, 232-240.