MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo 839 April, 1985

OPTIMAL BAYESIAN ESTIMATORS FOR IMAGE SEGMENTATION AND SURFACE RECONSTRUCTION.

J. L. Marroquin

ABSTRACT: A very fruitful approach to the solution of image segmentation and surface reconstruction tasks is their formulation as estimation problems via the use of Markov random field models and Bayes theory. However, the Maximum a Posteriori (MAP) estimate, which is the one most frequently used, is suboptimal in these cases. We show that for segmentation problems, the optimal Bayesian estimator is the maximizer of the posterior marginals, while for reconstruction tasks, the thresholded posterior mean has the best possible performance. We present efficient distributed algorithms for approximating these estimates in the general case. Based on these results, we develop a maximum likelihood procedure that leads to a parameter-free distributed algorithm for restoring piecewise constant images. To illustrate these ideas, the reconstruction of binary patterns is discussed in detail.

© Massachusetts Institute of Technology (1985).

This report describes research done within the Laboratory for Information and Decision Systems and the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology. Support for the A.I. Laboratory's research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research Contract N00014-80-C-0505. The author was supported by the Army Research Office under contract ARO-DAAG29-84-K-0005.

1. Introduction.

A very powerful approach for solving computational early vision problems is to formulate them as estimation problems. The estimates are based on probabilistic models that embody both our prior generic knowledge about the behavior of the solution, as well as the process by which the observations are obtained. In particular, various researchers have used Bayesian estimation and Markov Random Field (MRF) models to perform image segmentation (Elliot et. al. (1983); Hansen and Elliot (1982); Geman and Geman (1983); Marroquin (1984b); Cohen and Cooper (1984)) and surface reconstruction tasks (Marroquin (1984a)). Most of these researchers have assumed that the Maximum a Posteriori (MAP) estimate, which is defined as the configuration which maximizes the posterior distribution, provides the best possible solution to these problems, and several methods (such as dynamic programming and "simulated annealing" (Kirkpatrick, et. al. (1983)), a form of stochastic relaxation) have been proposed to compute it.

Let us consider, however, the example portrayed in figure 1.

Panel (c) represents the MAP estimate of the binary MRF (a) (whose parameters are assumed to be known) from the noisy observations (b). It is clear that the estimates shown in (d) and (e) (which we will describe later) are better than the MAP estimate from almost any viewpoint. This suggests the use of criteria other than the value of the posterior probability to measure the performance of an estimator. In particular, we propose the use of the following criteria for the two classes of problems mentioned above:

1.1. Error Criterion for the Segmentation Problem.

Consider a field x with N elements each of which can belong to one of a finite set Q_i of classes. Let x_i denote the class to which the i^{th} element belongs. The segmentation problem is to estimate x from a set of observations $\{y_1, \ldots, y_p\}$. Note that x_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), etc.

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

$$e_s(x, \hat{x}) = \sum_{i=1}^{N} (1 - \delta(x_i - \hat{x}_i)) \qquad , x_i, \hat{x}_i \in Q_i \tag{1}$$

where

$$\delta(a) = \begin{cases} 1, & \text{if } a = 0 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

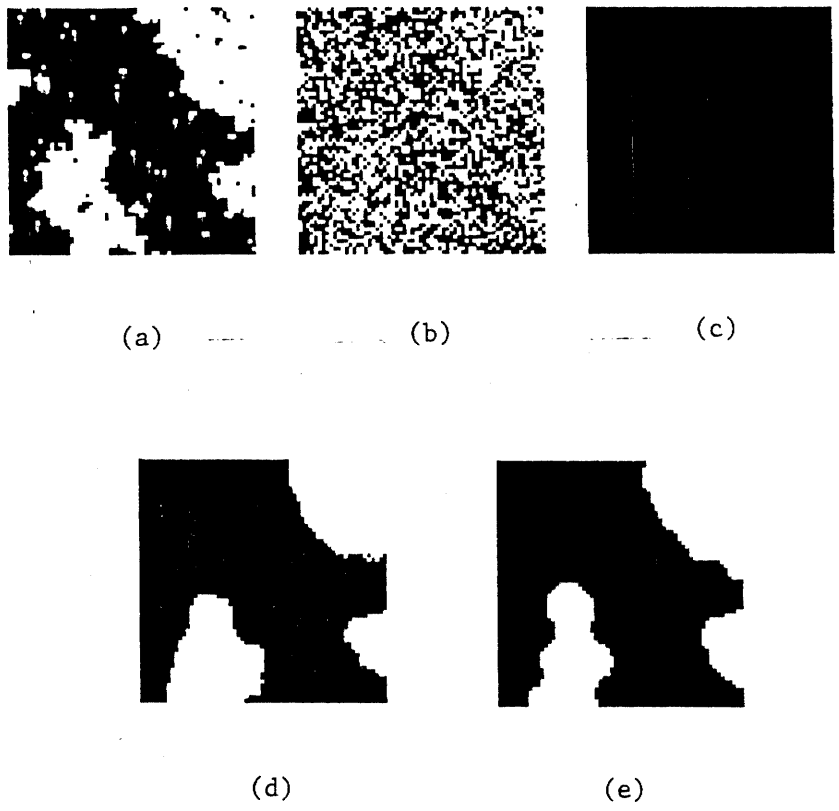

Figure 1. (a) Sample function of a binary MRF. (b) Output of a binary symmetric channel (error rate: 0.4) (c) MAP estimate. (d) Monte Carlo approximation to the MPM estimate. (e) Deterministic approximation to the MPM estimate.

1.2. Error Criterion for the Reconstruction Problem.

In this case, we also consider a field x with N elements which can take values on finite sets $\{Q_i\}$, but now we assume specifically that x_i represents the intensity of an image (or the height of a surface) at site i. This suggests that an estimate $\hat{x}$ should be considered "good" if it is close to x in the ordinary sense, so that the total squared error:

$$e_r(x, \hat{x}) = \sum_{i=1}^{N} (x_i - \hat{x}_i)^2 \tag{3}$$

will be a reasonable measure for its performance.

Let us now derive the optimal estimators for these error criteria.

2. Optimal Bayesian Estimators.

In this section we introduce two estimators based on the posterior distribution $P_{x|y}$: The Maximizer of the posterior marginals $\hat{x}_{MPM}$, which is obtained by maximizing separately the posterior marginal distribution of each element, and the Thresholded posterior mean $\hat{x}_{TPM}$. Their formal definition is as follows:

$$\hat{x}_{MPM}(i) = q \quad : \quad P_{i|y}(q;y) = \sup_{r \in Q_i} \{P_{i|y}(r;y)\} \tag{4}$$

where $P_{i|y}$ is the posterior marginal distribution of the i^{th} element:

$$P_{i|y}(q;y) = \sum_{x:x_i=q} P_{x|y}(x;y) \tag{5}$$

The TPM estimate is:

$$\hat{x}_{TPM}(i) = q \quad : \quad (q-\bar{x}_i)^2 = \inf_{r \in Q_i} \{(r-\bar{x}_i)^2\} \tag{6}$$

where $\bar{x}$ is the posterior mean:

$$\bar{x} = \sum_x x P_{x|y}(x;y) \tag{7}$$

We will show that the MPM and TPM estimators are optimal with respect to the error criteria (1) and (3) respectively. In what follows, we will assume that the prior and conditional distributions P_x, $P_{y|x}$ are known, so that the posterior distribution can be obtained from Bayes rule.

Proposition 1: $\hat{x}_{MPM}$ is optimal with respect to the segmentation error e_s.

This is not a new result (see for example Abend, 1968); however, we present here a simple proof:

Let $\hat{x} = \hat{x}_{MPM}$, and let z denote *any* other estimate for x. We will show that

$$E[e_s(x,\hat{x})] \leq E[e_s(x,z)]$$

where the expectation is taken over all x and all possible observations y.

For any fixed y we have, from (1):

$$\bar{e}_s(\hat{x} \mid y) = E[e_s(x,\hat{x}) \mid y] = \sum_x \sum_{i=1}^{N} (1-\delta(x_i-\hat{x}_i)) P_{x|y}(x;y)$$

$$= \sum_{i=1}^{N} \sum_{x:x_i \neq \hat{x}_i} P_{x|y}(x;y) = \sum_{i=1}^{N} (1 - P_{i|y}(\hat{x}_i;y))$$

with $P_{i|y}$ given by (5). Also,

$$\bar{e}_s(z \mid y) = \sum_{i=1}^{N} (1 - P_{i|y}(z_i; y))$$

Therefore,

$$\bar{e}_s(\hat{x} \mid y) - \bar{e}_s(z \mid y) = -\sum_{i=1}^{N} [P_{i|y}(\hat{x}_i; y) - P_{i|y}(z_i; y)] \tag{8}$$

But by the definition (4) of $\hat{x}$, each term on the right hand side of (8) is non-positive, so that

$$\bar{e}_s(\hat{x} \mid y) - \bar{e}_s(z \mid y) \leq 0 \qquad \text{for all } y$$

and therefore,

$$E[e_s(x, \hat{x})] \leq E[e_s(x, z)] \qquad \blacksquare$$

Note that we have not made any assumption about P_x, $P_{y|x}$, or $\{Q_i\}$, so that this is a completely general result. In particular, the sets of valid classes Q_i do not have to be the same for all i , so that this result holds if, as in Geman and Geman (1983), x is the pair $\{f, l\}$, where f denotes the intensity of a piecewise constant image, and l the boundaries between uniform regions.

Proposition 2: $\hat{x}_{TPM}$ is optimal with respect to the reconstruction error e_r.

Proof: Again, we put $\hat{x} = \hat{x}_{TPM}$, and let z denote any other estimate. For any fixed y we have, from (3):

$$\bar{e}_r(\hat{x} \mid y) = \sum_x \sum_{i=1}^{N} (x_i - \hat{x}_i)^2 P_{x|y}(x; y)$$

$$= \sum_{i=1}^{N} \sum_x (x_i^2 - 2x_i\hat{x}_i + \hat{x}_i^2) P_{x|y}(x; y)$$

$$= \sum_{i=1}^{N} [(\bar{x}_i - \hat{x}_i)^2 + \overline{x_i^2} - \bar{x}_i^2]$$

where

$$\bar{x}_i = \sum_x x_i P_{x|y}(x; y)$$

$$\overline{x_i^2} = \sum_x x_i^2 P_{x|y}(x; y)$$

also,

$$\bar{e}_r(z \mid y) = \sum_{i=1}^{N} [(\bar{x}_i - z_i)^2 + \overline{x_i^2} - \bar{x}_i^2]$$

so that

$$\bar{e}_r(\hat{x} \mid y) - \bar{e}_r(z \mid y) = \sum_{i=1}^{N} [(\bar{x}_i - \hat{x}_i)^2 - (\bar{x}_i - z_i)^2] \leq 0$$

from the definition (6) of $\hat{x}$ ∎

Remarks:

1. In the limit of fine quantization the variables x_i become continuous (i.e., $Q_i = \mathbf{R}$ for all i). In this case we recover the classical result that establishes that the posterior mean minimizes the mean squared error; for the particular case of Gaussian (Habibi, 1972; Nahi and Assefi, 1972) or isotropic fields (Levy and Tsitsiklis, 1983), it is then possible to construct efficient algorithms that compute this quantity.

2. The surface reconstruction problem treated by Marroquin (1984a) can be considered a particular instance of the reconstruction problem defined here. If we let x be the pair $\{f, l\}$, where f is the (continuous valued) depth process and l is a binary line process, the optimal estimator will be $\hat{x}_{TPM}$ (note that for a binary variable, $\hat{x}_{TPM}(i) = \hat{x}_{MPM}(i)$).

3. These results can also be applied to the case of an m-ary line process (like the one suggested by Geman and Geman (1983)) by defining an appropriate mixed error:

$$e_m(f, l, \hat{f}, \hat{l}) = \sum_{i=1}^{N_f} (f_i - \hat{f}_i)^2 + \lambda \sum_{i=1}^{N_l} (1 - \delta(l_i - \hat{l}_i))$$

with δ as defined in equation (2). One can show, using arguments similar to those of the proofs above, that the optimal estimate, for *any* positive value of λ will be $\{\hat{f}_{TPM}, \hat{l}_{MPM}\}$. Note that the posterior mean for f and the posterior marginal distributions for l must be computed by averaging over all possible values of both f and l:

$$\bar{f}_i = \sum_f \sum_l f_i P_{f,l|y}(f, l; y)$$

$$P_{l_i}(q) = \sum_f \sum_{l:l_i=q} P_{f,l|y}(f, l; y)$$

3. Algorithms.

As in the case of the MAP estimate, the exact computation of the optimal MPM and TPM estimates is a formidable task, except for very small N. It is possible, however, to define general distributed Monte Carlo procedures that will allow us to approximate these values as precisely as we want.

The general idea is to use the Metropolis algorithm (Metropolis, 1953) to generate global configurations of the field x which are, asymptotically, distributed

according to $P_x \mid y$. Since the Markov chain defined by the Metropolis algorithm is known to be ergodic (Geman and Geman, 1983), we can approximate $\bar{x}$ by:

$$\bar{x}_i \approx \frac{1}{n-k} \sum_{t=k}^{n} x_i^{(t)} \tag{9}$$

and the posterior marginals by:

$$P_{i|y}(q) \approx \frac{1}{k-n} \sum_{t=k}^{n} \delta(x_i^{(t)} - q) \tag{10}$$

where $x^{(t)}$ 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, $\hat{x}_{MPM}$ and $\hat{x}_{TPM}$ can be easily computed using (4) and (6).

If we compare this procedure with the use of "Simulated annealing" (Kirkpatrick, et.al., 1983) for approximating the MAP estimate, we note that even from a computational viewpoint it has some distinct advantages:

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 ($T = 1$), 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.

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).

With respect to the relative performance, we point out that for high signal to noise ratios, the MAP estimate is usually close to the optimal one (which explains its good performance in many cases). If the noise level is high, however, the difference in the performances of the two estimators may be dramatic (see fig. 1).

Finally, from a conceptual viewpoint, the definition of the solution of the estimation problem in terms of the posterior mean (or the posterior marginals) may lead in some cases to the construction of efficient deterministic algorithms for approximating the solution. To make these ideas concrete, we now analyze in detail a specific instance of the segmentation problem.

4. Optimal Segmentation of Binary MRF's.

In this section we will consider the following particular example of the general segmentation problem:

Let f represent a binary pattern on a rectangular lattice L with N elements, which is a particular realization of a Markov Random Field with potentials given by:

$$V(f_i, f_j) = \begin{cases} -1, & \text{if } f_i = f_j \quad \text{and } \|i-j\| \in (0,1] \\ 1, & \text{if } f_i \neq f_j \quad \text{and } \|i-j\| \in (0,1] \\ 0, & \text{otherwise} \end{cases}$$

where i and j are sites of L and $\|i-j\|$ is the distance betweeen i and j (i.e., f is a realization of a two-dimensional Ising net). Without loss of generality, we will assume that $f_i \in \{0,1\}$ for all $i \in L$.

The probability density of the configurations F is given by:

$$\Pr(F = f) = \frac{1}{Z} \exp[-\frac{1}{T_0} U(f)] \tag{11}$$

where Z is a normalizing constant; the energy $U(f)$ is:

$$U(f) = \sum_{i,j \in N_N} V(f_i, f_j)$$

where $i, j \in N_N$ means that i and j are nearest neighbors, and T_0 is the natural temperature of the system.

Suppose that f is sent through a binary symmetric channel with error rate ϵ, so that the conditional distribution of the observations g is:

$$P(g_i \mid f_i) = \begin{cases} (1-\epsilon), & \text{if } g_i = f_i \\ \epsilon, & \text{if } g_i \neq f_i \end{cases} \quad , f_i, g_i \in \{0,1\}$$

The posterior distribution is:

$$P_{f|g}(f;g) = \frac{1}{Z_P} e^{-U_P} \tag{12}$$

where Z_P is a constant, and the posterior energy U_P is:

$$U_P = \frac{1}{T_0} \sum_{i,j} V(f_i, f_j) + \alpha \sum_i (1 - \delta(f_i - g_i)) \tag{13}$$

with

$$\alpha = \ln\left(\frac{1-\epsilon}{\epsilon}\right).$$

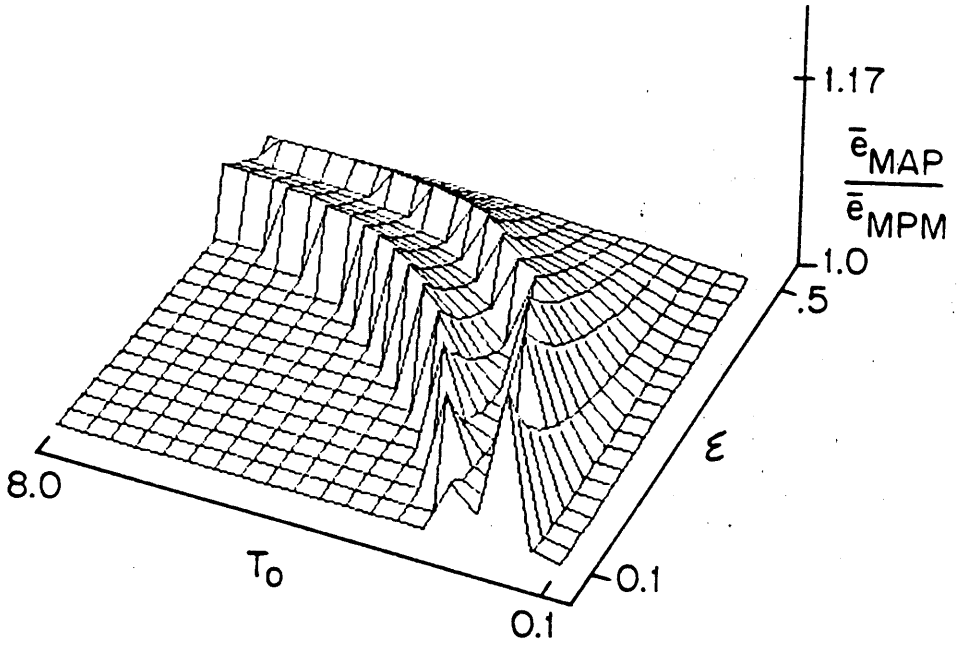

Figure 2. Ratio of the average errors of the MAP and MPM estimators for a 2×2 Ising net.

4.1. Error Analysis.

Using equations (1) and (12), it is not difficult to write down the exact expressions for the average segmentation error corresponding to the MAP and MPM estimators for this problem. However, the numerical evaluation of these expressions is feasible only for small values of N.

In figure 2 we show a plot of the ratio:

$$r = \frac{E[e_{MAP}]}{E[e_{MPM}]}$$

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 $\epsilon = 0.1$ and $T_0 = 0.2$) 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 0.5 (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.

4.2. Example.

We now return to the example presented in figure 1, and examine it in more detail. Panel (a) represents a typical realization of a 64×64 Ising net with free boundaries, using a value of $T_0 = 1.74$ (0.75 times the critical temperature of the lattice); panel (b), the output of a binary symmetric channel with error rate $\epsilon = 0.4$; panel (c) the MAP estimate, and panel (d) an approximation to the MPM estimate (which we will label "MPM (M.C.)") obtained using the Metropolis algorithm and equation (10) to estimate the posterior density. The corresponding values of the posterior energy U_P (equation (13)) and the relative segmentation error ($e_s/64^2$) are shown on table 1.

Table 1

	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
Seg. Error	–	0.4	0.33	0.128	0.124

4.3. A Fast Algorithm.

For this problem, it is possible to construct a fast deterministic algorithm whose experimental performance (in terms of the average segmentation error) is equivalent to the Monte Carlo method discussed above. It is based on the following ideas:

First, we note that for a binary pattern, the MPM and TPM estimates coincide. We will approximate the posterior mean of (12) by that of a Gaussian distribution P_G with the property:

$$P_G(h) = \frac{1}{Z_P} e^{-U_P(h)} \qquad \text{for all } h \in \{0,1\}.$$

In particular, we use:

$$P_G(h) = \frac{1}{Z_G} \exp[-\frac{1}{T_0} \sum_{i,j \in N_N} (h_i - h_j)^2 - \alpha \sum_i (h_i - g_i)^2].$$

For this distribution, $\bar{h}$ is the (unique) minimizer of the convex function:

$$U_G(h) = -\frac{1}{T_0} \sum_{i,j \in N_N} (h_i - h_j)^2 - \alpha \sum_i (h_i - g_i)^2$$

which corresponds to the unique fixed point of the system:

$$h_i^{(k+1)} = \frac{\sum_{j \in N_i} h_j^{(k)} + \alpha T_0 g_i}{|N_i| + \alpha T_0} \tag{14}$$

where

$$N_i = \{j \in L \quad : \quad \|i - j\| = 1\}.$$

We could now approximate our estimate by putting:

$$\hat{f}_i^* = \Theta(\bar{h}_i)$$

where

$$\Theta(x) = \begin{cases} 1, & \text{if } x \geq \frac{1}{2} \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

There is an additional consistency condition that $\hat{f}^*$ must satisfy, however. It can be shown that when the posterior distribution has the form given by (12) and (13), the MPM estimate $\hat{f}$, which by definition satisfies:

$$P_{i|g}(\hat{f}_i; g) > P_{i|g}((1 - \hat{f}_i); g)$$

also satisfies:

$$P_{i|g}(\hat{f}_i; \hat{f}) > P_{i|g}((1 - \hat{f}_i); \hat{f}) \tag{16}$$

which means that if we replace the observations by the MPM estimate, and compute a new MPM estimate for this modified problem, we should get the same result. Translating this condition to the case of $\hat{f}^*$, we get that it must satisfy:

$$\hat{f}_i^* = \Theta(h_i^*) \tag{17}$$

where h^* satisfies:

$$h_i^* = \frac{\sum_{j \in N_i} h_j^* + \alpha T_0 \Theta(h_i^*)}{|N_i| + \alpha T_0}$$

In practice, we get h^* as the fixed point of the system:

$$h_i^{*(k+1)} = \frac{\sum_{j \in N_i} h_j^{*(k)} + \alpha T_0 \Theta(h_i^{*(k)})}{|N_i| + \alpha T_0} \tag{18}$$

with

$$h^{*(0)} = \bar{h}$$

Note that the function:

$$U_h(h) = \sum_{i,j \in N_i} (h_i - h_j)^2 + \alpha T_0 \sum_i (h_i - \Theta(h_i))^2$$

acts as a Lyapunov function for the system (18), which is therefore (locally) stable (Vidyasagar, 1978).

This algorithm can be visualized as operating in two steps: In the first one, we extract all the information that we need from the observations and encode it in $\bar{h}$ (which is continuous-valued), and in the second one, we find the closest binary pattern that satisfies the consistency condition (16).

To illustrate the performance of this approximation, we show $\hat{f}^*$, for the example discussed above, in panel (e) of figure 1, and its corresponding energy and segmentation error in the last column of table 1 (labeled "MPM det.").

4.4. Dimensionality of the Parameter Space.

Equations (14) and (18), which describe the deterministic approximations to $\hat{f}_{MPM}$ depend on the parameters of the system, ϵ and T_0, only through the product:

$$\gamma = \alpha T_0 = T_0 \log\left(\frac{1-\epsilon}{\epsilon}\right) \tag{19}$$

which means that the behaviour of the algorithm is completely characterized by the single parameter γ. In the case of the Monte Carlo approximation (10), if we fix the value of γ, the value of T_0 cannot be chosen arbitrarily, since it has to satisfy the consistency condition:

$$T_0 = \frac{\gamma}{\hat{\alpha}}$$

with

$$\hat{\alpha} = \log\left(\frac{1-\hat{\epsilon}}{\hat{\epsilon}}\right) \quad ;$$

$$\hat{\epsilon} = \frac{1}{N} \sum_{i=1}^{N} z_i \tag{20}$$

where z is the residual process defined as:

$$z_i = \begin{cases} 1, & \text{if } \hat{f}_i \neq g_i \\ 0, & \text{otherwise} \end{cases} \tag{21}$$

This means that, given γ, the correct value of T_0 can, in principle, be determined in an adaptive way, so that in this case too, the behaviour of the approximation depends effectively only on γ.

5. Simultaneous Estimation of the Field and the Parameters.

The above considerations suggest that we can estimate simultaneously the original field *and* the parameters of the system by looking for the supremum of an appropriate likelihood function over the one-dimensional subspace parametrized by γ. We proceed as follows:

For a given value of γ, we can approximate the corresponding MPM estimate $\hat{f}$ using the methods developed in the previous section, and compute the residual process z and the conditional (on γ) Maximum Likelihood Estimate of the error rate ϵ using equations (20) and (21). The corresponding conditional estimate for T_0 will be:

$$\hat{T}_0 = \frac{\gamma}{\hat{\alpha}} \tag{22}$$

To measure the "likelihood" of the estimate $\hat{f}$, we use the degree of uniformity (or "whiteness") of the residual process z. This property can be quantified by the variance of the local noise density, which we estimate as follows:

We cover the lattice with a set $\{S_j\}$ of m non-overlapping squares (say, 8 pixels wide). For each square S_j, the relative variance of the noise density is:

$$\sigma_j = \left(\frac{\bar{z}_j - \hat{\epsilon}}{\hat{\epsilon}} \right)^2 \tag{23}$$

with

$$\bar{z}_j = \frac{1}{|S_j|} \sum_{i \in S_j} z_i$$

where $|S_j|$ is the area of the j^{th} square.

The desired likelihood function is:

$$L(\hat{f}) = -\sum_{j=1}^{m} \sigma_j \tag{24}$$

Note that a more conventional likelihood function, such as the conditional likelihood proposed by Besag (1972), will not work in this case; this function is defined as:

$$L(\hat{f}) = \frac{L_1(\hat{f}) + L_2(\hat{f})}{2} \qquad \text{with}$$

$$L_k(\hat{f}) = \prod_{i \in C_k} P(\hat{f}_i \mid \hat{f}_j, j \in N_i, \hat{T}_0) =$$

$$= \prod_{i \in C_k} \left[\frac{\exp[-\frac{1}{\hat{T}_0} \sum_{j \in N_i} V(\hat{f}_i, \hat{f}_j)]}{\exp[-\frac{1}{\hat{T}_0} \sum_{j \in N_i} V(\hat{f}_i, \hat{f}_j)] + \exp[-\frac{1}{\hat{T}_0} \sum_{j \in N_i} V(1 - \hat{f}_i, \hat{f}_j)]} \right]$$

$$= \prod_{i \in C_k} (1 + \exp[\frac{2}{\hat{T}_0} \sum_{j \in N_i} V(\hat{f}_i, \hat{f}_j)])^{-1}, \qquad k = 1, 2$$

where the "codings" C_1 and C_2 are sets of conditionally independent sites defined as:

$$C_1 = \{i \ : \ (x_i \text{ is odd and } y_i \text{ is even}) \text{ or } (x_i \text{ is even and } y_i \text{ is odd})\}$$

$$C_2 = \{i \ : \ (x_i \text{ is odd and } y_i \text{is odd}) \text{ or } (x_i \text{ is even and } y_i \text{ is even})\}$$

with (x_i, y_i) denoting the row and column indices of site i. In our case, we find that as γ decreases, $\hat{f}$ becomes more and more uniform, while $\hat{T}_0$ remains almost constant. It is not difficult to see that as a result, the conditional likelihood L will decrease monotonically with γ, which renders it useless for our purpose.

The range of values $[\gamma_0, \gamma_M]$ of the parameter γ that corresponds to the class of systems of interest can be determined as follows:

One can show that for $\gamma > 8$ we will always have $\hat{f}_{MPMi} = g_i$ for all i, so that we can use $\gamma_M = 8$. The value of γ_0 can be obtained from an upper bound for ϵ and a lower bound for T_0. For example, assuming that $\epsilon \leq .45$ and $T_0 \geq .5T_c$, we get $\gamma_0 = .23$. (Note that when the natural temperature T_0 of a first order, isotropic MRF is below 0.5 times T_c (the critical temperature of the lattice; see Kindermann and Snell, 1980), the patterns become practically uniform (i.e., f_i =constant for all i), while for values of T_0 greater than $1.5T_c$, we get patterns that are practically indistinguishable from white noise. Therefore, the assumption $T_0 \geq .5T_c$ covers practically all the interesting cases).

The complete estimation procedure is as follows:

Maximum Likelihood Estimation Algorithm:

1: Sample the interval $[\gamma_0, \gamma_M]$ at the points

$$\gamma_0 \leq \gamma_1, \ldots \gamma_n \leq \gamma_M$$

2: For each $\gamma \in Q = \{\gamma_1, \ldots \gamma_n\}$:

2.1: Find $\hat{f}(\gamma)$ using (14) and (18).

2.2: Compute z using (21).

2.3: Compute $\hat{\epsilon}$ using (20). If $\hat{\epsilon} = 0$, set $\mathcal{L}(\hat{f}(\gamma)) = -\infty$ and proceed with the next value of γ. Otherwise, compute $\hat{\alpha}$ and go to 2.4.

2.4: Compute $\hat{T}_0$ using (22).

(a) (b) (c)

Figure 3. (a) Synthetic image. (b) Noisy observations. (c) Maximum Likelihood Estimate.

2.5: Compute $L(\hat{f}(\gamma))$ using (23) and (24).

3: Compute the optimal estimate $\hat{f}^*$ using:

$$\hat{f}^* = \hat{f}(\gamma^*) \quad : \quad L(\hat{f}(\gamma^*)) = \sup_{\gamma \in Q} L(\hat{f}(\gamma)) \tag{25}$$

The corresponding $\hat{\epsilon}^*, \hat{T}_0^*$ will be the optimal estimates for ϵ and T_0, respectively.

Remarks:

1. This estimation algorithm allows us to reconstruct a binary 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, but neither the natural temperature T_0 nor the error rate ϵ have to be known in advance. In practice, this means that we can apply it to restore any binary image with uniform granularity, even if it has not been generated by a Markov random process. We have used this algorithm to reconstruct a variety of binary images with excellent results. In figure 3 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).

2. This procedure can be easily extended to handle any one-parameter noise corruption process (such as zero mean, additive white Gaussian noise). The extension

to the case of N-ary fields, i.e., to the restoration of piecewise constant images, is also straightforward (using the general algorithm (10) together with equation (4) instead of (14) and (18) in step 2.1).

3. We have found that the likelihood function (24) is reasonably well behaved as a function of γ. This permits us to perform the one-dimensional search for its supremum in an economical way, by first determining its approximate location using a coarse sampling pattern, and then refining its position by a fine sampling of a reduced interval. In practice, it is possible to get very good results using on the order of 15 samples.

4. The whole procedure is highly distributed, so that it is possible to implement it in parallel hardware in a very efficient way.

6. Conclusions.

A very fruitful approach to the solution of image segmentation and surface reconstruction tasks is their formulation as estimation problems, via the use of MRF models and Bayes theory. However, the MAP estimate, which has been often used, will not always give the best solution. We have shown that, for segmentation problems, the optimal Bayesian estimator is the maximizer of the posterior marginals (MPM), while for reconstruction tasks, the thresholded posterior mean (TPM) has the best possible performance. In some cases, these estimators will give results that are similar to those obtained by the MAP procedure, but in many interesting situations (in particular for low signal to noise ratios) their performance will be significantly better.

The exact computation of these estimates (including the MAP) is computationally unfeasible in most practical situations. We have presented a general Monte Carlo procedure which approximates their value as precisely as desired (at the expense of increased computational cost). This procedure can be easily implemented in parallel hardware, and it has the advantage that coarse solutions are computed very fast, and are then progressively refined. We have also shown how in particular cases, it is possible to use the conceptual framework provided by these estimators to construct fast algorithms with very good experimental performance.

Finally, we have developed a maximum likelihood procedure for the simultaneous estimation of a piecewise constant field and the parameters of the system. This construction leads to a parameter-free distributed algorithm for reconstructing piecewise constant images from noisy observations.

Acknowledgements: I want to thank Prof. S.K. Mitter for many valuable suggestions that led to these results. I also want to thank him and Prof. T. Poggio for their encouragement and enthusiasm.

References.

Abend K. "Compound Decision Procedures for Unknown Distributions and for Dependent States of Nature". in Pattern Recognition pp. 207-249 L. Kanal, ed. Thompson Book Co. Washington, D.C. (1968).

Besag J. "Spatial Interaction and the Statistical Analysis of Lattice Systems". J. Royal Stat. Soc. B 34 75-83 (1972).

Cohen F.S. and Cooper D.B. "Simple Parallel Hierarchical and Relaxation Algorithms for Segmenting Noncausal Markovian Random Fields" Brown University Laboratory for Engineering Man/Machine Systems. Tech. Report LEMS-7 (1984).

Elliot H., Derin R., Christi R. and Geman D. "Application of the Gibbs Distribution to Image Segmentation". Univ. of Massachusetts Technical report (1983).

Geman S. and Geman D. "Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images". Technical Report. Brown University (1983).

Habibi A. "Two Dimensional Bayesian Estimation of Images" Proc. IEEE 60, 878–883 (1972).

Hansen A.R. and Elliot H. "Image Segmentation using Simple Markov Field Models" Comp. Graphics and Image Proc. 20, 101–132 (1982).

Kindermann R. and Snell J.L. "Markov Random Fields and their Applications". Vol 1. Amer. Math. Soc. (1980).

Kirkpatrick S. Gelatt C.D. and Vecchi M.P. "Optimization by Simulated Annealing". Science 220 (1983) 671–680.

Levy B.C. and Tsitsiklis J.N. "An Inverse Scattering Procedure for the Linear Estimation of Two–Dimensional Isotropic Random Fields" Laboratory for Information and Decision Systems LIDS-P-1192 MIT (1983).

Marroquin J.L. "Surface Reconstruction Preserving Discontinuities" Artificial Intelligence Lab. Memo No. 792. MIT (1984a).

Marroquin J.L. "Bayesian Estimation of One–Dimensional Discrete Markov Random Fields" Laboratory for Decision and Information Systems LIDS-R-1423. MIT (1984b).

Metropolis N. et. al. "Equation of State Calculations by Fast Computing Machines". J. Phys. Chem. 21 6 (1953) 1087.

Nahi N.E. and Assefi T. "Bayesian Recursive Image Estimation" IEEE Trans. on Computers 21, 734–738 (1972).

Vidyasagar M. "Nonlinear Systems Analysis" Prentice Hall (1978).