MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo 792

August, 1984

SURFACE RECONSTRUCTION PRESERVING DISCONTINUITIES.

J. L. Marroquin

ABSTRACT: This paper presents some experimental results that indicate the plausibility of using non–convex variational principles to reconstruct piecewise smooth surfaces from sparse and noisy data. This method uses prior generic knowledge about the geometry of the discontinuities to prevent the blurring of the boundaries between continuous subregions.

We include examples of the application of this approach to the reconstruction of synthetic surfaces, and to the interpolation of disparity data from the stereo processing of real images.

© Massachusetts Institute of Technology (1984).

This report describes research done within the Artificial Intelligence Laboratory, and the Laboratory for Information and Decision Systems 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.

In the processing of two-dimensional signals, often arises the problem of reconstructing a piecewise smooth surface from noisy observations taken at sparse locations. In this reconstruction, it is 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 of the surface. 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, etc.

The most successful approaches to this problem [T2] consist of, first, interpolating an everywhere smooth surface 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, to re-interpolate 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.

One way around this difficulty is to perform the boundary detection and interpolation tasks *at the same time*. In the method we will present, this is done by generating a variational principle that includes our prior knowledge about the smoothness of the surface and about the geometry of the discontinuities, as well as the information provided by the observations. The global minimum of this (non-convex) "energy" functional is then found by a stochastic approximation scheme.

This approach is based on the work of Geman and Geman [G1]. Before describing it, let us formulate the problem in a more precise way.

Imagine a region Ω of the plane which is formed by a number of subregions separated by boundaries which are known to be piecewise smooth curves. Suppose that within each of these subregions, some property f (in what follows, we will refer to f as "depth") varies in a smooth fashion, presenting, at the same time, abrupt jumps across most of the boundaries. Suppose also that we have measurements for the values of f at some discrete set of sites S; these measurements will, in general, be corrupted by some form of noise.

Our problem is then to estimate the values of f on some finite lattice of points $L \subseteq \Omega$, and to find the position of the boundaries, using all the available information in an optimal way.

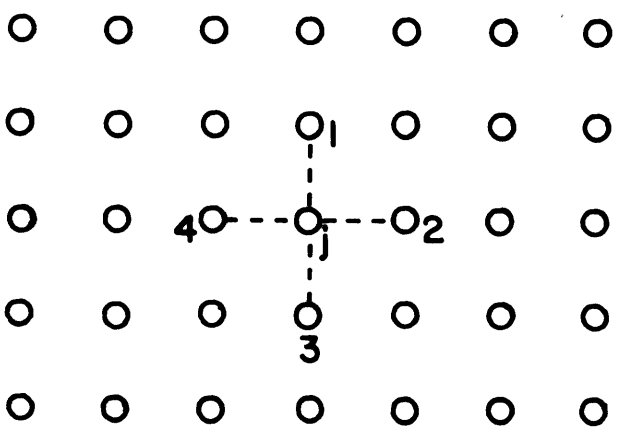

Figure 1. Sites 1, 2, 3 and 4 are the neighborhood of site j

2. Geman's Work on Bayesian Image Restoration.

2.1 Images as Markov Random Fields.

The first idea on which this approach is based, is that an image formed by regions of constant intensity separated by piecewise smooth boundaries can be modeled as a sample function of a stochastic process based on the Gibbs distribution, that is, as a Markov Random Field (MRF) [1] (sample functions of actual MRF's can be found in his paper. See also [C1] and [G2]). This means that the conditional probability of a given pixel j having a particular value f_j, given the values of f in all the remaining sites of the lattice, is identical to the conditional probability given the values of f in a small set of sites which we will call the neighborhood of j.

Given a system of neighborhoods on a lattice, we define a "clique" C as a set of sites such that all the sites that belong to C are neighbours of each other. For example, on a 4-connected lattice (Fig. 1), the sites 1, 2, 3 and 4 form the neighborhood of site j, and the cliques are sets consisting either of single sites, or of two (vertically or horizontally) adjacent sites (nearest neighbours; see Fig. 2).

It can be shown [P1], that for a MRF with a given system of neighborhoods, the probability density of the images f generated by this process is of the form:

$$P_f(f) = \frac{1}{Z} e^{-\frac{1}{\beta} U(f)}$$

where Z is a normalizing constant, β is a parameter, and the "Energy function" $U(f)$ is of the form:

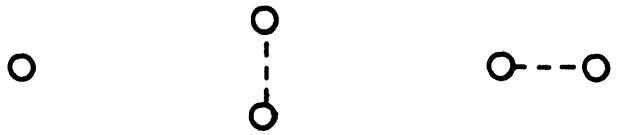

Figure 2. Cliques for the 4-connected lattice of Fig. 1.

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

where C ranges over the cliques associated with the given neighborhood system, and the potentials $V_C(f)$ are functions supported on them. Thus, in our example of a 4-connected lattice, U would be:

$$U(f) = \sum_j V_1(f_j) + \sum_{i,j \in N_N} V_2(f_i, f_j)$$

where $i, j \in N_N$ means that i and j are nearest neighbours, and V_1 and V_2 are some functions [2].

In particular, if we want a Markov random process that generates piecewise constant surfaces, we may use potentials:

$$V_1(f) = 0$$

$$V_2(f_i, f_j) = \begin{cases} -1, & \text{if } f_i = f_j \\ 1, & \text{otherwise} \end{cases} \tag{1}$$

The parameter β can be interpreted as the natural temperature of the system, and controls the expected size of the regions formed by the process, larger regions being formed at low temperatures.

2.2 Bayesian Estimation of Markov Random Fields.

Suppose that we have observations g that can be modeled as:

$$g_j = \Psi(H_j(f), n_j) \qquad j \in S \tag{2}$$

where n_j is a white noise process uncorrelated with f; S is some set of sites; H_j is an operator with local support (representing blurring, for example), and Ψ is invertible with respect to n_j. (Ψ may represent, for example, noise addition or multiplication followed by a pointwise transformation).

The conditional probability $P(g \mid f)$ is given by:

$$P(g \mid f) = \prod_{j \in S} P_n(\Psi^{-1}(g_j - H_j(f)))$$

where P_n is the probability density function of the noise.

The posterior distribution is found by Bayes rule:

$$P(f \mid g) = \frac{P_f(f) \cdot P(g \mid f)}{P(g)}$$

Replacing the expressions for $P_f(f)$ and $P(g \mid f)$, taking logarithms, and remembering that $P(g)$ is a constant for a given set of observations, we get that the Maximum a Posteriori (MAP) estimate for f is found by minimizing:

$$E(f) = \frac{1}{\beta} U(f) - \sum_{j \in S} \ln[P_n(\Psi^{-1}(g_j - H_j(f)))] \tag{3}$$

In particular, if n is a zero–mean Gaussian white noise stationary process with power spectral density σ^2, and

$$g_j = H_j(f) + n_j = f_j + n_j \quad ,$$

then,

$$P(g \mid f) = \frac{1}{\sqrt{2\pi}\sigma} exp[-\frac{1}{2\sigma^2} \sum_{j \in S} (g_j - f_j)^2]$$

For our example of piecewise constant surfaces, with potentials given by (1), the MAP estimate will be obtained by minimizing:

$$E(f) = \sum_{i,j \in N_N} V_2(f_i, f_j) + \frac{\beta}{2\sigma^2} \sum_{j \in S} (f_j - g_j)^2 \tag{4}$$

As we can see, this expression has two terms: One that measures the agreement of the estimate with the observations, and another that corresponds to the constraints imposed by our prior knowledge about the nature of the solution. The tradeoff

between them is controlled by the parameter

$$\alpha = \frac{\beta}{2\sigma^2}$$

which corresponds to the signal to noise ratio.

It is interesting to note that the general form of this expression is similar to the variational principles obtained by regularization methods for solving ill-posed problems, α playing the role of the regularization parameter. In the standard regularization methods, however, the functional is convex (because of the choices of norms and stabilizing functionals, see [P2,T1]), whereas in the case of equation (4), the high non-linearity of V_2 makes $E(f)$ non-convex. For this reason, its minimization becomes computationally much more expensive.

2.3 Simulated Annealing and the Minimization of $E(f)$.

In a recent paper, Kirkpatrick, et. al. [K1] proposed a stochastic approximation method for solving combinatorial optimization problems, which can be used for our minimization problem. It is based on an algorithm invented by Metropolis [M1] that simulates the behaviour of many-particle systems in thermal equilibrium:

Consider a system with N particles, each of which may be in any one of a discrete number of allowable states. Let f_j denote the current state of the j^{th} particle; T, the temperature, and let $E(f)$ be the total energy of the system. Suppose that we visit the particles of the system in some random sequential order. When a particle j is visited, we update its state as follows:

(i) Choose a new state $\hat{f}_j$ randomly from the set of allowable states (excluding the current state f_j) , using a uniformly distributed random number.

(ii) Compute the increment in energy ΔE_j that results from moving the state of the j^{th} particle from f_j to $\hat{f}_j$.

(iii) If $\Delta E_j \leq 0$, make the move, i.e., set $f_j = \hat{f}_j$.

If $\Delta E_j > 0$, generate a new random number r, uniformly distributed between 0 and 1.

If $r \leq e^{-\Delta E_j/T}$, set $f_j = \hat{f}_j$.

If $r > e^{-\Delta E_j/T}$, leave f_j unchanged.

It can be shown [M1,K2] that if every particle is visited infinitely often, this procedure will generate global states for f distributed according to Gibbs distribution, i.e., as the number of iterations goes to infinity, we will have:

$$P(\omega = f) = \frac{1}{Z} e^{-\frac{E(f)}{T}}$$

As T goes to 0, this distribution will tend to an impulse (or set of impulses) corresponding to the state (or states) of minimum energy, that is, to the value of f that minimizes $E(f)$ globally.

One serious difficulty, however, is that attaining thermal equilibrium might take a very long time at low temperatures. Kirkpatrick's idea was to start at a relatively high temperature (where thermal equilibrium is reached very fast), and then, to slowly cool the system, until "freezing" occurs and the state stops changing.

Geman & Geman were able to show that if the temperature is lowered at the rate:

$$T = \frac{C}{\log(n+1)} \tag{5}$$

where n is the number of iterations, and C is a constant, this algorithm will in fact converge (in probability) to the set of states of minimal energy (see also [G3]). Also, they proved that the non-homogeneous Markov chain that corresponds to the annealing procedure is (asymptotically) stationary and ergodic [3], so that we may use time statistics to estimate the final state. We will return to this in the next section.

Unfortunately, the value of the constant C for which Geman and Geman were able to guarantee convergence is in general very high [4] (so that the convergence time becomes impractically slow), but it has been found experimentally that a value of $C = 3\beta \log 2$ (where β is the natural temperature of the system) normally produces a reasonable convergence behaviour (of the order of 1000 iterations).

The computational efficiency of Kirkpatrick's algorithm depends on whether the increment in energy ΔE_j associated with a change in the state of the j^{th} variable is easy to compute. Fortunately, the Markov property of f, and the localization condition on the support of the operator H_j (equation (2)) guarantee that, for our problem, this will be the case, so that simulated annealing is a practical, although expensive way of finding the minimum of (3).

2.4 The Line Process.

Another powerful idea in the Geman's work is the introduction of an unobservable line process l into the image model. The variables associated with this process are located at the sites of the dual lattice of lines that connect the sites of the original (pixel) lattice (see Fig. 3). These variables may be binary (indicating the presence or absence of a boundary between two pixels), or may take more values to indicate the orientation of the boundary as well. In both cases, their function is to decouple adjacent pixels, reducing the total energy if the intensities of these pixels are different.

The introduction of this process is particularly important for the discontinuity-preserving reconstruction of surfaces from sparse observations, since it allows us to include into the estimation problem our prior knowledge about the geometry of

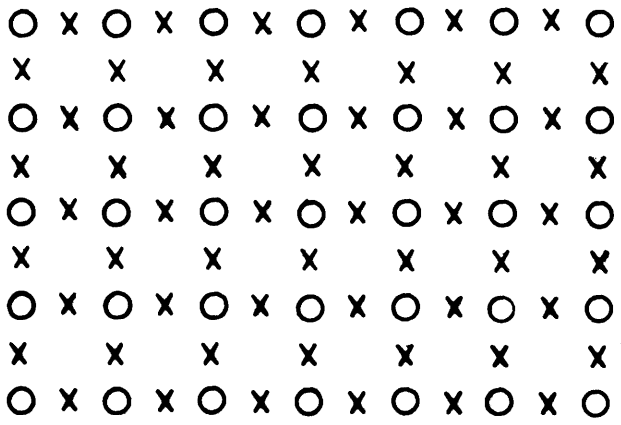

Figure 3. Dual lattice of line elements (sites denoted by x)

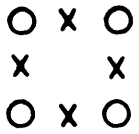

Figure 4. Cliques for the line process

the boundaries between regions in an explicit way. This is done by modifying the energy function; the new expression is:

$$E(f,l) = \sum_{i,j \in N_N} V_f(f_i, f_j, l_{ij}) + \sum_{C_l} V_{C_l}(l) + \frac{\beta}{2\sigma^2} \sum_{j \in S} (f_j - g_j)^2 \qquad (6)$$

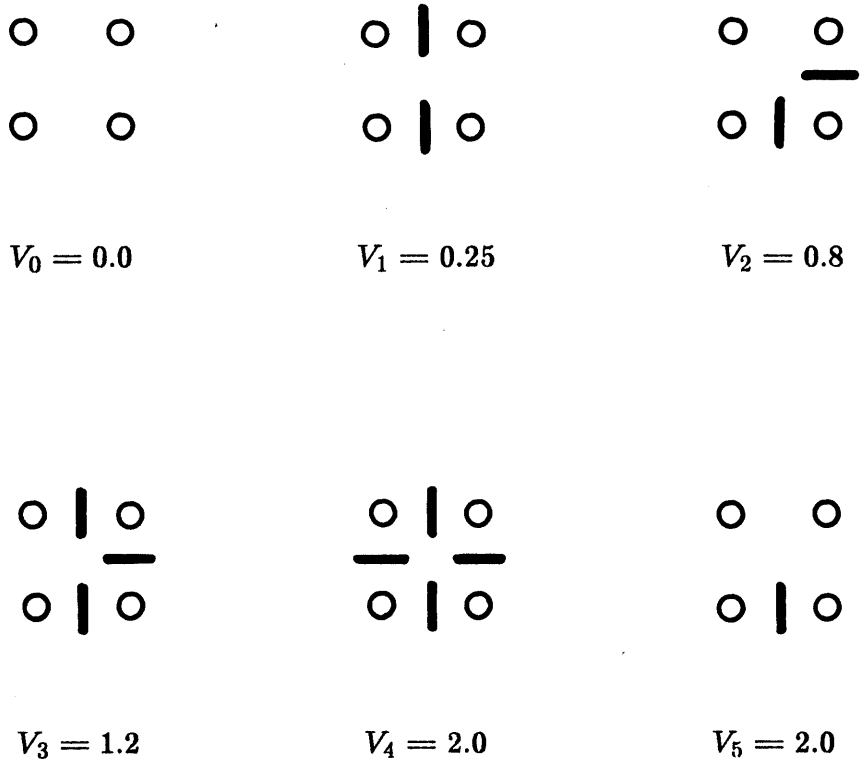

Figure 5. Potentials for the different configurations of a line process

where

$$V_f(f_i, f_j, l_{ij}) = \begin{cases} 0, & \text{if } l_{ij} \text{ is "on"} \\ V_2(f_i, f_j), & \text{otherwise} \end{cases}$$

V_2 is defined in eq. (1); l_{ij} is the line element between sites i and j, and the line potentials V_{C_l} have as supports cliques of size 4, such as the one shown in Fig. 4. Every line element (except at the boundaries of the lattice) belongs to 2 such cliques. The values of the potentials associated with each possible configuration of lines within a clique must be specified. Thus, for example, if we know that straight horizontal and vertical boundaries are likely to be present, we may use a binary process, and potential values as those of Fig. 5 (rotational invariance is assumed). If we want to handle more general situations (such as piecewise smooth boundaries), it is necessary to allow more states for the line elements, corresponding to different orientations, augmenting consequently the table of values for the potentials.

3. Extension to the Boundary Preserving Interpolation of Piecewise Smooth Surfaces.

3.1 Continuous Intensity Values.

Geman & Geman successfully applied their method to the restoration of piecewise constant images corrupted by additive white Gaussian noise, assuming that the grey levels corresponding to the constant subregions were either known in advance, or easily obtainable from the data (as peaks of a histogram, for example). However, if we want to apply this construction to more practical cases, we must relax the piecewise constant condition, so as to include tilted planes, smooth gradients, and in general, piecewise smooth surfaces.

To do this, we must allow the depth variables f_j to take any real value. However, this would require, in general, to replace Metropolis algorithm by a process capable of handling continuous variables (such as a diffusion. See [G2]), and in this case, the convergence of the annealing process is uncertain (rigorous convergence results have not been established yet). It is also possible, of course, to discretize the set of allowable values for f_j, and to use simulated annealing in a conventional way, but if this discretization is fine enough to guarantee sufficient precision in the results, the price that one would have to pay, in terms of computational efficiency, may be too high.

One way around this difficulty, is to define a "mixed" annealing strategy, in which the continuous variables are updated in a deterministic way, while the stochastic updating defined by Metropolis algorithm affects only the discrete variables. To do this, we replace the depth potential V_f in eq. (6) by one that guarantees that, for any given state of the line process, the resulting conditional energy function $E(f \mid l)$ is convex, so that the use of a deterministic gradient descent iteration for updating the state of the depth process is justified in some sense (this is not a rigorous argument; we present it only to provide some motivation for the definition of this strategy which, at this point, is justified only by our experimental results).

This can be achieved by using a quadratic potential:

$$\hat{V}_f(f_i, f_j, l_{ij}) = \begin{cases} 0, & \text{if } l_{ij} \text{ is "on"} \\ (f_i - f_j)^2, & \text{otherwise} \end{cases} \tag{7}$$

Since the resulting conditional energy function is of the form:

$$E(f \mid l) = \sum (f_i - f_j)^2 + \frac{\beta}{2\sigma^2} \sum_{j \in S} (f_j - g_j)^2 + K \tag{8}$$

for some positive constant K, we have that $E(f \mid l) > 0$ for any $f \neq 0$, and by an appropriate change of coordinates, it can be put in the form:

$$E(f \mid l) = u^T A u$$

with $u \in \mathcal{R}^{|L|}$, and A a non-negative definite matrix.

Now, for any $t \in (0,1)$, and any $u \neq v$, we have,

$$tE(u \mid l) + (1-t)E(v \mid l) - E(tu + (1-t)v \mid l) =$$

$$= t(1-t)(u-v)^T A(u-v) \geq 0$$

so that $E(f \mid l)$ is a convex function of f, and an iteration of a gradient descent algorithm will move towards a global minimum of this function. Note however, that there may be degenerate cases in which some region Q within which there are no observations is isolated from the rest of the lattice by the line process. In this case, any solution for which

$$f_j = \text{constant}\,, \qquad j \in Q$$

will be a fixed point of the gradient descent algorithm, and although all these solutions have the same energy, some of them are obviously more desirable than others. Experimentally, we have found that a good strategy to prevent the formation of undesirable "islands", is to use the global minimum of $E(f \mid l = 0)$ (which is the unique fixed point of the gradient descent algorithm) as the starting configuration for the mixed annealing process. This configuration can be interpreted physically as a membrane that is coupled to the data points (observations) by means of springs with constants equal to $\frac{\beta}{2\sigma^2}$, and is in a position that minimizes the total energy stored in itself, and in the springs (see [T2]).

3.2 The Mixed Annealing Process.

The scheme we are proposing is as follows:

Every global iteration (that is, for every fixed temperature), all the line and intensity sites are visited sequentially. When a line site is visited, its state is updated using Metropolis algorithm. The corresponding increment in energy ΔE_l is computed as follows:

Let C_1 and C_2 be the two cliques to which the line element belongs. Let l_{ij} be its current state, and $\hat{l}_{ij}$ be the candidate state (if l is a binary process, $\hat{l}_{ij} = 1 - l_{ij}$; otherwise, it is chosen at random from the set of allowable states different from l_{ij}). Let l be the current global line configuration, and $\hat{l}$ be obtained from l by replacing l_{ij} by $\hat{l}_{ij}$. We have:

$$\Delta E_l = V_{C_1}(\hat{l}) - V_{C_1}(l) + V_{C_2}(\hat{l}) - V_{C_2}(l) +$$

$$+\text{sgn}(l_{ij} - \hat{l}_{ij})(f_i - f_j)^2 \qquad (9)$$

where

$$\text{sgn}(x) = \begin{cases} 1, & \text{if } x > 0 \\ -1, & \text{if } x < 0 \\ 0, & \text{if } x = 0 \end{cases}$$

When an intensity site j is visited, its new state f'_j is obtained deterministically by the formula:

$$f'_j = \frac{\sum_{i:|i-j|\in(0,1]}(1 - \mathrm{sgn}(l_{ij}))f_i + \alpha q_j g_j}{\sum_{i:|i-j|\in(0,1]}(1 - \mathrm{sgn}(l_{ij})) + \alpha q_j} \tag{10}$$

where $\alpha = \frac{\beta}{2\sigma^2}$, and $q_j = 0$, unless there is an observation at site j (with value g_j), in which case, $q_j = 1$.

The temperature is lowered using (5).

3.3 Parameter Values.

Unfortunately, we do not have, at this point, a clean way for selecting the values for the parameters of the system, other than a trial and error procedure. A few considerations about their meaning are in order:

The parameter α controls the degree of smoothing that one applies to the data, and must be related to the quality of the observations. Large values of α (≈ 10) will force the interpolated surface to pass through the observations, while small values (≈ 1) will smooth away all but the largest fluctuations.

The parameters associated with the line potentials control both the shape and the number of boundaries that the algorithm will detect. For example, for the binary line process of Fig. 5, if h is the height of the smallest jump that we want to consider as a boundary, and d is the largest gradient in the smooth regions, we must have for V_1 (the potential of a straight boundary)

$$d^2 < V_1 < h^2 \tag{11}$$

The ratio of V_1 to the potentials for the rest of the configurations, represents our prior knowledge about the relative likelihood of corners, "T" junctions, etc.

4. Experiments

We now present some experimental results that support the use of this approach for surface reconstruction and image segmentation tasks. In these examples, we assume that we have the following prior knowledge about the nature of the surfaces we are trying to reconstruct:

(i) The region under consideration can be segmented into a small number of subregions.

(ii) Within each subregion the surface is smooth (the gradient is less than 0.5).

(iii) The boundaries between regions are piecewise horizontal or vertical. There are relatively few corners.

(iv) The average height of the discontinuities across boundaries is greater than 0.8.

(v) The observations are corrupted by an additive white Gaussian noise process, and we have some estimate of its intensity.

This knowledge is embodied in the model for the line process, and in the numerical value of the parameters. For our experiments, we have used the binary model of Fig. 5, with the values for the potentials for the different configurations indicated there.

4.1 Experimental Results.

In the first set of experiments, we generated sparse observation points at 200 random locations of a 30 X 30 rectangular grid. Figures 6, 7, 8 and 9 show (with height coded by grey level) the observations (a); the boundaries found by the algorithm (b); the configuration obtained by interpolation with no boundaries (c), and the final reconstructed surface (d), for:

(i) A square at height 2.0 over a background at constant height = 1.0 (Fig. 6).

(ii) A triangle, with the same characteristics (Fig. 7).

(iii) A tilted square plane (slope = 0.1) over a constant height background with white Gaussian added noise ($\sigma = 0.1$)(Fig. 8).

(iv) Three rectangles at different (constant) heights over a uniform background (Fig. 9).

In the second set of experiments, the same algorithm was used for a boundary detection / image segmentation task. In this case, we have observations (corrupted by white Gaussian noise) at almost every point in the lattice. The original figure is a 10 X 10 tilted square plane of slope 0.2 located at the center of the lattice (Fig. 10). Note that in this case, the tilted plane cuts across the uniform background, so that the vertical steps at both sides have opposite signs, while the horizontal steps change sign at the center of the figure, where, in fact, there is no discontinuity. As in the previous experiments, the results speak for themselves.

Finally, we present an example of the application of this algorithm to the processing of real images. We use it to interpolate disparity data, obtained along the zero–crossing contours of the convolution of a stereo pair of aerial photographs with a "Difference of Gaussians" operator, by Grimson's implementation of the Marr–Poggio stereo algorithm [G4,M2] (the data array was rotated, so that assumption (iii) held). The results are shown in figure 11. We believe that they will improve when we implement the extensions to this method outlined in section 5.2.

4.2 Convergence of the Annealing Process.

We have found that the mixed annealing algorithm, with annealing schedule given by (5), eventually converges to a low energy solution; however, its convergence

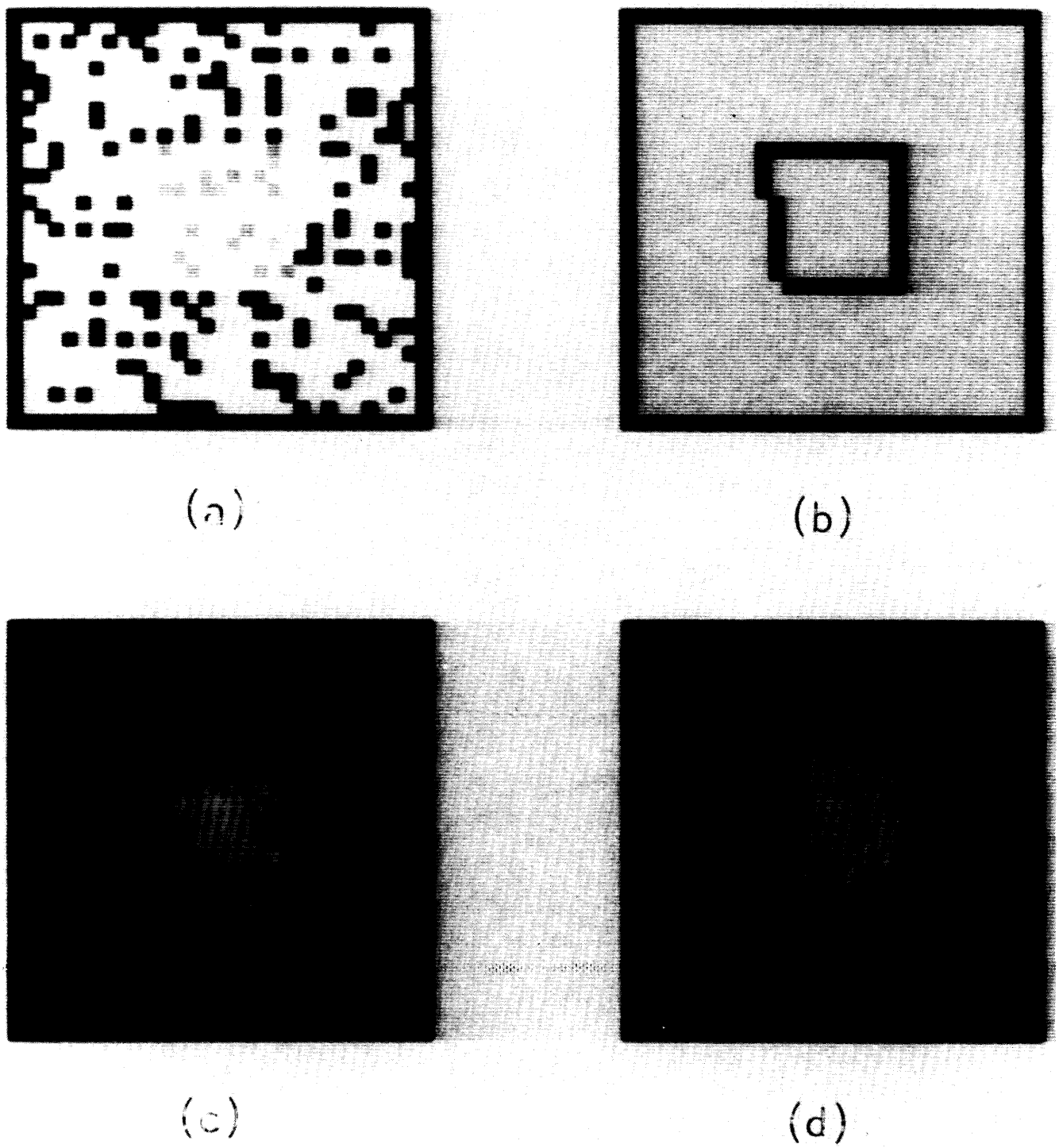

Figure 6. (a) Observations of a square at height 2.0 over a background at height 1.0 (a white pixel means that the observation is absent at that point). (b) Boundaries found by the Algorithm. (c) Interpolation with no boundaries. (d) Reconstructed surface.

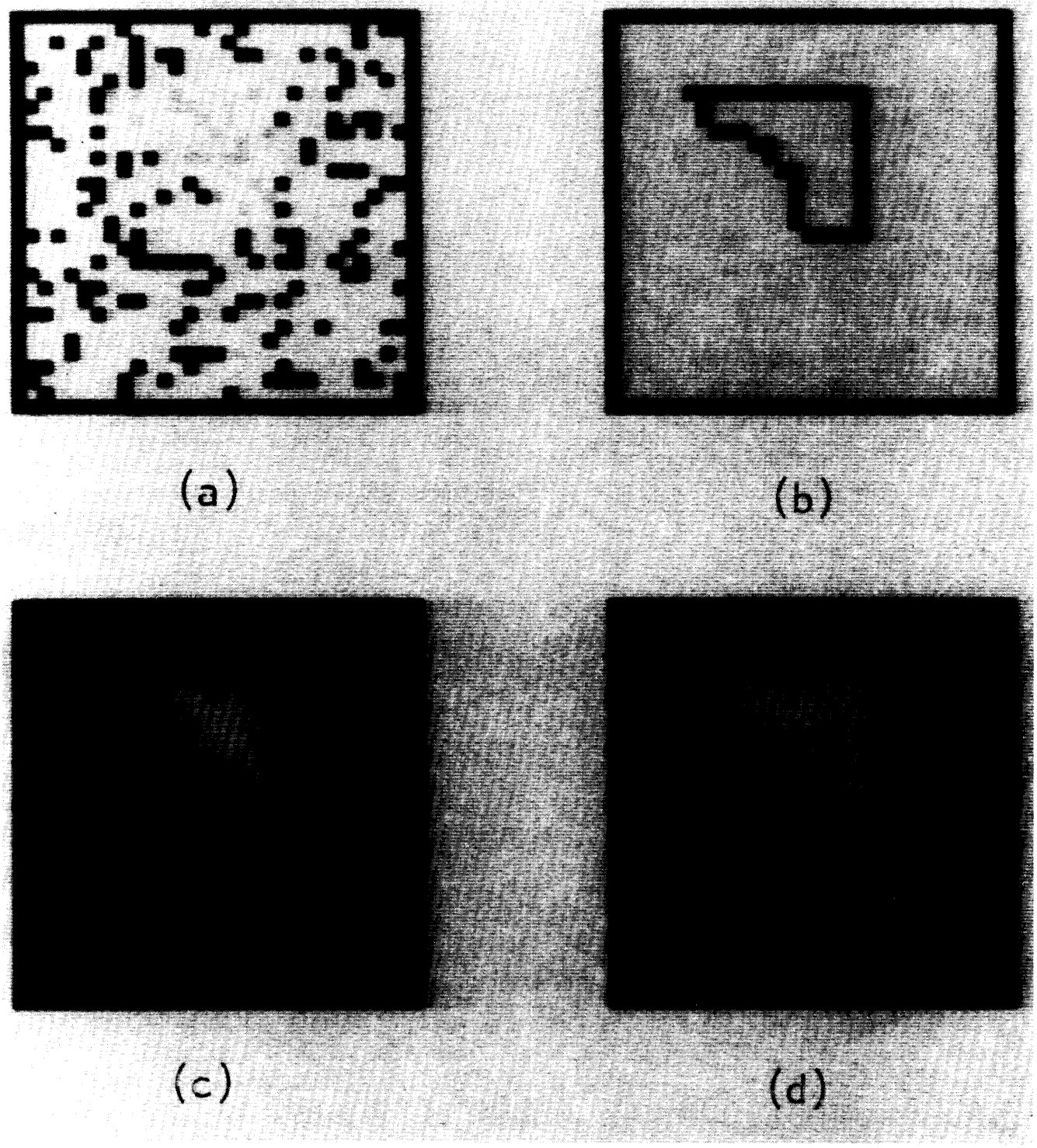

Figure 7. (a) Observations of a triangle at height 2.0 over a background at height 1.0. (a white pixel means that the observation is absent at that point). (b) Boundaries found by the Algorithm. (c) Interpolation with no boundaries (d) Reconstructed surface.

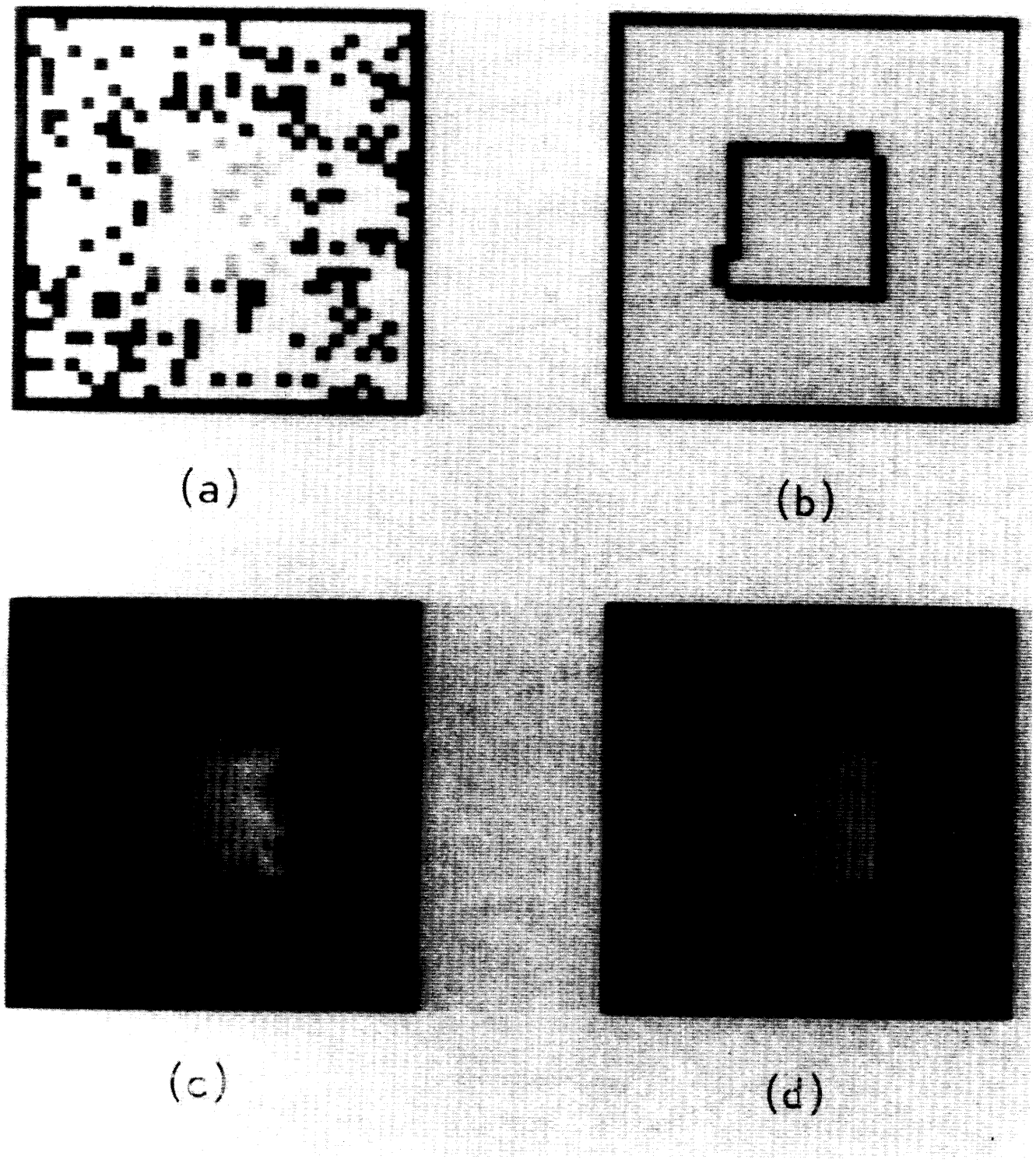

Figure 8. (a) Observations of a tilted square (slope = 0.1) over a background at height 1.0 with added white Gaussian noise ($\sigma = 0.1$) (a white pixel means that the observation is absent at that point). (b) Boundaries found by the Algorithm. (c) Interpolation with no boundaries. (d) Reconstructed surface.

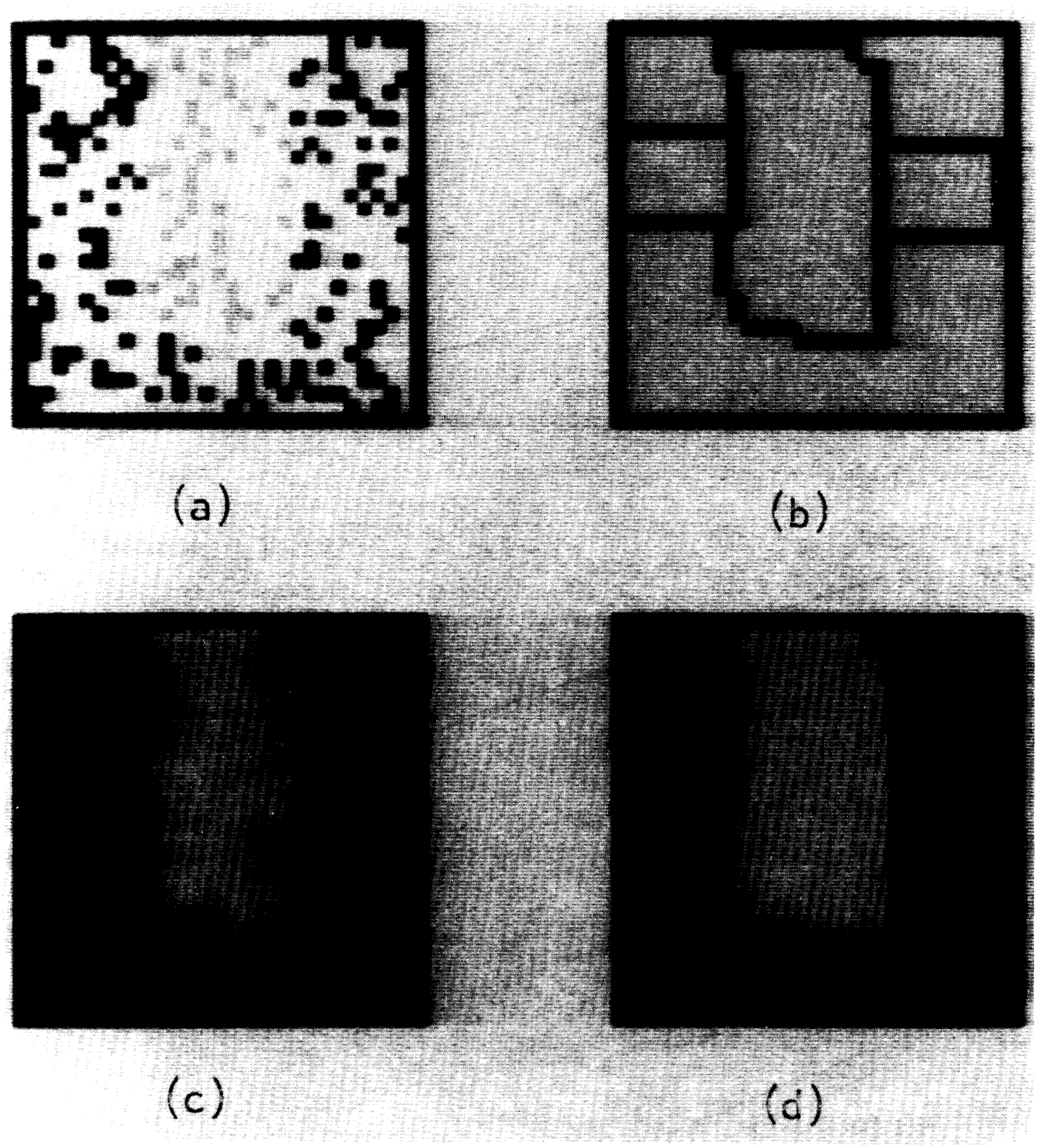

Figure 9. (a) Observations of 3 rectangles at heights 2.0, 2.0 and 3.0 over a background at height 1.0 (a white pixel means that the observation is absent at that point). (b) Boundaries found by the Algorithm. (c) Interpolation with no boundaries. (d) Reconstructed surface.

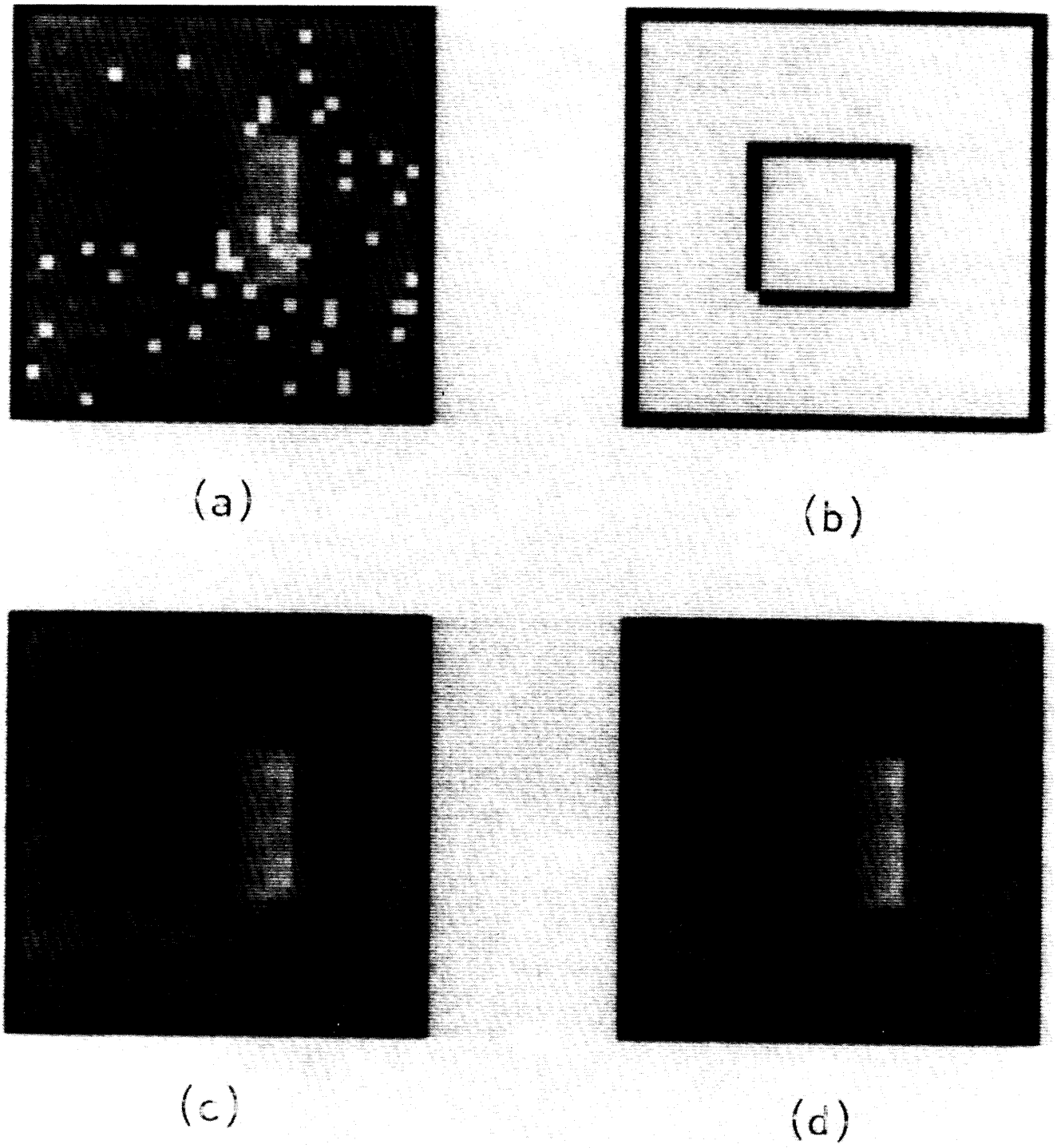

Figure 10. (a) Observations of a tilted square (slope = 0.2) over a background at height 1.0 with added white Gaussian noise ($\sigma = 0.1$). White pixels denote missing observations. (b) Boundaries found by the Algorithm. (c) Interpolation with no boundaries. (d) Reconstructed surface.

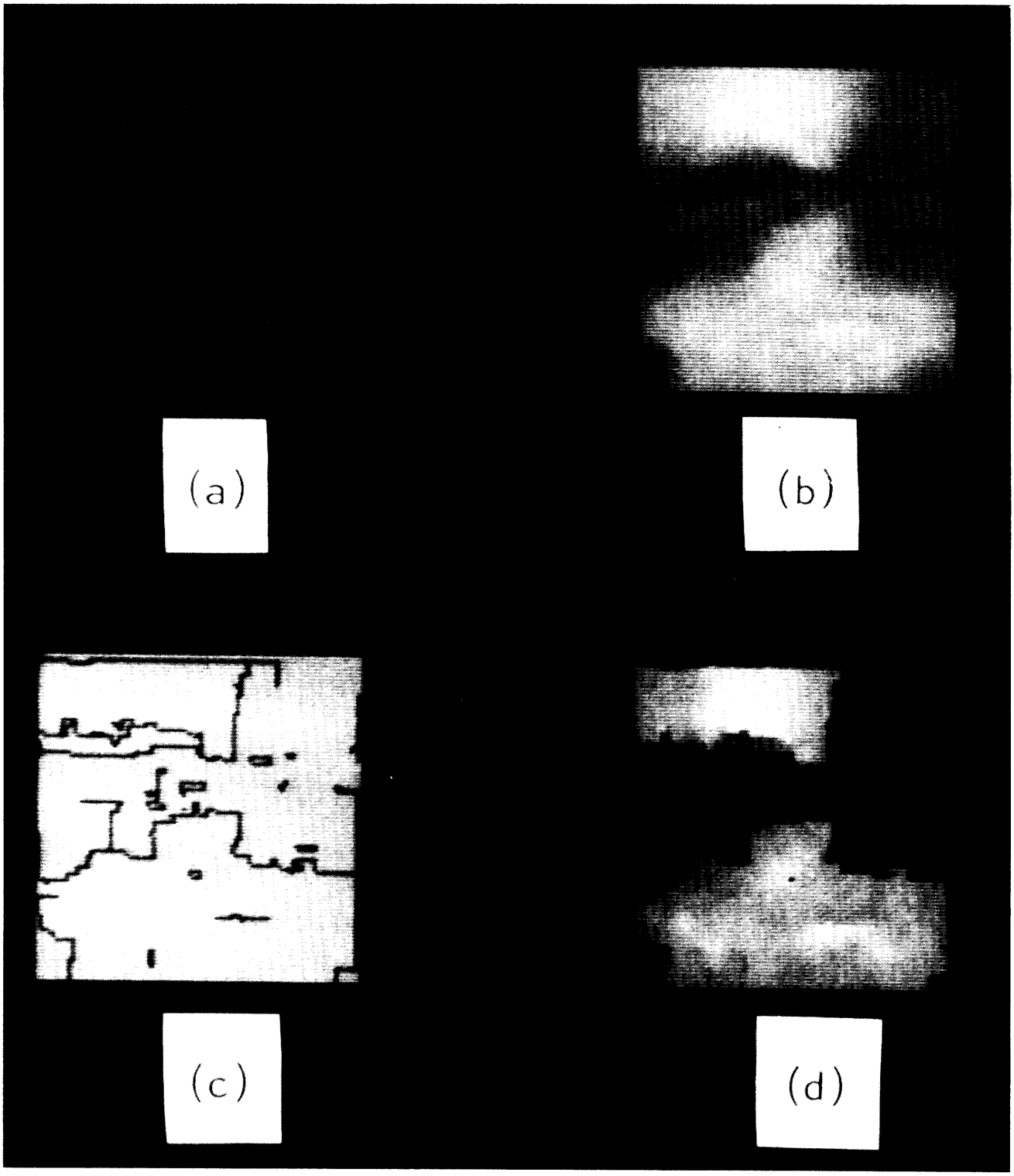

Figure 11. (a) Disparity data for a stereo pair of aerial photographs.(b) Interpolation with no boundaries. (c) Boundaries found by the Algorithm (d) Reconstructed surface.

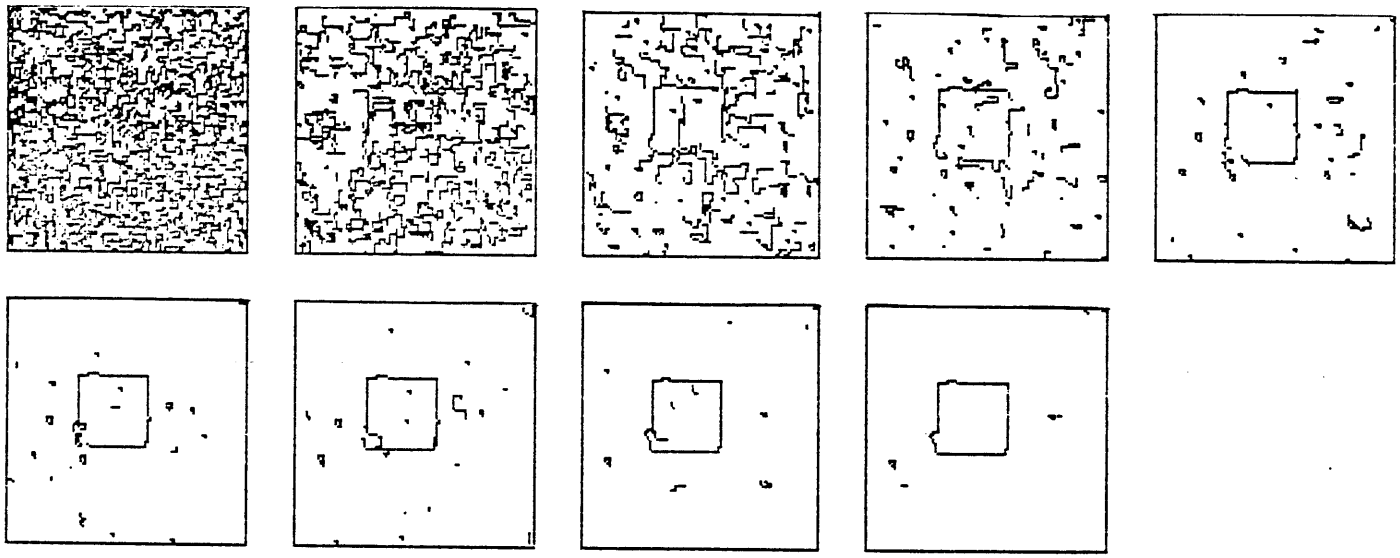

Figure 12. Snapshots of the Annealing process.

can be greatly improved if we periodically set the state of the line process using an estimate of the final (lowest energy) state.

To get this estimate, we use the ergodicity of the process ([G1], theorem C), and compute time statistics about its evolution. Specifically, we estimate the marginal distributions of the states of every "line" cell by computing, within a time window, the percentage of the time for which this cell is "on". If it is more than half of the time, the state of this cell in the annealing network is forced to 1; otherwise, it is forced to 0. As the temperature decreases, the probability distribution of the global states becomes peaked, and the maximum probability estimate obtained from the marginals gets closer to the true final state.

The results discussed in the previous section were obtained as steady states of this modified process, using a time window of 10 iterations (for the experiment of Fig. 10, an additional window of 100 iterations was used). In all cases, the steady state was obtained after approximately 450 iterations. Fig. 12 shows snapshots, taken every 50 iterations, of the state of the line process for the surface reconstruction problem on a 100 X 100 lattice, with 2000 observations, portraying a central square figure at constant height over a uniform background.

5. Discussion.

The results we have presented indicate the plausibility of using variational principles that include our prior knowledge about the geometry of the discontinuities of a piecewise smooth surface, to reconstruct it from sparse and noisy data, preserving the boundaries between continuous subregions. This method may also have a more

general significance, since most problems in early vision are ill-posed and must be regularized [P2]. The regularization methods used so far are based on smoothness assumptions, and for this reason, all have problems when dealing with functions that are only piecewise continuous. In this sense, the approach we have presented may be regarded as an extension of standard regularization methods for handling discontinuities (see [P2]). The algorithms presented are admittedly slow. Their computational efficiency, however, can be greatly improved by the use of parallel hardware, particularly in view of the local nature of the support for the state updating operations (equations (9) and (10)). It would be interesting to investigate the implementation of this algorithm on the "Connection Machine", currently under construction at the A.I. laboratory [H1] (see also [F1]).

5.1 Relation to other Boundary Preserving Surface Reconstruction Techniques.

The use of MRF models is closely related to the use of physical models for the surfaces that one wants to reconstruct. Thus, for example, Terzoupulos [T2] proposes the use of a thin plate model to embody the knowledge about the smoothness of the surface. A threshold on the bending moment of this plate is then used to locate the discontinuities, and the plate is allowed to break at these points. This technique is computationally more efficient, since the energy functionals involved are convex; however the use of the method we are proposing seem to have some definite advantages:

(i) From a conceptual viewpoint, it is better to perform the interpolation and boundary detection tasks at the same time, rather than approximating an everywhere smooth surface first, since this operation hides the discontinuities that one then tries to find in the second phase.

(ii) In our method, the values of the parameters depend only on the average height of the jumps that one wants to consider as boundaries *in the reconstructed surface*, and thus, they are independent of the location of the observations. If these are sparsely located, the bending moment of the thin plate may be low, even when the discontinuity is relatively large, and the threshold method may fail.

(iii) A priori knowledge about the shape, orientation and position of the discontinuities can be easily incorporated by choice of the potentials of the line process.

(iv) The same algorithm can be used for surface interpolation, noise elimination (smoothing) and boundary detection.

5.2 Extensions and Open Problems.

This method can be easily extended to the case of piecewise smooth (not necessarily straight) boundaries by using a 4-valued line process to include edge orientation (see [G1]). This extension will increase its practical value, and permit

its use in a variety of Computer Vision tasks (processing of depth, motion and brightness data) as well as in other fields (geophysics, medicine, etc.)

Another interesting extension is the use of the knowledge about the position of the boundaries provided by some process (for example, the output of a conventional edge detector operating on the components of a stereo pair) to influence the reconstruction of a different related surface (such as the disparity surface obtained from a stereo matcher). This can be easily accomplished, in principle, by modifying the potentials of the line process, so that the presence of a related edge at a line location lowers the energy of the corresponding configurations. We are currently working on these developments.

An important problem that has to be addressed to make this technique more practical, is the improvement of its computational efficiency. One needs to accelerate the convergence of the annealing scheme, and also, see if it is possible to exploit the structure of the energy functional of this particular problem, to develop more efficient (possibly deterministic) algorithms to find its global minimum.

From a theoretical point of view, it would be interesting to establish in a rigorous way the conditions for the convergence of the "mixed" annealing strategy, since it may be applicable to a wider class of optimization problems.

Another open question is the determination of the optimal values for the parameters of the energy functionals. It will be interesting to explore in this connection the use of statistical methods [B1,K3], regularization techniques [P2,T1,W1], and learning algorithms [H2].

Acknowledgements: I want to thank Whitman Richards, Alan Yuille, and specially Sanjoy Mitter for useful discussions. Tomaso Poggio suggested the use of Geman's approach for the case of surface reconstruction, and the connection between this technique and regularization methods. He also provided the motivation for writting this paper, and, together with Shimon Ullman, made many valuable suggestions concerning its final form. The data for the example of Fig. 11 were provided by Eric Grimson.

Notes.

[1] A Markov random field, a generalization of the concept of a Markov chain, is formally defined as follows (see [G1,K2]):

Let S be a set of sites, and $G = \{G_s, s \in S\}$ be a neighborhood system for S, i.e., a collection of subsets of S for which:

(i) $s \notin G_s$.

(ii) $s \in G_r$ if and only if $r \in G_s$.

Let $F = \{F_s, s \in S\}$ be any family of random variables indexed by S, and Ω be the set of possible configurations for F. F is a MRF with respect to G if:

(i) $P(F = f) > 0$, for all $f \in \Omega$.

(ii) $P(F_s = f_s \mid F_r = f_r \quad r \neq s) = P(F_s = f_s \mid F_r = f_r \quad r \in G_s)$.

for every $s \in S$.

[2] Note that whereas the functions defining valid conditional probabilities for a MRF cannot be chosen arbitrarily, and are, in general, very difficult to specify [B1,G1], the form of the potentials V_C is not restricted in any way, and can be used freely to specify the required behaviour of the field f (which is what one does in practice). The relation between these potentials and the conditional probabilities is given by the following formula (which follows from Bayes rule):

$$P(F_i = f_i \mid F_j = f_j, j \neq i) = \frac{\exp[-\frac{1}{\beta} \Sigma_{C:i \in C} V_C(f)]}{\Sigma_{q \in Q} \exp[-\frac{1}{\beta} \Sigma_{C:i \in C} V_C(f^q)]}$$

where Q is the set of allowable values for the state of F_i, and f^q is the configuration which is equal to q at site i, and coincides with f everywhere else.

[3] The annealing process can be considered as a Markov chain (the discrete time parameter is the number of global iterations) with states corresponding to the global states of the field. Since the transition probabilities (defined by Metropolis Algorithm) are not constant (they change with the temperature which in turn depends on the iteration number n), this chain is non-homogeneous. It is asymptotically ergodic in the sense that for any real valued function Y of the global state at time t, $f(t)$, we have [G1]:

$$\lim_{n \uparrow \infty} \frac{1}{n} \sum_{t=1}^{n} Y(f(t)) = \int_{\Omega} Y(\omega) dP_f(\omega)$$

where Ω is the set of allowable global states. This means that we can use time averages to estimate ensemble averages.

[4] Geman & Geman showed that convergence can be guaranteed if C is greater than $N\Delta$, where N is the total number of sites in the lattice, and Δ is the largest absolute difference in energies associated with pairs of global configurations that differ at only one site.

References.

[B1] Besag J. "Spatial Interaction and the Statistical Analysis of Lattice Systems". J. Royal Stat. Soc. B 34 75-83 (1972).

[C1] Cross G.C. and Jain A.K. "Markov Random Field Texture Models". IEEE Trans. PAMI 5 25–39 (1983).

[D1] Davis L. "A Survey of Edge Detection Techniques". Computer Graphics and Image Processing. 4 (1975) 248-270.

[F1] Fahlman S.E. Hinton G.E. and Sejnowski T.J. "Massively Parallel Architectures for AI: NETL, Thistle and Boltzmann Machines". Proc. Nat. Conf. on A.I. AAAI-83, Washington, D.C. (1983) 109-113.

[G1] Geman S. and Geman D. "Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images". Unpublished manuscript. Brown University (1983).

[G2] Grenander U. "Analyzing Images". Unpublished Manuscript. Brown University (1984).

[G3] Gidas B. "Non–Stationary Markov Chains and the Convergence of the Annealing Algorithm". Unpublished Manuscript. Rudgers University (1984).

[G4] Grimson W.E.L. "A Computer Implementation of a Theory of Stereo Vision". Phil. Trans. Royal Soc. Lond. B 298 (1982).

[H1] Hillis D. "The Connection Machine". A.I. Memo 646 MIT (1981).

[H2] Hinton G.E. Sejnowski T.J. and Ackley D.H. "Boltzmann Machines: Constraint Satisfaction Networks that Learn" Tech. Rep. CMU-CS-84-119 Carnegie-Mellon Univ. Dept. of Comp. Sc. (1984).

[K1] Kirkpatrick S. Gelatt C.D. and Vecchi M.P. "Optimization by Simulated Annealing". Science 220 (1983) 671–680.

[K2] Kindermann R. and Snell J.L. "Markov Random Fields and their Applications". Vol 1. Amer. Math. Soc. (1980).

[K3] Kashyap R.L. and Chellappa R. "Estimation and Choice of Neighbors in Spatial Interaction Models of Images". IEEE Trans. on Info. Theory 29 60-72 (1983).

[M1] Metropolis N. et. al. "Equation of State Calculations by Fast Computing Machines". J. Phys. Chem. 21 6 (1953) 1087.

[M2] Marr D. and Poggio T. "A Theory of Human Stereo Vision". Proc. Roy. Soc. Lond. B 204 (1979) 301–328.

[P1] Preston C.J. "Gibbs States on Countable Sets". Camb. Univ. Press (1974).

[P2] Poggio T. and Torre V. "Ill–Posed Problems and Regularization Analysis in Early Vision". A.I. Memo 773 MIT (1984).

[T1] Tikhonov A.N. and Arsenin V.Y. "Solutions of Ill–Posed Problems". Winston & Sons. (1977).

[T2] Terzoupulos D. "Multiresolution Computation of Visible-Surface Representations". Ph. D. Thesis. Dept. of E.E. & C.S. MIT (1984).

[T3] Torre V. and Poggio T. "On Edge Detection". A.I. Memo 768 MIT (1983).

[W1] Wahba G. "Ill-posed Problems: Numerical and statistical methods for mildly, moderately and severely ill-posed problems with noisy data". Tech. Rep. 595 Univ. of Wisconsin, Madison (1980).