International Journal of Computer Vision, 3, 17–32 (1989)
© 1989 Kluwer Academic Publishers, Boston, Manufactured in The Netherlands

Stochastic Stereo Matching over Scale*

STEPHEN T. BARNARD
Artificial Intelligence Center, SRI International, Menlo Park, California 94025

Abstract

A stochastic optimization approach to stereo matching is presented. Unlike conventional correlation matching and feature matching, the method provides a dense array of disparities, eliminating the need for interpolation. First, the stereo-matching problem is defined in terms of finding a disparity map that satisfies two competing constraints: (1) matched points should have similar image intensity, and (2) the disparity map should vary as slowly as possible. These constraints are interpreted as specifying the potential energy of a system of oscillators. Ground states are approximated by a new variant of simulated annealing, which has two important features. First, the microcanonical ensemble is simulated using a new algorithm that is more efficient and more easily implemented than the familiar Metropolis algorithm (which simulates the canonical ensemble). Secondly, it uses a hierarchical, coarse-to-fine control structure employing Gaussian or Laplacian pyramids of the stereo images. In this way, quickly computed results at low resolutions are used to initialize the system at higher resolutions.

1 Introduction

Few problems in computational vision have been investigated more vigorously than stereo. Compared to other modes of depth perception, stereo vision seems relatively straightforward. The images received by two eyes are slightly different due to binocular parallax; that is, they exhibit a disparity that varies over the visual field, and that is inversely related to the distance of imaged points from the observer. If we can determine this disparity field, we can measure depth and mimic human stereo vision.

This paper describes an approach to stereo in which the matching problem is posed as computational analogy to a thermodynamic physical system. The state of the system encodes a disparity map that specifies the correspondence between the images. Each such state has an energy that provides a heuristic measure of the quality of the correspondence. To solve the stereo-matching problem, one looks for the ground state; that is, the state (or states) of lowest energy.

The remainder of this section briefly discusses the major approaches to stereo matching. In section 2 the model system for a stochastic optimization method is defined. Section 3 describes how a stochastic technique called simulated annealing can be used to perform the optimization and introduces a new, more efficient variety of simulated annealing that operates over a sequence of increasingly finer scales. Several experimental results are given in section 4. Section 5 concludes with some observations.

1.1 Background

The development of computational models of stereo vision has been guided by both scientific

*Support for this work was provided by the Defense Advanced Research Projects Agency under contracts DCA 76-85-C-0004 and MDA 903-83-C-0084.

and technological motivations. The modularity of stereopsis in the human visual system, conclusively demonstrated by random-dot stereograms, indicates that this perceptual function can be studied in isolation. If the same computational principles used for stereo also apply to other modes of perception a successful model of stereo could suggest models for other vision problems. Stereo finds important practical applications in mapping and robot sensing.

1.2 Correlation

Perhaps the most obvious approach to stereo matching, loosely called "correlation," is to choose intensity patches in one image and then to search for the best matching location in the other image, typically using normalized cross-correlation as a measure of similarity or mean-square difference as a measure of dissimilarity. Many variations of this basic theme have been explored.

This general approach suffers from some difficult problems.

1. The size of the patches affects the likelihood of false matches. A patch must be large enough to contain the information necessary to specify another patch unambiguously; or, failing this, some additional means of disambiguating false matches must be used.
2. At the same time, the patches must be small compared to the variation in the disparity map. If the patches are too large the system will be insensitive to significant relief in the scene. These problems have motivated the use of scale hierarchies. (See 1.4 below.)
3. In typical images much of the area consists of uniform or slowly varying intensity, and correlation will not be sensitive in such cases. In practice, a correlation method can provide only a relatively sparse set of correspondences, from which a dense map must then be interpolated.

1.3 Feature Matching

Another approach is to attempt matching only on information-rich points. Even in correlation methods, an interest operator is often used to screen patches. The feature-matching approach seeks to establish correspondences directly between discrete sets of points—typically, the output of an edge detector, such as zero-crossing contours.

This approach suffers from similar difficulties:

1. The support of the feature detector affects the likelihood of false matches. Zero-crossings from high-frequency bands will probably have many ambiguous matches for significant ranges of disparity.
2. The support of the feature detector must be small compared to the variation in the disparity map (caused by relief in the scene) if the 2D features are to locate 3D features accurately.
3. Feature matching provides sparse matches by definition.

1.4 Scale Hierarchy

Disparity scales linearly. This suggests that a stereo matcher can begin its search at a coarse scale, find coarsely quantized disparities, use this result to initialize its search at a finer scale, and so on. In addition to improving efficiency by limiting the effective search space of the matcher, this technique ameliorates the false-target problem.

Computational vision models using hierarchies of scale are too numerous to list here. Coarse-to-fine control strategies have been used in both correlation and feature-matching models of stereo vision. Terzopoulos [1] gives a clear and concise description of multigrid relaxation methods. In particular, he discusses the difficulties of applying the methods to nonconvex variational problems, and suggests the possibility of stochastic multigrid methods.

1.5 Lattice and Variational Models

Several models of stereo vision fit neither the correlation nor the feature-matching paradigms; instead, they pose the matching problem in terms of optimizing a global measure [2, 3, 4]. To take one example, Julesz [3] proposed a model consisting

of two lattices of spring-loaded magnetic dipoles, representing the two images of a random-dot stereogram. The polarity of the dipoles represents whether pixels in the left and right images are black or white. A state of global fusion is achieved in the ground state, with the attraction or repulsion of the dipoles balanced by the forces of the springs.

More recently, Poggio et al. [5] have proposed a regularization criterion based on minimizing the following quantity:

$$\mathscr{E} = \iint \{[\nabla^2 G \circ (I_L(x, y) - I_R(x + D(x, y), y))]^2 + \lambda(\nabla D)^2\} dx\, dy \qquad (1)$$

where I_L and I_R are continuous intensity functions in the left and right visual fields, $\nabla^2 G$ is a linear bandpass filter (Laplacian of a Gaussian), $\circ$ is the convolution operator, ∇D is the gradient of disparity, and λ is a constant. (Note that $(\nabla D)^2$ is interpreted as $\nabla D \cdot \nabla D$, or the square of the magnitude of the disparity gradient.) Equation (1) can be justified in terms of two heuristics: the first term in the integrand is a measure of photometric difference; the second is a measure of the first-order variation in the disparity map. In this heuristic sense it is similar to the Julesz spring-dipole model, with the two terms corresponding to the potential energy of the dipoles and the springs, respectively. Of course, equation (1) has the advantage of being precise, as well as addressing the case of continuous intensity.

Witkin et al. [6] described a method for optimizing a generalization of (1) that is essentially a sophisticated form of gradient descent which tracks the solution over increasingly finer scales. The hope is that $\mathscr{E}$ is convex at a coarse scale and that relatively coarse intermediate solutions will place the system in the correct convex region at finer scales. They report that the method is prone to error when it encounters bifurcations in its trajectory. As the scale becomes finer the system must "choose" which path to follow, and it cannot recover from a mistake because $\mathscr{E}$ may never increase. The solution is therefore critically dependent on initial conditions. This paper presents an alternative stochastic method that can cope with this problem.

2 The Model System

2.1 Epipolar Camera Model

We assume that two coplanar images $\mathscr{L}(x,y)$ and $\mathscr{R}(x,y)$ are formed by central projection with focal length f, and with the centers of projection separated by distance B along a baseline parallel to a common focal plane. The camera coordinate systems are right handed, with origins in the focal plane and z axes pointing toward the observer. For convenience, we assume that $0 \leqslant x,y < 1$. If a point (x,y,f) in the left image matches point (x',y,f) in the right image we say that it has disparity $d = x - x'$. Note that the 3D coordinates of the imaged point with respect to the left camera are

$$\mathbf{p} = \left(\frac{B}{d}\, x, \frac{B}{d}\, y, \frac{B}{d}\, f\right)$$

Under these conditions, the disparities are restricted to the horizontal (x) direction. This assumption involves no loss of generality, because if the relative positions and orientations of the two cameras are known, as well as the internal camera parameters, correspondences are restricted to epipolar lines. If the epipolar lines are not horizontal the images can be mapped into a normal stereo pair in which they are.

2.2 Cyclopean Disparity Map

At this point we identify $\mathscr{L}$ and $\mathscr{R}$ with $\nabla^2 G \circ I_L(x,y)$ in (1). We seek a disparity map, $\mathscr{D}(x, y)$, defined over the same interval as $\mathscr{L}$ and $\mathscr{R}$, which specifies the correspondence between $\mathscr{L}$ and $\mathscr{R}$. The cyclopean representation (see [7]) defines $\mathscr{D}$ with the following relation:

$$\mathscr{L}\left[x - \frac{\mathscr{D}(x,y)}{2}, y\right]$$

corresponds to

$$\mathscr{R}\left[x + \frac{\mathscr{D}(x,y)}{2}, y\right]$$

The major advantage of the cyclopean representation is that, by defining disparity without prefer-

ence to either image, it allows a more uniform treatment of occlusion boundaries.

Rewriting the integrand of equation (1) in the cyclopean form we have:

$$\mathcal{E}(x,y) = \left[\mathcal{L}\left(x - \frac{\mathcal{D}(x,y)}{2}, y\right) - \mathcal{R}\left(x + \frac{\mathcal{D}(x,y)}{2}, y\right)\right]^2 + \lambda[\nabla\mathcal{D}(x,y)]^2 \quad (2)$$

$$= \mathcal{E}_1(x,y) + \lambda\mathcal{E}_2(x,y) \quad (3)$$

and the quantity to be minimized is

$$\mathcal{E}(\mathcal{D}) = \iint \mathcal{E}(x,y)dx\,dy \quad (4)$$

2.3 A Spring Model

To establish a concrete idea of the meaning of (3) and (4), consider the spring model illustrated in one dimension in figure 1.

The model consists of two surfaces, $\mathcal{R}(x,y)$ below and $\mathcal{L}(x,y) + S_1$ above. Midway between these surfaces is a lattice of pivot points, and at each such point is an elastic lever arm, with rest length S_1 and spring constant k_1. The lever arms are free to rotate in the (x,z) plane (i.e., in epipolar planes), while their endpoints are constrained to lie on the two surfaces. The lever arms are connected to their neighbors by other springs with spring constant k_2 which exert torques over moment arm A. The angles of the lever arms represent disparity on an M^2 cyclopean lattice:

$$D(i,j) = \mathcal{D}(x_i,y_j), \quad 0 \leqslant i,j < M$$

with

$$D(i,j) \approx S_1 \sin\theta_{i,j} \approx S_1\theta_{i,j}$$

The potential energy stored in a lever arm is approximately[1]

$$\mathcal{H}_1(i,j) \approx \frac{1}{2}k_1\left[\mathcal{L}\left(x - \frac{D(i,j)}{2}, y\right) - \mathcal{R}\left(x + \frac{D(i,j)}{2}, y\right)\right]^2$$

and in a connecting spring is approximately

$$\mathcal{H}_2(i,j,k,l) \approx \frac{1}{2}\ k_2\left(\frac{A}{S_1}\right)^2\ [D(i,j) - D(k,l)]^2$$

The energy associated with a single lattice point is

$$\mathcal{H}(i,j) \approx \mathcal{H}_1(i,j) + \frac{1}{2}\sum_{(k,l)\in\mathcal{V}_{i,j}} \mathcal{H}_2(i,j,k,l) \quad (5)$$

where $\mathcal{V}_{i,j}$ is the set of neighbors of (i,j). The factor of $\frac{1}{2}$ in the second term is due to the fact that the energy of each connecting spring is shared by two lattice points (neglecting boundary conditions for simplicity).

If we take $\mathcal{V}$ to be the four nearest neighbors, the sum in (5) approximates the squared magnitude of the disparity gradient:

$$\sum_{(k,l)\in\mathcal{V}_{i,j}} \mathcal{H}_2(i,j,k,l) \approx k_2\left(\frac{A}{S_1}\right)^2[\nabla D(i,j)]^2$$

The energy of the entire system is

$$\mathcal{H}(D) = \sum_{i,j} \mathcal{H}(i,j)$$

Comparing terms between $\mathcal{H}$ and $\mathcal{E}$, we have approximately

$$\mathcal{H}(D) \propto \mathcal{E}(\mathcal{D})$$

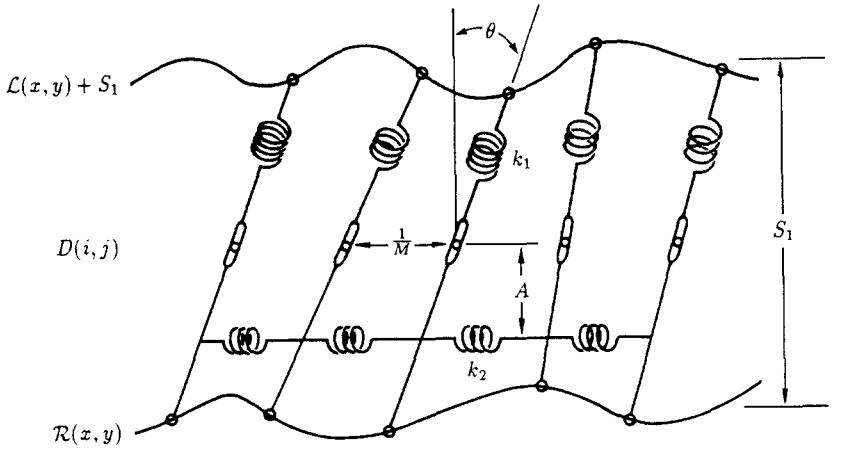
with

$$\lambda = \left(\frac{k_2}{k_1}\right)\left(\frac{A}{S_1}\right)^2$$

We can therefore interpret $\mathcal{E}$ as a Hamiltonian specifying the energy of the spring system, neglecting kinetic energy terms. The constant λ is proportional to the relative stiffness of the two types of springs.

A physical realization of the spring model would be a dynamic system of oscillators that would follow a trajectory through a $2M^2$ dimensional phase space. (Each lever arm has two degrees of freedom: θ and $\dot{\theta}$.) We could flesh out

[1]These formulas require a small-angle approximation $\sin(\theta \approx \theta$. Note that the angles can be made arbitrarily small by increasing S_1.



vertical springs: spring constant k_1, rest length S_1

horizontal springs: spring constant k_2, rest length $S_2 = \frac{1}{M}$

Fig. 1. Spring model

this model by specifying the moments of inertia and damping coefficients of the lever arms. We could also add a periodic forcing function to add energy to the system, balancing the energy dissipated by damping. Having done this, we could write the differential equations of motion describing the model's deterministic dynamic behavior. In principle, we could trace the trajectory of the system through its phase space, gradually reducing the amplitude of the forcing function while keeping the system in dynamic equilibrium. There is little point in simulating the dynamics in such detail, however, since we know that even low-dimensional forced oscillators have chaotic attractors [8]. The dynamics will be effectively stochastic.

An alternative and much less expensive approach, which we shall take in the next two sections, is to explicitly acknowledge the stochastic, ergodic nature of the model. Kinetic energy will be modeled as heat.

3 Stochastic Optimization

We have already partially discretized equation (1) by defining the lattice D on $\mathcal{Y}$. At this point we similarly define lattices L and R on $\mathcal{Y}$ and $\mathcal{R}$. D now has integer values and is interpreted as:

$$L(i - \left[\frac{D(i,j)}{2}\right], j)$$

corresponds to

$$R(i + \left[\frac{D(i,j)}{2}\right], j)$$

Equation (2) becomes

$$E(i,j) = \left[L\left(i - \left[\frac{D(i,j)}{2}\right], j\right) - R\left(i + \left[\frac{D(i,j)}{2}\right], j\right)\right]^2 + \lambda[\nabla D(i,j)]^2 \quad (6)$$

The total potential energy is

$$E = \sum_{i,j} E(i,j)$$

In terms of the spring model, the ends of the lever arms are now constrained to lie on a finite number of positions on the two surfaces. Although this system is finite, it is also high-

dimensional, and the problem of finding minimal-energy states is still difficult because the number of possible states is vast (exponential in M^2). Furthermore, because L and R are in general nonlinear functions, there is no reason to believe that E is convex, and therefore no reason to believe that a discrete iterative-descent algorithm would work.

3.1 Standard (Canonical) Annealing

Simulated annealing is a fairly new technique for solving such combinatorial optimization problems. In section 3.3 a new variety of simulated annealing (called microcanonical annealing) is presented which has several advantages for computer implementation. In this section the basic principles of the standard form of simulated annealing are described to set a context for the introduction of microcanonical annealing. A review of simulated annealing is given by Laarhoven and Aarts [10].

The most fundamental result of statistical physics is the Boltzmann (or Gibbs) distribution:

$$Pr(E_i) = \frac{\exp(-E_i/kT)}{Z(T)}$$

which gives the probability of finding a system in state i with energy E_i, assuming that the system is in equilibrium with a large heat bath at temperature kT (k is Boltzmann's constant). The normalizing quantity in the denominator, called the partition function, is a sum over all accessible states v:

$$Z(T) = \sum_v \exp(-E_v/kT) \tag{7}$$

Physicists are generally interested in calculating macroscopic properties of model systems at various temperatures. The average value of some macroscopic variable A (which may be the average energy of the system, for example) can be written

$$\langle A \rangle = \sum_v A_v Pr(A_v) = \frac{\Sigma_v A_v \exp(-E_v/kT)}{Z(T)}$$

Unfortunately, the partition function is usually impossible to calculate.

In 1953 Metropolis et al. [9] described a Monte Carlo algorithm that generates a sequence of states which converges to the Boltzmann distribution in the limit (figure 2.) This method, which simulates the effect of allowing the system to interact with a much larger heat bath, samples what is called the canonical ensemble. Macroscopic parameters can then be calculated without knowledge of the partition function by averaging over long sequences.

The Metropolis algorithm begins in an arbitrary state and then successively generates candidate state transitions ($v \rightarrow v'$) at random. A transition is accepted with the following probability:

$$Pr(v \rightarrow v' | v, v') = \begin{cases} 1 & \text{if } \Delta E < 0 \\ \exp(-\Delta E/kT) & \text{otherwise} \end{cases} \tag{8}$$

1. Begin with the system in an arbitrary state v.
2. Make a small change to the state, typically by changing the system in only one degree of freedom. Call the new state v'.
3. Evaluate the resulting change in energy: $\Delta E = E_{v'} - E_v$.
4. If $\Delta E < 0$ (that is, the change takes the system to a state of lower energy) accept the change.
5. If $\Delta E \geqslant 0$ accept the change with the probability $\exp(-\Delta E/kT)$.
6. Repeat steps (2) through (5) until the system reaches equilibrium.

Fig. 2 The Metropolis Algorithm.

where $\Delta E = E_{v'} - E_v$. Asymptotic convergence of the Metropolis algorithm to the Boltzmann distribution is guaranteed if the process for generating candidate state transitions is ergodic.

Kirkpatrick et al. [11] and Cerny [12] independently recognized a connection between the Metropolis technique and combinatorial optimization problems. If the energy of a state is considered as an objective function to be minimized, the minimum can be approximated by generating sequences at decreasing temperatures, until finally a ground state, or a state with energy very close to a ground state, is reached at $kT = 0$. This is analogous to the physical process of annealing.

There are results showing the existence of annealing schedules (i.e., the rate of decrease of temperature) that guarantee convergence to ground states in finite time [13], but these schedules are too slow for practical use. Faster ad hoc schedules have been used in many problems with good average-case performance. While these faster schedules may not find an optimal state, they can converge to states that are very close to optimal.

The application of standard simulated annealing to the stereo matching problem is straightforward. An early version is described in [14]. (Marroquin [15] and Divko and Schulten [16] have independently described similar methods.) In the following sections a more efficient version is presented.

3.2 Annealing over Scale

Simulated annealing could be applied directly to a pair of stereo images at the finest scale, but the convergence would be rather slow if the images had a large range of disparity. This was the approach reported in [14] (using the standard annealing algorithm and a slightly different energy function).

A more efficient method is to use the coarse-to-fine strategy that has been found to be so effective in other image-matching work. At a coarse level of resolution the number of lattice sites and the range of disparity are small; therefore, the size of that state space is relatively small. We should be able to compute an approximate ground state quickly, and then use it to initialize the annealing process at the next, finer level of resolution, and so on.

The Laplacian pyramid, originally developed as a compact image- coding technique [17] offers an efficient representation for hierarchical annealing. In a Laplacian pyramid an $n \times n$ image (for convenience, assume n is a power of 2) is transformed into a sequence of bandpass-filtered copies, I_k, $k = 0, \ldots, n$, where I_k is an image of size $2^{n-k} \times 2^{n-k}$. Each image is therefore smaller than its predecessor by a factor of 1/2 in linear dimension and a factor of 1/4 in area. We will refer to I_k as the image at level k. The center frequency of the passband is reduced by one octave between levels. This transform can be computed efficiently by recursively applying a small generating kernel to create a Gaussian (low-passed) pyramid, and then differcing successive low-passed images to construct the Laplacian pyramid. The difference- of-Gaussians gives a good approximation to the $\nabla^2 G$ filter.

After constructing Laplacian pyramids from the original stereo images, disparity is reduced by a factor of 1/2 in successive scales. Therefore, at some level, disparity is small everywhere. For typical stereo images, we can take this to be level $n - 3$. (For example, if the original images were a power of 2 in linear dimension, the Laplacian images at level $n - 3$ would be 8×8 pixels. Disparities in the range of 0 to 63 pixels in a pair of 512×512 images would be reduced to 0, with truncation.) We shall start annealing at this level, find an approximate ground state, and then expand the solution to the next scale. To make this coarse-to-fine strategy work, however, we must specify how a low- resolution result is used to start the annealing process at the next-higher scale.

Expanding a low-resolution result to the next level presents a problem. Obviously, one should begin by simply doubling the size of the low-resolution lattice and doubling the disparity values. Having done this, however, the new state has a low energy but the system is not close to equilibrium. Every odd disparity value is "unoccupied," and the new map is therefore more uniform than it should be. This spurious uniformity, which is solely due to the quantization of the previous result, is likely to place the system near a

local minimum from which it will not recover. Fortunately, there is an easy solution to this problem: destroy this uniformity by adding heat. Simply run the annealing algorithm "in reverse" by adding energy instead of removing it. Heating may proceed much faster than cooling because the system relaxes to equilibrium quickly at high temperatures.

Choosing parameters of this procedure—heating and cooling rates, termination conditions, and so on—remains an art, as in virtually all applications of simulated annealing. The results in section 4 were generated with a common parameter set which was determined empirically from tests on a wide variety of images. Some of the issues governing the choice of parameters will be discussed in section 4.

3.3 Microcanonical Annealing

Creutz [18] has described an interesting alternative to the Metropolis algorithm, outlined in figure 3. Instead to simulating the effect of a large heat bath, the Creutz algorithm simulates a thermally isolated system in which energy is conserved. Samples are drawn from the microcanonical ensemble. One can imagine the difference between the Metropolis algorithm and the Creutz algorithm as follows. The Metropolis algorithm generates a "cloud" of states, each with, in general, different energies, which fills a volume of phase space. As temperature decreases this volume contracts to one or more ground states. The Creutz algorithm, by contrast, generates states on a constant-energy surface in a somewhat larger phase space. As energy decreases these surfaces shrink to the same set of ground states.

The simplest way to accomplish this is to augment the system with one additional degree of freedom, called a *demon,* which carries a variable amount of energy, E_D. This demon holds the kinetic energy of the system and, in effect, replaces the heat bath. The total energy of the system is now

$$E_{\text{total}} = E_{\text{potential}} + E_{\text{kinetic}}$$
$$= E + E_D$$

The demon energy, being kinetic, is constrained to be nonnegative. The algorithm accepts all transitions to lower energy states, adding $-\Delta E$ (the energy given up) to E_D. Transitions to higher energy are accepted only when $\Delta E < E_D$, and the energy gained is taken away from E_D. Total energy remains constant.

Microcanonical annealing simply replaces the Metropolis algorithm with the Creutz algorithm. Instead of explicitly reducing temperature, the microcanonical annealing algorithm reduces energy by gradually lowering the value of E_D. Standard arguments can be used to show that at equilibrium E_D assumes a Boltzmann distribution over time [18]:

$$Pr(E_D = E) \propto \exp(E/kT)$$

1. Begin with the system in an arbitrary state v.
2. Make a small change to the state, typically by changing the system in only one degree of freedom. Call the new state v'.
3. Evaluate the resulting change in energy: $\Delta E = E_{v'} - E_v$
4. If $\Delta E < 0$ accept the change and increase the demon energy ($E_D \leftarrow E_D - \Delta E$).
5. If $\Delta E \geqslant 0$ accept the change contingent upon E_D:
   - If $\Delta E < E_D$ accept the change and decrease the demon energy ($E_D \leftarrow E_D - \Delta E$).
   - Otherwise, reject the change.
6. Repeat steps (2) through (5) until the system reaches equilibrium.

Fig. 3. The Creutz Algorithm.

Temperature therefore emerges as a statistical feature of the system:

$$kT = \langle E_D \rangle \tag{9}$$

This simple version of microcanonical annealing, using only one demon, is not suited to a parallel implementation. Each decision to accept or reject a state transition depends on the value of E_D, and therefore on the previous decision. The computation can be made parallel by using a lattice of demons. Temperature is still measured with (9), but using the distribution of E_D over space rather than time.[2]

There is a minor complication in using a lattice of demons. The single-demon algorithm visits sites at random and the demon allows energy to be transferred throughout the lattice. Similarly, in the lattice-of-demons algorithm the demons must be mixed throughout the lattice. If this is not done energy will be transferred between sites very slowly, only through the nearest-neighbor interactions of (6). We use a complete random permutation of the demons after every lattice update, but more local methods are also adequate.

Microcanonical annealing has several advantages over standard annealing:

1. It does not require the evaluation of the transcendental function exp (x). Of course, in practice this function can be stored in a table, but we would like our algorithm to be suited to fine-grained cellular automata with very limited local memory.
2. It is easily implemented with low-precision integer arithmetic; again, a significant advantage for simple hardware implementation.
3. In the Metropolis algorithm a state transition is accepted or rejected by comparing exp ($-\Delta E/kT$) to a random number drawn from a uniform distribution over [0,1], and these numbers should be accurate to high precision. The Creutz algorithm does not require high-quality random numbers.

Experiments indicate that the Creutz method can be programmed to run an order of magnitude faster than the conventional Metropolis method for discrete systems [19].

[2]Statistics can be sampled over both time and space, if desired.

In standard annealing it is not clear how to determine when the system reaches equilibrium. One can examine fluctuations in the average energy, which should be of order $1/M^2$ at equilibrium, but this may require many extra iterations to get adequate statistics because one does not know in advance what average value to expect. In microcanonical annealing there is a simpler way. Let r_{eq} be the ratio of the observed average demon energy to the standard deviation of the same observed distribution:

$$r_{eq} = \frac{\langle E_D \rangle}{\rho(E_D)} \tag{10}$$

At equilibrium $r_{eq} \approx 1$.

As with the Metropolis algorithm, the Creutz algorithm converges to the Boltzmann distribution in the limit for any ergodic process generating candidate state transitions. Of course, different state-transition schemes will affect the rate of convergence. We have found the following simple method to be adequate:

$$Pr(d \rightarrow d') = \begin{cases} 0.5 & \text{if } |d - d'| = 1 \\ 0 & \text{otherwise} \end{cases}$$

In other words, the disparities increase or decrease by one lattice position as the system follows a Brownian path on its phase-space surface of constant energy. Only one bit it required to specify each transition.

4 Experimental Results

This section presents experimental results for three distinct cases: a medium-resolution, oblique, ground-level scene with prominent occlusions, a sparse random-dot stereogram, and a high-resolution aerial stereo pair. The method has been tested on over 30 real images; these examples have been chosen to indicate a variety of conditions.

Identical parameters were used for all three cases. Four nearest neighbors were used for $\mathcal{N}$. We used a value of 64 for λ, which works well for images quantized into eight-bit values. A schedule for heating and cooling was established to yield about 200 complete scans at each scale, with about 85 percent of the cycle devoted to cooling. A "complete scan" means that exactly one random-

state transition was considered for each lattice site. The transitions can either be generated "on the fly" or, for efficiency, stored as a small number of binary arrays, and similarly for the permutations of the demon lattice. We heat the system in increments of 10 units until the temperature (see equation (9)) exceeds 300. We cool the system in increments of 3 units until the temperature is less than 30. This choice of parameters represents an empirical worst-case. A better theoretical grounding for the parameters would be desirable, but is beyond the scope of this paper.

Figure 4 is distinguished by sharp discontinuties of depth. Figure 5 shows some intermediate results of annealing process. The hottest and coldest states for each level of resolution are shown (except for the very coarse 8 × 8 level, in which the ground state is one of uniform disparity). The method has done a reasonably good job of separating the tree from the background, which is somewhat surprising since it has no explicit representation for occlusions. In an attempt to model occlusions we have experimented with line processes, like the one used by Geman and Geman [13], and with nonlinear "springs" that weaken as they deform. Our results so far have not justified the added complexity.

The two graphs in figure 6 trace the evolution of temperature and r_{eq}. Note that the plot of r_{eq} indicates that the system moves away from equilibrium during the relatively fast heating cycles, but relaxes quickly back to equilibrium after cooling starts. The system appears to drop away from equilibrium at low temperatures according to the r_{eq} plot, but this effect is actually because there are relatively few energy levels available to the demons near the ground state.

Figure 7 is a 10%, 128×128 random-dot stereogram with four depth planes separated by intervals of 2 pixels of disparity. Figure 8 shows the final disparity assignments.

Figure 9 is a relatively high-resolution aerial stereo pair. The left image is 512×512, while the right is somewhat wider to ensure that the every pixel in the left image matches some point in the right. Figure 10 shows the matching results in two forms. Figure 10(a) is the disparity map with contours at every 5th disparity level. Figure 10(b) is a synthetic view of the scene that was created by transforming the disparity map to terrain elevations, texture mapping the left image onto this surface, and then rendering the surface. This data, which has a disparity range of 72 pixels, is the largest problem we have attempted so far. It required about 12 hours of processing on a Symbolics 3645. Inspection of this map indicates that it is accurate to 1 pixel over the entire field.

In earlier work [14, 20] we used the absolute dif-

Fig. 4. An oblique stereogram with occlusion.

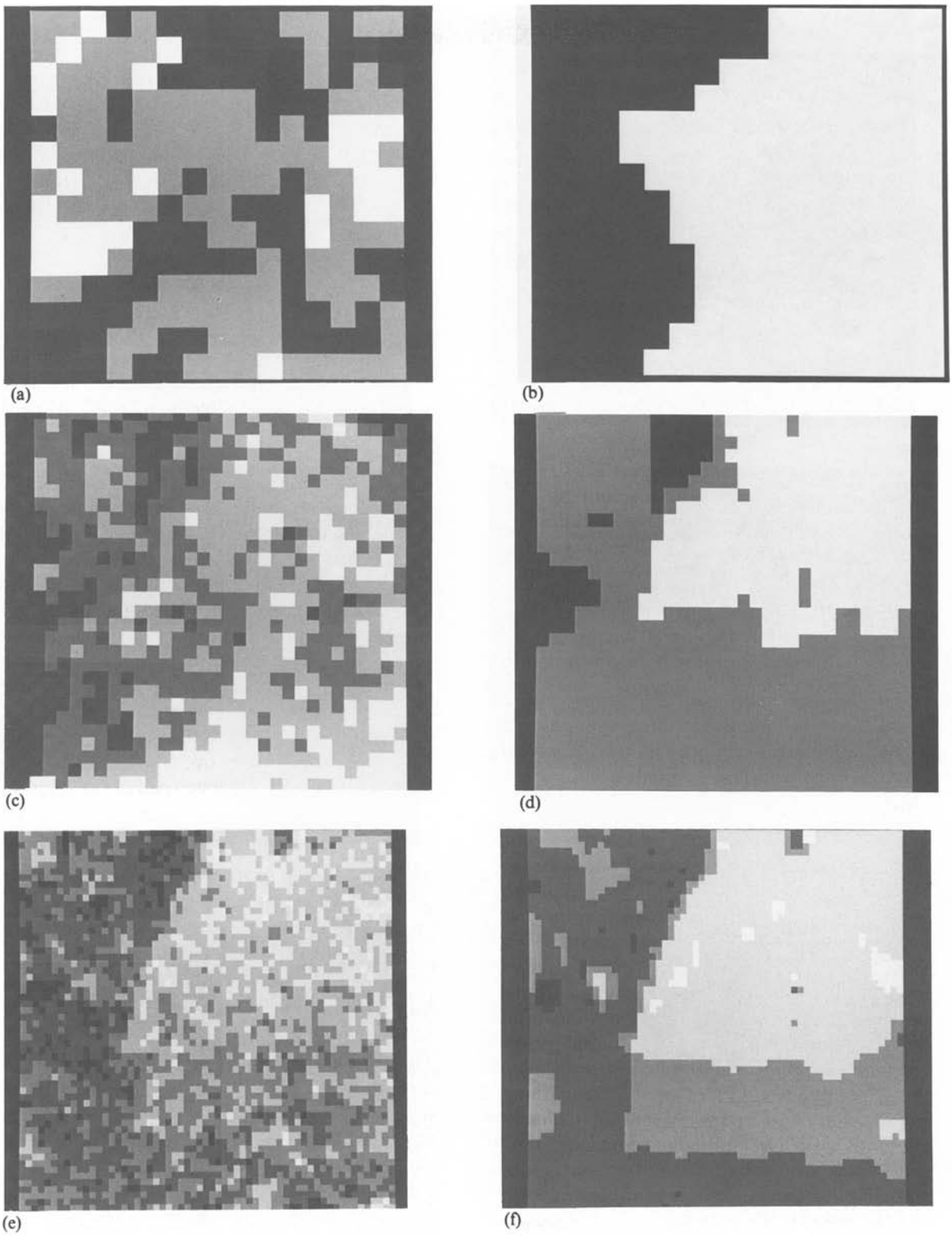
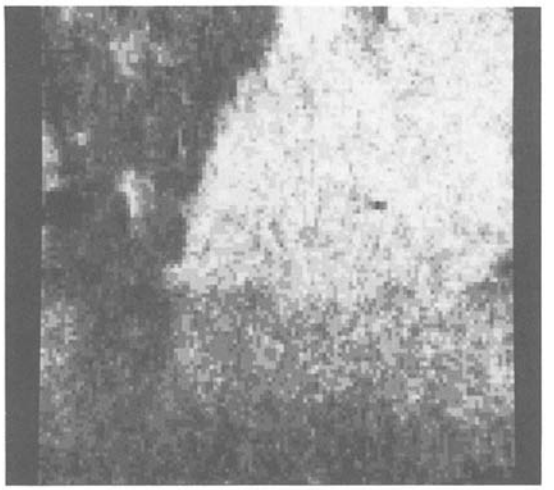
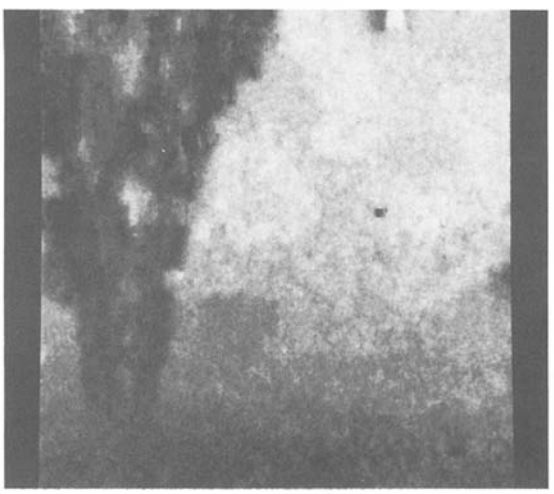


Fig. 5. Results for figure 4. (a) Hot 16×16. (b) Cold 16×16. (c) Hot 32×32. (d) Cold 32×32. (e) Hot 64×64. (f) Cold 64×64. (g) Hot 128×128. (h) Cold 128×128. (i) Hot 256×256. (j) Cold 256×256.



(g)

(h)



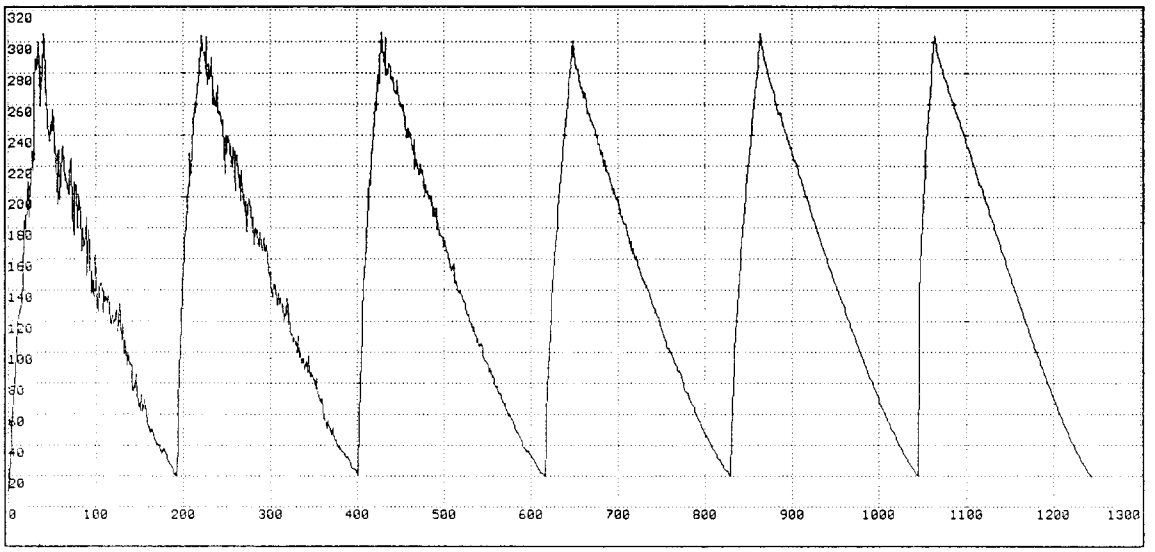
(i)

(j)

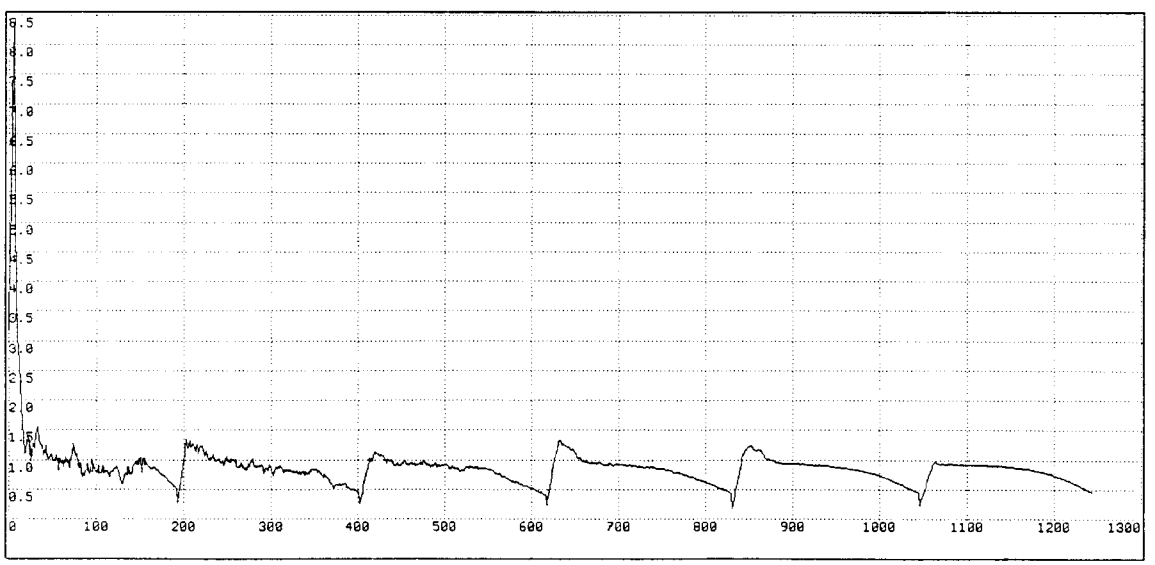
ference instead of the squared difference in (3). This is slightly more efficient to implement and makes little difference in the results. In terms of the spring model, this would correspond to using springs that exert a constant force in opposite direction of their deformation. Of course, a different value of λ is required. We have also experimented with eight-neighbor versions of $\mathcal{N}$, but no significant improvement in performance was observed.

5 Conclusions

The major conclusion we can draw is that equation (1) is an adequate criterion for stereo matching, even in scenes with abrupt occlusions. The results in figures 4 and 5 indicate that solutions can accommodate very abrupt changes in depth. Residual energy in the near-optimal states is concentrated along steep disparity contours rather than spread over large areas.



(a)



(b)

Fig. 6. Plots of T and r_{eq} for figures 4 and 5. (a) T vs. number of iterations. (b) r_{eq} vs. number of iterations.

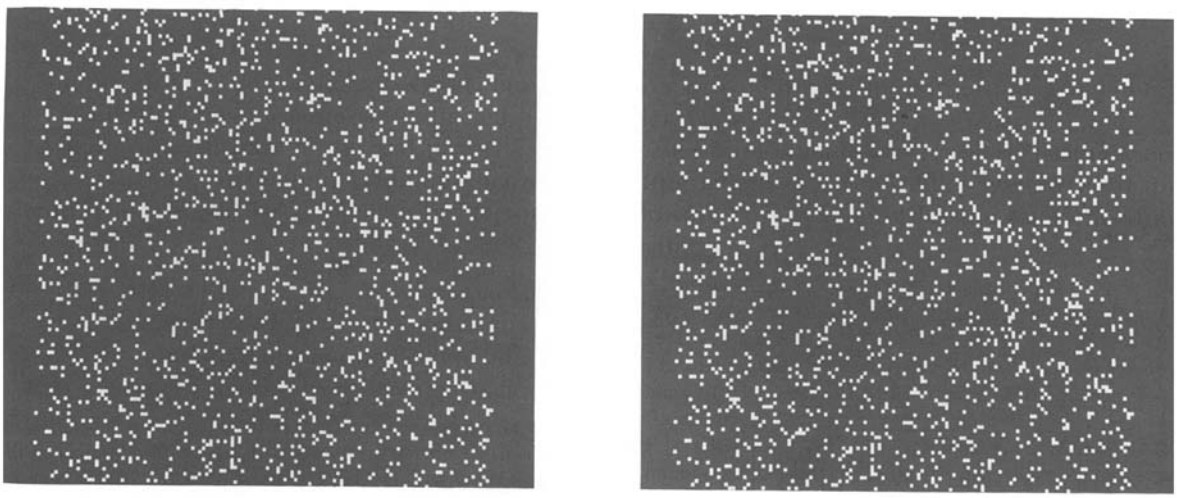


Fig. 7. A 10% random-dot stereogram.

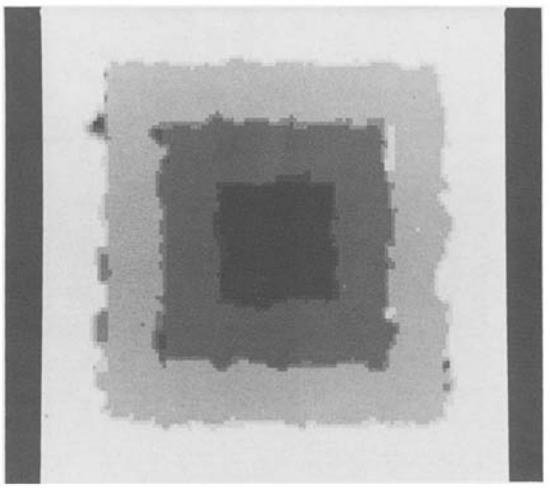


Fig. 8. Results for figure 7.

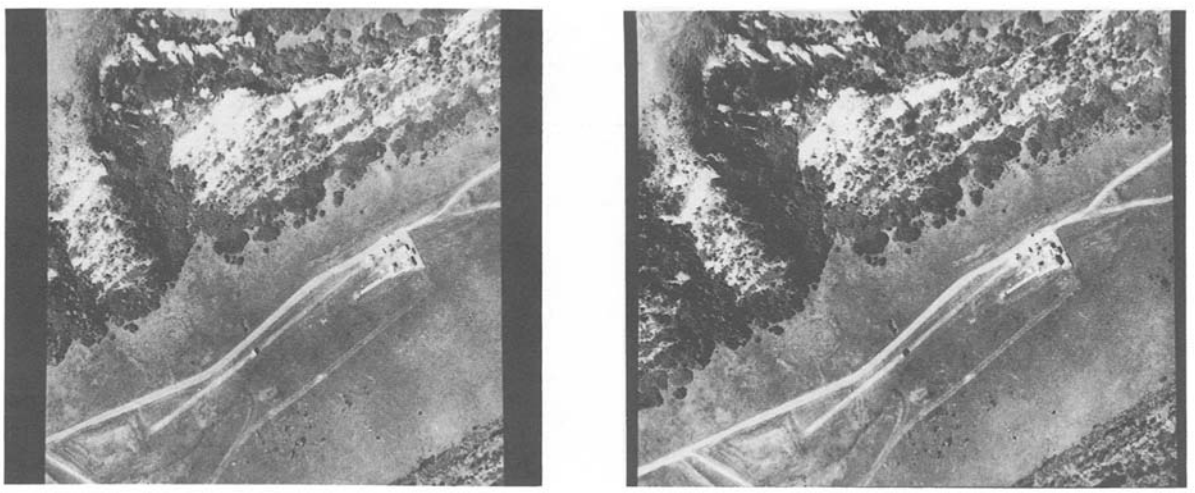


Fig. 9. A 512×512 aerial stereogram.

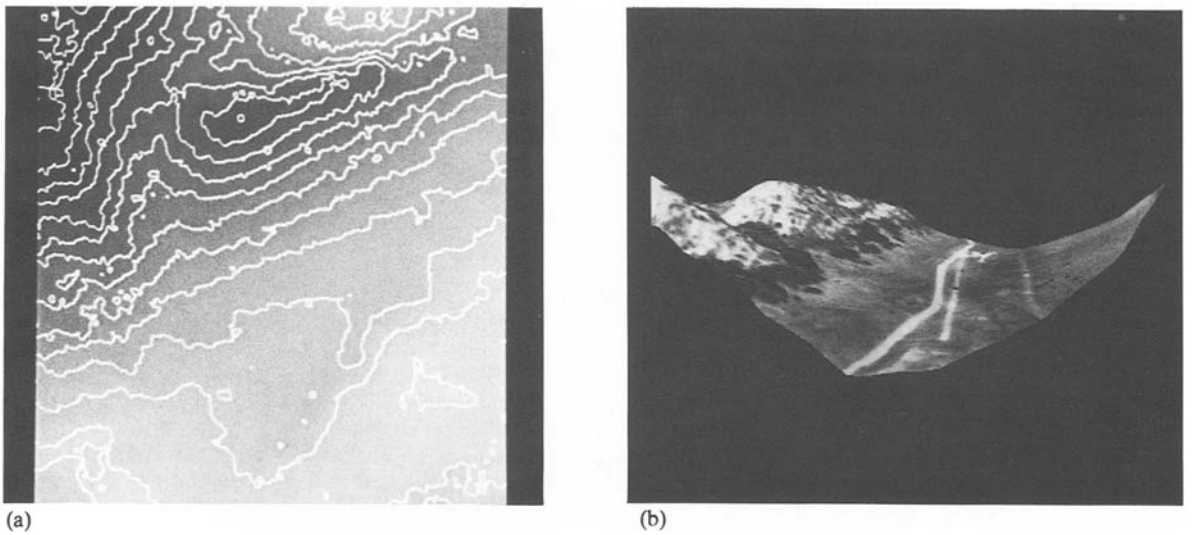


Fig. 10. Results for figure 9. (a) Disparity showing every 5th contour. (b) Synthetic view of terrain using (a).

The use of a scale hierarchy dramatically increases the efficiency of the method, especially for large problems such as that illustrated in figures 9 and 10. An additional benefit of using a scale hierarchy is that the solution is less sensitive to small amounts of vertical disparity, which is eliminated at coarser scales. (Uncertainty in the camera model will usually cause some vertical disparity in high-resolution images.) A Gaussian low-pass hierarchy works as well as the Laplacian hierarchy if the images are recorded with equivalent sensors. The benefit of bandpass filtering is to eliminate the low-frequency variation caused by uncalibrated photometry.

Annealing provides a way to bridge the gap between scales. The microcanonical annealing algorithm appears to be an improvement over canonical annealing for reasons discussed in section 3.3. It is certainly much easier to implement in cellular automata. Theoretical results showing convergence in finite time do not necessarily carry over to microcanonical annealing, but the requirements of these results are never met in practice anyway.

Canonical annealing and pure single-demon microcanonical annealing are at opposite ends of a spectrum. In canonical annealing the heat bath is much larger than the model system, and is not represented explicitly. In pure microcanonical annealing the heat bath—that is, the single demon—is much smaller than the system, and it is represented explicitly. The lattice-of-demons algorithm is midway between these extremes, with the heat bath and the model system having comparable sizes. In a sense, this is a classical space/time trade-off. By representing the heat bath explicitly we can avoid the evaluation of complicated functions.

Comparison with the scale-space continuation method [6] is difficult because the nature of the data affects the smoothness of the energy landscape. In some stereo pairs the data will be so clear that this more direct form of optimization will work well. (For example, dense, random, greyscale stereograms with intensities chosen over a broad range of values can be solved even by a "greedy" algorithm; that is, a Monte Carlo optimization accepting only transitions that lower energy.) Both methods use multiscale representations, but the stochastic approach uses multiple scales only for efficiency, while the gradient descent method uses them in an attempt to avoid the local minima that would be encountered when working only at the finest scale. A comparative study is needed to determine when the additional overhead of annealing is justified.

References

1. D. Terzopoulos, "Image analysis using multigrid relaxation methods," *IEEE TRANS. PAMI,* 8:129-139, 1986.
2. G. Sperling, "Binocular vision: A physical and neural theory," *J. Am. Psychol.* 83:461-534, 1970.
3. B. Julesz, *Foundations of Cyclopean Perception,* Univ. of Chicago Press: Chicago, IL, 1971.
4. D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science* 194:283-287, 1976.
5. T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature* 317:314-319, 1985.
6. A. Witkin, D. Terzopoulos, and M. Kass, "Signal matching through scale space," *Intern. J. Computer Vision* 1:133-144, 1987.
7. B.K.P. Horn, *Robot Vision,* M.I.T. Press: Cambridge, MA, 1986.
8. G.H. Walker and J. Ford, "Amplitude instability and ergodic behavior for conservative nonlinear oscillator systems," *Phys. Rev.* 188: 416-432, 1969.
9. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *J. Chem. Phys.* 21: 1087-1092, 1953.
10. P.J.M. van Laarhoven and E.H.L. Aarts, *Simulated Annealing: Theory and Applications,* D. Reidel Publishing: Dordrecht, Holland, 1987.
11. S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi, "Optimization by simulated annealing," *Science* 220:671-680, 1983.
12. V. Černy, "Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm," *J. Opt. Theory Appl.* 45:41-51, 1985.
13. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images," *IEEE Trans. PAMI* 6:721-741, 1984.
14. S. Barnard, "A stochastic approach to stereo vision," in *Proc. Nat. Conf. Artif. Intell.,* Philadelphia, pp. 676-680, 1986.
15. J.L. Marroquin, "Probabilistic solution of inverse problems," Tech. Rept. 860, M.I.T. Artificial Intelligence Lab., Cambridge, MA, 1985.
16. R. Divko and K. Schulten, "Stochastic spin models for pattern recognition," Physik-Department, Technische Universität München, personal correspondence, 1987.
17. P. Burt, "The Laplacian pyramid as a compact image code," *IEEE Trans. Communicaiions* 31:532-540, 1983.
18. M. Creutz, "Microcanonical Monte Carlo simulation," *Physical Rev. Lett* 50:1411-1414, 1983.
19. G. Bhanot, M. Creutz, and H. Neuberger, "Microcanonical simulation of Ising systems," *Nuclear Physics* B235[FS11], pp. 417-434, 1984.
20. S. Barnard, "Stereo matching by hierarchical, microcanonical annealing," *Proc. 10th Intern. Joint Conf. Artif Intell.,* Milan, Italy, pp. 832-835, 1987.