A STOCHASTIC APPROACH TO
STEREO VISION

Stephen T. Barnard

Artificial Intelligence Center
SRI International
333 Ravenswood Avenue
Menlo Park, California 94025

Abstract

A stochastic optimization approach to stereo matching is presented. Unlike conventional correlation matching and feature matching, the approach 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 be smooth. These constraints are expressed in an "energy" function that can be evaluated locally. A simulated annealing algorithm is used to find a disparity map that has very low energy (i.e., in which both constraints have simultaneously been approximately satisfied). Annealing allows the large-scale structure of the disparity map to emerge at higher temperatures, and avoids the problem of converging too quickly on a local minimum. Results are shown for a sparse random-dot stereogram, a vertical aerial stereogram (shown in comparison to ground truth), and an oblique ground-level scene with occlusion boundaries.

1 Introduction

To solve the stereo matching problem, one must assign correspondences between points on two lattices (the left and right images), such that corresponding points are the projections of the same point in the scene. The problem can be viewed as a complex optimization in which two criteria must be satisfied simultaneously. First, the corresponding points should have similar local features (in particular, similar intensity). Secondly, the spatial distribution of disparities, or, equivalently, the spatial distribution of depth estimates, should be plausible with respect to the depths likely to be observed in real scenes. Several authors have noted that, because surfaces are spatially coherent, the result of the stereo process should also be coherent, except at the relatively rare occlusion boundaries (for example, see Julesz [1] and Marr and Poggio [2]). The first criterion — similarity of local features — is insufficient because stereo correspondences are locally ambiguous. The second criterion, which is sometimes called the *smoothness constraint*, provides a heuristic for deciding which of the many combinations of feature-preserving correspondences are best.

The two major conventional approaches to stereo matching — feature matching and area correlation — suffer from two serious problems:

- Areas of nearly homogeneous image intensity are difficult to match because they lack local spatial structure. Edge-matching approaches never even attempt to match in such areas because no edges are found, and area correlation approaches fail because no significant peaks appear in the correlation surface. For most stereo vision applications, however, a dense matching is required. Dense estimates of depth are also more consistent with the subjective quality of human stereo experience, as revealed, for example, in random-dot stereograms. To obtain dense depth maps with the conventional approaches, one must resort to a postmatching interpolation step.

- Even where local structure is abundant, stereo correspondences may be ambiguous. Small-scale periodic structures are particularly difficult to match. To resolve these ambiguities, stereo matchers usually rely on a propagation of information, either from nearby areas, or from matching at larger scales, or both.

This paper describes an approach to stereo matching that is quite different from conventional area-based and feature-based matching. It is essentially an undirected Monte Carlo search that simulates the physical process of annealing, in which a physical system composed of a large number of coupled elements is reduced to its lowest energy configuration (or *ground state*) by slowly reducing the temperature while maintaining the system in thermal equilibrium. The system is composed of the lattice sites of the left image, and the state of each site encodes a disparity assignment. The total energy of the system is the sum of the energies of the local lattice sites. The local energy, which is a function of the states of the lattice site and its neighbors, has two terms: one term is proportional to the absolute intensity difference between the matching points, and the other term is proportional to the local variation of disparity (that is, to the *lack* of smoothness). The effect of a heat bath is simulated by considering local random state changes and accepting or rejecting them depending on the change in energy and the current temperature.

2 Simulated Annealing

Simulated annealing is a stochastic optimization technique that was inspired by concepts from statistical mechanics [3], [4]. It has been applied to a wide variety of complex problems that involve many degrees of freedom and do not have convex solution spaces. See Carnevali [5] for examples of image-processing applications. At the heart of simulated annealing is the Metropolis algorithm [6], which samples states of a system

in thermal equilibrium. When a system is in thermal equilibrium, its states have a Boltzman distribution:

$$P(E) = \exp(-E/T) \qquad (1)$$

where E is energy, $P(E)$ is the probability of a state having energy E, and T is the temperature of the system.[1] The Metropolis algorithm takes the system to equilibrium by considering random, local state transitions on the basis of the change in energy that they imply: if the change is negative, the transition is accepted; whereas, if the change is positive, the transition is accepted with probability $\exp(-\Delta E/T)$.

Starting at a very high temperature, simulated annealing uses the Metropolis algorithm to bring the system to equilibrium. Then the temperature is lowered slightly and the procedure is repeated until a very low temperature is achieved. If the temperature is lowered too quickly, the system may get stuck in locally optimal configurations and the ground state may not be reached. The algorithm is shown in Figure 1.

Select a random state S.
Select a sufficiently high starting temperature T.
while $T > 0$ **do**
    Make a random state change $S' \leftarrow R(S)$.
    $\Delta E \leftarrow E(S') - E(S)$
    ; *Accept lower energy states.*
    if $\Delta E \leq 0$ **then** $S \leftarrow S'$
    ; *Accept higher-energy states with probability P.*
    else
        $P \leftarrow \exp(-\Delta E/T)$
        $x \leftarrow$ random number in $[0, 1]$
        if $x < P$ **then** $S \leftarrow S'$
    if there has been no significant decrease in E
        for many iterations
    then lower the temperature T.

Figure 1: The Simulated Annealing Algorithm

Simulated annealing tends to exhibit good average-case performance. It has the advantage of being a very simple algorithm that is inherently massively parallel. Furthermore, the parallelism is easily implemented because the processors need only short interconnections, may run asynchronously, and can even be unreliable. To be a good candidate for simulated annealing, a problem should follow the analogy of physical annealing. The function to be optimized should be expressed as an analog to the energy of a system composed of many local elements, and the interaction between the local elements should be short-range. A small random change in the state of the system should be possible by switching the microstate of a local element, and the resulting change in energy should be quickly computed by evaluating only the effects of the element's neighbors.

3 Stochastic Stereo Matching

If the relative positions and orientations of the two cameras are known, as well as the internal camera parameters, we can use the *epipolar constraint* to restrict the correspondences to the epipolar lines [7]. With no loss of generality, we can assume that the epipolar lines are parallel to the horizontal lines of lattice sites.[2] The correspondence problem then reduces to the assignment of a single horizontal disparity to each pixel in, say, the left image lattice.

Suppose that we have left and right image lattices, L_k and R_k, with $k = \{i, j\}, 0 \leq i, j \leq n - 1$, that constitute a stereo pair with horizontal epipolar lines. The intensity of the left lattice point L_k is $I_L(k) = I_L(i, j)$, and similarly for right lattice points. For every k there is a (horizontal) disparity $D(k)$ such that the lattice point $L_k = L_{i,j}$ in the left image matches the point $R_{k'} = R_{i,j+D(k)}$ in the right image. The problem is to find an assignment of disparities to lattice points that satisfies the two criteria discussed in Section 1: similar intensity and smoothness. We assume that the upper and lower limits of disparity, D_{min} and D_{max}, are known. Furthermore, we consider only integer values of disparity. Even with these restrictions, the system has $N = (D_{max} - D_{min} + 1)^{n^2}$ possible states. Typical values in our examples are $D_{max} = 9$, $D_{min} = 0$, and $n = 128$, in which case $N = 10^{16384}$. Exhaustive search is obviously out of the question.

The disparity map should satisfy two criteria that are, to some extent, incompatible. The first criterion, which we call the *photometric constraint*, dictates that the disparity assignments should map points in L to points in R with comparable intensity: $I_L(k) \approx I_R(k')$. The second criterion is the *smoothness constraint*, which limits the variation in the disparity map.

Both criteria cannot be perfectly satisfied except in trivial situations. The photometric constraint can only be approximately satisfied due to sensor noise, quantization, slight lighting differences, and the presence of areas in one image that are occluded in the other. As discussed above, areas of homogeneous intensity will lead to ambiguous disparities based on photometry alone. The smoothness constraint will be perfectly satisified only with a uniform disparity map.

In an attempt to satisfy the two criteria simultaneously, we minimize a function of the form:

$$E = \sum_k \|I_L(k) - I_R(i, j + D(k))\| + \lambda \|\nabla D(k)\| \ . \qquad (2)$$

The first term inside the sum represents the photometric constraint and the second term the smoothness constraint. The constant λ determines their relative importance. We implement the $\|\nabla D(k)\|$ operator as the sum of the absolute differences between disparity $D(k)$ and the disparities of the kth lattice point's eight neighbors. Equation (2) is similar to the nonquadratic Tikhonov stabilizer proposed for stereo by Poggio, et. al. [8].

Following the simulated annealing algorithm, the system begins in a state chosen at random. Individual lattice points are considered in scan-line order, new disparities are selected at random, and the changes in energy are computed from equa-

---

[1]The Boltzman distribution is usually written as $\exp(-kE/T)$, where k is Boltzman's constant. Because we define energy and temperature as pure numbers, no constant is necessary.

[2]If the epipolar lines are not horizontal, the images can be mapped into a rectified stereo pair.

tion (2). Instead of monitoring the energy distribution to test for thermal equilibrium, we use a fixed annealing schedule.

4 Results

We have tested the stochastic matching algorithm on a variety of images, including random-dot stereograms, vertical aerial stereograms, and oblique ground-level stereograms. Identical parameters were used for all the examples shown in this section. In particular, the intensity ranged between 0 and 255, and we used $\lambda = 5$. We used a fixed annealing schedule: the temperature begins at $T = 100$ and is repeatedly reduced by 10% until it falls below $T = 1$. A total of ten scans through the lattice are performed for each temperature in this sequence.

Figure 2 shows a four-level "wedding cake" random-dot stereogram composed of 10% white and 90% black pixels. The background has zero disparity, and each successive layer has an additional two pixels of disparity. The figure shows the results with disparities encoded as grey values. Pixels with higher disparity are "closer" and are displayed as brighter values. Intermediate results for $T = 47$ and $T = 25$ and the final result for $T = 0$ are shown.

Figures 2c-e illustrate an important advantage in stochastic matching: the large-scale structure of the scene begins to emerge at higher temperatures, and as the temperature decreases finer structures become apparent. Temperature therefore provides a mechanism for dealing with problems of scale that is simpler than the complex search strategies employed by conventional methods. Note that the final disparity map is dense and that it corresponds very well to the three-dimensional wedding-cake shape. The errors are confined to the occlusion boundaries.

The next example, shown in Figure 3, is a vertical aerial stereogram supplied by the Engineering Topographic Laboratory (ETL). The original images have been bandpassed to remove the DC component. Again, intermediate results for $T = 47$ and $T = 25$ and the final result for $T = 0$ are shown. In addition, a disparity map supplied by ETL is shown for comparison. [3] The stochastic matching algorithm produces a result that is quite similar to the ETL data, although it is somewhat smoother. To some extent, this difference can be explained by the fact that the ETL result was produced from higher-resolution stereo images. The errors on the right border of Figure 3e are due to the fact that the stereo images do not have 100% overlap.

The final example, shown in Figure 4, is an oblique view of an outdoor scene containing a number of trees in both the foreground and background. The result in Figure 4e is certainly plausible, although we do not have a quantitative disparity model to compare it with, as in the previous examples. The matching algorithm seems to have smoothed over the foreground trees more than necessary, although we must be careful when relying on our subjective impressions of depth. When we interpret a scene like this one, we do not use stereo exclusively.

---

[3] The ETL disparity map was made with an interactive digital correlation device that depends on a human operator to detect and correct errors. The disparity map in Figure 3f has been sampled from a larger map compiled from much higher-resolution imagery.
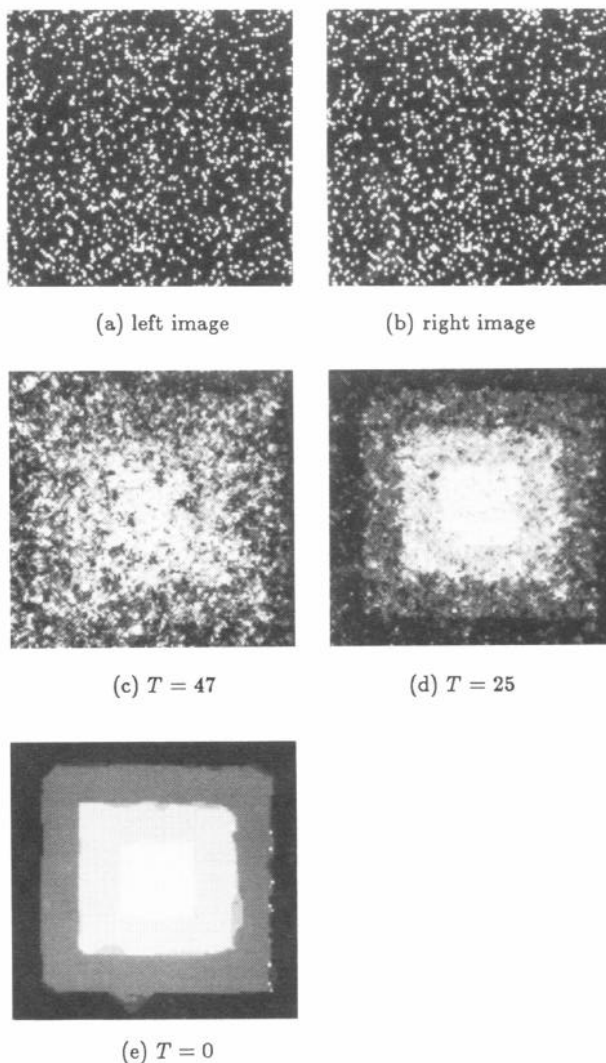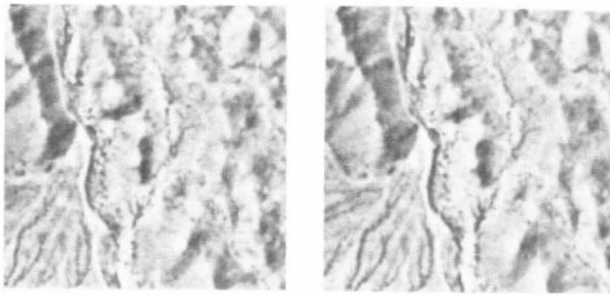

(a) left image   (b) right image

(c) $T = 47$ (d) $T = 25$

(e) $T = 0$

Figure 2: A 10% Random Dot Stereogram.

5 Conclusions

Stochastic stereo matching provides an attractive alternative to conventional stereo-matching techniques in several respects. The algorithm is simple, and, with suitable parallel hardware, can be very fast. Unlike conventional approaches, it produces a dense disparity map.

As noted by Geman and Geman [9], stochastic optimization by simulated annealing is in some ways similar to relaxation labeling [10]. In both approaches, objects are classified in such a way as to be consistent with a global context and to satisfy local constraints. There are, however, important differences. Relaxation labeling is a nonstochastic approach that, unlike simulated annealing, finds the local optimum closest to the initial state. Simulated annealing is intended to find the global optimum, or at least a local optimum nearly as good as the global one. Relaxation labeling has no counterparts for two important concepts in simulated annealing: temperature and thermal equilibrium.
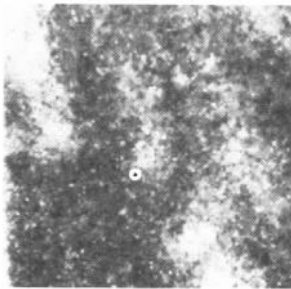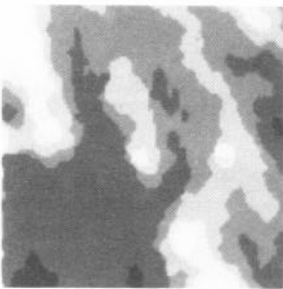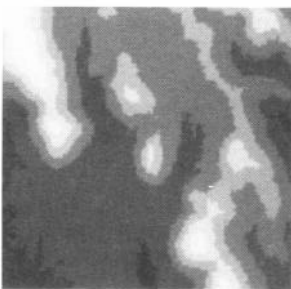
(a) left image

(b) right image
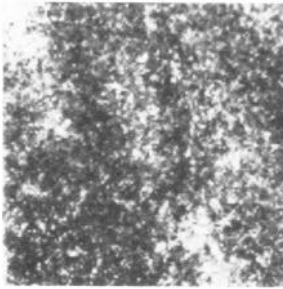
(c) $T = 47$

(d) $T = 25$

(e) $T = 0$

(f) ETL disparity map

Figure 3: A Vertical Aerial Stereogram.

(a) left image

(b) right image

(c) $T = 47$

(d) $T = 25$

(e) $T = 0$

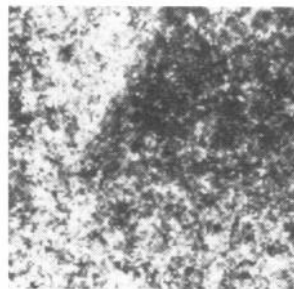Figure 4: An Oblique Stereogram.

The concept of temperature in simulated annealing provides a way to handle different scales in the problem instance. At higher temperatures, objects are only weakly coupled, and long-range interactions among large collections of objects can dominate the behavior of the system. At lower temperatures, local interactions take over. This effect was clearly seen in the examples of Section 4. Some physical systems exhibit a phase transition at some critical temperature. When simulating such systems, one must be careful to lower the temperature very slowly in the vicinity of the critical temperature. We have not observed phase transitions in the stereo problem and have been able to use fixed annealing schedules.

We are considering two extensions of the simple model presented here. First, the effective range of disparity could be increased by using lattices of several scales, allowing the coarser ones to bias the finer, in a manner similar to the hierarchical control structures used in many other matching techniques. Second, following Geman and Geman [9], a "line process" could be used to model depth discontinuities; although, in addition to lines, the process would also model occluded areas.
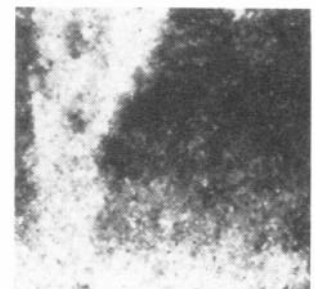
References
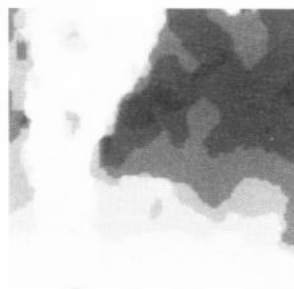
[1] B. Julesz, *Foundations of Cyclopean Perception*, Univ. of Chicago Press, Chicago, Ill., 1971.

[2] D. Marr and T. Poggio, Cooperative computation of stereo disparity," *Science*, **194**, (1976), pp. 283-287.

[3] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi, Optimization by simulated annealing, *Science*, vol. 220, no. 4598, May 13, 1983, pp. 671-680.

[4] S. Kirkpatrick and R.H. Swendsen, Statistical mechanics and disordered systems, *Comm. ACM*, vol. 28, no. 4, April 1985, pp. 363-373.

[5] P. Carnevali, L. Coletti, and S. Patarnello, Image processing by simulated annealing, *IBM. J. Res. Develop.*, vol. 29, no. 6, November 1985, pp. 569-579.

[6] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, Equations of state calculations by fast computing machines, *J. Chem. Phys.*, vol. 21, no. 6, June 1953, pp. 1087-1092.

[7] S.T. Barnard and M.A. Fischler, Computational stereo, *Computing Surveys*, vol. 14, no. 4, December 1982, pp. 553-572.

[8] T. Poggio, V. Torre, and C. Koch, Computational vision and regularization theory, *Nature*, vol. 317, September 1985, pp. 314-319.

[9] S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images, *IEEE Transactions Pattern on Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, November 1984, pp. 721-741.

[10] R.A. Hummel and S.W. Zucker, On the foundations of relaxation labeling processes, *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. PAMI- 5, May 1983, pp. 267-287.