MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

Working Paper 253 July, 1983

DESIGN OF COOPERATIVE NETWORKS

J. L. Marroquin

ABSTRACT: In this paper we analyse several approaches to the design of Cooperative Algorithms for solving a general problem: That of computing the values of some property over a spatial domain, when these values are constrained (but not uniquely determined) by some observations, and by some a priori knowledge about the nature of the solution (smoothness, for example).

Specifically, we discuss the use of: Variational techniques; stochastic approximation methods for global optimization, and linear threshold networks. Finally, we present a new approach, based on the interconnection of Winner–take–all networks, for which it is possible to establish precise convergence results, including bounds on the rate of convergence.

1. Introduction.

The purpose of this paper is to discuss several approaches to the design of algorithms for the solution of a general problem, which often arises in the context of modeling perceptual processes: That of computing the values of some property over a spatial domain, given observations that place some constraints on these values, but do not define them uniquely.

Well known examples of this problem are, for instance, the computation of stereo disparity [MP1], [MP2] and of visual flow [H1]; the computation of structure from motion [U1] and the problem of region discrimination [D1].

We will be interested specifically in the design of "Cooperative Algorithms" [MP1] —Networks of locally connected simple processors which cooperate to perform a global computation.

First of all, let us introduce some notation, and make this definitions more precise.

2. The Problem.

Consider a finite set Ω (which in general corresponds to the discretization of some region of $\mathcal{R}^2$), and some property C, which can take values from a finite set $Q = \{C_1, \ldots, C_N\}$.

Suppose we have some observations which define a "compatibility function" $\phi: Q \times \Omega \mapsto \mathcal{R}$ which has the following property:

> *$\phi(C_1, x) > \phi(C_2, x)$ iff assigning C_1 to x is more compatible with the observations than assigning C_2.*

We suppose that for some $x \in \Omega$, ϕ is a multimodal function of C, and so, it is not possible to assign a unique value of C to every x, based only on the observations.

Our problem (to which we will refer to as "Problem P1") is to reduce this ambiguity as much as possible using some measure of the compatibility of the assignment $C(x)$ with $C(y)$, for $y \in N_x$ (some neighbourhood of x), which we will denote by $g(x, C)$ (we will refer to g as the "smoothness constraint").

To perform this task, we will construct a network by associating with each $x \in \Omega$, either an element C_x which can be in one of N states: $C_x \in Q$ (in the case of an N-ary network), or a set of N elements $\{C_{x,d}: d \in Q\}$, with $C_{x,d} \in \{0, 1\}$ (binary network).

In any case, a Cooperative Algorithm is a rule for updating the state of the network C. It can be represented formally as:

$$C_{x,d}(t+1) = f_{x,d}(C(t), t) \quad (or \quad C_x(t+1) = f_x(C(t), t),$$

with the additional requirement that each element is locally connected, that is:

$$f_{x,d}(C(t),t) = f_{x,d}(\{C_{y,s}(t): y \in N_x, s \in N_d\}, t)$$

The algorithm will be adequate for solving problem P1, if given some initial state, it converges reasonably fast to a fixed point C^* which satisfies "as much as possible" the constraints:

a) Uniqueness.

b) Compatibility with the observations.

c) Smoothness.

Remarks:

(i) In general, it will not be possible to satisfy completely and simultaneously all these constraints. Therefore, any algorithm will have to compromise. If we are using this algorithm to model a perceptual process, its behaviour will be satisfactory if this compromise is the same as the one made by the biological system.

(ii) In the case of a binary network, it is possible to have ambiguous configurations at any time t (i.e., states for which $C_{x,d}(t) = 1$ for more than one $d \in Q$) and therefore, it is necessary to include the uniqueness constraint explicitly in the design. For an N-ary network, of course, this constraint is implicitly incorporated and always satisfied.

(iii) The operation of the network will be Synchronous if all its elements are updated in parallel at the same time, and Asynchronous if they are updated sequentially, one at a time. Note that one synchronous iteration is equivalent to $|C|$ (the number of elements of C) asynchronous ones (we will refer to $|C|$ succesive iterations as a Global Iteration), and that the evolution of the asynchronous network will depend, in general, on the order in which its elements are updated.

In the next sections, we discuss several approaches to the design of cooperative networks whose steady states correspond to solutions of problem P1. Specifically, we will explore the following:

a) Variational approaches.

b) Linear Threshold Networks.

c) Winner-take-all Networks.

To make this discussion more specific, we will use throughout as a particular instance of problem P1, a simplified version of the Stereo Disparity Problem (See, for example, [MP1] for a detailed description):

Given two "retinal" binary images $R(x)$, $L(x)$, $x \in \Omega$, which correspond to the right and left eye views of a surface whose distance from the observer

is piecewise constant, the problem is to find, for each x, a disparity $d(x)$ which satisfies the following conditions almost everywhere:

(i) $R(x_1 + d(x_1, x_2), x_2) = L(x_1, x_2)$ (Compatibility with the observations).

(ii) $d(y) = d(x)$ for $y \in N_x$ (Smoothness).

Note that condition (i) is violated in regions close to the vertical boundaries of planes lying at different depths (occluded regions), and condition (ii) is violated at every depth discontinuity.

3. Variational Approaches to the Design of Cooperative Networks.

The formulation of problem P1, suggests the definition of an "Energy" function of the form:

$$E(C) = \sum_{x \in \Omega} [-g(x, C) - B\phi(C_x, x)] \tag{1}$$

for some constant B. Then, the solution to P1, will correspond to the variational problem of finding a function $C_x = C(x){:}\Omega \mapsto Q$ that minimizes (1).

To find a formal solution to this problem, we generate a set of simultaneous equations, either by putting:

$$\frac{\partial E}{\partial C(x)} = 0$$

for all $x \in \Omega$, or by considering the continuous problem that is generated as the high resolution limit of (1):

$$\min_{C(x)} E(C) = \int_{\Omega} [-g(x, C) - B\phi(C(x), x)]dx \tag{2}$$

and then discretizing the resulting differential equations (the answer should be the same in both cases).

To be specific, consider the following example, for the stereo disparity computation:

Let $C(x)$ be the disparity surface we are looking for, and let $g(x, C)$ be a measure of the curvature of the surface, such as:

$$g(x, C) = -\frac{1}{2}|grad\, C(x)|^2 \tag{3}$$

and let ϕ be defined as:

$$\phi(c, x) = -\frac{1}{2}[R(x_1 + c, x_2) - L(x_1, x_2)]^2 \tag{4}$$

Note that even though R and L are in general formally discontinuous functions, one could argue that due to the blurring produced by any optical system, they will become continuous in the high resolution limit. For a fixed x, ϕ will look somewhat like Fig. 1.

The Euler equations for the variational problem (2) are:

$$\nabla^2 C + B\frac{\partial \phi(C(x), x)}{\partial C(x)} = 0 \tag{5}$$

with natural boundary conditions:

$$\frac{\partial C}{\partial n} = 0 \text{ on the boundary } \partial\Omega$$

Using a finite difference approximation for the derivatives, we get, for the interior points, the non–linear equations:

$$\frac{C(x_1+1, x_2) + C(x_1-1, x_2) + C(x_1, x_2+1) + C(x_1, x_2-1)}{4} - C(x)$$

$$+\frac{B}{2}[\phi(C(x)+1, x) - \phi(C(x)-1, x)] = 0 \tag{6}$$

which in principle could be solved with the following relaxation scheme:

$$C_x(t+1) = \frac{1}{4}\Big[C_{x_1+1,x_2}(t) + C_{x_1-1,x_2}(t) + C_{x_1,x_2+1}(t) + C_{x_1,x_2-1}(t)\Big]$$

$$+\frac{B}{2}[\phi(C_x(t)+1, x) - \phi(C_x(t)-1, x)] \tag{7}$$

Note that if $\phi(c, x)$ were not worse than quadratic on c, (5) would generate a sparse system of linear equations, whose unique (under some mild assumptions) solution could then be obtained by a relaxation scheme which would correspond to an N–ary cooperative network [T1]. However, the multimodality of ϕ is crucial to the definition of problem P1, and in this case, it is clear that it is not possible to guarantee the convergence of (7) to the global minimum of (1).

4. Global Optimization and Statistical Thermodynamics.

Since a straightforward application of variational methods to P1 will not work in general, one could try to find the global minimum of (1) directly, using a stochastic approximation scheme.

We will now describe a method, recently published by Kirkpatrick et. al. [K1], which may be useful in this connection.

Consider a many-body system C, where $C_x(t)$ represents the state of the x^{th} element at time t, and suppose that to each configuration C corresponds an energy

function $E(C)$; if the elements C_x change their state randomly, in such a way that every feasible configuration is equally probable, when the system reaches thermal equilibrium at a given temperature T (which is related to the variance of the fluctuations in the state of each element), the probability of finding the system at a state C, will be proportional to $\exp(-E(C)/kT)$, where k is Boltzmann's constant (see,for example [M2]).

On the other hand, Metropolis et. al. [M1] have shown that it is possible to simulate the behaviour of such systems, for arbitrary energy functions, so that if we change the state according with certain rules, we will get each configuration C with a probability given by the Boltzmann distribution.

Since as $T \downarrow 0$ this distribution collapses into the lowest energy state or states, Metropolis Algorithm provides us, in principle, with a method for finding the global minimum of an arbitrary function of many variables.

Metropolis Algorithm, for updating the state of C_x is:

1: Choose $u \in Q - \{C_x(t)\}$ at random. (Q is the set of admissible values for the state).

2: Compute $\Delta E = E(C(t)) - E(C'(t))$, where C' is obtained from C by replacing C_x by u.

3: If $\Delta E < 0$, put $C_x(t+1) = u$. Otherwise, put $C_x(t+1) = u$ with probability $\exp(-\Delta E/kT)$, and $C_x(t+1) = C_x(t)$ with prob. $(1 - \exp(-\Delta E/kT)$.

In Kirkpatrick's scheme, the system is initially "melted", by allowing it to reach thermal equilibrium at high temperature. (This has the effect of providing a randomized initial state). Then, the temperature is lowered very slowly, allowing the system to stay at each temperature long enough as to reach a steady state ("simulated annealing"), until the system "freezes" and no further changes occur.

In the particular case of (1), since for the stereo disparity computation $C_x \in Q$, Q being a finite set, we can apply Kirkpatrick's Algorithm —which can be considered as a stochastic N-ary asynchronous cooperative network— directly. Note that the computation of ΔE is local in this case: For example, for g and ϕ defined by (3) and (4), we get:

$$\Delta E = (C_x - u)\Big[C_{x_1-1,x_2} + C_{x_1,x_2-1} + C_{x_1+1,x_2} + C_{x_1,x_2+1}\Big] +$$

$$+2(u^2 - C_x^2) + B[\phi(C_x, x) - \phi(u, x)] \qquad (8)$$

This solution,however, is not completely satisfactory, because of the following reasons:

a) The convergence of Kirkpatrick's Algorithm for "melted" initial states is very slow.

b) In general, we can only guarantee the convergence of the algorithm to a configuration C that will give, with high probability, a value of $E(C)$ close to the global minimum, and it is not easy to characterize these configurations in terms of the solution to P1 (See appendix A).

5. "Linear Threshold" Networks.

A different approach to the solution of P1 is to specify first the form of the cooperative network, and then try to find a set of parameters that will cause the algorithm to have the desired behaviour. In particular, we will discuss in this section binary algorithms of the form:

$$C_i(t+1) = \sigma(p_i)$$

$$\text{with} \quad p_i = \left[\sum_j C_j(t) w_{ij} + \eta_i - \theta\right], \qquad i,j \in \Omega \times D; \tag{9}$$

$$\sigma(p) = \begin{cases} 1, & \text{if } p \geq 0 \\ 0, & \text{otherwise} \end{cases};$$

$$w_{ij} \text{ satisfying} \quad w_{ij} = w_{ji}, \quad \text{for all} \quad i,j \in \Omega \times Q$$

$$\text{and} \quad C_i \in \{0,1\}, \quad \text{for all} \quad i$$

The parameters w_{ij}, η_i and θ must be chosen in such a way that the constraints to the solution of P1 are implemented locally.

5.1 Synchronous Algorithms.

For the case of the stereo computation, Marr and Poggio [MP1] have shown that it is possible to define a binary network $\{C_{x,d}\}$, where $d \in Q = \{d_1, \ldots d_7\}$ is the disparity associated with point $x \in \Omega$, to solve P1.

The smoothness constraint is implemented by defining:

$$w_{x,d,y,d} = 1, \quad \text{for} \quad y \in N_x$$

where N_x is an excitatory neighbourhood of x. The uniqueness constraint, by:

$$w_{x,d,y,d'} = -\epsilon, \quad \text{for} \quad (y,d') \in M_{x,d}$$

with $M_{x,d}$ an inhibitory neighbourhood corresponding to multiple matches at x (see [MP1] for a precise definition of these neighbourhoods), and

$$w_{x,d,y,d'} = 0 \quad \text{elsewhere.}$$

The compatibility with the observations is enforced by putting

$$\eta_{x,d} = C^0_{x,d} = \begin{cases} 1, & \text{if } R(x_1 + d, x_2) = L(x_1, x_2) \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

Although it has not been possible to this date to find a rigorous proof for the convergence of this algorithm, numerical experiments and a probabilistic analysis [MP2] show that the synchronous network defined above will converge to reasonably good solutions for random dot stereograms portraying piecewise constant surfaces. It is not clear, however, how to extend this formulation to the more interesting cases of slowly varying disparities — although some ideas are suggested in [MP2] —, and different sized elements placed in points that do not correspond to a regular lattice.

5.2 Asynchronous Algorithms.

We now consider algorithms of the form (9) that operate asynchronously. In this case, it has been shown [H2,H3] that if we choose the parameters in such a way that p_i is never 0 (this can be done, for example, if w_{ij} and η_i are integers, by giving θ a non-integer value), the "Energy" function:

$$E(C) = -\frac{1}{2}\sum_{i,j} w_{ij} C_i C_j - \sum_i C_i(\eta_i - \theta) \tag{11}$$

will decrease monotonically at every global iteration of the asynchronous algorithm in which the state of every element is updated, unless the network is at a fixed point.

It is interesting to note that with the parameter definitions given above for the stereo problem, we can identify the smoothness constraint:

$$g(x, C) = \frac{1}{2} C_{x,d} \sum_{y \in N_x} C_{y,d} \quad ;$$

the compatibility with the observations:

$$\phi(C, d, x) = C_{x,d} C^0_{x,d}$$

and the uniqueness constraint:

$$U(x, C) = -C_{x,d}\left[\theta + \frac{\epsilon}{2} \sum_{y,d' \in M_x,d} C_{y,d'}\right]$$

Then, the asynchronous operation of (9) may be considered as an algorithm for finding a local minimum of (11), which is a particular case of the function:

$$E(C) = -\sum_{x,d \in \Omega \times Q} [g(x, C) + \phi(C, d, x) + U(x, C)] \tag{12}$$

It is also possible to use Kirkpatrick's Algorithm, and define an asynchronous stochastic binary network that will converge to a value of (11) close to its global minimum [H2]. However, it is possible to show (see appendix A) that although the correct solution is a local minimum of (11), in general it is not the global minimum, and there are many local minima with values of E close to it that correspond to incorrect solutions.

6. Winner-take-all Networks.

Linear threshold networks are not the only form of local implementation of the constraints generated by P1. A different possibility is to associate with each point $x \in \Omega$ a binary "Winner-take-all" network [F1] with $|Q|$ cells: $\{C_{x,d}: d \in Q\}$.

The input $u(x, d)$ to each cell will be of the form:

$$u(x, d) = g(x, C) + B\phi(d, x) \tag{13}$$

The output (the new value of $C_{x,d}$) is given by:

$$C_{x,d} = \begin{cases} \sigma(\phi(d, x) - \theta), & \text{if } u(x, d) = \max_{d' \in Q} u(x, d') \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

with σ defined as in (9). This means that $C_{x,d}$ will be "on" at time $t + 1$ only if it is maximally stimulated with respect to its neighbours at time t, and if it is "compatible enough" with the observations.

The main advantage of this design is that it is possible to prove its fast convergence to the correct solution of P1. As an example, we will analize the case of the stereo disparity computation. We will need the following definitions:

1. Ω will be defined as a connected set of points lying on a square lattice.
2. For every point $x \in \Omega$, we define N_x as:

$$N_x = \Omega \bigcap \{y: |x - y| < 2\} \tag{15}$$

For the interior points of Ω, N_x is shown in Fig. 2.

3. Given a connected region $R \subseteq \Omega$, we define the set of its interior points $I(R)$ as:

$$I(R) = \{x \in R: |N_x \bigcap R| = |N_x|\} \tag{16}$$

In a similar way we define:

$$I^2(R) = I(I(R))$$

and so on. We call the points $x \in R - I(R)$, Boundary points of R.

4. Given a connected region $R \subseteq \Omega$, we define its Diameter $D(R)$ as the smallest integer such that:

$$I^{D(R)}(R) = \emptyset$$

Now consider the following algorithm for the stereo disparity problem:

$$C_{x,d}(t+1) = \begin{cases} C^0_{x,d}, & \text{if } u_{x,d}(t) = \max_{d' \in Q} u_{x,d'}(t) \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

$$\text{where } u_{x,d}(t) = \alpha C^0_{x,d} + \sum_{y \in N_x} C_{x,d}(t)$$

with $C^0_{x,d} = C_{x,d}(0)$ given by (10), and α a constant > 8.

We have the following:

Theorem 1: Given a random dot stereogram of a domain Ω portraying n non-overlapping constant depth regions $\{R_1 \ldots R_n\}$ with disparities $\{d_1 \ldots d_n\}$ satisfying the condition:

$$if \quad |N_x \bigcap R_j| < 8 \quad and C^0_{x,d'} = 1, \quad then \quad |N_x \bigcap R_j| > |N_x \bigcap H_{d'}|, \tag{18}$$

$$\text{for} \quad d' \neq d_j; \text{for all} \quad x \in R_j, \quad \text{and all} \quad R_j \in \Omega \quad \text{with} \quad H_d = \{x : C^0_{x,d} = 1\},$$

then, algorithm (17) will converge to a fixed point C^* in which:

(i) For every x on a non occluded region at disparity d, $C^*_{x,d} = 1$ and $C^*_{x,d'} = 0$ for $d' \neq d$.

(ii) For x on an occluded region, either $C^*_{x,d} = 0$ for all d, or $C^*_{x,d} = 1$ for one or more isolated d, corresponding to false targets.

Further, the convergence to C^* will take less than K iterations, where K is the diameter of the largest connected cluster of "on" cells on a wrong layer of $C^0_{x,d}$.

Proof:

Consider a non-occluded region R at disparity d. For any interior point $x \in R$, we have:

$$u_{x,d}(0) = \max_{d'} u_{x,d'}(0) \tag{21}$$

and therefore, $C_{x,d}(1) = 1$.

For a point x near the boundary, $|N_x| < 8$, and condition (18) guarantees that (21) holds for this point too.

Now consider a cluster $A \subseteq H_{d'} \bigcap R$, for some $d' \neq d$. For all its boundary points y, we will have

$$u_{y,d'}(0) < u_{y,d}(0),$$

since y is either interior to R, in which case

$$\sum_{x \in N_y} C_{x,d}(0) > \sum_{x \in N_y} C_{x,d'}(0),$$

or it is in the boundary of R, in which case (18) holds. Therefore, we have:

$$A(1) \subseteq I(A) \quad \text{and} \quad R(1) = R$$

$$\text{where} \qquad A(1) = \{x \in A : C_{x,d'}(1) = 1\}$$

$$\text{and} \qquad R(1) = \{x \in R : C_{x,d}(1) = 1\}.$$

The recursive application of this reasoning establishes the first part of the theorem.

Finally, for occluded regions, there will be no dense sets in C^0at any disparity, and since the form of (17) precludes the growth over regions with $C^0 = 0$, if there are any isolated points for which $C^0_{x,d} = 1$, they will remain "on" in C^*, and otherwise, $C^* = 0$ uniformly over these regions. ■

Remarks:

1. Condition (18) means that there should not be dense clusters of "on" cells on the wrong layers of C^0 along the boundaries of any region R_j. If this happens, limit cycles, involving some cells along the problematic segments of these boundaries, may appear. However, if we extend the definition of C^* to include configurations which are invariant, except for these limit cycles, it is clear that Theorem 1 still holds for $\Omega' \subseteq \Omega$, where Ω' is a subdomain in which (18) holds. The final configurations in this case, will have some boundaries misplaced, and there may be some leftover ambiguity ($C^*_{x,d} = 1$ for more than one d) along the boundary points. It is interesting to note that the human visual system exhibits an analogous behaviour in similar situations.

2. Algorithm (17) will not grow regions into occluded (uncorrelated) areas. Psycophysical experiments show that these areas should be included with the adjacent region that is at the greatest depth. It can be verified that an algorithm such as the following:

$$C_{x,d}(t+1) = \begin{cases} 1, & \text{if } \sum_{y \in N_x} C_{y,d}(t) > 2C_{x,d'}(t)\left[\sum_{y \in N_x} C_{y,d'}(t)\right], \quad d' \neq d \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

with $C_{x,d}(0) = C^*_{x,d}$ (the fixed point of (17)), will converge to a solution in which these regions are correctly filled in, provided there are no wrong clusters in the occluded regions, and that each layer of constant d is allowed to converge separately, starting with $d = dmin = \min(d \in Q)$.

3. When defining a stereo computation, one can work either on right (or left) eye–centered coordinates, or on "object–centered" coordinates (see Fig. 3). This distinction becomes particularly important near the boundaries of partially occluded regions, since a single–valued (discontinuous) surface on object–centered coordinates can become multivalued on eye–centered ones. The strictly correct version of algorithm (17) should be considered to be defined with x expressed in object–centered coordinates. In this case, d is not disparity, but depth, and C^0 should be defined by:

$$C^0_{x,d} = \begin{cases} 1, & \text{if } F_R(x,d) = F_L(x,d) \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

where $F_R(x,d)$ and $F_L(x,d)$ are the points on the right and left retinas corresponding to the point (x,d) (x_L and x_R in Fig. 3). However, for a random dot stereogram, the behaviour of the algorithm on non-occluded regions is the same if we work on, say, right eye–centered coordinates, and its description is simpler, so we adopt it here. Theorem 1 holds for both cases.

4. Note that even when $(x_1, x_2) \in \Omega, (x_1 + d, x_2)$ may not be, and so, if we load the network using (10), some cells near the boundaries of Ω may remain undefined, and (17) may give incorrect results. Therefore, we implicitly assume the existence of a larger region $\Omega_0 \supseteq \Omega$ such that for all $x \in \Omega, C^0_{y,d}$ is defined for $y \in N_x \cup \{x\}$ and $d \in Q$. Also, the operation of (17) should be understood in a modified sense, so that $C_{x,d}(t) = C^0_{x,d}$ for all $x \in \Omega_0 - \Omega$, all $d \in Q$, and all t.

One useful corollary, which follows directly from remark (1) is that if we have a stereogram with sparsely located tokens, algorithm (17) will not misplace them, although condition (18) may be violated in the blank areas. In precise terms, we have:

Corollary 1: Suppose that within each region R_j of constant disparity d_j there is a set of sparse points $X_j \subseteq R_j$, and suppose that it is possible to find a region R_j' (not necessarily connected) such that $X_j \subseteq R_j' \subseteq R_j$, and that (18) holds for every R_j'.

Then, for every $x \in X_j, C^*_{x,d_j} = 1$ and $C^*_{x,d} = 0$, for $d \neq d_j$, where C^* must be taken in the sense of remark (1). ■

A second corollary establishes that it is not necessary to process all Ω at the same time, but that a complete representation can be built up by defining local networks corresponding to windows $W \subseteq \Omega$, provided the boundary conditions are handled correctly.

Let $C^t_\Omega(x,d)$ and $C^t_W(x,d)$ be the state of the (x,d) cell at time t in the complete and local network respectively. We have:

Corollary 2: Suppose (18) holds in Ω, and consider the sets $W_1 \subseteq W \subseteq \Omega$. Suppose that W and W_1 are chosen in such a way that for every $x \in W_1, C^0_W(y,d) = C^0_\Omega(y,d)$, for all $y \in N_x \cup \{x\}$, and all $d \in Q$.

Suppose further that either W_1 intersects at least two regions at different disparities or, if $W_1 \subseteq R_j$, for some j, then, for some $x \in W_1$,

$$|N_x \bigcap W_1| > |N_x \bigcap H_d \bigcap W_1|, \quad \text{for all} \quad d \neq d_j$$

(i.e., the stereogram is not completely ambiguous inside W_1).

Then, algorithm (17), modified in such a way that $C^t_W(x,d) = C^0_W(x,d)$ for all t, all $x \in W - W_1$, and all $d \in Q$, will converge to a fixed point C^* for which $C^*_W(x,d) = C^*_\Omega(x,d)$ for all x belonging to unoccluded regions inside W_1.

Proof:

Consider a region R of constant disparity d such that $R' = R \bigcap W_1 \neq \emptyset$, and let $B_1 = R \bigcap B(W_1)$. For every point $x \in R' - B_1$, $C^1_W(x,d) = 1$, by the same arguments as in Theorem 1. For $x \in B_1$, $C^1_W(x,d) = 1$ too, since $C^0_W(y,d) = C^0_\Omega(y,d)$ for $y \in N_x$, and (18) holds in Ω. Therefore, $R'(1) = R'$.

On the other hand, for any cluster $A \subseteq R' \bigcap H_{d'}, d' \neq d$, $A(1) \subseteq A$ and $A(1) \neq A$, since the stereogram is not completely ambiguous inside W_1. Applying this reasoning recursively, we get, for every $x \in R'$, that $C^*_W(x,d) = 1$, and $C^*_W(x,d') = 0, d' \neq d$, which, together with Theorem 1, completes the proof. ■

Note that $W - W_1$ defines the overlap that should exist among local windows, so that the complete representation, defined by

$$\Omega = \bigcup_j W_1^{(j)}$$

is correctly formed.

6.1 Extensions.

Algorithm (17) can be extended to the case where depth is not constant, but varies slowly within each region R_j, as long as the local depth variation within each region is smaller than the local variation across boundaries between regions. Formally, we can express this condition by requiring the existence of a known real number β such that:

$$\max_{\substack{x \in R_j \\ y \in N_x \bigcap R_j}} |d_x - d_y| \leq \beta < \min_{\substack{x \in R_j \\ w \in N_x \bigcap R_i \\ i \neq j}} |d_x - d_w|, \quad \text{for all} \quad j \qquad (22)$$

where d_x is the depth at x.

It is not necessary to have the sites over which the matching is done distributed on a regular lattice, as long as there is enough information to solve the ambiguities at every point. In this case, N_x has the form:

$$N_x = \{y : |x - y| < r\} \tag{23}$$

and we require that:

$$|N_x| > 1 \quad \text{for every} \quad x \in \Omega \quad \text{such that}$$

$$\sum_{d'} u_\phi(x, d') > 1 \tag{24}$$

where u_ϕ is a function that replaces $C^0_{x,d}$ on (17). It may be defined as:

$$u_\phi(x, d) = \begin{cases} 1, & if\, \phi(x, d) < \theta \\ 0, & \text{otherwise} \end{cases} \tag{25}$$

Note that the measure of compatibility with the observations ϕ need not be as simple as (7), but may have a more complex form that incorporates the matching of other attributes, such as color, orientation, etc.

Finally, the extended version of (17) will be:

$$C_{x,d}(t+1) = \begin{cases} u_\phi(x, d), & \text{if} \quad u_{x,d}(t) = \max_{d' \in Q} u_{x,d'}(t) \\ 0, & \text{otherwise} \end{cases} \tag{26}$$

where:

$$u_{x,d}(t) = B u_\phi(x, d) + \sum_{\substack{y \in N_x \\ d' \in N_\beta(d)}} C_{x,d'}(t) \quad ;$$

$$N_\beta(d) = \{d' \in Q : |d - d'| \leq \beta\} \quad ;$$

$$B > \max_{x \in \Omega} |N_x| \quad \text{and} \quad C_{x,d}(0) = 0, \quad \text{for all} \quad x, d.$$

Condition (18) becomes:

$$If \quad |N_x \bigcap R_j| < |N_x|, \quad then \quad |N_x \bigcap R_j| > |N_x \bigcap H_{d'}| \tag{27}$$

$$\text{for} \quad d' \notin N_{2\beta}(d_j), \quad \text{for all} \quad x \in R_j \quad \text{and} \quad R_j \subseteq \Omega,$$

$$\text{where} \quad H_{d'} = \{x : u_\phi(x, d_1) = 1, \text{for some} \quad d_1 \in N_\beta(d)\}$$

With these modifications, Theorem 1 still holds for the modified algorithm (26).

6.2. Numerical Results.

To test the performance of algorithm (17) with random dot stereograms, a simulator was implemented in a Lisp–Machine. Figure 8 shows the fixed points corresponding to dense and sparse stereograms portraying a pyramid. As predicted by the theory, the convergence to the correct solution is fast (less than 4 iterations) in both cases. In the case of the sparse stereogram, the boundaries are slightly misplaced, but, as can be verified by direct inspection of the stereogram, all the dots are correctly located. The fixed point corresponding to the synchronous operation of (9) is also presented, for comparison.

7. Conclusions.

We have discussed in this paper several approaches to the design of cooperative networks for a general class of computational problems that consists on finding the value of some property C for every point on a set, given that the available observations do not define this value unambiguously, using as a constraint the relations that must hold between the value of C for a given point and for its neighbours.

We found that even when this problem leads naturally to the formulation of an associated variational problem (via the definition of an appropriate "Energy" function), the multimodality of the functions involved (the ambiguity of the observations) preclude the direct use of standard variational techniques (which are based on the satisfacion of necessary conditions for a minimum), that would lead to the formulation of relaxation schemes.

On the other hand, it is conjectured that stochastic approximation methods for global optimization —Specifically, Kirkpatrick's algorithm — should be able to find configurations that correspond to values of the energy close to the global minimum (even though the rate of convergence might be slow). However,there is no guarantee that these configurations will correspond to correct solutions of the original problem, unless the energy function is defined in a very precise way (paraphrasing Alan Oppenheim, any configuration is optimal, given an appropriate choice of the energy function), and the problem of defining precisely the required function, may be as difficult as finding the solution to the original problem directly.

A more promising approach seems to be to define the form of the cooperative algorithm based on the local enforcement of the original computational goals. In this context, we discussed the use of Linear Threshold networks, which for specific computations (the stereo disparity problem) give reasonably good solutions and rates of convergence. Their main disadvantage is that it has not been possible to analyse them rigurously in order to characterize their fixed points, bound their rate of convergence, etc.

A recent attempt, based on the definition of a monotonically decreasing energy function for the asynchronous operation of these networks, was found to

be unsatisfactory, mainly because of the difficulties encountered in characterizing the configurations that correspond to minima of this function.

Finally, a new scheme for the implementation of local constraints, using interconnected Winner–take–all networks was introduced. This design seems very attractive for the following reasons:

(i) It is possible to prove rigurously its convergence to the desired solution.

(ii) It is possible to find an upper bound for the number of iterations needed to reach the solution. This number is in general much smaller than the average number required by other schemes.

(iii) It is possible to extend these results to more general situations, such as: non–regular lattices, different types of matching attributes, etc.

Appendix A: Experimental Analysis of the Stereo Disparity Computation Using Linear Threshold Networks.

For the experimental study of these algorithms, a simulator was implemented on a Lisp–Machine of the Artificial Intelligence Laboratory, with the following features:

a) The state of the network is displayed on the screen and updated on line, so that it is possible to observe its evolution continuously.

b) It is possible to simulate both synchronous and asynchronous modes of operation. In this last case, several ordering schemes for updating the state of the elements are possible:

(i) "Gauss-Seidel" ordering. This corresponds to three nested loops, for d, x_1 and x_2: The innermost loop corresponds to x_2, and the outermost one to d.

(ii) Random ordering: Three uniformly distributed pseudo–random numbers corresponding to d, x_1 and x_2 are generated each time.

(iii) Randomly permuted ordering: In this case, the three–dimensional array $C_{x_1,x_2,d}$ is scanned along its diagonals, but before updating an element, its indices are mapped into a new set using random permutations of the sequences $\{0,1,\ldots 64\}$, $\{0,1,\ldots 64\}$ and $\{-3,-2,\ldots 3\}$. In this form, a "disordered" sequence is generated, but it is guaranteed that each element will be updated exactly once every global iteration.

c) For each type of asynchronous iteration it is possible to specify a "temperature" greater than zero, and simulate an "annealing" scheme using Hinton's algorithm [H2].

A1. Results

The "stimulus" used for the set of experiments performed, was a random dot stereogram portraying a square of 21×21 elements floating at disparity -2 in front of a flat background at disparity 0 (Fig. 4–a).

Both the synchronous and asynchronous algorithms (except for the "Gauss–Seidel" ordering, which gives large clusters of "on" cells on wrong layers at its fixed point) converge to configurations that roughly correspond to a central "squarish" area at disparity -2, and a flat background at disparity 0. These configurations may differ considerably along the boundary between regions (Figs. 4–c,d). The asynchronous algorithm tends to produce some small clusters in wrong disparity layers, particularly along the edges of the background (Fig. 5–b). The rate of convergence is in general better for the asynchronous algorithm (particularly for the randomly permuted ordering) than for the synchronous one.

In both cases, the behaviour of the algorithm shows two distinct phases: In the first iteration, most of the elements that are "on" on the wrong layers (and some

on the correct ones) are turned "off". As a result of this, at succeding iterations, the probability of having a cluster capable of growing is relatively high for the correct regions, which begin to fill in, and very small for the wrong ones, for which the remaining "on" cells are turned "off".

This form of operation causes the convergence behaviour described above: The precise shape of the boundaries between regions will depend on the exact shape and location of the random clusters that are formed after the first iteration on the correct layers. Also, it is easy to see that the form of the inhibitory neighbourhood [MP1] causes the cells lying on wrong layers along a narrow band near the edges of the background to be on the average less inhibited by the "on" elements in the correct layers (which in turn are less stimulated) than the interior points, making thus more likely the formation of wrong stable clusters in these regions. This effect is more pronounced in the asynchronous case, since a wrong cell that is left "on", can increase the excitation of a neighbouring one on the same global iteration, increasing the likelihood of a stable cluster, whereas on the synchronous case, all the cells of the cluster must be left "on" ot the same time.

A.2. Behaviour of the "Energy" function.

For the values used for the parameters ($\epsilon = 2, \theta = 3.5$) the energy defined in (11) decreases monotonically at each global iteration of the asynchronous network, and thus, it converges to a configuration that is a local minimum of (11). On the other hand, it can be shown that the correct solution is a local minimum of (11) as well. In fact, all the stable configurations described above correspond to local minima whose value is very close to that of the correct solution.

Furthermore, it can be easily shown that a uniform layer of "on" cells at disparity 0, will give a lower energy value than the correct solution, if the ratio of area/perimeter of the central figure is less than a critical value (for the current values of the parameters, this critical ratio is approximately 13).

These observations mean that the function (11) exhibits, in phase space, a large valley with a lot of local minima in it, which correspond to "squarish" central figures at the correct disparities (and maybe some small incorrect clusters), and with high probability, with the initial conditions given by (15), the algorithm will converge to some minimum in this valley.

A.3. Effect of Temperature.

The simulation of "annealing", in the sense of Kirkpatrick, is difficult to perform on a Lisp–machine, due to the large amount of computation involved. However, the results obtained so far (Fig. 6) and the above considerations, indicate that the algorithm will converge to a local minimum located on the "valley of the squarish figures" (note that the convergence to the global minimum may be undesirable in some cases).

For the particular case of the stereo problem, however, the initial conditions already have a lot of structure which does not make sense to destroy by "melting" the system (Fig. 6–a). Instead, the following modified version may be used:

1: Load the network using (10).

2: Allow the asynchronous algorithm to converge at zero temperature (the fixed point may contain stable clusters in wrong layers).

3: Rise the temperature a little (say, put $T = 1$) and perform a few iterations (this will eliminate the wrong clusters, and make some holes in the correct layers).

4: Allow the network to converge at zero temperature (this gets rid of the holes).

This procedure is illustrated in Fig. 7.

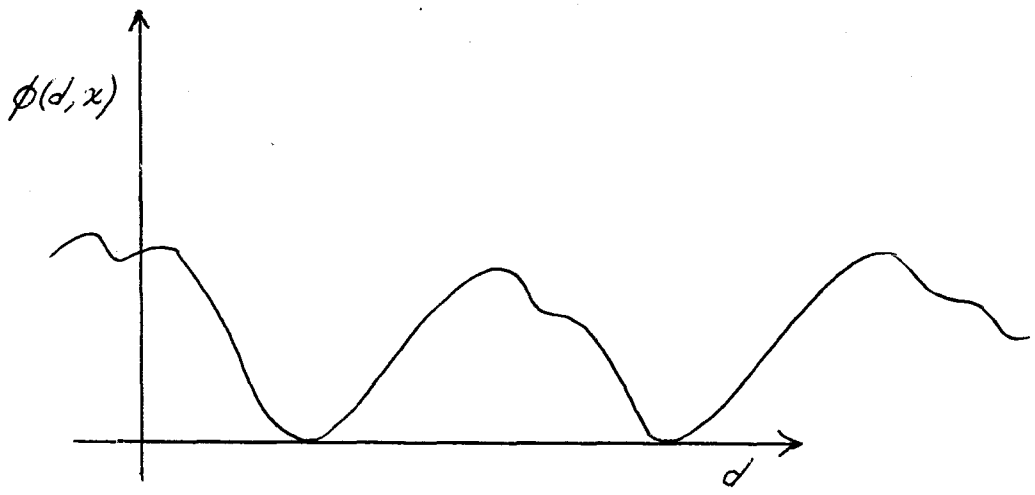

Fig. 1: Function ϕ for eq. (4)

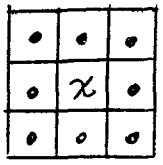

Fig. 2: N_x for eq. (15)

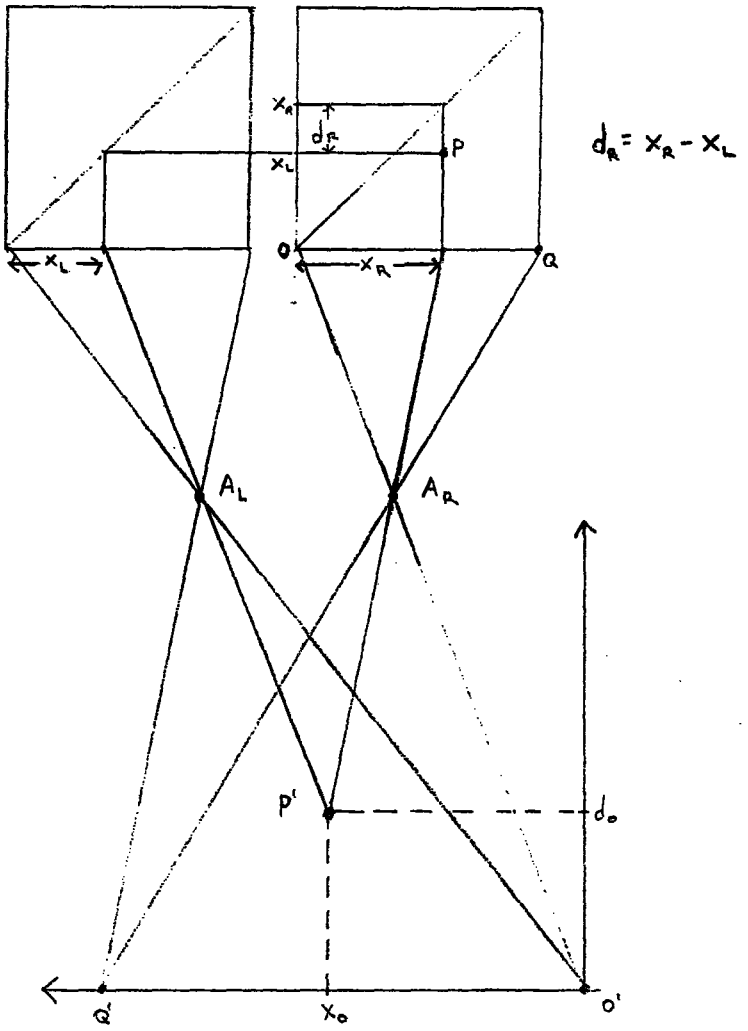

Fig. 3: Relation between Right-eye-centered coordinates (x_R, d_R) and Object-centered ones (X_o, d_o) for point P. The position of the projection points A_R, A_L depends on the position of the right and left eyes.

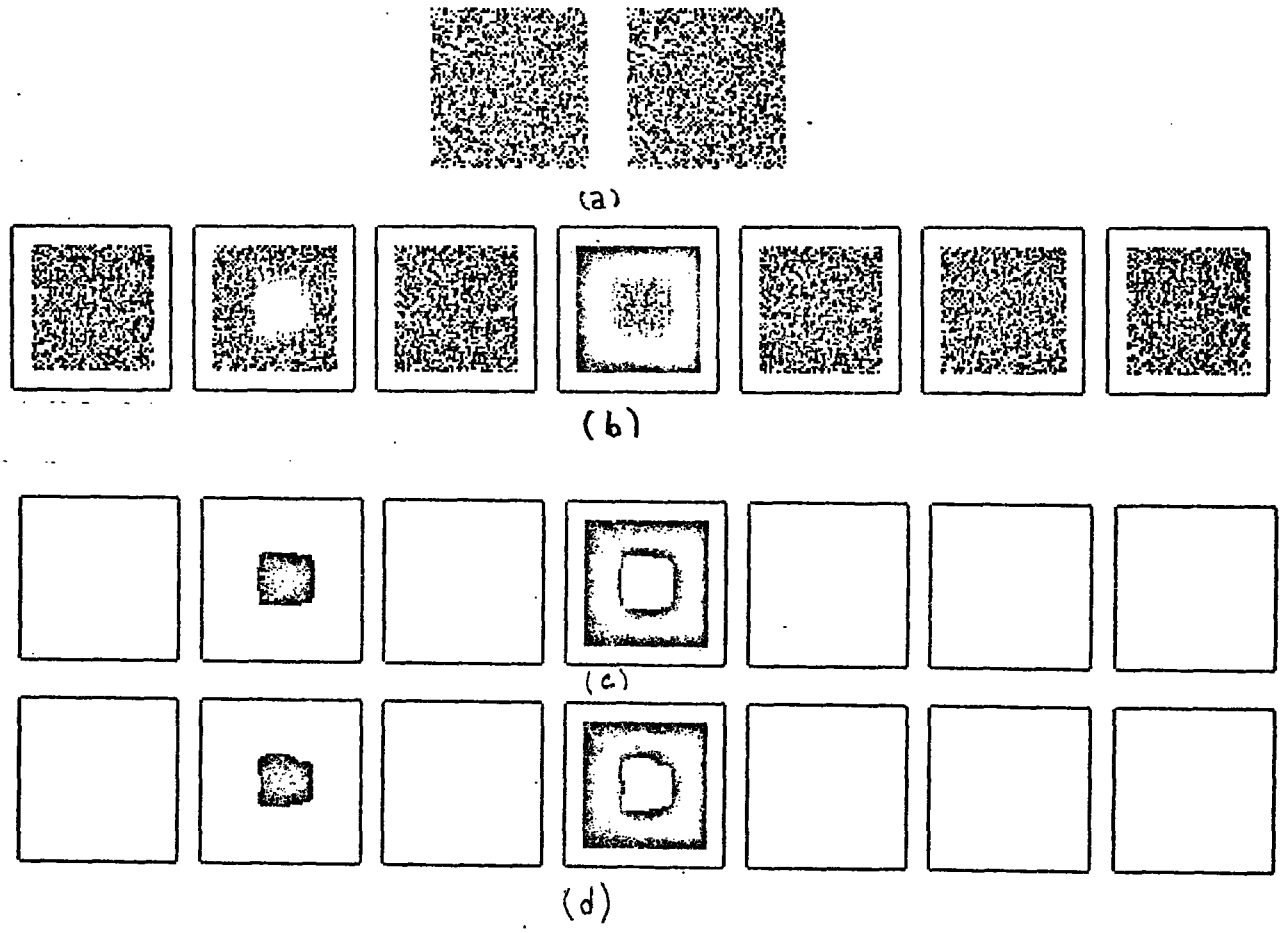

Fig. 4: (a) Random dot stereogram portraying a 21 × 21 square at disparity −2. (b) Initial state of the network for loading rule (10). (c) Fixed point for the Synchronous Algorithm. (d) Fixed point for the Asynchronous Algorithm.

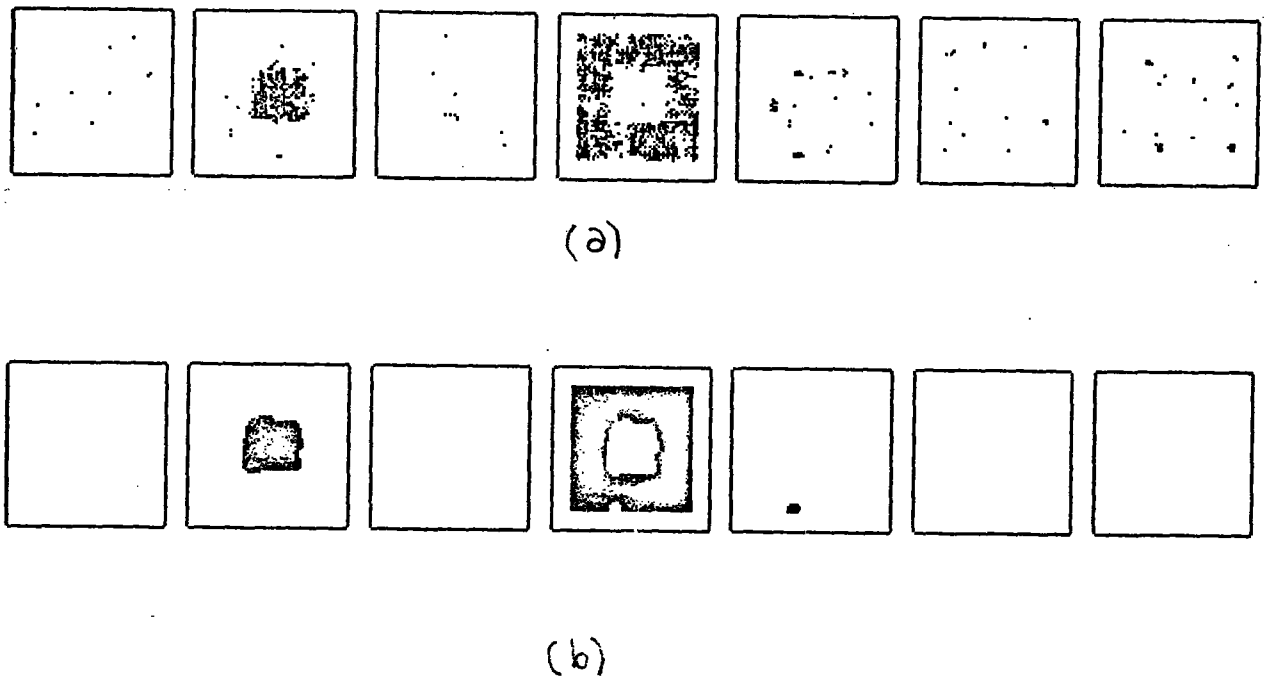

Fig. 5: (a) State of the network after the first iteration of the Asynchronous Algorithm. (b) Fixed point showing a wrong cluster at disparity 1.

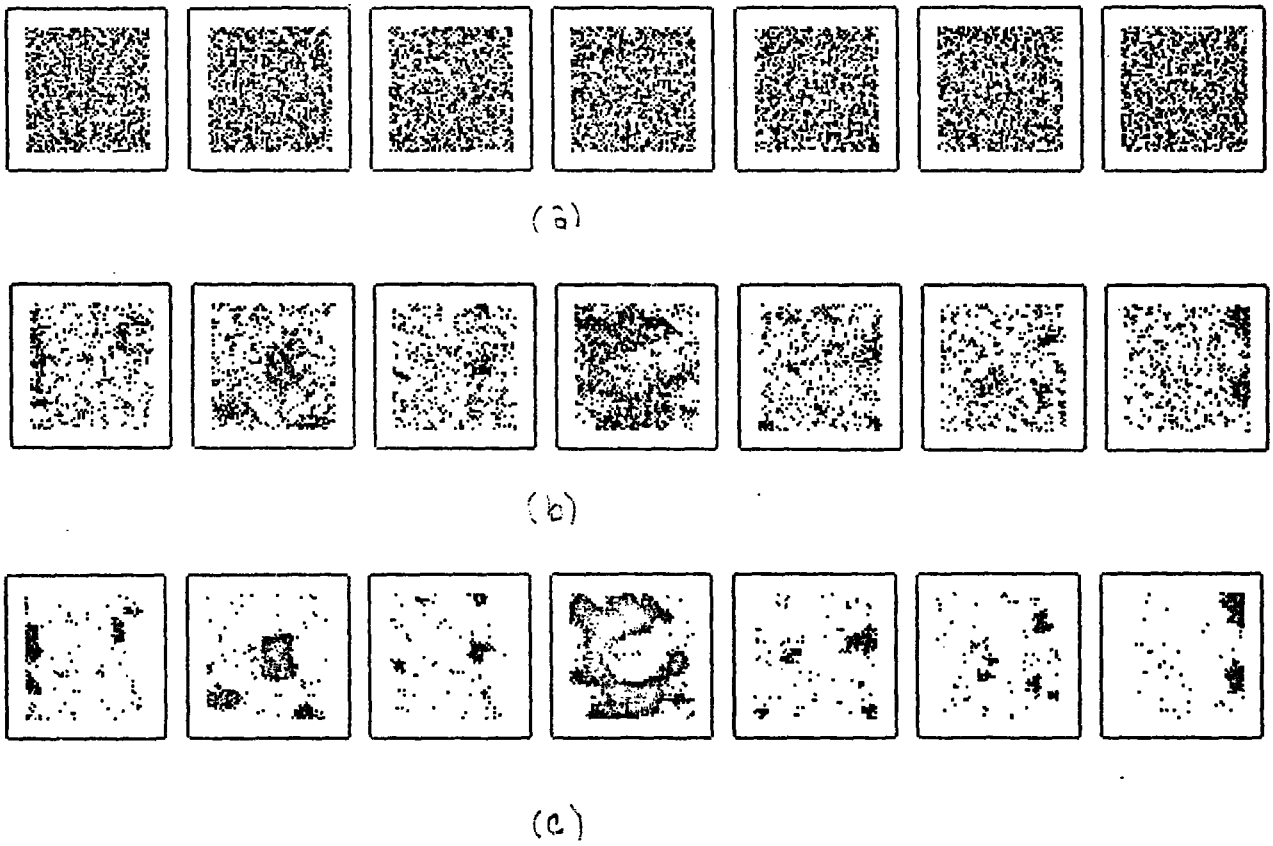

Fig. 6: State of the network after 5 iterations at different temperatures: (a) $T = 50$ (Initial state as in Fig. 4-b). (b) $T = 4$ (Initial state: (a)). (c) $T = 1.5$ (Initial state: (b)).

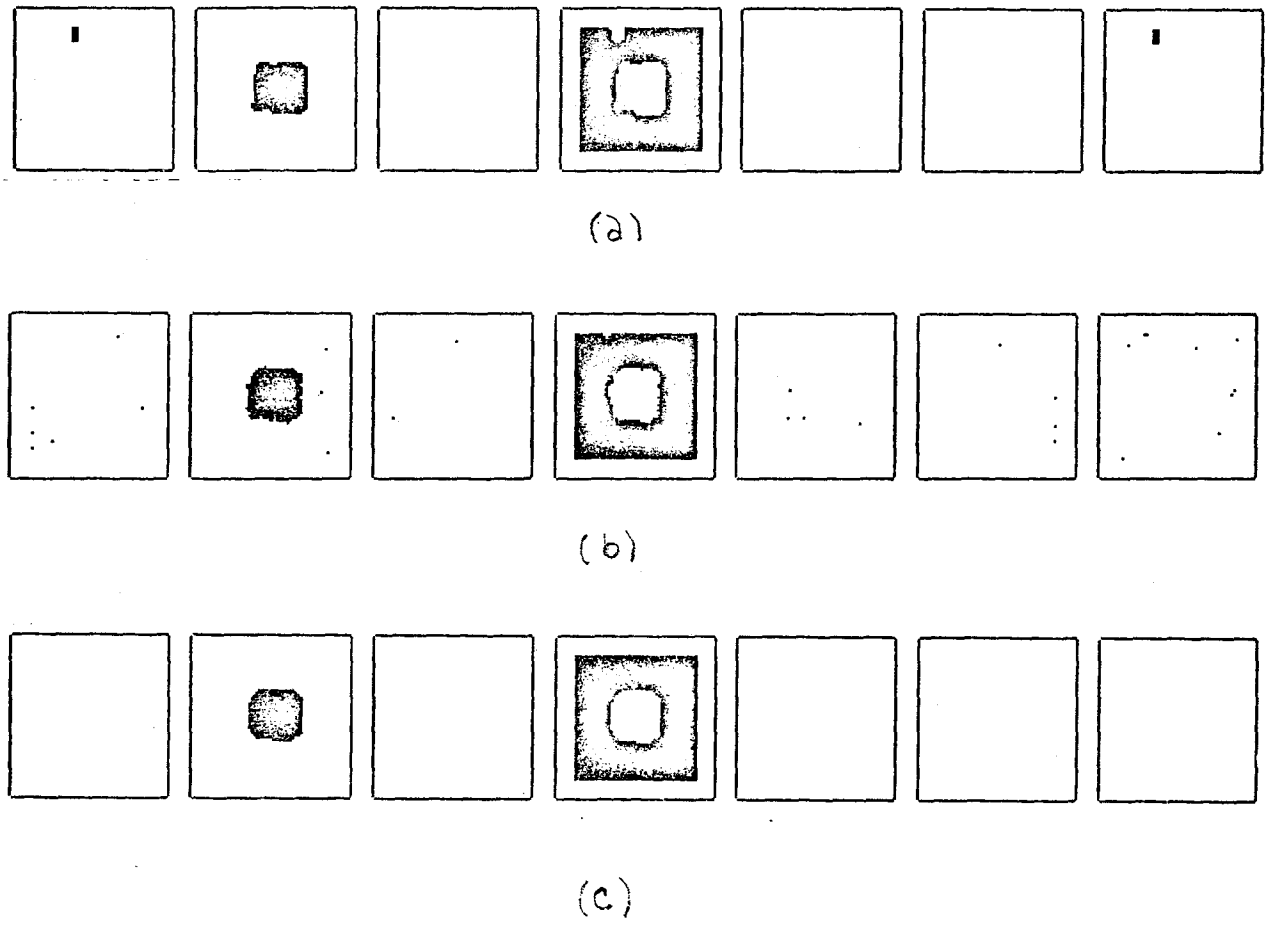

Fig. 7: (a) Fixed point at $T = 0$. (b) State after 4 iterations at $T = 1$. (c) Fixed point at $T = 0$ with (b) as initial state.

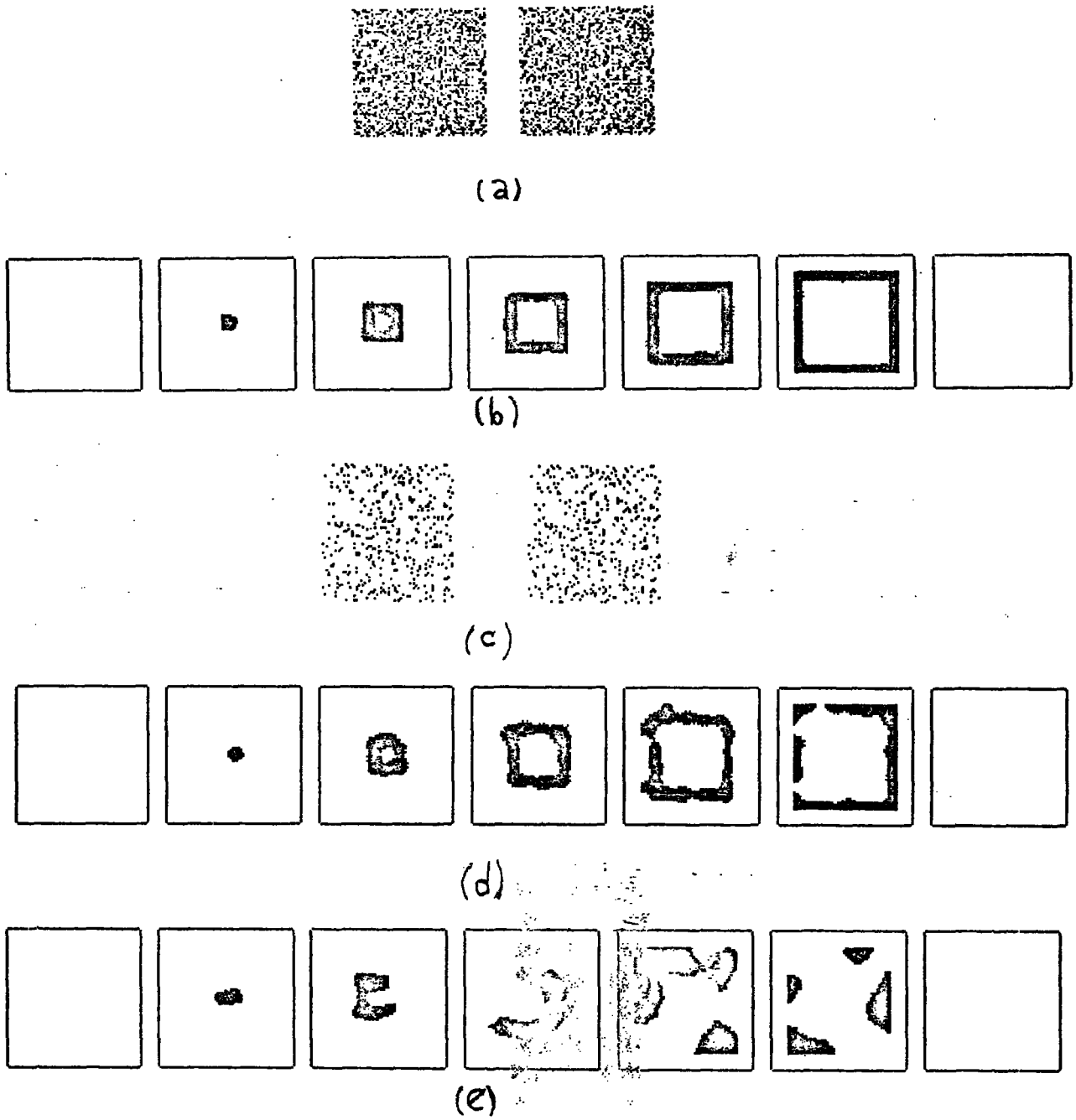

Fig. 8: (a) Dense Stereogram (density = 0.4) portraying a pyramid. (b) Fixed point for algorithm (17) (c) Sparse stereogram (density = 0.1) portraying a pyramid. (d) Fixed point for algorithm (17). (e) Fixed point for the Synchronous algorithm (9).

References

[D1] Davis L. and Rosenfeld A. "Cooperating Processes for Low Level Vision: A Survey" Art. Intell. 17, 245 (1981)

[F1] Feldman J. "Memory and Change in Connection Networks" TR96 C.S. Dept. Univ. of Rochester. (1981)

[H1] Hildreth E. and Ullman S. "The Measurement of Visual Motion" AIM 699 AI Lab. MIT (1982)

[H2] Hinton g. and Sejnowski T. "Optimal Perceptual Inference" Proc. IEEE CVPR83 Washington, D.C. (1983)

[H3] Hopfield J. "Neural Networks and Physical Systems with Emergent Computational Abilities" Proc. Nat. Ac. Sc. USA 79, 2554 (1982)

[K1] Kirkpatrick S. et al. "Optimization by Simulated Annealing" Science 220, 4598, 671 (1983)

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

[M2] Mc Quarrie D. "Statistical Thermodynamics" Harper & Row (1973)

[MP1] Marr D. and Poggio T. "Cooperative Computation of Stereo Disparity" AIM 671. AI Lab. MIT (1976)

[MP2] Marr D., Palm G. and Poggio T. "Analysis of a Cooperative Stereo Algorithm" Biol. Cyb. 28, 223 (1978)

[T1] Terzopoulos D. "Multilevel Reconstruction of Visual Surfaces" AIM 671 AI Lab. MIT (1982)

[U1] Ullman S. "The Interpretaion of Visual Motion" MIT Press (1979)