MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo 364

June 1976

COOPERATIVE COMPUTATION OF STEREO DISPARITY

by

D. Marr and T. Poggio*

Abstract: The extraction of stereo disparity information from two images depends upon establishing a correspondence between them. This article analyzes the nature of the correspondence computation, and derives a cooperative algorithm that implements it. We show that this algorithm successfully extracts information from random-dot stereograms, and its implications for the psychophysics and neurophysiology of the visual system are briefly discussed.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the Laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-75-C-0643.

* Permanent address: Max-Planck Institut fur Biologische Kybernetik,
74 Tubingen 1, Spemannstrasse 38, Germany.

Introduction

Perhaps one of the most striking differences between a brain and today's computers is the amount of wiring. In a digital computer, the ratio of connexions to components is about three, whereas for the mammalian cortex it lies between 10 and 10,000 (1).

Although this fact points to a clear structural difference between the two, it is important to realise that this distinction is not fundamental to the nature of the information processing that each accomplishes, merely to the particulars of how it does it. In Chomsky's terms (2), it affects theories of performance but not theories of competence, because the nature of a computation that is carried out by a machine or a nervous system depends only on the problem to be solved, not on the available hardware (3). Nevertheless one can expect a nervous system and a digital computer to use different types of algorithm, even when performing the same underlying computation. Algorithms with a parallel structure, requiring many simultaneous local operations on large data arrays, are expensive for today's computers but probably well-suited to the highly interactive organization of nervous systems.

The class of parallel algorithms includes an interesting and not precisely definable subclass which we may call *cooperative algorithms* (3). Such algorithms operate on many "input" elements and reach a global organisation *via* local, interactive constraints. The term "cooperative" refers to the way in which local operations appear to cooperate in forming global order in a well-regulated manner. Cooperative phenomena are well-known in physics (4, 5), and it has recently been proposed that they may play an important role in biological systems as well (4, 6, 7, 8, 9, 10). One of the earliest suggestions along these lines is due to Julesz (11), who maintains that stereoscopic fusion is a cooperative process. His spring and dipoles model represents a suggestive metaphor for this idea. Besides its biological relevance, the extraction of stereoscopic information is an important and yet unsolved problem in visual information processing (12). For this reason -- and also as a case-in-point -- it seems interesting to describe a cooperative algorithm for this computation.

In this article, we shall (a) analyse the computational structure of the stereo-disparity problem, stating the goal of the computation and characterising the associated local constraints; (b) describe a cooperative algorithm that implements this computation; and (c) exhibit its performance on random-dot stereograms. Although the problem addressed here is not directly related to the question of how the brain extracts disparity information, we shall briefly mention some questions and implications for psychophysics and neurophysiology.

Computational Structure of the Stereo-disparity Problem

Because of the way our eyes are positioned and controlled, our brains usually receive similar images of a scene taken from two nearby points at the same horizontal level. If two objects are separated in depth from the viewer, the relative positions of their images will differ in the two eyes. Our brains are capable of measuring this disparity, and using it to estimate depth.

Three steps are involved in measuring stereo disparity: *(S1)* a particular

location on a surface in the scene must be selected from one image; *(S2)* that same location must be identified in the other image; and *(S3)* the disparity in the two corresponding image points must be measured.

If one could identify a location beyond doubt in the two images, for example by illuminating it with a spot of light, steps *S1* and *S2* could be avoided and the problem would be easy. In practise one cannot do this (see figure 1), and the difficult part of the computation is solving the correspondence problem. Julesz found that we are able to interpret random-dot stereograms, which are stereo pairs that consist of random dots when viewed monocularly, but which fuse when viewed stereoscopically to yield patterns separated in depth. This might be thought surprising because when one tries to set up a correspondence between two arrays of random dots, false targets arise in profusion (see figure 1). Yet we are able to determine the correct correspondence. We need no other cues.

In order to formulate the correspondence computation precisely, we have to examine its basis in the physical world. Two constraints of importance may be identified (13): *(C1)* A given point on a physical surface has a unique position in space at any one time; *(C2)* Matter is cohesive, it is separated into objects, and the surfaces of objects are generally smooth compared with their distance from the viewer.

These constraints apply to locations on a physical surface. Therefore when we translate them into conditions on a computation we must ensure that the items to which they apply there are in (1-1) correspondence with well-defined locations on a physical surface. To do this, one must use surface markings, normal surface discontinuities, shadows *etc.*, which in turn means using predicates that correspond to changes in intensity. One solution is to obtain a primitive description (like the primal sketch (15)) of the intensity changes present in each image, and then to match these descriptions. Line and edge segments, blobs, termination points, and tokens obtained from these by grouping, usually correspond to items that have a physical existence on a surface.

The stereo problem may thus be reduced to that of matching two primitive descriptions, one from each eye. One can think of the elements of these descriptions as carrying only position information, like the black squares in a random-dot stereogram, although in practise there will exist rules about which matches between descriptive elements are possible, and which are not. The two physical constraints *C1* and *C2* can now be translated into two rules for how the left and right descriptions are combined:

(R1) Uniqueness. Each item from each image may be assigned at most one disparity value. This condition relies on the assumption that an item corresponds to something that has a unique physical position.

(R2) Continuity. Disparity varies smoothly almost everywhere. This condition is a consequence of the cohesiveness of matter, and it states that only a small fraction of the area of an image is composed of boundaries that are discontinuous in depth.

It is important to stress that in real life, *R1* cannot be applied simply to grey-level points in an image. The simplest counter-example is that of a goldfish swimming in a bowl, because many points in the image receive contributions from the bowl and from

FIGURE 1. There is ambiguity in the correspondence between the two retinal projections. In this figure, each of the four points in one eye's view could match any of the four projections in the other eye's view. Of the 16 possible matchings only four are correct (filled circles), while the remaining 12 are "false targets" (open circles). It is assumed here that the targets (filled squares) correspond to "matchable" descriptive elements obtained from the left and right images. Without further constraints based on global considerations, such ambiguities cannot be resolved. Redrawn after Julesz (ref. 12 figure 4.5-1).

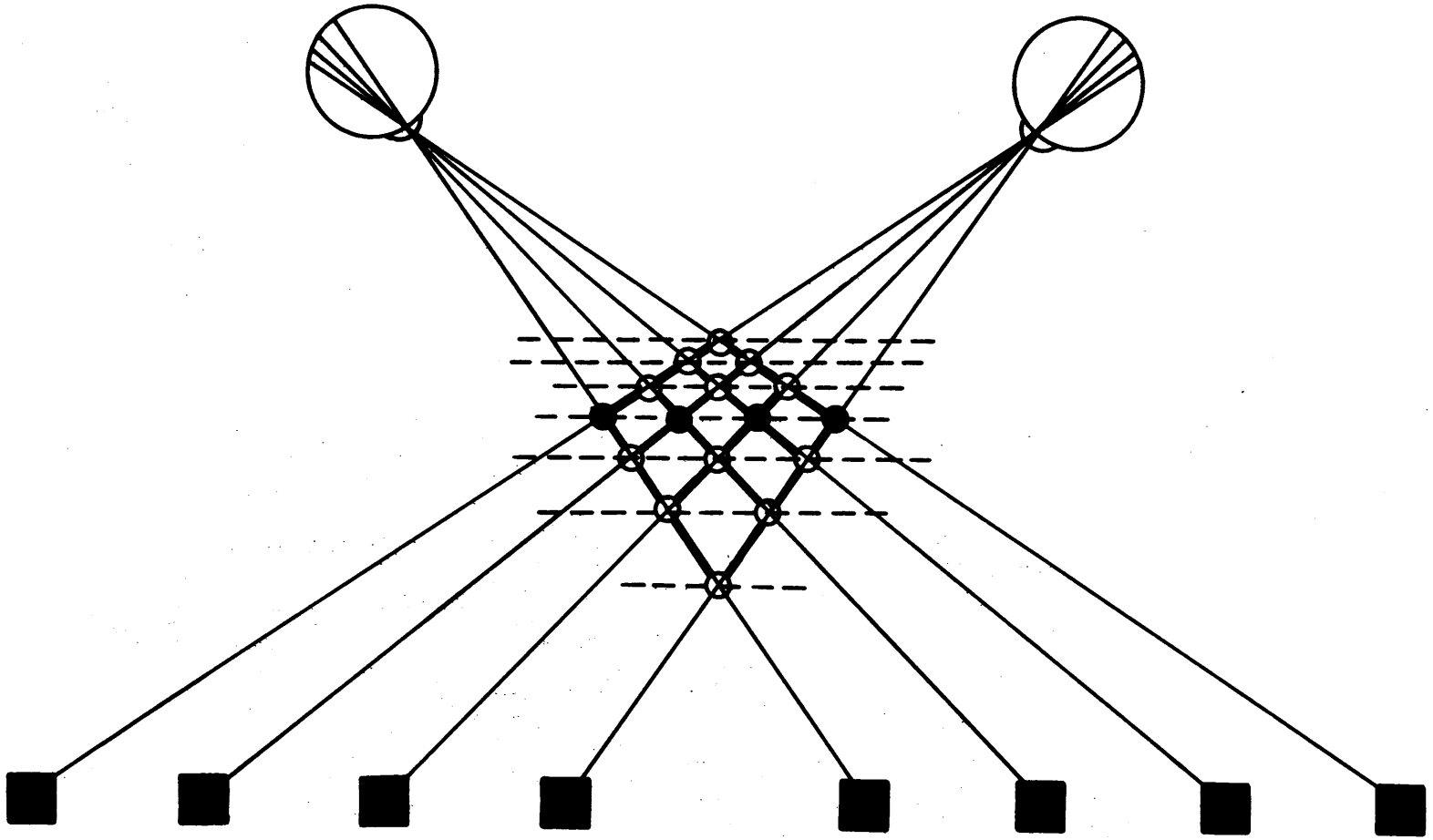

FIGURE 1

FIGURE 2. Figure 2a shows the explicit structure of the two rules *R1* and *R2* for the case of a one-dimensional image, and it also represents the structure of a network for implementing the algorithm described by equation 2. Solid lines represent "inhibitory" interactions, and dotted lines represent "excitatory" ones. 2b gives the local structure at each node of the network 2a. This algorithm may be extended to two-dimensional images, in which case each node in the corresponding network has the local structure shown in 2c. Such a network was used to solve the stereograms exhibited in figures 3 - 6.

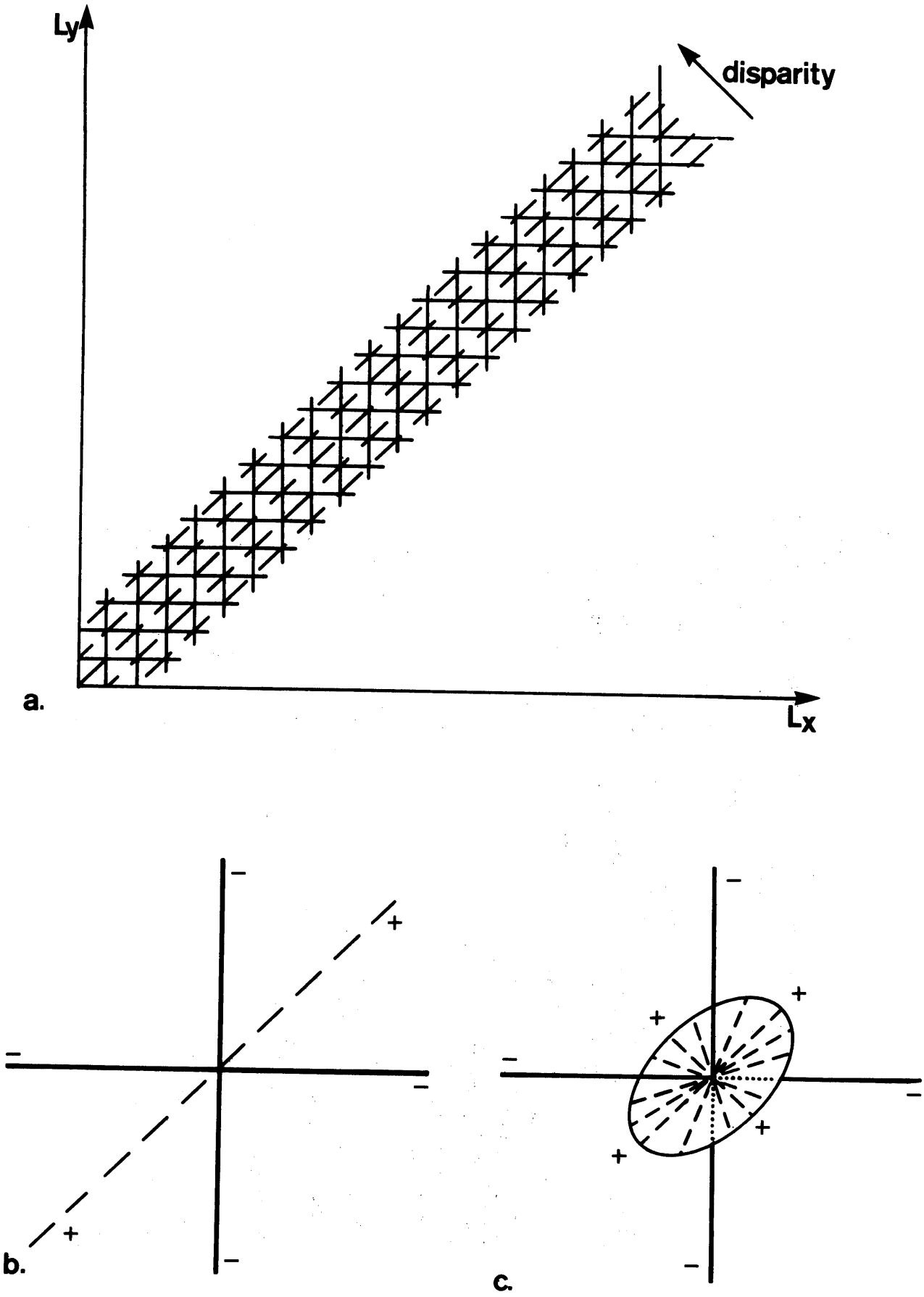

FIGURE 2

the goldfish. Here, and in general, a grey-level point is in only implicit correspondence with a physical location, and it is therefore impossible to ensure that grey-level points in the two images correspond to exactly the same physical position. Sharp changes in intensity are usually due either to the goldfish, or to the bowl, or to a reflexion, and therefore define a single physical position precisely.

A Cooperative Algorithm

By constructing an explicit representation of the two rules, we can derive a cooperative algorithm for the computation. Figure 2a exhibits their geometry in the simple case of a one-dimensional image. *Lx* and *Ly* represent the positions of descriptive elements on the left and right images. The thick vertical and horizontal lines represent lines of sight from the left and right eyes, and their intersection points correspond to possible disparity values. The dotted diagonal lines connect points of constant disparity.

The uniqueness rule *R1* states that only one disparity value may be assigned to each descriptive element. If we now think of the lines in figure 2a as a network, with a node at each intersection, this means that only one node may be switched on along each horizontal or vertical line.

The continuity rule *R2* states that disparity values vary smoothly almost everywhere. That is, solutions tend to spread along the dotted diagonals.

If we now place a "cell" at each node (figure 2b), and connect it so that it inhibits cells along the thick lines in the figure, and excites cells along the dotted lines, then provided the parameters are appropriate the stable states of such a network will be precisely those in which the two rules are obeyed. It remains only to show that such a network will converge to a stable state, and we were able to carry out a combinatorial analysis (as in refs. 9 & 15) which established its convergence for random-dot stereograms (16).

This idea may be extended to two-dimensional images simply by making the local excitatory neighbourhood two-dimensional. The structure of each node in the network for two-dimensional images is shown in figure 2c.

A simple form of the resulting algorithm (3) is given by the following set of difference equations:

$$(1) \quad C^{(n+1)} = \sigma\{\Xi(C^{(n)}) + C^{(0)}\} \quad , \text{ i.e.}$$

$$(2) \quad C^{(n+1)}_{xyd} = \sigma\{\sum_{x'y'd' \in S(xyd)} C^{(n)}_{x'y'd'} - \epsilon \sum_{x'y'd' \in O(xyd)} C^{(n)}_{x'y'd'} + C^{(0)}_{xyd}\}$$

where $C^{(n)}_{xyd}$ represents the state of the node or cell at position (x, y) with disparity d at iteration n, Ξ is the linear operator that embeds the local constraints (S and O are the circular and thick line neighborhoods of the cell xyd in figure 2c), and ϵ is the "inhibition" constant. σ is a sigmoid function with range [0, 1]. The state $C^{(n+1)}_{xyd}$ of the corresponding node at time $(n+1)$ is thus determined by a nonlinear operator on the output of a linear transformation of the states of neighbouring cells at time n.

The desired final state of the computation is clearly a fixed point of this algorithm, and moreover any state that is inconsistent with the two rules is not a stable fixed point. Our combinatorial analysis of this algorithm shows that, when σ is a simple

threshold function, the process converges for a rather wide range of parameter values (16). The specific form of the operator is apparently not very critical.

Non-iterative local operations cannot solve the stereo problem in a satisfactory way (11). Recurrence and non-linearity are necessary to create a truly cooperative algorithm that cannot be decomposed into the superposition of local operations (17). General results concerning such algorithms seem to be rather difficult to obtain, although we believe that one can usually establish convergence in probability for specific forms of them.

Examples of Applying the Algorithm

Random-dot stereograms offer an ideal input for testing the performance of the algorithm, since they enable one to bypass the costly and delicate process of transforming the intensity array received by each eye into a primitive description (14). When we ourselves view a random-dot stereogram, we probably compute a description couched in terms of edges rather than squares, whereas the inputs to our algorithm are the positions of the white squares. Figures 3, 4, 5 and 6 show some examples in which the iterative algorithm successfully solves the correspondence problem, thus allowing disparity values to be assigned to items in each image. Presently, its technical applications are limited only by the preprocessing problem.

This algorithm can of course be realised by various mechanisms, but parallel, recurrent, nonlinear interactions, both excitatory and inhibitory, seem the most natural. The difference equations set out above would then represent an approximation to the differential equations that describe the dynamics of the network.

Implications for Biology

We have hitherto refrained from discussing the biological problem of how stereopsis is achieved in the mammalian brain. Our analyses of the computation, and of the cooperative algorithm that implements it, raise several precise questions for psychophysics and physiology. An important preliminary point concerns the relative importance of neural fusion and of eye-movements for stereopsis. The underlying question is, are there many disparity "layers" (as our algorithm requires), or are there just three "pools" (18) -- crossed, uncrossed and zero disparity. Most physiologists and psychologists seem to accept the existence of numerous, sharply tuned binocular "disparity detectors", whose peak sensitivities cover a wide range of disparity values (19, 20). We do not feel that the available evidence is decisive (21), but an answer is critical to the biological relevance of our analysis. If for example there were only three pools or layers with a narrow range of disparity sensitivities, the problem of false targets is virtually removed, but at the expense of having to pass the convergence plane of the eyes across a surface in order to achieve fusion. Psychophysical experiments are presently under way to gain some insight into this problem, but we believe that only physiology ✲:●!%is capable of providing a clear-cut answer.

If this preliminary question is settled in favour of a "multi-layer" cooperative algorithm, there are several obvious implications of the network (figure 2) at the physiological level: (a) the existence of many sharply tuned disparity units, that are rather insensitive to the nature of the descriptive element to which they may refer; (b) their

FIGURE 3. This and the following figures show the results of applying the algorithm defined by equation 2 to two random-dot stereograms. The initial state of the network C is defined by the input such that a node takes the value 1 if it occurs at the intersection of a 1 in the left and right eyes (see figure 2), and it has value 0 otherwise. The network iterates on this initial state, and the parameters used here, as suggested by the combinatorial analysis, were $\theta = 3.0$, $\epsilon = 2.0$ and $M = 5$, where θ is the threshold and M is the diameter of the "excitatory" neighborhood illustrated in figure 2c. The stereograms themselves are labelled LEFT and RIGHT, the initial state of the network as 0, and the state after n iterations is marked as such. To understand how the figures represent states of the network, imagine looking at it from above. The different disparity layers in the network lie in parallel planes spread out horizontally, so that the viewer is looking down through them. In each plane, some nodes are on and some are off. Each of the seven layers in the network has been assigned a different gray level, so that a node that is switched on in the top layer (corresponding to a disparity of +3 pixels) contributes a dark point to the image, and one that is switched on in the lowest layer (disparity = -3) contributes a lighter point. Initially (iteration 0) the ☼☻!%network is disorganized, but in the final state, stable order has been achieved (iteration 14), and the inverted wedding-cake structure has been found. The density of this stereogram is 50%.

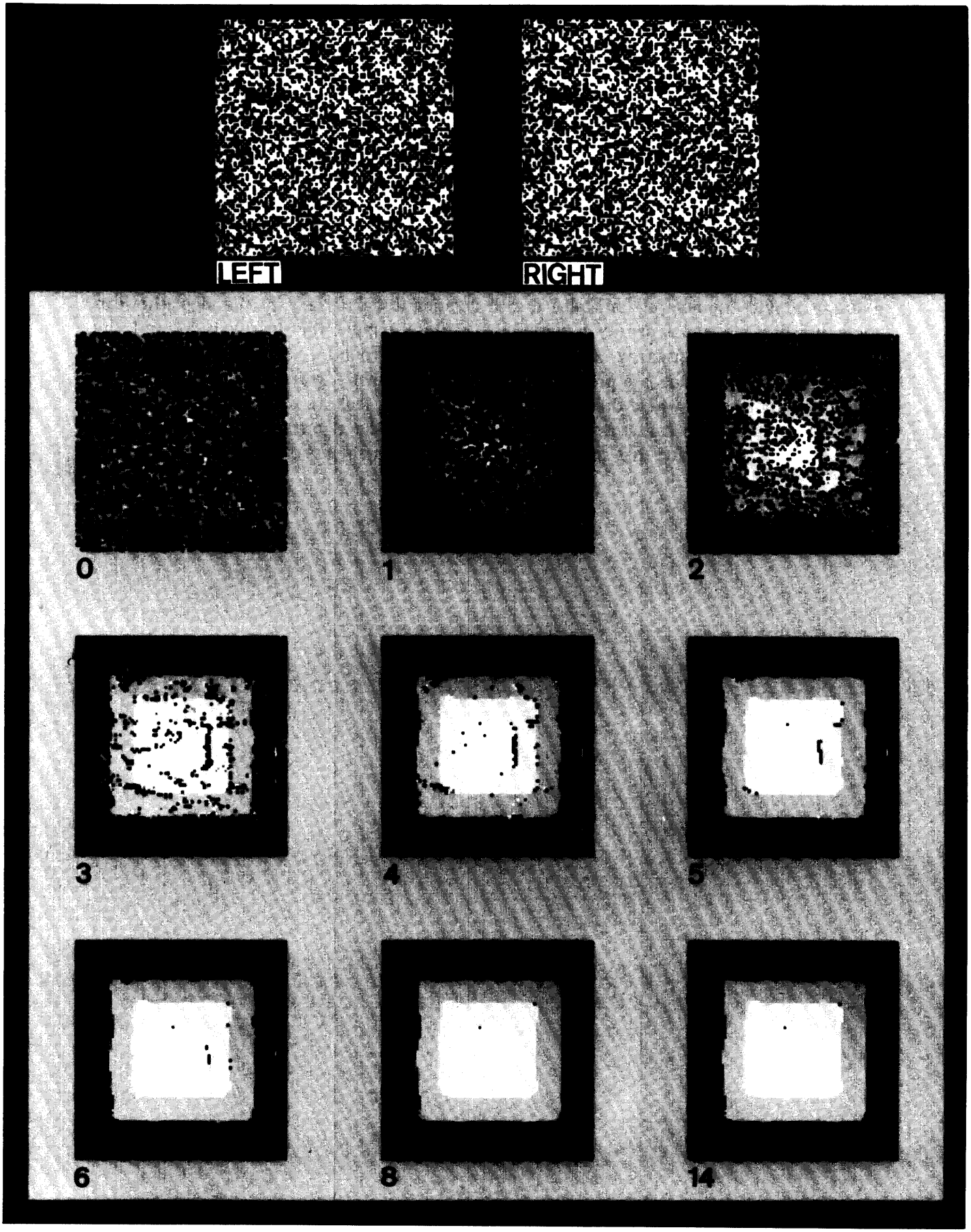

FIGURE 3

FIGURE 4. The algorithm of equation 2, with parameter values given in the legend to figure 3, is capable of solving random-dot stereograms with densities from 50% down to less than 10%. For this and smaller densities, the algorithm converges increasingly slowly. If a simple homeostatic mechanism is allowed to control the threshold θ as a function of the average activity (number of "on" cells) at each iteration (compare ref. 15), the algorithm can solve stereograms whose density is very low. In this example, the density is 5% and the central square has a disparity of +2 relative to the background. The algorithm "fills in" those areas where no dots are present, but it takes several more iterations to arrive near the solution than in cases where the density is 50%. When we look at a sparse stereogram, we perceive the shapes in it as cleaner than those found by the algorithm. This seems to be due to subjective contours that arise between dots that lie on shape boundaries.

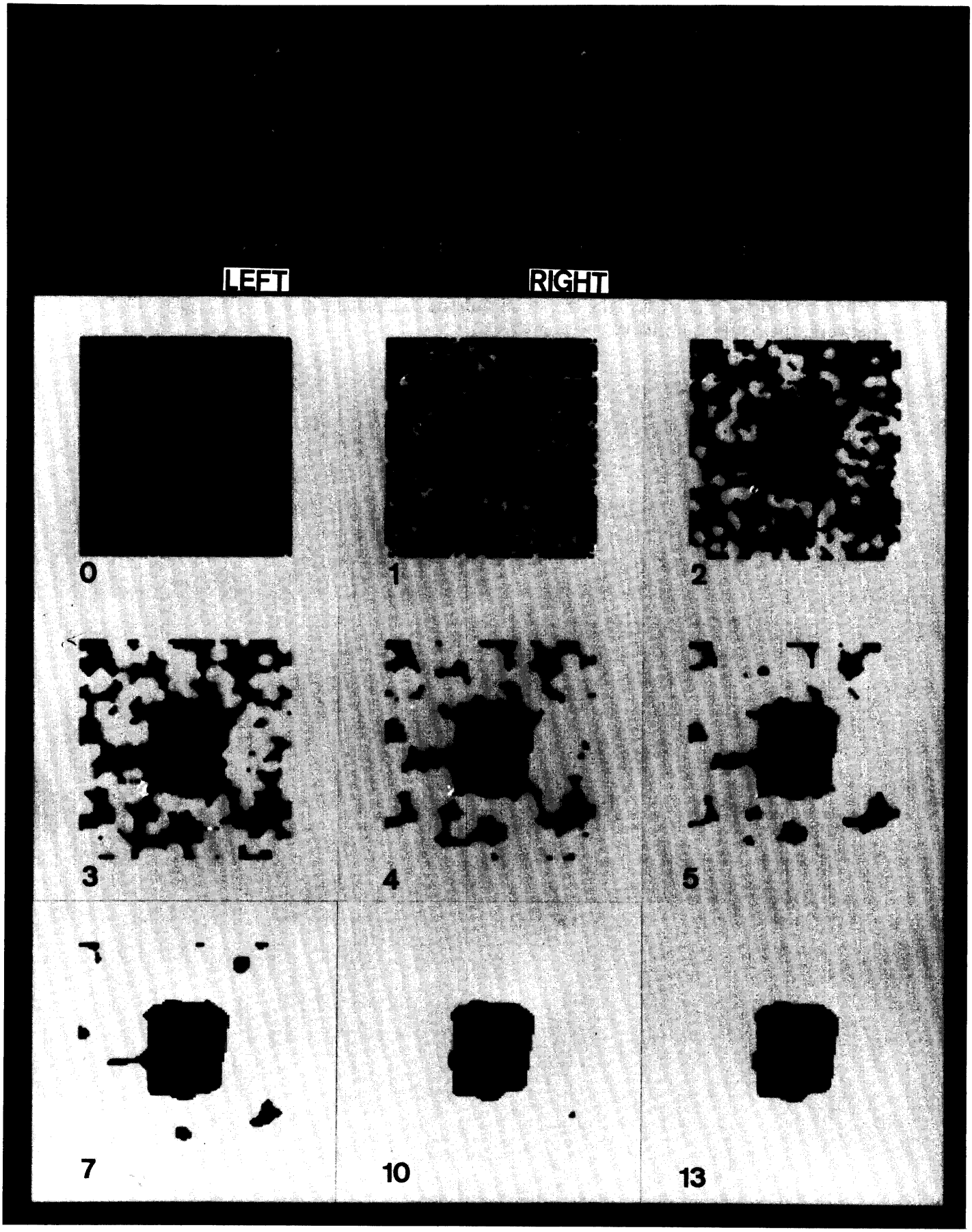

FIGURE 4

FIGURE 5. The disparity boundaries found by the algorithm do not depend on their shapes. In figures a, b and c we give examples of a circle, an octagon (notice how well the difference between them is preserved) and a triangle. The fourth example (d) shows a square in which the correlation is 100% at the boundary, but diminishes to 0% in the center. When one views this stereogram, the center appears to shimmer in a peculiar way. In the network, the center is unstable.

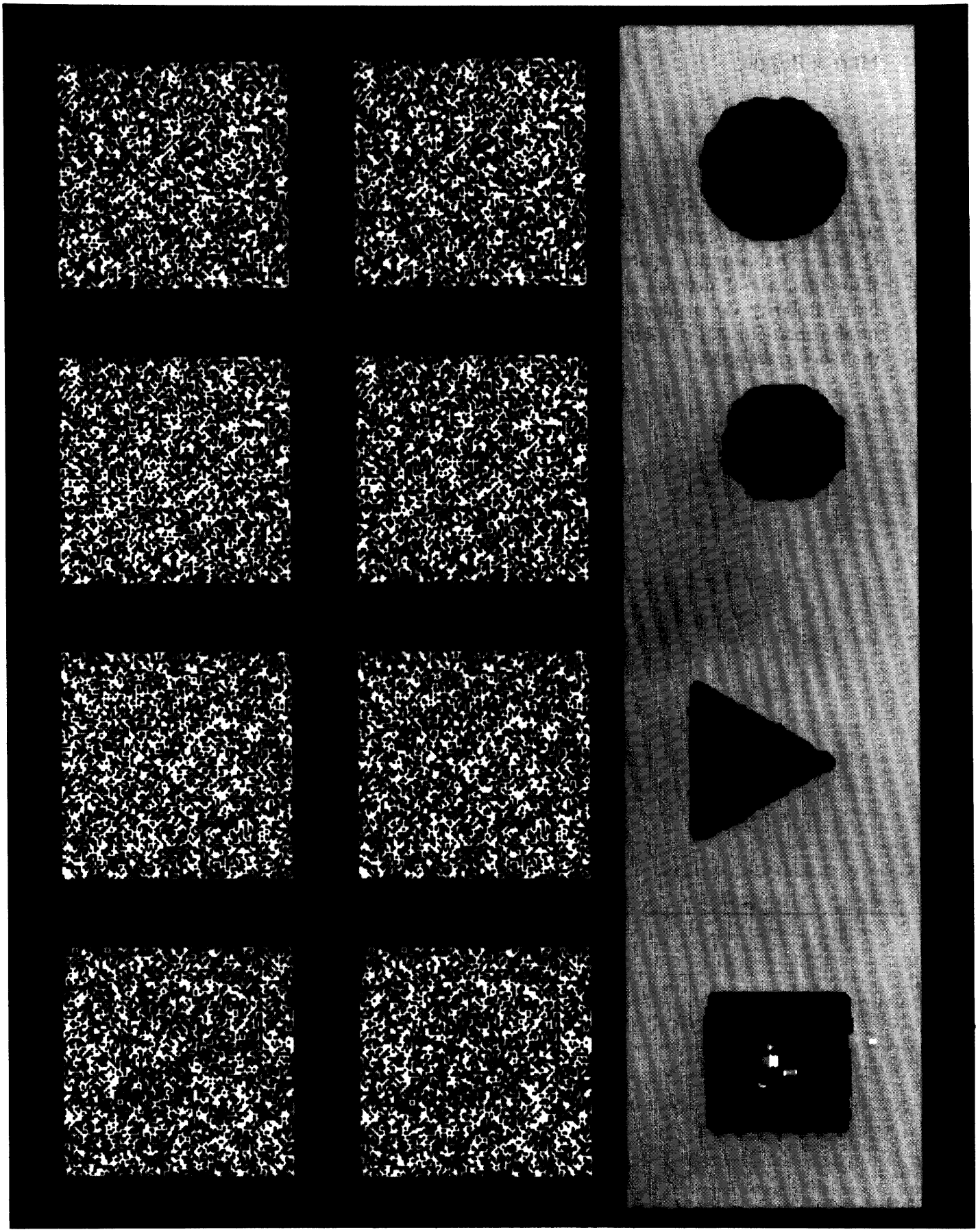

FIGURE 5

FIGURE 6. The width of the minimal resolvable area increases with disparity. In all four stereograms the pattern is the same, and consists of five circles with diameters 3, 5, 7, 9 and 13 dots. The disparity values exhibited here are +1, +2, +3 and +6, and for each pattern, we show the state of network after 10 iterations. As far as the network is concerned, the last pair (disparity +6) is uncorrelated, since only disparities from -3 to +3 are present in our implementation. After 10 iterations, information about the lack of correlation is preserved in the two largest areas.

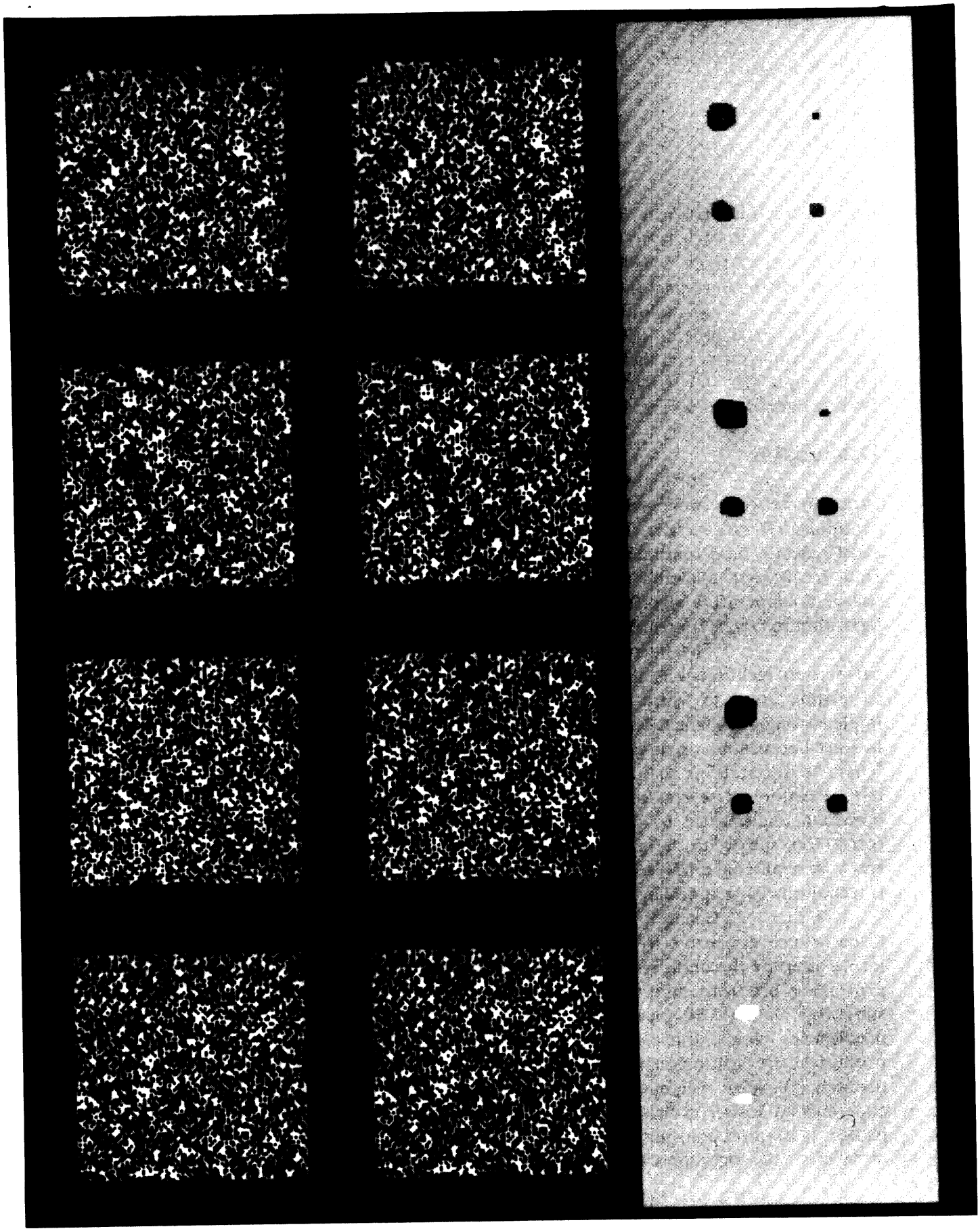

FIGURE 6

organisation into disparity layers (or stripes or columns); (c) the presence of reciprocal excitation within each layer; (d) the presence of reciprocal inhibition between layers along the two lines of sight. Ideally, the inhibition should exhibit the characteristic "orthogonal" geometry of the thick lines in figure 2, but slight deviations may be permissible (16).

At the psychophysical level, several experiments (under stabilized image conditions) could provide critical evidence for or against the network: (a) results about the size of Panum's area and the number of disparity "layers"; (b) results about "pulling" effects in stereopsis (20); (c) results about the relationship between disparity and the minimum fusable pattern size (see fig. 6).

Discussion

Our algorithm performs a computation that finds a correspondence function between two descriptions, subject to the two constraints of uniqueness and continuity. More generally, if one has a situation where "allowable" solutions are those that satisfy certain local constraints, a cooperative algorithm can often be constructed so as to find the "nearest" allowable state to an initial one. Provided that the constraints are local, use of a cooperative algorithm allows the representation of global order, to which the algorithm converges, to remain implicit in the network's structure.

The interesting difference between this stereo algorithm and standard correlation techniques is that one is not required to specify minimum or maximum correlation areas, to which the analysis is subsequently restricted. Previous attempts at implementing automatic stereocomparison through local correlation measurement have failed in part because no single neighbourhood size is always correct (12). The absence of a "characteristic scale" is one of the most interesting properties of this algorithm, and it is a central feature of several cooperative phenomena (22). We conjecture that the matching operation implemented by the algorithm represents in some sense a generalised form of correlation, subject to the *a priori* requirements imposed by the constraints. The idea is easily generalisable to different constraints and to other forms of equations (1) or (2), and it is technically quite appealing.

Cooperative algorithms may have many useful applications, (for example to best-match associative retrieval problems (15)), but their relevance to early processing of information by the brain remains an open question (23). Although a range of early visual processing problems might yield to a cooperative approach ("filling-in" phenomena, subjective contours (24), grouping, figural reinforcement, texture "fields", the correspondence problem for motion), it is important to emphasize that in problems of biological information processing, the first important and difficult task is to formulate the underlying computation precisely (3). After that, one can study good algorithms for it. In any case, we feel that an experimental answer to the question of whether depth perception is actually a cooperative process is a critical prerequisite to further attempts at analysing other perceptual processes in terms of similar algorithms.

References and Notes

1. D. A. Sholl, *The Organisation of the Cerebral Cortex* (Methuen, London, 1956). The comparison depends of course on what is meant by a component. We refer here to the level of a gate and of a neuron, respectively.
2. A. N. Chomsky, *Aspects of the Theory of Syntax.* (M.I.T. Press, Cambridge Mass., 1965).
3. D. Marr and T. Poggio, in *The Visual Field: Psychophysics and Neurophysiology. Neurosciences Research Program Bulletin*, E. Poeppel *et al.*, Eds. (in the press). Also available as *M. I. T. A. I. Lab. Memo 357.*
4. H. Haken, Ed. *Synergetics-Cooperative Phenomena in Multicomponent Systems.* (Teuber, Stuttgart, 1973).
5. H. Haken, *Rev. of Mod. Phys. 47*, 67 (1975).
6. J. D. Cowan, The problem of organismic reliability, in *Progress in Brain Research*, N. Wiener & J. P. Schade Eds., vol. 17. Amsterdam, Elsevier (1965).
7. H. R. Wilson and J. D. Cowan, *Kybernetik 13*, 55 (1973).
8. M. Eigen, *Naturwissenschaften 58*, 465 (1971).
9. P. H. Richter, *Die Phenomenologie der Immune Antwort.* (Contribution to a competition of the Bavarian Academy of Science, Max-Planck-Institut fur Biophys. Chemie, 1974).
10. A. Gierer and H. Meinhardt, *Kybernetik 12*, 30 (1972).
11. B. Julesz, *Foundations of Cyclopean Perception* (Univ. of Chicago Press, Chicago, 1971).
12. K. Mori, M. Kidode and H. Asada, *Comp. Graphics and Image Processing 2*, 393 (1973).
13. D. Marr, *M. I. T. A. I. Lab. Memo 327* (1974).
14. D. Marr, Early processing of visual information, *Phil. Trans. Roy. Soc. B*, (in the press).
15. D. Marr, *Phil. Trans. Roy. Soc. B 252*, 23 (1971). See especially section 3.1.2.
16. D. Marr and T. Poggio, in preparation.
17. T. Poggio and W. Reichardt, Visual control of orientation behaviour of the fly, part II. *Quarterly Rev. in Biophysics*, (in the press).
18. W. Richards, *J. Opt. Soc. Amer. 62*, 410 (1971).
19. H. B. Barlow, C. Blakemore and J. D. Pettigrew, *J. Physiol. (Lond.) 193*, 327 (1967); J. D. Pettigrew, T. Nikara and P. O. Bishop, *Exp. Brain Res. 6*, 391 (1968); C. Blakemore, *J. Physiol. (Lond.) 209*, 155 (1970).
20. B. Julesz and J-J. Chang, *Bio. Cybernetics 22*, 107 (1976).
21. D. H. Hubel and T. N. Wiesel, *Nature 225*, 41 (1970).
22. K. G. Wilson, *Rev. of Modern Physics 47*, 773 (1975).
23. Julesz (11), Cowan (6), and Wilson & Cowan (7) were the first to discuss explicitly the cooperative aspect of visual information processing. A large literature has recently been accumulating on possible cooperative processes in nervous systems, ranging from the "catastrophe" literature (E. C. Zeeman, *Scientific American 234*, 65, April 1976) to various attempts of more doubtful credibility. There has hitherto been no careful study of a cooperative algorithm in the context of a carefully defined computational problem (but see ref. 15), although algorithms that may be interpreted as cooperative were discussed, for instance, by P. Dev, *Int. J. Man-Machine Studies 7*, 511 (1975); and by A. Rosenfeld, R. A. Hummel and S. W. Zucker, *IEEE Trans. SMC-6*, 420 (1976). In particular neither Dev nor

J. I. Nelson, *J. theor. Biol. 49*, 1 (1975) formulated the computational structure of the stereo-disparity problem. As a consequence, the resulting geometry of the inhibition between their disparity detectors does not correspond to ours (see figure 2c) and apparently fails to provide a satisfactory algorithm.
24. S. Ullmann, *M. I. T. A. I. Lab. Memo 367* (1976).
25. We thank Whitman Richards for valuable discussions, Henry Lieberman for making it easy to create stereograms, and Karen Prendergast for preparing the figures. T. P. acknowledges the support of the Max-Planck-Gesellschaft during his visit to M.I.T.