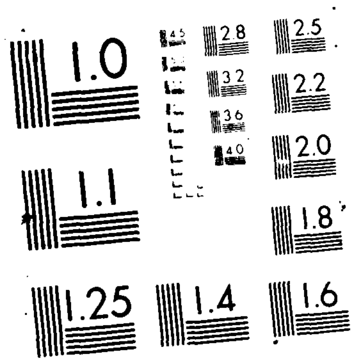
AD A185 041 MASSIVELY PARALLEL IMPLEMENTATIONS OF THEORIES FOR 1/1
APPARENT MOTION(U) MASSACHUSETTS INST OF TECH CAMBRIDGE
ARTIFICIAL INTELLIGENCE L.. N M GRZYWACZ ET AL. JUN 87
UNCLASSIFIED AI-M-888 N00014-85-K-0124 F/G 14/4 NL



1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(12)

AD-A185 841

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: A. I. Memo 888	2. GOVT ACCESSION NO.
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Massively Parallel Implementations of Theories for Apparent Motion	5. TYPE OF REPORT & PERIOD COVERED: AI-Memo; 1987
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Norberto M. Grzywacz and Alan L. Yuille	8. CONTRACT OR GRANT NUMBER(s): N00014-85-K-0[illegible]4
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Artificial Intelligence Laboratory, 545 Technology Square, Cambridge, MA 02139	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Advanced Research Projects Agency, 1400 Wilson Blvd., Arlington, VA 22209	12. REPORT DATE: June 1987
	13. NUMBER OF PAGES: 38
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Office of Naval Research, Information Systems, Arlington, VA 22217	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DTIC ELECTE NOV 0 6 1987 S D

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Analog networks — 3-D structure
Rigidity — Vision

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We investigate two ways of solving the correspondence problem for motion using the assumptions of minimal mapping and rigidity. Massively parallel analog networks are designed to implement these theories. Their effectiveness is demonstrated with mathematical proofs and computer simulations. We discuss relevant psychophysical experiments/

DTIC FILE COPY

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY
and
CENTER FOR BIOLOGICAL INFORMATION PROCESSING
WHITAKER COLLEGE

A.I. Memo No. 888 June 1987
C.B.I.P. Memo No. 016

MASSIVELY PARALLEL IMPLEMENTATIONS OF THEORIES FOR APPARENT MOTION

Norberto M. Grzywacz and Alan L. Yuille[1]

ABSTRACT We investigate two ways of solving the correspondence problem for motion using the assumptions of minimal mapping and rigidity. Massively parallel analog networks are designed to implement these theories. Their effectiveness is demonstrated with mathematical proofs and computer simulations. We discuss relevant psychophysical experiments.

Acknowledgments. This report describes research done within the Artificial Intelligence Laboratory and the Center for Biological Information Processing (Whitaker College) at the Massachusetts Institute of Technology. Support for the A.I. 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 N00014-85-K-0124. Support for this research is also provided by a grant from the Office of Naval Research, Engineering Psychology Division.

[1] Current address is Harvard University Department of Applied Mathematics, G12e Pierce Hall, Cambridge, MA 02138.

A-1

Table of Contents

1 Introduction

One of the most important roles of the early human visual system is the extraction of the three-dimensional (3-D) structure of surfaces (Marr, 1982). It has been proposed that the system deals with this task through different modules, each analyzing a different type of image information. One of the most important of these modules is the one that recovers the 3-D shape of objects from their motion cues. Indeed humans are capable of recovering structure from motion, under both orthographic and perspective projection, and in the absence of all other cues to 3-D structure (for examples of the early work see Wallach and O'Connell, 1953; Gibson and Gibson, 1957; White and Mueser, 1960; Green, 1961; Braunstein, 1962; Johansson, 1964; for a review of the psychophysical literature see Hildreth, Inada, Grzywacz and Adelson, 1987).

The problem of the recovery of structure from motion is underconstrained because the image information available in the retina is two-dimensional (2-D), and therefore, not enough to determine the 3-D shape of the visual world. To solve this problem, Ullman (1979) proposed that the human visual system uses assumptions about the world, such as rigidity of objects, to constrain the solution. His ideas led to a large body of computational work testing the validity of different assumptions directed to solve the structure from motion problem (for examples of the early work see Ullman, 1979; Clocksin, 1980; Prazdny, 1980; Longuet-Higgins, 1981; Longuet-Higgins and Prazdny, 1981; Tsai and Huang, 1981; for a review of the computational literature see Grzywacz and Hildreth, 1987).

Ullman used psychophysical data to argue that the process is divided into two stages. The first is solving the so-called *correspondence problem*, which consists of matching tokens, such as points or straight lines, between different image frames (see explanation below). He suggested that once this matching is done the second stage assumes rigidity of the object's structure in order to recover its 3-D shape. (Later, Ullman relaxed the assumption of rigidity in favor of a scheme in which the transformations of structure from frame to frame would be as rigid as possible, although not strictly rigid; Ullman, 1984.)

It is not necessary to postulate a solution of the structure from motion problem in terms of isolated features. In fact, optical flow approaches to the problem have been suggested (e.g. Prazdny, 1980; Longuet Higgins and Prazdny, 1981; Hoffman, 1982; Waxman and Ullman, 1985). There are reasons, however, to consider feature-based schemes. The main reason is that the optical flow field (a 2-D field that can be associated with the variation of the image brightness pattern) and the 2-D motion field (the projection on the image plane of the 3-D velocity field of a moving scene), rarely coincide. For some analytic models of surface reflectance this can be proven (Verri and Poggio, 1986). The problem stems from the fact that image brightness patterns and their changes do not correspond directly to physical entities and their motion (Ullman, 1979). Not surprisingly, however, it turns out from Verri and Poggio's work, that the optical flow and motion field nearly coincide at brightness edges and thus at the most elementary type of features.

Another reason to consider the feature based schemes is that a reliable recovery of structure from motion seems to require, a simultaneous inspection of image frames that have large separations in time (Wallach and O'Connell, 1953; White and Mueser, 1960; Green, 1961; Braunstein and Andersen, 1984; Doner, Lappin and Perfetto, 1984; Andersen and Siegel, 1986; Braunstein, Hoffman, Shapiro, Andersen and Bennett, 1986; Hildreth et al., 1987, Grzywacz, Hildreth, Inada and Adelson, 1987). This requirement brings back the correspondence problem mentioned above. In simple words, this is the problem of matching parts in different image frames such that matched primitives correspond to the same features in the viewed object.

The human visual system is able to solve the correspondence problem even when the motion is presented in discrete frames which have large separations in time. This is the phenomenon of long-range apparent motion. (Two distinct processes for the measurement of motion seem to exist in the human visual system (Braddick, 1974, 1980), one dealing with large separations in space and time, the long-range motion process, and the other dealing with small separations, the short range motion process.) Apparent motion has been studied extensively in the psychophysical literature (see, for example, Wertheimer, 1912; Korte, 1915; Kolers, 1972; Attneave, 1974; Braddick, 1980; Ullman, 1979; Anstis, 1980; Green, 1983, 1986; Mutch, Smith and Yonas, 1983; Ramachandran and Anstis, 1983, a,b,c, 1985; Anstis and Mather, 1985; Mather, Cavanagh and Anstis, 1985; Ramachandran, 1985; Anstis and Ramachandran, 1986; Green and Odom, 1986; von Grunau, 1986; Grzywacz, 1986, 1987; Prazdny, 1986; Ramachandran, Inada and Kiama, 1986; Watson, 1986; Finlay and Dodwell, 1987).

Ullman (1979) proposed a computational theory for apparent motion, which he called the *Minimal Mapping Theory.* Minimal mapping is the process by which features in a given frame are matched to features in another frame such that the sum of the distances traveled is minimal. (For psychophysical evidence supporting minimal mapping as an important factor in apparent motion see Ullman, 1979; Williams and Sekuler, 1984; Green and Odom, 1986.) This theory proposes, therefore, to solve the correspondence problem through the minimization of a cost function. (However, note that strictly speaking Ullman's theory does not require the minimization of the sum of Euclidian distances, but it allows for most abstract distances such as difference of orientation or brightness of the features. In this paper we consider only the Euclidian version of the theory.)

Finding the correct cost function, however, is only half the problem. We need a fast and reliable method of minimizing it. If the cost function is convex there exist many fast and reliable methods for finding the global minimum. For non-convex cost functions stochastic relaxation strategies like the Metropolis (Metropolis, Rosenbluth, Rosenbluth, Teller and Teller, 1953) or the simulated annealing algorithms (Kirkpatrick, Gelatt and Vecchi, 1983) will generally find the global minimum, but reportedly take a long time to do so. (For examples of the use of stochastic relaxation methods in computational vision see, Ballard, Hinton and Sejnowski, 1983; Hinton and Sejnowski, 1983; Geman and Geman, 1984; Marroquin, 1984; Divko and Schulten, 1986; Kienker, Sejnowski, Hinton and Schumacher, 1986; O'Toole and Kersten, 1986; Sereno, 1986.) Ullman (1979) used a linear programming method to solve the correspondence problem, and although this always converged correctly it did so very slowly (Ullman, pers. comm.) Instead of a slow algorithm that always converges to the right answer it may often be a better strategy to use a fast algorithm that converges to almost the right answer most of the time. This suggests implementing the problem in terms of deterministic analog networks with parallel architecture (for examples of the use of deterministic analog networks in computational vision, see Amari, 1977; Dev, 1975; Marr and Poggio, 1976; Ullman, 1979; Feldman and Ballard, 1982; Poggio, Torre and Koch, 1985; Fukushima, 1986; Grzywacz and Yuille, 1986; Hutchinson and Koch, 1986; Koch, Marroquin and Yuille, 1986; Rummelhart, Hinton, Williams, 1986; Little, Bulthoff and Poggio, 1987.)

An important example of nonlinear analog networks studied in the literature [illegible] tems whose elementary units are built out of resistors, capacitors and inductances, and whose elementary units are connected through devices that implement a static nonlinearity. If this nonlinearity is a sigmoidal input–output relationship, similar to those implemented by synapses, then these networks are called "neural networks" (Hopfield, 1982, 1984; Hopfield and Tank, 1985) since its units may be regarded as simplified models of neurons. We emphasize, however, that real neurons are complex computational devices (von Neumann, 1958; Koch, Poggio and Torre, 1982; Crill and Schwindt, 1983; Kuffler, Nichols and Martin, 1984) and that the name "neural–network" is used here only as a metaphor.

Currently, research is being done to construct electronical devices that implement such networks. If built, they will perform calculations extremely fast because of their parallel, analog nature. Hopfield and Tank (1985) have shown that these networks are capable of calculating good approximate solutions to complex minimization problems, such as the Traveling Salesman Problem. Koch, Marroquin and Yuille (1986) successfully applied them to the surface interpolation problem of early vision.

The present paper proposes and studies massively "neural–network" implementations designed to solve the correspondence problem in apparent motion (where "massively" means that every two elementary units are interconnected).

In Section 2 we describe a "neural–network" implementation of a version of the Minimal Mapping Theory. In the same section we give examples of computer simulations of this implementation, and show that it accounts for the basic psychophysical apparent motion phenomenology. This section also presents a demonstration of the speed of the "neural–network" implementation and of the fact that even for very complex, nonrigid motion, a nearly optimal solution is obtained. In Section 3 we prove theorems about the convergence of the network and show that for some situations the system will always find the correct solution. In the same section we will discuss how we chose the network parameters for our computer simulations.

Section 4 is directed to another question. It is natural to ask whether errors are caused by dividing the structure from motion process into two stages; first solving the correspondence problem and then using the correspondence information to recover the 3–D shape of objects. Both processes are solved using different assumptions and it is possible that these conflict for some stimuli. In this section we use the same mathematical formalism used in the preceding sections to determine whether rigidity alone (the basic assumption used to recover the 3–D structure from motion) is sufficient to solve the correspondence problem (and simultaneously the structure from motion problem). We show that further constraints are usually needed to obtain the correct answers. This result gives a computational argument in favor of a division of the structure from motion process in the above two stages. We will also discuss a theory that combines the minimal mapping and rigidity assumptions and is able to solve the correspondence and the structure from motion problems simultaneously.

2 The Minimal Mapping Theory for Apparent Motion

This section will begin with a formal introduction to the Minimal Mapping Theory and propose a "neural network" implementation of this theory (Section 2.1). We then proceed to demonstrate that this implementation simulates the basic apparent motion psychophysical phenomenology (Section 2.1.1), i.e. ambiguous and unambiguous 2–D motions, wagon-wheel type illusions, and transparent and opaque 3–D motions. We also analyze the convergence time of the network in comparison with the time constant of its basic units and discuss the quality of the solutions obtained. These solutions are not strictly correct since the minimization procedure may become trapped in local minima. We show, however, that those solutions are near optimal. Our main result in this section is this: provided that the motion is sufficiently small, network parameters can be chosen such that convergence to the optimal solution is guaranteed.

2.1 A Network Implementation

In the Minimal Mapping Theory (Ullman, 1979), the image of an object with N features is described by the 2–D coordinates of point on the object, $(x_i(t), y_i(t))$, $i = 1, \cdots, N$. Let images be given at two instants, $t - \delta t$ and t, and let us begin by assuming that the number of features in the two instants are identical. We now define a set of binary correspondence variables V_{ia} such that if feature i in the first frame maps to feature a in the second frame then $V_{ia} = 1$, otherwise $V_{ia} = 0$. From the assumptions of the Minimal Mapping Theory we want to define a matching cost function, E_{MM}, which is minimized only when the total distance traveled by the features is minimal. We follow Ullman and let:

$$E_{MM} = \sum_{i,a}^{N} V_{ia} d_{ia}, \tag{2.1}$$

where,

$$d_{ia} = \left((x_a(t) - x_i(t - \delta t))^2 + (y_a(t) - y_i(t - \delta t))^2 \right)^{1/2}. \tag{2.2}$$

To find the correspondence, the Minimal Mapping Theory proposes to minimize E_{MM} with respect to V_{ia} requiring a bijective mapping, i.e. that all features in the first frame are matched exactly to one feature in the second frame.

In order to perform a fast minimization we adapt in this paper a "neural network" method proposed by Hopfield and Tank (1985). Consider a system with N^2 neural like elementary units symmetrically connected to each other. Each unit will represent a possible correspondence between feature i at instant $t - \delta t$ and feature a at instant t.

We first define a new array of variables, $[U_{ia}]$, which will represent the internal voltage of the "neural" units. These are internal variables of the new problem and have a monotonically increasing relationship to V_{ia} (which will represent the output of these units):

$$V_{ia} = \frac{1}{1 + e^{-2\lambda U_{ia}}}, \tag{2.3}$$

$$U_{ia} = \frac{1}{2\lambda} log \frac{V_{ia}}{1 - V_{ia}}. \tag{2.4}$$

where λ is a positive parameter of the problem. Although $-\infty < U_{ia} < \infty$, one can see from Eq. 2.3 that V_{ia} is still bounded between 0 and 1. We next define the *full energy function* to be

$$
\begin{aligned}
E = & \frac{A}{2}\sum_{a=1}^{N}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} V_{ia}V_{ja} + \sum_{i=1}^{N}\sum_{a=1}^{N}\sum_{\substack{b=1\\ b\neq a}}^{N} V_{ia}V_{ib} \\
& + \frac{B}{2}\left(\sum_{i=1}^{N}\sum_{a=1}^{N} V_{ia} - N\right)^2 + \frac{C}{2}E_{MM} \\
& + \frac{1}{2\lambda\tau}\sum_{i=1}^{N}\sum_{a=1}^{N}\left(V_{ia}\log(V_{ia}) + (1 - V_{ia})\log(1 - V_{ia})\right),
\end{aligned}
\tag{2.5}
$$

where A, B, C and τ are positive parameters of the problem. (We will informally identify each of the terms of the right hand side of Eq. 2.5 by the parameter leading it.) Minimization of the first component of the A term forces each feature in the second frame to maintain correspondence with as few features as possible in the first frame, (and vice versa for the second component). Minimization of the B term tends to force the total number of correspondences to be N. Thus the terms A and B together will tend to force a one–to–one correspondence between features in the two frames. The τ term is necessary to give a time constant for convergence of the network, as will be seen below. Finally, the parameter C serves to provide scaling for the physical dimensions, i.e. if the image of a given object is just an expansion of the image of another, then the network will obtain the same solution for the two objects, provided that C is scaled properly.

Perceptually, if the two image frames have a different number of features, say N_1 and N_2, usually splitting and fusion will take place, such that no feature will be left alone. It is easy to incorporate this effect into the energy function by substituting N in the B term of Eq. 2.5 by $N_{max} = \max(N_1, N_2)$. This was done for a few of our computer simulations.

Observe that if the U_{ia} variables are updated according to the differential equations:

$$
\frac{dU_{ia}}{dt} = -\frac{\partial E}{\partial V_{ia}}, \qquad 1 \leq i \leq N, \quad 1 \leq a \leq N, \tag{2.6}
$$

then the system will stop in a point of the solution space in which the function E is at one of its minima. To see this, observe that because of the monotonicity between U_{ia} and V_{ia} expressed in Eq. 2.3, the update rule, Eq. 2.6, will tend to force V_{ia} to descend down the gradient of E. Note that if λ is large enough the variables V_{ia} will tend to be either 0 or 1 and thus, in spite of the fact that the search process is in a continuous space, it will tend to force a binary decision to determine whether a correspondence is to be established or not. In fact using the chain rule for differentiation and Eq. 2.6 we find

$$
\frac{dE}{dt} = -\sum_{ia}\frac{\partial V_{ia}}{\partial U_{ia}}\frac{\partial E}{\partial V_{ia}}\frac{\partial E}{\partial V_{ia}}. \tag{2.7}
$$

From Eq. 2.4 we calculate

$$
\frac{\partial V_{ia}}{\partial U_{ia}} = \frac{2\lambda}{(1 + e^{-2\lambda U_{ia}})^2}. \tag{2.8}
$$

Therefore $dE/dt \leq 0$, which together with the fact that $E \geq 0$ proves that the system will reach equilibrium, and in that situation E will be at a minimum. Technically this means that E is a Liapunov function of the system (see also Hopfield, 1984).

The solution of Eq. 2.6 can be implemented by a "neural-network". To calculate the symmetric connection strength, $T_{ia,jb}$, between unit $i\ a$ and unit $j\ b$, and the external input currents, I_{ia} (data), we substitute Eq. 2.5 into Eq. 2.6:

$$\frac{dU_{ia}}{dt} = -A(V_i^{COL} + V_a^{ROW} - 2V_{ia}) + B(N - V) - Cd_{ia} - \frac{U_{ia}}{\tau}. \tag{2.9}$$

Here we have introduced a new notation. $V = \sum_{ia} V_{ia}$, $V_i^{COL} = \sum_a V_{ia}$ and $V_a^{ROW} = \sum_i V_{ia}$. Equation 2.9 is the equation of motion of the system and was what we simulated in the computer. Note that the time constant is τ. That implies that the internal resistivity and capacitance of the network units can be set constant, equal to each other and independent of the problem to be solved.

$T_{ia,jb}$ is the contribution to the rate of change of U_{ia} (the voltage of unit $i\ a$) by V_{jb} (the output of unit $j\ b$) and can therefore be readily obtained from Eq. 2.9:

$$T_{ai,bj} = -A\,(\delta_{ab}\,(1 - \delta_{ij}) + \delta_{ij}\,(1 - \delta_{ab})) - B. \tag{2.10}$$

Similarly I_{ia} is the contribution to the rate of change of U_{ia} which is independent of the state of other units:

$$I_{ia} = BN - Cd_{ia}. \tag{2.11}$$

The A term in Eq. 2.10 represents inhibitory connections within each row and each column of $[U_{ia}]$. The B term in Eq. 2.10 represents a global inhibition between every pair of units. Therefore, every two units are mutually connected, with a total of $N^4 - N^2$ connections.

The B term on Eq. 2.11 is the excitation bias and is equally applied to every unit. The C term in Eq. 2.11 is the inhibitory current through which the data is provided to the system. The larger the d_{ia}, the more a feature would have to travel between place i in the first frame to place a in the second frame, and the less favorable this connection should be, therefore more inhibition is applied to the corresponding "neural unit".

It is important to note that in contrast with Hopfield and Tank's method for the traveling salesman problem (Hopfield and Tank, 1985), the data enters into our system as applied currents and not as modifications of the connectivities between units.

In the next section we present the results of our computer simulations by the numerical integration of Eq. 2.9.

2.1.1 Computer Simulations

We simulated this network on a Symbolics 3600 LISP machine. In our simulations we did not try to optimize the parameters A, B, C, τ and λ in any sense. (Although for the simulations reported in this paper, we took into account the rules discussed in Section 3.) Instead we found that the asymptotic behavior of the system was the same for a large range of parameter values (few orders of magnitude), and that a given set of parameters would give correct simulations to problems with a different number of features. For all the simulations reported in this paper (unless reported otherwise) we used $A = 10^2, B = 10^4, C = 1, \tau = 1$ and $\lambda = 1$, and the maximal distance between features in an object was always 1. Finally, we used homogeneous initial conditions for our simulations, i.e.:

$$V_{ia}(t = 0) \;=\; \frac{1}{N}. \tag{2.11}$$

The first simulations showed that the network can correctly replicate apparent motion percepts. Figure 1 illustrates the matching predicted for a 10 feature object rotating by 10°. (Our simulations extended to objects containing up to 20 features.) In Fig. 1b the same object translates slightly. In our figures the features in the first frame are always represented by squares and those in the second frame by triangles. The labels for the features are maintained after the motion, so that the expected values for the $[V_{ia}]$ matrix at equilibrium should be close to 1 at the diagonal, and close to 0 off diagonal. The temporal evolution for this matrix in the rotation case of Fig. 1 is shown in a 3-D plot in Fig. 2. (A similar temporal evolution was obtained for the translation.) The solid lines in Fig. 1, and in similar figures afterwards, indicate the established correspondences, i.e. the maxima of the $[V_{ia}]$ arrays. The durations of network computation for this figure were 0.06τ and 0.045τ for the rotation and translation respectively. (We point out that the dependence on the complexity of the problem, of the convergence time of the simulated parallel network, is different than that of the CPU time of the computers in which the simulation was performed. This is because these computers were serial. Thus the CPU times were irrelevant for our conclusions and were not monitored.)

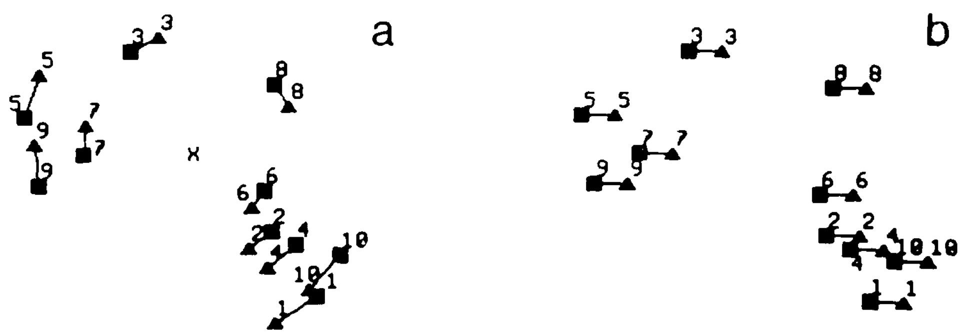


Figure 1. The network matching predictions for a moving object of 10 features. The positions of the features are represented in the first frame by squares and in the second by triangles. A specific feature is indicated by the same index in the two frames and the solid lines indicate the correspondences established by the network. a. The object is rotated by 10° around the optic axis. The x indicates the center of the rotation. b. The object is translated to the right. The correct correspondences were established in both cases. They are expected to be correct when the displacement between frames is small.

Note that the correct correspondence was obtained, i.e. the diagonal of the array $[V_{ia}]$ was preferred (Fig. 2). Incorrect matches were suppressed to several orders of magnitude below the correct ones. In the rotation case, even for feature number 10, which by simple proximity would prefer to match features 1, 2 or 4 (Fig. 1), the global consensus held and the correct correspondence was made.

Note in Fig. 2, that at $t = 0$ the array is flat, which indicates the lack of preference for any particular correspondence. Afterwards, a competition between the correspondences is initiated until the diagonal is preferred ($t = 0.00375\tau, 0.0075\tau, 0.015\tau$). Only after this diagonal is chosen, the last false matches are eliminated ($t = 0.03\tau, t = 0.06\tau$).

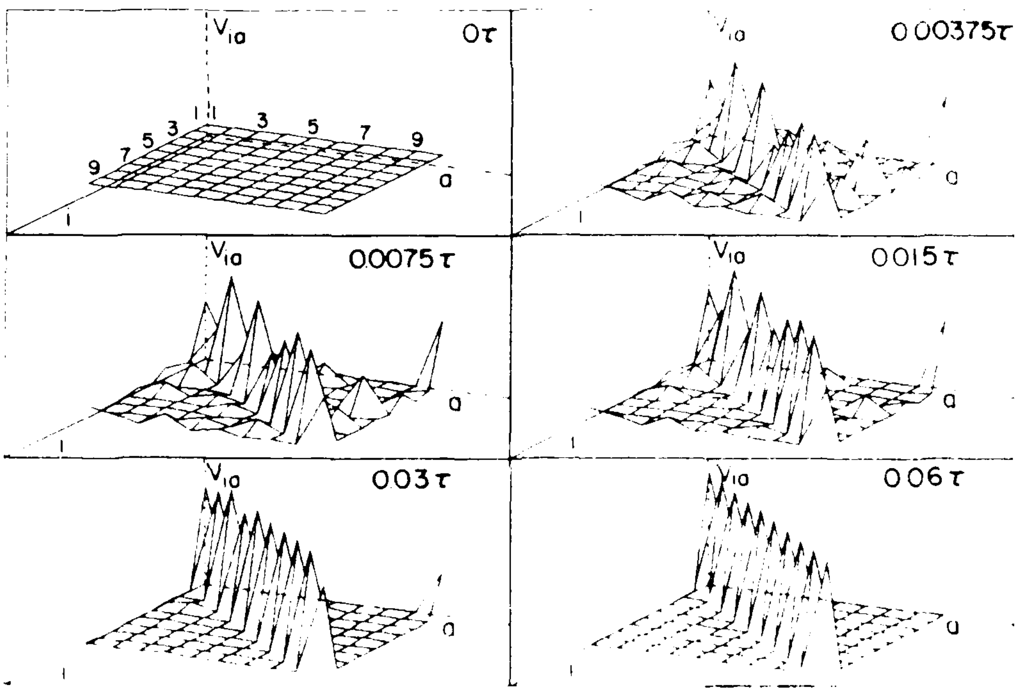


Figure 2. A 3 dimensional plot of the time evolution of the correspondence array for the rotation case of Fig. 1. In the six graphs, the V_{ia} axis represents the value of the array (ranging from 0 to 1), and the i and a axis represent the features indices in the first and second frames respectively. The times of computation for the arrays are displayed on the upper right corner of each graph. The correctedness of the correspondence found in Fig. 1 is illustrated here by the convergence of the array to a diagonal form.

Another result of interest in Figs. 1 and 2 is that the times of convergence were shorter than the time constant of the elementary units of the network ($< 0.06\tau$ and $< 0.015\tau$ for the rotation and the translation respectively). In Section 3, we prove that even in equilibrium the variables V_{ia} are different from 0 or 1, although they can approach these values arbitrarily closely. It follows that for practical purposes a criterion threshold has to be arbitrarily set to define convergence. For Fig. 1, for example, we set this threshold at $V_{ia} < 0.05$ or $V_{ia} > 0.95$, $1 \le i, a \le N$. That is, after $t = 0.06\tau$ in the rotation case and $t = 0.015\tau$ in the translation case, all the array elements were either below 0.05 or above 0.95. (This criterion was used for all figures in this paper.) The ...

that the convergence of the system was faster than the time constant of the elementary units means that the variables V_{ia} can pass the threshold criterion very fast, although technically they will reach equilibrium only after a time constant or so had elapsed. At any rate, the time of convergence of the system is limited only by τ, and can be very short. In all the figures in this paper the convergence time was much shorter than τ.

The example of Fig. 1 is such that the extent of motion is small. In Section 3, we prove a theorem which states that for short motions a choice of parameters can be made such that a convergence to the correct solution is guaranteed. The result in Fig. 1 confirms this theorem.

Not only for small motions, however, does the network simulate psychophysical percepts. In the case of large rotations, for example, perceptual illusions often occur. This is because in these situations, features can travel large distances, and may approach positions in the second frame that originally were occupied by other features. Such an example is the wagon-wheel illusion, a well known motion picture effect, in which a spoked wagon wheel seems to rotate in the direction opposite to its real sense of rotation. This illusion is also obtained by the network, and is illustrated in Fig. 3. In this example, eight features disposed in the corners of a perfect octagon rotate $11°15'$ in one case (Fig. 3a) and $33°45'$ in another case (Fig. 3.b). The 3-D plot of the matrices $[V_{ia}]$ at the convergence time are shown in Fig. 3c and 3d for Figs. 3a and 3b respectively. The convergence time for this figure was 0.02τ.

The wagon-wheel illusion is established by the incorrect correspondences that happen in the large rotation. (Instead of the diagonal, a rotation permutation of the array $[V_{ia}]$ was selected.) Once again, the incorrect matches were suppressed by many orders of magnitude.

The network can also deal in a psychophysically appropriate way with ambiguous situations, i.e. cases of perceptual metastability. An example of such a situation is shown in Fig. 4 and has been studied extensively in the psychophysical literature (Von Schiller, 1933; Gengerelli, 1948; Ramachandran and Antis, 1983, a,b,c; 1985). It consists of two features disposed at the end of an imaginary rigid rod. The rod rotates at each new frame by $90°$ around its center. The features in the second frame are equidistant to each one of the features in the first frame. It follows that a given feature in the first frame is equally likely to match both features in the second frame, thus giving rise to a metastable situation. The numerical values in the matrix $[V_{ia}]$ at the time of convergence are given in the figure. The time of convergence was $1.6 \times 10^{-4}\tau$, and the array did not change even after 10τ.

The metastability of the motion display is expressed in the fractional results computed by the network. This is possible, because the variables are not binary (Eq. 2.3), although often tend to 0 or 1 at equilibrium. The interpretation of these fractional results should be in probabilistic terms: i.e. a given feature in the first frame has a probability close to 0.5 of matching a given feature in the second frame. Indeed, when noise intervenes in the data to the network, i.e. when there is a random modulation of the distance between the features, the system no longer converges to 0.5, but rather, a one-to-one matching choice is made by the network. Finally, we point out that the sum of the matching probabilities for a feature reported by the network is less than 1, since all the $V_{ia} = 0.4975 < 0.5$. This result is not a numerical artifact, as in Section 3 we prove analytically that $V < N$ (where V was defined in Eq. 2.9). We also prove in the same section, however, that a choice of network parameters can be made such that V is arbitrarily close to N.

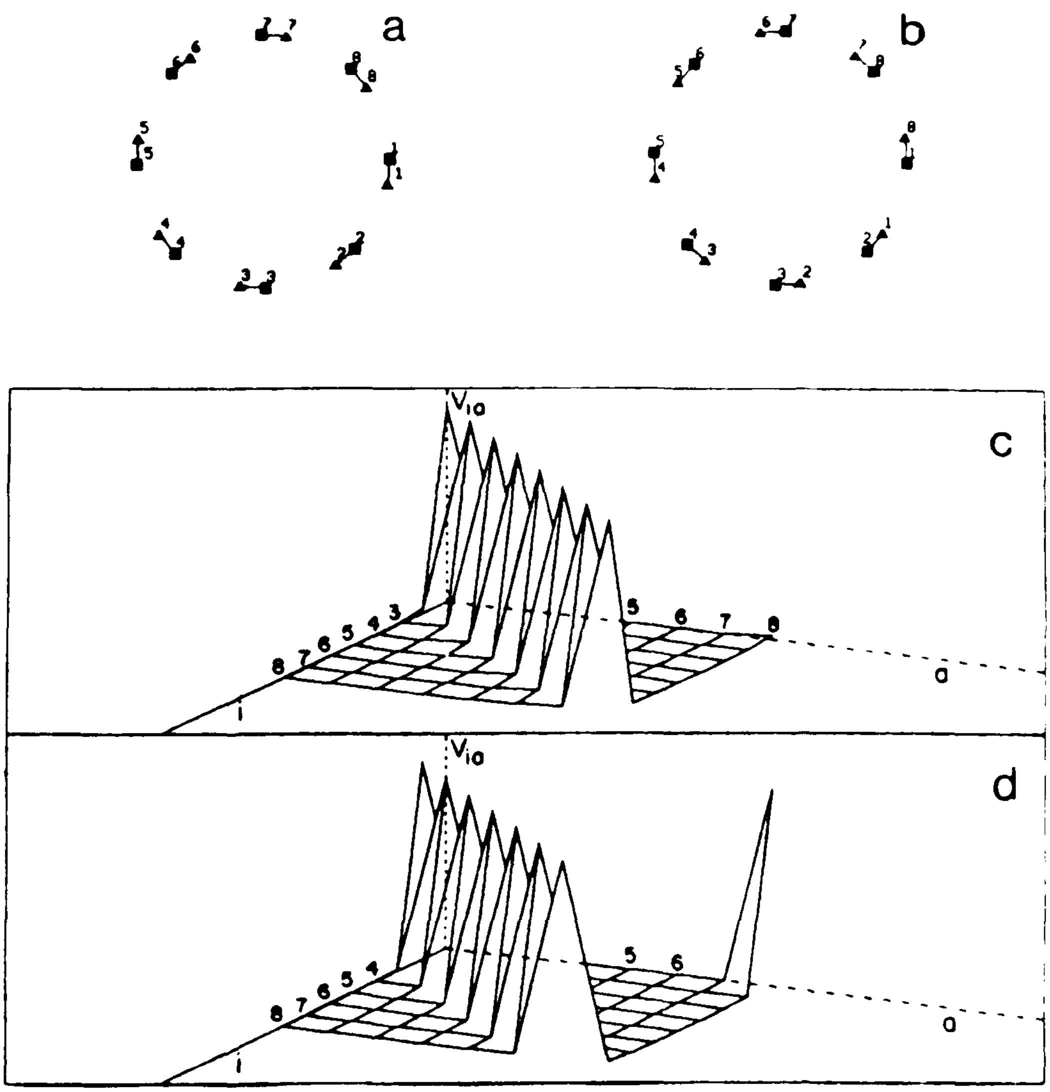
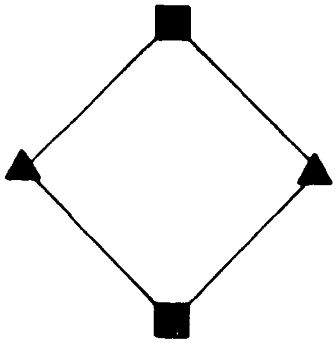


Figure 3. The wagon–wheel illusion. The symbols in Figs. a and b are the same as in Fig 1, and the axes of Figs. c and d are the same as in Fig. 2. An object whose features lie on the corners of a perfect octagon is rotated around the optic axis. The rotations were 11°15′ and 33°45′ in Figs. a and b, respectively. The established correspondence was correct for the small rotation but incorrect for the large one; the reported direction of rotation was reversed as is the case for humans. Figures c and d show the correspondence array at the time of convergence for the small and large rotations respectively. The illusion corresponds to the network converging to a diagonal form in the first case, but to a non-diagonal form in the second case.

(In humans, if the visual display of Fig. 4 is presented repeatedly, the percept is either of oscillation or rotation depending on the temporal parameters of the stimulus (Ramachandran and Antis, 1983, a,b,c; 1985). However, the percept predicted by the Minimal Mapping theory, and thus by our network, is random from presentation to presentation. In fact, it can be shown that



$$[V_{ia}] = \begin{pmatrix} 0.4975 & 0.4975 \\ 0.4975 & 0.4975 \end{pmatrix}$$

Figure 4. An ambiguous situation. The symbols are the same as in Fig. 1. The two features of the first frame are equally likely to match either feature of the second frame. The network deals with this problem by converging to values that are neither 0 nor 1. For all of the previous examples the networks converged to binary values. The matrix shown in the figure is the final value reached by the correspondence array. Its values are close to 0.5, and therefore close to the probability that a particular match is made. For humans such a display is bistable. The reason why the result is not exactly 0.5 is not a numerical artifact and is explained in the text.)

the solution $V_{ia} \approx 0.5$ is unstable, and any noise pushes the final values to 0 or 1. This discrepancy between the predictions and the psychophysics is accounted by the Minimal Mapping Theory's omission of information about the past motion of the features; see the Discussion section for more details on the limitations of the Minimal Mapping Theory.)

As pointed out in the introduction, Ullman (1979) suggested that the main role of apparent motion is to serve as the first stage in the process of recovering the 3-D structure of objects from their motion. It follows, therefore, that the apparent motion mechanism has to cope with perceptual oddities due to 3-D motion, particularly nonrigidity in the image, and appearance and disappearance of features due to occlusions. Figure 5 illustrates how the network deals with these problems and shows that its solutions are similar to those of the visual system.

In the figure, a 3 Dimensional 5-feature object is rotated by 27° around an axis which is perpendicular to the viewing axis, and which belongs to the plane that divides the head between left and right. From a bird's eye view, the features of the object lie on the corners of a perfect pentagon (Fig. 5a), and are projected orthographically into the image plane. This projection is shown in Figs. 5 b and c under the assumption that the object is transparent and opaque respectively. In the opaque case it is assumed that only the front features can be seen by the observer (see Fig. 5a).

In the transparent case all five features are seen, and the relative distance between features in the image change, because features in different positions in the surface have different velocities. Note in Fig. 5b that this image nonrigidity does not disturb the ability of the network to solve the correspondence problem. The convergence time for this figure was 0.12τ.

In the opaque case only three of the features are seen in the first frame and two in the second. The other features are occluded by the surface. The main problem that the network faces in this case is that the first frame has more features than the second. Perceptually this leads

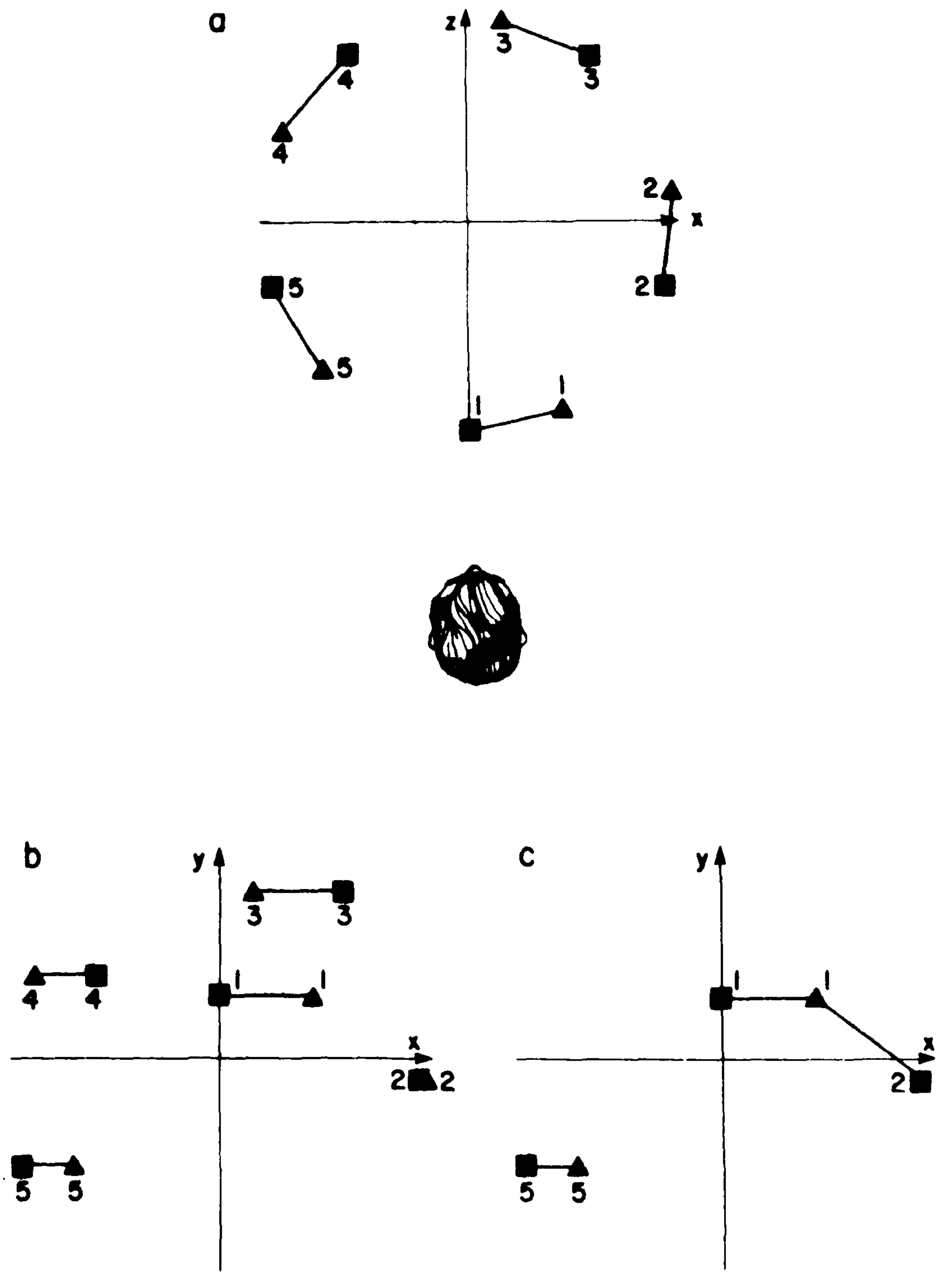


Figure 5. Nonrigidity and the appearance and disappearance of features. a. A bird's eye view of an object rotating by 27° around an axis perpendicular to the viewing axis and vertical in relation to the head of the observer (shown schematically in the figure). The features of the object lie on the corners of a perfect pentagon. b. The object is assumed transparent. The correspondences are computed correctly, in spite of the nonrigidity of the image, i.e. features travel by different amounts. c. The features are assumed to lie on the surface of an opaque cylinder. Note that feature 2 appears in the first frame, but not in the second. The solution of the network matches that of the human visual system, and features 1 and 2 fuse in the second frame.

to fusion, i.e. two features or more from the first frame match one in the second (Kolers, 1972). The network also obtained this solution (Fig. 5c, $t = 0.1\tau$) when N in Eq. 2.5 was substituted by N_{max} as explained in Section 2. Note also, that the fusion obtained by the network had the minimal mapping property, i.e. features tended to travel as little as possible. The same strategy (i.e. substituting N by N_{max}) leads to splitting, i.e. a feature in the first frame matches two or more in the second, if the number of features in the second frame is larger than that of the first. (This result is again similar to human perception; Kolers, 1972. Fusion and splitting, however, have been shown to disappear if the knowledge of occlusion is present; Ramachandran and Anstis, 1983,b.)

We show in Section 3, that for short motions, the right parameters can be chosen, such that the correct solution is obtained by the network. This seems to be the reason for the success of the network in the simulation of perceptual data (Figs. 1–5). This fact does not imply, however, that the network converges in general to the global minimum of the energy function given in Eq. 2.5. In fact we illustrate in Figs. 6 and 7 that for random motions an incorrect matching may be found. We also show, however, that even if the correspondence is incorrectly established, it is near optimal.

For Fig. 6 a computational experiment with 450 runs was done. For each run the first and second frame consisted of two objects of 6 features each, randomly placed in a disc of radius 1. The correct match, i.e. the one that minimizes the total distance traveled by the features, was established by exhaustive search. The network was then applied for the 450 runs and the number of cases that fell in each of the following four categories was observed: 1. correct answers, 2. incorrect answers but one–to–one matching, 3. lack of one-to-one matching but six matches, and 4. less than six matches. The frequency histogram is shown in Fig. 6.

Note that a one–to–one mapping was always established (and consequently the number of matches was always six). In this experiment, however, only 58.4% of the solutions computed by the network corresponded to minimal mapping.

In the other 41.6% of the cases, an incorrect answer was found. These incorrect solutions, however, were near optimal as seen in Fig. 7. Four motions for which a incorrect mapping was established are displayed in Figs. 7 a–d. In these figures the correct matches, as found by exhaustive search, are marked by the dotted lines, and the predictions of the network are marked by the solid lines. Note that the solutions found by the network were almost identical to the optimal ones, and the errors were each time the switching of only one pair of correspondences.

The histograms in Figs. 7, e–h, correspond to Figs. 7, a–d, respectively. They plot the distribution of the total distance traveled by the features, for the $6! = 720$ possible cases of one–to–one matching. The arrows in these histograms show the total distance traveled for the answer given by the network. Note that as predicted by Figs. 7 a–d, the network results fell in near optimal positions, i.e. many standard deviations away from the mean of the distribution.

Another fact of interest related to the experiment in Fig. 6, and which may provide a psychophysically testable prediction for such types of networks, is that the time of convergence is much longer on average for incorrect matches than it is for correct ones. In fact, for the last 150 runs of the experiment in Fig. 6, the mean time of convergence for cases where correct matches were predicted was $0.16\tau \pm 0.001\tau$ (standard error), and the mean time for the incorrect cases was $0.366\tau \pm 0.013\tau$. Errors are due to a conflict between the necessity for minimization of the total

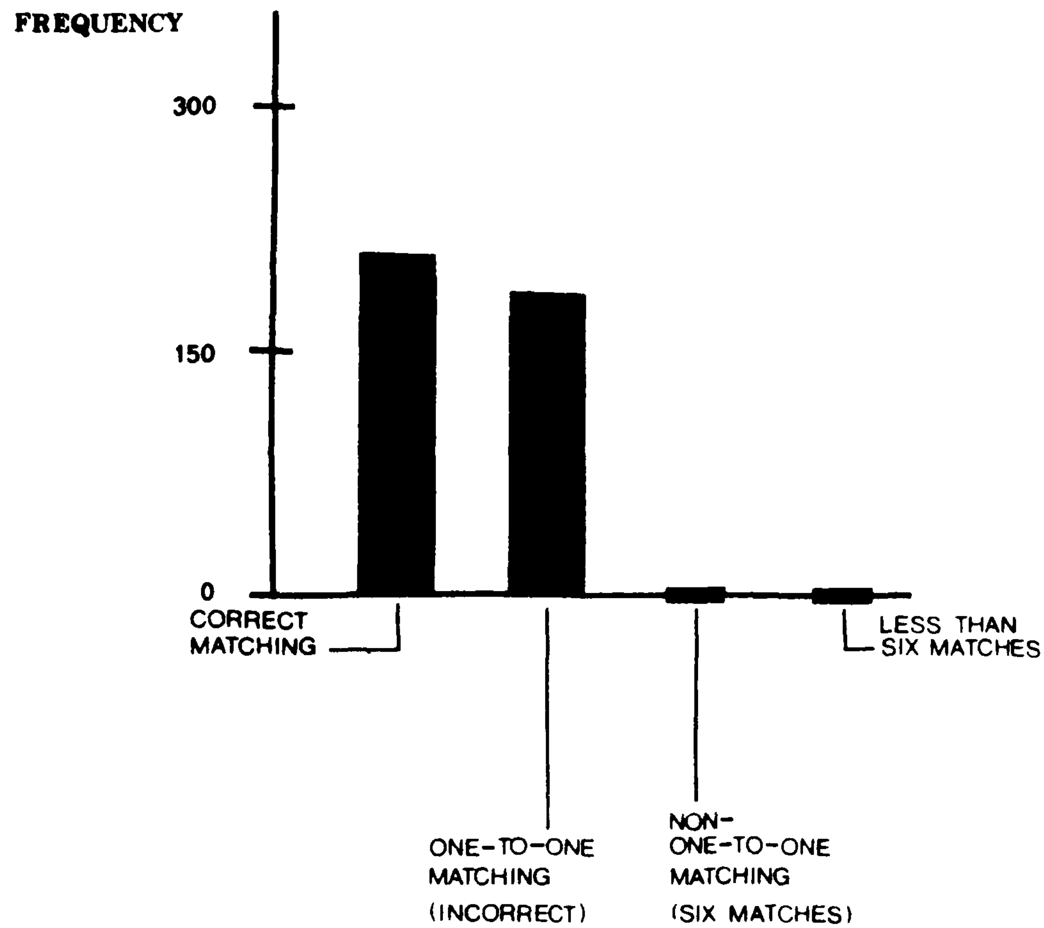


Figure 6. A frequency histogram for correct vs. incorrect matchings. For 450 runs the first and the second frames consisted of two random objects of six features each (the features were randomly placed on a disc of radius 1). The first column corresponds to the cases where a true minimal mapping was found by the network, i.e. the sum of the distances traveled by the features is minimal as verified by an exhaustive search. The second column corresponds to the cases where the minimal mapping was not found by the network, but a one-to-one matching was still made. There was not any case where a one to one match failed to appear (third and fourth columns of the histogram). Thus, the correct solution is not always obtained.

---

distance traveled and the necessity for one-to-one matching. These conflicts often cause a delay in the decision process of the network. In Fig. 8 we illustrate this fact for the paradigm of Fig. 7 d. Similarly to Fig. 2, we show the temporal evolution for the $[V_{ia}]$ array.

Note that at $t = 0.06\tau$, the values of V_{32} and V_{33} begin to rise, mainly driven by the proximity of feature 3 in the first frame to features 2 and 3 in the second frame (see Fig. 7d). Given the imposition of one-to-one matches, this leads to a slow competition between V_{32} and V_{33} ($t = 0.12\tau, 0.24\tau$). In the meantime the values of V_{21}, V_{46}, V_{55} and V_{64} raised and converged to 1 at about $t = 0.24\tau$. From the exhaustive search we found that the optimal solution implied $V_{62} \approx 1$

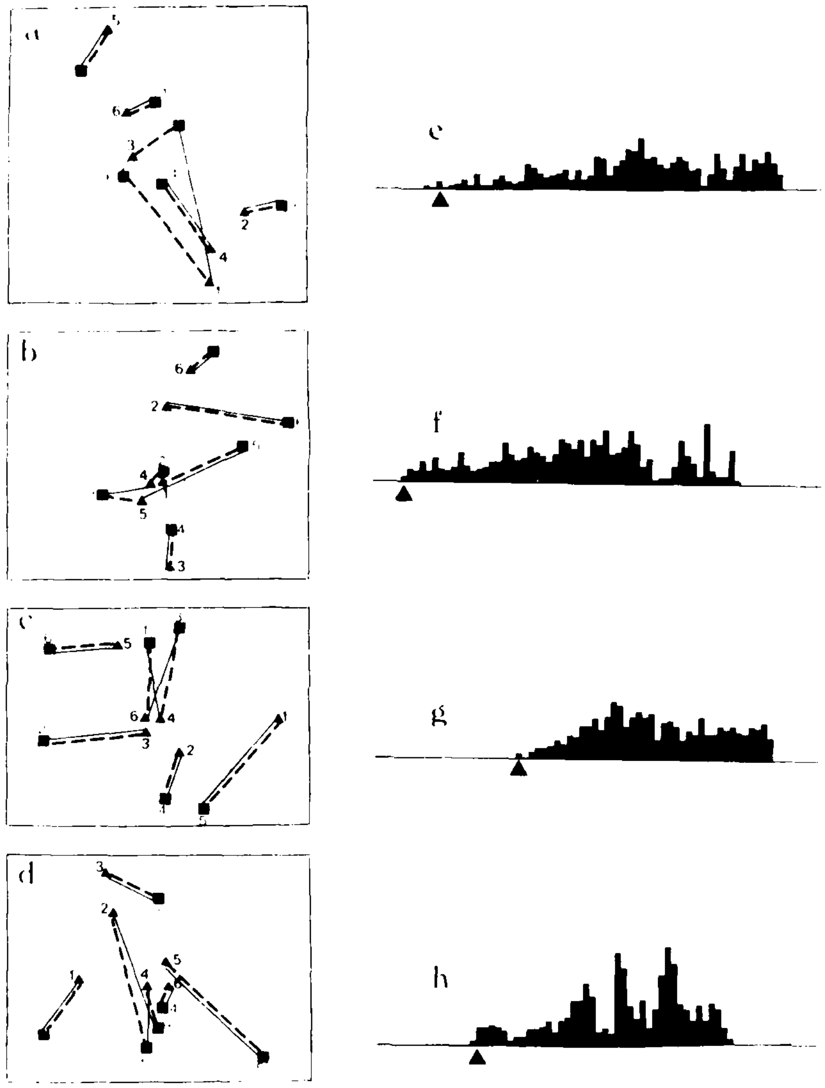


Figure 7. Near optimal computations by the network. a–d. The four cases were taken from the experiment done in Fig. 6, and show examples where minimal mapping was not found by the network. The symbols are similar to those of Fig. 1. The dotted lines represent the correct minimal mapping as found by an exhaustive search. The mistakes made by the network were always the switching of only one pair of correspondences. e–h correspond to a–d respectively. These histograms show the distribution of the total distance traveled by the features for all of the possible cases of one to one mapping. The abscissa has arbitrary scale (but equal in all histograms). The histograms have the same area; $6! = 720$ matching cases. The arrows indicate the total distance traveled for the solution obtained by the network (in figures f and g this value was contained by the left most bin of the histogram). In the cases where errors were made, the solution was nevertheless near optimal.

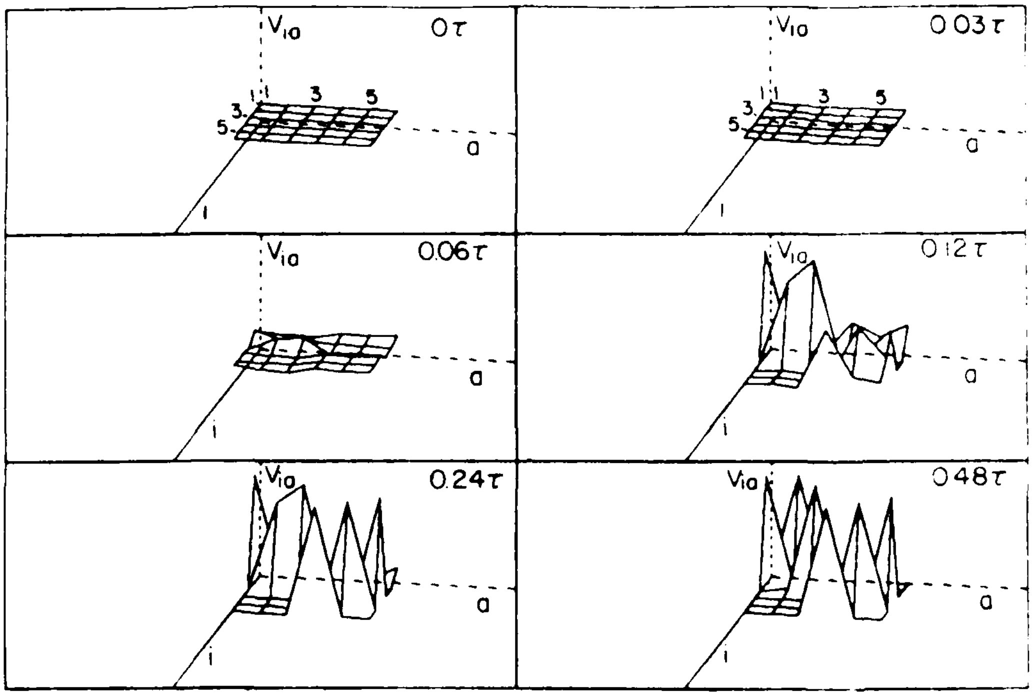


Figure 8. How errors are made by the network. The figure shows the time evolution of the correspondence array for the example shown in Fig. 7 d. For an explanation of the details see Fig. 2. The mistake is made because of the conflict between minimal mapping and one-to-one matching. From a minimal mapping point of view, the matches V_{32} and V_{33} would be preferred. This, however, goes against the one to one matching requirement. While V_{32} and V_{33} compete, other matches, which are not necessarily correct from a minimal mapping point of view, develop.

This was an impossible solution for the network after $t = 0.24\tau$, because $V_{64} \approx 1$. It followed that the network could not reach an optimal solution anymore and had settled to a nearly optimal one, in which $V_{32} \approx 0$ and $V_{12} \approx V_{33} \approx 1$. The long time of convergence was due to the inability of V_{62} to rise due to the imposition of one-to-one matching and to the weak capacity of the network to increase V_{12} because of the large distance between feature 1 in the first frame and feature 2 in the second.

The main reason for building an implementation of the Minimal Mapping Theory in terms of "neural-networks" is to obtain a fast convergence to the solution. This was the case for the examples showed so far, in which the convergence happened in a *fraction of the time constant* of the elementary units of the network. We now bring evidence that this fastness persists even when the number of features in motion increases. In order to demonstrate this we performed an experiment whose results are plotted in the graph of Fig. 9. For each entry in the graph a few runs were performed. Each run consisted of an object of a given number of features (abscissa) randomly placed on a disc of radius 1. The object was identical in the first and second frames to guarantee that a correct solution would be obtained. The average time of convergence and the standard error

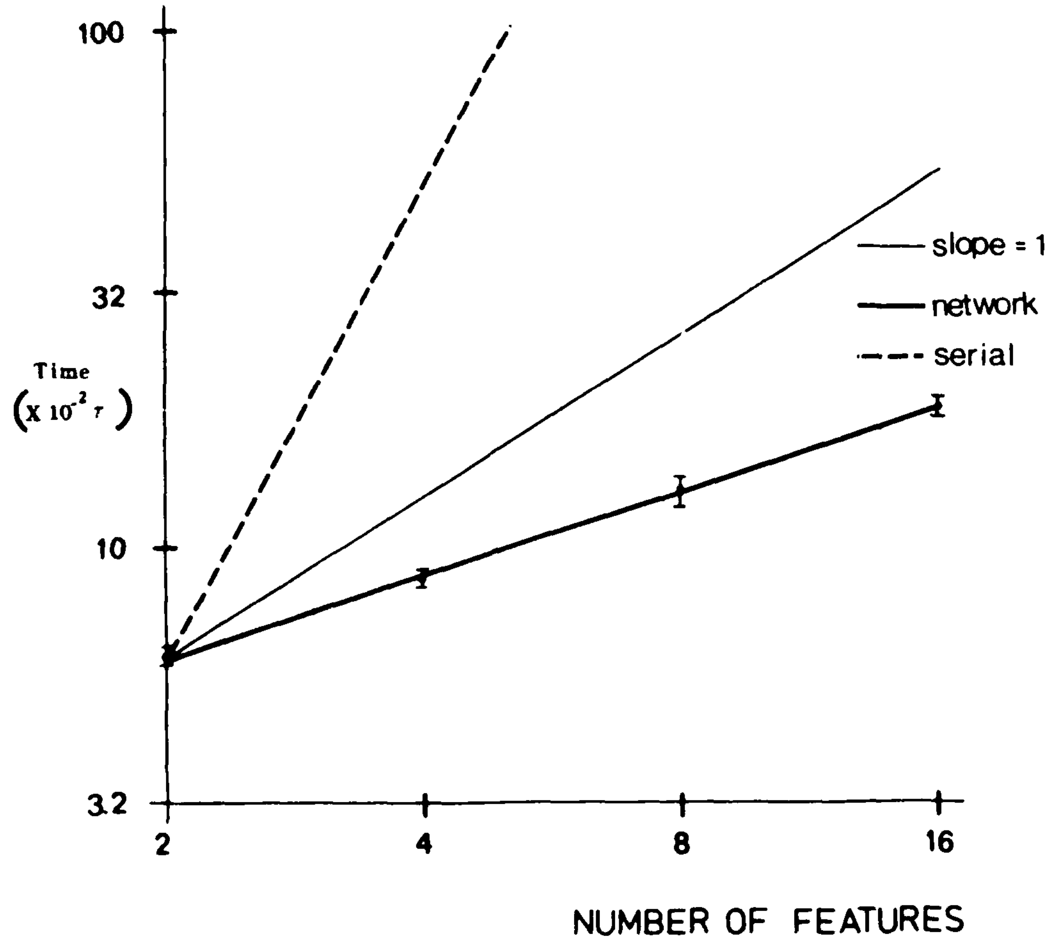


Figure 9. Convergence time of the network vs. serial algorithms. The data points show the average time of convergence (and standard error) of the network as a function of the number of object features. The features were randomly positioned on a disc of radius 1, and the image in the first and the second frame were identical to guarantee a correct solution by the network. The thick line is fit to the data and corresponds to a power law (Eq. 2.12), with a power of about 0.52. The thin line is drawn for comparison and has a slope of 1. The dashed line has the same slope as the theoretically calculated worst-case time of convergence for serial algorithms solving the same problem. Similar slopes were obtained for average times of convergence for related algorithms (Lawer, Lenstra, Rinnooy Kan and Shmoys, 1985.) The network dependence on the number of features is mild and much weaker than serial algorithms.

for these runs were measured (ordinate).

The results are plotted in a log log scale in Fig. 9. The thick solid line shows the results of the experiment. The fact that this curve was a straight line in a log log plot implies that the dependence of the convergence time, T_c, on the number of features, N, was a power law, i.e.

$$T_c = FN^{\gamma}, \tag{2.12}$$

where F and γ are positive constants. For comparison the thin solid line shows a linear dependence (adjusted to be equal to the experiment for the two features case), i.e. $\gamma = 1$. Note that the dependence of the solution obtained by the network is sublinear. In fact its power was about $\gamma = 0.52$. (This means that from the point of view of discrete optimization the network method has a complexity of about $O(n^{1/2})$.) One sees, therefore, that the convergence time of the "neural network" scales weakly (square root) with the number of features in motion.

The strength of this result is emphasized if one considers good serial algorithms to solve the same problem. Mathematically, minimal mapping is a discrete optimization problem known as the linear assignment problem (Burkard, 1979). Some of the best serial algorithms proposed to solve this problem scaled with the third power of the number of features, (Dinic and Kronrad, 1969; Tomizawa, 1971), i.e. $\gamma = 3$. (Once again, this implies that from the point of view of discrete optimization these methods have a complexity of about $O(n^3)$.) The relatively strong dependence of the serial methods are illustrated by the dashed line of Fig. 9. Note the much steeper slope of the serial algorithms, compared to the network implementation. (There are not at the present time, as far as we know, studies of the complexities of other parallel solutions for the correspondence or related problems. Therefore a comparison between our network with other parallel methods was not possible.)

In conclusion we have shown evidence that the convergence time of the "neural network" implementation of the Minimal Mapping Theory scales weakly with the number of features in motion, and therefore, remains short even for cases with a large number of features. This is due to the massive nature of the connectivity of the network, which allows information to travel at high rates from unit to unit in the network.

In the next section we prove theoretical results related to the quality of convergence of the "neural network" implementation of the Minimal Mapping Theory.

3 Theoretical results

Hopfield and Tank (1985) demonstrated good solutions to the Traveling Salesman Problem for up to thirty cities. It seems that for a larger number of cities the solutions become less good (Hopfield, pers. comm.). We have reasons to believe that the network reported in this paper behaves similarly. Our problem, however, is different in an important aspect. The size of the $d_{i\alpha}$'s depend on the time between matched image frames. We prove this theorem: provided that the extent of motion is sufficiently small the network will always obtain the correct match. Therefore, an increase in the number of features to be matched can be compensated for by reducing the time between frames.

In order to show this result we prove that if the diagonal terms of the $[d_{i\alpha}]$ matrix are sufficiently small compared to the off diagonal terms, then one we can choose the parameters of the system such that it will always converge to the correct solution. At the end of the section, we will use this and other results to explain how choices of parameters were made in this work.

We will first show, however, that the strength of matches, $V_{i\alpha}$, are never exactly 0 or 1, but can only approach these values arbitrarily closely. In the proof for this claim we will also provide a derivation of an analytic expression for the equilibrium solutions of the network.

As shown in Section 2, Eq. 2.5 is a Liapunov function for the system. Therefore the solutions of the system are asymptotically stable. It follows that at equilibrium $dU_{ia}/dt = 0$ or from Eq. 2.9:

$$U_{ia} = \tau \left(B \left(N - V \right) - C d_{ia} - A \left(V_i^{COL} + V_a^{ROW} - 2V_{ia} \right) \right), \tag{3.1}$$

which is an analytic expression for the equilibrium solution of the system. The values of V_{ia} are bounded; $0 \leq V_{ia} \leq 1$ (Eq. 2.3). It follows that the right wing of Eq. 3.1 is bounded from below and above. Indeed:

$$\tau \left(B \left(N - N^2 \right) - C d_{ia} - 2NA \right) \leq U_{ia} \leq \tau \left(BN - C d_{ia} \right). \tag{3.2}$$

This proves that at equilibrium, $0 < V_{ia} < 1$, because by Eq. 2.3, $V_{ia} \rightarrow 1$ (0) if and only if $U_{ia} \rightarrow +\infty$ $(-\infty)$.

The values of V_{ia} are different than 0 and 1 not only for equilibrium. Indeed, differentiating Eq.2.4 and substituting in Eq. 2.6 yields:

$$\frac{dV_{ia}}{dt} = -2\lambda V_{ia}(1 - V_{ia}) \frac{\partial E}{\partial V_{ia}}. \tag{3.3}$$

It follows that if at $0 \leq t' < \infty$, $V_{ia} = 1$ (0), then $dV_{ia}/dt = 0$. (One can show that $\partial E/\partial V_{ia}$ is always finite.) Therefore, if at a given instant, $V_{ia} = 1$ (0), then it remains there forever.

Let us now state the main result of this section.

THEOREM: For given A and $N \geq 2$, if $d_{ii} < d_{jb}$, $1 \leq i, j, b \leq N$, $j \neq b$, then for any $1 > \epsilon > 0$, there are $B_0 > 0$ and $C_0 > 0$, such that if $B > B_0$ and $C > C_0$, it follows that at equilibrium $1 - V_{ii} < \epsilon$ and $V_{jb} < \epsilon$.

In the process of proving this theorem we will provide bounds for B_0 and C_0 in terms of A, the data parameters and ϵ. We begin our proof with three short lemmas.

LEMMA 1: At equilibrium $N > V$.

Proof:

Consider the update Eq. 2.9. This can be written as

$$\frac{d}{dt}\left(U_{ia}e^{t/\tau}\right) = e^{t/\tau}\left(-A\left(V_i^{COL} + V_a^{ROW} - 2V_{ia}\right) + B(N - V) - Cd_{ia}\right). \tag{3.4}$$

If $N - V = 0$, then $U_{ia}\exp(t/\tau)$ decreases, because the sum of the terms on the right-hand side of 3.4 is negative. This implies that U_{ia} and consequently V_{ia} and V decrease. The assertion of the lemma then follows from the fact that at $t = 0$, $V = N$ (see initial conditions in Eq. 2.11).

•

LEMMA 2: For given A, if $d_{ii} < d_{jb}$, $1 \leq i, j, b \leq N$, $j \neq b$, then for any $\alpha > 0$, there is $C_0 > 0$, such that if $C > C_0$, it follows that at equilibrium $U_{ii} - U_{jb} > \alpha\tau$.

Proof:

From Eq. 3.1 one obtains that at equilibrium:

$$\begin{aligned}(U_{ii} - U_{jb}) = \\ & - A\left(V_i^{COL} + V_i^{ROW} - V_j^{COL} - V_b^{ROW}\right. \\ & \left. -2V_{ii} + 2V_{jb}\right) + C\left(d_{jb} - d_{ii}\right) \\ \geq & - NA + Cd^*,\end{aligned} \tag{3.5}$$

where $d^* = \min_{i,j\neq b}(d_{jb} - d_{ii})$. This inequality holds because by Lemma 1, $| V_k^{COL} + V_l^{ROW} - 2V_{kl} | \leq N$. Let $C > C_0 = (\alpha + (NA))/d^*$, then:

$$(U_{ii} - U_{jb}) > \alpha\tau. \tag{3.6}$$

•

LEMMA 3: For given A,C and $N \geq 2$, and for any $\epsilon > 0$, there is $B_0 > 0$ such that if $B > B_0$ then at equilibrium $N - V < \epsilon$.

Proof:

From Eq. 3.1 and by Lemma 1, the following inequality can be written at equilibrium:

$$\frac{U_{ia}}{\tau} \geq -AN + B(N - V) - Cd^{**}, \tag{3.7}$$

where $d^{**} = \max_{i,a} d_{ia}$. Let $B > B_0 = (AN + Cd^{**})/\epsilon$. Then $N - V < \epsilon$. This is because, if on the contrary $N - V \geq \epsilon$, it follows:

$$U_{ia} \geq 0. \tag{3.8}$$

From Eq. 2.3 this implies that $V_{ia} \geq 1/2$ or:

$$V \geq \frac{N^2}{2} \geq N, \tag{3.9}$$

which is in contradiction to Lemma 1 and implies $N - V < \epsilon$.

•

We now proceed with the proof of the theorem.

Proof of the Theorem:
Let

$$B > B_0 = \frac{2(AN + Cd^{**})}{\epsilon}, \tag{3.10}$$

and

$$C > C_0 = \frac{\log\left(\left((2N-\epsilon)\left(2\left(N^2-N\right)-\epsilon\right)\right)/\epsilon^2\right) + 2\lambda\tau AN}{2\lambda\tau d^*}. \tag{3.11}$$

We want to prove that $V_{jb} < \epsilon$ and $V_{ii} > 1 - \epsilon$. In the first case we will prove a stronger result, namely $V_{jb} < \epsilon/(2(N^2 - N))$. Suppose on the contrary that $V_{jb} \geq \epsilon/(2(N^2 - N))$. In that case

$$\sum_{j \neq b} V_{jb} \geq \frac{\epsilon}{2}, \tag{3.12}$$

and from Eq. 2.4:

$$U_{jb} \geq \frac{1}{2\lambda} \log \frac{\epsilon}{2(N^2 - N) - \epsilon}. \tag{3.13}$$

From the proof of Lemma 2 and Condition 3.11, one obtains:

$$U_{ii} - U_{jb} > \frac{1}{2\lambda} \log\left(\frac{(2N-\epsilon)\left(2\left(N^2-N\right)-\epsilon\right)}{\epsilon^2}\right). \tag{3.14}$$

Combining Eqs. 3.13 and 3.14 and substituting the result into Eq. 2.3 one obtains:

$$V_{ii} > 1 - \frac{\epsilon}{2N}, \tag{3.15}$$

or

$$\sum_i V_{ii} > N - \frac{\epsilon}{2}. \tag{3.16}$$

However,

$$V = \sum_i V_{ii} + \sum_{j\neq b} V_{jb}. \tag{3.17}$$

Thus, from Eqs. 3.12 and 3.16 one obtains that $V > N$, which is a contradiction to Lemma 1. This implies $V_{jb} < \epsilon/(2(N^2 - N)) < \epsilon$.

Let us now prove that $V_{ii} > 1 - \epsilon$. Because $V_{jb} < \epsilon/(2(N^2 - N))$, we obtain

$$\sum_{j\neq b} V_{jb} < \frac{\epsilon}{2}. \tag{3.18}$$

Also, from Condition 3.10 and the proof of Lemma 3:

$$V > N - \frac{\epsilon}{2}. \tag{3.19}$$

From Eqs. 3.17, 3.18 and 3.19 one obtains:

$$\sum_k V_{kk} > N - \epsilon. \tag{3.20}$$

But $V_{kk} \leq 1$, thus:

$$V_{ii} > N - \epsilon - \sum_{k\neq i} V_{kk} > 1 - \epsilon, \tag{3.21}$$

which is the desired result.

•

We have shown, therefore, that the network is capable of exactly solving the correspondence problem for motions smaller than the internal distances of the object. This is particularly important for non dense objects, i.e. those containing small to medium numbers of features (e.g. Fig. 1). Our computer simulations confirm this result, and indicate that for such objects, a near optimal match is obtained for complex large motions (Fig. 7).

The development of the theorem, and other results, suggest rules of thumb for the choice of the network's parameters. Consider the energy function in Eq. 2.5. For given N, a proportional change of parameters A, B, C and $1/\lambda$ will only scale the shape of E, and thus, will not change the equilibrium solutions of the system. Also, the dynamics of convergence will not be changed, because a modulation of these parameters will cause an inversely proportional change in λ, leaving the equation of motion unmodified. (To understand this claim more easily see the equation of motion in the form expressed in Eq. 3.3.) It follows, contrary to what was concluded by Hopfield (1984), that the absolute value of the parameter λ is irrelevant; only its relative value to the other parameters matters. In all of our simulations and in the rest of this discussion, λ was set to 1.

A few extra rules of thumb can also be derived from our results. Equation 3.10 suggests that the parameter B has to be high compared to AN and Cd^*. The equation gives formulas for how large B should be in terms of the precision required in the problem (ϵ). Equation 3.11 suggests that Cd^* should be relatively high compared to AN for short motions. Simulations showed that AN should be high if the system has to solve ambiguous situations in which multiple matches to a given feature are possible (Fig. 4).

4 The Structural Theory for Apparent Motion

In this work so far, we developed and analyzed a "neural-network" implementation of the Minimal Mapping Theory. The justification for the Minimal Mapping Theory is based on Ullman's argument (1979) that the structure from motion process is divided in two stages; first solving the correspondence problem, then using the correspondence information to recover the 3 D shape of objects. In this section the same mathematical formalism of the preceding sections is used, i.e. that of the "neural networks", to bring some support to Ullman's two-stage hypothesis. We study whether rigidity alone (the basic assumption used to recover the 3 D structure from motion) is sufficient to solve the correspondence problem (and simultaneously the structure from motion problem). We assume rigidity in the form used by Ullman (1984). We call the theory based on rigidity alone the *Structural Theory* for apparent motion. It is shown that further constraints are usually needed to help this theory obtain correct answers.

4.1 A Network Implementation

In this section we do not use the assumption of strict rigidity, but rather Ullman's incremental rigidity scheme, which allows for nonrigid motions (Ullman, 1984; Grzywacz and Hildreth, 1986, 1987; Grzywacz, et al., 1987; Hildreth et al. 1987). In the incremental rigidity scheme an object with N features is described by a model $(x_i(t), y_i(t), z_i(t))$, for $i = 1, \ldots, N$. The x, y components are directly observable (assuming orthographic projection) and the z components are to be deduced. At $t = 0$ the z components are set to zero. Then, at each instant, one uses the previous values of the z's, $z_i(t - \delta t)$ to calculate the new ones, $z_i' = z_i(t)$. This calculation minimizes deviations of the object's rigidity, ΔR, between frames. ΔR may be defined as follows. First define $L_{ij}(t)$ by

$$L_{ij}(t) = (x_i(t) - x_j(t))^2 + (y_i(t) - y_j(t))^2 + (z_i(t) - z_j(t))^2. \tag{4.1}$$

Then define

$$\Delta R = \sum_{i,j}^{N} (L_{ij}(t) - L_{ij}(t - \delta t))^2. \tag{4.2}$$

The Structural Theory proposes to solve simultaneously the correspondence and the structure from motion problems. This is to be done by finding the correspondences, which upon application of the incremental rigidity scheme, yield the minimal ΔR. We now use the set of binary correspondence variables V_{ia} to define a new matching cost function E_R, whose minimization is equivalent to that proposed by the Structural Theory:

$$E_R = \sum_{i,j,a,b}^{N} (L_{ab}(t) - L_{ij}(t - \delta t))^2 V_{ia} V_{jb}, \tag{4.3}$$

To find the correspondence and structure simultaneously by using incremental rigidity, we minimize E_R with respect to z_i' and V_{ia}, requiring that all features in the first frame are matched to exactly one feature in the second. The method is similar to the one described for the Minimal Mapping Theory. It begins by substituting the E_{MM} term of Eq. 2.5 by E_R of Eq. 4.3. It proceeds by updating the U_{ia} variables (see definition in Eq. 2.4) by using simultaneously the equations of motion 2.6 and

$$\frac{dz_i'}{dt} = -\beta \frac{\partial E}{\partial z_i'}, \qquad 1 \leq i \leq N. \tag{4.4}$$

where β is a positive parameter of the problem. As in the case of the Minimal Mapping Theory, E is a Liapunov function of the system. This is because for the Structural Theory Eq. 2.7 can be rewritten as:

$$\frac{dE}{dt} = -\sum_{ia} \frac{\partial V_{ia}}{\partial U_{ia}} (\frac{\partial E}{\partial V_{ia}})^2 - \beta \sum_i (\frac{\partial E}{\partial z_i'})^2, \tag{4.5}$$

which together with Eq. 2.8 proves that $dE/dt \leq 0$. It follows that also for the Structural Theory the system will stop in a point of the solution space in which the function E is at one of its minima.

The next section illustrates the results of our simulations with the equations of motion 2.6 and 4.4, and compares the results to those obtained for the Minimal Mapping Theory. It also discusses a theory which is a hybrid between th Structural and the Minimal Mapping theories, and which seems to give rise to better behaviors than any of the isolated theories.

4.2 Comparison with the Minimal Mapping Theory

Despite extensive experimentation with the parameters, the system based on the Structural Theory rarely converged to the correct answer, unless given a hint of the correct matches. The system made, however, some interesting mistakes. It would sometimes choose matches and depth values for the features, in such a way that the model of the object for the second frame had almost the same 3-D structure as the model for the first frame, but such that the motion between frames was complicated. We illustrate this phenomenon in Fig. 10.

In the example shown in this figure, a three feature object was rotated around an axis perpendicular to the $x - z$ plane by 30°. (It can be shown that in this case, if we use a matching cost function of the form expressed in Eq. 4.3., the y coordinates of the features are irrelevant to the problem.) When observed from a bird's eye view the object looked like a rectangular triangle of sides 3, 4 and 5 (solid straight lines of Fig. 10a). The x coordinates of the three features in the first frame where 0, 0 and 4 for features A, B and C respectively. The z coordinates for the same

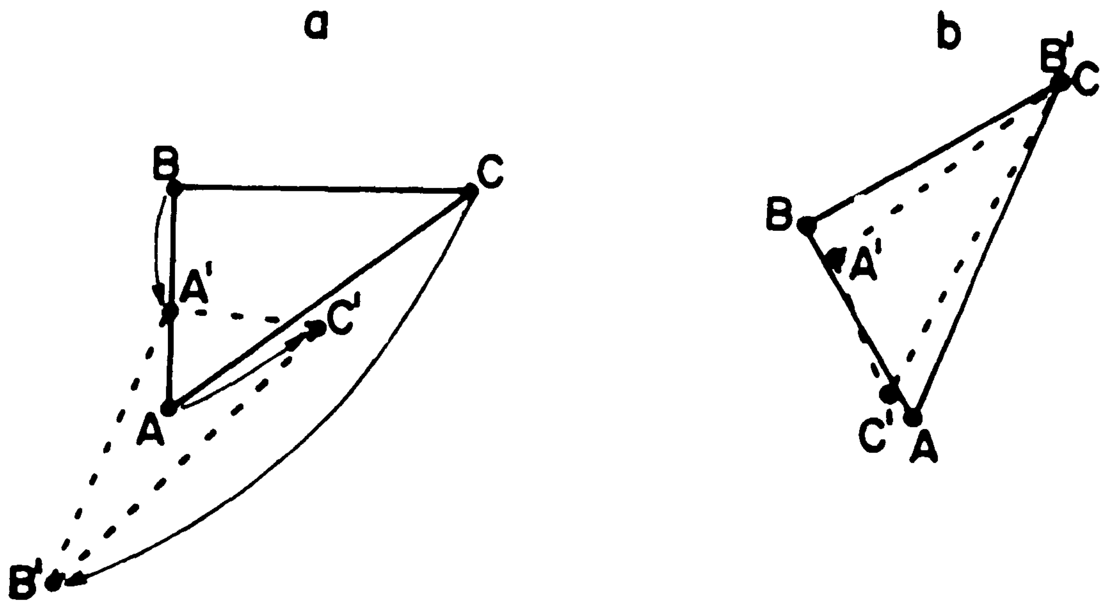


Figure 10. The errors of the Structural Theory. The solid triangles are the bird's eye views of the moving object. a. shows the first frame and b. the second. The dashed triangle in a. is the triangle computed by the network implementation of the Structural Theory. The image coordinates of A', B' and C' are the same as the image coordinates in the second frame of A, B and C, respectively. The curved arrows show the computed correspondences. The computed structure and correspondences were incorrect. However, if the computed structure is superimposed on the true structure, while forcing their corresponding corners to be close, then they are shown to be similar (Fig. b). Thus, such a theory may be able to compute a rough estimate of the structure of an object, without having to solve the correspondence problem.

features were 0, 3 and 3. The rotation was anticlockwise (with feature A fixed), when observed from the bird's eye view. The solid lines of Fig. 10 b show the position of the object in the second frame from this view. The values of x were directly measurable by the observer. We assumed that the observer knew the values of z in the first frame. The values of z for the second frame and the values of the V_{ia}'s were calculated by integrating the equations of motion 2.6 and 4.4. (The parameters used in this display were $A = 5000, B = 10000, C = 10, \tau = 1, \lambda = 50$ and $\beta = 10$. The initial values of the z coordinates in the second frame were close to zero, but randomly chosen. In this example these coordinates were $0.01, -0.01$ and 0.005 for features A, B and C, respectively.)

A bird's eye view of the solution is shown in the dotted lines of Fig. 10a. The curved arrows indicate the motions observed (as shown by the correspondence variables, V_{ia}). Note that these motions were incorrect and very complicated. The 3-D structure of the new triangle, however, was not very different from the original one. The dotted lines of Fig. 10 b represent the dotted triangle of Fig. 10 a, but with the sides rotated and "mirror imaged". These transformations were done in such a way that the matched corners in the two frames were now close in space. Note the similarity of structures between the original and computed triangles. This indicates that although the "neural network" implementation of the Structural Theory is unable to compute the matches correctly, it may be used in some situations to bypass the correspondence problem altogether, and make a fast (but rough) estimation of the parameters of 3-D structure of the object.

The failure of this system to obtain the correct correspondences does not imply that the Structural Theory would fail for any implementation. On the contrary, for most rigid motions, an exhaustive search based on the Structural Theory would give the correct answer. This is because the right correspondences and structure of the object is often the only situation where the energy function is exactly 0. The above failures, however, are to be taken as a serious handicap of the Structural Theory. It shows that the solution space explored by this theory is complex, i.e. it has many local minima. This argument shows that only very elaborate, and therefore, slow methods can find the global minimum. The Minimal Mapping Theory, on the other hand, would only yield the correct matches for translations or relatively short rotations, independently of the implementation. As we have shown, however, for the Minimal Mapping Theory a very fast "neural network" implementation is always possible. The evidence that apparent motion in humans is mainly based on minimal mapping, therefore, seems to point out, that their solution of the motion correspondence problem gives up precision under all circumstances in favor of speed.

We call the attention to the fact that the complexity of the solution space in the Structural Theory is not due to the use of two equations of motion, Eqs. 2.6 and 4.4, instead of only one used by the Minimal Mapping Theory. This complexity is because of the more complicated *dependence* of E_R on the correspondence variables, V_{ia}, than of E_{MM} (compare Eqs. 2.1 and 4.3). In fact, Grzywacz (1986) has demonstrated that problems similar to those illustrated in Fig. 10 still exist in a 2-D version of the Structural Theory. In this version a search for depth values (equation of motion 4.4) is not necessary.

Besides being able to bypass the correspondence problem under some circumstances (Fig. 10b), the Structural Theory may also turn out to be useful in cases for which minimal mapping fails. Such situations may include large rotations and motion of features past occluding boundaries of an object. We found in our simulations that a theory that is a hybrid between the Structural and the Minimal Mapping theories can often handle these situations. Our implementation of this hybrid theory was done by including both the E_{MM} and the E_R terms in the energy function (Eq. 2.5). This hybrid theory proved to be the best of both worlds, being able to compute simultaneously and correctly the correspondences of the features in motion and their depth. We conclude that although the rigidity assumption used by the Structural Theory has serious drawbacks when used alone to solve the correspondence problem, it can significantly help when used in conjunction with the minimal mapping assumption.

5 Discussion

This paper has described methods of implementing theories of motion correspondence using massively parallel networks. Our emphasis has been on networks that are fast and which obtains the correct result most of the time rather than on networks that are infallible but slow. We showed how to design a network implementing Ullman's theory of minimal mapping and demonstrated its effectiveness. We proved some convergence results for this network. Next we questioned whether rigidity alone was sufficient to determine correspondence and tested a theory based on this assump tion. This theory behaved poorly but a hybrid version incorporating some elements of the Minimal Mapping Theory worked well.

An aim of our work was to see if rigidity alone was sufficient to solve the correspondence problem. There are a number of ways that rigidity could be used and it is infeasible to test all of them. Instead we concentrated on a method based on the incremental rigidity scheme (Ullman, 1984), and conjectured that other schemes would give similar results. Our results suggest that rigidity alone is unable to solve the correspondence problem, but there are two reservations. Firstly it is possible that other methods of using rigidity may give better results. Secondly it is possible that the fault lay in the use of our choice of network and that other implementations would succeed. To check this second possibility we designed a scheme based on simulated annealing (Kirkpatrick et al., 1983). Trial runs indicated that the convergence of the Structural Theory did not improve. The energy function seems to have a number of minima of similar depth and so no method, even simulated annealing, will succeed in a reasonable time.

There are some simple psychophysical experiments that could be done to see if rigidity is used for correspondence. Consider a triangle in space lying in a plane along the line of sight of the viewer so that the projections of the three vertices onto the image plane lie in a straight line. As the triangle is rotated the order of vertices in the projection will reverse. In these situations minimal mapping will give the wrong answer. The modified version of the Structural Theory (including minimal mapping terms) will give the correct answer. Informal psychophysics suggests that human perception may be wrong in this case, but the results are not conclusive.

We were able to prove that our minimal mapping network converged to the right answer only if the displacement of the features between frames was smaller than the average distance between features. There are probably few situations for which minimal mapping would give the correct answer if the displacement of features is larger than the average distance between them. It would be interesting to devise examples of these situations and do psychophysics experiments.

Minimal mapping is an elegant theory that gives a good description of a range of physical phenomena. Recently, however, two psychophysical effects have been discovered that the theory cannot account for without modifications. The first is *motion inertia* (Ramachandran and Anstis 1983,1987; Eggleston, 1984; Grzywacz, 1987). This shows that the matching of features between two frames is influenced by their matching in previous frames; features have *inertia* and tend to prefer matches in the directions in which they have been moving. In contrast the *Motion capture* effects can be dramatically illustrated by Ramachandran's moving leopard analogy. If the boundary of the leopard is invisible then the spots on the leopard are matched to their nearest neighbor. If the boundary is visible then it "captures" the spots and their matches are different. Effects like this can be demonstrated by experiments in which dot stimuli are captured by surrounding contours, moving periodic gratings or other dots (Mackay, 1961; Ramachandran and Anstis, 1983,b; Ramachandran and Inada, 1985; Williams, Philip and Sekuler, 1986). These experiments show that minimal mapping has limitations and some modifications are needed.

The main reason for using a massively parallel network is the reduction in computation time. The advantage arises because many problems are parallelizable, and with such a network we can exploit the trade off between the number of elements and the time of computation. Currently, research is being done to construct electronical devices that implement such networks. This massive parallelism may also lead to fault tolerance. Networks are attractive because they offer a method of turning a problem with discrete elements into one with continuous ones, thereby making it possible to solve a decision problem with an analog machine. Another method of turning a discrete problem into a continuous one has been described by Marroquin (Marroquin, 1987).

A further advantage of networks of this type is their possible biological plausibility. This argument, however, must be used cautiously. The network is composed of simple electrical components that could simulate the dynamics of the membrane of simple neurons. Moreover there is similarity between the sigmoid input-output relations of the network elements and the behavior of the synapses of neurons. However there are a number of important differences: real neurons are very complex (von Neumann, 1958; Koch et al., 1982; Crill and Schwindt, 1983; Kuffler et al., 1984) and certainly do not have symmetric synaptic connections. Moreover the brain is not one large homogeneous network and instead has many different levels of organization. The interconnections between neurons are constrained to be local, although well defined fiber tracts exist for long distance communication. Therefore networks of the type we have been considering can only model local regions of the brain.

Our networks make fast decisions, but not always the right ones. It can be argued that sometimes it is more important to obtain fast approximate solutions to problems rather than slow accurate ones. This is curiously similar to the arguments of Simon in decision theory (Simon, 1979). The claim being that a decision maker should, and in practice does, make quick approximate decisions rather than being perfectly rational and finding the best possible decision regardless of the time it takes to compute it.

Acknowledgments

We would like to thank E. Hildreth, M. Drumheller and T. Poggio for helpfull sugestions on the manuscript. C. Correa and L. Ardrey saved this work by making wonderful drawings. This research was supported by a Fairchild Fellowship and by grants from the Sloan Foundation and the Office of Naval Research, Engineering Psychology Division.

References

Andersen, R.A., and R.M. Siegel (1986) Two- and Three-Dimensional Structure from Motion Sensitivity in Monkeys and Humans. Soc. Neurosci. Abstr. 12, Part 2:1183.

Anstis, S.M. (1980) The Perception of Apparent Motion. Phil. Trans. R. Soc. Lond. B, 290:153-168.

Anstis, S.M., and G. Mather (1985) Effects of Luminance and Contrast on the Direction of Ambiguous Apparent Motion. Perception, 14:167-179.

Anstis, S.M., and V.S. Ramachandran (1986) Entrained Path Deflections in Apparent Motion. Vision Res. 10:1731-1739.

Anstis, S.M., and V.S. Ramachandran (1987) Visual Inertia in Apparent Motion. Vision Res. In Press.

Arbib, M.A. (1975) Artificial Intelligence and Brain Theory. Annals of Biomedical Engineering, 3:238-274.

Attneave, F. (1974) Apparent Movement and the What-Where Connection. Psychologia, 17:108-120.

Ballard, D.H., G.E. Hinton, T.J. Sejnowsky (1983) Parallel Visual Computation. Nature, 306:21-26.

Braddick, O.J. (1974) A Short Range Process in Apparent Motion. Vision Res. 14:519-527.

Braddick, O.J. (1980) Low Level Process in Apparent Motion. Phil. Trans. R. Soc. Lond. B. 290:137-151.

Braunstein, M. L. (1962) Depth Perception in Rotation Dot Patterns: Effects of Numerosity and Perspective. J. Exp. Psychol. 6:41-420.

Braunstein, M.L., and G.J. Andersen (1984) A Counterexample to the Rigidity Assumption in the Visual Perception of Structure from Motion. Perception, 13:213-217.

Braunstein, M.L., D.D. Hoffman, L.R. Shapiro, G.J. Andersen, and B.M. Bennett (1986) Minimum Points and Views for the Recovery of Three-Dimensional Structure. Studies in the Cognitive Sciences. 41. Univ. California, Irvine.

Burkard, R.E. (1979) Traveling Salesman and Assignement Problems: A Survey. In Discrete Optimization I, Eds. Hammer, P.L., E.L. Johnson, and B.H.Korte. North-Holand Publishing Company. Amsterdam, 193–215.

Clocksin, W.F. (1980) Perception of Surface Slant and Edge Labels from Optical Flow: A Computational Approach. *Perception* 9:253–269.

Crill, W.E., and P.C. Schwindt (1983) Active Currents in Mammalian Central Neurons. Trends in Neurosci. 6:236–240.

Dev, P. (1975) Perception of Depth Surfaces in Random Dot Stereograms. International Journal of Man–Machine Studies, 7:511–528.

Dinic, E.A., and M.A. Kronrad *(1969) An Algorithm for the Solution* of the Assignement Problem. Soviet Math. Dokl. 10:1324–1326.

Divko, R., and K. Schulten (1986) Stochastic Spin Models for Pattern Recognition. Proceedings of the American Institute of Physics Conference on Neural Networks for Computing. 151:129–134.

Doner, J., J.S. Lappin, and G. Perfetto (1984) Detection of Three Dimensional Structure in Moving Optical Patterns. J. Exp. Psychol.: Human Perc. Perf. 10:1–11.

Eggleston, R.G. (1984) Apparent Motion and Prior Correspondence Effects in Visual Perception. Diss. Abstr. Int. 44:2581–2582.

Feldman, J.A., and D.H. Ballard (1982) Connectionist Models and their Properties. Cognitive Science, 6:205–254.

Finlay, D.J., and P.C. Dodwell (1987) Speed of Apparent Motion and the Wagon Wheel Effect. Percept. & Psychophys. 41:29–34.

Fukushima, K. (1986) A Neural Network Model for Selective Attention in Visual Pattern Recognition. Biol. Cybern. 55:5 15.

Geman, S., and D. Geman (1984) Stochastic Relaxation, Gibbs Distributions, and the Baysian Restoration of Images. IEEE Transaction on Pattern Analysis and Machine Intelligence, 6:721 741.

Gengerelli, J.A. (1948) Apparent Movement in Relation to Homogeneous and Heterogeneous Stimulations of the Cerebral Hemispheres. J. Exp. Psychol. 38:592 599.

Gibson, J.J., and E.J. Gibson (1957) Continuous Perceptive Transformations and the Perception of Rigid Motion. J. Exp. Psychol. 54:129-138.

Green, B.F. (1961) Figure Coherence in the Kinetic Depth Effect. J. Exp. Psychol. 62:272-282.

Green, M. (1983) Inhibition and Facilitation of Apparent Motion by Real Motion. Vision Res. 23:861 865.

Green, M. (1986) What Determines Correspondence Strength in Apparent Motion. Vision Res. 26:599 607.

Green, M., and J.V. Odom (1986) Correspondence Matching in Apparent Motion: Evidence for Three-Dimensional Representation. Science, 233:1427-1429.

Grunau, M.W. von (1986) A Motion Aftereffect for Long Range Stroboscopic Apparent Motion. Percept. & Psychophys. 40:31 38.

Grzywacz N.M. (1986) Properties of Generalized Minimal Mapping Theories for Apparent Motion. Technical Digest of the Optical Society of America Meeting.

Grzywacz N.M. (1987) Interactions Between Minimal Mapping and Inertia in Long Range Apparent Motion. Invest. Ophtalmol. Vis. 28.

Grzywacz, N.M., and E.C. Hildreth (1987). Incremental Rigidity Scheme for Recovering Structure from Motion: Position Based versus Velocity Based Formulations. J. Opt. Soc. Am. 4:503 518.

Grzywacz N.M., E.C. Hildreth, V.K. Inada and E.H. Adelson (1987) The Temporal Integration of 3-D Structure from Motion: a Computational and Psychophysical Study. To appear in "Brain Theory", the proceedings of the second Trieste meeting on Brain Theory.

Grzywacz N.M., and A. Yuille (1986) Motion Correspondence and Analog Networks. Proceedings of the American Institute of Physics Conference on Neural Networks for Computing, 151:200-205.

Hildreth, E.C., V.K. Inada, N.M. Grzywacz and E.H. Adelson (1987) The Effects of Temporal and Spatial Extents on the Recovery of Structure from Motion. MIT Artificial Inteligence Memo, 920.

Hinton, G.E., and T.J. Sejnowski (1983) Optimal Perceptual Inference. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 448-453.

Hoffman, D.D. (1982) Inferring Local Surface Orientation from Motion Fields. J. Opt. Soc. Am. 72:888-892.

Hopfield, J.J. (1982) Neural Networks and Physical Systems with Emergent Collective Computational Abilities. Proc. Natl. Acad. Sci. USA, 79:2554-2558.

Hopfield, J.J. (1984) Neurons with Graded Response Have Collective Computational Properties Like Those of Two-State Neurons. Proc. Natl. Acad. Sci. USA, 81:3088-3092.

Hopfield, J.J.,and D.W. Tank (1985). Neural Computation in Optimization Problems. Biol. Cybern. 52:141-152.

Hutchinson, J.M., and C. Koch (1986) Simple Analog and Hibrid Networks for Surface Interpolation. Proceedings of the American Institute of Physics Conference on Neural Networks for Computing, 151:235-240.

Johansson, G. (1964) Perception of Motion and Changing Form. Scand. J. Psychol. 5:181-208.

Kienker P.K., T.J. Sejnowski, G.E. Hinton, and L.E. Schumacher (1986) Separating Figure from Ground with a Parallel Network. Perception, 15:197-216.

Kirkpatrick, S., C. Gellatt, Jr., and M. P. Vecchi (1983) Optimization by Simulated Annealing. Science, 220:671-680.

Koch, C., T. Poggio, and V. Torre (1982) Retinal Ganglion Cells: A Functional Interpretation of Dendritic Morphology. Philos. Trans. R. Soc. B, 298:227-264.

Kolers, P.A. (1972) Aspects of Motion Perception. Pergamon Press. Oxford.

Korte, A. (1915) Kinematoscopische Untersuchungen. Z Psychol. 72:193-206.

Kuffler, S.W., J.G. Nicholls, and A.R. Martin (1984) From Neuron to Brain. Sinauer Associates. Sunderland, USA.

Lawler, E.L., J.K. Lenstra, A.H.G. Rinnooy Kan, and D.B. Shmoys (1985) The Traveling Salesman Problem. John Wiley & Sons. Chichester, USA.

Little, J., H. Bulthoff, and T. Poggio (1987) Parallel Optical Flow Computation. Proceedings of the SAIC Image Understanding Workshop, Ed. Baumann, L.

Longuet-Higgins, H. C. (1981) A Computer Algorithm for Reconstructing a Scene from Two Projections. Nature, 293:133-135.

Longuet Higgins, H. C., and K. Prazdny (1981) The Interpretation of Moving Retinal Images. Proc. R. Soc. London Ser. B, 208:385-397.

Mackay, D.M. (1961) Visual Effects of Non-Redundant Stimulation. Nature, 192:739-740.

Marr, D. (1982) Vision. W.H. Freeman and Company. New York.

Marr, D., and T. Poggio (1976) Cooperative Computation of Stereo Disparity. Science, 194:283-287

Marroquin, J. (1984) Surface Reconstruction Preserving Discontinuities. MIT Artificial Inteligence Memo. 792.

Marroquin, J. (1987) Deterministic Bayesian Estimation of Markovian Random Fields with Applications to Computational Vision. Proceedings of the First International Conference on Computer Vision, London.

Mather, G., P. Cavanagh, and S.M. Anstis (1985) A Moving Display which Opposes Short Range and Long-Range Signals. Perception, 14:163-166.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953) Equation of State calculations by Fast Computing Machines. J. Phys. Chem. 21:1087-1091.

Mutch, K., I.M. Smith, and A. Yonas (1983) The Effect of Two-Dimensional and Three-Dimensional Distance on Apparent Motion. Perception, 12:305-312.

von Neumman, J. (1958) The Computer and the Brain. Yale University Press. New Haven.

O'Toole, A., and D. Kersten (1986) Adaptive Connectionist Approach to Structure from Stereo. Technical Digest of the Optical Society of America Meeting.

Poggio, T., V. Torre, and C. Koch (1985) Computational Vision and Regularization Theory. Nature, 317:314-319.

Prazdny, K. (1980) Egomotion and Relative Depth Map from Optical Flow. Biol. Cyber. 36:87-102.

Prazdny, K. (1986) What Variables Control (Long-Range) Apparent Motion. Perception, 15:37-40.

Ramachandran, V.S. (1985) Apparent Motion of Subjective Surfaces. Perception, 14:127-134.

Ramachandran, V.S., and S.M. Anstis (1983,a) Extrapolation of Motion Path in Human Visual Perception. Vision Res. 23:83-85.

Ramachandran, V.S., and S.M. Anstis (1983,b) Displacement Thresholds for Coherent Apparent Motion. Vision Res. 23:1719-1724.

Ramachandran, V.S., and S.M. Anstis (1983,c) Perceptual Organization in Moving Patterns. Nature, 304:529-531.

Ramachandran, V.S., and S.M. Anstis (1985) Perceptual Organization in Multistable Apparent Motion. Perception, 14:135-143.

Ramachandran, V.S., and V. Inada (1985) Spatial Phase and Frequency in Motion Capture of Random-Dot Patterns. Spatial Vision, 1:57 67.

Ramachandran, V.S., V. Inada, and G. Kiama (1986) Perception of Illusory Occlusion in Apparent Motion. Vision Res. 26:1741 1749.

Rummelhart, D.E., G.E. Hinton, and R.J. Williams (1986) Learning Representations by Back Propagating Errors. Nature, 323:533 536.

von Schiller, P. (1933) Stroboskopische Alternativversuche. Psych. Forsch. 17:179 211.

Sereno, M.E. (1986) Neural Network Model for the Measurement of Visual Motion. Technical Digest of the Optical Society of America Meeting.

Simon, H. (1979) Rational Decision Making in Business Organization. American Economical Review, 69:493-513.

Tomizawa, N. (1971) On Some Techniques Useful for Solution of Transportation Network Problems. Networks, 1:173 194.

Tsai, R.Y.,and T.S. Huang (1981) Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces. Univ. Illinois Urbana-Champaign, Coordinated Science Laboratory Report, R-921.

Ullman, S. (1979) The Interpretation of Visual Motion. MIT Press, Cambridge.

Ullman, S. (1984) Maximizing Rigidity: the Incremental Recovery of 3 D Structure from Rigid and Rubbery Motion. Percept. 13:255 274.

Verri, A., and T. Poggio (1986) Motion Field and Optical Flow: Differences and Qualitative Properties. MIT Artificial Inteligence Memo, 917.

Wallach, H., and D.N. O'Connell (1953) The Kinetic Depth Effect. J. Exp. Psych. 45:205 217.

Watson, A.B. (1986) Apparent Motion Occurs Only Between Similar Spatial Frequencies. Vision Res. 26:1727 1730.

Waxman, A.M., and S. Ullman (1985) Surface Structure and Three Dimensional Motion from Image Flow Kinematics. J. Robotics Res. 4:72-94.

Wertheimer, M. (1912) Experimentelle Studien Über das Sehen von Bewegung. Z. Psychol. 61:161-278.

White, B.W., and G.E. Mueser (1960) Accuracy in Reconstructing the Arrangement of Elements Generating Kinetic Depth Displays. J. Exp. Psychol. 60:1-11.

Williams, D.W., and R. Sekuler (1984) Coherent Global Motion Percepts from Stochastic Local Motions. Vision Res. 24:55-62.

Williams, D.W., G. Philips, and R. Sekuler (1986) Hysteresis in the Perception of Motion Direction as Evidence for Neural Cooperativity. Nature, 324:253-255.

END

12-87

DTIC