MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo 987. November, 1987

Energy functions for early vision and analog networks.

A. Yuille

Abstract. This paper describes attempts to model the modules of early vision in terms of minimizing energy functions, in particular energy functions allowing discontinuities in the solution. It examines the success of using Hopfield-style analog networks for solving such problems. Finally it discusses the limitations of the energy function approach.

© Massachusetts Institute of Technology, 1987

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 N00014-85-K-0124.

Introduction.

This paper has two main themes. It tries to model early vision processes in terms of minimizing energy functions. Secondly it examines methods of minimizing these functions, in particular analog style networks.

The first section gives some background on the use of energy functions and neural networks in vision. The next few sections describe work done by the author and collaborators on a number of vision problems. The final sections discusses limitations to this approach.

1. Energy Functions and Networks.

It is convenient to divide vision up into two stages. In the first stage the visual scene is analysed, segmented, and properties such as depth, colour and texture are extracted. In the second stage objects are recognized and high level information is used. The output of the first stage is a representation of the scene in terms of depth values, colour and so on. This representation can be called a $2 - 1/2$ D sketch (Marr 1982) or an Intrinsic Image (Barrow and Tennenbaum 1981). It is generally assumed (Marr 1982, Horn 1986, Ullman 1979) that the construction of such a representation does not involve any knowledge of the world (or of the task being performed) more sophisticated than low level assumptions, such as the rigidity of objects. This representation is produced by a number of independent modules, such as stereo, structure

from motion, shape from shading. This paper will confine itself entirely to the modules of early vision.

A number of these modules have been modelled in terms of energy functions. Ullman (1979) described a theory of motion involving solving the correspondence problem between image frames by minimizing a cost function. Ikeuchi and Horn (1981) describe a theory of shape from shading using a variational principle and Ikeuchi (1980) uses a similar technique for shape from texture. Ikeuchi (1980) describes how this technique is able to impose smoothness constraints on the object and draws the analogy with imposing constraints in Artificial Intelligence. These methods can be illustrated by work on optical flow by Horn and Schunk (1981). Let the brightness function be $I(x_1, x_2, t)$. Then, assuming that points with the same image intensity over time correspond, the velocity field (v_1, v_2) obeys

$$\frac{\partial I}{\partial x_i} v_i + \frac{\partial I}{\partial t} = 0 \tag{1.1}$$

where we use the summation convention over repeated indices (for example $a_i b_i = a_1 b_1 + a_2 b_2$). Now (1.1) is a single equation for the two unknowns (v_1, v_2) and does not specify them uniquely. To obtain a unique solution Horn and Schunk assume continuity of the velocity field and minimize a function

$$E = \int \frac{\partial \underline{v}}{\partial x_i} \frac{\partial \underline{v}}{\partial x_i} + \lambda \int (v_i \frac{\partial I}{\partial x_i} + \frac{\partial I}{\partial t})^2. \qquad (1.2)$$

Here the first term corresponds to requiring smoothness of the velocity field and the second to enforcing (1.1) . They define an iterative algorithm to minimize (1.2) and obtain good results.

Many other vision problems have been treated in a similar way and we mention a few examples. Hildreth (1984) used a similar method to solve the aperture problem for motion based on zero crosssing contours. Grimson (1981) uses a similar approach to interpolate a surface through sparse stereo data. Terzopoulos (1984) extended Grimson's work using more sophisticated techniques. Poggio and Torre (1984) descovered the similarity of these methods to a branch of mathematics called regularization theory (Tikhonov 1977) and proposed a unified framework. These methods all had an important property that was both a weakness and a strength; they usually imposed continuous solutions and smoothed over discontinuities. Regularization theory (Poggio and Torre 1984) indeed required that the solution to a problem depended smoothly on the data. Inserting a discontinuity in the solution would require a yes/no decision, and hence could not depend continuously on the data. *

* Terzopoulos (1984) suggested the surface could be interpolated smoothly and then the boundaries found by an edge detection operation measuring the "tension" in the smoothed surface.

This inability to deal with discontinuities however had important practical advantages, the energy functions tended to be convex and not have local minima. Thus they could be minimized by simple methods such as gradient descent. More sophisticated techniques could be used to speed up the convergence. For example, Terzopoulos (1984) adapted a multi-layer algorithm due to Brandt (1977).

To deal with discontinuities a new approach was needed. Geman and Geman (1984) did work on image segmentation using line processors. These are illustrated in figure 1. There are two lattices, the standard space lattice and an additional line processor lattice. The line processor elements are either on or off. When a line processor is on it breaks the constraints between the adjacent space pixels. Similar work was reported by Blake (1983) who used the idea of weak constraints * , that is to say constraints which must be satisfied almost everywhere but which can be broken at a cost. The binary nature of the line processors means that discontinuities can be dealt with. However the energy functions are no longer convex and new strategies are needed to minimize them. Various methods have been tried. Geman and Geman (1984) use simulated annealing while Blake (1983) uses a method which systematically approximates the energy function by a convex one, *graduated non-convexity*..

---

* Based on work by Hinton (1979).

Marroquin, (1985) Marroquin et al (1987) interpret the energy functions in terms of probability theory using the Clifford Hamersley theorem, a connection described in Geman and Geman (1984). Instead of minimizing the energy function they use probabilistic algorithms to estimate the maximum a posteriori Bayesian estimate of the solution. Apart from simulated annealing, which takes a long time, none of these methods are guaranteed to converge to the correct result.

In this paper we will describe an alternative approach to minimizing these energy functions based on analog networks of the Hopfield type. Mathematically this gives a method of smoothing the energy function reducing the number of local minima. It is also implemented by a network which could be built in V.L.S.I. and which could possibly be implemented by real neurons. The V.L.S.I. network would be massively parallel and could minimize the energy function orders of magnitude faster than serial, or parallel computers. Hopfield networks were originally designed to be an associative parallel memory (1982, 1984). In Appendix (1) we give a simple introduction to Hopfield networks and then show how their formalism can be extended to allow some generalizations. Although the Hopfield networks are nonlinear we can usually write analytic closed forms for the solutions they converge to. There is empirical evidence that they often converge to the correct result. Moreover

we can prove mathematically that they will always converge to a solution of the mean field theory equations and thus represent a deterministic method to approach the probabilistic solutions (see also Marroquin 1987). There are also similarities with the graduated non-convexity approach of Blake (1983).

Many vision algorithms were designed to be implemented on neuronally plausible networks. For example Horn's work on colour (1973) was partially intended as a possible model of the human colour system. Cooperative stereo algorithms by Arbib and Dev (1975) and Marr and Poggio (1977) were also implemented by simple neural-like elements. * Terzopoulos (1984) suggested the use of analog networks for surface interpolation, but did not implement them. Poggio et al (1985) developed analog networks for quadratic regularization energy functions. They note that, supposing we are considering motion smoothing, minimizing the energy function $E(v_i, \partial v_i \partial x_j)$ given by (1.2) is equivalent to solving the associated Euler Lagrange equations (Courant and Hilbert 1953)

---

* Interestingly the Marr and Poggio network can be considered as of discrete Hopfield network. Discrete Hopfield networks, however, are less effective than continuous networks for minimizing functions since the continuous networks smooth the energy function removing local minima.

$$\frac{\partial E}{\partial(\partial v_i/\partial x_j)} = \frac{\partial}{\partial x_j}\frac{\partial E}{\partial v_i}. \tag{1.3}$$

Since $E(v_i, \partial v_i \partial x_j)$ is quadratic in $v_i, \partial v_i \partial x_j$ the equations (1.3) are linear in v_i and its derivatives. Any system of linear equations can be modelled by an analog network involving resistances, capacitors and inductances (Kaplus 1958) * and hence (1.3) can be solved for such a network.

Hopfield style networks could also be adapted to minimize (1.2) and it is interesting to consider the differences with the networks described above. Hopfield networks have a dynamical update rule, which for this problem corresponds directly to a continuous form of steepest descent,

$$\frac{dv_i}{dt} = -\frac{\partial E}{\partial v_i}. \tag{1.4}$$

It follows from the chain rule of differentiation that the energy function $E(t)$ will continuously decrease with time

$$\frac{dE}{dt} = -\frac{\partial E}{\partial v_i}\frac{\partial E}{\partial v_i} \leq 0. \tag{1.5}$$

E is bounded below, by 0, and so with this dynamics the system has to converge to a minimum of E. Thus E is a *Lyaponov function.*

The analog networks of Poggio et al (1985), and the networks of Hopfield,

---

*Special tricks are needed to get negative resistors

can all be constructed out of simple electronic elements (resistors, operational amplifiers, etc). They are also compatible with the existing knowledge of the electrical behaviour of the dendrites and axons of neurons. Thus neural hardware would be capable of implementing such networks, although the evidence suggests that neurons are considerably more complicated.

2. Surface Interpolation

Surface interpolation is a good example for illustrating the difference between energy functions requiring smoothness and those allowing discontinuities. Following the work of Geman and Geman for image restoration, an energy function for surface interpolation can be written (Marroquin 1985), *

$$E(f,l) = \sum_i (f_i - f_{i+1})^2(1 - l_i) + C_d \sum_i (f_i - d_i)^2 + C_l \sum_i l_i. \qquad (2.1)$$

Here the d_i correspond to the depth data, the f_i to the desired answer and the l_i to the line process elements. C_d and C_l are constants. The line process

* For simplicity of the mathematics we write the energy function in one dimension. All the results described also hold for two-dimensional surfaces.

lattice is interposed with the space lattice, see figure 1. The l_i can take values 0 or 1. If the l_i are all set equal to zero then E reduces to standard surface interpolation with a membrane * . This formulation allows the smoothness constraint, enforced by the first term of (2.1), to be broken at a cost. If the surface gradient, $f_i - f_{i+1}$, becomes too large the line process term l_i can switch on at a cost of C_l in the third term. See figure 1.

This energy function can be generalized to two dimensions in a straightforward manner (Marroquin 1985, Blake 1983, Koch, Marroquin and Yuille 1986). Additional terms can be added to the energy function to impose continuity of lines. The interpolation can be generalized from membranes to thin plates by including terms with second order derivatives. This will require introducing *surface gradient processors* corresponding to discontinuities in the surface gradient.

There has been much work on surface interpolation and several people have used energy functions of form (2.1). Different strategies were used to minimize (2.1), Blake uses a technique called graduated non-convexity in which a non-convex energy function is gradually transforned into a convex one. This method is not guaranteed to find the global minima of the original energy function and may only work for dense data, nethertheless it gives good

* Membranes and thin plates correspond to minimizing the first and second derivatives of a surface respectively.

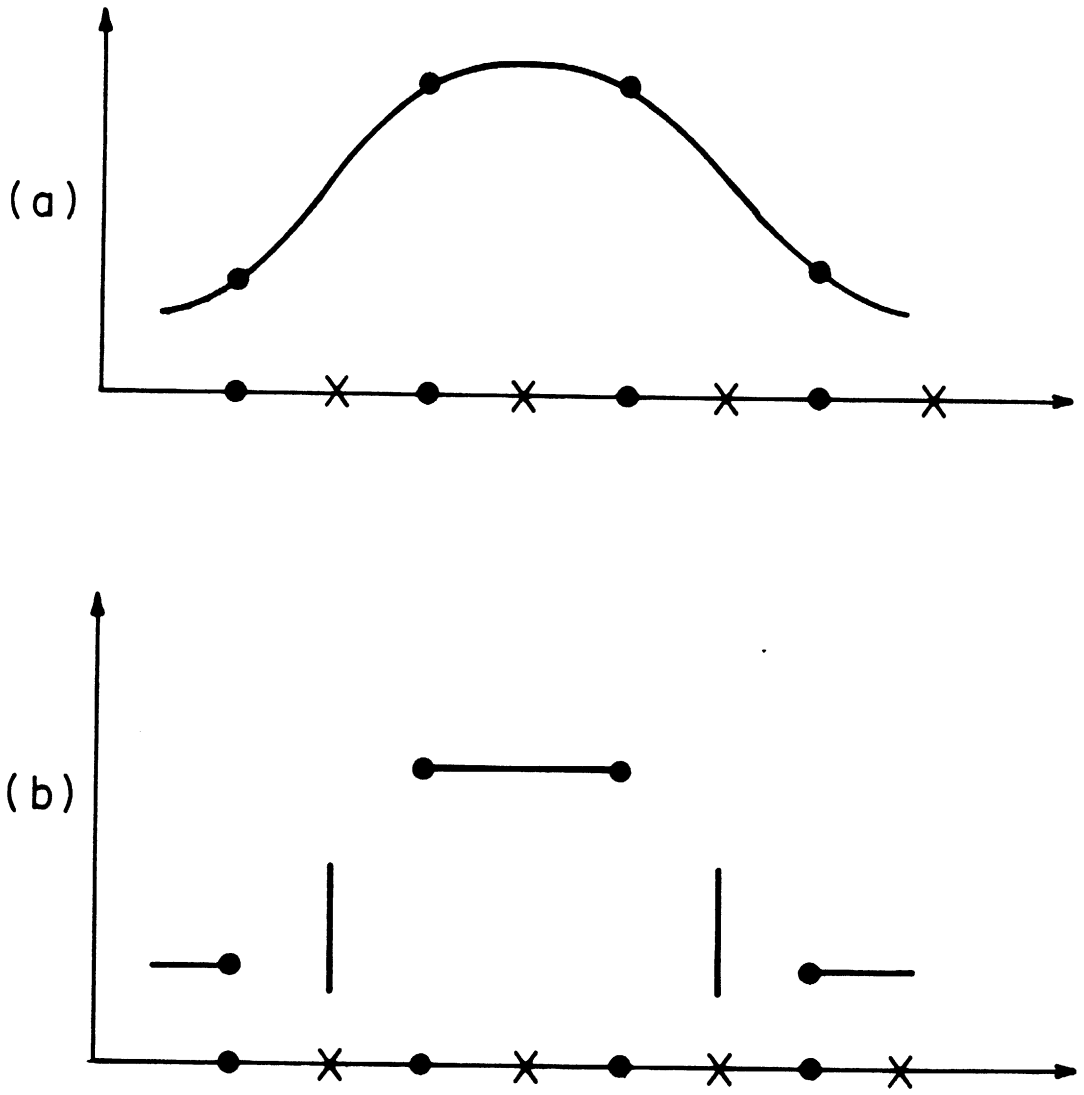

Figure 1 *The line processors. In (a) there are no line processes and the interpolation process assumes that the points lie on a single surface. In (b) the line processes switch on and break the surface where the gradient is large.*

empirical results. Marroquin used simulated annealing and a number of other statistical and deterministic algorithms. Again these were not guaranteed to always succeed but give good results.

Since there are no guaranteed methods to find the minima of (2.1) we would like a technique that may not always suceed but which is very fast. In a series of papers Hopfield (1982, 1984, 1985) describes networks made up of simple analog devices. These networks have two important properties. Firstly they could be implemented in V.L.S.I. and perform parallel computations at speeds orders of magnitude faster than existing parallel (or serial) computers. Secondly the networks may be biologically plausible.

From our viewpoint these networks have a third advantage; they are able to find good estimates of the global minima of non-convex energy functions. Hopfield and Tank (1985) demonstrated that they could find close solutions to the Travelling salesman problem (to within a few percent of the optimum length) for up to thirty cities. The work described below was first reported in Koch, Marroquin, Yuille (1986). We now show how to design a network to minimize (2.1).

The key point of the Hopfield approach is to replace the binary variables l_i by continuous variables V_i lying in the region $[0, 1]$. Each V_i is related to an internal variable U_i, which is unbounded, by a gain function g, $V_i = g(U_i)$. A typical choice for g is

$$g(u_i) = \frac{1}{1 + e^{-2\lambda u_i}}. \tag{2.2}$$

This is a sigmoid function monotonically increasing from 0 as $u_i \mapsto -\infty$ to 1 as $u_i \mapsto \infty$. The parameter λ controls its "sharpness". As $\lambda \mapsto \infty$ it becomes a Heaviside (step-edge) function. The energy function can now be written

$$E(f,V) = \sum_i (f_i - f_{i+1})^2(1-V_i) + C_d \sum_i (f_i - d_i)^2 + C_l \sum V_i + C_g \sum \int_{V=0}^{V=V_i} g^{-1}(V)dV. \tag{2.3}$$

where the last term is a gain function term. The dynamical equations are

$$\frac{dU_i}{dt} = -\frac{\partial E}{\partial V_i} \tag{2.4a}$$

$$\frac{df_i}{dt} = -\frac{\partial E}{\partial f_i}. \tag{2.4b}$$

Observe that since U_i is related to V_i by the gain function (2.2) the V_i^s will always lie in $[0,1]$. This type of dynamics is gradient descent for the f_i. If we substitute for U_i in (2.4a) it becomes

$$\frac{dV_i}{dt} = -2\lambda V_i(1-V_i)\frac{\partial E}{\partial V_i} \tag{2.5}$$

and so is a form of gradient descent for the V_i with a weight factor. It can easily be checked that E is a Lyaponov function for the system

$$\frac{dE}{dt} = -\sum_i \left(\frac{\partial E}{\partial f_i}\right)^2 - \sum \left(\frac{\partial E}{\partial V_i}\right)^2 \frac{dV_i}{dU_i}. \tag{2.6}$$

Since V_i is a monotonic function of U_i the right hand side of (2.6) is always negative. Thus E always decreases and is bounded below, so the system coverges to a minima.

Koch, Marroquin and Yuille (1986) describe the implementation details. The system was simulated on a Symbolics 3600 LISP machine. The system was tested first in 1-D and then in 2-D. Some experimentation was needed to find suitable values for the parameters C_d, C_l, C_g. The system performed well even with noisy data and with sparse sampling (sometimes as low as five percent of the points were sampled).

The λ parameter controls the degree of smoothing of the gain function. For high λ the V_i are essentially forced to be either 0 or 1 and the continuous system (2.3) is close to the discrete system (2.1). For small λ the energy function becomes convex (this can be verified by calculating the Hessian of (2.2)). Thus λ corresponds to the degree of smoothing of the problem. Altering the value of λ to change the degree of convexity of the energy function is analogous to graduated non-convexity (Blake 1983). For our simulations the only difference in running the networks with small or large λ was the convergence time. The smaller λ, the longer it took to converge, without any significant effect

on the final solution. The convergence time of the networks was usually a few time constants (using the adiabatic expansion it can be shown that networks obeying equations (2.4) reach a final state after a few time constants).

The fact that the network converges even with high λ suggests a hybrid strategy to minimize (2.1). The f_i are continuous and are updated by gradient descent while the V_i are discrete. The V_i are sampled at random and changed if this reduces the energy. Some experimental success is reported with this (Koch, Marroquin and Yuille 1986). For further work see Hutchinson and Koch (1986), Marroquin (1985), Marroquin et al (1987). Clearly such a strategy will only work if the energy function (2.1) has few local minima.

We now perform a new analysis of the network and prove results showing that it converges to a solution of the mean field theory equations. If we use a probabilistic algorithm, like the Metropolis algorithm, in the final state of the system each line process l_i will be on with a certain probability $p_i(T)$, where T is the temperature. The network will converge to a state where the line process elements take deterministic values $p_i(T)$ where T is inversely proportional to λ. It is in this sense that the network is a solution of the mean field theory equations. Note that, because of the coupling with the depth field f_i, there will be several solutions to the mean field theory equations and we cannot guaurantee that we will find the one with least energy. Another

deterministic method, based on probabilistic considerations, of finding solutions of the mean field theory equations for this problem has been proposed by Marroquin (1987).

Using (2.6) and substituting for V we see that the energy function E decreases at the rate

$$\frac{dE}{dt} = -\sum_i \left(\frac{\partial E}{\partial f_i}\right)^2 - \sum \left(\frac{\partial E}{\partial U_i}\right)^2 \frac{dU_i}{dV_i}. \tag{2.7}$$

The energy is bounded below so the system converges to a state with $dE/dt = 0$. From (2.7) we see that this implies (note that dU_i/dV_i is always positive and non-zero)

$$\frac{\partial E}{\partial f_i} = 0, \tag{2.8a}$$

$$\frac{\partial E}{\partial U_i} = 0. \tag{2.8b}$$

Note that

$$\frac{\partial V_i}{\partial U_i} = \frac{2\lambda}{(e^{\lambda U_i} + e^{-\lambda U_i})^2}. \tag{2.9}$$

We calculate

$$\frac{\partial E}{\partial U_i} = \frac{2\lambda}{(e^{\lambda U_i} + e^{\lambda U_i})^2}\left(-(f_{i+1} - f_i)^2 + C_l + 2\lambda C_g U_i\right). \tag{2.10}$$

This function has zeros at $C_g U_i = (f_{i+1} - f_i)^2 - C_l$ and at $\pm\infty$. Calculating the second derivatives of E with respect to U_i we see that the zeros at $\pm$ inf are maxima and the zero at $C_g U_i = (f_{i+1} - f_i)^2 - C_l$ is a minimum. Thus although E is not convex with respect to U_i it has a unique minimum. Note that none of the V_i will be exactly 0 or 1. The true energy minima will therefore have

$$C_g U_i = (f_{i+1} - f_i)^2 - C_l. \tag{2.11}$$

Recall that, for large λ, the sign of U_i determines whether $V_i = 0$ or 1. Thus a discontinuity will be imposed only if

$$(f_{i+1} - f_i)^2 \geq C_l. \tag{2.12}$$

.

The depth terms f_i will obey

$$(f_i - f_{i+1})(1 - V_i) + (f_i - f_{i-1})(1 - V_{i-1}) + (f_i - d_i) = 0. \tag{2.13}$$

We rewrite (2.11) in terms of l_i as

$$l_i = \frac{1}{1 + exp{-2\lambda\big((f_{i+1} - f_i)^2 - C_l\big)/C_g}}. \tag{2.14}$$

We have shown that if the interpolated surface gradient exceeds a threshold then a discontinuity will be inserted. It is not clear however that the discontinuity will be inserted at exactly the correct place. Observe, however, that in the absence of depth data there is no clear criterion for where the edge should be.

A stochastic algorithm (for example Marroquin 1985) would use gradient descent for the f_i, as in (2.4a), and a probabilistic update rule for the line processors. The line processors are examined seperately and updated as follows: (i) Calculate the change in energy $\Delta E(l_i)$ resulting from changing the state of the line process l_i. (ii) If $\Delta E(l_i) \leq 0$ make the change. (iii) If $\Delta E(l_i) \geq 0$ make the change with probability $1/(1 + exp{-2\Delta E/T})$.

This system will converge to a state of thermal equilibrium with

$$p(l_i = 0) = Aexp(-(f_{i+1} - f_i)^2/T), \tag{2.15a}$$

$$p(l_i = 1) = Aexp(-C_l/T), \tag{2.15b}$$

where A is a normalization factor to ensure $p(l_i = 0) + p(l_i = 1) = 1$. Thus

$$p(l_i = 1) = \frac{1}{1 + exp{-\big((f_{i+1} - f_i)^2 - c_l\big)/T}}. \tag{2.16}$$

Comparing (2.16) with (2.14) we see that the network indeed finds a solution satisfying the mean field theory equations.

3. Motion Smoothing and Segmentation

We now briefly describe some work on motion smoothing and segmentation. This work was done in collaboration with C. Koch. Mathematically it is very similar to the work on surface interpolation.

The problem we are tackling involves segmenting an image using motion flow (with correspondence based on image intensity). For example, detecting an object moving over a textured background.

A method for obtaining the motion field was described in section 1. This assumes continuity of the motion field and therefore would break down at the boundaries of the object. It is straightforward, however, to modify the energy function to include line processes. The line processes should switch on at the boundaries of the object thereby both segmenting the scene and preventing the velocity field being distorted by smoothing over the boundaries. Inside these boundaries the energy function should give the velocity field as before.

The energy function can be written as

$$E(v^x_{i,j}, v^y_{i,j}, h_{i,j}, v_{i,j}) = \sum_{i,j} \left(v^x_{i,j}(I_{i+1,j} - I_{i,j}) + v^y_{i,j}(I_{i,j+1} - I_{i,j})\right)^2$$

$$+ \sum_{i,j} \left((v^x_{i+1,j} - v^x_{i,j})^2 + (v^y_{i+1,j} - v^y_{i,j})^2\right)(1 - h_{i,j}) + \left((v^x_{i,j+1} - v^x_{i,j})^2\right.$$

$$\left.+(v^y_{i,j+1} - v^y_{i,j})^2\right)(1 - v_{i,j}) + \sum_{i,j} V(h, v). \tag{3.1}$$

The horizontal and vertical line processors are given by $h_{i,j}$ and $v_{i,j}$ respectively. The $V(h, v)$ term corresponds to the cost for the vertical and horizontal line processes including the terms enforcing local continuity.

This energy function is very similar to that for surface interpolation. Once again the line processes produce local minima in the energy function and no algorithms are guaranteed to converge.

It is straightforward to design a Hopfield network to minimize (3.1). The $h_{i,j}$ and $v_{i,j}$ become continuous variables related to new variables $p_{i,j}$ and $q_{i,j}$ by $h_{i,j} = g(p_{i,j})$ and $v_{i,j} = g(q_{i,j})$ respectively. Gain function terms for h and v are added to the energy. The dynamics are defined as in the previous section

$$\frac{dv^x}{dt} = -\frac{\partial E}{\partial v^x} \tag{3.2a}$$

$$\frac{dv^y}{dt} = -\frac{\partial E}{\partial v^y} \tag{3.2b}$$

$$\frac{dp_{i,j}}{dt} = -\frac{\partial E}{\partial h_{i,j}} \tag{3.2c}$$

$$\frac{dq_{i,j}}{dt} = -\frac{\partial E}{\partial v_{i,j}}. \tag{3.2d}$$

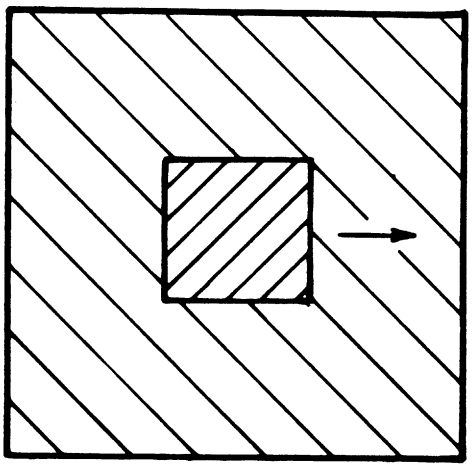

Figure 2 *An example of an object moving relative to a fixed background*

This network is still being tested but preliminary results are encouraging. It is able to segment a textured object moving over a textured background field. It can be mathematically analysed, as for surface interpolation, and similar results hold.

4 Motion correspondence

We now consider a rather different type of vision problem namely motion correspondence and the recovery of structure. This work is described in detail in Grzywacz and Yuille (1986).

Ullman (1979) proposed dividing the structure from motion process up into two different stages. The first consists of matching tokens, such as points or straight lines, between different image frames; solving the so called correspondence problem. Once this matching is done the second stage assumes rigidity to recover the structure of the object. Ullman1984 later suggested an alternative method of finding the structure (see also Grzywacz and Hildreth 1987) capable of dealing with non-rigid motion. He again used psychophysics to argue that the three dimensional structure of an object was not perceived immeadiately but developed gradually over time. He proposed that the visual system constructed an internal model of the object, initially flat, which was updated over time by assuming the minimal change of rigidity between sucessive image frames, the so called *incremental rigidity scheme.*

It is natural to ask whether errors are caused by dividing the process into two stages. Both are solved using different assumptions and it is possible that these conflict for some stimuli. It is also interesting to see if rigidity alone is sufficient to solve the correspondence problem. To investigate this

we define a cost function that minimizes incremental rigidity and solves the correspondence problem simultaneously.

The incremental rigidity scheme maintains a model of the object at any given time $M(t) = \big(x_i(t), y_i(t), z_i(t)\big)$, $i = 1, ..., N$. At the next time frame, at $t + \delta t$, the model is updated taking into account the new information available (we are assuming orthographic projection onto the x, y plane). The model is updated to $M(t + \delta t) = \big((x_i(t + \delta t), y_i(t + \delta t), z_i(t + \delta t))\big)$, $i = 1, ..., N$ where $z_i(t+\delta t)$ is determined by minimizing the change in rigidity between $M(t)$ and $M(t + \delta t)$. The object is initially assumed flat, i.e. $M(0) = \big((x_i(0), y_i(0), 0)\big)$ $i = 1, ..., N$.

We first investigate using rigidity to solve the correspondence problem and determine the structure simultaneously. A measure of rigidity is

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

The $z_i(t + \delta t)$ are defined to minimize the change in rigidity ΔR between frames

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

We now define a set of binary *correspondence* variables $\big(V_{ia}\big)$. If point i in the first frame goes to point a in the second frame then $V_{ia} = 1$, otherwise

$V_{ia} = 0$. We can define a matching cost E_R by

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

To find the correspondence and structure we minimize E_R with respect to $(z_a(t + \delta t))$ and (V_{ia}) requiring that all points in the first frame are matched to exactly one point in the second.

We use a method developed by Hopfield and Tank (1985) for the Travelling Salesman Problem. Again we first define a new array of variables, (U_{ia}). These, are internal variables of the new problem and have a monotonically increasing relationship to V_{ia}:

$$V_{ia} = \frac{1}{1 + e^{-2\lambda U_{ia}}}, \tag{4.4}$$

where λ is again a parameter of the problem. We next define the full energy function to be:

$$\begin{aligned} E = \frac{A}{2} \sum_{a=1}^{N} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} V_{ia} V_{ja} + \frac{B}{2} \sum_{i=1}^{N} \sum_{a=1}^{N} \sum_{\substack{b=1 \\ b \neq a}}^{N} V_{ia} V_{ib} \\ + \frac{C}{2} ((\sum_{i=1}^{N} \sum_{a=1}^{N} V_{ia} - N))^2 + \frac{D}{2} E_R \\ + \frac{F}{2\lambda} \sum_{i=1}^{N} \sum_{a=1}^{N} ((V_{ia} \log(V_{ia}) + ((1 - V_{ia})) \log((1 - V_{ia})))), \end{aligned} \tag{4.5}$$

where A, B, C, D, F are positive parameters of the problem. (We will informally identify each of the terms of the right hand side of Eq. 4.5 by the

parameter leading it.). Minimization 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 B term). Minimization of the C term, forces the total amount of correspondences to be N. Thus the terms A and C will force N correspondences of strength 1 to be established in such a way that each feature of the first frame will tend to have a correspondence with a feature in the second frame and so that the correspondences will be evenly distributed among the features of the second frame. It follows that the process will tend not to leave any feature unmatched.

The F term is necessary to give a time constant for convergence of the network. We define the usual update laws

$$\frac{dU_{ia}}{dt} = -\frac{\partial E}{\partial V_{ia}}, \qquad 1 \leq i \leq N, \quad 1 \leq a \leq N, \tag{4.6}$$

Provided that λ is large enough the variables V_{ia} will tend to be either 0 or 1 and thus it will tend to force a binary decision to determine whether a correspondence is to be established or not.

We simulated this network on a Symbolics 3600 LISP machine. The results are described in detail in Grzywacz and Yuille (1986). To summarize them: despite extensive experimentation with the parameters the system rarely converged to the correct answer unless given a hint of the correct

matches. The system made some interesting mistakes, it would sometimes choose matches which were almost rigid but which corresponded to complicated motion of the object between different frames. This suggested that rigidity alone was not a strong enough constraint and we should introduce another term in the energy function corresponding to smoothness of motion between frames. After some experimentation we fell back on the energy function E_{MM} used by Ullman (1979) to solve the correspondence problem.

$$E_{MM} = \frac{M}{2} \sum_{i=1}^{N} \sum_{a=1}^{N} V_{ia}^2 d_{ia}^2. \tag{4.7}$$

When we added the E_{MM} term to the energy E the network gave consistently good results for a wide range of data Grzywacz and Yuille (1986). It gave a high percent of correct matches for systems of up to thirty points.

To see what contribution the rigidity term made to the matching we removed it and ran the system using the minimal mapping term. The system gave identical results suggesting that the rigidity term was usually uneccessary for matching. A possible exception is at the occluding boundary of an object. Here the order of points can reverse between frames and the minimal mapping scheme gave incorrect results. For small angles and for some values of the parameters the rigidity term obtained the correct matching.

In fact it is easy to see that, with the correct matching, the minimal

mapping energy will be zero for rigid motion in a straight line. This supports the idea that rigidity may be important for correspondence for rotational motions. We plan psychophysical experiments to investigate this case.

In our simulations using the E_{MM} term we did not try to optimize the parameters A, B, C, F, M and λ in any sense. Instead we found that the asymptotic behavior of the system was the same for a large range of parameter values (few orders of magnitude). Typical values used during the course of this research were $A = B = 50000, C = 500000, F = 1, M = 50$ and $\lambda = 1$, where the distances between features in a given frame ranged from 1 to 10. We used homogeneous initial conditions for our simulations, i.e.:

$$V_{ia}(t = 0) = \frac{1}{N}. \tag{4.8}$$

We also tested our network on simple situations such as dot splitting, when there are two dots in the second frame equidistant from a dot in the first. Interestingly our network gave the correct psychophysical result; the dots split into two with $V_{ai} = 1/2$. This suggests psychophysical experiments comparing the predictions of the networks to that of observers for other simple stimuli.

There is an important practical advantage for V.L.S.I. circuits in only using the minimal mapping term for matching. It can be shown (Grzywacz

and Yuille 1986) that the MM term can be choosen to be linear in the matching term,i.e.

$$E_{MM} = \frac{M}{2} \sum_{i=1}^{N} \sum_{a=1}^{N} V_{ia} d_{ia}^2. \tag{4.9}$$

This gives an energy function

$$\begin{aligned} E = \frac{A}{2} \sum_{a=1}^{N} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} V_{ia} V_{ja} + \frac{B}{2} \sum_{i=1}^{N} \sum_{a=1}^{N} \sum_{\substack{b=1 \\ b \neq a}}^{N} V_{ia} V_{ib} \\ + \frac{C}{2} ((\sum_{i=1}^{N} \sum_{a=1}^{N} V_{ia} - N))^2 + E_{MM} \\ + \frac{F}{2\lambda} \sum_{i=1}^{N} \sum_{a=1}^{N} ((V_{ia} \log(V_{ia}) + ((1 - V_{ia})) \log((1 - V_{ia})))), \end{aligned} \tag{4.10}$$

This means that the d_{ia}'s do not affect the connection strengths between elements. Thus the connection strengths will not have to be changed with each time frame.

The energy function described above is rather different from those in the previous two sections. It looks considerably more complicated and a lot more care is needed to find the correct parameters. Unlike the previous problems the value of λ is important, it has to be small compared to the other parameters of the problem. Finally the similarity of the energy function to that used by Hopfield and Tank for the Travelling Salesman Problem suggests it contains many local minima. Given all this, the success of Hopfield style networks for this problem is encouraging. Hopfield networks give progressively worse

results for T.S.P.'s with more than thirty cities. We can prove, however, (Grzywacz and Yuille 1986) that our network will always yield the correct answer if the motion between frames is sufficiently small compared to the seperation of the dots. Thus if time between frames is small enough the correct matching will be made. Other ways of dealing with large numbers of points will be discussed in section 6.

As for surface interpolation we can derive an analytic expression for the solution. Using the chain rule for differentiation we find

$$\frac{dE}{dt} = -\sum_{ia} \frac{\partial U_{ia}}{\partial V_{ia}} \frac{\partial E}{\partial U_{ia}} \frac{\partial E}{\partial U_{ia}}. \tag{4.11}$$

From (4.4) we calculate

$$\frac{\partial U_{ia}}{\partial V_{ia}} = \frac{cosh^2 \lambda U_{ia}}{2\lambda}. \tag{4.12}$$

This term is always positive definite as it is bounded below by $1/2\lambda$. The energy is bounded below and so the system reaches a final state with $dE/dt = 0$. Using (4.11) and (4.12) we see that a necessary and sufficient condition for such a state is

$$\frac{\partial E}{\partial U_{ia}} = 0. \tag{4.13}$$

Inverting (4.4)

$$U_{ia} = \frac{1}{2\lambda} log \frac{V_{ia}}{1 - V_{ia}}. \tag{4.14}$$

Differentiating (4.10)

$$\frac{\partial E}{\partial U_{ia}} = \frac{2\lambda}{cosh^2 \lambda U_{ia}} \big(A(V_i^{COL} + V_a^{ROW} - 2V_{ia}) + C(V - n) + d_{ck}^2 + FU_{ia} \big). \tag{4.15}$$

Here we have introduced new notation. $V = \sum_{ia} V_{ia}$, $V_i^{COL} = \sum_a V_{ia}$ and $V_a^{ROW} = \sum_i V_{ia}$. We have also absorbed the constant $M/2$ into the definition of d_{ia}.

Observe that (4.15) vanishes at $U_{ia} = \pm\infty$. But we can show that these roots do not correspond to minima and hence are not solutions. To prove this we must show that the Hessian of E with respect to the U_{ia} is not positive definite there. We calculate

$$\frac{\partial^2 E}{\partial U_{ia}^2} = \frac{2\lambda}{cosh^2 \lambda U_{ia}} \big((\frac{4\lambda}{cosh^2 \lambda U_{ia}} + F)$$

$$-2\lambda tanh \lambda U_{ia} \big(A(V_i^{COL} + V_a^{ROW} - 2V_{ia}) + C(V - n) + d_{ck}^2 + FU_{ia} \big) \big). \tag{4.16}$$

For large U_{ia} the dominant term inside the bracket is $-2\lambda F U_{ia} tanh \lambda U_{ia}$. Thus $\partial^2 E / \partial U_{ia}^2$ will tend to zero from below as $U_{ia} \mapsto \pm\infty$. Hence the Hessian cannot be positive definite there and the roots at infinity are not minima.

Now consider the other roots of (4.15). These obey

$$A(V_i^{COL} + V_a^{ROW} - 2V_{ia}) + C(V - n) + d_{ck}^2 + FU_{ia} = 0 \tag{4.17}$$

and are also roots of $\partial E/\partial V_{ia} = 0$. At such points

$$\frac{\partial^2 E}{\partial U_{ia}\partial U_{jb}} = \frac{\partial^2 E}{\partial V_{ia}\partial V_{jb}}\frac{\partial V_{ia}}{\partial U_{ia}}\frac{\partial V_{jb}}{\partial U_{jb}} \tag{4.18}$$

and so the Hessian of E with respect to U_{ia} will be positive definite if and only if the Hessian of E with respect to V_{ia} is positive definite. We calculate

$$\frac{\partial^2 E}{\partial V_{ia}\partial V_{ia}} = C + \frac{F}{V_{ck}(1 - V_{ck})} \tag{4.19a}$$

$$\frac{\partial^2 E}{\partial V_{ia}\partial V_{ib}} = A + 2C \tag{4.19b}$$

$$\frac{\partial^2 E}{\partial V_{ia}\partial V_{ja}} = A + C \tag{4.19c}$$

$$\frac{\partial^2 E}{\partial V_{ia}\partial V_{jb}} = C. \quad i \neq j, a \neq b. \tag{4.19d}$$

which is positive definite.

Thus the system will converge to a minima of the energy function which are given by

$$U_{ia} = \frac{2C(n - V) - d_{ck}^2 - A(V_i^{COL} + V_a^{ROW} - 2V_{ia})}{F}. \tag{4.20}$$

Unlike the surface interpolation case we have found no simple interpretation of these equations in general, although results can be obtained for particular cases (Grzywacz and Yuille 1986). Again notice that the final V_{ia} must lie between the limits 0 and 1.

5 Stereo.

A natural way of writing stereo in the form of an energy function is as follows (Barnard 1986, Horn 1986, Gennert 1987)

$$E(d) = \sum_i \left(L_i - R_{i+d(i)}\right)^2 + \mu \sum_i \left(d(i+1) - d(i)\right)^2. \qquad (5.1)$$

Here $d(i)$ is the disparity between the images. L_i and R_i are measures of the left and right images and can be continuous or discrete. For example they could be the image intensities, or they could be the positions of zero crossings. The first term in (5.1) matches the left and right images in such a way as to minimize the disparity gradient represented by the second term. Line processors can be introduced to prevent the disparity gradients from becoming

too large and allows distinct objects to influence each others matching. *. This yields

$$E(d,l) = \sum_i \big(L_i - R_{i+d(i)}\big)^2 (1-l_i) + \mu \sum_i \big(d(i+1) - d(i)\big)^2 (1-l_i) + \tau \sum_i l_i. \tag{5.2}$$

The l_i performs two functions in (5.2). It prevents the disparity gradient from becoming too large but it also prevents matching in the first term if the difference between the images is too great. This latter effect may help deal with occluding situations when one eye sees a region which the other cannot. It is simple to generalize this energy function to two dimensions.

It is straightforward to design a Hopfield net for this problem in the usual manner. We simulated this on two types of examples (in two dimensions). The first consisted on a sine wave with the central square displaced and the second was a standard random dot stereogram. The results were disappointing in both cases. Although it was possible to get roughly the right answer for the sine wave the random dot pattern gave consistently poor results. This occured despite filtering the images with a variety of filters, gaussians and difference of gaussians at various scales.

We believe that this bad behaviour is due to the complicated structure

---

* This work was done in collaboration with T. Poggio

of the energy as a function of d. The behaviour of this function is crucially dependent on the structure of R. For example if R is a linear function of its arguments the energy will be quadratic in d and therefore well behaved. If R is more complex, in particular if it arises from a random dot stereogram, the energy function can be a complicated function of d with many local minima. To apply this method to a general scene would require smoothing the images (by an amount determined by pre-processing) until the function R was sufficiently well behaved, avoiding smoothing the image too much to destroy its interesting features. While this is conceivable it seems that alternative stereo algorithms are likely to be more sucessful.

A possible approach is to attempt to minimize (5.2) using other algorithms. Some success (Barnard 1986) has been reported for using simulated annealing for (5.1) (in two dimensions). So far attempts to use simulated annealing, and other stochastic methods, to minimze (5.2) have not been sucessful. Although it has been possible to hand tune the parameters to get the correct result for one class of stimuli, sign waves with displaced centres, it will not work on others, such as random dots. Various refinements of the methods have been tried including a coarse to fine strategy of convolving the image with a large scale gaussian, minimizing the energy at this scale and using this to guide the matching at smaller scales.

This result is negative and by no means conclusive. More sophisticated algorithms could be tried including perhaps methods of estimating the parameters of the energy function directly from the image. Alternative energy functions could be tried. We believe, however, that this may represent a limit for the practical use of energy functions. If an energy function is too complicated for straightforward algorithms to solve then the problem is badly posed and more heuristic methods should be used. This will be discussed further in the next section.

6 Limitations of the energy function approach.

The previous four sections described attempts at modelling problems in terms of energy functions using analog nets, The first three attempts were reasonably sucessful and the last one failed. We argued that this reflected the relative complexities of the energy functions being minimized.

It is clearly possible to write any vision problem in terms of minimizing an energy function * . It is less clear that this is a good strategy. For non-trivial

* In the same way that all the laws of physics can be summarized in one equation

problems it often leads to energy functions with many local minima depending on large numbers of parameters. These parameters will often depend on the image being viewed and might even have different values in different regions of the image * . As yet there is no reliable way to estimate these parameters, or of minimizing energy functions with many local minima.

Energy function methods seem a natural idea for matching representations of images containing a large number of similar primitives. Random Dot stereograms (introduced by Julesz 1971) are ideally suited for this type of algorithm. Realistic images, however, contain many different features of varying sizes. A good strategy for stereo could involve matching the most salient features and using this to guide the ambiguous features. The work on stereo by Mitchison and McKee (1987) shows that for one-dimensional sterograms the positions of the endpoints have an important effect on the matching of the interior points. Another interesting example of this type is psychophysics for motion correspondence illustated by Ramachandran's analogy of a moving leopard (Ramachandran 1985, Ramachandran and Anstis 1983). If the outline of the leopard is not visible the leopard's spots are matched to nearby neighbours and no motion is seen. If the outline is also visible then its mo-

by summing the squares of all the individual laws (Feynman 1963).

* For example in the way the noise thresholds are determined locally for the Canny edge detector.

tion "captures" the spots and they are matched correctly. This suggests that random dot stereograms are a limited paradigm and that although humans have the capacity to match them correctly they may not be the strategy used for real images when more information is available.

Moreover the energy function approach, at least in its most naive form, tends to ignore some of the structure of the problem. For example, the motion smoothing and segmentation algorithm described in section 3 would, in theory, be able to detect the boundaries of the leopard but it would not be able to use the speed of the boundary directly to influence the internal matching. There would certainly be an indirect influence, since the motion is required to be smooth, but this would be weak and depend on the distance from the boundary. One can contrast this with a more heuristic approach which would analyse the scene, detect the object boundaries, estimate their velocities and use this as initial data for matching the interior of the object. The algorithm could be designed in terms of several different networks connected together and certainly individual parts of this method could be implemented by energy functions. For example Grzywacz (private communication) has shown that the network described in section 4 will correctly match the leopard's spots if the estimated motion is available as initial data.

The stereo energy function (5.2) also does not capture some of the im-

portant features of the problem. Places where the line-processors should be on correspond to boundaries of objects, and therefore should correspond to edges in the image. Thus it would be more sensible to find these edges by simple processing of the image rather than by minimizing (5.2).

For any given module, such as stereo, there are many possible algorithms some matching image intensity, others matching edge-like features. The relative effectiveness of these different algorithms will depend on the images being viewed, an edge-based stereo algorithm would be ineffective in a scene containing few or weak edges. For some scenes it would be natural to try to find and then match salient features, such as the occluding boundaries of objects or regions of high texture density. Heuristics like coarse to fine matching ,as used by Marr and Poggio (1979) for stereo, could also be used. For realistic images a stereo system might have to use a number of different algorithms, or submodules, interacting with one another and combining to give the solution. These submodules might each be implemented in terms of networks minimizing energy functions but the most important part of the calculation, and the hardest part of designing such as system, would lie in the control strategy for combining the different submodules.

Thus although minimizing energy functions is a useful technique for early vision it has definite limitations. If the energy has too many local minima it

may be better to try heuristic methods to avoid them rather than to use complicated search techniques. Minimizing an energy function is only one of the many different search strategies used in Artificial Intelligence research and is only effective for certain problems.

7 Conclusion.

We described how a number of problems in early vision could be described in terms of minimizing energy functions. We showed that Hopfield style networks were able to give fast reliable answers for some of these. We discussed the limitations of this approach.

Acknowledgements I would like to acknowledge conversations with Norberto Grzywacz, and Christof Koch. Berthold Horn and Tommy Poggio

gave helpful comments on drafts of the paper.

Appendix

A.1. Hopfield's Formalism

Hopfield describes a system of n neurons with output's V_i and an external input I_i, where i runs from 1 to n. He defines an update rule by

$$C_i \frac{du_i}{dt} = \sum_j T_{ij} V_j - \frac{u_i}{R_i} + I_i. \qquad (A.1.1)$$

Here T_{ij} is a measure of the strength of the connection between neuron i and neuron j. A priori, every T_{ij} can be positive, negative or zero. C_i is a capacitance and $R_i = 1/\sum_j T_{ij}$ a resistance associated with every neuron. u_i represents some internal function of the neuron i, for example its somatic potential, and is given by

$$u_i = g_i^{-1}(V_i) \qquad (A.1.2)$$

where $g(x)$ is a monotonic increasing, but bounded function. This model is deterministic, its final state will depend on the initial conditions and the inputs I_i. Hopfield argues that it embodies a content addressable memory. To do this he defines an "energy" function

$$E(V_i) = -\frac{1}{2}\sum_{i,j} T_{ij}V_iV_j + \sum_i \frac{1}{R_i}\int_0^{V_i} g_i^{-1}(V)dV + \sum_i I_iV_i \qquad (A.1.3)$$

with the restrictions $T_{ij} = T_{ji}$ and $T_{ii} = 0$. Differentiating (A.1.3) using the chain rule and substituting (A.1.1) and (A.1.2) we have

$$\frac{dE}{dt} = -\sum_i C_i g_i^{-1\prime}(V_i)\left(\frac{dV_i}{dt}\right)^2 \qquad (A.1.4)$$

which is always negative. E is hence a Lyaponov function and by standard results of Statistical Mechanics any solution to (A.1.1) will converge to one of a fixed number of stable points, provided E is bounded below. The precise fixed point (A.1.1) converges to will depend on the initial conditions and the external inputs. The potential of the system contains a large number of local minima and the minima the system ends up in is determined by the initial conditions. The system can therefore be thought of as a content addressable memory. The requirements that the connectivity matrix T is symmetric is needed in order to assure that the update of u_i has the form of equation (A.1.1).

A.2 Extending Hopfield's networks: We will now proceed to propose a more general class of networks and update rules, The essence of Hopfield's networks can be described as follows. First we define an "energy" function

$$E = E(V_i). \tag{A.2.1}$$

We should emphasize the quotation marks round "energy". It is not necessarily the energy of the physical system but merely a Lyapunov function. It is bounded below. We will reformulate our update rule as

$$\frac{dV_i}{dt} = -\sum_j A_{ij} \frac{\partial E}{\partial V_j}. \tag{A.2.2}$$

This is a generalization of (A.1.1). In order for E to be a Lyapunov function, its temporal derivative must be everywhere negative or zero. In other words, the "energy" must always decrease or at most remain constant, but never increase. Differentiating (A.2.1) using (A.2.2) yields

$$\frac{dE}{dt} = -\sum_{i,j} A_{ij}^{-1} \frac{dV_i}{dt} \frac{dV_j}{dt} = -\sum_{i,j} A_{ij} \frac{\partial E}{\partial V_i} \frac{\partial E}{\partial V_j}. \tag{A.2.3}$$

E is a Lyapunov function (and hence the system has a content addressible memory) if and only if A_{ij} is positive definite, i.e. $x^T A x > 0$ for all vectors x. For any arbitrary function E there are therefore an infinite number of possible updating rules and so an infinite number of possible systems.

It is straightforward to check that Hopfields network defined by (A.1.1), (A.1.2) and (A.1.3) can be obtained by setting

$$A_{ij}^{-1} = C_i g_i^{-1\prime}(V_i)\delta_{ij}. \qquad (A.2.4)$$

We will show in A.3 that Hopfield's energy function will still be a Lyapunov function if we set

$$A_{ij}^{-1} = C_i g_i^{-1\prime}(V_i) l_i \delta_{ij}, \qquad (A.2.5)$$

where the l_i are an arbitrary set of positive numbers. This enables us to relax the symmetry constraint on the T_{ij} to

$$T_{ij} l_i = T_{ji} l_j, \qquad (A.2.6)$$

where there is no summation over the indices. We prove in the next section that for an $n \times n$ matrix this gives us an additional $n - 1$ degrees of freedom.

It is important to note that the number of constraints is less important than the form of the constraints. In general, the matrix T_{ij} will be sparse, since most connections do not exist. This becomes especially true if n is large. The l_i's can then be applied to the remaining non-zero T_{ij}. Hopfield's symmetry conditions implies, that once the top left half of T_{ij} is specified, the bottom half will be completely determined, since $T_{ij} = T_{ji}$, independent of the number of zeros in T_{ij}. In our extension, however, $n - 1$ of the **non-**

zero entries can be specified at will, as long as T_{ij} and T_{ji} have the same sign. This condition prohibits the use local inhibitory interneurons, that is neurons which are excited ($T_{ij} > 0$) but which inhibit in turn ($T_{ji} < 0$). If the network is used as a content addressible network without allowing too many errors during "recall", only $\approx 6\%$ of the possible $n^2 - n$ entries are different from zero (Hopfield, 1982). Thus, for $n = 30$ and using Hopfield's symmetric network, only about 26 T_{ij}'s can be specified while the remaining ones are fixed. Our extension implies, however, that all T_{ji} can take on arbitrary values — as long as the sign of transposed elements is the same. Notice that these conditions in no way constrain the diagonal terms T_{ii}. *

A.3 The full extension:

Formally, the integrability condition (A.2.1) in E is analogous to the existence of potentials, or state functions, in Thermodynamics. If this condition is not met, then the value of E depends on the path taken in V_i space and hence cannot be a Lyapunov function.

There are many other ways of constructing Lyapunov functions for a system with a given update rule.

In order to study the class of connectivity matrices T_{ij} leading to converging behaviour, we will now express Hopfield's update rule as

---

*An alternative to our proof is scaling u_i and V_i independently (Hopfield, private communication).

$$\frac{dV_i}{dt} = \frac{1}{C_i g_i^{-1\prime}(V_i)}\left(\sum_j T_{ij}V_j - \frac{g_i^{-1}(V_i)}{R_i} - I_i\right) \qquad (A.3.1)$$

By inverting equation (A.2.2) we can define the appropriate Lyapunov function

$$\frac{\partial E}{\partial V_i} = -\sum_j A_{ij}^{-1}\frac{dV_j}{dt} \qquad (A.3.2)$$

or,

$$E(V_i) = -\int \sum_{i,j} A_{ij}^{-1}\frac{dV_i}{dt}dV_j. \qquad (A.3.3)$$

Substituting from (A.3.1) gives an integrand I

$$I = -\sum_{i,j} A_{ij}^{-1}\frac{1}{C_i g_i^{-1\prime}(V_i)}\left(\sum_k T_{ik}V_k - \frac{g_i^{-1}(V_i)}{R_i} - I_i\right)dV_j. \qquad (A.3.4)$$

If this expression is integrable, that is its value is independent of the path along which the integral was evaluated, then E is well-defined and a Lyapunov function. A function I is integrable if, and only if, $dI = 0$ where d is the exterior derivative operator (Misner, Thorne, Wheeler, 1977). Define B_{ij} and h_i by $A_{ij}^{-1} = C_i g_i^{-1\prime}(V_i)B_{ij}$ and $h_i = g_i^{-1}(V_i)/R_i$. The integrand is then

$$I = \sum_{i,j,k} B_{ij}T_{ik}V_k dV_j - \sum_{i,j} B_{ij}h_i dV_j - \sum_{i,j} B_{ij}I_i dV_j. \qquad (A.3.5)$$

If B_{ij} and T_{ij} are independent of the V_i's then I is integrable provided

$$dI = \sum_{i,j,k} B_{kj}T_{ki}dV_i \times dV_j - \sum_{i,j} B_{ij}\frac{dh_i}{dV_i}dV_i \times dV_j = 0. \qquad (A.3.6)$$

Since $dV_i \times dV_j$ is antisymmetric in i and j this will hold provided $\sum_k B_{kj}T_{ki} + B_{ij}\frac{dh_i}{dV_i}$ is symmetric in i and j. If h_i is non-linear this will only be possible provided B_{ij} does not contain off-diagonal terms. Thus, with

$$B_{ij} = l_i\delta_{ij}, \qquad (A.3.7)$$

where the l_i's are positive numbers, equation (A.3.4) changes into

$$\sum_{i,j} l_i\delta_{ij}(\sum_k T_{ik}V_k - \frac{g_i^{-1}(V_i)}{R_i} - I_i)dV_j. \qquad (A.3.8)$$

If l_iT_{ij} is symmetric, then (A.3.8) can be integrated since it consists only of terms like $d(V_iV_j)$ and $g_i^{-1}(V_i)dV_i$ which are well-defined. Thus, we can generalize Hopfields's result to all matrices T_{ij}, provided l_iT_{ij} is symmetric and A_{ij} is positive definite. Following (A.3.7), matrix A must be diagonal and therefore its eigenvectors are equal to the l_i's, each of which can be any arbitrary positive numbers. Whatever the values of the l_i's there are definite relations which must hold between the T_{ij}. In particular, for all i, j, k, we have

$$\frac{T_{ij}}{T_{ji}} \cdot \frac{T_{ki}}{T_{ik}} \cdot \frac{T_{jk}}{T_{kl}} = 1. \qquad (A.3.9a)$$

with the auxiliary conditions

$$T_{ij} \cdot T_{ji} > 0. \qquad (A.3.9b)$$

More complicated relations can be deduced but they can all be obtained by combining relations of type (A.3.9).

We now examine the constraints in more detail. Suppose we specify the values T_{ij}, $i < j$ in the upper right half of the matrix. Then the values in the lower left half are given by

$$T_{ji} = T_{ij}\frac{l_j}{l_i}. \qquad (A.3.10)$$

We will ignore the case when all elements in column j are zero (and therefore also all elements in row j), since the order of the matrix will simply be reduced by one. Otherwise the lower left half of the matrix is determined by the l_i/l_j. Now all the l_i/l_j can be determined from a basic $n-1$ elements

$$\frac{l_1}{l_2}, \frac{l_2}{l_3}, ..., \frac{l_{n-1}}{l_n} \qquad (A.3.11)$$

$$\frac{dE}{dt} = \sum_j \frac{\partial E}{\partial V_j}\frac{dV_j}{dt} \leq 0, \qquad (A.3.12)$$

for all motions. Requiring the update rule to be related to E by (A.3.2) will ensure (A.3.6). There are many other ways of enforcing (A.3.12). For Hopfield's updating rule, however, we have been unable to find any other integrable Lyapunov function.

References

Arbib, M.A. "Artificial Intelligence and Brain Theory". Annals of Biomedical Engineering, 3:238–274. 1975.

Barnard, S. *Proc. Image Understanding Workshp*, Los Angeles. 1986.

Barrow, H.G. and Tennenbaum, J.M. "Computational Vision", *Proc. IEEE.* **69**, 5. 572-595. 1981.

Blake, A. "The least-disturbance principle and weak constraints", *Patt. Recogn. Lett.*, **1**, 393-399, 1983.

Brandt, A. "Multi-level Adaptive Solutions to Boundary-value Problems", *Math. Comp.* **31**. 333-390. 1977.

Canny, J.F. "Finding edges and lines in images". MIT A.I.Lab Technical Report No. 720. 1983.

Courant, R. Hilbert, D. "Methods of mathematical physics", Vol.1. Interscience Publishers. Inc. New York. 1953.

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

Feynman, R. "The Feynman Lectures on Physics". Addison-Wesley. 1963.

Gennert, M. "A Computational Framework for understanding Problems in Stereo Vision". MIT Artificial Intelligence Laboratory PhD. Thesis. In preparation. 1987.

Geman, S. and Geman, D. "Stochastic Relaxation, Gibbs Distributions asnd the Bayesian Restoration of Images", *IEEE Trans, Pattern Analysis and Machine Intelligence,* **6**, 721-741, 1984.

Grimson, W.E.L. "From Images to Surfaces: A computational study of the human early visual system", MIT Press, Cambridge, Ma. 1981.

Grzywacz, N.M. Hildreth, E. "Incremental rigidity scheme for recovering structure from motion". *Journal of the Optical Society.* Vol 4, pp 503-518. 1987.

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

Hildreth, E.C. "The Measurement of Visual Motion", MIT Press, Cambridge, Ma. 1984.

Hinton, G. "Relaxation and its role in vision", *Ph.D. Thesis.* University of Edinburgh. 1979.

Hopfield, J.J. "Neural networks and physical systems with emergent collective computational abilities", *Proc. Natl. Acad. Sci. USA*, **79**: 2554-2558. 1982.

Hopfield, J.J. "Neurons with graded response have collective computational properties like those of two-state neurons",*Proc. Natl. Acad. Sci. USA*, **81**: 3088-3092. 1984.

Hopfield, J.J. and Tank, D.W. "Neural computation in optimization problems", *Biological Cybernetics.* 1985.

Horn, B.K.P. "Robot Vision". MIT Press, Cambridge, Ma. 1986.

Horn, B.K.P. "Determining lightness from an image", *Computer Graphics and Image Processing.* **3**. 277-299. 1974.

Horn, B.K.P. and Schunk, B.G. "Determining optical flow", *Artificial Intelligence,* **17**, 185-203. 1981.

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

Ikeuchi, K. "Numerical shape from shading and occluding contours in a single view", *A.I. Memo 566.* AI Lab, MIT. 1980.

Ikeuchi, K. "Shape from Regular Patterns,"M.I.T. Artificial Intelligence Laboratory Memo 567. 1980.

Ikeuchi, K. and Horn, B.K.P. "Numerical shape from shading and occluding boundaries," *Artificial Intelligence.17.* 1981.

Julesz, B. "Foundations of Cyclopean Perception", The University of Chicago Press. Chicago. 1971.

Karplus, W.J. "Analog Circuits: Solutions of Field Problems,". McGraw-Hill. Ney York. 1958.

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

Koch, C., Marroquin, J. and Yuille, A.L. "Analog "Neuronal" Networks in Early Vision". Proc. Natl. Acad. Sci. USA, 83:4263–4267. 1986.

Koch, C., and T. Poggio. "Biophysics of Computation: Neurons, Synapses and Membranes". MIT Artificial Inteligence Memo, 795. 1986.

Marr, D. "Vision". W.H. Freeman Press. San Fransisco. 1982.

Marr, D. Poggio, T. "Cooperative computation of stereo disparity", *Science,* **195**, 283-287. 1977.

Marr, D. Poggio, T. "A theory of human stereo vision", *Proc. Roy. Soc. Lond. B.*, **204**, 301-328. 1979.

Marroquin, J. "Probabilistic solution of inverse problems", *Ph.D. Thesis..* Department of Electrical Engineering and Computer Science, MIT. 1985.

Marroquin, J. "Deterministic Bayesian Estimation of Markovian Random

fields with Applications to Computational Vision". Proceedings of the First International Conference on Computer Vision, London. 1987

Marroquin, J., Mitter, S. and Poggio, T. "Probabilistic solutions of ill-posed problems in computational vision". *Journal of the American Statistical Association.* Vol 82, no. 397. 1987.

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

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

Misner, C.W., Thorne, K.S. and Wheeler, J.A. "Gravitation". W.H. Freeman. 1973.

Mitchison, G.J. and McKee, S. "The resolution of ambiguous stereoscopic matches by interpolation". *Vision Res.*. Vol 27, no 2, pp 285-294. 1987.

Poggio, T. and Torre, V. "Ill-posed problems and regularization analysis in early vision". *Artificial Intelligence Laboratory Memo.* No 773. M.I.T. 1984.

Poggio, T., V. Torre, and C. Koch "Computational Vision and Regularization Theory". *Nature*, 317:314–319. 1985.

Ramachandran, V.S. "Apparent Motion of Subjective Surfaces". Percep-

tion, 14:127–134. 1985.

Ramachandran, V.S., and Anstis, S.M. "Displacement Thresholds for Coherent Apparent Motion". *Vision Res.* 23:1719–1724. 1983.

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

Terzopoulos, D. "Multiresolution computation of visible-surface representations", *Ph.D. Thesis.* Department of Electrical Engineering and Computer Science, MIT. 1984.

Tikhonov, A.N. and Arsenin, V.Y. "Solutions of Ill-posed Problems", Winston and Sons, Washington. D.C. 1977.

Ullman, S. "Relaxation and constrained optimization by local processes", *Comp. Graphics Image Proc.*, **10**, 115-125. 1979.

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

Ullman, S. 1984. "Incremental Rigidity". MIT Artificial Intelligence Laboratory Memo. 1984.

This blank page was inserted to preserve pagination.

CS-TR Scanning Project
Document Control Form

Date : 5/26/95

Report # AIM-987

Each of the following should be identified by a checkmark:
Originating Department:

☒ Artificial Intellegence Laboratory (AI)
☐ Laboratory for Computer Science (LCS)

Document Type:

☐ Technical Report (TR) ☒ Technical Memo (TM)
☐ Other:_____

Document Information

Number of pages: 57 (58-IMAGES)
Not to include DOD forms, printer instructions, etc... original pages only.

Originals are:
☒ Single-sided or
☐ Double-sided

Intended to be printed as :
☐ Single-sided or
☒ Double-sided

Print type:
☐ Typewriter ☐ Offset Press ☒ Laser Print
☐ InkJet Printer ☐ Unknown ☐ Other:_____

Check each if included with document:

☒ DOD Form ☐ Funding Agent Form ☐ Cover Page
☐ Spine ☐ Printers Notes ☐ Photo negatives
☐ Other: _____

Page Data:

Blank Pages(by page number):_____

Photographs/Tonal Material (by page number):_____

Other (note description/page number):

Description :	Page Number:
IMAGE MAP (1) UN #'ED TITLE PAGE	
(2-53) PAGES #'ED 1-52	
(54-55) SCANCONTROL, DOD	
(56-58) TRGT'S (3)	

Scanning Agent Signoff:
Date Received: 5/26/95 Date Scanned: 6/5/95 Date Returned: 6/8/95

Scanning Agent Signature: Michael W. Cook

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AI Memo 987	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER AD-A190384
4. TITLE (and Subtitle) Energy Fundtions for Early Vision and Analog Networks		5. TYPE OF REPORT & PERIOD COVERED memorandum 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Alan Yuille		8. CONTRACT OR GRANT NUMBER(s) N00014-85-K-0124
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 November 1987 13. NUMBER OF PAGES 53
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)

18. SUPPLEMENTARY NOTES

None

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

energy functions
line processors
analog networks

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

This paper describes attempts to model the modules of early vision in terms of minimizing energy functions, in particular energy functions allowing discontinuities in the solution. It examines the success of using Hopfield-style analog networks for solving such problems. Finally it discusses the limitations of the energy function approach.

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)

Scanning Agent Identification Target

Scanning of this document was supported in part by the **Corporation for National Research Initiatives,** using funds from the **Advanced Research Projects Agency** of the **United states Government** under Grant: **MDA972-92-J1029.**

The scanning agent for this project was the **Document Services** department of the **M.I.T Libraries.** Technical support for this project was also provided by the **M.I.T. Laboratory for Computer Sciences.**

Scanned

Date: 6/5/1995

M.I.T. Libraries
Document Services