Network: Computation in Neural Systems **7** (1996) 635–669. Printed in the UK

Neural model of visual stereomatching: slant, transparency and clouds

Jonathan A Marshall†, George J Kalarickal† and Elizabeth B Graves†

Department of Computer Science, CB 3175, Sitterson Hall, University of North Carolina, Chapel Hill, NC 27599-3175, USA

Received 3 September 1996

Abstract. Stereomatching of oblique and transparent surfaces is described using a model of cortical binocular 'tuned' neurons selective for disparities of individual visual features and neurons selective for the position, depth and 3D orientation of local surface patches. The model is based on a simple set of learning rules. In the model, monocular neurons project excitatory connection pathways to binocular neurons at appropriate disparities. Binocular neurons project excitatory connection pathways to appropriately tuned 'surface patch' neurons. The surface patch neurons project reciprocal excitatory connection pathways to the binocular neurons. Anisotropic intralayer inhibitory connection pathways project between neurons with overlapping receptive fields. The model's responses to simulated stereo image pairs depicting a variety of oblique surfaces and transparently overlaid surfaces are presented. For all the surfaces, the model (i) assigns disparity matches and surface patch representations based on global surface coherence and uniqueness, (ii) permits coactivation of neurons representing multiple disparities within the same image location, (iii) represents oblique slanted and tilted surfaces directly, rather than approximating them with a series of frontoparallel steps, (iv) assigns disparities to a cloud of points at random depths, like human observers and unlike Prazdny's (1985) method, and (v) causes globally consistent matches to override greedy local matches. The model represents transparency, unlike the model of Marr and Poggio (1976), and it assigns unique disparities, unlike the model of Prazdny.

1. Introduction

This paper describes a neural network model for stereomatching, a major component of stereopsis. The model is constrained by anatomical, neurophysiological and psychophysical data on stereopsis. Parts of the model can be identified with certain areas of animal brains, based on neurophysiological data. An additional biological constraint is self-organization. The self-organization constraint is important because stereopsis appears to develop postnatally (Fox *et al* 1980).

The neural network model for stereomatching presented here is based on an analysis of self-organization in response to exposure to ordinary stereo image pairs, using a simple, neurobiologically plausible set of learning rules (Marshall 1990a, 1992, 1995a). The same rules also lead to the self-organization of neural mechanisms for a variety of other visual processing functions (Gupta and Marshall 1995, Hubbard and Marshall 1994, Marshall 1989, 1990a, b, c, d, 1992, 1995b, Marshall and Alley 1993, Schmitt and Marshall 1995). Thus, the stereomatching model fits within a unified theory of vision based on self-organization.

† E-mail addresses: marshall@cs.unc.edu, kalarick@cs.unc.edu, graves@cs.unc.edu

0954-898X/96/040635+35$19.50 © 1996 IOP Publishing Ltd

In the remainder of this section, the problem of stereomatching is discussed in more detail, and some of the hardest stereomatching tasks performed by humans are described. Two important stereomatching algorithms are discussed. Biological constraints and the shortcomings of these algorithms motivate the new model.

1.1. The correspondence problem and random dot stereograms

Identification of corresponding retinal images of some points or features in the visual scene is necessary to measure disparity. Identification of corresponding parts in the two retinal images is termed the *correspondence problem* or the *stereomatching problem*. The necessity of disparity computation for stereopsis was demonstrated by Julesz (1971) using random dot stereograms (RDSs). The lack of monocular depth cues in RDSs poses a challenge for the design of computational algorithms for stereomatching. A large number of possible matches between dots in the two images needs to be checked to eliminate *false matches* and extract surfaces in depth. Human observers can perceive objects of arbitrary shape in RDSs with only the disparity information defining the shapes.

1.2. Objectives

This paper describes a novel neural network based algorithm that computes stereomatches in images that were problematic for previous algorithms.

Previous stereomatching algorithms cannot represent transparently overlaid surfaces simultaneously (Frisby and Pollard 1991, Marr and Poggio 1976), cannot represent surfaces of arbitrary orientation directly (Marr and Poggio 1976, Qian and Sejnowski 1989), and cannot assign unique disparity values to features (Prazdny 1985).

To emphasize the distinction between the network presented here and the other algorithms, the new stereomatching network is called the *exclusive grouping* (EG) model because it separates the feature pairs into exclusive groups that define different surfaces and assigns unique disparity values to the feature pairs.

To motivate the EG model for stereomatching, the next section discusses the Marr–Poggio (1976) and Prazdny (1985) algorithms.

1.3. Previous models: strengths and shortcomings

Marr and Poggio (1976) developed an early stereomatching algorithm that implemented constraints based on an analysis of opaque objects. In contrast, the Prazdny (1985) algorithm was motivated by an analysis of transparently overlaid surfaces.

1.3.1. The Marr–Poggio stereomatching algorithm. The Marr–Poggio (1976) stereomatching algorithm uses uniqueness and continuity constraints to eliminate false matches and obtain global consistency. The uniqueness constraint requires that each item from each image be assigned at most one disparity value. This is based on the assumption that a feature corresponds to something that has a unique physical position in space. The continuity constraint requires disparity to vary smoothly almost everywhere. However, this constraint is valid only for opaque objects.

Their algorithm iteratively seeks solutions to a system of constraints imposed by:

- the stereo image data
- lateral inhibition between neurons representing different disparities along the same left- or right-eye lines of sight, and

- lateral excitation between neurons representing the same disparity at adjacent image locations.

The stereo image data provide the specific primitives or features (e.g. white or black dots) to be matched. The inhibition implements a uniqueness constraint; the neurons representing different possible matches for each left- or right-image dot compete with one another, thereby preventing more than one disparity from being assigned to each dot. The lateral excitation implements the continuity constraint, whereby a match at a given disparity tends to cause the same disparity value to be assigned to adjacent image locations.

The Marr–Poggio algorithm has two drawbacks that will be discussed below. Both problems are caused by the strictly frontoparallel lateral excitation. The lateral excitation is an attempt to express the continuity constraint.

First, the Marr–Poggio algorithm cannot correctly represent stereoimages of

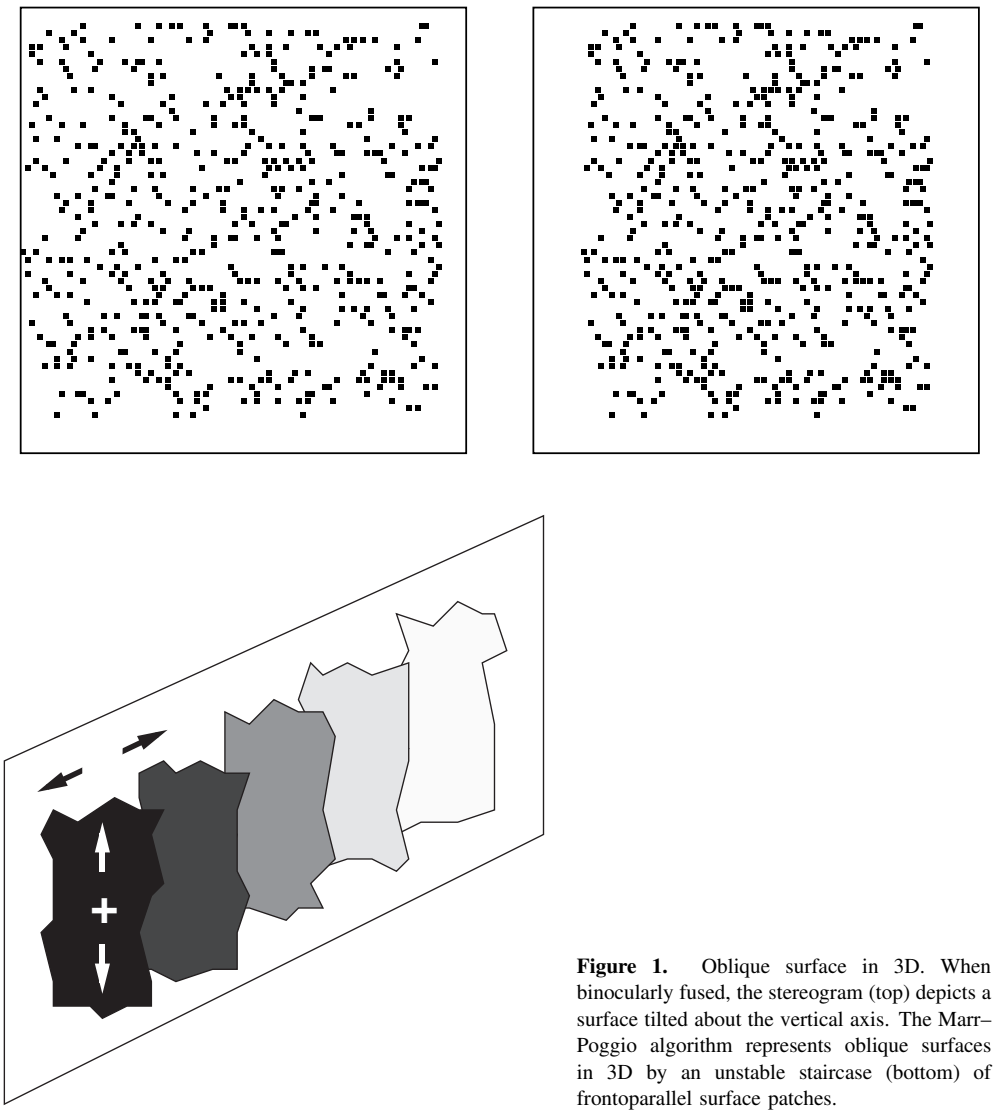

Figure 1. Oblique surface in 3D. When binocularly fused, the stereogram (top) depicts a surface tilted about the vertical axis. The Marr–Poggio algorithm represents oblique surfaces in 3D by an unstable staircase (bottom) of frontoparallel surface patches.

transparently overlaid surfaces in depth. In RDSs of transparently overlaid surfaces, multiple surfaces occupy the same 2D spatial region in image space. Therefore, the continuity constraint is invalid in this situation. Human observers perceive transparently overlaid surfaces in these RDSs (Prazdny 1985). However, these RDSs are often somewhat harder to fuse than RDSs defining opaque surfaces.

The frontoparallel lateral excitation also activates neurons that do not receive direct input from the stereo image features. Thus, in response to stereo input containing sparse disparity information for transparently overlaid surfaces in depth, the dense surface interpolation via lateral excitation causes neurons representing different surfaces in depth to inhibit one another laterally. This results in an unstable representation for transparently overlaid surfaces.

Second, a flat surface that recedes in depth can be represented approximately by an unstable staircase of frontoparallel surface patches (figure 1).

1.3.2. The Prazdny stereomatching algorithm. Prazdny (1985) addressed the representation of transparently overlaid surfaces by eliminating inhibition and adding a neighbourhood support function that computes the evidence for the presence of surfaces passing through each point in depth. Removing the inhibition allows the network to represent multiple disparities within an image location (figure 2).

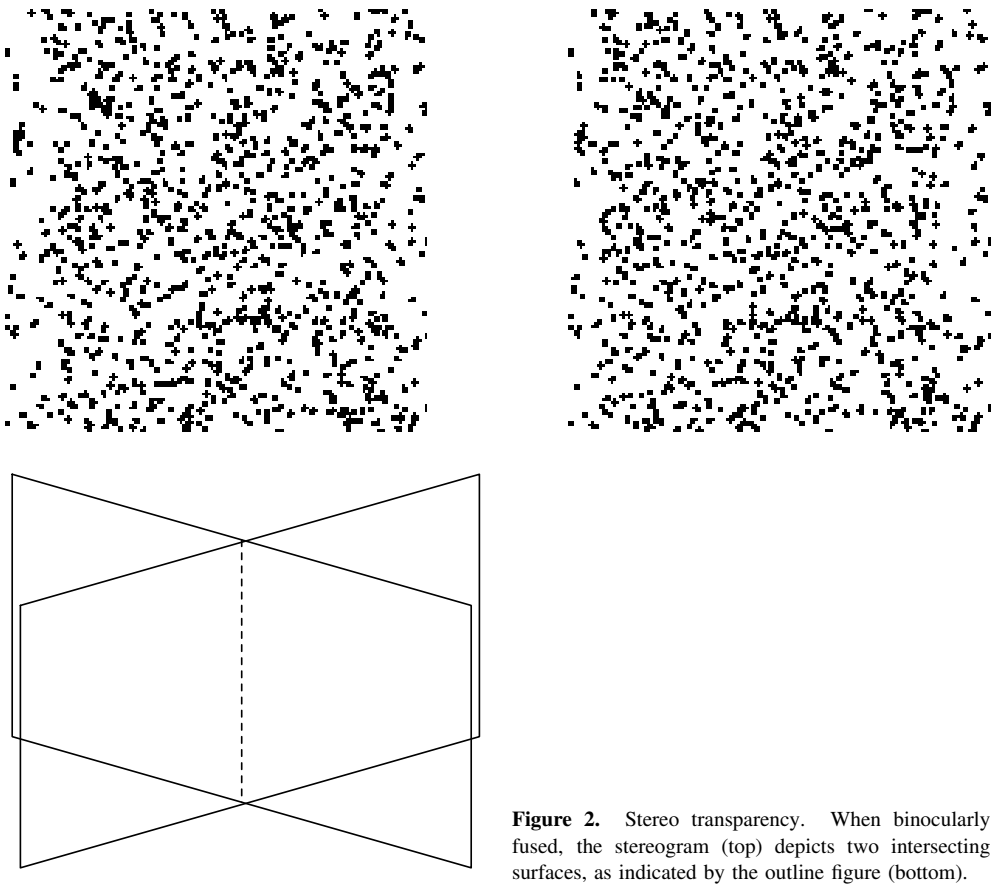

Figure 2. Stereo transparency. When binocularly fused, the stereogram (top) depicts two intersecting surfaces, as indicated by the outline figure (bottom).

The Prazdny algorithm collects evidence for surfaces in depth without assigning specific left–right correspondences. A surface is detected and represented if the number of potential feature matches defining the surface is sufficiently large. However, the Prazdny algorithm lacks a uniqueness constraint; as a result, some features can be assigned to more than one surface at very different disparities. In the worst case, periodically repeated image features (e.g. a chain-link fence) are represented as containing multiple surfaces simultaneously (Weinshall 1989).

When RDSs with pairs of dots with random disparities are viewed stereoscopically, observers perceive a cloud of random dots. They do not perceive a surface. Nevertheless, the individual dots are perceived in depth. In the Prazdny algorithm, however, no surface receives strong excitation and the dots cannot be assigned specific disparities.

1.4. Overview of the EG stereomatching algorithm

In the previous section it was noted that the Marr–Poggio algorithm assigns at most one disparity value to each dot but represents oblique and transparently overlaid surfaces incorrectly. On the other hand, the Prazdny algorithm represents oblique and transparently overlaid surfaces but cannot assign unique disparity values to the dots. The EG stereomatching algorithm retains the advantages of the Marr–Poggio and Prazdny algorithms but overcomes some of their disadvantages. This section outlines the similarities and differences between the EG stereomatching algorithm and the Marr–Poggio and Prazdny algorithms.

The EG stereomatching algorithm has lateral inhibition between neurons representing different disparities along the same left- or right-eye lines of sight, like the Marr–Poggio algorithm and unlike that of Prazdny. However, the EG algorithm does not have the direct lateral excitation of the Marr–Poggio algorithm, between neurons representing the same disparity at adjacent image locations.

In addition, the EG algorithm uses a local support function similar to but more general than that of the Prazdny algorithm, to group dot pairs into surfaces. The grouping is exclusive because dots can belong to only one group, unlike in the Prazdny algorithm.

Thus, the EG algorithm eliminates the features of the Marr–Poggio and Prazdny algorithms that led to their shortcomings, namely the strictly frontoparallel excitation in the Marr–Poggio algorithm and the multiple disparity value assignment in the Prazdny algorithm.

The grouping capability is further exploited in solving ambiguous RDSs. In ambiguous RDSs several valid correspondences exist. Human observers in these cases tend to perceive the scene in a globally consistent manner. For example, if an RDS can be matched to represent a single surface or two surfaces, the single surface is more likely to be perceived.

1.5. Overview of the paper

Section 2 describes the neural network architecture and its operation for stereomatching. The key features of the model are selective inhibition and controlled top-down feedback. Section 3 presents the network implementation details. The stereomatching results of the network in response to several RDSs are presented in section 4. Section 5 summarizes the novel and salient properties of the network and suggests future work. The appendix contains the parameters used in the simulations.

2. Network architecture and operation

The EG neural network architecture, introduced in the preceding section, is based on the layered organization of the visual cortex. The inter- and intralayer connectivity patterns used in the model may correspond to the feedforward and feedback excitatory connections and lateral inhibitory connections found in the visual cortex. The different layers in the model are identified with cortical areas based on similarity in their functions. Thus, the abstract model may have a natural correspondence with parts of animal visual cortex.

The EG neural network is hierarchically organized into three layers. Layer 1 is the input layer; it receives input signals that have been preprocessed through several stages of early visual processing. Layer 1 neurons send activations to layer 2 neurons, which in turn serve as inputs to layer 3 neurons. The layer 3 to layer 2 feedback connections represent a novel feature of the network. In addition, there are lateral inhibitory connections within layers 2 and 3.

This section describes the excitatory and inhibitory connection patterns in the network and how these produce the desired functional properties of disparity selectivity, uniqueness and grouping required for stereomatching.

2.1. Network architecture

The model is organized into three layers of neurons. Layer 1 neurons are sensitive to *monocular* image features, i.e. layer 1 neurons respond to some image features in one eye's image and do not respond at all to the other eye's image. These neurons can be arranged in alternating ocular dominance 'stripes' (figure 3) similar to area V1 of monkey visual cortex (Hubel and Wiesel 1977). The features could be spots (corresponding to the outputs of centre/surround neurons), edges (corresponding to the outputs of oriented simple cells) (Hubel and Wiesel 1977), or other visual elements.

The layer 2 neurons are *binocular*. They receive feedforward excitatory input connections from both left- and right-eye sensitive layer 1 neurons. Their input connections have a fixed spatial weighting for each image; they are excited selectively by the disparity of a match between potentially corresponding left- and right-eye visual features. The layer 2 neurons are thus analogous to the 'tuned' binocular neurons of monkey areas V1 and V2 (Poggio and Fischer 1977) and the 'tuned' binocular neurons of cat areas 17 and 18 (Bishop and Pettigrew 1986, DeAngelis *et al* 1991, Ferster 1981, Ohzawa *et al* 1990).

The neurons in layer 3 receive feedforward excitatory input connections from many layer 2 neurons. Their input connections have a fixed spatial weighting that corresponds to a 'fuzzy' 2D planar local *surface patch* in 3D. Layer 3 neurons are thus tuned to position, depth and 3D orientation of local surface patches defined by the binocular matches across layer 2. These model neurons may correspond to certain neurons in a form-sensitive area such as MT (Bradley and Andersen 1995). Several studies have suggested that surface representation may be formed in early stages of visual processing (He and Nakayama 1994, Nakayama and Shimojo 1992, Nakayama *et al* 1989, Shimojo *et al* 1989).

Lateral inhibitory connections within layer 2 and layer 3 are organized according to a common-input principle (Marshall 1995a). The weights of the lateral inhibitory connections between any pair of neurons depend upon the degree to which their feedforward excitatory input connections project from the same source neurons. That is, neurons with common inputs inhibit each other.

Finally, there are feedback excitatory connections from layer 3 neurons to layer 2 neurons. The weights of these connections are approximately symmetric with those of their reciprocal layer 2 to layer 3 connections.

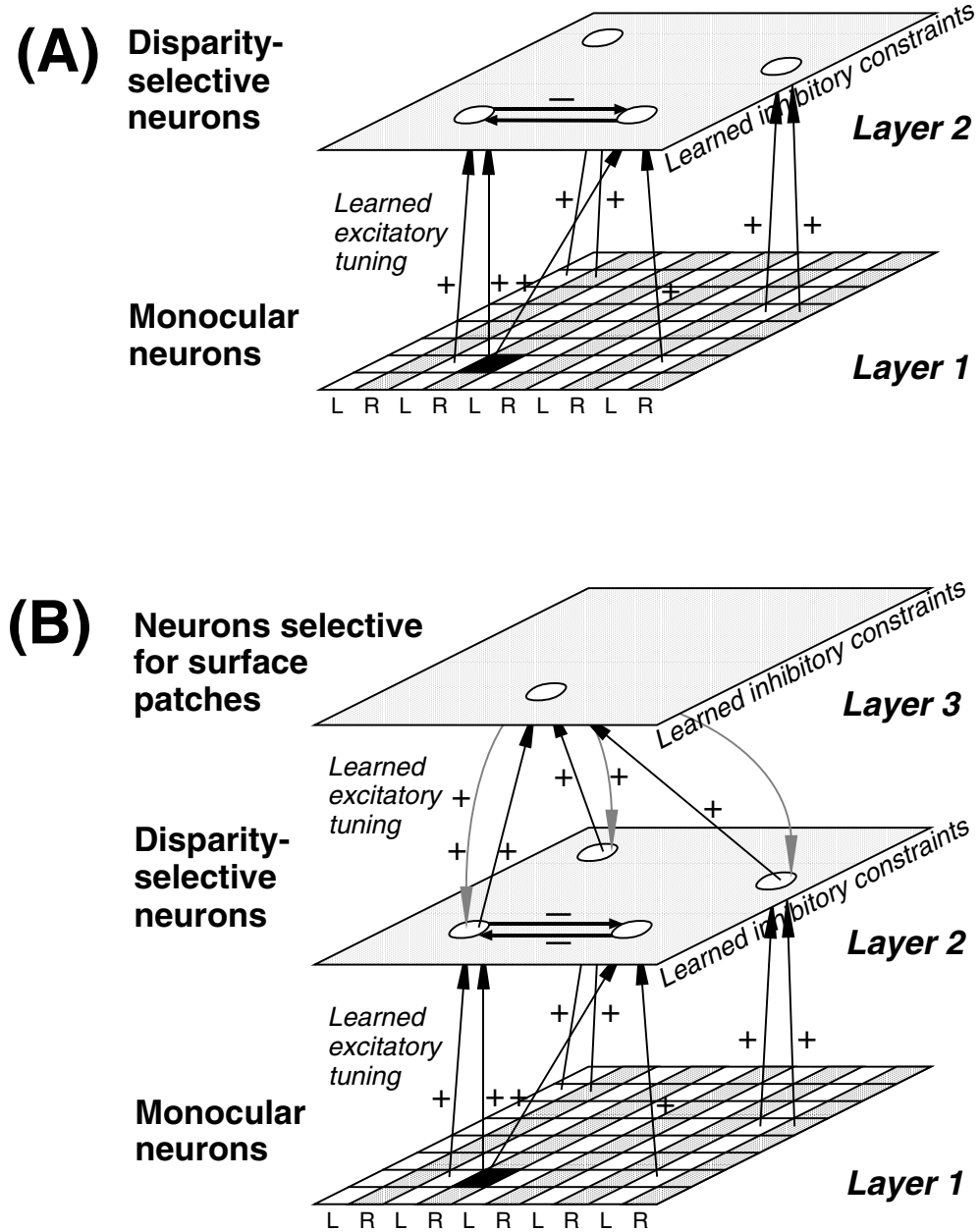

Figure 3. (A) Stereo disparity constraint network. Layer 1 contains input neurons with monocular receptive fields, arranged retinotopically in alternating left–right ocular dominance columns. These neurons project excitatory connections to neurons in layer 2, so that each layer 2 neuron is tuned for stimuli at a particular position and disparity. (B) Stereo grouping network. Each layer 2 neuron is tuned for stimuli at a particular position and disparity. Neurons in layer 3 are selective for correlated combinations of neuron activations across layer 2, i.e. for local surface patch information.

2.2. The operation of the layers

The critical features of the EG stereomatching algorithm are the anisotropic lateral inhibition within layer 2 and the feedback excitatory connections from layer 3 to layer 2. The anisotropic lateral inhibition is based on a common-input principle. The layer 1 to layer 2 excitatory connections and the anisotropic lateral inhibitory connections within layer 2 form the basis for the disparity selectivity and the uniqueness constraint in layer 2. The feedforward–feedback loop between layers 2 and 3 groups input features to represent coherent surface patches and resolves ambiguous matches using global information.

This section describes how intralayer anisotropic lateral inhibition and layer 2 to layer 3 feedback operate to implement disparity selectivity, the uniqueness constraints and grouping. Thus, physiology of the visual cortex is related to their possible function in the context of stereomatching.

2.2.1. Layers 1 and 2: disparity selectivity and uniqueness. The operation of layers 1 and 2 (figure 3) is similar to that of the Marr–Poggio stereomatching algorithm. The feedforward excitatory connections from layer 1 to layer 2 determine binocular disparity tuning; each layer 2 neuron responds selectively to its preferred image position and disparity. A single binocular feature stimulus at a given disparity *half excites* many neurons whose left or right receptive fields include the stimulated region, and *fully* excites only the single neuron (or single population of neurons) whose left and right receptive fields both include the stimulated region.

The circuitry in layer 2 is similar to the inhibitory uniqueness-constraint circuitry of the Marr–Poggio algorithm. Unlike the Marr–Poggio algorithm, however, there are no lateral excitatory connections within layer 2.

All layer 2 neurons receiving input from a common layer 1 source neuron project inhibitory connections to one another. These layer 2 neurons represent the potential stereomatches for a stimulus in the layer 1 neuron's receptive field. Thus, the lateral inhibitory connections within layer 2 implement a uniqueness constraint (Marr and Poggio 1976, Marshall 1992). A visual feature in one eye's image can be matched with at most one visual feature in the other eye's image.

Layer 2 neurons that do not receive input from common layer 1 source neurons project relatively weak lateral inhibitory connections to one another. Thus, these particular combinations of neurons can be simultaneously active. This characteristic is what allows the EG network to represent multiple transparently overlaid surfaces simultaneously within any image region.

Figure 3(A) depicts four of the (many) layer 2 neurons. For illustration purposes, three of these neurons have similar disparity sensitivities (the narrow angle between their left and right layer 1 input arrows), and the fourth is tuned to a larger disparity (broader angle between its two input arrows). Within layer 2, neurons project reciprocal lateral inhibitory connections to one another. The inhibitory connection weights within layer 2 are strong only between neurons that receive common layer 1 input. Two such neurons are shown in layer 2 as receiving strong excitatory input connections from one layer 1 neuron (solid black); their strong reciprocal inhibition is shown as horizontal arrows (−). Between neurons that do not receive common layer 1 input, the inhibitory connections are weak or non-existent. The strong inhibitory connections implement a uniqueness constraint, because representations of possible correspondence pairings of a given visual feature in one eye's image with features in the other eye's image cannot become active simultaneously. The figure describes the excitatory and inhibitory connections as 'learned' because of the self-organization properties discussed in section 3.

2.2.2. Layers 2 and 3: grouping and local surface support. Instead of lateral excitatory connections within layer 2, this neural network has an excitatory feedback loop between layers 2 and 3. When a locally planar surface is presented, the features on the surface excite layer 2 neurons representing the possible stereomatchings. Layer 3 neurons receive excitatory inputs from layer 2 neurons that together correspond to a particular local planar position, depth, slant and tilt. Thus, layer 3 neurons gather evidence for the presence of surfaces (figure 3(B)).

The layer 3 neuron (or group of neurons) whose tuning most closely corresponds to that of the input surface(s) becomes most active. That neuron (or group of neurons) feeds back excitation reciprocally to layer 2 neurons that correspond to the stereomatches which lie on the surface. Thus, the feedback connections from layer 3 to layer 2 add excitatory local neighbourhood support information to the activation of selected layer 2 neurons, in proportion to the strength of the surface on which the feature may lie. This extra feedback excitation favours representations of coherent surfaces over representations of isolated feature matches. The feedback excitation is useful when the point-by-point matching information for a visual feature is ambiguous (figure 3(B)).

In figure 3(B), a layer 3 neuron is depicted that receives strong excitatory connections from several layer 2 neurons, all of which are tuned to a particular disparity, at different image locations; it is therefore selective for a frontoparallel patch at that disparity. That layer 3 neuron does not receive strong excitatory input from the fourth depicted layer 2 neuron, because its preferred disparity lies off the plane defined by the disparities of the other three layer 2 neurons. Many surface patches, at different 3D orientations, disparities and positions, can be represented by the different layer 3 neurons. The excitatory feedback connections from layer 3 to layer 2 project reciprocally, from each layer 3 neuron to those layer 2 neurons that project strong feedforward connections to it. The feedback connections allow the pooled surface patch information at layer 3 to influence the choice of active layer 2 neurons; a local-support mechanism is thereby provided.

Care is taken in the network design (see section 4 below) to prevent uncontrolled excitation from occurring as a result of the positive gain of the layer 2↔layer 3 loop. Unlike the Marr–Poggio algorithm and the Hopfield (1982) neural network, this neural network does not allow the feedback (or lateral) excitation alone to activate the layer 2 neurons. The feedback is allowed only to amplify the feedforward excitation from layer 1 (Hirsch and Gilbert 1991, Marshall 1990a). This allows the network to represent the presence of multiple transparently overlaid surfaces simultaneously via simultaneous activations of multiple neurons within a spatial region.

Thus, feedback from layer 3 to layer 2 provides a flexible and effective implementation of the continuity constraint. Furthermore, the top-down feedback resolves ambiguous matches and represents oriented surface patches in 3D. The selective lateral inhibitory constraint network (uniqueness constraint) within layer 2 ensures that all pairings are globally consistent.

2.3. Stereomatching operation

The preceding sections presented the operation and functions of the individual network layers. This section summarizes how the interactions among the layers lead to stereomatching of the input images.

Visual stimulation of each eye gives rise to retinotopically mapped patterns of activation across layer 1. The layer 2 neurons that represent potential binocular matches then receive the strongest feedforward excitation. The layer 2 neurons that represent conflicting matches

send lateral inhibitory signals to one another. The pattern of activation across layer 2 causes the layer 3 neurons that best represent the surface patches in the input to become strongly active. The active layer 3 neurons then send feedback excitation to the layer 2 neurons that represent matches which lie on the surface patches represented by the layer 3 neurons. This feedback biases the lateral competition in which layer 2 neurons are engaged, in favour of matches consistent with the surface patches represented by the layer 3 neurons. The system operates in continuous time, so the best surface patches become increasingly well represented by layer 3 activation, subject to the matching constraints within layer 2, until equilibrium is reached. The final result is activation of the layer 3 neurons that best represent the individual feature matches on those surface patches. Each individual feature match is exclusively allocated (Bregman 1990, Marshall 1995a) to a surface patch.

The selective inhibition in layer 2 allows non-conflicting feature representations to be active simultaneously, even across the same image region, as in transparency, unlike the Marr–Poggio algorithm. It also assigns unique disparity values to the visual features, unlike the Prazdny algorithm. Furthermore, the network does not necessarily impose a frontoparallel bias, unlike both the Marr–Poggio and Prazdny algorithms.

The layer 2↔layer 3 feedback loop implements a form of *grouping* or *binding*. First, the loop collects evidence for all the possible surface groupings. Second, it uses the grouping evidence to resolve ambiguities in the choice of representation for individual visual features. The binding of monocular features into binocular stereomatches and surface patches is mediated by the exclusive allocation behaviour of layer 2 and is controlled by the cumulative support functions of layer 3.

3. Simulation implementation details

The previous sections discussed the motivation for the EG network by analysing previous algorithms and biological constraints. These constraints guide the hard wiring of the network.

Here we present the hard wiring of the connection weights between the neurons (section 3.1), the equations that control the neuron activations (section 3.2), and the construction of the input RDSs (section 3.3).

3.1. Hardwiring

This section presents the formulae used to assign weights to the connections in the hard-wired prototype network.

Let N_{1X} be the number of layer 1 monocular neurons selective to either left or right images at different positions along a horizontal epipolar line and N_{1Y} be the number of horizontal epipolar lines for which the layer 1 neurons are selective. Thus, there are $N_1 = N_{1X} \times N_{1Y}$ neurons in layer 1.

In layer 2, let there be N_{2X} binocular neurons sensitive to different positions along each of the N_{2Y} horizontal epipolar lines. Each layer 2 neuron is also sensitive to one of N_{2D} disparity values. Thus, the number of layer 2 neurons, N_2, is equal to $N_{2X} \times N_{2Y} \times N_{2D}$.

Each layer 3 neuron is selective for position, depth, slant and tilt of local planar patches in 3D space. The network can represent N_{3X} positions along each of the N_{3Y} horizontal scan lines, N_{3Z} depths, N_{3S} slants and N_{3T} tilts. Thus, the number of layer 3 neurons, N_3, is $N_{3X} \times N_{3Y} \times N_{3Z} \times N_{3S} \times N_{3T}$.

The neurons in layer 1 are assigned unique integers from 0 to $N_1 - 1$, those in layer 2 are assigned numbers from N_1 to $N_1 + N_2 - 1$, and those in layer 3 are given indices from

$N_1 + N_2$ to $N_1 + N_2 + N_3 - 1$.

A layer 1 neuron i is associated with position (X_i, Y_i), which corresponds to the centroid of its receptive field, and with ocularity LR_i. A layer 2 neuron i is associated with position (X_i, Y_i), which corresponds to the centroid of its left- and right-eye receptive fields, and with disparity D_i, its disparity selectivity. A layer 3 neuron i is associated with position (X_i, Y_i), depth Z_i, slant S_i and tilt T_i; these denote the surface patch to which the layer 3 neuron i is maximally sensitive.

A neuron with index i is assigned coordinate values (X_i, Y_i), where

$$X_i = \begin{cases} \left\lfloor \dfrac{i \bmod N_{1\mathrm{X}}}{2} \right\rfloor \mathrm{scale}_{1\mathrm{X}} + \mathrm{base}_{1\mathrm{X}} & 0 \leqslant i < N_1 \\ \left\lfloor \dfrac{i - N_1}{N_{2\mathrm{Y}} \times N_{2\mathrm{D}}} \right\rfloor \mathrm{scale}_{2\mathrm{X}} + \mathrm{base}_{2\mathrm{X}} & N_1 \leqslant i < N_1 + N_2 \\ \left\lfloor \dfrac{i - (N_1 + N_2)}{N_{3\mathrm{Z}} \times N_{3\mathrm{S}} \times N_{3\mathrm{T}} \times N_{3\mathrm{Y}}} \right\rfloor \mathrm{scale}_{3\mathrm{X}} + \mathrm{base}_{3\mathrm{X}} & N_1 + N_2 \leqslant i < N_1{+}N_2{+}N_3 \end{cases} \tag{1}$$

and

$$Y_i = \begin{cases} \lfloor i / N_{1\mathrm{X}} \rfloor \mathrm{scale}_{1\mathrm{Y}} + \mathrm{base}_{1\mathrm{Y}} & 0 \leqslant i < N_1 \\ \left(\left\lfloor \dfrac{i - N_1}{N_{2\mathrm{D}}} \right\rfloor \bmod N_{2\mathrm{Y}} \right) \mathrm{scale}_{2\mathrm{Y}} + \mathrm{base}_{2\mathrm{Y}} & N_1 \leqslant i < N_1 + N_2 \\ \left(\left\lfloor \dfrac{i - (N_1 + N_2)}{N_{3\mathrm{Z}} \times N_{3\mathrm{S}} \times N_{3\mathrm{T}}} \right\rfloor \bmod N_{3\mathrm{Y}} \right) \mathrm{scale}_{3\mathrm{Y}} + \mathrm{base}_{3\mathrm{Y}} & N_1{+}N_2 \leqslant i < N_1{+}N_2{+}N_3. \end{cases} \tag{2}$$

In the above equations, $\mathrm{base}_{1\mathrm{X}}$, $\mathrm{base}_{2\mathrm{X}}$, $\mathrm{base}_{3\mathrm{X}}$, $\mathrm{base}_{1\mathrm{Y}}$, $\mathrm{base}_{2\mathrm{Y}}$, $\mathrm{base}_{3\mathrm{Y}}$, $\mathrm{scale}_{1\mathrm{X}}$, $\mathrm{scale}_{2\mathrm{X}}$, $\mathrm{scale}_{3\mathrm{X}}$, $\mathrm{scale}_{1\mathrm{Y}}$, $\mathrm{scale}_{2\mathrm{Y}}$ and $\mathrm{scale}_{3\mathrm{Y}}$ are constants that are used to align the receptive fields of the neurons in the various layers within the common X–Y coordinate frame. The 'base' constants specify the lower bound for the X and Y coordinates and the 'scale' constants determine the distance between adjacent neurons along each dimension.

Associated with each layer 1 neuron i is the variable LR_i, where

$$\mathrm{LR}_i = \begin{cases} 1 & \text{if } i \text{ is even} \\ -1 & \text{if } i \text{ is odd.} \end{cases} \tag{3}$$

$\mathrm{LR}_i = 1$ means that neuron i is sensitive to left image features, and $\mathrm{LR}_i = -1$ means that neuron i is sensitive to right image features.

Each layer 2 neuron i is selective to a disparity D_i, where

$$D_i = ((i - N_1) \bmod N_{2\mathrm{D}}) \, \mathrm{scale}_{2\mathrm{D}} + \mathrm{base}_{2\mathrm{D}} \tag{4}$$

and $\mathrm{base}_{2\mathrm{D}}$ and $\mathrm{scale}_{2\mathrm{D}}$ are constants that determine the lower bound of disparity values and 'distance' between adjacent neurons along this dimension, respectively.

The depth represented by layer 3 neuron i, Z_i, is computed by

$$Z_i = \left(\left\lfloor \frac{i - (N_1 + N_2)}{N_{3\mathrm{S}} \times N_{3\mathrm{T}}} \right\rfloor \bmod N_{3\mathrm{Z}} \right) \mathrm{scale}_{3\mathrm{Z}} + \mathrm{base}_{3\mathrm{Z}} \tag{5}$$

where $\mathrm{base}_{3\mathrm{Z}}$ and $\mathrm{scale}_{3\mathrm{Z}}$ are constants that determine the lower bound of depth values and 'distance' between adjacent neurons along this dimension, respectively.

Associated with each neuron i in layer 3 are the slant index $S_i' \in \{0, \ldots, N_{3\mathrm{S}} - 1\}$ and the tilt index $T_i' \in \{0, \ldots, N_{3\mathrm{T}} - 1\}$, where

$$S_i' = (i - (N_1 + N_2)) \bmod N_{3\mathrm{S}} \tag{6}$$

and

$$T_i' = \left\lfloor \frac{i - (N_1 + N_2)}{N_{3S}} \right\rfloor \bmod N_{3T}. \tag{7}$$

Each pair in $\{0, \ldots, N_{3S} - 1\} \times \{0, \ldots, N_{3T} - 1\}$ is associated with a slant–tilt pair. Let the pair of actual slant–tilt values associated with the slant–tilt indices (S_i', T_i') be (S_i, T_i).

3.1.1. Feedforward excitatory connection weights. Monocular layer 1 neurons with horizontally displaced receptive fields in the two eyes send feedforward excitatory connections to binocular layer 2 neurons. The sign and magnitude of the horizontal displacements characterize the disparity selectivity of the layer 2 neurons.

Layer 2 neurons with receptive fields at corresponding positions are selective for zero disparity. Layer 2 neurons with positive (negative) disparity selectivity have their left-eye receptive field horizontally displaced to the left (right) of the left-eye receptive field of the zero-disparity neuron at that position, while their right-eye receptive field is horizontally displaced to the right (left) of the right-eye receptive field of the zero-disparity neuron at that position (figure 4).

The connection weights from left- and right-eye selective neurons in layer 1 to layer 2 neurons have a 2D anisotropic Gaussian profile along the 2D space of the positions of the retinotopically arranged layer 1 neurons. The connection strength tapers off more rapidly along the Y direction than along the X direction. Let i be a layer 1 neuron, $0 \leqslant i < N_1$ and j a layer 2 neuron, $N_1 \leqslant j < N_1 + N_2$. Then,

$$W_{ij} = \exp\left(-\left(\sigma_y^{(1)}(Y_i - Y_j)^2 + \sigma_x^{(1)}(X_i + \mathrm{LR}_i \times D_j - X_j)^2 + \mathrm{bias} \times \mathrm{abs}\,(D_j)\right)\right) \tag{8}$$

where the constants $\sigma_x^{(1)}$ and $\sigma_y^{(1)}$ determine the spread of the 2D Gaussian connection strengths along X and Y directions, respectively. The constant 'bias' biases the network towards zero disparity.

The feedforward excitatory connection from layer 1 neuron i to layer 2 neuron j, z_{ij}^{+12}, is given by

$$z_{ij}^{+12} = \begin{cases} W_{ij} & \text{if } W_{ij} > \theta_e^{12} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where θ_e^{12} is a threshold parameter.

Planar patches in 3D visual space after processing by layer 1 neurons are mapped into a 3D space of X position, Y position and disparity in layer 2. Layer 3 neurons are connected to layer 2 neurons in such a way that the pattern connectivity looks like a 'pancake' in the 3D space of X position, Y position and disparity. The pancake shape is obtained using an ellipsoidal Gaussian connection weight profile.

In the simulated network, the connection weights between layer 2 and layer 3 neurons are symmetric. The excitatory top-down feedback from layer 3 neuron j to layer 2 neuron i, z_{ij}^{+32}, is equal to z_{ji}^{+23}. Layer 3 neurons are selective for X and Y position, depth, slant and tilt (figures 5 and 6).

To obtain different slants and tilts, each pancake has different parameters. Let i be a layer 2 neuron and j be a layer 3 neuron, i.e. $N_1 \leqslant i < N_1 + N_2$ and $N_1 + N_2 \leqslant j < N_1 + N_2 + N_3$. Then

$$e_1 = \sigma_z^{(2)}\left(-\sin(T_i)\sin(S_i)(X_j - X_i) + \cos(T_i)\sin(S_i)(Y_j - Y_i) + \cos(S_i)(D_j - Z_i)\right) \tag{10}$$

$$e_2 = \sigma_y^{(2)}\left(\cos(T_i)(X_j - X_i) + \sin(T_i)(Y_j - Y_i)\right) \tag{11}$$

$$e_3 = \sigma_x^{(2)}\left(\sin(T_i)\cos(S_i)(X_j - X_i) + \cos(T_i)\cos(S_i)(Y_j - Y_i) + \sin(S_i)(D_j - Z_i)\right) \tag{12}$$

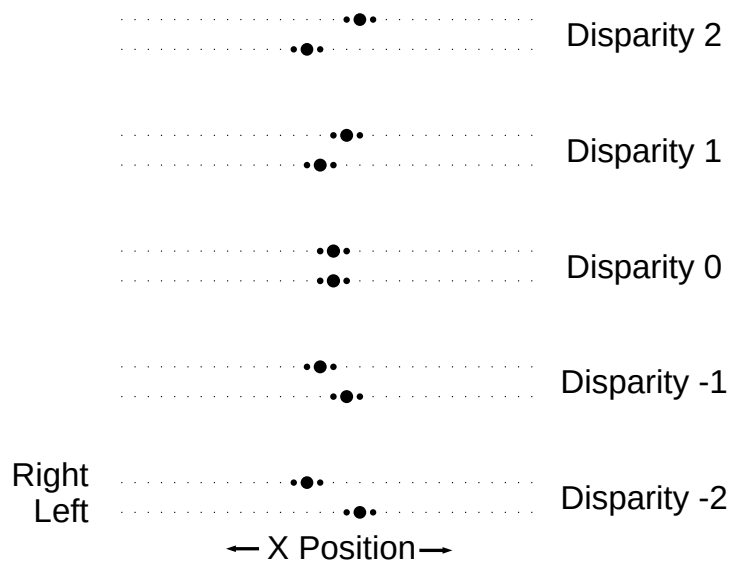

Figure 4. Layer 1 to layer 2 connection weights. The figure shows the layer 1 to layer 2 connectivity patterns to obtain disparity selectivity in layer 2 neurons. The figure depicts the connection weights from the monocular layer 1 neurons sensitive to left and right eyes at a Y position to layer 2 neurons at the same Y position. The layer 2 neurons shown have the same X and Y positions, but their disparity selectivity ranges from -2 to $+2$. The strength of the connection weights is represented by size of the dark dot. The connection weights marked by tiny light dots are zero.

and

$$W_{ij} = \exp\left(-(e_1^2 + e_2^2 + e_3^2)\right) \tag{13}$$

where the constants $\sigma_x^{(2)}$, $\sigma_y^{(2)}$ and $\sigma_z^{(2)}$ determine the spread of the 2D Gaussian connection weights along the X, Y and disparity axes, respectively. The weight of the connection from neuron i to neuron j, z_{ij}^{+23}, is

$$z_{ij}^{+23} = \begin{cases} W_{ij} & \text{if } W_{ij} > \theta_{\mathrm{e}}^{23} \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

where θ_{e}^{23} is a threshold parameter.

3.1.2. Lateral inhibitory connection weights. The formula used to assign lateral inhibitory weights between neurons follows a 'common-input' principle: lateral inhibitory weights between neurons are approximately proportional to the amount of overlap in the neurons' feedforward input. The rationale for this method is discussed in section 5.3.

Let neurons i and j be distinct and belong to the same layer l, where l is 2 or 3 and

$$W_{ij} = \sum_{k \in \text{layer}(l-1)} \min\left(z_{ki}^{+(l-1,l)}, z_{kj}^{+(l-1,l)}\right). \tag{15}$$

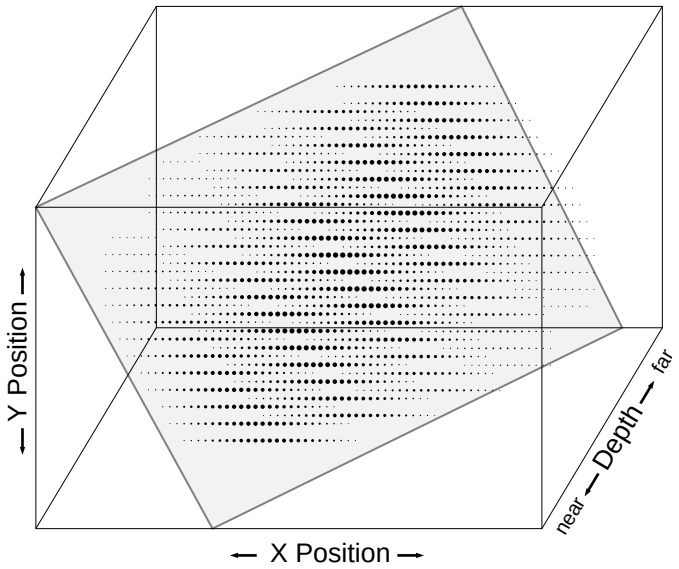

Figure 5. Layer 2↔layer 3 weights for representing surface patch with slant and tilt. The connectivity pattern between layer 2 neurons and a layer 3 neuron that represents a slanted and tilted surface in depth is shown. Layer 2 neurons selective for X and Y positions and disparity are mapped into the 3D space of X and Y positions and depth. Positive (negative) disparities are mapped to near (far) depths. The layer 3 neuron is represented by the shaded polygon. The coordinates of the polygon centre represent the X and Y position and depth selectivity of the layer 3 neuron, respectively. The slant and tilt of the polygon represent the slant and tilt selectivity of the layer 3 neuron, respectively. The magnitude of the connection strength is represented by the size of the dots. The dot placement within the 3D space of X and Y position and depth produces a 'pancake' shape. Layer 2 neurons with no connections to the layer 3 neuron are not shown.

Then, the lateral inhibitory connection weights, $z_{ij}^{-(l,l)}$ and $z_{ji}^{-(l,l)}$, between neurons i and j, are

$$z_{ij}^{-(l,l)} = z_{ji}^{-(l,l)} = \begin{cases} W_{ij} & \text{if } W_{ij} > \theta_i^{(l,l)} \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

The lateral inhibitory connection patterns for layers 2 and 3 are shown in figures 7 and 8, respectively.

The amount of lateral inhibition received by a layer 2 neuron decreases with distance. The lateral inhibition profile of a neuron is broader than the feedforward connection weight pattern. The Gaussian pattern of feedforward connection weights, in conjunction with Gaussian lateral inhibition centred at the same position, produces layer 2 neuron receptive fields that can be described as a difference of Gaussians.

3.2. The excitation, inhibition and activation rules

The excitation, inhibition and activation rules that govern the behaviour of the neurons in the different layers are described in this section.

The 'bottom-up' (feedforward) excitation received by the neuron j in layer l from

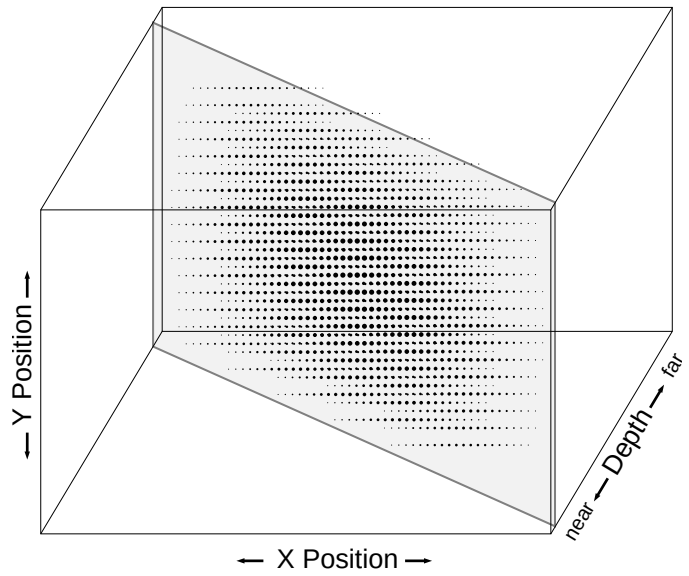

Figure 6. Layer 2↔layer 3 weights for representing surface patch with tilt. The figure depicts the connectivity pattern between layer 2 neurons and a layer 3 neuron that represents a tilted surface in depth. In the simulated network, the connection weights between layer 2 and layer 3 neurons are symmetric. See figure 5 for conventions.

neurons in layer $l-1$, $B_j^{(l)}$, is determined by

$$B_j^{(l)} = F\left(\sum_i z_{ij}^{+(l-1,l)} G\left(x_i^{(l-1)}\right)\right) \tag{17}$$

where $l \in \{2,3\}$, $z_{ij}^{+(l-1,l)}$ is the non-negative weight of the feedforward excitatory connection from the layer $(l-1)$ neuron i to the layer l neuron j and $x_i^{(l-1)}$ is the activation level of layer $(l-1)$ neuron i. The functions $F(\cdot)$ and $G(\cdot)$ are non-decreasing, non-negative functions.

The top-down facilitatory feedback from layer $(l+1)$ neurons to layer l neuron j is

$$T_j^{(l)} = I\left(\sum_i z_{ij}^{+(l+1,l)} H\left(x_i^{(l+1)}\right)\right) \tag{18}$$

where $z_{ij}^{+(l+1,l)}$ is the weight of the feedback excitatory connection between layer $(l+1)$ neuron i and layer l neuron j and $x_i^{(l+1)}$ is the activation level of layer $(l+1)$ neuron i. The functions $I(\cdot)$ and $H(\cdot)$ are non-decreasing, non-negative functions. In the simulated network, only layer 2 receives feedback from layer 3, i.e. in the above equation l is restricted to 2.

The bottom-up feedforward excitation and the top-down feedback excitation to layer l neuron j are combined to obtain the net excitation $E_j^{(l)}$ received by layer l neuron j, where

$$E_j^{(l)} = B_j^{(l)}(1 + \Gamma \times T_j^{(l)}). \tag{19}$$

In the above equation l is 2. $\Gamma \geqslant 0$ is a constant that controls the amount of feedback to layer 2 neurons. The feedback is shunted by the bottom-up excitation. Thus, the equation allows selective application of top-down feedback to layer 2 neurons.

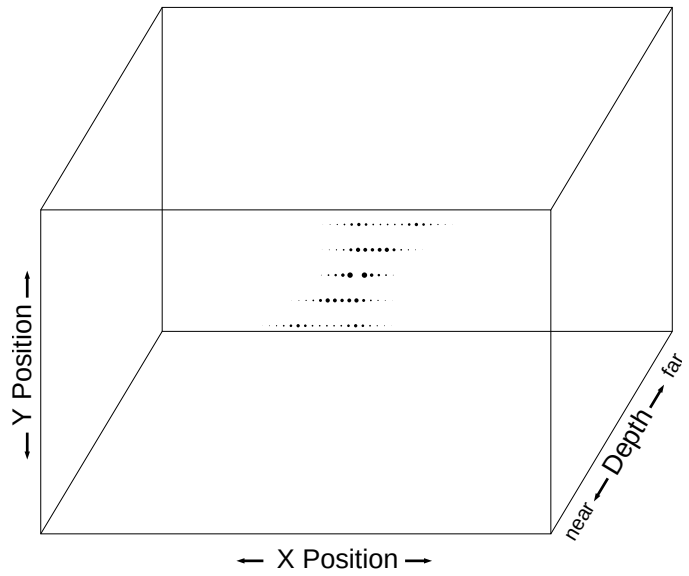

Figure 7. Lateral inhibitory connections within layer 2. Layer 2 neurons are selective for X and Y positions and disparity. In the figure, layer 2 neurons are arranged in the 3D space of X and Y positions and depth. Layer 2 neurons with positive (negative) disparities are mapped to near (far) depths. The reciprocal inhibitory connection weights shown are between the neuron selective to $X = 16$, $Y = 7$ and zero disparity and other layer 2 neurons selective to $Y = 7$. The dot size is proportional to the connection weight. Neurons do not inhibit themselves in the model. The connection weights are proportional to the size of feedforward weight from input neurons common to the layer 2 neurons.

The total inhibitory input to the jth layer l neuron is

$$I_j^{(l)} = K\left(\sum_{i\neq j} z_{ij}^{-(l,l)} J\left(x_i^{(l)}\right)\right). \tag{20}$$

In the above equation, $z_{ij}^{-(l,l)}$ is the non-negative weight of the lateral inhibitory connection from layer l neuron i to layer l neuron j, where $x_i^{(l)}$ is the activation level of the ith layer l neuron, $l \in \{2, 3\}$ and $i \neq j$. The functions $J(\cdot)$ and $K(\cdot)$ are non-decreasing, non-negative functions.

Activation changes for layer 2 and layer 3 neurons are governed by a shunting equation (Grossberg 1972, 1982):

$$\frac{\mathrm{d}}{\mathrm{d}t}x_i^{(l)} = \epsilon\left(-Ax_i^{(l)} + \beta\left(B - x_i^{(l)}\right)E_i^{(l)} - \gamma\left(C + x_i^{(l)}\right)I_i^{(l)}\right) \tag{21}$$

where A, B, C, β, γ and ϵ are constants. Because (21) is a shunting equation, the layer 2 and layer 3 neuron activation levels are forced to remain within a bounded range, between $-C$ and B. The constant ϵ controls the rate at which layer l neuron activations tend towards their equilibrium values.

3.3. The input patterns

The input stereo pairs provide the first constraint on stereomatching. The stereo pairs contain dots that define surfaces in depth. This section describes how the stereo pairs containing

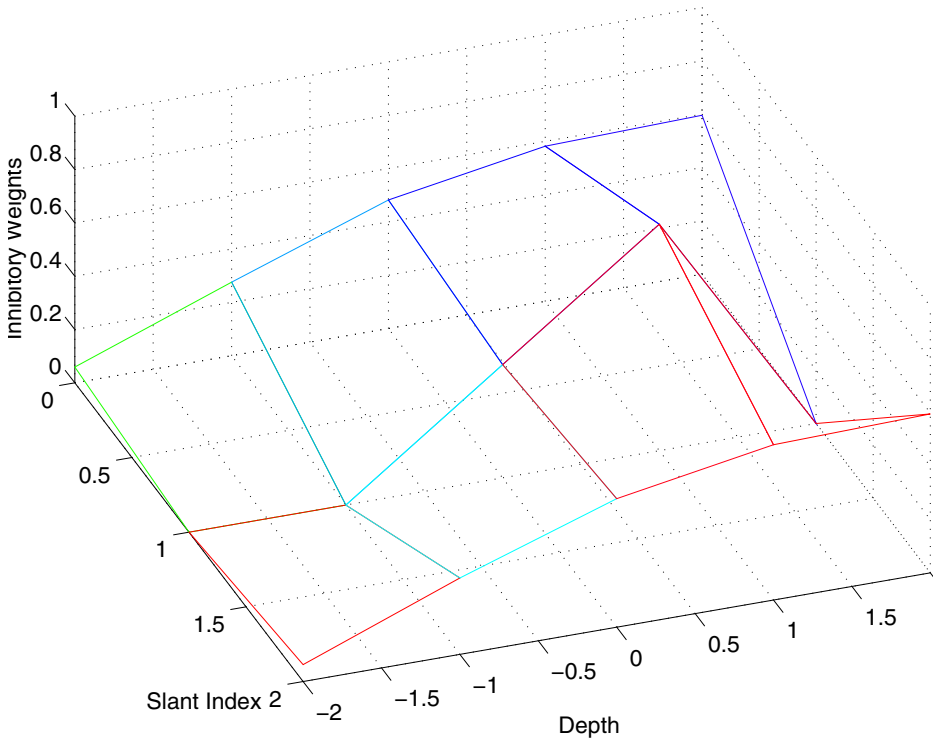

Figure 8. The figure shows the symmetric lateral inhibitory connections within layer 3. The connection weights are between the layer 3 neuron selective to surface at $X = 16$, $Y = 7$, depth $= 2$, slant index $= 2$ and tilt index $= 2$ (a frontoparallel surface at near depth) and other layer 3 neurons sensitive to surfaces at $X = 16$, $Y = 7$, tilt index $= 2$ (surfaces at different depths and rotated about the vertical axis). For a fixed slant index, the inhibitory weight decreases as the difference in the depth represented by the neurons increases. Layer 3 neurons do not inhibit themselves. The inhibitory weights are symmetric with respect to slant index for neurons at the same depth.

information on surfaces at different depths, slants and tilts were constructed.

In the simulated network, a continuous range of input disparities is obtained by distributing partial activations over several layer 1 neurons. For example, consider corresponding layer 1 neurons sensitive to left- and right-image features at horizontal position x on corresponding horizontal epipolar lines, which are labelled n_l and n_r, respectively. Let the neurons which are on the same horizontal epipolar line as n_l and n_r and have receptive fields closest to that of n_1 and n_r be $n1_l$, $n1_r$, $n2_l$ and $n2_r$. The neurons $n1_l$ and $n1_r$ are sensitive to features occurring at position $x - 1$ and $n2_l$ and $n2_r$ are sensitive to features occurring at position $x + 1$. To input a disparity of $+1$ at position x, neurons $n1_l$ and $n2_r$ are made fully active. In order to input disparity $-d$, where $0 \leqslant d \leqslant 1$, $d \times 100\%$ of the maximum activation level is applied to the neurons $n1_r$ and $n2_l$ and $(1 - d) \times 100\%$ of the maximum activation level is applied to neurons n_l and n_r. This method provides the input directly within a continuous range of disparities and positions.

4. Simulation results

Computer simulations were used to test the stereomatching properties of the EG network. The conventions used to present the results are presented below.

The simulated network was tested with inputs that illustrate its salient features. In figures 9–14 and figures 16–20 the input stereograms are shown at the bottom. The graphic at the top of each figure is interpreted as follows. The rectangular box encloses a part of the 3D space of X position, Y position and depth. The grey shade within the rectangular box represents layer 3 neurons sensitive to local surface patches that have the corresponding slant, tilt and spatial position. If a plane is shown within the box then the corresponding layer 3 neuron is active and its activation level is shown by the thickness of the border of the plane. The black dots within the box represent active layer 2 neurons. The size of the dots represents the activation level of the layer 2 neurons. The dots corresponding to the layer 2 neurons are positioned within the rectangular box according to the neurons' X and Y positions and disparity. The depth at which the layer 2 neurons are placed corresponds to the disparity selectivity of the neurons. The results can be verified by cross fusing the input stereograms.

The activation levels across figures 9–14 and figures 16–20 are represented with respect to a common maximum level of activation.

4.1. A frontoparallel surface

This example illustrates the ability of the network simulation to represent frontoparallel surfaces.

The right image of the stereogram in figure 9 is created by displacing all dots in the left image to the right by the same amount, thereby producing a positive disparity of one unit. The network assigns unique disparities to all the dot pairs from the two images and combines the local disparity information via the layer 2↔layer 3 feedback loop. The result is that the network strongly detects the presence of a frontoparallel surface patch, as symbolized by the thick border of the grey plane.

4.2. A surface slanted about the vertical axis

In this example, the input stereogram was created so that along each horizontal epipolar line the disparity of the corresponding dots in the left and right images increased from left to right (figure 10). The X positions are identical for all Y positions in the stereo pair. The simulation results shown in figure 10 indicate that the network has found a surface that fits the stereo image data.

4.3. A surface tilted about the horizontal axis

The input stereogram (figure 11) represents a planar patch that is tilted about the horizontal axis and recedes away in depth in cross fusion. In the stereogram, the disparity along each horizontal axis is constant. However, moving from bottom to top the disparities increase. The simulation results shown in figure 11 indicate that the network has found a surface that fits the stereo image data. The representation for that surface is strongly activated even though some of the individual dot matches lie slightly off the plane, as shown in the figure. This behaviour arises because each layer 3 neuron is sensitive to a pancake-shaped *range* of 3D spatial positions, not just to a single thin plane.

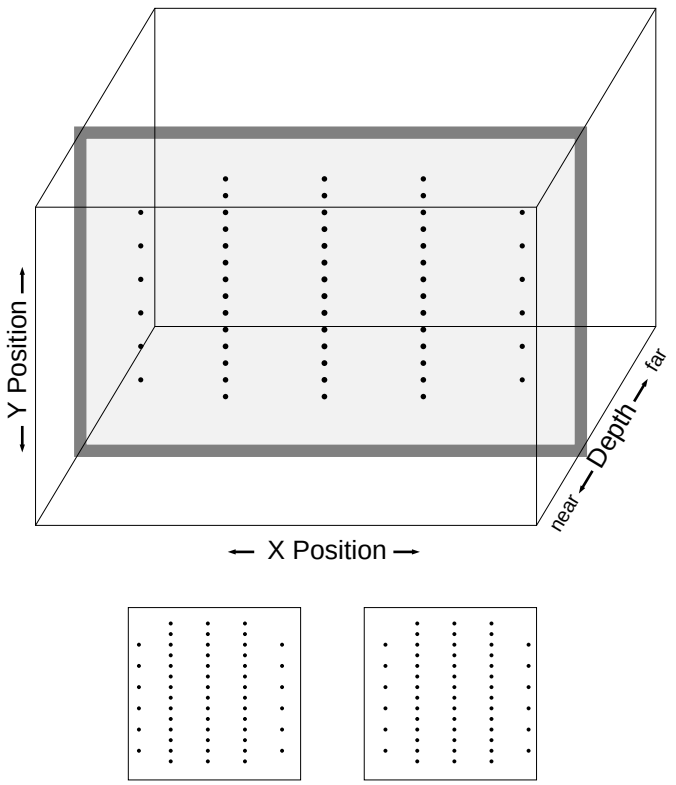

Figure 9. A frontoparallel surface. The input stereogram to the network represents a surface closer to the observer than the background in cross fusion. The rectangular box represents the 3D space of X position, Y position and depth. The grey rectangular plane within the rectangular box represents an active layer 3 neuron. The absence of any other grey planes within the rectangular box implies that no other layer 3 neurons have suprathreshold activation levels. The activation levels of layer 3 neurons are depicted by the border thickness of the grey plane. The dots within the box represent the activation levels of active layer 2 neurons. The dot sizes represent the activation levels of layer 2 neurons. The activation levels represented in the figure are the equilibrium levels. The position and orientation of the grey plane within the box indicate that the active layer 3 neuron represents a frontoparallel surface that is closer to the observer than the background. The plane is parallel to the XY plane and closer to the front face of the box than to its back face.

4.4. A slanted and tilted surface

In this example, the input surface is both slanted and tilted. The stereogram was constructed so that the disparity in corresponding dots in the left and right images increased along the X axis from left to right and along the Y axis from the bottom to the top of the image.

As shown in figure 12, the network simulation represents this surface just as directly as frontoparallel surfaces. It is clear from the figure that the uniqueness constraint is satisfied and that the layer 3 neuron selective to the slant, tilt and depth consistent with the active layer 2 neurons represents the oblique surface. Thus, this network is an improvement over the Marr–Poggio algorithm, which represents oblique surfaces only in approximation by an unstable staircase of frontoparallel surface patches. As in figure 11, the surface in figure 12 is also strongly activated even though some individual dot matches lie slightly off the plane.

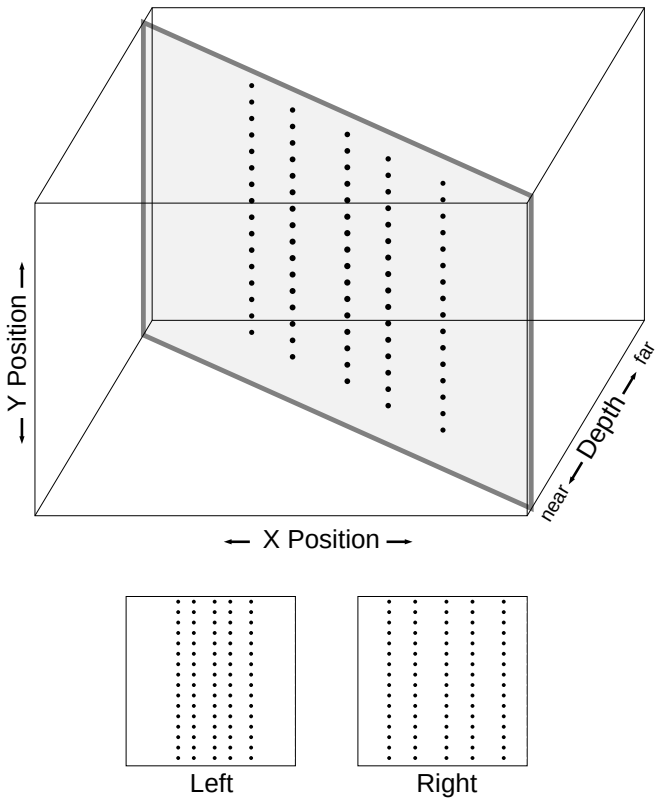

Figure 10. A surface slanted about the vertical axis. The input stereogram to the network represents a surface that is tilted about the vertical when cross fused. This example provides further evidence of the network's ability to represent non-frontoparallel surfaces. See figure 9 for figure conventions.

4.5. Two frontoparallel transparent surfaces

This example illustrates the ability of the network simulation to represent two transparent frontoparallel surfaces.

The input stereogram was constructed by interleaving the two disparities (figure 13). The important points to note in the figure are that every corresponding pair of dots from the left and right image is represented in layer 2 and that two layer 3 neurons are active simultaneously and receive bottom-up excitation from disjoint sets of active layer 2 neurons. The network allows simultaneous layer 3 neuron activations because inhibition between the active layer 3 neurons is weak.

4.6. Two intersecting transparent surfaces

In this example, the input stereogram (figure 14) represents two planar patches slanted about the vertical in different directions and intersecting in the middle. Along each horizontal epipolar line the disparities of the matching pair go from high negative disparity to high positive disparity or vice versa.

The simulation result (figure 14) shows two active layer 3 neurons, each representing one

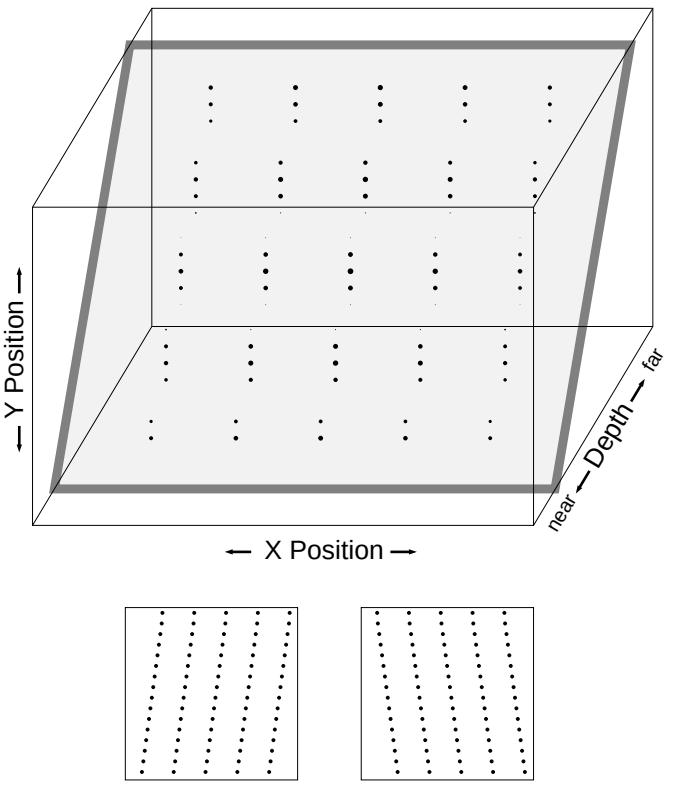

Figure 11. A surface tilted about the horizontal axis. The input stereogram to the network represents a surface that is tilted about the horizontal when cross fused. This example provides further evidence of the network's ability to represent non-frontoparallel surfaces. The conventions used are the same as in figure 9.

of the tilted surfaces. Although the two active layer 3 neurons receive input from common active layer 2 neurons, low inhibition between the two allows simultaneous activations. Figures 13 and 14 clearly demonstrate that the network represents transparent surfaces in depth.

4.7. Ternus configuration

This example shows the ability of the network simulation to prefer globally optimal matches over locally optimal ones.

The stereogram in figure 15 is arranged in a configuration corresponding to an apparent-motion display of Ternus (1926). The right image is created by shifting the four dots in the left image so that the three right dots in the left image and the three left dots in the right image have corresponding positions.

A locally consistent match would pair the three right dots in the left image with the three left dots in the right image. This match would activate neurons in layer 2 sensitive to zero disparity. The leftmost dot in the left image and the rightmost dot in the right image could also be paired (figure 15(A)). This locally consistent matching would provide evidence for the presence of a frontoparallel line segment at the same depth as the background and a dot

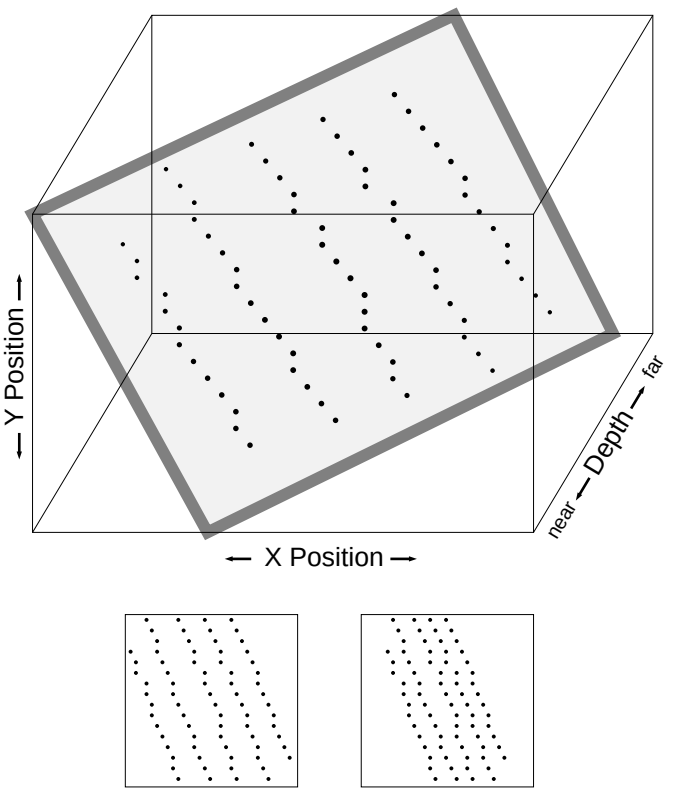

Figure 12. An oblique surface. The input stereogram to the network represents a surface whose centre is at the same depth as the background and slants and tilts away from the observer when cross fused. This example is chosen to demonstrate the ability of the network to represent non-frontoparallel surfaces just as efficiently as frontoparallel surfaces. See figure 9 for details about the figure.

at closer depth. Another locally consistent matching is shown in figure 15(B); this matching has four different disparities that do not define a coherent surface.

A globally consistent match would pair the dots in the order they appear (figure 15(C)). This would cause layer 2 neurons with the same disparity selectivity at different positions to become active and provide strong evidence for the presence of a longer frontoparallel line segment close to the observer (in cross fusion).

Figures 16 and 17 show the simulation behaviour in response to a Ternus-type stereogram consisting of several parallel rows of dots. In the initial stages of processing in the simulation, all possible dot pairings are represented in layer 2 through activation of the corresponding disparity selective neurons (figure 16). However, the layer 2 neuron activation levels are far from their asymptotic values. At this stage, two mechanisms are at play: the lateral inhibition among layer 2 neurons that receive common inputs, and top-down feedback from layer 3 neurons that are consistent with the surface that they represent. The combination of these two processes leads to the result depicted in figure 17.

Figure 16 shows the initial activation levels of a cross section of layer 2 neurons, sliced at $Y = 7$. The figure illustrates that the layer 2 neurons consider all possible pairwise matchings of dots in the input. The neurons with higher activation levels are those that

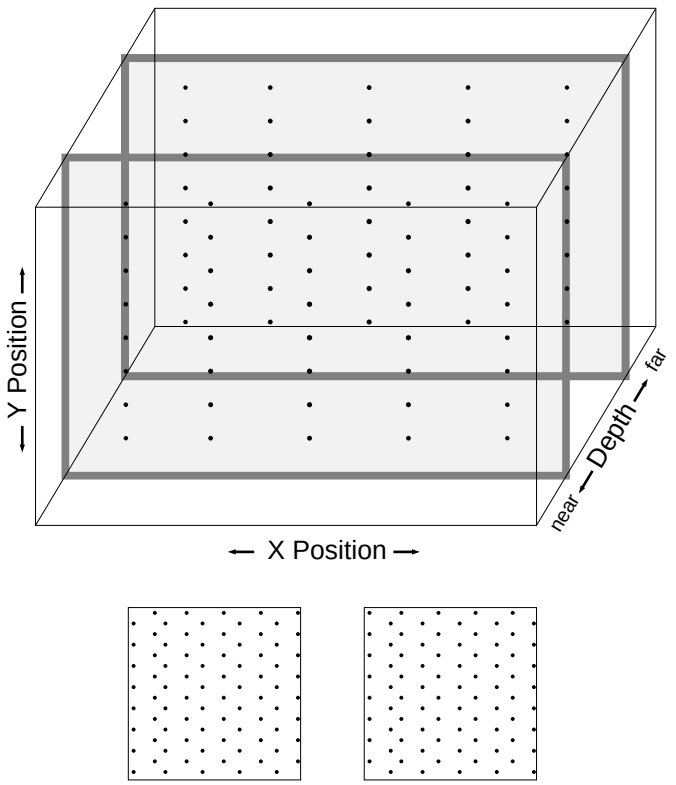

Figure 13. Two frontoparallel transparent surfaces. The input stereogram represents two frontoparallel surfaces, one closer to the observer than the other when cross fused. This simulation demonstrates the ability of the network to represent transparently overlaid surfaces at different depths. See figure 9 for details.

receive excitation from both images, while the neurons with lower activation levels are those that receive excitation from just one image. Layer 2 neurons that have overlapping inputs from layer 1 inhibit one another. The neurons that receive full excitation from both images are successful in inhibiting neurons that receive excitation from only one of the images.

Layer 2 neurons also receive top-down feedback from layer 3 neurons. High activation levels in layer 3 neurons signify the presence of globally consistent local matching by layer 2 neurons. Thus, layer 2 neurons that receive feedback from highly active layer 3 neurons have a competitive edge over layer 2 neurons which receive less feedback and have common layer 1 neuronal inputs with them.

Figure 17 shows the final state of the network simulation. The network performs a locally and globally consistent stereomatching and satisfies the uniqueness constraint within layer 2. The globally consistent surface representation wins over all the other surfaces, because it gives four units of feedback, rather than three or fewer units.

The Prazdny algorithm, unlike the results shown, would represent several surfaces, corresponding to the various possible matches, because it lacks a match disambiguation mechanism.

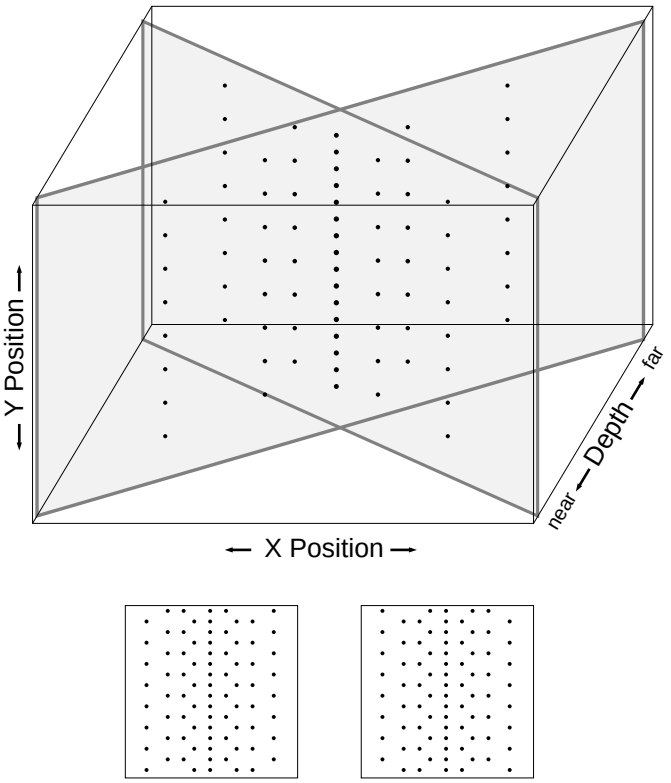

Figure 14. Two intersecting transparent surfaces. The input stereogram to the network represents two transparent surfaces that are tilted about the vertical and intersect in depth. The surfaces appear tilted as shown by the two planes within the rectangular box when cross fused. This examples shows the robustness of the network in handling transparent surfaces in depth. See figure 9 for the conventions used.

4.8. Resolution of ambiguity with layer 3 to layer 2 feedback

The layer 3 to layer 2 feedback imposes a global coherence constraint. The effectiveness of the top-down feedback in implementing the global constraint is described further in this example.

The right image of the stereogram in figures 18 and 19 is created by displacing all dots in the left image to the right by the same amount, thereby producing a positive disparity of two units. To create ambiguous matches, a fourth column of dots is added in the right image. This extra column is the second column in the right image and is at the same X position as the second column in the left image. Thus, the second column of dots in the left image has a strong potential to match the second or the third column in the right image, thereby creating ambiguous local matching.

Figure 18 shows the final response of the stereo network simulation without the layer 3 to layer 2 feedback connections. Without the top-down feedback, the stereomatching in layer 2 obeys only local constraints, imposed by lateral inhibition. In figure 18, the first and the third columns in the left image are matched with the first and fourth columns in the right image, respectively, thereby activating layer 2 neurons selective to $+2$ disparity and a layer 3 neuron selective to a frontoparallel surface at $+2$ (near) depth. Layer 2 neurons

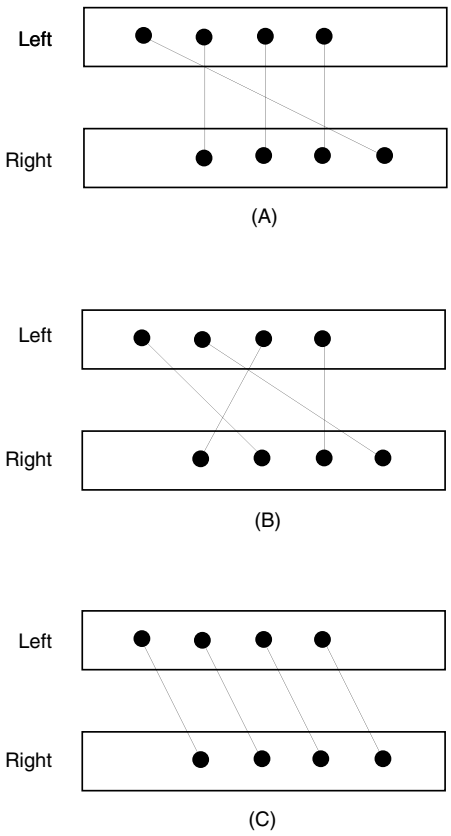

Figure 15. Different matching possibilities in a Ternus configuration. (A) and (B) Locally consistent matches. (C) Globally consistent matches. The left and right input features are represented by the black dots. The lines connecting left- and right-eye dots represent the matches.

selective to zero disparity match the second column in the left image with the third column in the right image. The second column in the right image has no match in the left image and most strongly activates layer 2 neurons selective for zero disparity. The possible matching of the second column in the left image with the third column in the right image is prohibited by the uniqueness constraint. The zero-disparity selective layer 2 neurons and the rightmost column of $+2$ disparity selective neurons activate a layer 3 neuron selective to a slanted surface at zero (intermediate) depth.

With top-down feedback the stereo network disambiguates the matching using global information. The final state of the network with top-down feedback in response to the ambiguous input is shown in figure 19. The only difference between the simulations shown in figures 18 and 19 is that the top-down feedback is enabled in the latter. In response to the input in figure 19, the network initially begins responding to all possible matches. Layer 2 neurons selective to $+2$ disparity are in the majority. This causes the layer 3 neuron selective for the frontoparallel surface at $+2$ depth (near) to be most active in layer 3. In addition, zero-disparity neurons detect the possible match of the second column in the left image with the second column in the right image. The top-down feedback causes the layer 2 neurons selective to $+2$ disparity to become more active and this results in the suppression of zero-disparity selective layer 2 neurons through lateral inhibition. Thus, the second column in the right image is not matched with any column in the left image. However, the second column in the right image partially excites different disparity selective layer 2 neurons selective to its position. Layer 2 neurons selective for $+2$ disparity which receive input from the second

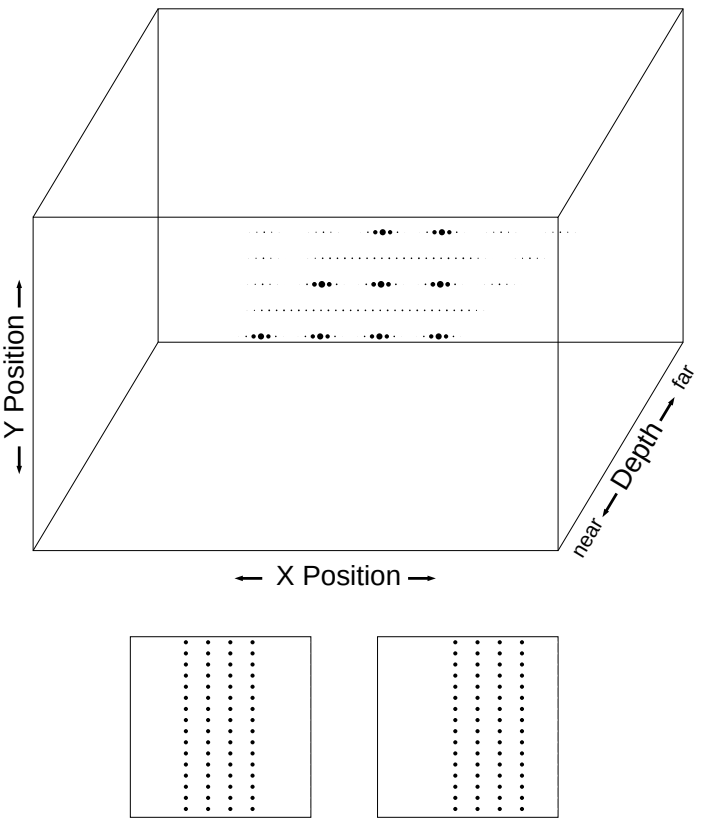

Figure 16. An early state of the network. The input stereogram represents a 'Ternus' surface closer to the observer than the background in cross fusion. The layer 2 neurons at $Y = 7$ are depicted within the rectangle. The activation level of the layer 2 neurons is represented by the size of the black dots. Inactive layer 2 neurons are not shown. The clusters of layer 2 neurons with low activation levels are caused by the overlapping receptive fields of these neurons. Clusters of comparatively high activation levels can be seen at disparity -2, disparity 0 and disparity $+2$.

column in the right image are grouped with the other $+2$ disparity selective neurons and suppress other neurons that were initially excited by the second column in the right image.

4.9. Random cloud of dots

The input stereogram in this example consists of dot pairs randomly placed in 3D (figure 20). The simulation result depicts three layer 3 neurons that are selective to frontoparallel surfaces at different depths. These neurons have very weak, subthreshold activation levels, which are shown by the thin borders of the grey planes. These grey planes illustrate the low relative activation levels within layer 3. The disparity assignments to corresponding pairs of dots are represented within layer 2. In this example, there is no globally consistent stereo pairing in the input. In the absence of a globally consistent matching, the network relies on locally consistent matching. This behaviour of the network is an improvement over the Prazdny algorithm, which would be unable to assign any specific disparities to the dots in the input pair.

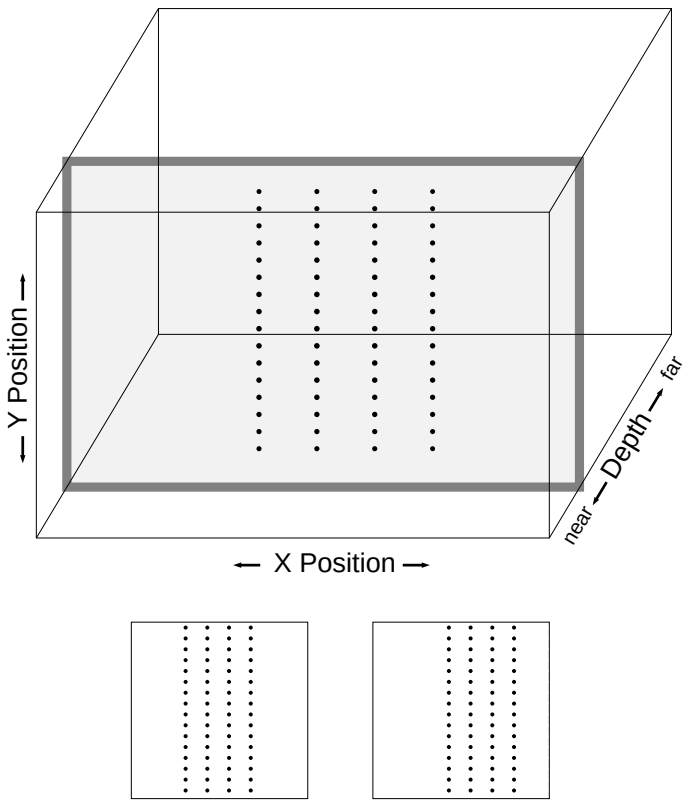

Figure 17. Ternus configuration. The stereogram represents a Ternus configuration. This example demonstrates the ability of the network to prefer globally consistent pairing over some other pairing that preserves only local consistency. See figure 9 for figure conventions. See also the text and figures 18 and 19 for the significance of this example.

5. Discussion

5.1. Weinshall's ambiguous stereograms (1989)

Weinshall (1989) created ambiguous stereograms in which multiple transparent surfaces, corresponding to non-unique local stereomatches, are perceived. These stereograms are a random dot version of the double-nail illusion (Krol and van de Grind 1980). In the double-nail illusion humans favour matches that define surfaces close to the frontoparallel plane.

A single surface is perceived when the stereomatches defining the surface have 100% correlation between the left and right images. Weinshall (1989) demonstrated that multiple surfaces are perceived when none of the surfaces defined by the different plausible stereomatches has 100% correlation, even though correlation in one surface is increased with respect to the other surfaces.

In the EG model, the stereomatches with 100% correlation between the left and right images most strongly activate the layer 3 neuron representing the corresponding surface, and therefore, the layer 2 neurons representing these stereomatches receive the strongest feedback from layer 3 to layer 2 and succeed in suppressing the other matches.

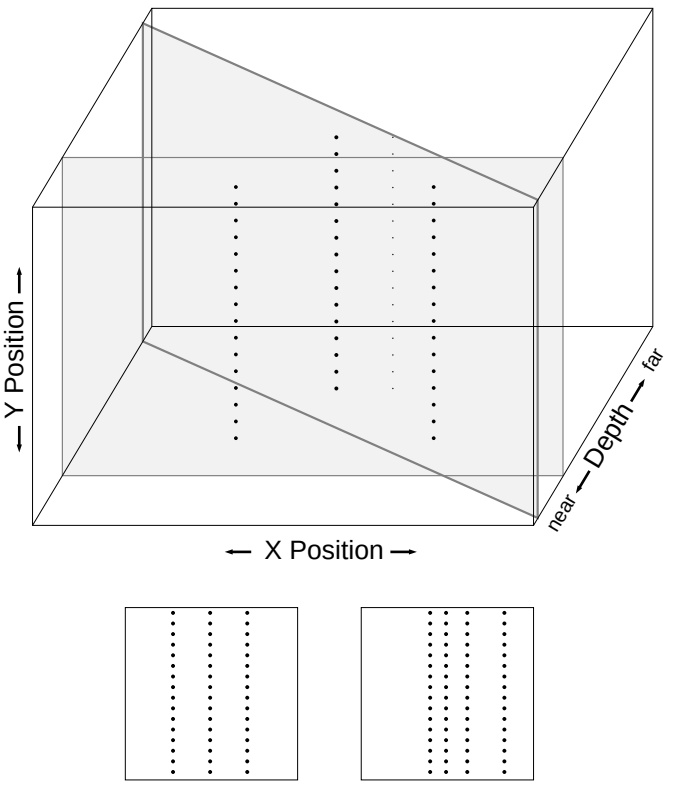

Figure 18. Network response to ambiguous input without feedback. The input stereogram is ambiguous because of the unequal number of columns in the left and right images. This figure shows the output of the stereo network simulation *without* the top-down feedback. The layer 2 neurons achieve only locally consistent matches and the layer 3 neurons represent two surfaces. See figure 9 for conventions.

In the EG model, with the ambiguous stereograms, layer 3 neurons representing the surfaces defined by all the plausible local stereomatches receive a comparable amount of excitation and hence, the layer 2 neurons representing the different stereomatches receive a similar amount of feedback from layer 3 to layer 2. The randomness in the distribution of the dots in the RDS may result in different stereomatches prevailing at different locations in layer 2. The layer 3 neurons representing the different surfaces can be coactive if they do not have much overlap in their feedforward connections, as specified in (15). Thus, the EG network can represent multiple surfaces corresponding to ambiguous stereomatches. This explanation is essentially the same as that of Pollard and Frisby (1990).

5.2. Stereoscopic depth attraction and repulsion

As the stereoscopic depth of a surface is moved away from that of a reference surface, the perceived depth of the surface changes from being closer to the reference surface than its objective depth to being further away from it than its objective depth. This is called the depth attraction–repulsion effect (Westheimer 1994).

In the EG network, the feedforward excitatory connections from layer 2 to layer 3 and the lateral inhibitory connections within layer 3 produce 'Mexican hat' shaped response

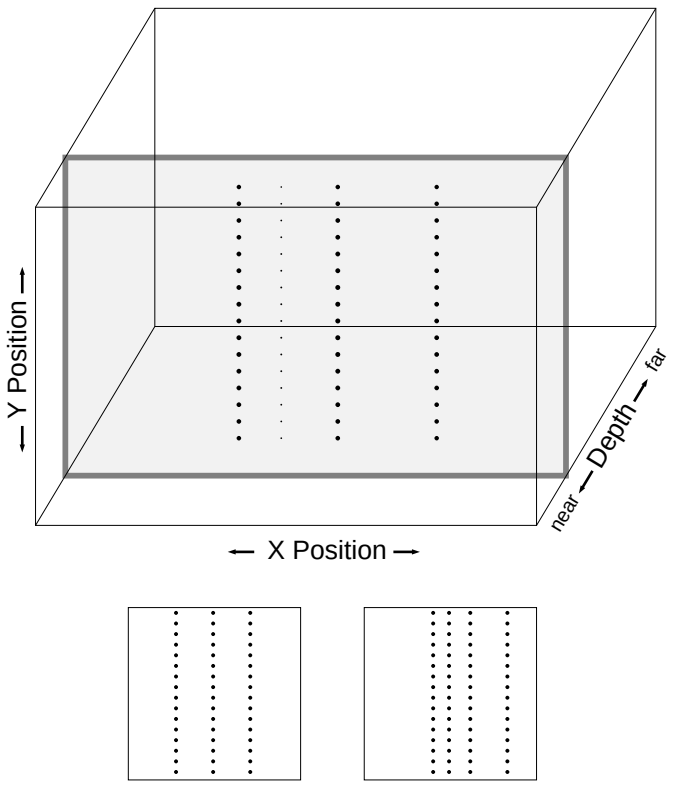

Figure 19. Network response to ambiguous input with feedback. This figure shows the output of the stereo network simulation *with* the top-down feedback, in response to an ambiguous input. The layer 2 neurons achieve locally and globally consistent matches, and the layer 3 neurons represent a single frontoparallel surface. See figure 9 for conventions.

profiles for the layer 3 neurons. Such receptive fields have been used to model depth attraction and repulsion (Lehky and Sejnowski 1990).

5.3. Relation of the model to self-organization

The motivation for and design of the EG stereomatching network grew from an analysis of self-organization. The way in which a neural network with unsupervised learning would develop in response to repeated exposure to many stereoimages was considered. As a result, the network's structure can be described compactly, as the emergent product of certain simple learning rules. The learning process will be sketched here and detailed in a later paper.

5.3.1. Development of layers 1 and 2. The network follows the 'EXIN' (excitatory + inhibitory) learning rules described by Marshall (1995a). A variant (Grossberg 1976a, b) of a Hebbian learning rule governs weight changes for the excitatory connections from layer 1 neurons to layer 2 neurons. The structure of the layer 1 to layer 2 excitatory connections emerges in a straightforward manner (Marshall 1990c, 1992), in a similar manner to the emergence of receptive fields in many models of other aspects of vision, such as oriented

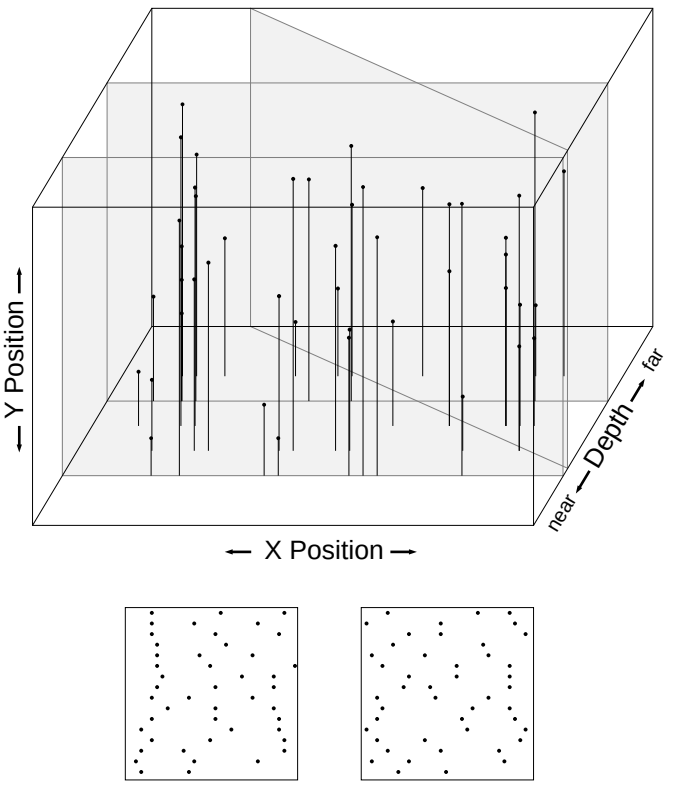

Figure 20. Random cloud of dots. The input stereogram to the network represents corresponding pairs of dots at random positions. This input demonstrates the network's ability to assign disparities to dot pairs even in the absence of a globally consistent configuration of the dots representing surfaces in depth. Each dot is drawn with a vertical tail that ends on the bottom surface of the rectangular box, to clarify the disparities. The layer 3 neurons shown here have subthreshold activation levels. They are shown to illustrate the relative activation levels of the layer 3 cells. Although, no layer 3 neuron reaches suprathreshold activation level, the layer 2 neurons perform locally consistent stereo correspondence.

and end-stopped receptive fields (Marshall 1990b, d) and motion-sensitive receptive fields (Hubbard and Marshall 1994, Marshall 1990a, 1995b, Schmitt and Marshall 1995). Each layer 2 neuron acquires a strong unique preference for a particular stereo disparity pair at a particular image location (Lee and Olshausen 1996, Marshall 1990c, 1992).

In addition to the Hebbian excitatory learning rule, the EXIN model imposes an anti-Hebbian *inhibitory* learning which operates on the lateral inhibitory connections between neurons within layer 2. In layer 2 of the stereomatching network, the inhibitory learning rule tends to strengthen the inhibition between neurons which are frequently coexcited—those that receive common excitatory input from layer 1 neurons. The inhibition is thus strengthened between conflicting disparity pairs and weakened between non-conflicting disparity pairs (Marshall 1990c, 1992) (figure 3).

5.3.2. Development of layers 2 and 3. The feedforward and feedback excitatory connection weights between layer 2 and layer 3 are again shaped by the EXIN excitatory learning rule. Because the layer 2 neurons are usually coactive in sets that define surfaces in depth,

the layer 3 neurons in the stereomatching network become selective for local *3D oriented surface patches in depth*.

The EXIN inhibitory learning rule weakens the layer 3 lateral inhibitory connection weights between neurons representing dissimilar (e.g. different depths) surface patches. It strengthens the layer 3 lateral inhibitory connection weights between neurons representing similar (i.e. substantially intersecting) surface patches. Thus, inhibition within layer 3 is also based on pattern overlap, as in layer 2.

The EXIN excitatory learning rule causes the feedback connection weights to become approximately symmetric to the feedforward connection weights. Thus, every layer 3 neuron learns to excite all the layer 2 neurons that tend to excite it.

5.4. Future work

The EG network solves only the stereomatching problem involved in binocular vision. The network is based on an analysis of self-organization. Further work is required to generate the stereomatching network from an initially non-specific network as a result of binocular stimulation and self-organizing learning rules, as outlined in section 5.3.

Half occlusion occurs when opaque objects occlude others; one eye sees parts of the occluded object not seen by the other eye. These features do not have a match in the other eye. Thus, they are called half-occluded features. In RDSs of one opaque surface occluding another, the half-occluded features are generally perceived as belonging to the occluded object (Anderson and Nakayama 1994, Shimojo and Nakayama 1990, 1994). This occurs even though the half-occluded dots are close to the edge of the occluding object. The edges of the occluder appear sharp. In the EG network the grouping of the half-occluded dots is ambiguous.

In the EG network, only one matching feature (dots) is used. A complete model of stereomatching should handle matching primitives such as grey-scale correlation (white and black dots), dot cluster patterns, edges, orientation, etc (Anderson and Nakayama 1994). The EG network must be extended to represent surfaces defined by disparities arising from different matching primitives.

The EG network was tested using RDSs that contained surfaces whose position, depth, slant and tilt were similar to those encoded by the layer 3 neurons. Surfaces with position, depth, slant and tilt intermediate to those encoded by individual layer 3 neurons can be represented in a distributed fashion by simultaneous partial activation of multiple neurons with nearby selectivities.

The current model exploits the grouping at a scale set by the hard-wired layer 2 to layer 3 feedforward connections to resolve ambiguities arising at a smaller scale in the stereomatching stage. To represent surfaces of different sizes and curvatures, neurons selective for a continuum of sizes and curvatures would be required in layer 3. Furthermore, additional layers employing the same architecture for exclusive allocation and grouping may be required to group nearby surfaces into a coherent whole.

The presence of different matching primitives can lead to binocular rivalry. When different features are presented to corresponding locations of the eyes, perception can alternate between the features presented to the eyes. For example, when a vertical line is presented to the left eye and a horizontal line to the right eye at corresponding retinal locations, vertical and horizontal segments are perceived alternately. There exist periods when the vertical or horizontal is seen exclusively. There is considerable debate on whether binocular rivalry and stereopsis are parallel processes or whether binocular rivalry ensues when stereopsis fails (Blake 1989, Blake and O'Shea 1988, Wolfe 1986, 1988).

The EG algorithm and the other algorithms of stereomatching (e.g. Frisby and Pollard 1991, Marr and Poggio 1976, Prazdny 1985) compute absolute disparities. In animals absolute disparities are difficult to determine because of eye, head and body movements. The disparity of an object's retinal images changes with eye, head and body movements. Thus, absolute disparities are unreliable information from which to extract surfaces.

Van Ee and Erkelens (1995) showed that human stereoscopic depth perception is insensitive to slant without a visual reference. They attribute stable stereoscopic depth perception during whole field disparity changes during eye, head and body motion to the lack of a visual reference. They hypothesize that relative disparity is used in stereoscopic depth perception.

5.5. Conclusions

This paper describes a novel neural network architecture for stereomatching, called the EG network. The salient features of the neural architecture are selective inhibition in layer 2 and top-down feedback from layer 3 to layer 2. In addition, layer 2 neurons are not activated by top-down excitation alone. Computer simulations show that the network represents oblique surfaces in 3D, multiple transparently overlaid surfaces, and random dot clouds directly.

The model has the following properties:

- *Transparency representation.* The selectively reduced inhibition between the representations of non-conflicting stereomatches enables the network to represent transparently overlaid surfaces (not just opaque surfaces), unlike the Marr–Poggio algorithm. Furthermore, unique disparity values are assigned to every visual feature, unlike the Prazdny algorithm.
- *Oblique surface representation.* The network represents and supports obliquely-oriented surfaces in depth directly (not indirectly as an approximation to frontoparallel surfaces), unlike both the Marr–Poggio and the Prazdny algorithms. The feedback loop between layer 2 and layer 3 allows the support functions to be oblique when needed, not just frontoparallel.
- *Global overcomes local.* The combination of lateral inhibition and feedback excitation allows the network to represent globally consistent, globally optimal stereomatches (as in the Ternus display), rather than merely locally optimal ones.
- *Grouping/binding.* The lateral inhibition within layers 2 and 3 and the feedback loop between layers 2 and 3 exclusively groups or binds the layer 1 monocular features into a representation of surfaces in depth.
- *Self-organization.* The model is motivated by and based on a set of simple learning rules, which rules have also been used to model other aspects of visual processing (e.g. orientation selectivity and motion selectivity). Thus, the stereomatching network is part of a unified theory of vision based on activity-guided development.

Despite the open questions described in section 5.4, the EG network contributes to the understanding of stereopsis and other visual processes. The network constitutes a general solution for resolving conflicts in grouping and transparency representation. In addition to the applicability of the exclusive grouping property to stereomatching, the EG model has also been used to represent motion grouping and motion transparency (Schmitt and Marshall 1995). Furthermore, the relationship of the stereomatching network to self-organization provides the possibility of compactly describing cortical function and

development. The EG algorithm solves problems in stereomatching not handled by previous algorithms. The same EXIN rules have been used to describe the self-organization of a variety of visual processing functions (Gupta and Marshall 1995, Hubbard and Marshall 1994, Marshall 1989, 1990a, b, c, d, 1992, 1995b, Marshall and Alley 1993, Schmitt and Marshall 1995). Thus, the EG algorithm for stereomatching is part of a more general class of algorithms that emerge from a common neural self-organization process.

Acknowledgment

Some of the results presented here were previously reported at the 1995 Annual Meeting of the Association for Research in Vision and Ophthalmology (ARVO) (Marshall and Kalarickal 1995). This work was supported in part by ONR Grant N00014-93-1-0208 and by NIH Grant EY09669.

Appendix

The parameter values used to compute the weights were chosen as follows:

$N_{1X} = 64$, $N_{1Y} = 16$, $N_1 = 1024$, $N_{2X} = 64$, $N_{2Y} = 16$, $N_{2D} = 5$, $N_2 = 5120$, $N_{3X} = 1$, $N_{3Y} = 1$, $N_{3Z} = 5$, $N_{3S} = 3$, $N_{3T} = 3$, $N_3 = 45$, $\text{base}_{1X} = 0$, $\text{base}_{2X} = 0$, $\text{base}_{3X} = 15.5$, $\text{base}_{1Y} = 0$, $\text{base}_{2Y} = 0$, $\text{base}_{3Y} = 7$, $\text{base}_{2D} = -2.0$, $\text{base}_{3Z} = -2.0$, $\text{scale}_{1X} = 1$, $\text{scale}_{2X} = 0.5$, $\text{scale}_{3X} = 16$, $\text{scale}_{1Y} = 1$, $\text{scale}_{2Y} = 1$, $\text{scale}_{3Y} = 16$, $\text{scale}_{2D} = 1$, $\text{scale}_{3Z} = 1$, $\sigma_x^1 = 0.69$, $\sigma_y^1 = 2.00$, $\sigma_x^2 = 0.058$, $\sigma_y^2 = 0.123$, $\sigma_z^2 = 1.185$, $\theta_e^{12} = 0.18$, $\theta_e^{23} = 0.18$, $\theta_i^{22} = 0.01$, $\theta_i^{33} = 0.02$, bias $= 0.02$.

The slant and tilt of layer 3 neuron i were assigned as follows:

$$
\begin{array}{lll}
S_i' = 0 & T_i' = 0 \rightarrow S_i = \arctan \frac{4}{15} & T_i = 45^\circ \\
S_i' = 0 & T_i' = 1 \rightarrow S_i = -\arctan \frac{4}{15} & T_i = 135^\circ \\
S_i' = 0 & T_i' = 2 \rightarrow S_i = -\arctan \frac{1}{5} & T_i = 90^\circ \\
S_i' = 1 & T_i' = 0 \rightarrow S_i = -\arctan \frac{4}{15} & T_i = 0^\circ \\
S_i' = 1 & T_i' = 1 \rightarrow S_i = \arctan \frac{4}{15} & T_i = 0^\circ \\
S_i' = 1 & T_i' = 2 \rightarrow S_i = 0^\circ & T_i = 90^\circ \\
S_i' = 2 & T_i' = 0 \rightarrow S_i = \arctan \frac{4}{15} & T_i = 45^\circ \\
S_i' = 2 & T_i' = 1 \rightarrow S_i = \arctan \frac{4}{15} & T_i = 135^\circ \\
S_i' = 2 & T_i' = 2 \rightarrow S_i = \arctan \frac{1}{5} & T_i = 90^\circ.
\end{array}
$$

The parameters and the functions used to compute the excitations, inhibitions and activations were as follows. The notation $[x]$ denotes $\max(0, x)$.

For layer 2 neurons, $G(x) = x$, $F(x) = x^2$, $H(x) = [x]$, $I(x) = x$, $\Gamma = 1.0$, $J(x) = [x]$, $K(x) = x$, $\epsilon = 1.0$, $\beta = 0.15$, and $\gamma = 9.3$.

For layer 3 neurons, $G(x) = x^2$, $F(x) = x^2$, $\Gamma = 0.0$, $J(x) = [x]$, $K(x) = x$, $\epsilon = 0.005$, $\beta = 0.06$, and $\gamma = 6.3$.

For layer 2 and layer 3 neurons the following parameters were the same. $A = 1$, $B = 2$, and $C = 0.2$.

The activation equations for layer 2 and layer 3 neurons were integrated numerically using the Euler method with a time step of 0.05. The network reached an equilibrium state by time $= 80$.

References

Anderson B L and Nakayama K 1994 Toward a general theory of stereopsis: binocular matching, occluding contours, and fusion *Psychol. Rev*. **101** 414–45

Bishop P O and Pettigrew J D 1986 Neural mechanisms of binocular vision *Vision Res*. **26** 1587–600

Blake R 1989 A neural theory of binocular rivalry *Psychol. Rev*. **96** 145–67

Blake R and O'Shea R P 1988 'Abnormal fusion' of stereopsis and binocular rivalry *Psychol. Rev*. **95** 151–4

Bradley D C and Andersen R A 1995 Stereospecificity of the nonclassical receptive field surround in macaque area MT *Investigative Ophthalmol. Visual Sci*. **36** S873

Bregman A S 1990 *Auditory Scene Analysis: the Perceptual Organization of Sound* (Cambridge, MA: MIT Press)

DeAngelis G C, Ohzawa I and Freeman R D 1991 Depth is encoded in the visual cortex by a specialized receptive field structure *Nature* **352** 156–9

Ferster D L 1981 A comparison of binocular depth mechanisms in areas 17 and 18 of the cat visual cortex *J. Physiol*. **311** 623–55

Fox R, Aslin R N, Shea S L and Dumais S T 1980 Stereopsis in human infants *Science* **207** 323–4

Frisby J P and Pollard S B 1991 Computational issues in solving the stereo correspondence problem *Computational Models of Visual Processing* ed M S Landy and J A Movshon (Cambridge, MA: MIT Press) pp 331–58

Grossberg S 1972 Neural expectation:cerebellar and retinal analogs of cells fired by learnable or unlearnable pattern classes *Kybernetik* **10** 49–57

——1976a On the development of feature detectors in the visual cortex with applications to learning and reaction-diffusion systems *Biol. Cybern*. **21** 145–59

——1976b Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors *Biol. Cybern*. **23** 121–34

——1982 *Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control* (Boston, MA: Reidel Press)

Gupta V S and Marshall J A 1995 Development of triadic neural circuits for anticipatory receptive field shifts under intended eye movements *Investigative Ophthalmol. Visual Sci*. **36** S378

He Z J and Nakayama K 1994 Perceived surface shape not features determines correspondence strength in apparent motion *Vision Res*. **34** 2125–35

Hirsch J and Gilbert C D 1991 Synaptic physiology of horizontal connections in the cat's visual cortex *J. Neurosci*. **11** 1800–9

Hopfield J J 1982 Neural networks and physical systems with emergent collective computation abilities *Proc. Natl Acad. Sci. USA* **79** 2554–8

Hubbard R S and Marshall J A 1994 Self-organizing neural network model of the visual inertia phenomenon in motion perception *Technical Report 94-001* Department of Computer Science, University of North Carolina, Chapel Hill, p 26

Hubel D H and Wiesel T N 1977 Functional architecture of macaque monkey visual cortex *Proc. R. Soc. Lond.* B **198** 1–59

——1984 *Self-Organization and Associative Memory* (New York: Springer)

Julesz B 1971 *Foundations of Cyclopean Perception* (Chicago, IL: University of Chicago Press)

Krol J D and van de Grind W A 1980 The double-nail illusion: experiments on binocular vision with nails, needles, and pins *Perception* **9** 651–69

Lee C W and Olshausen B A 1996 A nonlinear Hebbian network that learns to detect disparity in random-dot stereograms *Neural Comput*. **8** 545–66

Lehky S R and Sejnowski T J 1990 Neural model of stereoacuity and depth interpolation based on a distributed representation of stereo disparity *J. Neurosci*. **10** 2281–99

Marr D and Poggio T 1976 Cooperative computation of stereo disparity *Science* **194** 283–7

Marshall J A 1989 Self-organizing neural network architectures for computing visual depth from motion parallax *Proc. Int. Joint Conf. on Neural Networks (Washington, DC)* vol II (Piscataway, NJ: IEEE) pp 227–34

——1990a Self-organizing neural networks for perception of visual motion *Neural Networks* **3** 45–74

——1990b Development of length-selectivity in hypercomplex-type cells *Investigative Ophthalmol. Visual Sci*. **31** S397

——1990c A self-organizing neural network for computing stereo disparity and transparency *Technical Digest, Opt. Soc. Am. Annual Meeting* p 268

——1990d Adaptive neural methods for multiplexing oriented edges *Intelligent Robots and Computer Vision IX: Neural, Biological, and 3-D Methods (Boston, MA)* ed D P Casasent *Proc. SPIE* **1382** 282–91

——1992 Unsupervised learning of contextual constraints in neural networks for simultaneous visual processing of multiple objects *Neural and Stochastic Methods in Image and Signal Processing (San Diego, CA)* ed S S Chen *Proc. SPIE* **1766** 84–93

Marshall J A 1995a Adaptive perceptual pattern recognition by self-organizing neural networks: context, uncertainty, multiplicity, and scale *Neural Networks* **8** 335–62

——1995b Motion perception: self-organization *The Handbook of Brain Theory and Neural Networks* ed M Arbib (Cambridge, MA: MIT Press) pp 153–7

Marshall J A and Alley R K 1993 A self-organizing neural network that learns to detect and represent visual depth from occlusion events *Proc. AAAI Fall Symp. on Machine Learning and Computer Vision* ed K Bowyer and L Hall pp 70–4

Marshall J A and Kalarickal G J 1995 A neural circuit for stereomatching of transparent and oblique surfaces *Investigative Ophthalmol. Visual Sci*. **36** S366

Nakayama K and Shimojo S 1992 Experiencing and perceiving visual surfaces *Science* **257** 1357–63

Nakayama K, Shimojo S and Silverman G H 1989 Stereoscopic depth: its relation to image segmentation, grouping, and the recognition of occluded objects *Perception* **18** 55–68

Ohzawa I, DeAngelis G C and Freeman R D 1990 Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors *Science* **249** 1037–41

Poggio G F and Fischer B 1977 Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkeys *J. Neurophysiol*. **40** 1392–405

Pollard S B and Frisby J P 1990 Transparency and the uniqueness constraint in human and computer stereo vision *Nature* **347** 553–5

Prazdny K 1985 Detection of binocular disparities *Biol. Cybern.* **52** 93-9

Qian N and Sejnowski T J 1989 Learning to solve random-dot stereograms of dense and transparent surfaces with recurrent backpropagation *Proc. 1988 Connectionist Models Summer School* ed D Touretzky *et al* (San Mateo, CA: Morgan Kaufman) pp 435–43

Schmitt C P and Marshall J A 1995 Development of a neural circuit for motion grouping and disambiguation *Investigative Ophthalmol. Visual Sci*. **36** S377

Shimojo S and Nakayama K 1990 Real world occlusion constraints and binocular rivalry *Vision Res*. **30** 69–80

——1994 Interocularly unpaired zones escape local binocular matching *Vision Res*. **34** 1875–81

Shimojo S, Silverman G H and Nakayama K 1989 Occlusion and the solution to the aperture problem for motion *Vision Res*. **29** 619–26

Ternus J 1926 Experimentelle Untersuchungen über phänomenale Identität *Psycholog. Forschung* **7** 81–136

van Ee R and Erkelens C J 1995 Binocular perception of slant about oblique axes relative to a visual frame of reference *Perception* **24** 299–314

Weinshall D 1989 Perception of multiple transparent planes in stereo vision *Nature* **341** 737–9

Westheimer G 1994 The Ferrier lecture, 1992. Seeing depth with two eyes: stereopsis *Proc. R. Soc. London* B **257** 205–14

Wolfe J M 1986 Stereopsis and binocular rivalry *Psychol. Review* **93** 269–82

——1988 Parallel ideas about stereopsis and binocular rivalry: a reply to Blake and O'Shea 1988 *Psychol. Rev.* **95** 155–8