1996-11

Cortical Computation of Stereo Disparity

McLoughlin, Niall P.

Boston University Center for Adaptive Systems and Department of Cognitive and Neural Systems

Cortical computation of stereo disparity

Niall McLoughlin and Stephen Grossberg

November, 1996

Technical Report CAS/CNS-1996-022

Boston University Center for Adaptive Systems
and
Department of Cognitive and Neural Systems
677 Beacon Street
Boston, MA 02215

CORTICAL COMPUTATION OF STEREO DISPARITY

by

Niall P. McLoughlin[1] and Stephen Grossberg[2]

[1] Department of Neurobiology
Harvard Medical School
220 Longwood Ave., Room B I-550
Boston, MA 02115
niall@cns.bu.edu
617-432-4438
FAX: 617-734-7557

[2] Department of Cognitive and Neural Systems
Center for Adaptive Systems
Boston University
677 Beacon Street
Boston, MA 02215
steve@cns.bu.edu
617-353-7858
FAX: 617-353-7755

November 1996
Revised: February 1997

Technical Report CAS/CNS-TR-96-022

Requests for reprints should be sent to:
Stephen Grossberg
Department of Cognitive and Neural Systems
Boston University
677 Beacon Street
Boston, MA 02215

---

[1] Supported in part by the Defense Advanced Research Projects Agency (ONR N00014-92-J-4015), the Air Force Office of Scientific Research (AFOSR 90-0175), the Office of Naval Research (ONR N00014-91-J-4100), and the James S. McDonnell Foundation (94-40).

[2] Supported in part by the Defense Advanced Research Projects Agency and the Office of Naval Research (ONR N00014-95-1-0409 and ONR N00014-95-1-0657).

Acknowledgments: The authors wish to thank Carol Jefferson and Robin Locke for their valuable assistance in the preparation of the manuscript.

ABSTRACT

Our ability to see the world in depth is a major accomplishment of the brain. Previous models of how positionally disparate cues to the two eyes are binocularly matched limit possible matches by invoking uniqueness and continuity constraints. These approaches cannot explain data wherein uniqueness fails and changes in contrast alter depth percepts, or where surface discontinuities cause surfaces to be seen in depth although they are registered by only one eye (da Vinci stereopsis). A new stereopsis model explains these depth percepts by proposing how cortical complex cells binocularly filter their inputs and how monocular and binocular complex cells compete to determine the winning depth signals.

INTRODUCTION

One of the great challenges in contemporary science is to explain how the brain transforms the scintillating patterns of light that impinge on our two two-dimensional retinas into three-dimensional percepts of objects seen in depth. In order to accomplish this, the brain needs to determine which image features on the two retinas belong together, despite the fact that the positions of these features are different on each eye, and depend upon how far away an object is and on where the eyes are looking. Through this binocular matching process, the brain converts the positionally disparate features on the two retinas into single object locations seen in depth, and then organizes these individual locations in depth into the boundaries that surround the objects that we see. The brain's problem is complicated by the fact that whole regions of a scene may be visible to only one eye. Nevertheless, these monocularly defined regions are still perceived at the correct depths. The present work describes a model of how this binocular matching process takes place in the visual cortex and uses the model to explain recent psychophysical and neural data that previous models have not accommodated. The model also clarifies how constraints on binocular matching and boundary formation, that may at first seem to be at odds with one another, can be reconciled.

For many years, random dot stereograms have been used to probe how the brain does stereo matching (Julesz, 1971). In such a stereogram, random dots seen by one eye are paired with positionally shifted dots that are seen by the other eye. The binocular disparities of the paired dots are used by the brain to compute percepts of relative depth. The contrast polarity of the dots with respect to their background can greatly alter the depth percept. For example, random dots that are presented to the two eyes with opposite contrast polarities are treated as statistically independent and are not matched binocularly (Harris and Parker, 1995). The same dots can, however, be fused if their background is changed so that both sets of dots appear brighter or darker than the background. Additional psychophysical studies have used bars rather than dots to provide examples wherein a single feature seen by one eye can be non-uniquely matched with more than

one feature seen by the other. More generally, these studies show how the number of matches and the depths at which they are perceived depend upon the patterns of image contrast that are seen by both eyes, as in studies of dichoptic masking (McKee *et al.*, 1994) and variants of Panum's limiting case (McKee *et al.*, 1995; Smallman and McKee, 1995) (Figure 1). These instructive properties of binocular matching are not accounted for by classical models of stereopsis (Sperling, 1970; Julesz, 1971; Nelson, 1975; Marr and Poggio, 1976).

Figure 1

Previous models also fail to explain the depth percepts that occur in the presence of surface discontinuities. For example, when surfaces in a scene abruptly terminate, as at occluding walls of a room, one or both eyes may detect regions that are not registered by the other eye. Although they generate no binocular disparities, these "half-occluded" (Belhumeur and Mumford, 1992; Anderson and Nakayama, 1994) regions are attributed their correct binocular depth. This da Vinci stereopsis phenomenon imposes a number of constraints upon the design of the visual system. For example, unmatched monocular regions must be able to survive whatever form of binocular filtering occurs in the cortex. Computationally, this suggests the inclusion of monocular cells that incorporate eye-of-origin information (Nakayama and Shimojo, 1990; Tyler, 1983) within the binocular matching stage of vision.

METHODS

Our new neural model of binocular vision explains phenomena such as da Vinci stereopsis by analyzing how the visual system copes with surface discontinuities (Grossberg, 1994; Grossberg and McLoughlin, 1995; McLoughlin and Grossberg, 1994). We show herein how this model explains key data about the non-uniqueness and contrast-sensitivity of binocular matching. The model hereby shows how, by appropriately renouncing the uniqueness and continuity constraints of previous models, it can account for many more psychophysical and neural data about stereo matching.

The model accomplishes this by proposing how cortical complex cells carry out binocular matching, and how monocular and binocular cortical cells compete to determine the winning match or matches. Complex cells are modeled herein as disparity-sensitive cells that pool signals from cortical simple cells that are sensitive to similar orientations but opposite contrast polarities in the image (Hubel and Wiesel, 1962). Hence a complex cell can fire to either light or dark image features that are presented at the cell's preferred disparity. This property of complex cells helps to explain how the brain computes percepts of object boundaries even if the contrast of the object with respect to its background reverses as one traverses the boundary (Grossberg, 1994). Since complex cells pool signals from both contrast polarities, subsequent cortical processing stages can track the boundary independent of its direction of contrast.

Further structure is needed to explain binocular matching, however, since it is known that while light/light and dark/dark binocular matches are effective, light/dark and dark/light matches are not (Belhumeur and Mumford, 1992; von Helmholtz, 1910/1925; Ohzawa, DeAngelis, and Freeman, 1990). They cancel at the matching stage. This property helps to explain how the brain matches left and right eye signals that may be derived from the same object feature.

How does the brain reconcile the seemingly conflicting requirements of binocularly matching like-polarity signals from the same object features, and building object boundaries that pool across contrast polarity? The model proposes that like contrast polarities are binocularly matched before the matched contrasts from opposite contrast polarities are pooled together. The model also shows how this matching process incorporates simple cells with even and odd receptive fields, both of which contribute to complex cell firing (Ohzawa, DeAngelis, and Freeman, 1990; Pollen and Ronner, 1981). Pooling even and odd cell signals eliminates a number of spurious binocular matches which could otherwise lead to incorrect depth estimates (Grossberg and McLoughlin, 1995; McLoughlin and Grossberg, 1994).

Figure 2

Figure 2a summarizes a complex cell model circuit that realizes these properties. Here, left and right eye simple cells with the same symmetry and contrast polarity are binocularly matched. Disparity is encoded by a horizontal shift in the centers of the left and right eye receptive fields. No binocular matching occurs between the inputs to the monocular complex cells (Figure 2b). Instead, the monocular simple cell activity is passed onto the matching stage. In either case, cells with opposite symmetry and contrast polarity inhibit each other at this matching stage. The net activity from each match is then half-wave rectified to generate outputs from all four combinations of symmetry and polarity that summate at complex cells. Ohzawa, DeAngelis, and Freeman (1990) presented a similar model of binocular complex cell summation. A small difference between our model and theirs is that we code disparity by horizontal shifts in the left and right eye receptive field centers while they use phase differences. A large difference is that they do not address how false matches are suppressed among monocular and binocular cells, nor how relative contrast influences binocular matching.

It is worth noting that the model circuit depicted in Figure 2a, which closely resembles Figure 3b of Ohzawa *et al.* (1990), is capable of accounting for the Harris and Parker (1995) data. Harris and Parker demonstrated that noisy random dot stereograms composed of dots lighter and darker than the background are more efficiently fused than noisy random dot stereograms composed of only light or dark dots. They explained their results by suggesting that dark dots are matched only with dark dots and that light dots are matched only with light dots. Hence, a light dot in one eye has less chance of matching a "noisy" dot in the other eye if the noise dots are both light and dark than just light alone. This property is instantiated within our model circuit by having opposite polarity matches inhibit each other at the matching stage. Opposite polarity dots will not match, although the same mechanism will match either light to light or dark to dark. Harris and Parker also found that if only one set of the light or dark dots had noise added, then efficiency was intermediate to either set alone. They proposed that "although contrast polarity may be used to assist binocular matching in a population of disparity-selective neurons, the signal delivered by the output of such neurons may reflect only the disparity values and may fail to indicate which feature

generated any particular disparity value" (pp. 810). By combining the half-wave rectified outputs of our matching process, the output stage of Figure 2a exemplifies this property. In fact, Figure 2a accounts for the Harris and Parker data set without resorting to separate ON and OFF binocular complex cells.

Figure 3

The present model handles false matches, or binocular combinations of input features that are not perceived (Julesz, 1971), as follows. Each binocular scene generally contains many false matches as well as a smaller number of perceived matches; see Figure 1. Within the model, false matches are suppressed by competition along the line of sight. In this manner, cells that attempt to code the same input compete for activation. This mechanism has been utilized by a number of previous models (Julesz, 1971; Marr and Poggio, 1976). Unlike previous models, we include monocular cell responses within the competition to help deal with surface discontinuities (Grossberg, 1994; Grossberg and McLoughlin, 1995). Complex cells, whether monocular or binocular, must be sufficiently active before they can begin to inhibit their competitors. Our complex cell model also frees us from imposing uniqueness constraints on the selection process. Each feature is capable of making and maintaining multiple binocular matches at the complex cells unless one match is much stronger than the rest. A complete set of equations detailing our implementation of this model is presented in the Appendix.

Tyler (1983) was perhaps the first to propose the inclusion of monocular cortical units into an explanation of the physiological basis of fusion. However, unlike the current model, which enforces competition between binocular and monocular cortical units, Tyler suggested that monocular and binocular units integrate their responses together unless their visual directions differed too greatly. The proposed integration was used to explain why the monocular half-images are not seen in binocularly fused presentations.

RESULTS

This fusion of ideas about complex cell filtering, the role of monocular cells, and line-of-sight inhibition allows a single implementation to account for how subjects perceive Panum's limiting case (Figure 4a), Panum's not-so-limiting case (Figure 4b), contrast variants thereof (Figures 5a–d) and dichoptic masking (Figure 6a–c). No previous model has had this explanatory range. Seven pools of disparity selective complex cells, corresponding to horizontal pixel shifts of -15, -5, -2, 0, +2, +5, and +15, along with monocular complex left and right eye cell pools were simulated in each case. Each pool of complex cells completely tiled the input images. In all the following figures, outputs from the seven disparity pools are collapsed together to form a single disparity map, with the positive *d* direction corresponding to farther than the fixation plane and the negative *d* direction corresponding to nearer than the fixation plane. Outputs from the monocular complex cells are shown along side for comparison.

Figure 4

In Figure 4a, two equal contrast-defined bars are presented to the right eye while one bar is presented to the left (McKee *et al.*, 1995). Both right eye bars make equivalently good matches with the left eye bar. Binocular complex cells coding the near and far disparities preserve these matches, since neither match is strong enough to suppress the other. The model response to this input pattern is presented underneath the input. All three bars are matched binocularly. Two fused bars are perceived at near and far disparities, and there is no activity at the left or right eye monocular complex cells. In Figure 4b, a second bar is added to the left eye input. In this case, binocular cells corresponding to the far disparity encode all four bars. The binocular complex cells corresponding to the near match seen in Figure 4a are suppressed by the combined effects of the two binocular matches at the far disparity. This occurs as both far matches share common inputs with the near match perceived previously.

Figure 5

In Figure 5a, a second bar is once again added to the left eye, but it is of much lower contrast relative to the bars in the right eye (Smallman and McKee, 1995). Binocular complex cells once again attempt to encode the inputs into two far matches and the one near match. However, this time one of the two far matches (the additional bar with the left-most left eye bar) is composed of two inputs of very dissimilar contrast. This far match is significantly weaker and is suppressed before it reaches threshold by the strong near match with which it shares common inputs. The model thus responds as shown. The high contrast bars make equal and opposite binocular matches which suppress all possible binocular matches of the low contrast bar. The low contrast bar is represented instead by monocular left eye complex cells, which are not inhibited, as they do not share inputs with either of the binocular matches.

Smallman and McKee (1995) investigated this effect extensively using a variety of contrasts for the additional bar and for the matching bars. We attempted to model their results by varying the contrast of the additional bar over a wide range of values as depicted in Figure 5b–d. If the contrast of the additional bar is only slightly less (Figure 5b) or slightly more (Figure 5c) than the matching bars, both binocular far matches are retained. This occurs because, once the imbalanced binocular far complex cells exceed their threshold, their inhibitory effects combine with the balanced far match to suppress the complex cells encoding the balanced near match. Two matches suppress one and the results are depicted in Figure 5b and c. If, however, the additional bar is of much higher contrast than the matching bars, a significant imbalance occurs once again (Figure 5d). This time the monocular left eye cells are more strongly activated by the high contrast stimulus than are any of the binocular complex cells that attempt to encode it. The monocular complex cells suppress all binocular matches of the high contrast left eye bar. This frees the lower contrast matching bars from the influence of the additional bar, and hence they fuse into the Panum's limiting case arrangement as before. These examples illustrate how contrast and monocular cells can influence the binocular matching process.

Figure 6

In Figure 6a, detection of a low contrast bar (probe) to the right eye is greatly reduced by simultaneous presentation of a high contrast bar at the same retinal location in the left eye. This effect is known as dichoptic masking (Legge, 1979). As in the previous examples, because the two bars have very different contrasts, the model does not fuse them. The higher contrast left eye bar activates monocular complex cells which suppress all possible binocular matches with right eye stimuli. This causes the low contrast right eye bar to be coded by right eye monocular cells. This model of binocular fusion represents the initial stages of a more complete computational model of how three-dimensional surface representations are generated (Grossberg, 1994; Grossberg and McLoughlin, 1995). Within this expanded model, the monocular left and right eye complex cell activities are pooled together at subsequent stages. In particular, monocular and binocular complex cells group together to form elongated boundary contours which are fed into a filling-in stage which generates surfaces from enclosed regions. As the left and right eye bars are represented at the same retinal coordinates within the monocular complex cell pools, the final surface percept will be a mixture of the two monocular views.

When in Figure 6b, a second high contrast bar is added to the right eye, binocular complex cells encode the two bright bars at a near disparity and the low contrast bar is easier to detect. This happens because the two high contrast bars make the best binocular match and this match suppresses all other matches of these bars. The low contrast bar is once again represented by right eye monocular complex cells. Since the binocularly matched masking stimulus is shifted in depth from the low contrast probe, detectability of the latter increases. As there is no left eye monocular activity, the low contrast probe is unmasked as the final surface percept will originate solely from the right eye input. In Figure 6c, the contrast of the additional right eye bar is changed to that of the low contrast bar. Now the high contrast left eye bar makes its best match once again with the left eye monocular cell pool. This suppresses all binocular matches of the left eye bar. The two right eye bars are thus coded within the right eye monocular cell pool. Since the high contrast left

eye bar falls in the same retinal location as the probe, poorer detectability of the probe again ensues as the monocular left and right eye complex cell activities are pooled at subsequent stages of the model.

CONCLUSION

In conclusion, we have presented a new model of stereopsis which includes: (a) a contrast-sensitive matching scheme in which (i) even and odd symmetric receptive fields summate, (ii) same direction-of-contrast subunits match while opposite polarities inhibit each other, and (iii) matched opposites summate at the complex cell level; (b) monocular cells compete with binocular cells; and (c) competition occurs only between cells which code the same input (line of sight inhibition). This model helps to explain how surface discontinuities lead to monocularly viewed regions being perceived at the depth of a neighboring binocularly viewed region during da Vinci stereopsis (Grossberg, 1994; Grossberg and McLoughlin, 1995), and in so doing, can also account for key data on non-uniqueness and contrast-sensitivity during binocular matching.

APPENDIX

Inputs were initially passed through ON (on-center, off-surround) and OFF (off-center, on-surround) cells, not depicted in figures, that model lateral geniculate cell types (Gove, Grossberg, and Mingolla, 1995; Schiller, 1992). Steady-state responses of cell membrane equations are, for ON cells:

$$X_{ij}^{R/L+} = \left[\frac{(U+B)\left(\sum_{pq}\left(C_{pq} - S_{pq}\right)I_{(i+p)(j+q)}\right)}{a + \sum_{pq}\left(C_{pq} + S_{pq}\right)I_{(i+p)(j+q)}} \right]^{+}, \quad (1)$$

and for OFF cells:

$$X_{ij}^{R/L-} = \left[\frac{(U+B)\left(\sum_{pq}\left(S_{pq} - C_{pq}\right)I_{(i+p)(j+q)}\right)}{\alpha + \sum_{pq}\left(C_{pq} + S_{pq}\right)I_{(i+p)(j+q)}} \right]^{+}, \quad (2)$$

where R and L denote left and right eye inputs, i and j the center of the cell's response field, α a decay parameter (10); U and B reversal potentials that bound upper and lower cell activity levels (1, 1); I_{pq} the input at position (p,q); $[x]^{+}$ is the half-wave rectification operator $max(x,0)$; and C_{pq} and S_{pq} the center and surround kernels (2-D Gaussians with standard deviations of 0.5 and 1.5, respectively). Kernels C_{pq} and S_{pq} are balanced so that their areas are equal.

Inputs to the complex cells were defined as:

$$M_{ijd} = \left| \left(\sum_l W_{dl}^{L\,even} E_{(i+l)j}^{L+\,even} + \sum_r W_{dr}^{R\,even} E_{(i+r)j}^{R+\,even} \right) - \left(\sum_l W_{dl}^{L\,even} E_{(i+l)j}^{L-\,even} + \sum_r W_{dr}^{R\,even} E_{(i+r)j}^{R-\,even} \right) \right|$$
$$+ \left| \left(\sum_l W_{dl}^{L\,odd} E_{(i+l)j}^{L+\,odd} + \sum_r W_{dr}^{R\,odd} E_{(i+r)j}^{R+\,odd} \right) - \left(\sum_l W_{dl}^{L\,odd} E_{(i+l)j}^{L-\,odd} + \sum_r W_{dr}^{R\,odd} E_{(i+r)j}^{R-\,odd} \right) \right|, \quad (3)$$

where |x| is the full-wave rectification operator, and the E_{ij} represent even and odd symmetric simple cell receptive fields defined as follows:

$$E_{ij}^{R/L+/-even} = \sum_{pq} \sin\left(\frac{2\pi k}{\Omega}\right) \exp\left(-\frac{1}{2}\left(\frac{p^2}{\sigma_p^2} + \frac{q^2}{\sigma_q^2}\right)\right) X_{(i+p)(j+q)}^{R/L+/-}, \tag{4}$$

$$E_{ij}^{R/L+/-odd} = \sum_{pq} \cos\left(\frac{2\pi k}{\Omega}\right) \exp\left(-\frac{1}{2}\left(\frac{p^2}{\sigma_p^2} + \frac{q^2}{\sigma_q^2}\right)\right) X_{(i+p)(j+q)}^{R/L+/-}. \tag{5}$$

The W_d weigh the left *(L)* and right *(R)* eye contributions to complex cell responses at disparity *d*. Monocular complex cells were constructed by setting the summation $W^{even/odd}$ equal to zero for one eye and scaling the remaining $W^{even/odd}$ to be approximately equal to $2W_0^{even/odd}$, where $W_0^{even/odd}$ is one of the two summation kernels of a zero-disparity binocular complex cell. Ω defines the period of the simple cells (2π); *k* defines their orientation (for this study only vertically oriented cells were investigated, so *k* was set equal to *p*, the horizontal dummy variable), and σ_p and σ_q their extent (2, 1.5).

The complex cell potentials $C_d(t)$ of both monocular and binocular cells interact with each other via a membrane equation.

$$\frac{dC_{ijd}}{dt} = -\beta C_{ijd} + \gamma(A_C - C_{ijd})M_{ijd} - (B_C + C_{ijd})\sum_{pe} f(C_{(i+p)jd})N_{pde}, \tag{6}$$

where β and γ are scale constants (.01, 15); reversal potentials A_C and B_C bound the activity of the complex units (1, 1); feedback signal $f(x)$ is a threshold-linear function: $f(x) = x - T$ if $x > T$;

otherwise $f(x) = 0$ $(T = .025)$; and N_{pde} is the inhibitory kernel between the complex units of different disparities d and e. Monocular complex units are described by zero disparity kernels:

$$N_{pde} = D_{de} \exp(-\sigma_i (p + K_{de})^2) \text{ for } d \neq e; \text{ else } N_{pdd} = D_d [\exp(-\sigma_c p^2) - \exp(-\sigma_s p^2)]. \quad (7)$$

Inhibition between cells tuned to different disparities is defined by a shifted Gaussian kernel whose shift K_{de} varies with the difference between d and e, $\sigma_i = .025$, and $D_{de} = 1$. Inhibition between cells coding the same disparity is defined by the difference of two Gaussians with $D_d = 2.5$, $\sigma_c = .015$, and $\sigma_s = .15$. Complex cell potentials were integrated through time using the 4th order Runge-Kutta method until a steady state was reached, as depicted in Figures 4–6.

REFERENCES

Anderson, B.L. and Nakayama, K. (1994). Toward a general theory of stereopsis: Binocular matching, occluding contours, and fusion. *Psychological Review*, **101**, 414–445.

Belhumeur, P.N. and Mumford, D. (1992). A Bayesian treatment of the stereo correspondence problem using half-occluded regions. In **Proceedings of the Institute of Electrical and Electronics Engineers Conference on Computer Vision and Pattern Recognition**. Los Alamitos, CA: IEEE, 506–512.

Gove, A., Grossberg, S. and Mingolla, E. (1995). Brightness perception, illusory contours, and corticogeniculate feedback. *Visual Neuroscience*, **12**, 1027–1052.

Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception and Psychophysics*, **55**, 48–120.

Grossberg, S. and McLoughlin, N.P. (1995). Cortical dynamics of 3–D surface perception: Binocular and half-occluded scenic images. **Technical Report CAS/CNS-TR-95-022.** Boston, MA: Boston University. *Neural Networks*, Submitted.

Harris, J.M. and Parker, A.J. (1995). Independent neural mechanisms for bright and dark information in binocular stereopsis. *Nature*, **374**, 808–811.

Hubel, D.H. and Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, **160**, 106–154.

Julesz, B. (1971). **Foundations of cyclopean perception**. Chicago: University of Chicago Press.

Legge, G. E. (1979). Spatial frequency masking in human vision: Binocular interactions. *Journal of the of the Optical Society of America*, **69**, 838–847.

Marr, D. and Poggio, T. (1976). Cooperative computation of stereo disparity. *Science*, **194**, 283–287.

McKee, S.P., Bravo, M.J., Smallman, H.J., and Legge, G.E. (1995). The "uniqueness constraint" and binocular masking. *Perception*, **24**, 49–65.

McKee, S.P., Bravo, M.J., Taylor, D.G., and Legge, G.E. (1994). Stereo matching precedes dichoptic masking. *Vision Research*, **34**, 1047–1060.

McLoughlin, N.P. and Grossberg, S. (1994). How are monocularly viewed surfaces perceived in depth during da Vinci stereopsis? *Investigative Ophthalmology and Visual Science (Supplement)*, **35**, 2110.

Nakayama, K. and Shimojo, S. (1990). DaVinci stereopsis: Depth and subjective occluding contours from unpaired image points. *Vision Research*, **30**, 1811–1825.

Nelson, J.I. (1975). Globality and stereoscopic fusion in binocular vision. *Journal of Theoretical Biology*, **49**, 1–88.

Ohzawa, I., DeAngelis, G.C., and Freeman, R.D. (1990). Stereoscopic depth discrimination by the visual cortex: Neurons ideally suited as disparity detectors. *Science*, **249**, 1037–1041.

Pollen, D.A. and Ronner, S.F. (1981). Phase relationships between adjacent simple cells in the visual cortex. *Science*, **212**, 1409–1411.

Schiller, P. (1992). The ON and OFF channels of the visual system. *Trends in Neurosciences*, **15**, 86–92.

Smallman, H.S. and McKee, S.P. (1995). A contrast ratio constraint on stereo matching. *Proceedings of the Royal Society of London, B*, **260**, 265–271.

Sperling, G.J. (1970). Binocular vision: A physical and neural theory. *Journal of the American Psychological Society*, **83**, 461–534.

Tyler, C.W. (1983). Sensory processing of binocular disparity. In C.M. Schor and K.J. Ciuffreda (Eds.), **Vergence eye movements: Basic and clinical aspects**. London: Butterworths, (pp. 199–295).

von Helmholtz, H. (1910/1925). **Treatise on physiological optics**. New York: Dover Press.

FIGURE CAPTIONS

Figure 1. Illustrative binocular stimuli: (a) Unambiguous binocular stimulus composed of two distinct bars presented dichoptically (each eye is presented with a shifted version of the two bars). Two bright bars are perceived floating in front of a black background. The shaded circles indicate the geometry of the binocular percept as viewed from above. Each shaded circle corresponds to a perceived edge of a bar. The unshaded circles represent some of the false matches that are not perceived. (b) In an example of Panum's limiting case, one bar is presented to the right eye while two shifted bars are presented to the left eye. The binocular disparity between the right eye bar and each of the left eye bars is equal and opposite in sign. The right eye bar fuses with both left eye bars. Two bars are perceived floating in depth, one in front of and the other behind the fixation plane (zero disparity by definition).

Figure 2. Model circuits of the inputs to cortical complex cells. (a) Binocular complex cells receive inputs from horizontally displaced monocular simple cells. Pairs of simple cells with shifted receptive fields of each contrast polarity and symmetry are binocularly matched. Opposite polarity matches inhibit each other, at the series of circles, and the half-wave rectified results are passed onto the binocular complex cell for summation. Binocular and monocular complex cells then compete with each other to encode the simple cell inputs. See Figure 3. (b) Monocular complex cells receive input from only one eye's simple cells (either the left or the right). Opposite contrast simple cell inputs inhibit each other at the matching stage, and once again the half-wave rectified outputs of this stage are passed onto the complex cell for summation. It is possible to imagine the initial matching stage occurring at intermediate binocular simple cells whose outputs compete and summate to form the binocular complex cell. Here it is implemented by convergence of simple cell inputs onto the complex cell's dendritic field.

Figure 3. Competition between complex cells. (a) Caricature of the inhibition between binocular and monocular complex cells. Complex cells compete to encode the incoming

stimulation. Each input can make many possible matches some of which are depicted by the circular spots in (a). Monocular complex cell responses are depicted at zero disparity as filled black (left eye) or white (right eye) circles. Striped circles indicate binocular complex cell responses. Complex cells coding near (negative) and far (positive) disparities along with the monocular complex cells are depicted. Inhibition occurs along the projection lines of the inputs depicted in (a). In particular, all complex cells which lie along any particular line attempt to encode the same monocular input. This competition includes monocular complex cells to cope with da Vinci stereopsis. (b) A more detailed overview of the inhibition between complex cells. Monocular inputs are presented for simplicity as two overlapping patterns of spatial activity. As can be seen, for a single left or right eye input, many complex cells are activated to some extent. Inhibition between different disparities takes allelotropia into consideration; that is, the fact that a binocularly fused stimulus is perceived to lie approximately halfway between its monocular half-images. Explicit equations which instantiate such a process are presented in the Appendix. Within-disparity inhibition helps to spatially sharpen the complex cells response. Note only positive disparities are shown for clarity.

Figure 4. Simulations of some binocular stimuli from McKee *et al.* (1995). For each example, left and right eye inputs are displayed above the computed outputs. Inputs are presented schematically using a gray scale code which represents the absolute luminance at each point. Outputs from the model are presented beneath each set of inputs. In each case, two sets of monocular (left and right), and seven sets of binocular cells compete to encode the inputs. The binocular cell pools are broken up into three sets of increasingly positive (far) disparity tuned cells, three sets of increasingly negative (near) disparity tuned cells, and cells tuned to the fixation plane (zero disparity). As in the psychophysical study, the model fixates the background. Monocular outputs from each simulation are presented on either side of the collapsed binocular disparity map. The vertical axis *(d)* of the disparity map represents the disparity at which key features in the input scenes are fused. (a) Panum's limiting case. Two contrast defined bars with Michelson contrasts, defined as $C_M = (L_{MAX} - L_{BACK})/(L_{MAX} +$

L_{BACK}), where L_{MAX} is the luminance of the bar, and L_{BACK} is the background luminance, of .43 are presented to the right eye while one bar is presented to the left. As described in the text, the model matches the single left eye bar to both right eye bars producing the depicted disparity map. (b) Panum's not-so-limiting case. When a second bar of $C_M = .43$ is added to the left eye's input, all bars are matched uniquely behind the fixation plane. See text for details.

Figure 5. Contrast effects of the additional bar. Smallman and McKee (1995) analyzed the effect of varying the contrast of the second left eye bar. They found that, in general, stimuli of different contrasts fuse only over a limited range of contrast differences. Here, we replicate their basic finding that stimuli, in this case bars, must be of similar contrast to be fused. (a) If the additional bar is of much lower contrast than the matching bars (.10 *vs* .43), the additional bar is effectively ignored and the brighter left eye bar fuses with both right eye bars. See text for details. However, if the additional bar is of lower (b) but similar contrast (.36), or higher (c) but similar contrast (.47), the four bars fuse uniquely behind the fixation plane. (d) If the additional bar is of much higher contrast (.73), it matches the monocular complex cells better than any binocular complex cells, and in doing so excludes it from being matched with either of the right eye bars. This results in non-unique matching of the lower contrast left eye bar. See text for details.

Figure 6. Dichoptic masking stimuli from McKee *et al.* (1994). (a) Basic dichoptic masking stimulus. A high contrast bar (.43) is presented to the left eye while a low contrast (.10) bar is presented at the same retinal location in the right eye. Both bars remain unfused as their contrasts differ too much. Dichoptic masking occurs as a result of pooling the left and right eye monocular complex cell outputs together at a subsequent stage of processing. In particular, the responses of monocular and binocular complex cells are grouped together to form elongated boundary contours which are fed into a filling-in stage which generates surfaces from enclosed regions. As the left and right eye bars are presented at the same retinal coordinates, the final surface percept is a mixture of the two monocular views. (b) A second high contrast bar is

added to the right eye. Both high contrast bars match and are encoded by the near complex cell pool. The low contrast right eye bar continues to be encoded by the right eye monocular complex cell pool. Unmasking of this low contrast probe occurs as the binocularly fused bar is shifted relative to the monocular bar, and there is no activity in the monocular complex left eye cell pool. (c) Once again, if the additional right eye bar is of a very different contrast (.10), the left and right eye inputs remain unfused. Dichoptic masking returns as the probe, which is encoded in the monocular complex right eye cells, is pooled with the left eye mask, which is encoded at the same retinal location in the monocular complex left eye cell pool, as before.

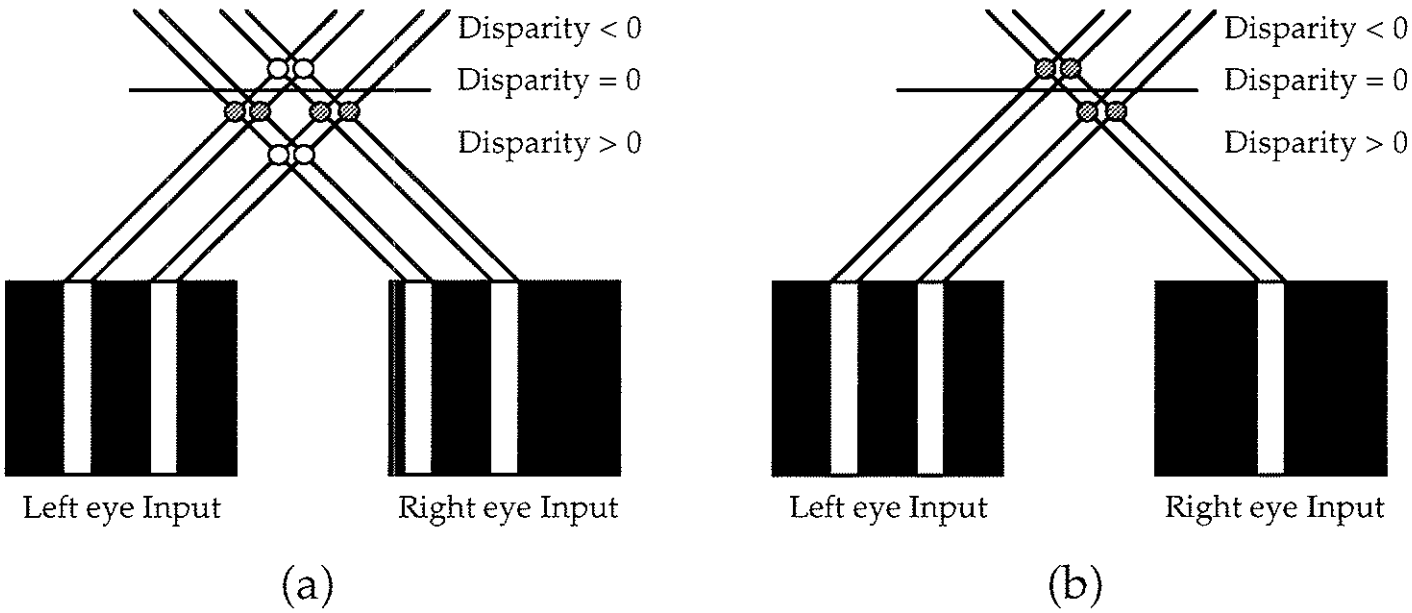

Disparity < 0
Disparity = 0
Disparity > 0
Left eye Input
Right eye Input
(a)
Disparity < 0
Disparity = 0
Disparity > 0
Left eye Input
Right eye Input
(b)

Figure 1

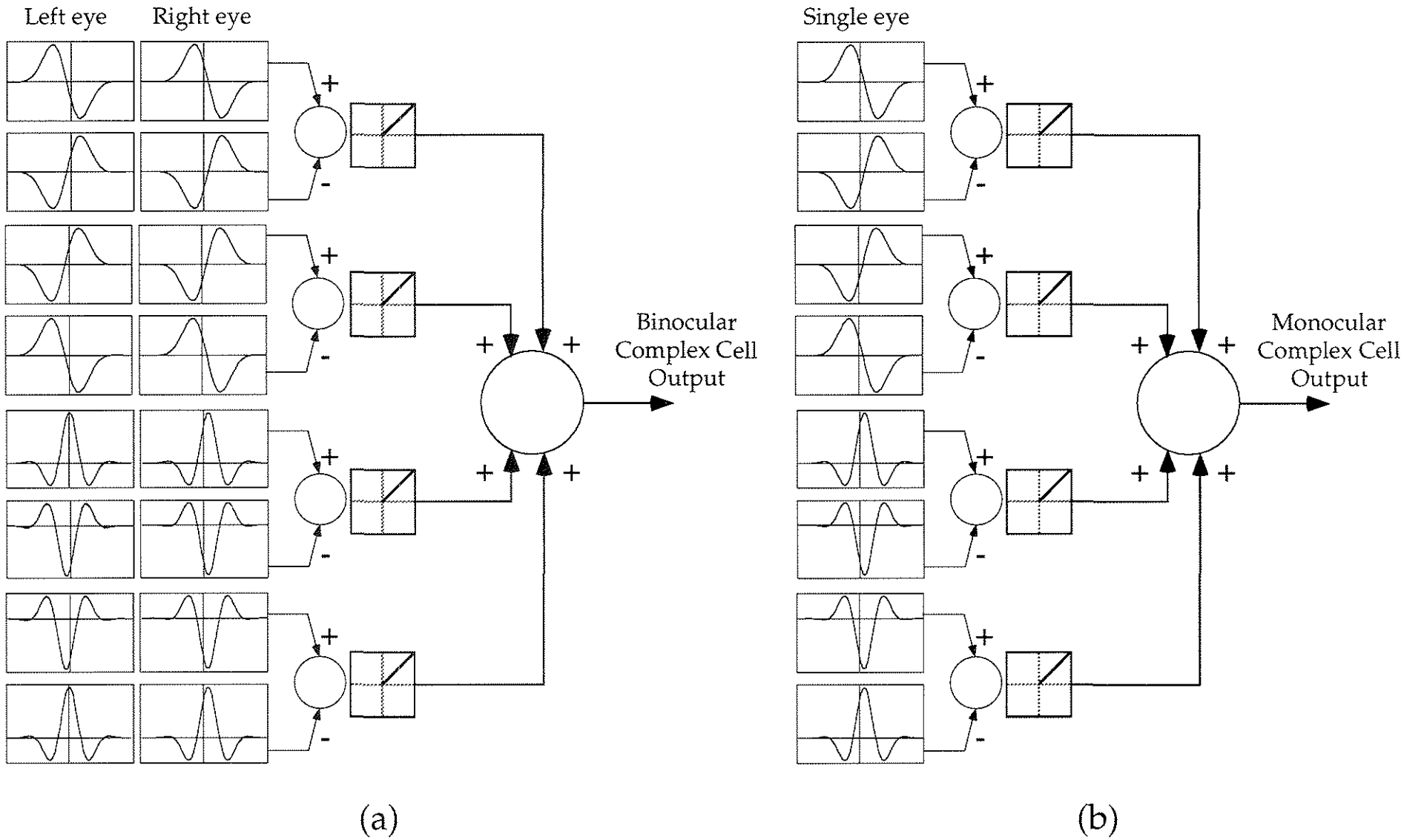
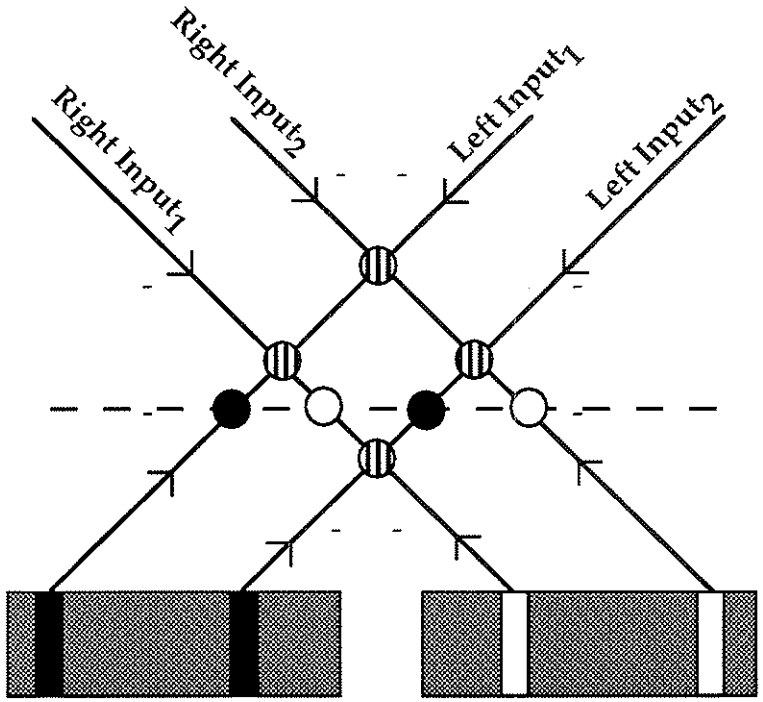
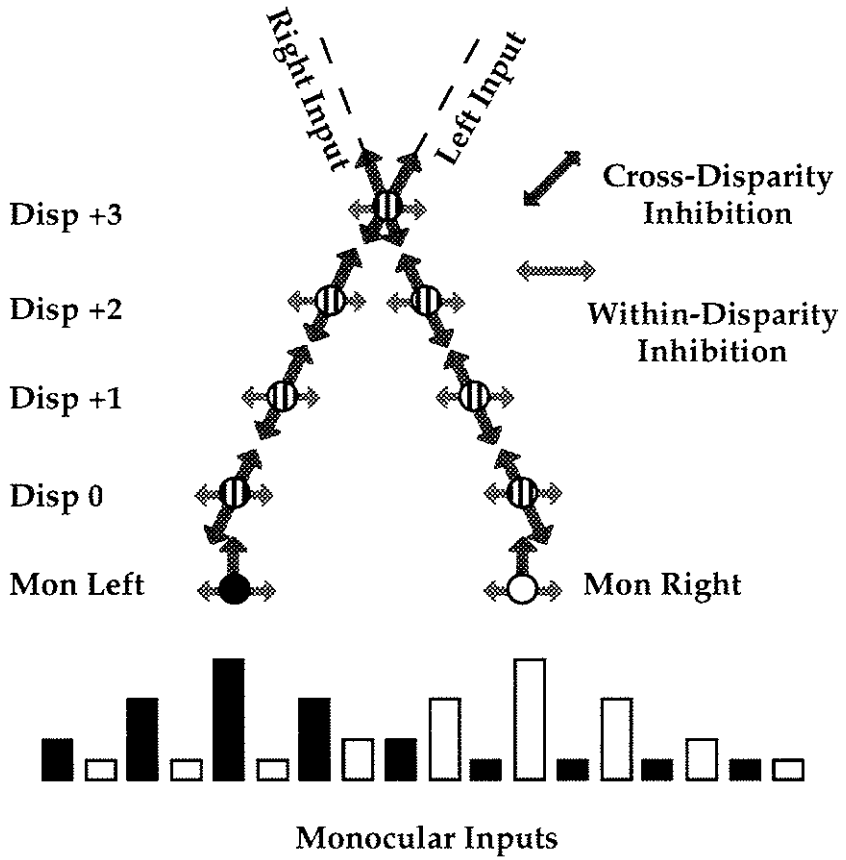

Left eye
Right eye
Binocular
Complex Cell
Output
(a)
Single eye
Monocular
Complex Cell
Output
(b)

Figure 2

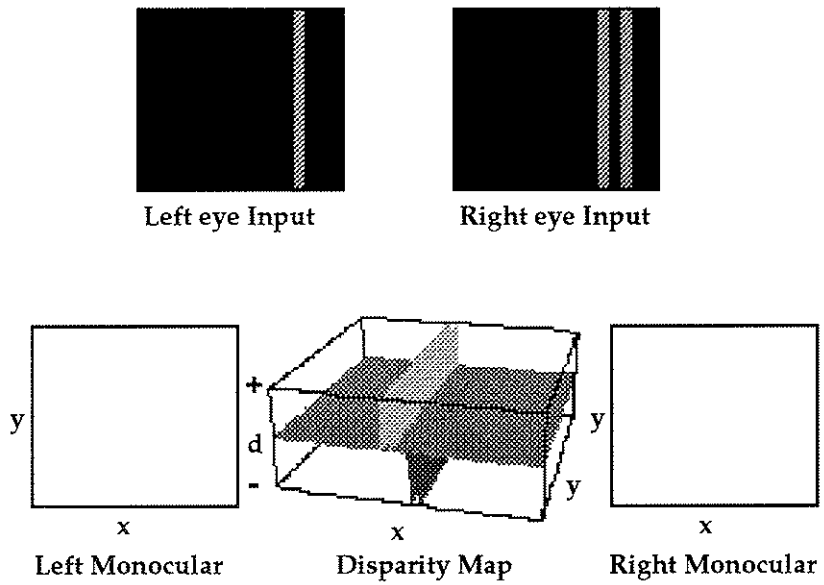

Right Input1
Right Input2
Left Input1
Left Input2
Left eye Inputs
Right eye Inputs

(a)

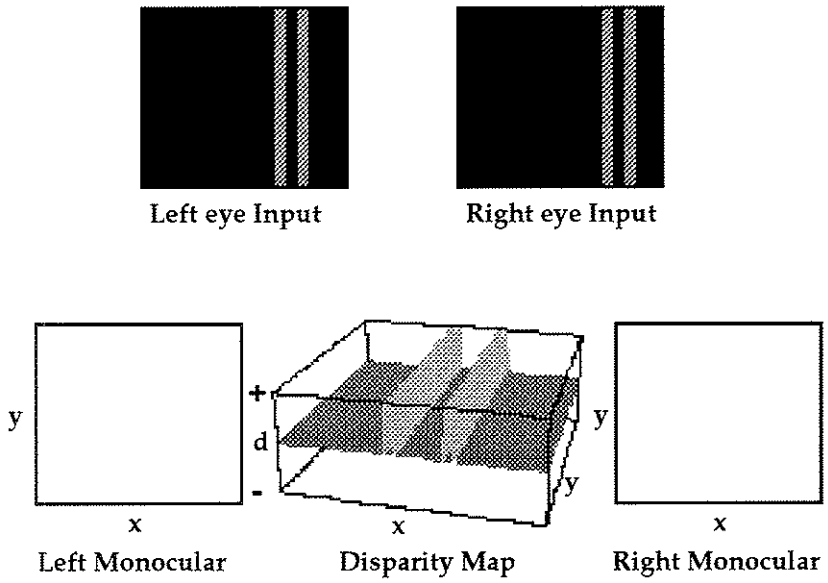

Right Input
Left Input
Cross-Disparity Inhibition
Within-Disparity Inhibition
Disp +3
Disp +2
Disp +1
Disp 0
Mon Left
Mon Right
Monocular Inputs

(b)

Figure 3


Left eye Input
Right eye Input
y
x
+
d
-
y
x
y
x
Left Monocular
Disparity Map
Right Monocular

(a)


Left eye Input
Right eye Input
y
x
+
d
-
y
x
y
x
Left Monocular
Disparity Map
Right Monocular

(b)

Figure 4

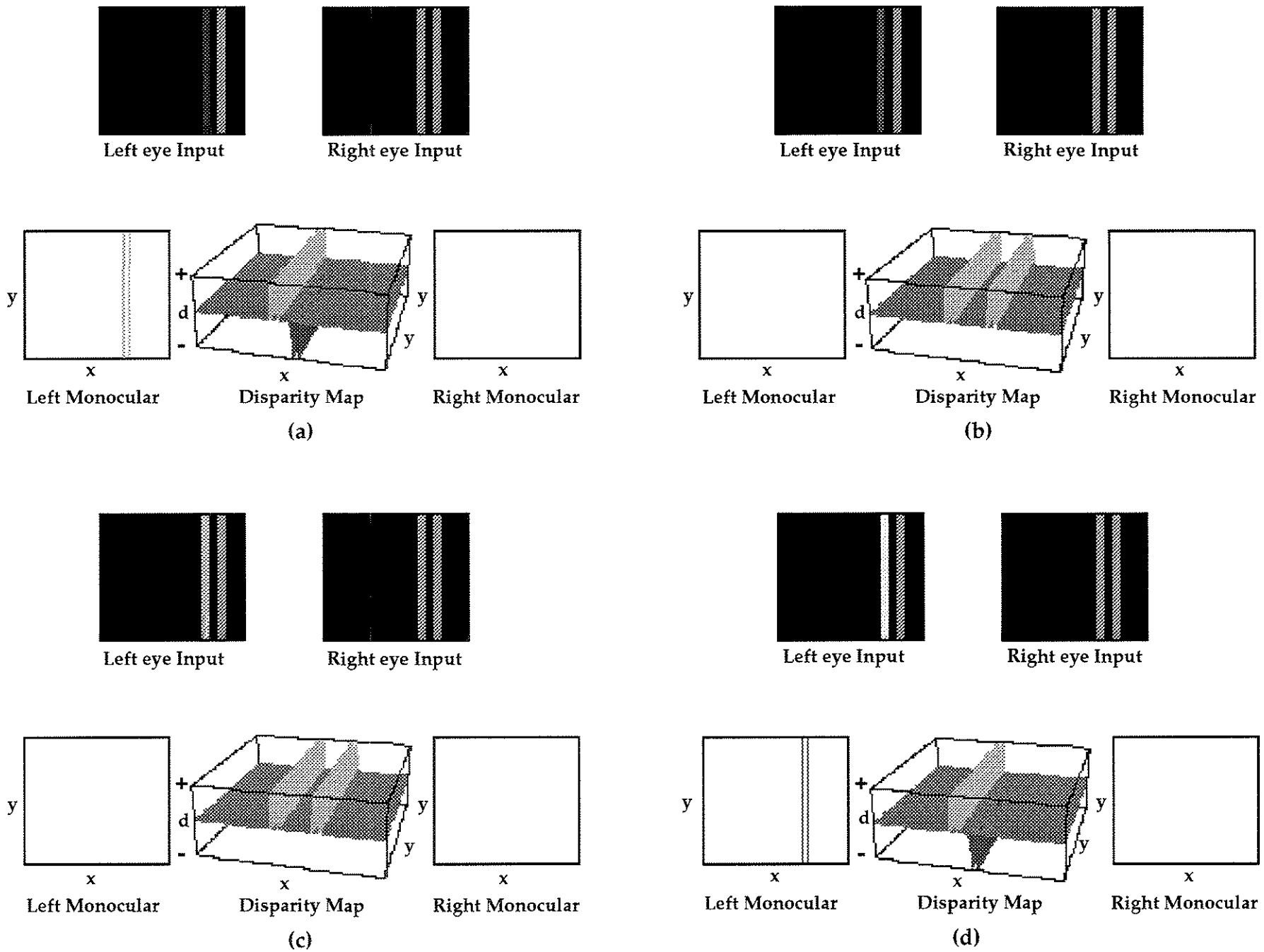

Left eye Input
Right eye Input
y
x
Left Monocular
+
d
-
y
y
x
Disparity Map
x
Right Monocular
(a)
Left eye Input
Right eye Input
y
x
Left Monocular
+
d
-
y
y
x
Disparity Map
x
Right Monocular
(b)
Left eye Input
Right eye Input
y
x
Left Monocular
+
d
-
y
y
x
Disparity Map
x
Right Monocular
(c)
Left eye Input
Right eye Input
y
x
Left Monocular
+
d
-
y
y
x
Disparity Map
x
Right Monocular
(d)

Figure 5

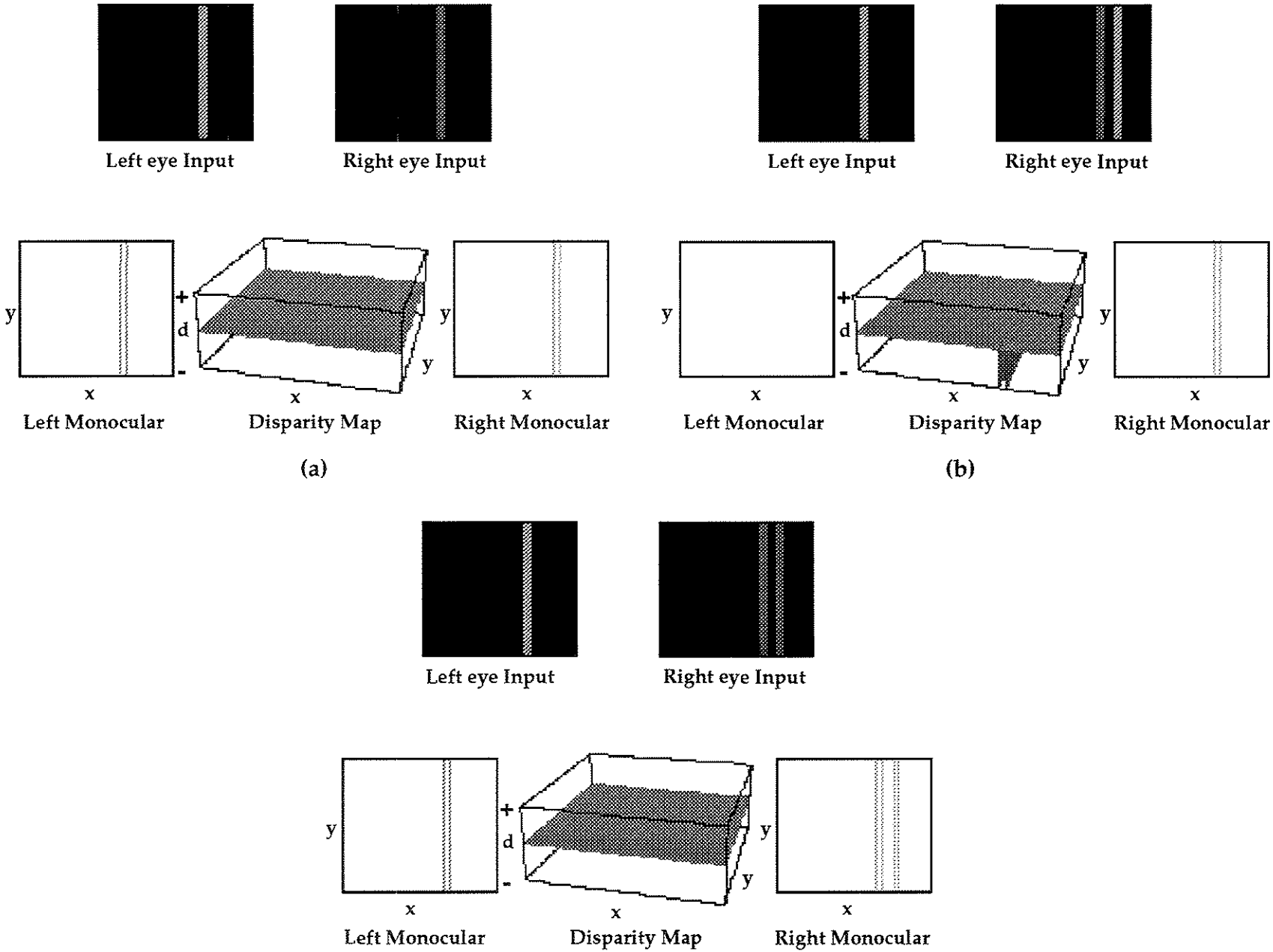

Left eye Input
Right eye Input
Left eye Input
Right eye Input
y
x
+
d
-
y
x
y
x
Left Monocular
Disparity Map
Right Monocular
(a)
y
x
+
d
-
y
x
y
x
Left Monocular
Disparity Map
Right Monocular
(b)
Left eye Input
Right eye Input
y
x
+
d
-
y
x
y
x
Left Monocular
Disparity Map
Right Monocular
(c)

Figure 6