Neural network–expert system models of relative-depth perception

STEPHEN J. GOTTS and FREDERICK J. BREMNER
Trinity University, San Antonio, Texas

Degree of binocular, horizontal disparity was used by two hybrid neural network/expert system computer models to make relative-depth judgments for pairs of stimulus points. These judgments were then correlated with the actual depth relationships of the points. Results from Simulation 1 showed that horizontal disparity could be computed by the shift in activated cortical hypercolumns evoked by a particular stimulus, and that, in general, multiple disparities could be compared to make accurate judgments about relative depth. However, these results also indicated that stimuli toward the periphery of the visual field were inaccurately perceived as being more distant. Simulation 2 corrected for this inaccuracy by appropriately weighting a stimulus point's disparity value as a function of its horizontal position in the visual field.

Our ability to perceive and judge depth stems largely from the horizontally differing perspectives of two eyes; each perspective provides a slightly shifted view of the world. Many theorists (e.g., Barlow, Blakemore, & Pettigrew, 1967; Hubel & Wiesel, 1962; Marr, 1982) have been interested in what the process of seeing with two eyes, known as *stereopsis*, entails and exactly how it allows us to perceive depth.

STEREOPSIS

Two distinct explanations of binocular depth perception evolved in psychophysical circles as early as 1850. Whereas David Brewster advocated a convergence theory, Charles Wheatstone supported a disparity theory (Gulick & Lawson, 1976). Brewster argued that the successive convergence of the optic axes on two points of discrepant distance is the process that we use to perceive depth. Wheatstone, on the other hand, believed that the difference between the images formed on the retinae of the two eyes results in the perception of depth. Although these two arguments seem to force disparity and convergence into a state of mutual exclusiveness, it is likely that both concepts function in concert to provide depth cues (Foley, 1980). The optic axes of convergence must cross at some point in the visual field, forming a depth reference area known as the *horopter* (Ogle, 1964). Relative depths of visual stimuli positioned off the horopter are probably judged by comparing horizontal disparities.

Two simulations were designed using a hybrid neural network–expert system approach in order to show how convergence and horizontal disparity might work together to provide cues that are helpful in making relative-depth judgments. In both simulations, the optic angles of convergence were held theoretically fixed. Stimuli were then chosen from the set of points located off the horopter, making it possible to assess the accuracy of judgments based solely upon horizontal disparity.

SIMULATION 1

While it is clear that single cells in the visual cortex are disparity selective (Barlow et al., 1967), it is still unclear exactly how this information is used by the brain at later stages of computation. Several theorists (DeAngelis, Ohzawa, & Freeman, 1991; Guillemot et al., 1993; Nomura, 1994) have proposed models suggesting uses of the cues provided by these binocular neurons. The hypothesis underlying our neural network–expert system model is that the cortex, in order to make relative-depth judgments, performs the equivalent of a difference calculation, given the horizontal shift in activated cortical hypercolumns (Hubel & Wiesel, 1962) due to stereopsis. Although this strategy is an obvious departure from the established view, it provides a computationally simple mechanism inspired by the basic topographic arrangement of the hypercolumns.

Method

Apparatus
An 8-Mb Compudyne (IBM-compatible) 486DX personal computer operating at 33 MHz was used to implement the simulation. The back-propagation neural network was written in Microsoft C/C++ Version 7.0 (Microsoft Corporation, 1991). Training shapes were generated from a diagram of the left and right visual hemifields found in Kelly (1985).

Input Layer
The input patterns to the network represented the activations of retinal ganglion cells caused by stimulus points in the visual field.

Correspondence should be addressed to S. J. Gotts, 2555 NE Loop 410 #1213, San Antonio, Texas 78217 (e-mail: sgotts@trinity.edu).

Julesz's (1971) work with random-dot stereograms established the legitimacy of using single-point stimuli in studies of depth perception. Only points from the binocular, right visual hemifield were used to train the network, as this was sufficient to demonstrate the feasibility of the proposed mechanism. The input layer to the network consisted of a vector 80 units long, the first 40 units representing the ganglia of the nasal hemiretina of the right eye, and the second 40 units representing the ganglia of the temporal hemiretina of the left eye. Each unit of the vector corresponded to a sector of one eye's visual field and symbolized one retinal, X, on-centered ganglion cell. X retinal ganglia seem to be involved in perceiving detail, while the other two main types of retinal ganglia—Y and W—seem to be more involved in detecting motion in the visual field and in providing information necessary to move the eyes (Sterling, 1990). As no motion is involved in the current task, the X pathway is clearly the most germane of the three.

Hidden Layer

The hidden layer of the neural network consisted of a vector 125 units long. The closest neurological correlate to this layer of neurons is the lateral geniculate nucleus (LGN), which relays information from the retina to the cortex while maintaining contralateral and ipsilateral pathways (Kelly, 1985). It is important to mention, however, that this analogy is somewhat loosely constructed. The connections with the hidden layer are not constrained to maintain the same special lateral relationships: A single hidden neuron is connected to all the neurons at the input and output layers.

Output Layer

The output layer of the neural network consisted of a 2×20 matrix representing hypercolumns found in Area V1 of the visual cortex (Kandel, 1985). The two rows of the matrix represented the contralateral and ipsilateral dimensions of each hypercolumn, while the 20 columns denoted each of the 20 hypercolumns.

Procedure

Network training. The back-propagation neural network was allowed to self-organize until it correctly mapped the entire set of 144 input points to corresponding outputs. The network converged as soon as the error for each output unit was less than .10. As there was no danger of activation versus inactivation misclassification (because .8 was the lowest possible value for an active classification, whereas .2 was the highest possible value for an inactive classification), the .10 error level was assessed as being sufficiently low.

Network testing. After the training phase, a set of 20 running-fact pairs was presented to the model. Each of the 20 pairs consisted of two running facts, representing two stimulus points of discrepant depth. Each of the first 10 running-fact pairs involved points with dissimilar horizontal positions, while the last 10 each involved points with similar horizontal positions.

Expert system. For each of the stimulus points, the hypercolumn discrepancies were calculated by the formula (discrepancy = column number of hypercolumn activated in ipsilateral row − column number of hypercolumn activated in contralateral row). The difference values were compared using a series of inequality formulae, and judgments were made on the basis of these formulae. If the first point's disparity was greater than that of the second point, the first point was judged closer (a judgment value of −1). On the other hand, if the second point's disparity was greater than that of the first, the second point was judged closer (a judgment value of +1). Points eliciting the same disparity values were judged as having equal distance (a judgment value of 0). The system's relative-depth judgments were written to a file, and the judgments were then correlated with the correct depth relationships. Figure 1 shows a logical flowchart for the hybrid model and
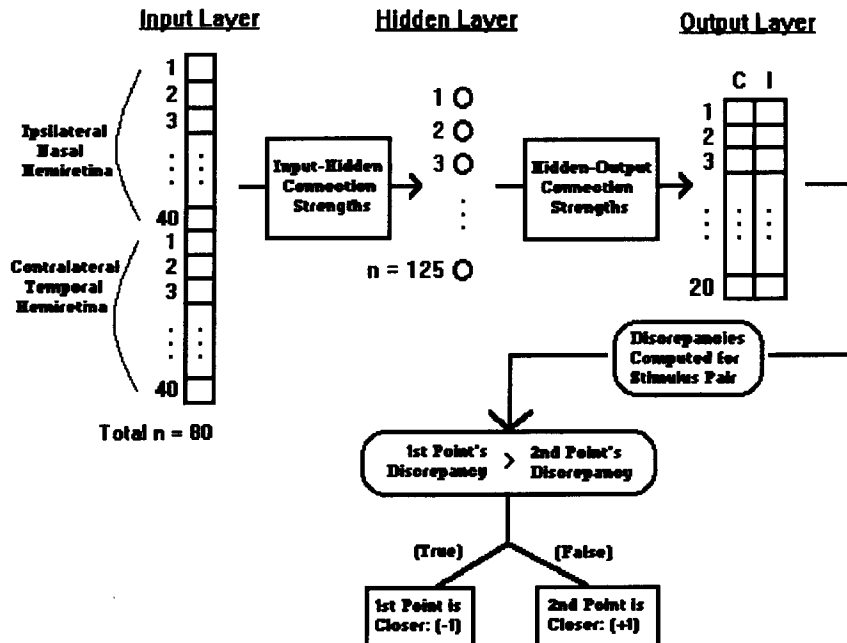

Figure 1. Flowchart of the hybrid model. The neural-network component interacts with the expert-system component by supplying it with the hypercolumnar activation information it needs to make its judgments.

clarifies the relationship between the neural-network and expert-system components.

Results and Discussion

The neural network was able to converge and correctly map all input patterns to their corresponding output patterns with error less than .10. The total training time was 12 min; 36 passes of the training facts were necessary. The Pearson correlation between relative-depth judgment data and correct depth relationships was found to be significant [$r(18) = .48, p < .05$]. These results establish that the proposed mechanism of relative-depth perception is indeed feasible. The judgment values and actual depth relationships are provided in Table 1.

While the model perfectly discriminated relative depth for points from the last 10 running-fact pairs (which involved comparing points from similar horizontal positions), it had trouble judging points from the first 10 pairs (which involved comparing points from dissimilar horizontal positions). Note, in particular, the lack of correspondence for Pairs 2, 5, 6, 8, and 10. The small inaccuracy of the model was due to a slight shift of the stimulus points for a given hypercolumn disparity toward the observer as the visual hemifield was traversed from left to right (refer to Figure 2). The shift is most pronounced at the far right, or periphery, of the hemifield. When a close point near the right border of the binocular hemifield was compared with a more distant point near the left border, the expert-system component sometimes mistakenly judged the far-left point as the closer of the two presented. However, when presented with points having similar horizontal positions, the system performed flawlessly. The shift is clearly present in any system that consists of two overlapping, polar-coordinate systems, each having a unique origin. The human brain obviously employs such a system. However, whether the human brain compensates for this shift is unclear.

SIMULATION 2

An inaccuracy such as the one described in Simulation 1 is neither desirable nor necessary. The shift phenomenon is the direct result of employing overlapping, polar-coordinate systems, and it can be corrected easily with a simple weighting strategy. The hybrid model in this simulation combines the strategy employed by the first simulation with a function that corrects each stimulus point's disparity value on the basis of its horizontal position in the visual field.

Method

Apparatus

The machine used in the current simulation was the same as that used in Simulation 1. Additionally, the C computer code was the same as that written for Simulation 1, with the exception of the difference calculation; the exact calculation is described later (see Procedure).

Network Design

To determine whether compensating for the shift was improving the system's relative-depth judgments, the design, training facts, and running facts from Simulation 1 were used.

Procedure

As in Simulation 1, the neural network went through a period of training to learn the relationship between stimulus points in the right visual hemifield and activation of hypercolumns in the left visual cortex. As soon as the network was able to map correctly each of the 144 training points to its corresponding hypercolumn activations with error less than .10 on each output, the network was presented with the set of 20 running-fact pairs, for which the teacher values were not provided. For this simulation, a different formula from the one given in Simulation 1 was used to calculate the hypercolumn discrepancies (disparity = column number of activated ipsilateral row − column number of activated contralateral row). The column numbers (from the output matrix) of the activated hypercolumns for each point were left unaltered. However, the ipsilateral column number was weighted according to which horizontal position was involved (the contralateral eye was arbitrarily selected to observe varying horizontal positions; the ipsilateral eye could have been used with equal facility). If a point was located on the right side of the hemifield, its ipsilateral column number was weighted more than if it had been located on the left side. This larger weighting would result in a larger disparity value, effectively compensating for the visual shift. In Simulation 1, when a point from the left side of the visual hemifield was presented with an equidistant point from the right side, the point on the right was actually judged as being more distant because its disparity value was too small. By increasing its value with the weighting strategy, it was moved closer to the visual-shift line for that particular disparity, and the depth judgment was corrected. Refer to Figure 2 for a graphic explanation. The formula for computing any stimulus point's disparity was given by [disparity = ipsilateral column ∗ (1 + contralateral column ∗ shift) − contralateral column], where shift = .004611. It is important to note here that this formula, in itself, is trivial; it enabled the authors to establish the

Table 1
Judged and Correct Depth Relationship Values for Simulations 1 and 2

| Running-Fact No. | Judgments | | Correct Relationship |
	Model 1	Model 2	
1	−1	−1	−1
2	+1	−1	−1
3	−1	−1	−1
4	−1	−1	−1
5	+1	−1	−1
6	0	−1	−1
7	−1	−1	−1
8	0	−1	−1
9	−1	−1	−1
10	0	−1	−1
11	−1	−1	−1
12	−1	−1	−1
13	−1	−1	−1
14	−1	−1	−1
15	−1	−1	−1
16	−1	−1	−1
17	−1	−1	−1
18	−1	−1	−1
19	−1	−1	−1
20	−1	−1	−1

Note—−1 indicates that the first point is closer; +1 indicates that the second point is closer; 0 indicates that the points have the same depth.

Line along which
discrepancy values
are the same

Line along which
all points have
equal depth

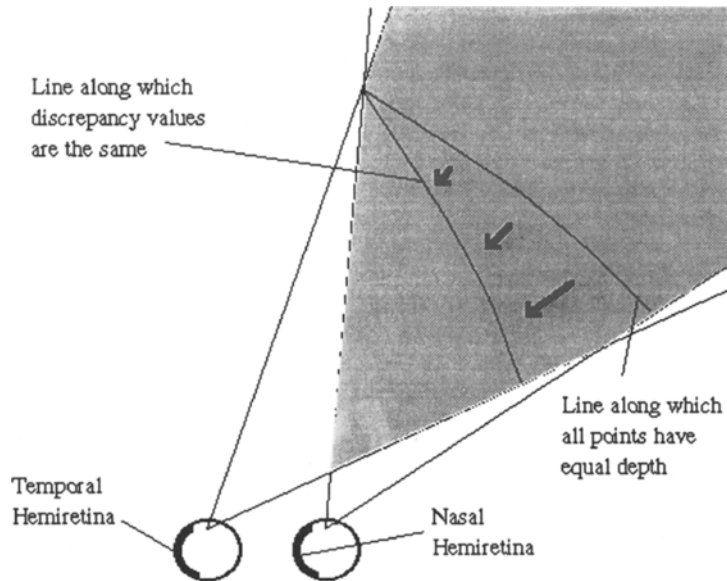Temporal
Hemiretina

Nasal
Hemiretina

Figure 2. A diagram of the line along which the model in Simulation 1 perceives points as having the same depth. Note the shift of the line toward the observer at the right edge of the visual hemifield. The correction, made by the model in Simulation 2, allows for more accurate relative-depth judgments.

feasibility of our particular disparity computation. The shift problem addressed by the new formula is one of employing stereopsis as a mechanism for relative-depth perception, not just a problem of our particular computation. Any system that accurately judges relative depth must somehow solve the shift problem.

Once computed, the new disparity values were compared using the inequality formulas from Simulation 1 to judge relative depth for each pair of stimulus points. The system tested to see whether the first point's disparity was greater than that of the second point. If so, the first point was judged closer; otherwise, the second point was judged closer. There was no real chance that the system would judge the points as equidistant, on account of the acuity of the weighting. It would only judge two points as having equal distance if they were the same point. The system's relative-depth judgments were written to a file and then correlated with the correct depth relationships.

Results and Discussion

The neural network implemented in Simulation 2 was able to converge and correctly map all input patterns to their corresponding output patterns with error less than .10. The total training time for this version of the network was 15 min; 40 passes of the training facts were necessary. These varied slightly from the training results found in Simulation 1, simply because the pseudorandom number algorithm generated a different set of values every time the program was executed; the algorithm used the computer's clock time as a unique seed value. Different random numbers were assigned as hidden weight values in this simulation, which resulted in a slightly different training time.

The Pearson correlation between the system's relative-depth judgments and the correct depth relationships was found to be highly significant $[r(18) = 1.00, p < .01]$. The system was 100% accurate at making relative-depth

judgments (an apparent improvement on the accuracy of Simulation 1). The data from the current simulation (see Table 1) establish a successful modeling of the perception of relative depth. Furthermore, the data establish that our hypothesis concerning how disparity is computed is entirely feasible.

GENERAL DISCUSSION

Use of a hybrid neural network/expert system approach in this particular problem is valuable because it allows us to model both the visual-tract physiology and the higher cognitive processing of depth cues in the most attractive way possible. While not clear for single-point stimuli, the benefit of the neural network component will be fully realized when more complex stimuli are involved. Future iterations of the model will be able to render disparity values easily, while a more rule-based approach would be extremely tedious to implement and less physiologically valid. The rule-based approach of the expert-system component, however, makes quick and explicit decisions on the basis of the information fed to it by the neural network. These decisions are made in a way that more closely mimics the higher cognitive processes of the human brain. In this particular setting, the hybridization of the neural-network and expert-systems approaches yields a useful model demonstrating that a hypercolumnar-disparity computation is indeed feasible.

REFERENCES

BARLOW, H. B., BLAKEMORE, C., & PETTIGREW, J. D. (1967). The neural mechanism of binocular depth discrimination. *Journal of Physiology*, **193**, 327-342.

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1991). Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature*, **352**, 156-159.

Foley, J. M. (1980). Binocular distance perception. *Psychological Review*, **87**, 411-434.

Guillemot, J.-P., Paradis, M.-C., Samson, A., Ptito, M., Richer, L., & Lepore, F. (1993). Binocular interaction and disparity coding in area 19 of visual cortex in normal and split-chiasm cats. *Experimental Brain Research*, **94**, 405-417.

Gulick, W. L., & Lawson, R. B. (1976). *Human stereopsis: A psychophysical analysis*. New York: Oxford University Press.

Hubel, D. H., & 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.

Kandel, E. R. (1985). Processing of form and movement in the visual system. In E. R. Kandel & J. H. Schwartz (Eds.), *Principles of neural science* (2nd ed., pp. 366-383). New York: Elsevier.

Kelly, J. P. (1985). Anatomy of the central visual pathways. In E. R. Kandel & J. H. Schwartz (Eds.), *Principles of neural science* (2nd ed., pp. 356-365). New York: Elsevier.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman.

Microsoft Corporation (1991). *Microsoft C/C++* (Version 7.0) [Computer program]. Redmond, WA: Author.

Nomura, M. (1994). A model for neural representation of binocular disparity in striate cortex: Distributed representation and veto mechanism. *Biological Cybernetics*, **69**, 165-171.

Ogle, K. N. (1964). *Researches in binocular vision*. New York: Hafner.

Sterling, P. (1990). Retina. In G. M. Shepard (Ed.), *The synaptic organization of the brain* (3rd ed., pp. 170-213). New York: Oxford University Press.