Cooperative Stereo Matching Using Static and Dynamic Image Features

M.A. Mahowald and T. Delbrück
Department of Computer Science
California Institute of Technology
Pasadena, California, 91125
e-mail: stereo@hobiecat.caltech.edu

Visual experience is intrinsically subjective. The manifest unity of our perceptions belies the indeterminacy of our sensations. A single pattern of excitation on the retinal receptors is consistent with many possible worlds of objects. The simple problem of determining object brightness exemplifies the ambiguities inherent in the retinal image. The photons incident on a given receptor may indicate the presence of a bright or dark object depending on the overall level of illumination. In mediating automatic gain control, the horizontal cells of the retina express an assumption about the nature of the illuminant. These neurons in the most peripheral part of the visual system take the first step toward interpretation of the image. By a process of lateral inhibition, they discount the effect of illumination and determine the brightness of an object relative to that of nearby objects. At all levels of complexity, the visual system interprets each data point within the context of the scene. This interpretation is consistent with the input data and with the internal structure of the system. Through evolution, the structure of the visual system provides correspondence between the mental image and the objective world.

As a result of the work of many engineers and scientists, artificial visual systems are also evolving. System synthesis can lead to a better understanding of natural systems, since it demands a concrete formulation of the relationships among representation, system architecture, and the visual world. Marr, who pioneered a computational approach to vision, has been a major influence in this field (Marr, 1982). He viewed vision as a form of information processing that must be based on constraints that come from consideration of the visual world. He described a process as taking place within a definite representational framework according to a specific algorithm.

In 1976, Marr and Poggio proposed a collective algorithm for stereopsis that could perform feature matching between two images (Marr, 1976). The crux of the problem is that images with dense, homogeneous textures contain many identical features. Each feature can be matched to several features in the corresponding image, but only one match is correct. It is not possible to find the correct match without considering several features simultaneously. Constraints

derived from an analysis of space, objects, and projective geometry specify interactions among features that permit identification of a correct match. These constraints are based on assumptions about the nature of the visual world.

Marr's algorithm describes a locally connected, fedback, nonlinear system suitable for realization in an electronic medium. The system architecture embodies the assumptions of the algorithm. Using this architecture, we have synthesized an analog CMOS circuit that finds regions of correspondence between two one-dimensional images in real time. Experiments on simplified images demonstrate the circuit's tolerance to transistor mismatch resulting from the collective nature of the algorithm.

We have fabricated and tested two chips that use this correspondence circuit. One of these chips performs stereo matching based on static contrast features in the image. The other chip takes advantage of natural time constants of the analog system, and uses time derivatives of intensity to drive the correspondence circuitry.

Evolving electronic artificial vision increases our awareness of the role of the physical nature of the system. Algorithms for vision must be consistent not only with the external world, but also with the properties of the computational medium.

STEREOPSIS

The review presented here will be brief. A more complete description of the problem of stereopsis can be found in (Julesz, 1971; Poggio and Poggio, 1984).

Binocular vision generates two images of a scene, one from each eye. Because the two eyes regard the scene from different points of view, they may differ in their impression of the relationships between objects. Figure 1 shows two eyes of an observer in cross-section. The lens of the eye projects an image of targets in three-dimensional space onto the the surface of the retina. The position of a feature in the projected image depends on the visual direction of the target. Visual direction is the angle of the line of sight between the retina and the target. When the eyes fixate on a point, the locus of points in space having equal visual direction in both eyes is called the *horopter*. All targets on the horopter are said to have zero disparity, because they project to corresponding points on the two retinas. Targets closer to the viewer than the horopter have crossed (negative) disparity, and targets more distant than the horopter have uncrossed (positive) disparity. The principle task of stereopsis is to determine the disparity between corresponding regions of the images from the two eyes. This information, along with the state of vergence of the eyes and the eyes' separation in the head, specifies the distance between the target and the viewer.

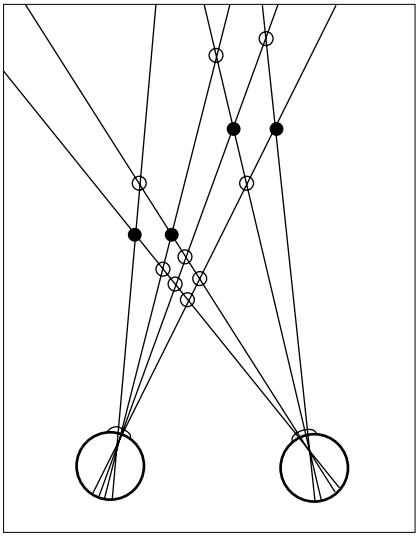


Figure 1 Stereopsis. This figure illustrates the projection of an image of four identical targets (dark disks) onto the right and left eyes of an observer. The lines going through the lens connecting each target with the retina are lines of sight. The intersections of the lines of sight indicate possible target positions in space. False targets (transparent disks) are located at the intersections of lines of sight that originate from different targets in the two eyes.

Calculation of disparity requires the determination of correspondence between features on the two retinas. This task would be straightforward if features could be identified uniquely. However, the random-dot stereograms developed by Bela Julesz (Julesz, 1960) demonstrate that the human visual system can compute disparity even when there are many identical features in close proximity (Figure 2). It has been shown that occlusion events are primary cues to the determination of depth (Shimojo *et al.* 1985). Researchers in artificial vision, however, have focused on the matching of stationary features visible to both eyes in attacking the problem of stereopsis (Poggio and Poggio, 1984).

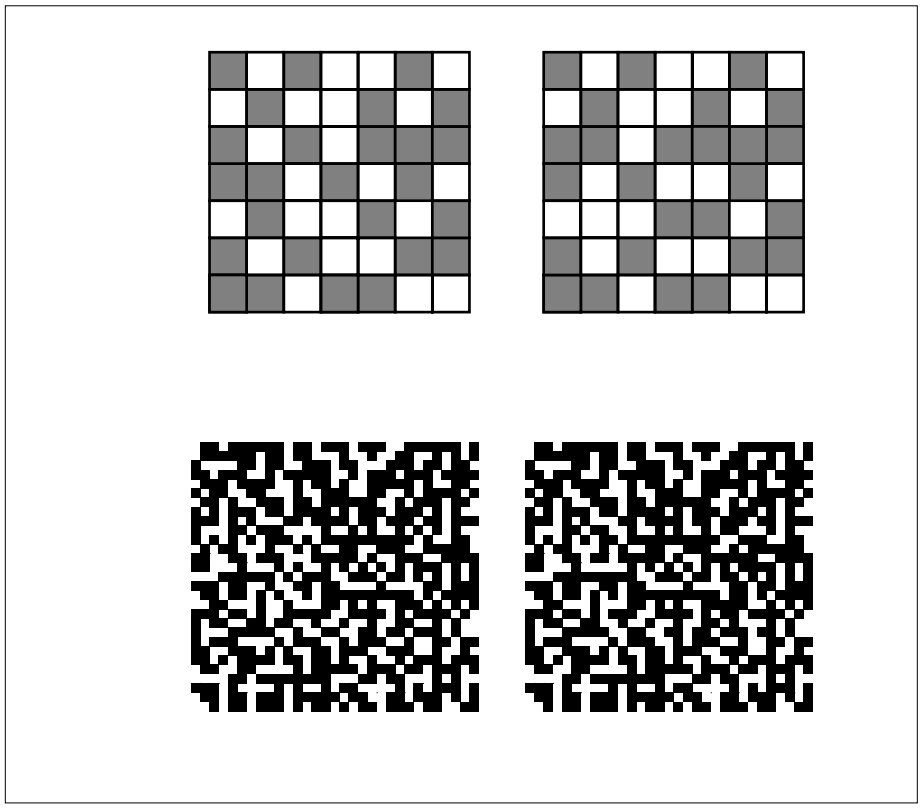


Figure 2 Random-dot stereograms. **(a)** Making a random-dot stereogram. A random pattern of 1s and 0s is generated to be presented to the left eye. An identical copy of the pattern is made for the right eye, except that a central square region within the image (labeled with As and Bs) is displaced to the right. When the two images are fused, this square region will appear closer than the background. Occluded areas (areas having no counterpart in the opposite eye's image) are labeled with X's and Y's. (Modified from (Julesz, 1971)). **(b)** A random-dot stereogram showing a raised square. You can fuse the stereogram by letting your eyes diverge as though you were looking at infinity. Your left eye should see the pattern on the left, and your right eye should see the pattern on the right. The primary difficulty is focusing on the paper while your eyes are diverged. Myopic readers will find it helpful to remove their glasses.

THE MARR/POGGIO STEREO ALGORITHM

Marr and Poggio (Marr, 1976) describe a collective stereo algorithm that succeeds in finding the disparity in random-dot stereograms. Their algorithm is designed to find the disparity in a region of limited depth around the horopter. The algorithm can be used for stereo matching in one- or two-dimensional images. One-dimensional stimuli can be used to test many of the basic features of a stereo matching algorithm, because the eyes are displaced from each other along a line; therefore, only disparities along a line contribute to depth perception. We can understand the algorithm by considering three simple rules derived from the properties of images and of physical surfaces. These rules allow the correct correspondence of features on two retinas to be found in the presence of false matches.

The first rule is that features in the two images must be similar to correspond to each other. For example, dark features within one image should correspond to dark features in the other image. This rule is called the *compatibility* constraint. Psychophysical evidence suggests that the human visual system obeys some form of compatibility constraint. For example, people are unable to fuse images of reversed contrast (Julesz, 1971).

The second rule is that a feature from one image should correspond uniquely to one feature from the other image. This constraint is derived from the fact that a point on a surface has only one spatial location at a given time. The *uniqueness* constraint is violated in the case of transparent surfaces, when an image feature is a combination of points from two physical surfaces.

The third rule is based on the observation that objects, being cohesive, occupy a localized region of depth. The *continuity* constraint used in the algorithm assumes that surfaces of objects are oriented parallel to the viewer so that changes in disparity will be rare, occurring only at surface boundaries.

The representational framework of the algorithm is depicted in Figure 3. Input to the algorithm is provided by two one-dimensional retinas that encode the horizontal positions of features in the image. In the output representation, real space is divided into a grid of discrete positions. The spatial dimensions of distance (disparity) and horizontal position are encoded by a two dimensional array of correlators. Activation of one of these correlators indicates the presence of target in the region of space to which it corresponds.

The constraints of the algorithm are embodied in the connections among elements. The compatibility constraint is implemented by feedforward input to the correlators from the retinas. The correlators receive two inputs, one from each eye. A correlator is stimulated by a nonlinear combination of its inputs; the two inputs must signal the presence of a similar feature in order to drive the correlator. False matches may be formed between features that do not correspond — features that are not actually generated by the same target in space. False matches are suppressed by feedback connections within the correlator array that implement the continuity and uniqueness constraints.

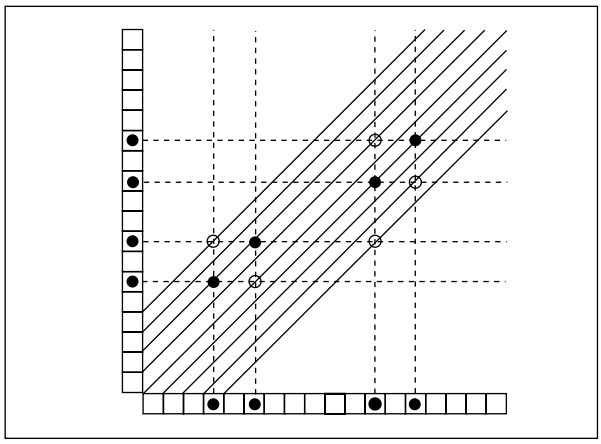


Figure 3 The representational framework of the cooperative algorithm. Depicted is the algorithm's response to the scene illustrated in Figure 1. The left and right retinas are represented on the vertical and horizontal axes. Correlators (not shown explicitly) are arranged in lines that are angled 45 degrees from the axes. Each line corresponds to a single disparity. The outputs of each retina are transmitted along lines of sight running perpendicular to the retinal axis. Each correlator receives input from the two pixels whose lines of sight intersect at its location. Wherever there are similar features from the right and left eye, they form a match (circle) at the intersection of their two lines of sight. Targets at the same distance in space form matches located on the same disparity plane. Open circles are false matches, and filled circles are true matches. Inhibitory interaction among correlation elements run along lines of sight (dotted lines). Solid lines along disparity planes indicate positive coupling between correlators.

Positive coupling between correlators at the same disparity encourages the continuity of the solution. The positive coupling is implemented in the original algorithm by a fixed set of connections among correlators in the same disparity plane. If a correlator is active, it drives the correlators to which it is coupled. Correlators at the correct disparity are all driven by retinal inputs with matching features and so tend to be active. Because correlators at the true disparity are surrounded by correlators that are likely to be active, they receive not only positively correlated retinal inputs but large inputs via positive coupling between correlators. The false matches receive less input from neighboring correlators.

The uniqueness constraint states that each feature from one eye can be matched to only one feature from the other eye. This constraint is implemented

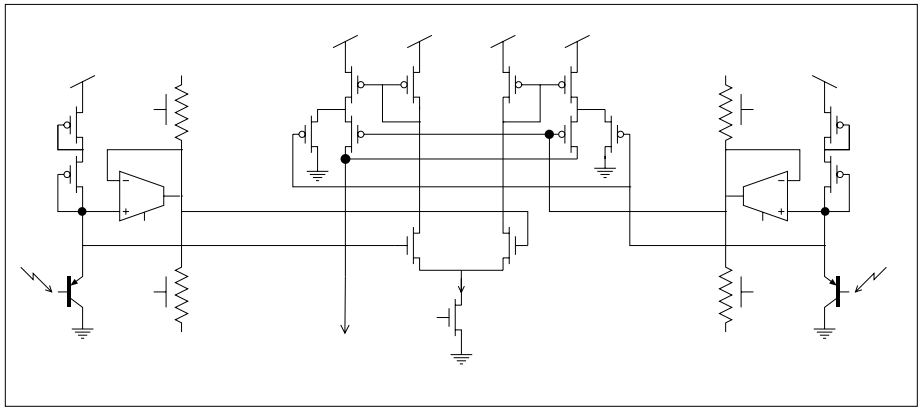
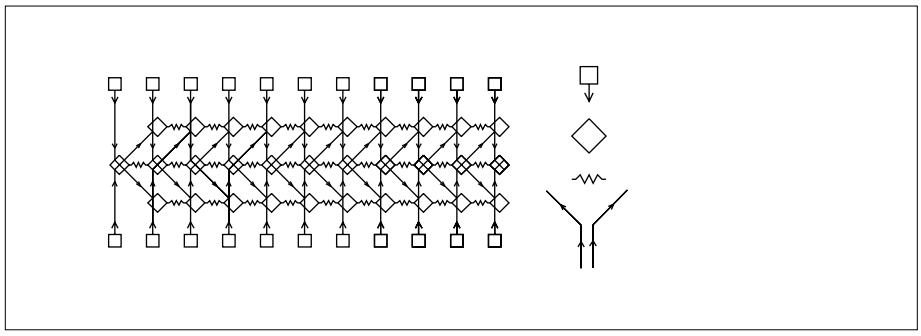


Figure 5 Retinal circuitry and modified Gilbert multiplier. The left pixel is located in the left retina, and the right pixel is located in the right retina. The multiplier is located in a correlator cell. Each correlator has its own multiplier, which receives input from a unique combination of pixels.

by an inhibitory interaction among correlators that receive input from the same pixel. This interaction performs a winner-take-all function that selects true matches by suppressing all but the maximally driven correlator. The final state of activation of the correlator array indicates the positions of genuine targets in space.

THE CHIP

The Marr/Poggio stereo correspondence algorithm is well suited to a circuit implementation because it requires only local connectivity. The system architecture leads to a physical structure that instantiates the representational framework of the algorithm. Figure 4 provides a simplified view of the chip architecture. The chip correlates the outputs of two one-dimensional retinas, and the representation of the solution expands into the second dimension of the silicon surface. The circuits are analog nonlinear elements that compute the correspondence between regions of the retinal images in real time.



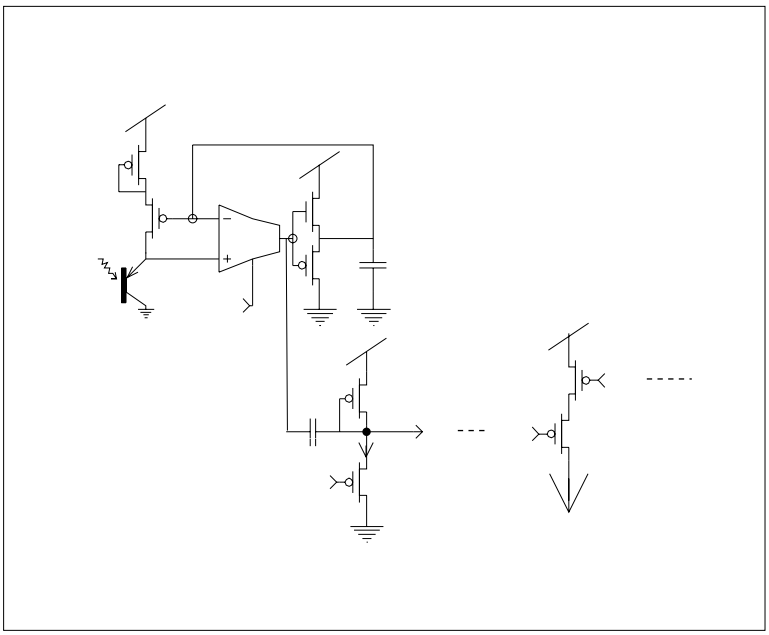


Figure 6 Time derivative pixel and series connected transistors. The half-wave current-rectifier circuit bias current is labeled I_b.

The multiplier performs a four quadrant multiplication biased around $\frac{I_b}{2}$:

$$I_{\text{mult}} = (\frac{I_b}{2})(1 + \tanh(V_{\text{r}}) \tanh(V_{\text{l}})).$$

V_{r} and V_{l} are the differential voltages from the right and left retinas. The resistive network acts as a reference level for the computation; intensities brighter than the average are "white," whereas those darker than the average are "black." The output of a black pixel multiplied by the output of another black pixel results in a current into the correlator that is greater than $\frac{I_b}{2}$, whereas the output of a black pixel multiplied by the output of a white pixel results in a current that is less than $\frac{I_b}{2}$. The magnitude of the signal coming into the multiplier is related to the contrast of the feature.

Another variant of the chip uses time derivatives of intensity as features for the correlation. The retinas consist of arrays of change-sensitive pixels, called *hysteretic receptors* (Delbrück and Mead, 1989). The response of one of these pixels to an increase in light intensity is a sharp and transient decrease in the output voltage. The response to a decrease in light intensity is very small. The hysteretic receptor has a gain for transients that is about 100 times larger than the gain for the steady-state illumination level.

The output of the hysteretic element is capacitively coupled to a half-wave current-rectifier circuit biased to generate a small current when the intensity is unchanging (Lazzaro, this volume). This bias also controls the time constant of the circuit. Even small increases in intensity increase the current through the rectifier. The voltage output of the rectifier is broadcast to the correlator array along a line of sight.

The correlation of signals from the right and left pixels is performed in the correlator cell by means of two serially connected transistors. In subthreshold, these two transistors compute the function:

$$I_{\text{mult}} = \frac{I_{\text{r}} I_{\text{l}}}{I_{\text{r}} + I_{\text{l}}}.$$

I_{r} and I_{l} are the currents through the rectifiers in the right and left pixels, respectively. This operation is a normalized multiplication of the two retinal inputs. If either retinal input is small, the current into the correlator is small. When the images on the retinas are unchanging, the input to the correlator is proportional to the bias current of the rectifiers. When there is a change in the input, the current into the correlator can be up to three orders of magnitude larger than the bias current (Figure 7).

The time-derivative chip has the advantage that the correlator cell is compact. In addition, the derivative-based correlator input has a range larger than the multiplier current in the static contrast chip. However, the operation of the correlator array is easier to conceptualize in the chip that uses static contrast as a matching primitive. Most of our discussion will focus on the operation of the static contrast-sensitive chip; description of the change-sensitive chip will be postponed until the section on experimental results.

The operation of the correlator array depends on cooperative interactions among correlators. The connections between correlators implement the continuity and uniqueness constraints of the original algorithm. Figure 8 schematically depicts the interactions at a correlation element. The correlator itself is simple; it is a single electrical node. The computations performed at a correlator node are complex: Currents from saturating nonlinear sources are summed to create a voltage; this voltage is used to control a nonlinear conductance; and the conductance determines the extent of electrotonic coupling. Although the number of interactions at this node are small by neural standards, the correlator circuit demonstrates the high computational density available in an analog medium.

The uniqueness constraint is implemented via a winner-take-all (WTA) circuit shown in Figure 9 (Lazzaro *et al.* 1989). The WTA circuit establishes a competitive feedback interaction between all the correlators along a line of sight. Each correlator will participate in two competitive groups; one group for a pixel from the left retina and another group for a pixel from the right retina. On the stereo correspondence chip, the number of channels participating in each WTA competition is equal to the number of disparity planes.

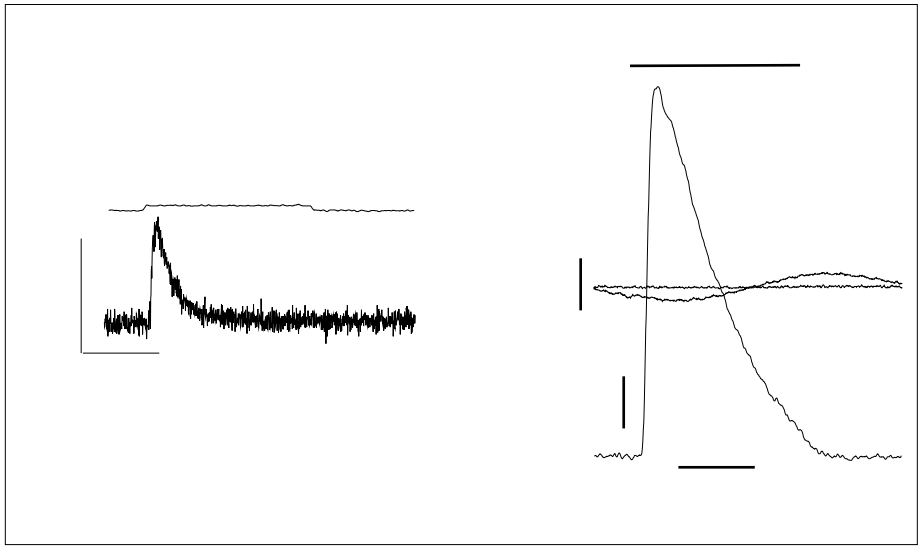


Figure 7 The retinal response and correlation performed by the motion-sensitive Marr chip. **(a)** The retinal signal generated by a pixel in response to a flashing LED. The LED signal is shown in the top trace. The magnitude of the modulation is 5 percent. The output of the pixel is shown in the bottom trace. The current maximum is approximately 4 nA. The zero input bias into the correlator is 100 pA. **(b)** Response of a single correlator to a flashing LED. The traces labeled "Right Only " and "Left Only" show the output of the correlator when only one retina is stimulated. The trace labeled "Both" shows the correlator response when both retinas are stimulated.

The competitive interaction is optimized to suppress false matches. Because competition takes place along lines of sight, all false matches are competing with true matches. A simple analysis of the current flowing through the WTA circuits of the correlators receiving retinal inputs shows that false matches can never suppress true matches.

The matching process can be broken down into competitions between two targets at a time. Imagine that there are two targets, A and B. Denote the pixel response to target A as a and the pixel response to target B as b. Figure 10(a) shows the system response to a two target stimulus in which $a > b$. Input to the correlator that correlates the response of pixel a_r in the right retina and pixel a_l

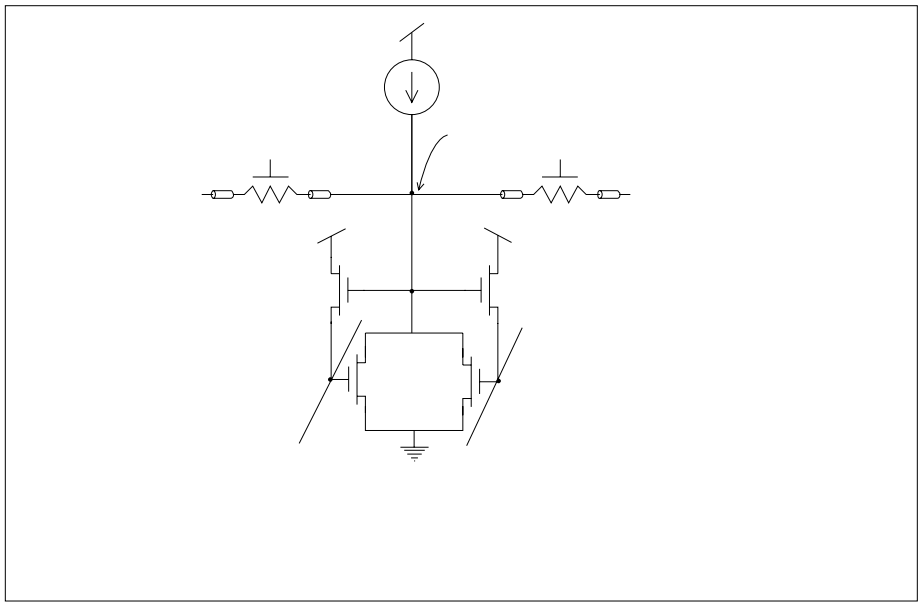


Figure 8 Simplified version of the electrical interactions at a correlation element. The correlator output voltage, $V_{i,j,d}$, is determined by the sum of the currents flowing into the node. The primary input, $I(i,j)$, is a current that is a nonlinear AND–type function of the signals from pixel i in the right retina and pixel j in the left retina. Correlation elements in the same disparity plane are coupled by resistive elements. The schematic of the resistive element indicates that the resistance is nonlinear and controllable. Two transistors drawing current to ground provide inhibition. The voltages V_i^{R} and V_j^{L} are a function of the correlator voltages along the line of sight of pixel i and pixel j respectively.

in the left retina is a^2. If $a > b$, this correlator sets the WTA circuits associated with $\mathrm{a_r}$ and $\mathrm{a_l}$, so that each WTA circuit sinks a current $\frac{a^2}{2}$ (a total of a^2). The retinal input to the false match between $\mathrm{a_r}$ and $\mathrm{b_l}$ is ab. The current available for this false match to suppress the true match between $\mathrm{b_r}$ and $\mathrm{b_l}$ is $ab - \frac{a^2}{2}$. The false match can suppress the true match only if $ab - \frac{a^2}{2} > \frac{b^2}{2}$. Figure 10(b) shows the amount of current by which the false match can suppress the true match, as a function of the ratio of the contrast of the targets, $\frac{a}{b}$. Intuitively, the false matches AB have a higher correlation input than the true match BB, but the suppression of the AB matches by the AA match is always sufficient to let the BB match win. The false match never has enough current to win

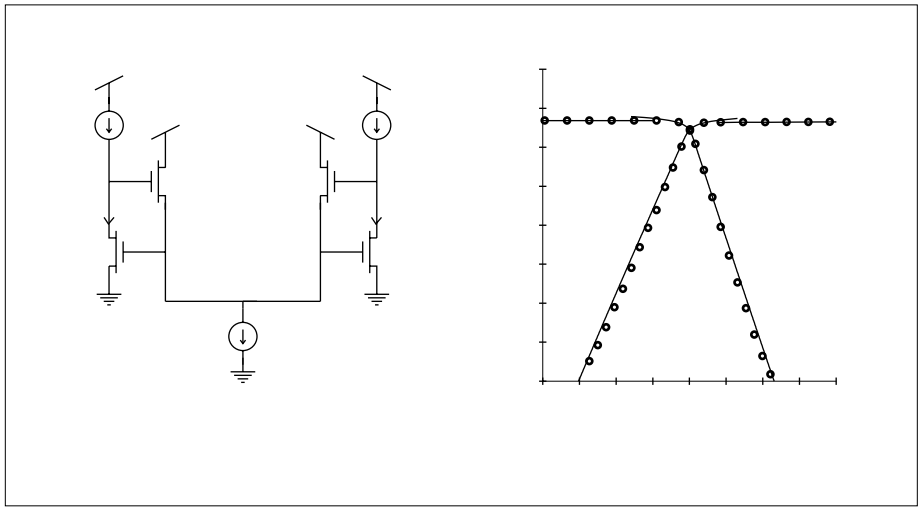


Figure 9 The winner-take-all circuit. **(a)** Schematic of a simple two-channel winner-take-all circuit. To understand how the circuit works in subthreshold, imagine that the circuit is in equilibrium and that each channel is receiving an identical input current. In this configuration, $I_1 = I_2 = I_{out_1} = I_{out_2}$. The voltage on the common line, V_c, is therefore constrained to be logarithmic in the input current. The voltages on the output nodes, V_1 and V_2, are constrained to supply the bias current to the common line through source-follower transistors, $T2_1$ and $T2_2$. These voltages are above the common line voltage by an amount that is logarithmic in the bias current. To make one channel win over the other, we increase its input current. Increasing the current to one channel charges up that channel's output node. The voltage on the common line will follow the output voltage of the winning channel with a voltage difference set by the bias current. The output node will stop charging when the current through its T_1 transistor is equal to the new input current. The output voltage of the winning channel will increase logarithmically with input current. The loser will be suppressed. **(b)** Current-voltage characteristic of the two channel WTA circuit. The voltage output of the two channels is plotted against the ratio of their input currents. Since the voltage on the common line, V_c, controls the current out of both channels, the capacitor of the channel with less current will be discharged until its T_1 transistor draws only its input current. For current differences between the channels of more than a few percent, the T_1 transistor of the losing channel will come out of saturation; the output voltage will be within a few $\frac{kT}{q}$ of ground. When the current difference between channels is small, the output voltage on the losing channel is determined by the Early voltage of the T_1 transistor and by the level of the input current.

over the true match. Correlations between two features of equal value present the most difficult case. False matches then generate the same correlator inputs

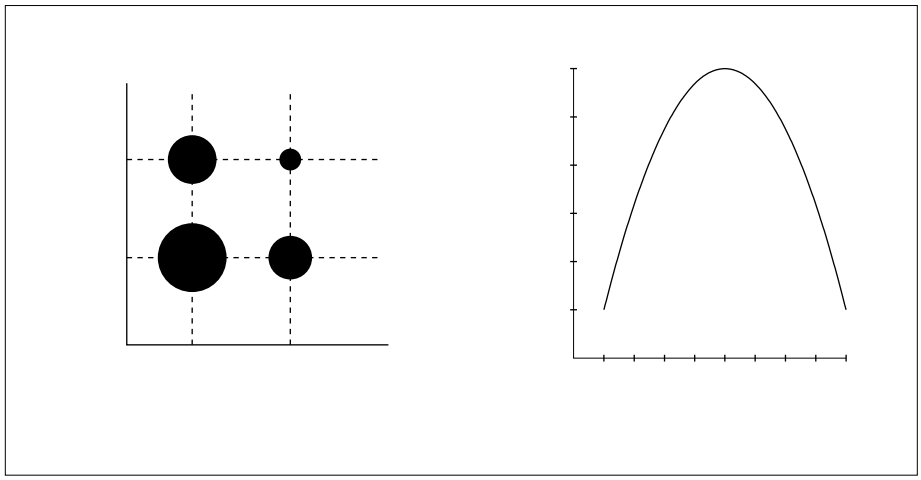


Figure 10 Suppressing false targets. **(a)** The computation of the correspondence between two targets, A and B, is illustrated using the representation explained in Figure 3. The magnitude of the retinal input to the correlator is shown as the area of a filled circle. Pixels associated with the lines of sight are labeled. **(b)** The function $f(\alpha) = -\frac{1}{2}\alpha^2 + \alpha - \frac{1}{2}$, where α is the ratio of the contrast of the targets, $\frac{a}{b}$. The function $f(\alpha)$ represents the amount of current by which the false match can suppress the true match in units of b^2. If the value of $f(\alpha)$ is negative, that much extra current would need to be supplied to the false match in order for it to tie with the true match in the WTA competition.

as do true matches. We can suppress these false matches only by invoking the continuity constraint.

The continuity constraint is mediated by resistive elements that couple the correlation nodes within a disparity plane. The resistor is implemented with the same circuit that is used in the static retina to compute the local average illumination level. The resistor circuit has the saturating current-voltage characteristic described by

$$I = G \tanh(\frac{V}{2}).$$

The units of voltage are $\frac{kT}{q}$. The parameter G is controlled by an external voltage.

In the correlator array, resistive coupling locally averages the activity level within a disparity plane by providing a path for current flow. Within a uniform disparity region of the image, the correlators at the correct disparity plane will all be receiving positively correlated retinal input. Other disparity planes will be receiving some anticorrelated input. A correlator that is receiving anticorrelated input will draw current from neighboring correlators within its disparity plane. The WTA circuit is thus able to suppresses the correlators in the disparity planes signaling false targets.

The strength of coupling must be limited if the chip is to function properly. The saturating characteristic of the resistor is important in allowing the disparity to change as a function of horizontal image position. The current that can be drawn from the winning disparity region across a disparity discontinuity is limited by the nonlinearity of the resistor. Saturation of the resistor allows a large voltage difference to form across the edge (Hutchinson, 1988).

We can write a system of equations that capture most of the operation of the system. These equations assume that the transistors in the circuits are being operated in subthreshold. The drain–source current I_{ds} of a transistor operated in subthreshold is given by:

$$I_{\mathrm{ds}} = I_0 e^{\kappa V_{\mathrm{g}} - V_{\mathrm{s}}} (1 - e^{V_{\mathrm{d}} - V_{\mathrm{s}}} + \frac{V_{\mathrm{d}} - V_{\mathrm{s}}}{V_0})$$

where V_{g} is the gate voltage, V_{s} is the source voltage, and V_{d} is the drain voltage. The constant I_0 is about 10^{-15} amps. The constant κ represents the body effect (Mead, 1989) and is about 0.7. The constant V_0 is the Early voltage and is typically a few tens of volts (Mead, 1989). All voltages are in units of $\frac{kT}{q}$.

The dynamical equation describing the interactions at a correlator node is

$$\begin{aligned} C\frac{d}{dt}V_{i,j,d} = {} & I(I_i^R, I_j^L) \\ & + G\tanh\Big(V_{i+1,j+1,d} - V_{i,j,d}\Big) \\ & + G\tanh\Big(V_{i-1,j-1,d} - V_{i,j,d}\Big) \\ & - I_0 e^{\kappa V_i^R}\Big(1 - e^{-V_{i,j,d}} + \frac{V_{i,j,d}}{V_0}\Big) \\ & - I_0 e^{\kappa V_j^L}\Big(1 - e^{-V_{i,j,d}} + \frac{V_{i,j,d}}{V_0}\Big) \end{aligned} .$$

In this equation, C is the capacitance on a correlator node. $V_{i,j,d}$ is the correlator voltage, and the subscripts refer to the inputs from pixel i on the right retina, pixel j from the left retina, and disparity plane d. The disparity planes run from $-d_0$ to $+d_0$, zero disparity referring to alignment of the two retinas. $I(I_i^R, I_j^L)$ is the correlation input from the right and left pixels. The next

two terms are the resistor currents to and from neighboring correlators at the same disparity. The constant G is controllable externally. The final two terms represent the nonlinear inhibition performed by the WTA circuit.

The common-line voltages on the winner-take-all circuits are given by

$$V_i^R = \ln\left(\sum_{d=-d_0}^{d=+d_0} \frac{I_0}{I_b} e^{\kappa V_{i,i-d,d}}\right)$$

for common lines emmanating from the right retina. A similar expression holds for the left retina. This equation represents a summation over a line of sight. The WTA bias current is denoted I_b.

It is difficult to gain a quantitative understanding of this system, since the circuit elements are nonlinear and extensively cross-coupled. For a given stimulus, the total positive current into each disparity plane is a fixed quantity, depending only on the retinal stimulus and not on the state of the correlator array. Current can only leave the disparity plane through the WTA circuits. The current–voltage characteristic of the WTA circuit determines the correlator voltage. The distribution of correlation input among the correlators in a disparity plane determines the state of the system. The spread of current within a single disparity plane is a function of the voltages of the correlators in the other disparity planes. In a simple resistive network, the space constant is set by the relationship between the lateral resistance and the conductance to ground. In the correlator array, the WTA inhibition sets the strengths of conductances between the output nodes and ground, based on the maximum correlator voltages along each line of sight. Consequently, the space constant of the network depends on the data. A current may propagate laterally for a long distance in one context, yet the same current may be quickly shunted to ground under another set of inputs. Context dependence makes the circuit (and the algorithm (Marr *et al.* 1978)) very difficult to characterize. This report describes the function of the chip by looking at specific examples.

EXPERIMENTAL RESULTS

Presentation of data to the system is simplified by the fact that photosensors are integrated on the chip; a single lens focused on the surface of the silicon projects an image onto two parallel, one-dimensional retinas. We can generate artificial disparities by using two images of bars, one for each retina. When the two images are identical, the shift of one image relative to the other determines the disparity. We can create images with multiple disparity regions by shifting portions of the images relative to each other. The output of the retinas and of the two dimensional array of correlators is scanned serially off the chip using

the method described in (Sivilotti *et al.* 1987). The output voltages of the correlators produce an output current by controlling the gate of a transistor. The output current is sensed by an off-chip current-sense amplifier.

The response of the chip to a simple pattern with two disparity regions is shown in Figure 11. The image is clearly segmented into two depth planes. In this example, the correlation is performed based on dense retinal signals. The space constant of the resistive network in the retina is set to be large, so that its principle function is to act as the reference value for determining positive and negative contrast. Because the stimulus is high contrast, most of the retinal outputs are saturated either high or low. Many false matches occur, since several consecutive pixels on the retinas are black or white.

To limit the number of false targets generated by large areas of uniform contrast, we can use the retina to enhance edges in the image. Figure 12 shows the chip's response to an edge-enhanced input. The computed disparity of the wide bar is continuous, even though the input is present mainly at the edges. When all the disparity planes are tied at a low value due to lack of retinal input, the current through the resistor can spread along a disparity plane to fill in the correct solution. When the disparity planes have similar retinal inputs, the impedance of a losing correlator node is set by the Early voltage of its T_1 transistor. If the lateral resistance is small, the space constant of the losing disparity plane is large, so the solution spreads a long way.

In addition to filling in the solution in regions of low retinal input, the resistors help to suppress false matches. Ideally a false match can never be bigger than a true match, so the resistors do not have to draw much current for the false matches to be suppressed. Unfortunately, the circuit elements we are using are not ideal. Transistor mismatches in the multipliers and pixel elements introduce random variations in the correlator inputs. Offsets may result in a false match being up to twice as big as a true match (Mead, 1989). The resistive coupling between correlation nodes helps reduce the effect of circuit offsets.

The ability of the resistors to suppress unwanted signals is a function of their strength. Decreasing the resistance couples correlators within a disparity plane more strongly and helps to suppress false matches and offsets. The strength of the resistors determines the area over which retinal inputs are averaged. The disparity plane with the largest average input wins. If the correlators within a disparity plane are too strongly coupled, then they act as a unit. The retinal input to the entire disparity plane is averaged, and the plane with the largest total input current wins. To allow breaks in disparity, we must limit the resistance.

We can investigate the effects of changing the coupling between correlators by examining the area around the discontinuity, in an image with two disparities. Figure 13 shows the output of the correlators on both of the winning disparity planes. When the lateral resistance is large, very little current flows,

Figure 11 Finding two disparity planes. **(a)** The chip was tested using black and white bar patterns that could be tilted, as shown by the shaded bars. The degree of tilt determines the disparity of the pattern between the two retinas; when the bars are perpendicular to the retinas they are at zero disparity. We can construct a disparity discontinuity by combining two such bar patterns at different degrees of tilt. Such a stimulus is shown in dotted outline superimposed over the array. **(b)** Retinal response to this input pattern. The space constant of the resistive network is longer than the bar width, so the retinal output is not edge enhanced. **(c)** The product of the two retinal outputs, computed off-chip and displayed in the same format as the actual chip output. It is an approximation to what the array of correlators is receiving as input. High correlation values are dark. High correlations away from the correct disparity plane are false matches. **(d)** Analog voltage output from the chip, encoded by gray levels, shows segmentation into two disparity planes. The cooperative interactions among correlators suppresses false correlations.

and the winning correlator clearly dominates the loser. Since the resistor saturates, the voltage difference across the break can increase without drawing more current across the edge.

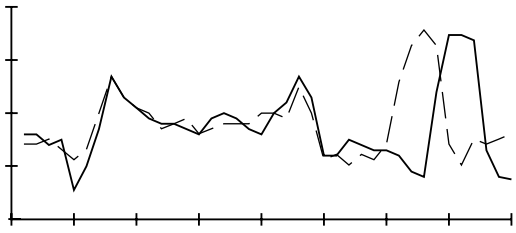


Figure 12 Filling in the solution. **(a)** Retinal response to this input pattern. The space constant of the resistive network is shorter than the width of the large bar. Due to the averaging properties of the resistive net, the response to the edge of the wide bar is smaller than the response to the narrow bar. The response is enhanced at the edge of the wide bar. In the central region of the wide bar, the retinal output is near zero. **(b)** Analog output from the chip. The solution breaks into two disparity regions. The central area of the bar is filled in.

As the resistance decreases relative to the multiplier bias, the disparity discontinuity becomes less sharp. The resistors draw current from the winning correlator at the break. Figure 13 illustrates the effect of decreases in the lateral resistance. When the ratio of the multiplier bias current to the resistor saturation current is 0.30 (small resistance case), the current flowing through the resistor is able to supply some current to all the correlators across the break. The solution propogates across the change in disparity, so the winning disparity plane is less distinguishable from the losing plane. When the ratio is larger, the losing plane is clearly distinguished from the winning plane.

Figure 13 Correlator outputs for an image with two regions of different disparity. The input image has one region at disparity +2 next to another region of disparity -1. The boundary between the regions occurs in the center of the image. The analog outputs of the correlators on the +2 and -1 disparity planes are shown for different multiplier bias currents. The off-chip current-sense amplifier saturates at 3.7 volts. The zero-point reference voltage is 2.4 volts. The ratio, r, of the multiplier bias current to the resistor bias current is shown next to the curve, along with the disparity plane, d.

The variant of the chip that uses motion signals as features for correlation has a larger input signal and smaller offsets than has the static contrast-sensitive chip. There is only one current mirror between the output voltage generated by the retina and the input current to the correlator. These mirrors are mismatched between correlators, but the mismatches are smaller than those generated by the cascaded mirrors in the Gilbert multiplier. Because of its stronger input signals, the time derivative chip is more able to fill in areas of the solution that are not receiving retinal input, while still finding multiple disparity regions. Using a 40-pixel input array, the time derivative chip is able reliably to find correct solutions for images with three disparity regions, whereas the static contrast chip can discriminate only two disparity regions (Figures 14, 15 and 16). The larger correlation input, however, means that the correlators are less able to suppress transient false matches (Figure 15).

For images with uniform intensity areas, the motion chip eliminates many false targets while maintaining a large input signal. The spatial extent of the time-derivative signal coming from the retinas is set by the speed of motion

and the time constant of the derivative elements. An element generates few false matches by correlating with its previous position (rather than its current position) in the other retina if the time constant is short. As in the case of the edge-enhancing static retina, the retinal signal is small in areas of uniform illumination. Spatially uniform intensity regions do not generate time-derivative signals when the image is displaced. The retinal input to the correlators generated at moving edges is very large compared to the retinal input in regions of unchanging intensity. The current supplied by a saturated resistor can be large compared to the static correlator input, but still small compared to the motion-generated signals. The solution can propagate effectively in regions of no image motion, whereas the signals from moving edges are easily able to break the solution into multiple disparity regions.

The integrative properties of the correlators may aid in filling in the solution. The dynamics of the correlators are set by the WTA circuits. The winning channel is discharged logarithmically in time by what is essentially a diode-connected transistor. In the absence of any competing input, the winning channel takes a long time to decay (Figure 16). Therefore, the solution is integrated spatially as the stimulus moves over the pixel array. The fact that the winner stays active for a long time does not generate any false matches. In contrast, many false matches would be generated if the retinal derivative circuit had a large time constant.

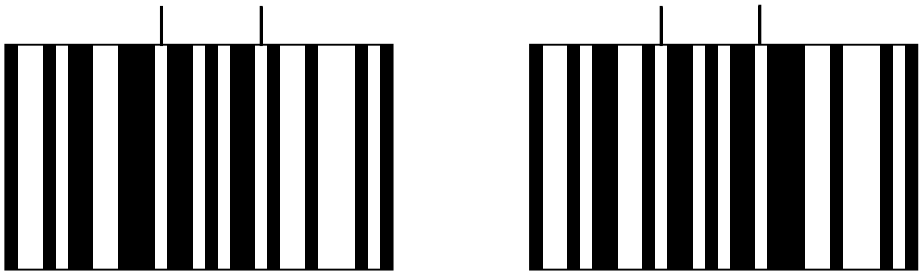


Figure 14 Image with three constant-disparity regions. Boundaries between regions are indicated by vertical markers. This binocular pair, when fused, shows a central region containing 4 white bars standing out over a surrounding background.

Figure 15 Scanned correlator-array output for motion sensitive chip, showing solution divided into three regions. Image is being moved slowly across the retinas. Bright areas signify a match. The picture was taken 0.25 seconds after the start of image motion. Transient false matches are not well suppressed.

Figure 16 **(a)** Scanned correlator array output showing response of the chip to the still input pattern. When there are no motion signals from the retina, the correlator array outputs are averaged over the whole disparity plane. Each disparity plane appears as a horizontal bar. **(b)** Scanned correlator array output, 0.65 seconds after stopping motion of the input pattern. The solution is still visible because the state of the winning correlators decays slowly.

DISCUSSION

Computational and synthetic approaches to artificial vision have a synergistic relationship. The stereo matching circuit profits from many of the strengths of Marr's computational algorithm. For example, resistive coupling between correlators enables the system to function in spite of large offsets. To average out offsets and to suppress false matches, we must set the resistors at a low value. Strong coupling results in a limited number of disparity regions that can be correctly identified in a single image. It is not yet understood what the intrinsic limitations are on the number of different disparity planes the chip can find. That the chip can find any solution in the presence of large offsets is a testament to the robust nature of the collective algorithm.

The chip architecture, which is an embodiment of the representational framework of the algorithm, uses a value-unit encoding (Ballard, 1986). Rather than encoding target depth by the magnitude of a voltage, activation of a correlator unit represents the presence of a target in a discrete region of three-dimensional space. Disparity is determined by the pattern of activity in the correlator array. The WTA inhibition is analog in nature; if two correlation elements are more or less equally stimulated by retinal input and cooperative interactions, then their outputs are comparable. If the disparity of the image is between two disparity planes, both are activated. When properly interpreted, the value-unit encoding achieves a resolution that is finer than a single pixel, even though the individual circuit elements are imprecise.

Implementation of algorithms in a physical medium stimulates their development in unforseen directions. For example, the use of optically acquired images precipitated the development of a variety of input features for stereo matching (Nishihara, 1984). These features are more robust to noise and vertical offsets between images than are the abstract binary tokens used as matching features in the original algorithm.

The static contrast-sensitive chip uses the analog value of a center-surround computation, which is easy to implement, as a primitive for stereo matching. The computation performed by the retinas on the chip retains information about the contrast between objects and is biologically plausible (Mahowald and Mead, 1989). The use of a center-surround matching primitive has been investigated by Mayhew and Frisby (1981). The use of time derivatives of intensity as a matching primitive has not been previously explored. Retinal neurons that generate transient responses, similar to those generated by the time-derivative silicon retinas, are known to project, via several relays, to disparity-tuned cells in layer IVb of the primary visual cortex (Poggio, 1984). The use of transient matching primitives introduces time as a representational dimension in the stereo correspondence computation.

Little research has been done in time-based algorithms for stereopsis because it is difficult to simulate temporal functions using traditional methods. It

is known that time is an intrinsic part of the disparity computation in natural systems. Perceptual psychologists have shown that binocular time delay and disparity can be substituted for each other in moving stimuli (Burr and Ross, 1979). Binocular time delay has been used to characterize disparity sensitive neurons in visual cortex (Gardner *et al.* 1985). Signals that are time delayed between the two eyes result from motion in a complex environment in which surfaces occlude one another. (Shimojo *et al.* 1985).

Parallel, analog hardware greatly facilitates real-time processing of complex inputs. The time-derivative correlating chip is a first step toward a real-time, interactive system for stereopsis. Using time derivatives has improved the performance of the chip by reducing the effect of offsets. In the future, we hope to expand the chip's representation to include binocular time delay. By taking advantage of the properties of the analog medium, we hope to gain further insights into the problem of stereopsis.

Acknowledgements

We thank Jin Luo and Tor Lande for technical assistance. We thank Hewlett-Packard for computing support, and DARPA and MOSIS for chip fabrication. This work was sponsored by the Office of Naval Research and the System Development Foundation.

References

Ballard, D.H. (1986). Cortical connections and parallel processing: Structure and function. *The Behavioral and Brain Sciences* **9:** 67–120.

Burr, D.C., and Ross, J. (1979). How does binocular delay give information about depth? *Vision Research* **19:** 523–532.

Delbrück, T., and Mead, C.A. (1989). An electronic photoreceptor sensitive to small changes in intensity. In Touretsky, D. S. (ed), *Advances in Neural Information Processing Systems* 1., pp 712–727, San Mateo, CA: Morgan Kaufman.

Gardner, J.C., Douglas, R.M., and Cyander, M.S. (1985). A time-based stereoscopic depth mechanism in the visual cortex. *Brain Research* **328:** 154–157.

Hutchinson, J., Koch, C., Luo, J., and Mead, C. (1988). Computing motion using analog and binary resistive networks. *IEEE Computer* March pp. 52–63.

Julesz, B. (1960). Binocular depth perception of computer-generated patterns. *Bell Syst. Tech. J.* **39:** 1125–1162.

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

Lazzaro, J., Ryckebusch S., Mahowald, M.A., and Mead, C.A. (1989). Winner-Take-All circuits of $O(n)$ complexity. In Touretsky, D.S. (ed), *Advances in Neural Information Processing Systems* 1. pp. 703–711, San Mateo, CA: Morgan Kaufman.

Mahowald, M.A. and Mead, C.A. (1989). Silicon retina. In Mead, C.A. *Analog VLSI and Neural Systems*, pp. 257–278, Reading, MA: Addison-Wesley.

Marr, D., Palm, G., and Poggio, T. (1978). Analysis of a cooperative stereo algorithm. *Biological Cybernetics* **28:** 223–239.

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

Marr, D. (1982). *Vision*, New York: W. H. Freeman.

Mayhew, J. and Frisby, J.P. (1981). Psychophysical and computational studies towards a theory of human stereopsis. *Artificial Intelligence* **17:** 349–385.

Mead, C.A. (1989). *Analog VLSI and Neural Systems.* Reading, MA: Addison-Wesley.

Mead, C.A. and Mahowald, M.A. (1988). A silicon model of early visual processing. *Neural Networks.* **1:** 91–97.

Nishihara, H. (1984). Practical real-time imaging stereo matcher. *Optical Engineering* **23:** 536–545.

Poggio, G. (1984). Processing of stereoscopic information in primate visual cortex. In Edelman, G. M., Gall W. E., and Cowan, W. M. (eds), *Dynamic aspects of neocortical function*, pp. 613–635, New York: John Wiley & Sons.

Poggio, G. and Poggio, T. (1984). The analysis of stereopsis. *Annual Review of Neuroscience* **7:** 379–412.

Shimojo, S., Silverman, G.H., and Nakayama, K. (1985). An occlusion-related mechanism of depth perception based on motion and interocular sequence. *Nature* **333:** 265–268.

Sivilotti, M.A., Mahowald, M.A., and Mead, C.A. (1987). Real-time visual computations using analog CMOS processing arrays. In Losleben, (ed), *Advanced Research in VLSI, Proceedings of the 1987 Stanford Conference,* pp. 295–312, Cambridge, MA : MIT Press.