Binocular Disparity and the Perception of Depth

Review

Ning Qian
Center for Neurobiology and Behavior
Columbia University
New York, New York 10032

We perceive the world in three-dimensions even though the input to our visual system, the images projected onto our two retinas, has only two spatial dimensions. How is this accomplished? It is well known that the visual system can infer the third dimension, depth, from a variety of visual cues in the retinal images. One such cue is binocular disparity, the positional difference between the two retinal projections of a given point in space (Figure 1). This positional difference results from the fact that the two eyes are laterally separated and therefore see the world from two slightly different vantage points.

The idea that retinal disparity contributes critically to depth perception derives from the invention of the stereoscope by Wheatstone in the 19th century, with which he showed conclusively that the brain uses horizontal disparity to estimate the relative depths of objects in the world with respect to the fixation point, a process known as stereoscopic depth perception or stereopsis. Because simple geometry provides relative depth given retinal disparity, the problem of understanding stereo vision reduces to the question: How does the brain measure disparity from the two retinal images in the first place?

Since Wheatstone's discovery, students of vision science have used psychophysical, physiological, and computational methods to unravel the brain's mechanisms of disparity computation. In 1960, Julez made the important contribution that stereo vision does not require monocular depth cues such as shading and perspective (see Julez, 1971). This was demonstrated through his invention of random dot stereograms. A typical stereogram consists of two images of randomly distributed dots that are identical except that a central square region of one image is shifted horizontally by a small distance with respect to the other image (see Figure 6a for an example). When each image is viewed individually, it appears as nothing more than a flat field of random dots. However, when the two images are viewed dichoptically (i.e., the left and right images are presented to the left and right eyes, respectively, at the same time), the shifted central square region "jumps" out vividly at a different depth. This finding demonstrates that the brain can compute binocular disparity without much help from other visual modalities.

The first direct evidence of disparity coding in the brain was obtained in the late 1960s, when Pettigrew and coworkers recorded disparity selective cells from the striate cortex in the cat, the primary visual area (see Bishop and Pettigrew, 1986). The result came as a surprise. Few people at the time expected to find disparity tuned cells so early in the brain's visual processing stream. A decade later, Poggio and collaborators reported similar findings in awake behaving monkeys (see Poggio and Poggio, 1984). They classified cells into a few discrete categories, although it now appears that these categories represent idealized cases from a continuous distribution (LeVay and Voigt, 1988). More recently, Ohzawa et al. (1990, 1996) provided detailed quantitative mapping of binocular receptive fields of the cat visual cortical cells and suggested models for simulating their responses. While these and many other experiments have demonstrated the neural substrates for disparity coding at the earliest stage of binocular convergence, they leave open the question of how a population of disparity selective cells could be used to compute disparity maps from a pair of retinal images such as the stereograms used by Julez. What is needed, in addition to experimental investigations, is a computational theory (Marr, 1982; Churchland and Sejnowski, 1992) specifying an algorithm for combining the neuronal signals into a meaningful computational scheme.

Although there have also been many computational studies of stereo vision in the past, until recently, most studies had treated disparity computation mainly as a mathematics or engineering problem while giving only secondary considerations to existing physiological data. Part of this tradition stems from a belief advanced paradoxically by David Marr, one of the most original thinkers in vision research, that physiological details are not important for understanding information processing tasks (such as visual perception) at the systems level. Marr (1982) argued that a real understanding will only come from an abstract computational analysis of how a particular problem may be solved under certain mathematical assumptions, regardless of the neuronal implementations in the brain. Although the importance of Marr's computational concept cannot be overstated, the main problem with ignoring physiology is that there is usually more than one way to "solve" a given perceptual task. Without paying close attention to physiology, one often comes up with algorithms that work in some sense but have little to do with the mechanisms used by the brain. In fact, most previous stereo vision algorithms contain nonphysiological procedures that could not be implemented with real neurons (see Qian, 1994, for a discussion).

To understand visual information processing performed by the brain instead of by an arbitrary machine, one obviously has to construct computational theories of vision based on real neurophysiological data. Such a realistic modeling approach to stereo vision has been proven possible recently. Disparity-tuned units, based on the response properties of real binocular cells, can be shown to effectively compute disparity maps from stereograms. Moreover, the stereo algorithm can be extended to include motion detection and provide coherent explanations for some interesting depth illusions and physiological observations. In the discussion that follows, I attempt to recreate the line of thought that led to these models and use them as examples to discuss the general issue of how models such as these can be useful in interpreting relevant experimental data.
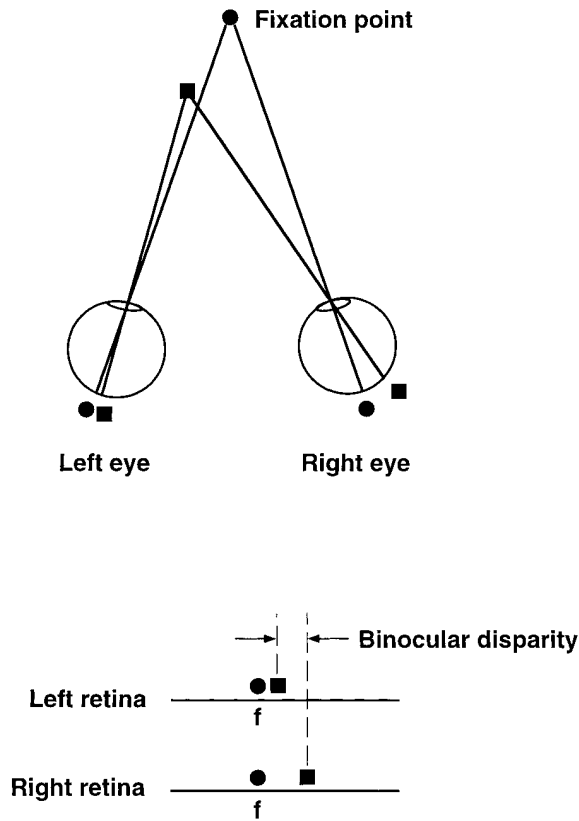
Figure 1. The Geometry of Binocular Projection (Top) and the Definition of Binocular Disparity (Bottom)

The fixation point projects to the two corresponding foveas (fs) and has zero disparity by definition. It can be shown that all zero disparity points in the plane fall on the circle passing through the fixation point and the two eyes (or more accurately, the nodal points of the two lens systems). All other points do not project to corresponding locations on the two retinas and have non-zero disparities. The magnitude of disparity is usually expressed in terms of visual angle.

Physiological Computation of Binocular Disparity

As mentioned above, disparity-sensitive cells have been found in the very first stage of binocular convergence, the primary visual cortex. In their classical studies, Hubel and Wiesel (1962) identified two major classes of cells in this area: simple and complex. Simple cells have separate on (excitatory) and off (inhibitory) subregions within their receptive fields that respond to light and dark stimuli, respectively. In contrast, complex cells respond to both types of stimuli throughout their receptive fields. Hubel and Wiesel suggested a hierarchy of anatomical organization according to which complex cells receive inputs from simple cells, which in turn receive inputs from LGN cells. Although the strict validity of this hierarchy is debatable, as some complex cells appear to receive direct LGN inputs, it is generally agreed that the majority of simple and complex cells in the primary visual cortex are binocular (i.e., they have receptive fields on both retinas) and disparity tuned (i.e., they respond differently to different stimulus disparities; see Figure 2). What, then, are the roles of simple and complex cells in disparity computation?
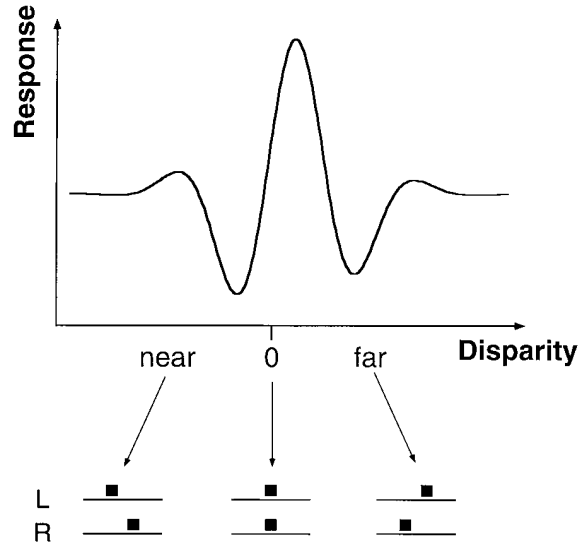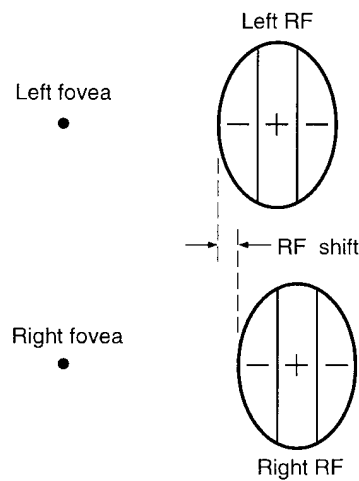

Figure 2. A Schematic Drawing of a Disparity Tuning Curve for a Binocular Cell in the Visual Cortex

Three representative retinal stimulus configurations are shown at the bottom. They correspond to a point in space that is nearer than, at, or further away from the fixation point.

Early physiological experiments (Bishop and Pettigrew, 1986) suggested that to achieve disparity tuning, a binocular simple cell has an overall shift between its left and right receptive fields as illustrated in Figure 3a. Others (Ohzawa et al., 1990, 1996) have suggested that the shift is between the excitatory and inhibitory subregions within the aligned receptive field envelopes (Figure 3b). For simplicity, both of these alternatives will be referred to as "receptive field shift" when it is not essential to distinguish them. These receptive field structures of binocular simple cells could arise from orderly projections of LGN cells with concentric receptive fields, as originally proposed by Hubel and Weisel (Figure 4).

Since disparity is nothing but a shift between the two retinal projections (Figure 1), one might expect intuitively that such a simple cell should give the best response when the stimulus disparity matches the cell's left–right receptive field shift. In other words, a simple cell might prefer a disparity equal to its receptive field shift. A population of such cells with different shifts would then prefer different disparities, and the unknown disparity of any stimulus could be computed by identifying which cell gives the strongest response to the stimulus. The reason that no stereo algorithm has come out of these considerations is that the very first assumption–that a binocular simple cell has a preferred disparity equal to its receptive field shift–is not valid. Simple cells cannot have a well-defined preferred disparity because their responses depend not only on the disparity but on the detailed spatial structure of the stimulus (Ohzawa et al., 1990; Qian 1994; Zhu and Qian, 1996; Qian and Zhu, 1997). Although one can measure a disparity tuning curve from a simple cell, the peak location of the curve (i.e., the preferred disparity) changes with some simple manipulations (such as a lateral displacement) of the stimuli. This property is formally known as Fourier phase dependence because the spatial structure of an image

(a) Position-shift Model

(b) Phase-difference Model

LGN RFs on
the left retina

Binocular
simple cell

LGN RFs on
the right retina

Figure 4. A Hubel–Wiesel Type of Model Shows How the Receptive Field Structure of the Binocular Simple Cell in Figure 3a Could Arise from the Orderly LGN Projections to the Cell

On-center and off-center LGN cells are assumed to project to the on and off subregions of the simple cell, respectively (only on-center LGN cells are shown for clarity). A similar scheme may be proposed for the cell in Figure 3b. Recent studies support Hubel and Wiesel's idea but indicate the importance of intracortical interactions in shaping cortical receptive field properties (see, for example, Somers et al., 1995).
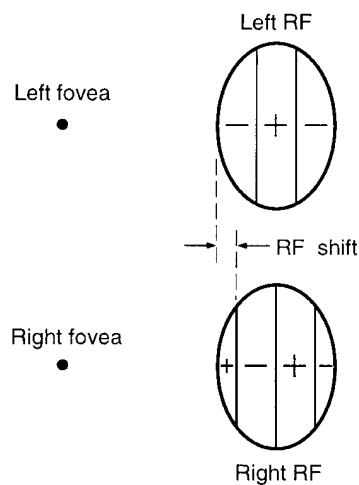
Figure 3. Schematic Drawings Illustrate the Left–Right Receptive Field (RF) Shift of Binocular Simple Cells

The + and − signs represent the on and off subregions within the receptive fields, respectively. Two different models for achieving the shift have been suggested by physiological experiments.
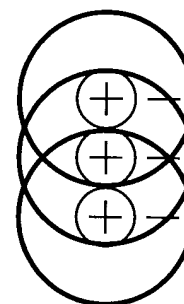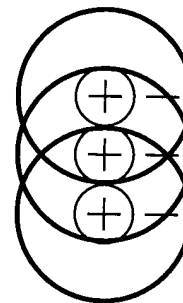(a) Position-shift model. According to this model, the left and right receptive fields of a simple cell have identical shapes but have an overall horizontal shift between them (Bishop and Pettigrew, 1986).
(b) Phase-difference model. This model assumes that the shift is between the on–off subregions within the left and right receptive field envelopes that spatially align (Ohzawa et al., 1990, 1996). The fovea locations on the left and right retinas are drawn as a reference point for vertically aligning the left and right receptive fields of the simple cell.

is reflected in the phase of its Fourier transform. The Fourier phase dependence of simple cell responses is obviously not desirable from the point of view of extracting a pure disparity signal from which to compute disparity maps.

The phase dependence of simple cell responses can

be easily understood by considering the disparity tuning of a simple cell to a vertical line. The Fourier phase of the line is directly related to the lateral position of the line, which will affect where its projection falls on the left and right receptive fields of the simple cell. The line with a given disparity may evoke a strong response at one line position because it happens to project onto the excitatory subregions of both the left and right receptive fields but may evoke a much weaker response at a different position because it now stimulates some inhibitory portions(s) of the receptive fields. Therefore, the response of the simple cell to a fixed disparity changes with the changing stimulus Fourier phases, and, consequently, it cannot have a well-defined preferred disparity. There is direct experimental evidence supporting this conclusion. For example, Ohzawa et al. (1990) found that disparity tuning curves of simple cells measured with bright bars and dark bars (which have different Fourier phases) are very different. The Fourier phase dependence of simple cell responses can also explain an observation by Poggio et al. (1985), who reported that simple cells show no disparity tuning to dynamic

random dot stereograms. Each of the stereograms in their experiment maintained a constant disparity over time but varied its Fourier phase from frame to frame by constantly rearranging the dots. Simple cells lost their disparity tuning as a result of averaging over many different (phase-dependent) tuning curves (Qian, 1994).

Although simple cells are not suited for disparity computation, complex cell responses have the desired phase independence as expected from their lack of separate excitatory and inhibitory subregions within their receptive fields (Skottun et al., 1991). To build a working stereo algorithm, however, one needs to specify how this phase independence is achieved and how an unknown stimulus disparity can be recovered from these responses. Most physiology experiments approach stereo vision from the opposite perspective and measure the responses of visual cells to a set of stimuli with known disparities in order to obtain the cells' disparity tuning curves. These curves alone are not very useful from a computational point of view because a response can be read from a disparity tuning curve only when the stimulus disparity is already known. We need a quantitative procedure for computing an unknown disparity in a pair of retinal images from the responses of complex cells to the images.

Fortunately, a method for determining the responses of binocular complex cells has recently been proposed by Ohzawa et al. (1990) based on their quantitative physiological studies (see also Ferster, 1981). These investigators found that a binocular complex cell in the cat primary visual cortex can be simulated by summing up the squared responses of a quadrature pair of simple cells, and the simple cell responses, in turn, can be simulated by adding the visual inputs on their left and right receptive fields (see Figure 5). (Two binocular simple cells are said to form a quadrature pair if there is a quarter-cycle shift between the excitatory–inhibitory subregions for both their left and right receptive fields; Ohzawa et al., 1990.) The remaining questions are whether the model complex cells constructed this way are indeed independent of stimulus Fourier phases and if so, how their preferred disparities are related to their receptive field parameters.

These issues have recently been investigated through mathematical analyses and computer simulations (Qian, 1994; Zhu and Qian, 1996; Qian and Zhu, 1997). The complex cell model was found to be independent of stimulus Fourier phases for some types of stimuli, including the bars used in the physiological experiments of Ohzawa et al. (1990), and its preferred disparity is equal to the left–right receptive field shift within the constituent simple cells. For more complicated stimuli, such as random dot stereograms, however, the complex cell constructed from a single quadrature pair of simple cells is still phase sensitive, although less so than simple cells. This problem can be easily solved by considering the additional physiological fact that complex cells have somewhat larger receptive fields than those of simple cells (Hubel and Wiesel, 1962). This fact is incorporated into the model by averaging over several quadrature pairs of simple cells with nearby and overlapping receptive fields to construct a model complex cell (Zhu and Qian, 1996; Qian and Zhu, 1997). The resulting complex cell is largely phase independent for any stimulus,
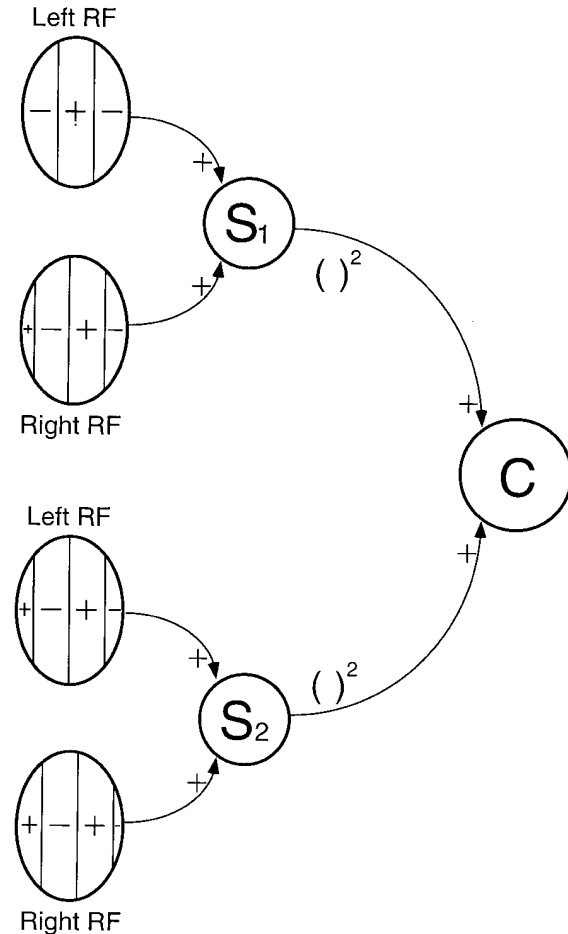

Figure 5. The Model Proposed by Ohzawa et al. (1990) for Simulating Real Binocular Complex Cell Responses

The complex cell (labeled C) sums squared outputs of a quadrature pair of simple cells (labeled S_1 and S_2). Each simple cell in turn sums the contributions from its two receptive fields on the left and right retinas. The squaring operation could presumably be performed by a multiplication-like operation on the dendritic tree of the complex cell (Mel, 1993). Ohzawa et al. (1990) pointed out that to consider the fact that cells cannot fire negatively, each simple cell should be split into two with inverted receptive field profiles, so that they can carry the positive and negative parts of the firing, respectively. The resulting complex cell model will contain four half-wave rectified simple cells. It is, however, exactly equivalent to the model shown here with two simple cells.

and its preferred disparity still equals the receptive field shift within the constituent simple cells.

With the above method for constructing model complex cells with well-defined preferred disparities, we are finally ready to develop a stereo algorithm for computing disparity maps from stereograms. By using a population of complex cells with preferred disparities covering the range of interest, the disparity of any input stimulus can be determined by identifying the cell in the population with the strongest response (or by calculating the population averaged preferred disparity of all cells weighted by their responses). An example of applying this algorithm to a random dot stereogram is shown in Figure 6. The result demonstrates that a population of complex
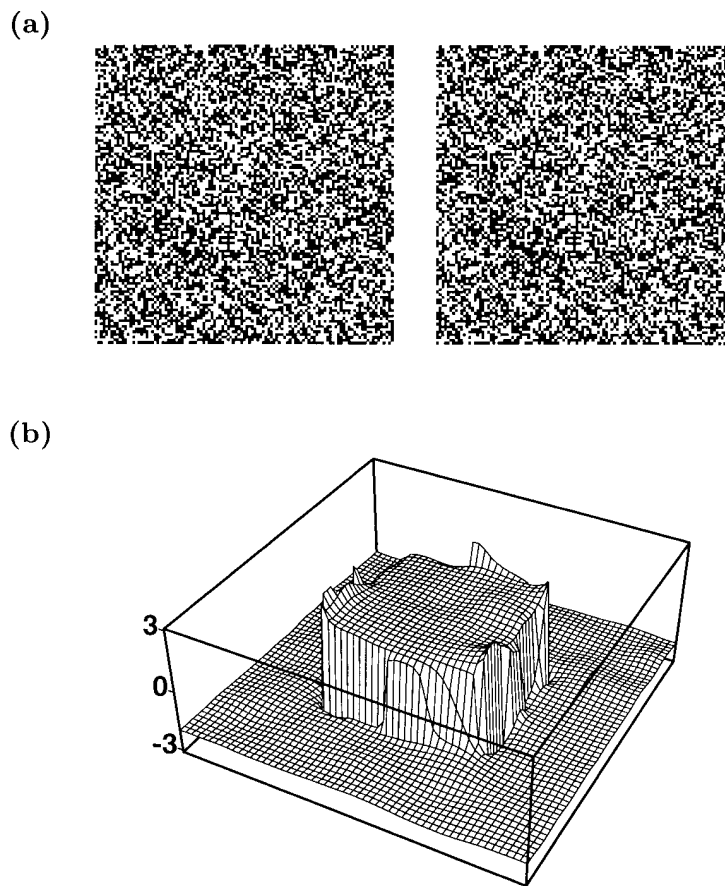
(a)

(b)

Figure 6. Disparity Computation with Binocular Cells

(a) A 110 × 110 random dot stereogram with a dot density of 50% and dot size of 1 pixel. The central 50 × 50 area and the surround have disparities of 2 and −2 pixels, respectively. When fused with uncrossed eyes, the central square appears further away than the surround.

(b) The disparity map of the stereogram computed with eight complex cells at each location using the method outlined in the text. The distance between two adjacent sampling lines represents a distance of two pixels in (a). Negative and positive values indicate near and far disparities, respectively. The eight complex cells have the same preferred spatial frequency of 0.125 cycles/pixel and receptive field Gaussian width of 4 pixels. Similar disparity maps can be obtained by using complex cells at different spatial scales or by averaging results across several scales. The results obtained with either of the two receptive field models shown in Figure 3 are also very similar (Qian and Zhu, 1997).

cells can effectively compute the disparity map of the stereogram via a distributed representation.

There is, as yet, no direct anatomical evidence supporting the quadrature pair method for constructing binocular complex cells from simple cells. However, based on the quantitative physiological work of Ohzawa et al. (1990), the method is at least valid as a phenomenological description for a subset of real complex cell responses. In addition, the analyses indicate that the same phase-independent complex cell responses can be obtained by appropriately combining the outputs of many simple or LGN cells without requiring the specific quadrature relationship (Qian, 1994; Qian and Andersen, 1997).

The Correspondence Problem

To measure binocular disparity, the visual system must solve the correspondence problem: it must determine which parts on the two retinal images come from the same object in the world. How does the above stereo algorithm address the correspondence problem? Historically, it has been suggested that the visual system solves the problem by matching up image features between the two retinas. In the case of random dot stereograms, the correspondence problem is often stated as identifying which dot in the left image matches which dot in the right image. Since all dots in the two images are of identical shape, it is often argued that any two dots could match, and the visual system is faced with

an enormously difficult problem of sorting out the true matches from the huge number of false ones. This argument is the starting point for a whole class of stereo algorithms (Marr and Poggio, 1976; Prazdny, 1985; Qian and Sejnowski 1989; Marshall et al., 1996). It is not physiological, however, because the left and right receptive fields of a typical binocular cell can be much larger than a dot in a stereogram. Even the cells in monkey striate cortex that represents the fovea, the area of greatest visual acuity and smallest receptive fields, have a receptive field size of about 0.1 degree (Dow et al., 1981). This dimension is more than twice as large as a dot in the stereogram in Figure 6 when viewed at a distance of >35 cm. A closely related fact is that most cells are broadly tuned to disparity; even the most sharply tuned cells have tuning widths of about 0.1–0.2 degree (Poggio and Poggio, 1984; Lehky and Sejnowski, 1990). It is therefore difficult to imagine how a cell could match a specific pair of dots while ignoring many others in its receptive fields. It appears more reasonable to assume that a binocular cell tries to match the two image patches covered by its receptive fields–each may contain two or more dots–instead of operating on fine image features such as individual dots. Since each image patch is likely to contain a unique dot distribution, it can be best matched by only one (corresponding) patch in the other image. Therefore, for an algorithm that avoids operating at the level of individual dots, the false-match problem is practically nonexistent (Sanger, 1988). The stereo

model of the previous section demonstrates that binocular complex cells as described by Ohzawa et al. (1990) have the right physiological property for matching the image patches in its receptive fields. A careful mathematical analysis of the model complex cells reveals that their computation is formally equivalent to summing two related cross-products of the band-pass-filtered left and right image patches (Qian and Zhu, 1997). This operation is related to cross-correlation, but it overcomes some major problems with the standard cross-correlator.

Disparity Attraction and Repulsion

A good model should explain more than it is originally designed for. To evaluate the stereo model above, it has been applied to other problems of stereopsis. For example, the model has been used to explain the observation that we can still perceive depth when the contrasts of the two images in a stereogram are very different, so long as they have the same sign (Qian, 1994). Recently, the model has been applied to a psychophysically observed depth illusion. In 1986, Westheimer first described that when a few isolated features are viewed on the fovea, the perceived depth of a given feature depends not only on its own disparity but on the disparity of neighboring features. Specifically, two vertical line segments at different disparities, separated laterally along the horizontal frontoparallel direction, influence each other's perceived depth in the following way: when the lateral distance between the two lines is small (<~5 min), the two lines appear closer in depth as if they are attracting each other. At larger distances, this effect reverses, and the two lines appear further away from each other in depth (repulsion). When the distance is very large, there is no interaction between the lines. To model these effects, the responses of a population of complex cells centered on one line were examined as a function of how they are influenced by the presence of the other line at various lateral distances (Qian and Zhu, 1997, ARVO, abstract). The interaction between the lines in the model originates from the lines' simultaneous presence in the cells' receptive fields, and this can naturally explain Westheimer's observation without introducing any ad hoc assumptions (Figure 7). Thus, the psychophysically observed disparity attraction–repulsion phenomenon may be viewed as a direct consequence of the known physiological properties of binocular cells in the visual cortex.

Binocular Receptive Field Models and Characteristic Disparity

The stereo model also helped interpret a recent physiological observation by Wagner and Frost (1993). Recording from the visual Wulst of the barn owl, Wagner and Frost found that for some cells, the peak locations of a cell's disparity tuning curves to spatial noise patterns and sinusoidal gratings of various frequencies approximately coincide at a certain disparity. They called this disparity the characteristic disparity (CD) of the cell.

Wagner and Frost (1993) attempted to use their data to distinguish two well-known binocular receptive field models in the literature (see Figure 3). As mentioned above, early physiological studies suggested that there
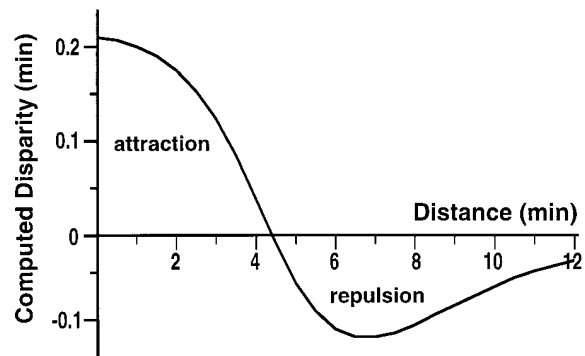

Figure 7. The Computed Disparity of One Line (Test Line) Plotted as a Function of Its Separation from the Second Line (Inducing Line)

The actual disparities of the test and the inducing lines are fixed at 0 and 0.5 min, respectively. Positive computed disparities indicate attractive interaction while negative values indicate repulsion between the two lines. The result was obtained by averaging across cell families with different preferred spatial frequencies and frequency tuning bandwidths using physiologically determined distributions for these parameters (DeValois et al., 1982).

is an overall positional shift between the left and right receptive fields in binocular simple cells (Bishop and Pettigrew, 1986). The shapes of the two receptive field profiles of a given cell have been assumed to be identical. In contrast, more recent quantitative studies by Ohzawa et al. (1990) have found that the left and right receptive field profiles of a simple cell often possess different shapes. They accounted for this finding by assuming that the receptive field shift is between the on–off subregions within the identical receptive field envelops that spatially align. This receptive field model is referred to as the phase-difference or phase-parameter model to distinguish it from the position-shift model that preceded it historically. Wagner and Frost (1993) concluded that their CD data favor the position-shift type of receptive field model but not the phase-shift type. In other words, they argued that only cells with identical left and right receptive field shapes can have the experimentally observed CDs. Unfortunately, their conclusion is not based on a careful mathematical analysis, and it turns out to be unfounded.

To examine this issue more accurately, the simple and complex cell responses to the stimuli used in Wagner and Frost's experiments were analyzed and simulated. It was found that the existence of approximate CDs in real cells cannot be used to distinguish between the two types of receptive field models described above (Zhu and Qian, 1996). Specifically, simple cells constructed from either type of receptive field model cannot have a CD because, as we have seen earlier, they lack well-defined disparity tuning curves due to their dependence on stimulus Fourier phases. Models of complex cells constructed from the position-shift type of receptive field model have a precise CD, while those from the phase-difference type of receptive field have an approximate CD with a systematic deviation similar to that found in some real cells. From this analysis follows the testable prediction that cells with CDs must all be complex cells. The analysis also provides methods for correctly distinguishing the two receptive field models experimentally. For example, one

method is to examine whether the peaks of disparity tuning curves to grating stimuli align precisely at the CD location, or whether the alignment is only approximate and deviates systematically with the spatial frequencies of the gratings. It is also possible that real visual cortical cells use a hybrid of phase difference and positional shift to code binocular disparity (Ohzawa et al., 1996; Zhu and Qian, 1996; Fleet et al., 1996). Such a hybrid model may be necessary for explaining the observed correlation between the perceived disparity range and the dominant spatial frequency in the stimulus (Schor and Wood, 1983; Smallman and MacLeod, 1994). Under reasonable assumptions, the stereo algorithm described earlier works equally well with either the phase-difference or position-shift type of model (or their hybrid) for the front-end simple cell receptive fields.

Motion-Stereo Integration

There is increasing psychophysical and physiological evidence indicating that motion detection and stereoscopic depth perception are processed together in the brain. Numerous psychophysical experiments have demonstrated that stereo cues strongly influence motion perception and vice versa (Regan and Beverley, 1973; Adelson and Movshon, 1982; Nawrot and Blake, 1989; Qian et al., 1994a). Physiologically, many visual cortical cells, especially those along the dorsal visual pathway, are tuned to both motion and disparity (Maunsell and Van Essen, 1983; Roy and Wurtz, 1990; Bradley et al., 1995; Ohzawa et al., 1996). It is therefore desirable to construct an integrated model that encompasses these two visual modalities. Although it has long been recognized that at an abstract level, both motion and stereo vision can be formulated as solving a correspondence problem (the motion correspondence problem is similar to the stereo one stated earlier; it is about determining which image regions on successive time frames come from the same object in the world), this characterization does not suggest how the two visual functions are processed together by a population of cells tuned to both motion and disparity, or how the two modalities interact. It has been demonstrated recently that under physiologically realistic assumptions about the spatiotemporal properties of binocular cells, the stereo model described earlier can be naturally combined with motion energy models (a class of physiologically plausible models for motion detection; see, for example, Adelson and Bergen, 1985) to achieve motion–stereo integration (Qian and Andersen, 1997). The complex cells in the model are tuned to both motion and disparity just like physiologically recorded cells, and a population of such cells could simultaneously compute stimulus motion and disparity.

Interestingly, complex cells in the integrated model are much more sensitive to motion along constant disparity planes than motion in depth toward or away from the observer (Qian, 1994). This property is consistent with the physiological finding that few cells in the visual cortex are tuned to motion in depth (Maunsell and Van Essen, 1983; Ohzawa et al., 1996) and with the psychophysical observation that human subjects poorly detect

motion in depth based on disparity cues alone (Westheimer, 1990). Because of this property, motion information can help reduce the number of possible stereoscopic matches in an ambiguous stereogram by making certain matches more perceptually prominent than others, a phenomenon that has been demonstrated psychophysically (Qian et al., 1993, Soc. Neurosci. abstract). The integrated model has also been used to explain the additional psychophysical observation that adding binocular disparity cues into a stimulus can help improve the perception of multiple and overlapping motion fields in the stimulus (a problem known as motion transparency) (Qian et al., 1994b).

Pulfrich Depth Illusions

Another interesting application of the integrated motion–stereo model is a unified explanation for a family of depth illusions associated with the name Carl Pulfrich. The classical Pulfrich effect refers to the observation that a pendulum oscillating back and forth in a frontoparallel plane appears to rotate along an elliptical path in depth when a neutral density filter is placed in front of one eye (see Figure 8a) (Morgan and Thompson, 1975). The direction of rotation is such that the pendulum appears to move away from the covered eye and toward the uncovered eye. By reducing the amount of light reaching the covered retina, the filter introduces a temporal delay in transmitting visual information from that retina to the cortex (Mansfield and Daugman, 1978; Carney et al., 1989). The traditional explanation of this illusion is that since the pendulum is moving, the temporal delay for the covered eye corresponds to a spatial displacement of the pendulum, which produces a disparity between the two eyes and therefore a shift in depth. This interpretation is problematic, however, because the Pulfrich depth effect is present even with dynamic noise patterns (Tyler, 1974; Falk, 1980), which lack the coherent motion required to convert a temporal delay into a spatial disparity. Furthermore, the effect is still present when a stroboscopic stimulus is used to prevent the two eyes from simultaneously viewing a target undergoing apparent motion (Burr and Ross, 1979). Under this condition, the traditional explanation of the Pulfrich effect fails because no conventionally defined spatial disparity exists. It has been suggested that more than one mechanism may be responsible for these phenomena (Burr and Ross, 1979; Poggio and Poggio, 1984).

Recent mathematical analyses and computer simulations indicate that all of these observations can be explained in a unified way by the integrated motion–stereo model (Qian and Andersen, 1997). Central to the explanation is the mathematical demonstration that a model complex cell with physiologically observed spatiotemporal properties cannot distinguish an interocular time delay (Δt) from an equivalent binocular disparity given by:

$$d \approx \frac{\omega_t^\circ}{\omega_x^\circ} \Delta t \qquad (1)$$

where ω_t° and ω_x° are the preferred temporal and spatial frequencies of the cell. This relation holds for any arbitrary spatiotemporal pattern (including pendulum, dynamic noise, and stroboscopic stimuli) that can significantly activate the cell. By considering the population
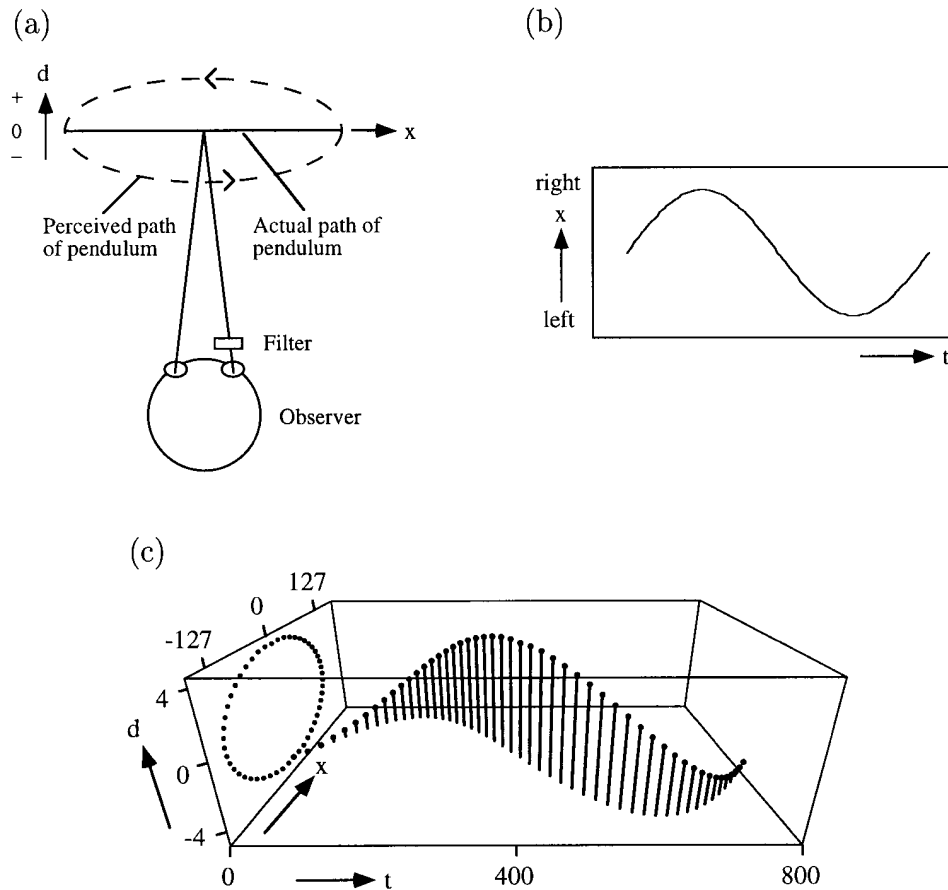
Figure 8. Pulfrich's Pendulum

(a) A schematic drawing of the classical Pulfrich effect (top view). A pendulum is oscillating in the frontoparallel plane indicated by the solid line. When a neutral density filter is placed in front of the right eye, the pendulum appears to move in an elliptical path in depth as indicated by the dashed line.

(b) The horizontal position of the pendulum as a function of time for one full cycle of oscillation.

(c) The computed equivalent disparity as a function of horizontal position and time. The data points from the simulation are shown as small closed circles. Lines are drawn from the data points to the x–t plane in order to indicate the spatiotemporal location of each data point. A time delay of 4 pixels is assumed for the right receptive fields of all the model cells. The pendulum has negative equivalent disparity (and therefore is seen as closer to the observer) when it is moving to the right and has positive equivalent disparity (further away from the observer) when it is moving to the left. The projection of the 3-D plot onto the d–x plane forms a closed path similar to the ellipse in (a).

responses of a family of cells with a wide range of disparity and motion parameters, all major observations regarding Pulfrich's pendulum and its generalizations to dynamic noise patterns and stroboscopic stimuli can be explained (Qian and Andersen, 1997). An example of simulation on Pulfrich's pendulum is shown in Figure 8c.

Two testable predictions can be made based on the analysis. First, the response of a binocular complex cell to an interocular time delay should be approximately matched by a binocular disparity according to equation 1. To test this prediction, a cell's tuning curves to binocular disparity and to interocular time delay should be established, then the preferred spatial frequency (ω_x°) and temporal frequency (ω_t°) should be determined for the same cell. If the model explanation is valid, the two tuning curves will be related to each other by the scaling factor $\omega_t^\circ/\omega_x^\circ$ along the horizontal axis. The second prediction is also based on equation 1. The equation predicts

that cells with different preferred spatial to temporal frequency ratios will individually "report" different apparent Pulfrich depths for a given temporal delay. If we assume that the perceived depth corresponds to the disparities reported by the most responsive cells in a population (or by the population average of all cells weighted by their responses), then the perceived Pulfrich depth should vary according to equation 1, as we selectively excite different populations of cells by using stimuli with different spatial and temporal frequency contents.

Conclusions

The brain is complex with many levels of organization. It consists of a multitude of systems that can perform sophisticated information-processing tasks such as visual perception, motor control, and learning and memory. A complete understanding of any such system requires not only experimental investigations but a

computational theory that specifies how the signals carried by a large number of neurons in the brain can be combined to accomplish a given task. Without quantitative modeling, our intuitions may often be incomplete or even wrong and have only limited power in relating and comprehending a large amount of experimental data. Equally important, computational models of neural systems must be based on real physiological data; otherwise, their relevance to understanding brain function will be very limited.

In this review, I have used stereo vision as an example to illustrate that given an appropriate set of experimental data, a physiologically realistic modeling approach to neural systems is feasible and fruitful. The experimental and theoretical studies discussed here suggest that although disparity sensitivity in the visual cortex originates from the left–right receptive field shifts of simple cells, it is at the level of complex cells that stimulus disparity is reliably coded in a distributed fashion. The models help increase our understanding of visual perception by providing unified accounts for some seemingly different physiological and perceptual observations and suggesting new experiments for further tests of the models. Indeed, without modeling, it would be rather difficult to infer that random dot stereograms could be solved by a family of binocular complex cells, that the psychophysically observed disparity attraction–repulsion phenomenon could be a direct consequence of the underlying binocular receptive field structure, or that different variations of the Pulfrich depth illusion could all be uniformly explained by the spatiotemporal response properties of real binocular cells. Physiology-based computational modeling has the potential of synthesizing a large body of existing experimental data into a coherent framework.

It is also important to note that how much physiological details should be included in a model depends on what questions one seeks to address (Churchland and Sejnowski, 1992). For example, it is adequate to start with the physiologically measured receptive fields when our goal is to investigate how a population of the cells may solve a certain perceptual problem. Useful models can be constructed before knowing exactly how these receptive fields come about. On the other hand, if one is interested in the origin of the receptive fields, the morphology of the cells' dendritic trees and the distribution of the synaptic inputs onto the trees will then have to be considered. If one would further like to understand the mechanisms that generate the synaptic currents, the relevant membrane channel molecules and intracellular second message systems will then have to be included in the model. One can imagine a chain of successive steps of reduction that points all the way to fundamental physical processes.

Despite the progress made by the recent experimental and theoretical studies reviewed here, many open questions in stereo vision remain. A major limitation of the models discussed here is that they only consider short-range interactions within the scope of the classical receptive fields of primary visual cortical cells. Long-range connections between these cells and influences outside the classical receptive fields have been documented physiologically (Gilbert, 1992). In addition, many cells in the extrastriate visual areas, where the receptive fields are much larger, are also disparity selective. How to incorporate these experimental findings into the models to account for perceptual phenomena involving long-range interactions (Spillman and Werner, 1996) requires further investigation. Another limitation is that the stereo model described here cannot explain our ability to detect the so-called second-order disparities. These disparities are not defined by the correlation between luminance profiles in the two retinal images but by higher-order image properties such as subjective contours (Ramachandran et al., 1973; Sato and Nishida, 1993; Hess and Wilcox, 1994). A second parallel pathway with additional nonlinearities must be added to the model for the detection of the second-order disparities (Wilson et al., 1992). Similarly, the stereo model cannot explain the the perceived depth in stereograms with unmatched monocular elements caused by occlusion (Nakayama and Shimojo, 1990; Liu et al., 1994). Finally, it is known that we can perceive multiple and overlapping depth planes defined by disparities (Prazdny, 1985). To solve this so-called stereo transparency problem using physiologically realistic mechanisms with relatively large receptive fields remains a major challenge. A close interplay between experimental and computational approaches may hold the best promise for resolving these and other outstanding issues of stereo vision in the future.

References

Adelson, E.H., and Movshon, J.A. (1982). Phenomenal coherence of moving visual patterns. Nature *300*, 523–525.

Adelson, E.H., and Bergen, J.R. (1985). Spatiotemporal energy models for the perception of motion. J. Opt. Soc. Am. [A] *2*, 284–299.

Bishop, P.O., and Pettigrew, J.D. (1986). Neural mechanisms of binocular vision. Vision Res. *26*, 1587–1600.

Bradley, D.C., Qian, N., and Andersen, R.A. (1995). Integration of motion and stereopsis in cortical area MT of the macaque. Nature *373*, 609–611.

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

Carney, T., Paradiso, M.A., and Freeman, R.D. (1989). A physiological correlate of the Pulfrich effect in cortical neurons of the cat. Vision Res. *29*, 155–165.

Churchland, P.S., and Sejnowski, T.J. (1992). The Computational Brain. (Cambridge, Massachusetts: MIT Press).

DeValois, R.L., Albrecht, D.G., and Thorell, L.G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. Vision Res. *22*, 545–559.

Dow, B.M., Snyder, A.Z., Vautin, R.G., and Bauer, R. (1981). Magnification factor and receptive field size in foveal striate cortex of the monkey. Exp. Brain Res. *44*, 213–228.

Falk, D.S. (1980). Dynamic visual noise and the stereophenomenon: interocular time delays, depth and coherent velocities. Percept. Psychophys. *28*, 19–27.

Ferster, D. (1981). A comparison of binocular depth mechanisms in areas 17 and 18 of the cat visual cortex. J. Physiol. *311*, 623–655.

Fleet, D.J., Wagner, H., and Heeger, D.J. (1996). Encoding of binocular disparity: energy models, position shifts and phase shifts. Vision Res. *36*, 1839–1858.

Gilbert, C.D. (1992). Horizontal integration and cortical dynamics. Neuron *9*, 1–13.

Hess, R.F., and Wilcox, L.M. (1994). Linear and non-linear filtering in stereopsis. Vision Res. *34*, 2431–2438.

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

Julez, B. (1971). Foundations of Cyclopean Perception. (University of Chicago Press).

Lehky, S.R., and Sejnowski, T.J. (1990). Neural model of stereoacuity and depth interpolation based on a distributed representation of stereo disparity. J. Neurosci. *10*, 2281–2299.

LeVay, S., and Voigt, T. (1988). Ocular dominance and disparity coding in cat visual cortex. Vis. Neurosci. *1*, 395–414.

Liu, L., Stevenson, S.B., and Schor, C.W. (1994). Quantitative stereoscopic depth without binocular correspondence. Nature *367*, 66–68.

Mansfield, R.J.W., and Daugman, J.D. (1978). Retinal mechanisms of visual latency. Vision Res. *18*, 1247–1260.

Marr, D. (1982). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. (San Francisco: W.H. Freeman and Company).

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

Marshall, J.A., Kalarickal, G.J., and Graves, E.B. (1996). Neural model of visual stereomatching: slant, transparency, and clouds. Network: Comp. Neural Sys., in press.

Maunsell, J.H.R., and Van Essen, D.C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey II. Binocular interactions and sensitivity to binocular disparity, J. Neurophysiol. *49*, 1148–1167.

Mel, B.W. (1993). Synaptic integration in an excitable dendritic tree. J. Neurophysiol. *70*, 1086–1101.

Morgan, M.J., and Thompson, P. (1975). Apparent motion and the Pulfrich effect. Perception *4*, 3–18.

Nakayama, K., and Shimojo, S. (1990). da Vinci stereopsis: depth and subjective occluding contours from unpaired image points. Vision Res. *30*, 1811–1825.

Nawrot, M., and Blake, R. (1989). Neural integration of information specifying structure from stereopsis and motion. Science *244*, 716–718.

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

Ohzawa, I., DeAngelis, G.C., and Freeman, R.D. (1996). Encoding of binocular disparity by simple cells in the cat's visual cortex. J. Neurophysiol. *75*, 1779–1805.

Poggio, G.F., and Poggio, T. (1984). The analysis of stereopsis. Annu. Rev. Neurosci. *7*, 379–412.

Poggio, G.F., Motter, B.C., Squatrito, S., and Trotter, Y. (1985). Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. Vision Res. *25*, 397–406.

Prazdny, K. (1985). Detection of binocular disparities. Biol. Cybern. *52*, 93–99.

Qian, N. (1994). Computing stereo disparity and motion with known binocular cell properties. Neural Comp. *6*, 390–404.

Qian, N., and Sejnowski, T.J. (1989). Learning to solve random-dot stereograms of dense and transparent surfaces with recurrent backpropagation. Proc. 1988 Connectionist Models Summer School 435–443.

Qian, N., and Andersen, R.A. (1997). A physiological model for motion–stereo integration and a unified explanation of the Pulfrich-like phenomena. Vision Res., in press.

Qian, N., and Zhu, Y. (1997). Physiological computation of binocular disparity. Vision Res., in press.

Qian, N., Andersen, R.A., and Adelson, E.H. (1994a). Transparent motion perception as detection of unbalanced motion signals I: psychophysics. J. Neurosci. *14*, 7357–7366.

Qian, N., Andersen, R.A., and Adelson, E.H. (1994b). Transparent motion perception as detection of unbalanced motion signals III: modeling. J. Neurosci., *14*, 7381–7392.

Ramachandran, V.S., Rao, V.M., and Vidyasagar, T.R. (1973). The role of contours in stereopsis. Nature *242*, 412–414.

Regan, D., and Beverley, K.I. (1973). Disparity detectors in human depth perception: evidence for directional selectivity. Nature *181*, 877–879.

Roy, J., and Wurtz, R.H. (1990). The role of disparity-sensitive cortical neurons in signaling the direction of self-motion. Nature *348*, 160–162.

Sanger, T.D. (1988). Stereo disparity computation using gabor filters. Biol. Cybern. *59*, 405–418.

Sato, T., and Nishida, S. (1993). Second order depth perception with texture-defined random-check stereograms. Invest. Opthalmol. Vis. Sci. Suppl. (ARVO) *34*, 1438.

Schor, C.M., and Wood, I. (1983). Disparity range for local stereopsis as a function of luminance spatial frequency. Vision Res. *23*, 1649–1654.

Skottun, B.C., DeValois, R.L., Grosof, D.H., Movshon, J.A., Albrecht, D.G., and Bonds, A.B. (1991). Classifying simple and complex cells on the basis of reponse modulation. Vision Res. *31*, 1079–1086.

Smallman, H.S., and MacLeod, D.I. (1994). Size–disparity correlation in stereopsis at contrast threshold. J. Opt. Soc. Am. A *11*, 2169–2183.

Somers, D.C., Nelson, S.B., and Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. J. Neurosci. *15*, 5448–5465.

Spillman, L., and Werner, J.S. (1996). Long-range interaction in visual perception. Trends Neurosci. *19*, 428–434.

Tyler, C.W. (1974). Stereopsis in dynamic visual noise. Nature *250*, 781–782.

Wagner, H., and Frost, B. (1993). Disparity-sensitive cells in the owl have a characteristic disparity. Nature *364*, 796–798.

Westheimer, G. (1986). Spatial interaction in the domain of disparity signals in human stereoscopic vision. J. Physiol. *370*, 619–629.

Westheimer, G. (1990). Detection of disparity motion by the human observer. Optom. Vis. Sci. *67*, 627–630.

Wilson, H.R., Ferrera, V.P., and Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. Vis. Neurosci. *9*, 79–97.

Zhu, Y., and Qian, N. (1996). Binocular receptive fields, disparity tuning, and characteristic disparity. Neural Comp. *8*, 1647–1677.