PART I DEPTH PROCESSING AND STEREOPSIS

2

Physiologically based models of binocular depth perception

NING QIAN AND YONGJIE LI

2.1 Introduction

We perceive the world as three-dimensional. The inputs to our visual system, however, are only a pair of two-dimensional projections on the two retinal surfaces. As emphasized by Marr and Poggio (1976), it is generally impossible to uniquely determine the three-dimensional world from its two-dimensional retinal projections. How, then, do we usually perceive a well-defined three-dimensional environment? It has long been recognized that since the world we live in is not random, the visual system has evolved and developed to take advantage of the world's statistical regularities, which are reflected in the retinal images. Some of these image regularities, termed depth cues, are interpreted by the visual system as depth. Numerous depth cues have been discovered. Many of them, such as perspective, shading, texture, motion, and occlusion, are present in the retina of a single eye, and are thus called monocular depth cues. Other cues are called binocular, as they result from comparing the two retinal projections. In the following, we will review our physiologically based models for three binocular depth cues: horizontal disparity (Qian, 1994; Chen and Qian, 2004), vertical disparity (Matthews *et al.*, 2003), and interocular time delay (Qian and Andersen, 1994; Qian and Freeman, 2009). We have also constructed a model for depth perception from monocularly occluded regions (Assee and Qian, 2007), another binocular depth cue, but have omitted it here owing to space limitations.

Vision in 3D Environments, ed. L. R. Harris and M. Jenkin. Published by Cambridge University Press.

2.2 Horizontal disparity and the energy model

The strongest binocular depth cue is the horizontal component of the binocular disparity, defined as the positional difference between the two retinal projections of a given point in space (Figure 2.1). It has long been recognized that the brain uses the horizontal disparity to estimate the relative depths of objects in the world with respect to the fixation point, a process known as stereovision or stereopsis (Howard, 2002). With retinal positions expressed as visual angles,

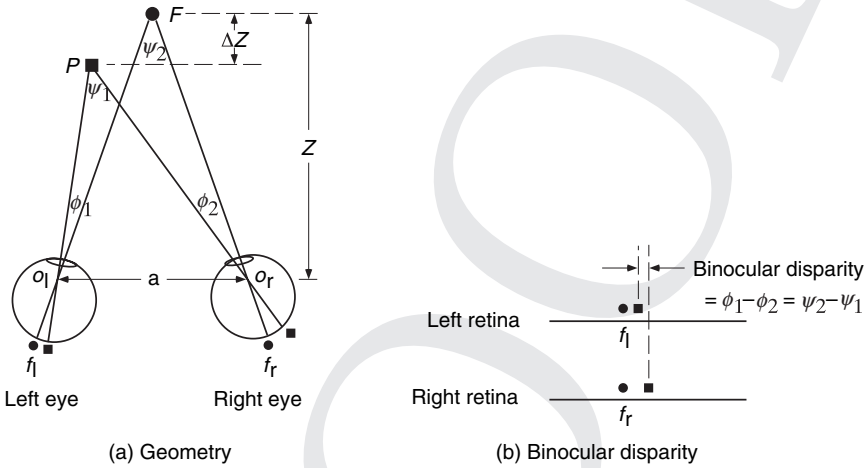


Figure 2.1 The geometry of binocular projection (a) and the definition of binocular disparity (b). For simplicity, we consider only the plane of regard defined by the instantaneous fixation point (F) and the optical centers (o_l and o_r) of the two eyes (i.e., the points in the eyes' optical system through which the light rays can be assumed to pass in straight lines). The two foveas (f_l and f_r) are considered as corresponding to each other and thus have zero disparity. To make clear the positional relationship between other locations on the two retinas, one can imagine superimposing the two retinas with the foveas aligned (bottom). The fixation point F in space projects approximately to the two corresponding foveas (f_l and f_r), with a near-zero disparity. The disparity of any other point in space can then be defined as $\phi_1 - \phi_2$, which is equal to $\psi_2 - \psi_1$. It then follows that all zero-disparity points in the plane fall on the so-called Vieth–Müller circle passing through the fixation point and the two optical centers, since all circumference angles corresponding to the same arc ($o_l o_r$) are equal. Other points in the plane do not project to corresponding locations on the two retinas, and thus have nonzero disparities. Each circle passing through the two optical centers defines a set of isodisparity points. When the fixation distance is much larger than the interocular separation and the gaze direction is not very eccentric, the constant-disparity surfaces can be approximated by frontoparallel planes.

the horizontal disparity H for an arbitrary point P in Figure 2.1a is defined as $\phi_1 - \phi_2$, which is equal to $\psi_2 - \psi_1$ based on geometry. From further geometrical considerations, it can be shown that if the eyes are fixating at a point F at a distance Z at a given instant, then the horizontal disparity H of a nearby point P at a distance $Z + \Delta Z$ is approximately given by

$$H \approx a\frac{\Delta Z}{Z^2} \tag{2.1}$$

where a is the interocular separation, and H is measured in radians of visual angle.[1] The approximation is good provided that the spatial separation between the two points is small compared with Z. The inverse square relationship in Eq. (2.1) can be easily understood. ϕ_1 and ϕ_2 are the visual angles spanned by the separation *PF* at the two eyes, and are thus inversely proportional to the fixation distance Z plus higher-order terms. Since $H = \phi_1 - \phi_2$, the $1/Z$ term is canceled by the subtraction and the next most important term is thus proportional to $1/Z^2$.

Because simple geometry provides relative depth given retinal disparity, one of the main problems of stereovision is how the brain measures disparity from the two retinal images in the first place. Many algorithms for disparity computation have been proposed. Most of them, however, have emphasized the ecological, mathematical, or engineering aspects of the problem, while often ignoring relevant neural mechanisms. For example, a whole class of models are based on Marr and Poggio (1976)'s approach of starting with all possible matches between the features (such as dots or edges) in the two half images of a stereogram and then introducing constraints to eliminate the false matches and compute the disparity map. These models literally assume that there are cells that respond to only a specific match and nothing else. In reality, even the most sharply tuned binocular cells respond to a range of disparities (Nikara *et al.*, 1968; Maske *et al.*, 1984; Bishop and Pettigrew, 1986; Poggio and Fischer, 1977; Poggio and Poggio, 1984). If these models are revised to use realistic, distributed disparity representation, then it is not known how to implement the constraints needed for disparity computation (Assee and Qian, 2007). The style of disparity computation in the brain seems to be fundamentally different from these models (Qian, 1997).

In an effort to address this shortcoming, we have constructed physiologically based algorithms for disparity computation according to the quantitative properties of binocular cells in the visual cortex reported by Ohzawa and coworkers

[1] The disparity of the fixation point itself is usually very small (McKee and Levi, 1987; Howard, 2002) and can be assumed to be zero when it is not the subject of study.

(Ohzawa *et al.*, 1996; Freeman and Ohzawa, 1990; DeAngelis *et al.*, 1991; Ohzawa *et al.*, 1990, 1997). These investigators mapped binocular receptive fields (RFs) of primary visual cortical cells in detail and proposed a model for describing their responses.

Let us first consider simple cells. Two different models for describing binocular simple-cell RFs have been proposed. Early physiological studies suggested that there is an overall positional shift between the left and right RFs of binocular simple cells (Bishop and Pettigrew, 1986). The shapes of the two RF profiles of a given cell were assumed to be identical (Figure 2.2a). In contrast, later quantitative studies by Ohzawa *et al.* (1990) have found that the left and right RF profiles of a simple cell often possess different shapes. These authors accounted for this finding by assuming that the RF shift is between ON/OFF subregions within

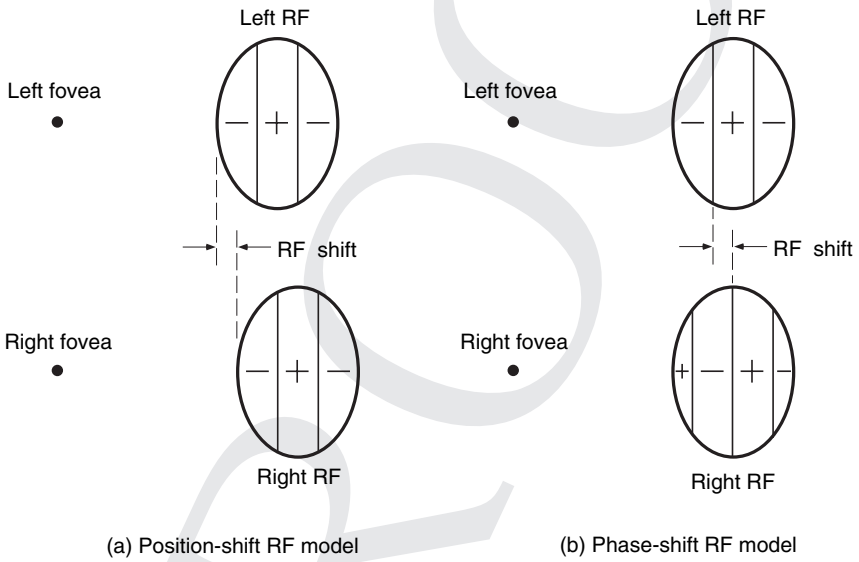


Figure 2.2 Schematic drawings illustrating the shift between the left and right receptive fields (RFs) of binocular simple cells. The "+" and "−" signs represent the ON and OFF subregions, respectively, within the RFs. Two different models for achieving the shift have been suggested by physiological experiments. (a) Position-shift model. According to this model, the left and right RFs of a simple cell have identical shapes but have an overall horizontal shift between them (Bishop and Pettigrew, 1986). (b) Phase-shift model. This model assumes that the shift is between ON/OFF subregions within the left and right RF envelopes that spatially align (Ohzawa *et al.*, 1990, 1996; DeAngelis *et al.*, 1991). The fovea locations on the left and right retinas are drawn as a reference point for vertically aligning the left and right RFs of a simple cell. Modified from Figure 2 of Qian (1997).

identical RF envelopes that spatially align (Figure 2.2b). This new RF model is often referred to as the phase shift, phase difference, or phase parameter model. For ease of description, the shift, expressed in terms of the visual angle, in both of these alternatives will be referred to as the "RF shift" (Figure 2.2) when it is not essential to distinguish between them. Later, we will discuss important differences between these two RF models. Figure 2.2 only shows the ON and OFF subregions of the RFs schematically. As is well known, the details of the RF profiles of simple cells can be described by the Gabor function, which is a Gaussian envelope multiplied by a sinusoidal modulation (Marcělja, 1980; Daugman, 1985; McLean and Palmer, 1989; Ohzawa *et al.*, 1990; DeAngelis, Ohzawa and Freeman, 1991; Ohzawa *et al.*, 1996; Anzai *et al.*, 1999b). The Gaussian envelope determines the overall dimensions and location of the RF, while the sinusoidal modulation determines the ON and OFF subregions within the envelope.

Since disparity itself is a shift between the two retinal projections (Figure 2.1), one might expect that a binocular simple cell would give the best response when the stimulus disparity happens to match the cell's left–right RF shift. In other words, a simple cell might *prefer* a disparity equal to its RF 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 because the very first assumption – that a binocular simple cell has a preferred disparity equal to its RF shift – is not always valid; it is only true for simple patterns (such as bars or gratings) undergoing coherent motion, and not for any static patterns, nor for moving or dynamic stereograms with complex spatial profiles (such as random-dot patterns) (Qian, 1994; Chen *et al.*, 2001). Simple cells cannot generally have a well-defined preferred disparity, because their responses depend not only on the disparity but also 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 location of the peak 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 is reflected in the phase of its Fourier transform. Because of the phase dependence, simple-cell responses cannot explain the fact that we can detect disparities in static stereograms and in complex dynamic stereograms.

The phase dependence of simple-cell responses can be understood intuitively by considering the disparity tuning of a simple cell to a static 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 in the left and right RFs of the simple cell. A line with a given disparity may evoke a strong response at one lateral position because it happens to project onto the excitatory subregions of both the left and the right RFs, but may evoke a much weaker response at a different lateral position because it now stimulates some inhibitory portions of the RFs. Therefore, the response of the simple cell to a fixed disparity changes with changes in the Fourier phases of the stimulus 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 the disparity tuning curves of simple cells measured with bright bars and dark bars (whose Fourier phases differ by π) were 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 its Fourier phase was varied 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; Chen *et al.*, 2001).

While simple cells are not generally suited for disparity computation, owing to their phase dependence, the responses of complex cells do have the desired phase independence, as expected from their lack of separate ON and OFF subregions within their RFs (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. Fortunately, a model for describing the responses of binocular complex cells has been proposed by Ohzawa and coworkers based on their quantitative physiological studies (Ohzawa *et al.*, 1990, 1997; Anzai *et al.*, and Freeman, 1999c). The model is known as the disparity energy model, since it is a binocular extension of the well-known motion energy model (Adelson and Bergen, 1985; Watson and Ahumada, 1985). Ohzawa *et al.*, (1990) found that a binocular complex cell in the cat primary visual cortex can be simulated by summing 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 from their left and right RFs (see Figure 2.3). (Two binocular simple cells are said to form a quadrature pair if there is a quarter-cycle shift between the ON/OFF subregions of their left and right RFs (Ohzawa *et al.*, 1990; Qian, 1994).)

The remaining questions are whether the model complex cells constructed in this way are indeed independent of the Fourier phases of the stimulus and, if so, how their preferred disparities are related to their RF parameters. We have

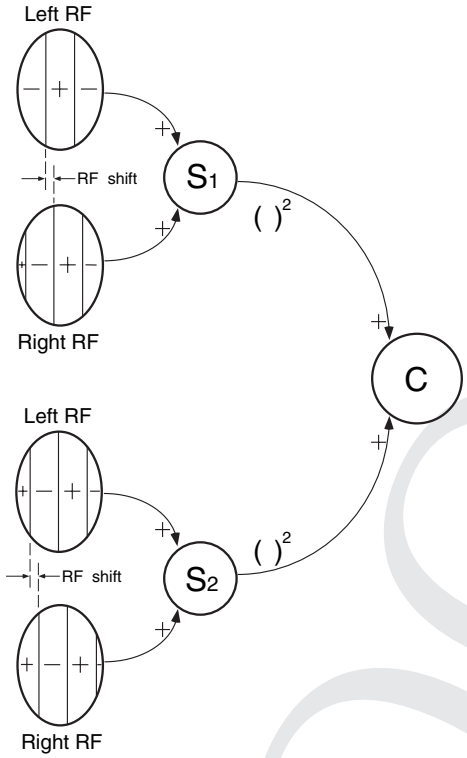


Figure 2.3 The model proposed by Ohzawa *et al.*, (1990) for describing the response of binocular complex cells. The complex cell (labeled C in the figure) sums the squared outputs of a quadrature pair of simple cells (labeled S_1 and S_2). Each simple cell, in turn, sums the contributions from its two RFs on the left and right retinas. The left RF of S_2 differs from the left RF of S_1 by a quarter-cycle shift. Likewise, the two right RFs also differ by a quarter-cycle shift. Several mathematically equivalent variations of model are discussed in the text. Reproduced from Figure 5 of Qian (1997).

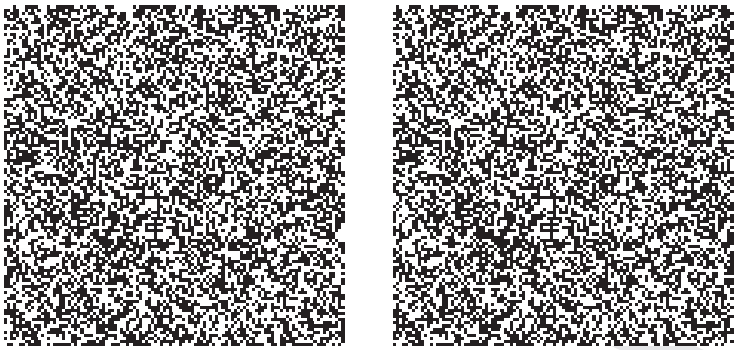
investigated these issues through mathematical analyses and computer simulations (Qian, 1994; Zhu and Qian, 1996; Qian and Zhu, 1997). The complex-cell model was found to be independent of the Fourier phases of the stimulus for simple stimuli, including the bars used in the physiological experiments of Ohzawa *et al.*, (1990), and its preferred disparity was approximately equal to the left–right RF shift within the constituent simple cells. For more complicated stimuli such as random-dot stereograms, however, a complex cell constructed from a single quadrature pair of simple cells is still phase-sensitive, albeit less so than simple cells. This problem can be easily solved by considering the additional physiological fact that complex cells have somewhat larger RFs than those

of simple cells (Hubel and Wiesel, 1962). We incorporated this fact into the model by spatially pooling *several* quadrature pairs of simple cells with nearby and overlapping RFs to construct a model complex cell (Zhu and Qian, 1996; Qian and Zhu, 1997). The resulting complex cell was largely phase-independent for any stimulus, and its preferred disparity was still approximately equal to the RF shift within the constituent simple cells.

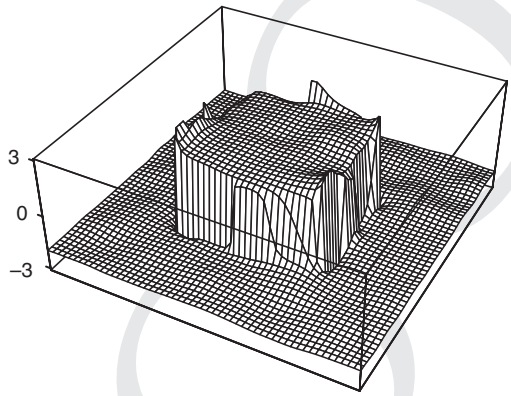
With the above method for constructing reliable complex-cell responses, and the relationship derived by that method between the preferred disparity and the RF parameters, we were finally ready to develop, for the first time, a stereo algorithm for solving stereograms using physiological properties of binocular cells (Qian, 1994; Zhu and Qian, 1996; Qian and Zhu, 1997). By using a population of complex cells tuned to the same preferred spatial frequency and with their preferred disparities covering the range of interest, the disparity of an input stimulus could 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 the application of this algorithm to a random-dot stereogram is shown in Figure 2.4.

A mathematical analysis of these 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 essentially an efficient version of cross-correlation (Qian and Zhu, 1997; Qian and Mikaelian, 2000). Since the disparity is a shift between two retinal projections, it is certainly reasonable to use a cross-correlation-like operation to compute it. Qian and Mikaelian (2000) also compared this energy-based algorithm with the so-called phase algorithm in computer vision (Sanger, 1988; Fleet *et al.*, 1991) (which should not be confused with the phase-shift RF model).

It has been demonstrated experimentally that complex cells receive monosynaptic inputs from simple cells but not vice versa (Alonso and Martinez, 1998), as required by the model. On the other hand, 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 and coworkers (DeAngelis *et al.*, 1991; Ohzawa *et al.*, 1990, 1996, 1997), the method is at least valid as a phenomenological description of a subset of real complex-cell responses. In addition, our analyses indicate that the same phase-independent complex-cell responses can be obtained by combining the outputs of many simple cells to average out their phase sensitivity, without requiring the specific quadrature relationship (Qian, 1994; Qian and Andersen, 1997; Qian and Mikaelian, 2000). Two other variations of the model also lead to the same complex-cell responses. The first considers the fact that cells cannot fire negatively. Therefore, each simple cell in Figure 2.3 should be split into



(a) Random-dot stereogram



(b) Computed disparity

Figure 2.4 A random-dot stereogram (a) and the computed disparity map (b). The stereogram has 110×110 pixels with a dot density of 50%. The central 50×50 area and the surrounding area have disparities of 2 and -2 pixels, respectively. When fused with uncrossed eyes, the central square appears further away than the surround. The disparity map of the stereogram was computed with eight complex cells (with the same spatial scale but different preferred disparities) at each location. The distance between two adjacent sampling lines represents a distance of two pixel spacings in the stereogram. Negative and positive values indicate near and far disparities, respectively. The disparity map can be improved by combining information across different scales (Chen and Qian, 2004). Modified from Figures 4 and 8 of Qian and Zhu (1997).

a push–pull pair with inverted RF profiles, so that they can carry the positive and negative portions of the original responses without using negative firing rates (Ohzawa *et al.*, 1990). In the second variation, the squaring operation in Figure 2.3 is considered to occur at the stage of simple-cell responses and the

complex cell simply sums the simple-cell responses (Heeger, 1992; Anzai *et al.*, 1999b,c; Chen *et al.*, 2001).

Although the disparity energy model was originally proposed based on data from cats (Ohzawa *et al.*, 1990), later studies indicate that the same approach can be used to describe the responses of monkey binocular cells as well (Livingstone and Tsao, 1999; Cumming and DeAngelis, 2001). One difference, though, is that monkeys have a much smaller fraction of simple cells than cats do; most monkey V1 cells appear to be complex. The energy model, however, requires that there be more simple cells than complex cells. This difficulty could be alleviated by assuming that for many complex cells in monkeys, a stage similar to the simple-cell responses happens in the dendritic compartments of complex cells. In other words, simple-cell-like properties could be constructed directly from inputs from the lateral geniculate nucleus to a dendritic region of a complex cell. The simple-cell-like responses from different regions of the dendritic tree are then pooled in the cell body to give rise to complex-cell response properties. This scheme is also consistent with the observation that some complex cells seem to receive direct inputs from the lateral geniculate nucleus (Alonso and Martinez, 1998).

2.3 Disparity attraction and repulsion

After demonstrating that our physiologically based method could effectively extract binocular-disparity maps from stereograms, we then applied the model to account for some interesting perceptual properties of stereopsis. For example, the model can explain the observation that we can still perceive depth when the contrasts of the two images in a stereogram are different, so long as they have the same sign, and the reliability of depth perception decreases with the contrast difference (Qian, 1994; Smallman and McKee, 1995; Qian and Zhu, 1997; Qian and Mikaelian, 2000). We also applied the model to a psychophysically observed depth illusion reported by Westheimer (1986) and Westheimer and Levi (1987). These authors found that when a few isolated features with different disparities are viewed foveally, the perceived disparity between them is smaller (attraction) or larger (repulsion) than the actual value, depending on whether their lateral separation is smaller or larger than several minutes of arc. If the separation is very large, there is no interaction between the features. We showed that these effects are a natural consequence of our disparity model (Mikaelian and Qian, 2000). The interaction between the features in the model originates from their simultaneous presence in the cells' RFs, and by pooling across cells tuned to different frequencies and orientations, the psychophysical

results can be explained without introducing any ad hoc assumptions about the connectivity of the cells (Lehky and Sejnowski, 1990).

2.4 Vertical disparity and the induced effect

We have focused so far on the computation of horizontal disparity – the primary cue for stereoscopic depth perception. It has been known since the time of Helmholtz that vertical disparities between the two retinal images can also generate depth perception (Howard, 2002). The mechanism involved, however, is more controversial.

The best-known example of depth perception from vertical disparity is perhaps the so-called induced effect (Ogle, 1950): a stereogram made from two identical images but with one of them slightly magnified vertically (Figure 2.5a) is perceived as a slanted surface rotated about a vertical axis (Figure 2.5b). The surface appears further away on the side with the smaller image, and the apparent axis of rotation is the vertical meridian through the point of fixation (Ogle, 1950; Westheimer and Pettet, 1992). To better appreciate this phenomenon, we indicate in Figure 2.6a the signs of the depth and disparity in the four quadrants around the point of fixation for the specific case of a left-image magnification. The features in the left image (filled dots) are then outside the corresponding features in the right image (open dots), as shown. The perceived slant is such that the first and fourth quadrants appear far and the second and third quadrants appear near with respect to the fixation point. It then follows that

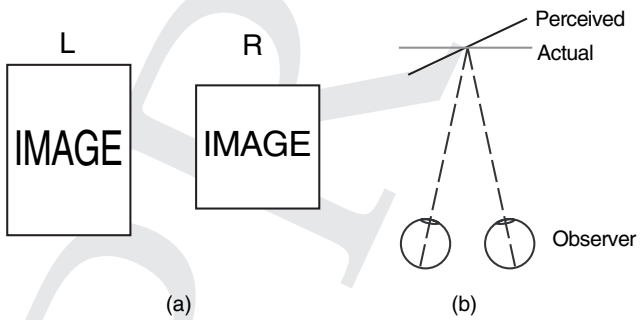


Figure 2.5 (a) A schematic stereogram for the induced effect (Ogle, 1950). The left eye's view (L) is magnified vertically with respect to the right eye's view (R). (b) With a stereogram like that in (a), a slanted surface is perceived, shown schematically in the top view, as if the right image had been magnified horizontally (Ogle, 1950). Reproduced from Figure 1 of Matthews *et al.* (2003).

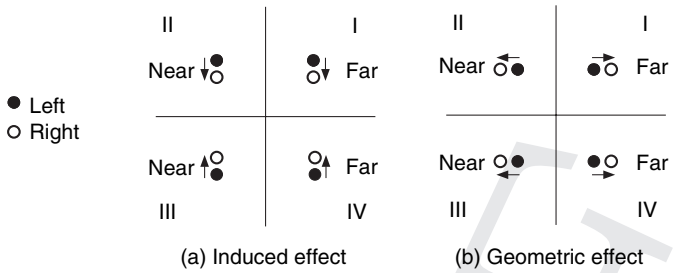


Figure 2.6 The signs of the disparity and depth for (a) the induced effect (vertical disparity) and (b) the geometric effect (horizontal disparity). For clarity, the features in the left and right images are represented schematically by filled and open dots, respectively. In each panel, the fixation point is at the center of the cross, which divides the space into four quadrants. The arrows indicate the signs of the disparity in the four quadrants caused by (a) a vertical magnification in the left eye and (b) a horizontal magnification in the right eye. The sign of the perceived depth (near or far) in each quadrant is also indicated. Note that the depth sign of the vertical disparity is quadrant-dependent (Westheimer and Pettet, 1992), while that of the horizontal disparity is not. Reproduced from Figure 2 of Matthews *et al.* (2003).

the opposite vertical-disparity signs in the first and fourth quadrants generate the same depth sign (far), and that the same vertical-disparity signs in the first and second quadrants generate opposite depth signs (far and near, respectively). In other words, the depth sign of a given vertical disparity depends on the quadrants around the fixation point (Westheimer and Pettet, 1992). To generate the same kind of surface slant with horizontal disparity (termed the "geometric effect" by Ogle (1950)), one would have to magnify the *right* image horizontally. Unlike the case for the vertical disparity, however, the depth sign of the horizontal disparity is fixed and independent of the quadrant (Figure 2.6b).

These and other considerations have led to the widely accepted notion that the role of vertical disparity is fundamentally different from that of horizontal disparity. In particular, since the vertical disparity is large at large retinal or gaze eccentricity and does not have a consistent local depth sign, and since the effect of vertical disparity can be best demonstrated with large stimuli (Rogers and Bradshaw, 1993; Howard and Kaneko, 1994) and appears to be averaged over greater areas than that of horizontal disparity (Kaneko and Howard, 1997), it is believed that the effect of vertical disparity is global, while the effect of horizontal disparity is local. Numerous theories of vertical disparity have been proposed (Ogle, 1950; Koenderink and van Doorn, 1976; Arditi *et al.*, 1981; Mayhew and Longuet-Higgins, 1982; Gillam and Lawergren, 1983; Rogers and

Bradshaw, 1993; Howard and Kaneko, 1994; Liu *et al.*, 1994; Gårding *et al.*, 1995; Banks and Backus, 1998; Backus *et al.*, 1999); many of them employ some form of global assumption to explain the induced effect. For example, Mayhew and Longuet-Higgins (Mayhew, 1982; Mayhew and Longuet-Higgins, 1982) proposed that the unequal vertical image sizes in the two eyes are used to estimate two key parameters of the viewing system: the absolute fixation distance and the gaze angle. Since the horizontal disparity is dependent on these parameters, the estimated parameters will modify the interpretation of horizontal disparity globally, and hence the global depth effect of vertical disparity. There are, however, several challenges to this theory. First, the predicted depth-scaling effect of vertical disparity cannot be observed with display sizes ranging from 11° (Cumming *et al.*, 1991) to 30° (Sobel and Collett, 1991). The common argument that these displays are simply not large enough is unsatisfactory because the induced effect *can* be perceived with these display sizes. Furthermore, even with stimuli as large as 75°, the observed scaling effect is much weaker than the prediction (Rogers and Bradshaw, 1993). Second, the predicted gaze-angle shift caused by vertical magnification is never perceived, and additional assumptions are needed to explain this problem (Bishop, 1996). Third, to account for the results under certain stimulus conditions, the theory has to assume that multiple sets of viewing-system parameters are used by the visual system at the same time, an unlikely event (Rogers and Koenderink, 1986).

A general problem applicable to all purely global interpretations of vertical disparity, including the theory of Mayhew and Longuet-Higgins, is that vertical disparity *can* generate reliable (albeit relatively weak) local depths even in small displays that are viewed foveally (Westheimer, 1984; Westheimer and Pettet, 1992; Matthews *et al.*, 2003). One might argue that functionally, the depth effect of vertical disparity in small displays is not as important as the induced effect in the case of large stimuli because the vertical disparity is usually negligible near the fovea, while full-field vertical size differences between the eyes can occur naturally with eccentric gaze. However, as pointed out by Farell (1998), the vertical disparity can be quite large even near the fovea when oriented contours in depth are viewed through narrow vertical apertures. This situation is illustrated in Figure 2.7a. When the apertures are narrow enough, the horizontal disparity is largely eliminated and subjects have to rely on vertical disparity to make local depth judgments.

We have proposed a new theory for depth perception from vertical disparity (Matthews *et al.*, 2003) based on the oriented binocular RFs of visual cortical cells (Ohzawa *et al.*, 1990, 1996, 1997; DeAngelis *et al.*, 1991; Anzai *et al.*, 1999b,c) and on the radial bias of the preferred-orientation distribution in the cortex (Bauer *et al.*, 1983; Leventhal, 1983; Bauer and Dow, 1989; Vidyasagar and Henry, 1990).

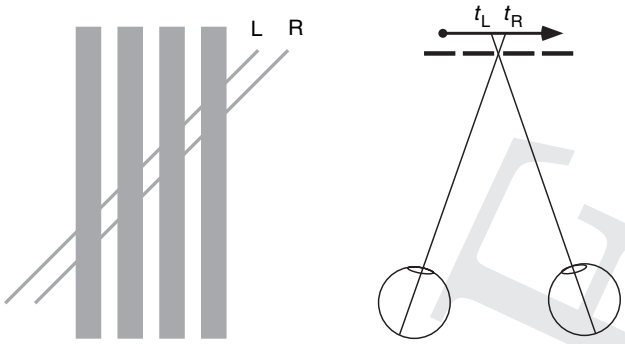


Figure 2.7 (a) An illustration of how vertical disparity can arise from horizontal disparity carried by oriented contours (Farell, 1998). The vertical occluders have zero disparity, while the diagonal line has a far horizontal disparity between its left (L) and right (R) images. The visible segments between the occluders have disparities mainly in the vertical dimension. (b) An analogous illustration of how interocular time delay can arise from horizontal disparity carried by moving targets (i.e., oriented contours in the spatiotemporal space) (Burr and Ross, 1979). The moving dot in the figure has a far horizontal disparity, but when viewed through the apertures between the occluders, it appears at the same spatial locations (i.e., the locations of the apertures) but at different times. If the y axis in (a) represents time, then (a) is the spatiotemporal representation of (b). Reproduced from Figure 4 of Matthews *et al.* (2003).

It can be shown within the framework of the disparity energy method that cells with preferred horizontal and vertical spatial frequencies ω_x^0 and ω_y^0 (and thus the same preferred orientation θ) may treat a vertical disparity V in the stimulus as an equivalent horizontal disparity given by (Matthews *et al.*, 2003)

$$H_{\text{equiv}} = \left(\frac{\omega_y^0}{\omega_x^0}\right) V = -\frac{V}{\tan\theta}. \tag{2.2}$$

The second equality holds because $\tan\theta = -\omega_y^0/\omega_x^0$ when θ is measured counterclockwise from the positive horizontal axis.[2]

Figure 2.8 provides an intuitive explanation of why oriented cells may treat a vertical disparity as an equivalent horizontal disparity. An orientation-tuned cell with a vertical offset between its left and right RFs can be approximately

[2] The negative sign is needed because when $\tan\theta$ is positive as in Figure 2.8, ω_x^0 and ω_y^0 have opposite signs according to the formal conventions of the Fourier transform.

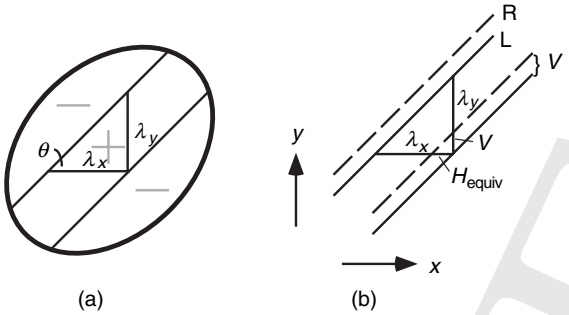


Figure 2.8 A geometric explanation of Eq. (2.2). (a) Parallel lines are drawn through the boundaries of the ON and OFF subregions of an RF profile. The horizontal and vertical distances between these lines are approximately equal to half of the preferred horizontal spatial period and half of the preferred vertical spatial period, respectively, of the cell. (b) If the left and right RFs have a vertical shift V, an equivalent horizontal shift of H_{equiv} is introduced.

viewed as having an equivalent horizontal offset instead. Therefore, the cell may treat a vertical disparity in the stimulus as an equivalent horizontal disparity because, most of the time, horizontal disparity is more significant than vertical disparity owing to the horizontal separation of the eyes. To determine the equivalent horizontal disparity, note that the horizontal and vertical distances between the two adjacent parallel lines marking the ON and OFF subregions of the RFs are approximately equal to half of the preferred horizontal spatial period λ_x and half of the preferred vertical spatial period λ_y respectively, of the cell (Figure 2.8). Now suppose there is a vertical shift of V between the left and right RFs (Figure 2.8b). It is obvious that the equivalent horizontal shift is given by

$$H_{\text{equiv}} = \left(\frac{\lambda_x}{\lambda_y}\right) V = \left(\frac{\omega_y^0}{\omega_x^0}\right) V = -\frac{V}{\tan\theta},$$

which is the same as Eq. (2.2). The second equality holds because spatial periods are inversely related to the corresponding spatial frequencies. The negative sign in Eq. (2.2) is a consequence of the fact that we define both positive horizontal and positive vertical disparities in the same way (for example, as the right image position minus the left image position). For the oriented RFs shown in Figure 2.8, a positive V must lead to a negative H_{equiv}.

How can Eq. (2.2) account for the perceived depth in stereograms containing vertical disparities? According to Eq. (2.2), cells with a preferred orientation θ would treat a vertical disparity V as an equivalent horizontal disparity

$(-V/\tan\theta)$. For stimuli without a dominant orientation, such as random textures, cells tuned to all orientations, with both positive and negative signs of $\tan\theta$, will be activated. These cells will report equivalent horizontal disparities of different signs and magnitudes, and the average result across all cells should be near zero. The only possibility of seeing depth from vertical disparity in stimuli *without* a dominant orientation arises when certain orientations are overrepresented by the cells in the visual cortex and, consequently, their equivalent horizontal disparities are not completely averaged out after pooling across cells tuned to all orientations. On the other hand, if the stimuli do have a strong orientation θ_s, the cells with preferred orientation $\theta = \theta_s$ will be maximally activated and the equivalent horizontal disparity they report should survive orientation pooling. Therefore, depth perception from vertical disparity should be most effective for stimuli with a strong orientation, but will usually be less effective than horizontal disparity (Westheimer, 1984), since most stimuli will activate cells tuned to different orientations, and pooling across orientations will make the equivalent horizontal disparities weaker. A near-vertical orientation of the stimulus, however, will not easily allow cortical cells to convert a vertical disparity into an equivalent horizontal disparity, because vertically tuned cells have $\omega_y^0 = 0$ in Eq. (2.2). Similarly, a near-horizontal orientation will not be effective either, since the equivalent horizontal disparity will be too large (owing to the vanishing of ω_x^0) to be detected (unless V approaches zero). Therefore, the theory predicts that the best orientation of a stimulus for perceiving depth from vertical disparity should be around a diagonal axis.

A critical test of our theory is whether it can explain the well-known induced effect (Ogle, 1950): a stereogram made from two identical images but with one of them slightly magnified vertically is perceived as a surface rotated about the vertical axis going through the point of fixation (Figure 2.9a). First note that the induced effect can be observed in stimuli having no dominant orientation, such as random textures (Ogle, 1950). Therefore, according to the above discussion, a reliable equivalent horizontal disparity could be generated only by an overrepresentation of certain orientations in the brain. Remarkably, physiological experiments have established well a radial bias of preferred orientations around the fixation point in the cat primary visual cortex (Leventhal, 1983; Vidyasagar and Henry, 1990) and in the supragranular layers of the monkey area V1 (Bauer *et al.*, 1983; Bauer and Dow, 1989). (The supragranular layers are known to project to higher visual cortical areas (Felleman and Van Essen, 1991), and are thus more likely to be relevant than the other layers for perception.) That is, although the full range of orientations is represented for every spatial location, the orientation connecting each location and the fixation point is over-represented at that location (Figure 2.9b). This is precisely what is needed

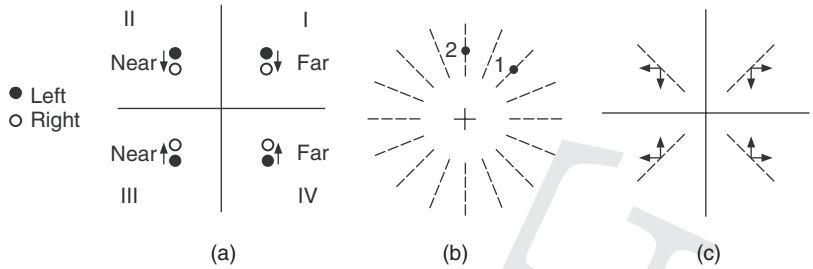


Figure 2.9 Our explanation for the induced effect and the related quadrant dependence of vertical disparity. (a) The signs of the point disparity and depth in the four quadrants around the fixation point caused by a magnification of the left image (as in Figure 2.6a). Features in the left and right images are represented by filled and open dots, respectively. The signs of the vertical disparities are indicated by arrows, and the depth signs are marked as "near" or "far". (b) The radial bias (dashed lines) of the preferred orientations around the fixation point (central cross) found in the visual cortex. For example, the 45° orientation and the vertical orientation are overrepresented for spatial locations 1 and 2, respectively. (c) Conversion of vertical disparity into equivalent horizontal disparity by the overrepresented cortical cells in the four quadrants. The four vertical-disparity arrows are copied from (a), and the four horizontal arrows indicate the signs of the equivalent horizontal disparities according to the overrepresented orientations (dashed lines) and Eq. (2.2). Reproduced from Figure 6 of Matthews *et al.* (2003).

for explaining the induced effect and the related quadrant dependence of the vertical disparity for stimuli without a dominant orientation (Figure 2.9c).

To be more quantitative, let the fixation point be the origin and assume that the left image is magnified vertically by a factor of $k\,(>1)$. Then, the vertical disparity at the stimulus location (x,y) is $V(x,y)=(k-1)y$. The radial bias means that the cortically over-represented orientation for the location (x,y) is given by $\tan\theta = y/x$. Then, according to Eq. (2.2), the corresponding equivalent horizontal disparity should be

$$H_{\text{equiv}}(x,y) = -\frac{(k-1)y}{\tan\theta} = -(k-1)x. \tag{2.3}$$

Therefore, although the vertical magnification of the left image by a factor of k creates a vertical disparity of $(k-1)y$ at the location (x,y), the over-represented equivalent horizontal disparity is $-(k-1)x$, and could be mimicked by magnifying the right image horizontally by a factor of k. The perceived surface should thus be rotated around the vertical axis going through the fixation point, which is consistent with psychophysical observations (Ogle, 1950). Note that the radial bias does not affect the depth perceived from real *horizontal*

disparity, since unlike vertical disparity, horizontal disparity is not subject to an orientation-dependent conversion.

We mentioned that the quadrant dependence of the vertical disparity means that the vertical disparity does not have a consistent local depth sign, and this may seem to imply that the induced effect can be explained only by global considerations. However, we have shown above that our local theory can account for the phenomenon very well through an orientation-dependent conversion of vertical disparity into an equivalent horizontal disparity. Our theory is consistent with the finding that vertical disparity is more effective at larger display sizes (Rogers and Bradshaw, 1993; Howard and Kaneko, 1994) and with the related observation that vertical disparity appears to operate at a more global scale than horizontal disparity (Kaneko and Howard, 1997). This is because the radial bias of cells' preferred orientations is stronger at higher eccentricities (Leventhal, 1983), although the bias is also present for foveal cells in monkey area V1 (Bauer *et al.*, 1983; Bauer and Dow, 1989). Larger displays cover more eccentric locations, and are therefore more effective.

For small displays, the effect of vertical disparity is harder to observe because of the weaker radial orientation bias in the brain; however, our theory predicts that the effect can be made stronger by using a near-diagonal orientation of the stimulus. Our theory predicts further that when there is both horizontal and vertical disparity, the total horizontal disparity should be equal to the actual horizontal disparity plus the equivalent horizontal disparity generated by the vertical disparity. Therefore, these two types of disparity should locally enhance or cancel each other depending on their depth signs. We have tested and confirmed these predictions using diagonally oriented stimuli (Matthews *et al.*, 2003).

Our theory also makes specific physiological predictions. First, there should be a population of V1 cells that shows *both* disparity tuning *and* orientation bias, and the bias should be stronger at greater eccentricity. Second, V1 cells' responses to a given vertical disparity should depend on their preferred orientation. These predictions have been confirmed in a subsequent physiological study by Durand *et al.*, (2006) who concluded that "our results directly demonstrate both assumptions of this model."

2.5 Relative versus absolute disparity

The disparity defined in Figure 2.1 – the positional difference between the left and right retinal projections of a point in space – is more precisely called the absolute disparity. The difference between the absolute disparities of two points is termed the relative disparity between those points. Since the fixation

point disparity is usually very small and stable (McKee and Levi, 1987; Howard and Rogers, 1995), the absolute disparity of a point is approximately equal to the relative disparity between that point and the fixation point. It is therefore difficult to distinguish between the two types of disparity under most normal viewing conditions, where many points with different disparities are present and one of the points is the fixation point at any given instant. One might hope to create a situation without relative disparity and with only absolute disparity by presenting a stimulus with a single disparity. However, under this condition, the stimulus will trigger a vergence eye movement, which quickly reduces the absolute disparity to near zero.

In the laboratory, it is possible to use a feedback loop to maintain a constant absolute disparity (Rashbass and Westheimer, 1961). With such a procedure, it has been shown that V1 cells encode absolute disparity (Cumming and Parker, 1999, 2000). Since binocular depth perception is known to rely mainly on relative disparity (Westheimer, 1979; Howard and Rogers, 1995), it is thus possible that a higher visual cortical area converts absolute disparity into relative disparity through simple subtraction (Cumming and DeAngelis, 2001; Neri *et al.*, 2004; Umeda *et al.*, 2007). Although we have constructed our models based on V1 RF properties, we do not infer that binocular depth perception necessarily happens in V1; later stages may have similar RF properties, or may simply inherit and refine V1 responses to generate perception. On the other hand, it is neither economical nor necessary for the brain to encode relative disparity across the entire binocular visual field. Assume that the brain has computed absolute disparities at N points in a scene. Since there are $N(N-1)$ ordered pairs of the N points, a much greater amount of resources would be required for the brain to convert and store all the $N(N-1)$ relative-disparity values. An alternative possibility is that the brain might simply use absolute disparity across the whole field as an implicit representation of the relative disparity, and compute the relative disparity explicitly only for the pair of points under attentional comparison at any given time. The fact that depth perception from a single absolute disparity is poor may be a simple reflection of poor depth judgment from vergence.

One might argue that a relative-disparity map is more economical because, unlike absolute disparity, it does not change with vergence and thus does not have to be recomputed with each vergence eye movement. However, since saccades and head/body movements are frequent, and the world is usually not static, the brain has to recompute the disparity map frequently anyway. Also, the fact that V1 encodes absolute disparity suggests that it might be too difficult to compute relative disparity directly without computing absolute disparity first.

2.6 Phase-shift and position-shift RF models and a coarse-to-fine stereo algorithm

We mentioned earlier that two different models for binocular simple-cell RFs have been proposed: the position-shift and phase-shift models (Figure 2.2). Much of what we have discussed above applies to both RF models. However, there are also important differences between them (Zhu and Qian, 1996; Qian and Mikaelian, 2000; Chen and Qian, 2004). For example, we have analyzed disparity tuning to sinusoidal gratings and broadband noise (such as random-dot stereograms) for the position- and phase-shift models (see Eqs. (2.11)–(2.15) in Zhu and Qian (1996) and related work in Fleet *et al.*, (1996)). For a complex cell with a phase shift $\Delta\phi$ between the left and right RFs and a preferred spatial frequency ω_0, its peak response to noise occurs at the preferred disparity

$$D^{\text{phs}}_{\text{noise}} = \frac{\Delta\phi}{\omega_0}. \tag{2.4}$$

Around this disparity, one peak in the periodic response to a sinusoidal grating with spatial frequency Ω occurs at

$$D^{\text{phs}}_{\text{sin}} = \frac{\Delta\phi}{\Omega} = D^{\text{phs}}_{\text{noise}}\frac{\omega_0}{\Omega}. \tag{2.5}$$

In contrast, for a cell with a positional shift d, these peaks are all aligned at d, the cell's preferred disparity:

$$D^{\text{pos}} = d. \tag{2.6}$$

Therefore, near a cell's preferred disparity for noise stimuli, the preferred disparity for sinusoidal gratings depends on the spatial frequency Ω of the grating for phase-shift RFs but not for position-shift RFs (Zhu and Qian, 1996). Such a dependence has been observed in the visual Wulst of the barn owl (Wagner and Frost, 1993), supporting the phase-shift model originally proposed for the cat V1 (Ohzawa *et al.*, 1990). On the other hand, the preferred disparity of phase-shift cells is limited to plus or minus half of the preferred spatial period of the cells (Blake and Wilson, 1991; Freeman and Ohzawa, 1990; Qian, 1994; Smallman and MacLeod, 1994), and some real cells in the barn owl do not follow this constraint strictly (Zhu and Qian, 1996). It thus appears that both the phase- and the position-shift RF mechanisms are used to code disparity. Later physiological experiments on cats and monkeys have confirmed that a mixture of the two RF models is the best description of the binocular cells in these species (Anzai *et al.*, 1997, 1999a; Cumming and DeAngelis, 2001).

The above discussion of the two RF models prompted Chen and Qian (2004) to ask "what are the relative strengths and weaknesses of the phase- and the position-shift mechanisms in disparity computation, and what is the advantage, if any, of having both mechanisms?" With appropriate parameters, either type of RF model (or a hybrid of them) can be used as a front-end filter in the energy method for disparity computation described earlier (Qian and Zhu, 1997). However, our analysis and our simulations over a much wider range of parameters reveal some interesting differences between the two RF models in terms of disparity computation (Chen and Qian, 2004). The main finding is that the phase-shift RF model is, in general, more reliable (i.e., less variable) than the position-shift RF model for disparity computation. The accuracy of the computed disparity is very good for both RF models at small disparity, but it deteriorates at large disparity. In particular, the phase-shift model tends to underestimate the magnitude of the disparity owing to a zero-disparity bias (Qian and Zhu, 1997). Additionally, both RF models are only capable of dealing well with disparity within plus or minus half of the preferred spatial period of the cells. This was known earlier for the phase-shift model (see above). It turns out that the position-shift model has a similar limitation: although position-shift cells can have large preferred disparities, the responses of a population of them for disparity computation often has false peaks at large preferred disparities (Chen and Qian, 2004).

These results and the physiological data of Menz and Freeman (2003) suggest a coarse-to-fine stereo algorithm that takes advantage of both the phase-shift and the position-shift mechanisms (Chen and Qian, 2004). In this algorithm, disparity computation is always performed by the phase-shift mechanism because of its higher reliability over the entire disparity range. Since the phase-shift model is accurate only when the disparity is small, the algorithm iteratively reduces the magnitude of the disparity through a set of spatial scales by introducing a constant position-shift component for all cells to offset the stimulus disparity. Specifically, for a given stereogram, a rough disparity map is first computed with the phase-shift model at a coarse scale using the energy method (Qian, 1994). The computed disparity at each spatial position is then used as a constant position-shift component for all cells at the next, finer scale. At the next scale, different cells all have the same position-shift component but different phase-shift components so that the disparity computation is still done by the reliable phase-shift mechanism. The amount of disparity that the phase-shift component has to deal with, however, has been reduced by the common position-shift component of all cells, and the new disparity estimated from the phase-shift component will thus be more accurate. The process can be repeated across several scales. We have implemented such a coarse-to-fine algorithm and

found that it can indeed greatly improve the quality of computed disparity maps (Chen and Qian, 2004). This coarse-to-fine algorithm is similar in spirit to the one originally proposed by Marr and Poggio (1979), but with two major differences. First, we have used the position-shift component of the RFs to reduce the magnitude of the disparity at each location, while Marr and Poggio (1979) used vergence eye movement, which changes the disparity globally. Second, at each scale we have used the energy method for disparity computation, while Marr and Poggio (1979) used a nonphysiological, feature-matching procedure.

2.7 Are cells with phase-shift receptive fields lie detectors?

Recently, Read and Cumming (2007) asked the same question of why there are both phase- and position-shift RF mechanisms in the brain, but reached a different conclusion. They argued that cells with position-shift RFs code real, physical disparities while those with phase-shift RFs code impossible, nonphysical disparities and are thus "lie detectors." In particular, they believe that cells with phase-shift RFs "respond optimally to [impossible] stimuli in which the left and right eye's images are related by a constant shift in Fourier phase." It is not clear how they reached this conclusion. The phase-shift model assumes that the sinusoids of the Gabor functions for the left and right RFs have a phase shift; mathematically, however, this is *not* equivalent to a constant phase shift of the RFs' Fourier components.

Read and Cumming (2007) defined an impossible stimulus as a visual input that "never occurs naturally, ... even though it can be simulated in the laboratory." They considered a cell as coding impossible stimuli, and thus as a lie detector, if the cell responds better or shows greater response modulation to impossible stimuli than to naturally occurring stimuli (see also Haefner and Cumming, 2008). Unfortunately, this definition is problematic because, according to it, nearly all visual cells should be classified as lie detectors coding impossible stimuli. To begin with, most visual cells have retinally based RFs. To stimulate these cells optimally, the stimulus has to match the retinal location and size of the RFs. This means that the stimulus has to move with the eyes, have the right size, and be placed at the right location and distance from the eyes. Such stimuli never happen naturally. We therefore conclude that the notion of dividing cells into those coding physical and those coding impossible stimuli is not compelling. Visual cells generally respond better to artificial stimuli tailored to match their RF properties than to naturally occurring stimuli. That does not mean that they are designed to code impossible stimuli.

Read and Cumming (2007) also disputed Chen and Qian (2004)'s conclusion that phase-shift cells are more reliable than position-shift cells for disparity

computation. They correctly pointed out that the distribution of the computed disparity depends on whether the stimulus disparity is introduced symmetrically or asymmetrically between the two eyes. However, our recent simulations (Yongjie Li, Yuzhi Chen, and Ning Qian, unpublished observations) show that regardless of the symmetry, the disparity distribution computed with position-shift cells always has more outliers, and consequently a much larger standard deviation, than has the distribution computed using the phase-shift RF model. We thus maintain our conclusion that phase-shift cells are more reliable than position-shift cells for disparity computation. Read and Cumming (2007) emphasized that the population response curve for the position-shift model is symmetric when the stimulus disparity is introduced symmetrically. However, this is only true for stimuli containing a single, uniform disparity and is thus not useful for general disparity computation.

Finally, Read and Cumming (2007) proposed a new algorithm for disparity computation. A close examination reveals that this algorithm and the earlier algorithm of Chen and Qian (2004) search for the same goal in a space covered by cells with various combinations of phase shifts and position shifts, but with different search strategies. The common goal is a set of cells all having the same position-shift component, equal to the stimulus disparity, and whose phase-shift component encodes zero disparity. When multiple scales are considered, Chen and Qian (2004)'s coarse-to-fine algorithm is more efficient as it involves only a single disparity computation with phase-shift cells at each scale, while Read and Cumming (2007)'s algorithm involves multiple disparity computations, also with phase-shift cells, at each scale. Interestingly, Read and Cumming (2007)'s algorithm employs far more phase-shift-based computation than position-shift-based computation and thus also takes advantage of the better reliability of the phase-shift RF mechanism.

2.8 Motion–stereo integration

There is increasing psychophysical and physiological evidence indicating that motion detection and stereoscopic depth perception are processed together in the brain (Regan and Beverley, 1973; Nawrot and Blake, 1989; Qian *et al.*, 1994a; Maunsell and Van Essen, 1983; Bradley *et al.*, 1995; Ohzawa *et al.*, 1996). We have demonstrated that under physiologically plausible assumptions about the spatiotemporal properties of binocular cells, the stereo energy model reviewed above can be naturally combined with the motion energy model (Adelson and Bergen, 1985; Watson and Ahumada, 1985) to achieve motion–stereo integration (Qian and Andersen, 1997). The cells in the model are tuned to both motion and disparity just like physiologically observed cells, and a

population of complex cells covering a range of motion and a range of disparity combinatorially could simultaneously compute the motion and disparity of a stimulus.

Interestingly, the complex cells in the integrated model are much more sensitive to motion along constant-disparity planes than to motion in depth towards or away from the observer because the left and right RFs of a cell have the same spatiotemporal orientation (Qian, 1994; Ohzawa *et al.*, 1996, 1997; Qian and Andersen, 1997; Chen *et al.*, 2001). This property is consistent with the physiological finding that few cells in the visual cortex are truly tuned to motion in depth (Maunsell and Van Essen, 1983; Ohzawa, *et al.*, 1996, 1997) and with the psychophysical observation that human subjects are poor at detecting motion in depth based on disparity cues alone (Westheimer, 1990; Cumming and Parker, 1994; Harris *et al.*, 1998). Because of this property, motion information could help reduce the number of possible stereoscopic matches in an ambiguous stereogram by making stereo matches in frontoparallel planes more perceptually prominent than matches of motion in depth. The integrated model has also been used to explain the additional psychophysical observation that adding binocular-disparity cues to a stimulus can help improve the perception of multiple and overlapping motion fields in the stimulus (i.e., motion transparency) (Qian *et al.*, 1994b). In this explanation, it is assumed that transparent motion is usually harder to perceive than unidirectional motion because in area MT, motion signals from different directions suppress each other (Snowden *et al.*, 1991; Qian and Andersen, 1994). The facilitation of transparent-motion perception by disparity can then be accounted for by assuming that the suppression in area MT is relieved when the motion signals from different directions are in different disparity planes (Qian *et al.*, 1994a,b). This prediction of disparity-gated motion suppression in area MT has subsequently been verified physiologically (Bradley *et al.*, 1995). Finally, the integrated motion–stereo model has allowed us to explain many temporal aspects of disparity tuning (Chen *et al.*, 2001).

2.9 Interocular time delay and Pulfrich effects

Another interesting application of the integrated motion–stereo model is a unified explanation for a family of Pulfrich-like depth illusions. The classical Pulfrich effect refers to the observation that a pendulum oscillating back and forth in a frontoparallel plane appears to move along an elliptical path in depth when a neutral density filter is placed in front of one eye (Figure 2.10). The direction of apparent rotation is such that the pendulum appears to move

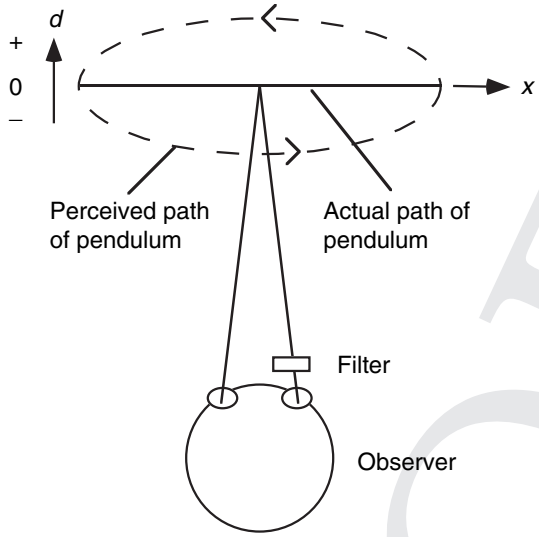


Figure 2.10 A schematic drawing of the classical Pulfrich effect (top view). A pendulum is oscillating in the frontoparallel plane indicated by the solid line. When a neutral density filter is placed in front of the right eye, the pendulum appears to move along an elliptical path in depth, as indicated by the dashed line. The direction of rotation is such that the pendulum appears to move away from the covered eye and towards the uncovered eye. Reproduced from Figure 1 of Qian and Andersen (1997).

away from the covered eye and towards the uncovered eye. It is known that by reducing the amount of light reaching the covered retina, the filter introduces a temporal delay in the transmission of 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, when the uncovered eye sees the pendulum at one position, the eye with the filter sees the pendulum at a different position back in time. In other words, the coherent motion of the pendulum converts the interocular time delay into a horizontal disparity at the level of stimuli. However, the Pulfrich depth effect is present even with dynamic noise patterns (Tyler, 1974; Falk, 1980), which lack the coherent motion required for this conversion. Furthermore, the effect is still present when a stroboscopic dot undergoing apparent motion is used such that the two eyes see the dot at exactly the same set of spatial locations but slightly different times (Morgan and Thompson, 1975; 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).

The stroboscopic version of the Pulfrich effect can occur in the real world when a target moves behind a set of small apertures (Morgan and Thompson, 1975; Burr and Ross, 1979) (Figure 2.7b). Without the occluders, the moving target has a horizontal disparity with respect to the fixation point. With the occluders, the target appears to the two eyes to be at the same aperture locations but at slightly different times. For example, in Figure 2.7b, the target appears at the location of the central aperture at times t_L and t_R. In this type of situation, the brain has to rely on interocular time delay to infer the depth of the target.

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

$$H_{\text{equiv}} = \frac{\omega_t^0}{\omega_x^0}\Delta t, \tag{2.7}$$

where ω_t^0 and ω_x^0 are the preferred temporal and horizontal spatial frequencies of the cell. This relation is analogous to Eq. (2.2), where a vertical disparity is treated as an equivalent horizontal disparity by binocular cells. It holds for any arbitrary spatiotemporal pattern (including a coherently moving pendulum, dynamic noise, and stroboscopic stimuli) that can significantly activate the cell. By considering the population 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 a simulation for a stroboscopic pendulum is shown in Figure 2.11.

Two testable predictions were made based on the analysis (Qian and Andersen, 1997). First, the responses of a binocular complex cell to interocular time delay and binocular disparity should be related according to Eq. (2.7). This prediction was confirmed by later physiological recordings by Anzai *et al.*, (2001), who concluded that "our data provide direct physiological evidence that supports the [Qian and Andersen] model." The second prediction is also based on Eq. (2.7). 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

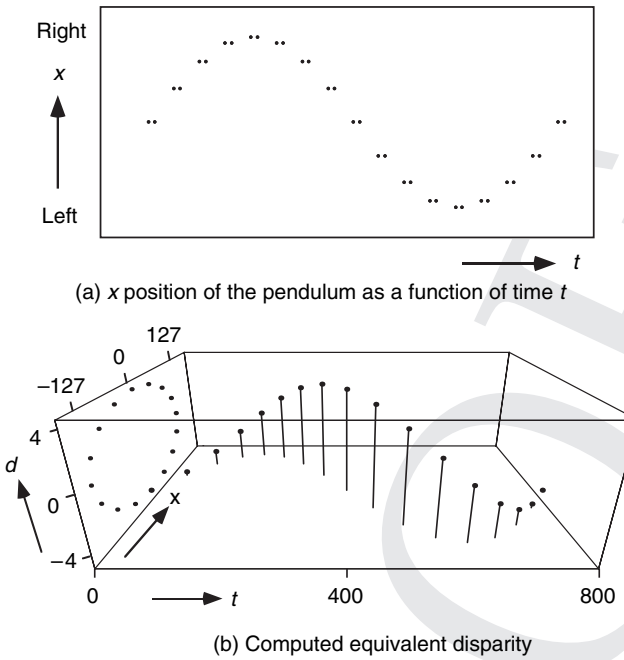


Figure 2.11 (a) A spatiotemporal representation of a stroboscopic pendulum for one full cycle of oscillation. The two dots in each pair are for the left and the right eye respectively; they are presented at exactly the same spatial location (i.e., the same x) but slightly different times. The time gap between the two sets of dots and the duration of each dot (i.e., the size of a dot along the time axis) are exaggerated in this drawing for the purpose of illustration. (b) The computed equivalent disparity as a function of horizontal position and time. The data points from the simulation are shown as small solid circles. Lines are drawn from the data points to the x–t plane in order to indicate the spatiotemporal location of each data point. 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 (it is seen as further away from the observer) when it is moving to the left. The projection of the 3D plot onto the d–x plane forms a closed path similar to the ellipse in Figure 2.10. The units are arbitrary, measured by the pixel sizes along the space and time dimensions used in the simulation. Reproduced from Figure 4 of Qian and Andersen (1997).

the perceived Pulfrich depth should vary according to Eq. (2.7) as we selectively excite different populations of cells by using stimuli with different spatial- and temporal-frequency contents. Psychophysical data are consistent with this prediction (Wist *et al.*, 1977; Morgan and Fahle, 2000).

Our Pulfrich model (Qian and Andersen, 1997) has since been known as the joint motion–disparity coding model. Despite its success, the model was

questioned by Read and Cumming (2005a,b), who argued that all Pulfrich effects can be explained by a model that codes motion and disparity separately. Read and Cumming focused on the S-shaped curves of perceived disparity as a function of interocular time delay in the stroboscopic Pulfrich effect (Morgan, 1979). However, we have recently demonstrated fundamental problems with Read and Cumming's work in terms of causality, physiological plausibility, and definitions of joint and separate coding, and we have compared the two coding schemes under physiologically plausible assumptions (Qian and Freeman, 2009). We showed that joint coding of disparity and either unidirectional or bidirectional motion selectivity can account for the S curves, but unidirectional selectivity is required to explain direction–depth contingency in Pulfrich effects. In contrast, separate coding can explain neither the S curves nor the direction–depth contingency. We conclude that Pulfrich phenomena can be logically accounted for by joint encoding of unidirectional motion and disparity.

2.10 Concluding remarks

Above, we have reviewed some of our work on physiologically based models of binocular depth perception. Our work was aimed at addressing the limitations of the current experimental and computational methods. Although experimental studies are fundamental to our understanding of visual information processing, these studies do not directly provide algorithms for how a population of cells with known properties may be used to solve a difficult perceptual problem. For example, knowing that there are tuned near- and far-disparity-selective cells in the visual cortex does not tell us how to compute disparity maps from arbitrary stereograms with these cells. Without quantitative modeling, our intuition may often be incomplete or even wrong, and it has only limited power in relating and comprehending a large amount of experimental data.

On the other hand, most computational studies of visual perception have typically been concerned with the ecological or engineering aspects of a task, while giving little or at best secondary consideration to existing physiological data. This tradition appears to stem from David Marr's overemphasis on separating computational analyses from physiological implementations (Marr, 1982). Although purely computational approaches are highly interesting in their own right, the problem is that without paying close attention to physiology, one often comes up with theories that work in some sense but have little to do with the mechanisms used by the brain. In fact, most computer vision algorithms contain nonphysiological procedures.

In this chapter, we have used examples from binocular depth perception to illustrate that given an appropriate set of experimental data, a physiologically plausible approach to the modeling of neural systems is both feasible and fruitful. The experimental and theoretical studies reviewed here suggest that although the disparity sensitivity in the visual cortex originates from left–right RF shifts in simple cells, it is at the level of complex cells that stimulus disparity is reliably coded in a distributed fashion. These studies suggest further that depth perception from vertical disparity and interocular time delay can be understood through vertical disparity and interocular time delay being treated as equivalent horizontal disparities by visual cortical cells. 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 these models. Indeed, without modeling, it would be difficult to infer that random-dot stereograms could be effectively solved by a population of binocular complex cells without resorting to explicit matching, that the psychophysically observed disparity attraction/repulsion phenomenon under different stimulus configurations could be a direct consequence of the underlying binocular RF structure, or that different variations of the Pulfrich depth illusion could all be uniformly explained by the spatiotemporal response properties of binocular cells. Physiology-based computational models have the potential to synthesize a large body of existing experimental data into a coherent framework. They can also make specific, testable predictions and, indeed, several of our key predictions have been confirmed by later experiments, as we have discussed above. Therefore, a close interplay between the experimental and computational approaches holds the best promise for resolving outstanding issues in stereovison (Qian, 1997; Chen *et al.*, 2001), and for achieving a deeper understanding of neural information processing in general.

Acknowledgments

We would like to thank our collaborators Drs. Richard Andersen, Andrew Assee, Yuzhi Chen, Julián Fernéndez, Ralph Freeman, Nestor Matthews, Xin Meng, Samuel Mikalian, Brendon Watson, Peng Xu, and Yudong Zhu for their contributions to the work reviewed here. This work was supported by NIH grant #EY016270.

References

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

Alonso, J. M. and Martinez, L. M. (1998). Functional connectivity between simple cells and complex cells in cat striate cortex. *Nature Neurosci.*, 1: 395–403.

Anzai, A., Ohzawa, I., and Freeman, R. D. (1997). Neural mechanisms underlying binocular fusion and stereopsis: position vs. phase. *Proc. Natl. Acad. Sci. USA*, 94: 5438–5443.

Anzai, A., Ohzawa, I., and Freeman, R. D. (1999a). Neural mechanisms for encoding binocular disparity: receptive field position vs. phase. *J. Neurophysiol.*, 82: 874–890.

Anzai, A., Ohzawa, I., and Freeman, R. D. (1999b). Neural mechanisms for processing binocular information: I. Simple cells. *J. Neurophysiol.*, 82: 891–908.

Anzai, A., Ohzawa, I., and Freeman, R. D. (1999c). Neural mechanisms for processing binocular information: II. Complex cells. *J. Neurophysiol.*, 82: 909–924.

Anzai, A., Ohzawa, I., and Freeman, R. D. (2001). Joint-encoding of motion and depth by visual cortical neurons: neural basis of the Pulfrich effect. *Nature Neurosci.*, 4: 513–518.

Arditi, A., Kaufman, L. and Movshon, J. A. (1981). A simple explanation of the induced size effect. *Vis. Res.*, 21: 755–764.

Assee, A. and Qian, N. (2007). Solving da Vinci stereopsis with depth-edge-selective v2 cells. *Vis. Res.*, 47: 2585–2602.

Backus, B. T., Banks, M. S., van Ee, R., and Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vis. Res.*, 39: 1143–1170.

Banks, M. S. and Backus, B. T. (1998). Extra-retinal and perspective cues cause the small range of the induced effect. *Vis. Res.*, 38: 187–194.

Bauer, R. and Dow, B. M. (1989). Complementary global maps for orientation coding in upper and lower layers of the monkey's foveal striate cortex. *Exp. Brain Res.*, 76: 503–509.

Bauer, R., Dow, B. M., Synder, A. Z., and Vautin, R. G. (1983). Orientation shift between upper and lower layers in monkey visual cortex. *Exp. Brain Res.*, 50: 133–145.

Bishop, P. O. (1996). Stereoscopic depth perception and vertical disparity: neural mechanisms. *Vis. Res.*, 36: 1969–1972.

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

Blake, R. and Wilson, H. R. (1991). Neural models of stereoscopic vision. *Trends Neurosci.*, 14: 445–452.

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? *Vis. 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. *Vis. Res.*, 29: 155–165.

Chen, Y. and Qian, N. (2004). A coarse-to-fine disparity energy model with both phase-shift and position-shift receptive field mechanisms. *Neural Comput.*, 16: 1545–1577.

Chen, Y., Wang, Y., and Qian, N. (2001). Modeling V1 disparity tuning to time-dependent stimuli. *J. Neurophysiol.*, 86: 143–155.

Cumming, B. G. and DeAngelis, G. C. (2001). The physiology of stereopsis. *Annu. Rev. Neurosci.*, 24: 203–238.

Cumming, B. G. and Parker, A. J. (1994). Binocular mechanisms for detecting motion-in-depth. *Vis. Res.*, 34: 483–495.

Cumming, B. G. and Parker, A. J. (1999). Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity. *J. Neurosci.*, 19: 5602–5618.

Cumming, B. G. and Parker, A. J. (2000). Local disparity not perceived depth is signaled by binocular neurons in cortical area V1 of the macaque. *J. Neurosci.*, 20: 4758–4767.

Cumming, B. G., Johnston, E. B., and Parker, A. J. (1991). Vertical disparities and perception of three-dimensional shape. *Nature*, 349: 411–414.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A*, 2: 1160–1169.

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

Durand, J. B., Celebrini, S., and Trotter, Y. (2006). Neural bases of stereopsis across visual field of the alert macaque monkey. *Cereb. Cortex*, 17: 1260–1273.

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

Farell, B. (1998). Two-dimensional matches from one-dimensional stimulus components in human stereopsis. *Nature*, 395: 689–693.

Felleman, D. J. and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex*, 1: 1–47.

Fleet, D. J., Jepson, A. D., and Jenkin, M. (1991). Phase-based disparity measurement. *Comput. Vis. Graphics Image Proc.*, 53: 198–210.

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

Freeman, R. D. and Ohzawa, I. (1990). On the neurophysiological organization of binocular vision. *Vis. Res.*, 30: 1661–1676.

Gårding, J., Porrill, J., Mayhew, J. E. W., and Frisby, J. P. (1995). Stereopsis, vertical disparity and relief transformations. *Vis. Res.*, 35: 703–722.

Gillam, B. and Lawergren, B. (1983). The induced effect, vertical disparity, and stereoscopic theory. *Percept. Psychophys.*, 34: 121–130.

Haefner, R. M. and Cumming, B. G. (2008). Adaptation to natural binocular disparities in primate V1 explained by a generalized energy model. *Neuron*, 57: 147–158.

Harris, J. M., McKee, S. P., and Watamaniuk, S. N. J. (1998). Visual search for motion-in-depth: stereomotion does not "pop out" from disparity noise. *Nature Neurosci.*, 1: 165–168.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Vis. Neurosci.*, 9: 181–197.

Howard, I. P. (2002). *Basic Mechanisms*. Vol. 1 of *Seeing in Depth*. Toronto: Porteous.

Howard, I. P. and Kaneko, H. (1994). Relative shear disparity and the perception of surface inclination. *Vis. Res.*, 34: 2505–2517.

Howard, I. P. and Rogers, B. J. (1995). *Binocular Vision and Stereopsis*. New York: Oxford University Press.

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.

Kaneko, H. and Howard, I. P. (1997). Spatial limitation of vertical-size disparity processing. *Vis. Res.*, 37: 2871–2878.

Koenderink, J. J. and van Doorn, A. J. (1976). Geometry of binocular vision and a model for stereopsis. *Biol. Cybern.*, 21: 29–35.

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.

Leventhal, A. G. (1983). Relationship between preferred orientation and receptive field position of neurons in cat striate cortex. *J. Comp. Neurol.*, 220: 476–483.

Liu, L., Stevenson, S. B., and Schor, C. W. (1994). A polar coordinate system for describing binocular disparity. *Vis. Res.*, 34: 1205–1222.

Livingstone, M. S. and Tsao, D. Y. (1999). Receptive fields of disparity-selective neurons in macaque striate cortex. *Nature Neurosci.*, 2: 825–832.

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

Marcělja, S. (1980). Mathematical description of the responses of simple cortical cells. *J. Opt. Soc. Am. A*, 70: 1297–1300.

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

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

Marr, D. and Poggio, T. (1979). A computational theory of human stereo vision. *Proc. R. Soc. Lond. B*, 204: 301–328.

Maske, R., Yamane, S., and Bishop, P. O. (1984). Binocular simple cells for local stereopsis: comparison of receptive field organizations for the two eyes. *Vis. Res.*, 24: 1921–1929.

Matthews, N., Meng, X., Xu, P., and Qian, N. (2003). A physiological theory of depth perception from vertical disparity. *Vis. Res.*, 43: 85–99.

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.

Mayhew, J. E. W. (1982). The interpretation of stereo-disparity information: the computation of surface orientation and depth. *Perception*, 11: 387–403.

Mayhew, J. E. W. and Longuet-Higgins, H. C. (1982). A computational model of binocular depth perception. *Nature*, 297: 376–379.

McKee, S. P. and Levi, D. M. (1987). Dichoptic hyperacuity: the precision of nonius alignment. *Vis. Res.*, 4: 1104–1108.

McLean, J. and Palmer, L. A. (1989). Contribution of linear spatiotemporal receptive field structure to velocity selectivity of simple cells in area 17 of cat. *Vis. Res.*, 29: 675–679.

Menz, M. D. and Freeman, R. D. (2003). Stereoscopic depth processing in the visual cortex: a coarse-to-fine mechanism. *Nature Neurosci.*, 6: 59–65.

Mikaelian, S. and Qian, N. (2000). A physiologically-based explanation of disparity attraction and repulsion. *Vis. Res.*, 40: 2999–3016.

Morgan, M. J. (1979). Perception of continuity in stereoscopic motion: a temporal frequency analysis. *Vis. Res.*, 19: 491–500.

Morgan, M. J. and Fahle, M. (2000). Motion–stereo mechanisms sensitive to inter-ocular phase. *Vis. Res.*, 40: 1667–1675.

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

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

Neri, P., Bridge, H., and Heeger, D. J. (2004). Stereoscopic processing of absolute and relative disparity in human visual cortex. *J. Neurophysiol.*, 92: 1880–1891.

Nikara, T., Bishop, P. O., and Pettigrew, J. D. (1968). Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex. *Exp. Brain Res.*, 6: 353–372.

Ogle, K. N. (1950). *Researches in Binocular Vision*. Philadelphia, PA: W. B. Saunders.

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.

Ohzawa, I., DeAngelis, G. C., and Freeman, R. D. (1997). Encoding of binocular disparity by complex cells in the cat's visual cortex. *J. Neurophysiol.*, 77: 2879–2909.

Poggio, G. F. and Fischer, B. (1977). Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *J. Neurophysiol.*, 40: 1392–1405.

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. *Vis. Res.*, 25: 397–406.

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

Qian, N. (1997). Binocular disparity and the perception of depth. *Neuron*, 18: 359–368.

Qian, N. and Andersen, R. A. (1994). Transparent motion perception as detection of unbalanced motion signals II: physiology. *J. Neurosci.*, 14: 7367–7380.

Qian, N. and Andersen, R. A. (1997). A physiological model for motion-stereo integration and a unified explanation of Pulfrich-like phenomena. *Vis. Res.*, 37: 1683–1698.

Qian, N. and Freeman, R. D. (2009). Pulfrich phenomena are coded effectively by a joint motion–disparity process. *J. Vis.*, 9: 1–16.

Qian, N. and Mikaelian, S. (2000). Relationship between phase and energy methods for disparity computation. *Neural Comput.*, 12: 279–292.

Qian, N. and Zhu, Y. (1997). Physiological computation of binocular disparity. *Vis. Res.*, 37: 1811–1827.

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.

Rashbass, C. and Westheimer, G. (1961). Disjunctive eye movements. *J. Physiol.*, 159: 339–360.

Read, J. C. A. and Cumming, B. G. (2005a). All Pulfrich-like illusions can be explained without joint encoding of motion and disparity. *J. Vis.*, 5: 901–927.

Read, J. C. A. and Cumming, B. G. (2005b). The stroboscopic Pulfrich effect is not evidence for the joint encoding of motion and depth. *J. Vis.*, 5: 417–434.

Read, J. C. A. and Cumming, B. G. (2007). Sensors for impossible stimuli may solve the stereo correspondence problem. *Nature Neurosci.*, 10: 1322–1328.

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

Rogers, B. J. and Bradshaw, M. F. (1993). Vertical disparities, differential perspectives and binocular stereopsis. *Nature*, 361: 253–255.

Rogers, B. J. and Koenderink, J. (1986). Monocular aniseikonia: a motion parallax analogue of the disparity-induced effect. *Nature*, 322: 62–63.

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

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 response modulation. *Vis. 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.

Smallman, H. S. and McKee, S. P. (1995). A contrast ratio constraint on stereo matching. *Proc. R. Soc. Lond. B*, 260: 265–271.

Snowden, R. J., Treue, S., Erickson, R. E., and Andersen, R. A. (1991). The response of area MT and V1 neurons to transparent motion. *J. Neurosci.*, 11: 2768–2785.

Sobel, E. C. and Collett, T. S. (1991). Does vertical disparity scale the perception of stereoscopic depth? *Proc. R. Soc. Lond. B*, 244: 87–90.

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

Umeda, K., Tanabe, S., and Fujita, I. (2007). Representation of stereoscopic depth based on relative disparity in macaque area V4. *J. Neurophysiol.*, 98: 241–252.

Vidyasagar, T. R. and Henry, G. H. (1990). Relationship between preferred orientation and ordinal position in neurons of cat striate cortex. *Vis. Neurosci.*, 5: 565–569.

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

Watson, A. B. and Ahumada, A. J. (1985). Model of human visual-motion sensing. *J. Opt. Soc. Am. A*, 2: 322–342.

Westheimer, G. (1979). Cooperative neural processes involved in stereoscopic acuity. *Exp. Brain Res.*, 36: 585–597.

Westheimer, G. (1984). Sensitivity for vertical retinal image differences. *Nature*, 307: 632–634.

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.

Westheimer, G. and Levi, D. M. (1987). Depth attraction and repulsion of disparate foveal stimuli. *Vis. Res.*, 27: 1361–1368.

Westheimer, G. and Pettet, M. W. (1992). Detection and processing of vertical disparity by the human observer. *Proc. R. Soc. Lond. B*, 250: 243–247.

Wist, E. R., Brandt, T., Diener, H. C., and Dichgans, J. (1977). Spatial frequency effect on the Pulfrich stereophenomenon. *Vis. Res.*, 17: 371–397.

Zhu, Y. and Qian, N. (1996). Binocular receptive fields, disparity tuning, and characteristic disparity. *Neural Comput.*, 8: 1611–1641.