



A physiological theory of depth perception from vertical disparity

Nestor Matthews¹, Xin Meng, Peng Xu², Ning Qian^{*}

*Department of Physiology and Cellular Biophysics, Center for Neurobiology and Behavior, Columbia University, P.I. Annex Room 730,
722 W. 168th Street, New York, NY 10032, USA*

Received 24 July 2001; received in revised form 25 July 2002

Abstract

It has been known since the time of Helmholtz that vertical differences between the two retinal images can generate depth perception. Although many ecologically and geometrically inspired theories have been proposed, the neural mechanisms underlying the phenomenon remain elusive. Here we propose a new theory for depth perception from vertical disparity based on the oriented binocular receptive fields of visual cortical cells and on the radial bias of the preferred-orientation distribution in the cortex. The theory suggests that oriented cells may treat a vertical disparity as a weaker, equivalent horizontal disparity. It explains the induced effect, and the quadrant and size dependence of vertical disparity. It predicts that horizontal and vertical disparities should locally enhance or cancel each other according to their depth signs, and that the effect of vertical disparity should be orientation dependent. These predictions were confirmed through psychophysical experiments.

© 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Binocular disparity; Stereopsis; Disparity energy model; Orientation dependence; Binocular cells

1. Introduction

Binocular disparity is defined as the positional difference between the two retinal projections of a given point in space. Since the retinal surface is two-dimensional (2D), disparity is generally a 2D vector with a pair of independent components. Due to the lateral separation of the eyes, disparity is mainly along the horizontal dimension defined by the interocular axis. The vertical component of disparity is usually much smaller—except at large retinal eccentricities (Howard & Rogers, 1995). It is well established that the brain uses horizontal disparity to estimate the relative depths of objects with respect to the fixation point. It has also been known that vertical disparity can generate binocular depth perception as well (Howard & Rogers, 1995), although the mechanism involved is much more controversial.

The best-known example of depth perception from vertical disparity is the so-called induced effect (Ogle, 1950): A stereogram made of two identical images but with one of them slightly magnified vertically (Fig. 1a) is perceived as a slanted surface rotated about a vertical axis (Fig. 1b). The surface appears further away on the side with the smaller image, and the apparent axis of rotation is the vertical meridian going through the point of fixation (Ogle, 1950; Westheimer & Pettet, 1992). To better appreciate this phenomenon, we indicate in Fig. 2a the depth and disparity signs in the four quadrants around the point of fixation. Assuming a left-image magnification, the features on 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 the opposite vertical disparity signs in the first and fourth quadrants generate the same depth sign (far), and that the same vertical disparity sign in the first and second quadrants generates opposite depth signs (far and near, respectively). That is, the depth sign of a given vertical disparity depends on the quadrants around the fixation point (Westheimer & Pettet, 1992). To generate the same kind of surface slant with horizontal disparity (termed

^{*} Corresponding author. Tel.: +1-212-543-5213; fax: +1-212-543-5161.

E-mail address: nq6@columbia.edu (N. Qian).

URL: <http://brahms.cpmc.columbia.edu>.

¹ Present address: Department of Psychology, Denison University, Granville, OH 43023, USA.

² Present address: SONY Electronics Inc., 3300 Zanker Rd, SJ3C3, San Jose, CA 95134, USA.

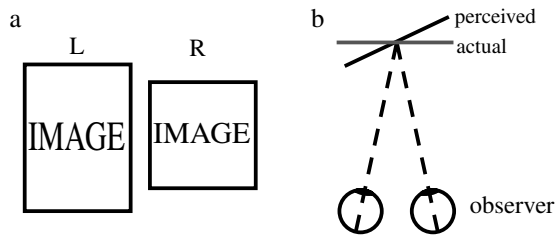


Fig. 1. (a) A schematic stereogram for the induced effect. The left eye's view (L) is magnified vertically with respect to the right (R). (b) With a stereogram like that in (a), a slanted surface is perceived, shown schematically here in the top view, as if the right image had been magnified horizontally.

the geometric effect by Ogle (1950)), one will have to magnify the *right* image horizontally. Unlike vertical disparity, the depth sign of horizontal disparity is fixed and independent of quadrants (Fig. 2b).

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 vertical disparity is larger at greater eccentricities and does not have a consistent local depth sign, and since the effect of vertical disparity can be best demonstrated with large stimuli (Howard & Kaneko, 1994; Rogers & Bradshaw, 1993) and appears to be averaged over greater areas than horizontal disparity (Kaneko & Howard, 1997), it is believed that vertical disparity acts globally while horizontal disparity acts locally. Numerous theories of vertical disparity have been proposed (Arditi, Kaufman, & Movshon, 1981; Backus, Banks, van Ee, & Crowell, 1999; Banks & Backus, 1998; Garding, Porrill, Mayhew, & Frisby, 1995; Gillam & Lawergren, 1983; Howard & Kaneko, 1994; Koenderink & van Doorn, 1976; Liu, Stevenson, & Schor, 1994; Mayhew & Longuet-Higgins, 1982; Ogle, 1950; Rogers & Bradshaw, 1993); many of them employ some form of global assumption to explain the induced

effect. For example, an elegant theory for the interpretation of vertical disparity has been proposed by Mayhew (1982) and Mayhew and Longuet-Higgins (1982). According to this theory, 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 gaze angle. Since horizontal disparity is dependent on these parameters, the estimated parameters will modify horizontal disparity globally, and hence the 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, Johnston, & Parker, 1991) to 30° (Sobel & 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 & 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 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 & Koenderink, 1986).

A general problem applicable to all purely global interpretations of vertical disparity, is that vertical disparity *can* generate reliable (albeit relatively weak) local depths even in small displays that are viewed foveally (Westheimer, 1984; Westheimer & Pettet, 1992). Fig. 3 provides a demonstration of this effect. One might argue that functionally, the depth effect of vertical disparity in small displays is not as important as the induced effect in large stimuli because vertical disparity is usually negligible near the fovea, while full-field vertical size differences between the eyes can occur naturally with

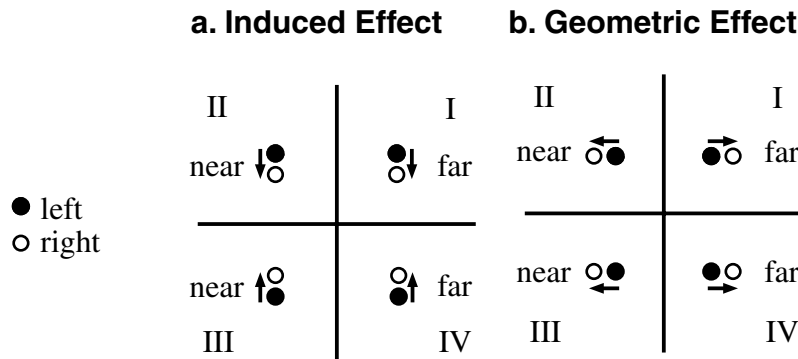


Fig. 2. The disparity and depth signs for (a) the induced effect (vertical disparity) and (b) the geometric effect (horizontal disparity). For clarity, the left and right image features are represented schematically by filled and open dots, respectively. In each panel, the fixation point is at the center of the cross which divides the screen into four quadrants. The arrows indicate the signs of 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 perceived depth sign (near or far) in each quadrant is also marked. Note that the depth sign of vertical disparity is quadrant dependent, while that of horizontal disparity is not.

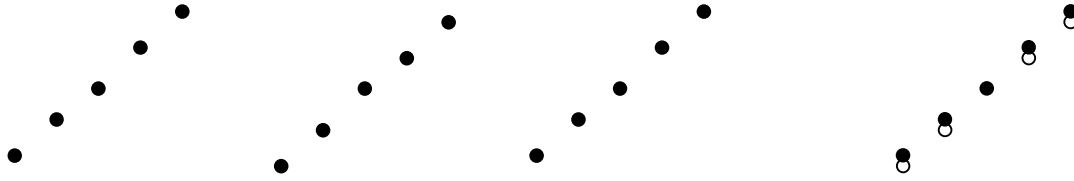


Fig. 3. A stereogram demonstrating the depth effect of vertical disparity. The first and the third images are identical, and can be paired with the second image for uncrossed and crossed fusion, respectively. The right-most column shows the superposition of the monocular images. The central dot of the 5-dot stereogram has zero disparity. The four surrounding dots all have the same vertical disparity, and appear to be slightly further away than the central dot. Since only relatively small vertical disparities can be well fused, readers may need to hold the figure at a sufficiently large distance to see the effect.

eccentric gaze. However, as pointed out recently by Farell (1998), 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 Fig. 4a. When the apertures are narrow enough, the horizontal disparity will be largely eliminated and subjects will have to rely on vertical disparity to make local depth judgments. This is analogous to the situation where an interocular time delay is generated by a target moving behind apertures (Fig. 4b), which has been suggested as a real-world realization of the Pulfrich effect (Burr & Ross, 1979; Morgan & Thompson, 1975). Without the occluders, the moving target has a horizontal disparity with respect to the fixation. With the occluders, the target appears to be at the same aperture locations to the two eyes but at slightly different times.

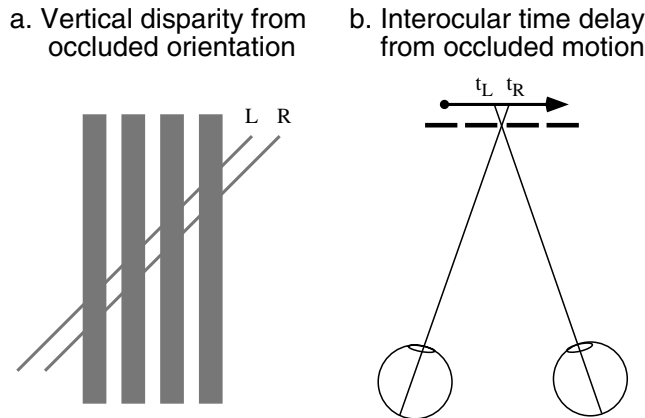


Fig. 4. (a) An illustration of how vertical disparity can arise from horizontal disparity carried by oriented contours. The vertical occluders have zero disparity while the diagonal line has a far horizontal disparity between its left (L) and right (R) images. The visible segments between the occluders have disparities mainly along 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). 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. For example, the dot appears at the central aperture at times t_L and t_R , respectively.

Another general problem with most existing theories is that they are only about the *interpretation* of vertical disparity, and do not address the issue of how vertical disparities are *extracted* from stereograms by visual cortical cells in the first place. As such, the theories are limited to ecological or geometrical considerations and shed little light on the underlying physiological mechanisms.

In this paper, we propose a new theory for depth perception from vertical disparity based on the oriented binocular receptive fields (RFs) of visual cortical cells (Anzai, Ohzawa, & Freeman, 1999a,b; DeAngelis, Ohzawa, & Freeman, 1991; Ohzawa, DeAngelis, & Freeman, 1990, 1996, 1997) and on the radial bias of the preferred-orientation distribution in the cortex (Bauer & Dow, 1989; Bauer, Dow, Synder, & Vautin, 1983; Leventhal, 1983; Vidyasagar & Henry, 1990). The theory naturally integrates the measurement and interpretation of vertical disparity, explains the induced effect and the local depth effects of vertical disparity, and suggests a unified framework for understanding the relationship between vertical and horizontal disparities. We have also confirmed psychophysically two key predictions of the theory. Some preliminary results were reported previously in abstract form (Xu, Matthews, & Qian, 2000).

2. Theory

Our theory is based on the typical 2D spatial RFs of visual cortical cells. According to quantitative physiological studies (Anzai et al., 1999a; DeAngelis et al., 1991; Ohzawa et al., 1990, 1996), the left and right RFs of a binocular simple cell can be written as

$$f_l(x, y) = g(x, y) \cos(\omega_x x + \omega_y y + \phi_l) \quad (1)$$

$$f_r(x, y) = g(x, y) \cos(\omega_x x + \omega_y y + \phi_r) \quad (2)$$

where $g(x, y)$ is a 2D Gaussian function that determines the RF envelopes, ω_x and ω_y are the preferred horizontal and vertical spatial frequencies of the cell, and ϕ_l and ϕ_r are the phase parameters for left and right RFs, respectively. The preferred orientation (θ) of the cell is

parallel to the contours defined by $\omega_x x + \omega_y y = \text{constant}$; therefore

$$\tan \theta = -\omega_x / \omega_y \quad (3)$$

where θ is measured counterclockwise from the positive x axis.

The physiological data also indicate that the response of a simple cell can be modeled by binocular linear filtering followed by half-squaring, and the response of a complex cell can be simulated by combining simple cell responses in the framework of the disparity energy model (Anzai et al., 1999b; Chen, Wang, & Qian, 2001; DeAngelis et al., 1991; Ohzawa et al., 1990, 1997; Qian, 1994). Now consider a subset of complex cells all with the same ω_x and ω_y (and thus the same preferred orientation θ) but with the phase-parameter difference $\phi_- \equiv \phi_r - \phi_l$ covering the full 2π range. It can be shown (see Appendix A) that a stereogram with horizontal disparity H and vertical disparity V will maximally excite the cell whose ϕ_- is given by

$$\hat{\phi}_- \approx \omega_x H + \omega_y V \equiv \vec{\omega} \cdot \vec{D} \quad (4)$$

where $\vec{\omega} \equiv (\omega_x, \omega_y)$ and $\vec{D} \equiv (H, V)$ are the frequency and disparity vectors, respectively. The $\vec{\omega}$ so defined is perpendicular to the cell's preferred orientation θ because the dot product between (ω_x, ω_y) and $(1, \tan \theta)$ is zero.

For the special case of no vertical disparity (i.e., $V = 0$), Eq. (4) becomes $\hat{\phi}_- \approx \omega_x H$, and horizontal disparity H can be estimated from the parameters of the most responsive cell according to $H \approx \hat{\phi}_- / \omega_x$, or through weighted population averaging, as we proposed previously (Qian, 1994, 1997; Qian & Zhu, 1997). For the general case, first note that a family of cells all with the same preferred orientation θ (therefore the same $\vec{\omega}$ direction) can only estimate the component of \vec{D} perpendicular to θ (parallel to $\vec{\omega}$) because of the dot product in Eq. (4). This is the disparity aperture problem, analogous to the same problem in motion, and it has been discussed recently in the psychophysical literature (Farell, 1998; Morgan & Castet, 1997). Second, for stimuli containing a single orientation, all highly activated cells are tuned to that orientation and their $\vec{\omega}$ vectors must all point in the same perpendicular direction. Therefore, the aperture problem cannot be solved under this degenerate condition, and only the disparity component perpendicular to the stimulus orientation, D^\perp (or its equivalent along the horizontal axis, h), can be computed according to Eq. (4) (see Fig. 5). The actual horizontal (H) and vertical (V) components of \vec{D} cannot be recovered in this case. This result is consistent with the psychophysical finding of Morgan and Castet (1997). Finally, for stimuli containing more than one orientation, both horizontal and vertical disparity components could be estimated by using two or more populations of cells with different preferred orienta-

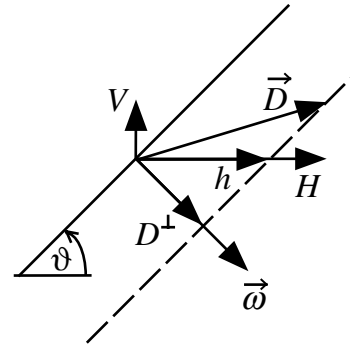


Fig. 5. The aperture problem in stereo vision for a population of cells, all with the same preferred orientation. The solid line with an angle θ from the horizontal axis represents the preferred orientation, and the $\vec{\omega}$ vector points in the perpendicular direction. For a given disparity vector \vec{D} , these cells can only detect the component D^\perp perpendicular to the preferred orientation of the cells according to $D^\perp = \hat{\phi}_- / \|\vec{\omega}\|$, or its equivalent along the horizontal axis $h = D^\perp / \sin \theta$. The cells cannot distinguish those \vec{D} vectors ending on the dashed constraint line, and are therefore incapable of recovering the real horizontal (H) and vertical (V) components of \vec{D} .

tions (and therefore different $\vec{\omega}$), similar to the various methods for solving the motion aperture problem (Adelson & Movshon, 1982; Heeger, 1987; Hildreth, 1984; Horn & Schunck, 1981; Simoncelli & Heeger, 1998; Wilson, Ferrera, & Yo, 1992).

The estimated vertical as well as horizontal disparities could then be fed into any of the previous computational models for the perceptual interpretation. However, we would like to propose a new scheme for interpreting vertical disparity which renders the explicit extraction of disparity components unnecessary. Our interpretation comes naturally by rewriting Eq. (4) as

$$\hat{\phi}_- / \omega_x \approx H + (\omega_y / \omega_x) V \quad (5)$$

Since horizontal disparity H is usually much larger than vertical disparity V in the central visual field, the visual system may have evolved to simply use $\hat{\phi}_- / \omega_x$, or the related population averaging, for estimating H in all cases. Then, when there is a significant vertical disparity V in the stimulus, Eq. (5) implies that V will be “mistaken” as an equivalent horizontal disparity:

$$H_{\text{equiv}} = (\omega_y / \omega_x) V = -V / \tan \theta \quad (6)$$

by the cells with preferred horizontal and vertical spatial frequencies ω_x and ω_y (and preferred orientation θ). This result is analogous to our previous demonstration that spatiotemporally oriented cells may treat an interocular time delay as an equivalent horizontal disparity (Qian & Andersen, 1997).

How can Eq. (6) account for the perceived depth in stereograms containing vertical disparities? First consider the case without real horizontal disparity ($H = 0$) in the stimulus. According to Eq. (6), cells with preferred orientation θ would treat a vertical disparity V as an

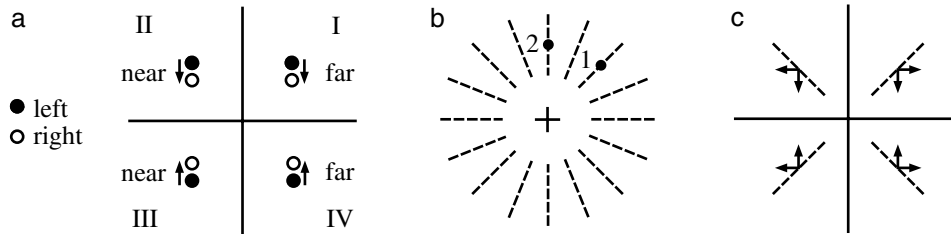


Fig. 6. Our explanation of the induced effect and the related quadrant dependence of vertical disparity. (a) The disparity and depth signs in the four quadrants around the fixation caused by a left-image magnification (same as Fig. 2a). The left and right image features are represented by filled and open dots, respectively. The signs of 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 (central cross) found in the visual cortex. For example, the 45° orientation and the vertical orientation are over-represented for spatial locations 1 and 2, respectively. (c) Conversion of vertical disparity into equivalent horizontal disparity by the over-represented 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 over-represented orientations (dashed lines) and Eq. (6).

equivalent horizontal disparity ($-V/\tan\theta$). For stimuli without dominant orientations, 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 magnitude, and the average result across all cells should be near zero. The only possibility of seeing depth from vertical disparity in stimuli *without* dominant orientations is that certain orientations are over-represented 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 stimulus-orientation, however, will not easily allow cortical cells to convert a vertical disparity into an equivalent horizontal disparity because vertically tuned cells have $\omega_y = 0$ in Eq. (6). Similarly, a near horizontal stimulus-orientation will not be effective either since the equivalent horizontal disparity will be too large (due to the vanishing ω_x) to be detected. Therefore, the best stimulus orientation for perceiving depth from vertical disparity should be around a diagonal axis, as in Fig. 3.

Let us define the disparity of a point as positive when its position in the right image minus the corresponding position in the left image is positive.³ With this defini-

tion, a positive and a negative horizontal disparity generate a depth perception of far and near, respectively. If the first and the second images in Fig. 3 are used for uncrossed fusion, then the outer four dots have a negative vertical disparity. Since the dominant stimulus orientation is 45°, and $\tan(45^\circ) = 1$, Eq. (6) indicates that the equivalent horizontal disparity for the vertical disparity of the outer four dots should be positive, in agreement with their “far” perception with respect to the central dot. The relative depth order of the central and the outer dots should not be affected by vertical vergence. For example, if the vertical vergence is such that the outer four dots now have zero disparity, then the central dot will have a positive vertical disparity. The equivalent horizontal disparity for the vertical disparity of the central dot will then be negative, in agreement with its “near” depth relative to the outer dots. If the dots are arranged along the 135° orientation instead, the depth sign of the same vertical disparity should be reversed because $\tan(135^\circ) = -1$ (see Experiment 2 below). From the functional perspective, the mechanism we proposed here should allow the visual system to perceive the correct depth relationship from vertical disparity in occlusion situations like Fig. 4a.

A critical test of our theory is whether it can explain the well-known induced effect (Ogle, 1950): A stereogram made of two identical images but with one of them slightly magnified vertically is perceived as a rotated surface about the vertical axis going through the point of fixation (Fig. 6a). First note that the induced effect can be observed in stimuli having no dominant orientations, such as random textures (Ogle, 1950). Therefore, according to the above discussion, a reliable equivalent horizontal disparity could only be generated by an over-representation of certain orientations in the brain. Remarkably, physiological experiments have suggested a radial bias of preferred orientations around the fixation in the cat primary visual cortex (Leventhal, 1983; Vidyasagar & Henry, 1990) and in the supra-granular layers of monkey area V1 (Bauer & Dow, 1989;

³ Here we mean external stereo images on a screen or paper. For retinal images, the disparity signs should be reversed.

Bauer et al., 1983). 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 (Fig. 6b). This is precisely what is needed for explaining the induced effect and the related quadrant dependence of vertical disparity for stimuli without a dominant orientation (Fig. 6c). Specifically, the signs of $\tan \theta$ for the over-represented orientations are positive in the first and third quadrants and negative in the second and fourth quadrants. Therefore, according to Eq. (6), vertical disparities of the same sign in the first and second quadrants should be treated as equivalent horizontal disparities of opposite signs, and vertical disparities of opposite signs in the first and fourth quadrants should be treated as equivalent horizontal disparities of the same sign.

Quantitatively, let the fixation point be the origin and assume the left image is magnified vertically by a factor of k (>1). Then, the vertical disparity (defined as the right position minus the corresponding left position) at the stimulus location (x, y) is $V(x, y) = (1 - k)y$. The radial bias means that the cortically over-represented orientation for location (x, y) is given by $\tan \theta = y/x$. Then, according to Eq. (6), the corresponding equivalent horizontal disparity should be

$$H_{\text{equiv}}(x, y) = -(1 - k)y / \tan \theta = (k - 1)x \quad (7)$$

Therefore, although the vertical magnification of the left image by a factor of k creates a vertical disparity of $(1 - k)y$ at 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, consistent with the psychophysical observations (Ogle, 1950). We predict that the radial bias of the preferred orientations also holds for the human visual cortex. Note that the radial bias does not affect the perceived depth from real *horizontal* disparity, since unlike vertical disparity, horizontal disparity is not subject to an orientation-dependent conversion.

The quadrant dependence of vertical disparity means that vertical disparity does not have a consistent local depth sign. This may seem to imply that the induced effect could only be explained 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 equivalent horizontal disparity. Our theory is consistent with the finding that vertical disparity is more effective at larger display sizes (Howard & Kaneko, 1994; Rogers & Bradshaw, 1993) and with the related observation that vertical disparity appears to operate at

a more global scale than does horizontal disparity (Kaneko & Howard, 1997). This is because the radial bias of cells' preferred orientations was found to be stronger at larger eccentricities (Leventhal, 1983) although the bias is also present for foveal cells in monkey area V1 (Bauer & Dow, 1989; Bauer et al., 1983). Larger displays cover more eccentric locations, and are therefore more effective. Besides, cells tuned to more eccentric locations have bigger RFs that can detect larger vertical disparities normally present in the periphery. In addition, since vertical disparity is weaker than horizontal disparity over a local region, vertical disparity is more likely to benefit from spatial summation over a larger display size.

We finally consider the case with both horizontal and vertical disparities in the stimuli. Eq. (5) predicts that these two types of disparities should locally enhance or cancel each other depending on their depth signs. This is a new prediction because most previous theories are global in nature and are mainly concerned with the slant perception of an entire surface. Those theories only predict that *overall* horizontal and vertical magnifications can enhance or cancel each other in generating the perceived surface slant. Our theory makes this prediction too (see Eqs. (5)–(7)), but in addition, it predicts that *local* horizontal and vertical disparities (instead of overall magnifications) can also enhance or cancel each other in generating regional depth structure (other than slant) when vertical disparity is made effective through a near-diagonal stimulus orientation. A previous experiment failed to find such local interactions between the two types of disparities (Westheimer, 1978) presumably because small, I-shaped patterns containing only vertical and horizontal orientations were used. We therefore tested the predicted enhancement and cancellation using diagonally oriented stimuli.

A demonstration consistent with the above prediction is shown in Fig. 7. The central dot of each of the three 5-dot stereograms has zero disparity. The four outer dots in the top stereogram have both a far horizontal disparity and a far vertical disparity, while those in the bottom stereogram have a far horizontal disparity but a near vertical disparity. Since vertical disparity is not as effective as horizontal disparity at small display sizes, the vertical disparity is four times as large as the horizontal disparity in the figure. As predicted, the relative depth between the four outer dots and the central dot is the largest in the top stereogram, but smallest in the bottom stereogram. The four outer dots in the middle stereogram have a far horizontal disparity only. It serves as a reference, generating an intermediate amount of relative depth.

In the following sections, we report two psychophysical experiments that we performed using similar patterns.

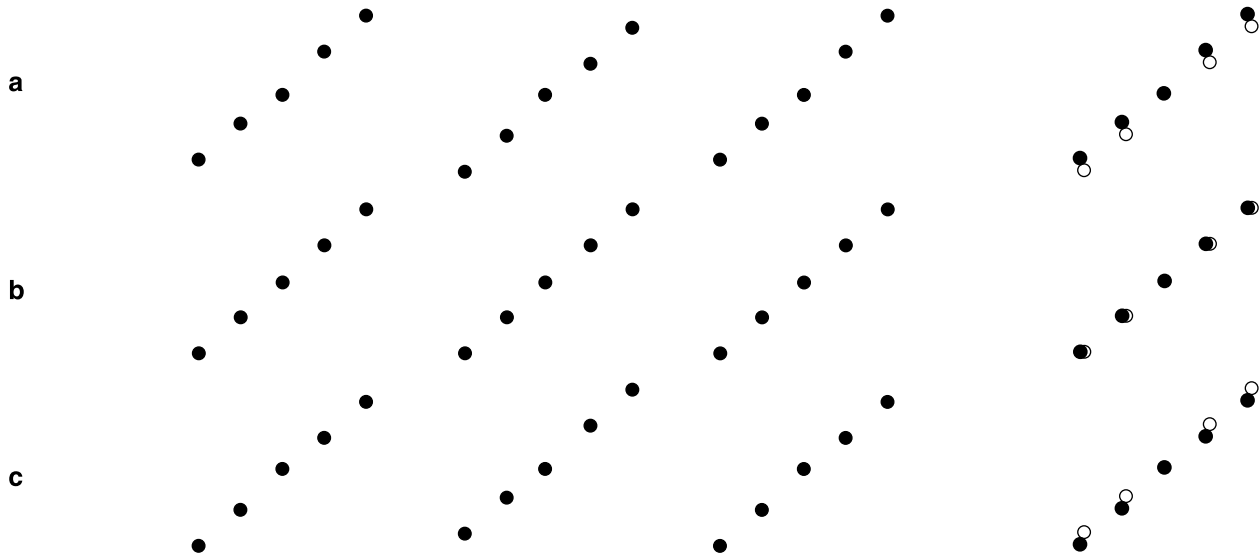


Fig. 7. Three stereograms demonstrating enhancement and cancellation between horizontal and vertical disparities in small displays. As in Fig. 3, the first and the third images are identical, and can be paired with the second images for uncrossed and crossed fusion, respectively. The right-most column shows the superposition of the monocular images. The central dot of each 5-dot stereogram has zero disparity. The four outer dots all have the same horizontal disparity, but different vertical disparities in different stereograms. Specifically, the middle stereogram (b) has zero vertical disparity, while the top (a) and bottom (c) stereograms have vertical disparity of the same magnitude (4 times the horizontal disparity magnitude) but opposite signs. The two types of disparities in the top stereogram enhance each other while those in the bottom cancel each other so that the relative depth between the central dot and the four outer dots is greatest in the top stereogram and smallest in the bottom stereogram. Since only relatively small vertical disparities can be well fused, readers may need to hold the figure at a sufficiently large distance to see the effect.

3. Experiment 1

In this experiment, we tested the predicted enhancement and cancellation between horizontal and vertical disparities using stimuli with a 45° orientation.

3.1. Methods

3.1.1. Apparatus

The experiment was conducted on a 21 in. (53.3 cm) ViewSonic PT810 monitor that was controlled by a Macintosh computer and software from the psychophysics toolbox (Brainard, 1997; Pelli, 1997). The vertical refresh rate of the monitor was 120 Hz, and the spatial resolution was 1024×768 pixels. In a well-lit room, subjects foveally viewed the stimuli through a mirror stereoscope. A chin rest helped to stabilize head position at 64.5 cm.

3.1.2. Subjects

Four subjects, two of whom were naive observers (YC and MX), participated in this experiment. (MX was not the coauthor XM.) All had normal or corrected-to-normal acuity, and could resolve disparities of $30''$ (Randot stereo test).

3.1.3. Stimuli

The stimuli were 5-dot stereograms similar to those in Fig. 7. The central dot had zero disparity and was used

as the fixation point. The four outer dots had various combinations of horizontal and vertical disparities (see below). Eye vergence state was stabilized by both horizontal and vertical nonius lines (McKee & Levi, 1987), which were present on the screen at all times. Generated by our anti-aliasing program, the stimuli and the nonius lines appeared black (1.23 cd/m^2) within a white (104 cd/m^2) surround. The Michelson contrast was 97.7%. Each dot had a 0.3° diameter, and the center-to-center separation between adjacent dots was 0.85° . The horizontal (vertical) nonius line was vertically (horizontally) aligned with the fixation dot. Each nonius line was 2.4° long and $1.8'$ wide, and its nearest end was separated from the fixation dot by 2° .

3.1.4. Experimental procedure

On each trial, two stereograms were presented simultaneously one above the other. One of the stimuli, which we call the 1D-disparity pattern, was similar to Fig. 7b and contained only a horizontal disparity of either $1.4'$, $1.8'$ or $2.2'$ for the four outer dots. For the other stimulus, which we call the 2D-disparity pattern (similar to Fig. 7a or c), the four outer dots always contained a horizontal disparity of $1.8'$, but also one of the seven vertical disparities from $-5.4'$ to $5.4'$ in $1.8'$ increments. The spatial order of the 1D and 2D patterns in a trial was randomized. The observers' task was to indicate which of the two patterns on each trial appeared to have greater relative depth, i.e., a larger depth difference between the

central dot and the four outer dots. Subjects were instructed to switch fixation between the central dots of the two stimuli as often as desired before making their judgment. No feedback was given at any time.

There were a total of 21 stimulus conditions (seven 2D patterns paired with three 1D patterns). A trial block comprised 2 sets of the 21 conditions in random order. Each observer completed 10 such blocks, i.e., 20 trials per stimulus condition. For each observer, the effect of vertical disparity was assessed by three psychometric curves, one for each of the three horizontal disparities in the 1D patterns. Each psychometric curve plotted the fraction of trials that greater relative depth had been seen in the 2D-disparity pattern as a function of vertical disparity. A least-squares procedure was then used to fit the data with a sigmoid of the form $1/(1 + \exp[-K(V - V_0)])$, where K and V_0 determine the slope and midpoint of the curve respectively. We tested whether the fit was significant (i.e., whether the depth perception varied reliably with vertical disparity) using the Pearson correlation coefficient (r) between the data and the fitted curve as the statistic (5 degrees of freedom, i.e., 7 data points minus 2 estimated parameters (slope and midpoint)).

3.2. Results

The results are shown in Fig. 8a where the fraction of trials on which the 2D-disparity patterns were judged to have greater relative depth than the 1D-disparity pattern was plotted against the vertical disparity in the 2D-disparity patterns. The dashed, solid, and dotted curves are for the horizontal disparity of the 1D patterns equal to 1.4', 1.8' and 2.2', respectively. If the perceived depth were determined by horizontal disparity alone (Westheimer, 1978), all the curves would be flat since the horizontal disparities in the 1D-disparity and 2D-disparity patterns being compared were fixed across vertical disparities (and were identical for the solid curve). Instead, the fraction of trials on which greater depth was seen in the 2D-disparity patterns increased systematically as the vertical disparity in these patterns changed from negative to positive, suggesting that the perceived depth depended on the addition or subtraction between the two types of disparities. The observed systematic shifts among the three curves are also consistent with a combined contribution from horizontal and vertical disparities. Each curve in Fig. 8a was fitted significantly ($p < 0.01$) with a sigmoidal function.

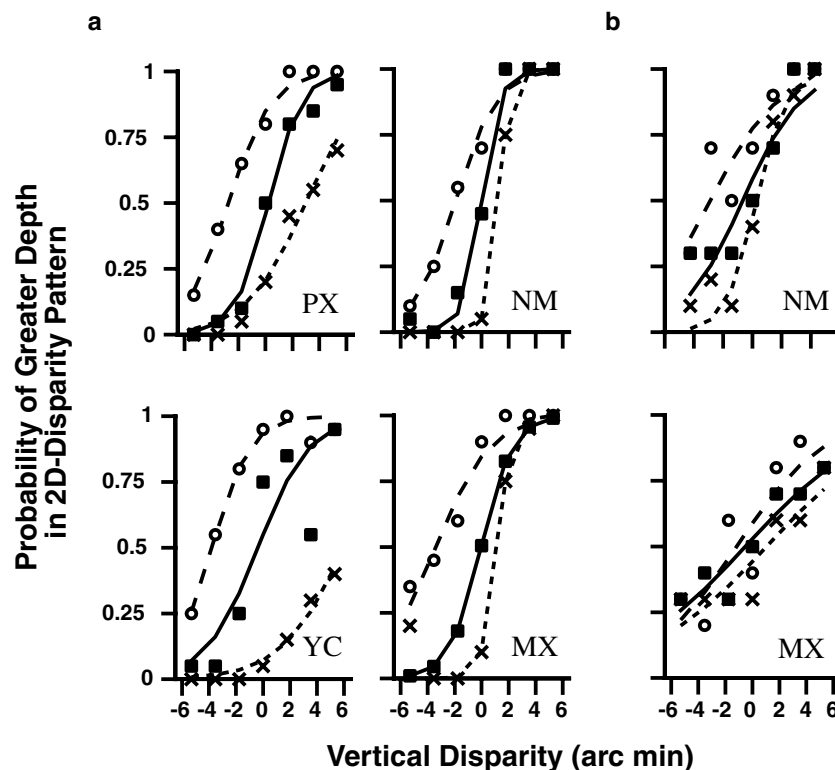


Fig. 8. Psychophysical data of enhancement and cancellation between horizontal and vertical disparities. The fraction of trials in which the 2D-disparity pattern was perceived as having more relative depth than the 1D-disparity pattern is plotted as a function of the vertical disparity component in the 2D-disparity patterns. The three curves in each panel correspond to horizontal disparities of 1.4' (dashed), 1.8' (solid), and 2.2' (dotted) in the 1D-disparity pattern, respectively. There were 20 trials for each datum. (a) Results from the unlimited viewing condition for four subjects. (b) Results from brief (200 ms) viewing condition for two subjects who could perform this difficult task. See text for further details.

To rule out the possible role of eye movements in these results, we repeated the experiment with the same stimuli but with a stimulus duration of only 200 ms (McKee, Levi, & Bowne, 1990; McKee, Welch, Taylor, & Bowne, 1990). The brief stimulus duration made it necessary to present the two patterns in each trial sequentially. To help each observer fuse the stimuli, the central dot (at zero disparity for all stimuli) as well as the nonius lines were present all the time. After fusing the central dot, the observer initiated each trial by pulling a trigger to present the four outer dots of the first stimulus for 200 ms. During a subsequent interval of 600 ms, the observer maintained fusion of the central dot. The observer then pulled the trigger again to present the four outer dots of the second stimulus for 200 ms. The observer indicated at the end of each trial whether the first or second pattern appeared to have greater relative depth.

Due to the difficulty associated with such brief presentation (McKee, Levi, et al., 1990), only one of the authors and one of the naive observers could perform this task. The results are shown in Fig. 8b. Again non-flat curves were obtained, and each curve was fitted significantly ($p < 0.01$) with a sigmoidal function. The main differences between Fig. 8a and b are: (1) in Fig. 8b, the three curves from each subject are closer to each other, indicating that it is harder for the subjects to differentiate the three horizontal disparities in the stimulus with the brief presentation time; and (2) the slopes of the curves in Fig. 8b are shallower, indicating that it is also harder for the subjects to discriminate the seven vertical disparities in the stimulus with the brief presentation time. These degradations are certainly expected when the task becomes harder (McKee, Welch, et al., 1990). Nevertheless, the key feature that the perceived depth changes monotonically with vertical disparity is preserved. If vertical disparity did not contribute to the local depth structure, or if it only contributed in a non-specific way (e.g., by making horizontal disparity harder to perceive), one would not obtain the observed monotonic curves.

The enhancement and cancellation between the two types of disparities demonstrated here are analogous to the similar interactions found between interocular time delay and horizontal disparity in the Pulfrich effect (Burr & Ross, 1979).

4. Experiment 2

In this experiment, we confirmed the results in Experiment 1 with a different procedure. More importantly, we tested another key prediction of the model, namely the orientation dependence of the vertical disparity effect (Eq. (6)). A factorial design with seven

vertical disparities, three horizontal disparities and four orientations was used.

4.1. Methods

4.1.1. Apparatus

The apparatus was the same as in Experiment 1 except that a 21 in. ViewSonic P817 monitor was used and the viewing distance was changed to 80 cm.

4.1.2. Subjects

The participants were two of the authors, and a third subject (LD) who was a psychophysicist in the same lab but was not informed about the specific purpose or design of this experiment. All had normal or corrected-to-normal acuity, and could resolve disparities of 30" (Randot stereo test).

4.1.3. Stimuli

The stimuli were the 5-dot patterns that were used in Experiment 1, but with the following exceptions. First, all stimuli were presented for just 200 ms. Second, the seven vertical disparities ranged from $-9.6'$ to $+9.6'$ in $3.2'$ steps, and the three horizontal disparities were $-0.8'$, 0 , $+0.8'$. Much larger disparity steps were used here in order to have a better separation of the curves than in Fig. 8b. Third, the 5-dot stimuli were presented at four different orientations of 0° , 45° , 90° , or 135° . Finally, the dot size, center-to-center distance, and nonius line dimensions were 90% of the values in Experiment 1 due to the distance and monitor change.

4.1.4. Experimental procedure

To help each observer fuse the stimuli, the nonius lines and the central dot (at zero disparity for all stimuli) were present all the time. After fusing the central dot, the observer initiated each trial by pulling a trigger to present the four outer dots of the stimulus for 200 ms. Unlike Experiment 1, here only a single 5-dot pattern was presented on each trial. The observer's task was to indicate whether the four outer dots appeared "nearer" or "farther" in depth than the central dot. The experimental design was $4 \times 3 \times 7$: Four orientations by three horizontal disparities by seven vertical disparities. A single trial block comprised two sets of these 84 stimulus conditions in random order, and each subject completed 12 blocks (24 trials per stimulus condition). Due to the difficulty associated with the brief presentation time and the relatively weak depth effect of vertical disparity, all three subjects went through a few sessions of practice before data collection. No feedback was given at any time.

As in Experiment 1, the data were analyzed by constructing psychometric curves. Specifically, for each observer, orientation, and horizontal disparity, we plotted the proportion of nearer responses as a function

of vertical disparity, and used a least-squares procedure to fit the data with a sigmoid. As before, we used the significance of the fit as a measure of whether the vertical disparity had a depth effect. Changes in the significance of the fit and in the sign of the slope across the orientation conditions marked the orientation dependence of the vertical-disparity effect.

4.2. Results

The results are shown in Fig. 9, where the fraction of trials in which the outer four dots were perceived as nearer than the central dot is plotted as a function of the

vertical disparity. The data from different stimulus orientations and subjects are represented in different columns and rows, respectively. The three curves in each panel correspond to horizontal disparities of $-0.8'$, 0 and $0.8'$, respectively. The results in the first column resemble those in Fig. 8. The main difference is that the three curves in each panel are more separated here due to the larger horizontal disparity step in this experiment. Another difference is that in Experiment 1, the three horizontal disparities were all positive while here they had negative, zero and positive values. The systematic dependence of the curves on the vertical and horizontal disparities again demonstrates the enhancement and

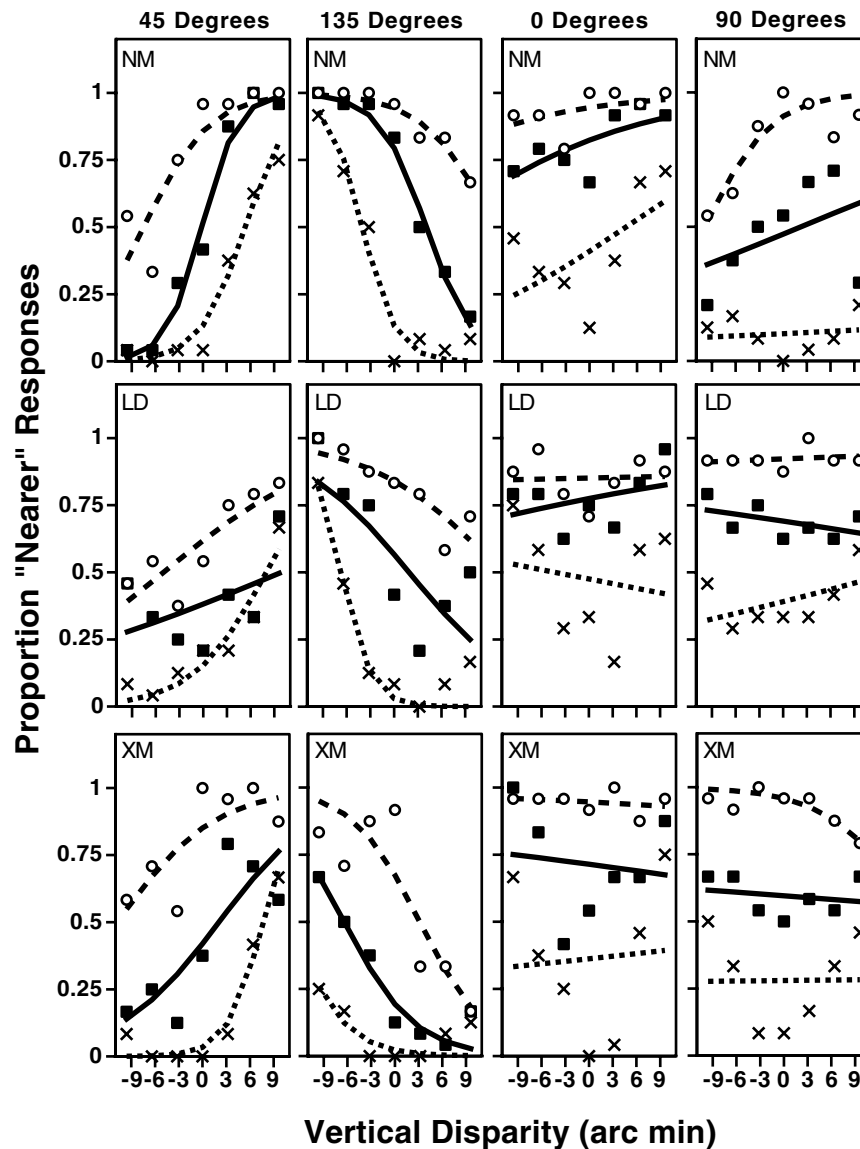


Fig. 9. The interaction between vertical and horizontal disparities at different stimulus orientations. Within each curve, the fraction of trials in which the four outer dots were seen as nearer than the zero-disparity central dot is plotted as a function of vertical disparity. Separate curves in a panel are drawn for the $-0.8'$ (dashed), 0 (solid), and $+0.8'$ (dotted) horizontal disparities. Each column of panels corresponds to a different stimulus orientation. When the stimulus orientation changed from 45° (first column) to 135° (second column), the sign of the slope reversed. When the stimulus orientation was 0° or 90° (third and fourth columns), the slopes of the curves flattened. Across all orientations, the midpoints of the curves shifted reliably downward (i.e., the proportion of nearer responses decreased) as the horizontal disparity changed from negative (near) to positive (far).

cancellation between the two types of disparities in generating depth perception.

In the discussion following Eq. (6) above, we predicted the dependence of the vertical-disparity effect on stimulus orientation. In particular, a given vertical disparity should have opposite depth signs in stimuli with 45° and 135° orientations, and should be ineffective in stimuli with 0° and 90° orientations. The results in Fig. 9 are consistent with the prediction. Specifically, the proportion of nearer responses increased (positive slope) with vertical disparity when the stimuli were oriented at 45° (Fig. 9, first column), and decreased (negative slope) with vertical disparity when the stimuli were oriented at 135° (second column). Moreover, at the cardinal orientations (0° and 90°) the proportion of nearer responses was relatively flat across the range of vertical disparities (Fig. 9, third and fourth columns). Among the 18 curves for the two diagonal orientations in Fig. 9, 17 are fitted significantly by a sigmoidal function at the level of $p < 0.05$. In contrast, only 2 of the 18 curves for the two cardinal orientations are fitted significantly by a sigmoidal function. Interestingly, at the cardinal orientations when the horizontal disparity is $0.8'$, the data (crosses) tend to be arranged in a V shape with the end points near the 50% chance level. This is probably due to a weakened depth perception at large vertical disparity magnitudes. However, for unknown reasons, the V shape does not appear when the horizontal disparity is $-0.8'$.

The depth effect of horizontal disparity is also readily apparent in Fig. 9. Specifically, the midpoint of the psychometric functions reliably shifted downward, i.e., the proportion of nearer responses decreased, as the horizontal disparity changed from negative (dashed curves) through zero (solid curves) to positive (dotted curves), as expected. This pattern can be seen across all stimulus orientations.

5. Discussion

We have proposed a new theory of depth perception from vertical disparity based on existing physiological data. The theory can account for the local depth effect of vertical disparity, the induced effect from overall vertical magnification, and the quadrant and display-size dependence. It also explains that under identical conditions, vertical disparity is usually less effective than horizontal disparity in generating depth perception. The theory has two main ingredients. First, an oriented cell treats a vertical disparity as an equivalent horizontal disparity according to Eq. (6). Consequently, for a given vertical disparity, cell families with different preferred orientations “report” different equivalent horizontal disparities and the value reported by the most responsive (or most represented) orientation will not be

completely averaged out after pooling across all orientations. Second, there is a radial bias of orientation preference in the brain. For stimuli with a dominant, near-diagonal orientation such as those in Figs. 3 and 7, the first ingredient by itself predicts a depth effect from vertical disparity because those cells whose preferred-orientation matches the stimulus orientation are most responsive. On the other hand, for stimuli without a dominant orientation such as random textures commonly used in induced-effect experiments, the second ingredient (radial bias) becomes essential as it makes a particular orientation most represented at a given location.

We have also confirmed two predictions of our theory through psychophysical experiments: (1) vertical and horizontal disparities can locally enhance or cancel each other according to their depth signs, and (2) the depth effect of the vertical disparity is orientation dependent. The orientation dependence is closely related to the quadrant dependence reported by Westheimer and Pettet (1992). When a stimulus has a dominant orientation, that orientation should be used in Eq. (6) to determine the depth sign of a given vertical disparity, leading to the orientation dependence. When a stimulus does not have a dominant orientation, its location in different quadrants will activate different cortically over-represented orientations that should be used in Eq. (6), resulting in the quadrant dependence. A related point is that if a stimulus is not only oriented but the orientation is also radial (i.e., going through the fixation point), as in Figs. 3 and 7, then the stimulus orientation and the over-represented orientation agree and should reinforce each other. It is therefore best to test the orientation dependence of vertical disparity with oriented stimuli going through the fixation point, as we did in this study. On the other hand, if the stimulus orientation is orthogonal to the radial orientation (e.g., a 45° orientation in the second quadrant or a 135° orientation in the first quadrant), then the two factors conflict with each other and the depth effect of vertical disparity should be weaker in this case. This prediction is consistent with our informal observations. However, a formal test may be difficult because stimuli with different orientations in the same quadrant, or the same orientation in different quadrants, also have different eccentricity distributions.

It should be noted that although we constructed our theory based on V1 RF properties, we do not imply that binocular depth perception necessarily happens in V1; later stages may have similar properties, or may simply inherit and refine V1 responses to generate perception. Also note that the reported radial orientation bias is from cats and monkeys, and is considered as a secondary phenomenon by some physiologists. It is not known whether the bias also exists in the human visual cortex. Thus, the proposed role of the bias in our theory is only a hypothesis. The explanatory and predictive power of

our theory provides some indirect support for the hypothesis.

5.1. Vergence eye movement

Recent experiments indicate that the neural signals for binocular depth perception and for vergence eye movement are likely to be different (Cumming & Parker, 1997; Masson, Busetini, & Miles, 1997). Therefore, the relation we proposed between vertical and horizontal disparities in depth perception does not imply a similar relation between vertical and horizontal vergence eye movements. A different population of cells is likely to be responsible for vergence, and based on the relevant psychophysical results (e.g., see Chapter 10 of Howard & Rogers (1995)), those cells must treat horizontal vergence and vertical vergence differently. Interestingly, in monkey area V1, only cells in the supragranular layers that project to higher visual cortical areas have the radial orientation bias required by our theory, while cells in the infragranular layers that project to subcortical eye-movement structures do not have the radial bias (Bauer & Dow, 1989; Bauer et al., 1983).

It is also interesting to note that our theory works well when the eyes change their direction or vergence. This is because the radial bias of cortical orientation preference is always relative to the foveal representation corresponding to the fixation point at any given instant, and therefore will not be destroyed by eye movements.

5.2. The relationship between our theory and that of Arditi et al.

Our theory is related to but different from the theory of Arditi et al. (1981), who assumed that vertical magnification of one image generates *real* horizontal disparities in the stimulus. Their theory is only valid for stimuli containing continuous, oriented contours (Gillam & Lawergren, 1983; Mayhew & Frisby, 1982), as shown in Fig. 10a. For stimuli with dotted orientations as in Fig. 10b, or with no orientation at all, a vertical disparity between the two eyes does not generate any real horizontal disparity in the stimulus. In contrast, our theory assumes that at the level of cortical-cell responses (instead of at the level of stimuli), a vertical disparity is treated as an *equivalent* horizontal disparity according to Eq. (6). Since orientation-tuned cells will respond to the dotted orientation (though not as strongly as to continuous orientation), Eq. (6) is still valid for Fig. 10b. For stimuli without a dominant orientation, the radial bias of cortical orientation preference generates an equivalent horizontal disparity from vertical disparity. Thus, our theory can explain the depth effect of vertical disparity in stimuli with or without continuous orientations.

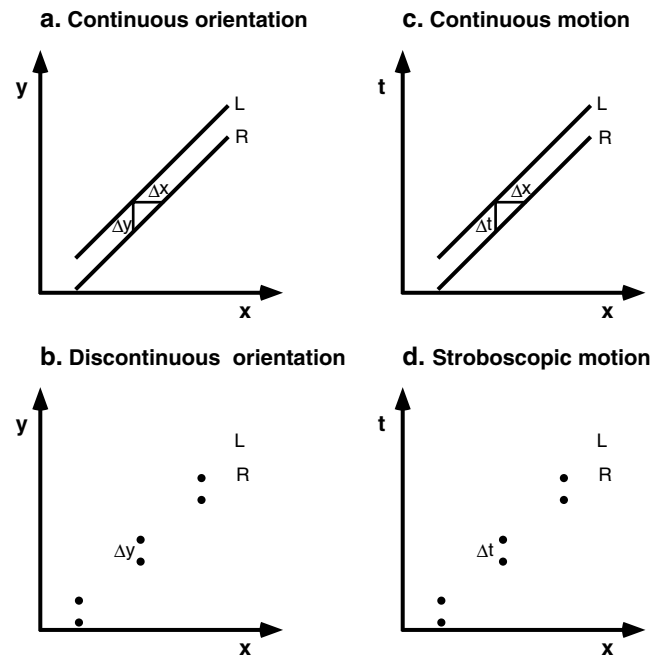


Fig. 10. x - y (a,b) and x - t (c,d) plots for spatial orientation and motion, respectively, illustrating the analogy between vertical disparity and interocular time delay. (a) For a continuous spatial orientation, a vertical shift Δy between the left and the right images can be viewed as generating a real horizontal disparity Δx . The same argument applies to a vertical magnification of one eye's image with respect to the other; the only difference is that Δy varies with y in that case. (b) The argument in (a) does not hold, however, if the orientation is discontinuous. (c) For a continuous motion (or equivalently, a spatiotemporal orientation), an interocular time difference Δt between the left and right images can be viewed as generating a real horizontal disparity Δx . (d) The argument in (c) does not hold if the motion is stroboscopic. The situations in (b) and (d) can be realized in the real world through occlusions, as shown in Fig. 4a and b, respectively.

Interestingly, if the vertical dimension is replaced by time (and therefore vertical disparity by interocular time difference, and spatial orientation by spatiotemporal orientation or motion), the explanation of Arditi et al. simply becomes the standard explanation for Pulfrich's pendulum, which asserts that an interocular time delay and continuous pendulum motion generate a *real* binocular disparity in the stimulus (see Fig. 10c). The standard explanation cannot account for the Pulfrich effect in stroboscopically presented dots (see Fig. 10d) which completely eliminate binocular disparity in the stimulus (Burr & Ross, 1979; Morgan & Thompson, 1975), just as Arditi et al.'s explanation fails when there are no continuous contours in the stimuli. We have shown previously that different versions of the Pulfrich effect can all be explained by assuming that cortical cells treat an interocular time delay as an equivalent horizontal disparity (Qian & Andersen, 1997). Here we suggest that a similar equivalence between vertical and horizontal disparities at the level of cortical-cell responses can explain the depth effect of vertical disparity.

From the functional point of view, it has been noted that an object moving behind small apertures can generate an interocular time difference without binocular disparity (Fig. 4b), and the Pulfrich effect provides a solution for seeing the correct depth (Burr & Ross, 1979). Analogously, vertical disparity can be of significant magnitude even near the fovea when spatially oriented contours in depth are viewed through narrow vertical apertures (Farell, 1998) (Fig. 4a), and our theory explains how depth is perceived in this case.

5.3. Local vs. global roles of vertical disparity

It is often argued that horizontal disparity acts locally whereas vertical disparity acts globally. We believe that while horizontal disparity does appear to have a relatively more local role than vertical disparity (Kaneko & Howard, 1997), there is evidence indicating that the effect of horizontal disparity is not absolutely local (Brookes & Stevens, 1989; Glennerster & McKee, 1999), and the effect of vertical disparity is not completely global (Kaneko & Howard, 1996; Rogers & Koenderink, 1986). There are ecological reasons for both a global and a local role of vertical disparity. Globally, vertical disparity can in principle help determine the viewing geometry (Backus et al., 1999; Banks & Backus, 1998; Garding et al., 1995; Gillam & Lawergren, 1983; Howard & Kaneko, 1994; Mayhew & Longuet-Higgins, 1982; Ogle, 1950; Rogers & Bradshaw, 1993); Locally, vertical disparity can be an important depth cue when oriented contours are viewed through narrow vertical apertures as illustrated in Fig. 4a (Farell, 1998). The quadrant dependence and display-size dependence of vertical disparity may appear to argue for a solely global role of vertical disparity. But as we have explained, these features are actually highly consistent with our theory based on local disparity detectors. Indeed, since our theory predicts that the vertical-disparity effect is size dependent, and is less strong than horizontal disparity over small regions, it is not surprising that perceptually, the effect of vertical disparity is less local than that of horizontal disparity. We therefore conclude that both horizontal and vertical disparities can generate local depth effects, and at the same time can be influenced by global context. Since vertical disparity is weaker than horizontal disparity over a small region, it can be more readily influenced than the horizontal disparity by the global context and thus appears less local. A theory that assumes global pooling of vertical disparity at the very first stage cannot explain relatively local aspects of vertical disparity. Thus, it is best to model binocular depth perception by an initially local processing stage, as we propose here, and adding a more global stage later (which we have not yet formally proposed) to introduce interactions among local detectors. Needless to say, our local model does not exclude a global role of vertical

disparity. In fact, the global disparity information is fully contained in the spatial distribution of our local disparity detectors, which could feed into the next stage for further processing (such as global pooling or the extraction of viewing geometry).

5.4. Concluding remarks

To our knowledge, our theory provides the first physiological explanation of depth perception from vertical disparity, and is consistent with most extant experimental findings on vertical disparity. Our theory is not mutually exclusive with previous theories, since our theory is at the physiological level whereas most previous theories are at the ecological or geometrical level. They do not necessarily contradict each other, just as an ecological explanation of perceptual grouping does not necessarily contradict an explanation based on long-range connections among V1 cells tuned to similar properties. Our theory has the advantage of being parsimonious as the measurement and the interpretation stages are associated with the same population of cells. Most previous theories, in contrast, are only concerned with interpretation, and require a measurement stage to provide proper inputs. In addition, our theory, unlike most previous theories, can explain why vertical disparity is generally weaker than horizontal disparity in generating depth perception and why the difference between the two types of disparities decreases with increasing display size. Finally, our theory avoids some difficulties of those previous theories which assume that vertical disparity is used to estimate the gaze angle and/or fixation distance. In particular, the gaze-angle-shift predicted by those models has never been observed and our theory does not make such a prediction.

On the other hand, our physiological theory is not yet as comprehensive as some ecological or geometrical theories, which, for example, can account for related phenomena such as the weighted combination of different depth cues from retinal and extra-retinal sources (Backus et al., 1999; Banks & Backus, 1998) and effects of half-occlusions (Anderson, 1994; Malik, Anderson, & Charowhas, 1999). Much work is needed in the future for finding plausible physiological implementations for all the important components in the ecological and geometrical theories.

Acknowledgements

We would like to thank Dr. Suzanne McKee for suggesting the brief presentation paradigm, and Drs. Ben Backus, Yuzhi Chen, Bruce Cumming, Vince Ferrera, John Koester, Zili Liu, and Fred Miles for helpful discussions and comments. This work was supported by

NIH grants #R01 MH54125 and F32 EY06969, and by the McDonnell Foundation.

Appendix A

We derive Eq. (4) in this appendix. Eq. (4) is obtained under the assumption that both H and V are significantly smaller than the RF sizes. It is easy to see the validity of Eq. (4) under the special case of a point stimulus at location (x, y) with horizontal disparity H and vertical disparity V . According to Eqs. (1) and (2), the left and right RF responses in this case are

$$r_l = g(x, y) \cos(\omega_x x + \omega_y y + \phi_l) \quad (\text{A.1})$$

$$r_r = g(x - H, y - V) \cos[\omega_x(x - H) + \omega_y(y - V) + \phi_r] \quad (\text{A.2})$$

Since H and V are much smaller than the RF sizes, we have $g(x - H, y - V) \approx g(x, y)$. In the disparity energy model, the complex cell reaches maximum response when the monocular responses of the constituent simple cells are matched. This happens when the arguments of the two cosine terms above are identical:

$$\phi_l = -\omega_x H - \omega_y V + \phi_r \quad (\text{A.3})$$

Rearrangement of this expression gives Eq. (4). The general, formal derivation of Eq. (4) is analogous to our previous work on the Pulfrich effect (Qian & Andersen, 1997), where Eq. (13) shows that among the complex cells with the same preferred horizontal spatial frequency ω_x and temporal frequency ω_t but different phase-parameter difference ϕ_- , the cell maximally excited by a stimulus with a real horizontal disparity H and an interocular time difference T satisfies the condition:

$$\phi_- \approx \omega_x H + \omega_t T \quad (\text{A.4})$$

Eq. (4) can be obtained by substituting time for the vertical spatial dimension in the derivation.

References

- Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, *300*, 523–525.
- Anderson, B. L. (1994). The role of partial occlusion in stereopsis. *Nature*, *367*, 365–368.
- Anzai, A., Ohzawa, I., & Freeman, R. D. (1999a). Neural mechanisms for processing binocular information: I. Simple cells. *Journal of Neurophysiology*, *82*, 891–908.
- Anzai, A., Ohzawa, I., & Freeman, R. D. (1999b). Neural mechanisms for processing binocular information: II. Complex cells. *Journal of Neurophysiology*, *82*, 909–924.
- Arditi, A., Kaufman, L., & Movshon, J. A. (1981). A simple explanation of the induced size effect. *Vision Research*, *21*, 755–764.
- Backus, B. T., Banks, M. S., van Ee, R., & Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research*, *39*, 1143–1170.
- Banks, M. S., & Backus, B. T. (1998). Extra-retinal and perspective cues cause the small range of the induced effect. *Vision Research*, *38*, 187–194.
- Bauer, R., & Dow, B. M. (1989). Complementary global maps for orientation coding in upper and lower layers of the monkey's foveal striate cortex. *Experimental Brain Research*, *76*, 503–509.
- Bauer, R., Dow, B. M., Synder, A. Z., & Vautin, R. G. (1983). Orientation shift between upper and lower layers in monkey visual cortex. *Experimental Brain Research*, *50*, 133–145.
- Bishop, P. O. (1996). Stereoscopic depth perception and vertical disparity: Neural mechanisms. *Vision Research*, *36*, 1969–1972.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Brookes, A., & Stevens, K. A. (1989). Binocular depth from surfaces versus volumes. *Journal of Experimental Psychology—Human Perception and Psychophysics*, *15*, 484–497.
- Burr, D. C., & Ross, J. (1979). How does binocular delay give information about depth? *Vision Research*, *19*, 523–532.
- Chen, Y., Wang, Y., & Qian, N. (2001). Modeling V1 disparity tuning to time-dependent stimuli. *Journal of Neurophysiology*.
- Cumming, B. G., & Parker, A. J. (1997). Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature*, *389*, 280–283.
- Cumming, B. G., Johnston, E. B., & Parker, A. J. (1991). Vertical disparities and perception of three-dimensional shape. *Nature*, *349*, 411–414.
- DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1991). Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature*, *352*, 156–159.
- Farrell, B. (1998). Two-dimensional matches from one-dimensional stimulus components in human stereopsis. *Nature*, *395*, 689–693.
- Garding, J., Porrill, J., Mayhew, J. E. W., & Frisby, J. P. (1995). Stereopsis, vertical disparity and relief transformations. *Vision Research*, *35*, 703–722.
- Gillam, B., & Lawergren, B. (1983). The induced effect, vertical disparity, and stereoscopic theory. *Perception and Psychophysics*, *34*, 121–130.
- Glennerster, A., & McKee, S. P. (1999). Bias and sensitivity of stereo judgements in the presence of a slanted reference plane. *Vision Research*, *39*, 3057–3069.
- Heeger, D. J. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America A*, *4*(8), 1455–1471.
- Hildreth, E. C. (1984). Computations underlying the measurement of visual motion. *Artificial Intelligence*, *23*, 309–355.
- Horn, B. K. P., & Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, *17*, 185–203.
- Howard, I. P., & Kaneko, H. (1994). Relative shear disparity and the perception of surface inclination. *Vision Research*, *34*, 2505–2517.
- Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis*. New York, NY: Oxford University Press.
- Kaneko, H., & Howard, I. P. (1996). Relative size disparities and the perception of surface slant. *Vision Research*, *36*, 1919–1930.
- Kaneko, H., & Howard, I. P. (1997). Spatial limitation of vertical-size disparity processing. *Vision Research*, *37*, 2871–2878.
- Koenderink, J. J., & van Doorn, A. J. (1976). Geometry of binocular vision and a model for stereopsis. *Biological Cybernetics*, *21*, 29–35.
- Leventhal, A. G. (1983). Relationship between preferred orientation and receptive field position of neurons in cat striate cortex. *Journal of Comparative Neurology*, *220*, 476–483.
- Liu, L., Stevenson, S. B., & Schor, C. W. (1994). A polar coordinate system for describing binocular disparity. *Vision Research*, *34*, 1205–1222.
- Malik, J., Anderson, B. L., & Charowhas, C. E. (1999). Stereoscopic occlusion junctions. *Nature*, *2*, 840–843.

- Masson, G. S., Busetini, C., & Miles, F. A. (1997). Vergence eye movements in response to binocular disparity without depth perception. *Nature*, *389*, 283–286.
- 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., & Frisby, J. P. (1982). The induced effect: Arguments against the theory of Arditi, Kaufman and Movshon. *Vision Research*, *22*, 1225–1228.
- Mayhew, J. E. W., & Longuet-Higgins, H. C. (1982). A computational model of binocular depth perception. *Nature*, *297*, 376–379.
- McKee, S. P., & Levi, D. M. (1987). Dichoptic hyperacuity: The precision of nonius alignment. *Vision Research*, *4*, 1104–1108.
- McKee, S. P., Levi, D. M., & Bowne, S. F. (1990). The imprecision of stereopsis. *Vision Research*, *30*, 1763–1779.
- McKee, S. P., Welch, L., Taylor, D. G., & Bowne, S. F. (1990). Finding the common bond: Stereoacuity and the other hyperacuties. *Vision Research*, *30*, 879–891.
- Morgan, M. J., & Castet, E. (1997). The aperture problem in stereopsis. *Vision Research*, *37*, 2737–2744.
- Morgan, M. J., & Thompson, P. (1975). Apparent motion and the Pulfrich effect. *Perception*, *4*, 3–18.
- Ogle, K. N. (1950). *Researches in binocular vision*. Philadelphia, PA: W B Saunders.
- Ohzawa, I., DeAngelis, G. C., & 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., & Freeman, R. D. (1996). Encoding of binocular disparity by simple cells in the cat's visual cortex. *Journal of Neurophysiology*, *75*, 1779–1805.
- Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1997). Encoding of binocular disparity by complex cells in the cat's visual cortex. *Journal of Neurophysiology*, *77*, 2879–2909.
- Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Qian, N. (1994). Computing stereo disparity and motion with known binocular cell properties. *Neural Computation*, *6*, 390–404.
- Qian, N. (1997). Binocular disparity and the perception of depth. *Neuron*, *18*, 359–368.
- Qian, N., & Andersen, R. A. (1997). A physiological model for motion-stereo integration and a unified explanation of Pulfrich-like phenomena. *Vision Research*, *37*, 1683–1698.
- Qian, N., & Zhu, Y. (1997). Physiological computation of binocular disparity. *Vision Research*, *37*, 1811–1827.
- Rogers, B. J., & Bradshaw, M. F. (1993). Vertical disparities, differential perspectives and binocular stereopsis. *Nature*, *361*, 253–255.
- Rogers, B. J., & Koenderink, J. (1986). Monocular aniseikonia: A motion parallax analogue of the disparity-induced effect. *Nature*, *322*, 62–63.
- Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Research*, *38*, 743–761.
- Sobel, E. C., & Collett, T. S. (1991). Does vertical disparity scale the perception of stereoscopic depth? *Proceedings of the Royal Society of London B*, *244*, 87–90.
- Vidyasagar, T. R., & Henry, G. H. (1990). Relationship between preferred orientation and ordinal position in neurons of cat striate cortex. *Visual Neuroscience*, *5*, 565–569.
- Westheimer, G. (1978). Vertical disparity detection: Is there an induced size effect. *Investigative Ophthalmology and Visual Science*, *17*, 545–551.
- Westheimer, G. (1984). Sensitivity for vertical retinal image differences. *Nature*, *307*, 632–634.
- Westheimer, G., & Pettet, M. W. (1992). Detection and processing of vertical disparity by the human observer. *Proceedings of the Royal Society of London B*, *250*, 243–247.
- Wilson, H. R., Ferrera, V. P., & Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience*, *9*, 79–97.
- Xu, P., Matthews, N., & Qian, N. (2000). Depth perception from vertical disparity, and interactions between vertical and horizontal disparities. *Investigative Ophthalmology and Visual Science Supplement (ARVO)*, *40*, 1238.