Available online at www.sciencedirect.com
SCIENCE DIRECT®
PERGAMON
Vision Research 43 (2003) 2479–2492
Vision Research
www.elsevier.com/locate/visres

Rapid Communication

Stereoscopic depth perception from oblique phase disparities ☆

Saumil S. Patel [a,b,*], Michael T. Ukwade [c], Scott B. Stevenson [b,c], Harold E. Bedell [b,c], Vanitha Sampath [c], Haluk Ogmen [a,b]

[a] College of Engineering, University of Houston, Houston, TX 77204, USA
[b] Center for Neuro-Engineering and Cognitive Science, University of Houston, Houston, TX 77204, USA
[c] College of Optometry, University of Houston, Houston, TX 77204, USA

Received 24 June 2002; received in revised form 12 March 2003

Abstract

In order to understand the role of oblique retinal image disparities in the perception of stereoscopic depth, we measured the depth perceived from random dot stereograms in which phase disparities were introduced in a selected band of stimulus orientations. A band of orientation was defined by a center orientation that ranged from 7.5 (near vertical) to 82.5 o[rientation]deg and by a bandwidth that was defined as the difference between the highest and the lowest orientation in the band. The bandwidths tested were 15, 30 and 45 odeg. A constant phase disparity of 90 p[hase]deg was introduced in all of the oriented spatial frequency components within the orientation band and the perceived depth of each stimulus was matched using a small square binocular probe. For each bandwidth, perceived depth increased with an increase in the center orientation up to approximately 60 odeg. This suggests that the human stereovision system derives a large proportion of information about perceived stereoscopic depth from oblique phase disparities. Simulations using an energy model of stereoscopic depth perception indicate that oblique phase disparities are unlikely to be processed by neural mechanisms tuned to near-vertical orientations within the stimulus. Our results therefore suggest that oblique retinal disparities are initially detected as oblique phase disparities by binocular mechanisms tuned to oblique orientations. Because the perceived depth from oblique phase disparities is consistent with the trigonometrically determined equivalent horizontal disparities, we presume that the information from oblique phase disparities is included in the visual system's computation of the horizontal retinal disparity.

1. Introduction

A three-dimensional perception of the visual world is derived from horizontally disparate views of each eye (Wheatstone, 1838). When an object is seen by both eyes, its horizontal position disparity can be computed either by matching inter-ocular features (e.g. luminance, edges, lines, curves etc.) or by computing inter-ocular spatial phase differences between the various spatial frequency components in the two images (see Howard & Rogers, 1995). Consider a planar object that contains multiple spatial frequency components of vertical orientations (i.e., an object produced from the Fourier synthesis of vertical gratings). A horizontal position disparity in this object with respect to a fixation point results in spatial phase differences in each spectral component of the images of the two eyes (phase disparities). Further, these phase disparities vary linearly with the spatial frequency of each image component. In the case of a planar object that contains spatial frequency components of non-vertical (or oblique) orientations, the resulting phase disparities for a given spatial frequency vary also with the cosine of the orientation. It is possible to recover the object's horizontal position disparity from the spatial frequency components at oblique orientations. In order to do so, the phase disparity (which is measured along the direction of contrast modulation for the given spatial frequency) has to be scaled by the spatial frequency and the cosine of the oblique orientation (see Fig. 1). Based on physiological and psychophysical evidence (e.g. Anzai, Ohzawa, & Freeman, 1997; Mansfield & Parker, 1993), we assume that the binocular mechanisms responsible for

☆ Parts of this study were presented at the 1998 and 2001 annual meetings of ARVO.

* Corresponding author. Address: College of Engineering, Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77204, USA.
E-mail address: saumil@swbell.net (S.S. Patel).

0042-6989/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0042-6989(03)00464-4

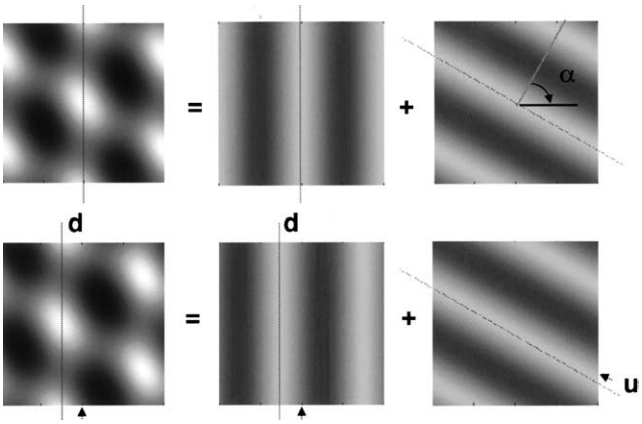


Fig. 1. Extraction of an object's position disparity from its vertical and oblique spatial frequency components. The top left panel shows the image of a planar object that consists of two spatial frequency components (f cpd) oriented at 0 odeg (top-middle) and α odeg (top-right; see methods for unit conventions). The long dotted lines indicate a reference position in corresponding images. A horizontal displacement of the object, relative to the viewing aperture, of d deg results in an image (bottom-left) in which the vertical (bottom-middle) and oblique (bottom-right) spatial frequency components are displaced differentially (compare d with u), along the direction parallel to luminance modulation. The horizontal displacement of the 0 odeg component is d deg. Using geometry, it can be shown that the oblique displacement u, of the α odeg component, is equal to $d\cos(\alpha)$ deg. If the top-left and bottom-left images are viewed by different eyes, a binocular phase detector tuned to f cpd and α odeg would signal a phase disparity of ϕ radians, which is equal to $2\pi fu$. Thus, the object's position disparity (or equivalent horizontal disparity), d, can be computed from this oblique phase disparity by dividing ϕ by $2\pi f\cos\alpha$.

stereoscopic depth perception are relatively narrowly tuned to both orientation and spatial frequency. If the visual system utilizes oblique disparities in the computation of relative depth then the following prediction can be made: the depth perceived from a stimulus that contains only a constant oblique phase disparity will increase with a change in the orientation from vertical (0 orientation degrees [odeg, see methods for units conventions]) to horizontal (90 odeg). This prediction arises because the recovery of an object's horizontal disparity requires that the oblique phase disparity be divided by the cosine of the oblique orientation, and the cosine of the oblique orientation decreases as it changes from vertical to horizontal.

Numerous psychophysical studies have indicated that stereoscopic depth perception is affected by the orientation of the binocular stimulus. Ebenholtz and Walchli (1965) measured stereothresholds for a pair of line targets as a function of head and object orientation. They found that the stereothreshold increased according to a cosine function as the orientation of the target on the retina changed from vertical to horizontal, regardless of whether the orientation change was produced by head or target tilt. Subsequently, Blake, Camisa, and Antoinetti (1976) confirmed that the relationship between stereothreshold and target orientation followed a cosine function. Remole, Code, Matyas, and Mcleod (1992) measured the perceived tilt of a frontoparallel plane formed by an array of parallel rods, as a function of the rod orientation. They found that the perceived tilt increased as the orientation of the rods changed from vertical to near horizontal, again according to a cosine function. Morgan and Castet (1997) measured stereothresholds for sinusoidal gratings and Gabor patches at different orientations. When expressed as spatial phase disparities, stereothresholds for 1 and 2 cpd gratings and for 8 cpd Gabor patches remained constant to an orientation of about 80 o[rientation]deg. This outcome is consistent with a cosine relationship between the stereothreshold and grating/Gabor orientation, if each oblique phase disparity is expressed in terms of an equivalent horizontal disparity (see Fig. 1 and Eq. (1)). Based on their results, Morgan and Castet suggested that stereo matching in some cases may be performed by orientationally tuned mechanisms.

Mansfield and Parker (1993) used spatially filtered random-dot targets to measure contrast thresholds for stereopsis. The random-dot stimuli consisted of two components: a signal component and a mask (or uncorrelated noise) component. Each component was band-limited in both spatial frequency and orientation. Contrast thresholds for stereopsis were measured for various center orientations of the signal and mask components. The results indicate that the contrast threshold for stereopsis is elevated when the center orientation of the mask matches the center orientation of the signal, especially for high center spatial frequencies. At these high spatial frequencies, a mask oriented orthogonal to the signal's center orientation has only a small effect on contrast threshold for stereopsis. This finding that the contrast threshold for stereopsis exhibits a tuning function with respect to the orientation of the mask is consistent with the suggestion that stereo matching is performed by orientation tuned mechanisms. Simmons and Kingdom (1995) also presented data to suggest that obliquely tuned neurons play a role in the perception of depth. In particular, a disparity energy model with phase disparity mechanisms tuned to oblique orientations, was able to account for the contrast thresholds that they measured at large disparities (>60 arc-min) for horizontally oriented isochromatic Gabor stimuli.

Sensitivity to oblique stimulus disparities has also been proposed as an explanation for the induced size effect. In the induced size effect, depth is perceived when images of slightly different vertical magnifications are presented to the two eyes. Arditi, Kaufman, and Movshon (1981) found that the induced effect is stronger for stimuli that contain oblique spatial frequency components compared to those that do not, suggesting a potential role for these oblique disparities in the

perception of depth. More recently, van Ee and Anderson (2001) showed that perceived depth in a stimulus that consists of a set of randomly oriented lines increases with the range of orientations that is represented. This result suggests that oblique disparities play a substantial role in the perception of depth. Finally, Stevenson and Schor (1997) determined that accurate judgments about the direction of depth are possible in a dynamic random-dot stereogram, even in the presence of substantial vertical disparities. Based on this result, they concluded that stereo-matching may be a two-dimensional process and is not restricted just to epipolar lines.

Many of the previous studies that attempted to relate perceived depth to oblique disparities used stimuli with salient features, making it difficult to determine whether the perceived depth was a result of low-level disparity processing or was a result of higher-level processes that derive depth from feature disparities. For example, Farell (1998) clearly demonstrated that the perceived stereoscopic depth from single spatial frequency gratings does not directly predict the depth perceived from a plaid. When the two gratings that comprise the plaid cohere, perceived depth depends on the horizontal disparity between features (e.g. inter-sections). If the gratings are not perceived to cohere, then the two gratings are seen with separate depths that are determined by their individual disparities. Another problem inherent in most previous studies that assessed perceived depth for stimuli with oblique disparities is that the stimuli also contained non-zero *vertical* disparities. For example, an oblique single spatial frequency grating that is viewed through an aperture contains vertical as well as horizontal retinal image disparities. Unless the grating is presented very briefly, these vertical disparities trigger reflexive vertical vergence eye movements that reduce the vertical misalignment of the two eyes' images (Howard, Allison, & Zacher, 1997). As a consequence of this vertical vergence, steady-state oblique disparities in the stimulus may be altered in the retinal representations.

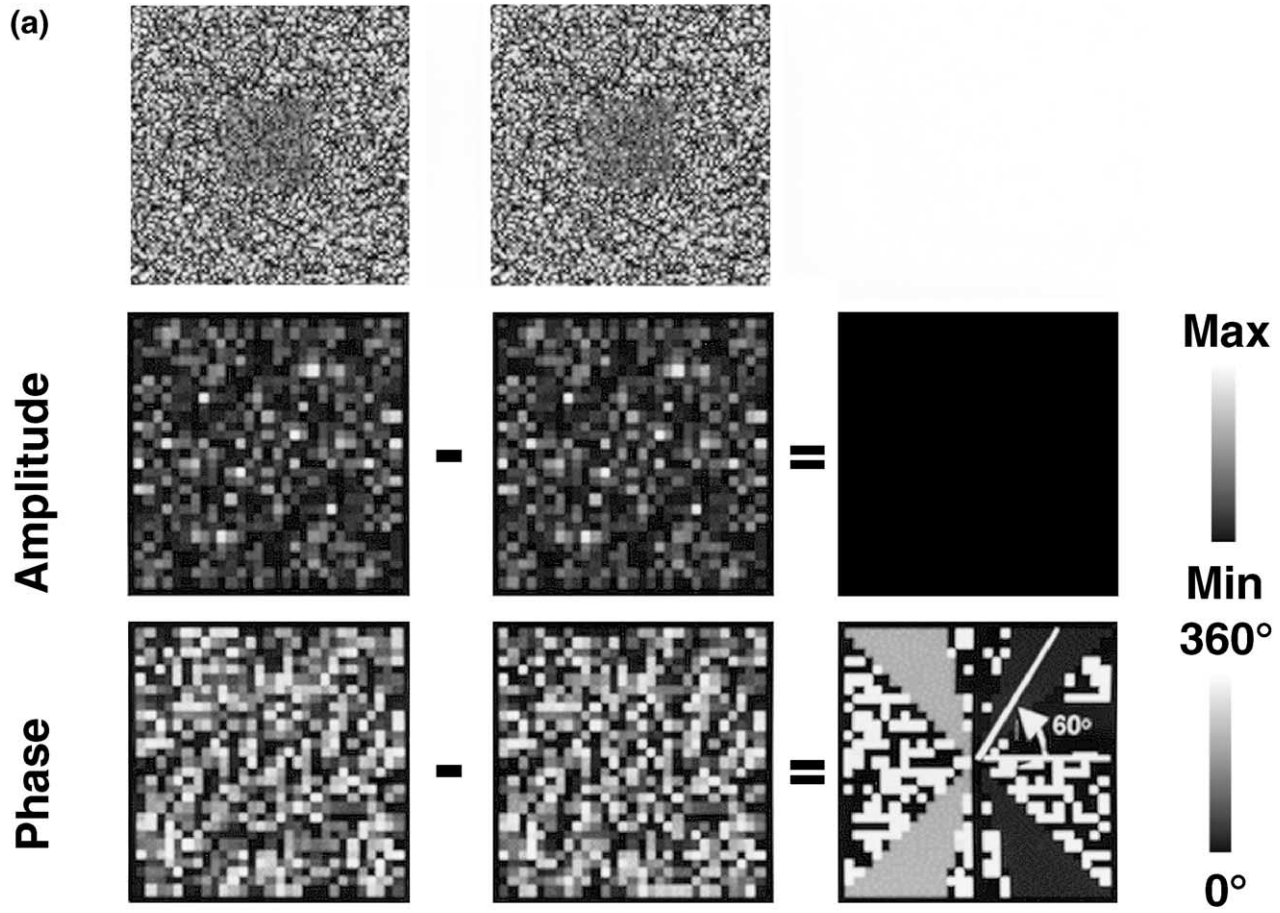
In this study, we measured perceived depth as a function of stimulus orientation by introducing constant phase disparities rather than constant horizontal position disparities in random-dot stimuli. To minimize involuntary vertical vergence movements, our random-dot stimuli were designed to be free of aggregate vertical disparities. To clarify the types of disparity mechanisms that are likely to respond to these stimuli, we ran simulations with our stimuli using an extension of a published disparity energy model for stereoscopic depth perception (Qian & Zhu, 1997). Our goal was not to distinguish between different mechanisms (position vs. phase) that might be responsible for depth perception, but rather to show clearly the significance of oblique phase disparities in the perception of depth.

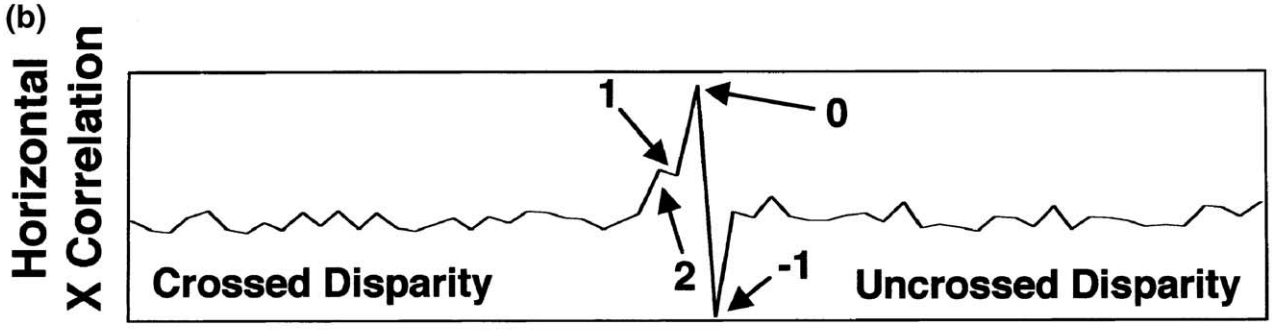
2. Methods

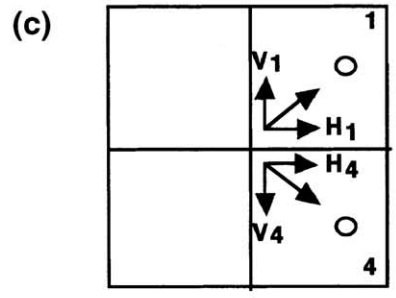
2.1. Stimuli

To directly assess the role of binocular disparities at various orientations in the perception of depth, we created broad-band random-dot binocular stimuli, in which the individual spatial frequencies within a band of orientations (*orientation band*) have a constant or coherent phase disparity and therefore different position disparities. Because the phase disparity threshold is approximately constant for a wide range of stimulus orientations (Morgan & Castet, 1997), a coherent phase disparity should stimulate putative phase disparity-sensitive mechanisms equally above their phase disparity thresholds. These stimuli do not contain salient visual features that can be matched by a feature-matching stereo mechanism (see below). Further, to prevent vertical vergence, our stimuli were orientationally balanced, i.e., the oblique phase disparities were always introduced in complementary orientations (e.g. 30 and −30 odeg). Thus, our stimuli have a purely horizontal aggregate position disparity when the vertical and horizontal components of position disparity that are equivalent to each phase disparity (equivalent position disparity; also see Eq. (1)) are averaged across all spatial frequencies and orientations. The use of coherent phase disparity has the additional advantage of eliminating the phase aliasing problem at higher spatial frequencies. This problem can be clarified with an example. A 4.5 min crossed *position* disparity in a broad-band stimuli introduces a crossed phase disparity of 27 pdeg in a 1 cpd spatial frequency component, and a disparity of 270 pdeg in a 10 cpd spatial frequency component. Because of the cyclic nature of phase disparity information, the 270 pdeg *crossed* phase disparity in the 10 cpd spatial frequency component is equivalent to an *uncrossed* phase disparity of 90 odeg. By introducing a constant phase disparity of 90 pdeg in all spatial frequency components, there is no ambiguity about the direction of depth that is signified by the phase disparities at high spatial frequencies. A sample pair of random-dot grayscale images used in our first experiment is shown in Fig. 2(a). Fig. 2(b) illustrates how the oblique phase disparities in complementary orientations were manipulated to yield stimuli with a purely horizontal position disparity.

The inner square of random-dots (1×1 deg) presented to the right eye was created from the inner square of random-dots presented to the left eye by adding a constant 90 pdeg phase angle to the Fourier spatial frequency (SF) components in two symmetrical bands of orientations (Fig. 2). The *orientation band* is characterized by a *center orientation* (midway between the maximum and minimum orientation in the band) and *bandwidth* (the difference between the maximum and minimum orientation in the band that contains the







Zero Net Vertical Disparity

constant phase disparity). A two-dimensional Fourier transform of the image of the left eye was first obtained using MatLab (Mathworks, Natick, MA). The amplitude and phase matrices were obtained from the complex Fourier matrix of the inner square viewed by the left eye. The portion of the right eye's phase matrix corresponding to the positive spatial frequencies (quadrants 1 and 4 of the square phase matrix; quadrant numbers are consistent with those used in a Cartesian coordinate system) was equal to that of the left eye plus a constant phase angle of 90 pdeg (Fig. 2, also see below for unit conventions). The portion of the right eye's complex Fourier matrix that corresponds to positive spatial frequencies was formed using the amplitude values from the left eye's amplitude matrix and the phase angles from the right eye's phase matrix. The portion of right eye's Fourier matrix that corresponds to negative spatial frequencies (quadrants 2 and 3 of the square phase matrix) was filled by the complex conjugate values from the corresponding elements in the positive frequency portion (Fig. 2). Thus, the phase of the right-eye spatial frequency components in the band of orientations was shifted to produce a net horizontal position disparity and zero vertical position disparity when averaged across all the orientations and spatial frequencies. The outline of the inner square region always remained at zero position disparity. The outer region of random dots (3.3×3.3 deg) was the same for both eyes and provided the reference plane for depth judgments of the inner square. Both eyes' inner squares had the same amplitude spectrum and, consequently, identical mean luminance (7 cd/m^2, when viewed through polarized filters) in the central regions. The mean luminance of the inner squares is the same for all pairs of images used in the experiments. The mean luminance of the outer region of each stimulus pair is equal to the average of the minimum and maximum luminances present in both the inner squares. This luminance scaling introduces a slight difference in contrast between the inner squares and the outer regions in both eye's images. Thus, the inner squares in the left and the right eyes have a similar appearance with respect to the outer squares. The root mean square contrast averaged across several images of the inner square is 0.29 ± 0.002 [$\pm$1SD], regardless of the orientation bandwidth. The root mean square contrast averaged across several images of the outer square is 0.45 ± 0.03, 0.49 ± 0.02 and 0.51 ± 0.01 for 15, 30 and 45 odeg orientation bandwidths, respectively. A normalized horizontal cross-correlation between the inner squares shown in Fig. 2(a) confirms that most of the horizontal disparity energy is at zero disparity.

2.2. Procedure

In a dark room, observers ($N = 3$) viewed orthogonally polarized halves of a computer screen (832 horizontal $\times$ 624 vertical; 15 inch diagonal; 256 gray levels; gamma corrected) from 50 cm through a pair of matched polarizing filters. The pair of images that constituted the stimulus was displayed on separate halves of the screen. The bandwidths of the oriented bands of phase disparity were 15, 30 and 45 odeg, which were presented in different experimental sessions. Within each session, pairs of images with a specific center orientation (from 7.5 to 82.5 odeg) were presented in random order. Except at noted explicitly below, all of the spatial

Fig. 2. (a) Sample pair of images and their Fourier amplitude and phase matrices. For details of how the images are created see Section 2. The center orientation of the band is 60 odeg (see Section 2 for units conventions) and the bandwidth is 30 odeg. The images in the left and right columns of the first row were presented to the right and left eyes respectively. If crossed free-fusion is achieved, viewers should see the inner square in front of the page. The Fourier amplitude and phase matrices corresponding to the inner squares of the image pair are shown in the second and third rows. In order to scale the spectral components for better illustration, the mean luminance was removed from the images prior to Fourier transformation. Note that the amplitude and phase spectra of both the inner squares are random. Also note that the difference between the amplitude spectra of the two inner squares is zero (coded as black in rightmost column of middle row). The dark gray bands in quadrants 1 and 4 in the rightmost column of the bottom row represent the spectral components having a 90 pdeg (see Section 2 for units conventions) phase disparity. The light gray bands in the left half of the matrix (negative spatial frequencies) indicate a phase difference of −90 (or 270) pdeg. This sign difference between the left and the right half plane of the phase-difference matrix occurs because the Fourier transform of a real matrix consists of a matrix whose symmetric elements in the positive and negative half planes are complex conjugates. Note that white represents phase difference of 360 pdeg which is the same as 0 pdeg which is coded as black. (b) The normalized horizontal cross-correlation between the inner squares of the images shown in the first row of panel (a). The normalized horizontal cross-correlation, which indicates the relative horizontal disparity energy, is computed by taking the average of the cross-correlation of individual pixel rows in the images of the inner squares. The numbers and arrows specify the disparities in pixels (1 pixel = 2 arc-min). Although a higher positive correlation exists in the crossed compared to the uncrossed disparity direction, the primary peak in the cross-correlation signal is at zero disparity. Therefore, no simple interpretation of the horizontal disparity energy can account for the substantial depth that is perceived from the stimuli in panel (a). (c) Illustration of how stimulus phase disparities yield net zero vertical disparity. The big square represents the Fourier matrix of the image of the right eye. The numbers in the corners of the big square indicate the quadrants. The circles represent a pair of Fourier components at complementary orientations (same magnitude but different signs of the orientation angle). Both of the Fourier components have a positive phase difference (or phase disparity) with respect to the corresponding components in the left eye's image. The oblique arrows represent the orientation of the spectral components. For a positive phase disparity, the direction of the arrows represents the direction of spatial shifts with respect to the corresponding components in the left eye's image. Each oblique spatial shift (or arrow) is decomposed into a vertical and a horizontal component (H, V). The vertical components of the spatial shift are equal and in opposite directions and hence the net vertical disparity is zero. The same would hold if the two spectral components had anegative phase disparity.

frequencies in the inner square had a constant phase disparity of 90 pdeg in a direction such that the inner square should have been perceived in front of the outer square. Observers maintained fixation at the center of the fused stimulus and matched the perceived depth of the inner square with the perceived depth of a superimposed white 8×8 arc-min probe. The perceived depth of the superimposed probe was varied by varying its horizontal position disparity using a joystick. The stimulus remained visible on the screen for an unlimited viewing duration. Except for SSP, observers were not aware of the conditions that were run in each session. Because stimuli with different center orientations were presented randomly in an experimental run, it is unlikely that any idiosyncratic biases in depth judgments could have affected the results in a center-orientation dependent manner.

2.3. Statistics

All statistics are a result of multivariate repeated measures ANOVAs performed in SuperANOVA software (Abacus Concepts, CA). Because each bandwidth condition had a different number of center orientations, separate repeated-measures ANOVAs were performed for each of the three bandwidths. Individual differences between center orientations for a given bandwidth were analyzed using post hoc contrasts. The *p*-values reported throughout this paper are Greenhouse–Geisser corrected values.

2.4. Units conventions

A units convention was adopted in which *deg* signifies visual angle, *pdeg* signifies phase angle and *odeg* signifies orientation angle. All orientation angles are specified with respect to the horizontal direction. The orientation of a spatial frequency grating component corresponds to the direction parallel to the direction of contrast modulation at that spatial frequency, i.e., 0 odeg corresponds to a vertical grating. Equivalently, the orientation of a spatial frequency grating component corresponds to the direction perpendicular to the direction of no contrast modulation.

3. Results and discussion

When observers viewed these filtered random-dot stimuli, all of them reported seeing the inner square as a single somewhat diffused surface in depth. Although this surface was not as crisp as that seen in conventional random-dot stereograms, neither was it as disorganized as might be predicted, considering that each spatial frequency component could have signaled a different depth (Boothroyd & Blake, 1984). Besides this diffused surface, observers also saw sparse structures floating at different depths. It is possible that the perception of the diffuse depth surface is due to computations performed by low-level stereo mechanisms and the perception of floating blobs is due to computations performed by feature-based stereo mechanisms.

As shown in Fig. 3, the meridian of center orientation significantly affects the perceived depth for all bandwidths (BW = 15, $F[5,10] = 15.08$, $p = 0.04$; BW = 30, $F[4,8] = 13.1$, $p = 0.02$; BW = 45, $F[3,6] = 21.79$, $p = 0.01$). As is apparent in both the average and individual data, perceived depth increases consistently as a function of center orientation for all three orientation bandwidths up to a center orientation of 60 odeg. Beyond 60 odeg, a pronounced reduction in the perceived depth occurs in the data for a bandwidth of 15 odeg. We fitted inverse cosine functions (Morgan & Castet, 1997) to the average data for bandwidths of 15 and 30 odeg as shown by the dotted lines in Fig. 3. The inverse cosine functions were forced to become minimum at the smallest center orientation for the 15 and 30 odeg bandwidths. For these two orientation bandwidths, the

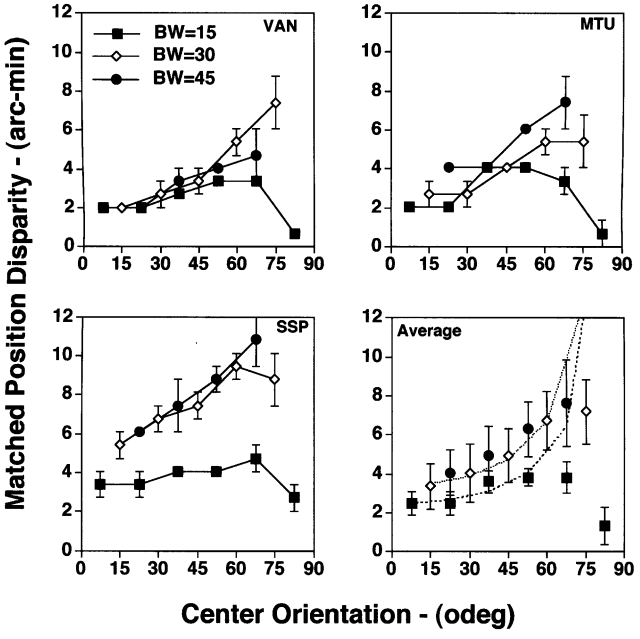


Fig. 3. Perceived depth as a function of center orientation for stimuli containing oblique phase disparities. The *x*-axis represents the center orientation of the orientation band that contains the 90 pdeg phase disparity. The *y*-axis represents the horizontal position disparity of a white square probe that was perceived to match the depth of the diffused surface of the inner square as seen by the observer. Data for three observers (VAN, MTU, SSP) and the data averaged across these observers are shown for bandwidths of 15, 30 and 45 odeg. In the lower right panel that shows the average data, the dotted lines are cosine functions shifted so that the minimum passes through the data point for a center orientation of 7.5 or 15 odeg, for bandwidths of 15 and 30 odeg respectively. The error bars represent ±1 standard error.

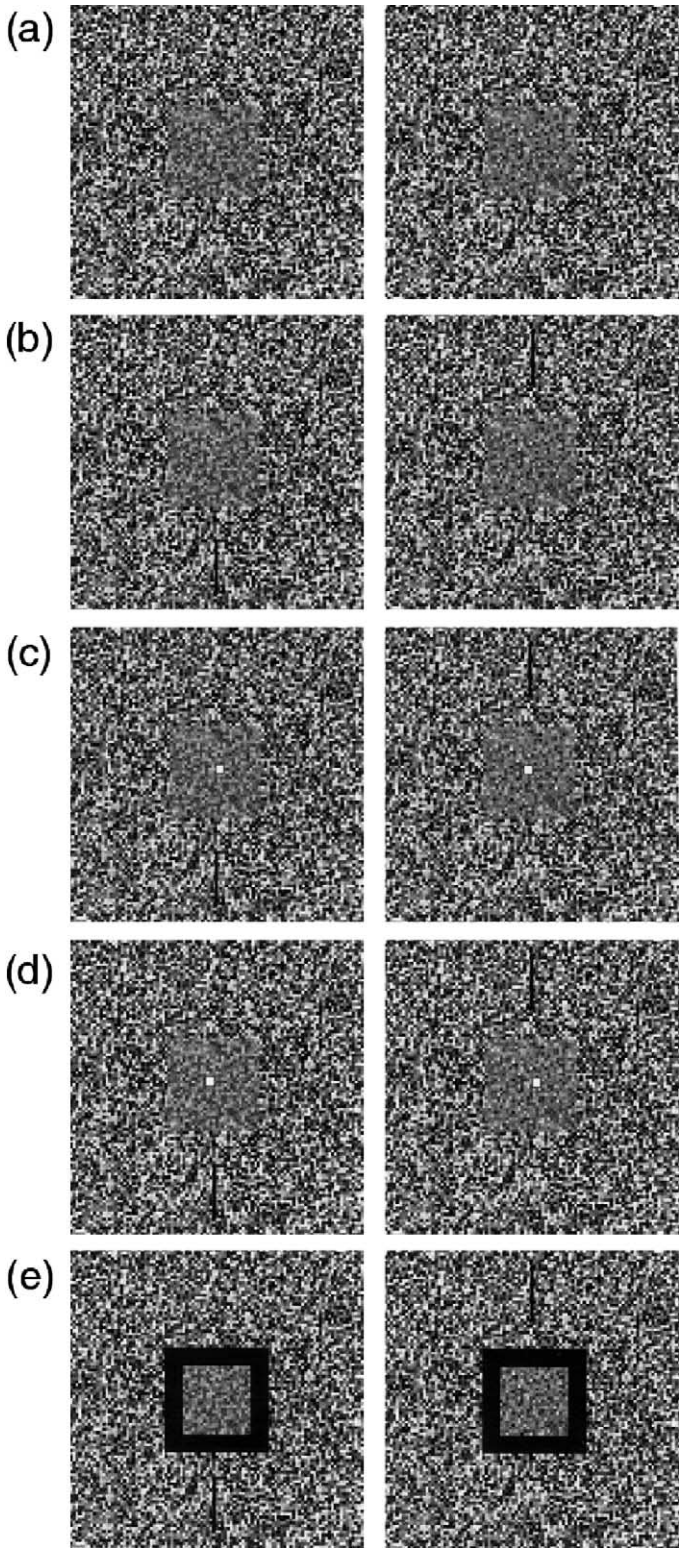


Fig. 4. Pairs of images to illustrate vergence posture, the effect of a superimposed depth probe, and the role of the edges of the inner square on perceived depth in our experiments. If the pair of images in each row is fused by crossing the two eyes, then the observer should see the inner square in front of the outer square. (a) Pair of images as presented in our experiment. The center orientation is 45 odeg and orientation bandwidth is 30 odeg. (b) The same images as in panel (a), with a pair of superimposed dark vertical Nonius lines. (c) The same images as in panel (a) with a square depth probe that should appear behind the inner square. (d) The same images as in panel (a) with a square depth probe that should appear in front of the inner square. (e) The same images as in panel (a), with pixels near the edges of the inner squares replaced by black dots. Four pixels inside each edge of the inner square are replaced by black dots. The inner square in this panel is therefore smaller than the inner square in other panels. Note that the perceived depth of the inner square with respect to the outer square is very similar in all panels.

increase in depth matches the fitted inverse cosine functions almost perfectly up to an orientation of approximately 60 odeg, but deviates substantially thereafter. This result suggests that the stereovision system may de-emphasize disparity information beyond an orientation of approximately 60 odeg. The data for 45 odeg bandwidth do not show a noticeable reduction in perceived depth for large orientations presumably because this bandwidth is too large to selectively activate independent orientation tuned mechanisms.

It is possible that the amount of perceived depth in our stimuli corresponds to the position disparity produced within a single "dominant" spatial frequency, for example, the spatial frequency at which the spatial contrast sensitivity function reaches its peak. By restricting the range of spatial frequencies that have constant phase disparities, results from additional experiments (manuscript in preparation) indicate that a single spatial frequency mechanism is not responsible for the depth perceived in the previous experiment. Therefore, the perceived depth in phase-coherent stimuli is instead likely to result from the averaging of disparity information across multiple spatial frequencies (Fleet, Wagner, & Heeger, 1996; Grossberg, 1994; Hess, Liu, & Wang, 2002; Rohaly & Wilson, 1994).

In our experiments, observers were asked to look at the center of the display and adjust the perceived depth of the probe to match the perceived depth of the inner square. The adjustment procedure took anywhere from 5 to 10 s. To illustrate that the vergence posture can be adequately maintained for this duration, some additional stimuli are shown in Fig. 4. The stimulus pair in panel 4(a) is similar to that used in our experiment. The stimulus pair in panel 4(b) is identical, except for the inclusion of a pair of Nonius lines (Fig. 4(b)). Upon fusing the images in Fig. 4(b), it can be seen that the relative horizontal position of the Nonius lines, which should be close to alignment, remains nearly constant for an extended duration. This observation suggests that a relatively stable vergence posture can be maintained for long durations.

It is possible that the disparity information in the superimposed probe might interact with the disparity information in the inner square, thereby causing a systematic error in the estimated depth of the inner square. Fig. 4(c) and (d) show the same stimuli as in Fig. 4(a) but with a superimposed depth probe either in front or behind the inner square. It is clear that the perceived depth of the inner square is very similar in Fig. 4(a) and Fig. 4(c) and (d), indicating that the neither the presence nor the disparity of the superimposed depth probe produces a substantial error in the estimation of perceived depth.

Although the edges of the inner squares in our stimuli were set to zero disparity, it is possible that perceived depth results from a conventional stereo cue produced by an interaction between the filtered random dots and the edges of the inner square. In order to remove any possible interaction cues, in Fig. 4(e), all four edges of the inner square in both eyes' images are occluded by a thick black frame. Despite removing the disparity information at the boundaries (particularly the vertical edges) between the inner and outer squares, the inner square is still perceived with substantial depth. We noted that the perceived depth of the inner square is slightly less in Fig. 4(e), compared to that in Fig. 4(a), possibly because of the reduced representation of low spatial frequency components in the smaller images. Nevertheless, the continued perception of robust planar depth in Fig. 4(e) illustrates that the depth seen in our stimuli is likely to reflect global processing of the stimulus, rather than the processing of local edge cues.

The data that are reported in Fig. 3 were obtained using only stimuli with crossed disparities. Previously, we obtained depth-matching data for stimuli with an orientation bandwidth of 45 odeg and center orientations of 22.5 and 67.5 odeg that were presented with both crossed and uncrossed phase disparity (1998 ARVO abstract). As expected, the magnitudes of perceived depth for stimuli with crossed and uncrossed *coherent* phase disparity are very similar, but in opposite directions (Fig. 5). In other experiments, constant phase disparities were introduced in all of the component orientations of the stimuli, similar to the stimuli reported here. Crossed and uncrossed constant phase disparities produced similar

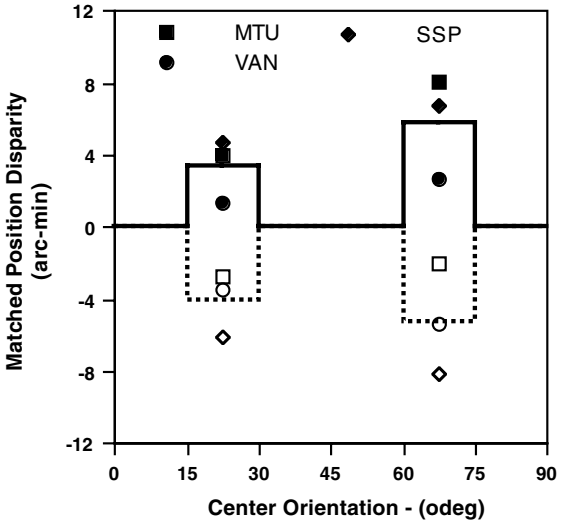


Fig. 5. Comparison of perceived depth of the inner square in stimuli with +90 (crossed) and −90 (uncrossed) pdeg *coherent* phase disparities. Different symbols represent the data from the three observers (MTU, SSP, VAN). The solid and dotted bars represent the matched position disparity averaged across the three observers. Results are shown for stimuli with an orientation bandwidth of 45 odeg and center orientations of 22.5 and 67.5 odeg. Notice that the average matched position disparity is very similar for 90 and −90 pdeg phase disparities for both center orientations.

magnitudes of perceived depth between the inner and outer squares (1998 ARVO). The reader can verify this observation qualitatively by switching from crossed to uncrossed fusion while viewing any of the stimulus pairs in Fig. 3 or Fig. 4.

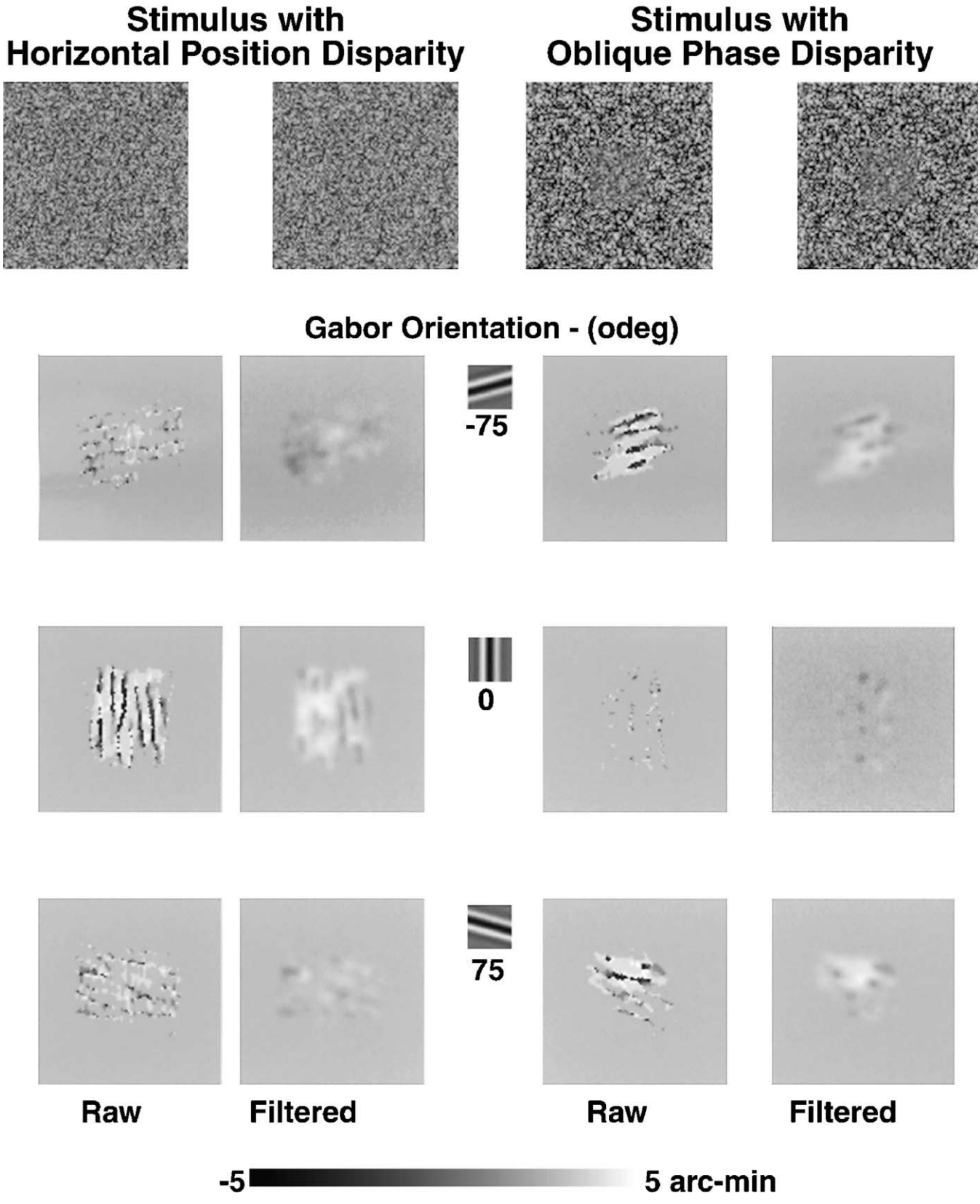
3.1. Model simulations

To test whether disparity signals can be extracted from our stimuli by disparity-sensitive mechanisms that are tuned to vertical stimulus orientations, we presented these stimuli to a standard energy model of stereoscopic depth perception. The model is similar but not identical to the one that was described by Qian and Zhu (1997). The main difference between their model and ours is in the nature of the monocular receptive fields (RFs). Whereas Qian and Zhu used one-dimensional RFs, our implementation of the model includes two-dimensional RFs for each monocular neuron. By making the RFs two-dimensional, the monocular neurons in our 'extended' model become selective for vertically oriented stimuli. In our estimation, this modification makes the extended model more consistent with neurophysiological studies, most of which suggest that mechanisms responsible for stereoscopic depth perception respond primarily to horizontal image disparities between vertically oriented components of the stimulus (Barlow, Blakemore, & Pettigrew, 1967; Burkhalter & Van Essen, 1986; Felleman & Van Essen, 1987; Gonzalez, Krause, Perez, Alonso, & Acuna, 1993; Hubel & Livingstone, 1987; Hubel & Wiesel, 1970; Maunsell & Van Essen, 1983; Poggio & Fischer, 1977; Poggio, Motter, Squatrito, & Trotter, 1985). In contrast, the binocular mechanisms in Qian and Zhu's model (1997) are isotropic, i.e., they do not exhibit any preferred orientation. A less important difference between our model and Qian and Zhu's model is the disparity smoothing function that is applied. In particular, we used a circularly symmetric Gaussian filter to spatially filter the output of the complex cells in the model. A description of our 'extended' model follows.

Our version of the model uses 16 binocular simple cells at each spatial location (eqv. image pixel location). Each simple cell sums the output of the monocular cells from corresponding retinal locations in the two eyes (for a 1-D example, see Eq. (3) in Qian & Zhu, 1997). The RF of each monocular cell is described by a two-dimensional oriented Gabor function. The spatial frequency of the Gabor was 3.75 cpd and the dispersion coefficient was 8 arc-min. The vertical and horizontal extent of the RF was approximately 6 times the dispersion coefficient. The possible spatial phase angles between the RFs of the monocular cells in the left and right eyes were −135, −112.5, −90, −67.5, −45, −22.5, 0, 22.5, 45, 67.5, 90, 112.5, and 135 pdeg. Together, these monocular cells formed eight simple cells with relative spatial phases of $(-135, 45)$, $(-112.5, 22.5)$, $(-90, 0)$, $(-67.5, -22.5)$, $(-45, -45)$, $(-22.5, -67.5)$, $(0, 90)$ and $(22.5, -112.5)$. Eight additional simple cells with relative monocular spatial phases of $(-45, 135)$, $(-22.5, 112.5)$, $(0, 90)$, $(22.5, 67.5)$, $(45, 45)$, $(67.5, 22.5)$, $(90, 0)$ and $(112.5, -22.5)$ formed the quadrature pairs needed to build complex model cells (see Eq. (4) in Qian & Zhu, 1997). The resulting eight complex cells at each spatial location were tuned to phase disparities distributed equally between −180 and 180 pdeg. At the end of the simulation, the preferred phase disparity of the most active complex cell was determined at each spatial location and converted to an equivalent position disparity by dividing its phase by the center spatial frequency of the Gabor RF. A two-dimensional map of position disparity was constructed from the position disparity signals at each spatial location. Subsequently, a circularly symmetric Gaussian filter (SD = 4 arc-min) was used to filter the map of position disparity.

Recently, Cumming (2002) reported that the monocular orientation preferences of some disparity-sensitive cortical cells do not predict the orientation of the optimal disparity response, which he found in these cells to be preferentially horizontal. Despite more robust responses to horizontal disparity, the disparity selectivity profile of these cells indicate that the subpopulation as a whole is more sensitive to changes in vertical disparity. Consequently, it remains unclear what roles these cells play in the processing of stereoscopic depth and/or the generation of vertical and horizontal vergence responses. Therefore, the apparent discrepancy between monocular and binocular response properties in this subpopulation of cortical cells should not call into question the models of stereopsis (such as the model of Qian & Zhu, 1997) that compute disparity energy from neurons with coupled monocular and binocular orientation preferences.

Fig. 6 shows the responses of our 'extended' energy model for a stimulus with a horizontal position disparity, and for a stimulus that contains a constant phase disparity within a band of oblique stimulus orientations. The simulations were performed in Matlab using steps that were identical to those illustrated in Fig. 1 of Qian and Zhu (1997). Independent simulations were performed for three orientations (−75, 0 and 75 odeg) of the monocular RFs, with respect to the binocular stimulus. Both the unfiltered and filtered disparity maps are shown for a given spatial location. We used only a single spatial scale in our simulations, and the performance of the model is expected to improve if additional spatial scales are included (Qian & Zhu, 1997). When a stimulus with oblique phase disparities is presented to the model (see right half in Fig. 6), the complex cells that are tuned to the vertical stimulus orientation remain essentially inactive. The strongest activation occurs when the complex cells are tuned instead to an oblique orientation.


Stimulus with
Horizontal Position Disparity
Stimulus with
Oblique Phase Disparity
Gabor Orientation - (odeg)
-75
0
75
Raw
Filtered
Raw
Filtered
-5
5 arc-min

In contrast, the stimulus with a horizontal position disparity most strongly activates the complex cells in the model when these cells are tuned to a vertical orientation (Gabor orientation = 0; see the left half of Fig. 6). If we assume that the binocular mechanisms responsible for the perception of depth are tuned to stimulus orientation, then these simulations indicate that a mechanism tuned to oblique orientations is necessary to account for the perception of depth in our filtered RD stimuli.

It should be noted that the model that we used is based on complex cells with different monocular RFs at the same spatial location in the two eyes (i.e., phase-based computations). Although our modeling results indicate that phase-based computations are *sufficient* to explain our experimental results, they cannot distinguish between phase-based and position-based computations of horizontal disparity in the visual system.

Consequently, we constructed a simple variant of our 'extended' energy model that is similar to the position-based model proposed by Fleet et al. (1996, see their Fig. 3(b)). Binocular simple cells were constructed from pairs of monocular RFs located at different horizontal locations, but with the same RF phase of either 0 or 90 pdeg in the two eyes. Seven simple cells were constructed, using monocular cells with different RF positions in the right eye (−3 to +3 pixels), and a left-eye RF with a spatial phase of 0 pdeg as the reference. Seven more simple cells, constructed from left eye and right eye RFs with a monocular spatial phase of 90 pdeg, formed the quadrature pairs needed to build complex model cells (see Fig. 3(b) in Fleet et al., 1996). The resulting seven complex cells at each spatial location were tuned to position disparities distributed equally between −3 and 3 pixels. The output of the simulation was a two-dimensional map of position disparity, constructed from the position disparity labels of the most active complex cell at each spatial location. This model was identical to the phase-based model in most respects and performed similarly to the phase-based model when the input stimulus was a single spatial frequency sinusoidal grating (Fig. 7, top row). However, the simulation results of the position-based model were extremely poor when a broad-band random-dot stimulus was used as the input (Fig. 7, middle and bottom rows). The noisy response of the position-based model to the random-dot background at zero disparity is illustrative of its inferior performance compared to the phase-based model. The poor extraction of stimulus disparity by the position-based model, even when the orientation of the monocular RFs was vertical and the random-dot stimuli contained a coherent horizontal position disparity (Fig. 7, middle row), suggest that a general equivalence between phase-based and position-based energy models is difficult to establish. Whereas the phase-based and position-based models appear to be mathematically equivalent for narrow-band stimuli, Fleet et al. (1996) suggested that the performance of these models depends on a variety of stimulus attributes, including texture. It may be possible to achieve better performance from the position-based model by reducing the spatial-frequency bandwidth of the monocular RFs and/or by combining information from various spatial frequencies and orientations, but a more thorough evaluation and comparison of the two models is beyond the scope of this paper.

Most natural stimuli contain a broad range of spatial frequencies and orientations. Consequently, the horizontal position disparity between natural 3-D objects results also in substantial disparity information within non-vertical stimulus orientations. As discussed above, the horizontal disparity between objects can readily be recovered from disparity information in these non-vertical stimulus orientations. Indeed, a large proportion of physiologically identified disparity-tuned neurons are tuned to non-vertical stimulus orientations (Anzai et al., 1997; Maske, Yamane, & Bishop, 1986; Ohzawa & Freeman, 1986). These neurons are sensitive to interocular phase disparities for non-vertical stimulus components and can act as the detection stage for oblique disparities (Anzai et al., 1997). However, in order for the phase disparities sensed by obliquely tuned neurons to contribute to an accurate perception of depth, they have to be converted to signals that are consistent with horizontal position disparity. This conversion process must take into account the preferred spatial frequency and the preferred orientation of the obliquely tuned neuron. Presently, very little is known as to how the signals from neurons in early visual processing that are sensitive to position and phase disparity are combined to signal horizontal position disparity and eventually to represent stereoscopic depth (Cumming & Parker, 2000).

Fig. 6. Simulation results for an 'extended' disparity energy model based on phase-disparity computations for stereoscopic depth perception. The model was similar to that proposed by Qian and Zhu (1997, see Fig. 4 in their paper). The left column shows a random dot stimulus (100×100 pixel outer square, 31×31 pixel inner square) with a coherent horizontal *position* disparity (top) and the model's responses below. The right column shows a stimulus (100×100 pixel outer square, 31×31 pixel inner square) with a constant *phase* disparity (90 pdeg) in oblique orientations from 60 to 89.9 odeg and −60 to −89.9 odeg (top) and the model's responses below. The size of a pixel was 2 arc-min. The small square icons in the middle column illustrate the RFs of the monocular cells used in the simulations. The row below each stimulus pair illustrates simulation results within a two-dimensional disparity map, in which zero disparity is represented as medium gray and crossed and uncrossed disparities are represented by lighter and darker shades of gray, respectively. The raw disparity maps (first and third columns) are the equivalent position disparity signals that correspond to the preferred phase disparity of the maximally active complex cells. Each row of disparity maps corresponds to a different orientation of the monocular RFs that were used in the simulations, and consequently to a different preferred orientation of the complex cells in the model.

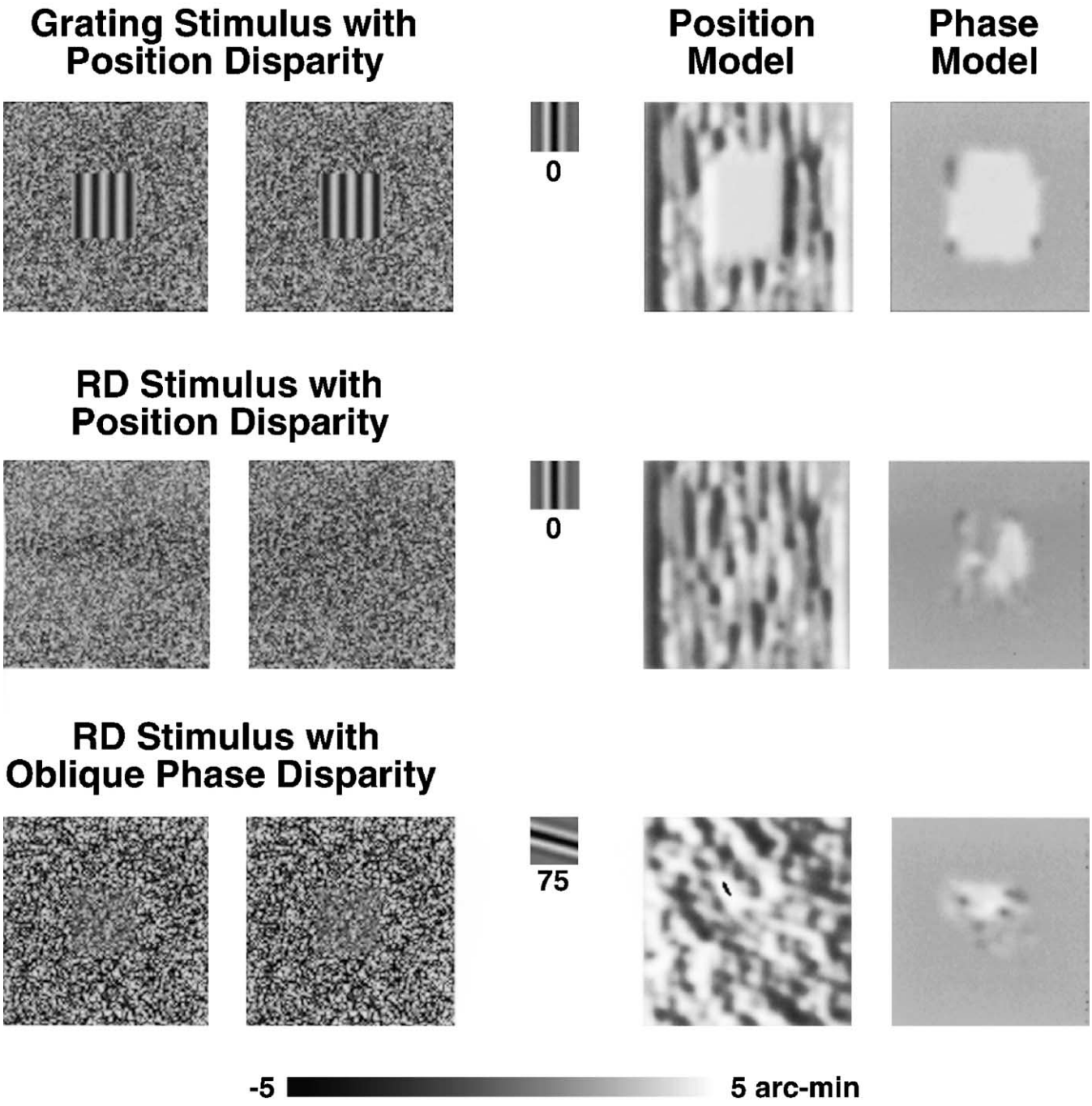


Fig. 7. Simulation results for an 'extended' disparity energy model based on position-disparity computations for stereoscopic depth perception. Simulation results are presented for three types of stimuli, shown in the two leftmost columns. The small square icons in the middle row represent the RF of the monocular cells used in the simulations. The two rightmost columns illustrate simulation results within a two-dimensional disparity map, in which zero disparity is represented as medium gray and crossed and uncrossed disparities are represented by lighter and darker shades of gray, respectively. The disparity maps in these two right columns indicate the position-or phase-disparity label of the maximally active complex cells, after the application of spatial filtering as in Fig. 6. The column labeled "Phase Model" shows the results of the same phase-based model that is used in Fig. 6. In addition, the stimuli with oblique disparities in the bottom row are the same as the stimuli shown in Fig. 6.

The mathematical formulation that computes the equivalent horizontal position disparity from a given oblique phase disparity is:

$$d_{\mathrm{p}} = \frac{\phi}{2\pi f \cos(\alpha)}, \qquad (1)$$

where, f is the preferred spatial frequency (in cpd), α is the preferred orientation (in odeg) and ϕ the phase disparity (in radians). At each spatial scale, the net horizontal position disparity may be obtained as a weighted average of the equivalent horizontal position

disparities from various orientations. A possible orientation weighting function for all spatial scales is computed from the average data in Fig. 4 (pooled for orientation bandwidths of 15 and 30 odeg) and is shown in Fig. 8. It can be seen clearly that beyond 60 odeg, the weighting function declines very sharply. Note that the mathematical formulation for the conversion of phase disparity to equivalent horizontal position disparity (Eq. (1)) is singular $\alpha = 90$ odeg. If the orientation weighting function decreases to zero faster than the cosine function beyond 60 odeg, then this orientation weighting will solve the problem of singularity. Such a weighting function also makes sense from an optimum signal-to-noise ratio standpoint. For a constant horizontal position disparity in a stimulus, the resulting phase disparities in various spatial frequency components of the stimulus decreases as the component's orientation increases from 0 (vertical grating component) to 90 odeg (horizontal grating component). Thus for a stereovision system in which phase disparity is considered a 'signal', for a particular horizontal position disparity, the strength of this 'signal' decreases as the preferred stimulus orientation of the phase disparity detector changes from vertical to horizontal. The orientation weighting function would therefore reduce the contribution of phase disparity signals that have low reliability towards the computation of horizontal disparity. Previous modeling results suggested that disparity detectors in the human brain may pool information over various orientations and spatial locations to improve the performance of the stereovision system (Fleet et al., 1996; Grossberg, 1994; Jones & Malik, 1992; Qian & Zhu, 1997; Simmons & Kingdom, 1995). Indeed, a weighted average across the spatial frequency components of the stimulus has been proposed to account for the upper limit of perceived apparent motion (Bischof & Di Lollo, 1991). If the equivalent horizontal position disparity signals are combined across various orientations by the statistical rules of averaging, then the final estimate of horizontal disparity should have substantially lower noise than an estimate based only on the position disparity signals from vertically oriented mechanisms.

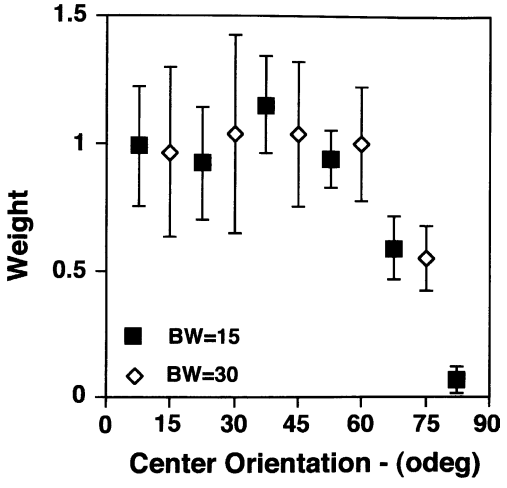


Fig. 8. Estimated weights for the averaging of equivalent horizontal position disparities as a function of the orientation of the oblique phase disparities. Because the phase disparity within each orientation band in experiment 1 was held constant at 90 pdeg, the orientation weighting function (±1 SE) was computed by dividing the average matched position disparity for stimuli with 15 and 30 odeg bandwidth by the fitted inverse cosine functions in Fig. 3.

In summary, our experimental results indicate that phase disparities at oblique orientations contribute significantly toward determining the magnitude of perceived stereoscopic depth. An energy model consisting of binocular mechanisms that are tuned only to vertically oriented stimuli is insensitive to oblique disparities, and therefore cannot account for the perception of depth in our stimuli. Because our stimuli are constructed from featureless random dots, and because the perception of depth from oblique phase disparities is consistent with the averaging of horizontal position disparity across many orientations and spatial frequencies, our results are consistent with the presence of a low-level mechanism that converts oblique phase disparities to equivalent horizontal position disparities. The perception of approximately planar depth from stereo displays that do not contain a coherent position disparity is also consistent with the averaging of disparity information across many stimulus orientations and spatial frequencies prior to the computation of depth. The use of non-horizontal as well as horizontal information for the perception of depth would be expected to improve substantially the signal-to-noise ratio of the human stereovision system.

Acknowledgements

This work was supported by University of Houston Institute for Space Systems Operation, and research grants R01 EY-10531 and R01 EY-05068 from the National Eye Institute. We would like to thank the reviewers for making valuable suggestions to improve this paper.

References

Anzai, A., Ohzawa, I., & Freeman, R. D. (1997). Neural mechanisms underlying binocular fusion and stereopsis: position vs. phase. *Proceedings of the National Academy of Sciences, 94*, 5438–5443.

Arditi, A., Kaufman, L., & Movshon, J. A. (1981). A simple explanation of the induced size effect. *Vision Research, 21*, 755–764.

Barlow, H. B., Blakemore, C., & Pettigrew, J. D. (1967). The neural mechanism of binocular depth discrimination. *Journal of Physiology, 193*, 327–342.

Bischof, W. F., & Di Lollo, V. (1991). On the half-cycle displacement limit of sampled directional motion. *Vision Research, 31*, 649–660.

Blake, R., Camisa, J. M., & Antoinetti, D. N. (1976). Binocular depth discrimination depends on orientation. *Perception and Psychophysics, 20*, 113–118.

Boothroyd, K., & Blake, R. (1984). Stereopsis from disparity of complex grating patterns. *Vision Research, 24*, 1205–1222.

Burkhalter, A., & Van Essen, D. C. (1986). Processing of color, form and disparity information in visual areas VP and V2 of ventral extrastriate cortex in macaque monkey. *Journal of Neuroscience, 6*, 2327–2351.

Cumming, B. G. (2002). An unexpected specialization for horizontal disparity in primate visual cortex. *Nature, 418*, 633–665.

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

Ebenholtz, S. M., & Walchli, R. M. (1965). Stereoscopic thresholds as a function of head- and object-orientation. *Vision Research, 5*, 455–461.

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

Felleman, D. J., & Van Essen, D. C. (1987). Receptive field properties of neurons in area V3 of macaque monkey extrastriate cortex. *Journal of Neurophysiology, 57*, 889–920.

Fleet, D. J., Wagner, H., & Heeger, D. J. (1996). Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision Research, 36*, 1839–1859.

Gonzalez, F., Krause, F., Perez, R., Alonso, J. M., & Acuna, C. (1993). Binocular matching in monkey visual cortex: single cell responses to correlated and uncorrelated dynamic random dot stereograms. *Neuroscience, 52*, 933–939.

Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception and Psychophysics, 55*, 48–121.

Hess, R. F., Liu, C. H., & Wang, Y. (2002). Luminance spatial scale and local stereo-sensitivity. *Vision Research, 42*, 331–342.

Howard, I. P., Allison, R. S., & Zacher, J. E. (1997). The dynamics of vertical vergence. *Experimental Brain Research, 116*, 153–159.

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

Hubel, D. H., & Livingstone, M. S. (1987). Segregation of form, color and stereopsis in primate area 18. *Journal of Neuroscience, 7*, 3378–3415.

Hubel, D. H., & Wiesel, T. N. (1970). Stereoscopic vision in macaque monkey. *Nature, 225*, 41–42.

Jones, D. G., & Malik, J. (1992). Computational framework for determining stereo correspondence from a set of linear spatial filters. *Image and Vision Computing, 10*, 699–708.

Mansfield, J. S., & Parker, A. J. (1993). An orientation-tuned component in the contrast masking of stereopsis. *Vision Research, 33*, 1535–1544.

Maske, R., Yamane, S., & Bishop, P. O. (1986). Stereoscopic mechanisms: binocular responses of the striate cells of cats. *Proceedings of the Royal Society of London B, 229*, 227–256.

Maunsell, J. H. R., & 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. *Journal of Neurophysiology, 49*, 1148–1167.

Morgan, M. J., & Castet, E. (1997). The aperture problem in stereopsis. *Vision Research, 37*, 2737–2744.

Ohzawa, I., & Freeman, R. D. (1986). The binocular organization of complex cells in the cat's visual cortex. *Journal of Neurophysiology, 56*, 221–242.

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

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

Qian, N., & Zhu, Y. (1997). Physiological computation of binocular disparity. *Vision Research, 37*, 1811–1827.

Remole, A., Code, S. M., Matyas, C. E., & Mcleod, M. A. (1992). Multimeridional apparent frontoparallel plane: relation between stimulus orientation angle and compensating tilt angle. *Optometry and Vision Science, 69*, 544–549.

Rohaly, A. M., & Wilson, H. R. (1994). Disparity averaging across spatial scales. *Vision Research, 34*, 1315–1325.

Simmons, D. R., & Kingdom, F. A. A. (1995). Differences between stereopsis with isoluminant and isochromatic stimuli. *Journal of the Optical Society of America A, 12*, 2094–2104.

Stevenson, S. B., & Schor, C. M. (1997). Human stereo matching is not restricted to epipolar lines. *Vision Research, 37*, 2717–2723.

van Ee, R., & Anderson, B. L. (2001). Motion direction, speed and orientation in binocular matching. *Nature, 410*, 690–694.

Wheatstone, C. (1838). Contributions to the physiology of vision—Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London Series B, 2*, 371–393.