

## Orientation Disparity: A Cue for 3D Orientation?

**Hal S. Greenwald**

*hal.greenwald@gmail.com*

**David C. Knill**

*knill@cvs.rochester.edu*

*Center for Visual Science, University of Rochester, Rochester, NY 14627, U.S.A.*

**Orientation disparity, the difference in orientation that results when a texture element on a slanted surface is projected to the two eyes, has been proposed as a binocular cue for 3D orientation. Since orientation disparity is confounded with position disparity, neither behavioral nor neurophysiological experiments have successfully isolated its contribution to slant estimates or established whether the visual system uses it. Using a modified disparity energy model, we simulated a population of binocular visual cortical neurons tuned to orientation disparity and measured the amount of Fisher information contained in the activity patterns. We evaluated the potential contribution of orientation disparity to 3D orientation estimation and delimited the stimulus conditions under which it is a reliable cue. Our results suggest that orientation disparity is an efficient source of information about 3D orientation and that it is plausible that the visual system could have mechanisms that are sensitive to it. Although orientation disparity is neither necessary nor sufficient for estimating slant, it appears that it could be useful when combined with estimates from position disparity gradients and monocular perspective cues.**

### 1 Introduction ---

Whether placing an object on a table or walking up a hill, humans rely on binocular information to estimate the orientations of surfaces so they can interact with and navigate through the environment. Binocular cues, which arise from the slight differences in the views from the two eyes, are important for estimating three-dimensional (3D) surface orientation, and there are multiple aspects of binocular input that could be useful for producing these estimates. The first, position disparity, is the difference in the retinal positions at which corresponding image features appear in the two eyes (this is what people typically imagine when they think about binocular disparity). Position disparity provides information about depth, and the gradient of position disparity across a surface provides information about its 3D orientation. Orientation disparity, the difference in the orientations at which a texture element on a slanted surface appears when projected to

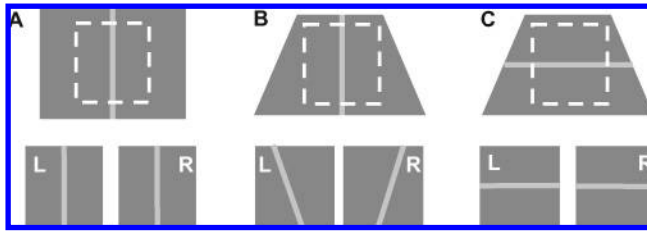


Figure 1: Binocular views of a line on a slanted surface as projected to each eye for (A) a frontoparallel surface and slanted surfaces with the line (B) perpendicular to the axis of rotation and (C) parallel to the axis of rotation. Surface *B* is the only surface of the three with nonzero orientation disparity.

both eyes, has been proposed as an additional binocular cue that may be useful for estimating 3D orientation. When viewing a frontoparallel surface containing a single vertical line, the projected line appears vertical to both eyes. However, if the surface is rotated away from the viewer about its horizontal axis so that the top edge is farther away than the bottom edge, the line will appear to rotate counterclockwise in the left eye and clockwise in the right eye. The orientations at which the line appears when projected to each eye change as a function of slant, and the difference between these orientations is orientation disparity (see Figure 1). In general, the mapping between the images projected to the two eyes is an affine transformation (Koenderink & van Doorn, 1976), and the orientation differences reflect one component of this deformation.

There is disagreement in the literature about whether the visual system has mechanisms that are tuned to orientation disparity. The reason this has been so difficult to resolve is that orientation disparity and position disparity are confounded (Bridge & Cumming, 2001). von der Heydt, Hänni, and Dürsteler (1981) claimed that uncorrelated dichoptic stimuli with consistent orientation disparity information induced the percept of slant; they used this to argue in favor of orientation disparity being an independent slant cue and that effects due to orientation disparity and position disparity gradients could be separated. However, these data have never been published in a peer-reviewed journal, and to our knowledge, the effect has not been replicated. Psychophysical studies of orientation disparity have had conflicting interpretations (Gillam & Rogers, 1991; Gillam & Ryan, 1992; Cagenello & Rogers, 1993; Heeley, Scott-Brown, Reid, & Maitland, 2003). In general, these studies have found evidence that humans can perceive slant from binocular images related by geometric distortions that are consistent with orientation disparity and that sensitivity to slanted surfaces depends on the orientations of the texture elements. However, the stimuli in these experiments also contained position disparity information and often monocular

perspective information. Also, they primarily tested at only very low slants, where orientation disparities are smallest and vary the least across different slants. Heeley et al. (2003) tested slant thresholds using bandpass stimuli with varying orientation bandwidths and found that performance for stimuli with 180 degree bandwidths was no worse than for narrower bandwidths. They claimed that since stimuli that are broadband in orientation lack a clear directional structure, slant estimates from these stimuli must have been based entirely on position disparity. However, as we will show, because these stimuli have orientation content that is correlated between the left and right retinal images, they produce orientation disparities when slanted.

Two additional psychophysical studies found indirect evidence that pointed toward orientation disparity being a useful binocular cue. Ninio (1985) created stereograms with slight biases added to the position disparities and orientation disparities at the tips of small slanted needles in an attempt to dissociate position and orientation disparity and found that slant percepts were stronger when orientation disparity was consistent with the slant. Adams and Mamassian (2002) attempted to manipulate the responses of monocular orientation-dependent mechanisms that would support binocular mechanisms tuned to orientation disparity by having subjects adapt to Gabor patterns in each eye that were oriented  $\pm 6$  degrees apart. The adaptation produced significant decreases in slant perception, which suggested that monocular orientation sensitivity is a basis for slant perception, but it is also possible that other mechanisms for slant also adapted. Blakemore, Fiorentini, and Maffei (1972) found neurons in the cat striate cortex that responded well to binocular stimulation using small differences in the orientation of a line presented to both eyes and concluded that orientation disparity was a binocular mechanism. Bridge and Cumming (2001) performed a physiological study of macaque V1 in which they also found binocular neurons that appeared to be sensitive to orientation disparity, but they claimed that this resulted from offsets between the stimuli and the receptive fields of these cells. It has also been impossible to disambiguate whether the computations by parietal neurons tuned to 3D orientation are based on position disparity gradients or orientation disparity (Taira, Tsutsui, Jiang, Yara, & Sakata, 2000). Consequently, whether the visual system has mechanisms for 3D orientation that are tuned to orientation disparity remains an open question.

Most computational models of disparity detection have dealt only with position disparity: the difference in the retinal positions at which corresponding features appear in binocular image pairs (Ohzawa, DeAngelis, & Freeman, 1990; Lippert & Wagner, 2002; Read, 2002; Chen & Qian, 2004). Typically surface curvature and slant have been derived from approximating surfaces with small frontoparallel patches and computing position disparity gradients, a method that seems consistent with physiological

and behavioral measures (Nienborg, Bridge, Parker, & Cumming, 2004). Most models have not accounted for local distortions like rotation and compression (Koenderink & van Doorn, 1976). An exception to this is Jones and Malik (1992), in which the authors used a set of linear filters tuned to different orientations and scales to estimate surface slant from orientation disparity, but their algorithm was not limited by biological constraints.

We simulated the activity of a population of visual cortical cells tuned to orientation disparity in response to textured surfaces presented at different 3D orientations and analyzed the output to quantify the amount of Fisher information present as a function of slant, local orientation content, and axis of rotation. Our primary question was whether orientation disparity provides sufficient information about surface slant for the visual system to have mechanisms that are sensitive to it or whether there is too little information present for it to be an efficient cue. While this did not address whether humans use orientation disparity, which is difficult due to confounds between position disparity and orientation disparity, it allowed us to sketch out the range of slants at which orientation disparity might be useful based on the information it provides.

One way to evaluate the performance of our model is to compare it with humans' measured slant thresholds. If the model performs similarly to humans, then this would suggest that orientation disparity could be a useful cue for 3D orientation. However, very few studies have examined how well humans can discriminate slant from binocular disparity. Knill and Saunders (2003) reported that binocular 75% slant thresholds decreased from around 15 degrees near frontoparallel to 10 degrees for targets slanted at 40 degrees and to about 6.5 degrees for targets slanted at 70 degrees. These were probably overestimates since Greenwald and Knill (2009) found that mean 84% binocular slant thresholds around a base slant of 35 degrees were less than 5 degrees. Hillis, Watt, Landy, and Banks (2004) also found that one author's estimated thresholds decreased with increasing slant when stimuli were slanted about the vertical axis. Their slant thresholds for surfaces within  $\pm 45$  degrees of frontoparallel were within the 5 to 10 degree range at a viewing distance similar to what we used. There are several reasons that psychophysical measurements of slant estimation are not ideal benchmarks for the performance of our theoretical model. First, stimulus properties such as size can have an impact on thresholds, and the image patches presented to our model as input subtended about 1.2 degrees of visual angle. Larger stimuli allow spatial integration, which can reduce uncertainty. A more significant concern is that great care is typically taken to eliminate perspective cues when measuring binocular slant thresholds (e.g., using random dot stimuli), and orientation disparity is a binocular cue that depends on perspective distortions. Nevertheless, these studies provide at least ballpark estimates for how much information orientation disparity would need to contribute to be an effective cue.

## 2 Properties and Limitations of Orientation Disparity

To what extent and for which slants is orientation disparity theoretically useful? If orientation disparity is an effective cue for 3D orientation, its values should follow a broad distribution in natural, everyday scenes and vary across different slants at levels that are within the human visual system's ability to resolve the differences. We describe how the geometric properties of the scene and local texture properties influence orientation disparity, review published estimates of human orientation acuity, consider how the disparity gradient limit might impose bounds on the slants at which orientation disparity could be useful, and estimate the distribution of orientation disparity in natural environments.

The magnitude of orientation disparity depends on the geometry specified by the viewing distance, the interpupillary distance, and the slant, tilt, and spin of the surface being viewed. Slant is the angle between the surface normal and the line of sight and specifies the rotation about the tilt axis. An object with 0 degrees of slant would be frontoparallel relative to the viewer. Tilt determines the axis about which the surface is slanted, and it is the angle between the projection of the surface normal into the image plane and a fixed vector lying in the surface plane (Witkin, 1981; Stevens, 1983). By convention, a surface with a tilt of 0 degrees rotates counterclockwise about the vertical axis when slanted. Finally, spin specifies the rotation of the surface about its normal; a horizontal grating would have a spin of 0 degrees, and a vertical grating would have a spin of 90 degrees. Since lines oriented 180 degrees apart are identical, orientation disparity ranges only between  $-90$  degrees and  $90$  degrees, and values wrap around to the other side once they exceed this range. The equation for orientation disparity (OD) is

$$OD = \tan^{-1} \left( \frac{\cos \phi \sin^2(\frac{\alpha}{2}) \sin 2\tau - \sin \phi (\cos^2 \tau + \cos \alpha \sin^2 \tau)}{\cos(\tau - \phi) \sin \frac{v}{2} \sin \alpha - \cos \frac{v}{2} (\cos \alpha \cos \tau \cos(\tau - \phi) + \sin \tau \sin(\tau - \phi))} \right) \\ - \tan^{-1} \left( \frac{-2 \cos \tau \cos \phi \sin^2(\frac{\alpha}{2}) \sin \tau + \cos^2 \tau \sin \phi + \cos \alpha \sin^2 \tau \sin \phi}{\cos(\tau - \phi) \sin \frac{v}{2} \sin \alpha + \cos \frac{v}{2} (\cos \alpha \cos \tau \cos(\tau - \phi) + \sin \tau \sin(\tau - \phi))} \right), \quad (2.1)$$

where  $\alpha$  is the slant,  $\tau$  is the tilt,  $\phi$  is the spin, and  $v$  is the vergence angle (see the supplementary material for the derivation, available online at <http://www.mitpressjournals.org/doi/suppl/10.1162/neco.2009.08-08-848>). The vergence angle depends on the interpupillary distance (PD) and viewing distance ( $d$ ) and is specified by

$$v = 2 \tan^{-1} \left( \frac{PD}{2d} \right). \quad (2.2)$$

Orientation disparity values increase as the vergence angle grows (when surfaces are nearer or the eyes are farther apart); thus, typical magnitudes

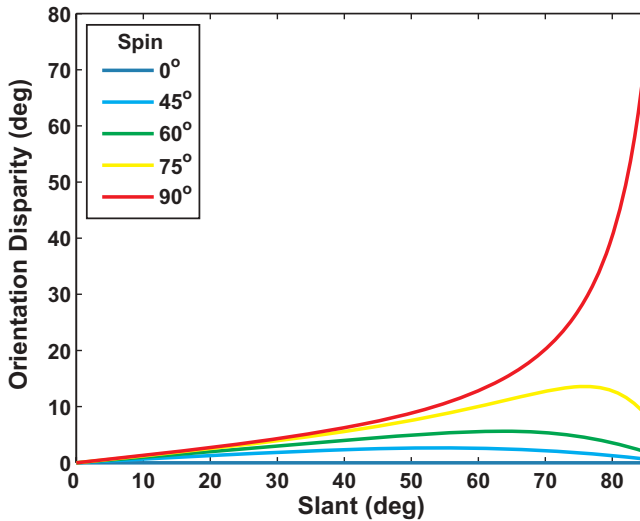


Figure 2: Orientation disparity as a function of line orientation for surfaces rotated about the horizontal axis (tilt = 90 degrees). A horizontal line would have a spin of 0 degrees, and a vertical line would have a spin of 90 degrees. A line oriented at 0 degrees produces 0 degrees of orientation disparity at all slants. Orientation disparities for surfaces slanted about the horizontal axis are symmetric for spins relative to 90 degrees.

vary across viewing situations and individuals. We always assumed a viewing distance of 50 cm and an interpupillary distance of 65 mm, which is within the normal range for humans.

Figure 2 shows how orientation disparity varies as a function of slant and spin for a surface slanted about the horizontal axis. For frontoparallel surfaces, orientation disparity is always 0 degrees regardless of tilt or spin, and it generally increases with increasing surface slant. Orientation disparity is highly dependent on the local texture orientation relative to the axis of rotation specified by tilt, and texture elements that are parallel to the axis of rotation or horizontal relative to the eyes never result in any orientation disparities since the rotations do not cause any perspective distortions in the images. For a surface textured with vertical lines and rotated about the horizontal, orientation disparity initially grows gradually with slant and accelerates its growth as slant increases; slants of 60, 70, and 85 degrees produce orientation disparities of 12.8, 20.2, and 73.1 degrees, respectively. In general, one would expect orientation disparity to provide better information at higher slants because magnitudes are larger and change more rapidly as slant varies. As the texture elements are oriented away from 90 degrees (vertical), orientation disparity magnitudes are not as large, do

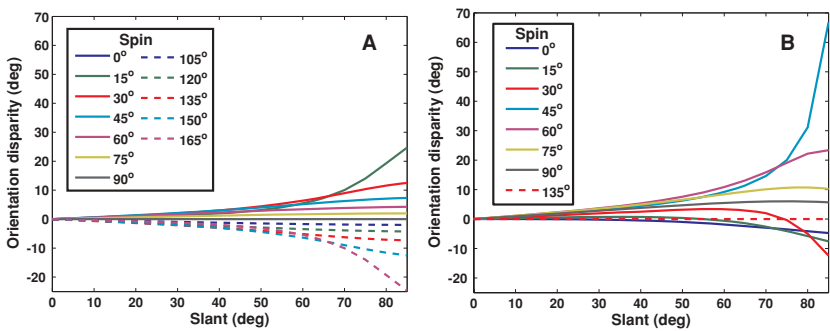


Figure 3: Orientation disparity for line elements on surfaces as a function of slant, spin, and two different tilts. (A) Tilt = 0 degree. The axis of rotation for slant is at 90 degrees, and texture elements oriented at 0 and 90 degrees always produce 0 degrees of orientation disparity. (B) Tilt = 45 degrees. The axis of rotation for slant is at 135 degrees. Texture elements oriented at 135 degrees always produce orientation disparities of 0 degree.

not change as quickly, and no longer increase monotonically with slant (at least for slants up to 85 degrees), although larger slants are still associated with relatively larger orientation disparities. For lines oriented at 45 or 135 degrees, the maximum orientation disparity is only about 2.6 degrees, and the distribution of values starts to become too narrow to support good slant estimates. Figure 3 shows how orientation disparity values change as a function of slant and spin (texture element orientation) for tilts of 0 and 45 degrees. While the maximum orientation disparity magnitudes are not as large at these tilts as for rotation about the horizontal axis (see Figure 2), there is a larger range of spins that produce relatively large orientation disparities. At a tilt of 90 degrees, only spins within about 15 degrees of vertical produce orientation disparities above 10 degrees, whereas the range of spins that exceed this value for tilts of 0 and 45 degrees is larger. When neither slant nor tilt is known a priori, slant estimation from orientation disparity suffers from an ambiguity that is similar to the aperture problem for motion perception first described by Stumpf in 1911 (see Todorovic, 1996) because any pattern of orientation disparities may be consistent with multiple combinations of slant and tilt. Once the tilt is known based on estimates from other information present in the stimuli (such as from the surface contours or position disparity), the slant can be ascertained using orientation disparity. In all of our simulations, we assumed that the tilt was known.

We estimated the expected distribution of orientation disparities in natural scenes by simulating surfaces with random slants, tilts, and spins and computing the orientation disparity that would result when a line on each surface is projected to both eyes. For orientation disparity to be useful for estimating slant, it must produce nonzero values, and its distribution

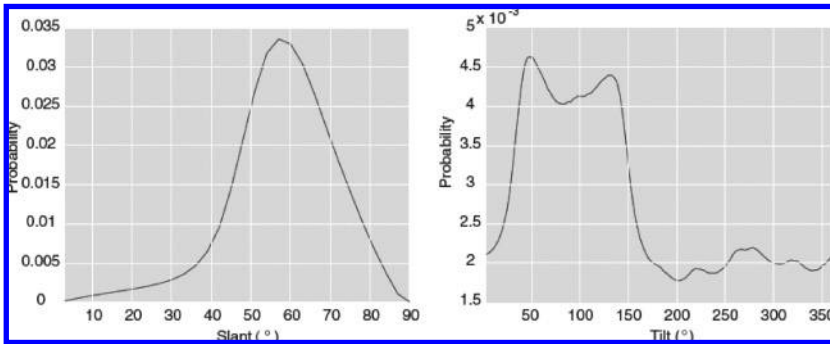


Figure 4: Probability distributions of local slant (left) and tilt (right) in natural scenes. (Reproduced with permission from Yang & Purves, 2003.)

should be sufficiently broad to allow a reasonable mapping between slant and orientation disparity. In contrast, if a wide range of 3D orientations produced a narrow range of orientation disparities, it would suggest that orientation disparity is not a good cue because the variance of the estimates based on it would be large due to the high degree of overlap across different slants. The 250,000 simulated surfaces with which we approximated the distribution of orientation disparity were represented by a vector that was oriented in depth according to a random combination of slant, tilt, and spin. Slant and tilt were chosen independently according to empirical probability distributions of local surface orientations in natural scenes (Yang & Purves, 2003; see Figure 4), and spin was selected from a uniform distribution. The simulated surfaces were positioned 50 cm from the virtual viewer and were projected binocularly under parallel projection using an interpupillary distance of 6.5 cm. Figure 5 shows a scatter plot based on this simulation that represents the distribution of orientation disparity as a function of the orientation in the left eye. Overall, we obtained samples across the full 180 degrees range of orientation disparities, and there was underlying structure within the data that reflected different parameter values. The mean orientation disparity magnitude was 3.8 degrees with a standard deviation of 4.1 degrees, and only about 25% of the simulated surfaces produced orientation disparity magnitudes of less than 1 degree. The data suggest that orientation disparities of up to about 20 degrees in either direction commonly occur in natural scenes and that the distribution of values is sufficiently large for orientation disparity to be a plausible cue for 3D orientation.

Human two-dimensional (2D) orientation acuity may introduce a limit on the resolution at which slant can be estimated from orientation disparity, particularly at low slants, where orientation disparity magnitudes are small. Psychophysical estimates of orientation discrimination thresholds vary, but the limit appears to be 0.5 to 1.0 degrees. Published 2D orientation



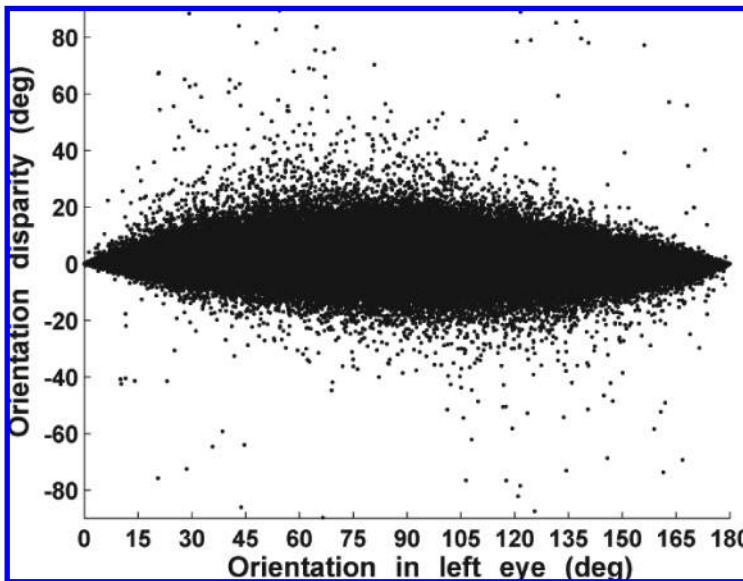


Figure 5: Scatter plot of orientation disparity as a function of orientation in the left eye for simulated surfaces with random slants, tilts, and spins chosen from distributions that occur in natural scenes. Orientations in the left eye of 0/180 degrees were horizontal, and orientations of 90 degrees were vertical.

acuity levels differ as a function of stimulus size (Orban, Vandenbussche, & Vogels, 1984; Henrie & Shapley, 2001; Sally & Gurnsey, 2004), contrast (Salley & Gurnsey, 2004), and orientation (Heeley et al., 2003). Large, high-contrast stimuli improve orientation discrimination, and performance is better at horizontal and vertical orientations than at oblique orientations. For approximately frontoparallel surfaces rotated about the horizontal axis, an orientation discrimination threshold of 1 degree translates to a maximum resolution of about 8 degrees of slant using orientation disparity, which is similar to measured thresholds for frontal surfaces defined by monocular and binocular cues (Knill & Saunders, 2003). For a surface slanted 70 degrees away from the viewer, a 1 degree orientation discrimination threshold should permit slant estimates from orientation disparity with 1 degree resolution, which is better than measured human slant thresholds. Although orientation acuity may limit the resolution of slant judgments from orientation disparity, the bounds it could impose would permit slant estimates that meet or exceed normal human performance. Also, orientation acuity would be a limiting factor only if neurons that encode orientation disparity depend on explicit 2D orientation estimates; other possible algorithms, including those based on binocular image correlations like the disparity energy model described in section 3.1, might not be subject to this constraint.

The disparity gradient limit could also restrict the utility of orientation disparity. The disparity gradient is the ratio of the binocular position disparity difference between two points to the magnitude of the visual angle separating them. Burt and Julesz (1980) found that once this ratio surpassed approximately 1.0 (when disparity exceeded angular separation), subjects could not fuse the images from the two eyes. The actual limit could exceed 1.0 and may vary as a function of the stimulus properties (Prazdny, 1985), but it applies to both random dot and line stereograms (Tyler, 1974). Assuming that binocular fusion is necessary for accurate stereoscopic estimates, the disparity gradient limit places an upper bound on the slants at which any binocular information, including orientation disparity, is useful. A potential explanation for why the disparity gradient limit restricts binocular vision is that a high disparity gradient indicates that the monocular views are related by a large horizontal distortion, which reduces the correlations between the retinal images (Banks, Gepshtein, & Landy, 2004). At a viewing distance of 50 cm and a tilt of 90 degrees, the disparity gradient exceeds 1.0 only for slants above 82.6 degrees; therefore, the disparity gradient limit affects only the highest slants.

### 3 Method

---

To quantify the amount of information provided by orientation disparity for estimating 3D orientation, we built a biologically plausible model of a population of visual cortical neurons tuned to orientation disparity and analyzed the encoded activity patterns across a variety of slants, tilts, spins, and texture classes. Since the confound between position disparity and orientation disparity appears to make direct psychophysical examination of orientation disparity impossible, we developed a model that uses low-level mechanisms that are believed to be present in the primary visual cortex to create a slant detector specifically tuned to different orientations in the two eyes; this enabled us to estimate how well humans might perform in a behavioral test using only orientation disparity. First, we generated slanted surfaces with randomly generated textures and produced binocular image pairs from these using perspective projection. Then we presented these image pairs as input to the model, which simulated binocular processing in the primary visual cortex and produced outputs for individual units tuned to different combinations of preferred 2D orientations in the two eyes. Finally, we estimated the level of Fisher information provided by the activity patterns of the population of units and compared the sensitivity of the model to threshold estimates for normal slant perception. All simulations were performed using Matlab, and we used a PC computing cluster for most of the computations.

**3.1 Model.** The model we used to process the binocular image pairs was a standard disparity energy model (Ohzawa et al., 1990; Bridge, Cumming,

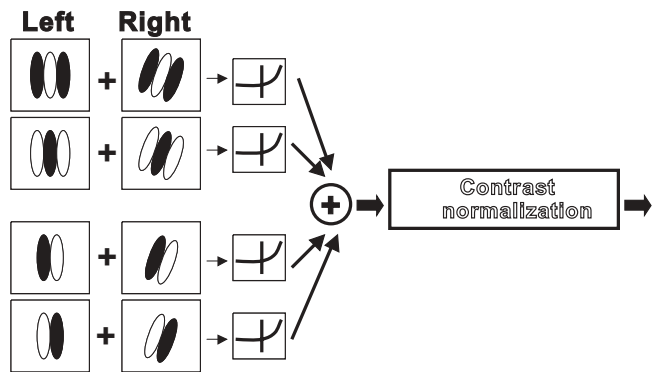


Figure 6: Disparity energy model tuned to orientation differences in the two eyes. This schematic shows the biologically plausible squaring operation. See the text for details.

& Parker, 2001) modified to be sensitive to binocular orientation differences instead of the position or phase differences typically used in models tuned to position disparity (see Figure 6). The model first computed the outputs of linear, monocular subunits centered on the fixation point in response to binocular image pairs. These receptive fields were constructed from Gabor filters, which are sinusoids convolved with gaussian functions. They were  $20 \times 20$  pixels in size and had a spatial frequency bandwidth of 1 octave and an orientation bandwidth of 28 degrees. Pixel sizes used in our model were based on a display with  $1024 \times 768$  pixel resolution ( $3779.5$  pixels =  $1$  m). Half of the units were tuned to a preferred spatial frequency of  $0.1$  cycles per pixel, which matched the kernels used to generate the textured stimuli, and a second set of units was sensitive to a spatial frequency range 1 octave higher to capture the higher spatial frequencies that occurred at larger slants due to texture compression. Each eye had receptive fields with phases that were  $90$  degrees apart (in quadrature), which allowed the model to be sensitive to both oriented bars (the “even” receptive fields shown on top in Figure 6) and oriented edges (the “odd” receptive fields on bottom). The receptive field responses encoded contrast polarity using signed outputs,<sup>1</sup> and these outputs from the left and right eyes were summed and squared (see equation 3.1). This created binocular units that were sensitive to particular orientation disparities specified by differences between the preferred orientations of the monocular receptive fields.

<sup>1</sup>Since neurons cannot have negative firing rates, different cells must respond to opposite contrast polarities. A biologically plausible implementation of the squaring operation would be to half-wave-rectify the outputs of receptive fields with opposite preferred contrast polarities, sum them, and square them. Mathematically, our approach was equivalent.

The model's final step was to apply contrast normalization to each unit by dividing the total computed binocular energy by the sum of the mean squared outputs of all of the monocular receptive fields over all preferred orientations. The contrast normalization removed the effects of local contrast variations across different images and ensured that the responses primarily reflected orientation content. It also helped isolate the correlations in the left and right images. The final output of the model was a measure of orientation disparity "energy" for each unit that should correlate with the firing rates of binocular visual cortical neurons with similar tuning properties. We computed energy using the following equation:

$$Energy(\theta_L, \theta_R) = \frac{(L_E(\theta_L) + R_E(\theta_R))^2 + (L_O(\theta_L) + R_O(\theta_R))^2}{\langle L_E(\theta)^2 \rangle + \langle L_O(\theta)^2 \rangle + \langle R_E(\theta)^2 \rangle + \langle R_O(\theta)^2 \rangle}. \quad (3.1)$$

$L_E(\theta_L)$  is the response of the left monocular, even-symmetric subunit tuned to orientation  $\theta_L$ , and  $R_O(\theta_R)$  is the response of the right monocular, odd-symmetric subunit tuned to orientation  $\theta_R$ . The terms in the denominator within the angle brackets represent averaged responses across all sampled preferred orientations.

The receptive fields receiving input from the left eye were tuned to orientations from 0 to 170 degrees in 10 degree increments, and we varied the preferred orientation in the right eye in 5 degree increments relative to the preferred orientation in the corresponding left receptive field to create preferred orientation disparities of up to 80 degrees. This resulted in 458 combinations of preferred orientations in the two eyes at the two different spatial frequency bands for a total of 916 units. This method of sampling orientation space was not biologically grounded but allowed us to sample a wide range of 2D orientations and orientation disparities with a manageable number of simulated units.

For a control condition, we modified the orientation disparity energy model to use only monocular input. This allowed us to separate the performance of the binocular model that was due to orientation disparity from performance due to unintended correlations between slant and local texture orientation or spatial frequency. The monocular energy model used receptive fields from the left eye at the same 18 preferred orientations as before (0–170 degrees in 10 degree increments) and at the same two preferred spatial frequencies for a total of 36 units. Both the monocular and binocular models used the same textures as input. The equation for monocular energy is provided below; the nomenclature is the same as for the previous equation:

$$Energy(\theta_L) = \frac{L_E(\theta_L)^2 + L_O(\theta_L)^2}{\langle L_E(\theta)^2 \rangle + \langle L_O(\theta)^2 \rangle}. \quad (3.2)$$

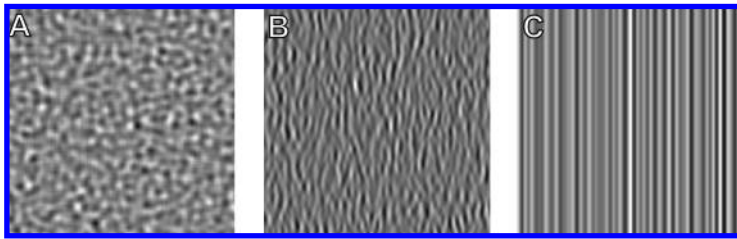


Figure 7: Examples of (A) broadband noise, (B) bandpass noise, and (C) oriented grating textures that were used as stimuli.

To increase the biological plausibility of our results, we corrupted the outputs of both models with simulated zero-mean gaussian neural noise. Current theories suggest that neuronal activity is subject to noise that is proportional to the firing rate of each neuron. Gur and Snodderly (2006) found that the median Fano factor, the ratio of the variance of the noise to the mean neural activity, in the primary visual cortex of awake, behaving monkeys was approximately 0.3, so we ran simulations using Fano factors of 0, 0.15, and 0.3. We selected the variance parameters for the internal noise distributions independently for each unit and for every sample based on the resulting energy levels.

**3.2 Stimuli.** The stimuli used as input to our model were textured planes slanted about their horizontal axis (tilt = 90 degrees) and positioned 50 cm in front of a virtual viewer with an interpupillary distance of 65 mm. The surfaces were sufficiently large so that  $40 \times 50$  pixel (10.6 mm  $\times$  13.2 mm) regions could be removed from the centers of the left and right images after the surface was spun, slanted, and projected to each eye under perspective projection without including any borders. They were slanted by up to 85 degrees in either direction away from frontoparallel. Performance was symmetric about frontoparallel, but including both directions enabled us to show that orientation disparity can be used to estimate slant over the full 180 degree range. The model's receptive fields for both eyes were positioned on the center of the stimuli, where position disparities were the smallest. The resulting binocular image pairs were presented as input to a binocular disparity energy model tuned to orientation disparity, as described above.

We used three types of textures to study how the orientation content of different textures modulates the reliability of orientation disparity. Our prediction was that the amount of information provided by orientation disparity might vary depending on the spectral properties of the surface texture and that textures composed of line elements would result in better performance than random dots. The broadband textures (see Figure 7A)

contained a uniform distribution across all 2D orientations and were generated by convolving gaussian noise ( $\sim N(0,1)$ ) with a two-dimensional difference-of-gaussians filter. We created the bandpass textures (see Figure 7B) by convolving gaussian noise with a two-dimensional Gabor filter, and they contained a dominant central orientation with an orientation bandwidth of about 60 degrees. Textures in the third class, oriented gratings (see Figure 7C), contained only a single orientation. To generate these, we filtered a one-dimensional gaussian noise vector with a one-dimensional Gabor function and duplicated the result to produce the rows, which varied randomly in luminance. Although the texture classes differed in orientation content, their spatial frequency content was otherwise identical. All of the textures had a central spatial frequency of 0.1 cycles per pixel (3.3 c/deg) and a bandwidth of 1 octave relative to this frequency.

**3.3 Analysis of Fisher Information.** To quantify the reliability of 3D orientation information from orientation disparity, we used an approach based on linear discriminant analysis (Duda & Hart, 1973). Linear discriminant analysis is primarily a classification technique; one finds a hyperplane within the multidimensional data space that best separates two classes, and samples to be classified are projected onto a vector that is normal to that hyperplane. Any samples that fall on one side of the discriminant hyperplane are classified as belonging to one category, and samples that fall on the other side are classified as belonging to the other category. We computed the local Fisher information at a slant  $S$  by first finding the linear discriminant ( $w$ ) that separated the population of neural responses ( $A$ ) to stimuli at  $S - 2.5^\circ (\alpha_1)$  from the population of neural responses to stimuli at  $S + 2.5^\circ (\alpha_2)$ . We then used the following equation to compute the Fisher information:

$$I = \frac{(\langle (A(\alpha_1) - \mu) \cdot w \rangle - \langle (A(\alpha_2) - \mu) \cdot w \rangle)^2}{\frac{(\sigma_{(A(\alpha_1) - \mu) \cdot w}^2 + \sigma_{(A(\alpha_2) - \mu) \cdot w}^2)}{2} (\alpha_1 - \alpha_2)^2} = \frac{1}{\sigma_{Estimator}^2}. \quad (3.3)$$

This is equivalent to the squared  $d'$  between the population responses to stimuli at the true slants normalized by the slant difference. It is inversely proportional to the Cramer-Rao lower bound on the variance of the best unbiased slant decoder from the neural population response (assuming the optimal decoder is linear within the 5 degree range spanned by the test stimuli).

One problem with the standard approach for finding linear discriminants is that it involves inverting a covariance matrix, which requires sufficient training data to avoid singular covariance matrices. Since our model included 916 units, the covariance matrices contained 839,056 elements, and it would have required over 3 million training samples at each slant to reasonably estimate these matrices. This was not feasible in terms of time or

computer resources (computer time, memory, and data storage), so we approximated the linear discriminants by minimizing an error function using conjugate gradient descent (Hestenes & Stiefel, 1952). Our error function arbitrarily set  $-1$  as the target value for one slant and  $+1$  for the slant 5 degrees higher, and we found a set of weights ( $w$ ) that minimized the value of

$$Error = \sum_{\alpha \in \alpha_1, \alpha_2} \sum_{N=1}^{10000} (T(\alpha) - (A_{\alpha, N} - \mu) \cdot w)^2, \quad (3.4)$$

where  $A_{\alpha, N}$  was the simulated activity pattern from training example  $N$  at slant  $\alpha$ ,  $\mu$  was the combined mean activity at slants  $\alpha_1$  and  $\alpha_2$ , and  $T(\alpha)$  was the target value. We generated 20,000 training examples at each of the slants 2.5 degrees above and below each main slant, which were between  $-80$  degrees and  $80$  degrees in 10 degree increments, and split each training set into two sets of 10,000 examples; one set was used to train the weight vectors, and we used performance on the other set as a stopping criterion. The weights were initialized to zero for a bias toward low weights, and the algorithm terminated when performance on the stopping set no longer improved. We computed the linear discriminant weights separately for the two frequency bands because the algorithm used differences in spatial frequency content to distinguish among slants when we computed the weights for all of the units at once.

The gradient descent algorithm produced a set of weights at 10 degree slant intervals that could be used to classify an unknown surface as belonging to one of two slants that were 5 degrees apart. Next, we quantified the amount of information present in the activity patterns using additional 10,000 sample testing sets. We used the values of  $\mu$  computed for each slant interval during training to center the data before projecting them onto the linear discriminant vectors. Using separate data for training and testing (cross-validation) ensured that the observed performance was not a result of overfitting the data.

## 4 Results

---

**4.1 Effects of Sensorineural Noise.** We computed the Fisher information extracted from the model's simulated activity patterns and also converted this into predicted standard deviations that indicate the expected performance of an unbiased decoder. Using the broadband textures as input, we examined the effects of sensorineural noise on the information contained in the activity patterns our model produced; the results are shown in Figure 8. The solid gray lines indicate the Fisher information captured by the binocular model, which primarily used orientation disparity but may also have relied on monocular information and a negligible amount of coarse position disparity information. We minimized the effects of position

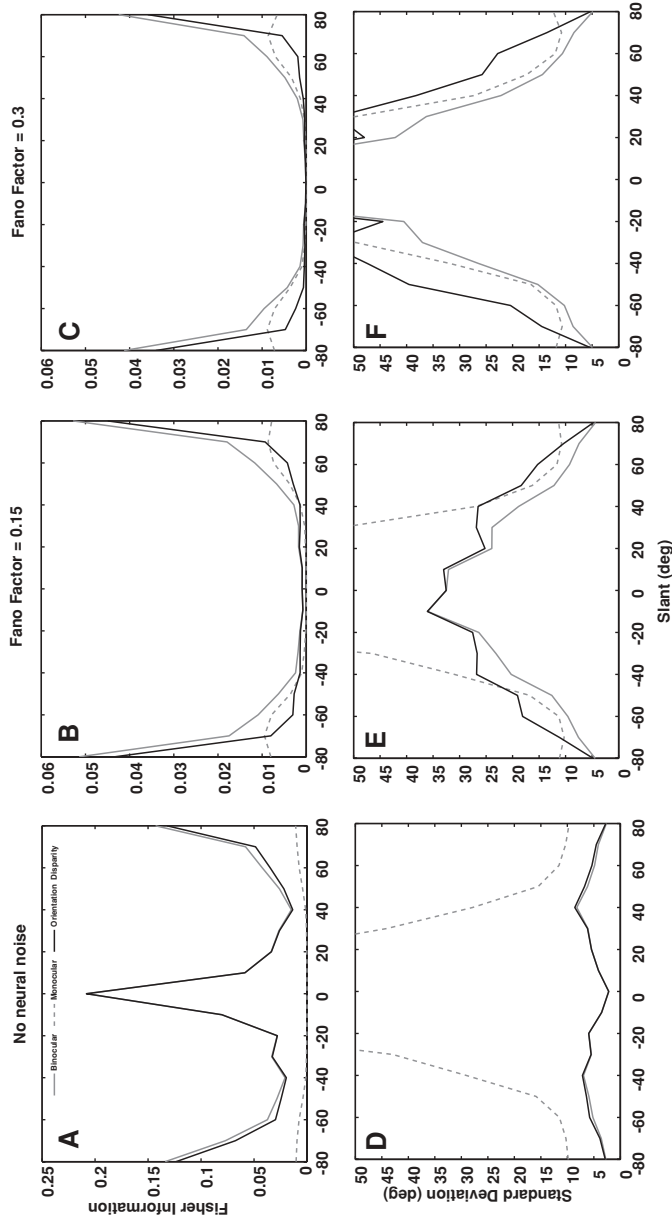


Figure 8: Fisher information and expected performance for broadband textures as a function of sensorineural noise. (A–C) Estimated Fisher information. (D–F) The same information expressed as a lower bound on standard deviation. The three columns correspond to different neural noise levels as indicated. The  $y$ -axis for  $A$  is scaled differently from  $B$  and  $C$  because the range of values was much larger.



disparity by using image patches at fixation and centered on the axis of rotation. The dashed gray lines show the performance of the monocular model, which correlated orientation content and spatial frequency content with surface slant. Since the binocular energy model contained some monocular information, we estimated the Fisher information resulting from orientation disparity by subtracting the monocular information from the total information estimated from the binocular model; the solid black lines show these estimates. Figures 8A to 8C show the amount of Fisher information contained in the activity patterns, and Figures 8D to 8F express the same data as a lower bound on the standard deviation of an unbiased estimator as described earlier. The standard deviation of a uniform distribution over the 180 degree range is about 52 degrees, so standard deviations of this magnitude or higher would indicate chance performance.

For the data presented in Figures 8A and 8D, the only variability in the output of the model was due to random variations in the input patterns; no additional sensorineural noise was included. This provided an upper bound on performance but was not realistic because the visual system has internal noise. Performance improved at extreme slants as expected, but the model performed even better for frontoparallel stimuli, which was not expected because these stimuli contained minimal orientation disparity. There were two apparent trends in the linear discriminant weights based on these stimuli: units tuned to positive and negative orientation disparities tended to be assigned weights with opposite signs, and the magnitudes of the weights tended to be higher for units tuned to orientations in the two eyes that were symmetric about 90 degrees. Although the preference for the symmetric orientations may have suggested that the linear discriminants were sensitive to patterns associated with vertical texture elements, which would have been the most informative for a tilt of 90 degrees, it seems more likely that they relied on binocular correlations between the retinal image pairs, to which the disparity energy model was also sensitive. As slant increased, perspective transformations caused the textures to rotate in opposite directions when projected to the eyes, and the correlations would have been the highest at preferred orientations that were symmetric about the vertical axis. Since the transformations were not pure rotations, the correlations also decreased with increasing slant. Without any noise added to the units, these small differences that occurred at low slants were large enough to produce excellent performance. However, as the other figure panels show, the unexpected high performance on stimuli at low slants was not robust to noise. The monocular model performed poorly relative to the binocular model; this indicated that the performance of the binocular model, which produced standard deviations of 2.7 degrees at slants of  $\pm 80$  degrees, was due to orientation disparity rather than monocular information.

The data shown in the remaining panels resulted from adding noise to the outputs of the simulated neurons. Internal noise can actually be helpful because it regularizes the data and ensures that the system relies on the

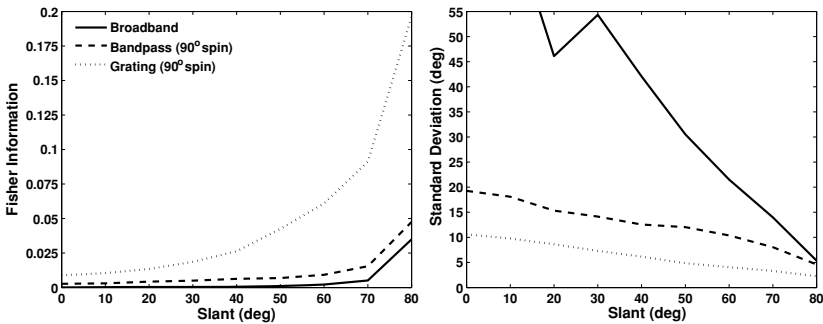


Figure 9: Performance as a function of the type of texture provided as input to the model with a Fano factor of 0.3. (Left) Estimated Fisher information. (Right) Lower bounds on standard deviation predicted from the Fisher information.

most robust cues. We used a Fano factor of 0.15 in Figures 8B and 8E, which was half the amount of noise added to the units represented in Figures 8D and 8F. As previously stated, the Fano factor of 0.3 used for the simulations shown in Figures 8D and 8F matched the median Fano factor estimated in the primary visual cortex of awake, behaving monkeys (Gur & Snodderly, 2006). The sample textures used as input were identical across all three noise conditions. Adding noise universally reduced performance, but it affected low slants much more than high slants. The high performance we observed for low slants in the noise-free condition disappeared when noise was added, after which these slants produced the lowest measures of Fisher information. In subsequent simulations, we always added noise to the simulated activity patterns using a Fano factor of 0.3.

**4.2 Effects of Texture and Orientation Content.** One of our predictions was that the variations in the orientation content of the different texture classes would produce differences in the performance of our model. Specifically, we predicted that grating patterns, which have a clear orientation, would produce better discriminability than textures that are broadband in orientation content. We compared performance using random luminance grating patterns with textures that were bandpass or broadband in orientation content (see Figure 7). The gratings and bandpass textures were spun so that their primary orientation was vertical (spin was irrelevant for the broadband textures due to their isotropy). Consequently, our analysis of different textures measured the ideal performance for each stimulus class.

A comparison of the information content from different texture classes is presented in Figure 9. In this and all subsequent figures, we averaged the data from equal slant magnitudes to help smooth the data since the data were symmetric about 0 degrees. Also, we show only performance due to orientation disparity, which we estimated using the differences between

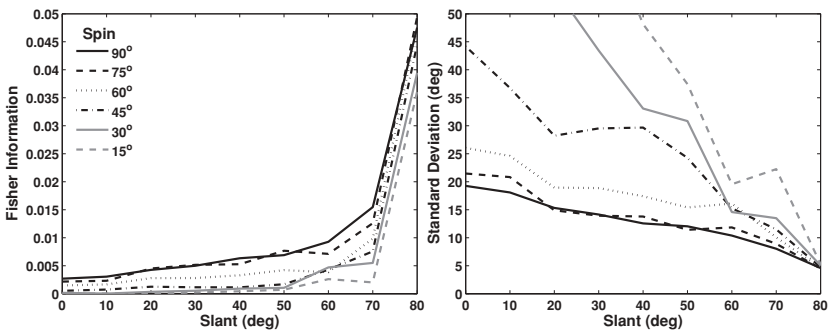


Figure 10: Performance from orientation disparity on bandpass textures with the central orientation spun to the orientations indicated as shown by (left) Fisher information and (right) predicted standard deviations based on Fisher information. Spin: 0 ° = horizontal, 90° = vertical.

the Fisher information estimates from the binocular and monocular models. The data from the broadband patterns are the same as previously displayed in Figures 8C and 8F, and the dashed and dotted patterns show performance on vertically oriented bandpass textures and vertical grating patterns, respectively. As expected, the measured Fisher information was highest for the vertical grating patterns. When the spin of a grating pattern is known and there is no internal noise, there is a bijection between the orientations seen by the two eyes and the slant of the target that allows the monocular model to perform as well as the binocular model. Adding internal noise to the model obscured this mapping and revealed the advantage of binocular processing. The amount of Fisher information extracted from the broadband and bandpass textures was much lower than that from the gratings. Bandpass textures outperformed broadband textures across the entire orientation range we analyzed, although the difference was not very large at the highest slants. For bandpass textures, performance was well above chance across the entire range of slants we used, whereas performance on broadband textures was at chance levels up to about 40 degrees from frontoparallel.

A better comparison would be to randomly spin the grating and bandpass textures before presenting them to the model, but there was a nonlinear interaction between slant and spin in these cases that would have caused problems for the linear discriminant analysis. Average performance on randomly spun patterns should be lower than what we found for the vertical patterns because the specific orientations present in the texture make a difference. To explore this further, we quantified how information from orientation disparity changed as a function of texture orientation.

Figure 10 shows how the information provided by orientation disparity changed as the dominant orientation of bandpass textures varied from 15

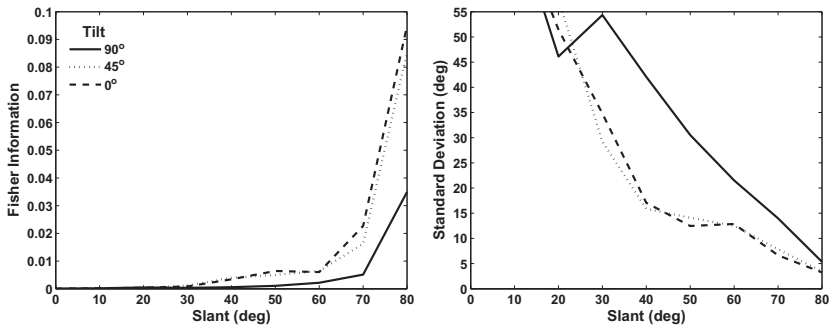


Figure 11: Performance on broadband textures as a function of tilt shown using estimated (left) Fisher information and (right) standard deviations.

to 90 degrees in 15 degree increments. The 30 degree bandwidths of the bandpass textures produced overlapping orientation content for neighboring spin conditions. The data for a spin of 90 degrees (vertical) is the same as shown in Figure 9. There should be no information present when the only orientation present is 0 degrees since the orientations of horizontal lines do not change when slanted about the horizontal. The bandpass textures spun to 0 degrees provided some information from their nonhorizontal components, but this was minimal, and the conjugate gradient-descent algorithm did not work well on those samples. The results indicated monotonically decreasing performance as the orientation content shifted from vertical. This shows that even though the vertical bandpass and grating patterns outperformed the broadband patterns, their performance would be more similar when averaged across spins.

**4.3 Effects of Tilt.** Up to this point, all of the results from our model have been based on surfaces slanted about the horizontal axis, but it is also important to examine how the information content varies over a range of surface orientations. Specifically, if the model could estimate slant only for surfaces tilted at 90 degrees, this would be evidence that the human visual system should not use orientation disparity. Consequently, we used our model to compare performance on broadband stimuli tilted at 90 degrees (rotation about the horizontal axis), 45 degrees (rotation about an oblique axis), and 0 degrees (rotation about the vertical axis). A tilt of 90 degrees can produce the largest orientation disparities (see Figures 2 and 3), but the results of our simulations shown in Figure 11 indicated that slanting stimuli about the horizontal axis does not produce the best slant estimates from orientation disparity. Instead, tilts of 45 and 0 degrees produced better performance. The reason for this, as suggested earlier, is that the wider distribution of orientation disparities across different spins at these tilts allowed better discrimination of slant from orientation disparity. These

results show that orientation disparity provides useful information about 3D orientation across the full range of tilts.

## 5 Discussion

---

In all of the simulations, we found that, as predicted, performance was best for high slants and worst for low slants, which have been the primary foci of previous orientation disparity studies. The only exception was that a simulation lacking internal noise produced excellent performance near frontoparallel, but this was due to artifacts in the stimuli that did not persist with even small amounts of internal noise. The binocular models consistently extracted more information from the stimuli than the monocular models, which were relatively uninformative except at the highest slants, where the signal-to-noise ratio for spatial frequency differences became sufficiently high. The results indicated that orientation disparity carries useful slant information but mostly at higher slants beyond about 45 degrees. This was true even for strongly oriented stimuli.

Our estimates of the information that orientation disparity provides about 3D orientation depend on the level of sensorineural noise. We compared performance as a function of noise levels and found that increasing the noise had a very small impact on performance at high slants but a comparatively large impact on low slants. At low slants, there were artifacts that helped the system discriminate between surfaces that were near frontoparallel, but these effects disappeared with the addition of a small amount of noise. Even for textures that were broadband in orientation content, performance at high slants was remarkably robust to noise. How well we can use orientation disparity to discriminate among different low slants depends considerably on the levels of internal noise found in the visual system, and we have outlined an expected range of performance.

The orientation spectra of different textures have a substantial influence on how much information orientation disparity can contribute, and we used a range of textures to describe a range of performance. Broadband textures contain a random distribution of orientations and are more similar to textures commonly encountered in natural environments, whereas oriented gratings, which contain a single orientation, are at the other extreme because they are helpful for testing the limits of the visual system but are less common in natural scenes. Not surprisingly, we found that the information from vertical gratings was much higher than the information obtained using broadband textures. Bandpass textures, which contained a broader range of orientations than the gratings but still had a dominant vertical orientation, produced performance that was between the performances on the vertical gratings and broadband textures. We also used the bandpass stimuli to test the effect of spin, and performance decreased as the patterns were rotated away from the vertical orientation, which was what we predicted. This showed that spin is a relevant factor for oriented textures. If the grating

patterns were spun randomly, the visual system's performance would be expected to decline as the angle of the texture elements away from vertical increased. Average performance for the oriented textures across all spins would be lower than our results using the vertical textures, and it would not seem unreasonable to expect performance to approach levels similar to those we found using broadband stimuli.

As shown in Figures 2, 3, and 10, tilt also affects orientation disparity, and there are often multiple combinations of slants and tilts that can produce the same orientation disparities. For the purposes of our model, we assumed that the tilt was known, but since slant and tilt are confounded in orientation disparity, our measures slightly overestimated the amount of information that orientation disparity provided. In general, additional mechanisms must determine the tilt of a surface using other visual cues before orientation disparity can help estimate its slant. This information could be derived from monocular cues including contour compression, texture gradients, or shading or from binocular disparity gradients. Once the tilt is known, orientation disparity can be a useful source of information about 3D orientation. We tested our model on stimuli using different tilts and found that orientation disparity provides useful information about slant over the entire range of tilts.

Our model performed at levels that were generally within the range of normal human acuity for estimating slant from binocular vision, particularly at high slants. This showed that a biologically plausible model tuned only to orientation disparity could estimate surface slant with performance that is similar to that of human observers. While we cannot say for certain whether the visual system actually uses orientation disparity to estimate slant, the levels of Fisher information we measured did not eliminate it as a useful cue, and our results may be helpful for guiding future psychophysical and neurophysiological investigations. It is possible that better decoding schemes could extract more information from the simulated activity patterns and produce better overall performance. Also, our model used only single small image patches, and integrating over multiple image patches would produce further performance gains. The standard deviations we predicted from the estimated Fisher information are therefore not absolute bounds on slant estimation from orientation disparity, but any visual cortical mechanism tuned to orientation disparity could conceivably estimate slants at the levels we found and should show similar effects of slant, tilt, orientation content, and internal noise levels. We want to emphasize that orientation disparity is not a necessary or sufficient cue for slant estimation and that we are not proposing that the visual system uses it instead of position disparity. Instead, orientation disparity is just one of several cues for 3D orientation, including position disparity gradients and monocular perspective cues like aspect ratio and texture compression. The confound between position disparity and orientation disparity suggests that they may be two aspects of the same binocular cue, and combining them could result in better estimates of 3D orientation.

## Acknowledgments

---

We acknowledge Alex Pouget, Jeff Beck, and others for helpful discussions regarding our analyses and the Center for Visual Science for the use of its computing cluster, without which this research would not have been possible. This research was supported by National Institutes of Health grant EY017939.

## References

---

- Adams, W. J., & Mamassian, P. (2002). Common mechanisms for 2D tilt and 3D slant after-effects. *Vision Research*, 42, 2563–2568.
- Banks, M. S., Gepshtein, S., & Landy, M. S. (2004). Why is spatial stereoresolution so low? *Journal of Neuroscience*, 24(9), 2077–2089.
- Blakemore, C., Fiorentini, A., & Maffei, L. (1972). A second neural mechanism of binocular depth discrimination. *Journal of Physiology*, 226, 725–749.
- Bridge, H., & Cumming, B. G. (2001). Response of macaque V1 neurons to binocular orientation differences. *Journal of Neuroscience*, 21(18), 7293–7302.
- Bridge, H., Cumming, B. G., & Parker, A. J. (2001). Modeling V1 neuronal responses to orientation disparity. *Visual Neuroscience*, 18(6), 879–891.
- Burt, P., & Julesz, B. (1980). A disparity gradient limit for binocular fusion. *Science*, 208(4444), 615–617.
- Cagenello, R., & Rogers, B. J. (1993). Anisotropies in the perception of stereoscopic surfaces: The role of orientation disparity. *Vision Research*, 33(16), 2189–2201.
- Chen, Y., & Qian, N. (2004). A coarse-to-fine disparity energy model with both phase-shift and position-shift receptive field mechanisms. *Neural Computation*, 16, 1545–1577.
- Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. New York: Wiley.
- Gillam, B., & Rogers, B. (1991). Orientation disparity, deformation, and stereoscopic slant perception. *Perception*, 20, 441–448.
- Gillam, B., & Ryan, C. (1992). Perspective, orientation disparity, and anisotropy in stereoscopic slant perception. *Perception*, 21, 427–439.
- Greenwald, H. S., & Knill, D. C. (2009). Cue integration outside central fixation: A study of grasping in depth. *Journal of Vision*, 9, 1–16.
- Gur, M., & Snodderly, D. (2006). High response reliability of neurons in primary visual cortex (V1) of alert, trained monkeys. *Cerebral Cortex*, 16, 888–895.
- Heeley, D. W., Scott-Brown, K. C., Reid, G., & Maitland, F. (2003). Interocular orientation disparity and the stereoscopic perception of slanted surfaces. *Spatial Vision*, 16(2), 183–207.
- Henrie, J. A., & Shapley, R. (2001). The relatively small decline in orientation acuity as stimulus size decreases. *Vision Research*, 41(13), 1723–1733.
- Hestenes, M. R., & Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49(6), 409–436.
- Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, 4(12), 967–992.

- Jones, D. G., & Malik, J. (1992). Determining three-dimensional shape from orientation and spatial frequency disparities. In *Proceedings of the European Conference on Computer Vision* (pp. 661–669). Berlin: Springer.
- Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, 43(24), 2539–2558.
- Koenderink, J. J., & van Doorn, A. J. (1976). Geometry of binocular vision and a model for stereopsis. *Biological Cybernetics*, 21(1), 29–35.
- Lippert, J., & Wagner H. (2002). Visual depth encoding in populations of neurons with localized receptive fields. *Biological Cybernetics*, 87, 249–261.
- Nienborg, H., Bridge, H., Parker, A. J., & Cumming, B. G. (2004). Receptive field size in V1 neurons limits acuity for perceiving disparity modulations. *Journal of Neuroscience*, 24(9), 2065–2076.
- Ninio, J. (1985). Orientational versus horizontal disparity in the stereoscopic appreciation of slant. *Perception*, 14, 305–314.
- Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science*, 249(4972), 1037–1041.
- Orban, G. A., Vandenbussche, E., & Vogels, R. (1984). Human orientation discrimination tested with long stimuli. *Vision Research*, 24(2), 121–128.
- Prazdny, K. (1985). On the disparity gradient limit for binocular fusion. *Perception and Psychophysics*, 37(1), 81–83.
- Read, J. C. A. (2002). A Bayesian approach to the stereo correspondence problem. *Neural Computation*, 14, 1371–1392.
- Sally, S. L., & Gurnsey, R. (2004). Orientation discrimination across the visual field: Matching perceived contrast near threshold. *Vision Research*, 44(23), 2719–2727.
- Stevens, K. A. (1983). Slant-tilt: The visual encoding of surface orientation. *Biological Cybernetics*, 46, 183–195.
- Taira, M., Tsutsui, K. I., Jiang, M., Yara, K., & Sakata, H. (2000). Parietal neurons represent surface orientation from the gradient of binocular disparity. *Journal of Neurophysiology*, 83, 3140–3146.
- Todorovic, D. (1996). A gem from the past: Pleikart Stumpf's (1911) anticipation of the aperture problem, Reichardt detectors, and perceived motion loss at equilibrium. *Perception*, 25(10), 1235–1242.
- Tyler, C. W. (1974). Depth perception in disparity gratings. *Nature*, 251, 140–142.
- von der Heydt, R., Hännny, P., & Dürsteler, M. R. (1981). The role of orientation disparity in stereoscopic perception and the development of binocular correspondence. *Advances in Physiological Sciences*, 16, 461–470.
- Witkin, A. P. (1981). Recovering surface shape and orientation from texture. *Artificial Intelligence*, 17, 17–45.
- Yang, Z., & Purves, D. (2003). Image/source statistics of surfaces in natural scenes. *Network*, 14(3), 371–390.



**This article has been cited by:**