Behavioral/Cognitive

The Limits of Human Stereopsis in Space and Time

David Kane,[1] Phillip Guan,[2] and Martin S. Banks[1,2]

[1]Vision Science Program, University of California, Berkeley, California 94720; and [2]Graduate Program in Bioengineering, University of California, Berkeley, California 94720 and University of California, San Francisco, California 94143

To encode binocular disparity, the visual system determines the image patches in one eye that yield the highest correlation with patches in the other eye. The computation of interocular correlation occurs after spatiotemporal filtering of monocular signals, which leads to restrictions on disparity variations that can support depth perception. We quantified those restrictions by measuring humans' ability to see disparity variation at a wide range of spatial and temporal frequencies. Lower-disparity thresholds cut off at very low spatiotemporal frequencies, which is consistent with the behavior of V1 neurons. Those thresholds are space–time separable, suggesting that the underlying neural mechanisms are separable. We also found that upper-disparity limits were characterized by a spatiotemporal, disparity-gradient limit; to be visible, disparity variation cannot exceed a fixed amount for a given interval in space–time. Our results illustrate that the disparity variations that humans can see are very restricted compared with the corresponding luminance variations. The results also provide insight into the neural mechanisms underlying depth from disparity, such as why stimuli with long interocular delays can still yield clear depth percepts.

Introduction

To encode luminance contrast, the visual system uses neurons that respond differentially to spatiotemporal variations in luminance (Movshon et al., 1978a,1978b; DeValois et al., 1982). The substructure of the neurons' receptive fields determines the spatial and temporal properties of their preferred luminance stimulus. The encoding of binocular disparity is fundamentally different because the visual system must determine which parts of the two retinal images correspond to one another. This is accomplished by determining the displacement of a patch in one eye's image that yields the highest correlation with a patch in the other eye. Disparity estimation is similar to windowed cross-correlation, a technique used successfully in computer vision (Kanade and Okutoni, 1994) and in modeling human vision (Cormack et al., 1991; Fleet et al., 1996; Banks et al., 2004). Windowed cross-correlation is fundamentally similar to the disparity–energy calculation characteristic of binocular interaction in visual cortex (Ohzawa et al., 1990; Anzai et al., 1999). The cortical neurons performing this computation respond preferentially to the average disparity and not to variations in disparity across the receptive field (Nienborg et al., 2004).

Estimating disparity by correlation imposes significant limits on the encoding of spatial and temporal variations in disparity. Previous work has shown that the finest spatial variation in disparity that can be seen is much coarser than the finest visible variation in luminance (Tyler, 1974; Bradshaw and Rogers, 1999; Banks et al., 2004) and that the fastest detectable rate of change in disparity is much slower than the fastest visible luminance change (Richards, 1972; Norcia and Tyler, 1984; Patterson et al., 1992; Lankheet and Lennie, 1996). Moreover, large disparities—those exceeding the disparity-gradient limit—do not yield reliable depth percepts; specifically, depth from disparity collapses whenever the change in disparity for a given change in position exceeds 1–1.5 (Tyler, 1973; Burt and Julesz, 1980).

We measured the minimum disparity required to perceive depth from disparity for a wide range of spatiotemporal frequencies. We also measured the maximum disparity that supports depth perception at these frequencies. The minimum thresholds are space–time separable. The maximum disparity limits reveal that disparity variation cannot be greater in space–time than a critical value embodied by a spatiotemporal, disparity-gradient limit. These results reveal the set of spatial and temporal variations in disparity that can be seen. The minimum threshold and maximum limit data are both consistent with a cross-correlation model with separable spatial and temporal windowing functions. Together, the data and modeling provide significant insight into the spatiotemporal properties of the neural mechanisms that underlie the perception of depth from disparity.

Materials and Methods

Main experiment

Observers. Four subjects (two males and two females) 22–30 years of age participated. All had corrected-to-normal vision. Two were authors; the other two were unaware of the experimental hypotheses.

Apparatus. The stimuli were presented on a mirror stereoscope with two CRT displays (HM204DT; Iiyama). The lines of sight from the two eyes were reflected from mirrors near the eyes such that they were colinear with a normal from the center of the respective CRTs. The experiment was conducted in a dark room, so the CRTs provided the only measurable light input to the eyes. The CRTs were set to a spatial resolution of

Received April 17, 2013; revised Nov. 7, 2013; accepted Dec. 6, 2013.

Author contributions: D.K., P.G., and M.S.B. designed research; D.K. and P.G. performed research; D.K., P.G., and M.S.B. analyzed data; D.K., P.G., and M.S.B. wrote the paper.

This work was supported by the National Institutes of Health (Grant EY012851). We thank Hany Farid for assistance in the execution of the modeling and Jenny Read, Pascal Mamassian, and Cliff Schor for comments on an earlier version of the manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to Martin S. Banks, University of California, Berkeley, 360 Minor Hall, Berkeley, CA 94720-2020. E-mail: martybanks@berkeley.edu.

DOI:10.1523/JNEUROSCI.1652-13.2014

 0270-6474/14/341397-12$15.00/0

800 × 600 pixels. At the 115 cm viewing distance, pixels subtended 1.5 minutes of arc (arcmin). Using anti-aliasing, we could create much smaller disparities. We estimate that the smallest displayable disparity was 2 seconds of arc (arcsec); the smallest disparity we presented in the main experiment was 7 arcsec. Vergence distance was 125 cm, slightly different from the 115 cm optical distance from each eye to a CRT because of the angle between the two limbs of the stereoscope and the rotations of the mirrors. The refresh rate was 200 Hz.

Stimuli and procedure. Between trials, identical dynamic random-dot patterns were presented to the two eyes. In addition, a fixation target was presented that was composed of two binocular horizontal line segments and two dichoptic vertical line segments. By monitoring the apparent alignment of the dichoptic segments, observers could make sure that fixation was accurate before initiating a stimulus presentation. They were told to maintain fixation on the fixation target during the stimulus presentation as well. All reported that they did so in part because the task became more difficult if they moved their eyes. In each presentation, a signal stimulus and a no-signal stimulus were presented simultaneously for 1 s to the left and right of fixation. They were dynamic random-dot stereograms generated using the Psych-Toolbox (Brainard, 1997; Pelli, 1997). The signal stimulus specified a horizontal triangular-wave corrugation in depth that drifted upward or downward (chosen randomly) at a chosen speed. Figure 1 provides a static example. We used a triangular-wave corrugation because its disparity gradient is well defined. The no-signal stimulus had an identical distribution of disparities over time, but the spatial order was scrambled, yielding an incoherent appearance. Because the signal and no-signal stimuli had the same distribution of disparities over time, the observer had to perceive the spatiotemporal waveform to perform the discrimination task reliably. We intended to measure two thresholds: (1) a lower-disparity threshold, the smallest disparity required to perceive the signal waveform and thereby distinguish it from the no-signal stimulus, and (2) an upper-disparity limit, the largest disparity before disparity processing collapses and the signal waveform is not perceived. To measure upper-disparity limits, it was crucial to have a suitable no-signal stimulus. A no-signal stimulus with zero disparity is insufficient because observers can use the presence of any depth variation in the signal stimulus to perform the discrimination task when that stimulus exceeds the disparity-gradient limit. Our spatially incoherent no-signal stimulus appeared identical to the signal stimulus once the gradient limit was exceeded. At the limit, performance fell to chance, allowing us to measure upper-disparity limits reliably.

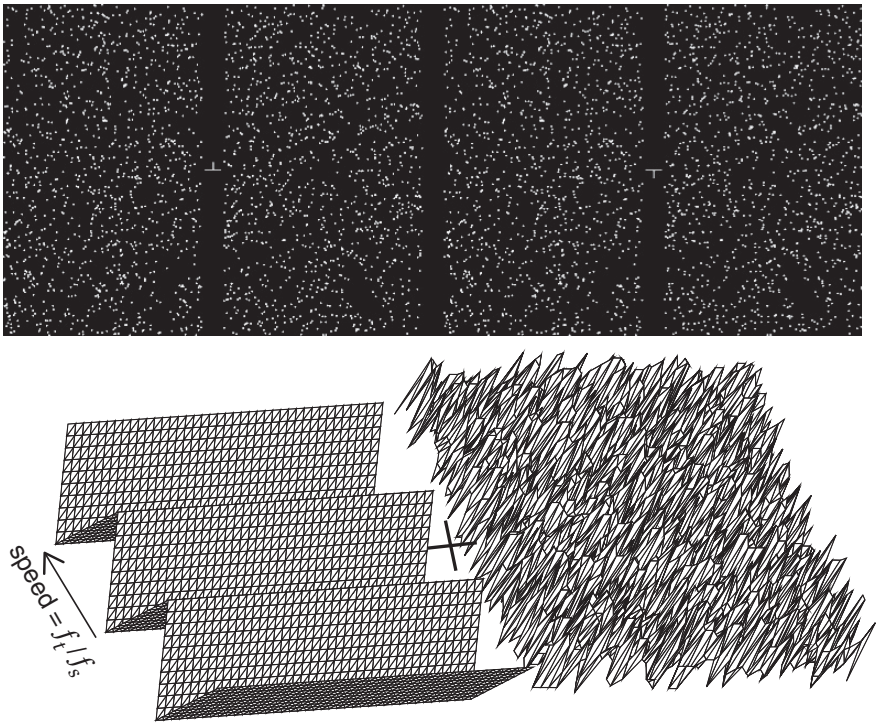

Figure 1. Signal and no-signal stimuli. Top row: Random-dot stereogram stimulus. Cross-fuse to see the stimulus in depth. The stereograms created a triangular wave in depth that moved upward or downward (left) and an incoherent pattern that had the same distribution of disparities (right). In the experiment, the dots were replaced on each frame (5 ms), thereby creating a dynamic random-dot stereogram. The no-signal stimulus was incoherent in space and time. Subjects fixated the dichoptic cross in the middle. Bottom row: Depiction of the stimulus in depth. The triangular-wave signal is on the left. It moved upward or downward at a speed of f_t/f_s. The incoherent no-signal stimulus is on the right.

The signal and no-signal patches were each 15° tall and 9° wide. The inner edges of the patches were 0.5° from the center of the fixation target. The dots in the stereograms were 3 arcmin in diameter. Dots were refreshed at 200 Hz. It was critical to refresh quickly to be sure that our measurements of temporal resolution were not confounded by long dot lifetimes. Dot density was 9 dots/deg^2, yielding a Nyquist frequency of 1.5 cpd for each frame (Banks et al., 2004). However, new dots consistent with the simulated waveform were presented every 5 ms, so the effective dot density (and therefore the effective Nyquist frequency) was much higher due to visual persistence (Lankheet and Lennie, 1996). Specifically, if the visual system integrated the information in n frames, the effective Nyquist frequency would be $1.5n$ (i.e., 6 cpd for integration of 4 frames, which is 20 ms). After each presentation, the observer indicated with a key press whether the signal had appeared on the left or right. Incorrect responses were indicated by an audible tone.

Eight temporal frequencies (0, 1, 2, 4, 5.6, 8, 11.3, and 16 Hz) and 7 spatial frequencies (0, 0.06, 0.12, 0.25, 0.5, 1, and 1.5 cpd) were presented in all combinations except 0 Hz, 0 cpd. None of the observers could perform the task reliably at 16 Hz, so those data were discarded. Disparity amplitude for each condition was manipulated using the method of constant stimuli. All conditions were randomly interleaved within an experimental session.

There were unmatched dots (i.e., seen by one eye but not the other) near the left and right edges of the stimuli and the number of such dots increased with disparity amplitude. Therefore, the center of the signal stimulus looked like a coherent drifting triangular wave and the edges appeared noisy. The central portion was large enough to make the discrimination between signal and no-signal stimuli easily.

Data analysis. The psychometric data were generally non-monotonic because performance was constrained by a lower-disparity threshold and an upper-disparity limit. Specifically, when disparity amplitude was increased from a very small value, percent correct rose above the 50% chance rate to 100%. However, as amplitude was increased yet further, the percent correct fell back to 50%. Examples are provided in Figure 2. Because the psychometric data had this form, we needed to fit two functions to it. We did so by modeling the psychometric function as the product of two cumulative probability distributions, one rising and one falling with increasing amplitude (Equations 1–3) as follows:

$$P_l(a) = 1 - \left[erf\left(\frac{\log(a) - \mu_l}{\sqrt{2\sigma_l^2}} \right) \right] \quad (1)$$

$$P_u(a) = 1 + \left[erf\left(\frac{\log(a) - \mu_u}{\sqrt{2\sigma_u^2}} \right) \right] \quad (2)$$

$$P_c(a) = 0.5 + \frac{P_l(a)\, P_u(a)}{2} \quad (3)$$

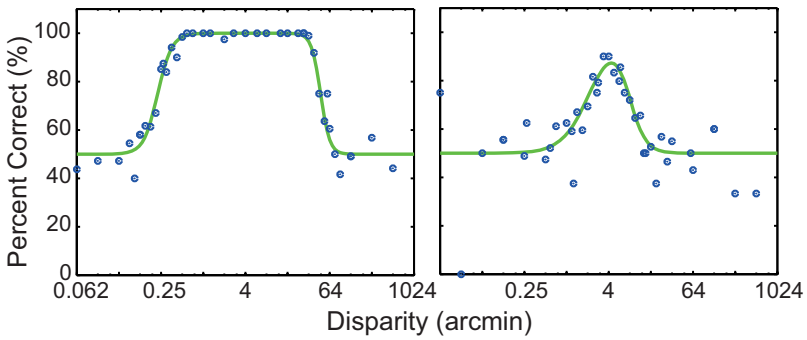

Figure 2. Psychometric data from two experimental conditions. Percent-correct performance is plotted as a function of disparity amplitude. With increasing disparity, performance typically first increased and then decreased. We fit the data with the product of two cumulative Gaussians (Equation 3), one increasing with increasing disparity (Equation 1), and one decreasing (Equation 2). The green curves represent the best-fitting functions. The psychometric data on the left are an example of what we define as a reliable threshold estimate. The data on the right are an example of a marginally reliable threshold estimate.

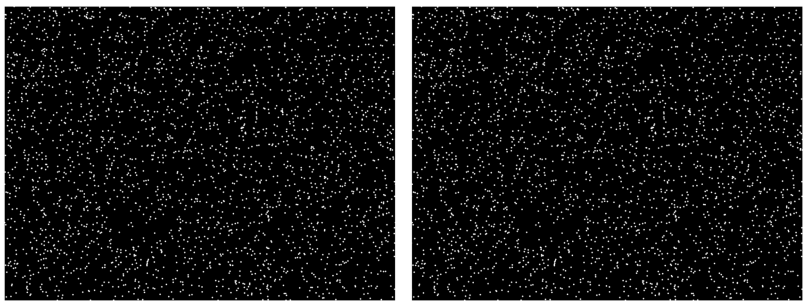

Figure 3. Stimuli in the spatial-step experiment. Cross-fuse to see the stimulus in depth. Dynamic random-dot stereograms created a horizontal ridge in depth. The disparities on one side of the ridge were convolved with a Gaussian, thereby blurring the disparity-defined edge. The blur on the lower side has a SD of 44 arcmin from a viewing distance of 50 cm. Here, only one frame is shown so the Nyquist frequency is much lower than in the experiment.

where a is the disparity amplitude; P_l, P_u, and P_c are, respectively, the distributions for the lower threshold, upper limit, and the combined distribution; μ_l and μ_u are the means for the lower and upper functions; and σ_l and σ_u are the SDs. Therefore, fits to the data involved four parameters. The values we report are the means of the two best-fitting distributions corresponding to the 75% correct points of the two distributions considered independently. We will refer to the lower thresholds from P_l as the lower-disparity thresholds and to the upper thresholds from P_u as the upper-disparity limits. When the two distributions did not overlap, we could reliably estimate two thresholds from the data. When the distributions overlapped, we could not necessarily estimate two values. If a fitted distribution exceeded 95% before intersecting the other distribution (e.g., when the lower distribution exceeded 95% before intersecting the upper distribution), we regarded the 75% estimate as reliable. If a distribution did not exceed 75% before intersecting the other, we regarded the estimate as unreliable and discarded the data for further use. If the intersection fell between 75% and 95%, we regarded the 75% estimate as marginally reliable. The data from the two retained categories—reliable and marginally reliable—are treated separately in the figures.

Spatial-step experiment

Five subjects (three males and two females) from 25 to 65 years of age participated. One of them also participated in the main experiment. All had corrected-to-normal vision. Two were authors; the other three were unaware of the experimental hypotheses.

The apparatus was the same as in the main experiment, so we will focus on the differences. The dynamic random-dot stimuli depicted a horizontal ridge in front of a background plane. Dot density was 36 dots/deg^2 per frame, so the Nyquist frequency was 3 cpd per frame. Dots were refreshed in each 5 ms frame, so the effective Nyquist frequency was much more than 3 cpd. Figure 3 provides a static example of the stimulus. The spatial disparity profile of the top or bottom edge of the ridge was convolved with a Gaussian with a SD of σ_s. The disparity of the ridge was 8 arcmin and the ridge appeared for 1 s. The ridge's vertical position and height were varied by $\pm$22.5 arcmin to ensure that subjects could not use position cues to perform the task. Observers indicated whether the top or bottom edge contained the Gaussian blur. No feedback was provided. The discrimination could not be made monocularly. σ_s was varied according to the method of constant stimuli. A total of 500–660 trials were presented per observer. The resulting psychometric data were fit with a cumulative Gaussian using a maximum-likelihood criterion and the 75% point was defined as the threshold value of σ_s.

Temporal-step experiment

Four subjects (two males and two females) from 25 to 65 years of age participated. Three of them also participated in the spatial-step experiment. All had corrected-to-normal vision. Two were authors; the others were unaware of the experimental hypotheses.

The apparatus was the same as in the other two experiments. The dynamic random-dot stimuli depicted a horizontal ridge that emerged from a uniform background plane. Dot density was 36 dots/deg^2 per frame. Dots were refreshed every 5 ms frame. In some cases, the change in disparity was convolved with a temporal Gaussian with an SD of σ_t. Two stimuli appeared in succession, one changing from 0 to 8 arcmin in 1 frame (5 ms) and the other changing more slowly due to the temporal Gaussian. The onset of the temporal change was randomly jittered by $\pm$100 ms so that observers could not use onset time as a cue. Observers indicated which interval contained the slower change. Feedback was provided. We used the method of constant stimuli to vary σ_t and thereby determine discrimination threshold (75% point on the best-fitting cumulative Gaussian). A total of 750–960 trials were presented per observer.

We computed the amplitude spectra of the sharp and blurred edges of the spatial-step stimulus and from those calculated the difference spectrum. The stimulus profiles and spectra are shown in Figure 4. The difference spectrum has a clear peak at a spatial frequency of $15/\sigma_s$ cpd with σ_s expressed in minutes of arc. A similar analysis of the temporal-step stimulus shows that the difference spectrum has a clear peak at a temporal frequency of $250/\sigma_t$ Hz with σ_t expressed in milliseconds. Therefore, from the values of σ_s and σ_t at threshold, we could determine the spatial and temporal frequencies that mediated performance.

Results

Main experiment

The individual subject data were very similar to one another, so we averaged across subjects. Figure 5*A*, left and right, plot the average data as a function of spatial frequency and temporal frequency, respectively, and Figure 5*B* plots the same data 3D. As disparity increased from a very small value, the waveform became visible; the disparities at which this occurred define the lower-disparity thresholds (inverted triangles in Fig. 5*A*; blue surface in Fig. 5*B*). As disparity increased further, the waveform again became indiscriminable; the disparities at which this happened define the upper-disparity limits (circles; purple surface). When the

upper limit was exceeded, the coherent corrugation and incoherent stimuli had identical appearances of noisy depth. The transition from a coherent to an incoherent percept was often precipitous in that it occurred over a relatively small change in disparity.

The volume between the two surfaces in Figure 5*B* represents the disparity variations that yield coherent depth percepts. Disparities lying within the volume are visible and those outside are not. The visible volume is very restricted compared with the analogous window of visibility for luminance modulation (Robson, 1966; Kelly, 1979). For example, the highest resolvable spatial frequency was ~3 cpd, whereas the analogous cutoff for luminance contrast is nearly 20 times higher at 50 cpd (Campbell and Green, 1965). The highest resolvable temporal frequency was ~8 Hz and the analogous limit for luminance is nearly 10 times higher at ~70 Hz (de Lange, 1958). Figure 6 demonstrates the large difference in the area of visible spatial frequencies in the disparity domain compared with the luminance domain when the temporal frequency is zero.

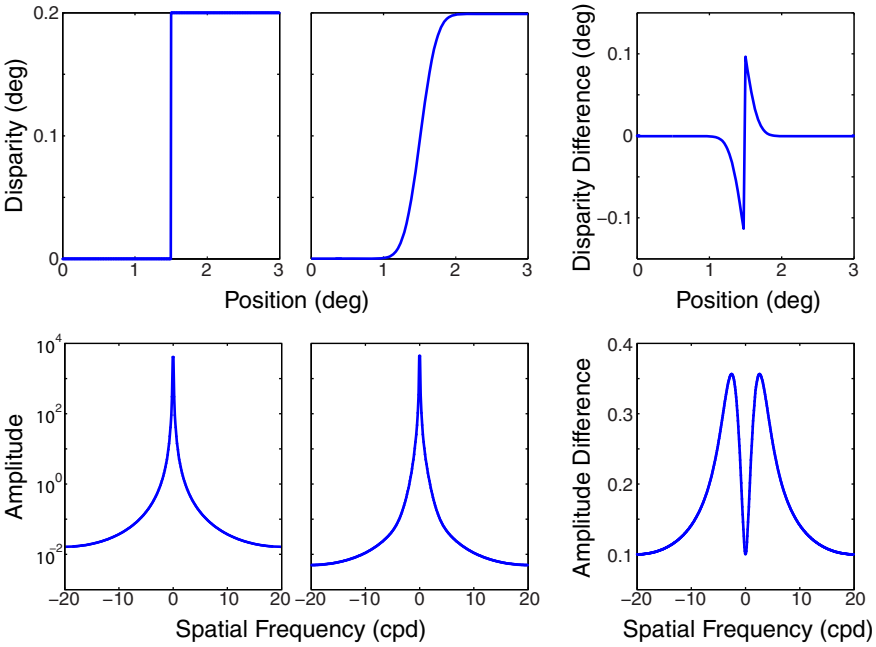

Figure 4. Disparity profiles and amplitude spectra of spatial-step stimuli. Top, Left, middle, and Right, Disparity profiles—with disparity as a function of position—for the sharp edge, blurred edge, and the difference between the two when σ_s equals 9 arcmin, respectively. Bottom, Left, middle, andRright, Amplitude spectra, respectively, for the disparities of the sharp edge, blurred edge, and the difference between the two. The difference spectrum has a clear peak at $15/\sigma_s$ cpd, which is 1.7 cpd when σ_s is expressed in arcmin.

At a temporal frequency of 0 Hz, the variation in lower threshold with spatial frequency is band-pass with the lowest threshold at ~0.3 cpd. The highest frequency we presented was 1.5 cpd, but from inspection of Figure 5*A*, frequencies of ~3 cpd would have been resolvable. Previous investigators reported similar resolutions ranging from 1.6–4 cpd (Tyler, 1973, 1974; Bradshaw and Rogers, 1999; Banks et al., 2004). Figure 5 also shows that threshold at a given spatial frequency depends on temporal frequency. Specifically, as temporal frequency increases, the minimum threshold at a given spatial frequency increases. At a spatial frequency of 0 cpd, the variation of lower threshold as a function of temporal frequency is low-pass: the threshold increases monotonically with increasing frequency until the stimulus becomes indiscriminable above 11.3 Hz. Previous researchers have reported similar cutoff frequencies of 6–12 Hz when the task required more than detecting the presence of depth (Richards, 1972; Norcia and Tyler, 1984; Patterson et al., 1992; Lankheet and Lennie, 1996).

We also examined the slopes of the psychometric functions for each experimental condition (Fig. 7). The slopes for the lower-disparity threshold data were nearly a constant fraction of the threshold value across all spatial and temporal frequencies, and the slopes for the upper-disparity limit data were also a nearly constant fraction (except when the spatial frequency was 0 cpd, where the upper-limit and lower-threshold slopes were noticeably shallower). Interestingly, the slopes for the upper-limit data were much steeper than the slopes for the lower-threshold data, which is consistent with our observation that increasing disparity beyond the disparity-gradient limit yields a precipitous fall in the ability to perceive the disparity-specified waveform. The large difference in slopes suggests that the lower-disparity thresholds and upper-disparity limits have different underlying causes.

Spatial- and temporal-step experiments

We wanted to know whether the observed limitations in spatial and temporal resolution apply to other stimuli, such as the appearance of spatial and temporal steps. There are plausible reasons why the results with the periodic stimuli of the main experiment would not apply to other stimuli. First, the inability to perceive periodic waveforms at high spatial or temporal frequency is determined in large part by the convergence of the lower-disparity threshold and upper-disparity limit at those frequencies. As Figure 5 shows, the thresholds (triangles) and limits (circles) approach one another as spatial and temporal frequency increase, which means that the ability to perceive a high-frequency waveform is compromised for two reasons. Spatial and temporal steps may not be subject to the same dual constraint. Second, generalizing findings with periodic waveforms to nonperiodic waveforms requires an assumption of linearity and disparity processing may not obey that assumption sufficiently.

For these reasons, we checked the results of the main experiment using nonperiodic stimuli and a different discrimination task. We also thought that the appearance of spatial and temporal steps might be of more general interest because the natural environment contains many instances of steps in depth but few instances of periodic depth variations. We presented spatial and temporal steps in disparity using dynamic, random-dot stereograms, as described in Materials and Methods.

The threshold values of σ_s for the 5 observers ranged from 1.04–3.45 arcmin. When σ_s was smaller than the threshold value, observers reported that the step looked sharp. We combined all the data into one psychometric function and the threshold value was 1.40 arcmin (95% confidence intervals of ±0.22 arcmin). In similar experiments in the luminance domain—i.e., discriminating sharp from blurred luminance edges—σ_s at threshold is 0.4–0.5 arcmin (Watson and Ahumada, 2011), ~3 times smaller than we observed in the disparity domain. This difference is smaller

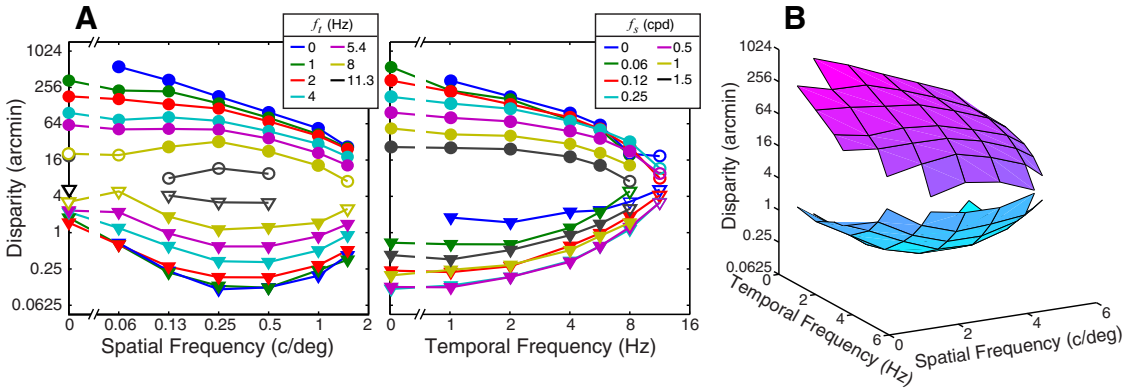

Figure 5. Results averaged across subjects. The average data were calculated by fitting curves to all of the psychometric data from all of the subjects together. ***A***, Left, Plots of the disparity amplitude at threshold as a function of spatial frequency. The lower data points, plotted as triangles, represent the lower thresholds and the upper points, plotted as circles, represent the upper limits. Blue, green, red, purple, cyan, yellow, and black symbols represent the data when temporal frequency was 0, 1, 2, 4, 5.4, 8, and 11.3 Hz, respectively. Filled symbols represent thresholds estimated from nonoverlapping psychometric functions (reliable estimates; see Materials and Methods); unfilled symbols represent thresholds estimated from overlapping functions (marginally reliable estimates). Right, Plots of the disparity amplitude at threshold as a function of temporal frequency. The lower and upper data points represent the lower thresholds and upper limits, respectively. Blue, green, red, purple, cyan, yellow, and black symbols represent the data when spatial frequency was 0, 0.06, 0.12, 0.25, 0.5, 1, and 1.5 cpd, respectively. Filled and unfilled symbols again represent thresholds measured from nonoverlapping and overlapping psychometric functions, respectively. ***B***, Average data plotted in 3D. The disparity amplitudes for which subjects could just discriminate the spatiotemporal waveform are plotted as a function of spatial and temporal frequency. The blue surface represents the lower-disparity thresholds and the purple surface the upper-disparity limits.

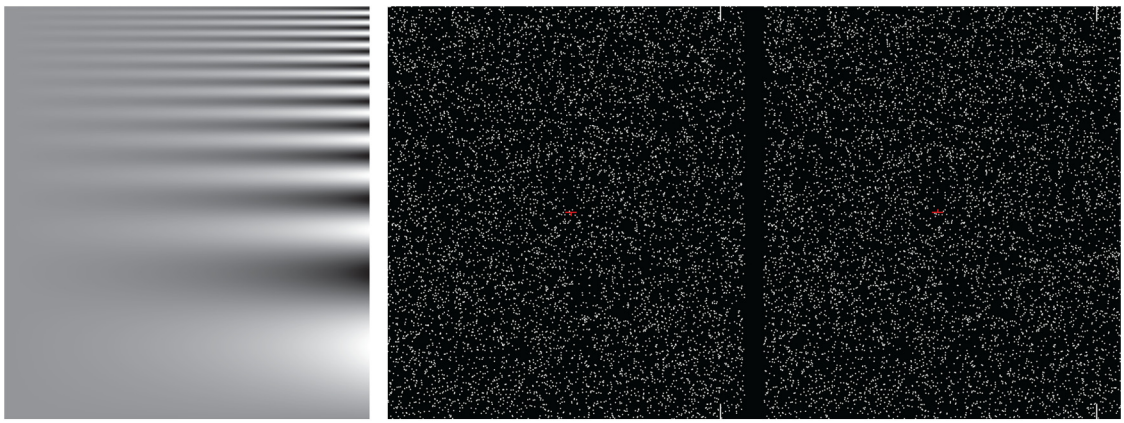

Figure 6. Images demonstrating the striking differences in spatial and temporal sensitivity associated with luminance and disparity processing. The spatial frequency units were calculated for a viewing distance of 3.5 times picture height. Left, Luminance grating with spatial frequency increasing from 0.06 to 3 cpd from bottom to top and contrast increasing from ~0 to ~1 from left to right. The grating is visible at high contrast at all spatial and temporal frequencies (right side) and is less visible or invisible at low contrast. There is little effect of spatial frequency because the range of frequencies is small relative to the spatiotemporal contrast sensitivity function. Right, Random-dot stereogram depicting a disparity-defined corrugation with the same spatial frequencies as the luminance grating. Cross-fuse to see stereoscopically. The spatial frequencies are correct when the viewing distance is 3.5 times picture height. Corrugation spatial frequency increases from 0.06 to 3 cpd from bottom to top. Disparity amplitude increases from 0 to 1.75° from left to right. Unlike the luminance grating, the disparity corrugation cannot be seen in the upper right corner, where the disparity amplitude exceeds the disparity-gradient limit. When the gratings modulate in counterphase, the region where disparity grating becomes invisible increases in size as temporal frequency increases because visibility is limited by the spatiotemporal disparity-gradient limit. The luminance grating remains unchanged when flickering in counter-phase at the same temporal frequencies because it is not affected by a spatiotemporal gradient limit.

than the 15- to 20-fold difference observed for the finest visible periodic waveform. From Figure 4, we observed that the peak spatial frequency in the spectrum of the difference between a sharp edge and blurred edge is $15/\sigma_s$, which for the threshold values of σ_s corresponds to 11 and 33 cpd for disparity and luminance, respectively, again suggesting a smaller difference than observed with periodic stimuli. The spatial-step results show that the ability to differentiate sharp from blurred steps in disparity is indeed compromised compared with the corresponding case in the luminance domain, but not as much as one would expect from the data with periodic waveforms.

The temporal-step experiment yielded rather different results. The threshold values of σ_t for the four subjects ranged from 46.9–84.2 ms. All observers reported that the steps that occurred more rapidly than the threshold value looked like instantaneous transitions. When we combined all the data into one psychometric function, the threshold value was 69.0 ms (95% confidence intervals of ±7.2 ms). To our knowledge, the corresponding experiment has not been done in the luminance domain. From the analysis in Figure 4, we note that the peak temporal frequency in the spectrum of the difference between an instantaneous and slower step is $250/\sigma_t$ Hz, which corresponds to 3.6 Hz, a strikingly low value. The temporal resolution limit observed in the main experiment leads to the expectation that the ability to distinguish instantaneous from slower temporal transitions should be quite compromised. The temporal-step results show that it is indeed compromised, indeed even more compromised than predicted by the simple analysis in Figure 4.

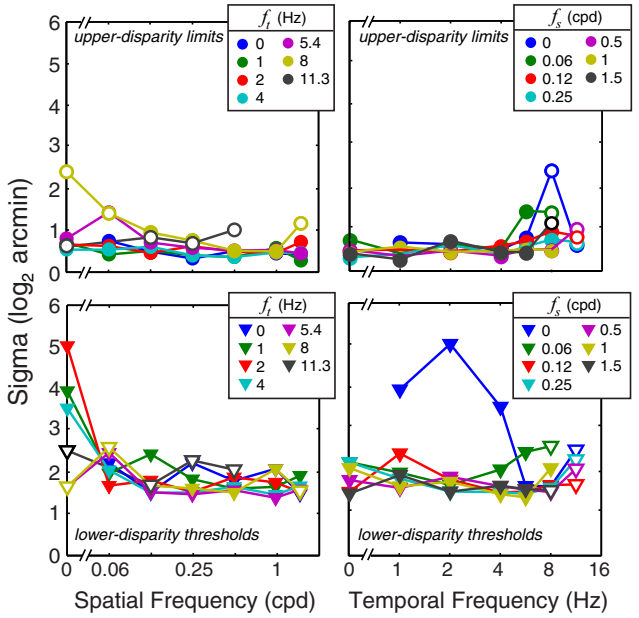

Figure 7. Slopes of psychometric functions for lower-disparity thresholds and upper-disparity limits. Each panel plots σ of the cumulative Gaussian (in units of $\log_2$ arcmin) that provided the best fit to the psychometric data for that condition. Top, σ for the upper-disparity limits. In those cases, the slopes were actually negative because percent correct decreased with increasing disparity. We inverted the signs for plotting. Bottom, σ for the lower-disparity thresholds. In those cases, the slopes were positive because percent correct increased with increasing disparity. Left and Right, Values when plotted against spatial and temporal frequency, respectively. Filled and unfilled symbols correspond to conditions where reliable and marginally reliable fits were obtained. Because the σ values are in log units, constant values mean that psychometric functions in semilog coordinates (percent correct as a function of log disparity) have constant slope. For nearly all spatial and temporal frequencies, the psychometric functions were much steeper for the upper limits than for the lower thresholds. σ values are ~0.5 for the upper-disparity limits and ~2.0 for the lower-disparity thresholds; the only notable exception is when the spatial frequency (f_s) is 0, where σ tends to be much higher.

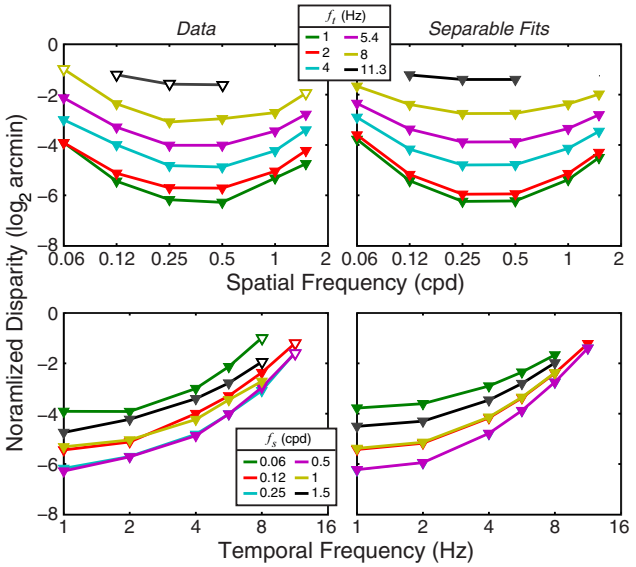

Figure 8. Separability of lower-disparity thresholds. Left, Lower-disparity data from Figure 5. Filled and unfilled symbols represent reliable and marginally reliable data, respectively. The data at $f_s = 0$ cpd and $f_t = 0$ Hz have been excluded (see text). Right, Products of functions in spatial frequency and temporal frequency (see text for best-fitting equations). Top, Plot of data and fits as a function of spatial frequency. Bottom, Plots as a function of temporal frequency. The ordinate values are disparity in $\log_2$ arcmin after the largest thresholds were set to 1 arcmin.

In conclusion, the results of the spatial- and temporal-step experiments show that the low spatial and temporal resolutions observed in the main experiment do in fact generalize to other stimuli and tasks. However, the generalization is not quantitatively predicted by a simple analysis that assumes linearity; rather, the ability to assess the sharpness of a spatial step is better than predicted, whereas the ability to assess the speed of a temporal step is poorer than predicted.

Lower-disparity thresholds

The lower-threshold data, $L(f_s, f_t)$, appear to be separable; that is, well fit by the product of two 1D functions in spatial and temporal frequency. We investigated this by fitting the product of two second-order polynomials with arguments of $\log f_s$ and $\log f_t$ —$G(\log f_s) = m(\log f_s)^2 + n(\log f_s) + o$ and $H(\log f_t) = p(\log f_t)^2 + q(\log f_t) + r$ —to the data and then assessing the goodness of fit between the data and the product. We used the average lower-disparity threshold data after excluding unreliable points and points at $f_s = 0$ and $f_t = 0$ because there can be no data at $(f_s, f_t) = (0,0)$. We also first normalized the data such that the highest threshold value was 1. The best-fitting parameter values were $(m–r) = (0.097, 0.294, -1.258, -0.301, 0.107, 4.278)$. The smallest RMS error was 0.257. Figure 8 shows the results graphically. The data (left) and predictions from separable functions (right) are very similar. Therefore, the lower-disparity threshold data are separable, suggesting that the underlying neural mechanisms are also separable. Of course, it might be possible that such data could be consistent with inseparable mechanisms, but separable mechanisms provide the most parsimonious account.

We conducted a similar analysis on spatiotemporal luminance contrast sensitivity (Kelly, 1979). The smallest RMS value (after data normalization to allow comparison with the disparity results) was 0.995, which shows that luminance contrast sensitivity is not nearly as separable in space and time as the disparity thresholds.

Upper-disparity limits

We first determined the upper limits as a function of spatial frequency by looking at the data when temporal frequency is 0 Hz. Then, we determined the upper limits as a function of temporal frequency by looking at the data when spatial frequency is 0 cpd. Finally, we considered the case where both spatial frequency and temporal frequency are non-zero.

The upper limits at 0 Hz could in principle be determined by the largest fusable disparity (i.e., an amplitude limit) or by a spatial disparity-gradient limit (Tyler, 1973; Burt and Julesz, 1980). The spatial disparity gradient for our triangular-wave stimulus is as follows:

$$\nabla d_s = 2af_s \quad (4)$$

where a is peak-to-trough disparity amplitude and f_s is spatial frequency of the corrugation. The data are consistent with a spatial disparity-gradient limit of $\nabla d_s = 1.2$. Therefore, when temporal frequency was 0 Hz, the upper limit was reached whenever the disparity changed by >1.2° for every 1° change in spatial position. This is very consistent with previous estimates of a spatial disparity-gradient limit (Tyler, 1973; Burt and Julesz, 1980; Ziegler et al., 2000; Filippini and Banks, 2009) and not consistent with an amplitude limit.

The spatial disparity-gradient limit is a byproduct of using windowed correlation to estimate disparity (Banks et al., 2004; Filippini and Banks, 2009; Allenmark and Read, 2010). Each eye's image is sampled by a spatial windowing function and then cor-

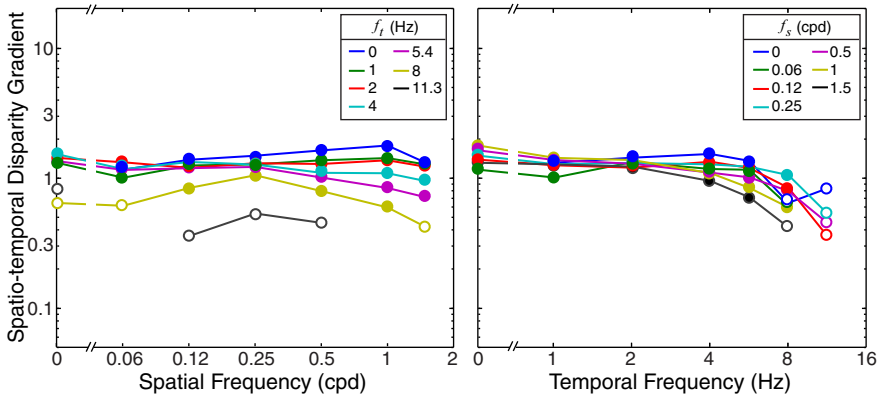

Figure 9. Spatiotemporal disparity gradient (∇d_{st}) as a function of spatial and temporal frequency. Left, Spatiotemporal disparity gradient associated with upper-disparity limits as a function of the spatial frequency of the corrugation waveform. The data have been averaged across subjects. Blue, green, red, purple, cyan, yellow, and gray symbols represent the data when temporal frequency was 0, 1, 2, 4, 5.4, 8, and 11.3 Hz, respectively. Right, Spatiotemporal disparity gradient associated with upper limits as a function of temporal frequency. Again, the data have been averaged across subjects. Blue, green, red, purple, cyan, yellow, and gray symbols represent the data when spatial frequency was 0, 0.06, 0.12, 0.25, 0.5, 1, and 1.5 cpd, respectively.

related. The samples are shifted horizontally to find the relative position yielding the highest correlation, and that horizontal shift corresponds to the disparity estimate in that region of the stimulus. The correlation between the two eyes' images is greatest when the images are identical; that is, when the stimulus is a frontoparallel surface. If the stimulus surface is slanted, the two eyes' images differ and correlation falls. In other words, as the spatial disparity gradient increases, the correlation decreases and the brain's ability to estimate depth from disparity decreases correspondingly. Once the gradient exceeds a critical value (in our case, ~1.2), the correlation signal cannot be extracted from background noise and the depth percept collapses.

Temporal variations in disparity also cause differences in the two eyes' images and such variations should also affect correlation. Once the variation exceeds a critical value, one would expect a collapse of the depth percept. To determine whether this occurs, we examined the effect of temporal frequency on upper-disparity limits by looking at the data when spatial frequency is 0 cpd. Those upper limits are also well predicted by a disparity-gradient limit: in this case, a temporal disparity-gradient limit as follows:

$$\nabla d_t = 2af_t = 11 \tag{5}$$

where f_t is temporal frequency and the units of ∇d_t are deg/s. The observation of a temporal disparity-gradient limit means that disparity estimation breaks down whenever disparity changes at a rate >11°/s (or 0.66 arcmin/msec). Similar to the spatial disparity-gradient limit, the visual system cannot compute depth from disparity when disparity changes too quickly in time. The temporal disparity-gradient limit is a necessary byproduct of estimating disparity by correlating the two eyes' images over time.

What happens with various combinations of spatial and temporal frequency? Is there a fundamental limit that describes the breakdown of disparity processing for all spatiotemporal frequencies? Such a limit would mean that depth perception would collapse whenever the spatiotemporal disparity gradient exceeded a certain value, presumably ~1.2. To investigate this, we need to know how many cycles of the disparity-defined waveform pass through a block of space–time for different combinations of spatial and temporal frequency (f_s and f_t, respectively). It is useful to have the two frequencies in the same units, so we multiply f_t by a constant k with units of sec/deg: $\hat{f}_t = kf_t$. The periods in space and transformed time are then, respectively, $p_s = 1/f_s$ and $\hat{p}_t = 1/\hat{f}_t$. The period for a spatiotemporal waveform p_{st} is given by the following:

$$\left(\frac{1}{p_{st}}\right)^2 = \left(\frac{1}{p_s}\right)^2 + \left(\frac{1}{\hat{p}_t}\right)^2 \tag{6}$$

Substituting for p_s and $\hat{p}_t$ and taking the square root:

$$f_{st}\sqrt{f_s^2 + \hat{f}_t^2}, \tag{7}$$

the spatiotemporal disparity gradient is then:

$$\nabla d_{st} = 2af_{st} = 2a\sqrt{f_s^2 + \hat{f}_t^2}. \tag{8}$$

Figure 9 plots the spatiotemporal disparity gradient ∇d_{st} as a function of spatial and temporal frequency for the upper-limit data when $k = 0.12$ (the value that minimizes differences between data points). The upper limit does indeed occur at a gradient of ~1.2 for nearly all combinations of spatial and temporal frequency. The only noteworthy deviation occurs at 11.3 Hz (black symbols in Fig. 9, left, and rightmost symbols in Fig. 9, right) where the lower-disparity thresholds started to impinge on the upper limits. The disparity amplitude at which the spatiotemporal disparity-gradient limit is exceeded is as follows:

$$a = \frac{1.2}{2\sqrt{f_s^2 + \hat{f}_t^2}} \tag{9}$$

Note that this equation is inseparable, unlike the lower-disparity thresholds.

We conclude, therefore, that a spatiotemporal, disparity-gradient limit exists and that it is a fundamental constraint on the set of stimuli for which disparity can be estimated.

Neural mechanisms underlying disparity estimation

As mentioned earlier, the visual system's use of correlation to estimate disparity imposes constraints on the spatial and temporal variations in disparity that can be reliably perceived. We tested this claim by constructing a simple spatiotemporal correlation model and investigating whether its behavior is consistent with our human data. Even though the lower-disparity thresholds and upper-disparity limits (summarized by the spatiotemporal disparity-gradient limit) are quite different phenomena, we examined the hypothesis that both are a byproduct of estimating disparity by correlation. The model is a spatiotemporal windowed cross-correlator the computation of which is analogous to the disparity-energy calculation performed in visual cortex (Ohzawa et al., 1990; Anzai et al., 1999; Prince and Eagle, 2000).

The stimuli presented to the model were the same random-dot stereograms used in the main experiment. The stimuli $L(x, y, t)$ and $R(x, y, t)$ were first convolved with the eye's point-spread function (PSF; Geisler and Davila, 1985). The effect of the PSF on the results was trivial, so we do not include that operation in the rest of the mathematical description. Because our lower-disparity thresholds were space–time separable (Fig. 5), we modeled the spatiotemporal windowing before cross-correlation as separable windows. Therefore, the windowing was the product of spatial and temporal windowing functions as follows:

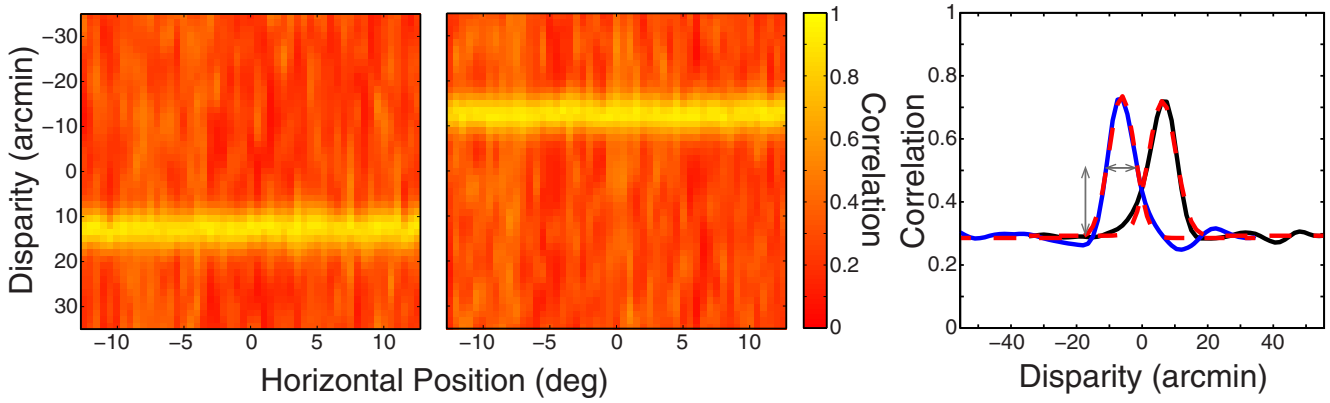

Figure 10. Correlation as a function of horizontal position and disparity for an example stimulus. Left, Correlation plotted as a function of horizontal position and disparity for the vertical stimulus position that contains a corrugation peak. Lighter colors represent higher correlations. Middle, The same but for a corrugation trough. Right, Averaged vertical cross-sections of the correlation plots on the left and right. The black and blue curves represent cross-correlation outputs averaged across horizontal position; the blue one for the trough and black one for the peak. The dashed red curves represent four-parameter Gaussians fit to those two parts of the output. The vertical arrow indicates the half maximum and the horizontal arrow indicates the full width at the half maximum.

$$w_L(x, y, t) = u_L(x, y)v(t)$$
$$w_R(x, y, t) = u_R(x, y)v(t) \quad (10)$$

The initial input to the cross-correlator is a video sequence of random-dot stereograms. A temporal window was implemented to convert a stack of video frames into a single image. The window was a fast-rise, slow-decay function similar to temporal filters in the luminance domain (de Lange, 1958; Kelly, 1961) and was identical in the two eyes:

$$v(t) = \left(\frac{t}{\tau}\right)^2 \exp\left(\frac{-t}{\tau}\right) \quad (11)$$

The temporal windowing function weighted individual video frames and integrated the weighted frames into one image. Because each eye received its own video stream, the result after temporal windowing was two temporally smeared images, one for each eye. Those images were then spatially windowed in preparation for cross-correlation. Spatial windowing was implemented with isotropic 2D Gaussians:

$$u_L(x, y) = \exp\left[-\left(\left(\frac{x - x_L}{\sqrt{2}\sigma}\right)^2 + \left(\frac{y - y_0}{\sqrt{2}\sigma}\right)^2\right)\right]$$
$$u_R(x, y) = \exp\left[-\left(\left(\frac{x - x_R}{\sqrt{2}\sigma}\right)^2 + \left(\frac{y - y_0}{\sqrt{2}\sigma}\right)^2\right)\right] \quad (12)$$

The horizontal positions of the windows in the two eyes were determined by x_L and x_R. The vertical positions were always the same (y_0). The functions u_L and u_R therefore determined the regions in the two eyes' images that would be correlated.

Window position was fixed in the left eye's image and moved horizontally in the right eye's image. For each pair of positions, correlation was computed as follows:

$$c(\delta_x) = \frac{\sum_{(x,y)\in w_L}(L(x, y) - \mu_L)(R(x - \delta_x, y) - \mu_R)}{\sqrt{\sum_{(x,y)\in w_L}(L(x, y) - \mu_L)^2}\sqrt{\sum_{(x,y)\in u_R}(R(x - \delta_x, y) - \mu_R)^2}} \quad (13)$$

where $\delta_x = x_L - x_R$. Normalization by the mean luminance within both eyes' windows ensured that correlation was between −1 and 1. The disparity estimate was the horizontal offset δ_x between the two subregions that yielded the highest correlation.

The free parameters in Equations 10 and 11 are σ and τ, which represent the spatial and temporal extents of the windowing functions, respectively. We determined the values of σ and τ that produced lower-disparity thresholds and upper-disparity limits that were most similar to our psychophysical data. The model was run using 20 values of τ and 39 values of σ, resulting in 780 σ–τ combinations.

We had to restrict the psychophysical data that were used to evaluate the model. From the lower-disparity thresholds, we excluded spatial frequencies of 0, 0.06, and 0.12 cpd because the low-frequency attenuation evident in the human data (Fig. 5) cannot be caused by spatial windowing. (Notably, such attenuation is also not observed in physiological recordings from area V1; Nienborg et al., 2004). From the upper-disparity limit data, we excluded spatial frequency of 0 cpd. Therefore, 28 combinations of spatial and temporal frequency were used to model the lower thresholds and 42 combinations to model the upper limits.

For each spatial frequency, temporal frequency, and disparity of the stimulus, cross-correlation was performed at 51 evenly spaced horizontal positions on a 1200-pixel (30°)-wide stimulus. The two vertical positions at which correlation was computed corresponded to a peak and a trough in the corrugation waveform. For gratings with non-zero temporal frequency, the peak of the temporal window was aligned with the video frame corresponding to a triangular waveform with phase equal to zero. Figure 10, left and middle, provide examples of the correlations as a function of horizontal position and disparity. The 51 correlation distributions were averaged across horizontal position to create the average correlation distributions, one for the corrugation peak and one for the trough (Figure 10, right). The average distributions were fit with four-parameter Gaussians; the parameters correspond to maximum correlation disparity at which the maximum occurred, SD, and a uniform pedestal representing a noise floor. We varied the disparity of the stimulus and then assessed discriminability by comparing the difference in the means of the two Gaussians—one for the corrugation peak and one for the trough—relative to their full width at half maximum. A stimulus was considered visible when the difference of means was larger than the average of their full widths at half maximum.

The model behaved differently for the lower-disparity thresholds and upper-disparity limits. For the lower-disparity thresholds, an increase in stimulus disparity yielded a monotonic increase in the difference between the two correlation distri-

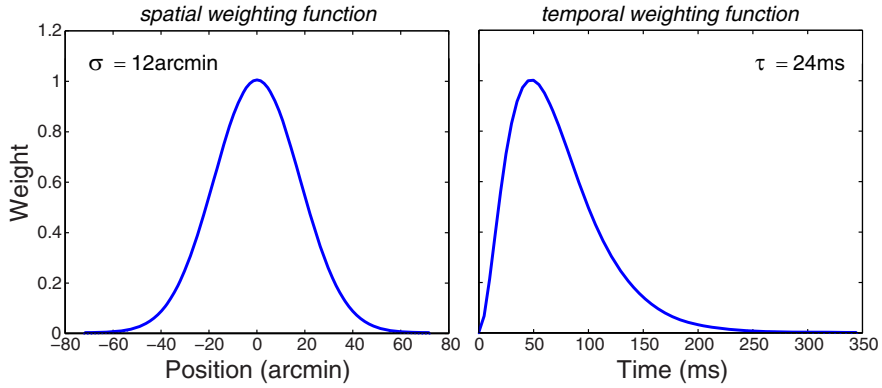

Figure 11. Best-fitting spatial and temporal windowing functions. For the spatial window, we assumed a 2D Gaussian with a spread parameter of σ (Equation 11). For the temporal window, we assumed a fast-rise, slow-decay function with a spread parameter of τ (Equation 10). Left and Right, Spatial ($\sigma = 12$ arcmin) and temporal ($\tau = 24$ ms) functions that provided the best fit to the data, respectively.

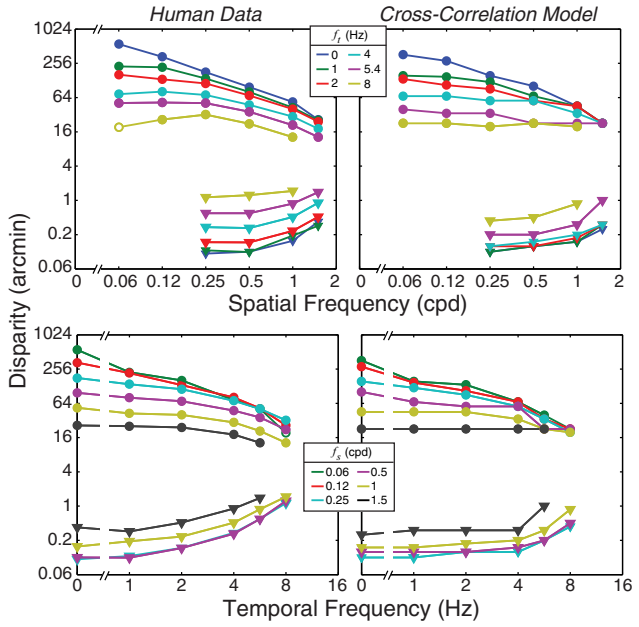

Figure 12. Comparison of human data and cross-correlation model. Left, Human data from the main experiment. The top plane plots the data as a function of spatial frequency and the bottom as a function of temporal frequency. Triangles represent the lower-disparity thresholds and circles the upper-disparity limits. The data at $f_s = 0$ have been excluded. The lower-disparity thresholds at $f_s = 0.06$ and 0.12 have also been excluded (see text for explanation). The data at a few other combinations of spatial and temporal frequency are not plotted here because the cross-correlation model did not find solutions for them, so they were assigned either a lower-disparity threshold of 4 arcmin or an upper-disparity limit of 1024 arcmin. Those points were included in the error minimization to find the optimal values of σ and τ. Right, Lower-disparity thresholds and upper-disparity limits obtained with $\sigma = 12$ arcmin and $\tau = 24$ ms. The ordinate values have been normalized such that they correspond to the same units in the human data.

butions relative to their widths (i.e., correlation peaks were consistently well defined, but superimposed at small stimulus disparities). Occasionally, we could not obtain a lower threshold, so we set the threshold in those cases to a very high value (1024 arcmin). The upper-disparity limits behaved differently. As the disparity amplitude was increased beyond a critical value, the correlation distributions no longer exhibited clear correlation peaks. Therefore, the widths of the Gaussians for the peak and trough of the waveform increased dramatically and discrimination by the full-width, half-maximum rule fell precipitously. We defined the disparity at which this occurred as the model's upper-disparity limit. For a few conditions, we could not obtain an upper limit, so we set the value very low (4 arcmin).

We ran the model for 780 combinations of σ and τ and generated separate outputs for lower thresholds and upper limits. We determined the quality of fit between the model and human data for each σ–τ combination. Before calculating error between the human data and model output, both sets of values were first normalized by the maximum observed disparity values within each dataset. Errors between the data and model were calculated using the absolute value of the errors between the logarithm of the human data and the model's corresponding thresholds or limits. We used the same method for the lower-threshold data except that we inverted the disparity values before normalizing. Finally, we found the values of σ and τ that minimized the product of the errors for the lower thresholds and upper limits. The optimal values were 12 arcmin and 24 ms. Interestingly, we obtained approximately the same optimal values when we fit only the lower-disparity thresholds or only the upper-disparity limits. This means that the same model with similar parameters provides a good fit to both sets of data. The spatial and temporal windows associated with the optimal values are shown in Figure 11. These best-fitting functions manifest spatiotemporal filtering in early pathways and windowing at the stage of binocular correlation (Nienborg et al., 2005).

Figure 12 shows the human data (left) and cross-correlation output (right) with the optimal values of σ and τ. The model yielded upper-disparity limits that were quite similar to those exhibited by humans. The model and human observers both exhibited a spatiotemporal disparity-gradient limit, because correlation broke down whenever disparity changed too much per unit space–time. The model produced lower-disparity thresholds that were reasonably similar to human lower thresholds. We did not attempt, however, to fit the rise in human thresholds at low spatial frequencies because a model of this form will not produce band-pass sensitivity; the lower-frequency sensitivity loss evident in Figure 5 must have some other cause. We note that attenuation at low spatial frequency is also not observed in physiological recordings from area V1 (Nienborg et al., 2004).

The similarity between the data and model output strongly suggests that the lower-disparity thresholds and upper-disparity limits are both determined by the same spatiotemporal windowing involved in correlating the input to the two eyes. We claim, therefore, that the volume of visible spatial and temporal variations in disparity is determined to large degree by the spatiotemporal windowing function used in disparity estimation.

Discussion

Comparison with V1 neurons

The responses of primate V1 neurons to spatial modulations in disparity are low-pass; that is, responsiveness is similar at low spatial frequencies and declines at higher frequencies (Nienborg et al., 2004). Furthermore, most neurons cannot respond to corrugation frequencies >1 cpd. Unfortunately, few neurons near

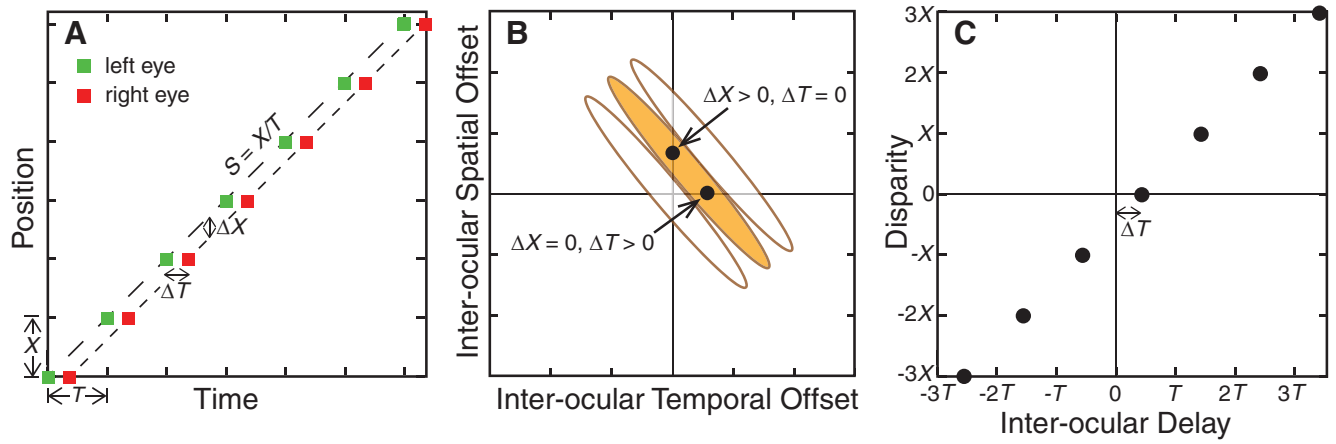

Figure 13. Stimulus and models of stereopsis with interocular delays. ***A***, Space–time stimulus that gives rise to the Mach-Dvorak effect. The horizontal positions of the stimuli presented to the left and right eyes are plotted as a function of time. The green and red squares represent the presentations to the left and right eyes, respectively. Stimulus speed is X/T where T is the temporal interval at which stimuli are repeated and X is the displacement for each repetition. Each left–right stimulus pair is presented at the same position and therefore has a spatial disparity of 0, but they are presented at different times to the two eyes with an interocular delay of ΔT. The depth perceived in this stimulus is often commensurate with ΔX. ***B***, Joint disparity-motion encoding model. The orange region represents the excitatory part of an example receptive field and the white regions the inhibitory parts. Joint sensors are inseparable in space and time so the preferred spatial disparity (ordinate) varies as a function of the interocular delay (abscissa). The black dot at $\Delta X = 0$, $\Delta T > 0$ represents a stimulus with a non-zero temporal delay, but a spatial disparity of 0, such as the stimulus in ***A***. The black dot at $\Delta X > 0$, $\Delta T = 0$ represents a stimulus with no temporal delay and a non-zero spatial disparity. The response of the joint disparity-motion sensor to those two stimuli would be the same. ***C***, Spatiotemporal filtering model. A stimulus like the one in ***A*** is presented. Due to the extended temporal window associated with stereo matching (Fig. 11), a number of potential matches across time affect the estimated disparity. The black circles represent a series of such matches relative to one left-eye image in ***A***. They have different spatial disparities depending on the interocular delay associated with the match. The disparity estimate is given by the weighted average of the various matches distributed over time with the greatest weight assigned to an interocular delay of 0. The stimulus would appear displaced in depth in this example because the time-weighted average is negative.

the fovea have been sampled, so one cannot compare directly the spatial resolution we observed psychophysically with these physiological data. However, the lower resolution in primate V1 surely imposes a constraint on performance and is thus consistent with our observation of low spatial-frequency resolution in humans.

Interestingly, neurons with larger receptive fields have lower stereoresolution (Nienborg et al., 2004). This phenomenon can be explained with consideration of cross-correlation. As the receptive-field size of a neuron increases, a larger proportion of the period of a corrugation waveform is captured and the neuron's response tends toward the response to a uniform disparity. Therefore, the size of a receptive field imposes a limit on the highest discriminable spatial frequency. This differs from the determinants of resolution in the luminance domain, where there is essentially no correlation between receptive-field size and acuity due to substructure in the receptive field (Nienborg et al., 2004).

The response of primate V1 neurons to temporal modulations of disparity has also been investigated (Nienborg et al., 2005). Spatial corrugations of low spatial frequency were presented at various temporal frequencies. The highest temporal frequency to which reliable responses were obtained was generally <10 Hz. Low temporal resolution for disparity could be a byproduct of cross-correlating temporally filtered inputs from the two eyes (Nienborg et al., 2004). We modeled this with a temporal window with a relatively long time constant.

Stereopsis with interocular time delays

A horizontally moving stimulus presented stroboscopically to the same positions in the two eyes is perceived as moving in the frontal plane. When the presentation to one eye is delayed (Fig. 13*A*), the perceived trajectory is displaced in depth by an amount approximately proportional to the interocular delay and object speed (Lee, 1970; Ross and Hogben, 1974; Burr and Ross, 1979; Morgan, 1979; Read and Cumming, 2005a; Hoffman et al., 2011). This Mach-Dvorak effect is interesting because the temporal delay between the eyes causes a spatial disparity of zero to be interpreted as non-zero. The effect has regained attention recently because it is frequently experienced in stereo 3D cinema and television (Hoffman et al., 2011).

For stereo matching to occur, the interocular delay in the Mach-Dvorak effect must be shorter than the integration time of the underlying computation. Interocular delays up to 50–80 ms yield reliable depth percepts; longer delays do not (Lee, 1970; Burr and Ross, 1979; Morgan, 1979; Hoffman et al., 2011). Therefore, the temporal window for stereoscopic matching must be as long as 50–80 ms. The window estimated from our data (Fig. 11) falls to half its peak value 50 ms after the peak, so it is consistent with the observed effects. A different technique revealed a time constant of 21 ms for a Gaussian temporal window (Read and Cumming, 2005a); this, too, is consistent with the duration estimated from the analysis of our data.

The neural computation underlying the Mach-Dvorak effect and the related Pulfrich effect has been controversial. Some have argued that these effects are the byproduct of joint disparity-motion sensors with receptive fields that are rotated relative to the space–time axes (i.e., they are inseparable in space and time; Qian and Andersen, 1997; Anzai et al., 2001; Qian and Freeman, 2009). Because of the rotation, the preferred spatial disparity varies with interocular delay (Fig. 13*B*). Therefore, a temporal delay with a zero spatial disparity ($\Delta X = 0$, $\Delta T > 0$) yields the same response as no temporal delay with a non-zero spatial disparity ($\Delta X > 0$, $\Delta T = 0$). Therefore, both combinations yield the perception of displacement in depth. Such joint disparity-motion sensors have been observed physiologically in areas 17 and 18 of the cat (Anzai et al., 2001) and areas MT and MST in the monkey (Pack et al., 2003). Others have argued that the Mach-Dvorak and Pulfrich effects are caused by spatiotemporal filtering (and/or windowing). Filtering by space–time-separable neurons yields an estimate of spatial disparity that differs from zero because multi-

ple stereo matches with different interocular times and different disparities affect the estimated disparity (Fig. 13*C*) (Read and Cumming, 2005a; Hoffman et al., 2011). Indeed, most neurons in area V1 of macaque exhibit space–time separability and therefore have the same spatial disparity preference for all interocular delays (Pack et al., 2003; Read and Cumming, 2005b). We found that the boundary conditions for estimating disparity (both lower thresholds and upper limits) are consistent with a cross-correlation model with space–time separable windows. Unfortunately, this finding is not directly relevant to the joint disparity-motion hypothesis because all of the data were collected with zero interocular delay. Without delay, rotated and nonrotated receptive fields behave the same and cannot be distinguished.

Finally, we note that the model used by Read and Cumming (2005a) to account for the Mach-Dvorak effect cannot explain an important aspect of our data. Their model assumes that the visual system averages disparities over time according to a Gaussian weighting function (see their Fig. 3). This model elegantly accounts for several aspects of the Mach-Dvorak phenomenon and could in principle also account for the low-pass nature of our lower-disparity thresholds when plotted as a function of temporal frequency (Fig. 5*A*). However, it cannot account for the manner in which the upper-disparity limits change with temporal frequency. Specifically, it cannot explain why increasing disparity amplitude at a given temporal frequency causes a decrease in performance. To explain this requires consideration of how interocular correlation is computed over time and how that necessarily leads to a spatiotemporal disparity-gradient limit.

Appearance

Using spatiotemporal correlation, the brain is unable to estimate disparities that change too much over a small space–time interval. This means that the visual system cannot derive a depth percept from disparities that frequently arise in the natural environment: object boundaries, multilayered scenes, etc. Despite this, we perceive depth changes as sharp in space and time, suggesting that we are unaware of the missing information. This apparent discrepancy can be understood with an analogy in color vision. Spatial and temporal resolution for hue changes is notably poorer than for luminance changes (van der Horst and Bouman, 1969), yet hue and luminance are generally perceived in sharp spatial and temporal registration. Several illusions show that the hue of a surface tends to spread perceptually to the nearest surrounding luminance edges (Pinna et al., 2003; Kanai et al., 2006). As a result, the hue boundary appears relatively sharp, making us unaware of the poor resolution associated with hue processing. Something like this seems to occur in stereopsis: we are unaware of relatively poor spatial and temporal resolution because other depth cues such as occlusion influence the computation, yielding an experience of sharpness in space and time. Consistent with this idea, the perceived location of an edge is determined more by the position of a luminance step than by the position of a disparity step (Robinson and MacLeod, 2013).

Video compression

We can only see a small fraction of the information incident on our eyes. The inability to see detail finer than 50 cpd has had significant impact on the design of visual displays. For example, the pixel density of high-definition television is slightly $<$50 cpd when the television is viewed at the prescribed distance (Poynton, 2012); there would be no point in presenting more pixels because viewers could not resolve them. We are also limited in the ability to perceive variations in luminance over time, and this temporal limit affects the design of lighting and visual displays, which typically have refresh rates of 60–70 Hz (Farrell et al., 1987).

Data compression involves encoding information using fewer bits than in the original representation, thereby reducing the use of resources such as data storage space and transmission capacity. Video-compression algorithms such as MPEG reduce the amount of required data by only storing differences between video frames and by taking advantage of properties of human vision, including the spatiotemporal contrast-sensitivity function and its dependence on luminance and chromatic variation (van der Horst and Bouman, 1969; Watson et al., 1986; Poynton, 2012). Because the visual system's resolution for chromatic information is relatively low, hue information can be represented with fewer bits than luminance information.

Our data show that humans are very limited in the ability to perceive spatial and temporal disparity variations. Knowing these limitations—summarized in Figures 5 and 6—could prove advantageous for compression of stereoscopic video. The input data from two cameras could be split into luminance/hue data and disparity data. Having split the data in this fashion, compression algorithms could then take advantage of limitations in processing luminance and hue separately from limitations in processing disparity. In particular, the disparity data could be spatiotemporally filtered to eliminate spatial and temporal frequencies that cannot be perceived. The compressed luminance/hue and compressed disparity data could then be transmitted or stored with significant savings in bandwidth.

Conclusion

The range of spatial and temporal variations in disparity that are perceivable is quite limited. Two phenomena demarcate the boundary conditions: (1) a minimum-disparity threshold below which no depth is perceived and (2) a spatiotemporal disparity-gradient limit above which no coherent depth can be perceived. A simple cross-correlation model with a spatiotemporal windowing function—analogous to the disparity-energy model—is sufficient to explain these seemingly distinct properties of stereopsis.

Notes

Supplemental material for this article is available at http://bankslab.berkeley.edu/projects/projectlinks/disparitygradient.html. This URL links to a video illustrating the spatiotemporal disparity gradient limit. This material has not been peer reviewed.

References

Allenmark F, Read JC (2010) Detectability of sine- versus square-wave disparity gratings: a challenge for current models of depth perception. J Vis 10:17. Medline

Anzai A, Ohzawa I, Freeman RD (1999) Neural mechanisms for processing binocular information. I. Simple cells. J Neurophysiol 82:891–908. Medline

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

Banks MS, Gepshtein S, Landy MS (2004) Why is spatial stereoresolution so low? J Neurosci 24:2077–2089. CrossRef Medline

Bradshaw MF, Rogers BJ (1999) Sensitivity to horizontal and vertical corrugations defined by binocular disparity. Vision Res 39:3049–3056. CrossRef Medline

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–436. CrossRef Medline

Burr DC, Ross J (1979) How does binocular delay give information about depth? Vision Res 19:523–532. CrossRef Medline

Burt P, Julesz B (1980) A disparity gradient limit for binocular fusion. Science 208:615–617. CrossRef Medline

Campbell FW, Green DG (1965) Optical and retinal factors affecting visual resolution. J Physiol 181:576–593. Medline

Cormack LK, Stevenson SB, Schor CM (1991) Interocular correlation, luminance contrast and cyclopean processing. Vision Res 31:2195–2207. CrossRef Medline

de Lange H (1958) Research into the dynamic nature of the human fovea-cortex systems with intermittent and modulated light. I. Attenuation characteristics with white and colored light. J Opt Soc Am 48:777–784. CrossRef Medline

De Valois RL, Albrecht DG, Thorell LG (1982) Spatial frequency selectivity of cells in macaque visual cortex. Vision Res 22:545–559. CrossRef Medline

Farrell JE, Benson BL, Haynie CR (1987) Predicting flicker thresholds for video display terminals. Proc SID 28:1–5.

Filippini HR, Banks MS (2009) Limits of stereopsis explained by local cross-correlation. J Vis 9:8.1–18. CrossRef Medline

Fleet DJ, Wagner H, Heeger DJ (1996) Neural encoding of binocular disparity: Energy models, position shifts and phase shifts. Vision Res 36:1839–1857. CrossRef Medline

Geisler WS, Davila KD (1985) Ideal discriminators in spatial vision: two-point stimuli. J Opt Soc Am A 2:1483–1497. CrossRef Medline

Hoffman DM, Karasev VI, Banks MS (2011) Temporal presentation protocols in stereoscopic displays: flicker visibility, perceived motion and perceived depth. J Soc Inf Disp 19:271–297. CrossRef Medline

Kanade T, Okutoni M (1994) A stereo matching algorithm with an adaptive window: theory and experiment. IEEE Trans Pattern Anal Machine Intell 16:920–932. CrossRef

Kanai R, Wu D, Verstraten FJA, Shimojo S (2006) Discrete color filling beyond luminance gaps along perceptual surfaces. J Vis 6:1380–1395. CrossRef Medline

Kelly DH (1961) Visual responses to time-dependent stimuli. I. Amplitude sensitivity measurements. J Opt Soc Am 59:422–429. Medline

Kelly DH (1979) Motion and vision. II. Stabilized spatio-temporal threshold surface. J Opt Soc Am 69:1340–1349. CrossRef Medline

Lankheet MJ, Lennie P (1996) Spatio-temporal requirements for binocular correlation in stereopsis. Vision Res 36:527–538. CrossRef Medline

Lee DN (1970) Spatio-temporal integration in binocular-kinetic space perception. Vision Res 10:65–78. CrossRef Medline

Morgan MJ (1979) Perception of continuity in stroboscopic motion: A temporal frequency analysis. Vision Res 19:491–500. CrossRef Medline

Movshon JA, Thompson ID, Tolhurst DJ (1978a) Spatial summation in the receptive fields of simple cells in the cat's striate cortex. J Physiol 283:53–77. Medline

Movshon JA, Thompson ID, Tolhurst DJ (1978b) Receptive field organization of complex cells in the cat's striate cortex. J Physiol 283:79–99. Medline

Nienborg H, Bridge H, Parker AJ, Cumming BG (2004) Receptive field size in V1 neurons limits acuity for perceiving disparity modulation. J Neurosci 24:2065–2076. CrossRef Medline

Nienborg H, Bridge H, Parker AJ, Cumming BG (2005) Neuronal computation of disparity in V1 limits temporal resolution for detecting disparity modulation. J Neurosci 25:10207–10219. CrossRef Medline

Norcia AM, Tyler CW (1984) Temporal frequency limits for stereoscopic apparent motion processes. Vision Res 24:395–401. CrossRef Medline

Ohzawa I, DeAngelis GC, Freeman RD (1990) Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. Science 249:1037–1041. CrossRef Medline

Pack CC, Born RT, Livingstone MS (2003) Two-dimensional substructure of stereo and motion inter-actions in macaque visual cortex. Neuron 37:525–535. CrossRef Medline

Patterson R, Ricker C, McGary J, Rose D (1992) Properties of cyclopean motion processing. Vision Res 32:149–156. CrossRef Medline

Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437–442. CrossRef Medline

Pinna B, Werner JS, Spillmann L (2003) The watercolor effect: a new principle of grouping and figure-ground organization. Vision Res 43:43–52. CrossRef Medline

Poynton C (2012) Digital video and HD: algorithms and interfaces. Waltham, MA: Morgan Kaufman.

Prince SJ, Eagle RA (2000) Stereo correspondence in one-dimensional Gabor stimuli. Vision Res 40:913–924. CrossRef Medline

Qian N, Freeman RD (2009) Pulfrich phenomena are coded effectively by a joint motion-disparity process. J Vis 9:24.1–16. CrossRef Medline

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

Read JC, Cumming BG (2005a) The stroboscopic Pulfrich effect is not evidence for the joint encoding of motion and depth. J Vis 5:417–434. CrossRef Medline

Read JC, Cumming BG (2005b) Effect of interocular delay on disparity-selective V1 neurons: relationship to stereoacuity and the Pulfrich effect. J Neurophysiol 94:1541–1553. CrossRef Medline

Richards W (1972) Response functions to sine- and square-wave modulations of disparity. J Opt Soc Am 62:907–911. CrossRef

Robinson AE, MacLeod DI (2013) Depth and luminance edges interact. J Vis 13:11.1–13. CrossRef Medline

Robson JG (1966) Spatial and temporal contrast-sensitivity functions of the visual system. J Opt Soc Am 56:1141–1142. CrossRef

Ross J, Hogben JH (1974) Short-term memory in stereopsis. Vision Res 14:1195–1201. CrossRef Medline

Tyler CW (1973) Stereoscopic vision: cortical limitations and a disparity scaling effect. Science 181:276–278. CrossRef Medline

Tyler CW (1974) Depth perception in disparity gratings. Nature 251:140–142. CrossRef Medline

van der Horst GJ, Bouman MA (1969) Spatiotemporal chromaticity discrimination. J Opt Soc Am 59:1482–1488. CrossRef Medline

Watson AB, Ahumada AJ (2011) Blur clarified: a review and synthesis of blur discrimination. J Vis 11:pii:10. CrossRef Medline

Watson AB, Ahumada AJ, Farrell JE (1986) Window of visibility: psychophysical theory of fidelity in time-sampled visual motion displays. J Opt Soc Am 3:300–307. CrossRef

Ziegler LR, Hess RF, Kingdom FA (2000) Global factors that determine the maximum disparity for seeing cyclopean surface shape. Vision Res 40:493–502. CrossRef Medline