

# Contributions of monocular and binocular cues to distance discrimination in natural scenes

**Brian C. McCann**

Texas Advanced Computing Center,  
Center for Perceptual Systems and Department of  
Psychology, University of Texas at Austin, Austin, TX, USA



**Mary M. Hayhoe**

Center for Perceptual Systems and Department of  
Psychology, University of Texas at Austin, Austin, TX, USA

**Wilson S. Geisler**

Center for Perceptual Systems and Department of  
Psychology, University of Texas at Austin, Austin, TX, USA

Little is known about distance discrimination in real scenes, especially at long distances. This is not surprising given the logistical difficulties of making such measurements. To circumvent these difficulties, we collected 81 stereo images of outdoor scenes, together with precisely registered range images that provided the ground-truth distance at each pixel location. We then presented the stereo images in the correct viewing geometry and measured the ability of human subjects to discriminate the distance between locations in the scene, as a function of absolute distance (3 m to 30 m) and the angular spacing between the locations being compared (2°, 5°, and 10°). Measurements were made for binocular and monocular viewing. Thresholds for binocular viewing were quite small at all distances (Weber fractions less than 1% at 2° spacing and less than 4% at 10° spacing). Thresholds for monocular viewing were higher than those for binocular viewing out to distances of 15–20 m, beyond which they were the same. Using standard cue-combination analysis, we also estimated what the thresholds would be based on binocular-stereo cues alone. With two exceptions, we show that the entire pattern of results is consistent with what one would expect from classical studies of binocular disparity thresholds and separation/size discrimination thresholds measured with simple laboratory stimuli. The first exception is some deviation from the expected pattern at close distances (especially for monocular viewing). The second exception is that thresholds in natural scenes are lower, presumably because of the rich figural cues contained in natural images.

## Introduction

Despite nearly two centuries since the invention of the stereoscope, and more than two millennia since the formalization of perspective geometry, the precision with which we can resolve the closer and further of two points in the natural world remains largely unknown, especially at far distances. There are two main reasons for this. First, the emphasis in perception research has been on simple laboratory stimuli designed to characterize the effect of single cues on depth discrimination (e.g., see Howard & Rogers, 2012) and/or to characterize how pairs of cues combine or interact (e.g., Ernst & Banks, 2002; Knill & Saunders, 2003; for reviews, see Geisler, 2011; Kersten, Mamassian, & Yuille, 2004). This approach allows for strong experimental control and has provided a vast wealth of knowledge about the various mechanisms of depth discrimination (e.g., see Howard & Rogers, 2012) but leaves open the question of how well all the mechanisms together support depth discrimination under natural conditions and how the different kinds of mechanisms contribute to depth perception under natural conditions. Second, there are significant technical and logistical difficulties in collecting large numbers of depth judgments between pairs of locations in natural scenes, especially in outdoor scenes where distances are relatively large. More specifically, it is difficult to measure the ground-truth distances to large numbers of points, it is difficult to indicate to the subject the specific points to be compared, and it is difficult to make measurements in a large enough number of different settings. Previous efforts to characterize depth, size, or distance discrimination in environments meant as a proxy for real-

Citation: McCann, B. C., Hayhoe, M. M., & Geisler, W. S. (2018). Contributions of monocular and binocular cues to distance discrimination in natural scenes. *Journal of Vision*, 18(4):12, 1–15, <https://doi.org/10.1167/18.4.12>.



world conditions have had the subject make just a few judgments in one or two settings (Allison, Gillam, & Vecellio, 2009; Holway & Boring, 1941; McKee & Taylor, 2010; Palmisano, Gillam, Govan, Allison, & Harris, 2010). Some similar studies compare real-world performance to matched conditions in synthetic virtual-reality scenes (Creem-Regehr, Stefanucci, Thompson, Nash, & McCardell, 2015; Creem-Regehr, Willemsen, Gooch, & Thompson, 2005; Lin et al., 2011; Loomis & Knapp, 2003), which are generally different in many ways from real scenes (Mon Williams & Wann, 1998; Thompson et al., 2004; Wann, Rushton, & Mon-Williams, 1995). The logistics of stimulus presentation in real, outdoor scenes limits the possibility of using the standard two-alternative forced-choice paradigms commonly preferred for measuring thresholds in the laboratory. Instead, most outdoor studies have relied on methods such as verbal report (Knapp & Loomis, 2004; E. J. Gibson & Bergman, 1954; E. J. Gibson, Bergman, & Purdy 1955) triangulation (Fukushima, Loomis, & Da Silva, 1997), or open loop walking and pointing (Creem-Regehr et al., 2005; Knapp & Loomis, 2004; Loomis, Da Silva, Fujita, & Fukushima, 1992). Some experiments have involved physical interval comparisons, such as by fractionation (Purdy & Gibson, 1955) or comparison of frontoparallel intervals with depth intervals (Loomis et al., 1992). Still, these experiments did not focus on measuring the precision of depth discriminations but instead focused on exploring systematic biases in the accuracy of spatial-interval estimates or the self-consistency of various reporting measures and display methodologies. Consequently, relatively little is known about the precision of depth judgments made at long distances in outdoor scenes. Furthermore, although it has been shown that binocular mechanisms contribute to depth judgments even at far distances (Cormack, 1984; Palmisano et al., 2010), this was achieved by specifically removing the influence of monocular information. Consequently, we still do not know the relative sensitivity of binocular and monocular mechanisms as a function of distance in natural conditions. Measuring the precision of depth judgments in outdoor natural scenes is necessary for understanding how the vast literature of published laboratory studies relates to performance in the real world (Geisler, 2008).

Here, we measured distance discrimination in natural outdoor scenes using a calibrated stereo display system. We collected a set of high-resolution calibrated stereo images and high-precision range images of outdoor scenes. The range images provided an accurate ground-truth measure of the scene distance at each image pixel location. The stereo images were displayed at the appropriate scale and viewing distance, so that the retinal images closely matched those that would have been formed had the subject been positioned at

the location where the stereo images were collected. We were then able to measure distance discrimination performance for large numbers of points in natural scenes, as a function of the absolute distance of the points and as a function of the angular spacing between the points. To evaluate the relative contribution of binocular and monocular cues as a function of distance, we compared binocular and monocular performance on the same stimuli.

Figure 1 shows the range of absolute distance bins (3–30 m; i.e., 10–100 ft.) that were tested. We chose this range because it is the range of distances for which the range scanner is highly accurate, and it represents a range over which we would expect binocular and monocular cues to tradeoff in their usefulness. We also chose this range so that we could place the display screen at a distance of 3 m. In this configuration, the display was literally equivalent to viewing the outdoor scene through a large window at a distance of 3 m from the window. Another advantage of this configuration is that there were essentially no defocus cues in the images, in agreement with what an observer would experience when fixating objects that are at a distance 3 m or further—fixating at the distance of the display screen would produce a blur of only 1/3 diopter for an object at infinity (threshold for detecting defocus blur in natural image patches is in the range of 0.1 to 0.4 diopters; e.g., Sebastian, Burge, & Geisler, 2015).

The red and blue curves in Figure 1 give some sense of how binocular-stereo and monocular-perspective cues might contribute to depth discrimination as a function of absolute distance. If binocular-stereo thresholds are on the order of 16 arcsec (Blakemore, 1970; Ogle, 1956), then the red curve shows the expected Weber fraction (in percentage) for discriminating distance as a function of absolute distance. (In computing these Weber fractions, we assumed crossed disparities; see Figure 1A and Appendix.) The Weber fraction is low at small distances and rises rapidly as a function of distance.

The blue curve in Figure 1B shows the expected Weber fraction for monocular-perspective cues under the assumption that the primary basis for these cues is local changes in the size or spacing between image features (e.g., the size and spacing of texture elements). Size discrimination for sine waves is in the neighborhood of 5% (Burbeck & Regan, 1983; Caelli, Brettel, Rentschler, & Hiltz, 1983; Hirsch & Hylton, 1982; Skottun, Bradley, Sclar, Ohzawa, & Freeman, 1987), and discrimination of the spacing between line segments is in the neighborhood of 3% at the optimal base separation (Westheimer & McKee, 1977). As the blue curve shows, this threshold for size discrimination should translate into a constant Weber fraction for distance discrimination as a function of absolute distance (see the Discussion section and Appendix for

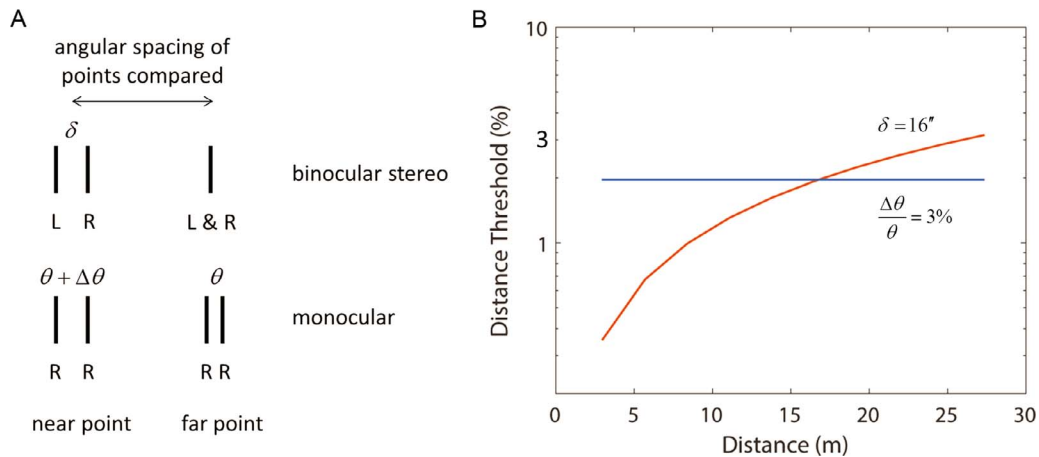


Figure 1. Expected usefulness of binocular and monocular cues for depth discrimination. (A) Geometry of stimuli in simple binocular stereo-discrimination and monocular separation-discrimination experiments. (B) Expected distance discrimination thresholds plotted as a function of absolute distance in meters, for an angular spacing between points being compared of  $1.5^\circ$ . The vertical axis plots the Weber fraction (in percentage) for distance discrimination. The red curve gives the Weber fraction expected from binocular-stereo cues alone, given a binocular disparity threshold of 16 arcsec (Blakemore, 1970; Ogle, 1956). The blue curve gives the Weber fraction that might be expected from monocular-perspective cues given a size/spacing discrimination Weber fraction of 3% (Burbeck & Regan, 1983; Caelli et al., 1983; Hirsch & Hylton, 1982; Skottun et al., 1987; Westheimer & McKee, 1977).

more details). At near distances, the monocular perspective cues should be much less useful than the binocular-disparity cues, but that relationship should reverse at large distances. One of our aims was to estimate where this crossover occurs in real-world conditions.

## Methods

### Registered camera and range images

High-resolution stereo camera and range images were obtained with a Nikon D700 digital camera mounted on a Riegl VZ-400 3D laser range scanner. To collect stereo images and to align the nodal points of the camera and laser scanner, the camera and scanner were mounted on a custom portable robotic gantry having four degrees of freedom: translation in x, y, and z and rotation about the vertical (y) axis. The specific sequence for image capture was as follows: (a) capture the first range image, (b) translate vertically (perpendicular to earth) to capture a photograph with the camera's nodal point aligned with the range image, (c) translate horizontally (parallel to earth) 6.5 cm to capture a photograph from the second vantage point, and (d) translate vertically to align the origin of the rangefinder with the second vantage point. Occasionally, either the photographs or range images would need to be recaptured due to disturbances obvious in the field (e.g., pedestrians, cars, or major changes in lighting). In these cases, capture could be repeated

without moving the robotic gantry, preserving the relative calibration.

Each digital camera image was  $4,284 \times 2,844$  pixels, with a bit depth of 14 bits per RGB color channel. Each pixel subtended about  $0.02^\circ$ . The camera spectral sensitivities were measured and are available at <http://natural-scenes.cps.utexas.edu/db.shtml>. The range scanner provides accurate depth measurements ( $\pm 5$  mm) over the range of approximately 2 m to 200 m. Custom software was used to register the images. Inspection of various test cases shows that we obtained very good registration ( $\pm 1$  pixel).

We obtained 96 high-quality registered stereo camera and range images from around the University of Texas campus. Images were captured from the typical eye height of a six-foot male, 66 inches above the ground. Gaze was approximately earth parallel. Fixation was at infinity. We then cropped the images to  $27^\circ \times 15^\circ$  of visual angle ( $1,280 \times 720$ ) to minimize the potential effects of camera lens distortions (e.g., barrel distortion). Of the 96 images, 15 were removed because they contained some artifacts due to image motion (e.g., wind). Thumbnails of the cropped regions from the remaining 81 “left eye” camera and range images are shown in Figures 2 and 3. The entire image set is available at <http://natural-scenes.cps.utexas.edu/db.shtml>. More details of the measurements and calibration are available in Burge, McCann, and Geisler (2016), which used a version of this image set cropped to a higher resolution/field of view, and in notes found at the above website.

Figure 4A shows the distribution of measured distances in the range images (note the bins have fixed





Figure 2. Thumbnails of the images used in the psychophysical experiments. These cropped images were a fraction of the size of the raw images because of the screen's limited field of view.



Figure 3. Range images corresponding to the experimental stimuli in Figure 2. Measuring ground-truth allows for objective evaluation of performance. The color scale is given in Figure 4A.



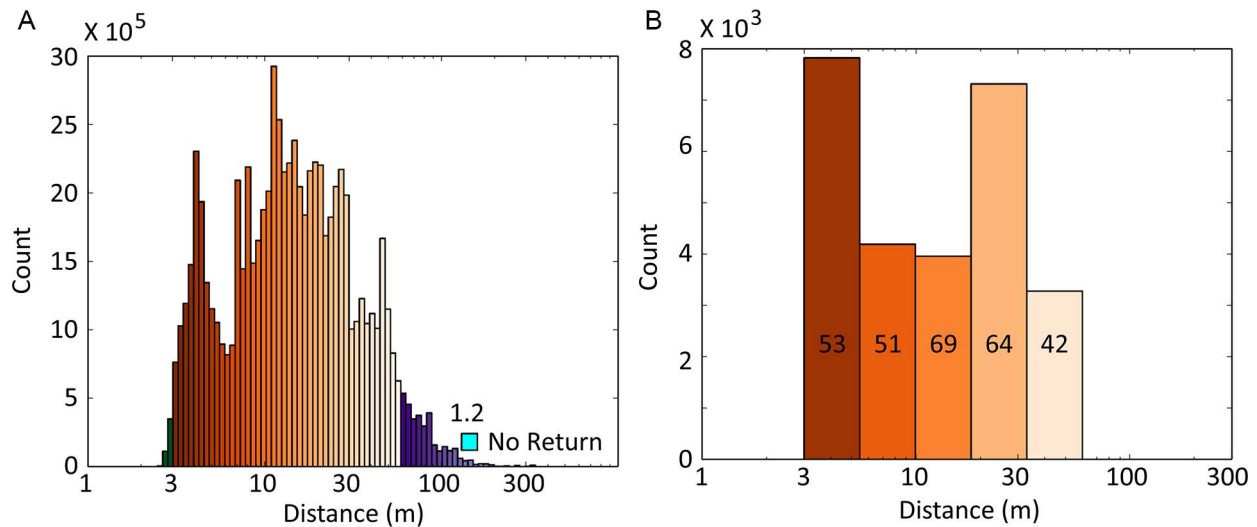


Figure 4. Distribution of distances. (A) Histogram of distances for all range images in Figure 3. (B) Histograms of distances tested in the experiment. The number on each bar indicates the number of images (out of the 81) that contributed trials to the tested distances. Note that the bin widths and x-axes in these plots are logarithmic.

log-unit widths). The distances are divided into four regions: distances closer than 3 m (shades of green), distances from 3–60 m (shades of orange), distances greater than 60 m (shades of blue), and locations where the device was unable to make a measurement (light blue). The display screen in the experiment described below was at a distance of 3 m, and thus we did not test distances closer than 3 m. The distances tested in the experiment were selected from the range of 3–60 m (orange-shaded pixels).

## Psychophysical methods

### Apparatus overview

A display system was designed to appear as a window (an aperture) through which the scenes could be viewed stereoscopically (see Figure 5). Because the screen was at the distance of the nearest possible object (3 m), all objects in the depicted scene should appear behind the window. The room was dark, and the edges of the screen were obscured with black felt obscuring the walls around the edges of the display. The depth

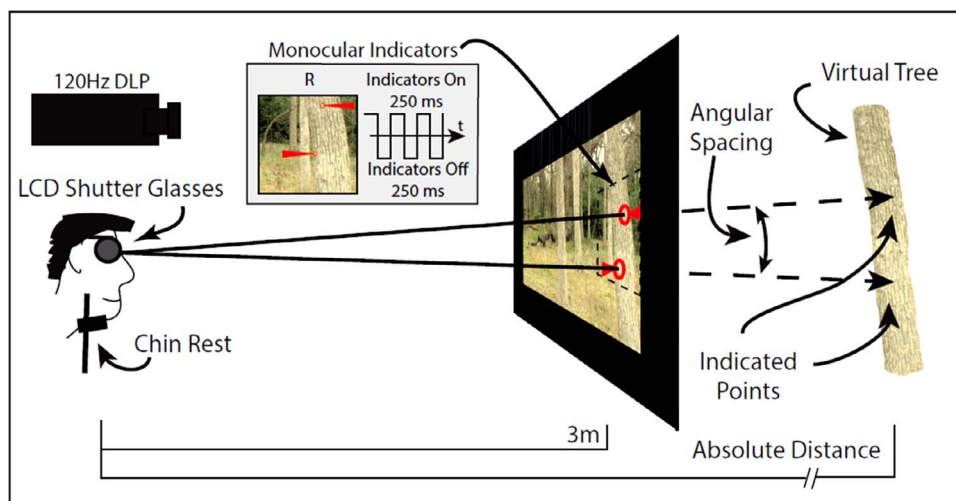


Figure 5. Schematic of apparatus and psychophysical task. The observers judged the nearer of two locations indicated in the virtual scene and were free to look back and forth between the locations. Distance discrimination threshold between the two locations was measured as a function of the absolute distance of the further of the two locations and of the angular spacing between the two locations. The locations to be compared were indicated with two monocular pointers (see inset). The subjects could toggle the pointers off and on and were instructed to make their judgments when the pointers were off. (The dashed box shows the location of the image in the inset and was not in the actual display.)

illusion was convincing, minimizing the impact of the planar display on the depth percept, consistent with many recommendations in Nagata (1991).

The projector was a 120 Hz, single-chip DLP, active-stereoscopic projector (DepthQ HDs3D-1). Combined with synchronized liquid crystal shutter glasses (nVidia 3d Vision), the system produced a 60-Hz effective stereoscopic frame rate at the native 720 p (1,280 × 720) resolution.

The screen was rectangular, 1.43 m wide and 0.8 m tall, on a tensioned frame. At the viewing distance of 3 m, this gave a visual angle of 27° horizontally and 15° vertically. The images were displayed at the native 720 p resolution of the projector. Therefore, pixels subtended about 0.02°, corresponding to a Nyquist frequency of 25 cycles/deg (roughly half human limits). Although higher resolution would be ideal, this sampling rate was matched to the camera.

The displayed images are geometrically correct from only one viewing location. Therefore, a chinrest was used to minimize head movements. Furthermore, the images are only strictly correct if the observer's interpupillary distance (IPD) is the same as the distance between the nodal points (65 mm) used in making the scene measurements. Accordingly, we selected observers with IPDs of approximately 65 mm. The result of carefully controlling the viewing geometry was a subjectively highly convincing fused stereoscopic view of the scene viewed through a window.

### **Psychophysical task**

The task was a near-far distance discrimination task. On a given trial, two scene points were indicated to the observer (see Figure 5, shown in the inset). The observer pressed the right arrow if the right scene point was judged to be nearer and the left arrow if the left scene point was judged to be nearer. Because the distances between points in the scene are known, performance could be evaluated objectively.

To indicate the scene points that the observer should compare, we used a triangle pointing toward the center of a small circle denoting a surface. The point at the very center of the circle was the point being judged. To avoid introducing any additional stereoscopic information with the indicator, it was presented in the right eye's image only. Because the indicator was an abstract shape presented monocularly, it had no real apparent depth in the scene. Trials were self-timed, and fixation was uncontrolled. The observer could view the indicators at any time by pressing the up-arrow key. If the key was held down, the indicators would flicker on and off at 2 Hz. If the key was released, the indicators remained off. Observers were instructed to make their judgments while the indicators were off.

For the monocular version of the task, the left eye was patched for all subjects. Shutter glasses were still worn in case the glasses themselves degraded the image in any way. With head rested in the chinrest, and with the left eye patched, the illusion of a window was still convincing (all the monocular cues were correct given the position of the eye).

Four male subjects (ages 22–29 years), one of whom was an author and three who were naive, were tested using an institutional review board–approved protocol from the University of Texas at Austin. All subjects had normal acuity and scored at the maximum level on the Randot stereo test.

### **Sampling procedure**

An important consideration in designing the experiment was how to sample the scene points to be compared. There are many possible stimulus dimensions that could affect depth discrimination in natural scenes, and it is impossible to control more than a few of the potentially relevant dimensions within a single study. Here we sampled points along three dimensions: (a) depth, the difference in distance between the two points, (b) the mean depth distance to the points, and (c) the visual angle separating the points. All other potential stimulus dimensions were left uncontrolled.

The first dimension is the natural measure of difficulty, the difference in distance between the two points—the smaller the difference, the more difficult the discrimination. In our initial measurements, we selected points from a range of fixed disparity bins, based on the stereo acuity literature. After inspecting results of the first two subjects, it was clear that using the dioptric difference instead of disparity provided better coverage of the psychometric function over the ranges studied. For the subsequent two subjects, new points were selected according to the new difficulty metric, and additional easy trials were added near saturation of performance. Importantly, these measures are essentially equivalent—both can be converted into the difference in distance between the two points. It should also be noted that the nearer point was always considered the correct selection, even in instances when the disparity was zero or had an incongruous sign resulting from the horizontal positioning of the targets and the eyes. These conditions were relatively rare but more common near threshold.

The second dimension was the mean absolute distance to the points being compared. The absolute distance to the scene points is expected to affect performance because of both perspective projection and the degradation of stereoscopic cues to depth. Absolute distances were sampled from the range of 3 to 60 m (see Figure 4). In computing thresholds, the

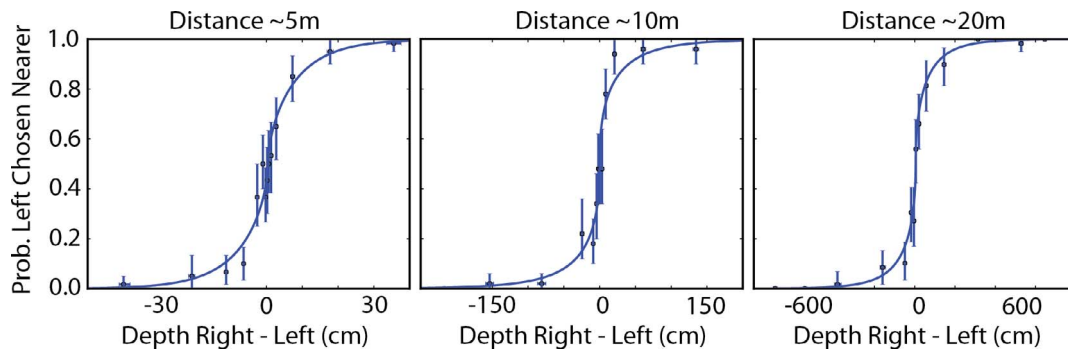


Figure 6. Example psychometric functions from one subject. All of these measurements were for a visual angle spacing of about  $5^\circ$ . Curves represent maximum likelihood fits to the data. Note the scale change on the x-axis with increasing absolute distance; thus, the distance threshold (in cm) increases with absolute distance.

absolute distances were binned with bin width that was approximately proportional to absolute distance.

The third dimension was the visual angle separating the two scene points. Measurements were made for bins centered on  $2^\circ$  ( $1.5^\circ$ – $2.5^\circ$ ),  $5^\circ$  ( $4^\circ$ – $6^\circ$ ), and  $10^\circ$  ( $8^\circ$ – $12^\circ$ ) of visual angle. Substantially larger spacing was not feasible because of sampling constraints produced by the size of the images ( $27^\circ \times 15^\circ$ ). An effort was made to spread trials uniformly across all 81 images, although the prevalence of content at a given distance inevitably biased the sampling toward a subset of scenes in certain conditions. Nonetheless, many scenes contributed to the stimuli for any given experimental condition (see Figure 4B). Trials for the different images and conditions were randomly interleaved.

To obtain a given sample, a random point in a random scene was chosen based on its absolute distance. This first point was taken as the more distant of the two points, because crossed disparities are much easier to find than uncrossed disparities at large distances (Liu, Bovik, & Cormack, 2008). The second point was chosen based on the desired value along the other two dimensions. First, an annulus of pixels corresponding to the appropriate angular-spacing bin was selected. Each of these pixels was checked for delta-distances appropriate for the needed difficulty level (commensurate with a given disparity or dioptric difference). Next, we checked whether the region surrounding the first point and each potential second point was approximately planar. To do this, we analyzed the cloud of  $x,y,z$  points within a  $0.25^\circ$  diameter (five-pixel radius) of the point being considered. The mean distance of the points in the cloud was subtracted from each point (centering the cloud at a mean distance of zero), and then each point was divided by the absolute distance of the point being considered. We then performed singular-value decomposition to extract the three principle components. If a surface is perfectly planar, then the third principle component will be zero. The point being considered was rejected if the amplitude ratio of the third to

second component exceeded 0.75. Points were also excluded if surface slant was too steep (which may be a depth edge); a point was excluded if the amplitude ratio of the second to first component was less than 0.05. Finally, we checked that neither point (in the randomly selected pair) was partially occluded (i.e., visible to only one eye). All pairs of points that passed these tests were considered potential pairs. Once a pair of target points was selected, the neighboring image region was ruled out, and the process began again until all needed trial bins were filled. The end result was that the number of trials at each spatial separation was approximately 2,200 per subject: 1,100 for the binocular conditions and 1,100 for the monocular conditions.

## Results

### Psychometric functions

To estimate psychometric functions and thresholds, the trials for each subject, at a given spatial separation and mode of viewing (monocular or binocular), were rank ordered by absolute distance and then binned using a sliding window that was approximately  $1/3$  log unit wide (geometric binning with a smaller bin width yielded similar but noisier functions). Maximum likelihood methods were used to fit the psychometric functions using all of the trials in a bin, respecting their continuous location on the delta-distance axis. Example psychometric functions for one subject in three distance bins, for a spacing of approximately  $5^\circ$ , are shown in Figure 6. However, in these plots, the data points show binned data with horizontal and vertical error bars. (The horizontal error bars are mostly smaller than data points.) They are helpful for assessing the quality of the fit, but they were not used in obtaining the maximum-likelihood fit. Psychometric functions were measured for each observer in each

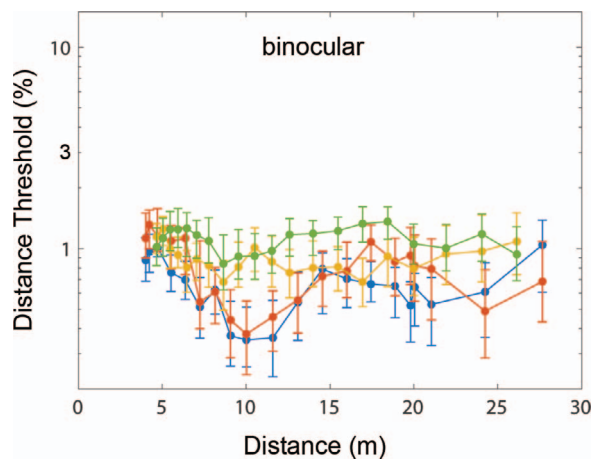


Figure 7. Distance discrimination thresholds (expressed as Weber fractions in percentage) of individual subjects as a function of absolute distance, in the binocular viewing condition, when the angular spacing between points being compared was in the range of  $1.5^{\circ}$ – $2.5^{\circ}$  (average spacing of  $2^{\circ}$ ). Error bars are bootstrapped 95% confidence intervals on the thresholds. Data points in this figure, and the subsequent figures, represent a running average with a bin width of approximately  $1/3$  log unit; thus, the scale appears somewhat different from Figure 4B.

condition, and from these functions, thresholds were computed.

### Binocular and monocular thresholds

Figure 7 plots distance-discrimination threshold (as a Weber fraction in percent) for all four observers, as a function of absolute distance, when the angular spacing between the scene points was about  $2^{\circ}$ . For all four observers, the thresholds, expressed as Weber fractions,

are relatively constant over the full range of absolute distances. In other words, distance discrimination is approximately proportional to absolute distance (Weber's law). Note that discrimination performance is very good, with Weber fractions as low as 0.3% for some subjects, and substantially better than that predicted from the estimates used in Figure 1.

Overall, we found that performance is consistent across the four observers. There looks to be some individual differences (e.g., at about 10 m); however, the two pairs of subjects saw different stimuli. The subjects who saw one set of stimuli (green and orange) were fairly consistent with each other, as were the subjects who saw the other set of stimuli (blue and red). The agreement between subjects within each pair, and the larger difference between the two sets, is consistent with stimulus differences as the cause. Although it is difficult to be confident of this with a small number of subjects, we chose to aggregate over subjects in the subsequent plots.

Figure 8A plots the distance-discrimination thresholds (aggregated across subjects) for the three values of angular spacing between the scene points. Threshold increases substantially (more than a factor of 3) with angular spacing. For the smallest spacing (blue data points), the threshold Weber fraction is fairly constant with absolute distance (Weber's law), but for the large spacing (green and orange data points), the threshold Weber fraction rises up to a distance of about 12–15 m and then levels out.

Figure 8B plots the distance discrimination thresholds measured under monocular conditions. Again, thresholds increase substantially with angular spacing. However, the thresholds now decrease (rather than increase) with absolute distance out to 10–15 m, before they level out. Interestingly, the initial drop in threshold with absolute distance is not what one would

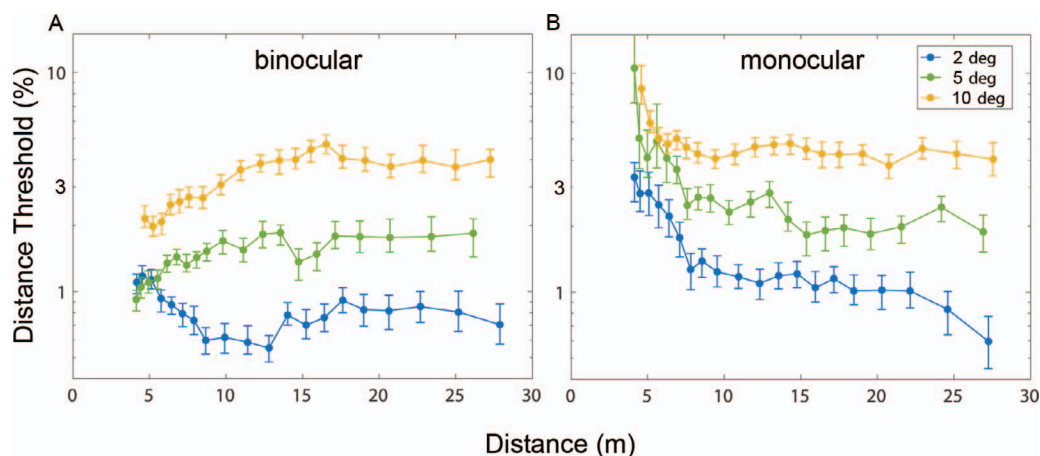


Figure 8. Distance discrimination thresholds (expressed as Weber fractions in percentage) as a function of absolute distance and angular spacing of the points being compared. Data points are the average across the four subjects. (A) Binocular viewing. (B) Monocular viewing. Error bars are bootstrapped 95% confidence intervals on the thresholds.



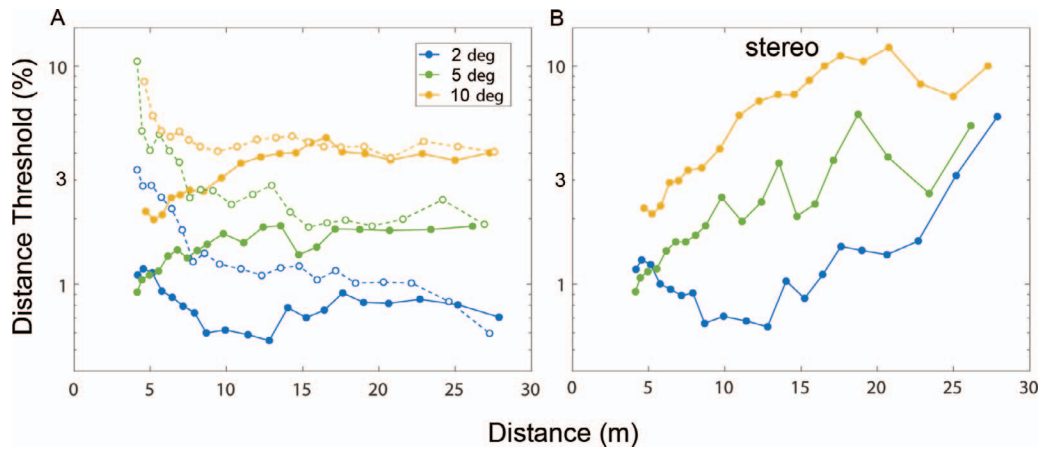


Figure 9. Comparison of monocular and binocular thresholds (expressed as Weber fractions in percentage). (A) Replot of the binocular (solid circles) and monocular thresholds (open circles) in Figure 8. The thresholds converge beyond a distance of 15 to 20 m, suggesting that binocular stereo cues contribute little beyond that distance. (B) Estimated thresholds for stereo cues alone (see text for details).

expect from the simple analysis in Figure 1, and thresholds are lower than expected from perspective cues alone. We will return to this result in the Discussion section.

The difference between the binocular and monocular thresholds is more clearly illustrated in Figure 9A, which plots the binocular thresholds from Figure 8A as solid symbols and the monocular thresholds from Figure 8B as open symbols. At near distances, the binocular thresholds are much smaller than the monocular thresholds. The curves converge at a distance of 15–20 m, suggesting that this is the range over which binocular stereo cues contribute substantially to distance discrimination.

### Estimation of thresholds for binocular stereo cues alone

The binocular distance-discrimination thresholds are based on both monocular cues and binocular-stereo cues. By analyzing the difference between the binocular and monocular distance-discrimination thresholds, it is possible to obtain an estimate of what the thresholds would be with binocular stereo cues alone. In the signal-detection-theory framework, the distance discriminability  $d'$  can be described as the signal (the change in distance)  $\Delta z$ , divided by the effective noise  $\sigma(z, \phi)$ , which depends on the absolute distance ( $z$ ) and angular spacing ( $\phi$ ) between the two scene points being compared. Thus, for the binocular case  $d'_B = \Delta z / \sigma_B(z, \phi)$ , for the monocular case  $d'_M = \Delta z / \sigma_M(z, \phi)$ , and for the stereo-alone case  $d'_\delta = \Delta z / \sigma_\delta(z, \phi)$ . Under the assumption that the effective noise for monocular and binocular-stereo cues is Gaussian and statistically independent, it can be

shown (see Appendix) that the thresholds for binocular stereo cues alone are given by

$$\Delta z_{T,\delta}(z, \phi) = \frac{\Delta z_{T,B}(z, \phi) \Delta z_{T,M}(z, \phi)}{\sqrt{\Delta z_{T,M}^2(z, \phi) - \Delta z_{T,B}^2(z, \phi)}} \quad (1)$$

where  $\Delta z_{T,B}(z, \phi)$  is the measured threshold under binocular viewing and  $\Delta z_{T,M}(z, \phi)$  is the measured threshold under monocular viewing.

Figure 9B shows the estimated distance-discrimination thresholds in natural scenes for binocular stereo cues alone, based on Equation 1. (Following Oruc, Maloney, & Landy, 2003, we also carried out this analysis under the assumption that the monocular and binocular noise was correlated by as much as 0.3 and found that it had little effect.) For the larger angular spacing between the points being compared (orange and green points), the estimated thresholds show the general pattern expected from the simple analysis in Figure 1. However, for smallest angular spacing (blue points), the estimated thresholds are relatively constant rather than rising over the first 15 m. Again, this is a result we will return to in the Discussion section.

## Discussion

There is little existing data on how well humans are able to discriminate distances under natural conditions. The primary aim of this study was to measure human ability to discriminate the distance between locations in natural scenes as a function of the distance of the further of the two locations and the angular spacing between the two locations. To do this, we obtained 81 calibrated stereo images together with co-registered range images

that provided the ground-truth distance at each pixel location, for both images in each stereo pair. The stereo image pairs were presented with the correct viewing geometry so that the stimuli were equivalent to viewing the scene through a  $27^\circ \times 15^\circ$  window with all scene points located behind the window. Measurements were made under both binocular and monocular viewing conditions. We found that binocular distance-discrimination thresholds, expressed as a Weber fraction, increase monotonically with distance up to a plateau and then remain relatively constant. On the other hand, the monocular thresholds decrease with distance down to a plateau that is aligned with the binocular plateau. In both cases, thresholds increase with the angular spacing between the locations being compared. From the binocular and monocular thresholds, we estimated the thresholds for the binocular stereo cues alone and found those to increase with distance and with angular spacing. For our natural scenes and viewing conditions, the results suggest that binocular stereo cues play a significant role in improving distance discrimination out to a distance of 15 to 20 m (on average).

It is important to recognize that there are limitations to the conclusions that can be drawn from the measurements reported here. First, the 81 scenes tested were captured around the campus at the University of Texas. Although they do contain a mixture of natural and human-made objects and were collected over a period of months, at different times of the day, it is possible that different results would be obtained in other kinds of environments. Second, the measurements were made with a field of view restricted to  $27^\circ \times 15^\circ$ . Although there are many natural situations where field of view is restricted, it is possible that different results would be obtained with a larger (or smaller) field of view. For example, Wu, Ooi, and He (2004) measured human ability to judge absolute distance for distances from 3 to 7 m and found that restricting the vertical field of view to less than  $30^\circ$  causes a decrement in the accuracy of blind walking to targets (see also Creem-Regehr et al., 2005). Similarly, Sinai, Ooi, and He (1998) and Wu, He, and Ooi (2007) found that the accuracy of absolute distance judgments is reduced if there is a gap or disruption of visibility of the ground plane, which occurred here because of the restricted field of view. Although these studies concerned absolute distance judgment, rather than distance discrimination, it is possible that similar effects would hold here.

An important question is how our measurements of human performance under natural conditions compare with measurements made with simple stimuli under controlled laboratory conditions. Consider first the distance thresholds estimated for binocular stereo cues alone (Figure 9B). Ogle (1956) measured stereo thresholds for vertical line targets as a function of the

angular spacing between the targets. He included conditions (like those in the present experiment) where the observer was allowed to look back and forth between to the locations, as well as conditions where fixation was held constant. In both cases, he found that thresholds approximately doubled from a spacing of  $2^\circ$  to  $5^\circ$  and approximately doubled again from a spacing of  $5^\circ$  to  $10^\circ$ . This rule held for all three observers, even though their absolute sensitivities varied by about 50%. In a related (but not as directly comparable) study, Blakemore (1970) measured stereo thresholds for pairs of line targets (at fixed spacing) as a function of retinal eccentricity. Similarly, he found that thresholds approximated doubled from  $0^\circ$  to  $5^\circ$  eccentricity and approximately doubled again from  $5^\circ$  to  $11^\circ$  eccentricity. Again, this rule held across both subjects, even though their sensitivities differed by a factor of 2. The solid curves in Figure 10A show the predictions of this rule under the assumption that stereo disparity thresholds are independent of absolute distance. (Note these curves are like the red curve in Figure 1, but with different values of the threshold disparity  $\delta$ .) Although there are some systematic deviations from the predictions, the overall trends are consistent with the classic psychophysical findings although more sensitive than the simple predictions in Figure 1. Interestingly, the overall thresholds of our subjects under natural viewing conditions are lower than those in the Ogle and Blakemore studies; the average threshold of Ogle's subjects at the smallest spacing is 14 arcsec and of Blakemore's subjects was 16 arcsec—about three to four times the average threshold of our subjects. Thus, our results suggest that human stereo thresholds are quite good under natural viewing conditions.

Next, consider the monocular thresholds in Figure 8B, which are replotted as points in Figure 10B. As noted in the Introduction, perspective geometry implies that (all other things equal) the angular separation between image features (or angular size of features) decreases linearly with distance. Thus, distance discrimination from perspective cues is likely to involve detecting changes in the angular separation between (or size of) image features. Given the approximate scale invariance of natural images (Burton & Moorhead, 1987; Field, 1987; Ruderman, 1997; Ruderman & Bialek, 1994), there should be features with a comparable range of angular separations at each distance and hence the angular discrimination information due to perspective should be similar at different absolute distances. If so, then one might expect a fixed angular separation/size threshold at all distances. This in turn implies a fixed Weber fraction for distance discrimination as function of distance (the horizontal lines in Figure 1 and Figure 10B). Assuming that separation/size discrimination threshold increases with spacing in a fashion similar to the stereo thresholds, we would



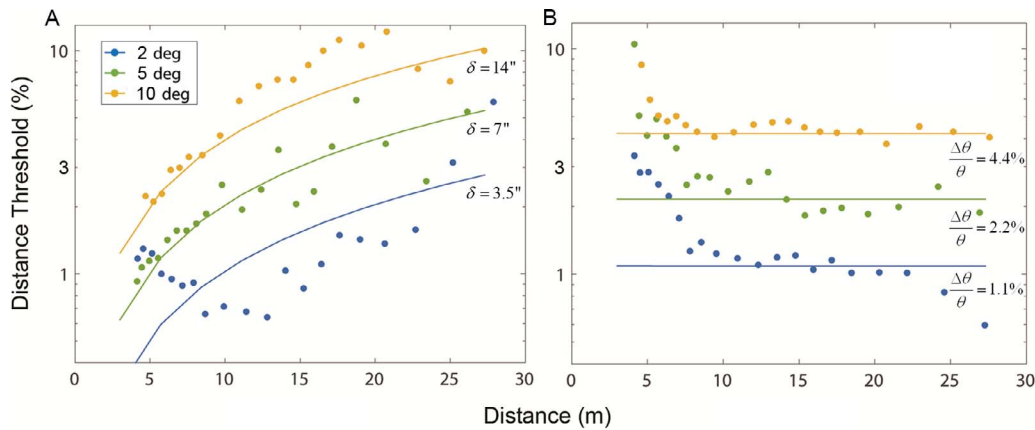


Figure 10. Comparison of thresholds in natural scenes with previous measurements using simple stimuli. (A) Comparison of the estimated distance thresholds from stereo cues in natural scenes (points from Figure 8B) with predictions from previous measurements (solid curves) for line targets (Blakemore, 1970; Ogle, 1956). The threshold disparities ( $\delta$ ) associated with the predictions for each angular spacing are in the ratios reported in the classic studies, but the absolute values of the threshold disparities are substantially lower, suggesting that human stereo thresholds in natural scenes are quite good. (B) Comparison of monocular distance discrimination thresholds (points from Figure 7B) with predictions from classic measurements of size and separation discrimination. The horizontal lines are the distance discrimination thresholds expected from monocular-perspective cues, given the size/separation Weber fractions shown. These angular discrimination thresholds (especially the ones for the smaller spacing between the points being compared) are smaller than those reported for simple stimuli, suggesting that the many perspective cues combine to give very good performance in natural scenes. It is possible that the rise in thresholds at near distances may be due in part to the somewhat restricted field of view.

expect threshold to approximately double when the spacing between points increases from  $2^\circ$  to  $5^\circ$  and again when the spacing increases from  $5^\circ$  to  $10^\circ$ . The horizontal lines in Figure 10B reflect this prediction. Again, however, there are some systematic deviations.

Also, the overall sensitivity appears to be greater than expected based on separations/size discrimination thresholds with simple stimuli. The Weber fraction for spatial-frequency discrimination (a form of size/separation discrimination) averages about 5% across a range of spatial frequencies (Burbeck & Regan, 1983; Caelli et al., 1983; Hirsch & Hylton, 1982; Skottun et al., 1987). Similarly, the smallest Weber fraction ( $\sim 3\%$ ) for separation discrimination of line segments (Figure 1A) seems to be obtained in hyperacuity tasks when the base separation is around 4 arcmin (Westheimer & McKee, 1977). The distance discrimination thresholds in Figure 10B (especially for the smaller spacing between the locations being compared) are consistent with considerably smaller Weber fractions, suggesting that monocular distance discrimination is quite good under natural conditions. This might be expected from the rich figural cues contained in natural images.

The most obvious deviation from the predictions from simple laboratory stimuli is the rise in threshold for monocular viewing at absolute distances less than 10 m. There are several possibilities. One possibility is that the density of pixel locations at the nearer distances is lower. A lower density may provide less useful contextual information for estimating the

relative distance at the two locations. A second possibility is that at nearer distances, the angular distance above the ground plane of the locations being compared is larger. A third possibility is that the window restricted the useful contextual information from the ground plane. In our display, the horizontal ground plane would be occluded up to a distance of 15 m. This is perhaps the most likely explanation. J. J. Gibson (1950), and others (Gillam, 1995; Meng & Sedgwick, 2001; Ooi et al., 2015; Wu, Zhou, Shi, He, & Ooi, 2015), have suggested (or shown) that the ground plane and surface contact relations are important for judging absolute distance and relative distance intervals. This may well underlie the elevation of monocular thresholds at near distances, although it is still uncertain how findings about absolute distance translate to distance discrimination thresholds.

Also, we note that the nonmonotonicity (elevation) of thresholds for the binocular data at the smallest separation is curious and may reflect some peculiarity of the viewing conditions similar to those that affect the monocular data. Nonetheless, if this nonmonotonicity is due to the window, then it seems to be less severe than for monocular viewing (compare Figure 10A and 10B). This might be partially explained by the results of Wu et al. (2015) and Ooi et al. (2015), who show that under binocular viewing (but not monocular viewing), subjects can use the visible parts of the ground plane to improve their absolute distance judgments of suspended objects (objects with no visible ground contact

point). Specifically, note that in our study, the ground plane was often visible within the window (see Figure 2). However, given the complexity of real-world scene structure, this is likely to be only one of multiple factors.

In this study, we exploited a database of registered range and stereo images to measure human ability to discriminate distance in natural scenes. This is just one of many possible measurements that might be made using this data set. An obvious extension of the current study would be to measure depth-interval discrimination in natural scenes. Another important topic in natural vision is how well humans are able to estimate (as opposed to discriminate) spatial intervals at large distances in natural scenes. Some studies asked subjects to judge relative depth intervals (Cormack, 1984; Lappin, Shelton, & Rieser, 2006) and others absolute depth intervals in physical units (Allison et al., 2009; E. J. Gibson & Bergman, 1954; He, Wu, Ooi, Yarbrough, & Wu, 2004; Loomis et al., 1992; Palmisano et al., 2010; Philbeck & Loomis, 1997). Many of these studies found strong biases that depend on the geometry and surface properties (e.g., texture) of the context in which the depth interval estimates are being made. However, the stimulus conditions in these studies were necessarily restricted, because the studies were conducted in real environments. Our database of registered range and stereo images (when displayed with the correct viewing geometry) is sufficiently accurate to substitute for directly viewing the real environment through a window. Thus, the database (available at <http://natural-scenes.cps.utexas.edu/db.shtml>) makes possible a more general and detailed parametric investigation of distance perception in natural scenes.

*Keywords: natural scenes, distance, threshold*

## Acknowledgments

Supported by National Institutes of Health grants EY 11747 (WSG) and EY 05729 (MMH).

Commercial relationships: none.

Corresponding author: Brian C. McCann.

Email: [brian.c.mccann@utexas.edu](mailto:brian.c.mccann@utexas.edu).

Address: Texas Advanced Computing Center, Center for Perceptual Systems and Department of Psychology, University of Texas at Austin, Austin, TX, USA.

## References

- Allison, R. S., Gillam, B. J., & Vecellio, E. (2009). Binocular depth discrimination and estimation beyond interaction space. *Journal of Vision*, 9(1):10, 1–14, <https://doi.org/10.1167/9.1.10>. [PubMed] [Article]
- Blakemore, C., (1970). The range and scope of binocular depth discrimination in man. *The Journal of Physiology*, 211(3), 599–622.
- Burbeck, C. A., & Regan, D. (1983). Independence of orientation and size in spatial discriminations. *Journal of the Optical Society of America*, 73, 1691–1694.
- Burge, J., McCann, B. C., & Geisler, W. S. (2016). Estimating 3D tilt from local image cues in natural scenes. *Journal of Vision*, 16(13):2, 1–25, <https://doi.org/10.1167/16.3.2>. [PubMed] [Article]
- Burton, G. J., & Moorhead, I. R. (1987). Color and spatial structure in natural scenes. *Applied Optics*, 26, 157–170.
- Caelli, T., Brettel, H., Rentschler, I., & Hilz, R. (1983). Discrimination thresholds in the two-dimensional spatial frequency domain. *Vision Research*, 23, 129–133.
- Creem-Regehr, S. H., Stefanucci, J. K., Thompson, W. B., Nash, N., & McCardell, M. (2015). Egocentric distance perception in the Oculus Rift (DK2). *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception–SAP '15*, 47–50, <https://doi.org/10.1145/2804408.2804422>.
- Creem-Regehr, S. H., Willemsen, P., Gooch, A. A., & Thompson, W. B. (2005). The influence of restricted viewing conditions on egocentric distance perception: Implications for real and virtual indoor environments. *Perception*, 34, 191–204, <https://doi.org/10.1068/p5144>.
- Cormack, R. H. (1984). Stereoscopic depth perception at far observation distances. *Perception & Psychophysics*, 35, 423–428.
- Ernst, M. O., & Banks, M. S. (2002, January 24). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429–433.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4, 2379.
- Fukushima, S. S., Loomis, J. M., & Da Silva, J. A. (1997). Visual perception of egocentric distance as assessed by triangulation. *Journal of Experimental Psychology. Human Perception and Performance*, 23, 86–100.
- Geisler, W. S. (2008) Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, 59, 167–192.
- Geisler, W. S. (2011). Contributions of ideal observer



- theory to vision research. *Vision Research*, *51*, 771–781.
- Gibson, E. J., & Bergman, R. (1954). The effect of training on absolute estimation of distance over the ground. *Journal of Experimental Psychology*, *48*, 473–482.
- Gibson, E. J., Bergman, R., & Purdy, J. (1955) The effect of prior training with a scale of distance on absolute and relative judgments of distance over ground. *Journal of Experimental Psychology*, *50*(2), 97–105.
- Gibson, J. J. (1950). *The perception of the visual world*. Boston, MA: Houghton-Mifflin.
- Gillam, B. (1995). The perception of spatial layout from static optical information. In W. Epstein & S. Rogers (Eds.), *Perception of Space and Motion*, (pp. 23–67). San Diego, CA: Academic Press.
- He, Z. J., Wu, B., Ooi, T. L., Yarbrough, G., & Wu, J. (2004). Judging egocentric distance on the ground: Occlusion and surface integration, *Perception*, *33*, 789–806.
- Hirsch, J., & Hylton, R. (1982). Limits of spatial-frequency discrimination as evidence of neural interpolation. *Journal of the Optical Society of America*, *72*, 1367–1374.
- Holway, A. H., & Boring, E. G. (1941). Determinants of apparent visual size with distance variant. *American Journal of Psychology*, *54*, 21, <https://doi.org/10.2307/1417790>.
- Howard, I. P., & Rogers, B. J. (2012). *Perceiving in depth* (Vol. 2–3). Oxford, UK: Oxford University Press.
- Kersten, D., Mamassian, P., & Yuille, A. L. (2004) Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271–304.
- Knapp, J. M., & Loomis, J. M. (2004). Limited field of view of head-mounted displays is not the cause of distance underestimation in virtual environments. *Presence: Teleoperators and Virtual Environments*, *13*, 572–577, <https://doi.org/10.1162/1054746042545238>.
- Knill, D. C., & Saunders, J. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, *43*, 2539–2558.
- Lappin, J. S., Shelton, A. L., & Rieser, R. J. (2006). Environmental context influences visually perceived distance. *Perception & Psychophysics*, *68*, 571–581.
- Lin, Q., Xie, X., Erdemir, A., Narasimham, G., McNamara, T. P., Rieser, J., & Bodenheimer, B. (2011). Egocentric distance perception in real and HMD-based virtual environments. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization–APGV '11* (Vol. 1, p. 75). New York, NY: ACM Press, <https://doi.org/10.1145/2077451.2077465>.
- Liu, Y., Bovik, A. C., & Cormack, L. K. (2008). Disparity statistics in natural scenes. *Journal of Vision*, *8*(11):19, 1–14, <https://doi.org/10.1167/8.11.19>. [PubMed] [Article]
- Loomis, J., & Knapp, J. (2003). *Virtual and adaptive environments* (L. Hettlinger & M. Haas, Eds.). New Jersey: CRC Press, <https://doi.org/10.1201/9781410608888>.
- Loomis, J. M., Da Silva, J. A., Fujita, N., & Fukusima, S. S. (1992). Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 906–921.
- McKee, S. P., & Taylor, D. G. (2010). The precision of binocular and monocular depth judgments in natural settings. *Journal of Vision*, *10*(10):5, 1–13, <https://doi.org/10.1167/10.10.5>. [PubMed] [Article]
- Meng, J. C., & Sedgwick, H. A. (2001). Distance perception mediated through nested contact relations among surfaces. *Perception & Psychophysics*, *63*, 1–15.
- Mon-Williams, M., & Wann, J. P. (1998). Binocular virtual reality displays: When problems do and don't occur. *Human Factors*, *40*, 42–49, <https://doi.org/10.1518/001872098779480622>.
- Nagata, S. (1991). How to reinforce perception of depth in single two-dimensional pictures. In S. R. Ellis, M. K. Kaiser, & A. C. Grunwald (Eds.), *Pictorial communication in virtual and real environments*, (Vol. 25, pp. 527–545). Bristol, PA: Taylor and Francis Inc.
- Ogle, K. (1956). Stereoscopic acuity and the role of convergence. *Journal of the Optical Society of America*, *3*, 269–273, <http://doi.org/10.1364/JOSA.46.000269>.
- Ooi, T. L., & He, Z. J. (2015). Space perception of strabismic observers in the real world environment. *Investigative Ophthalmology and Visual Science*, *56*(3), 1761–1768.
- Oruc, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, *43*, 2451–2468.
- Palmisano, S., Gillam, B., Govan, D. G., Allison, R. S., & Harris, J. M. (2010). Stereoscopic perception of real depths at large distances. *Journal of Vision*, *10*(6):19, 1–16, <https://doi.org/10.1167/10.6.19>. [PubMed] [Article]
- Philbeck, J. W., & Loomis, J. M. (1997). Comparison of two indicators of perceived egocentric distance

under full-cue and reduced-cue conditions. *Journal of Experimental Psychology: Human Perception and Performance*, 23(1), 72–85.

Purdy, J. G., & Gibson, E. J. (1955). Distance judgment by method of fractionation. *Journal of Experimental Psychology*, 50, 374–380.

Ruderman, D. L. (1997). Origins of scaling in natural images. *Vision Research*, 37, 3385–3398.

Ruderman, D. L., & Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Physical Review Letters*, 73, 814–817.

Sebastian, S., Burge, J., & Geisler, W. S. (2015). Defocus blur discrimination in natural images with natural optics. *Journal of Vision*, 15(5):16, 1–17, <https://doi.org/10.1167/15.5.16>. [PubMed] [Article]

Sinai, M. J., Ooi, T. L., & He, Z. J. (1998, October 1). Terrain influences the accurate judgement of distance. *Nature*, 395, 497–500.

Skottun, B. C., Bradley, A., Sclar, G., Ohzawa, I., & Freeman, R. D. (1987). The effects of contrast on visual orientation and spatial frequency discrimination: A comparison of single cells and behavior. *Journal of Neurophysiology*, 57, 733–786.

Thompson, W. B., Willemsen, P., Gooch, A. A., Creem-Regehr, S. H., Loomis, J. M., & Beall, A. C. (2004). Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence: Teleoperators and Virtual Environments*, 13, 560–571, <https://doi.org/10.1162/1054746042545292>.

Wann, J. P., Rushton, S., & Mon-Williams, M. (1995). Natural problems for stereoscopic depth perception in virtual environments. *Vision Research*, 35, 2731–2736, [https://doi.org/10.1016/0042-6989\(95\)00018-U](https://doi.org/10.1016/0042-6989(95)00018-U).

Westheimer, G., & McKee, S. P. (1977). Spatial configurations for visual hyperacuity. *Vision Research*, 17, 941–947.

Wu, B., He, Z. J., & Ooi, T. L. (2007). Inaccurate representation of the ground surface beyond a texture boundary. *Perception*, 36, 703–721.

Wu, B., Ooi, T. L., & He, Z. J. (2004, March 4). Perceiving distance accurately by a directional process of integrating ground information. *Nature*, 428, 73–77.

Wu, J., Zhou, L., Shi, P., He, Z. J., & Ooi, T. L. (2015). The visible ground surface as a reference frame for scaling binocular depth of a target in midair. *Journal of Experimental Psychology: Human Perception and Performance*, 41, 111–126, <https://doi.org/10.1037/a0038287>.

## Appendix

Here we describe (a) the expected relationship between simple size/spacing cues and distance discrimination thresholds, (b) the expected relationship between binocular stereo cues and distance-discrimination thresholds, and (c) the derivation of formula (Equation 1) for estimating distance-discrimination threshold based on binocular stereo cues alone, from the distance-discrimination thresholds measured under binocular and monocular viewing conditions. These are relatively standard formulas but are included here for completeness.

### Size/spacing cues for distance discrimination

The relationship between angular size/spacing and distance is given by the following two well-known equations:

$$2 \tan\left(\frac{\theta}{2}\right) = \frac{x}{z} \quad (\text{A1})$$

$$2 \tan\left(\frac{\theta + \Delta\theta}{2}\right) = \frac{x}{z - \Delta z} \quad (\text{A2})$$

where  $x$  is the physical separation between the pair of features,  $\theta$  is the angular separation of the further pair of features,  $\theta + \Delta\theta$  is the angular separation of the nearer pair of features,  $z$  is the distance of the further pair of features, and  $z - \Delta z$  the distance of the nearer pair of features. Combining these two equations shows that

$$\frac{\Delta z}{z} = 1 - \frac{\tan\left(\frac{\theta}{2}\right)}{\tan\left(\frac{\theta + \Delta\theta}{2}\right)} \quad (\text{A3})$$

This formula was used for the predictions in Figures 1 and 10B.

### Binocular stereo cues for distance discrimination

The relationship between disparity and distance is given by the following well-known pair of equations:

$$\theta = 2 \operatorname{atan}\left(\frac{a}{2z}\right) \quad (\text{A4})$$

$$\theta + \delta = 2 \operatorname{atan}\left(\frac{a}{2(z - \Delta z)}\right) \quad (\text{A5})$$

where  $\theta$  is the vergence angle,  $\delta$  is the (crossed)



disparity,  $a$  is the separation between the nodal points of the two eyes,  $z$  is the distance of the further feature, and  $z - \Delta z$  the distance of the nearer feature. Combining these two equations shows that

$$\frac{\Delta z}{z} = 1 - \frac{a}{2z \tan\left[\frac{\delta}{2} + \text{atan}\left(\frac{a}{2z}\right)\right]} \quad (\text{A6})$$

This formula was used for the predictions in Figures 1 and 10A.

### Binocular and monocular cue combination formula

Assuming the standard signal detection theory framework, the detectabilities (i.e., signal-to-noise ratio  $d'$ ) for the binocular viewing, monocular viewing, and binocular stereo cues alone are given, respectively, by the following three equations:

$$d_B^2 = \frac{\Delta z^2}{\sigma_B^2(z, \phi)} \quad (\text{A7})$$

$$d_M^2 = \frac{\Delta z^2}{\sigma_M^2(z, \phi)} \quad (\text{A8})$$

$$d_\delta^2 = \frac{\Delta z^2}{\sigma_\delta^2(z, \phi)} \quad (\text{A9})$$

where  $z$  is the distance to the further feature (or pair of features) and  $\phi$  is the angular spacing between the locations being compared.

Threshold is defined to be some criterion level of detectability  $d_0^2$ , and thus, the effective noise level for the binocular and monocular viewing conditions is given by

$$\sigma_B^2(z, \phi) = \frac{\Delta z_{T,B}^2(z, \phi)}{d_0^2} \quad (\text{A10})$$

$$\sigma_M^2(z, \phi) = \frac{\Delta z_{T,M}^2(z, \phi)}{d_0^2} \quad (\text{A11})$$

where  $\Delta z_{T,B}^2(z, \phi)$  is the threshold under binocular viewing and  $\Delta z_{T,M}^2(z, \phi)$  is the threshold under monocular viewing (see Figure 8A). Under the assumption that the effective noise for monocular cues and binocular stereo cues is statistically independent, we have

$$d_B^2(z, \phi) = d_\delta^2(z, \phi) + d_M^2(z, \phi) \quad (\text{A12})$$

It follows by setting  $d_\delta^2(z, \phi) = d_0^2$  that the binocular stereo threshold alone is given by

$$\frac{\Delta z_{T,\delta}^2(z, \phi)}{\sigma_B^2(z, \phi)} - \frac{\Delta z_{T,\delta}^2(z, \phi)}{\sigma_M^2(z, \phi)} = d_0^2 \quad (\text{A13})$$

Substituting equations (A10) and (A11) into (A13) and simplifying gives

$$\Delta z_{T,\delta}(z, \phi) = \frac{\Delta z_{T,B}(z, \phi) \Delta z_{T,M}(z, \phi)}{\sqrt{\Delta z_{T,M}^2(z, \phi) - \Delta z_{T,B}^2(z, \phi)}} \quad (\text{A14})$$

which is text Equation 1.