

Dichotomy between luminance and disparity features at binocular fixations

Yang Liu

Department of Electrical and Computer Engineering and
Center for Perceptual Systems, University of Texas at Austin,
Austin, Texas, USA



Lawrence K. Cormack

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



Alan C. Bovik

Department of Electrical and Computer Engineering and
Center for Perceptual Systems, University of Texas at Austin,
Austin, Texas, USA



Analysis of the statistics of natural scene features at observers' fixations can help us understand the mechanism of fixation selection and visual attention of the human vision system. Previous studies revealed that several low-level luminance features at fixations are statistically different from those at randomly selected locations. In our study, we conducted eye tracking experiments on naturalistic stereo images presented through a haploscope and found that fixated luminance contrast and luminance gradient are generally higher than randomly selected luminance contrast and luminance gradient, which agrees with previous literature, but the fixated disparity contrast and disparity gradient are generally lower than randomly selected disparity contrast and disparity gradient. We discuss the relevance of our findings in the context of the complexity of disparity calculations and the metabolic needs of disparity processing.

Keywords: stereopsis, fixation, eye tracking, natural scenes

Citation: Liu, Y., Cormack, L. K., & Bovik, A. C. (2010). Dichotomy between luminance and disparity features at binocular fixations. *Journal of Vision*, 10(12):23, 1–17, <http://www.journalofvision.org/content/10/12/23>, doi:10.1167/10.12.23.

Introduction

The spatial acuity of the human eye is highest at the foveola, falling off rapidly toward the periphery. The overwhelming amount of information that is presented to the human eye promotes the idea of selecting important regions for detailed examination, by directing the gaze to interesting locations. In this way, images collected from the areas of higher interest are sampled more densely by the fovea. Toward this end, the human visual system (HVS) uses a combination of steady eye fixations linked by rapid ballistic eye movements called saccades (Yabus, 1967). During fixations, high-resolution foveal image features are processed by the HVS. It is important to understand the nature of the information that is being processed by the HVS, as this will better enable modeling of visual function and may prove useful for applications such as active artificial vision. Since visual fixations are clearly not random, understanding the nature of the visual input requires an understanding of what draws visual attention. The statistics of the visual information that enters the HVS from the region of gaze shapes the internalization of the outside world in the brain. Yet, most studies of natural scene statistics have operated by

measuring image features over ensembles of entire images, without considering the fact not all scene information is equally represented by the HVS.

There has been some work regarding the nature of visual information at the point of fixation. A number of researchers have studied image features at fixations using eye tracking instruments. While visual fixations are certainly task-driven (Hayhoe & Ballard, 2005), low-level, precognitive features certainly play a role as well. Starting from a bottom-up point of view, the fundamental question is: what are the visual features that fixated regions possess that make them different (more attractive) than other places? Reinagel and Zador (1999) compared the statistics of natural images at point of gaze to the statistics of patches selected randomly from the same image sets. They found that the regions around human fixations tend to have higher spatial contrasts and spatial entropies than random fixation regions, which suggests that the human eye–brain system may try to select image regions that help maximize the information content transmitted to the visual cortex, by minimizing the redundancy in the image representation. By varying the patch sizes around fixations, Parkhurst and Niebur (2003) found that the largest difference between the luminance contrast of fixated regions and that of image shuffled (pseudo-

random) regions is observed when the patch size is about 1 degree. Studies on eye movements of macaque monkeys (Kayser, Nielsen, & Logothetis, 2006) also showed that luminance contrast is a strong cue that affects fixations. In a recent study (Rajashekar, van der Linde, Bovik, & Cormack, 2007), fixated image patches were foveated using an eccentricity-based model. Higher order statistics than in prior studies were analyzed. They found that band-pass contrast showed a notably larger difference between fixations and random patches than other higher order statistics. Other features found to attract fixations (in decreasing order of attractiveness) included band-pass luminance, RMS contrast, and luminance. These data were subsequently used to develop an image processing “fixation predictor” (Rajashekar, van der Linde, Bovik, & Cormack, 2008). In a visual search experiment with known targets masked by $1/f$ noise (Rajashekar, Bovik, & Cormack, 2006), observers were reported to fixate on regions with structures similar to the target’s shape. Along similar lines, Najemnik and Geisler (2005) and also Raj, Geisler, Frazor, and Bovik (2005) showed that low-level visual fixations can be viewed as an information gathering process.

There is now a body of literature on visual fixations on 2D luminance images that, while still incomplete, at least appears to be largely consistent across the studies that have been done. Yet, we live in a dynamic chromatic 3D world, and these additional dimensional attributes are likely to play a role in the way visual fixations are placed. In particular, there has been very little work done on analyzing the nature or statistics of images and scenes at the point of gaze in three dimensions. In our recent study (Liu, Bovik, & Cormack, 2008), we found that humans are exposed to a rich amount of suprathreshold disparities even in environments of large distances. Certainly, the three-dimensional attributes of the world affect the way we interact with it both visually and physically. One reason for the lack of studies on fixations in 3D space has been the dearth of 3D ground truth natural scene data. The few databases of naturalistic stereo images do not adequately fulfill this need, since what is needed are dense disparity maps for each scene. One such small database does exist containing a few stereo pairs with manually measured and labeled ground truth (the Middlebury database, Scharstein & Szeliski, 2002), but all of the scenes are indoor and are hardly naturalistic.

Motivated by our poor understanding of the nature of 3D visual features at fixations, we conducted a series of binocular eye tracking experiments using naturalistic stereo images. We presented preliminary results (Liu, Cormack, & Bovik, 2007) at the VSS meeting 2 years ago, where we showed that disparity variations at fixations are generally lower than those at other places. The disparity maps were computed using a local correlation method.

One very recent study (Jansen, Onat, & König, 2009) used natural scenes with ground truth disparity map

acquired from laser scanning. They found that disparity appears to be a salient feature that affects eye movements. In particular, they found that the presence of disparity information may affect saccade length but not duration, that disparity appears not to affect the saliency of luminance features, and that subjects tend to fixate at nearer objects earlier than distant objects. Each of these results is reasonably intuitive. The authors also found that in 3D noise images subjects weakly tend to fixate depth discontinuities more frequently than smooth depth regions. One might view this result as intuitive also; as with luminance images, such locations might be deemed interesting.

However, in our study, we obtain a different result that is counterintuitive to this notion. In brief, we find that when viewing naturalistic 3D scenes, humans tend to fixate away from large disparity gradients, preferring instead smooth depth regions. While the tendency of fixating on low disparity variation regions runs counter to the notions arrive at in luminance studies, we nevertheless suggest a logical explanation for this result: since disparity computations are difficult enough already, with heightened complexity near large gradients or discontinuities, the HVS chooses to steer away from these difficult points, settling on smooth, easy to compute 3D surfaces instead. The necessary 3D surface representations are probably adequately computed in this way for most visual tasks. Girshick, Burge, Erlikhman, and Banks (2008) reported that humans have a heavily biased prior toward zero slant, which is not observed in the half-cosine distribution of surface slant in natural scenes. This may suggest that a frontal parallel surface bias exists at fixations.

We used a database of naturalistic stereo images on which we performed eye tracking experiments. We studied 2D (luminance) and 3D image features at the measured fixations. Our results on luminance fixation statistics are concordant with what previous studies have found: luminance contrast and gradient at fixated patches is generally somewhat higher than at randomly selected patches. However, we found that the disparity contrast and gradient at fixated patches is generally significantly *lower* than that at random patches on natural images.

There are a number of immediate implications to this result. First, luminance and depth discontinuities often coincide. Our results show that while large-magnitude luminance gradients are attractors, large-magnitude depth gradients appear not to be (and may even be repellers). The frequent co-occurrence of luminance and depth discontinuities implies that the HVS must use additional criteria in deciding whether to fixate a discontinuous region. Second, the result has implications regarding the information that is used in stereopsis; it implies that active stereoscopic systems may operate more efficiently by building smooth surface representations of scenes without expending much effort to compute sharp delineations of 3D surface boundaries.

Methods

Observers

Three male observers participated in this study. All observers had normal or corrected-to-normal vision. Their stereo ability was tested to be normal by presenting them random dot stereograms. Subject LKC had extensive experience in psychophysical studies and knew the purpose of the experiment. Subjects JSL and CHY were naive observers.

Stimulus

We manually selected 48 grayscale stereoscopic outdoor scenes (Hoyer & Hyvärinen, 2000) that contained mountains, trees, water, rocks, bushes, etc., but avoided manmade objects (examples of these images are shown in Figure 1). We did not have the ground truth disparity data from these scenes. Instead, we relied on a simple yet biological inspired method to approximate ground truth disparities. Models of binocular complex neurons (Anzai, Ohzawa, & Freeman, 1999; Fleet, Wagner, & Heeger, 1996; Ohzawa, DeAngelis, & Freeman, 1990; Qian, 1994) commonly contain a cross-correlation term. For example, a binocular complex cell's response can be expressed as the sum of the squares of two quadrature simple cell responses, while the simple cell responses sum the dot

products of the receptive field and the image from both eyes:

$$r_1 = S_{LE} + S_{RE}, \quad (1)$$

$$r_2 = S_{LO} + S_{RO}, \quad (2)$$

$$r_c = r_1^2 + r_2^2 = S_{LE}^2 + S_{LO}^2 + S_{RE}^2 + S_{RO}^2 + 2S_{LE}S_{RE} + 2S_{LO}S_{RO}. \quad (3)$$

Here S_{AB} denotes inner products between left/right eye images and even/odd responses of simple cell receptive fields, where $A \in \{L, R\}$ indicates left or right eye and $B \in \{E, O\}$ indicates even or odd symmetry. The responses of even symmetric and odd symmetric simple binocular cells (quadrature pairs) are denoted by r_1 and r_2 , respectively. In Equation 3, the last two terms $S_{LE}S_{RE}$ and $S_{LO}S_{RO}$ are the cross-correlation between band-pass left image and right image. In the model, the outputs of a binocular complex neuron largely depend on this term. Aside from neurophysiological evidence, psychophysical studies (Cormack, Stevenson, & Schor, 1991) also showed that interocular correlation is a decisive factor in stereopsis. Furthermore, local correlations have been extensively used to resolve correspondences in numerous computational stereo algorithms too (for a good review, see Brown, Burschka, & Hager, 2003; Scharstein & Szeliski, 2002). Inspired by this neurobiological basis of stereopsis, Filippini and Banks (2009) built a local



Figure 1. Examples of naturalistic stereo images for eye tracking.

correlation stereopsis model. They conducted psychophysical experiments with human observers and compared human results with the results of their model under the same experiment setup. Using this model, they explained two well-known constraints of human stereopsis: the disparity-gradient limit, which is the inability to perceive depth when the change in disparity within a region is too large, and the limit of stereoresolution, which is the inability to perceive spatial variations in disparity that occur at too fine a spatial scale.

Our stereo correspondence algorithm is very similar to the method of Filippini and Banks (2009). Given a pixel (x_r, y_r) in the right image, we defined a search window centered on the same pixel location (preferring zero disparity) in the left image (Figure 2). The width of the search window in the left image was 161 pixels (3.2°), and the height of the window was 5 pixels (0.1°). Given a $1^\circ \times 1^\circ$ patch in the right image centered on the pixel (x_r, y_r) , the algorithm computes the normalized cross-correlation between the right patch and a candidate $1^\circ \times 1^\circ$ left patch centered on each pixel in the 161×5 search window, centered at (x_r, y_r) in the left image using the following equation:

$$C(d, s) = \frac{\sum (I_l(x, y) - \mu_l)(I_r(x + d, y + s) - \mu_r)}{\sqrt{\sum (I_l(x, y) - \mu_l)^2} \sqrt{\sum (I_r(x + d, y + s) - \mu_r)^2}}. \quad (4)$$

Here, I_l and I_r are the left and right images, while μ_l and μ_r are the mean luminance values of the left and right patches. The normalized interocular correlation $C(d, s)$ always takes a value between -1 and 1 , where 1 means the two patches are almost identical up to multiplicative scaling, and -1 means the two patches have reversed

luminance profiles (bright locations in the left are dark in the right).

The left patch yielding the largest cross-correlation was deemed to be the matched patch. The location (x_l, y_l) that it centered on was then the matched pixel for the right pixel (x_r, y_r) , and the horizontal disparity at pixel (x_r, y_r) was taken to be $D(x_r, y_r) = x_r - x_l$. While this disparity is not the conventional angular disparity defined in unit of degrees, it represents an angular difference given an assumed viewing geometry. The algorithm computed a disparity for every pixel in the right image, yielding a dense disparity map D . Naturally, we do not claim that this simple algorithm duplicates the disparity processing of the large population of neurons dedicated to the task. Nevertheless, it is an effective method and appears to be compatible with significant aspects of human stereopsis (Filippini & Banks, 2009).

It could be argued that since we are only interested in local scene statistics, why not record the movements of both eyes and use the disparity between the recorded left and right fixations to find the correct match? There are several reasons why this is not a practical solution. First, there are fixation disparities that occur between the right eye and the left eye. Most people have a fixation disparity that is less than 6 arcmin but can be as large as 20 arcmin with peripheral visual targets (Wick, 1985). When fixation disparity occurs, the image of an object point that a person is trying to fixate on does not fall on exactly corresponding points. Second, each Purkinje eye tracker has an accuracy of about 7 arcmin (the median offset from real fixations); hence the median error between two eye trackers is about 14 arcmin, which corresponds to about 12 pixels. This is quite a large error, considering that the stereo images are only 800×600 . Third, the fixation detection algorithm (associated with eye tracking) of the two eye paths can also introduce unwanted noise into the

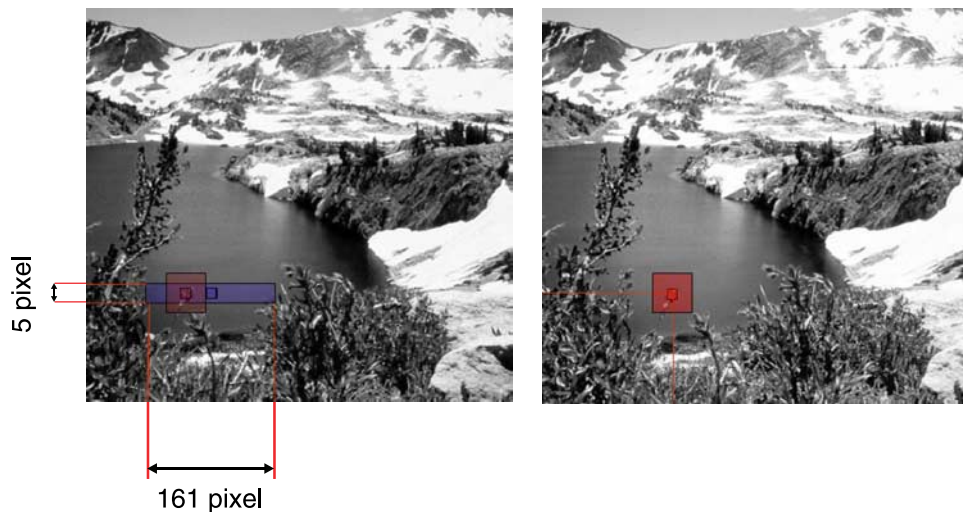


Figure 2. Given a pixel in the right image, a search window is defined in the left image. The best matched pixel is found by exhaustive search within the window.

corresponding fixation locations. Registered binocular eye tracking with a precision approximating that of stereoscopic processing is difficult to accomplish in practice. Lastly, since we are interested in disparity features within neighborhoods (not just points) of fixations, a dense disparity map is required for all points in the neighborhood. Hence, local processing of the type that our disparity algorithm accomplishes would be required anyway.

Equipment

Stereo images were displayed on two 17-inch gamma-calibrated monitors (Figure 3). The distance between the monitors and the observer was 124 cm. Each monitor's screen resolution was set at 800×600 pixels, corresponding to about 50 pixels per degree of visual angle. The total spatial extent of each display was thus about $16^\circ \times 12^\circ$ of visual angle.

A mirror stereoscope was placed between the two monitors and the observers to completely separate the displays from the left and right monitors. The left (right) eye could only access the view from the left (right)

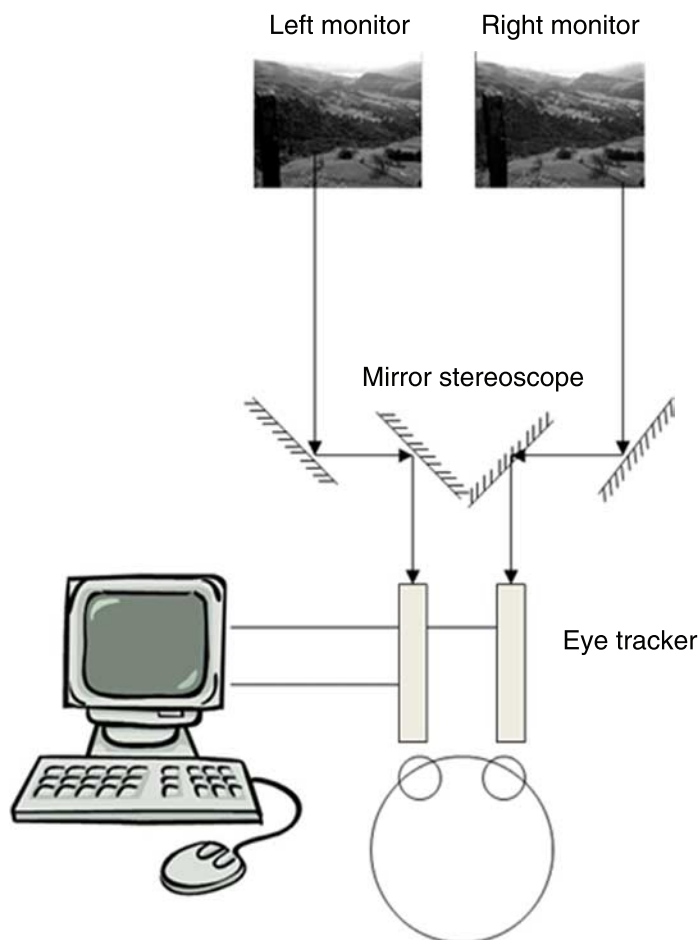


Figure 3. The experimental setup.

monitor. Eye movements were recorded by an SRI Generation V Dual Purkinje eye tracker. This eye tracker has an accuracy of $<10'$ of arc, a response time of under 1 ms, and bandwidth of DC to >400 Hz. The output of the eye tracker (horizontal and vertical eye position signals) was first low pass filtered by the hardware, then sampled at 200 Hz by a National Instruments data acquisition board in a Pentium IV host computer, where the data were stored for offline data analysis. The observers used a bite bar and a forehead rest to restrict their head movements.

Eye tracking

For each observer's visual comfort and stereoscopic viewing quality, we carefully adjusted the position and angle of the mirrors on the haploscope before every viewing session. A viewing session was composed of 48 stereo image pairs. At the beginning of each session, a $0.3^\circ \times 0.3^\circ$ crosshair was displayed on the centers of both monitors to help the observers to fuse by fixating on it. When the correct fuse (the crosshair was perceived single) was achieved, the observer pressed a button to start the calibration. Two 3×3 calibration grids were displayed on the monitors, respectively. After the observers visited all 9 dots, a linear interpolation was then done to establish the transformation between the output voltages of the eye tracker and the position of the subject's gaze on each computer display. The calibration also accounted for cross talk between the horizontal and vertical voltage measurements. After correct calibration, a $0.3^\circ \times 0.3^\circ$ crosshair was displayed on the centers to force all observers to start from the same center position. The stereo images were displayed on two monitors for 10 s during which the eye movements were recorded. Between two consecutive image pairs, two identical Gaussian noise images were displayed for 3 s on both monitors to help suppress afterimages corresponding to the previous stereo pairs that may otherwise have attracted fixations. Then, a $0.3^\circ \times 0.3^\circ$ crosshair was displayed on the centers of both monitors to help the observers to fuse before the presentation of the next stereo pair.

The calibration routine was repeated compulsorily every 10 images; moreover, a calibration test was done every 5 images. This required that the observer fixate for 500 ms within a 5-s time limit on a central square region ($0.3^\circ \times 0.3^\circ$) prior to progressing to the next image in the stimulus collection. If the calibration had drifted, the observer would be unable to satisfy this test, and the full calibration procedure was rerun.

Observers who became uncomfortable during the experiment were allowed to take a break of any duration they desired.

The ambient illumination in the experiment room was kept constant for all observers, with a minimum of 5-min luminance adaptation provided while the eye tracker was calibrated.

Results

The movements of the right eye were recorded and analyzed. Figure 4 shows an example of the eye path and fixations of one observer during the 10-s viewing time. The scan path is displayed using a red line, while the fixations are represented by green dots. Usually, the first fixation is the center of the image since all observers are forced to start from the center. We removed the first fixations from the analysis to cancel this bias.

Denote the right image as I , and the dense disparity map as D . We compute the luminance gradient map

$$G_l = \sqrt{\frac{\partial I^2}{\partial x} + \frac{\partial I^2}{\partial y}}, \quad (5)$$

and the disparity gradient map

$$G_d = \sqrt{\frac{\partial D^2}{\partial x} + \frac{\partial D^2}{\partial y}}, \quad (6)$$

using the MATLAB function `gradient(X)`. Here we define luminance contrast as the RMS contrast of a luminance patch

$$\sqrt{\frac{1}{N-1} \sum_{i=1}^N \left(\frac{I_i - \bar{I}}{\bar{I}} \right)^2}, \quad (7)$$

and likewise, disparity contrast as

$$\sqrt{\frac{1}{N-1} \sum_{i=1}^N (D_i - \bar{D})^2}, \quad (8)$$

which is the standard deviation of a disparity patch. All of the following analyses are based on these four scene features: luminance contrast, disparity contrast, luminance gradient, and disparity gradient.

Figure 5 depicts these scene features for one example stereo pair. Figure 5a shows the right luminance image. Figures 5c and 5e display the dense maps of luminance gradient and luminance contrast with a patch size of 0.5° (note that different patch sizes yield different luminance contrast maps).

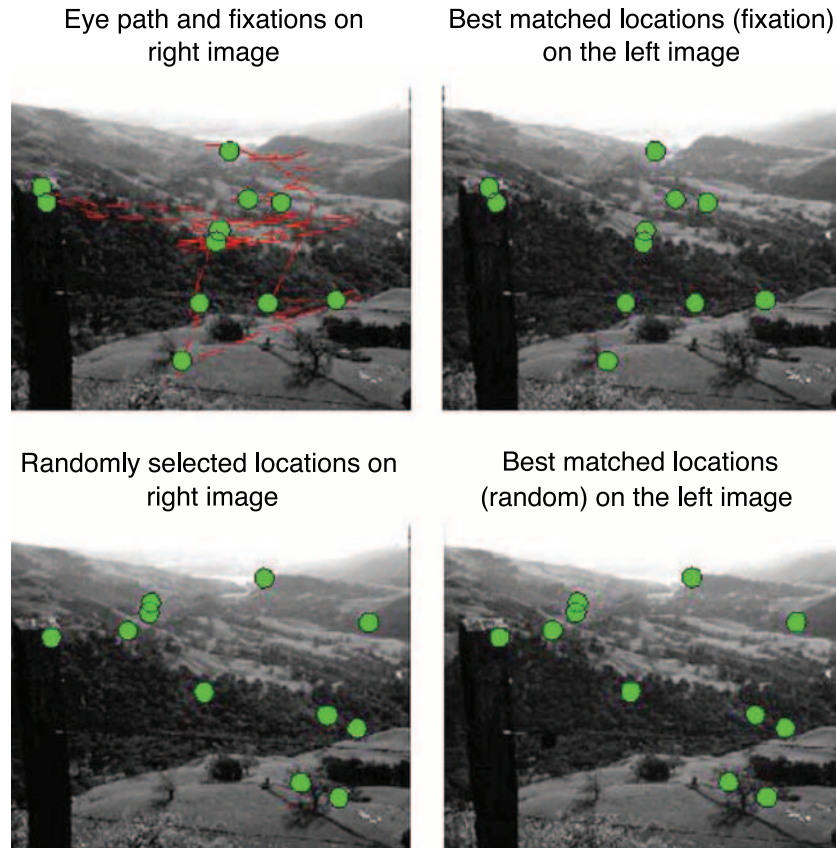


Figure 4. The gaze path and fixations of subject CHY on the right image are shown on the top left panel (for cross fusing). The best matched locations in the left image of these fixations are displayed on the top right panel. The bottom left and right panels show random selected locations and their matches for comparison.

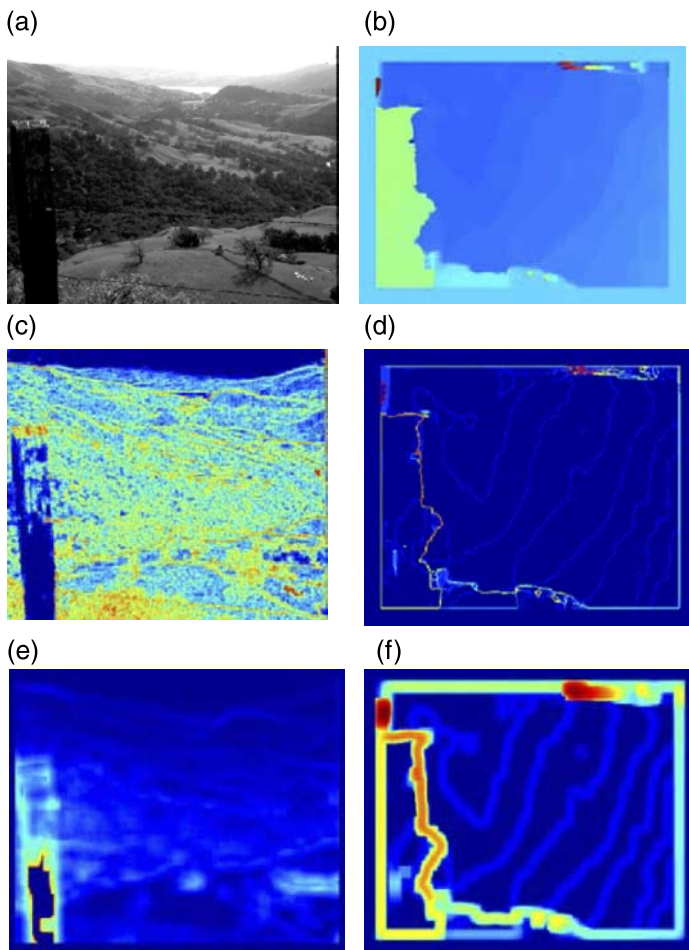


Figure 5. (a) Luminance image. (b) Disparity map. (c) Luminance gradient. (d) Disparity gradient. (e) Luminance contrast (patch size 0.5°). (f) Disparity contrast (patch size 0.5°). The borders in the disparity map and the luminance image were excluded from the analysis.

Figure 5b shows the disparity map derived from the maximum correlation method. Disparity gradient and disparity contrast maps are shown in Figures 5d and 5f. Note that the borders around the disparity map are caused by an inability of the correlation method: pixels outside the borders do not have a complete $1^\circ \times 1^\circ$ patch that centers on them, so they are not included in the stereo matching. In all of the following analyses, we excluded fixations falling outside the borders. We also manually eliminated fixations that fell on small regions where correspondences could not be resolved, viz., areas of very smooth luminance where stereo algorithms generally fail unless other information is used (e.g., the top right corner of Figure 5b).

Two methods to evaluate fixations

We adopt two different methods to compare scene statistics at fixations with those at other locations. The first

method selects random fixations that are uniformly distributed on the image plane. We compared several scene statistics at true fixations with these randomly picked locations. Because the stereo image database contains a wide variety of scene content and distance scales across images, this method has the advantage of minimizing interimage variability between two different scenes by only comparing fixations and random locations within the same image.

However, many researchers have noticed that there is a bias toward image center under different experimental conditions (Judd, Ehinger, Durand, & Torralba, 2009; Mannan, Ruddock, & Wooding, 1997; Parkhurst & Niebur, 2003, 2004; Tatler, 2007; Tatler, Baddeley, & Gilchrist, 2005). Given the current experimental setup, if the center bias is true, and if the image features have a non-homogenous distribution (for example, image features at the center are statistically different from all other locations), we cannot conclude that the differences between low-level scene features are relevant in fixation selection. This is why we use the second method: instead of randomly selecting locations, we randomly select images. We refer to this as the image shuffling method. In this method, we overlap the fixations on image A onto a randomly selected image B from the image database and compare the scene statistics of the real fixations on image A with those of “random” (overlapped) locations on image B.

The image-shuffled database therefore simulates a human observer who is not driven by the image features of that particular image, but otherwise satisfies all criteria of human eye movement statistics. Further, this methodology of simulating random fixations accounts for both known potential biases of human eye movements (such as the tendency of observers to fixate at the image center, and the log-normal distribution of saccade magnitudes) and unknown biases (such as possible correlations between the magnitude and the angle of the saccades). Tatler et al. (2005) provide a discussion of how such biases might influence the statistics of image features.

Method 1: Uniformly distributed random locations within an image

Suppose that an observer made f_i fixations on the i th image. Then, the total number of fixations that the observer made during a session is $\sum_i f_i$. We assume that a random observer also made the same number of fixations as the subject did. That is, for the i th image, the random observer also selected f_i fixations uniformly distributed on the image plane. For each human observer, we assume that there are 100 random observers, each making the same number of fixations on each image as does the human observer. For example, if the overall number of fixations that subject LKC made was 486 fixations on 48 images, then each random observer selected 486 random

locations as well. The total number of random locations is 48,600. We want to know whether or not there is a statistically significant difference between image features at fixations and those at randomly selected locations by comparing the human observer data and the 100 random observers' data.

We computed disparity gradient, disparity contrast, luminance gradient, and luminance contrast at fixations and random locations as defined previously. The results are shown in Figure 6.

The first three rows of Figure 6 show the mean disparity gradient, mean disparity contrast, mean luminance gradient, and mean luminance contrast of fixations (red) and random locations (blue) as a function of patch sizes (from 0.5 to 1.5 deg) for subjects LKC, JSL, and CHY. The bottom (fourth) row of Figure 6 displays the same plots for all subjects combined. The error bars show the 95% confidence intervals (CIs).

The most conspicuous difference between fixations and random locations are disparity features. The mean

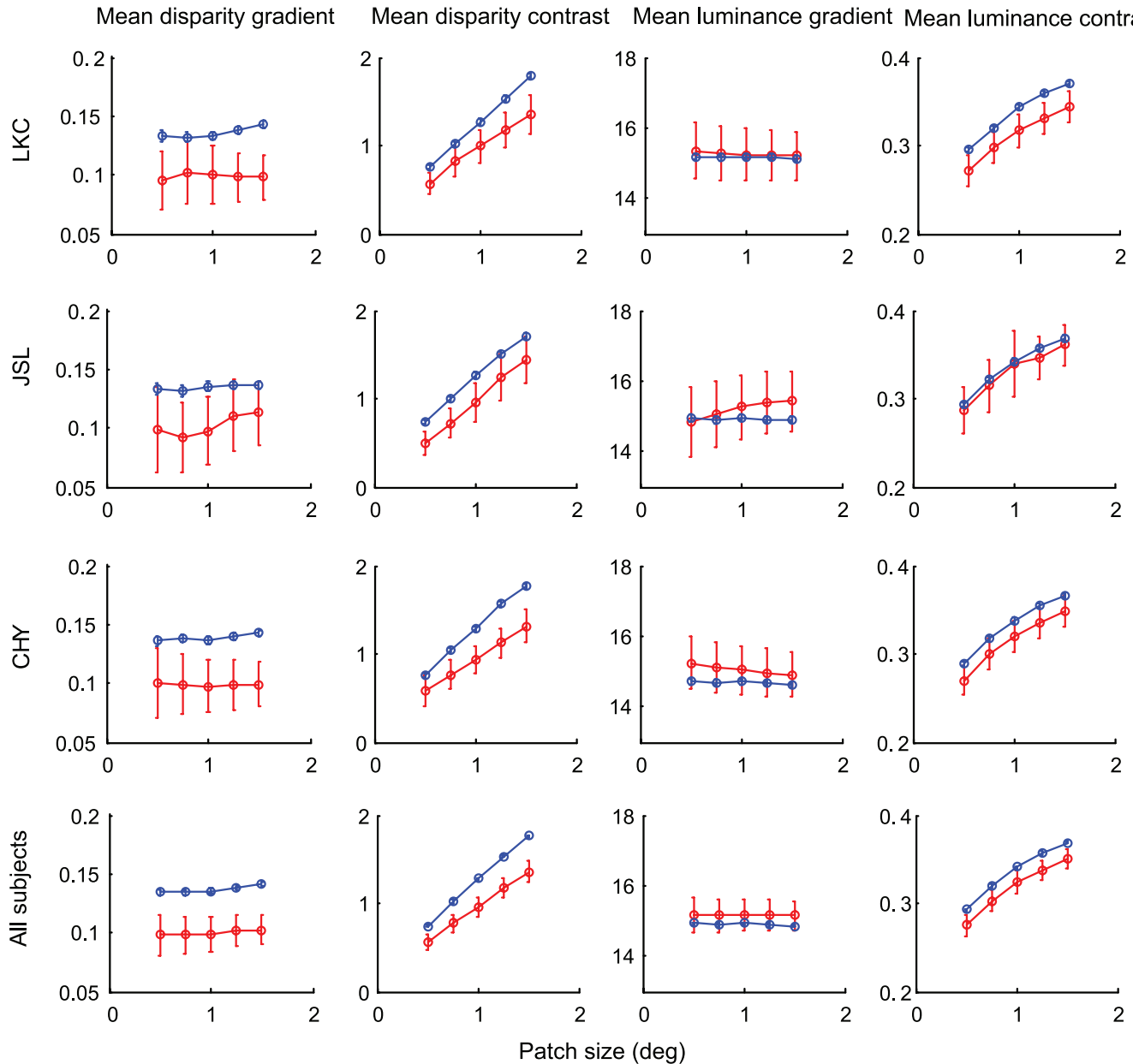


Figure 6. We plotted the mean disparity gradient, mean disparity contrast, mean image gradient, and mean image contrast of fixations (red) and random locations (blue) of several patch sizes in each row for each subject, and all subjects combined (the last row). The error bar shows the 95% confidence interval. We found that all subjects behave similarly. The mean disparity gradient and mean disparity contrast at fixations are smaller than those at random locations. Interestingly, the mean luminance gradients at fixations are larger than those at random locations, though they are not clearly separable (the blue plots are within the confidence interval of the red). However, the mean luminance contrasts at fixations are smaller than those at random locations.

disparity contrast and mean disparity gradient at fixations are clearly lower than those at random locations. The 95% CIs have no overlap. All subjects behaved consistently.

For luminance features, the results are generally smaller relative to the measurement error. The mean luminance gradient at fixations appears to be generally higher than that at random locations for all subjects. However, the 95% CIs of random locations are completely covered by the 95% CIs of fixations. While there might be a small statistical mean effect for observer type (random vs. human), it is very small relative to the luminance fluctuations across image locations. The mean luminance contrasts at fixations are lower than those at random locations for all subjects. However, the difference is not as strong as for the disparity features. For example, the 95% CIs of mean luminance contrast at random locations are completely overlapped by the CIs at fixations for subject LKC. In addition, the separation between the 95% CIs of random and fixation points seems to be less conspicuous than that of disparity features, as shown in column 1 and column 2 of Figure 6.

It is also interesting that the results for luminance gradient and luminance contrast measurements are dissimilar, given the expectation that both represent similar variations within the patch. We suspect that the mean of these scene features may obscure the rather large scene differences across images. The absolute values of luminance features and disparity features are definitely dependent on the image database and also vary across images. Since we are mostly interested in the relative difference between scene features at fixations and random locations, we computed the ratios of scene statistics at fixations to the same statistics at random locations on each image (Rajashekar et al., 2007).

Ratios of luminance gradient and luminance contrast

For the i th image, we compute the mean luminance contrast at the f_i fixations as

$$C_1^i = \frac{\sum_1^{f_i} C_1(x_j, y_j)}{f_i}, \quad (9)$$

where (x_j, y_j) is the location of the j th fixation, and C_1 is the luminance contrast map.

We also compute the mean luminance contrast at the f_i random locations as

$$C_{r1}^i = \frac{\sum_1^{f_i} C_1(u_j, v_j)}{f_i}, \quad (10)$$

where (u_j, v_j) is the location of the j th random location.

We define patch luminance gradient as the mean gradient of the patch:

$$\bar{G}_1 = \frac{\sum_1^N G_1(x, y)}{N}, \quad (11)$$

where G_1 is the luminance gradient map.

Similarly, we compute the mean patch luminance gradient of the f_i fixations for the i th image as

$$G_1^i = \frac{\sum_1^{f_i} \bar{G}_1(x_j, y_j)}{f_i}, \quad (12)$$

where (x_j, y_j) is the location of the j th fixation.

We also compute the mean patch gradient of the f_i random locations for the i th image as

$$G_{r1}^i = \frac{\sum_1^{f_i} \bar{G}_1(u_j, v_j)}{f_i}, \quad (13)$$

where (u_j, v_j) is the location of the j th random location.

For each image, we then define the fixation-to-random luminance contrast ratio $RC_1 = C_1^i / C_{r1}^i$ and the fixation-to-random luminance gradient ratio $RG_1 = G_1^i / G_{r1}^i$. If $RC_1 > 1$, it means that the fixated patches generally have a larger luminance contrast than randomly selected patches on the image being considered. If $RC_1 < 1$, then the meaning is reversed. The same meaning applies to the luminance gradient ratio.

The distribution of the ratios can be visualized using boxplots (Figure 7). The bottom and top edges represent the first quartile and the third quartile of the distribution. The red line at the center of the box shows the median. The length of the whisker is set to be 1.5 times the interquartile range (the length between the third quartile and the first quartile, that is, the height of the box). Those outliers that fell outside of the whiskers are plotted as red crosses.

For observer LKC, we show the boxplots of the luminance contrast ratio and the luminance gradient ratio at different patch sizes (0.5° – 1.5°) in Figure 7a. In order to make comparison with the unit level easier, we plot 1 as a blue circle overlapped on each boxplot. From the plots, it is easily seen that the medians (the red bar) fall well above unity for both the luminance contrast ratio and the luminance gradient ratio. The ratios of the luminance gradient and luminance contrast for subjects CHY and JSL display the same behavior, as shown in Figures 7b and 7c.

Ratios of disparity gradient and disparity contrast

The same analysis method that was used on the luminance contrast and luminance gradient was also applied for the analysis of disparity.

We calculate the mean disparity contrast C_d^i on the fixated patches, and the same quantity C_{rd}^i on the randomly selected patches. The ratio of disparity contrast between the fixated patches and the randomly selected patches is defined as $RC_d = C_d^i / C_{rd}^i$.

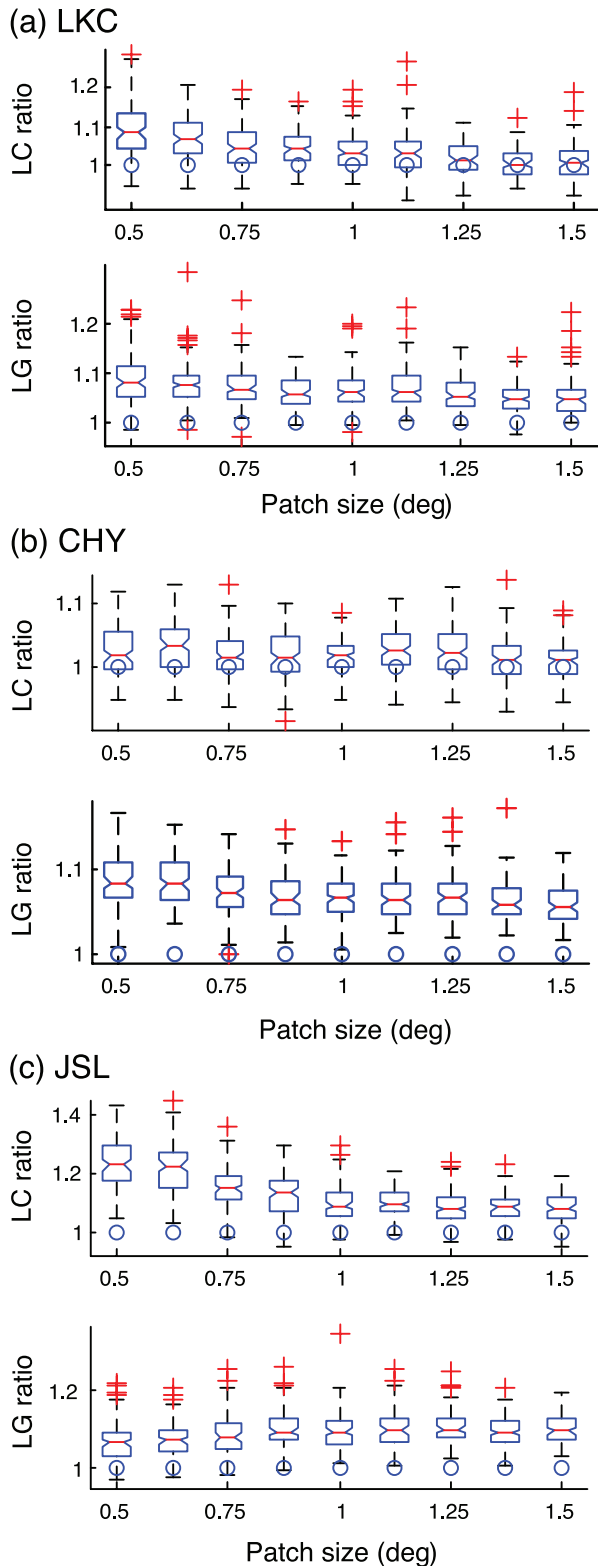
The mean patch disparity gradient at the fixated patches (G_d^i) and the randomly selected patches (G_{rd}^i) is also

calculated. The ratio of the disparity gradient between the fixated patches and the random patches is defined as $RG_d = G_d^i / G_{rd}^i$. As before, if the ratios are significantly greater than 1, then fixated patches tend to have a larger disparity contrast and gradient than randomly picked locations.

In Figure 8a, we show boxplots of the disparity contrast ratio and the disparity gradient ratio at different patch sizes for subject LKC. Quite surprisingly, the medians and the third quartiles are well below one for *all* patch sizes. This means that the fixated patches generally had a smaller disparity contrast and gradient than the random locations. The results of other observers all proved to be similar to that for LKC, as shown in Figures 8b (CHY) and 8c (JSL).

We ran 100 simulations for each image and plotted the mean luminance gradient ratio \overline{RG}_1 with 95% CIs and the mean disparity gradient ratio \overline{RG}_d with 95% CIs, for all subjects as shown in Figure 9a. For better comparison, we plot 1 as a straight horizontal line across all patch sizes. The red curves show the ratios of luminance gradient, and the blue curves showed the ratios of disparity gradients. Different markers are used to represent the observers: LKC (*), CHY (○), JSL (Δ). We made a similar ensemble comparison plot for the mean luminance contrast ratio and mean disparity contrast ratio, as displayed in Figure 9b.

From these figures, it is clear that all the mean ratios of luminance gradient/contrast and their 95% CIs fell well above 1, while the mean ratios of disparity gradient/contrast fell well below 1. All of the results we obtained lead to the conclusion that humans tend to fixate at regions of higher luminance variation and lower disparity variation.



Method 2: Random image shuffling

For one set of fixations in image i , we randomly select one image from the whole database. Overlapping the fixations onto the randomly picked image, we computed luminance contrast, luminance gradient, disparity contrast, and disparity gradient of the fixations and those of the

Figure 7. Boxplots of luminance contrast ratio and luminance gradient ratio for three observers: (a) LKC, (b) CHY, and (c) JSL. The bottom and top edges of each box represent the first quartile and the third quartile of the ratio distribution at each patch size. The red bars in the box show the medians of the ratio distributions. The whiskers extended from the top and bottom edges to the most extreme values within 1.5 times the interquartile range from the ends of the box. All other boxplots in Figures 8, 11, and 12 follow the same definition. All outliers outside of the whiskers are labeled as red crosses. The blue circles show 1 for reference. All of the luminance-related ratios have medians larger than 1.

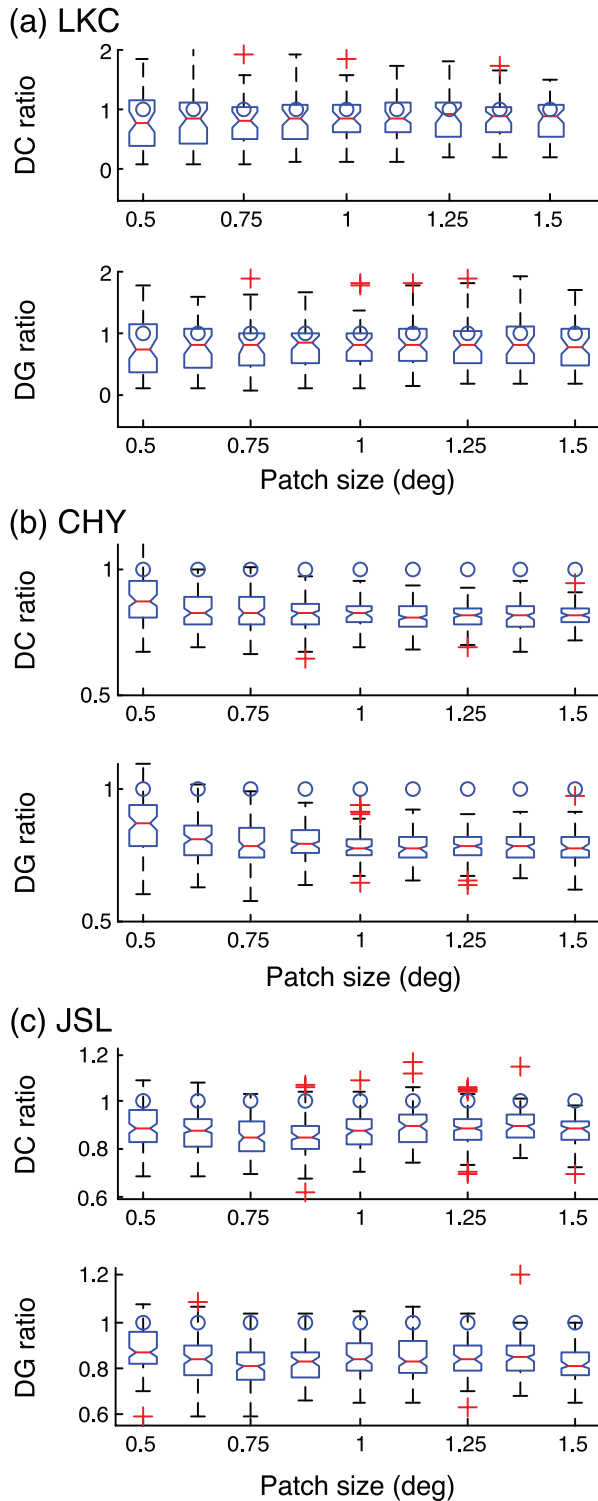


Figure 8. Boxplots of disparity contrast ratio and disparity gradient ratio for three observers: (a) LKC, (b) CHY, and (c) JSL. The blue circles show 1 for reference. All of the disparity-related ratios have medians smaller than 1.

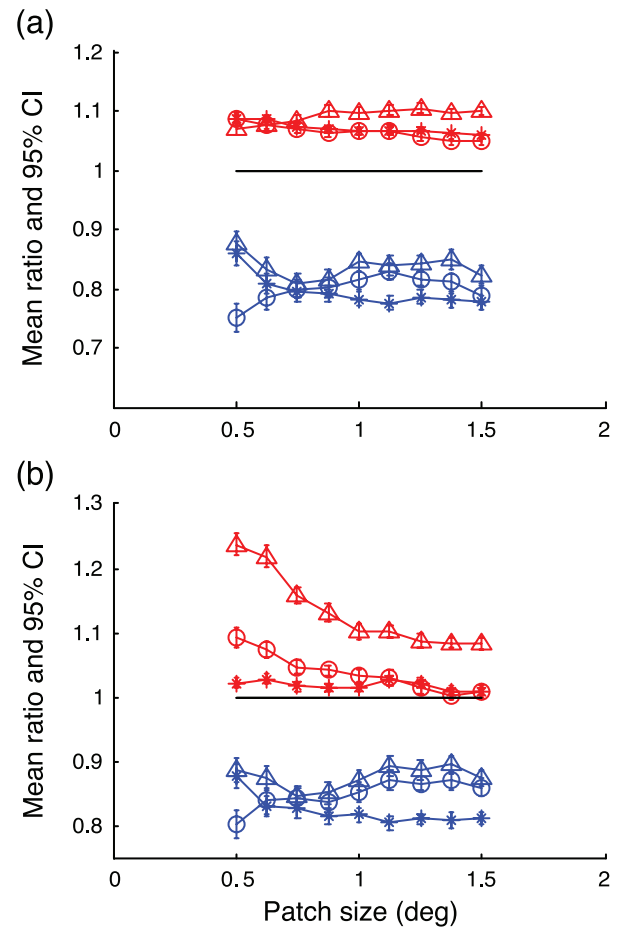


Figure 9. The red curves show (a) the mean luminance gradient ratios, (b) the mean luminance contrast ratios, and their 95% confidence intervals (CIs) of all patch sizes and all subjects. The blue curves show (a) the mean disparity gradient ratios, (b) the mean disparity contrast ratios, and their 95% CIs of all patch sizes and all subjects. Different markers represent different subjects: LKC ("***"), CHY ("o"), JSL (Δ). The black lines at 1 are plotted for reference. All luminance-related ratios are significantly larger than 1, while all disparity-related ratios are less than 1.

“fixations” on the random image. We ran the random picking from the image-shuffled database 100 times.

Using the previously defined scene features, we obtained four features from image i : C_l^i , G_l^i , C_d^i , and G_d^i . We denote the four features from a randomly picked image j as C_{rl}^j , G_{rl}^j , C_{rd}^j , and G_{rd}^j , where all the four scene features were computed by selecting image patches at the overlapped fixations on image j .

We show the mean luminance features and mean disparity features at fixated images and random images in Figure 10, organized identically as in Figure 6. The results are very similar to what we already observed in Figure 6. The fixations have higher mean values of luminance features but lower mean values of disparity features.

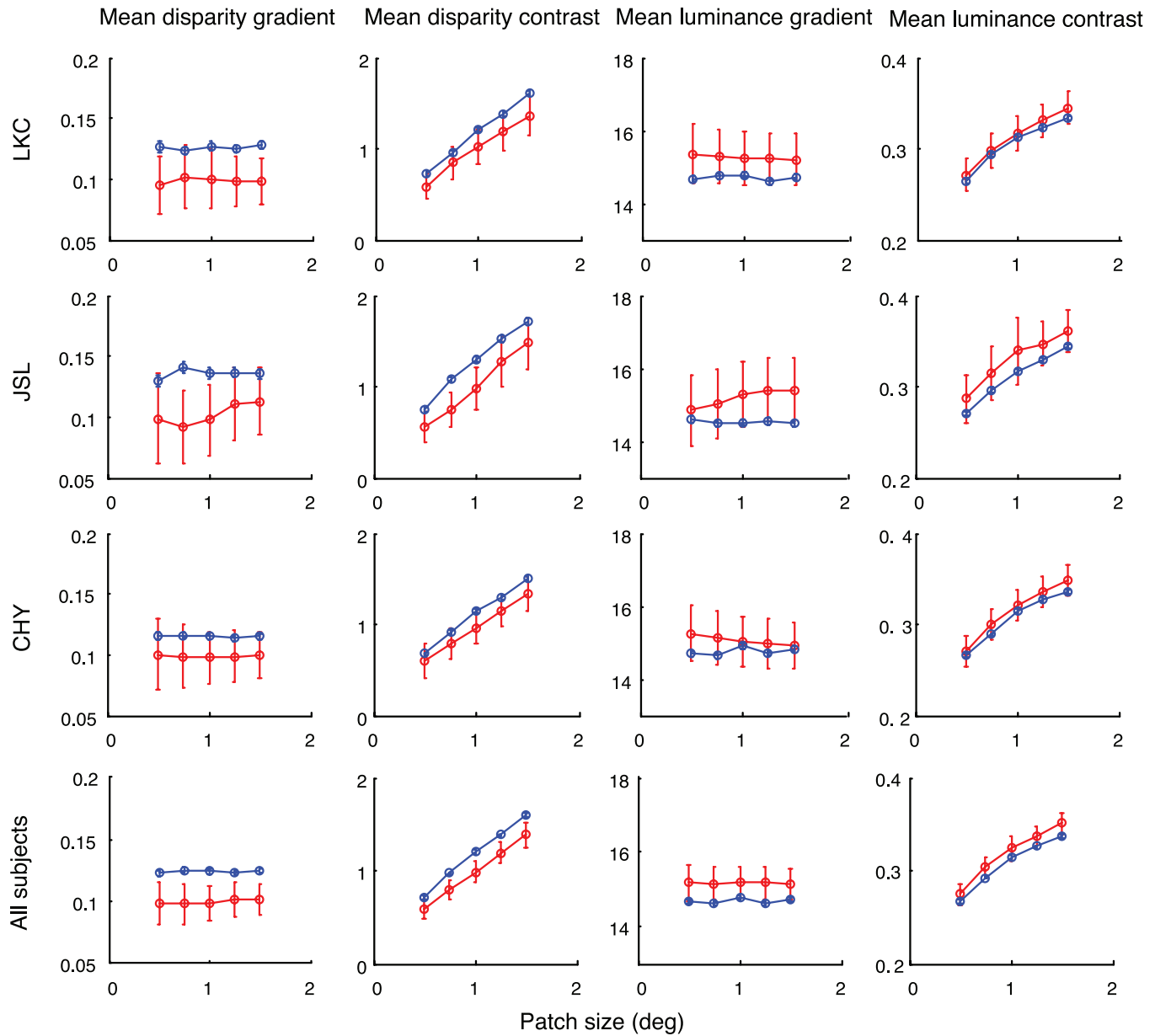


Figure 10. We plotted the mean disparity gradient, mean disparity contrast, mean image gradient, and mean image contrast of fixations (red) and random images (blue). The panel organization is the same as in Figure 6. The error bar shows the 95% confidence interval. We found that all subjects behave similarly. The mean disparity gradient and mean disparity contrast at fixations are less than those at random images, while the mean luminance gradients and mean luminance contrasts are also greater than those at random images.

Ratios of luminance gradient and luminance contrast

We computed the ratio of luminance gradient as

$$RG_1 = G_1^i / G_{\text{rd}}^j, \quad (14)$$

and the ratio of luminance contrast as

$$RC_1 = C_1^i / C_{\text{rd}}^j. \quad (15)$$

We show the boxplots of the fixation-to-random luminance contrast ratio and luminance gradient ratio for

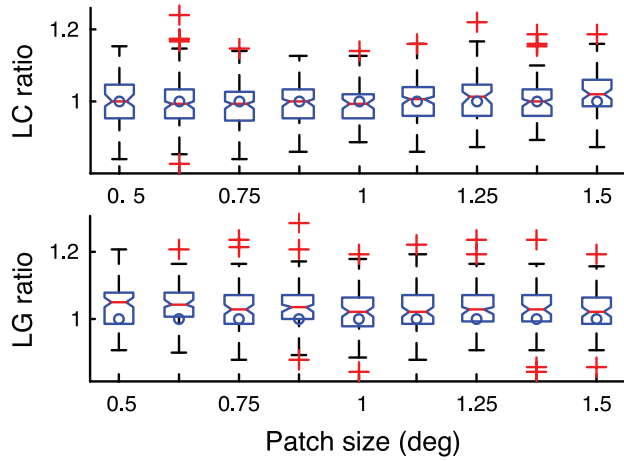
different patch sizes in Figure 11. The medians of luminance contrast ratios and luminance gradient ratios of JSL and CHY are above one in all cases. LKC's luminance gradient ratio also has a higher (>1) median. However, the median of luminance contrast ratio of LKC is very close to 1, which is the only exception.

Ratios of disparity gradient and disparity contrast

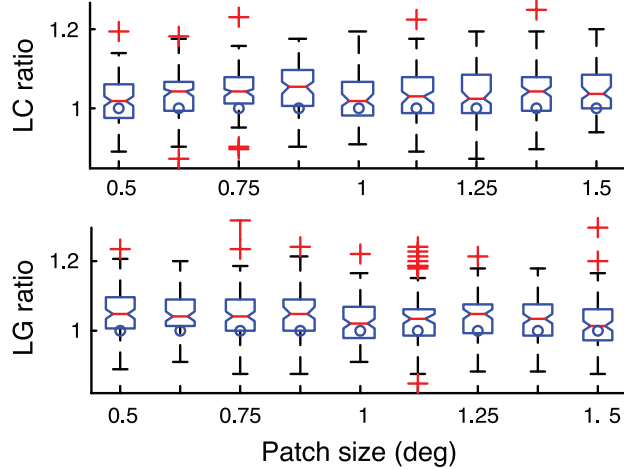
We compute the ratio of mean disparity gradient as

$$RG_d = G_d^i / G_{\text{rd}}^j, \quad (16)$$

(a) LKC



(b) CHY



(c) JSL

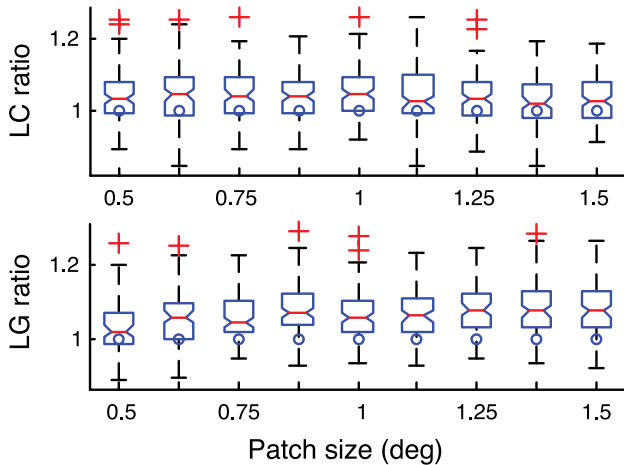
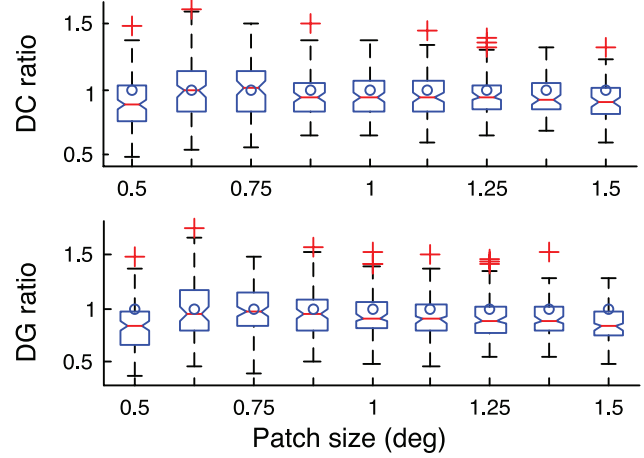


Figure 11. Boxplots of luminance contrast ratio and luminance gradient ratio for three observers: (a) LKC, (b) CHY, and (c) JSL. The blue circles show 1 for reference. Most of the luminance-related ratios have medians larger than 1. The only special case is where LKC has lower luminance contrast ratios, which are close to 1. However, these results largely agree with Figure 7.

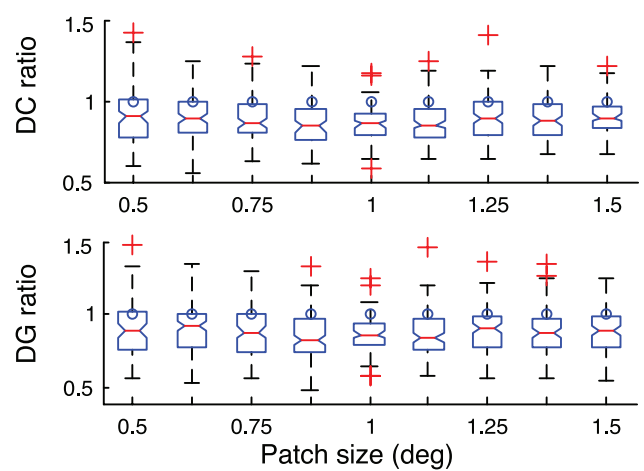
and the ratio of mean disparity contrast as

$$RC_d = C_d^i / C_{rd}^j. \quad (17)$$

(a) LKC



(b) CHY



(c) JSL

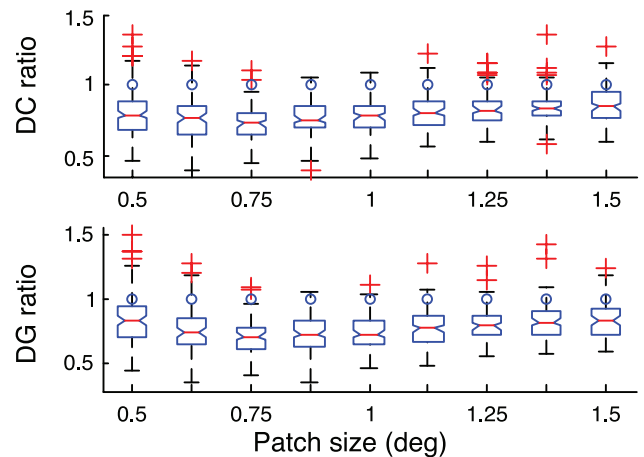


Figure 12. Boxplots of disparity contrast ratio and disparity gradient ratio for three observers: (a) LKC, (b) CHY, and (c) JSL. Again, these results agree with Figure 8.

The distributions of these ratios over different patch sizes are plotted in Figure 12. The medians of ratios are consistently below one for JSL and CHY. LKC also has most of the medians below one except for patch sizes of 0.5 deg and 0.75 deg.

We selected 100 random images for each fixated image and plotted the mean values of the fixation-to-random ratios and the 95% CIs of all these features in Figure 13, where the blue plots show the mean ratios of disparity features and the red plots show the mean ratios of luminance features. We observe very similar behavior as in Method 1. The mean ratios of disparity features were generally less than 1, while the mean ratios of luminance features were larger than 1, except for LKC at a few patch sizes.

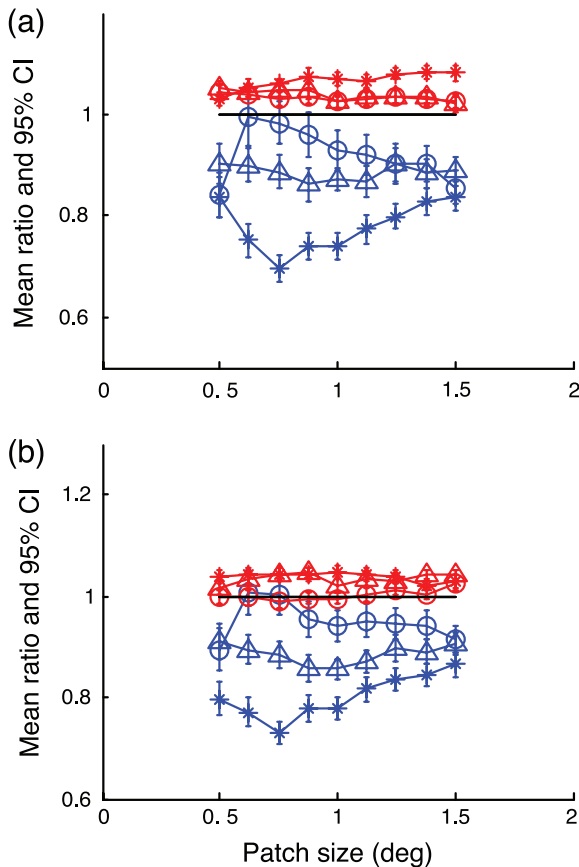


Figure 13. The red curves show (a) the mean luminance gradient ratios, (b) the mean luminance contrast ratios, and their 95% confidence intervals (CIs) for all patch sizes and all subjects. The blue curves show (a) the mean disparity gradient ratios, (b) the mean disparity contrast ratios, and their 95% CIs for all patch sizes and all subjects. Different markers represent different subjects: LKC ("o"), CHY ("*"), JSL (Δ). The black lines at 1 are plotted for reference. Most of the ratios of luminance features are larger than 1, and most of the ratios of disparity features are smaller than 1. LKC has ratios of luminance contrast very close to 1. However, LKC's ratios of luminance contrast are generally greater than his ratios of disparity contrasts. These results are also similar to Figure 9.

Regardless of whether we compared human fixations with randomized locations within the same image, or with the same locations of a randomized image, the results obtained suggest that the fixated locations generally have higher luminance variation and lower disparity variation. The results of image shuffling rule out the possibility of the fixation-to-random difference being caused by the center bias of fixations. Given the consistent results of the two different methods, we believe that luminance features and disparity features may indeed play different and perhaps unexpected roles under binocular conditions.

Discussion

Our findings on luminance statistics are consistent with the generally agreed-upon observation, repeated in several previous studies, that higher luminance variations may attract fixations. However, our finding that disparity gradient and contrast in fixated patches is generally smaller than in random patches has not been observed before and, we think, has not been mentioned.

The most immediate explanation for this behavioral phenomenon is that the binocular vision system actively seeks to fixate at scene points that will simplify or enhance stereoscopic perception. The nature of this enhancement is less clear. We suggest that the binocular visual system, unless directed otherwise by a higher level mediator, seeks fixations that simplify the computational process of disparity and depth calculation. In particular, we propose that the binocular vision system seeks to avoid, when possible, regions where the disparity computations are complicated by missing information, such as occlusions, or rapid changes in disparity, which may be harder to resolve.

When a binocular fixation is achieved, those physical points having zero disparity compose the horopter. Theoretically, the horopter consists of the VM circle passing through the two nodal points and the fixation point and a vertical line that passes through the fixation point. The empirical horopter can be measured by psychophysical methods. It turns out that the empirical horizontal horopter lies between the circle and the frontal parallel plane, while the empirical vertical horopter has a backward inclination (Helmholtz shear), with its top away from the subject. Panum's fusion area, the region wherein fusion occurs correctly, is a volume surrounding the empirical horopter. The shape of Panum's fusion area depends greatly on eccentricity, the characteristics of the stimulus, and the surrounding stimuli. When a point is located outside Panum's fusional area, dissimilar images fall on corresponding areas, giving an experience of binocular rivalry.

An interesting aspect is the expansion of Panum's fusional area with eccentricity. It has been reported that

the fusion limit is about $10'$ in the fovea, increasing to about $30'$ at an eccentricity of 6° (Mitchell, 1966; Palmer, 1961). The fovea has a small disparity tolerance, but the periphery has a larger operational range for fusion. This observation might support one of our hypotheses: When viewing an object, observers tend to fixate toward the center of the object, leaving the object boundaries (often associated with depth discontinuities) to peripheral disparity. Such a strategy would allow the fovea to operate under better posed stereo viewing conditions, leaving the periphery to deal with binocular rivalries caused by object boundaries.

The characteristics of the stimulus can also modify the shape of Panum's fusional area. For example, Burt and Julesz (1980) found limits on the fusibility of large disparity gradients. Note that, while the disparity gradient is often defined to be the ratio between the disparity and the relative angular difference between two images, the definition used here is practically the same but easier to understand in the context of luminance gradients. If the disparity gradient between two points becomes large, a "forbidden zone" is created about the fused point within which other image points have a large disparity gradient that cannot be fused. In other words, smooth surfaces are more likely to be fused than depth discontinuities. In light of our results, this supports the suggestion that the human eye–brain stereoscopic system prefers to fixate on smoother regions (having smaller disparity gradient and smaller disparity contrast), thus increasing the incidence of fusible features.

Another explanation of the apparent preference for fixating high luminance variations and low disparity variations might lie in the effect of luminance contrast on stereoacuity. Higher luminance contrast appears to promote stereoacuity. Cormack et al. (1991) used dynamic random dot stereograms and found a cube-root dependency of stereoacuity on luminance contrast above threshold and a square-root dependency for contrasts below about five times the threshold. Rohaly and Wilson (1999) found that perceived depth varies as a power law function of contrast. Cisarik and Harwerth (2008) reported that luminance contrast also has the same affect on disparity perception.

Stereoacuity also increases with interocular correlation and decreases with interocular difference. Smooth surfaces that contain smaller disparity gradients and disparity contrast generally have a large interocular correlation. Depth discontinuities and slants usually alter the two retinal images, thus reducing the interocular correlation. Depth discontinuities also typically accompany half-occlusions, which makes resolving stereo correspondences even more difficult. Harris, McKee, and Smallman (1997) argued that the correlator size at the fovea should be very small (4–5 arcmin in diameter). The occurrence of large disparities at the fovea could thus disrupt mechanisms for fine-scale stereo correspondence within the foveal region.

Another way to interpret our results arises from fMRI studies on the cortical activity of disparity processing. Tootell, Hamilton, Silverman, and Switkes (1988) found

enhanced neural activity as reflected in the uptake of deoxyglucose along cortical regions corresponding to disparity borders between different textures. The enhancement is only observable under binocular conditions. Backus, Fleet, Parker, and Heeger (2001) found that human cortical activity in V1 increased as the disparity increased above the stereoacuity threshold and dropped as the disparity approached the upper depth limit. Georgieva, Peeters, Kolster, Todd, and Orban (2009) showed that activity in occipital cortex and ventral IPS at the edges of the V3A complex was correlated with the amplitude of the disparity variation that subjects perceived in the stimuli. Brouwer, van Ee, and Schwarzbach (2005) presented stimuli with zero disparity, incongruent disparity (disparity cues inconsistent with other cues), and congruent disparity (more similar to natural disparities) to human observers. They reported that stimuli with congruent disparities correspond to the largest responses, while the zero disparity stimuli correspond to the smallest responses in V4d-topo, V7, and many other areas. All of these studies show that the processing of large, complex disparity shapes involves the allocation of more neuronal resources and energy than do smooth disparity fields. Most of the human fMRI data were collected by using simple, artificial stimuli. We believe that fMRI experiments using complex 3D natural scenes would be very fruitful toward understanding the overall computational mechanism of depth perception in the natural world.

Another fact worth mentioning is that the stimuli we used are of outdoor natural scenes containing few objects with regular, smooth geometrical shapes. We suspect that the higher luminance variation and low disparity variation effect at human fixations may be even stronger in indoor scenes, where the luminance content at the center of the objects often tends to contain significant high-level information (for example, characters printed on bottles), which could alter the fixation patterns greatly.

Jansen et al. (2009) showed that human fixations have higher disparity contrast than random locations do, when subjects were presented with white/pink noise stereograms. This would appear to be a result opposite from ours. However, they did not observe any significant difference between the disparity contrast at fixations as compared to other points, using naturalistic stereo images.

Our stimuli are naturalistic stereo images. Given the significant statistical and structural differences between natural images and white/pink noise, we think that the two sets of results are not comparable. Another major difference between the two studies was the method of disparity map computation. Jansen et al. (2009) computed disparity maps using a laser scanner and a geometric transformation, referencing the technical details to an unpublished conference poster. The exemplar disparity map shown in Figure 1 of Jansen et al. (2009) appears to be heavily smoothed compared to its luminance counterpart. Heavy smoothing of the disparity could significantly bias a disparity feature analysis.

There is one additional issue that is relevant to studies of this type. We maintain that binocular viewing conditions are essential to make concrete claims on fixated scene features. Work in this area is limited by the laboratory environment wherein stereo images are displayed on monitors at a constant distance and where the accommodation of the two eyes is also constant. Whereas the 3D viewing conditions that are experienced in daily life are much different, accommodation varies continuously with fixation distance. Objects that are farther or nearer than at the fixations exhibit some degree of blur, depending on the depth of field of the lens. This change of accommodation also alters the image statistics around the fixation and hence alters many functionalities of stereopsis (Hoffman, Girshick, Akeley, & Banks, 2008).

It remains a significant challenge for scientists to measure the real distribution of luminance and disparity features in real-world situations, given current eye tracking and displaying technology. We believe that our current experiments could yield even more interesting results by improving the binocular viewing setup, for example, by deploying a 3D display system with multi-focal planes such as proposed by Akeley, Watt, Girshick, and Banks (2004).

Acknowledgments

This research was supported by the National Science Foundation under Grants ITR-0427372 and IIS-0917175.

Commercial relationships: none.

Corresponding author: Yang Liu.

Email: young76@mail.utexas.edu.

Address: ENS 430, Department of ECE, The University of Texas at Austin, Austin, TX 78712, USA.

References

- Akeley, K., Watt, S., Girshick, A., & Banks, M. (2004). A stereo display prototype with multiple focal distances. *ACM Transactions on Graphics*, 23, 804, 813.
- Anzai, A., Ohzawa, I., & Freeman, R. D. (1999). Neural mechanisms for processing binocular information—II. Complex cells. *Journal of Neurophysiology*, 82, 909–924. [PubMed] [Article]
- Backus, B. T., Fleet, D. J., Parker, A. J., & Heeger, D. J. (2001). Human cortical activity correlates with stereoscopic depth perception. *Journal of Neurophysiology*, 86, 2054–2068. [PubMed]
- Brouwer, G. J., van Ee, R., & Schwarzbach, J. (2005). Activation in visual cortex correlates with the awareness of stereoscopic depth. *Journal of Neuroscience*, 25, 10403–10413. [PubMed] [Article]
- Brown, M., Burschka, D., & Hager, G. (2003). Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25, 993–1008.
- Burt, P., & Julesz, B. (1980). A disparity gradient limit for binocular fusion. *Science*, 208, 615–617. [PubMed]
- Cisarik, P. M., & Harwerth, R. S. (2008). The effects of interocular correlation and contrast on stereoscopic depth magnitude estimation. *Optometry and Vision Science: Official Publication of the American Academy of Optometry*, 85, 164–173. [PubMed]
- Cormack, L. K., Stevenson, S. B., & Schor, C. M. (1991). Interocular correlation, luminance contrast and cyclopean processing. *Vision Research*, 31, 2195–2207. [PubMed]
- Filippini, H. R., & Banks, M. S. (2009). Limits of stereopsis explained by local cross-correlation. *Journal of Vision*, 9(1):8, 1–18, <http://www.journalofvision.org/content/9/1/8>, doi:10.1167/9.1.8. [PubMed] [Article]
- Fleet, D. J., Wagner, H., & Heeger, D. J. (1996). Neural encoding of binocular disparity: Energy models, position shifts and phase shifts. *Vision Research*, 36, 1839–1857. [PubMed]
- Georgieva, S., Peeters, R., Kolster, H., Todd, J. T., & Orban, G. A. (2009). The processing of three-dimensional shape from disparity in the human brain. *Journal of Neuroscience*, 29, 727–742. [PubMed] [Article]
- Girshick, A., Burge, J., Erlikhman, G., & Banks, M. (2008). *Prior expectations in slant perception: Has the visual system internalized natural scene geometry?* Paper presented at the VSS 2008, Naples, FL.
- Harris, J. M., McKee, S. P., & Smallman, H. S. (1997). Fine-scale processing in human binocular stereopsis. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, 14, 1673–1683. [PubMed]
- Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9, 188–194. [PubMed]
- Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(3):33, 1–30, <http://www.journalofvision.org/content/8/3/33>, doi:10.1167/8.3.33. [PubMed] [Article]
- Hoyer, P. O., & Hyvärinen, A. (2000). Independent component analysis applied to feature extraction from colour and stereo images. *Network Computation in Neural Systems*, 11, 191–210. [PubMed]
- Jansen, L., Onat, S., & König, P. (2009). Influence of disparity on fixation and saccades in free viewing of natural scenes. *Journal of Vision*, 9(1):29, 1–19, <http://www.journalofvision.org/content/9/1/29>, doi:10.1167/9.1.29. [PubMed] [Article]

- Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009). *Learning to predict where humans look*. Paper presented at the IEEE International Conference on Computer Vision (ICCV).
- Kayser, C., Nielsen, K., & Logothetis, N. (2006). Fixations in natural scenes: Interaction of image structure and image content. *Vision Research*, 46, 2545, 2535. [PubMed]
- Liu, Y., Bovik, A. C., & Cormack, L. K. (2008). Disparity statistics in natural scenes. *Journal of Vision*, 8(11):19, 1–14, <http://www.journalofvision.org/content/8/11/19>, doi:10.1167/8.11.19. [PubMed] [Article]
- Liu, Y., Cormack, L. K., & Alan C. Bovik. (2007). Disparity statistics at point of gaze in 3D natural scenes [Abstract]. *Journal of Vision*, 7(9):827, 827a, <http://www.journalofvision.org/content/7/9/827>, doi:10.1167/7.9.827.
- Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation sequences made during visual examination of briefly presented 2D images. *Spatial Vision*, 11, 157–178. [PubMed]
- Mitchell, D. E. (1966). A review of the concept of “Panum’s fusional areas”. *Optometry and Vision Science*, 43, 387–401. [PubMed]
- Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, 434, 387–391. [PubMed]
- Ohzawa, I., DeAngelis, G., & Freeman, R. (1990). Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science*, 249, 1037–1041. [PubMed]
- Palmer, D. A. (1961). Measurement of the horizontal extent of Panum’s area by a method of constant stimuli. *Journal of Modern Optics*, 8, 151.
- Parkhurst, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision*, 16, 125–154. [PubMed]
- Parkhurst, D. J., & Niebur, E. (2004). Texture contrast attracts overt visual attention in natural scenes. *European Journal of Neuroscience*, 19, 783–789. [PubMed]
- Qian, N. (1994). Computing stereo disparity and motion with known binocular cell properties. *Neural Computation*, 6, 390–404.
- Raj, R., Geisler, W. S., Frazor, R. A., & Bovik, A. C. (2005). Contrast statistics for foveated visual systems: Fixation selection by minimizing contrast entropy. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, 22, 2039–2049.
- Rajashekar, U., Bovik, A. C., & Cormack, L. K. (2006). Visual search in noise: Revealing the influence of structural cues by gaze-contingent classification image analysis. *Journal of Vision*, 6(4):7, 379–386, <http://www.journalofvision.org/content/6/4/7>, doi:10.1167/6.4.7. [PubMed] [Article]
- Rajashekar, U., van der Linde, I., Bovik, A. C., & Cormack, L. K. (2007). Foveated analysis of image features at fixations. *Vision Research*, 47, 3160–3172. [PubMed]
- Rajashekar, U., van der Linde, I., Bovik, A. C., & Cormack, L. K. (2008). GAFFE: A gaze-attentive fixation finding engine. *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, 17, 564–573.
- Reinagel, P., & Zador, A. (1999). Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems*, 10, 341–350. [PubMed]
- Rohaly, A. M., & Wilson, H. R. (1999). The effects of contrast on perceived depth and depth discrimination. *Vision Research*, 39, 9–18. [PubMed]
- Scharstein, D., & Szeliski, R. (2002). A Taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computerized Vision*, 47, 7–42.
- Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7(14):4, 1–17, <http://www.journalofvision.org/content/7/14/4>, doi:10.1167/7.14.4. [PubMed] [Article]
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, 45, 643–659. [PubMed]
- Tootell, R. B., Hamilton, S. L., Silverman, M. S., & Switkes, E. (1988). Functional anatomy of macaque striate cortex—I. Ocular dominance, binocular interactions, and baseline conditions. *Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 8, 1500–1530. [PubMed] [Article]
- Wick, B. (1985). Forced vergence fixation disparity curves at distance and near in an asymptomatic young adult population. *American Journal of Optometry and Physiological Optics*, 62, 591–599. [PubMed]
- Yabus, A. (1967). *Eye movement and vision*. New York: Plenum Press.