Systems/Circuits

Perceived Depth in Natural Images Reflects Encoding of Low-Level Luminance Statistics

Emily A. Cooper and Anthony M. Norcia
Department of Psychology, Stanford University, Stanford, California 94305

Sighted animals must survive in an environment that is diverse yet highly structured. Neural-coding models predict that the visual system should allocate its computational resources to exploit regularities in the environment, and that this allocation should facilitate perceptual judgments. Here we use three approaches (natural scenes statistical analysis, a reanalysis of single-unit data from alert behaving macaque, and a behavioral experiment in humans) to address the question of how the visual system maximizes behavioral success by taking advantage of low-level regularities in the environment. An analysis of natural scene statistics reveals that the probability distributions for light increments and decrements are biased in a way that could be exploited by the visual system to estimate depth from relative luminance. A reanalysis of neurophysiology data from Samonds et al. (2012) shows that the previously reported joint tuning of V1 cells for relative luminance and binocular disparity is well matched to a predicted distribution of binocular disparities produced by natural scenes. Finally, we show that a percept of added depth can be elicited in images by exaggerating the correlation between luminance and depth. Together, the results from these three approaches provide further evidence that the visual system allocates its processing resources in a way that is driven by the statistics of the natural environment.

Key words: **binocular vision; depth perception; efficient coding; natural scene statistics; optimal coding; primary visual cortex**

Introduction

Natural scenes have statistical regularities that constrain the theoretically infinite space of visual input that strikes the retina. These regularities appear to be exploited by the brain as it encodes this visual input (Simoncelli and Olshausen, 2001; Geisler, 2008). According to two common frameworks for neural coding (efficiency and optimal-inference), sensory input from the environment is encoded according to its probability and relevance. The distribution of environmental probabilities is used to describe a prior assumption made by an optimal or efficient visual system. When this distribution is nonuniform, an efficient encoder up-weights the processing resources allocated for more likely input patterns, and this up-weighting is related to biases and improved performance on perceptual tasks (Brunel and Nadal, 1998; Geisler et al., 2009; Ganguli and Simoncelli, 2010; Girshick et al., 2011).

Seeing the world in three dimensions (3D) requires the visual system to infer an underlying scene layout from flat patterns of light. In making this inference, it has been proposed that the visual system should use prior assumptions consistent with statistical relationships between visual input and depth in nature. For example, color and luminance changes tend to occur along with depth changes, and convex edges tend to belong to near surfaces (Burge et al., 2010; Su et al., 2013). Previous work showed that natural scenes also contain a negative correlation between luminance and depth: darker regions tend to be farther away than brighter regions (Potetz and Lee, 2003, 2006; Samonds et al., 2012). This correlation suggests that luminance could provide the basis for another prior assumption about depth. However, a body of previous work using synthetic stimuli has reported that human depth perception does not reflect a "brighter is nearer" prior assumption (Farnè, 1977; Egusa, 1982; Schwartz and Sperling, 1983; O'Shea et al., 1994).

To investigate the utility of luminance information for depth perception, we first determined whether a biologically plausible computation that links depth and luminance would lead to useful priors. It is well known that precortical stages of visual processing segregate luminance into increments (relatively bright points) and decrements (relatively dark points) via parallel ON/OFF pathways (Werblin and Dowling, 1969; Nelson et al., 1978). We asked whether the correlation between depth and luminance in natural scenes could create differences between the depths encoded in these relative luminance pathways (for bright and dark). We separately measured the distribution of depths produced by the natural environment for light increments and decrements and found differences that are well matched to biases in the joint tuning preferences of cells in primary visual cortex (Samonds et al., 2012), suggesting a potential mechanism for implicitly encoding two separate prior distributions based on whether a point is bright or dark relative to the surroundings. We also conducted a

Received April 2, 2014; revised July 15, 2014; accepted July 21, 2014.

Author contributions: E.A.C. and A.M.N. designed research; E.A.C. performed research; E.A.C. analyzed data; E.A.C. and A.M.N. wrote the paper.

This work was supported by National Institutes of Health Grant 2R01EY018875-04A1 to A.M.N., E.A.C. was supported under a research contract between Sony Corporation and Stanford University. We thank Jason Samonds, Brian Potetz, and Tai Sing Lee for sharing their neurophysiology data, and Johannes Burge for helpful feedback on the manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Emily A. Cooper, Department of Psychology, Stanford University, Jordan Hall (Building 420), 450 Serra Mall, Stanford, CA 94305. E-mail: eacooper@stanford.edu.

DOI:10.1523/JNEUROSCI.1336-14.2014

 0270-6474/14/3411761-08$15.00/0

perceptual experiment using natural scenes, rather than synthetic patterns, as stimuli. We manipulated the luminance patterns of photographs to create versions that were biased either toward the environmental priors ("brighter is nearer") or against them ("brighter is farther"), as a test of whether these priors actually influence depth perception. Observers judged the more prior-consistent images as having more depth, regardless of the original scene content. These results suggest that humans do indeed use prior information about luminance/depth correlations when making depth judgments in complex natural scenes.

Materials and Methods

Natural scene statistics

To analyze the depth distributions of different relative luminance polarities in the natural environment, we selected 31 scenes from a database of coregistered natural image and range measurements (Potetz and Lee, 2003) using three criteria: (1) resolution between 2.4 and 3.3 arcminutes per pixel, (2) rural/natural setting, and (3) minimal missing image pixels. Red, green, and blue channel values (RGB) were transformed to light intensity using a standard conversion: 0.299*R + 0.587*G + 0.114*B. This weighting of the RGB values is consistent with a luminosity function peaking in the mid-range of the visible spectrum. Scenes were segmented into 3° diameter circular patches, overlapping by ½ diameter. Patches with fewer than 95% valid depths were excluded, yielding 13,393 patches and 40,045,314 pixels in all. Pixels were categorized as an increment or a decrement if their intensity was greater than or less than the average within the patch and their Michelson contrast was at least 5%. Using these criteria, 36% of pixels were increments and 46% were decrements.

We wanted to compare the distance distributions of these increment and decrement points to previously reported neuronal tuning properties. This required converting absolute distance from the range data into estimates of binocular disparity. Binocular disparity, the displacement of a point in the two eyes' images, is proportional to that point's depth relative to a reference fixation distance. Because the range data by design did not contain fixation distances, we first selected a small central region of each patch (~21 arcminutes wide) as the simulated fixation distance (z_0). This average central distance value was subtracted from the distance of each other pixel, resulting in a distribution of relative depths that simulated those seen by an observer looking at the patch center. We then converted the depth of each pixel to an estimate of the binocular disparity that would be cast on the retinas by a point at that depth. To do this, an observer was simulated with an interpupillary distance ρ of 0.064 m and a fixation distance of z_0. The approximate binocular disparity δ in radians of a point at distance z_1 was calculated for the distance of each pixel in the patch as follows:

$$\delta \approx \rho\left(\frac{1}{z_0} - \frac{1}{z_1}\right). \quad (1)$$

These disparities were then converted to arcminutes. We used a kernel-smoothing method to compute probability density distributions from these samples using MATLAB (MathWorks). Separate distributions for increments and decrements were computed in terms of relative depth and binocular disparity. We used normal kernel-smoothing windows with bandwidths of 0.09 m for relative depth and 0.03 arcminutes for disparity.

Physiology analysis

Neurophysiology data were reanalyzed from Samonds et al. (2012), who measured the tuning properties of neurons in macaque primary visual cortex for both relative luminance and depth (via binocular disparity); their Figure 3*d* shows histograms of disparity preference grouped by "luminance index" (a normalized ratio of a cell's mean firing rate for light increments vs decrements). From the original sample of 199 cells, we analyzed 189 cells that had a luminance index greater than 0.05 (increment preferred, 27 cells) or less than −0.05 (decrement preferred, 162 cells). The smaller number of cells tuned for increments is consistent with previous measurements (Yeh et al., 2009). For this reanalysis, the proportion of the total number of cells within each disparity bin (ranging between ±60 arcminutes in steps of 12 arcminutes) was calculated separately for increment and decrement populations. Statistical analyses (Wilcoxon rank-sum tests) on these cells tunings and on the natural scene statistics were performed in MATLAB using the ranksum function.

Participants

Twenty adults (age range 18–64 years; 9 females) participated in the main perceptual experiment designed to determine the way in which the human perceptual system is influenced by different patterns of luminance/depth correlation in natural scenes. Twenty additional adults (age range 18–35 years; 15 females) participated in the control experiment. Five additional participants completed the control experiment but were excluded from analysis because debriefing interviews revealed that they had not followed the instructions. Participants were recruited from the surrounding community and screened for normal visual acuity and stereoacuity. All participants were naive to the experimental hypotheses. The study protocol was approved by the Stanford University Institutional Review Board.

Stimuli

Fifteen photographs with registered depth information were selected from public datasets: Live Color+3D (Su et al., 2011; Su et al., 2013) and Middlebury Stereo (Hirchmuller and Scharstein, 2007; Scharstein and Pal, 2007). These smaller datasets were used for the perceptual experiment stimuli because they contain high-resolution, low-noise digital camera images. In the larger dataset used for the natural scene statistics analysis described above (Potetz and Lee, 2003), the use of a photosensor integrated within the laser range scanner resulted in images with an amount of noise that precluded their use as perceptual stimuli.

The scenes included a mixture of interior and exterior settings, as well as natural and man-made content. Each of the 15 photographs is shown in Figure 1*a*, along with a grayscale depth map. These scenes were selected to include a range of luminance/depth correlations from strongly negative (consistent with the "brighter is nearer" prior) to strongly positive (inconsistent with the prior). In Figure 1*b*, circle symbols indicate the original scenes' correlations (Pearson's *r*), sorted from most negative to most positive. The *y*-axis is inverted, so that negative correlations point upward. The selected photographs were each manipulated to create two new versions: one version that was shifted toward the prior (these all had a negative luminance/depth correlation, regardless of the original scene pattern) and one version that was shifted against the prior (these all had a positive luminance/depth correlation). The correlation values for these new versions are indicated by the up and down facing triangles in Figure 1*b*, respectively. In this way, participants could be shown the same scene content conveyed with luminance/depth patterns that either played into or violated the predicted "brighter is nearer" prior assumption.

Image manipulation

To create the new image versions, the original photographs were manipulated in MATLAB using the underlying depth maps as a guide. First, each image was cropped to a square region that contained minimal missing depth values. The depth maps were clipped to the 99th and first percentiles, and the remaining missing values were linearly interpolated. The RGB channels were gamma-corrected with a point-wise nonlinearity ($\gamma = 2.2$) (Stokes et al., 1996). This is an approximation to the inverse of the camera nonlinearities, which were unknown. Images were converted from RGB to hue/saturation/value representation. In this representation, the value channel carries the luminance-related image information. To create the "toward the prior" manipulation, a linear remapping was defined to transform the original value channel $V(x,y)$ to $V'(x,y)$ as follows:

$$V'(x,y) = V(x,y)(1 - \alpha Z'(x,y)) \quad (2)$$

for all rows x and columns y in the channel. A constant, α, was set to 0.75 for this experiment and $Z'(x,y)$, the normalized pixel depth, was defined as follows:

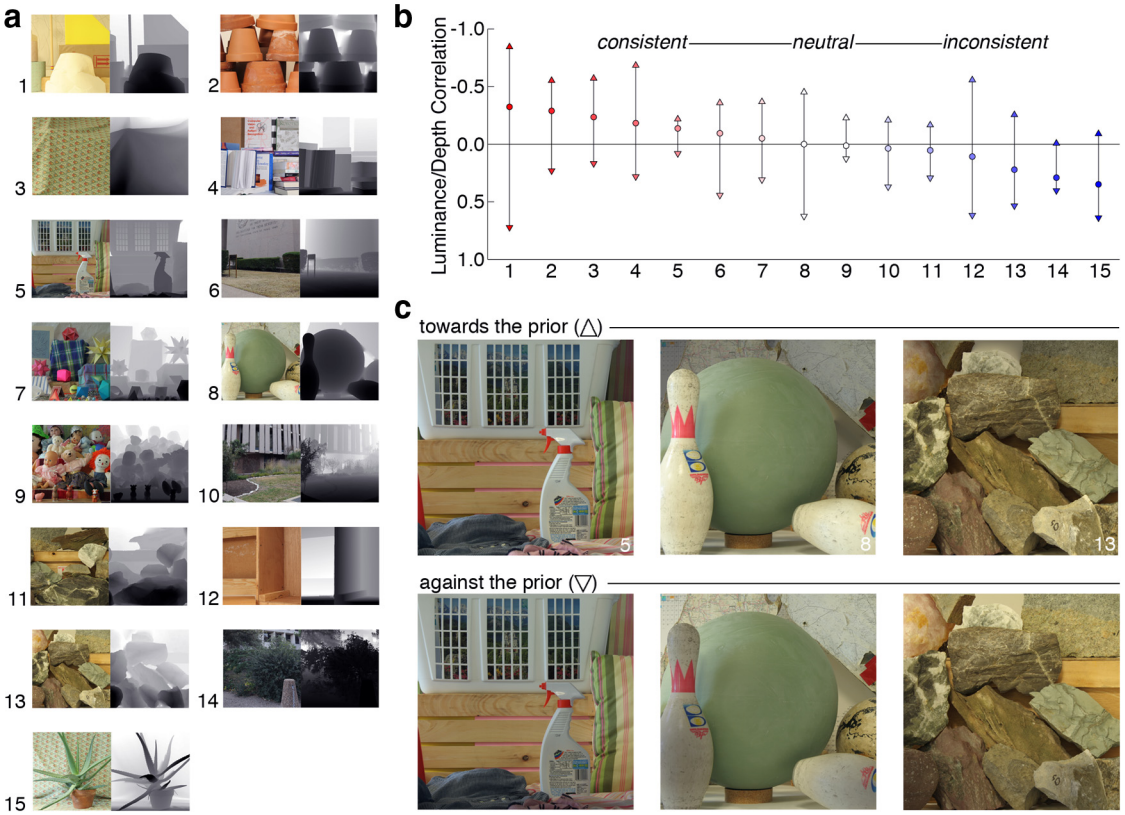

Figure 1. Stimuli for the perceptual experiment. ***a***, Photographs with registered ground-truth depth information. Grayscale depth maps are shown to the left, with darker values indicating near depths. ***b***, Correlations between luminance and depth in the experimental stimuli. Each circle represents one of the 15 scenes used in the perceptual experiment and shown in ***a***. The circle symbols represent the correlation (Pearson's *r*) between luminance and depth in the original scenes. Markers are colored to indicate the relative strength of the correlation: red represents negative and consistent with overall scene statistics; blue represents positive and inconsistent with overall scene statistics; white represents neutral, or near zero correlation. We performed an image manipulation to create two new versions of each scene. Upward and downward triangles represent the new correlations after manipulating the images toward the prior or against the prior, respectively. The *y*-axis is reversed so that negative correlations (consistent with prior from scene statistics) point upward. ***c***, Example photographs after manipulation. One new version was shifted toward the prior; these all have a negative luminance/depth correlation (upward triangle symbols in ***b***). The other new version was shifted against the prior; these all have a positive luminance/depth correlation (downward triangle symbols in ***b***).

$$Z'(x,y) = \begin{cases} 0 & Z(x,y) \le z_{med} \\ \dfrac{Z(x,y) - z_{med}}{z_{max} - z_{med}} & Z(x,y) > z_{med} \end{cases} \quad (3)$$

where z_{med} and z_{max} are the median and maximum depth values. The new values were substituted into the value channel, and the images were converted back to RGB and gamma encoded for display. This resulted in a new image with a smoothly darkening background. To create the "against the prior" manipulation, the depth maps were inverted and the same remapping was performed, so that the new images had a smoothly darkened foreground. Figure 1*c* shows three example scenes in their manipulated versions.

No high-level scene segmentation, lighting, or shape estimation was performed, and image alterations were kept small to avoid appearing artificial. Nonetheless, in dealing with natural scenes, it is always possible that preexisting shape cues could be disturbed by image manipulation. As can be seen in the three examples in Figure 1*c*, however, this manipulation created global differences in the images but also largely preserved important shape-from-shading and shadow cues, such as those seen in the folds in cloth and the overlapping rocks. Possible interactions between the manipulation and the original scene will be discussed in interpreting the experimental results.

Five additional scenes were included in the study but excluded from analysis because this subtle manipulation was insufficient to invert the luminance/depth correlation. Pilot testing suggested that the specific α did not fundamentally alter the perceptual effect, so that a value was chosen that created moderate but noticeable changes across all 15 scenes.

Stereo images

In addition to these image manipulations, each scene also had a stereo depth version. Stereo depth served as a benchmark for the 3D judgments made by the experiment participants. The stereo depth version was created using the original, unmanipulated images and showing slightly different views to the two eyes to create non-zero binocular disparities. For the Middlebury Stereo Dataset, left and right eye images were taken from two adjacent camera views. Live Color+3D did not include multiple views, so for these scenes the two views were synthesized from the single camera view using the Adobe Photoshop (CS3) displacement filter tool. This tool allows a grayscale depth-map image to be used as a displacement map for shifting image pixels to the left or right. In this case, the Live Color+3D depth maps were converted to grayscale images. Larger areas of missing pixels were painted in manually to avoid large distortions. To synthesize the left eye view, near pixels were shifted rightwards and far pixels leftwards. All stereo depth images had a compelling 3D appearance.

Main experiment: task and conditions

Participants were instructed that they would be taking part in a "3D experiment" conducted on a 3D display system. Before starting the experiment, they were shown a stereo depth view of one of the stimuli on the stereoscope apparatus (described below). They were told to focus on the 3D quality of the pictures and try to ignore other changes they may notice.

Stimulus presentation was controlled using MATLAB and the Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli, 1997). Each trial con-

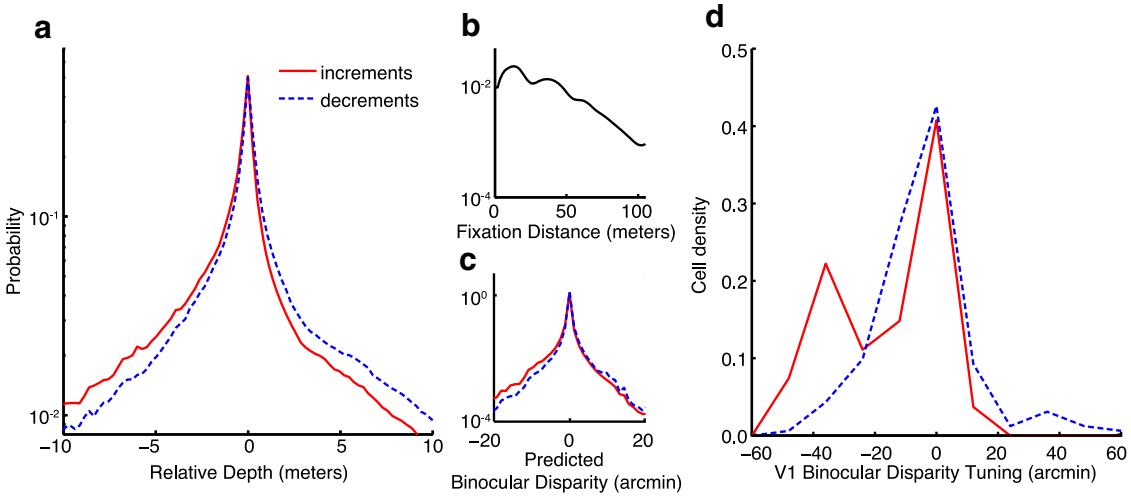

Figure 2. Estimating perceptual priors. ***a***, The probability density of relative depth for luminance increments (solid red) and decrements (dashed blue) within natural scene patches. The relative depth of each pixel is the distance from the mean depth of a small central region, a simulated fixation distance. Negative values are nearer than this distance; positive values are farther. ***b***, Probabilities for the absolute distance of the simulated fixations in the patch dataset. ***c***, Predicted approximate binocular disparities for increments and decrements seen by an observer viewing the natural scene patches and fixating at the central distance. ***d***, Binocular disparity tuning of V1 cells. Data are reanalyzed from Samonds et al. (2012). Cell density is the proportion of cells tuned to values of absolute binocular disparity between ±60 arcminutes. Density was calculated separately for cells preferring increments (solid red) and decrements (dashed blue). The shift in increment preferring cells toward greater cell density at near disparities mirrors the natural scene distribution shown in ***c***.

sisted of a sequential presentation of two different versions of the same scene (2 s each, with 0.5 s in between). After each trial, participants indicated with a key press which version of the image gave them a better sense of the 3D scene. Each scene version (original, stereo depth, toward the prior, and against the prior) was paired with each of the other versions four times, with the order of presentation counterbalanced. Participants only judged differences between two versions of the same scene, never between difference scenes. Trial order was randomized, and each participant made 60 judgments for each condition (15 scenes, 4 repetitions). Across all 20 participants, a total of 1200 3D judgments were collected for each condition. Participants were not instructed as to when a picture did or did not contain binocular disparities (stereo depth). The stereo depth trials were used to ensure that the participants understood the task and to measure their level of performance when a strong depth cue was added. We could then examine whether the image-based statistical manipulation affected 3D judgments on the trials where stereo depth was absent.

Control experiment: task and conditions

Although the images used in the main experiment only had minor photometric differences, we wanted to determine whether perceived 3D might be confounded with perception of low-level image differences. The procedure of the control experiment was the same as the main experiment with a few modifications. Instead of being asked to judge the 3D scene appearance, participants were asked to judge which image had greater contrast (the range between the brightest and darkest image areas). The stereo depth images were excluded from this experiment because they were photometrically identical to the original images. Images were presented as described for the main experiment. Pilot testing revealed that many participants would use the scene depth to judge the scene contrast (e.g., they indicated that they would intentionally select the scene with a darker background as having more contrast), so we displayed the images upside down to encourage participants to focus on contrast alone.

Apparatus

All stimuli were displayed on a stereoscope at a viewing distance of 71 cm. The display system consisted of a Sony PVM-2541 OLED panel and an Apple MacPro (mid-2010). Left and right eye views were separated by a septum. Participants viewed the images through 15 Diopter wedge prisms placed in front of each eye, and the images were shifted 10.7 cm each to the left and right of the screen center, creating a convergence angle consistent with the screen distance. In this setup, each pixel subtended ~1.4 arcminutes. Images were rescaled to subtend 13.2° in the visual field (580 × 580 pixels) and warped to remove keystoning and curvature distortions introduced by the prisms. This resulted in a field of view close to the original camera frustum for the Middlebury images (17° vertically) but smaller than the Live3D+Color images (62° vertically). For all conditions except stereo depth, both eyes were presented with the same image, so that the binocular disparities were zero. In the stereo depth condition, the left and right eye images differed.

Results

To investigate how the negative correlation between luminance and depth in natural scenes may be exploited as a prior assumption by the visual system, we simulated luminance segregation similar to the visual system's ON/OFF pathways and calculated probability distributions for depth separately for locally defined light increments (ON) and decrements (OFF). Figure 2*a* shows these two distributions: increments in solid red and decrements in dashed blue. Depth is defined in relative terms: zero indicates the simulated fixation distance, negative values are nearer than that distance, and positive values are farther. Both distributions peak near zero, indicating that the most likely depths are very close to the fixation distance. This makes sense because neighboring pixels in the scene image tend to belong to physically neighboring surfaces and therefore have similar depths. Both distributions fall off with increasing distance from fixation in both directions (near and far). However, whereas the decrement distribution falls off similarly in both directions, the distribution for increments is substantially asymmetric.

Previous neurophysiological work has identified cells in primary visual cortex (V1) that are jointly tuned for both relative luminance (increments and decrements) and depth (via binocular disparities). This previous work reported systematic biases in the joint tuning properties of these cells but did not directly compare these biases to the statistical distributions of luminance and depth in natural scenes (Samonds et al., 2012). We sought to compare these neurophysiology data with the information avail-

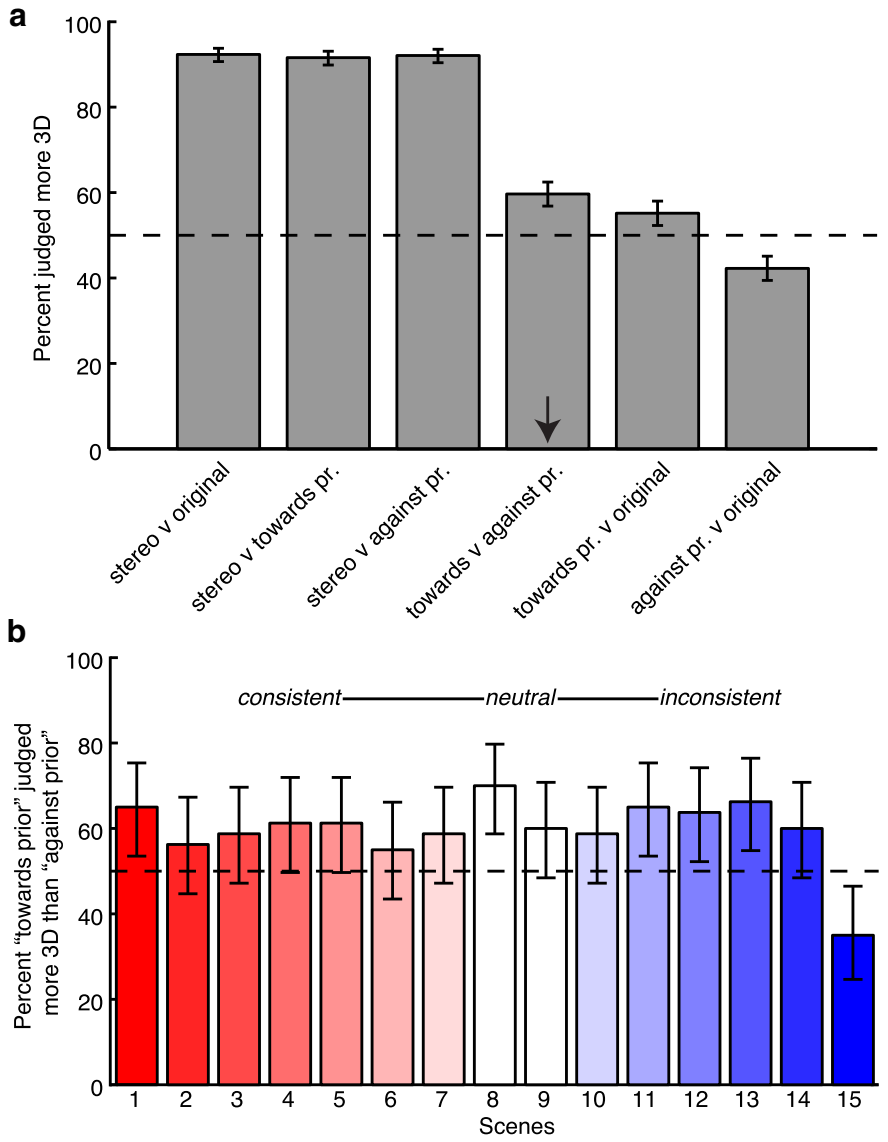

Figure 3. Perceptual experiment results. ***a***, Results for each condition collapsed across all scenes. Bar heights indicate the percentage of all trials in which the first-listed image version was judged as more 3D than the second-listed image version. pr, Prior. Error bars indicate 95% confidence intervals. ***b***, Results for the "toward the prior" versus "against the prior" comparison (arrow in ***a***) separately for each scene. Bar coloring is the same as Figure 1*b*.

able in natural scenes to determine whether these cell tunings indeed reflect an efficient encoding of the available depth information (Ganguli and Simoncelli, 2010). To do this, we converted the relative depths shown in Figure 2*a* to predicted binocular disparities for a simulated observer. The distribution of simulated fixation distances used for this analysis is shown in Figure 2*b*. Because this distribution contains many far distances (>50 m), the predicted binocular disparities tended to be quite small. The predicted binocular disparities are shown in Figure 2*c*. The disparity range shown here (±20 arcminutes) contains >99% of all of the predicted disparities.

Like the distributions for near and far depths, these binocular disparity distributions differ between increments and decrements, with a stronger near-bias for increments. In this simulation, 60% of the increment points belong to near disparities, compared with 49% of the decrement points. Additionally, a one-sided Wilcoxon rank-sum test indicated a relative bias toward near disparities; the increment distribution was shifted significantly lower than the decrement distribution: increment median = −0.09 (n = 14,219,793); decrement median = 0.01 (n = 18,311,324); rank sum = 3.15×10^{14}; z = 667.1; $p < 0.01$. The predicted disparity increment and decrement distributions in Figure 2*c* can now be compared with the distributions of disparity tunings for increment-preferred and decrement-preferred V1 cells shown in Figure 2*d*. The distribution of joint tunings has a pattern similar to the environment: cells that prefer luminance increments (red line) have a more asymmetric distribution, biased more toward near disparities (74% near-preferring) than cells that prefer decrements (blue line; 65% near-preferring). The overall bias toward near disparities across all cell tunings may be due to the fact that large far disparities are physically impossible at fixation distances beyond a few meters (Cooper et al., 2011). Despite this overall bias and a small sample size for increment-preferring cells, a rank-sum test also indicated a significant shift toward near disparities in the increment-preferring cells relative to the decrement-preferring cells: increment median = −0.18 (n = 27); decrement median = −0.06 (n = 162); rank sum = 1.60×10^4; z = 2.5; $p < 0.01$.

Both the natural scene probability distributions and the neuronal tuning distributions could be summarized as reflecting a "brighter is nearer" prior assumption. Using the rank sum statistics, we calculated the probability that an increment disparity value is lower than a decrement disparity value for both the natural scene and cell-tuning distributions (this is the Mann–Whitney U test statistic normalized by the total number of paired observations). These probabilities were 0.57 and 0.64, for the scenes and cells, respectively, indicating a similar level of overlap and overall shift between the natural scene and cell-tuning distributions. This suggests that the V1 cells described here might implicitly encode a prior, based on the environmental distribution, by up-weighting the processing resources for the most likely disparities to be processed in separate ON and OFF pathways.

Given this basis for a visual prior toward near depths for bright points, it is surprising that previous psychophysical studies have failed to clearly show a perceptual bias in this direction (Farnè, 1977; Egusa, 1982; Schwartz and Sperling, 1983; O'Shea et al., 1994). We wondered whether a perceptual bias to see bright points as nearer might be measured when people view natural scenes. The results of our perceptual experiment are shown in Figure 3. The percentage of trials in which one version of an image was judged more 3D than the other is shown in Figure 3*a*. Chance performance (50%) is plotted as a dashed line. As expected, when the stereo depth version was present, it was almost always (92% of trials) judged as more 3D than the other image (first three bars). When neither image contained stereo depth,

participants exhibited a significant preference for the "toward the prior" images, compared with both the "against the prior" images and the originals (60% and 55%, respectively). They also exhibited a preference for the original compared with the "against the prior" (which was only selected on 42% of trials). All of these results were statistically inconsistent with chance performance ($p < 0.05$) as determined by Clopper–Pearson confidence intervals for the binomial distribution.

Recall that some of the original scenes before manipulation were consistent with the predicted priors, and others were inconsistent (Fig. 1*b*). If the participants' behavior reflects implicit knowledge of overall scene statistics, we would predict that the "toward the prior" version should appear more 3D for all scenes, even when the original scene was inconsistent (i.e., scenes in which darker areas tended to be near instead of far). In contrast, if the image manipulation is simply exaggerating preexisting lighting and shape cues, we would predict for there to be an interaction, with "toward the prior" preferred for consistent scenes and "against the prior" preferred for inconsistent scenes. For example, according to this second prediction, a scene with a shadowed or darker foreground (inconsistent) would be preferred if this foreground was further darkened (against the prior). Figure 3*b* shows the results for this comparison for each scene. The "toward the prior" version was preferred for 14 of the 15 scenes. That is, with one exception, scenes appeared more 3D when they were represented as more consistent with overall scene statistics, even when the geometry of the original scene led to the opposite pattern. This suggests that luminance is indeed operating as an independent depth cue, rather than interacting with global scene structure, preexisting lighting, and shape cues.

Although all image manipulations were subtle, it would be less interesting if participants simply chose the scene with greater dynamic range (i.e., more overall perceived contrast between the darkest and lightest areas) as more 3D. We compared the results of this 3D experiment with the control experiment in which participants were asked to select the scene with greater apparent contrast. Across all nonstereo comparisons in the main experiment, there were 38 pairings in which one version was judged as more 3D at a rate of ≥5% above chance. Of these more 3D versions, 92% had a more negative luminance/depth correlation than their comparison, but only 61% were also judged as having more contrast. Using a second metric of image contrast (the variance of the image intensity values) also did not reveal a strong relationship between contrast and perceived 3D: 55% of scenes judged more 3D also had greater intensity variance. Although there was an overall tendency for participants to judge the "toward the prior" scenes as having more contrast than "against the prior" scenes (58% of trials), there was no correlation between the two judgments ($r = 0.01$), suggesting that perceived 3D in these natural scenes cannot be well predicted by their perceived contrast.

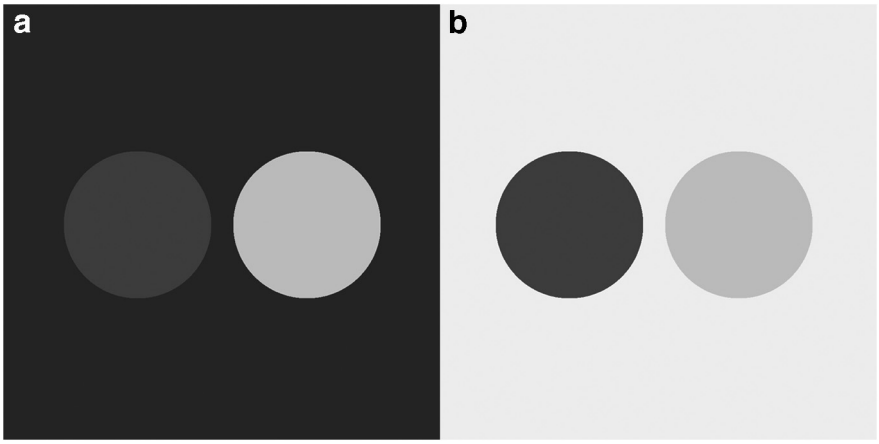

Figure 4. Example of illusions of depth from luminance and contrast. Dark and light gray circular patches are shown against a background that is either darker (***a***) or brighter (***b***) than both patches. Although the common observation is that brighter objects appear near, more recent work has shown that this illusion reverses against a bright background.

Discussion

Natural scene statistics and ON/OFF pathway segregation

The functional significance of separating out ON and OFF visual signals is typically explained as having benefits for encoding small contrast changes over a large dynamic range (Westheimer, 2007). This explanation, however, does not account for well-documented perceptual asymmetries between luminance increments and decrements in 2D spatial perception. For example, human observers have substantially better contrast discrimination thresholds for decrements compared with increments (Blackwell, 1946; Lu and Sperling, 2012). Better discrimination thresholds for decrements have been linked to efficient encoding of natural scenes, which contain more local decrements than increments, via up-weighting of the OFF pathways in early visual processing (Yeh et al., 2009; Ratliff et al., 2010; Baden et al., 2013).

It is often assumed that later-stage visual computations, such as motion and depth estimation, combine the ON and OFF streams and discard polarity information via energy-model type computations (Adelson and Bergen, 1985; Ohzawa et al., 1990; Edwards and Badcock, 1994; Harris and Parker, 1995). To the extent that the statistical distributions of the input to the ON and OFF pathways differ in natural scenes, it would make sense for the visual system to exploit the unique information that each pathways carries, rather than discard it completely. In motion processing, recent work has suggested that ON/OFF segregation is maintained and used to exploit higher-order patterns that are created by motion in natural scenes (Clark et al., 2014). In depth processing, as discussed in the Introduction, neurons involved in stereo-depth computations are jointly tuned for luminance with an overall correlation consistent with natural scenes (Samonds et al., 2012).

We have built on this previous work to specifically characterize how the distributions of depth signals carried in the ON and OFF visual pathways should differ from each other if these pathways evolved to efficiently encode natural scene correlations between luminance and depth. We then showed that these prior distributions are reflected in the tuning of macaque V1 cells. Specifically, we predicted from natural scene statistics that the ON pathway depth distribution should be biased toward near depths, whereas the OFF pathway distribution should be more symmetric. A simple observer model shows a qualitative match between scene disparities and cell disparity tuning. Maintaining the sign of luminance while processing disparity is critical if the visual system is to exploit the natural relationship between depth and luminance. There are thus now two cases, the disparity system as studied here and by Samonds et al. (2012) and the motion

system (Clark et al., 2014), in which the statistics of the natural environment appear to have led to a continuation of at least a partial segregation of the ON and OFF pathways into cortex.

Illusions of depth from brightness

Previous studies investigating the relationship between perceived brightness and perceived depth have largely used simple schematic stimuli and shown that the illusion of depth differences can be induced simply by introducing differences in luminance (Ashley, 1898; Ames, 1925; Taylor and Sumner, 1945; Johns and Sumner, 1948; Coules, 1955; Farnè, 1977; Egusa, 1982, 1983; Schwartz and Sperling, 1983; Dosher et al., 1986; O'Shea et al., 1994). Figure 4*a* shows an example image, for which the brighter patch is typically judged as nearer. With this type of visual stimulus, however, switching the surrounding area to be brighter than the two patches reverses the effect, and the darker patch appears nearer (Fig. 4*b*). This has led to the conclusion that it is contrast, or the amount of dynamic range, rather than the relative luminance (brighter or darker) that creates this illusion of nearness (Farnè, 1977; Egusa, 1982; Schwartz and Sperling, 1983; O'Shea et al., 1994). Other investigations have focused on specific lighting situations, such as diffuse illumination, in which there is a defined relationship between surface shape and the reflected luminance, but which are unlikely to be representative of overall visual experience (Langer and Zucker, 1994; Tyler, 1998; Langer and Bülthoff, 2000).

Our results differed from these previous reports because participants were biased to prefer a "brighter is nearer" scene, rather than a scene with greater perceived dynamic range, across a wide variety of images with natural lighting. Although simple, schematic stimuli can provide insight into the prior assumptions of the visual system, it is interesting to observe that those stimuli also represent atypical visual input. For instance, the images in Figure 4 contain only three different luminance values, two distinct objects, and largely appear quite flat compared with the images in Figure 1*a*, *c*. Perhaps in the case of the simplified illusions, object contrast is the best information for segmenting from the background because relative luminance does not provide enough discriminating information. Typically, it is assumed that the best way to measure a prior assumption is to present people with an impoverished, weak cue situation. We propose that an alternative method for investigating perceptual priors is to create scenes that are typical of the visual input (i.e., natural scenes) but that have properties that have been exaggerated to play into the prior.

It is important to note that abstract visual patterns are often the preferred stimuli in neurophysiology studies as well. Indeed, the luminance preferences of the V1 cell population described in this report were measured using bright and dark circular patches similar to those illustrated in Figure 4 (Samonds et al., 2012). It is difficult to apply our observations from human perceptual experiments to the current neurophysiology data, however, because the perceptual effects necessarily include both low- and high-level processes. The current neurophysiology measurements were restricted to primary visual cortex. It would be interesting to investigate whether the joint luminance/disparity preferences of such cells would be different if natural scene patches were used instead, and whether the joint tuning continues in later visual areas that also have disparity-selective cells.

Applications

Exaggeration of priors also lends itself to computational photography or videography applications in which it is desirable to enhance the 3D appearance of real-world content. This type of enhancement has been applied previously in traditional 3D computer-graphics rendering. The technique, which is called "depth cuing," works by using the underlying 3D model to add a correlation between luminance and depth to the rendered image. This method has been reported to be effective at enhancing the 3D appearance of computer graphics regardless of the sign of the correlation (Schwartz and Sperling, 1983; Foley et al., 1993). The current work builds on the small number of previous studies that have manipulated photographic content with similar methods (Rößing et al., 2012; Easa et al., 2013). These previous studies focused on either abstract photographic images (medical scans) or did not explicitly evaluate 3D appearance in the resulting images. The current results now show that depth cuing of complex photographic content can be effective at enhancing 3D appearance. In particular, a rendering algorithm that selectively darkens image backgrounds should have more enhanced 3D qualities than the reverse. This enhancement would be relatively easy to implement because a single processing algorithm appears to work on wide variety of scenes without manual intervention.

Conclusions

Seeing in 3D is typically understood as relying on a patchwork of canonical depth cues requiring demanding computations (e.g., stereo correspondence, object recognition, shape from shading) (Ramachandran, 1988; Arman and Aggarwal, 1993; Zhang et al., 1999; Riesenhuber and Poggio, 2000; Scharstein and Szeliski, 2002; Nieder, 2003). The availability of databases of natural scene and depth information have made it possible to expand our understanding of the variety of statistical depth information available in natural scenes beyond this classic taxonomy of depth cues (Potetz and Lee, 2003; Burge et al., 2010; Liu et al., 2010; Su et al., 2011, 2013). At the same time, natural scene stimuli open up the possibility for conducting controlled perceptual experiments using more typical visual content.

Here, we have shown that local luminance increments and decrements carry information about depth that could plausibly be extracted by the earliest stages of visual processing. Given the complexity of 3D inference, it would make sense for the visual system to exploit this low-cost information source. We have proposed that this could be accomplished via separate prior assumptions for depth in the ON/OFF visual pathways, and we have demonstrated that there is evidence for the implicit encoding of such priors in the distributions of cell tunings in V1. We also report a perceptual effect potentially caused by these priors: relative depth judgments are biased toward natural scenes that are shifted in the direction of the predicted prior. Future work could investigate how ON/OFF pathway segregation and integration are balanced to best exploit higher-order correlations in natural scenes.

Notes

Supplemental material for this article is available at http://purl.stanford.edu/yg499sy5636. The online Supplemental Material contains high-resolution versions of all images used in the perceptual experiments. Also included is MATLAB code for performing the analyses reported in the article. This includes the natural scene statistics analysis, the image manipulation, and the perceptual experiment analysis (raw response data from both experiments is provided). This material has not been peer reviewed.

References

Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. J Opt Soc Am 2:284–299. CrossRef Medline

Ames A (1925) Depth in pictorial art. Art Bull 8:4–24. CrossRef
Arman F, Aggarwal JK (1993) Model-based object recognition in dense-range images: a review. ACM Comput Surveys 25:5–43. CrossRef
Ashley ML (1898) Concerning the significance of intensity of light in visual estimates of depth. Psychol Rev 5:595–615.
Baden T, Schubert T, Chang L, Wei T, Zaichuk M, Wissinger B, Euler T (2013) A tale of two retinal domains: near-optimal sampling of achromatic contrasts in natural scenes through asymmetric photoreceptor distribution. Neuron 80:1206–1217. CrossRef Medline
Blackwell HR (1946) Contrast thresholds of the human eye. J Opt Soc Am 36:624–643. CrossRef Medline
Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436. CrossRef Medline
Brunel N, Nadal JP (1998) Mutual information, Fisher information and population coding. Neural Comput 10:1731–1757. CrossRef Medline
Burge J, Fowlkes CC, Banks MS (2010) Natural-scene statistics predict how the figure-ground cue of convexity affects human depth perception. J Neurosci 30:7269–7280. CrossRef Medline
Clark DA, Fitzgerald JE, Ales JM, Gohl DM, Silies MA, Norcia AM, Clandinin TR (2014) Flies and humans share a motion estimation strategy that exploits natural scene statistics. Nat Neurosci 17:296–303. CrossRef Medline
Cooper EA, Burge J, Banks MS (2011) The vertical horopter is not adaptable, but it may be adaptive. J Vis 11(3):20 1–19. CrossRef Medline
Coules J (1955) Effect of photometric brightness on judgments of distance. J Exp Psychol 50:19–25. CrossRef Medline
Dosher BA, Sperling G, Wurst SA (1986) Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. Vision Res 26:973–990. CrossRef Medline
Easa HK, Mantiuk RK, Lim IS (2013) Evaluation of monocular depth cues on a high-dynamic-range display for visualisation. ACM Trans Appl Percept 2:1–14.
Edwards M, Badcock DR (1994) Global motion perception: interaction of the ON and OFF pathways. Vision Res 34:2849–2858. CrossRef Medline
Egusa H (1982) Effect of brightness on perceived distance as a figure-ground phenomenon. Perception 11:671–676. CrossRef Medline
Egusa H (1983) Effects of brightness, hue, and saturation on perceived depth between adjacent regions in the visual field. Perception 12:167–175. CrossRef Medline
Farnè M (1977) Brightness as an indicator to distance: relative brightness per se or contrast with the background? Perception 6:287–293. CrossRef Medline
Foley JD, van Dam A, Feiner SK, Hughes JF, Phillips RL (1993) Introduction to computer graphics. Reading, MA: Addison-Wesley.
Ganguli D, Simoncelli EP (2010) Implicit encoding of prior probabilities in optimal neural populations. In: Advances in neural information processing systems (Lafferty JD, Williams CKI, Shawe-Taylor J, Zemel RS, Culotta A, eds), pp 658–666. Cambridge: MIT.
Geisler WS (2008) Visual perception and the statistical properties of natural scenes. Annu Rev Psychol 59:167–192. CrossRef Medline
Geisler WS, Najemnik J, Ing AD (2009) Optimal stimulus encoders for natural tasks. J Vis 9(13):17 1–16. CrossRef Medline
Girshick AR, Landy MS, Simoncelli EP (2011) Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. Nat Neurosci 14:926–932. CrossRef Medline
Harris JM, Parker AJ (1995) Independent neural mechanisms for bright and dark information in binocular stereopsis. Nature 374:808–811. CrossRef Medline
Hirchmuller H, Scharstein D (2007) Evaluation of cost functions for stereo matching. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp 1–8. Minneapolis.
Johns EH, Sumner FC (1948) Relation of the brightness differences of colors to their apparent distances. J Psychol 26:25–29. CrossRef Medline
Langer MS, Bülthoff HH (2000) Depth discrimination from shading under diffuse lighting. Perception 29:649–660. CrossRef Medline
Langer MS, Zucker SW (1994) Shape-from-shading on a cloudy day. J Opt Soc Am 11:467–478. CrossRef
Liu Y, Cormack LK, Bovik AC (2010) Dichotomy between luminance and disparity features at binocular fixations. J Vis 10(12):23 1–17. CrossRef Medline
Lu ZL, Sperling G (2012) Black-white asymmetry in visual perception. J Vis 12(10):8 1–21. CrossRef Medline
Nelson R, Famiglietti EV, Kolb H (1978) Intracellular staining reveals different levels of stratification for on- and off-center ganglion cells in cat retina. J Neurosci 41:472–483. Medline
Nieder A (2003) Stereoscopic vision: solving the correspondence problem. Curr Biol 13:R394–R396. CrossRef Medline
Ohzawa I, DeAngelis GC, Freeman RD (1990) Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. Science 249:1037–1041. CrossRef Medline
O'Shea RP, Blackburn SG, Ono H (1994) Contrast as a depth cue. Vision Res 34:1595–1604. CrossRef Medline
Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437–442. CrossRef Medline
Potetz B, Lee TS (2003) Statistical correlations between two-dimensional images and three-dimensional structures in natural images. J Opt Soc Am A 20:1292–1303. CrossRef Medline
Potetz B, Lee TS (2006) Scaling laws in natural scenes and the inference of 3D shape. In: Advances in neural information processing systems (Scholkopf B, Platt JC, Hoffman T, eds.), pp 1089–1096. Cambridge: MIT.
Ramachandran VS (1988) Perception of shape from shading. Nature 331:163–166. CrossRef Medline
Ratliff CP, Borghuis BG, Kao YH, Sterling P, Balasubramanian V (2010) Retina is structured to process an excess of darkness in natural scenes. Proc Natl Acad Sci U S A 107:17368–17373. CrossRef Medline
Riesenhuber M, Poggio T (2000) Models of object recognition. Nat Neurosci 3:1199–1204. CrossRef Medline
Rößing C, Hanika J, Lensch H (2012) Real-time disparity map-based pictorial depth cue enhancement. Eurographics 31:275–284.
Samonds JM, Potetz BR, Lee TS (2012) Relative luminance and binocular disparity preferences are correlated in macaque primary visual cortex, matching natural scene statistics. Proc Natl Acad Sci U S A 109:6313–6318. CrossRef Medline
Scharstein D, Pal C (2007) Learning conditional random fields for stereo. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp 1–8. Minneapolis.
Scharstein D, Szeliski R (2002) A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int J Computer Vision 47:7–42. CrossRef
Schwartz BJ, Sperling G (1983) Luminance controls the perceived 3-D structure of dynamic 2-D displays. Bull Psychon Soc 21:456–458. CrossRef
Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. Annu Rev Neurosci 24:1193–1216. CrossRef Medline
Stokes M, Anderson M, Chandrasekar S, Motta R (1996) A standard default color space for the internet—sRGB. In: Microsoft and Hewlett-Packard Joint Report.
Su CC, Bovik AC, Cormack LK (2011) Natural scene statistics of color and range. In: IEEE International Conference on Image Processing, pp 261–264.
Su CC, Cormack LK, Bovik AC (2013) Color and depth priors in natural images. IEEE Trans Image Processing 22:2259–2274.
Taylor IL, Sumner FC (1945) Actual brightness and distance of individual colors when their apparent distance is held constant. J Psychol 19:79–85. CrossRef
Tyler CW (1998) Diffuse illumination as a default assumption for shape-from-shading in the absence of shadows. J Imaging Sci Technol 42:319–325.
Werblin FS, Dowling JE (1969) Organization of the retina of the mudpuppy, *Necturus maculosus*: II. Intracellular recording. J Neurophysiol 32:339–355. Medline
Westheimer G (2007) The ON-OFF dichotomy in visual processing: from receptors to perception. Prog Retinal Eye Res 26:636–648. CrossRef
Yeh CI, Xing D, Shapley RM (2009) "Black" responses dominate macaque primary visual cortex V1. J Neurosci 29:11753–11760. CrossRef Medline
Zhang R, Tsai PS, Cryer JE, Shah M (1999) Shape from shading: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 21:690–706. CrossRef