

# Luminance correlations define human sensitivity to contrast resolution in natural images

Yury Petrov

*The Smith-Kettlewell Eye Research Institute, 2318 Fillmore Street, San Francisco, California, 94115*

Received May 25, 2004; revised manuscript received October 8, 2004; accepted October 19, 2004

Luminance patterns encode shape and surface structure of objects in our environment. Humans can detect gradations of 1%–2% of background luminance. Is this level of sensitivity to luminance gradations (contrast) determined by the amount of ecologically meaningful information available in natural scenes? In the first experiment, subjects discriminated natural images  $I$  from their “posterized” versions  $I(n)$ , in which the number of luminance gradations was reduced to  $n$ . In the second experiment, amplified residual images  $I_{\text{res}}(n) \propto I - I(n)$  were discriminated from white-noise images, which lack any luminance correlations and thus information content. Performance in the two experiments matched remarkably well. Furthermore, as a function of  $n$ , the signal detected in both experiments was well fitted by the mutual information between nearby image pixels in the residual image  $I_{\text{res}}(n)$ . This suggests that human sensitivity to luminance contrast is optimized to extract ecologically useful information encoded by the luminance patterns of natural scenes.

© 2005 Optical Society of America

OCIS codes: 330.1800, 330.5510, 100.2810, 100.5010.

## 1. INTRODUCTION

In the course of its evolutionary development, the visual system has been subject to many ecological constraints. Limited brain size and computational rate, along with the need for speedy responses to an ever-changing environment, have constrained the complexity of visual processing. Accordingly, our visual abilities are likely to be tailored to processing only ecologically valuable information contained in natural scenes.<sup>1–3</sup>

Sensitivity to variations of luminance (luminance contrast) is an aspect of visual processing crucial to understanding the shape and surface structure of objects. Historically, contrast sensitivity (in the form of contrast detection and discrimination thresholds) as an aspect of visual processing has been studied by vision scientists. Image information as applicable primarily to natural images has been studied by the image-processing community. Visual processing of natural images became a topic of active research only recently, and many questions regarding the role of natural environment in vision remain open. This paper focuses on the correlation between human sensitivity to contrast resolution (the number of shades of gray) of a natural image and its information content.

The effect of contrast resolution on the informative aspect of an image is illustrated in Fig. 1. On the left are “posterized” images  $I(n)$  of different contrast resolution derived from the original full-resolution image  $I$ . The number of shades of gray  $n$  increases exponentially with pixel depth  $b$  (the number of bits allocated for each pixel):  $n = 2^b$ . The residual images  $I_{\text{res}}(n) \propto I - I(n)$ , shown on the right, retain some coherent features of the original up to a certain pixel-depth threshold  $b_{\text{th}}$ . Therefore information is lost only when an image is sampled at a pixel depth below this threshold value. This value varies from

image to image and depends on the rate at which luminance varies across the image.

If the human visual system was designed optimally, its contrast sensitivity should not exceed contrast difference observed between two consecutive images in the left column of Fig. 1 at  $b_{\text{th}}$  (averaged over a representative set of images, of course). There is no justification for processing images at higher contrast resolution if no meaningful information can be extracted beyond that point. To test this hypothesis, I looked for a relationship between our ability to discriminate between images in the left column of Fig. 1 and to detect the residual coherent signal in its right column, using a large database of calibrated natural images.<sup>4</sup>

What cue is used to discriminate between images in the left column? Assuming that mean luminance for each image is set to a common value, the cue must be some measure of image contrast. The most salient features differentiating images within the column are luminance contours or luminance discontinuities arising as an artifact of posterization. Thus it is likely that some local measure of perceived contrast across the posterization contours is the relevant cue. The precise definition of this measure is not important for the present purpose, as long as this measure changes monotonically with contrast resolution  $n$  (or pixel-depth  $b$ ). One can see that this is indeed the case. Local contrast across posterization contours in Fig. 1 increases as contrast resolution is decreased. Thus, for comparing sensitivity to contrast resolution with sensitivity to the residual coherent signal (which is the purpose here), it is convenient to use pixel-depth  $b$  as the experimental variable that is correlated with both the local contrast signal and the coherence signal.

Previous work relevant to the question posed here is

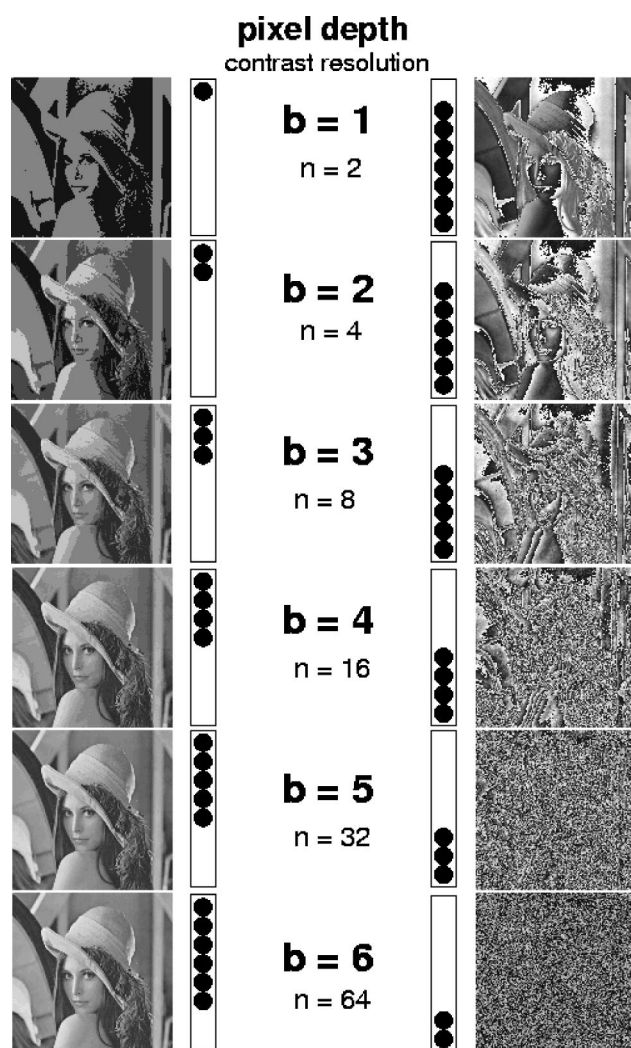


Fig. 1. Contrast resolution and information loss. “Lena” image at different pixel depths  $b$  (contrast resolutions  $n$ ) is shown on the left. The number of gray levels grows as  $2^b$ . Residual difference from the full-resolution ( $b = 8$ ) image is shown on the right (contrast was amplified). Stacked black disks illustrate which bits from the full-resolution image are retained in each case.

surprisingly scarce. It is limited to a few early studies coming from the image-processing community in which the effect of contrast resolution on the informative value of an image (as estimated by human observers) was investigated.<sup>5,6</sup> These studies used limited and rather artificial image sets and did not address the issue of contrast sensitivity. The latter has been studied extensively in the vision science community, but the proposed relationship between contrast sensitivity and information content of natural stimuli has not been investigated, to my knowledge. Early studies used simple stimuli such as luminance gratings<sup>7,8</sup> or bars<sup>9,10</sup> for which the notion of information content was irrelevant. More recent studies in which natural images were used<sup>11–16</sup> were focused largely on the complicated issue of finding the appropriate measure of contrast for such stimuli. As already mentioned, for the present purpose this issue can be obviated by measuring sensitivity to contrast resolution as a means to assess contrast sensitivity for natural images.

## 2. METHODS

Visual stimuli were generated on a PowerMac G4 computer, displayed on a high-resolution monitor (Hitachi Mc7515, RasterOps), and viewed in a dark room at high contrast (the black screen luminance was  $3 \text{ cd/m}^2$ , and the white luminance was  $105 \text{ cd/m}^2$ ). The BITS++ device (Cambridge Research Systems) was used to convert 8-bit red and green input to 13-bit gray-scale output. The monitor linearity and contrast resolution ( $>10$  bits) were confirmed with a Spectra Spotmeter photometer.

Each stimulus comprised two  $256 \times 256$  pixel gray-scale images aligned horizontally on a black background, symmetrically with respect to the center of the screen. Each image subtended  $5^\circ \times 5^\circ$  square region, and the two regions were separated by a  $1.5^\circ$  horizontal gap. The stimuli were obtained from a database of calibrated gray-scale natural images.<sup>4</sup> First, twelve representative images shown in Fig. 2(a) were selected. Each image was  $1536 \times 1024$  pixels in size; a 16-bit number specified luminance for each pixel. Although the pixels in the images are 16 bits deep, the camera digitized the original images as 12 bit. These images were then subdivided into  $256 \times 256$  pixel fragments to form the set of 288 natural images used in the psychophysics tests. One such fragment is outlined by a white square in the first column of Fig. 2(a) and is also shown in Figs. 2(b) and 2(c) to illustrate the experimental stimuli. Each of the 288 fragments was shown once for every contrast resolution tested. Experimental trials were grouped into 12 blocks, each block comprising 24 fragments from a given image in Fig. 2(a) shown at 6 contrast resolution levels. The order of fragments and their contrast resolution levels was chosen randomly within a block.

Subjects viewed the stimuli from a distance of  $\sim 0.6 \text{ m}$ . They responded to the task by pushing the “right” or “left” mouse button. This initiated the next trial, which followed after a 1-s blank interval. Viewing time was unrestricted to ensure that all available signal was assessed equally well in both experiments. Because subjects spent more time on the experiment 1 task than on that of experiment 2 (on average it was 10 and 4 s, respectively), the easiest solution was to let them do the tasks without time restrictions.

## 3. EXPERIMENT

I used a two-alternative forced-choice paradigm (2AFC): two images, test and reference, were displayed simultaneously, and an observer’s task was to indicate the test image. All observers found changes produced by posterization in the test image obvious and the task easy to understand. 2AFC is the optimal paradigm for tasks in which the distinction between test and reference stimulus (i.e., posterization artifacts in Fig. 1) is apparent.

In experiment 1, the full contrast resolution ( $>10$ -bit pixel depth) image  $I$  served as the reference (the video equipment used in the experiment supported 13-bit contrast resolution). The number of shades of gray  $n$  in the test image  $I(n)$  was effectively reduced to  $8 \leq n \leq 256$  by zeroing all but  $3 \leq b \leq 8$  highest-order bits  $b$  [Fig. 2(b)]. Here the bit order ( $b_0 b_1 b_2 b_3 \dots b_n$ ) is defined by



the order of terms in the binary representation of a natural number:  $N = b_0 2^0 + b_1 2^1 + b_2 2^2 + b_3 2^3 + \dots + b_n 2^n$ ; i.e., the least important bit  $b_0$  has the lowest order. Because zeroing the lower-order bits in the test image decreases its luminance (which can be used as a cue), the averaged luminance of the test image was adjusted to match with the reference image for each  $b$ . In this way only contrast cues were left.

One way to look at experiment 1 is to consider the lower-bit image component that was present only in the reference image as a signal. Because the signal was added to the higher-bit posterized component present in both images, the local pattern formed by this component provided pedestal contrast. Thus, in terms of conventional contrast sensitivity paradigms, experiment 1 can be categorized as contrast discrimination.

In experiment 2, the test image  $I_{\text{res}}$  was obtained from  $I$  by removing  $3 \leq b \leq 8$  highest-order bits and consecutively shifting the lower-order bits into their place. Thus the large contrast modulations encoded by the higher-order bits were lost, while fine contrast modulations specified by the lower-order bits were effectively amplified to produce a high-contrast image for each value of  $b$  [see

Fig. 2(c)]. The reference image in this experiment was white noise of the same pixel depth as the test image. Because white noise lacks any correlation between pixels, experiment 2 can be categorized as correlation detection.

Although stimuli in the two experiments were in some sense complementary to each other, the tasks were very different. Experiment 1 tested luminance contrast sensitivity, regardless of whether any information was conveyed by the signal in the lower-order bits. Conversely, experiment 2 tested detection of luminance correlations [perceived as residual image structure seen in Fig. 2(c)], whereas contrast sensitivity was largely immaterial.

#### 4. RESULTS AND DISCUSSION

Three experienced psychophysical observers participated in the study (HV and AMW were naïve to its purpose). In Fig. 3, the proportion of correct responses in experiments 1 (contrast discrimination) and 2 (correlation detection; shown with squares and circles, respectively) was plotted versus the number of higher-order bits  $b$  retained in experiment 1 and removed in experiment 2. In both experiments the test image was detected with near 100%

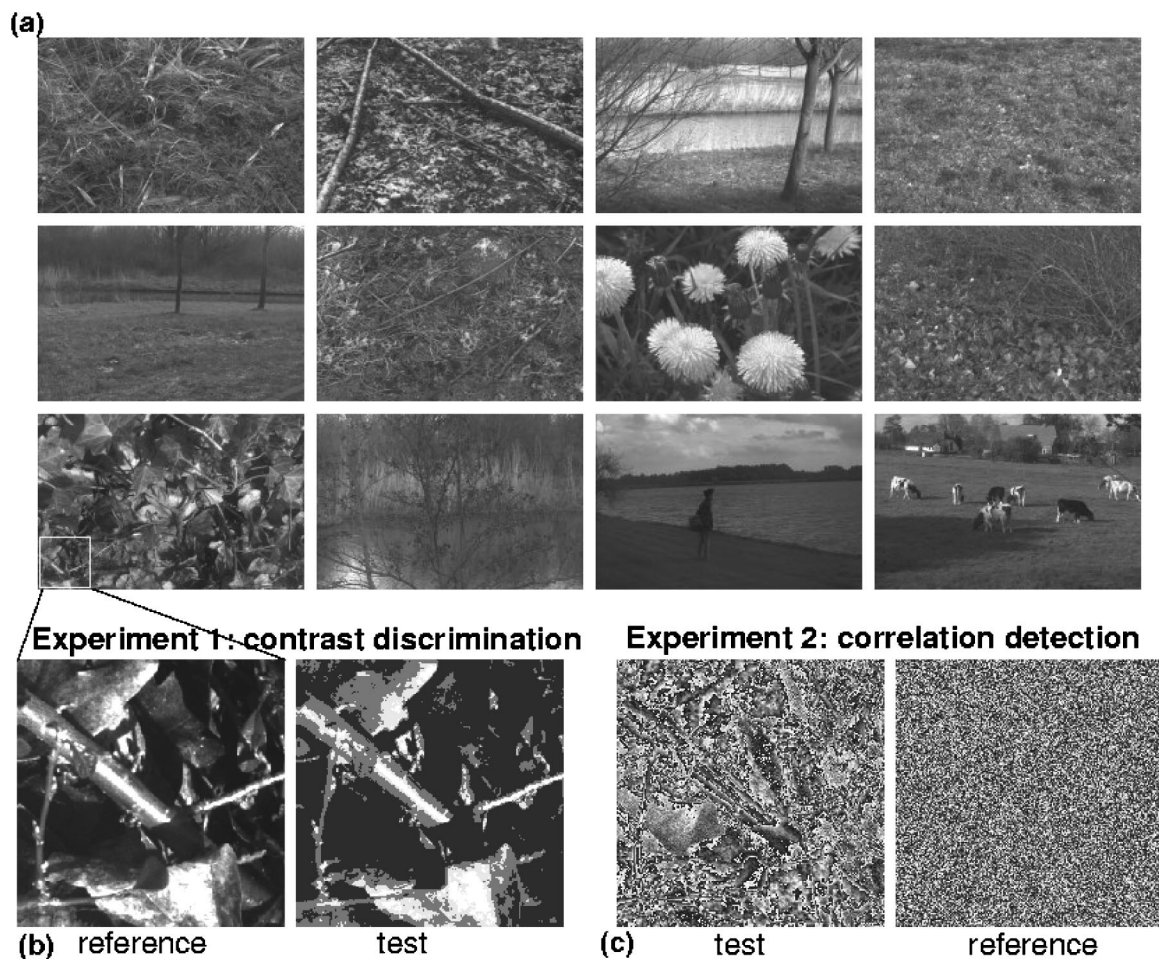


Fig. 2. Experimental stimuli. (a) The set of 12 images used in this study. Each image was subdivided into  $24 \times 256 \times 256$  fragments (illustrated with the white square in the left column), which were used as experimental stimuli. (b) In experiment 1, contrast resolution of a fragment was reduced by zeroing all but  $b$  highest-order bits for each pixel ( $b = 3$  for the case shown here). The resulting image (test) had to be discriminated from the original (reference). (c) In experiment 2, the  $b$  higher-order bits were zeroed and the residual image contrast amplified. The task was to discriminate the resulting image (test) from a white-noise image of the same pixel depth (reference).

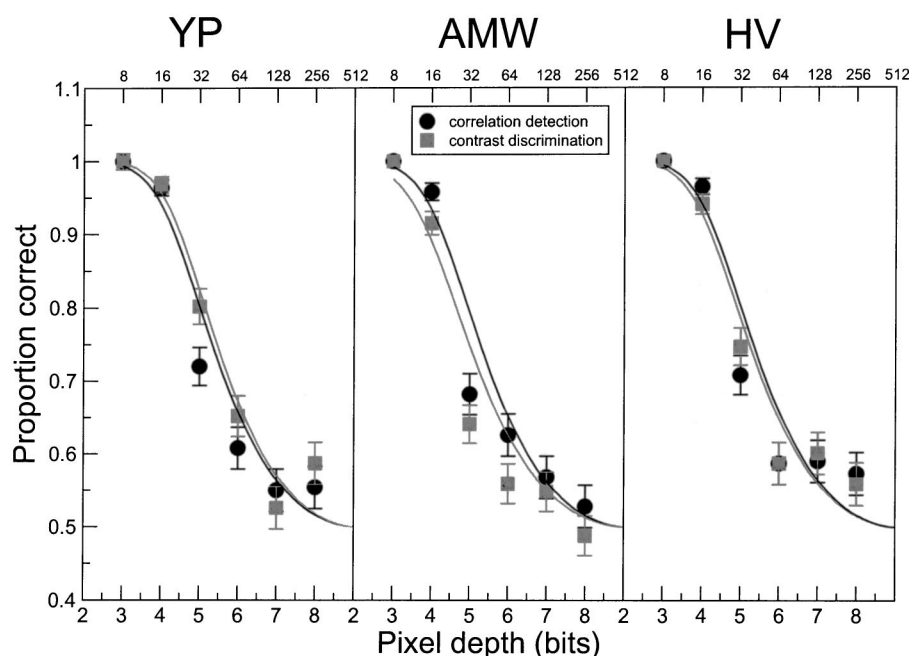


Fig. 3. Performance in contrast discrimination task (experiment 1, squares) and correlation detection task (experiment 2, circles). Pixel depth is plotted along the bottom  $x$  axis, the corresponding number of gray levels along the top  $x$  axis. Each datum shows performance averaged over the 288 image fragments. The error bars give the standard error of the mean computed from the binomial distribution. The fit (solid curves) was computed by using the mutual information measure shown in Fig. 4.

accuracy up to 4 bits/pixel (16 gray levels). Then performance quickly dropped to chance (0.5) by 7 bits/pixel (128 gray levels). Such a precipitous transition indicates that in both experiments, visual signals used to perform the experimental task were sharply reduced to an unperceivable level around 5–6 bits/pixel. The degree of overlap between the two data sets is striking considering the very different nature of the tasks.

The fact that sensitivity to contrast resolution in experiment 1 was so closely matched by sensitivity to residual correlations in experiment 2 is quite remarkable. This supports the hypothesis that human sensitivity to contrast resolution in natural images is tailored to the amount of ecologically important luminance correlations present. Since no correlations could be detected beyond 7 bits/pixel in experiment 2, no more information about objects in natural scenes can be extracted from the low-contrast component beyond this pixel depth. Therefore there is no ecological justification for higher contrast sensitivity in experiment 1.

As already mentioned in Section 1, the measure of contrast appropriately describing contrast perception for natural images is a matter of ongoing research. Measures such as Michelson contrast<sup>17</sup> or Weber fraction definition of contrast used for simple stimuli are not appropriate for natural images.<sup>11</sup> Measures such as the root-mean-square (rms) contrast,<sup>18</sup> although more applicable to natural images, lack local character needed to describe human perception adequately.<sup>19</sup> On the other hand, more elaborate measures (e.g., local band-limited contrast<sup>11</sup>) require a model of how contrast measures calculated for different bands and different image locations that observers attended to (these also need to be known) can be integrated to produce a single perceptual judgment. The strategy adopted here was to obviate that

complicated issue by using image contrast resolution (in the form of pixel-depth  $b$ ) as a contrast measure adequate for the purpose of comparison between sensitivity to contrast resolution and sensitivity to the residual coherent signal. The down side of such an approach is that it does not allow an easy conversion of psychophysical thresholds expressed in terms of pixel depth in experiment 1 to traditional contrast measures.

Another possible issue of concern is luminance of the stimuli used here. Although a high-contrast monitor was used, luminance of the displayed images was still an order of magnitude lower than it would have been for the original outdoor scenes. However, because pupil size adjusts to changes in a scene's luminance, the vast range of illumination in the outside world is drastically scaled down on the retina. Also, responses of mammalian retinal ganglion cells and of neurons in the lateral geniculate nucleus (LGN) depend on stimulus contrast, rather than the absolute levels of retinal illumination.<sup>20–22</sup> A control experiment in which monitor luminance was set to its lowest value did not reveal any significant change in performance, indicating that image luminance was not an important parameter in this study.

In an earlier study<sup>23</sup> we used the same database of natural images to calculate a measure of mutual information  $M$  contained in luminance correlations between nearby pixels (all 4212 images in the database were used). Mutual information is a simple measure of correlation between pixel intensities and as such is the straightforward choice to quantify the degree of image coherence. Two principal inputs to mutual information due to two-pixel and three-pixel correlations were calculated. These terms are shown in Fig. 4 as a function of pixel depth. They increase with pixel depth up to approximately 7 bits/pixel, where they apparently saturate.

Because the saturation threshold is close to the pixel depth where the breakdown of performance in experiment 2 occurred, it is likely that mutual information saturation accounts for this behavior. To pursue this hypothesis, I assumed that the relevant signal in experiment 2 was proportional to the amount of mutual information  $M_{\text{res}}(b) = M(16) - M(b)$  present in the test image for a given pixel depth  $b$ . Then, with the simplest form of signal detection theory, performance in experiment 2 can be fitted with a cumulative Gaussian  $F(d'/\sqrt{2})$ , where detectability  $d' = M_{\text{res}}(b)/\sigma = [M(16) - M(b)]/\sigma$ , and  $\sqrt{2}$  appears as a result of using the 2AFC paradigm. The only free parameter is  $\sigma$ , which describes a subject's precision in assessing mutual information for the given task. The total mutual information  $M(b)$  was approximated by the two-pixel and three-pixel terms  $M_2 + M_3$  shown in Fig. 4. Although higher-order (i.e., four-pixel, five-pixel, etc.) terms were not included, it is unlikely that they would have made a significant contribution.<sup>23</sup> The  $M_2 + M_3$  saturation value of 2.69 bits was used for the mutual information value at the 16-bit pixel depth,  $M(16)$ .

Figure 3 shows good agreement between the calculated (black curves) and experimental results. Note that such a close fit can not be explained by using a free parameter, because the fitted psychometric data is characterized by at least two parameters: threshold and steepness. Steepness describes the speed with which performance drops from near optimal (100%) to near chance (50%). It can be shown that goodness of fit depends critically on the shape of the mutual information curves in Fig. 4. The sigmoid shape of these functions is transformed into the large steepness of the resulting psychometric curves. The close fit confirms that mutual information provides a good approximation of the correlation measure used by human observers. Because the results of experiment 1 and 2 are so close, experiment 1 performance was fitted with the same model for illustrative purpose (gray curves). The fitted values of  $\sigma$  vary between 0.66 and 0.79 bits and are not significantly different between experiment 1 and experiment 2 ( $\sigma$  confidence intervals were

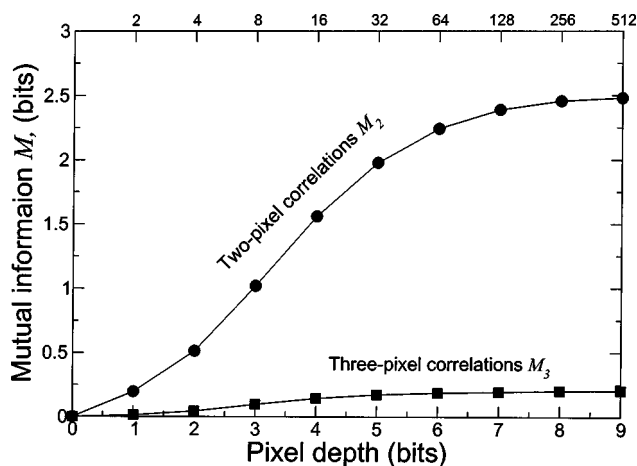


Fig. 4. Mutual information between nearby image pixels as a function of pixel depth. Mutual information terms due to correlation between two pixels  $M_2$  (circles) and three-pixels  $M_3$  (squares, excludes  $M_0$  terms) provide principal inputs to the total mutual information<sup>23</sup>; combined, they can be used as a measure of luminance correlations between nearby pixels.

computed by a bootstrap procedure, assuming binomial distribution for the measurement mean for each datum). This is not surprising, given that fitted data for the two experiments are almost identical.

There is another possible explanation for the near identical performance in experiments 1 and 2. Performance in experiment 1 dropped to chance at the same pixel depth as the signal in the lower-order bits (apparently) became noise. What if contrast sensitivity for the noisy lower-order image component were lower than it would have been for a coherent component? Then the degree of the residual image coherence would control performance in both experiments. Performance in experiment 1 would simply follow that of experiment 2 as the residual image became more and more noisy. To check for this possibility, experiment 1 was repeated with  $16 - b$  lower-order bits in the reference image replaced by white noise. If the above assumption were correct, the sensitivity threshold in the control experiment would have been at a lower pixel depth, because the  $16 - b$  lower-order bits would have then been encoding random noise for any  $b$ . However, performance did not change significantly, ruling out the above explanation.

It is important to note that noise in the lower-order bits is not an artifact of image acquisition and processing but rather a “natural” phenomenon. This has been shown for the image database used here though analysis of mutual information calculated for block-averaged images.<sup>23</sup> Also, clear-sky (near-uniform gray) fragments of natural images taken from the database retain a high degree of correlation up to the lowest bit. In fact, the (apparent) lower-order bit noise is not an exclusive characteristic of natural images but is also an element of artificial images, which are free of noise by default. It is easy to see why any image (of a certain minimal complexity) will appear noisy starting from a certain pixel depth. Consider the very lowest bit  $b_0$  for a given pixel in our database. It specifies whether the 16-bit natural number representing the pixel's luminance is even or odd. Unless luminance relief in the vicinity of this pixel is extremely flat, the number differs greatly even between the nearby pixels, with the effect of it's being even or odd with (almost) equal probability for a given pixel without any apparent (to human observers) correlation with the underlying image structure. The same argument applies to all lower-order bits that encode luminance gradations much smaller than typical local luminance variations in a natural image.

## 5. CONCLUSION

To summarize, this study showed that for natural images, contrast discrimination assessed as sensitivity to image contrast resolution closely follows luminance correlation detection. The results are best explained if ecologically relevant visual information determines human luminance contrast discrimination. This would follow the general strategy of computing cost minimization adopted by the visual system. After all, contrast discrimination serves to detect and measure luminance correlations that encode meaningful visual information in natural scenes. By means of psychophysics and statistical analysis this work provides a quantitative test of this strategy.



## ACKNOWLEDGMENTS

I thank Li Zhaoping, who helped to conceive the experiments and contributed significantly to the manuscript, and Andrew Glennerster for helpful discussions. This work was supported by the Wellcome Trust (grant 056657) and by Gatsby Charitable Foundation.

Correspondence and requests for materials should be addressed to Yury Petrov (e-mail: yury@ski.org).

## REFERENCES

1. H. B. Barlow, "The Ferrier Lecture 1980, 'Critical limiting factors in the design of the eye and visual cortex'," *Proc. R. Soc. London Ser. B* **212**, 1–34 (1981).
2. E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001).
3. N. J. Dominy and P. W. Lucas, "Ecological importance of trichromatic vision to primates," *Nature* **410**, 363–366 (2001).
4. J. H. Van Hateren and A. van der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," *Proc. R. Soc. London Ser. B* **265**, 359–366 (1998).
5. J. C. Bartleson and R. F. Witzel, "Source coding of image information," *Photograph. Sci. Eng.* **11**, 263–265 (1967).
6. J. V. Gaven and J. Tavitian, "The informative value of sampled images as a function of the number of gray levels used in encoding the images," *Photograph. Sci. Eng.* **14**, 16–21 (1970).
7. F. W. Campbell and D. G. Green, "Optical and retinal factors affecting visual resolution," *J. Physiol. (London)* **181**, 576–593 (1965).
8. F. W. Campbell and J. G. Robson, "Application of Fourier analysis to the visibility of gratings," *J. Physiol. (London)* **197**, 551–566 (1968).
9. G. E. Legge and D. Kersten, "Light and dark bars; contrast discrimination," *Vision Res.* **23**, 473–483 (1983).
10. P. Wittle, "Increments and decrements: luminance discrimination," *Vision Res.* **26**, 1677–1691 (1986).
11. E. Peli, "Contrast in complex images," *J. Opt. Soc. Am. A* **7**, 2032–2040 (1990).
12. Y. Tadmor and D. J. Tolhurst, "Discrimination of changes in the second-order statistics of natural and synthetic images," *Vision Res.* **34**, 541–554 (1994).
13. D. J. Tolhurst and Y. Tadmor, "Discrimination of changes in the slopes of the amplitude spectra of natural images: band-limited contrast and psychometric functions," *Perception* **26**, 1011–1125 (1997).
14. D. J. Tolhurst and Y. Tadmor, "Band-limited contrast in natural images explains the detectability of changes in the amplitude spectra," *Vision Res.* **37**, 3203–3215 (1997).
15. D. J. Tolhurst and Y. Tadmor, "Discrimination of spectrally blended natural images: optimisation of the human visual system for encoding natural images," *Perception* **29**, 1087–1100 (2000).
16. P. J. Bex and W. Makous, "Spatial frequency, phase, and contrast of natural images," *J. Opt. Soc. Am. A* **19**, 1096–1106 (2002).
17. A. A. Michelson, *Studies in Optics* (University of Chicago Press, Chicago, Ill., 1927).
18. G. S. Rubin and K. Siegel, "Recognition of low-pass filtered faces and letters," *Invest. Ophthalmol. Visual Sci. Suppl.* **25**, 71 (1984).
19. D. S. Loshin and T. A. Banton, "Local contrast requirements for facial recognition in patients with central field defects," *Invest. Ophthalmol. Visual Sci. Suppl.* **29**, 43 (1988).
20. H. B. Barlow and W. R. Levick, "Three factors limiting the reliable detection of light by retinal ganglion cells of a cat," *J. Physiol. (London)* **200**, 1–24 (1969).
21. K. Purpura, E. Kaplan, and R. M. Shapley, "Background light and the contrast gain of primate P and M retinal ganglion cells," *Proc. Natl. Acad. Sci. USA* **85**, 4534–4537 (1988).
22. G. H. Jacobs, "Responses of the lateral geniculate nucleus to light increment and decrement and the encoding of brightness," *Vision Res.* **6**, 83–87 (1966).
23. Y. Petrov and L. Zhaoping, "Local correlations, information redundancy, and sufficient pixel depth in natural images," *J. Opt. Soc. Am. A* **20**, 56–66 (2003).