Natural Systems Analysis

Wilson S. Geisler, Jeffrey S. Perry, and Almon D. Ing
Center for Perceptual Systems and Department of Psychology
University of Texas at Austin, Austin, TX 78712, USA

ABSTRACT

The environments we live in and the tasks we perform in those environments have shaped the design of our visual systems through evolution and experience. This is an obvious statement, but it implies three fundamental components of research we must have if we are going to gain a deep understanding of biological vision systems: (a) a rigorous science devoted to understanding natural environments and tasks, (b) mathematical and computational analysis of how to use such knowledge of the environment to perform natural tasks, and (c) experiments that allow rigorous measurement of behavioral and neural responses, either in natural tasks or in artificial tasks that capture the essence of natural tasks. This approach is illustrated with two example studies that combine measurements of natural scene statistics, derivation of Bayesian ideal observers that exploit those statistics, and psychophysical experiments that compare human and ideal performance in naturalistic tasks.

Keywords: natural scene statistics, contour grouping, region grouping, Bayesian ideal observer

1. Introduction

A fundamental principle of biology is that sensory and perceptual systems (including their associated mechanisms for learning and plasticity) evolve in the service of obtaining information about the environment that is relevant for the tasks the organism must perform in order to survive and reproduce. Thus, the design of a perceptual system is strongly influenced by the tasks the organism performs and by the physical/statistical properties of the environment, in addition to various biological constraints (e.g., limited numbers and dynamic ranges of neurons). This observation suggests that there may be great payoff in making the effort to study perceptual systems rigorously from the perspective of natural tasks and stimuli. Intuition suggests that such a "natural systems analysis" should have several parts: (a) identifying and characterizing natural tasks, (b) measuring and analyzing the environmental properties (e.g., natural scene statistics) relevant to those tasks, (c) determining how a rational (ideal) perceptual system would exploit those environmental properties to perform natural tasks, and (d) using the knowledge gained from (a)-(c) to formulate hypotheses for neural mechanisms and test them in physiological and behavioral studies that capture the essence of the natural task.

Ethology and behavioral ecology are concerned with identifying and characterizing natural tasks. Unfortunately, there have been few efforts to rigorously identify and characterize natural tasks in humans, although there have been some relevant studies in non-human primates (e.g., [1]). On the other hand, there has been a growing effort in recent years to measure and analyze natural scene statistics (e.g., for reviews see [2] and [3]). Furthermore, the tools for determining how a rational system would exploit natural scene statistics to perform complex tasks have been improving because of progress in the theory and application of Bayesian ideal observers [4-6] and because of the increases in computer performance. Finally, because of rapid advances in technology it is now possible to carry out rigorous experiments in fairly complex tasks. For example, it is now possible to precisely measure the positions of the eyes, limbs and body in natural tasks, to measure the activity of single neurons or of whole populations of neurons in the brain while an animal is performing natural tasks, and to create virtual environments containing visual, tactile, and auditory objects that can be precisely controlled and manipulated during the performance of natural tasks. Thus, most of the tools necessary for performing a rigorous natural systems analysis are now available.

This paper summarizes to two recent examples of this approach from our lab. The first example concerns the task of deciding whether particular pairs of contours passing under an occluding surface belong to the same or different physical contours in the environment. The second example concerns the task of identifying whether particular pairs of image patches belong to the same or different physical surfaces. In each example, we first define the natural task. Next, we measure the relevant natural scene statistics. Then, we derive the Bayesian ideal observer for that natural task based

Human Vision and Electronic Imaging XIII, edited by Bernice E. Rogowitz, Thrasyvoulos N. Pappas,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6806, 68060M, © 2008 SPIE-IS&T · 0277-786X/08/$18

on the measured natural scene statistics. Finally, we measure human performance in the natural task for stimuli based on the natural scene statistics and compare human performance to that of the ideal observer.

2. Example 1: Contour occlusion task

It is common in natural scenes for an object to be partially occluded by one or more other objects (Fig. 1). Such occlusions can provide useful depth and segmentation (figure-ground) information; for example, if the bounding contour of an object can be identified, then other contours intersecting that bounding contour are likely to be occluded, and hence likely to be at a greater distance and to derive from a different physical source than the bounding contour (e.g., a different object). However, the existence of occlusions can also greatly increase the difficulty of correctly interpreting natural images; for example, an occluding object necessarily obscures image features from the occluded objects, making identification of the occluded objects difficult.

The human visual system contains powerful contour grouping mechanisms that are thought to play an important role in helping the visual system both exploit occlusions and overcome the loss of features produced by occlusions. For example, contour grouping mechanisms allow us decide (correctly) that the two contours passing under the red leaf in Fig. 1 arise the same physical source (surface boundary). These contour grouping mechanisms undoubtedly evolved and/or develop in response to the properties of natural environments, and thus there have been recent efforts to directly measure the statistical properties of contours in natural images, with the aim of gaining a deeper understanding of the image information available to support contour grouping and of developing more refined models of contour grouping [7-9].

Fig. 1. Contour occlusion in natural images.

2.1. Contour statistics

Much of the procedure for measuring contour statistics is described elsewhere [7]. Briefly, we analyzed a set of natural images that were picked to be as diverse as possible, without containing human-made objects or structures. The images included close-up and distant views of different environments (i.e., forests, mountains, deserts, plains, seashore) and image constituents (e.g., water, sky, snow, plants, trees and rocks). Edge elements were extracted from each image using an automatic algorithm containing the following steps: (a) convert the image to gray scale, (b) filter the gray scale image with a non-oriented log Gabor filter (in the Fourier domain) having a spatial-frequency bandwidth of 1.5 octaves

and a peak spatial frequency of 0.1 c/pixel (the frequency was picked to provide a dense sampling of contours), (c) identify the locations of zero crossings in the filtered image, (d) at each zero crossing point in the (unfiltered) gray scale image apply a bank of odd and even log Gabor filters with a spatial-frequency bandwidth of 1.5 octaves and an orientation bandwidth of 40 deg, (e) normalize the filter responses by dividing by the sum of the responses across all orientations, (f) combine the odd and even responses to obtain an energy response, (g) find the peak of the energy response across orientation to determine the local contour orientation, (h) eliminate edge elements with peak normalized energy responses that do not exceed a low threshold, (i) interpolate the even and odd responses to better localize the edge position, (j) reapply an odd log Gabor filter at the estimated edge-element orientation and position in the gray scale image to determine the contrast polarity (the sign of the contrast) of the edge element. The last two steps were not applied in the original study [7]. The above edge extraction procedure was applied to synthetic test images with known contour positions, orientations and contrast polarities, and was found to be accurate for the test images. We note that the extraction of contrast polarity information is new to the current study. We chose not to examine contrast magnitude because the images were neither luminance nor color calibrated, and thus for these images we can only be confident about measurements of edge geometry and contrast polarity.

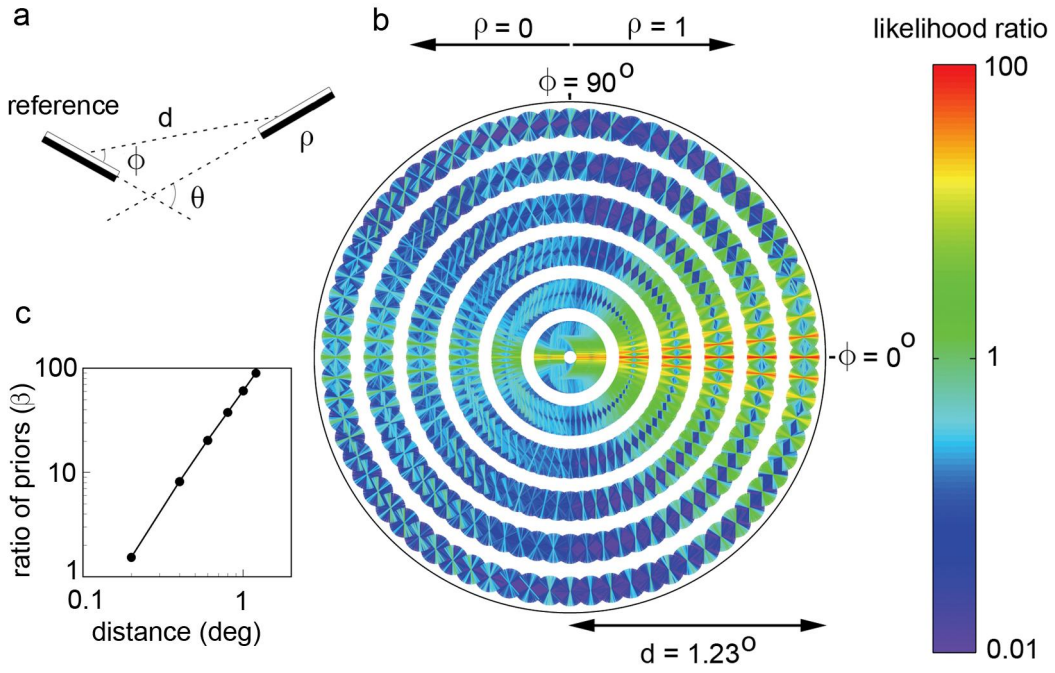

Fig. 2. Co-occurrence statistics of contour elements in natural images. **a**. Definition of parameters describing the geometrical and contrast relationship between a pair of contour elements. **b**. Plot of the likelihood ratio for a given relationship between pairs of contour elements. **c**. Ratio of the prior probabilities that pair of contour elements belong to different versus the same physical source, as a function of distance between the pair of elements. A likelihood ratio greater then 1.0 means (given equal priors) it is more likely that the elements belong to the same physical contour; a ratio less than 1.0 means it is more likely that the elements belong to different physical contours. [For each distance, direction and polarity, the orientation difference bins (line segments) are drawn in rank order starting from the lowest likelihood; thus, the highest likelihoods are the most visible in the plot.]

We measured pair-wise statistics. Specifically, for each pairing of extracted edge elements we considered one of them as the reference and described the geometrical and contrast relationship of the other element relative to the reference element (every edge element served as a reference element). The relationship between the elements is described by four parameters (see Fig. 2a): the distance between the centers of the edge elements (d), the direction of the second element from the reference element (ϕ), the orientation difference between the edge elements (θ), and difference between the edge elements in contrast polarity (ρ). (Note that ρ takes on only two possible values: 1 = same polarity and 0 = opposite polarity.) Thus, the pair-wise statistics can be described by a four dimensional probability density function,

$p(d,\phi,\theta,\rho)$. This function was estimated by binning the edge element pairs along the four dimensions (6 distances x 36 directions x 36 orientation-differences x 2 contrast polarities for a total of 15552 bins).

To obtain approximate ground truth we had two observers segment the extracted edge elements into groups that belong to the same physical contour. The observers viewed the images with each extracted edge element labeled by a single red pixel. They then selected those pixels that belong to the same contour. To aid them in the segmentation they were allowed to zoom in and out, toggle to the full color image, and toggle the colored pixels on and off. The fundamental premise is that most of the segmentations correspond to the physical ground truth (i.e., the grouped pixels do indeed arise from a common physical source—a surface/material boundary, a shadow/shading boundary, or surface marking boundary). The observers noted that some ambiguous cases arose, but that they were highly confident about most of the pixels assignments, which was supported by the high inter-observer agreement [7]. Given the segmentation data, it is then possible to estimate (by binning the edge element pairs) the probability distribution $p(c,d,\phi,\theta,\rho)$, where c takes on two possible values: $c=1$ if the edge elements belong to the same contour and $c=0$ if they belong to different contours.

The measured contour statistics are shown in Fig. 2b and 2c. In order to be consistent with the new optimal decision rule described below, the plotting conventions here are somewhat different from those in our earlier publication [7]. Specifically, Fig. 2b plots the distance-dependent likelihood ratio $l(\phi,\theta,\rho|d)$, and Fig. 2c plots the distance-dependent decision criterion (ratio of priors) $\beta(d)$. As can be seen, for any given distance the most likely geometrical relationship between edge elements is one consistent with approximate co-circularity (although there are some systematic deviations from co-circularity). This is true whether the polarity is the same or opposite; however the likelihoods (for co-circular geometrical relationships) are lower for opposite polarity edge elements. Perhaps not surprisingly, the ratio of the priors increases approximately in proportion to the square of the distance (Fig. 5c).

2.2. Contour occlusion experiment

In the contour occlusion experiment, we measured subjects' ability to identity whether a pair of edge elements passing under an occluder belonged to the same or different physical contours. On half the trials the edge elements belonged to the same physical contour and on half the trials to different contours. On each trial, a pair of edge elements was extracted directly from our database of natural images via the following procedure. First an image was randomly selected and then a single edge element was randomly selected from that image (e.g., one of the green highlighted edge elements in Fig. 3a). Second, we considered the set of all edge elements at a given distance from the selected edge element, where the distance equaled the occluder diameter for that trial. On "different" trials an edge element was randomly selected from those that belonged to a physical contour different from the contour containing the initially selected element. On "same" trials an element was randomly selected from those elements (usually just one or two elements) that belonged to the same physical contour as the initial selected element (e.g., in Fig. 3a, the second green highlighted element belongs to the same physical contour).

Once the edge elements were selected, they were displayed to the subject as shown in Fig. 3b. Specifically, the 768 x 768 pixel display subtended 16 deg in visual angle with a gray background luminance of 55 cd/m^2 and a gray, circular occluder of 60 cd/m^2. As in Fig. 3b, the 3.3 c/deg edge elements were always located on opposite ends of a line through the center of the occluder. The size of the edge elements in the display was the same as the size of the oriented filter kernels used to extract the edge elements when the pair-wise statistics were measured. The center pixel of the edge element sat on the occluder boundary. The contrast of the edge elements was set so that their locations and orientations were clearly visible (Michelson contrast = 0.6). On each trial, the display remained up until the subject responded, and the subject was free to make eye movements.

Performance was measured, in counterbalanced blocks, for occluder diameters of 20, 40, and 90 pixels (0.67 deg, 1.33 deg and 3 deg). (We note that it was not possible to measure performance for occluder diameters greater that 90 pixels because there are too few contours of sufficient length in the database.) In addition, performance was measured with and without the contrast polarity information; in the blocks where the polarity information was absent we replaced the odd symmetric edge elements with even symmetric elements. Finally, we also manipulated feedback. For the first 600

trials of the experiment no feedback was provided; for the second 600 trials feedback was provided on each trial (a tone indicated whether the response was correct or incorrect); for the third 600 trials no feedback was provided. We chose to manipulate feedback because of concern that our occlusion displays were so simplified (unnatural) that subjects may not be able to use the edge geometry and contrast information in the way that they would normally use that information in natural images. We reasoned that if the subjects show no improvement with feedback then it strongly suggests that they are able to apply their normal contour processing mechanisms to our simplified displays.

Seven subjects participated in the study. Two were familiar with the aims of the study and five were naïve. On the first day (prior to the main experiment), each subject completed a small block of trials designed to help them understand the task and display. Specifically, after each stimulus presentation and response (which was the same as in the main experiment), the subjects were shown a display like the one in Fig. 3a, which illustrated exactly how the simplified display was obtained from the original image. For the remainder of the study, the subjects only saw displays like the one in Fig. 3b. The study extended over a period of 3 to 6 days, depending on the subject.

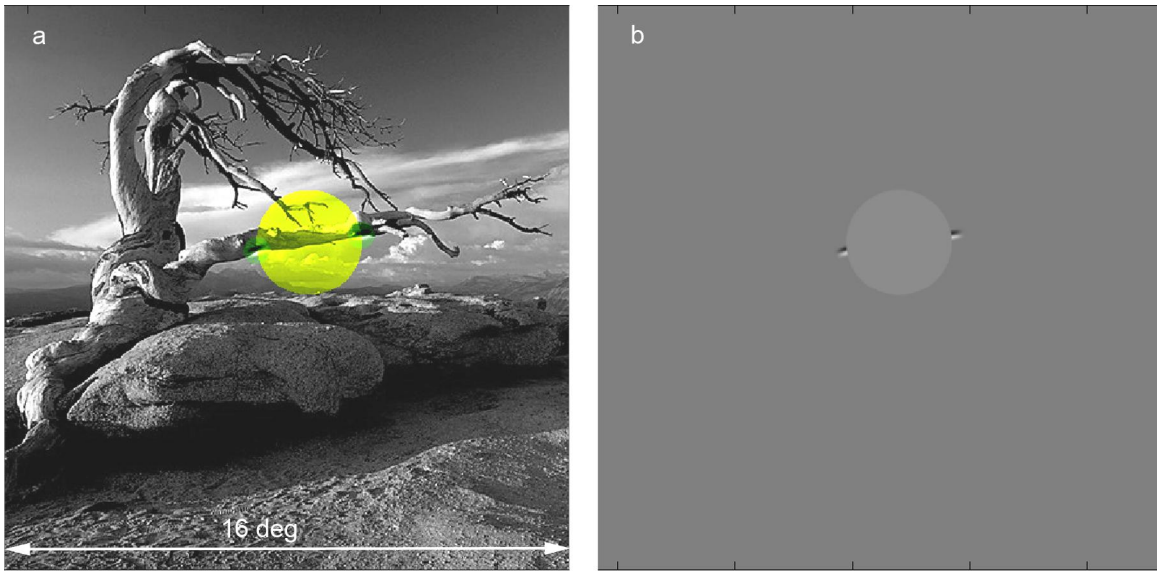

Fig. 3. Contour occlusion task stimuli. **a**. Illustration of how stimulus in b is obtained from a natural image (see text for details). **b**. Display on contour occlusion experiment. The largest occluder diameter was 90 pixels (shown here), corresponding 3 deg of visual angle.

2.3. Ideal observer for contour occlusion task

The only information available to perform the contour occlusion task is the geometrical relationship between the pair of edge elements, and in some conditions the geometrical and contrast-polarity relationship between the elements. Thus, using the measured contour statistics we can derive the performance of a rational (ideal) observer that has perfect knowledge of the natural scene statistics of edge element geometry and contrast polarity. An ideal observer that wishes to maximize accuracy will compare the posterior probability that the observed contour elements belong to the same physical contour with the posterior probability that they belong to different physical contours and then respond “same” if the former posterior probability is the larger:

$$\text{if } p\left(c=1\middle|d,\phi,\theta,\rho\right) > p\left(c=0\middle|d,\phi,\theta,\rho\right) \text{ then respond "same contour"} \qquad (1)$$

It can be shown that this decision rule is identical to this one: [1]

$$\text{if } \frac{p(\phi,\theta,\rho|d,c=1)}{p(\phi,\theta,\rho|d,c=0)} > \frac{p(c=0|d)}{p(c=1|d)} \text{ then respond "same contour"} \qquad (2)$$

The term on the left is a likelihood ratio: the probability of the observed direction, orientation and polarity relationship between the edge elements given the observed distance and that the elements belong to the same contour, divided by the probability of the same observed relationship between the elements given the observed distance and that they belong to different contours. We represent this likelihood ratio by $l(\phi,\theta,\rho|d)$, which is shown in Fig. 2b. The term on the right is the ratio of the prior probabilities that two edge elements separated by distance d belong to different versus the same contour. We represent this distance-dependent decision criterion by $\beta(d)$, which is shown in Fig. 2c. In the present contour occlusion task we forced the prior probability of the edge elements belonging to the same contour to be 0.5, and hence we forced the ideal criterion to be 1.0 ($\beta(d)=1.0$). The subjects were told that the prior probabilities were 0.5 before the start of the experiment. We applied equation (2) to each individual trial for each subject, which allowed us to compare human and ideal observer responses on a trial-by-trial basis, with no free parameters.

2.4. Comparison of human and ideal performance

The contour occlusion task was run first without feedback, then with trial-to-trial feedback, and finally again without feedback. We found that performance was relatively constant before, during, and after the feedback sessions, except perhaps for one subject. We conclude that the subjects started the experiment with stable decision criteria and that there is no evidence they were using different criteria from what they would use in natural scenes (for the specific dimensions of edge element geometry and contrast polarity). Therefore, we combined the data across the phases of feedback.

The average hits and false alarms of the human subjects and ideal observer, as a function of occluder diameter, are shown in Figs. 4a and 4b. The green symbols and curves are for the conditions where both edge element geometry and contrast polarity were displayed. The red symbols and curves are for the conditions where only the edge element geometry was displayed. Humans and ideal observers are affected similarly by occluder diameter and exclusion of contrast polarity information. As occluder diameter is increased hit rate declines and false alarm rate remains relatively constant. Excluding the contrast polarity information causes a slight reduction in hit rate (average of 1.3 % for humans and 0.5 % for ideal), but a more substantial increase in false alarm rate (average of 8.6 % for humans and 5.6 % for ideal). The similarity of human and ideal performance can be quantified by converting the hit and false alarm rates into d' values for real and ideal observers and plotting their ratio. A constant ratio means constant efficiency. As shown in Fig. 4c, efficiency is high and nearly constant with occluder diameter and perhaps slightly higher when contrast polarity information is presented. Near constant efficiency is also seen for the individual subjects (Fig. 4d). Overall, these results suggest that humans have good knowledge of the pair wise statistics of edge element geometry and contrast polarity in natural images and are able to use that knowledge efficiently.

A more detailed comparison of human and ideal decision rules can be obtained by examining the individual trials. Doing this we find that the subjects (like the ideal observer) do not show a strong bias on average; the hits and correct rejections are about equally frequent, and the false alarms and misses are about equally frequent. Also, remarkably, we find that humans respond very similarly to the ideal observer along the various stimulus dimensions (distance, direction, orientation difference and contrast polarity); there are not obvious regions of stimulus space where humans are particularly inefficient. In other words, the human visual system appears to have an accurate representation of the contour statistics in Fig. 2b and to implement something close to the optimal decision rule represented by applying equation (2) to the likelihood ratios in Fig. 2b.

---

[1] We note that representing the optimal decision rule in the form of equation (2) is an improvement over the formulation described in Geisler et al. [7]. The weakness of the previous formulation is that there is no meaningful way to measure ratio of the priors, which left the decision criterion β a free parameter. With equation (2) there are no free parameters.

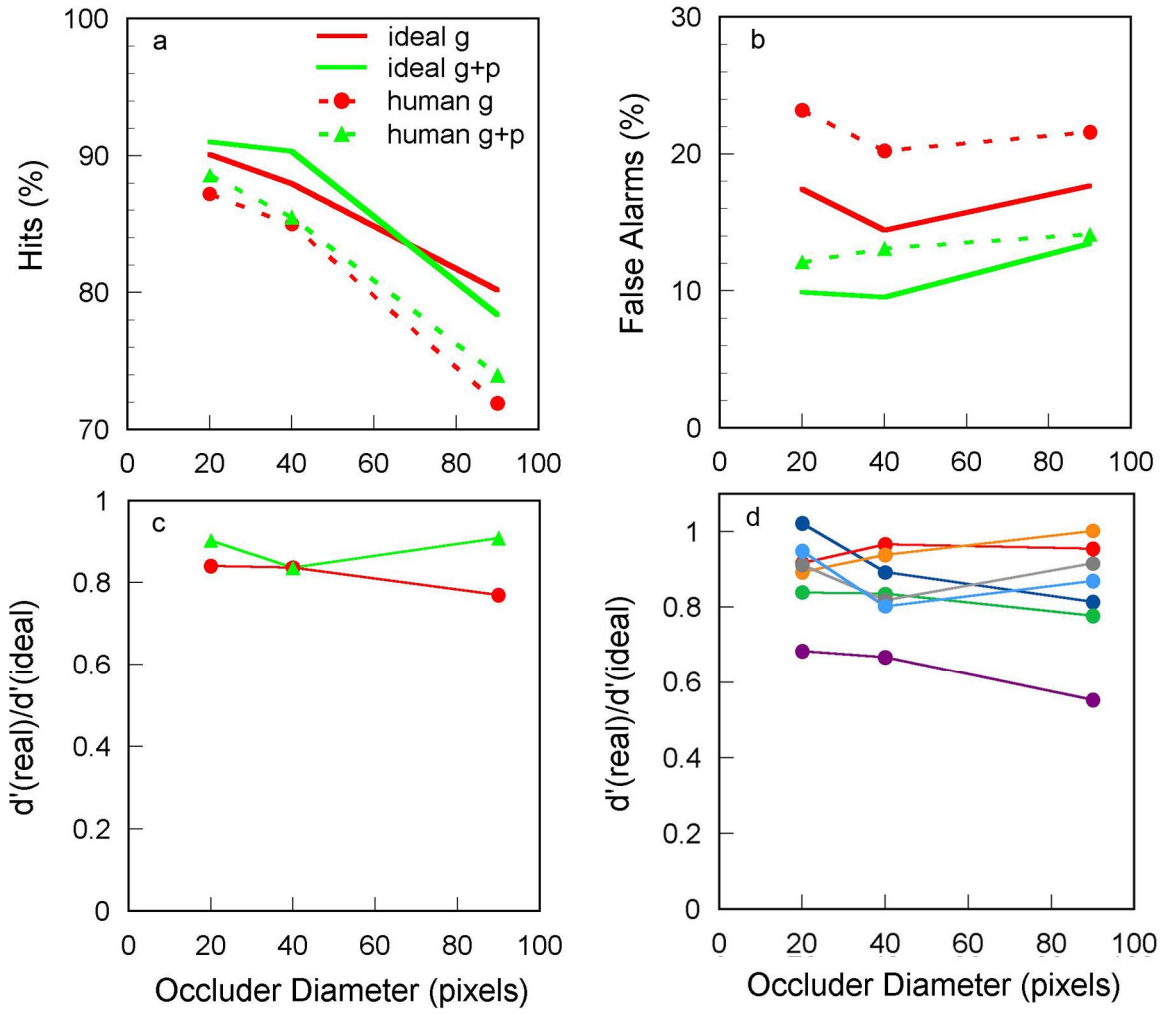

Fig. 4. Performance of seven subjects and the ideal observer in the contour occlusion experiment. **a**. Average percent hits for human and ideal observers plotted separately for tasks where the contrast polarity cue was present (green) and absent (red). **b**. Average percent false alarms for human and ideal observers plotted separately for contrast polarity cue present (green) and absent (red). **c**. Average ratio of sensitivity (d') of human and ideal observers plotted separately for contrast polarity cue present (green) and absent (red). **d**. Ratio of sensitivity (d') of human and ideal observers plotted separately for each of the human observers. Pixel size was 2 min of arc.

3. Example 2: Patch Classification Task

In the patch classification task, the observer is presented two equal size image patches sampled from a natural foliage image at some spatial separation and must decide whether the patches belong to the same or different physical surfaces. This task is roughly analogous to the contour occlusion task, except that it is relevant to surface/region grouping rather than contour grouping.

3.1. Patch statistics

To measure the relevant natural scene statistics we collected a diverse set of calibrated images of close-up foliage, which were then hand segmented by human observers. We chose to analyze close-up foliage because (i) foliage images are a major component of the environment and the dominant component in the natural environment of the macaque monkey (the primary animal model for human vision), (ii) close-up foliage images are relatively easy to hand segment,

and (iii) the statistical properties of distant foliage should be derivable from those of close-up foliage. In future studies, we propose to carry out similar analyses for other classes of image.

The images were captured with a 36-bit-per-pixel Kodak DCS720x digital CCD camera (12 bits per color sensor) that was calibrated to give the approximate responses in the human long (*L*), middle (*M*) and short (*S*) wavelength cones at each pixel location. Hand-segmentation was performed by paid undergraduate students at the University of Texas at Austin, using custom software. To obtain representative statistics it was necessary to segment a large number of images. However, careful segmentation is time consuming and thus it was not practical to perform a full segmentation of each image. Thus, prior to performing the segmentation, one of the authors (ADI) defined a circular region of interest within each image, and the undergraduates were instructed to segment all objects inside or touching the region of interest. By selecting a region from each image, we were able to obtain a dense segmentation for a relatively large sample of foliage images. The segmentation involved creating polygons that defined the boundary of each object's visible surface regions. Because of occlusions, multiple polygons were sometimes required to segment an object. Because an occluding object shares a common edge with the occluded object, polygon vertices could be shared by immediately adjacent polygons. The segmented objects were categorized for quality control. If anything about the segmentation was ambiguous or uncertain, it was deemed to be of *low quality*. The remaining objects were deemed *high quality*. Here, all analyses were performed only on the high quality segmented objects (there were 1,638). Although this restriction may limit the generality of the results to some extent, it guarantees that the segmentations closely approximate ground truth. Furthermore, even with this restriction, the regions of interest were densely segmented. Thus, we believe the measured statistics are highly relevant to natural tasks in the world of close-up foliage (and perhaps in other environments as well). Fig. 5a shows an example of a segmented image; the blue and red shaded objects are segmented leaves and the yellow shaded objects are segmented branches.

The patch statistics were measured by analyzing pairs of patches randomly sampled within objects and across object boundaries (see below).

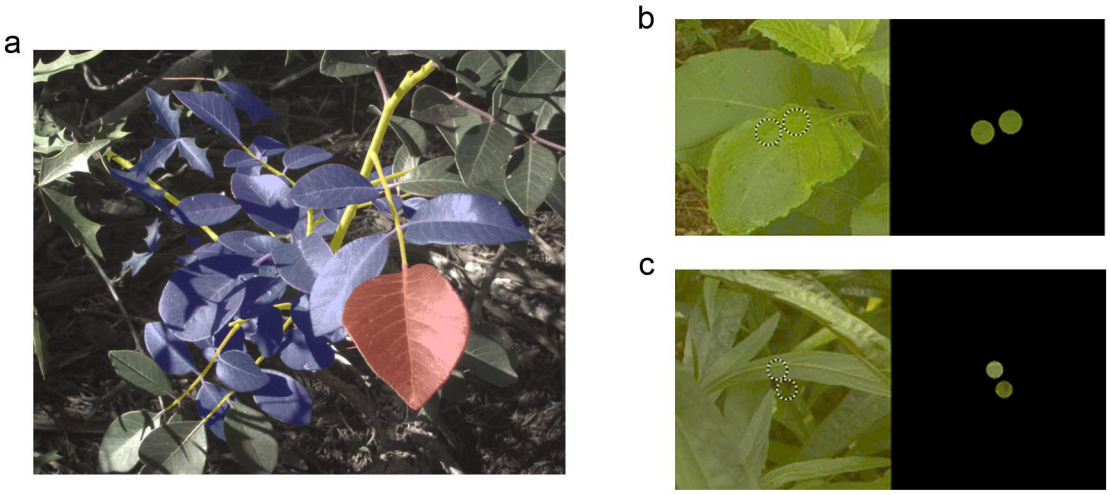

Fig. 5. Patch classification task stimuli. **a**. Example of a segmented calibrated image of close-up foliage used for measuring natural scene statistics and for generating stimuli for the patch classification task. Colored leaves and branches are segmented objects. **b**. Example patches randomly sampled from the same surface. The left image shows the context of the samples. The right image shows the display presented to observers. **c**. Example of patches randomly sampled from different surfaces.

3.2. Patch classification experiment

In the patch classification experiment, the observer was presented two equal size image patches sampled from a natural image at some spatial separation and was required to decide whether patches belong to the same or different physical surfaces. To generate a stimulus, a reference leaf was randomly sampled from the database. The "diameter" of the reference leaf was defined as the square-root as the total number of pixels contained in the leaf. One circular image patch inside the leaf was randomly selected. A second patch was selected so that its center was either 0.25, 0.5 or 1

object diameter from the first patch's center. (Distance and size were measured in fractions of object diameter so that the scene statistics would be roughly independent of viewing distance.) If the second patch was inside the leaf, the pair was a *same-surface* stimulus (Fig. 5b), if the second patch was outside the leaf, the pair was a *different-surface* stimulus (Fig. 5c). The diameter of the image patches was 0.2 of the reference leaf diameter. Neither patch was allowed to intersect a polygon boundary of the reference leaf. The two image patches were displayed on a black background (left panels in Fig. 5b and 5c), on a calibrated monitor.

For each distance, the patches were displayed in three different ways. In the "full" condition, both patches contained all of their original properties and were merely circular cut-outs from the image. In the "texture removed" condition, the texture was removed, only the chromaticity and luminance of the patches were preserved. In the "texture only" condition, the luminance and chromaticity differences were removed, only the spatial texture of the patches was preserved. Thus, there were 9 total conditions [(distance between patches = 0.25, 0.5, or 1 object diameter) $\times$ (patch type = full, texture removed, or texture only)].

On the first day of the experiment (before data collection), subjects browsed through the 96 images that made up the database. They looked at each image for approximately 5 seconds in order to develop an intuitive understanding of the kinds of images in the database. There were 620 trials per condition and subjects were informed that half were "same surface" stimuli and half were "different surface" stimuli. During the first 20 trials, a large portion of the original image was also displayed next to the patches. These trials served to demonstrate the nature of the task to the subjects. In the subsequent 600 trials, only the two image patches were displayed. On each trial, subjects viewed the patches for as long as they desired and pushed a button to categorize the patch as "same surface" or "different surface." The 9 conditions were run on separate days and feedback was not provided. In a second experiment, the same 9 conditions were run on the same subjects, but feedback was given on each trial.

3.3. Approximate ideal observer for patch classification task

Using randomly sampled pairs of patches from the database as a training set, we determined the approximate ideal observer (a model observer) for certain stimulus dimensions in the patch classification task. First, we converted the average L, M and S cones absorptions for each pixel into the $l\alpha\beta$ color space proposed by Ruderman [10], where $l = (\log L + \log M + \log S)/\sqrt{3}$, $\alpha = (\log L + \log M - 2\log S)/\sqrt{6}$, and $\beta = (\log L - \log M)/\sqrt{2}$. This space is useful because l, α and β are approximately statistically independent and Gaussian distributed in foliage rich natural scenes (we verified this for pixels in our database). We then defined six difference measures between image patches: Δl (difference in mean intensity value), $\Delta\alpha$ (difference in mean blue-yellow value), $\Delta\beta$ (difference in mean red-green value), $\Delta\sigma_l$ (difference in standard deviation of intensity values), $\Delta\sigma_\alpha$ (difference in standard deviation of blue-yellow values), and $\Delta\sigma_\beta$ (difference in standard deviation of red-green values). Note that the first three measures are differences in mean values (see [11]), and the last three measures are differences in contrast; also note that all of these measures are similar to Weber measures, because they involve differences of log values.

The probability distribution of these difference values (not shown here) for same- and different-surface patch pairs, as a function of the distance between patches constitutes the measured natural scene statistics. Using these measured natural scene statistics we derived the optimal quadratic decision bounds for the patch classification task. This was done in three ways: with all six measures, with the three mean-difference measures, and with the three contrast-difference measures.

3.4. Comparison of human and model observer performance

The symbols in Fig. 6 show the performance accuracy of the two naïve subjects for the initial phase of the experiment where no feedback was given. The gray symbols show the raw accuracy levels, and the blue symbols show the accuracy (PC max) corrected for bias, based on the standard signal detection analysis of the hit and false alarm rates [12]. The gray curves show the performance of the model observer. We note that only for the "texture removed" conditions is the model observer likely to be close to the true optimal; this is because in the "full" and "texture only" conditions the exact spatial contrast pattern of the patches was displayed, but the model observer only had access to the contrasts of the patches.

For both human and model observers, classification accuracy decreases with the distance between the patches. In the "full" and "texture removed" conditions performance was near 80% at the smallest distance and decreased to near 65% percent at the largest distance. This result is intuitive and shows that for region grouping, greater reliance should be placed on neighboring image regions. The fact the human performance parallels the model observer performance suggests that human efficiency is relatively high and constant (at least for the texture removed conditions where the model observer is nearest ideal). Furthermore, because no-feedback was given the results suggests that the human observers came into the experiment with an accurate representation of patch statistics, for scenes containing close-up foliage. This conclusion was supported by the second phase of the experiment where feedback was given; we found only small improvements in performance.

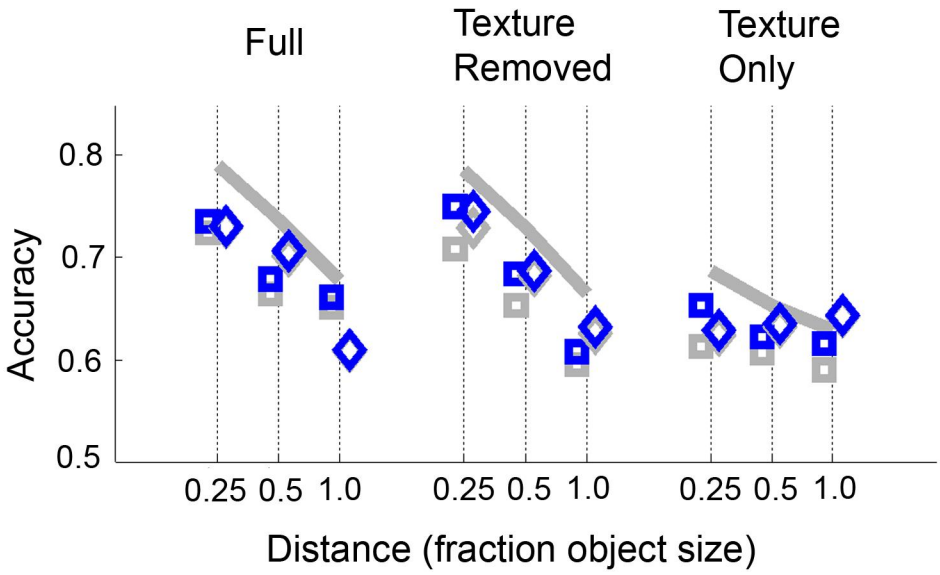

Fig. 6. Performance of two observers and model observers in patch classification experiment.

4. Conclusion

In sum, we performed an initial *natural systems analysis* for two basic visual tasks, contour grouping across occlusions, and region grouping of surface patches for leaf surfaces in close-up foliage scenes. We found that it is possible to quantitatively measure the relevant natural scene statistics and to derive parameter-free ideal observers (model observers) for these tasks. These ideal observers show how the visual system *should* represent and apply its (implicit) knowledge of the natural scene statistics to perform these tasks. We found that human performance closely parallels the performance of the ideal observers. Thus, slightly degraded versions of the ideal observer models can serve as candidate models for the corresponding neural coding and decision mechanisms of the human visual system. For example, a plausible working hypothesis is that the human visual system represents the structure in Figs. 2b and 2c in its contour grouping mechanisms.

ACKNOWLEDGEMENTS

Supported by NIH grant EY11747. Direct correspondence to WSG (geisler@psy.utexas.edu).

REFERENCES

[1]Regan,B.C.,Julliot,C.,Simmen,B.,Vienot,F.,Charles-Dominique,P. and Mollon,J.D.,"Fruits, foliage and the evolution of primate colour vision," Philosophical Transactions of the Royal Society of London B 356, 229-83 (2001).

[2]Simoncelli,E.P. and Olshausen,B.A.,"Natural image statistics and neural representation," Annual Review of Neuroscience 24, 1193-216 (2001).

[3]Geisler,W.S., "Visual perception and the statistical properties of natural scenes," Annual Review of Psychology 59, 167-192 (2008).
[4]Knill,D.C. and Richards,W.,[Perception as Bayesian Inference], Cambridge University Press,Cambridge, 1 (1996).
[5]Geisler,W.S. and Diehl,R.L.,"A Bayesian approach to the evolution of perceptual and cognitive systems," Cognitive Science 27, 379-402 (2003).
[6]Kersten,D.,Mamassian,P. and Yuille,A.L.,"Object perception as Bayesian inference," Annual Review of Psychology, 55, 271-304 (2004).
[7]Geisler,W.S.,Perry,J.S.,Super,B.J. and Gallogly,D.P.,"Edge co-occurrence in natural images predicts contour grouping performance," Vision Research 41, 711-724 (2001).
[8]Elder,J,H. and Goldberg,R.M.,"Ecological statistics for the Gestalt laws of perceptual organization of contours," Journal of Vision 2, 324-353 (2002).
[9]Martin,D.R.,Fowlkes,C.C. and Malik,J.,"Learning to detect natural image boundaries using local brightness, color and texture cues," IEEE Transaction on Pattern Analysis & Machine Intelligence. 26, 530-549 (2004).
[10]Ruderman,D.L., "The statistics of natural images," Computation in Neural Systems 5, 517-548 (1994).
[11]Fine,I.,MacLeod,D.I.A. and Boynton,G.M., "Surface segmentation based on the luminance and color statistics of natural scenes," Journal of the Optical Society of America A, 20, 1283-1291 (2003).
[12]Green,D.M. and Swets,J.A.,[Signal detection theory and psychophysics], Krieger,New York, 1 (1974).