Visual Perception and the Statistical Properties of Natural Scenes

Wilson S. Geisler

Center for Perceptual Systems and Department of Psychology, University of Texas
at Austin, Austin, Texas 78712-1062; email: geisler@psy.utexas.edu

Key Words

natural scene statistics, spatial vision, motion perception, color
vision, ideal observer theory

Abstract

The environments in which we live and the tasks we must perform
to survive and reproduce have shaped the design of our percep-
tual systems through evolution and experience. Therefore, direct
measurement of the statistical regularities in natural environments
(scenes) has great potential value for advancing our understanding
of visual perception. This review begins with a general discussion of
the natural scene statistics approach, of the different kinds of statis-
tics that can be measured, and of some existing measurement tech-
niques. This is followed by a summary of the natural scene statistics
measured over the past 20 years. Finally, there is a summary of the
hypotheses, models, and experiments that have emerged from the
analysis of natural scene statistics.

Contents

RATIONALE FOR MEASURING NATURAL SCENE STATISTICS

The process of natural selection guarantees a strong connection between the design of an organism's perceptual systems and the properties of the physical environment in which the organism lives. In humans, this connection is implemented through a mixture of fixed (hardwired) adaptations that are present at birth and facultative (plastic) adaptations that alter or adjust the perceptual systems during the lifespan.

The link between perceptual systems and environment is most obvious in the design of sensory organs. The physical properties of electromagnetic waves, acoustic waves, and airborne molecules and their relation to the properties of objects and materials are clearly a driving force behind the evolution of eyes, ears, and noses. Not surprisingly, central perceptual mechanisms that process the outputs of sensory organs also tend to be closely related to specific physical properties of the environment.

The design of a perceptual system is also constrained by the particular tasks the organism evolved to perform in order to survive and reproduce. For example, mammals that suffer high rates of predation have a strong need to detect predators and hence tend to have laterally placed eyes that maximize field of view, whereas mammals that are predators have a strong need to capture moving prey and hence tend to have frontally placed eyes that maximize binocular overlap (Walls 1942). Furthermore, there are purely biological constraints on the design of perceptual systems, including the biological materials available to construct the sensory organs and competition for space with other organs and systems within the body.

Our often-veridical perceptions of the world give the impression of a deterministic connection between perception and environment; however, this is largely an illusion. Most perceptual capabilities depend upon combining many very different sources of stimulus information, each of which is only probabilistically predictive in the task the organism is trying to perform. For example, our estimates of physical object size and shape are often based upon a combination of information sources, including lighting/shading, texture, occlusion, motion, and binocular disparity. Each of these sources is only probabilistically related to object shape and size, but together they provide us with very robust perceptions and perceptual performance. Furthermore, all visual measurements are noisy due to the inherent randomness of light absorption and chemical events within the photoreceptors. Consequently, the appropriate way to characterize natural stimuli is in statistical terms.

The primary aim of this review is to demonstrate the great potential value of analyzing the statistical properties of natural

scenes,[1] especially within the context of developing statistical models of perception. Another aim is to demonstrate that the Bayesian framework for statistical inference is particularly appropriate for characterizing natural scene statistics and evaluating their connection to performance in specific tasks. In principle, measuring natural scene statistics allows one to identify sources of stimulus information available for performing different perceptual tasks and to determine the typical ranges and reliabilities of those sources of information. Analyzing natural scene statistics within the Bayesian framework allows one to determine how a rational visual system should exploit those sources of information. This approach can be valuable for generating hypotheses about visual mechanisms, for designing appropriate experiments to test those hypotheses, and for gaining insight into why specific design features of the visual system have evolved or have been learned.

Roots of the Natural Scene Statistics Approach

The natural scene statistics approach originates in physics. Historically, physics has been concerned with topics of direct relevance to understanding the design of visual systems including the properties of light, the laws of image formation, the reflectance, scattering, and transmittance properties of natural materials, and the laws of motion and gravity. Against this backdrop, biologists began asking how visual systems are adapted to the physical environment and to the tasks that the organism performs. Most early work on the ecology and evolutionary biology of vision was concerned with the optics and the photoreceptors of eyes (e.g., Cronly-Dillon & Gregory 1991, Lythgoe 1979, Walls 1942). This early work emphasized the relationship between design,

function, and the properties of the environment, but because of the issues being investigated, gave little consideration to the statistical properties of natural stimuli.

Interest in the statistical properties of natural visual stimuli began with the discovery in physics of the inherent Poisson randomness of light (quantal fluctuations). Human and animal studies by early sensory scientists subsequently showed that under some circumstances behavioral and neural performance is limited by a combination of quantal fluctuations and internal sources of noise (e.g., Barlow 1957, Barlow & Levick 1969, Hecht et al. 1942). This work, along with parallel work in audition, led to the development of signal detection theory and Bayesian ideal observer theory (e.g., see Green & Swets 1966), which provides an appropriate formal framework for proposing and testing hypotheses about the relationship between perceptual performance and the statistical properties of stimuli and neural responses. However, early work on the statistical properties of visual stimuli and neural responses focused on simple detection and discrimination tasks, paying little attention to sources of stimulus variation other than quantal fluctuations and pixel noise.

Some early perception scientists (e.g., Gibson 1966, 1979) did appreciate the importance of the complex properties of natural stimuli for solving perceptual tasks, but they paid little attention to statistical variations of those properties in natural scenes. An exception was Brunswik (1956), who realized that the relationship between distal and proximal stimuli is inherently statistical; in fact, he demonstrated, by analyzing natural images, that perceptual biases such as the Gestalt rule of proximity have a statistical basis in natural scenes (Brunswik & Kamiya 1953).

Recent years have seen rapid growth in the statistical analyses of natural images (for a review, see Simoncelli & Olshausen 2001) as well as in the analysis and modeling of complex perceptual tasks within the framework of Bayesian ideal observer theory (for reviews,

---

[1]Natural scenes refer to real environments, as opposed to laboratory stimuli, and may include human-made objects. Most of the studies described here concern measurements of outdoor environments without human-made objects.

see Geisler & Diehl 2003, Kersten et al. 2004, Knill & Richards 1996). A central tenet of this review is that combining measurements of natural scene statistics with Bayesian ideal observer analysis provides an important new approach in the study of sensory and perceptual systems.

Within-Domain Statistics

Natural scene statistics have been measured at various stages (domains) along the path from physical environment to behavioral response. The simpler and more common kinds of measurements are what I call within-domain statistics (first column in **Figure 1A**). The purpose of within-domain statistics is to characterize the probability distribution of properties within a specific domain such as the (distal) environment or (proximal) retinal image. The more complex and less common kinds of measurements are what I call across-domain statistics (other columns in **Figure 1A**). Their purpose is to characterize the joint probability distribution of properties across specific domains. Across-domain statistics are essential for analyzing natural scene statistics within the Bayesian ideal observer framework.

In the case of within-domain statistics for the environment, a vector of physical scene properties ω is selected and then those properties are measured in a representative set of scenes in order to estimate the probability distribution of the properties $p(\omega)$. For example, ω might be the reflectance function at a point on a surface in a natural scene; that is, a vector giving the fraction of light reflected from the surface for a number of different wavelengths, $\omega = [\rho(\lambda_1), L, \rho(\lambda_n)]$. Making these physical measurements for a large number of surface points in natural scenes would make it possible to estimate the probability distribution of natural surface reflectance. Similarly, in the case of within-domain statistics for images, a vector of retinal image properties $\mathbf{s}$ is selected and their distribution measured. For example, $\mathbf{s}$ might be a vector rep-

resenting the wavelength spectrum at a retinal image location, $\mathbf{s} = [I(\lambda_1), \ldots, I(\lambda_n)]$ (see plots in **Figure 1B**). For the domain of neural response, a set of response properties for a population of neurons is selected and their distribution measured for a representative set of natural stimuli. In this case, $\mathbf{z}$ might be a vector of the spike counts of each neuron, $\mathbf{z} = [count_1, \ldots, count_n]$. Finally, for the domain of behavior, a vector of properties for some class of behavior is selected and their distribution measured for a representative set of natural stimuli. For example, $\mathbf{r}$ might be the eye fixation locations in a natural image during free viewing, $\mathbf{r} = [fixation_1, \ldots, fixation_n]$.

Measurements of within-domain statistics are highly relevant for understanding neural coding and representation. A plausible hypothesis is that the retina and other stages of the early visual pathway have evolved (or learned) to efficiently code and transmit as much information about retinal images as possible, given the statistics of natural images and biological constraints such as the total number of neurons and the dynamic range of neural responses. Variants of this efficient coding hypothesis have been widely proposed and evaluated (Atick & Redlich 1992; Attneave 1954; Barlow 1961, 2001; Field 1994; Laughlin 1981; van Hateren 1992). For example, the efficient coding hypothesis predicts many of the response characteristics of neurons in the retina directly from the joint probability distributions of the intensities at two pixel locations in natural images, $p(\mathbf{s}) = p(I_1, I_2)$, measured for various separations of the pixels in space and time. The measurement of within-domain statistics is central to the enterprise of testing the efficient coding hypothesis: To determine what would be an efficient code, it is essential to know the probability distribution of the image properties to be encoded.

Across-Domain Statistics

Within-domain statistics say nothing about the relationship between the domains listed in **Figure 1A**, such as the relationship

between properties of the environment and the images formed on the retina. Natural visual tasks generally involve making inferences about specific physical properties of the environment from the images captured by the eyes. These tasks include classifying materials, discriminating object boundary contours from shadow contours, estimating object shape, identifying objects, estimating the distance or motion of an object, or estimating one's own motion direction and speed. The relevant statistics for understanding how the visual system performs such tasks are the joint distributions of physical scene properties and image properties, $p(\omega, \mathbf{s})$, or equivalently, $p(\omega)$ and the conditional distribution $p(\mathbf{s}|\omega)$ for each value of ω. [Note that $p(\omega, \mathbf{s}) = p(\mathbf{s}|\omega)p(\omega)$.] Using Bayes' rule, these across-domain statistics specify the posterior probabilities of different states of the world (physical environment properties) given particular observed retinal image properties $p(\omega|\mathbf{s})$. It is the characteristics of the posterior probability distributions that visual systems evolve or learn to exploit in performing specific natural visual tasks.

Suppose an organism's task is to identify (for a given species of tree) whether a randomly selected location (small patch) in the retinal image corresponds to the surface of a fruit or the surface of a leaf, based solely on the information available in the wavelength spectrum at that randomly selected location (see **Figure 1B**). In this case, there are just two relevant states of the environment (distal stimuli): $\omega = fruit$ and $\omega = leaf$, and due to variations in lighting and reflectance, a large number of possible wavelength spectra (proximal stimuli) for each distal stimulus. In principle, $p(\omega)$ could be measured by randomly selecting lines of sight from a large number of example scenes and counting the proportion of times a fruit surface is the first surface encountered along that line of sight. Similarly, $p(\mathbf{s}|\omega)$ could be measured by recording the wavelength spectra for each of the randomly sampled lines of sight, sorting them according to whether they are from fruit or leaf, and then analyzing them separately to estimate the two conditional probability distributions. The statistical regularities represented by $p(\mathbf{s}|\omega)$ and $p(\omega)$ could be exploited by the visual system for identifying fruits and leaves from the wavelength spectra that reach the eye. On the other hand, knowing only the within-domain statistics, $p(\mathbf{s})$ and $p(\omega)$, is not useful for identifying fruits and leaves because the statistics do not specify the relationship between the image properties (wavelength spectra) and the physical objects (fruits and leaves) of relevance in the task. This example illustrates the use of across-domain statistics for characterizing the connection between environmental properties and image properties; comparable examples can be generated for the other kinds of across-domain statistics (see table in **Figure 1A**).

Bayesian ideal observer theory provides an appropriate formal framework for understanding how across-domain statistics might be exploited by the visual system to perform specific tasks (Geisler & Diehl 2003). The Bayesian approach in perception research has been discussed at length elsewhere and is only briefly summarized here, as a prelude to some of the studies described below. An "ideal observer" is a theoretical device that performs a task in an optimal fashion given the available information (and possibly other constraints). Deriving an ideal observer can be very useful because (*a*) the derivation usually leads to a thorough understanding of the computational requirements of the perceptual task, (*b*) the ideal observer provides the appropriate benchmark for comparison with behavioral performance, and (*c*) ideal observers often reduce to, or can be approximated by, relatively simple decision rules or procedures that can serve as initial hypotheses for the actual processing carried out in a perceptual system.

The logic behind deriving an ideal observer is straightforward. Consider an ideal observer that wishes to perform a specific task in its current environment and has access to some vector of properties in the retinal image. Upon receiving a particular stimulus vector

S, an ideal observer should make the response that maximizes the utility (gain/loss) averaged over all possible states of the environment,

$$\mathbf{r}_{opt}(\mathbf{S}) = \arg\max_{\mathbf{r}} \left(\sum_{\omega} \gamma(\mathbf{r}, \omega) p(\mathbf{S}|\omega) p(\omega) \right)$$

(1)

where $\gamma(\mathbf{r}, \omega)$ is the utility of making response $\mathbf{r}$ when the true state of the environment is ω, and the function arg max returns the response that maximizes the sum in the brackets. In other words, once the relevant across-domain statistics have been estimated and the utility function for the task has been specified, then Equation (1) can be used to determine (via simulation or calculation) the optimal performance in the task.

Think back to the hypothetical task described above: identifying whether a small patch of retinal image corresponds to a fruit or a leaf. There are two possible responses, $\mathbf{r} = \textit{fruit}$ and $\mathbf{r} = \textit{leaf}$. To maximize accuracy, an appropriate utility function is $\gamma(\mathbf{r}, \omega) = 1$, if $\mathbf{r} = \omega$, and $\gamma(\mathbf{r}, \omega) = -1$, if $\mathbf{r} \neq \omega$ (i.e., equal but opposite weights for corrects and errors). Substituting this utility function and the measured probability distributions into Equation (1) gives the (parameter-free) optimum decision rule. The performance accuracy of this decision rule can be determined by applying the rule to random samples $(\Omega, \mathbf{S})$ from the across-domain probability distribution, $p(\omega, \mathbf{s})$, or alternatively, by directly calculating the probability that $\mathbf{r}_{opt}(\mathbf{S}) = \Omega$. The optimal decision rule (or an approximation to it) could serve as a principled hypothesis about the perceptual mechanisms in the organism that discriminate fruits from leaves.

MEASURING NATURAL SCENE STATISTICS

A variety of devices and techniques has been used to measure natural scene properties. Spectrophotometric devices measure the wavelength distribution (radiance as a function of wavelength) of the light that reaches their sensors. They can be used to measure reflectance spectra of materials, irradiance spectra of light sources (illuminants), as well as radiance spectra that reach the eye. Spectrophotometers collect light over only one small patch at a time, making them impractical for collecting data from a large number of locations in a scene. Hyperspectral cameras can measure radiance spectra at each camera pixel location, but require relatively long exposure time, and thus are practical only for conditions where effects of object and shadow motion are minimized (e.g., long distances or indoor environments). The most common method of measuring natural scene properties has been to analyze images captured by digital still cameras and digital movie cameras. Digital cameras usually provide either 8-bit grayscale or 24-bit color (8 bits per color) images, although some high-end cameras provide 36-bit color (12 bits per color) images, which is desirable for some kinds of measurements. A weakness of standard digital cameras is that they cannot provide detailed spectral (chromatic) information, although with proper calibration it is possible to obtain images that give, for each pixel location, the approximate luminance and/or the approximate amount of radiant power absorbed in each of the three classes of cone photoreceptor, the long (L), middle (M), and short (S) wavelength cones. Many studies have analyzed uncalibrated camera images, which is justifiable if the scene statistics of interest (e.g., contour geometry) are little affected by monotonic transformations of the camera's color responses.

Another useful class of device is the range finder, which measures distance to each point in a scene by measuring return time for an emitted pulse of infrared laser light. These devices are accurate for large distances (a few meters to a kilometer or more). A related class of device is the three-dimensional scanner, which uses triangulation rather than time-of-flight, and is useful for making precise range measurements at near distances (e.g., measuring the shape of a face). A weakness of both devices is that the scans take

substantial time (typically seconds), and thus motion in the scene can produce significant distortions.

The above devices are the most common for measuring natural scene statistics, and they can be used in a fairly straightforward manner to measure within-domain statistics for the image or environment. Measuring across-domain statistics is more difficult because both image and environment properties must be measured for the same scene. One approach is to combine environment measurements from one instrument with image measurements from another instrument. For example, monocular across-domain statistics for depth can be measured by combining a calibrated camera image with a distance map obtained with a range finder.

An expedient approach for measuring across-domain statistics involves hand segmentation. The central assumption of this approach is that under some circumstances humans are able to make veridical assignments of image pixels to physical sources in the environment. When this assumption holds, the pixel assignments are useful measurements of environmental properties.

For example, consider the close-up image of foliage in **Figure 2A**. When observers are asked to segment individual leaves and branches that are within or touching the orange dashed circle, the result in **Figure 2B** is obtained. The colored leaves and branches show the segmented objects, the red/brown shaded leaf shows one individual segmented object. These segmentations are performed with high confidence and repeatability, and hence generally provide an accurate measurement of the physical source (the specific leaf or branch) that gave rise to a given pixel. Many across-domain statistics can be measured with a large set of such segmentation data. To take one simple example, it is straightforward to measure the posterior probability of the same or different object [ω = same or ω = different], given the distance between a pair of image pixels and their luminance values [$\mathbf{s} = (d_{12}, l_1, l_2)$].

Hand segmentation methods are useful for measuring across-domain statistics only to the extent that the segmentations are veridical (i.e., represent physical "ground truth"), and there are cases (e.g., distant images of foliage) where some image regions are ambiguous and difficult to segment.[2] In cases where hand segmentation methods fail, accurate ground truth measurements require more direct physical measurement. Another strategy for measuring across-domain statistics combines computer-graphics simulation with direct measurements of scene statistics.

NATURAL SCENE STATISICS

It is difficult to know ahead of time which specific statistics will prove most informative for understanding vision. At this time, progress is being made by selecting statistics based on intuition, historical precedence, and mathematical tractability. It is important to note that the number of samples required for estimating a probability distribution grows exponentially with the number of properties/dimensions ("the curse of dimensionality"), and hence most studies measure natural scene statistics for only one or a few properties at a time. This is a significant limitation because perceptual mechanisms exploit complex regularities in natural scenes that may only be fully characterized by measuring joint distributions over a substantial number of dimensions. Nonetheless, the published work has demonstrated that much can be learned from low-dimensional measurements and that there are useful methods for learning about the structure of probability distributions in high-dimensional spaces. This section presents a somewhat eclectic summary of some of the natural scene statistics that have been measured.

---

[2]Hand segmentation has also been used to measure how humans segment scenes into regions without specific instructions to be exhaustive or identify physical sources; in this case, the aim is not to precisely measure physical ground truth but rather to obtain a useful data set for training image-processing algorithms (e.g., Martin et al. 2004).

Luminance and Contrast

Luminance and contrast, fundamental stimulus dimensions encoded by visual systems, vary both within a given scene and across scenes. Most studies have involved measuring the statistics of luminance and contrast within images of natural scenes (e.g., Brady & Field 2000, Frazor & Geisler 2006, Laughlin 1981, Ruderman 1994, Tadmor & Tolhurst 2000). The distribution of local luminance within a given image is typically obtained by first dividing the luminance at each pixel by the average for the whole image.[3] Combining these distributions across images and then scaling to the average luminance of the images gives the distribution of luminance in the typical natural image. As shown in **Figure 3A**, this distribution is approximately symmetric on a logarithmic axis and hence positively skewed on a linear scale (Brady & Field 2000, Laughlin 1981, Ruderman et al. 1998). In other words, relative to the mean luminance, there are many more dark pixels than light pixels.

The distribution of local contrast within images has been measured using various definitions of contrast. **Figure 3B** shows the distribution of local root-mean-squared contrast (the standard deviation of luminance divided by the mean luminance in a small neighborhood) in the typical natural image. Another more specialized measure is an equivalent Michelson contrast—the Michelson contrast of a sine wave grating (sine wave amplitude divided by mean) that would produce the same contrast response as the local image patch, where the contrast response is from a filter designed to mimic a typical receptive field at some level of the visual system (Brady & Field 2000, Tadmor & Tolhurst 2000). These latter distributions tend to be similar in shape to the one in **Figure 3B**, but (as expected given

the selectivity of the filter) are shifted toward lower contrasts.

There are large variations of local luminance and local contrast in natural images, and these variations tend to be statistically independent. The average joint distribution of luminance and contrast has a slight negative correlation ($r = -0.2$) primarily due to the fact that sky regions tend to be both bright and low in contrast (**Figure 3C**). Low correlations between luminance and contrast are also observed within the constituents of natural images. For example, the joint distribution of luminance and contrast in purely foliage regions (**Figure 3D**) has a slight positive correlation ($r = 0.15$). As discussed below, the large variations in local luminance and contrast and their low correlation have important implications for neural coding.

Color

Interest in natural scene statistics was stimulated by the discoveries that the chromatic power spectra of natural light sources (Dixon 1978, Judd et al. 1964) and the reflectance spectra of natural materials (Maloney 1986, Maloney & Wandell 1986) are highly constrained and can be characterized with just a few numbers. These studies used the standard statistical technique of principal components analysis (PCA) to describe the structure of the probability distributions. For example, each reflectance spectrum can be represented by a single vector in a high-dimensional space, where each dimension of the space is the reflectance at a particular wavelength. A large set of reflectance spectra create a cloud of vectors in this space. Under the assumption of normality, PCA finds the principal axes of this cloud. The first principal axis is the one that accounts for the most variance in the distribution (i.e., it is the direction in the space along which the cloud is most spread out); the second principal axis is the one perpendicular to the first that accounts for most of the remaining variance, and so on. Principal components are unit vectors along the principal axes; in

---

[3]One could regard the ratio of pixel luminance to global luminance as a form of Weber contrast, but here the term "contrast" is reserved for measurements of luminance variation relative to the average luminance in a small neighborhood.
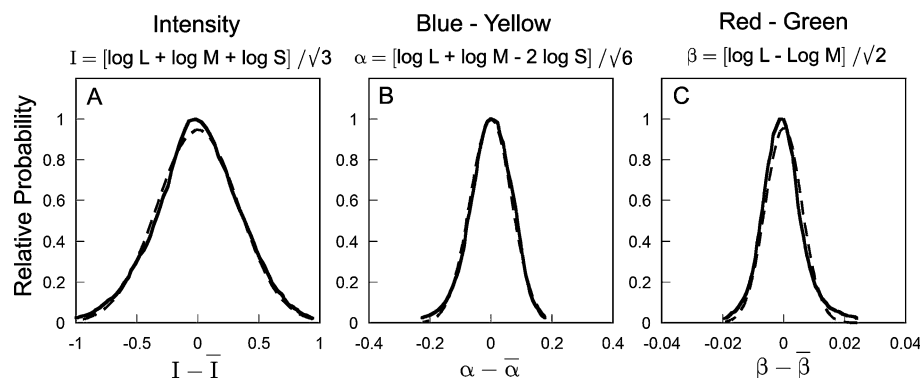
Intensity

$I = [\log L + \log M + \log S] / \sqrt{3}$

Blue - Yellow

$\alpha = [\log L + \log M - 2 \log S] / \sqrt{6}$

Red - Green

$\beta = [\log L - \log M] / \sqrt{2}$

Figure 4

Probability distributions for color in natural foliage scenes. (*A*) Distribution along the first principal axis in log cone response space, minus the average in the image. (*B*) Distribution along the second principal axis. (*C*) Distribution along the third principal axis. The dashed curves are best-fitting Gaussian distributions. The joint probability distribution is approximately the product of these three marginal distributions (adapted from Ruderman et al. 1998).

other words, they describe the directions of the principal axes.

The important discovery of these studies is that most of the variance is accounted for by the first few principal axes. The implications are that the probability distribution of natural irradiance spectra and of natural reflectance spectra can each be described in a rather small dimensional (tractable) space and that natural images can be coded relatively accurately with just a few classes of cone receptor, although three classes (the number in humans) is somewhat short of optimal (Brainard & Freeman 1997, D'Zmura & Iverson 1993, Maloney & Wandell 1986).

The chromatic properties of natural images captured with digital cameras are often described in terms of the number of photons absorbed in the three types of cones, at each image location. Thus, analogous to measuring the distribution of local luminance in natural images (**Figure 3***A*), one can measure the distribution of cone responses in natural images. As it turns out, the joint distribution of the logarithm of the cone responses in natural images is approximately Gaussian. A convenient way to describe Gaussian distributions is to determine the marginal distributions along the principal axes with PCA. This

is convenient because the joint distribution is then the product of the marginal distributions (note that the aim of using PCA is different here than in the case of analyzing natural irradiance and reflectance spectra). **Figure 4** shows the three principal axes (specified by the equations above each panel) and marginal distributions for foliage-dominated natural images obtained with a hyperspectral camera (Ruderman et al. 1998). Exactly the same principal components for close-up foliage images (e.g., **Figure 2**) were obtained with a calibrated digital camera (Ing & Geisler 2006).

Spatial Structure

Much of the information in retinal images is contained in the spatial pattern of luminance and color. One overall spatial statistic of natural images, which is relatively consistent across scenes, is the Fourier amplitude spectrum (or equivalently the spatial autocorrelation function). **Figure 5***A* shows the amplitude spectra of six different natural images plotted on logarithmic axes (each was obtained from the two-dimensional Fourier power spectrum by summing across orientation and then taking the square root). The solid line represents a slope of −1.0;

thus the amplitude spectra fall approximately as $1/f^n$, where f is spatial frequency and the exponent n is approximately 1.0 (Burton & Moorehead 1987, Field 1987, Ruderman & Bialek 1994). One consequence is that the amplitude spectra of natural images are relatively scale invariant, in the sense that scaling all frequencies by a factor (e.g., moving forward or backward in a scene) has little effect on the shape of the amplitude spectrum. For more discussion of scale invariance, see Ruderman (1997, 2002), Lee et al. (2001), and Balboa et al. (2001).

Based on **Figure 5A**, the simplest model of the spatial structure of natural images would be that produced by a $1/f$ amplitude spectrum and a random phase spectrum. An image created in this fashion is a sample of filtered Gaussian noise, often called $1/f$ noise. Such a sample of filtered noise does not contain recognizable objects or regions, but does contain complex random features. Thus, it is reasonable to ask if $1/f$ noise can serve as a statistical model of the local properties of images. One way to address this question is to compare how local sensors like those in visual cortex respond to $1/f$ noise versus actual natural images. A local sensor with an oriented receptive field (similar to an orientation-selective cortical neuron) computes a weighted sum across a small patch of image. Because $1/f$ noise is Gaussian distributed, any given weighted sum will also be Gaussian distributed (dashed curve in **Figure 5B**). However, in natural images the probability distribution of local sensor responses is generally not Gaussian, but rather is sharply peaked at zero response with heavy tails (solid curve in **Figure 5B**). In other words, for real images a given local sensor tends to respond relatively infrequently, but when it does respond it tends to produce a relatively large response (Field 1987, Olshausen & Field 1997).

One way that $1/f$ noise differs from natural images is that the local luminance distribution of natural images is not Gaussian on a luminance axis, but rather is approximately Gaussian on a log luminance axis (**Figures 3A** and **4**). Thus, a more realistic model of the spatial structure of natural images would be that of noise having a $1/f$ amplitude spectrum and a local luminance distribution that matches natural scenes. Frazor & Geisler (2006) call this first-order $1/f$ noise because the local luminance distribution is the first-order statistic of an image and the amplitude spectrum corresponds to the second-order statistic. Sensor responses to first-order $1/f$ noise peak more strongly at zero and have heavier tails than those for $1/f$ noise; however, the peaks and the tails of the distributions are not as pronounced as they are for natural images.

A more complete description of the spatial structure of natural images can be obtained by examining the joint statistics of responses from pairs of local sensors. As one would expect, there are significant correlations between pairs of nonoverlapping sensors. For example, if there is a large response from an oriented edge sensor, then it is likely that a neighboring colinear (or cocircular) edge sensor also has a large response (Geisler et al. 2001, Sigman et al. 2001). This occurs because natural scenes contain many contours that tend to be relatively smooth and to have significant spatial extent (e.g., see **Figure 2**). Similarly, if there is a large response from an oriented edge sensor, then it is likely that a neighboring parallel (but not colinear) edge sensor will have a large response (Geisler et al. 2001). This occurs because there is much parallel structure in natural images (e.g., see branches and leaf markings in **Figure 2**). Importantly, however, even when there is no correlation between the responses of a pair of sensors, the responses are often statistically dependent. For example, **Figure 5C** shows the distribution of responses of an orientation-selective sensor conditional on the response of another nonoverlapping orientation selective sensor at a nearby location (Schwartz & Simoncelli 2001). The responses are uncorrelated, but the variance of the response of one sensor (RF2) increases as function of the magnitude of the response of the

other sensor (RF1). In other words, strong features tend to cluster in natural images; thus, when a strong response occurs in a local sensor, the responses tend to be strong in other nearby sensors, although the sign of the response may be random (as in **Figure 5C**).

Another strategy for characterizing spatial structure in natural images is to measure general properties of the joint probability distribution of pixels within a spatial neighborhood of a given size. Two popular approaches have been to apply either PCA or independent components analysis (ICA) to a collection of patches sampled from natural images (for another approach, see Lee et al. 2003). If the patch size is, say, 12×12 pixels, then (ignoring color) each patch can be represented as a point in a 144-dimensional space where each dimension is the luminance at one of the 144 pixel locations. Applying PCA to natural image patches shows that a large number of principal axes are required to capture a substantial fraction of the variance in the distribution of image patches, and hence PCA does not lead to a compact summary of the spatial structure of natural images in the same way it does for natural irradiance and reflectance spectra.

ICA is a conceptually different approach. Rather than assuming, like PCA, that the whole space is described by a single multidimensional Gaussian distribution, ICA assumes that the space is described by a sum of statistically independent distributions, each representing a different unknown "source." The aim of ICA is to estimate, from a large collection of samples, the vector corresponding to each source (the direction of the primary axis of each source). This is an interesting approach because images probably are the result of independent physical sources (e.g., contours produced by different surface boundaries, surface markings, and shadows). ICA may discover some of those sources. Furthermore, the filters obtained from ICA, which recover (measure) the sources in an image, might be plausible candidates for neural receptive fields. Applying ICA to natural image patches yields filters similar in appearance to receptive fields of simple cells found in the visual cortex of mammals (Bell & Sejnowski 1997, Hyvarinen & Hoyer 2000, van Hateren & van der Schaaf 1998). (For ICA of natural auditory stimuli, see Lewicki 2002.) A conceptually related analysis producing similar results involves estimating independent sources, with the additional constraint that the sources are "sparse," in the sense that when a given source has a large value, other potential sources are constrained to have small values (Olshausen & Field 1997).

The studies of spatial structure described so far concern within-domain statistics, which are particularly relevant to issues of image coding (see below). There have also been attempts to measure across-domain statistics, which are more relevant to the performance of specific tasks. For example, Elder & Goldberg (1998, 2002) and Geisler et al. (2001, Geisler & Perry 2006) measured the pair-wise statistics of image contours that were hand segmented from natural images. Geisler et al. (2001, Geisler & Perry 2006) used an automatic algorithm to detect small edge elements from natural images and then had observers assign edge elements to physical contours in the image. Thus, with respect to the notation in **Figure 1A**, the state of the environment ω could take on one of two values ($\omega = c$ if two edge elements came from the same physical contour and $\omega = \sim c$ if two elements came from different physical contours). For each pair of edge elements, they determined the distance d between the elements, the direction θ of one of the elements with respect to the other element, the difference in orientation θ between the elements, and the difference in contrast polarity (same or opposite) ρ between the elements (**Figure 5D**). Thus, with respect to the notation in **Figure 1A**, the retinal image properties were defined by the vector $\mathbf{s} = (d, \phi, \theta, \rho)$. **Figure 5D** plots the ratio of the measured likelihood distributions for the cases where edge element pairs belong to the same contour and different contours:

ICA: independent components analysis

Access provided by New York University - Bobst Library on 08/19/15. For personal use only.

$l(d, \phi, \theta, \rho) = p(\mathbf{s}|\omega = c)/p(\mathbf{s}|\omega = \sim c)$. As can be seen, edge elements in natural images that are cocircular (which includes colinear) and of the same contrast polarity are more likely to come from the same physical contour; edge elements that deviate substantially from cocircularity or are of opposite polarity are likely to come from different physical contours. (Note, however, that even when the polarity is opposite, nearby edge elements are more likely to come from the same contour if they are approximately colinear; this presumably occurs because physical contours often cross backgrounds that modulate substantially in intensity.) These results have direct relevance for understanding the perceptual mechanisms underlying tasks such as contour grouping and contour completion.

Range

An important task for visual systems is estimating distance and three-dimensional shape from the two-dimensional images formed on each retina. A relevant statistic is the distribution of distances (ranges) in natural environments. **Figure 6A** shows a range image (over an extent of 259° horizontal and 80° vertical) measured in a forest environment with a laser range finder. In this image, lighter pixels denote greater distances. The solid curve in **Figure 6B** shows the probability of each
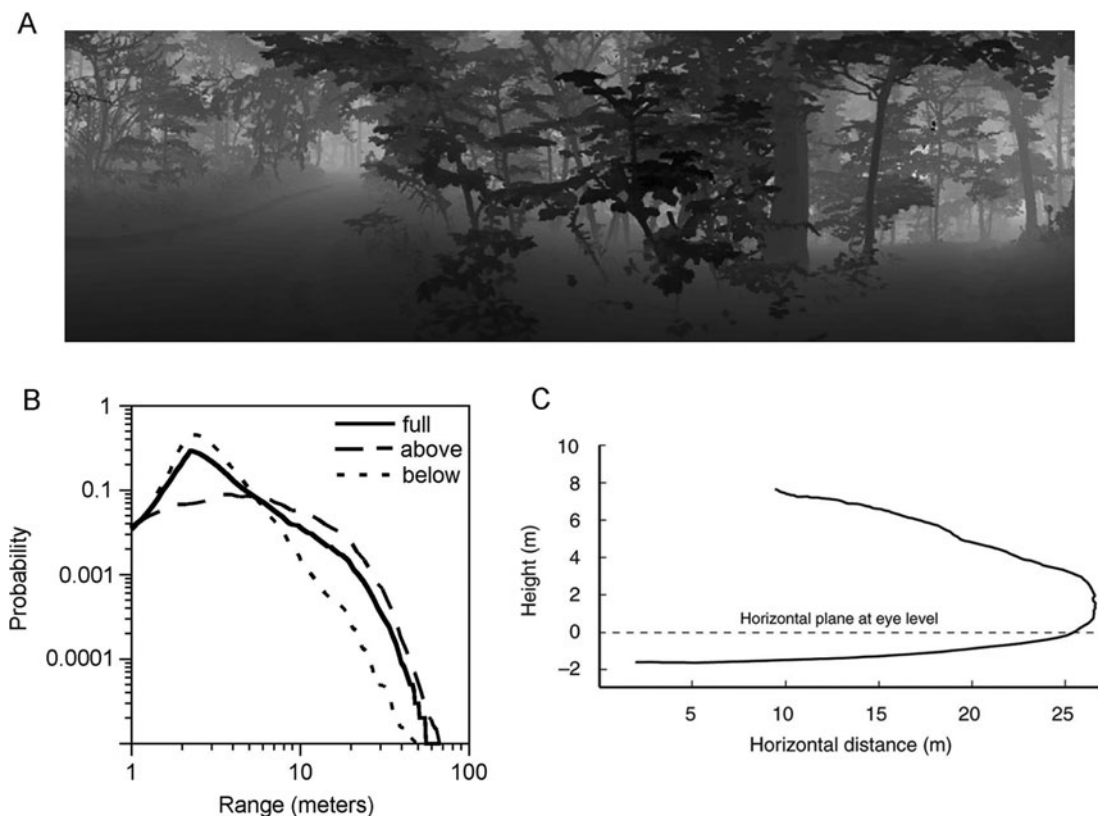
Figure 6

Range properties of forest environments. (*A*) Range image—the lighter the color, the greater the distance. (*B*) Histogram of range based upon 54 forest range images (adapted from Huang et al. 2000). (*C*) Average range as a function of elevation based upon a combination of 23 forest and 51 outdoor (Duke campus) range images (adapted from Yang & Purves 2003).

distance based upon 54 such forest range images (Huang et al. 2000). The distribution of distances is different above (dashed curve) and below (dotted curve) the line of sight. **Figure 6C** shows that average distance decreases rapidly below the line of sight due to the effect of the ground plane and decreases more gradually above the line of sight presumably due to the foliage canopy (Yang & Purves 2003). Huang et al. (2000) showed that their observed distributions of range in forest scenes can be approximated by a model consisting of a flat ground plane populated by a spatial Poisson distribution of cylinders (trees).

When coupled with other measurements, range statistics are of relevance for understanding depth and motion coding. For example, **Figure 7** shows the distribution of binocular disparity implied by the statistics of Yang & Purves (2003) when combined with the distribution of human fixation points measured as observers walk through forest environments (Cormack et al. 2005). Most binocular disparities fall within a modest range of ±1.5 degrees. Potetz & Lee (2003) measured range images along with coregistered grayscale images and discovered several modest correlations (e.g., range and luminance are negatively correlated).

Spatiotemporal Structure

The spatial structure of a retinal image changes over time due to self motion and to motion in the environment. The dynamical properties of natural retinal images are important to characterize because they contain useful information for many perceptual tasks, including heading estimation, image segmentation, distance estimation, and shape estimation. However, measuring and characterizing the spatiotemporal statistics of natural images are difficult because there is almost always some component of image motion that results from self-motion (i.e., the receptor array is almost always translating and/or rotating because of eye, head, and body movements).
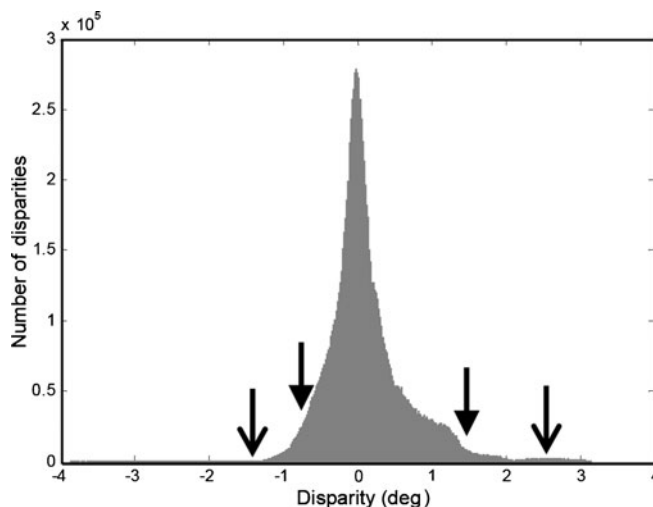

Figure 7

Distribution of binocular disparity for human observers when walking in forest environments. The solid arrows mark the upper and lower 2.5% quantiles; the open arrows mark the upper and lower 0.5% quantiles.

Thus, normal retinal image dynamics cannot be fully measured by a fixed camera or a camera attached to the head or body. Rather, one would like to move the image plane of the camera along the same trajectories typically undergone by the receptor array. This has been done for a flying insect by recording flight path and body orientation and then moving a camera along the same path with a robotic gantry (Boeddeker et al. 2005), but it has not been done for humans or for other mammals.

Nonetheless, systematic results have been obtained by analyzing video clips from movies and handheld cameras. For example, the symbols in **Figure 8** show the average spatiotemporal power spectra for natural image video reported by Dong & Atick (1995a). Contrast power decreases smoothly with increases in either spatial frequency or temporal frequency. The authors find the pattern of results can be fitted approximately (solid curves) by modeling the world as a collection of patches of spatial $1/f$ noise that are each undergoing translation at some random velocity. This is a plausible model for this image statistic because even nontranslational motion flow
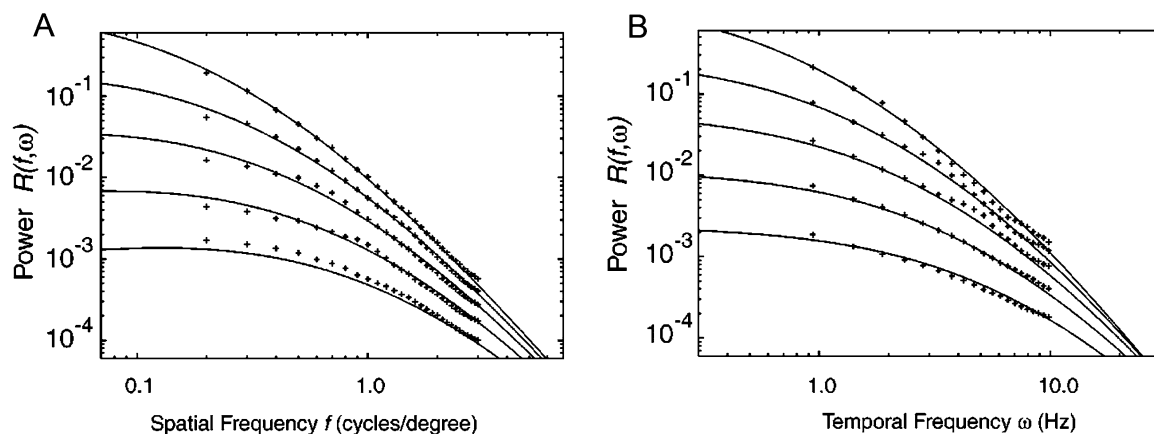
Figure 8

Spatiotemporal power spectra of natural images. (*A*) Spatial power spectra measured at different temporal frequencies (1.4, 2.3, 3.8, 6, and 10 Hz, from top to bottom). (*B*) Temporal spectra measured at different spatial frequencies (0.3, 0.5, 0.8, 1.3, and 2.1 cpd, from top to bottom). The solid curves show the expected spectra if the world is modeled as collection of patches of spatial $1/f$ noise that are each undergoing translation locally at some random velocity. (Adapted from Dong & Atick 1995a.)

patterns, such as those produced by moving through the three-dimensional environment, will most often produce approximately translational motion in a small neighborhood over a small time period.

These within-domain statistics are relevant for understanding motion coding. To measure across-domain statistics relevant for tasks such as heading estimation, it would also be necessary to measure the range at each pixel, as the image plane moves through space (Roth & Black 2005, Tversky & Geisler 2007).

Eye Movements and Foveation

The spatial resolution of the primate retina is high in the center of the fovea and falls off smoothly, but rapidly, as a function of retinal eccentricity. Information is collected from the environment by coupling this variable-resolution retina with a motor system that can rapidly direct the eyes in arbitrary directions, and indeed, most natural visual tasks involve selecting a series of locations to fixate. This implies that the relevant across-domain statistics for such tasks must take into account the foveated spatial resolution of the retina.

In other words, we need to measure the statistical relationship between environmental properties and properties of the retinal output and/or between properties of the retinal image and retinal output. In the terminology of **Figure 1***A* this means measuring $p(\mathbf{\omega}, \mathbf{z})$ and/or $p(\mathbf{s}, \mathbf{z})$, where $\mathbf{z}$ represents the specific properties of the retinal output that are of interest.

To make such measurements, Raj et al. (2005) modeled the spatial resolution (transfer function) of the human retina with a human contrast sensitivity function, which is also consistent with primate ganglion cell density and receptive field size as a function of eccentricity (Geisler & Perry 1998). They then considered the task of selecting fixations to maximally reduce total uncertainty about local contrast in natural images. The relevant across-domain statistics for this task are conditional probability distributions describing the probability of each possible local image contrast given a particular local contrast observed in the retinal output. **Figure 9** shows four of these conditional distributions. Note that the mode and variance of the distributions increase as function of both the retinal
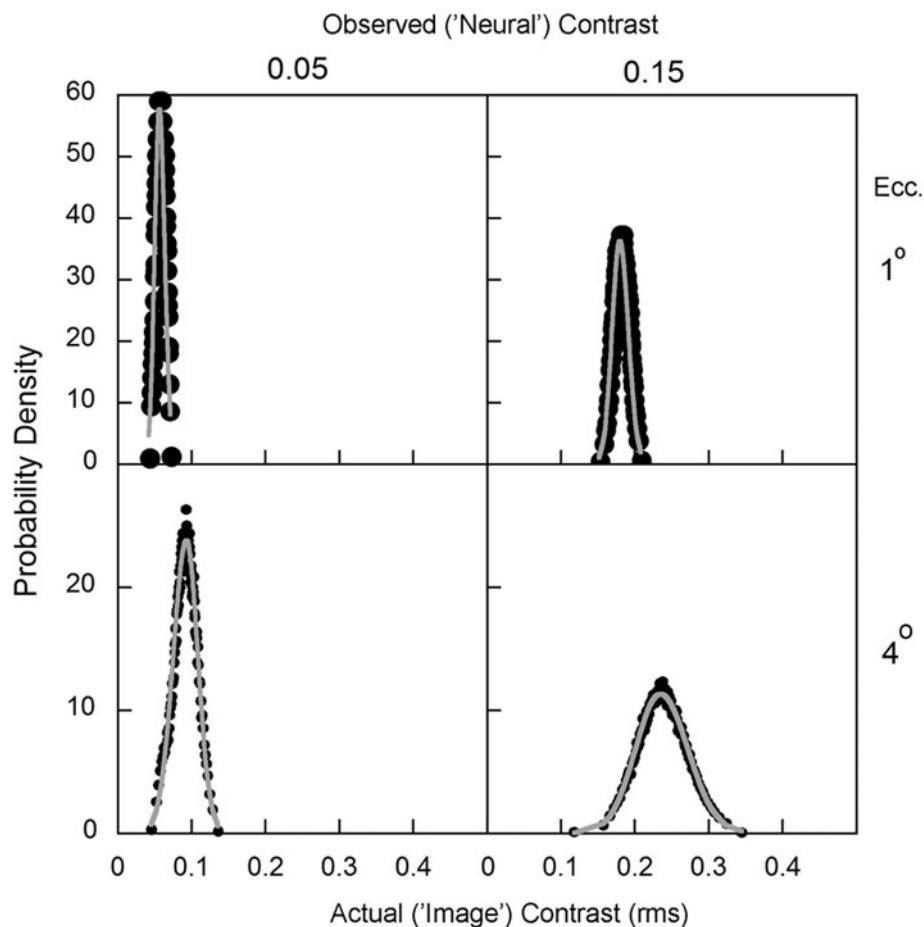
Figure 9

Example conditional probability distributions of local retinal image contrast given an eccentricity and an observed contrast at the output of a foveated retina, which was modeled after the primate retina. Such statistics are relevant for understanding how a rational visual system with a foveated retina should select fixation locations in order to accurately encode the image.

eccentricity and contrast observed in the retinal output. The increase in variance with eccentricity is intuitive because spatial resolution is decreasing with eccentricity. More surprising is that variance increases with observed contrast in the retinal output: the greater the response in the retinal output, the greater the uncertainty about the contrast in the retinal image. This finding may be related to the finding of Schwartz & Simoncelli (2001) that strong features tend to cluster in natural images (**Figure 5C**).

EXPLOITING NATURAL SCENE STATISTICS

The previous section summarized some of the measured natural scene statistics relevant for visual perception. This section considers how some of those statistical properties might be exploited by the human (or nonhuman primate) visual system. One aim is to convince the reader that measuring natural scene statistics can be useful for generating quantitative, testable hypotheses about perceptual mechanisms, can be useful for designing

experiments and stimuli, and can provide useful insight into design features of the visual system. This section focuses on a few examples from the topics of coding and representation, grouping and segregation, identification, and estimation.

Coding and Representation of the Visual Image

Within-domain statistics are the easiest to measure and are of particular relevance to coding and representation of the retinal image (the proximal stimulus). Hence more studies of natural scene statistics have been devoted to this topic than to any other. A central hypothesis about coding and representation is that neural resources are limited (in number of neurons, spikes per neuron, synaptic contacts per neuron, etc.), thus pushing visual systems to efficiently use those resources (Attneave 1954, Barlow 1961; for reviews, see Barlow 2001, Simoncelli 2003, Simoncelli & Olshausen 2001).

One important resource limitation is that neurons have a limited dynamic range, presumably for fundamental metabolic and biophysical reasons. A plausible hypothesis is that sensory neurons match their limited ranges to the dynamic range of the natural signals that they encode. This ensures that the full response range is used while minimizing the likelihood of overdriving or underdriving the neuron. One way to match the dynamic range of responses with that of the input signals is via histogram equalization: Adjust the shape of the neuron's response function so that all response levels occur equally often under natural stimulus conditions. Laughlin (1981) compared the probability distribution of local luminance (Weber contrast) in natural images with the luminance response functions of the large monopolar neurons in the eye of the blowfly and found that the response functions are consistent with the histogram equalization hypothesis. Subsequently, the hypothesis has been tested for the contrast responses of neurons in lateral geniculate nucleus (LGN) and

primary visual cortex (V1) of cats and primates (e.g., Brady & Field 2000, Clatworthy et al. 2003, Tadmor & Tolhurst 2000). Although the results depend somewhat on the specific definition of contrast used (Frazor & Geisler 2006), there is evidence for a rough match between the contrast response functions of neurons in the early visual system and the distribution of contrasts encountered by the eye in natural environments.

For natural images, there is a low correlation between luminance and contrast at the same retinal location (see **Figure 3C,D**); furthermore, it can be shown that for normal saccadic inspection of natural images there is little correlation in luminance at the same retinal location across fixations (Frazor & Geisler 2006) or in contrast at the same retinal location across fixations (Frazor & Geisler 2006, Reinagel & Zador 1999). The implication is that neurons whose receptive fields are spatially localized will typically receive random samples of luminance and contrast from distributions like those in **Figure 3**, several times per second. This fact raises questions about the coding of luminance and contrast in the visual system. The most obvious is how neurons in the visual system respond to simultaneous, frequent, and statistically independent variations in local luminance and contrast over their natural ranges. Geisler et al. (2007) measured response functions of individual neurons in V1 and found them to be separable in luminance and contrast; i.e., response as a function of luminance and contrast is the product of a single function for luminance and a single function for contrast: $r(C, L) = r_C(C)r_L(L)$. Similarly, Mante et al. (2005) measure responses of individual neurons in the LGN and found responses to be consistent with separable luminance and contrast gain control mechanisms. Separable responses are expected under the efficient coding hypothesis, if local luminance and contrast are statistically independent.

The results of Geisler et al. (2007) point to a new hypothesis about how local luminance is coded in the cortex. The classic hypothesis

is that most neurons in V1 respond poorly to luminance, and that luminance is coded by a specialized set of luminance-responding cells (for a review, see Peng & Van Essen 2005). Although most cortical neurons do not respond to uniform luminance stimulation, by parametrically varying luminance and contrast, Geisler et al. (2007) found that local luminance strongly modulates contrast response in the same separable fashion as other stimulus dimensions such as orientation, spatial frequency, and direction of motion. Thus, local luminance appears to be coded in the cortex in the same fashion as other well-known stimulus dimensions. Specialized luminance-responding cells may code uniform areas such as patches of sky. The studies of Mante et al. (2005) and Geisler et al. (2007) were motivated by measured natural scene statistics, and hence they demonstrate the potential value of natural scene statistics in guiding neurophysiology.

Another form of histogram equalization is to match the tuning characteristics across a population of neurons to the distribution of natural signals. For example, Cormack et al. (2005) found that the probability distribution of binocular disparities that occur during navigation through forest environments by humans (see **Figure 7**) corresponds reasonably well with the distribution of preferred disparities of single neurons in area MT of the macaque monkey (DeAngelis & Uka 2003). The results of Cormack et al. (2005) suggest that disparity tuning of neurons in monkey visual cortex conforms (in at least some ways) to an efficient coding of natural binocular images.

Another way to use neural resources efficiently is to remove redundant information across a population of neurons. A classic example concerns the coding of chromatic information with opponent color mechanisms. The spectral sensitivities of the L, M, and S cones overlap substantially (especially the L and M cones), creating highly correlated responses. Thus, if one observes a large response from, say, an L cone, it is very likely that a large re-

sponse will be observed from a spatially adjacent M cone. Rather than represent these two large responses with two high spike rates, it is more efficient (in terms of utilizing neural dynamic range) to transform the cone responses so that they are statistically independent. Interestingly, applying such a transformation predicts chromatic receptive fields that are similar to the color opponent mechanisms that have been estimated from psychophysical studies (Buchsbaum & Gottschalk 1983, Ruderman et al. 1998; see **Figure 4**). Similarly, spatial and temporal decorrelation (whitening) of the receptor responses to natural images predicts spatial (Atick & Redlich 1992, Srinivasan et al. 1982) and spatiotemporal receptive field shapes (Dong & Atick 1995b) similar to those found in the retina and LGN. Thus, in a number of ways, the design of the retina seems to be consistent with the efficient coding hypothesis.

The evidence is less clear with respect to coding and representation in V1. V1, like the retina and LGN, is laid out in a topographic map; however, for each ganglion cell or LGN relay neuron there are hundreds of V1 neurons, and thus V1 could potentially contain a large number of lossless (and efficient) representations of the retinal output, each tailored to a different task. Another hypothesis is that V1 provides a sparse, statistically independent representation of the retinal output. Field (1987) noted that (linear) receptive fields similar to those measured in V1 respond very infrequently to natural images and produce relatively large responses when they do respond. In other words, the probability distribution of responses to natural images is highly peaked at zero and has heavy tails (see **Figure 5B**). The implication is that a natural image will produce a pattern of strong responses in a rather sparse subset of cortical neurons. Olshausen & Field (1997) showed that simultaneously optimizing sparseness and statistical independence in the responses to natural images yields receptive fields similar to those of V1 neurons, suggesting that a sparse, statistical independent representation may be the goal of

V1 coding. However, this begs the question of the functional advantages of a sparse code. Possibly a sparse code provides a more meaningful (immediately interpretable) representation of the image than would highly compact codes, or possibly a sparse code facilitates subsequent processing (Olshausen 2003). Other possibilities are that a sparse code, which concentrates strong activity in a few neurons, might be less susceptible to neural noise arising in later cortical areas or might consume less metabolic energy than more distributed codes.

Interestingly, Schwartz & Simoncelli (2001) show that contrast normalization, an important nonlinear response property of V1 neurons (Albrecht & Geisler 1991, Carandini et al. 1997, Heeger 1991), should increase statistical independence by largely eliminating variance dependences such as the one illustrated in **Figure 5C**. However, contrast normalization has other potentially functional advantages that are not obviously related to efficient coding or natural image statistics, advantages such as contrast-invariant feature tuning (Albrecht & Geisler 1991, Heeger 1991) and enhanced feature identification performance (Geisler & Albrecht 1995). Perhaps all of these advantages have contributed to the evolution of contrast normalization mechanisms.

Other nonlinear response properties of retinal and cortical neurons have been inferred from measurements of their responses to natural images. In general, receptive fields estimated with linear systems analysis techniques (e.g., spike-triggered averaging) do not make accurate predictions of responses to natural images. There is not space here to review the literature directed at measuring and characterizing responses of neurons to natural stimuli; however, for recent reviews, see Reinagel (2001) and Wu et al. (2006), and for discussion, see Felson & Dan (2005) and Rust & Movshon (2005).

In addition to removing redundant information from the image, the retina removes nonredundant information by having a highly foveated retina with many fewer receptors than needed to fully code the image, and even fewer ganglion cells than receptors. The visual system compensates for the reduction in number of receptors and ganglion cells by having an eye movement system that can rapidly point the eye in desired directions. Thus, fully encoding the retinal image requires making a series of fixations. One hypothesis is that when the task is specifically to encode and remember an image, humans make eye movements that acquire as much image information as possible with the fewest number of fixations. Across-level statistics, such as those in **Figure 9**, can be combined with a Bayesian ideal observer analysis to determine the optimal procedure for selecting fixations in natural scenes. Raj et al. (2005) determined how to select successive fixations that maximally reduce the total uncertainty (entropy) about the contrast at every location in the image (see also Renninger et al. 2005). They then showed that the fixations selected by this algorithm are also near optimal for reducing total uncertainty about the detailed structure of the image. It remains to be seen how well human eye movements match the optimal eye movements, but it is likely that humans display (qualitatively) some of the optimal behaviors, which include moderate length saccades, avoidance of very-low-contrast regions of the image, a moderate percentage of fixations near high-contrast features, and avoidance of fixations near the boundaries of the image.

Grouping and Segregation

Efficiently encoding and representing the images falling on the retina may be the primary goal of initial visual processing, especially in the human retina, because the optic nerves create a severe information-transmission bottleneck and because the information transmitted up the optic nerve must support a very wide range of perceptual tasks. On the other hand, central processing is more likely to reflect specific tasks. The mechanisms of perceptual grouping and segregation fall into

this category. Their purpose is undoubtedly to group together image features that arise from the same physical source in the environment (e.g., the same surface or object) and segregate features that arise from different physical sources. In other words, grouping and segregation mechanisms are designed to make inferences about the environment from image information collected by the eyes, and hence across-domain statistics are particularly relevant.

Geisler et al. (2001) used the across-domain statistics in **Figure 5D** (averaged over contrast polarity) as the basis for a one-parameter model of contour grouping in random contour-element displays. In the model, contour elements are grouped together when the likelihood ratio given by **Figure 5D** exceeds a criterion. This model, based directly on natural scene statistics, was able to predict human contour grouping performance under a wide range of stimulus conditions. More recently, Geisler & Perry (2006) used the statistics in **Figure 5D** to derive a parameter-free Bayesian ideal observer for a contour completion task where the observer reports whether or not contour segments passing under an occluding surface belong to the same physical contour (source). The results showed that humans parallel optimal performance in all conditions and perform only slightly below optimal. These two studies strongly suggest that the human visual system incorporates and properly exploits the across-domain statistics represented in **Figure 5D**.

Fine et al. (2003) took an interesting approach to measuring across-domain color statistics relevant for region grouping/segregation. They assumed that the probability distribution of color differences ($\mathbf{s} = \Delta I, \Delta\alpha, \Delta\beta$) between adjacent pixels approximates the distribution of color differences between pixels from the same physical surface ($\omega = same$), and that the probability distribution of color differences between pixels taken from different natural images approximates the distribution of color differences for pixels from different physical surfaces ($\omega =$

different). (See **Figure 4** for the definition of I, α, β.) Starting from these distributions, they derived a Bayesian decision rule for segmenting pixels into regions. They compared segmentations using this decision rule with those of human observers and found a fairly high correlation. Although it is likely that nearby pixels usually belong to the same physical surface, their implicit assumption that the distribution of color differences within a surface does not vary with distance between pixels probably doesn't hold. Analyzing many hand-segmented images like the one in **Figure 2**, Wilson et al. (2006) found that the color difference distributions within surfaces vary substantially with distance. Nonetheless, the two studies show there is much useful information for region grouping and segregation in the across-domain statistics of color differences.

Identification

Although the retina may have evolved primarily to efficiently encode the retinal images, as mentioned above, there can be no doubt (even for general-purpose organisms like humans, who perform a wide range of tasks) that the design of the eye is constrained by the family of tasks the organism performs. Further, there may be specific aspects of retinal processing that are tailored to specific sources in the environment. For example, measurements of across-domain statistics suggest that the positioning of the cone spectral sensitivities along the wavelength axis may be optimal for identifying sources of food. Lythgoe & Partridge (1989) showed that in mammals, the S and M cones (most mammals have only these two cone types) are generally well positioned to discriminate between different kinds of foliage. Similarly, Osorio & Vorobyev (1996), Regan et al. (2001), and Parraga et al. (2002) show that the L and M cones in trichromatic primates are often well positioned for identifying fruit in foliage.

Identification of behaviorally relevant sources in the environment, such as specific

materials, surfaces, and objects, strongly depends on grouping and segregation mechanisms. Conversely, grouping and segregation undoubtedly depend on certain top-down and bottom-up identification mechanisms. Thus, grouping, segregation, and identification are best thought of as part of an integrated system for interpreting retinal images. A key goal of this integrated system is identifying the source of an image contour. In natural images, a contour could be the result of a surface boundary, a surface marking (i.e., a surface reflectance change), a cast shadow, or a shading boundary due to a change in surface orientation relative to the illumination. These are very different sources and their proper identification is crucial for arriving at the correct physical interpretation of the image. Measurements of across-level statistics could give us principled hypotheses for the brain mechanisms that identify contour sources given a retinal image.

To make a start at these measurements, Ing & Geisler (2006) analyzed a large set of hand-segmented, close-up foliage images like the one in **Figure 2**. They chose this class of images because they are relatively easy to segment and because foliage environments are the natural environment of the macaque monkey (the primary animal model for human vision). In addition to hand-segmenting leaves and branches, they also segmented shadow boundaries and surface markings. Using this segmentation data as the ground truth measurements of the sources, they measured the joint probability distribution of intensity and contrast differences (in I, α, β space; see **Figure 4**) across contours, given that the source contour was a surface boundary, a shadow boundary, or a surface-marking boundary. Finally, by combining these measured probability distributions with an ideal classifier, they demonstrated that for moderate length contours it is possible to discriminate between any two sources with 85% to 90% accuracy. This is a promising initial result because it shows that simple local image properties alone provide useful information for contour source identification. Further, the simple decision rules derived from the measured distributions provide concrete hypotheses for source identification mechanisms in the visual system, although global and top-down factors are also likely to contribute.

Estimation

Many natural perceptual tasks involve estimation of continuous environmental properties such as the reflectance spectrum, illumination spectrum, surface orientation, distance, or velocity, at one or more locations in the visual field. Across-level statistics are providing new hypotheses and valuable insights into such estimation tasks.

Under the assumption that retinal images can be approximated as the combination of a randomly selected natural illuminant and a collection of surfaces, each with a randomly selected natural reflectance spectrum, Maloney & Wandell (1986) showed it is possible to estimate the illuminant and reflectance spectra of surfaces with remarkably few classes of receptor (approximately four). (Note that estimating the reflectance function is equivalent to solving the color constancy problem.) This result may help explain why most organisms have relatively few classes of receptor. Subsequently, Brainard & Freeman (1997) described how to optimally estimate reflectance and illumination within the Bayesian ideal observer framework. Although these and similar studies provide valuable insight into the problem of color constancy, they assume uniform illumination and matt surface patches, neither of which occur very frequently in natural scenes (e.g., see Boyaci et al. 2006, Dror et al. 2004, Fleming et al. 2003, Khang et al. 2006). Thus, reflectance estimation given real illumination patterns and surface properties requires more complex perceptual mechanisms than the ones suggested by earlier work. However, real illumination and surface complexity may also provide additional information the visual system can use. For example, Sharan et al. (2005) and Motoyoshi et al. (2007) show

that humans can judge (with reasonable accuracy) the reflectance of natural materials from grayscale images that have been equated for mean luminance. Statistical analysis of the images shows that the shape of the pixel luminance distribution varies systematically with the reflectance of the material and that humans are able to exploit these statistics (and possibly other image statistics) when judging reflectance.

When retinal image information is poor, a rational visual system will put greater reliance on the prior probability distributions of different possible environmental states and bias its estimates accordingly (e.g., see Knill & Richards 1996, Torralba 2003). Weiss et al. (2002) show this principle could explain various motion illusions, under the assumption that the prior probability for local speed decreases monotonically with speed. This general assumption is likely to be true, but the specific predictions depend upon the shape of the prior probability distribution, which they did not measure (however, see Stocker & Simoncelli 2006). Yang & Purves (2003) applied a similar analysis in an attempt to predict apparent distance illusions. Using statistical measurements obtained with a range finder, they were able to qualitatively account for several distance-estimation biases that have been reported in the literature.

To circumvent some of the difficulties associated with measuring across-domain motion statistics, Tversky & Geisler (2007) combined graphics simulations with measured natural scene statistics. Specifically, they created model environments based on the measured range statistics of Huang et al. (2000; see **Figure 6B**) and the local $1/f$ statistics of natural image patches (**Figure 5A**). They then simulated various kinds of self motion through these environments and measured the scene statistics produced by these motions. These statistics were used to determine optimal integration area (receptive field size) of local motion sensors for heading estimation. The primary finding is that integration area should increase with the speed being estimated. This

hypothesis should be testable in physiological and psychophysical studies. Similar statistical measurements could potentially be used to estimate the approximate prior probability distribution of ground truth velocities needed for analyses like that of Weiss et al. (2002).

CONCLUSION

In the traditional approach to perception research, the scientist (*a*) thinks informally or casually about natural tasks and environments, (*b*) generates or modifies hypotheses about perceptual or neural mechanisms, (*c*) conducts controlled behavioral or physiological experiments to test those hypotheses and then cycles back to (*b*). Although there is nothing wrong with this approach (which has produced a vast amount of knowledge), a potential weakness is that hypotheses and experimental paradigms tend to be formulated from informal thinking about natural environments and stimuli rather than from principled physical measurements. By directly measuring statistical regularities of natural environments and stimuli it may be possible to derive novel and plausible hypotheses for perceptual mechanisms and to design experimental paradigms that better reflect the important characteristics of natural stimuli. Indeed, the studies described here demonstrate the value of measuring and characterizing natural scene statistics. Measurements of within-domain statistics have revealed much about the structure and variability of natural images. This has made it possible to test various forms of the efficient coding hypothesis and has led to novel models and experiments that would not have been conceived otherwise. Measurements of across-domain statistics, especially when combined with Bayesian ideal observer theory, are proving to be particularly useful in generating new insights and testable (parameter-free or nearly parameter-free) models for performance in tasks such as fixation selection, contour grouping, contour classification, motion estimation, and reflectance estimation.

Measuring and characterizing the statistical properties of natural environments and stimuli are difficult tasks, but the potential for payoff appears to be great.

ACKNOWLEDGMENTS

LITERATURE CITED

Albrecht DG, Geisler WS. 1991. Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Vis. Neurosci.* 7:531–46

Atick JJ, Redlich AN. 1992. What does the retina know about natural scenes? *Neural Comput.* 4:196–210

Attneave F. 1954. Some informational aspects of visual perception. *Psychol. Rev.* 61:183–93

Balboa RM, Tyler CW, Grzywacz NM. 2001. Occlusions contribute to scaling in natural images. *Vision Res.* 41:955–64

Barlow HB. 1957. Increment thresholds at low intensities considered as signal/noise discriminations. *J. Physiol. Lond.* 136:469–88

Barlow HB. 1961. Possible principles underlying the transformations of sensory messages. In *Sensory Communication*, ed. WA Rosenblith, pp. 217–34. Cambridge, MA: MIT Press

Barlow HB. 2001. Redundancy reduction revisited. *Network* 12:241–53

Barlow HB, Levick WR. 1969. Three factors limiting the reliable detection of light by retinal ganglion cells of the cat. *J. Physiol.* 200:1–24

Bell AJ, Sejnowski TJ. 1997. The "independent components" of natural scenes are edge filters. *Vision Res.* 37:3327–38

Boeddeker N, Lindemann JP, Egelhaaf M, Zeil J. 2005. Responses of blowfly motion-sensitive neurons to reconstructed optic flow along outdoor flight paths. *J. Comp. Physiol. A* 191:1143–55

Boyaci H, Doerchner K, Snyder JL, Maloney LT. 2006. Surface color perception in three-dimensional scenes. *Vis. Neurosci.* 23:311–21

Brady N, Field DJ. 2000. Local contrast in natural images: normalization and coding efficiency. *Perception* 29:1041–55

Brainard DH, Freeman WT. 1997. Bayesian color constancy. *J. Opt. Soc. Am. A* 14:1393–411

Brunswik E. 1956. *Perception and the Representative Design of Psychological Experiments*. Berkeley: Univ. Calif. Press

Brunswik E, Kamiya J. 1953. Ecological cue-validity of "proximity" and other Gestalt factors. *Am. J. Psychol.* 66:20–32

Buchsbaum G, Gottschalk A. 1983. Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proc. R. Soc. Lond. B Biol. Sci.* 220:89–113

Burton GJ, Moorehead IR. 1987. Color and spatial structure in natural scenes. *Appl. Opt.* 26:157–70

Carandini M, Heeger DJ, Movshon JA. 1997. Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci.* 17:8621–44

Clatworthy PL, Chirimuuta M, Lauritzen JS, Tolhurst DJ. 2003. Coding of the contrasts in natural images by populations of neurons in primary visual cortex (VI). *Vis. Res.* 43:1983–2001

Cormack LK, Liu Y, Bovik AC. 2005. Disparity statistics in the natural environment. *J. Vis.* 5(8):604 (Abstr.)

Cronly-Dillon JR, Gregory RL, eds. 1991. *Evolution of the Eye and Visual System, Vol. 2*. Boca Raton, FL: CRC Press

DeAngelis GC, Uka T. 2003. Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *J. Neurophysiol.* 89:1094–111

Dixon RE. 1978. Spectral distribution of Australian daylight. *J. Opt. Soc. Am.* 68:437–50

Dong DW, Atick JJ. 1995a. Statistics of natural time-varying images. *Netw. Comput. Neural Syst.* 6:345–58

Dong DW, Atick JJ. 1995b. Temporal decorrelation: a theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Netw. Comput. Neural Syst.* 6:159–78

Dror RO, Willsky AS, Adelson EH. 2004. Statistical characterization of real-world illumination. *J. Vis.* 4:821–37

D'Zmura M, Iverson G. 1993. Color constancy. I. Basic theory of two-stage linear recovery of spectral descriptions for lights and surfaces. *J. Opt. Soc. Am. A* 10:2148–65

Elder JH, Goldberg RM. 1998. *The statistics of natural image contours*. Presented at 1998 IEEE Comput. Soc. Workshop Percept. Org. Comput. Vis., Santa Barbara, Calif.

Elder JH, Goldberg RM. 2002. Ecological statistics for the Gestalt laws of perceptual organization of contours. *J. Vis.* 2:324–53

Felson G, Dan Y. 2005. A natural approach to studying vision. *Nat. Neurosci.* 8:1643–46

Field DJ. 1987. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* 4:2379–94

Field DJ. 1994. What is the goal of sensory coding? *Neural Comput.* 6:559–601

Fine I, MacLeod DIA, Boynton GM. 2003. Surface segmentation based on the luminance and color statistics of natural scenes. *J. Opt. Soc. Am. A* 20:1283–91

Fleming RW, Dror RO, Adelson EH. 2003. Real-world illumination and the perception of surface reflectance properties. *J. Vis.* 3:347–68

Frazor RA, Geisler WS. 2006. Local luminance and contrast in natural images. *Vis. Res.* 46:1585–98

Geisler WS, Albrecht DG. 1995. Bayesian analysis of identification performance in monkey visual cortex: nonlinear mechanisms and stimulus certainty. *Vis. Res.* 35:2723–30

Geisler WS, Albrecht DG, Crane AM. 2007. Responses of neurons in primary visual cortex to transient changes in local contrast and luminance. *J. Neurosci.* 27:5063–67

Geisler WS, Diehl RL. 2003. A Bayesian approach to the evolution of perceptual and cognitive systems. *Cogn. Sci.* 27:379–402

Geisler WS, Perry JS. 1998. A real-time foveated multi-resolution system for low-bandwidth video communication. In *Human Vision and Electronic Imaging, SPIE Proceedings*, ed. B Rogowitz, T Pappas, 3299:294–305. Bellingham, WA: SPIE

Geisler WS, Perry JS. 2006. Efficiency of contour grouping across occlusions in natural images. *J. Vis.* 6:336 (Abstr.)

Geisler WS, Perry JS, Super BJ, Gallogly DP. 2001. Edge co-occurrence in natural images predicts contour grouping performance. *Vis. Res.* 41:711–24

Gibson JJ. 1966. *The Senses Considered as Perceptual Systems*. Boston, MA: Houghton-Mifflin

Gibson JJ. 1979. *The Ecological Approach to Visual Perception*. Boston, MA: Houghton-Mifflin

Green DM, Swets JA. 1966. *Signal Detection Theory and Psychophysics*. New York: Wiley

Hecht S, Shlaer S, Pirenne MH. 1942. Energy, quanta, and vision. *J. Gen. Physiol.* 25:819–40

Heeger DJ. 1991. Computational model of cat striate physiology. In *Computational Models of Visual Perception*, ed. MS Landy, AJ Movshon, pp. 119–33. Cambridge, MA: MIT Press

Huang J, Lee AB, Mumford D. 2000. Statistics of range images. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1: 324–31. Piscataway, NJ: IEEE

Hyvarinen A, Hoyer P. 2000. Emergence of topography and complex cell properties from natural images using extensions of ICA. In *Advances in Neural Information Processing Systems*, ed. SA Solla, TK Leen, K-R Muller, pp. 827–33. Cambridge, MA: MIT Press

Ing AD, Geisler WS. 2006. Ribbon analysis of contours in natural images. *J. Vis.* 6:103 (Abstr.)

Judd DB, MacAdam DL, Wyszecki GW. 1964. Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am.* 54:1031–40

Kersten D, Mamassian P, Yuille AL. 2004. Object perception as Bayesian inference. *Annu. Rev. Psychol.* 55:271–304

Khang BG, Koenderink JJ, Kappers AML. 2006. Perception of illumination direction in images of 3-D convex objects: influence of surface materials and light fields. *Perception* 35:625–45

Knill DC, Richards W, eds. 1996. *Perception as Bayesian Inference*. Cambridge, MA: Cambridge Univ. Press

Laughlin SB. 1981. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch. C* 36:910–12

Lee AB, Mumford D, Huang J. 2001. Occlusion models for natural images: a statistical study of a scale-invariant dead leaves model. *Int. J. Comput. Vis.* 41:35–59

Lee AB, Pedersen KS, Mumford D. 2003. Nonlinear statistics of high-contrast patches in natural images. *Int. J. Comput. Vis.* 54:83–103

Lewicki MS. 2002. Efficient coding of natural sounds. *Nat. Neurosci.* 5:356–63

Lythgoe JN. 1979. *The Ecology of Vision*. New York: Oxford Univ. Press

Lythgoe JN, Partridge JC. 1989. Visual pigments and the acquisition of visual information. *J. Exp. Biol.* 146:1–20

Maloney LT. 1986. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *J. Opt. Soc. Am. A* 3:1673–83

Maloney LT, Wandell BA. 1986. Color constancy: a method for recovering surface spectral reflectance. *J. Opt. Soc. Am. A* 3:29–33

Mante V, Bronin V, Frazor RA, Geisler WS, Carandini M. 2005. Independence of gain control mechanisms in early visual system matches the statistics of natural images. *Nat. Neurosci.* 8:1690–97

Martin DR, Fowlkes CC, Malik J. 2004. Learning to detect natural image boundaries using local brightness, color and texture cues. *IEEE Trans. Pattern Anal. Mach. Intell.* 26:530–49

Motoyoshi I, Nishida1 S, Sharan L, Adelson EH. 2007. Image statistics and the perception of surface qualities. *Nature* 447:206–9

Olshausen BA. 2003. Principles of image representation in visual cortex. In *The Visual Neurosciences*, ed. LM Chalupa, JS Werner, pp. 1603–15. Cambridge, MA: MIT Press

Olshausen BA, Field DJ. 1997. Sparse coding with an overcomplete basis set: a strategy by V1? *Vis. Res.* 37:3311–25

Osorio D, Vorobyev M. 1996. Colour vision as an adaptation to frugivory in primates. *Proc. R. Soc. Lond. B Biol. Sci.* 263:593–99

Parraga CA, Troscianko T, Tolhurst DJ. 2002. Spatiochromatic properties of natural images and human vision. *Curr. Biol.* 12:483–87

Peng X, Van Essen DC. 2005. Peaked encoding of relative luminance in macaque areas V1 and V2. *J. Neurophysiol.* 93:1620–32

Potetz B, Lee TS. 2003. Statistical correlations between two-dimensional images and three-dimensional structures in natural scenes. *J. Opt. Soc. Am. A* 20:1292–303

Raj R, Geisler WS, Frazor RA, Bovik AC. 2005. Contrast statistics for foveated visual systems: fixation selection by minimizing contrast entropy. *J. Opt. Soc. Am. A* 22:2039–49

Regan BC, Julliot C, Simmen B, Vienot F, Charles-Dominique P, Mollon JD. 2001. Fruits, foliage and the evolution of primate colour vision. *Philos. Tran. R. Soc. Lond. B* 356:229–83

Reinagel P. 2001. How do visual neurons respond in the real world? *Curr. Opin. Neurobiol.* 11:437–42

Reinagel P, Zador AM. 1999. Natural scene statistics at the centre of gaze. *Comput. Neural Syst.* 10:1–10

Renninger LW, Coughlan J, Verghese P, Malik J. 2005. An information maximization model of eye movements. In *Advances in Neural Information Processing Systems 17*, ed. LK Saul, Y Weiss, L Bottou, pp. 1121–28. Boston, MA: MIT Press

Roth S, Black MJ. 2005. On the spatial statistics of optical flow. *Proc. 10th IEEE Int. Conf. Comput. Vis.* 1:42–49. Boston, MA: IEEE

Ruderman DL. 1994. The statistics of natural images. *Comput. Neural Syst.* 5:517–48

Ruderman DL. 1997. Origins of scaling in natural images. *Vis. Res.* 37:3385–98

Ruderman DL. 2002. Letter to the editor. *Vis. Res.* 42:2799–801

Ruderman DL, Bialek W. 1994. Statistics of natural images: scaling in the woods. *Phys. Rev. Lett.* 73:814–17

Ruderman DL, Cronin TW, Chiao C. 1998. Statistics of cone responses to natural images: implications for visual coding. *J. Opt. Soc. Am. A* 15:2036–45

Rust NC, Movshon AJ. 2005. In praise of artifice. *Nature Neurosci.* 8:1647–50

Schwartz O, Simoncelli EP. 2001. Natural signal statistics and sensory gain control. *Nat. Neurosci.* 4:8 19–25

Sharan L, Li Y, Adelson EH. 2005. Image statistics and reflectance estimation. *J. Vis.* 5(8):375 (Abstr.)

Sigman M, Cecchi GA, Gilbert CD, Magnasco MO. 2001. On a common circle: natural scenes and Gestalt rules. *Proc. Natl. Acad. Sci. USA* 98:1935–40

Simoncelli EP. 2003. Vision and the statistics of the visual environment. *Curr. Opin. Neurobiol.* 13:144–49

Simoncelli EP, Olshausen BA. 2001. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* 24:1193–216

Srinivasan MV, Laughlin SB, Dubs A. 1982. Predictive coding: a fresh view of inhibition in the retina. *J. R. Soc. London Ser. B* 216:427–59

Stocker A, Simoncelli EP. 2006. Noise characteristics and prior expectations in human visual speed perception. *Nat. Neurosci.* 9:578–85

Tadmor Y, Tolhurst DJ. 2000. Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes. *Vis. Res.* 40:3145–57

Torralba A. 2003. Modeling global scene factors in attention. *J. Opt. Soc. Am. A* 20:1407–18

Tversky T, Geisler WS. 2007. Optimal sensor design for estimating local velocity in natural environments. In *Human Vision and Electronic Imaging, SPIE Proceedings*, ed. B Rogowitz, T Pappas, S Daly, Vol. 6492. Bellingham, WA: SPIE. In press

van Hateren JH. 1992. A theory of maximizing sensory information. *Biol. Cybern.* 68:23–29

van Hateren JH, van der Schaaf A. 1998. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. R. Soc. London B Biol. Sci.* 265:359–66

Walls GL. 1942. *The Vertebrate Eye and Its Adaptive Radiation*. Bloomfield Hills, Mich.: Cranbrook Inst. Sci.

Weiss Y, Simoncelli E, Adelson EH. 2002. Motion illusions as optimal percepts. *Nat. Neurosci.* 5:598–604

Wilson JA, Ing AD, Geisler WS. 2006. Chromatic differences within surfaces and across surface boundaries. *J. Vis.* 6(6):559 (Abstr.)

Wu MC-K, David SV, Gallant JL. 2006. Complete functional characterization of sensory neurons by system identification. *Annu. Rev. Neurosci.* 29:477–505

Yang Z, Purves D. 2003. A statistical explanation of visual space. *Nat. Neurosci.* 6:632–40

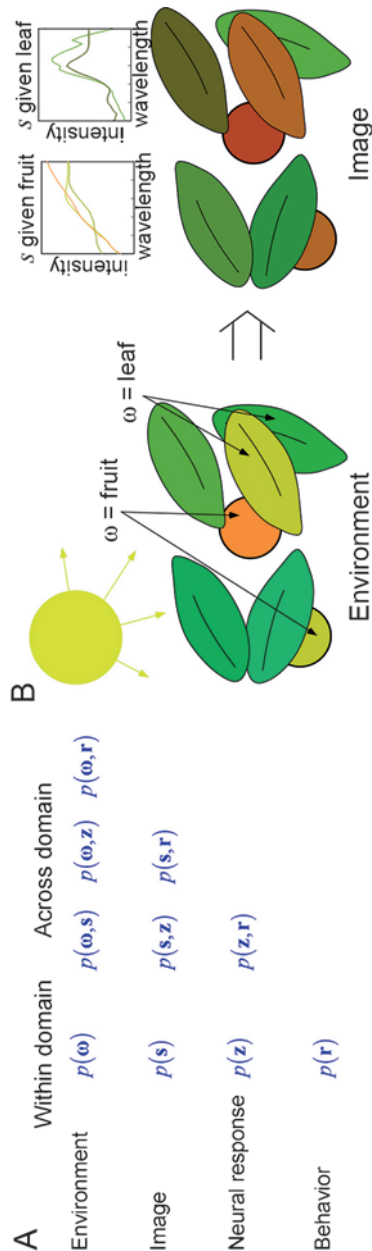**Figure 1**

Kinds of natural scene statistics. (*A*) There are two general kinds of natural scene statistics that can be measured at various stages (domains) from environment to behavior. Within-domain statistics describe the probability distribution of a vector of properties within a given domain. Across-domain statistics describe the joint probability distribution of properties across domains. (*B*) Hypothetical example. Suppose an organism's task is to identify whether a randomly selected location in the retinal image corresponds to the surface of a fruit or the surface of a leaf. The relevant property of the environment takes on just one of two possible values ($\omega = fruit$ or $\omega = leaf$). The relevant property of the image is a wavelength spectrum **s** (the plots show a few examples of spectra for fruits and leaves). Measuring the probability of fruit or leaf is an example of measuring $p(\omega)$. Measuring the probability distribution of wavelength spectra without regard to the source (leaf or fruit) is an example of measuring $p(\mathbf{s})$. Measuring the probability distribution of wavelength spectra separately for fruit and leaves and then combining with the probability of fruit or leaf is an example of measuring an across-domain statistic $p(\omega, \mathbf{s})$. See text for more details.

Figure 2

Using hand segmentation to measure across–domain statistics. (*A*) Close-up image of foliage obtained with a 36-bit calibrated camera. The orange circle indicates the region to be hand segmented. (*B*) The colored leaves and branches show the result of the hand segmentation. The orange shaded leaf is an individual segmented object.
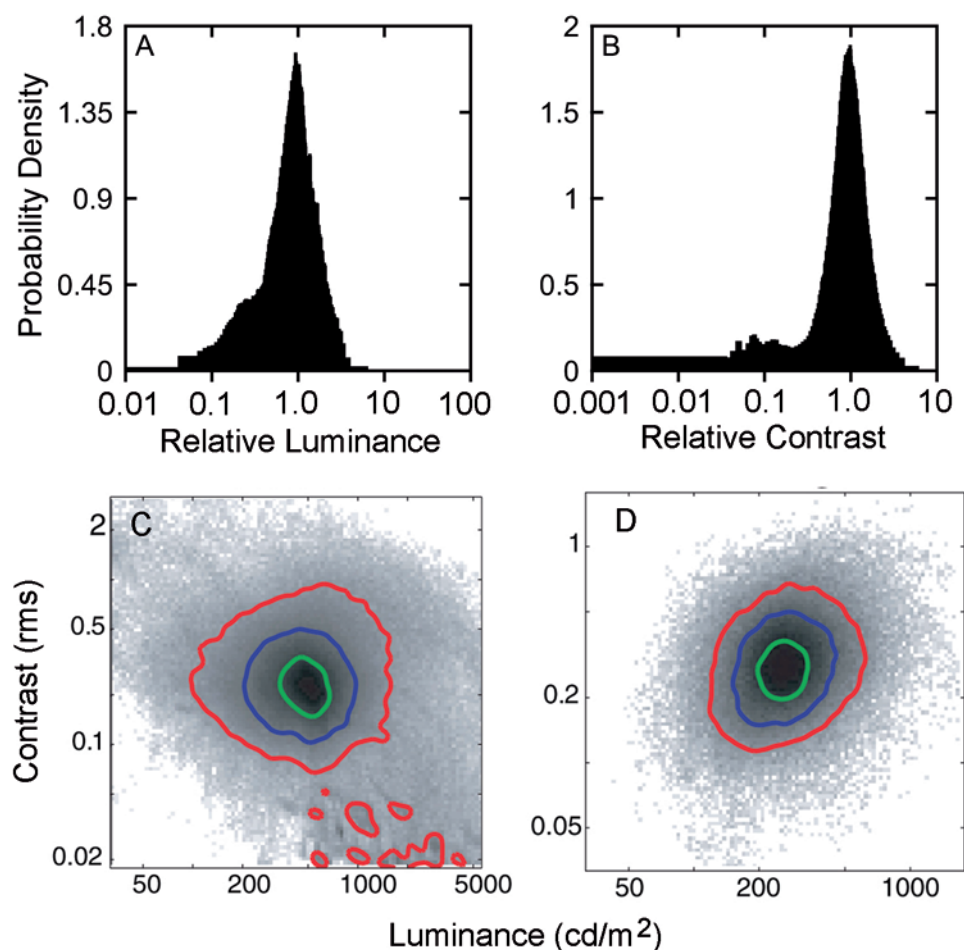
Figure 3

Luminance and contrast in natural images. (*A*) Average distribution of local luminance levels within a
natural image, relative to the mean luminance of the image. (*B*) Average distribution of local contrast in
natural images, relative to the mean contrast of the image. (*C*) Average joint distribution of local lumi-
nance and local contrast in natural images, scaled to the average luminance and average contrast in nat-
ural images. (Contours enclose 90%, 65%, and 40% of the observations.) (*D*) Average joint distribution
of local luminance and local contrast within foliage regions of natural images.
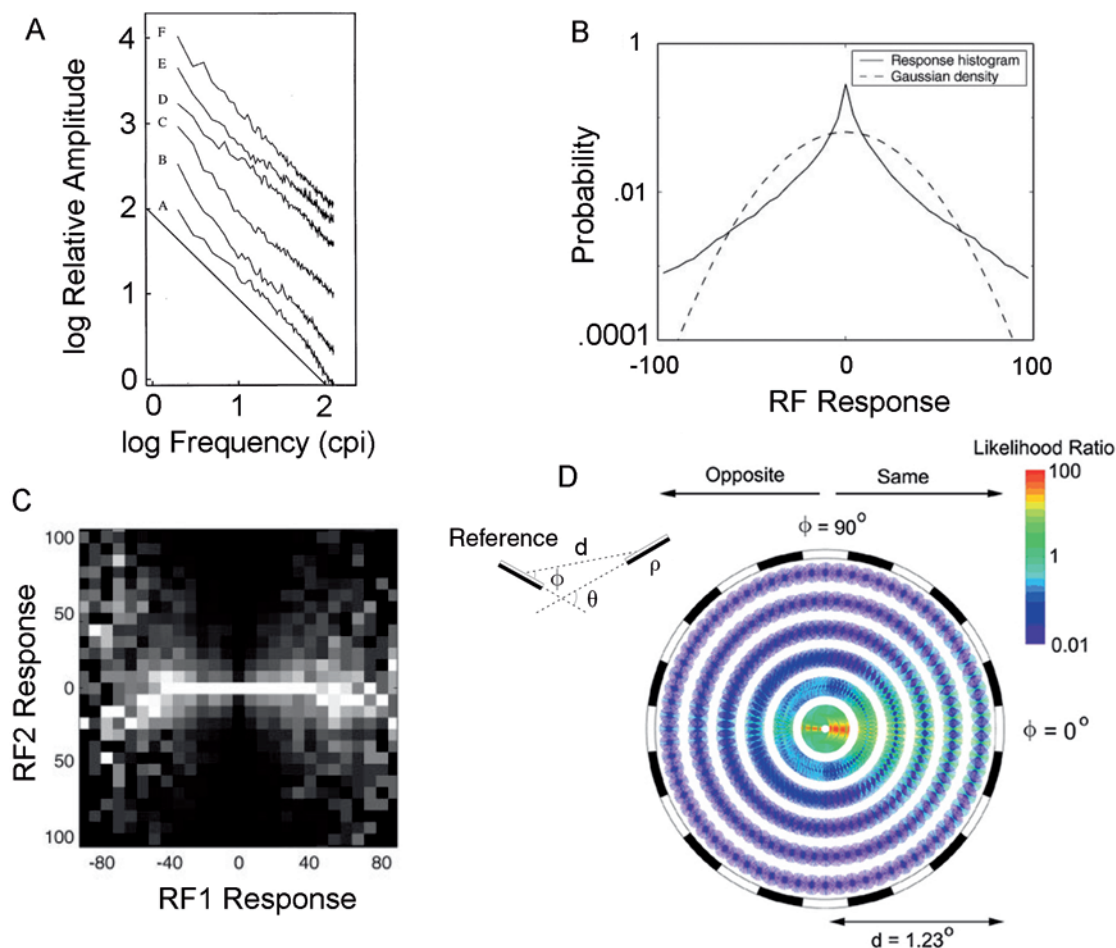
Figure 5

Spatial properties of natural images. (*A*) Amplitude spectra of six natural images (adapted from Field 1987). The spectrum for each image has been displaced vertically for display purposes. The diagonal line has a slope of -1.0. (*B*) Histogram of responses to natural images of a sensor having a receptive field profile representative of those in primary visual cortex. The dashed line shows the best-fitting Gaussian distribution. (*C*) Histograms of responses to natural images of a sensor (RF2) conditional on the response of a nearby but nonoverlapping sensor (RF1). The histograms are represented by the vertical columns of pixels (the brighter the pixel the greater the frequency). For display purposes, each column of pixels has been scaled to use the full grayscale range. (*D*) Histogram showing the ratio of the likelihood that a particular pair of edge elements belongs to the same physical contour to the likelihood that the pair belong to different physical contours. In this plot, the central horizontal line segment represents one of the pair of edge elements (the reference); each ring represents a distance bin *d*; each location around the diagram represents a direction bin ϕ; each line element plotted at a given distance and direction represents an orientation difference bin θ. The right side of the plot shows the likelihood ratios when the contrast polarity is the same, the left side when the contrast polarity is opposite.

Ⱥ

Contents

Errata

An online log of corrections to *Annual Review of Psychology* articles may be found at
http://psych.annualreviews.org/errata.shtml