Local statistics in natural scenes predict the saliency of synthetic textures

Gašper Tkačik[a,1], Jason S. Prentice[a], Jonathan D. Victor[b], and Vijay Balasubramanian[a]

[a]Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104; and [b]Department of Neurology and Neuroscience, Weill Cornell Medical College, New York, NY 10065

Edited* by Wilson S. Geisler, University of Texas at Austin, Austin, TX, and approved August 27, 2010 (received for review January 4, 2010)

The visual system is challenged with extracting and representing behaviorally relevant information contained in natural inputs of great complexity and detail. This task begins in the sensory periphery: retinal receptive fields and circuits are matched to the first and second-order statistical structure of natural inputs. This matching enables the retina to remove stimulus components that are predictable (and therefore uninformative), and primarily transmit what is unpredictable (and therefore informative). Here we show that this design principle applies to more complex aspects of natural scenes, and to central visual processing. We do this by classifying high-order statistics of natural scenes according to whether they are uninformative vs. informative. We find that the uninformative ones are perceptually nonsalient, while the informative ones are highly salient, and correspond to previously identified perceptual mechanisms whose neural basis is likely central. Our results suggest that the principle of efficient coding not only accounts for filtering operations in the sensory periphery, but also shapes subsequent stages of sensory processing that are sensitive to high-order image statistics.

natural scene statistics | psychophysics | vision

Many aspects of early visual processing appear to be shaped by a necessity for efficient representation of the information in natural stimuli. Examples include: (*i*) the center-surround receptive field of the retinal ganglion cell, which removes spatial correlations in natural images and decreases retinal redundancy (1–3), (*ii*) the twofold excess of retinal OFF pathways (encoding negative contrasts) as compared to ON pathways (encoding positive contrasts), which matches the asymmetric contrast structure of natural scenes (4), (*iii*) cone spectral sensitivities and color opponency in ganglion cells, which maximize chromatic information from natural scenes (5–7), (*iv*) overlaps of ganglion cell receptive fields within the retinal mosaic, which balance redundancy reduction against signal-to-noise ratio improvement (8, 9), and (*v*) the shapes of the nonlinear response functions of early sensory neurons, and their adaptation to stimulus variance, which have been related to the skewed intensity distributions that occur in natural stimuli (10, 11). In all cases, physiological and anatomical characteristics of the visual system are accounted for by a simple efficient coding principle: sensory systems invest their resources in relation to the expected gain in information (4).

All these examples refer to first-order image statistics (the distribution of light intensities at single pixels) or simple second-order image statistics (covariances of light intensities at pairs of pixels), and to processing within the retina. It is unknown whether such an explanatory framework extends to more complex image statistics, or to central visual processing. There are two reasons for this gap in knowledge. First, higher-order image statistics are challenging to analyze, because of their complexity and high dimensionality (12). Second, more is known about the filter-like properties of visual neurons, than about their sensitivity to higher-order features. Yet it is precisely these higher-order features that underlie the perception of lines, edges, and texture, so characteristic of the natural image ensemble (13).

Local texture in images is determined partly by the distribution of light intensities and partly by the spatial organization of light across pixels. Thus, we approach the problem of characterizing high-order natural image statistics by two complementary dimensionality-reduction approaches. To focus on intensity distributions, we analyze variations in local intensity histograms that arise due to spatial correlations of light. To focus on local spatial organization, we binarize images, and analyze fourth-order correlations of nearby pixels. We use both approaches to characterize image statistics according to their informativeness about the local structure of natural images, and we find that this characterization is robust across spatial scales.

Remarkably, in both cases, we find that the distinction between informative vs. uninformative high-order statistics corresponds closely to the perceptual sensitivities of the visual system. In the case of intensity distributions, the three most informative aspects of histogram statistics of natural images correspond to the three mechanisms that account for perception of spatially unstructured ("independent, identically-distributed") artificial textures, namely, mean, variance, and a quantity known as "blackshot" (14, 15). In the case of spatial organization, we find that the configurations of fourth-order correlations that are informative correspond to the configurations of fourth-order spatial correlations that are visually salient (16, 17). Moreover, sensitivity to the latter high-order correlations is known to arise in visual cortex (17–19).

These results suggest that the principle of "efficient coding" applies not only to the simple image statistics that shape peripheral processing, but also to high-order image statistics and to sensory processing within the central nervous system: cortical circuits are preferentially selective for image features that are more informative about the local structure of natural scenes.

Results

The Local Distribution of Light in Natural Scenes. Natural images have an inhomogeneous (20) and spatially correlated (21) distribution of light which makes pixels of similar intensities more likely to clump together. The resulting variations in the histograms of light intensity between local image patches contribute to the perception of texture. The clumping of similar intensities in local image patches is characterized by conditional distributions $P_R(\sigma_1|\sigma_0)$ for intensities σ_1 at pixels sampled at a distance R away from a central pixel of a intensity σ_0 (see *SI Appendix* for further details).

From a database of natural images discretized to have 16 equally likely intensity levels for each pixel, we sampled the conditional distributions $P_R(\sigma_1|\sigma_0)$, for each possible value of σ_0 and a large range of separations R between 2 and 2^{10} pixels

Author contributions: G.T., J.S.P., J.D.V., and V.B. designed research; G.T., J.S.P., J.D.V., and V.B. performed research; G.T. and J.S.P. analyzed data; and G.T., J.S.P., J.D.V., and V.B. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]To whom correspondence should be addressed. E-mail: gtkacik@sas.upenn.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.0914916107/-/DCSupplemental.

(see *Materials and Methods*, Fig. 1*A*). Because of spatial correlations, nearby pixels tended to have similar intensities, leading to peaked shapes for the intensity histograms for small R (Fig. 1*B*). At large separations, pixels tended towards statistical independence (Fig. 1*D*) so that the intensity histograms for large R became increasingly independent of the intensity of the central pixel. Because we discretized our images to have 16 equally probable intensities, the large R distributions tended towards uniformity. At small separations there is an asymmetry between bright and dark—dark pixels are more clumped, while bright pixels appear as small specks within darker areas (Fig. 1*B*, black vs white curve). This greater correlation between dark pixels is likely related to the reported excess of "dark regions" in natural scenes (4).

The Local Statistics of Light Predicts Perceptual Salience. To characterize the variations between the distributions $P_R(\sigma_1|\sigma_0)$, we carried out a principal components analysis (PCA) on the mean-subtracted ensemble of intensity histograms for all values of R and σ_0. The ensemble was sampled uniformly over the 16 possible intensities σ_0 and uniformly in $\log(R)$ (see *Materials and Methods*). We included a range of spatial scales R in the ensemble because there is no preferred distance from which a scene is viewed and thus no "typical" size at which to define a local neighborhood. We found that ~90% of the variance was explained by just three principal components v_j (Fig. 2*A*). Thus, most of the variation between intensity histograms of local image patches is explained by the differences in the strengths of the three coefficients in $P_R(\sigma_1|\sigma_0) \approx 1/16 + \sum_{j=1}^{3} c_R^j(\sigma_0)v_j(\sigma_1)$, where $1/16$ is the uniform distribution over the 16 intensity levels.

The above analysis shows that the intensity distributions in natural images are highly stereotyped: ~90% of their variance can be accounted for by linear admixtures of three elements, v_1, v_2, and v_3. Interestingly, previous psychophysical studies with synthetic textures have shown that human sensitivity to luminance distributions can also be accounted for by three mechanisms θ_1, θ_2, and θ_3 (15). Each of these mechanisms reports the projection of the luminance histogram onto one of three vectors: θ_1 projects onto $(\sigma_1 - 15/2)$, and thereby reports the mean intensity; θ_2 projects onto $(\sigma_1 - 15/2)^2$ and thereby reports the variance; and θ_3 (orthogonal to θ_1 and θ_2) projects onto a vector that is heavily weighted at low values, thereby reporting the fraction of dark pixels. The three θ_i can be linearly combined into the blackshot mechanism which is useful for discriminating between the darkest intensities (15) (see *SI Appendix*).

We therefore asked whether the three components derived from natural images (the v_j) span the same space as the three axes that define human sensitivity, the θ_j. Since the principal components decomposition into the v_j is only unique up to a coordinate rotation, we asked whether there was a correspondence at the subspace level, rather than whether each v_j matches the corresponding θ_j. Fig. 2*B* shows that there is such a correspondence. We demonstrated this result by finding a rotation within the v-subspace that transformed the v_j into another orthonormal set w_j, for which w_1 closely approximated θ_1, w_2 closely approximated θ_2, and w_3 closely approximated θ_3 (Fig. 2*B*) (closeness assessed by the sum of squared errors). We tested that a linear combination of the w_i can be selected to closely approximate the blackshot mechanism that is sensitive to fine gradations between the darkest pixels ((15) and *SI Appendix*). The identification of such a transformation is not at all guaranteed: the space of intensity histograms is 15-dimensional; within this, the v_j and the θ_j span approximately identical three-dimensional subspaces. Thus, humans are primarily sensitive to intensity histogram variations that match the principal histogram variations that actually occur in natural scenes. Fig. 2 *C–E* illustrate these histogram variations.

As controls for robustness, we also applied PCA to $P_R(\sigma_1|\sigma_0)$ at each R, and to uniform sampling in R (logarithmic sampling

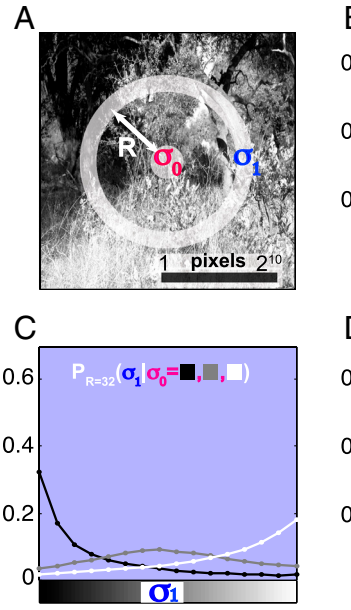

Fig. 1. Local distributions of light in natural scenes. (*A*) Natural images are discretized into 16 equipopulated grayscale levels. Central pixels with intensity σ_0 are chosen randomly, and the distribution of intensities σ_1 at a radius R from the center is sampled. The distance R is represented by the scale bar in pixels. (*B*), (*C*), and (*D*) Histograms of pixel intensities at different distances from the center (shown here for $R = 4, 32, 256$) if the central pixel σ_0 is black, gray, or white (black, gray, and white lines). Note the difference in histograms for black vs. white central pixel σ_0 for small R (see *SI Appendix* for $P(\sigma_1)$ as a function of R).

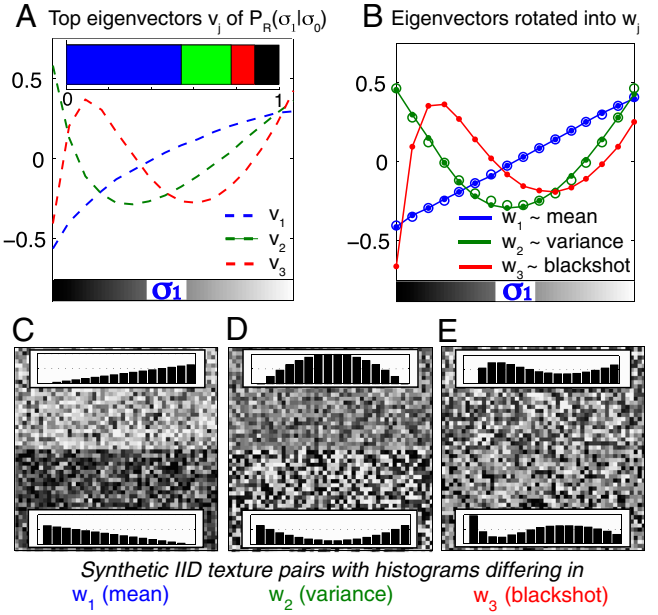

Fig. 2. Variability of local intensity histograms. (*A*) The three principal components of the ensemble of local intensity histograms $P_R(\sigma_1|\sigma_0)$ (blue, green, and red), along with the fraction of variance explained by each (bar chart inset). Together, the three components explain ~90% of the variance between histograms. (*B*) An orthogonal transformation rotates the three principal components, $\{v_j\} \rightarrow \{w_j\}$, so that w_1 is as close as possible to a linear function of σ_1, and w_2 to a quadratic function of σ_1 (open blue circles = linear function, $\theta_1(\sigma_1)$; open green circles = quadratic function, $\theta_2(\sigma_1)$). The three new axes $\{w_j\}$ relate to variations in mean, variance, and blackshot of the intensity histogram. (*C*), (*D*), and (*E*) IID texture pairs. The intensity of each pixel is chosen independently according to the inset distributions which vary from the uniform distribution by adding or subtracting one principal component w_j from *B*. These additions vary the mean (*C*), variance (*D*) or blackshot (*E*) of the texture. Only IID textures that vary in at least one of these three ways can be reliably discriminated by humans (15).

was used above). These procedures robustly gave the same eigenvectors, but the fraction of variance explained by w_2 and w_3 increased with decreasing R (*SI Appendix*). We also applied PCA to the ensemble of single-pixel intensity histograms (marginal intensity distributions; $P_R(\sigma_1)$) sampled from $R \times R$ pixel patches for all values of R. PCA on this ensemble (directly related to the experiments in (15)) also gave the same eigenvectors (*SI Appendix*). The observation that PCA on $P_R(\sigma_1)$ agrees with PCA on $P_R(\sigma_1|\sigma_0)$, confirms that the significant variations in local intensity histograms of natural scenes arise from clumping due to correlations.

To test the role played by higher-order image statistics in these results we repeated our analysis in synthetic image ensembles (*SI Appendix*) that matched natural images in power spectrum (21), but not in other respects. This synthetic ensemble required just two principal components (mean and variance) to explain more than 90% of variation in the local intensity histogram. Further, the skew towards dark intensities in the third blackshot component was absent. This suggests that higher-order correlations in natural scenes play a key role in making blackshot a perceptually salient image statistic.

In sum, we found a striking statistical regularity in the local intensity histograms of natural scenes: they can be accounted for by linear admixtures of three basic kinds of histogram variations. These three kinds of variations correspond to the three mechanisms that humans use to discriminate among synthetic independent, identically distributed (IID) textures (15). That is, it seems that humans discriminate intensity distribution variations that are frequent in nature, and are insensitive to the variations that occur rarely. Our results also suggest that the most common variations in natural scene patches occur partly because of the underlying correlations. This idea can be tested by generating *nonIID* images which vary only in the conditional pixel distributions we measured. We predict that humans discriminate such textures based largely on the three principal components we have measured in natural images.

Spatial Correlations and Local Textures in Natural Scenes. Textures in images also arise in part from correlations between many pixels at the same time. Such cross-correlations are difficult to characterize because they proliferate rapidly with the number of pixels. For example, with just four contiguous pixels there are four expectation values, six dipole (pair) correlations, four triplets, and one quadruplet (Fig. 3*A*). Even assuming that these are translation invariant, there are ten independent quantities. Moreover, lower-order correlations (e.g., pairwise) induce higher-order relations between multiple pixels, making it delicate to extract intrinsically higher-order structures. Because there are so many different ways in which multiple pixels can be related it is a challenge to find useful ways of characterizing higher-order correlations in natural scenes.

We devised a method for assessing such correlations, inspired by procedures for generating textures with higher-order correlations that are used in psychophysical studies (16, 17, 22–26). The generative approach begins with a "glider" $\mathcal{G}$, consisting of Q pixels in some geometrical arrangement; Fig. 3*B* displays eight such four-pixel gliders, $\mathcal{G}_1,\ldots,\mathcal{G}_8$. We allow each pixel to take one of L intensity levels. Consider a probability distribution $P_{\mathcal{G}}$ over the L^Q intensity assignments over the glider shape (e.g., Fig. 3*C* for a square glider with four binary pixels, i.e., $Q = 4$, $L = 2$, with $2^4 = 16$ possible colorings). It is possible to construct synthetic textures in which the only correlations are those implied by the distribution $P_{\mathcal{G}}$ (see *Materials and Methods* and two examples in Fig. 3*B*; (25, 26)). "Isodipole textures" generated from binarized gliders with four pixels containing fourth-order, but no second- or third-order correlations divide into two groups (e.g., Fig. 3*B*)—those in Group 1 are perceptually salient on a white binary noise background, and those in Group 2 are not (16–18, 22, 24). (Group 2 corresponds to the Group III of (17).)

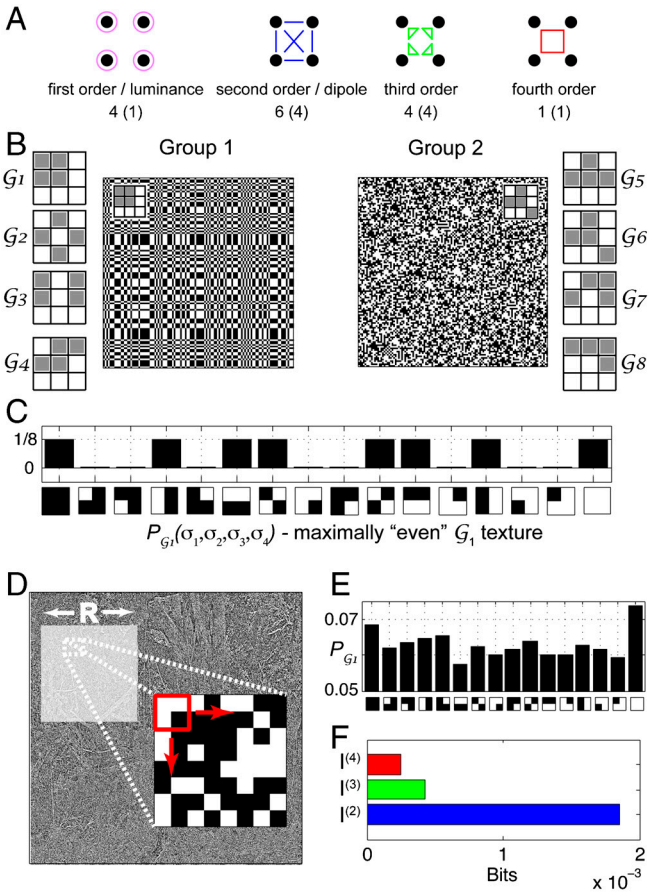

Fig. 3. Spatially correlated textures. (*A*) Correlations of different orders between four pixels. There are four mean pixel luminances (pink circles), six pairwise correlations (blue lines), four triplet correlations (green triangles), and one quadruplet (fourth-order) correlation (red square); translation invariance reduces the number independent quantities to 10 (numbers in parenthesis). (*B*) Examples of gliders and the textures they generate (see *Materials and Methods*). Both displayed textures have equally many white and black pixels, have no second- or third-order correlations, and a large fourth-order correlation. Gliders from Group 1 generate textures that are perceptually salient against a white binary noise background, while textures generated from gliders in Group 2 are not perceptually salient (16, 24). (*C*) An example of a distribution over binary patterns in a square glider. This distribution generates synthetic textures that have only fourth-order correlations (example texture in *3B*, left). (*D*) To measure the fourth-order correlations in natural scenes we select patches of $R \times R$ pixels from whitened natural scenes binarized to have equally many white and black pixels. Each of the eight gliders in *B* (a square glider shown here in red) is scanned across a patch, and the histogram of binary patterns encountered by the glider is accumulated. (*E*) Histogram of binary patterns encountered by a square glider scanning a 64×64 patch from a natural image. (*F*) The information about texture in a 64×64 binary image patch that is contained in second-, third-, and fourth-order correlations, extracted with a square glider.

We wanted a method of assessing how much of the local structure in natural scenes is explained by the presence of particular textures arising from higher-order correlations. We concentrated on the isodipole textures that have been the focus of psychophysical study. To begin to isolate higher-order correlations, we first removed the well understood scale invariant second-order correlations (21) by whitening our images, and then binarized pixels at the median of the image intensity distribution, so that half the pixels in each image were black and half were white (see *Materials and Methods*). (The binarization reintroduces a small amount of second-order correlation—see *SI Appendix*.) Then, we treated $R \times R$ pixel blocks of the images as texture patches (Fig. 3*D*), and accu-

mulated the histogram of intensities sampled by a given glider shape as it scanned over such texture patches. Thus, for each image patch of size R, each glider $\mathcal{G}$ yielded a histogram $P_{\mathcal{G}}(\vec{\sigma})$ over the $2^4 = 16$ possible ways to assign black or white to each of the four pixels in a glider. This histogram contained complete information about first-, second-, third- and fourth-order correlations between the four pixels of each glider in an image patch (results for a square glider and a 64×64 image patch are in Fig. 3*E*).

If there were no correlations of any kind between the pixels in the glider, then $P_{\mathcal{G}}$ in a given patch would be uniform and have a maximal entropy $S_{\mathcal{G}} = -\sum_{\vec{\sigma}} P_{\mathcal{G}}(\vec{\sigma}) \log_2 P_{\mathcal{G}}(\vec{\sigma})$, of four bits. Because pixels are not independent, the entropy $S_{\mathcal{G}}$ will in general be less than four bits; we can write $S_{\mathcal{G}} = Q - \sum_{\nu=1}^{Q} I_{\mathcal{G}}^{(\nu)}$ where $Q = 4$ is the number of binary pixels in the glider, and $I_{\mathcal{G}}^{(\nu)}$ measures the bits of entropy reduction caused by luminance bias ($\nu = 1$), and by pair ($\nu = 2$), triplet ($\nu = 3$), and quadruplet ($\nu = 4$) correlations (27).

This decomposition is general and can be used to isolate correlation of arbitrary order in natural scenes. In detail, we start by building a series of so-called "maximum-entropy" approximations to the true distribution $P_{\mathcal{G}}$: $P_{\mathcal{G}}^{(1)}(\vec{\sigma})$, $P_{\mathcal{G}}^{(2)}(\vec{\sigma})$, $P_{\mathcal{G}}^{(3)}(\vec{\sigma})$, $P_{\mathcal{G}}^{(4)}(\vec{\sigma})$, such that $P_{\mathcal{G}}^{(\nu)}$ is as random as possible while reproducing correlations up to order ν in $P_{\mathcal{G}}$ (see *SI Appendix*). Because our distributions have four pixels and thus a maximal correlation of fourth-order, $P_{\mathcal{G}}^{(4)}$ must identically equal the true distribution $P_{\mathcal{G}}$, and the series terminates. Each of these distributions has its associated entropy, $S_{\mathcal{G}}^{(\nu)} = -\sum_{\vec{\sigma}} P_{\mathcal{G}}^{(\nu)}(\vec{\sigma}) \log_2 P_{\mathcal{G}}^{(\nu)}(\vec{\sigma})$. Following (27), ν-th order correlations within the glider carry $I_{\mathcal{G}}^{(\nu)} = S_{\mathcal{G}}^{(\nu-1)} - S_{\mathcal{G}}^{(\nu)}$ bits of information about local texture. These information-theoretic quantities measure order in a texture that arises from correlations that involve *exactly* ν pixels. In this manner, we can isolate the impact of fourth-order correlation in natural textures despite the simultaneous presence of lower (e.g., second- or third-) order correlations, a characterization that is hard to achieve using the traditional moment-based correlation measures. An example of such a decomposition for a square glider sampling a 64×64 image patch is given in Fig. 3*F*.

Nonzero values of $I_{\mathcal{G}}^{(\nu)}$ indicate that the correlations among patches of ν pixels could not have been guessed from the correlations among smaller patches, i.e., that the correlations among ν pixels are informative. In gliders with $Q = 4$ pixels, the fourth-order correlation is *special*, because a single quadruplet ($\{\sigma_1,\sigma_2,\sigma_3,\sigma_4\}$) contributes to it, through a product $\Pi(\vec{\sigma}) = \sigma_1\sigma_2\sigma_3\sigma_4$ of all four pixels. Since pixels are binary ($\sigma = \pm 1$) this product is ± 1. Consequently (see *SI Appendix*) each glider distribution can be uniquely decomposed as

$$P_{\mathcal{G}}(\vec{\sigma}) = P_{\mathcal{G}}^{(3)}(\vec{\sigma}) + (\alpha_{\mathcal{G}}^{(4)}/16)\Pi(\vec{\sigma}), \qquad [1]$$

where $\Pi(\vec{\sigma}) = 1$ or -1 if the number of white pixels in the binary pattern $\vec{\sigma}$ is even or odd. In our ensemble, $|\alpha_{\mathcal{G}}^{(4)}|$ and the information measure $I^{(4)}$ are related (see *SI Appendix*). Conceptually, $\alpha_{\mathcal{G}}^{(4)}$ measures fourth-order correlation between binary pixels in a manner similar to a pairwise correlation coefficient. Positive (negative) $\alpha_{\mathcal{G}}^{(4)}$ denotes bias towards an even (odd) number of white pixels in a glider $\mathcal{G}$.

This formalism lays down the foundation for analysis of fourth-order correlation in natural scenes. Specifically, $I_{\mathcal{G}}^{(4)}$, computed over many texture patches, will tell us how much fourth-order correlation there is, on average, between four pixels arranged in a glider. If $I_{\mathcal{G}}^{(4)} = 0$, then fourth-order correlations are absent and $\alpha_{\mathcal{G}}^{(4)}$ must also be 0. If this quantity is significantly different from 0, the local fourth-order statistics are informative, i.e., they cannot be computed from lower-order ones. We validated our formalism by applying it to synthetic textures generated by specific gliders $\mathcal{G}$ (see *SI Appendix*). For such textures, correlations between pixels arranged according to $\mathcal{G}$ were highly informative, while correlations between pixels arranged according to other glider geometries were uninformative. Thus our analysis correctly recovers the structure present in synthetic textures.

The Local Statistics of Correlated Textures Predicts Perceptual Salience. To test how much information is conveyed about natural image textures by correlations of different orders, we constructed the quantities $I_{\mathcal{G}}^{(\nu)}$ and $\alpha_{\mathcal{G}}^{(4)}$ for each glider and many $R \times R$ image patches (computational details are given in *SI Appendix*). Fig. 4*A* shows that at all scales, second- and third-order correlations yield similar amounts of information about image patches seen through any glider. However, fourth-order correlations in natural scenes are much more informative when measured in the pixel arrangements of Group 1 gliders, which are also the ones that generate perceptually salient textures (16, 17). Correspondingly (Fig. 4*B*), $\alpha_{\mathcal{G}}^{(4)}$ becomes significantly positive for Group 1 gliders, but not for those of Group 2. The fact that $I_{\mathcal{G}}^{(4)}$ is significantly nonzero and $\alpha_{\mathcal{G}}^{(4)}$ is significantly positive for Group 1 gliders but not for those of Group 2 indicates that fourth-order correlations within Group 1 gliders are informative about natural scenes, while fourth-order correlations within Group 2 gliders can be inferred from lower-order correlations.

Above, we divided the gliders into groups based on psychophysical studies; next we show that this subdivision emerges from the image statistics themselves. To carry out this analysis, we compared the full distributions of $\alpha_{\mathcal{G}}^{(4)}$ over image patches generated by each glider, using the Jensen-Shannon distance measure (D_{JS})†. The Jensen-Shannon distance quantifies how discriminable two distributions are from each other; $D_{\mathrm{JS}} \to 0$ for identical distributions. In our context, D_{JS} assesses differences in fourth-order correlations in natural scenes seen through the lens of different gliders $\mathcal{G}$. Thus, we computed D_{JS} for each pair of the eight gliders in Fig. 3*B* sampling $R \times R$ image patches at three different scales R (Fig. 4*C*).

At sufficiently large R (e.g., $R \geq 64$ pixels) the eight gliders naturally cluster into two groups—the Jensen-Shannon distance is small within each group, and large between the groups. This clustering shows that, in natural textures, the correlations between pixel quadruplets differ qualitatively between Group 1 and Group 2 pixel arrangements. This separation into two groups, one perceptually salient and one not, was just as reported in perceptual studies ((17); see Fig. 3*B*). Here we are showing that the two groups also separate purely on the basis of natural scene statistics, without any reference to perceptual experiments. Group 1 gliders "sense" fourth-order correlations in natural scenes, while Group 2 gliders do not. We have checked that this separation into groups disappears in scrambled natural images that lack higher-order structure (see *SI Appendix*).

In sum, Fig. 4 *A*–*C* demonstrate that fourth-order correlations in natural scenes have a specific qualitative structure—only some patterns of four pixels are correlated. It is precisely these gliders (Group 1) for which fourth-order correlations are perceptually salient (17). In synthetic textures, these fourth-order correlations can be identified when present at low levels, and within a single 50 ms fixation. In contrast, the Group 2 correlations are only detected when present at high levels, if at all. Moreover, introduction and removal of Group 1 correlations from synthetic textures elicit a large visual evoked potential (VEP) (17, 19, 28); no comparable response is elicited by Group 2 correlations (17). Within Group 1 correlations, "even" configurations elicit a larger VEP than "odd" configurations; this too appears to correspond to a feature of natural image statistics—as shown by the positivity of $\alpha_{\mathcal{G}}^{(4)}$ (Fig. 4*B*), natural images contain a bias towards the "even" configurations of the Group 1 gliders.

†Given distributions p and q, let $m(x) = (p(x) + q(x))/2$. The Jensen-Shannon distance is: $D_{\mathrm{JS}} = 0.5 \int dx p(x) \log_2[p(x)/m(x)] + 0.5 \int dx q(x) \log_2[q(x)/m(x)]$. $D_{\mathrm{JS}} \to 0$ for identical, and $D_{\mathrm{JS}} \to 1$ for distinct p, q.

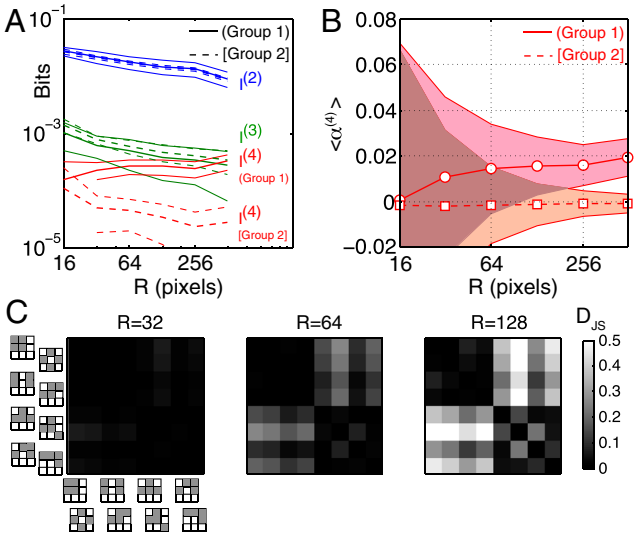

Fig. 4. Fourth-order correlations and perceptual salience. (*A*) Decomposition of textural information into second (blue) , third (green), and fourth (red) order for the two groups of gliders and many spatial scales (central line = mean, thin surrounding lines = std across gliders). In large image patches there is significantly more information about texture in the correlations between four pixels arranged in the patterns from Group 1 gliders, which also generate perceptually salient textures. Group 1 and Group 2 gliders have similar amounts of $I^{(2,3)}$. (*B*) Fourth-order correlations as measured by the parameter $\alpha_{\mathcal{G}}^{(4)}$ Eq. **1**. Results at each R are averaged across Group 1 gliders (solid, circles) and Group 2 gliders (dashed, squares), and across many $R \times R$ texture patches. The shaded areas show the standard deviation of $\alpha_{\mathcal{G}}^{(4)}$ across texture patches for the two groups. As R increases the correlations within the perceptually salient gliders acquire high statistical significance. (*C*) The Jensen-Shannon distance, D_{JS}, between the distributions of $\alpha_{\mathcal{G}}^{(4)}$ sampled across many $R \times R$ image patches, for all pairs of gliders (arrangements of four pixels, see Fig. 3*B*). As R increases, the gliders cluster into two sets, which respectively generate the perceptually salient (Group 1) and nonsalient (Group 2) textures determined by psychophysical studies (17).

Discussion

The concept of efficient coding is an organizing principle that accounts for many aspects of retinal processing (how the retina samples images, its chromatic sensitivity, its filter-like aspects, and intensity-response functions (1–11)) on the basis of simple statistics of natural scenes, such as their intensity and chromatic distributions and covariances. Some receptive field properties of neurons in primary visual cortex (V1) can also be viewed as adapted to the statistical structure of natural scenes (29–31). However, the applicability of the efficient coding hypothesis to later visual processing, where nonlinear feature extraction occurs, is as yet unclear. The key step in addressing this question is to characterize the higher-order statistics of natural images beyond intensity distributions and covariances. This is a challenging problem, due to the complexity of natural scenes and the intrinsic high dimensionality of the required statistics (12).

Our strategy for attacking this problem relies on a method to determine how much of the local structure in natural scenes is explained by a particular underlying texture. Traditional methods of quantifying structure, e.g., correlation coefficients, are not helpful because they do not quantify how much of local structure in scenes is explained by a particular kind of texture, and because they cannot easily disentangle correlations of various orders. We devised a simple, yet powerful, approach inspired by generative procedures for producing texture. We accumulated joint distributions of intensities of pixels arranged in specific geometric patterns (gliders). We measured how, and how much, these distributions varied from the random (uniform) distribution, and whether these distributions could be predicted from first- and second-order image statistics. We then used these deviations to characterize the high-order statistics of natural scenes.

The strategy was applied in two ways: one that focused on the kinds of gray level distributions that are typically present in local regions of natural scenes (where we used principal components analysis of intensity distributions), and one that focused on the kinds of local spatial organization that are present (where we used four-pixel gliders and a maximum-entropy formalism to analyze their probability distributions). In both cases we found that statistical variations that are informative about differences between natural image patches are precisely those that humans find salient. Our analysis does not provide a generative model of why only certain classes of textural variations occur in natural scenes, or give a causal account of texture discrimination. Nevertheless, it shows a striking correlation between the variations that occur naturally, and what we are able to perceive. Our results are robust—variations in sampling, discretization, and processing do not significantly affect the findings (see *SI Appendix*).

Our approach revealed regularities in natural scenes that go beyond the $1/f$ spectral distribution (21) and overall light intensity distribution (20). These regularities account for the blackshot sensitivity function, and for the separation of gliders into those that do and those that do not generate perceptually salient texture. It has been previously suggested that blackshot could enable fine discrimination in shaded regions during otherwise bright ambient illumination (15), but no quantitative argument for this has been put forward to date. In the case of fourth-order correlation, simple models based on the intrinsic symmetry, information, and geometric properties of the gliders likewise failed to explain perceptual results (17). Our analysis, on the other hand, finds an explanation for both classes of perceptual sensitivities from the statistics of natural scenes while developing a general methodology for linking complex natural scenes statistics to perceptual experiments with synthetic images.

The neural processing that underlies the perception of high-order spatial correlations is highly likely to be central. The relevant evidence is both theoretical and empirical. The theoretical evidence is that these correlations can be perceived even when they do not affect the first- and second-order statistics of the image, as shown by several psychophysical studies of isodipole textures (16, 17, 22, 24). Thus, their presence cannot be detected by analysis of the firing rates or mean-squared firing rates of banks of quasilinear neurons. The experimental evidence that this processing is central is that differential responses to such isodipole stimuli are absent in the lateral geniculate nucleus (19), but present in the cortex in cat (19), macaque (18), and human (17).

At first sight, our method of analysis seems to show that absolute amount of information concerning texture that is contained in specific higher-order correlations is quite small (Figs. 3 and 4). Why would the nervous system make selective investments for such apparently small gains? First, small differences can add up to a significant advantage, when summed over a large number of pixels. For example, 0.001 bit per pixel, accumulated over only a 30×30 image patch, yields 1 bit. Second, the actual textures in natural scenes combine correlations between different numbers of pixels arranged in many different kinds of patterns. Thus, correlations of any given type should only be expected to make a small contribution to the overall deviation from white noise. Nevertheless, it is precisely the sum of these small effects that gives rise to a natural image.

We did not attempt to account for sensitivities in vision related to lifestyles of specific animals, e.g., pathways tuned to the profiles of predators (32); we simply sampled exhaustively without bias across the whole ensemble. Our methods could be refined to focus on ethologically relevant aspects of images, by selecting segmented image patches containing visual features of behavioral interest. Our methods could also be refined to work with multiscale wavelet bases or other representations which inherently recognize that higher-order dependencies between many pixels are essential to the perception and generation of natural textures

(33–35). However, even without these refinements, we find a close correspondence between high-order statistics that are informative, and those that are visually salient.

Broadly, we identified statistical regularities of natural scenes, and showed (via comparison with earlier psychophysical experiments with artificial stimuli) that these regularities predicted the presence (and absence) of mechanisms sensitive to specific image statistics. We did not seek to account in general for texture segmentation in natural images. Rather, we used texture segmentation of artificial images as an assay for the kinds of image statistics to which the visual system is sensitive. To account for texture discrimination generally, we would have to extend our analysis to all kinds of image statistics, and also to cue combination between them, within, and across scales.

Our results provide evidence that among the universe of high-order statistics that can occur in synthetic images, the visual system is selectively sensitive to those that are informative in natural images. This finding suggests that an organizational principle recognized as applicable to simple image statistics and the sensory periphery also applies to complex image features and cortical visual processing: the brain invests resources to selectively extract those features that are informative about the structure of natural scenes. This principle predicts that visually salient third-order correlations, which are yet to measured, will be ones that are most informative about natural scenes.

Materials and Methods

Image Ensemble. Images were taken with a calibrated Nikon D70 camera, and comprise panoramic eye-level shots of a dry-season savannah habitat in the Okavango Delta, Botswana, during typical midday illumination. Trichromatic (red, green, and blue) images were converted into equivalent luminance images, by defining the luminance as proportional to the sum of the computed responses of the *L* and *M* cones. For details of calibration and image access, see ref. 7.

Synthetic Textures. Synthetic textures were constructed from a glider (a specified geometrical arrangement of pixels, $\mathcal{G}$) and a distribution over "glider colorings" (pixel intensities within the glider) $P_{\mathcal{G}}(\vec{\sigma})$. Given these data we selected a Q pixel glider within a texture patch and initialized $Q-1$ of its pixels randomly. We drew the Qth pixel according to the conditional distribution $P_{\mathcal{G}}(\sigma_Q|\sigma_1,\ldots,\sigma_{Q-1})$. We then shifted the glider and repeated the procedure for any unassigned pixels within the shifted glider. This procedure was repeated until all pixels in the image had been assigned intensities (25). The resulting texture was as random as possible subject to the constraint that the distribution of intensities in pixels arranged in the shape $\mathcal{G}$ will be $P_{\mathcal{G}}$ (26).

Analysis of Local Luminance Statistics. We selected 17 images with minimal portions of sky for the analysis (see *SI Appendix*). Pixels were discretized to 16 grayscale values ($\sigma_0 = 0\cdots15$) so that the distribution over intensities for each complete image was uniform. Then, the conditional distribution $P_R(\sigma_1|\sigma_0)$ of pixel intensities at radius R away from a randomly chosen central pixel of intensity σ_0 was sampled, for each σ_0. The values of R were chosen uniformly in $\log_2(R)$ for 18 values of R ranging from $R = 2^1 = 2$ to $R = 2^{9.5} = 724$ pixels. For each R, $5{\cdot}10^6$ pairs of pixels were included in the sample.

PCA of Luminance Distributions. We accumulated $P_R(\sigma_1|\sigma_0)$ for R and σ_0 and assembled the data into a 16×288 matrix (16 intensity levels for $\sigma_{1,0}$ at 18 different distances R; also see *SI Appendix*). To perform PCA, we subtracted the mean and then computed the covariance matrix of the resulting ensemble of histogram modulators. We diagonalized this matrix to find the eigenvalues and eigenvectors. The eigenvectors with the three largest eigenvalues are presented in *Results*. These eigenvectors were robust to variations in the strategy for sampling luminance distributions (see *SI Appendix*). The eigenvectors were also identical to those found by sampling intensities *within* (as opposed to *at*) a radius R of the central pixel.

Image Preprocessing for Isodipole Texture Analysis. Images were whitened by normalizing every Fourier component to the same magnitude; this flattened the power spectrum and removed second-order correlations, much like center-surround filtering in the retina. The resulting image is binarized so that black and white pixels are equal in number. Second-order correlations and luminance bias, averaged over the whole image, were thus removed, but residual correlations remain in local $R\times R$ image patches. Our analysis is scale invariant (checked by block-averaging the images prior to preprocessing—see *SI Appendix*).

ACKNOWLEDGMENTS. V.B. thanks the Aspen Center for Physics, and the IAS, Princeton for support as the Helen and Martin Chooljian Member. V.B. and J.D.V. thank the organizers of the *Perception to Action* workshop at the Institute for Advanced Studies, Jerusalem where this work was initiated. G.T. thanks Matthias Bethge for useful discussions. G.T., J.S.P., and V.B. were supported by National Science Foundation (NSF) Grants IBN-0344678 and EF-0928048 and National Institutes of Health (NIH) Grant R01 EY08124 and Grant T32-07035. J.D.V. was supported by NIH/National Eye Institute (NEI) Grants 2R01EY007977 and 2R01EY009314.

1. Barlow HB (1961) *Sensory Communication*, ed W Rosenblith (MIT Press, Cambridge, MA), pp 217–234.
2. Srinivasan MV, Laughlin SB, Dubs A (1982) Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society B (London)* 216:427–459.
3. Atick JJ, Redlich AN (1990) Towards a theory of early visual processing. *Neural Comput* 2:308–320.
4. Balasubramanian V, Sterling P (2009) Receptive fields and the functional architecture in the retina. *J Physiol* 587:2753–2767.
5. Atick JJ, Li Z, Redlich AN (1992) Understanding retinal color coding from first principles. *Neural Comput* 4:449–572.
6. Osorio D, Vorobyev M (1996) Color vision as an adaptation to fruigivory in primates. *Proceedings of the Royal Society B (London)* 263:593–599.
7. Garrigan P, et al. (2010) Design of a trichromatic cone array. *PLoS Comput Biol* 6: e1000677.
8. Borghuis BG, Ratliff CP, Smith RG, Sterling P, Balasubramanian V (2008) Design of a neuronal array. *J Neurosci* 28:3178–3189.
9. Liu YS, Stevens CF, Sharpee TO (2009) Predictable irregularities in retinal receptive fields. *Proc Nat'l Acad Sci USA* 106:16499–16504.
10. Laughlin SB (1981) A simple coding procedure enhances a neuron's information capacity. *Z Naturforsch* 36c:910–912.
11. Brenner N, Bialek W, van Steveninck RR (2000) Adaptive rescaling optimizes information transmission. *Neuron* 26:695–702.
12. Geisler WS (2008) Visual perception and the statistical properties of natural scenes. *Annu Rev Psychol* 59:167–192.
13. Oppenheim AV, Lim JS (1981) The importance of phase in signals. *Proceedings of the IEEE* 69:529–541.
14. Chubb C, Econopouly J, Landy MS (1994) Histogram contrast analysis and the visual segregation of IID textures. *J Opt Soc Am A* 11:2350–2374.
15. Chubb C, Landy MS, Econopouly J (2004) A visual mechanism tuned to black. *Vision Res* 44:3223–3232.
16. Julesz B, Gilbert EN, Victor JD (1978) VIsual discrimination of textures with identical third-order statistics. *Biol Cybern* 31:137–140.
17. Victor JD, Conte MM (1991) Spatial organization of nonlinear interactions in form perception. *Vision Res* 31:1457–1488.
18. Purpura KP, Victor JD, Katz E (1994) Striate cortex extracts higher-order spatial correlations from visual textures. *Proc Natl Acad Sci USA* 91:8482–8486.
19. Victor JD (1986) Isolation of components due to intracortical processing in the visual evoked potential. *Proc Natl Acad Sci USA* 83:7984–7988.
20. Richards W (1982) A lightness scale from image intensity distributions. *Applied Optics* 21:2569–2604.
21. Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. *J Opt Soc Am A* 4:2379–2394.
22. Julesz B (1981) Textons, the elements of texture perception, and their interactions. *Nature* 290:91–97.
23. Maddess T, Nagai Y (2001) Discriminating isotrigon textures. *J Vision* 1:151–151a.
24. Victor JD, Chubb C, Conte MM (2005) Interaction of luminance and higher-order statistics in texture discrimination. *Vision Res* 45:311–328.
25. Pickard DK (1980) Unilateral Markov fields. *Adv Appl Probab* 12:655–671.
26. Zhu SC, Wu Y, Mumford D (1998) Filters, random fields, and maximum entropy (FRAME): towards a unified theory for texture modeling. *International Journal of Computer Vision* 27:107–126.
27. Schneidman E, Still S, Berry MJ, 2nd, Bialek W (2003) Network information and connected correlations. *Phys Rev Lett* 91:238701.
28. Victor JD, Conte MM (1989) Cortical interactions in texture processing: scale and dynamics. *Visual Neruosci* 2:297–313.
29. Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381:607–609.
30. Bell AJ, Sejnowski TJ (1997) The "independent components" of natural scenes are edge filters. *Vision Res* 37:3327–3338.
31. Vinje WE, Gallant JL (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287:1273–1276.
32. Lythgoe JN (1980) *The ecology of vision* (Oxford Univerity Press, New York).
33. Portilla J, Simoncelli EP (2000) A parametric texture model based on joint statistics of complex wavelet coefficients. *Int J Comput Vision* 40:49–71.
34. Schwartz O, Simoncelli EP (2001) Natural signal statistics and sensory gain control. *Nat Neurosci* 4:819–825.
35. Karklin Y, Lewicki MS (2008) Emergence of complex cell properties by learning to generalize in natural scenes. *Nature* 457:83–87.