Method for estimating the relative contribution of phase and power spectra to the total information in natural-scene patches

David J. Field[1,*] and Damon M. Chandler[2]

[1]*Department of Psychology, Cornell University, Ithaca, New York 14853, USA*
[2]*School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, Oklahoma 74078, USA*
**Corresponding author: djf3@cornell.edu*

Received June 10, 2011; revised September 11, 2011; accepted October 22, 2011;
posted October 24, 2011 (Doc. ID 149099); published December 7, 2011

A wide variety of recent studies have argued that the human visual system provides an efficient means of processing the information in the natural environment. However, the amount of information (entropy) in the signal can be estimated in a number of ways, and it is has been unclear how much of the information is carried by the different sources of redundancy. The primary difficulty is that there has been no rational way to estimate the entropy of such complex scenes. In this paper, we provide a technique that uses a recent approach to estimating the entropy and dimensionality of natural scenes [D. M. Chandler and D. J. Field, J. Opt. Soc. Am. A **24**, 922–941 (2007)] to estimate the amount of information attributable to the power and phase spectra in natural-scene patches. By comparing the entropies of patches that have swapped phase spectra and fixed phase spectra, we demonstrate how to estimate both the amount of information in each type of spectrum and the amount of information that is shared by these spectra (mutual information). We applied this technique to small patches (4×4 and 8×8). From our estimates, we show that the power spectrum of 8×8 patches carries approximately 54% of the total information, the phase spectrum carries 56%, and 10% is mutual information ($54\% + 56\% - 10\% = 100\%$). This technique is currently limited to relatively small image patches, due to the number of patches currently in our collection (on the order of 10^6). However, the technique can, in theory, be extended to larger images. Even with these relatively small patches, we discuss how these results can provide important insights into both compression techniques and efficient coding techniques that work with relatively small image patches (e.g., JPEG, sparse coding, independent components analysis).

OCIS codes: 330.4060, 330.1880, 330.1800, 330.5510, 330.5020.

1. INTRODUCTION

In the last three decades, we have gained considerable insights into the properties of visual systems by considering the statistics of natural scenes (e.g., [1–6]). These approaches have demonstrated that many basic properties of the early visual system (both selectivity and tiling of visual neurons) contribute to make this neural representation an efficient code of the natural environment. Neural networks that minimize the dependencies between neurons using sparse coding or independent components analysis (ICA) produce neurons with basic linear properties quite similar to those of V1 neurons [4,7]. However, linear independent codes (i.e., factorial codes) cannot exist because the images of our environments are not a simple sum of independent functions. For natural scenes, it is quite likely that no code would be capable of removing all of the statistical dependencies.

One could argue that a goal of later stages of the visual system is to remove the statistical dependencies that remain after this early coding by V1. Indeed, efforts to model the temporal and neighborhood properties using efficient codes have provided more accurate accounts of V1 neurons (e.g., [8–11]). However, the full set of dependencies between pixels or neurons remains unknown. Some efforts have been made to capture the full statistical structure of high contrast 3×3 image patches [12]. The work provides an elegant and detailed study of this nine-dimensional state space. However, such results also point out the complexity of even these small patches.

Knowing the true information of a data set would provide an idea of just how compressible the data set is. Although lossy compression techniques like JPEG [13] and JPEG-2000 [14] have proven to be popular and successful strategies, there remains the question of the lower bound for lossless compression. Recently, we have described a method of estimating the true redundancy of a signal [15]. For images that are larger than 4×4, the method does involve some assumptions regarding how one might optimally extrapolate the available data. However, the technique is straightforward and involves relatively simple calculations and extrapolations of known data. Hosseini *et al.* [16] more recently described a technique for estimating the redundancy of natural scenes based on a measure of the mutual information between a pixel and its causal neighborhood of increasing size. Both the technique of Hosseini *et al.* [16] and our technique in [15] yielded an entropy rate of approximately 3.0 bits/pixel for natural scenes.

In this paper, we use our technique to investigate another important aspect of natural scenes: how much information is carried in the power spectrum versus the phase spectrum and how much information is common to both spectra (mutual information).

A. Importance of Power versus Phase for Visual Perception and Coding

It has been previously noted that the power spectrum of natural scenes falls with frequency f as approximately $1/f^{2\alpha}$ (amplitude spectrum by $1/f^{\alpha}$) [1,17]. Estimates of the parameter α for any given population typically vary from 0.7 to 1.5 with averages in the range of approximately 1.1 [2,18]. Such results imply that there exist factors that keep the images from true scale invariance (e.g., the limited depth of field of the camera, object motion, or choice of images). It has long been argued that the phase spectrum carries most of the information regarding an image (e.g., [19–21]). Exchanging the phase and power spectra of two natural scenes will typically result in images that appear much closer to the image with the correct phase spectrum [20,21]. However, this does not necessarily imply anything significant about visual processing (e.g., that we are more sensitive to the phase spectrum). If natural images have similar power spectra, then it would not be surprising that exchanging the power spectra has little effect on the perceived images. Rather, this would simply suggest that the power spectrum contained little information. As expected, Millane and Hsiao [21] demonstrated that for 512×512 images, the RMS difference between images is much larger when the phase spectra are swapped relative to when the amplitude spectra are swapped.

Morgan *et al.* [22] argued that for small image patches, the perceived image structure is well described by the power spectra. For larger image patches, the phase spectrum dominates. We will support this point in this paper and argue that this property is due to the sparse structure of images and the nature of the information in small patches. Specifically, most small image patches (e.g., 8×8 pixels) contain blank regions, single edges, or bits of texture. The power spectra for these small patches can be quite informative regarding which of these classes is present. For example, if a single edge is present, the power spectrum will provide information regarding the orientation of the edge. However, as the images become significantly larger, the images will typically contain a significant number of edges as well as textures and blank regions. For these larger patches, the phase spectrum is determined by the relative combination and positions of these features. With edges at a wide range of scales, the power spectrum converges toward the characteristic $\approx 1/f^2$ power spectrum. Indeed, it has been argued that the $1/f^2$ power spectrum is produced from a self-similar sum of local features [3,23,24].

As noted above, natural scenes show a consistent falloff in their power spectra. Even art, both landscape and abstract, shows this characteristic $1/f^2$ falloff (e.g., [25]). However, this consistency in the spectra should not lead one to conclude that the power spectrum contains no information. Certainly, if all spectra were truly identical, no information would be provided. However, both the rotationally averaged spectra [1–3,18] and the full two-dimensional spectra of natural scenes show significant variation from image to image (e.g., [26,27]). Furthermore, Torralba and Oliva [26] demonstrated that different image classes (photographs of street scenes, vistas, indoor scenes, etc.) have significantly different power spectra, and this difference allows for a significant degree of classification. Their results do not necessarily imply that the power spectrum is explicitly used when human observers classify images. It is possible that the variations in the power spectrum are highly correlated with the variations in the phase spectrum. In such a case, the power spectrum would provide little or no additional information to that found in the phase spectrum. This idea is further supported by the result that randomizing the phase spectrum of an image typically loses its identity, but fixing the power spectrum to a constant slope typically produces only relatively minor distortions.

It should also be noted that the power spectrum provides an alternative description of the pairwise correlations in an image (the power spectrum is the Fourier transform of the autocorrelation function). Therefore, techniques that rely on only the correlations (e.g., linear Hebbian learning) will make use of the statistics found only in the power spectrum. In contrast, techniques such as sparse coding and ICA (e.g., [4,7]) are dependent on the higher-order structure found in the phase spectrum. For these methods, the power spectrum is whitened to remove the pairwise correlations in the image.

In summary, the power spectrum of natural scenes is not identical from image to image. The spectrum contains information, and it has been demonstrated that this information is correlated with different image classes (e.g., [26]). However, the specific amount of information (entropy) that is attributable to the power and phase spectra of natural scenes, and the amount of information that is common to both spectra (mutual information), has yet to be reported. In this paper, we make an estimate of these entropies for patches of natural scenes.

Determining these entropies can provide insight into the relative effectiveness of one neural coding strategy versus another. For example, if most of the information in an image patch is due to the phase spectrum, then techniques that rely only on the power spectrum (such as linear Hebbian learning) would be expected to provide a less efficient code compared to strategies (such as ICA and sparse coding) that operate largely on the phase spectrum, and vice versa. On the other hand, if there is a large amount of mutual information between the power and phase spectra, then both types of coding strategies could, in theory, be equally effective, since knowledge of the power (or phase) spectrum also provides knowledge of the phase (or power) spectrum. Similar insights may also be gained regarding the performance of various image-processing and computer-vision algorithms. Determining these entropies may help explain why algorithms that utilize only the power spectrum or only the phase spectrum can still be successful at tasks that seemingly require the other type of spectrum.

B. How Much Information Is Attributable to Power versus Phase, and How Much Information Is Common to Both Spectra?

In this paper, we investigate the amount of information attributable to the power spectrum, the amount of information attributable to the phase spectrum, and the amount of information that is common to these spectra (mutual information). To this end, we apply the entropy estimator from [15] to compare the entropy of natural-scene patches to the entropies of patches created via two strategic manipulations of the phase spectra.

Fixed-phase manipulation: To investigate the amount of information attributable to the power spectrum, we create "fixed-phase" natural-scene patches in which the phase

spectra of all patches have been set to a single common phase spectrum. Here, because all patches have the same phase spectrum, there is no variability in (and thus no information attributable to) the phase spectrum. As a result, the entropy of these patches will reveal the amount of information in natural-scene patches attributable to the power spectrum.

Hybrid manipulation: To investigate the amount of information attributable to the phase spectrum and the amount of mutual information, we create "hybrid" natural-scene patches in which the phase spectrum of each patch has been replaced with the phase spectrum of another natural-scene patch selected at random. Here, each patch maintains its original power spectrum, but its phase spectrum now corresponds to that of another, randomly selected natural-scene patch. This manipulation removes any statistical dependence that might exist between the power and phase spectra. As a result, the entropy of these hybrid patches will reveal the total information that would exist if there were no mutual information between power and phase. The amount of information attributable to the phase spectrum can then be computed by comparing the entropy of the hybrid patches to the entropy of the fixed-phase patches. In addition, the amount of mutual information can be computed by comparing the entropy of the hybrid patches to the entropy of the original natural-scene patches.

The investigation we describe here uses relatively small patches (4×4 and 8×8 pixels). However, the technique employed can in theory be extended to larger images if the population of test images is sufficiently large. We believe our method provides a new approach to determining the relative contribution of different sources of information in an image—as well as showing the mutual information between those sources.

We also believe that the results for small patches is inherently interesting for two reasons: First, the method allows the investigation of local scene properties. Many forms of compression (e.g., JPEG [13]) and efficient neural coding strategies (e.g., ICA [7] and sparse coding [4]) are typically applied to relatively small patches. We believe our approach can provide insights into the different forms of information used by these efficient coding techniques. Second, by using smaller patches, our image population can be relatively large (e.g., 10^6 patches, whereas most databases contain at most 10^3 full-sized images). This large number allows a more accurate test of the methods described in this paper.

2. THEORY AND METHODS

This section describes an overview of the theory and experimental methods used to investigate the relative contributions of amplitude and phase information in natural-scene patches. First, we review the entropy estimator described in [15] (Sections 2.A and 2.B). Next, we describe how this technique can be used to estimate the information content in the power spectrum and phase spectrum and the amount of information shared between these spectra (mutual information) (Section 2.C). Finally, we provide details of the specific stimuli and procedures employed (Sections 2.D and 2.E, respectively).

A. Proximity Distributions

The technique we use for estimating information is based on the extrapolation of proximity distributions. As described in [15], proximity distributions can be used to estimate both the entropy and the dimensionality of an image set.

To measure the proximity distribution for a given set of image patches, the patches are first randomly divided into two groups:

- Group $\mathcal{T}$, which consists of patches that serve as the to-be-matched "target" patches.
- Group $\mathcal{N}$, which consists of patches that serve as the target's potential matches.

For each target patch in Group $\mathcal{T}$, an exhaustive search is performed to find the patch in Group $\mathcal{N}$ whose pixel values are closest in Euclidean distance to the pixel values of the target patch (the so-called "nearest neighbor" to the target patch). The proximity distribution is the average log nearest-neighbor distance computed as a function of the number of patches in Group $\mathcal{N}$ (the average is taken over all target patches).

Specifically, let T and N denote the number of patches in Groups $\mathcal{T}$ and $\mathcal{N}$, respectively. Let $\mathbf{X}^{(\mathcal{T}_t)}$, $t \in [1, T]$, denote the tth target patch, and let $\mathbf{X}^{(\mathcal{N}_n)}$, $n \in [1, N]$, denote the nth patch in Group $\mathcal{N}$. The Euclidean distance, $D_{n,t}$, between the pixel values of $\mathbf{X}^{(\mathcal{T}_t)}$ and the pixel values of $\mathbf{X}^{(\mathcal{N}_n)}$ is given by

$$D_{n,t} = \|\mathbf{X}^{(\mathcal{T}_t)} - \mathbf{X}^{(\mathcal{N}_n)}\|_{L_2} = \left(\sum_{i=1}^{k} (X_i^{(\mathcal{T}_t)} - X_i^{(\mathcal{N}_n)})^2\right)^{1/2}, \quad (1)$$

where X_i denotes the ith pixel of $\mathbf{X}$ and k denotes the total number of pixels (e.g., $k = 64$ for an 8×8 patch).

For each target patch $\mathbf{X}^{(\mathcal{T}_t)}$, we perform an exhaustive search for the minimum Euclidean distance between $\mathbf{X}^{(\mathcal{T}_t)}$ and each of the N patches in Group $\mathcal{N}$. This nearest-neighbor distance, denoted by $D^*_{N,t}$, is given by

$$D^*_{N,t} = \min_{n \in [1,N]} \{D_{n,t}\}. \quad (2)$$

Equation (2) is computed for each target patch, $\mathbf{X}^{(\mathcal{T}_t)}$, $t \in [1, T]$, resulting in a collection containing T values of $D^*_{N,t}$.

Finally, the average log nearest-neighbor distance, denoted by $E\{\log_2 D^*_N\}$, is estimated via the sample mean of the collection of $\log_2 D^*_{N,t}$ values:

$$E\{\log_2 D^*_N\} \approx \frac{1}{T} \sum_{t=1}^{T} \log_2 D^*_{N,t}. \quad (3)$$

The proximity distribution is $E\{\log_2 D^*_N\}$ computed as a function of N. In particular, a plot of $E\{\log_2 D^*_N\}$ versus N can reveal how the average log nearest-neighbor distance changes as more patches are added to Group $\mathcal{N}$. This evolution of $E\{\log_2 D^*_N\}$ as a function of N has been shown to be an effective tool for estimating the entropy and dimensionality of a relatively high-dimensional data set given a relatively small number of samples from that data set [15,28–31].

Figure 1 shows proximity distributions for 8×8 patches of Gaussian white noise, noise with an amplitude spectrum of $1/f$, and natural scenes from the van Hateren database [32]. The horizontal axis denotes N (the number of patches in Group $\mathcal{N}$). The vertical axis denotes the corresponding $E\{\log_2 D^*_N\}$. As N increases, the chance of finding a good

match (small Euclidean distance) to a given target patch improves, i.e., the Euclidean distance between the patches decreases. This general feature is not surprising; however, both the rate at which this function decreases and the asymptote allow us to estimate the dimensionality and entropy of this image population. In particular, the dimensionality of the image population is simply the negative inverse of the slope of the proximity distribution, and the Shannon information (entropy) is a function of the asymptote [15]; see also Appendix A.

Figure 1 also shows the proximity distribution for Gaussian white noise and for noise with a $1/f$ amplitude spectrum ($1/f^2$ power spectrum). The steeper proximity functions and lower asymptotes for the natural scenes relative to the noise are due to the relative redundancy of these three image classes. For data sets that are more redundant, one is more likely to find a patch that is closer to the target patch. The white noise population has the highest entropy (lowest redundancy) of any population—and the proximity distribution falls at the slowest possible rate. In [15], we demonstrated that the entropy of a population of images can be estimated from this distribution; an overview of this process is described in the following section.

B. Estimating Entropy and Dimensionality from the Proximity Distributions

Natural scenes are fundamentally noise-limited. Photon noise, sensor noise, and the noise from digital quantization limit the accuracy of any nearest-neighbor match. This noise limit forces each proximity distribution to converge on the slope for noise. For any class of k-pixel images $\mathbf{X} = [X_1, X_2, \ldots, X_k]$ (e.g., 4×4 images with $k = 16$), the slope asymptotes at $-1/k$ (e.g., $-1/16$). When the slope converges on this noise limit, the height of the proximity distribution curve determines the entropy of the image class. That is, for 64-pixel images (8×8), both curves will converge on a slope of $-1/64$. A 9-pixel image (3×3) will converge on a slope of $-1/9$.

The relative dimensionality ($-1/\text{slope}$) at different values of N may seem surprising. For example, as shown in Fig. 1, the white noise slope is not a slope of $-1/64$ when N is small. It converges on a slope of $-1/64$ when N is sufficiently large, but it does not begin with that slope. One property of a high-dimensional sphere is that most of the volume is found in the regions near the edge of the sphere. The higher the dimensionality of the sphere, the higher the proportion of the data t hat will be found near the edge. Therefore, when a high-dimensional sphere is sampled with a small number of samples, the sphere appears lower-dimensional. Only after a sufficient number of samples do we see the true dimensionality. In line with this, the proximity curve begins with a steeper curve but eventually converges on a slope that corresponds to the dimensionality.

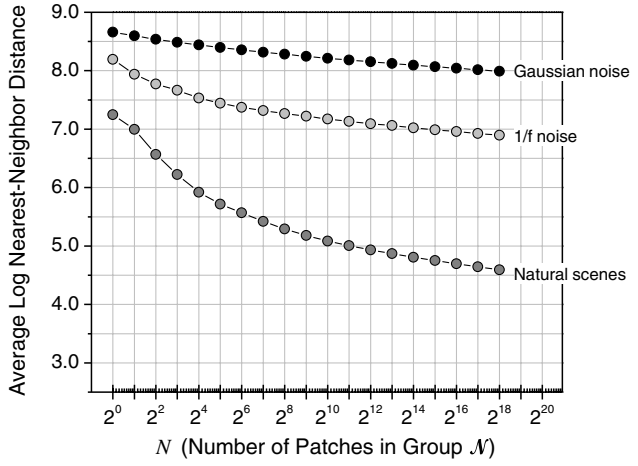

Fig. 1. Proximity distribution functions for 8×8 patches of Gaussian white noise, noise with an amplitude spectrum of $1/f$, and natural scenes. The horizontal axis denotes N (the number of patches in Group $\mathcal{N}$); the vertical axis denotes the corresponding $E\{\log_2 D_N^*\}$ estimated via a sample mean over all target patches. Black circles: Gaussian white noise; light-gray circles: $1/f$ noise; gray circles: natural scenes.

When the slope of the proximity distribution has converged, the (differential) entropy, h, can be approximated by

$$h \approx -kE\{\log_2 D_N^*\} + \log_2\left(\frac{A_k N}{k}\right) + \frac{\gamma}{\ln 2}, \tag{4}$$

where γ is the Euler constant and $A_k = k\pi^{k/2}/\Gamma(\frac{k}{2}+1)$ denotes the surface area of a k-dimensional hypersphere (see Appendix A and [29]).

This technique allows an estimate of the entropy of a data set using significantly fewer data samples than that required for a direct calculation. For a direct calculation, we would need to calculate the probability of every possible combination of pixels. Unfortunately, the total number of possible combinations increases exponentially with the image size, and it also depends on the possible shades of gray; thus, a direct calculation would require a prohibitively large number of samples, even for very small image patches. In general, for images containing $P \times P$ pixels and G shades of gray, the total number of possible combinations (possible patterns in the $P \times P$ array) is $2^{P \times P \times \log_2 G}$. For a 3×3 image with 256 shades of gray, the number of possible combinations is $2^{3\times3\times8} = 2^{72} = 4,722,366,482,869,645,213,696$ possible combinations of pixels. Even if it were possible to acquire and store this many patches, one would still need *multiple* instances of each combination of pixels in order to properly estimate the *probability* of each combination. In comparison, Eq. (4) provides an estimate of the entropy with just $2^{17} = 131,072$ samples [15]. Note that we are in no way implying that the visual system operates using this particular digital representation of grayscale values; rather, the value of $G = 256$ is used here only to exemplify the prohibitive number of patches required for a direct calculation of entropy.

Whereas for images up to approximately 3×3, the slope of the proximity distribution converges on the noise slope when N reaches $N \approx 2^{17}$, for larger images, the slope has not yet converged, even with an N of several million samples. This fact requires us to extrapolate the curve to provide an estimate of the entropy. In [15], we provided three different approaches to extrapolating the proximity distribution functions. To emphasize that this approach uses extrapolations, the term *XEntropy* was introduced. Our best estimate used something we called *XEntropy C*. Our estimate for the entropy rate of the 8×8 image patch population in [15] was 2.9 bits per pixel; i.e., we argued that the best compression rate that could be achieved with this image population is 2.9 bits per pixel. A very similar estimate of approximately 3.0 bits/pixel was reported by Hosseini *et al.* [16] using a very different method of estimating

entropy based on the mutual information between a pixel and its causal neighborhood of increasing size.

In this paper, we will again use XEntropy C to estimate the relative entropy of data sets. Here, we will not focus on the absolute entropy, but the difference in entropy between natural-scene patches containing strategic manipulations of their phase spectra. Even if our extrapolated entropy significantly underestimates or overestimates the entropy, as long as the error is in a consistent direction, this relative entropy still holds.

C. Estimating the Entropies of Amplitude and Phase Spectra

We seek the amount of information attributable to the amplitude (power) spectrum and the phase spectrum and the amount of mutual information that is common to these spectra. To this end, patches from three types of images were employed:

1. Natural-scene patches obtained from the van Hateren database [32].
2. Manipulated natural-scenes patches in which the phase spectra of all patches have been set to the same common phase spectrum.
3. Manipulated natural-scene patches in which the phase spectrum of each patch has been replaced with the phase spectrum of another natural-scene patch selected at random.

The first image class is an extended version of the data used in [15]. The second image class allows us to determine the amount of information in the amplitude spectrum since, by forcing all patches to have the same phase spectrum, there is no information in the phase spectra for these images. The third class allows us to determine the total information that these image classes would have if there were no mutual information between the amplitude and phase spectra; swapping the phase spectrum of each patch with the phase spectrum of a randomly selected natural-scene patch removes any dependence between the amplitude and phase spectra. By comparing the results of this latter class of image to the original images, we can determine the amount of mutual information in the phase and power spectra.

Specifically, let $H(A)$ denote the entropy of natural scenes attributable to the amplitude spectrum. Let $H(P)$ denote the entropy of natural scenes attributable to the phase spectrum. Let $H(A,P)$ denote the entropy of natural scenes (the joint entropy of A and P).

If the amplitude and phase spectra are statistically independent, then

$$H(A,P) = H(A) + H(P). \tag{5}$$

However, if there exists a statistical dependency between the amplitude and phase spectra, then $H(A,P) \le H(A) + H(P)$; specifically,

$$H(A,P) = H(A) + H(P) - I(A;P), \tag{6}$$

where $I(A;P) \ge 0$ denotes the mutual information between A and P. Note that $I(A;P) = 0$ only if A and P are independent.

The contributions of $H(A)$ and $H(P)$ to the joint entropy $H(A,P)$ are shown graphically in Fig. 2. If the amplitude and phase spectra are independent of each other [see Fig. 2(a)], the total entropy $H(A,P)$ is simply the sum of the entropy due to the amplitude spectrum $H(A)$ and the entropy due to the phase spectrum $H(P)$. On the other hand, if A and P are statistically dependent and therefore possess some amount of mutual information $I(A;P)$ [see Fig. 2(b)], then the total entropy $H(A,P)$ is determined by $H(A)$, $H(P)$, and the amount of overlap $I(A;P)$.

We seek estimates of $H(A)$, $H(P)$, and $I(A;P)$. To this end, we compare the entropy of patches from natural scenes $H(A,P)$ with the entropies of patches from scenes whose phase spectra have been strategically manipulated. Specifically, we created two conditions: (1) *fixed-phase* natural scenes and (2) *hybrid* natural scenes (see Fig. 3).

The fixed-phase natural scenes are those in which the phase spectrum of each patch has been fixed to the same constant phase spectrum. This manipulation results in the configuration shown in Fig. 3(b). Because all patches of all images now have the same phase spectrum, the only variation between these patches is due to the amplitude spectrum. Thus, the entropy of these fixed-phase patches $H_{\text{fixed-phase}}(A,P)$ is the entropy due to the amplitude spectrum: $H_{\text{fixed-phase}}(A,P) = H(A)$.

What we call "hybrid" natural scenes are those in which the phase spectrum of each patch has been replaced with the phase spectrum of another, randomly selected natural-scene patch, which creates the configuration shown in Fig. 3(c). This manipulation enforces statistical independence between the amplitude and phase spectra, while maintaining the individual entropies $H(A)$ and $H(P)$. Thus, the entropy of these hybrid patches $H_{\text{hybrid}}(A,P)$ is the sum of the entropies due individually to the amplitude spectrum and to the phase spectrum: $H_{\text{hybrid}}(A,P) = H(A) + H(P)$.

Via these manipulations, $H(P)$ can be computed by taking the difference between the entropy found in the hybrid condition and the entropy found in the fixed-phase condition:

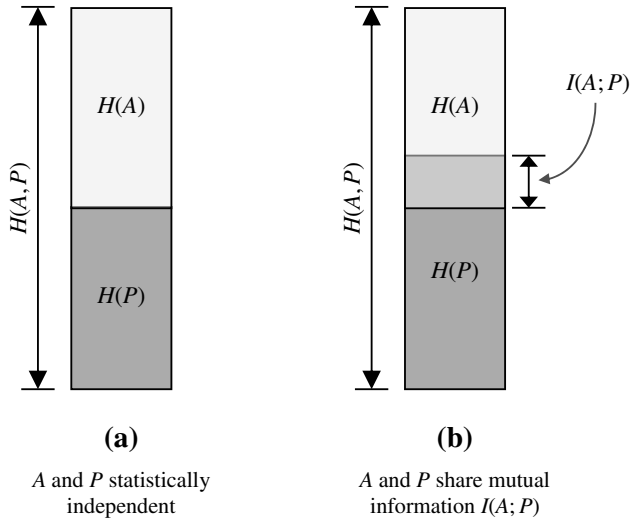

Fig. 2. Graphical illustration of the contributions of the information in the amplitude spectrum $H(A)$ and the information in the phase spectrum $H(P)$ to the total entropy $H(A,P)$. (a) If the amplitude and phase spectra are assumed to be independent of each other, then $H(A,P) = H(A) + H(P)$. (b) Illustration of the contributions of $H(A)$ and $H(P)$ to the total entropy when the amplitude and phase are statistically dependent and therefore possess some amount of mutual information (overlap) $I(A;P)$.

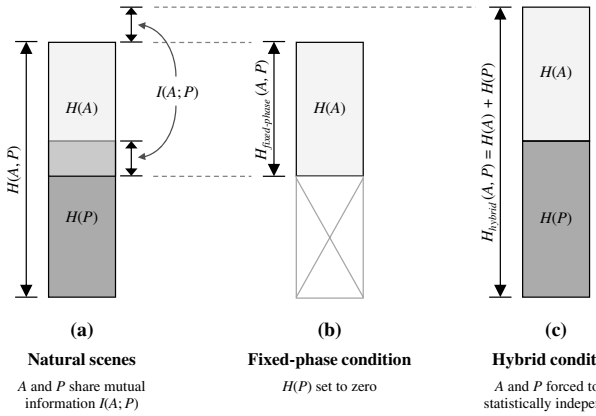

Fig. 3. Graphical illustration of the strategy used in the current study to compute $H(A)$, $H(P)$, and $I(A;P)$. (a) The configuration shows the assumed entropy of natural scenes $H(A,P)$ in which $H(A)$ and $H(P)$ possess some amount of mutual information (overlap) $I(A;P)$. (b) The configuration is achieved by setting the phase spectrum of all patches to a fixed-phase spectrum; this operation removes all variations due to phase, which forces $H(P) = 0$, and therefore $H(A)$ is given by the entropy of the fixed-phase patches [i.e., $H(A) = H_{\text{fixed-phase}}(A,P)$]. (c) The configuration is achieved by creating hybrid images in which the phase spectrum of each patch is replaced with the phase spectrum of another natural-scene patch selected at random; this operation removes any statistical dependence between amplitude and phase while maintaining the individual quantities $H(A)$ and $H(P)$. The entropy of the hybrid patches is therefore $H_{\text{hybrid}} = H(A) + H(P)$, which allows us to compute both $H(P)$ and $I(A;P)$ via $H(P) = H_{\text{hybrid}}(A,P) - H_{\text{fixed-phase}}(A,P)$ and $I(A;P) = H_{\text{hybrid}}(A,P) - H(A,P)$.

$$H(P) = H_{\text{hybrid}}(A,P) - H_{\text{fixed-phase}}(A,P), \tag{7}$$

$$= [H(A) + H(P)] - H(A). \tag{8}$$

Then, given $H(A,P)$ (natural scenes), $H(A)$ (fixed-phase), and $H(P)$, Eq. (6) can be rearranged as follows to compute the amount of mutual information between amplitude and phase:

$$I(A;P) = H(A) + H(P) - H(A,P). \tag{9}$$

The following section provides details of the natural scenes and manipulated (fixed-phase, hybrid) scenes used in this study.

D. Experimental Stimuli

One hundred thirty-seven digitized natural scenes were selected at random from the van Hateren database [32]. The original images were of size 1536×1024 and contained 16 bit pixel values. A 1024×1024 section was cropped from each image, and then the pixel values of that 1024×1024 section were converted to a floating-point representation. The (real-valued) pixels were offset and scaled to span the range 0–255 and then quantized to 8 bits (256 levels) of grayscale resolution via uniform scalar quantization [33] in which real-valued pixel x was mapped to its quantized (discrete-valued) version x_Δ via $x_\Delta = \lfloor x + \frac{1}{2} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the floor operator.

The 1024×1024 8 bit images were divided into patches in a nonoverlapping, sequential raster-scan order starting from the top-left corner of the image. Two patch sizes were employed in this study: 4×4 patches and 8×8 patches. The 137 images yielded 8,978,432 4×4 natural-scene patches and 2,244,608 8×8 natural-scene patches. Figure 4 depicts a random subset of the stimuli used in this study (only 8×8 patches are shown). The normal natural-scene patches are shown in the left part of Fig. 4, the fixed-phase patches are shown in the middle, and the hybrid patches are shown in on the right.

Patches for the fixed-phase condition were created from the natural-scene patches by replacing the phase spectrum of each patch with the phase spectrum obtained from a patch containing just the central pixel set to a value of 255 and all other pixels set to zero (see Fig. 5). (The specific phase spectrum employed here is not important; rather, the key is to assign this same phase spectrum to all patches.) Patches

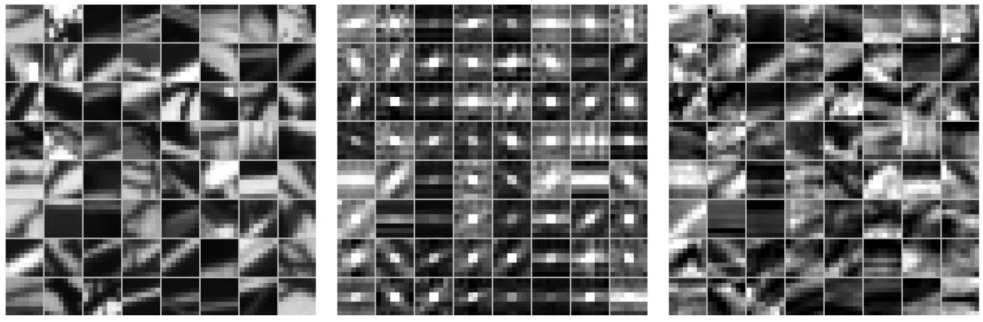

Fig. 4. Stimuli used in the experiment consisted of 4×4 and 8×8 natural-scene patches, fixed-phase patches, and hybrid patches (only 8×8 patches are shown). Each image depicts a random subset of sixty-four 8×8 patches. Left: Normal 8×8 natural-scene patches. Middle: Fixed-phase 8×8 patches generated by assigning the same fixed-phase spectrum to all patches; the amplitude spectrum was not adjusted. Right: Hybrid 8×8 patches generated by replacing the phase spectrum of each patch with the phase spectrum of another, randomly selected 8×8 natural-scene patch; the amplitude spectrum was not adjusted.

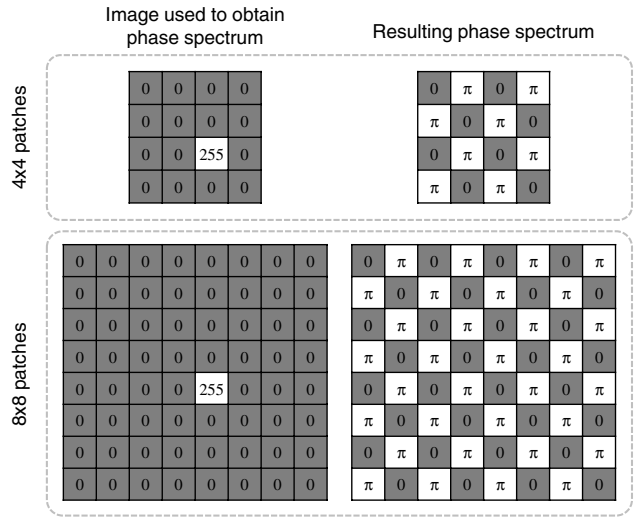

Fig. 5. The fixed-phase spectrum used for all patches in the fixed-phase condition was generated by computing the discrete Fourier transform of a 4×4 or 8×8 patch consisting of only the central pixel set to a value of 255. The phase spectra $\angle\mathbf{Z}$ shown on the right were computed via $\angle\mathbf{Z} = \tan^{-1}\left(\frac{\Im\{\mathbf{Z}\}}{\Re\{\mathbf{Z}\}}\right)$, where $\mathbf{Z}$ denotes the discrete Fourier transform of the central-pixel-on patches shown on the left. (The specific phase spectrum employed here is not important; rather, the key is to assign this same phase spectrum to all patches.)

for the hybrid condition were created from the natural-scene patches by replacing the phase spectrum of each patch with the phase spectrum obtained from another natural-scene patch selected at random.

Specifically, the fixed-phase and hybrid patches were created as follows: let $\mathbf{x}$ denote a 4×4 or 8×8 natural-scene patch, and let $\mathbf{y}$ denote a corresponding fixed-phase or hybrid patch. For each natural-scene patch $\mathbf{x}$, a two-dimensional discrete Fourier transform (DFT) was applied to obtain the spectrum $\mathbf{X}$. The amplitude spectrum $|\mathbf{X}|$ was then computed via

$$|\mathbf{X}| = \sqrt{\Re\{\mathbf{X}\}^2 + \Im\{\mathbf{X}\}^2}, \tag{10}$$

where $\Re\{\mathbf{X}\}$ and $\Im\{\mathbf{X}\}$ denote the real and imaginary components of $\mathbf{X}$, respectively. Next, the spectrum of the fixed-phase or hybrid patch, denoted by $\mathbf{Y}$, was synthesized via

$$\mathbf{Y} = |\mathbf{X}|e^{j\angle\mathbf{Z}}, \tag{11}$$

where $\angle\mathbf{Z} = \tan^{-1}(\frac{\Im\{\mathbf{Z}\}}{\Re\{\mathbf{Z}\}})$ denotes either the fixed-phase spectrum shown in Fig. 5 (for the fixed-phase condition) or the phase spectrum obtained from another randomly selected natural-scene patch (for the hybrid condition). Finally, a two-dimensional inverse DFT was applied to $\mathbf{Y}$ to obtain the fixed-phase or hybrid patch $\mathbf{y}$.

E. Nearest-Neighbor Search Procedure

The average log nearest-neighbor distance $E\{\log_2 D_N^*\}$ was estimated by using an exhaustive, brute-force search procedure as described in Section 2.A. To facilitate the implementation of the search, the patches for each condition were randomly separated into two groups: Group $\mathcal{T}$ (target patches) and Group $\mathcal{N}$ (patches to serve as potential matches to the target patches).

- *Group* $\mathcal{T}$: For the 4×4 patches, $T = 65,536$ target patches, selected at random, were placed into Group $\mathcal{T}$. For the 8×8 patches, $T = 16,384$ target patches, also selected at random, were placed into Group $\mathcal{T}$. These numbers of patches correspond to approximately 0.7% of the total number of 4×4 and 8×8 patches [34].
- *Group* $\mathcal{N}$: For both patch sizes, N randomly selected patches were placed into Group $\mathcal{N}$. The value of N was selected to take on power-of-two values up to 2^{19}, i.e., $N = 1, 2, 4, 8, 16, 32, \ldots, 2^{19}$. Thus, the number of patches in Group $\mathcal{N}$ was made variable and ranged from $N = 1$ to $N = 2^{19}$. We selected power-of-two values of N to obtain equally spaced data points on the proximity distribution in which the x axis (N) is plotted on a $\log_2$ scale.

As mentioned in Section 2.A, for each value of N, Eq. (3) was used to estimate the average log nearest-neighbor distance $E\{\log_2 D_N^*\}$. This estimate of $E\{\log_2 D_N^*\}$ was repeated for ten trials, where each subsequent trial began by replacing the T patches in Group $\mathcal{T}$ with another T randomly selected patches, and by replacing the N patches in Group $\mathcal{N}$ with another N randomly selected patches.

3. RESULTS AND ANALYSIS

As mentioned in Section 2.A, the proximity distribution specifies the average log nearest-neighbor distance $E\{\log_2 D_N^*\}$ as a function of N. The corresponding relative dimensionality specifies the inverse of the (negative) slope of the proximity distribution at each value of N. In this section, we first present and interpret the proximity distributions and relative dimensionality curves, and then we present and analyze the resulting entropy estimates.

A. Proximity Distribution and Relative Dimensionality

Figure 6 depicts the resulting proximity distribution curves for 4×4 and 8×8 patches. In each graph, the horizontal axis denotes N (the number of patches in Group $\mathcal{N}$), and the vertical axis denotes the corresponding value of $E\{\log_2 D_N^*\}$, averaged over all ten trials conducted for each value of N. Error bars denote ± 1 standard deviation of this average. For reference, Fig. 6 also shows data for 4×4 and 8×8 patches of Gaussian white noise; these data were computed analytically via Eq. (10) from [15].

Two important observations can be made from the proximity distributions for both 4×4 and 8×8 patches. First, the proximity distribution curve for the fixed-phase condition lies below the curve for natural scenes, despite the fact that both types possess the same amplitude spectra (and therefore the same degree of spatial correlations). These data indicate that, on average, fewer templates are required to describe a fixed-phase scene to the same level of accuracy as that achieved for a normal natural scene. As expected, by fixing the phase spectra of the image patches, we make the image set more redundant.

Second, the proximity distribution curve for the hybrid condition lies above the curve for the natural scenes, which indicates that more templates are required on average to describe a hybrid scene to the same level of accuracy as that achieved for a normal natural scene. Recall that in the hybrid condition, the amplitude and phase spectra were (by design) statistically independent. If the same were true for natural scenes, the two

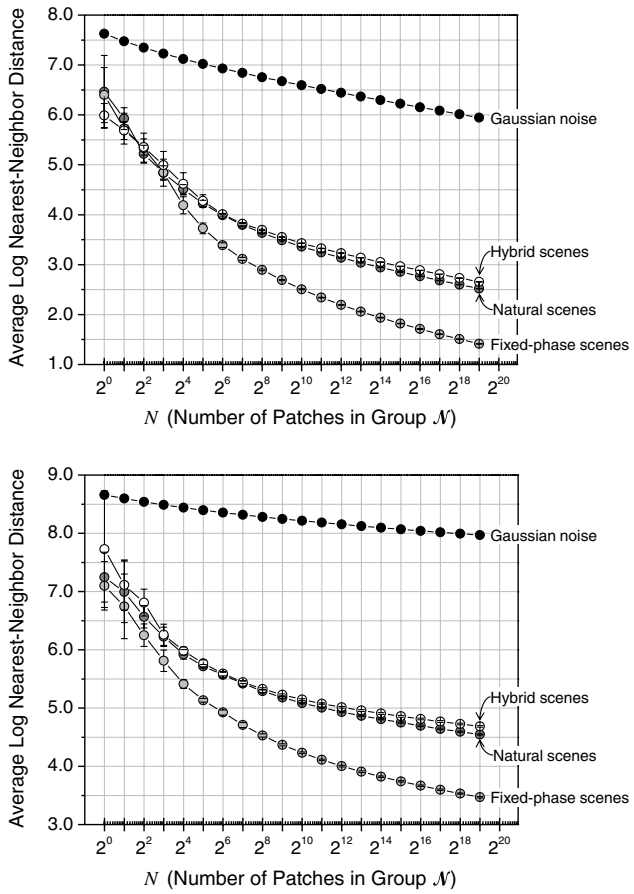

Fig. 6. Proximity distribution curves for 4×4 (top) and 8×8 (bottom) patches. In each graph, the horizontal axis denotes N (the number of patches in Group $\mathcal{N}$) and the vertical axis denotes the corresponding $E\{\log_2 D_N^*\}$ estimated via a sample mean over all target patches (average of 10 randomizations of Groups $\mathcal{T}$ and $\mathcal{N}$). Error bars denote ± 1 standard deviation. The data for Gaussian noise, provided for reference, were computed analytically via Eq. (10) from [15].

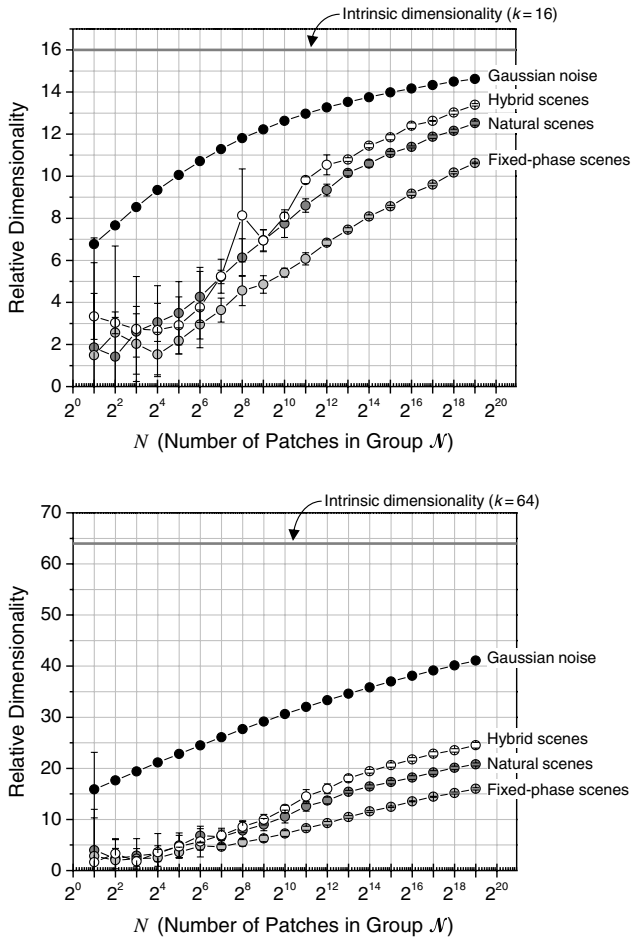

Fig. 7. Relative dimensionality curves for 4×4 (top) and 8×8 (bottom) patches. In each graph, the horizontal axis denotes N (the number of patches in Group $\mathcal{N}$) and the vertical axis denotes the corresponding relative dimensionality computed via $-d \log_2(N)/dE\{\log_2 D_N^*\}$ at each N. Error bars denote ± 1 standard deviation. The solid gray line denotes the intrinsic dimensionality of $k = 16$ (4×4) or $k = 64$ (8×8). The data for Gaussian noise were computed analytically via Eq. (10) from [15].

proximity distribution curves would be identical. The fact that the proximity distribution curve for natural scenes lies below the curve for the hybrid condition indicates the presence of mutual information between the amplitude and phase spectra in natural scenes.

Figure 7 shows the relative dimensionality curves for 4×4 and 8×8 patches. In each graph, the horizontal axis denotes N and the vertical axis denotes the inverse of the (negative) slope of the proximity distribution (i.e., $-d \log_2(N)/dE\{\log_2 D_N^*\}$), estimated via a difference in successive pairs of $E\{\log_2 D_N^*\}$. For reference, Fig. 7 also shows the relative dimensionality data for patches of Gaussian white noise computed analytically from Eq. (10) of [15].

Notice from both relative dimensionality graphs that for most values of N (in particular, for $N > 2^8$), fixed-phase patches appear lower-dimensional than natural-scene patches, and natural-scene patches appear lower-dimensional than hybrid patches, despite the fact that all of the 4×4 or 8×8 patches have the same intrinsic dimensionality of, respectively, $k = 16$ and $k = 64$. For the 4×4 patches, at $N = 2^{19}$ samples, the dimensionalities are approximately 15, 14, 13, and 11 for noise, hybrid, natural-scene, and fixed-phase patches, respectively. Thus, compared to natural scenes, fixing the phase spectrum decreases the relative dimensionality by 15%, and breaking the statistical dependence between amplitude and phase increases the relative dimensionality by 8%. For the 8×8 patches, greater percentage changes are observed: at $N = 2^{19}$ samples, the dimensionalities are approximately 42, 25, 21, and 16 for noise, hybrid, natural-scene, and fixed-phase patches, respectively. Compared to natural scenes, fixing the phase spectrum decreases the relative dimensionality by 24%, and breaking the statistical dependence between amplitude and phase increases the relative dimensionality by 19%.

B. Entropy Estimates

As demonstrated in [15], rational extrapolations can be applied to the relative dimensionality curves to extend the proximity distributions and thereby estimate entropy. Specifically, we defined the term *XEntropy* to denote this extrapolated entropy estimate. Here, we employ the XEntropy C estimator, which assumes that the relative dimensionality curves can be described by the same functional form as the relative dimensionality curve for Gaussian white noise. This functional form, RD(N), is given by

$$\mathrm{RD}(N) = -(\log N + b_0)^2/(a_2[\log N]^2 + 2a_2 b_0 \log N + a_1 b_0 - a_0), \quad (12)$$

where $a_2 = -1/k$, $a_0 = 65.05$, and the parameters a_1 and b_0 were adjusted to fit the measured data by using a trust-region search procedure [35].

We are in no way implying that Eq. (12) is an optimal extrapolation of these data. Rather, this functional form was chosen (1) for its relative simplicity (it is a rational function in $\log N$); (2) for the fact that it converges on the dimensionality of the data in the limit of large N, i.e., $\lim_{N\to\inf} \mathrm{RD}(N) = -1/a_2 = k$; and (3) because it provides decent fits to the data obtained both in the current study and in our previous study [15]. In [15], we showed that XEntropy C can provide accurate estimates of manipulated natural-scene patches with known entropy as long as the pixel intensities are fundamentally noise-limited (e.g., due to photon noise, quantization noise), which is the case for the patches used in the current study. However, although we believe that this is a rational extrapolation, we also believe that future work will allow more theoretically accurate extrapolations.

Figure 8 shows the extrapolated relative dimensionality curves, and Fig. 9 shows the corresponding entropy estimates, computed via Eq. (4) from each (extrapolated) value of $E\{\log_2 D_N^*\}$ at each N. The final estimates of entropy listed in Fig. 9 were computed by using Eq. (A3) with $k = 16$ (4×4) or $k = 64$ (8×8) and N chosen such that the relative dimensionality was within 1% of the intrinsic dimensionality (1% of k). The final entropy estimates for the natural scenes, fixed-phase scenes, and hybrid scenes are also listed in Table 1.

For the 4×4 patches, natural scenes have an entropy of 53 bits (entropy rate of 3.3 bits/pixel). The fixed-phase patches yield a lower entropy of 31 bits (1.9 bits/pixel). The hybrid patches yield the highest entropy of 57 bits (3.6 bits/pixel). As mentioned in Section 2.C, the entropy in the fixed-phase condition is the entropy due to the amplitude spectrum $H(A) = 31$ bits, whereas the entropy in the hybrid condition is $H(A) + H(P) = 57$ bits. The entropy due to the phase spectrum is therefore $H(P) = 57 - 31$ bits $= 26$ bits. Thus, for 4×4 patches, the amplitude spectrum contributes approximately 59% to the total information in natural-scene patches, and the phase spectrum contributes approximately 49%. Note that the sum of these two percentages is approximately 108%, indicating that about 8% of the information is common to both spectra (mutual information). As mentioned in Section 2.C, the mutual information $I(A;P)$ is given by difference between the entropy of hybrid patches $H(A) + H(P) = 57$ bits and the entropy of natural scenes $H(A,P) = 53$ bits; thus, $I(A;P) = H(A) + H(P) - H(A,P) = 57 - 53$ bits $= 4$ bits [8% of $H(A,P)$].

For the 8×8 patches, in line with what we found in [15], natural scenes have an entropy of 193 bits (3.0 bits/pixel) (184 bits and 2.9 bits/pixel were reported in [15], which used only a subset of the images employed here). The fixed-phase patches yield a lower entropy of 104 bits (1.6 bits/pixel). Again, the hybrid patches show the highest entropy of 213 bits (3.3 bits/pixel). Thus, $H(A) = 104$ bits and $H(P) = 213 - 104$ bits $= 109$ bits. These data reveal that for 8×8 patches, the amplitude spectrum contributes approximately 54% to the total information in natural-scene patches, the phase spectrum contributes approximately 56%, and there is approximately 10% mutual information ($I(A;P) = H(A) + H(P) - H(A,P) = 213 - 193$ bits $= 20$ bits [10% of $H(A,P)$]).

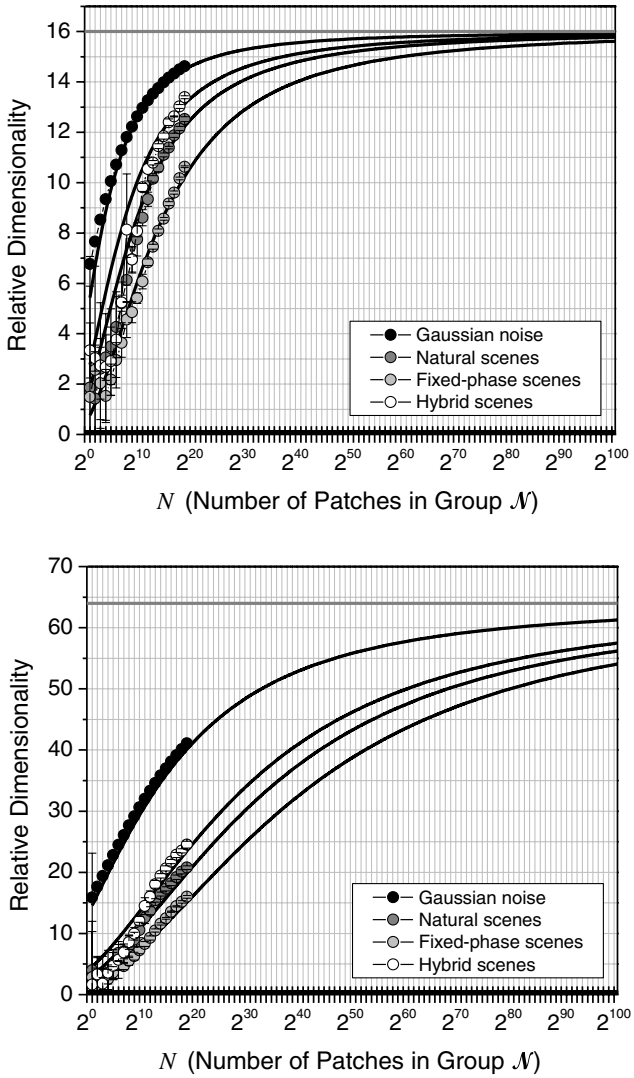

Fig. 8. Extrapolated relative dimensionality curves for 4×4 (top) and 8×8 (bottom) patches using the XEntropy C estimator from [15]. In each graph, the horizontal axis denotes N (the number of patches in Group $\mathcal{N}$) and the vertical axis denotes the corresponding relative dimensionality computed via $\mathrm{RD}(N) = -(\log N + b_0)^2/(a_2[\log N]^2 + 2a_2 b_0 \log N + a_1 b_0 - a_0)$, where $a_2 = -1/k$ and $a_0 = 65.05$, and where the parameters a_1 and b_0 were adjusted to fit the measured data. The solid gray line denotes the intrinsic dimensionality of $k = 16$ (4×4) or $k = 64$ (8×8).

Comparing the 4×4 data to the 8×8 data, the results reveal that as the patch size increases, the information due to the amplitude spectrum decreases while the information due to the phase spectrum increases. We expect that for even larger patches, the phase spectrum will continue to contribute more and more to the total information, whereas the amplitude spectrum will move closer to the characteristic $1/f$ trend and therefore contribute a progressively smaller percentage (see Section 4). However, the results also reveal an increase in the amount of mutual information as the patch size increases both in terms of the entropy (4 bits for 4×4 versus 20 bits for 8×8) and in terms of the entropy rate (0.25 bits/pixel for 4×4 versus 0.31 bits/pixel for 8×8).

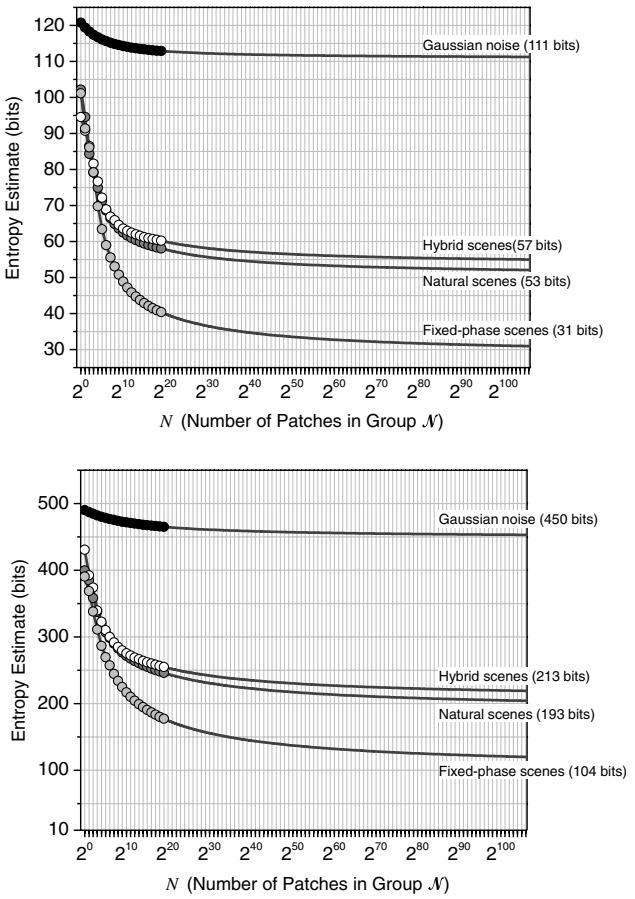

Fig. 9. Entropy estimates for 4×4 (top) and 8×8 (bottom) patches computed for each value of N using the XEntropy C estimator from [15]. The final entropy estimate listed for each curve was computed by using Eq. (A3) with $k = 16$ (4×4) or $k = 64$ (8×8) and N chosen such that the relative dimensionality was within 1% of the intrinsic dimensionality.

4. DISCUSSION

In this paper, we have used proximity distributions to compare the amounts of information provided by the amplitude (power) and phase spectra, and the amount of mutual information between these spectra, for patches of natural scenes.

A. Information Content of Amplitude versus Phase

For our 4×4 patches, the total entropy was estimated to be 53 bits or 3.3 bits/pixel. For this total, 59% is carried by the amplitude spectrum, 49% is carried by the phase spectrum, and 8% is part of the shared (mutual) information between the amplitude and phase spectra. For our 8×8 patches, the total entropy was estimated to be 193 bits or 3.0 bits/pixel. For this total, 54% is carried by the amplitude spectrum, 56% is carried by the phase spectrum, and 10% is part of the shared (mutual) information between the amplitude and phase spectra.

Although our technique makes it difficult to provide an accurate estimate with larger patches, we expect that the amplitude spectrum of larger patches will carry a smaller proportion of the total information (although larger in absolute bits). With larger patches, the average contrast will be more consistent (regression toward the mean). We expect that as the patch size increases, the slope of the amplitude spectrum's falloff will move closer to the typical $1/f$ found to be common in the literature [1,2,18]. With large images, we may well find that a higher proportion of the information in the image is found in the phase spectrum.

To illustrate this argument, Fig. 10 shows the distribution of rotationally averaged amplitude-spectrum slopes for 2000 image patches of size 8×8, 32×32, and 256×256 pixels. The slopes correspond to the parameter α obtained by fitting each patch's rotationally averaged amplitude spectrum with a function of the form $1/f^{\alpha}$, where f corresponds to radial frequency. As one can see, the variance of average slopes is much larger for the smaller patches. This is not surprising, since many of our 8×8 patches have very low contrast and are dominated by noise (a flat spectrum). However, for 256×256 patches, no image is of very low contrast or low mean intensity. With multiple features at different scales, these larger images converge on the mean amplitude spectrum.

B. Perceptual Importance of Amplitude versus Phase

The results of this study also allow us to address the ideas presented by Oppenheim and Lim [19], Piotrowski and Campbell [20], and Morgan *et al.* [22]. Each of these studies argued that in large images, the information is carried perceptually by the phase spectra. As we noted earlier, if the amplitude spectra of two different images are similar (e.g., close to $1/f$), then it would follow that exchanging the amplitude spectra would produce little perceptual change (simply because the amplitude spectra are similar).

However, Morgan *et al.* [22] noted that for small image patches, it was the amplitude spectrum that appeared to dominate perceptually. For small image patches, the patch is often dominated by a single edge. Under such circumstances, the orientation of the edge is well described by the amplitude spectrum with most of the energy at that orientation. It is therefore not surprising that the amplitude spectrum of a small patch provides a fairly good account of the perception of the edge.

Table 1. Entropy Estimates for 4×4 and 8×8 Patches[a]

		Nat. Scenes $H(A,P)$	Fixed Phase $H(A)$	Hybrid $H(A)+H(P)$	$H(P)$	$I(A;P)$
4×4 patches	bits	53	31	57	26	4
	bits/pixel	3.3	1.9	3.6	1.6	0.25
	% of $H(A,P)$	100%	58.5%	107.5%	49.1%	7.5%
8×8 patches	bits	193	104	213	109	20
	bits/pixel	3.0	1.6	3.3	1.7	0.31
	% of $H(A,P)$	100%	53.9%	110.4%	56.5%	10.4%

[a] $H(A,P)$ denotes the entropy of natural-scene patches; $H(A)$ denotes the entropy due to the amplitude spectrum; $H(P)$ denotes the entropy due to the phase spectrum; $I(A;P)$ denotes the mutual information between amplitude and phase; $H(A)+H(P)$ is the entropy that would result if the amplitude and phase spectra were independent [i.e., if $I(A;P)$ were zero].

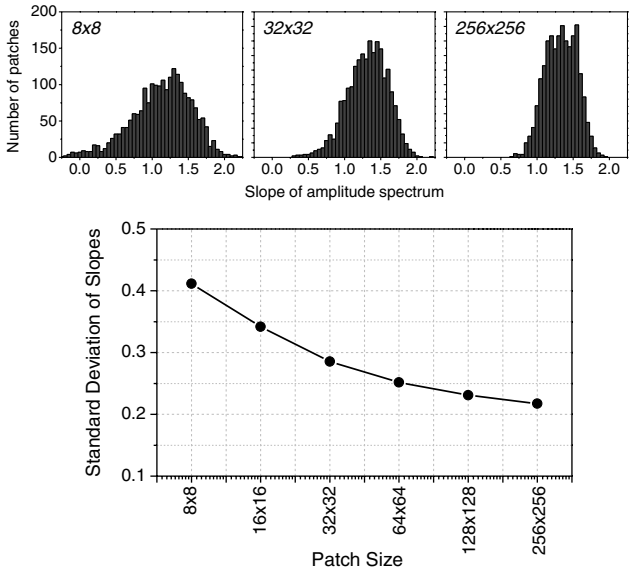

Fig. 10. Top: Distribution of rotationally averaged slopes for 2000 natural-scene patches of 8×8, 32×32, and 256×256 pixels. Bottom: Standard deviation of average slopes as a function of patch size. Notice that there is much more variation in slope for smaller patches. (The slope here refers to the parameter α obtained by fitting the rotationally averaged amplitude spectrum with a function of the form $1/f^{\alpha}$, where f corresponds to radial frequency; only values of $f > 0$ were used in the fitting.)

To illustrate this argument, Fig. 11 shows a random collection of 8×8 and 64×64 image patches that exceeded a fixed threshold contrast. As one can see, most of the 8×8 patches are dominated by a single orientation. However, for larger patches we typically see multiple edges within the patch. A change in the relative position of an edge is primarily determined by the phase spectrum. Therefore, large images with features at multiple orientations and scales are more likely to converge on the $1/f$ amplitude spectrum. It has previously been argued that the $1/f$ spectrum results from a simple sum of features at multiple scales or from a similar set of occluding surfaces [1,23].

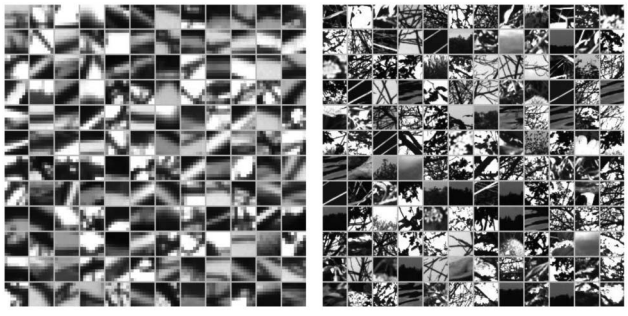

Fig. 11. Random collection of 8×8 (left) and 64×64 (right) patches with a standard deviation of pixel values greater than 35. For small patch sizes such as 8×8, most of the patches are dominated by a single edge; in this case, the orientation of the edge is well described by the power spectrum with most of the power at that orientation, and thus the power spectrum provides a fairly good account of the perception of the edge. However, for larger patches, features tend to occur at multiple orientations and scales, and thus the power spectrum is closer to $1/f$; in this latter case, the visual appearance of each patch is dictated more by its phase spectrum.

C. Role of Amplitude versus Phase in Efficient Neural Coding

As we mentioned in Section 1, because the amplitude spectrum is the square root of the Fourier transform of the autocorrelation function, the amplitude spectrum simply provides an alternative description of the pairwise correlations in an image. Neural coding strategies such as linear Hebbian learning that rely only on these correlations can therefore capture information only from the amplitude spectrum. On the other hand, techniques such as sparse coding [4] and ICA [7] employ a preprocessing stage that whitens the images, and thus these latter types of networks focus on the information provided by the phase spectrum.

Based on our results, most of the information in 4×4 patches is contained in the amplitude spectrum (59% amplitude versus 49% phase). Thus, for 4×4 patches, a linear Hebbian network or other network that relies on spatial correlations should, in theory, be able to discover a more efficient code than that obtained from a network that employs whitening. However, for 8×8 patches, which is a size more commonly used in efficient neural coding techniques, the reverse trend is observed. For 8×8 patches, the phase spectrum contains slightly more information than the amplitude spectrum (54% amplitude versus 56% phase). This finding may help provide a partial account for the success of techniques such as sparse coding [4] and ICA [7].

However, for both patch sizes, our results also show that there is some information common to both types of spectra. Approximately 8% and 10% of mutual information was found to exist between the amplitude and phase spectra for 4×4 and 8×8 patches, respectively. Thus, a neural coding strategy or an image-processing algorithm that uses only the amplitude spectrum or only the phase spectrum may be able to discover a feature that is seemingly attributable to the other type of spectrum. However, given the relatively small amount of mutual information and the relatively minor increase in mutual information observed when doubling the patch size, our results also suggest that an optimal block-based coding or processing strategy should ideally consider both types of spectra, at least when operating at the 4×4 or 8×8 level (in image processing, 8×8 is a very commonly used block size).

D. Limitations and Extensions

Although we believe that the technique employed here is a powerful method for measuring information in relatively high-dimensional data sets, we also need to provide caution. The numbers reported here apply only to our 4×4 and 8×8 image patches and should not be considered universal for all natural scenes. Our estimates are also based on extrapolations, and although we continue to explore the accuracy of these extrapolations, we can not guarantee that our extrapolations are optimal.

Our images are drawn from the van Hateren database [32]. Such images are not necessarily a good representation of images drawn from typical human experience, or from the web. The van Hateren database is attractive due to its widespread use; however, the results presented here are dependent on factors such as camera blur, scene content, noise, and the average contrast and intensity of the patches. Darker images have a lower absolute range that would reduce the number of bits. Although we cannot conclude that these

results are generally true of all image collections, the technique presented here can be applied to any particular image collection.

The patch sizes are also relatively small, which do not allow simple extensions to larger images. As the size of the image database increases, we will be able to provide more accurate estimates of the entropy of larger images. Each doubling of the size of the database allows us to move out one more step on the distribution function. Nonetheless, although our current method may seem to require a relatively large number of samples (2^{19} images), this is still far fewer than any direct method of Shannon entropy requiring probabilities on 2^{64} images.

In future work, we hope to extend this approach to larger images. We also believe that the approach can be used with techniques like sparse coding and ICA to determine how much mutual information remains after various stages of efficient coding.

5. CONCLUSIONS

This paper provides a new technique and a set of results for evaluating the relative proportions of different forms of information in natural scenes. The technique compares the information in natural-scene patches with the information in corresponding patches with altered phase spectra. Here, we applied this technique to estimate the information contained in the power and phase spectra of 4×4 and 8×8 natural-scene patches. For our 4×4 patches, we found that 59% of the information in the images is carried by the amplitude spectrum, 49% is carried by the phase spectrum, and 8% is part of the shared (mutual) information between these spectra. For our 8×8 patches, we found that 54% is carried by the amplitude spectrum, 56% is carried by the phase spectrum, and 10% is part of the mutual information. We discussed how these results provide important insights into the sources of redundancy that may allow for compression and efficient coding by the human visual system.

APPENDIX A

The estimation of entropy based on nearest-neighbor distances was initially proposed by Kozachenko and Leonenko [28], then was later applied to neural data by Victor [29] and subsequently to the estimation of mutual information by Kraskov *et al.* [30] and by Kybic [31]. This is a so-called *binless* estimator of differential entropy that operates by estimating $i_{\mathbf{X}}(\mathbf{x})\triangleq -\log_2 f_{\mathbf{X}}(\mathbf{x})$ via nearest-neighbor distances, where $\mathbf{X}$ denotes a (possibly vector-valued) random variable with corresponding probability density function $f_{\mathbf{X}}(\mathbf{x})$. In this formulation, differential entropy, $h(\mathbf{X})$, is the expected value of $i_{\mathbf{X}}(\mathbf{x})$:

$$h(\mathbf{X})\triangleq -\int_{\mathbf{x}\in\mathcal{A}} f_{\mathbf{X}}(\mathbf{x})\log_2 f_{\mathbf{X}}(\mathbf{x})\mathrm{d}\mathbf{x}=\int_{\mathbf{x}\in\mathcal{A}} f_{\mathbf{X}}(\mathbf{x})i_{\mathbf{X}}(\mathbf{x})\mathrm{d}\mathbf{x}$$
$$=E\{i_{\mathbf{X}}(\mathbf{x})\}\approx\frac{1}{M}\sum_{m=1}^{M}\hat{i}_{\mathbf{X}}(\mathbf{x}_m), \tag{A1}$$

where the final relation approximates the expectation in the third relation with the sample mean computed using M observed samples, $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_M$, drawn according to $f_{\mathbf{X}}$. Specifically, the approximation results from (1) replacing the integral with a sum, (2) assuming $f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}\approx\frac{1}{M}$, $\forall\mathbf{x}_m$, and (3) using $\hat{i}_{\mathbf{X}}(\mathbf{x}_m)$ as an estimator of $i_{\mathbf{X}}(\mathbf{x}_m)$.

The estimator $\hat{i}_{\mathbf{X}}(\mathbf{x})$ is computed based on the Euclidean distance D_N^* between $\mathbf{x}$ and its nearest neighbor among the remaining $N=M-1$ observations as

$$\hat{i}_{\mathbf{X}}(\mathbf{x})=kE\{\log_2 D_N^*\}+\log_2\left(\frac{A_kN}{k}\right)+\frac{\gamma}{\ln 2}, \tag{A2}$$

where γ is the Euler constant and $A_k=k\pi^{k/2}/\Gamma(\frac{k}{2}+1)$ denotes the surface area of a k-dimensional hypersphere. Combining Eqs. (A1) and (A2), $h(\mathbf{X})$ is approximated by

$$h(\mathbf{X})\approx-\frac{k}{M}\sum_{m=1}^{M}\log_2 D_{N,m}^*+\log_2\left(\frac{A_kN}{k}\right)+\frac{\gamma}{\ln 2}, \tag{A3}$$

where $D_{N,m}^*$ is the Euclidean distance between $\mathbf{x}_m$ and its nearest neighbor among the other $N=M-1$ observations.

ACKNOWLEDGMENTS

This material is based upon work supported by, or in part by, the U.S. Army Research Laboratory (USARL) and the U.S. Army Research Office (USARO) under contract/grant number W911NF-10-1-0015.

REFERENCES AND NOTES

1. D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," J. Opt. Soc. Am. A **4**, 2379–2394 (1987).
2. D. J. Field, "Scale-invariance and self-similar 'wavelet' transforms: an analysis of natural scenes and mammalian visual systems," in *Wavelets, Fractals and Fourier Transforms: New Developments and New Applications*, M. Farge, J. C. R. Hunt, and J. C. Vassilicos, eds. (Oxford University, 1993), pp. 151–193.
3. D. J. Field, "What is the goal of sensory coding?" Neural Comput. **6**, 559–601 (1994).
4. B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: a strategy employed by V1?" Vis. Res. **37**, 3311–3325 (1997).
5. E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," Annu. Rev. Neurosci. **24**, 1193–1216 (2001).
6. W. E. Vinje and J. L. Gallant, "Sparse coding and decorrelation in primary visual cortex during natural vision," Science **287**, 1273–1276 (2000).
7. A. Bell and T. Sejnowski, "The 'independent components' of natural scenes are edge filters," Vis. Res. **37**, 3327–3338 (1997).
8. P. O. Hoyer and A. Hyvärinen, "A multi-layer sparse coding network learns contour coding from natural images," Vis. Res. **42**, 1593–1605 (2002).
9. L. Wiskott and T. Sejnowski, "Slow feature analysis: unsupervised learning of invariances," Neural Comput. **14**, 715–770 (2002).
10. W. E. Vinje and J. L. Gallant, "Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1," J. Neurosci. **22**, 2904–2915 (2002).
11. P. Garrigues and B. A. Olshausen, "Learning horizontal connections in a sparse coding model of natural images," in *Advances in Neural Information Processing Systems 20*, J. C. Platt, D. Koller, Y. Singer, and S. Roweis, eds. (MIT Press, 2008), pp. 505–512.
12. A. B. Lee, K. S. Pedersen, and D. Mumford, "The nonlinear statistics of high-contrast patches in natural images," Int. J. Comput. Vis. **54**, 83–103 (2003).
13. W. B. Pennebaker and J. L. Mitchell, *The JPEG Still Image Data Compression Standard* (Van Nostrand Reinhold, 1993).

14. International Organization for Standardization, "Information technology—JPEG 2000 image coding system: core coding system," Tech. Rep. ISO/IEC FDIS15444-1:2000 (International Organization for Standardization, 2000).
15. D. M. Chandler and D. J. Field, "Estimates of the information content and dimensionality of natural scenes from proximity distributions," J. Opt. Soc. Am. A **24**, 922–941 (2007).
16. R. Hosseini, F. Sinz, and M. Bethge, "Lower bounds on the redundancy of natural images," Vis. Res. **50**, 2213–2222 (2010).
17. G. J. Burton and I. R. Moorhead, "Color and spatial structure in natural scenes," Appl. Opt. **26**, 157–170 (1987).
18. D. J. Tolhurst, Y. Tadmor, and T. Chao, "Amplitude spectra of natural images," Ophthalmol. Physiol. Opt. **12**, 229–232 (1992).
19. A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," Proc. IEEE **69**, 529–541 (1981).
20. L. N. Piotrowski and F. W. Campbell, "A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase," Perception **11**, 337–346 (1982).
21. R. P. Millane and W. H. Hsiao, "The basis of phase dominance," Opt. Lett. **34**, 2607–2609 (2009).
22. M. J. Morgan, J. Ross, and A. Hayes, "The relative importance of local phase and local amplitude in patchwise image reconstruction," Biol. Cybern. **65**, 113–119 (1991).
23. D. L. Ruderman, "Origins of scaling in natural images," Vis. Res. **37**, 3385–3398 (1997).
24. W. H. Hsiao and R. P. Millane, "Effects of occlusion, edges, and scaling on the power spectra of natural images," J. Opt. Soc. Am. A **22**, 1789–1797 (2005).
25. D. J. Graham and D. J. Field, "Statistical regularities of art images and natural scenes: spectra, sparseness and nonlinearities," Spatial Vis. **21**, 149–64 (2007). PMID: 18073056.
26. A. Torralba and A. Oliva, "Statistics of natural image categories," Network: Comput. Neural Syst. **14**, 391–412 (2003).
27. Y. Tadmor and D. Tolhurst, "Both the phase and the amplitude spectrum may determine the appearance of natural images," Vis. Res. **33**, 141–145 (1993).
28. L. F. Kozachenko and N. N. Leonenko, "A statistical estimate for the entropy of a random vector," Probl. Inf. Transm. **23**, 9–16 (1987).
29. J. D. Victor, "Binless strategies for estimation of information from neural data," Phys. Rev. E **66** 051903 (2002).
30. A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," Phys. Rev. E **69** (2004).
31. J. Kybic, "High-dimensional mutual information estimation for image registration," in *Proceedings of International Conference on Image Processing, 2004*, Vol. 3 (IEEE, 2005), pp. 1779–1782.
32. J. H. van Hateren and A. van der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," Proc. R. Soc. Lond. Ser. B **265**, 359–366 (1998).
33. R. M. Gray and D. L. Neuhoff, "Quantization," IEEE Trans. Inf. Theory **44**, 2325–2384 (1998).
34. The values of $T = 65,536$ and $T = 16,384$ and ten trials of the search procedure (using ten randomizations of Group $\mathcal{T}$) were sufficient to yield a consistent sample mean.
35. M. R. Celis, J. E. Dennis, and R. A. Tapia, "A trust region strategy for nonlinear equality constrained optimization," in *Numerical Optimization 1984*, P. T. Boggs, R. H. Byrd, and R. B. Schnabel, eds. (SIAM, 1985), pp. 71–82.