

# Relations between the statistical regularities of natural images and the response properties of the early visual system

Eizaburo Doi, Michael S. Lewicki

**Abstract**—Natural images are not random; instead, they exhibit statistical regularities. Assuming that our vision is designed for tasks on natural images, computation in the visual system should be optimized for such regularities. Recent theoretical investigations along this line have provided many insights into the visual response properties in the early visual system. In this article we review both the known statistical regularities of natural images, the extent to which low-level vision might be adapted to them, and the recent development in theoretical models to explain this relationship.

## I. INTRODUCTION

Natural images are highly variable. When we look at the texture of natural surfaces like tree bark, we immediately recognize it as such in spite of large variations in the actual images that fall on our retina. In order to accomplish this feat, our visual system must be able to pool the natural range of variability and to recognize an image as an instance of the same image type. Thus an important part of vision involves capturing the statistical variation of natural images.

At the same time, natural images are very specific. We know how different they are from random images. The characterization of their statistical regularities has been an important research topic in science and engineering fields. We can state its ultimate goal as developing a probability model that generates images indistinguishable from natural images. We are still far from this goal, although our knowledge of natural images has been increasing.

An important motivation of such an investigation is to better understand how our visual system codes visual information. Information theory states that the most statistically efficient codes are those that best capture the statistical regularities of the data. Because biological systems are under strong evolutionary pressure, it is hypothesized that the codes they use are highly efficient. Furthermore, knowledge of the statistical regularities of the data allows the system to perform important visual tasks such as finding interesting features and filling in missing information. More generally, the questions we need to answer are: what are the computational objectives of the early visual system, what are the computationally relevant biological constraints, and to what extent can the response properties be explained in relation to the statistical regularities of their input.

In this article we address side by side what we know about the statistical regularities of natural images and how the early visual system (low-level vision) can be explained in terms of them (recent reviews on this subject can be found

in [1]–[4]). First we examine the statistical regularities of a single pixel (intensity values) sampled from natural images. Next we examine the statistical regularities of image regions. Although these regions are, in most analyses, restricted to small image patches rather than whole natural scenes due to the computational complexity of the analysis, receptive field size in the visual system is also restricted, so this constraint is not as limiting as it might first seem.

When discussing theoretical approaches, it is helpful to keep in mind two types of models. One is *the inverse of the data generative model*, in which the probability model of generating the data is assumed and the task of the visual system is to infer the hidden causes that generate the data [5], [6]. The advantage of this approach is that it is conceptually well-defined and there is a clear connection to the probability model of natural images. The other type is *a representational model*, where the goal is not necessarily to infer the hidden variables of the generative process, but instead to satisfy some objectives for the representation, such as decorrelation or maximum information transfer.

## II. ONE-PIXEL STATISTICS

### A. Statistical regularities

The light intensity of natural scenes is highly biased towards small values (Fig. 1b). This analysis is based on natural images whose pixel values are linear to the light intensity [7]–[9], and look like Fig. 1a. Note that such data are appropriate as the input to the eye, while the images taken by a conventional film/CCD camera involve nonlinear transforms and therefore not appropriate for the analysis.

### B. Relation to response properties

Many researchers have assumed that the logarithm approximates the nonlinearity in the cone photoreceptor (e.g., [7], [8]). The logarithm transforms the data so that the transformed variable is more evenly distributed, which increases the contrast of the images. A more elaborate analysis reveals that the empirical cone nonlinearity model matches well to the cumulative histogram of the linear intensity natural images, implying that the cone nonlinearity serves as the so-called histogram equalization [9]<sup>1</sup>. The empirical model of cone nonlinearity is given by

$$x_{nl} = 1 - \exp(-k \cdot x_l) \quad (1)$$

where  $x_l$  is linear intensity,  $x_{nl}$  is its nonlinear cone response, and  $k$  is a free parameter that is adaptive to the luminance level

<sup>1</sup>This relationship was first reported for the contrast distribution in the fly's habitat and the nonlinearity of contrast-detecting cells in the fly [10].

E. Doi is with Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213, USA (e-mail: edoi@cnbc.cmu.edu).

M. S. Lewicki is with Center for the Neural Basis of Cognition and Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA (e-mail: lewicki@cnbc.cmu.edu).

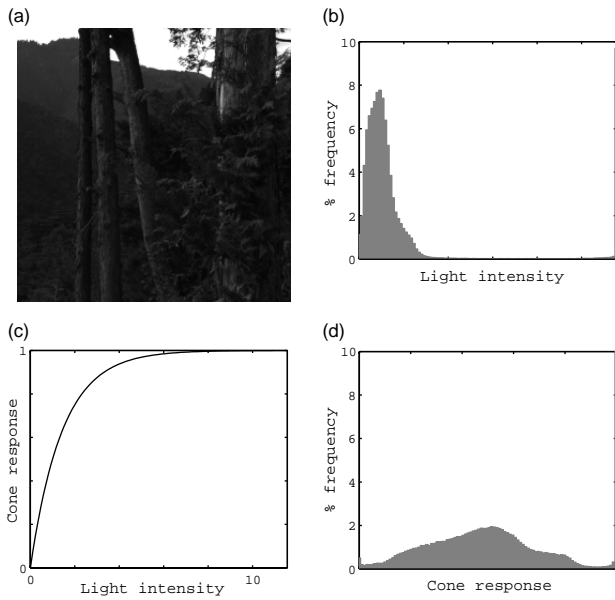


Fig. 1. The linear intensity of natural scenes is biased toward small values. (a) An example of linear intensity natural image. (b) The histogram of (a). (c) The cone nonlinearity model. (d) The histogram of cone nonlinear response to (a); The corresponding image is shown in Fig. 5 top-left panel.

[11]. The shape of eq. 1 is shown in Fig. 1c, whose parameter is determined for the image in Fig. 1a. The histogram of resulting cone responses is shown in Fig. 1d demonstrating that the pixel values are much more uniformly distributed compared to the raw intensity values. (The image in Fig. 1, when transformed into cone responses, also has a greatly enhanced visual contrast, Fig. 5.)<sup>2</sup>

C. Remarks

Since a uniform distribution conveys the maximum amount of information given a fixed range of the representation, the cone nonlinearity is regarded as the most efficient transformation that maximally utilizes the information capacity of the representation. In general, this objective – the transformation of the input so as to convey the maximum amount of information – is referred to as *efficient coding* (or *redundancy reduction*) [14], [15]. This is the basic underlying principle of the theoretical approaches we will consider in this article. Note that the optimality depends on the constraint: if the variance is fixed instead of the range, then the Gaussian distribution is the optimum; if the variable is positive and its mean is fixed, then the exponential distribution is the optimum [1]. There is no principled explanation, to our knowledge, regarding which constraint is appropriate for the early visual system; However, these results indicate that the photoreceptor nonlinearity is well explained given the fixed range of neuronal output.

<sup>2</sup>The slope of the empirical model given by eq. 1 is usually slightly less than the cumulative density of linear intensity, which makes the resulting histogram diverged from the uniform distribution as in Fig. 1. A model that minimizes the estimation error of the input given the intrinsic noise (the so-called channel noise) is reported to improve the fit over the histogram equalization model [12]. Another model takes into account adaptation mechanisms for the time-series of light intensity data, providing a more complete model [13].

III. SECOND-ORDER STATISTICS

To analyze the statistics of image regions, we first use second-order statistics; in the next section we consider higher-order statistics<sup>3</sup>. In the following analysis the mean value (i.e., first-order statistic) is set to zero, which is equivalent to assuming that information of the mean luminance level is discarded in the representation, and that the visual system is interested only in the variation about this reference.

A. Statistical regularities

The second-order statistics are completely described by the amplitude spectrum of natural images<sup>4</sup>, and the amplitude spectra of natural images are approximately proportional to  $1/f$ , where  $f$  is the spatial-frequency [16]. In Fig. 2 we show an example of the amplitude spectrum of natural images. We can also observe that the amplitude is slightly stronger along horizontal and vertical orientations; this is because of the dominance of vertical and horizontal structures in natural scenes (e.g., trees and horizons). The  $1/f$  amplitude spectrum means the correlated structure of pixel values in natural images; the correlation is stronger for lower spatial-frequency components. Were there no correlation in the image, the amplitude spectrum would be flat.

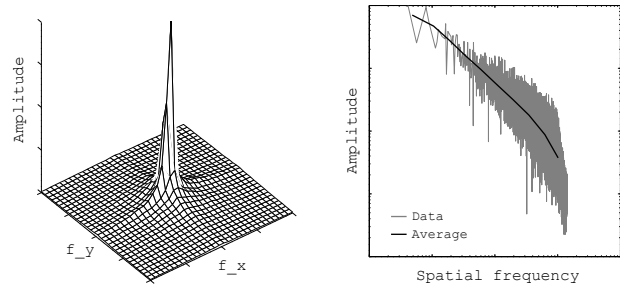


Fig. 2. The amplitude spectrum of natural images is proportional to  $1/f$ . The analyzed image data is shown in Fig. [5]. Left: two-dimensional spatial-frequency amplitude. Right: spatial-frequency amplitude obtained by integrating two-dimensional amplitude over orientation. Gray curve (“Data”) is the intact data points of all frequencies, and its smoothed version is shown with black curve (“Average”).

In addition to the general tendency of the  $1/f$  amplitude spectrum, natural images in different categories have their specific trends in the amplitude spectrum [17]. For instance, images of forests tend to have more high spatial-frequency components, while those of the beach tend to localize their amplitude in lower and horizontal frequencies, based on the analysis of hundreds of each categories. Similarly, the depth of natural images (i.e., close or distant view) can also be

<sup>3</sup>The definition of the  $i$ -th order statistic is  $\langle \prod_k^i x_k \rangle$ , where  $\langle \cdot \rangle$  is the ensemble average over samples and  $x_k$  is the pixel value at the location  $k$ . Assuming the translation invariance of natural images, the ensemble average can be replaced with the average over the index of location while keeping the relative position of the indices. The first-order statistic is the mean of single pixels, which is part of the analysis in the last section (note that one-pixel statistics entails all orders of statistics); second-order statistics are the covariance of pairs of pixels described by the covariance matrix; the rest of the statistical regularities are collectively referred to as higher-order statistics.

<sup>4</sup>This is because the covariance is the autocorrelation of natural images, and the Fourier transform of the autocorrelation function is given by the power spectrum (square of the amplitude spectrum) of natural images.

characterized by the amplitude spectrum: if images are taken at a close distance, the amplitude tends to be more isotropic in orientation and contain high spatial-frequency components, which are explained by the unconstrained orientation of view with respect to the objects and fine grains in the scenes, respectively; if the images are taken from afar, the amplitude tends to be strong along vertical and horizontal axes and localized to lower spatial-frequencies, which is explained by the dominance of vertical and horizontal structures in a distant view in natural scenes and the increase of larger and smoother structures such as fields, forests, and the sky.

### B. Relation to response properties

It has long been suggested that the goal of the early visual system is to remove the redundancy of the highly correlated pixel (or photoreceptor) representation by transforming it into a new representation in which the neuronal activities are (more) statistically independent, and therefore more efficient [14], [15]. Redundancy reduction in terms of second-order statistics means a transform by which the new variable has unit variance and is uncorrelated with the others. Such a transform is called “whitening” and can be implemented by a set of linear filters. Whitening with respect to natural image patches turns out to have a center-surround filter structure, which is in close agreement with receptive fields in the retinal ganglion cells (RGCs) and lateral geniculate nucleus (LGN) [18]–[21].<sup>5</sup>

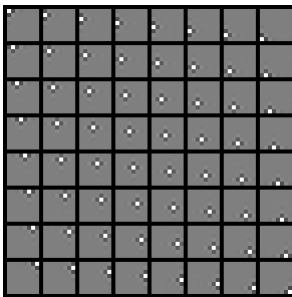


Fig. 3. Whitening filters for natural images show center-surround receptive field organization. Each small image corresponds to a single whitening filter for  $8 \times 8$  pixel image patches; here we show a whole set of 64 filters.

### C. Remarks

The model proposed in [18], [19] also takes into account the noise component in the input (the so-called sensory noise) and includes a noise filtering process prior to whitening. Assuming that the noise variance is proportional to the mean luminance level, the model yields a remarkable match to the receptive fields of RGCs that adaptively change its shape depending on the mean luminance level.

<sup>5</sup>Precisely, there exist infinite number of whitening filters. If  $\mathbf{W}_1$  is whitening filter matrix, the covariance matrix of its output should be equal to the identity,  $\langle \mathbf{W}_1 \mathbf{x} \cdot (\mathbf{W}_1 \mathbf{x})^T \rangle = \mathbf{I}$ , which implies that  $\mathbf{W}_1 \Sigma_{\mathbf{x}} \mathbf{W}_1^T = \mathbf{I}$ . Now, with an arbitrary orthogonal matrix  $\mathbf{U}$  let us consider a new set of filters  $\mathbf{W}_2 = \mathbf{U} \mathbf{W}_1$ ; we can see that it is also a whitening matrix since it satisfies  $\mathbf{W}_2^T \Sigma_{\mathbf{x}} \mathbf{W}_2 = \mathbf{I}$ . The whitening matrix that yields center-surround organization is the one given by  $\Sigma_{\mathbf{x}}^{-1/2}$  [20], [21].

Whitening is the operation to remove second-order statistical regularities. An alternative idea would be to utilize it, instead of removing. The categorical information in the amplitude spectrum described above may be utilized in the visual system, but no direct comparison has been made [17]. Another possibility is dimensionality reduction: principal component analysis (PCA) provides the optimal set of linear basis functions for reducing dimensionality in the new representation while minimizing mean squared errors of reconstruction. Fig. 4 shows a set of PCA filters, showing frequency patterns with no localized support in each filter, and hence, there is no clear correlate to the receptive field organization in the early visual system. However, PCA might have some functional relevance to our visual system. The data variance is highly concentrated in the subspace of a few principal components (e.g., approximately 20~30% of principal components explain 90% of the data variance in natural image patches). The visual input has huge dimensionality (5 million cone photo-receptors per retina), and thus it is likely that the visual system utilizes second-order structure for dimensionality reduction.

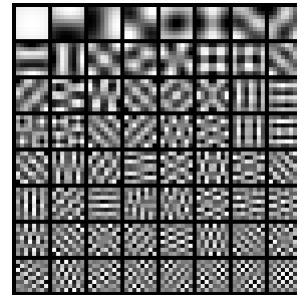


Fig. 4. PCA filters of natural image patches are not localized. This is a set of principal components for  $8 \times 8$  pixel natural image patches, in descending order of magnitude from top-left and row-wise.

## IV. HIGHER-ORDER STATISTICS

### A. Statistical regularities

Higher-order statistics of a natural image reside in the phase spectrum of its Fourier transform, the complement to the amplitude spectrum (i.e., second-order statistics). As has often been noted, the phase spectrum contains much more perceptually relevant information, because it carries the information of edges: at the location of an edge, the phases of different frequencies are aligned. This is demonstrated by randomizing the phase while preserving the amplitude spectrum of a natural image (Fig. 5). One can clearly see which frequency components are present and with what power (e.g., we can see the horizontal and low spatial-frequency component corresponding to the salient contrast pattern between the sky and the mountains). However, their location (i.e., phase) has nothing to do with the original image and there is no edge-like pattern due to the loss of phase alignment. Similarly, we can synthesize images with either random or fixed (uniform or  $1/f$ ) amplitudes while preserving the phase amplitude of the natural image: these images are recognizable and we can tell where the edges are. Somewhat surprisingly, the original phase spectrum with a fixed  $1/f$  amplitude yields an image

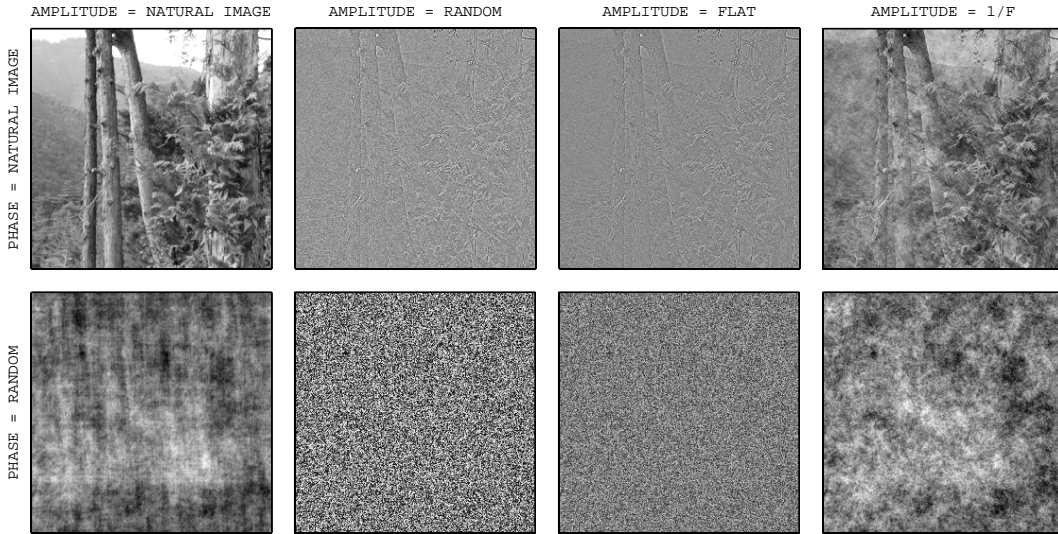


Fig. 5. The phase spectrum is more relevant to our perception than the amplitude spectrum. Each image is synthesized by combining amplitude and phase spectra. We extracted a phase spectrum from either natural or random images (row-wise); we obtained an amplitude spectrum from either natural or random image along with flat and  $1/f$  amplitude (column-wise). The {phase, amplitude} = {natural image, natural image} image corresponds to the original natural image; the {random, random} image is the original random image; and {random,  $1/f$ } image is the correlated random image whose amplitude spectrum is  $1/f$ . Note that the {natural, uniform} image is the whitened version of the natural image, equivalent to the output of RGCs/LGN model as described in sec. III.

much closer to the original (Fig. 5 top-right panel), suggesting that the exact amplitude information plays a lesser role in our perception.

The amplitude spectrum characterizes the covariance matrix of the data, and therefore, it can fully capture the statistical regularities if the data are drawn from a Gaussian distribution (given zero-mean). In the case of natural images, however, several observations indicate that the underlying distribution is highly non-Gaussian [16], [22]. In Fig. 6 we show an example of two-dimensional non-Gaussian data: assuming the data are Gaussian and analyzing the amplitude spectrum yields a model which clearly fails to capture the underlying structure of the data (a). For comparison, the true distribution is depicted in (b), which is what we would like to estimate from the data.

One piece of evidence for the non-Gaussianity of natural images is the observation of the sparseness of their distribution, conveniently measured by the (normalized) kurtosis:  $\tilde{\kappa}(x) = \langle x^4 \rangle / \sigma_x^2 - 3$ , a fourth-order statistic. The distribution is more sparse than Gaussian if  $\tilde{\kappa} > 0$ . There exist directions in the image space on which linear projections of the sample points are sparsely distributed [16], [22]. Note that such non-Gaussian distribution is impossible if the data are multivariate Gaussian, since any linear projection of multivariate Gaussian is Gaussian. Interestingly, a linear projection onto the Gabor function, which is proposed as the empirical model of simple-cells in V1 [23], [24] yields a sparse distribution, suggesting that the visual system captures the higher-order statistical regularities.

### B. Relation to response properties

Theoretical models of sparse coding and independent component analysis (ICA) have shown that by modeling the higher-order, non-Gaussian statistics of natural images it is possible to account for the oriented and localized structure

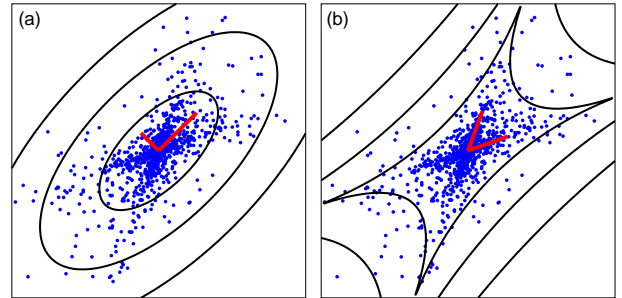


Fig. 6. Second-order statistics are not sufficient to characterize non-Gaussian data. Blue dots in both panels are the samples from a sparse non-orthogonal distribution. The contour plot shows the Gaussian fit to the data (a) and the true distribution of the data (b). The red bars shows the principal axes of the Gaussian or the basis functions of non-Gaussian distribution (see text for details).

of simple cell receptive fields [6], [21], [25]. In both models, the data point  $\mathbf{x}$  is assumed to be generated by the linear superposition of basis functions  $\mathbf{a}_i$  weighted by some hidden variables  $s_i$ ,

$$\mathbf{x} = \sum_{i=1}^m s_i \mathbf{a}_i = \mathbf{A} \mathbf{s} \quad (2)$$

where  $\mathbf{s} = (s_1, \dots, s_m)^T$  and  $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_m]$ .

The set of basis functions  $\mathbf{A}$  is adapted so as to maximize the data likelihood,

$$\mathcal{L} = \langle P(\mathbf{x}|\mathbf{A}) \rangle \quad (3)$$

where the data probability given  $\mathbf{A}$  is obtained by marginalizing over the hidden variables,

$$P(\mathbf{x}|\mathbf{A}) = \int P(\mathbf{x}|\mathbf{A}, \mathbf{s}) P(\mathbf{s}) d\mathbf{s}. \quad (4)$$

Now those models assume that 1) the hidden variables are statistically independent; 2) each hidden variable is drawn from sparse distribution such as the Laplacian (also referred to as double exponential) distribution, yielding the joint prior as

$$P(\mathbf{s}) = \prod_{i=1}^m P(s_i) \propto \prod_{i=1}^m \exp(-|s_i|). \quad (5)$$

The optimal  $\mathbf{A}$  is calculated using the gradient ascent to maximize eq. 3.

An assumption of the model is that the computational goal of the visual system is to infer the most probable representation of the (possibly noisy) stimulus, which is formalized by

$$\hat{\mathbf{s}} = \arg \max_{\mathbf{s}} P(\mathbf{s}|\mathbf{x}, \mathbf{A}). \quad (6)$$

The variable  $\hat{\mathbf{s}}$  corresponds to the neural response to the stimulus  $\mathbf{x}$ . ICA assumes that  $\mathbf{A}$  is square and invertible, yielding

$$\hat{\mathbf{s}} = \mathbf{A}^{-1}\mathbf{x} = \mathbf{W}\mathbf{x}, \quad (7)$$

that is,  $\hat{\mathbf{s}}$  is the linear transform of  $\mathbf{x}$  (and hence  $P(\mathbf{s}|\mathbf{x}, \mathbf{A}) = \delta(\mathbf{s} - \mathbf{A}^{-1}\mathbf{x})$ ). In sparse coding, on the other hand, the mapping of  $\mathbf{s}$  is not restricted to such a special case and we need to solve eq. 6, which is written as

$$\hat{\mathbf{s}} = \arg \max_{\mathbf{s}} \ln P(\mathbf{x}|\mathbf{s}, \mathbf{A}) + \ln P(\mathbf{s}) \quad (8)$$

$$= \arg \min_{\mathbf{s}} \frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{A}\mathbf{s}\|^2 + \sum_{i=1}^m |s_i|, \quad (9)$$

where  $\sigma^2$  defines the reconstruction error level. This means that  $\hat{\mathbf{s}}$  is given by the joint minimization of the squared reconstruction error (the first term) and the negative log-likelihood of the Laplacian prior (the second term).

In Fig. 7 we show ICA basis ( $\mathbf{A}$ ) and filter ( $\mathbf{W}$ ) functions for  $8 \times 8$  natural image patches (the results of sparse coding are similar). Filter functions correspond to the receptive fields that are mapped with the reverse correlation method, and we can see that all filters (except the one for the DC component) show localized and oriented organization similar to simple-cell receptive field in V1. The basis functions correspond to the patterns that the hidden variables encode, and look like edge patterns (except for the DC component). Note also that the ICA basis functions are non-orthogonal; if the basis functions were orthogonal, by definition they would be identical to the filters, as in PCA (one can visually confirm this with Fig. 7). These results show that simple-cell-like receptive fields can be obtained by seeking efficient representations, and that their neural activities might encode natural scenes in the sparse and statistically independent manner.

### C. Remarks

The statistical regularities of natural image patches we have mentioned so far are summarized as 1) significant second-order correlations, 2) existence of sparse directions; and 3) they are non-orthogonal. Note that the distribution shown in Fig. 6b is a two-dimensional example of such distributions.

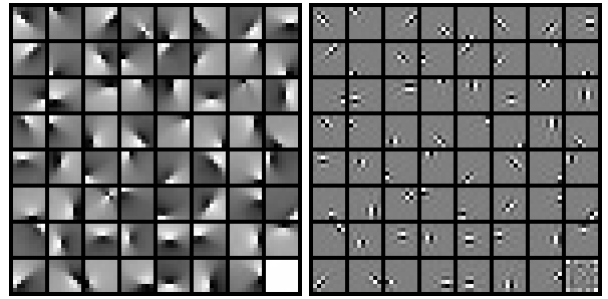


Fig. 7. ICA generates localized and oriented filters similar to V1 simple-cell receptive fields. These are a set of basis (left) and filter (right) functions for  $8 \times 8$  pixel natural image patches. The filter and basis functions correspond each other, and the one representing the DC component appears in the right-bottom corner.

Models like ICA and sparse coding form efficient codes, but, in contrast to whitening (which only removes correlations), they also remove higher-order redundancy. It can be summarized that the RGCs/LGN stage removes the second-order redundancy, and V1 stage removes the remaining higher-order redundancies. Note also that “independence” of the hidden variables is only achieved within the limits of the model, and higher-order dependencies certainly remain, which we will discuss in sec. VI.

## V. PRINCIPLED APPROACH TO COLOR VISION

Inspired by the sparse coding and ICA models, we examined what these models predict about color vision [9], [26]. The key is that the spatio-chromatic visual input for the visual nervous system is pre-processed by a trichromatic cone mosaic so that a single retinal location is sampled by only one of three classes of color photo-receptors [27]. This is of fundamental importance because the basic hypothesis states that computation in the visual system is optimized for the statistical regularities of its input. We first constructed a data set of cone mosaic responses to natural images that took into account the cone nonlinearity described above, as well as three types of photoreceptor spectral sensitivities and the spatial arrangement of the photo-receptors. Similar investigations have been done but without such considerations [28]–[31].

The whitening filters for cone mosaic responses show remarkable similarities to the parvo-cellular LGN cells. Namely, they have center-surround receptive field organizations, and they constitute two color axes (the so-called r-g and y-b axes) where the luminance information is multiplexed in the r-g channel. These results predict the visual response properties more accurately than the previous approach using PCA [7], [32].

ICA yielded simple-cell like receptive fields whose spectral sensitivity is given by the luminance function. In addition, the so-called double-opponent receptive fields emerged, the spatial organization of which is either center-surround or oriented, and the spectral sensitivity is along r-g or y-b axes. In Fig. 8 we show some of the receptive fields obtained by reverse correlation.

In addition to the similarity of the receptive field organization, the ICA model can reproduce the fMRI response of V1

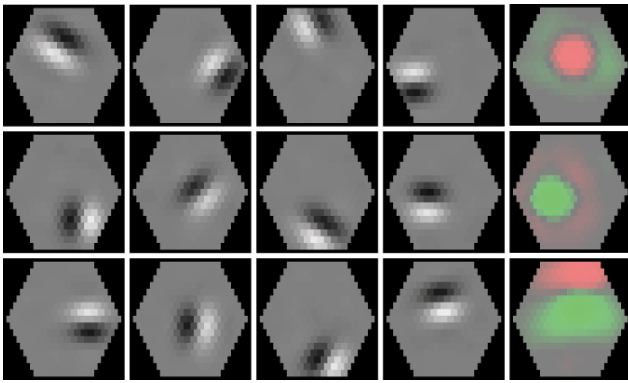


Fig. 8. ICA for cone mosaic responses to natural scenes reproduced spatio-chromatic receptive fields in V1 cells. The input consisted of 217 model cone photo-receptors arranged in the hexagonal array, and the number of coding units were  $2 \times 217 = 434$ . Here we show 15 examples of mapped receptive fields.

to cone contrast stimuli. V1 responds stronger to the red-green modulating stimuli compared to the luminance or yellow-blue modulation given the fixed cone contrast, which seems mysterious because the proportion of color cells in V1 is not more than 20%, and also because such responsiveness depends on a particular color axis [33]. The ICA model also reproduced the proportion of the cell types and the fMRI responses. We found that red-green modulation is unnatural compared to other modulations such as luminance or yellow-blue; and because such unnatural stimuli are not optimized to generate sparse representation they yield non-sparse representation, leading to the high activity of the population. This provides a novel and principled explanation to the observed phenomenon [34].

## VI. FURTHER EXTENSIONS

### A. Nonlinear models

Although ICA seeks statistically independent representations, the results on natural image patches exhibit residual dependencies [35], [36]. For instance, the variances of the basis function coefficients are dependent: when one coefficient has a large amplitude, other coefficients also tend to have large amplitudes. This implies that the data cannot be fully explained by the linear superposition of statistically independent basis functions.<sup>6</sup> One direct approach that reduces this type of redundancy is to normalize each coefficient by the magnitudes of other coefficients, leading to a more statistically independent representation [38]. Alternatively, we can extend the generative model to include higher-level hidden variables that serve as hyper-parameters and capture the variance dependencies among the ICA coefficients [36], [39]. These nonlinear models explain some of the more complex response properties of V1 neurons, such as gain control and phase invariance, further confirming the validity of the efficient coding hypothesis.<sup>7</sup>

<sup>6</sup>Such variance dependence is also observed among wavelet coefficients, which is another form of efficient linear codes of natural images [37].

<sup>7</sup>It has been proposed that the residual dependency could be utilized to organize topographic maps of ICA neurons, instead of removing the redundancy [35].

### B. Incorporation of biological constraints

Efficient coding can account for many aspects of visual response properties in the early visual system, but there remain important characteristics that have not been explained. One of these is the multi-scale nature of early visual representation, specifically that the spatial-frequency tuning of simple-cells in V1 is distributed across three octaves at each retinal eccentricity [40], [41]. In contrast, sparse coding and ICA yields single-scale representations as illustrated in Fig. 7 and 8.<sup>8</sup> Importantly, multi-scale representations are also found in RGCs and LGN [42]–[45], while whitening filters yield single-scale tuning (Fig. 3).<sup>9</sup>

One approach to this problem is to refine the theoretical model by incorporating more realistic biological constraints that could impact the form of the theoretically predicted optimal code. One assumption implicitly made by earlier models is that neurons have essentially infinite precision, because the neural activity is represented by floating point numbers. Real neurons, however, have limited information capacity, estimated to be as low as one bit per spike [47], [48]. In such noisy conditions, using efficient coding alone could be harmful because the redundancy is the only clue available to separate signal from noise. In this case, redundancy is useful and could be utilized to overcome the noise inherent in the limited neural capacity [49]. More generally, it should also be possible to increase the size of the neural population so that visual information is encoded more robustly. Physiologically, it is commonly believed that there is a bottleneck at the optic nerve fibers, and that the input should be compressed by decorrelation via PCA. However, as summarized in Table I, this is only true for the peripheral retinal region. In the foveal region, the number of cells always increases without any bottleneck, suggesting the possible existence of redundant representations.

Eccentricity	Cones	RGCs	LGN	V1 layer 4
0-10 deg	(1.0)	1.2	3.6	160
>10 deg	(1.0)	0.3	0.6	12

TABLE I

RATIO OF CELLS IN THE MACAQUE EARLY VISUAL SYSTEM.

Recently we developed theoretical models that take into account these computationally relevant biological constraints. We considered the optimal linear encoder and decoder<sup>10</sup> using noisy units similar to the neural representation [50], [51]. In this model, which we called “robust coding”, the computational objective is to minimize reconstruction errors instead of to infer the hidden variables, and therefore, the number

<sup>8</sup>Sparse coding is originally reported to yield multi-scale representation [6], but more detailed analyses using well-calibrated natural images provided by [8], [9] yielded single-scale representations similar to ICA results. Furthermore, in physiology there are many lower spatial-frequency tuning cells whereas there are a few, if any, in sparse coding or ICA representations.

<sup>9</sup>Other failures of prediction by efficient coding include the broad color tuning in V1 compared to an ICA model as discussed in [9] and the variety of the simple-cell receptive field’s shape compared to the prototypical receptive field organization in sparse coding and ICA [46].

<sup>10</sup>They correspond to filter and basis functions if the representation is complete, i.e., the number of coding units is equal to the input dimensionality.

of coding units is arbitrary instead of being restricted to the number of hypothetical hidden units. This allows us to model any retinal region where the ratio of neurons to photo-receptors are continuously changing from overcomplete (ratio>1) to undercomplete (ratio<1) representations. Analysis of the optimal solutions revealed that the optimal code changes its shape according to the channel noise level, covariance matrix of the input, and the number of coding units. Reconstruction subject to channel noise is dramatically improved over conventional transforms such as PCA, ICA, and wavelet, especially with  $8\times$  overcomplete representation (Fig.9). Theoretically, it can be shown that the reconstruction error can be made arbitrarily small by increasing the number of coding units [51].

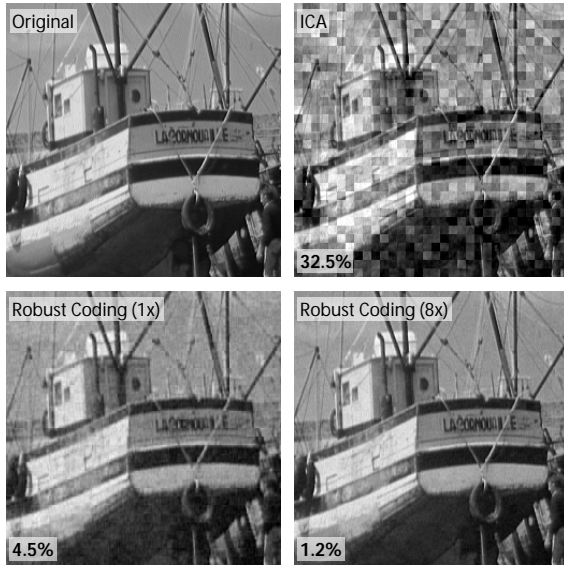


Fig. 9. Reconstruction results with limited capacity units. In this experiment channel noise is added such that the channel capacity of information is 1.0 bit for each coefficient. In each reconstruction image the percentage error of reconstruction is indicated.

In the above model, the cost function is defined by second-order statistics of data and noise. We augmented this model so that the representation should be sparse and could exploit the higher-order statistical regularities of input to improve coding efficiency. An example of resulting linear encoder and decoder is shown in Fig. 10: they look similar to those derived with ICA (Fig. 7); however, they consist of a larger number of coding units and also show broader spatial-frequency tuning similar to simple cells in V1. This model provides a basic framework to incorporate useful redundancy into efficient representation with arbitrary number of coding units.

VII. CONCLUSION

The examples we have reviewed in this article suggest that the response properties of the early visual system can be explained by statistical regularities of natural images, computational objectives such as efficient coding, and relevant biological constraints such as intrinsic noise. There remain, however, a number of outstanding issues. These include, along with those described in the last section, spatio-temporal representations [52], combining more biological constraints

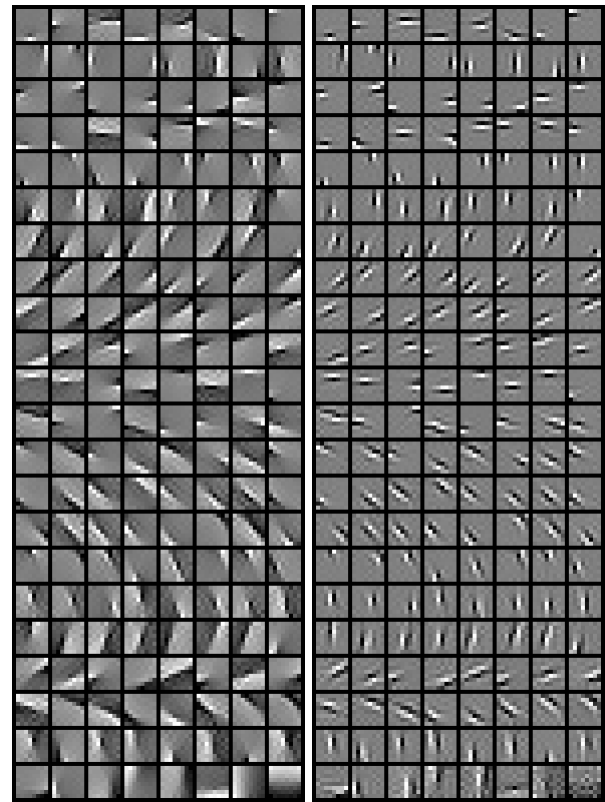


Fig. 10. Sparse overcomplete representation with noisy units yields broader spatial-frequency tuning. These are a subset of decoder (left) and encoder (right) vectors for  $8\times 8$  pixel natural image patches, sorted by spatial-frequency tuning. The full set consists of  $63\times 8 = 504$  coding units (intrinsic dimensionality is 63 because we removed the DC component; and this is  $8\times$  overcomplete representation).

(like energy efficiency or limited wiring length) with efficient coding ideas [3], the impact of tasks and behaviorally relevant sensory signals on the neural code [53], and simulations of realistic visual systems and environments (including the optics and the sensor arrangement of the eye, eye movements, etc).

The approach described in this article is generic and can be applied to any sensory modality. It has been suggested that the auditory neural code is adapted to efficiently encode natural sounds [53]. More recent results have shown that spikes themselves play an essential role in forming an efficient code for auditory signals [54], [55]. These and other results suggest that the principle of efficient coding, when combined with realistic biological constraints, promises to shed light on other forms of sensory coding and perception.

ACKNOWLEDGEMENT

We thank Yan Karklin for his helpful comments on the manuscript.

REFERENCES

- [1] E. P. Simoncelli and B. A. Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24:1193–216, 2001.
- [2] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu. On advances in statistical modeling of natural images. *Journal of Mathematical Imaging and Vision*, 18:17–33, 2003.
- [3] S. Laughlin and T. J. Sejnowski. Communication in neuronal networks. *Science*, 301(1870-1874), 2003.

- [4] B. A. Olshausen and D. J. Field. How close are we to understanding V1? *Neural Computation*, 17:1665–1699, 2005.
- [5] P. Dayan, G. E. Hinton, R. M. Neal, and R. S. Zemel. The Helmholtz machine. *Neural Computation*, 7:889–904, 1995.
- [6] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.
- [7] D. L. Ruderman, T. W. Cronin, and C.-C. Chiao. Statistics of cone responses to natural images: implications for visual coding. *Journal of Optical Society of America A*, 15:2036–2045, 1998.
- [8] 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. B*, 265:359–366, 1998.
- [9] E. Doi, T. Inui, T.-W. Lee, T. Wachtler, and T. J. Sejnowski. Spatio-chromatic receptive field properties derived from information-theoretic analyses of cone mosaic responses to natural scenes. *Neural Computation*, 15:397–417, 2003.
- [10] S. Laughlin. A simple coding procedure enhances a neuron’s information capacity. *Z. Naturforsch.*, 36(c):910–912, 1981.
- [11] D. A. Baylor, B. J. Nunn, and J. L. Schnapf. Spectral sensitivity of cones of the monkey macaca fascicularis. *Journal of Physiology*, 390:145–160, 1987.
- [12] T. von der Twer and D. I. A. Macleod. Optimal nonlinear codes for the perception of natural colors. *Network: Comput. Neural Syst.*, 12:395–407, 2001.
- [13] J. H. van Hateren and H. P. Snippe. Information theoretical evaluation of parametric models of gain control in blowfly photoreceptor cells. *Vision Research*, 41:1851–1865, 2001.
- [14] H. B. Barlow. Possible principles underlying the transformation of sensory messages. In W. A. Rosenblith, editor, *Sensory communication*, pages 217–234. MIT Press, MA, 1961.
- [15] J. J. Atick. Could information theory provide an ecological theory of sensory processing? *Network*, 3:213–251, 1992.
- [16] 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.
- [17] A. Torralba and A. Oliva. Statistics of natural image categories. *Network: Comput. Neural Syst.*, 14:391–412, 2003.
- [18] J. J. Atick and A. N. Redlich. Towards a theory of early visual processing. *Neural Computation*, 2:308–320, 1990.
- [19] J. J. Atick and A. N. Redlich. What does the retina know about natural scenes? *Neural Computation*, 4:196–210, 1992.
- [20] J. J. Atick and A. N. Redlich. Convergent algorithm for sensory receptive field development. *Neural Computation*, 5:45–60, 1993.
- [21] A. J. Bell and T. J. Sejnowski. The independent components of natural scenes are edge filters. *Vision Research*, 37:3327–3338, 1997.
- [22] D. J. Field. What is the goal of sensory coding? *Neural Computation*, 6:559–601, 1994.
- [23] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A*, 2:1160–1169, 1985.
- [24] J. P. Jones and L. A. Palmer. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58:1233–1258, 1987.
- [25] A. Hyvarinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley, 2001.
- [26] E. Doi and T. Inui. Self-organization of spatio-chromatic receptive fields in the early visual system by ICA. In *Technical report of IEICE*, volume NC98-170, pages 131–138, Tokyo, Japan, 1999.
- [27] A. Roorda and D. R. Williams. The arrangement of the three cone classes in the living human eye. *Nature*, 397:520–522, 1999.
- [28] D. R. Taylor, L. H. Finkel, and G. Buchsbaum. Color-opponent receptive fields derived from independent component analysis of natural images. *Vision Research*, 40:2671–2676, 2000.
- [29] P. O. Hoyer and A. Hyvarinen. Independent component analysis applied to feature extraction from color and stereo images. *Network*, 11:191–210, 2000.
- [30] T. Wachtler, T.-W. Lee, and T. J. Sejnowski. Chromatic structure of natural scenes. *Journal of Optical Society of America A*, 18:65–77, 2001.
- [31] M. S. Caywood, B. Willmore, and D. J. Tolhurst. Independent components of color natural scenes resemble V1 neurons in their spatial and color tuning. *Journal of Neurophysiology*, 91:2859–2873, 2004.
- [32] G. Buchsbaum and A. Gottschalk. Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proc. R. Soc. Lond. B*, 220:89–113, 1983.
- [33] S. Engel, X. Zhang, and B. Wandell. Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature*, 388:68–71, 1997.
- [34] E. Doi, T. Inui, T.-W. Lee, and T. J. Sejnowski. A model study of V1 simple-cells and color cells: close analyses of receptive fields and response properties. *unpublished manuscript*, 2003.
- [35] A. Hyvarinen, P. O. Hoyer, and M. Inki. Topographic independent component analysis. *Neural Computation*, 13:1527–1558, 2001.
- [36] Y. Karklin and M. S. Lewicki. Learning higher-order structures in natural images. *Network: Comput. Neural Syst.*, 14:483–499, 2003.
- [37] R. W. Buccirossi and E. P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE trans. on image processing*, 8:1688–1701, 1999.
- [38] O. Schwartz and E. P. Simoncelli. Natural signal statistics and sensory gain control. *Nature Neuroscience*, 4:819–825, 2001.
- [39] Y. Karklin and M. S. Lewicki. A hierarchical bayesian model for learning nonlinear statistical regularities in nonstationary natural signals. *Neural Computation*, 17:397–423, 2005.
- [40] R. L. De Valois, D. G. Albrecht, and L. G. Thorell. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22(545-559), 1982.
- [41] C. H. Anderson and G. C. DeAngelis. Population codes and signal to noise ratios in primary visual cortex. In *Society for Neuroscience Abstract*, page 822.3, 2004.
- [42] L. J. Croner and E. Kaplan. Receptive fields of p and m ganglion cells across the primate retina. *Vision Research*, 35:7–24, 1995.
- [43] B. B. Lee, J. Kremers, and T. Yeh. Receptive fields of primate retinal ganglion cells studied with a novel technique. *Visual Neuroscience*, 15:161–175, 1998.
- [44] A. M. Derrington and P. Lennie. Spatial and temporal contrast sensitivities of neurons in lateral geniculate nucleus of macaque. *Journal of Physiology*, 357:219–240, 1984.
- [45] M. J. McMahon, M. J. M. Lankheet, P. Lennie, and D. R. Williams. Fine structure of parvocellular receptive fields in the primate fovea revealed by laser interferometry. *Journal of Neuroscience*, 20:2043–2053, 2000.
- [46] D. L. Ringach. Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology*, 88:455–463, 2002.
- [47] A. L. Fairhall, G. D. Lewon, W. Bialek, and R. R. de Ruyter van Steveninck. Efficiency and ambiguity in an adaptive neural code. *Nature*, 412:787–792, 2001.
- [48] N. K. Dhingra and R. G. Smith. Spike generator limits efficiency of information transfer in a retinal ganglion cell. *Journal of Neuroscience*, 24:2914–2922, 2004.
- [49] H. B. Barlow. Redundancy reduction revisited. *Network: Comput. Neural Syst.*, 12:241–253, 2001.
- [50] E. Doi and M. S. Lewicki. Sparse coding of natural images using an overcomplete set of limited capacity units. In *Advances in Neural Information Processing Systems*, volume 17, pages 377–384. MIT Press, 2005.
- [51] E. Doi, D. C. Balcan, and M. S. Lewicki. A theoretical analysis of robust coding over noisy overcomplete channels. In *Advances in Neural Information Processing Systems*, submitted.
- [52] B. A. Olshausen. Sparse codes and spikes. In R. P. N. Rao, B. A. Olshausen, and M. S. Lewicki, editors, *Probabilistic models of the brain: perception and neural function*. MIT Press, 2002.
- [53] M. S. Lewicki. Efficient coding of natural sounds. *Nature Neuroscience*, 5:356–363, 2002.
- [54] E. C. Smith and M. S. Lewicki. Learning efficient auditory codes using spikes predicts cochlear filters. In *Advances in Neural Information Processing Systems*, volume 17. MIT Press, 2005.
- [55] E. C. Smith and M. S. Lewicki. Efficient auditory coding. submitted.