

Large-scale failures of $f^{-\alpha}$ scaling in natural image spectra

Michael S. Langer

NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540

Received January 19, 1999; accepted July 16, 1999; revised manuscript received August 11, 1999

Several studies have demonstrated that the power spectra of natural image ensembles scale as $f^{-\alpha}$. A stronger claim that has been made is that the power spectra of single natural images typically also scale as $f^{-\alpha}$. Results are presented that challenge this latter claim. The results are based on a method for estimating large-scale structure in single images that compares aliasing artifacts produced by image windows of different shape. Failures of $f^{-\alpha}$ scaling are found at large scales in many natural images. These failures cannot be accounted for by $f^{-\alpha}$ scaling models such as a linear superposition model or a model based on two-dimensional occlusions in the image plane. The results imply that claims about $f^{-\alpha}$ scaling in single natural images have been exaggerated. The results also offer insight into why such failures of $f^{-\alpha}$ scaling occur. © 2000 Optical Society of America [S0740-3232(00)00301-X]

OCIS codes: 100.2960, 100.2000, 070.2590, 070.4790, 330.5000, 330.6130.

1. INTRODUCTION

Natural images possess a remarkable amount of statistical regularity. There are spatial correlations within images, as nearby pixels tend to measure the luminances of nearby surface points in a scene and nearby surface points tend to have similar illumination and reflectance. There are also correlations between images as objects occur at varying sizes and positions from one image to the next. It has been argued that since there is no natural scale at which objects typically occur in images, natural image ensembles should contain roughly constant energy in constant-octave bands and hence the power spectrum of an ensemble of natural images should scale as f^{-2} , where f is spatial frequency.¹ One qualifier to this scale-invariance argument, however, is that occlusions between opaque objects can give rise to more general power-law scaling.² Such generalized $f^{-\alpha}$ scaling in the spectra of image ensembles was described in Refs. 3 and 4.

A stronger claim that has been made about natural image spectra is that single images also scale as $f^{-\alpha}$. Such $f^{-\alpha}$ scaling in single images was reported in Refs. 1 and 5 for the case of $\alpha = 2$, and much attention has been paid since to the range of best-fitting coefficients ($\alpha \neq 2$) that one finds over a set of natural images.^{3,6-9} The present study is motivated by a concern that the author had with these claims of $f^{-\alpha}$ scaling in single images, namely, that there was little accompanying evidence that a $f^{-\alpha}$ model provided a good fit to the single-image spectra. The only study that did address the goodness-of-fit issue concluded in the negative, namely, that single-natural-image spectra were often significantly better fitted by a quadratic model in log-log coordinates than by a linear model.⁷

The question of whether natural images typically scale as $f^{-\alpha}$ is an important one, especially if claims are to be made about how $f^{-\alpha}$ scaling has influenced the evolution of biological vision systems, a question that lies at the heart of studies of natural image spectra. Stated

broadly, scaling properties of image ensembles are relevant to the evolution of a visual system that occurs over time scales of generations,¹⁰ whereas scaling properties of single natural images are relevant to processes of contrast adaptation that occur over time scales of the order of seconds.^{11,12} To understand the different time scales over which adaptations of visual systems can occur and how these adaptations depend on the statistics of the environment, we need to better understand how scaling properties of single images differ from those of an ensemble of images.

This paper focuses on one neglected issue in studies of $f^{-\alpha}$ scaling of natural images, which is that the scaling behavior of single images is inherently difficult to estimate at large scales. In a single image, the number of spatial-frequency components in a given constant-octave bandwidth is proportional to f^2 , and hence the standard error of the mean amplitude is proportional to f^{-1} . For small f (that is, for large scales), estimates of mean amplitude will thus be relatively noisy. If a single image were drawn from an ensemble that failed to scale as $f^{-\alpha}$ at large scales, then this failure would be hidden by the noise at large scales. To the author's knowledge, this issue of how to estimate the scaling behavior at large scales in single images has not been addressed in previous studies.

This issue is addressed here by introducing a method for estimating large-scale structure in single images. The method compares two spectra obtained from a single image by using two different image windows. The idea is that different windows produce different aliasing artifacts from large-scale structure, and these artifacts are systematically related both to the shape of the window and to the power spectrum of the ensemble from which the image was drawn, in particular, to the amount of large-scale structure in the ensemble. By comparing the aliasing artifacts obtained with the use of two different windows, we

obtain a statistic that crudely measures the amount of large-scale structure in the image.

The paper is organized as follows. In Section 2 we review how image windows can yield biased power spectrum estimates, and we use this bias to motivate a statistic for measuring large-scale structure in a single image. The statistic compares the power spectra obtained by using two image windows: a square window and a diamond window. In Section 3 we apply this statistic to an ensemble of roughly 10,000 natural image tiles drawn from the van Hateren database and use the statistic to partition the ensemble of tiles based on values of the statistic. We find failures of $f^{-\alpha}$ scaling in a large percentage of the tiles. Two control experiments are then presented to show that the failures are not just an artifact of the method but rather reflect failures of $f^{-\alpha}$ scaling within single images in the ensemble.

2. THEORY

A. Large-Scale Aliasing and Bias

We begin by reviewing how image windows can yield biased power spectrum estimates. Consider a model image ensemble in which each image is defined on an $N \times N$ viewing torus. Real images are, of course, defined on a viewing sphere, not on a viewing torus, but since we will consider small windowed images only, we argue that there is no loss of generality in assuming a torus rather than a sphere. We assume a window of size $M \times M$ with $M \ll N$.

A torus and a sphere each have a maximum angular scale of 2π rad. A natural image may contain structure up to this maximum scale, but not larger. When an image is windowed, intensity structure at scales larger than M pixels is aliased to scales smaller than M pixels. This "large-scale aliasing" is quantified as follows. Assume that the autocorrelation function of the toroidal image ensemble is translation invariant, and so the power spectrum $R(\mathbf{f})$ of the ensemble is the discrete Fourier transform of the autocorrelation function, where $\mathbf{f} = (f_x, f_y)$ is the two-dimensional spatial frequency. Both f_x and f_y belong to $\{-N/2, \dots, N/2 - 1\}$.

Let $W(\mathbf{f})$ be the discrete Fourier transform of the $M \times M$ image window. It is easy to show that the power spectrum of the windowed image ensemble is the convolution product $R(\mathbf{f}) * |W(\mathbf{f})|^2$. The power spectrum of the windowed ensemble will differ from that of the nonwindowed ensemble in general, and this difference may be expressed by a bias, which we define as follows:

$$\text{bias}(\mathbf{f}) \equiv \log_2 \frac{R(\mathbf{f}) * |W(\mathbf{f})|^2}{R(\mathbf{f})}. \quad (1)$$

If the bias is positive for a given frequency \mathbf{f} , then the windowed ensemble has greater power at that frequency on average than the underlying toroidal ensemble. If the bias is negative for a given frequency \mathbf{f} , then the windowed image ensemble has less power at that frequency on average than the underlying toroidal ensemble. In what follows, we assume that N is a multiple of M , and we

restrict our attention to frequencies f_x and f_y that are multiples of N/M , which is the fundamental frequency of the image window.

To understand the bias that different windows can produce, we consider an image ensemble having power spectrum

$$R(\mathbf{f}) \equiv (c + |\mathbf{f}|)^{-\alpha}. \quad (2)$$

We consider several cases of the constants (c, α) to illustrate how bias depends on the large-scale structure in the ensemble and on the window. Figure 1 shows the bias produced for three cases of (c, α) and for two different windows, namely, a square window and a diamond window. The diamond window is just a square window that has been rotated by 45° and reduced in area by a factor of 2 so that it fits within the $M \times M$ square window.

First, we compare the bias of the square window with that of the diamond window. In all three cases of (c, α) shown, the square window produces a positive bias along the horizontal and vertical orientations but zero bias along diagonal orientations, yielding a + shape in the two-dimensional frequency domain. This + bias is well-known. It is due to the boundaries of the square window, which produce horizontal and vertical intensity edges in the image. Similarly, the diamond window produces a positive bias along diagonal orientations but zero bias along the horizontal and vertical frequencies, yielding a \times shape in the frequency domain. This \times bias is due to the boundaries of the diamond window, which produce diagonal intensity edges in the image.

Next, we compare how the bias depends on (c, α) . For the three cases shown, the least bias is produced in the first case, where $c > 0$. In this case $f^{-\alpha}$ scaling fails to occur at large scales because the spectrum is finite as f goes to zero, rather than infinite as in cases 2 and 3. Hence there is relatively little large-scale structure in the ensemble in case 1. The greatest bias is found in the third case, in which the ensemble has more energy at large scales owing to the larger value of α .

The idea that the bias depends on the large-scale structure will become important in Section 3. There we will estimate the amount of large-scale structure in single natural images by estimating the bias. In Subsection 2.B the statistic that will be used in these experiments is introduced.

B. Estimating Bias in a Single Image

We use the observations of Subsection 2.A to motivate a statistic that estimates the bias in a single image. The idea is to compare the spectra obtained by using two different windows. For a given window $w(\mathbf{x})$, the amplitude spectrum obtained by using that window is

$$A_w(\mathbf{f}) = \left| \sum_{\mathbf{x}} w(\mathbf{x}) [I(\mathbf{x}) - \mu] \exp(i\mathbf{f} \cdot \mathbf{x}) \right|.$$

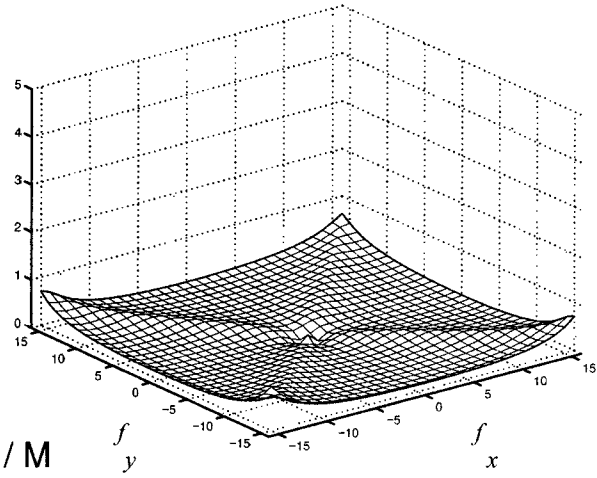
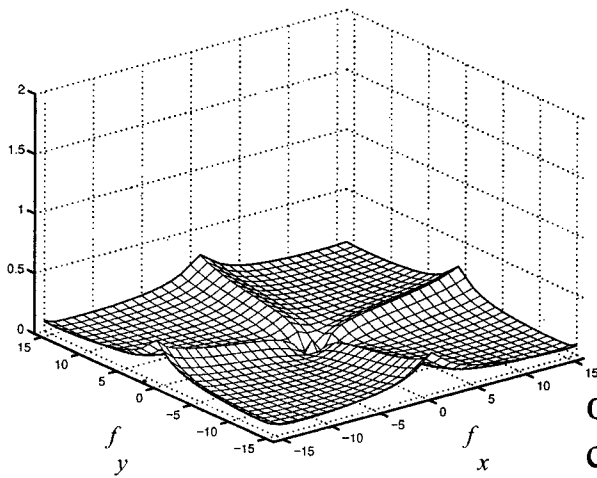
A constant μ has been subtracted from the image, where

$$\mu \equiv \frac{\sum_{\mathbf{x}} I(\mathbf{x}) w(\mathbf{x})}{\sum_{\mathbf{x}} w(\mathbf{x})},$$

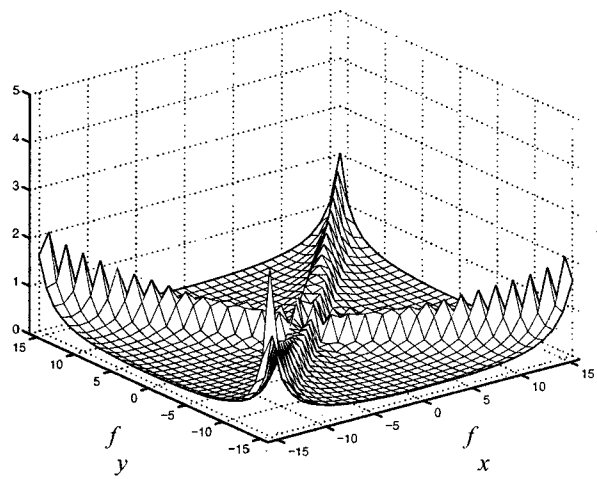
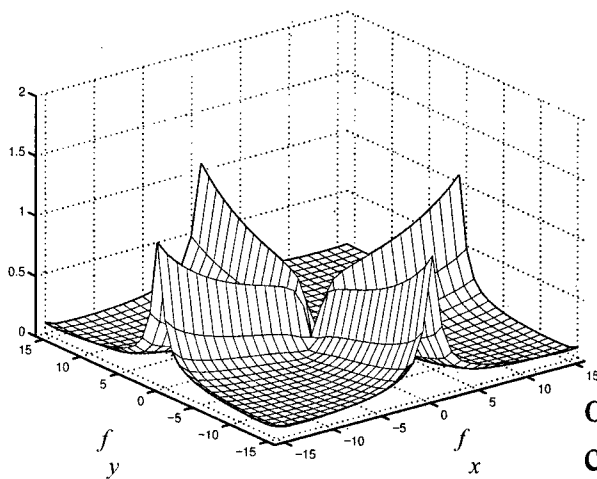
to reduce dc leakage.³

square window ('+' bias)

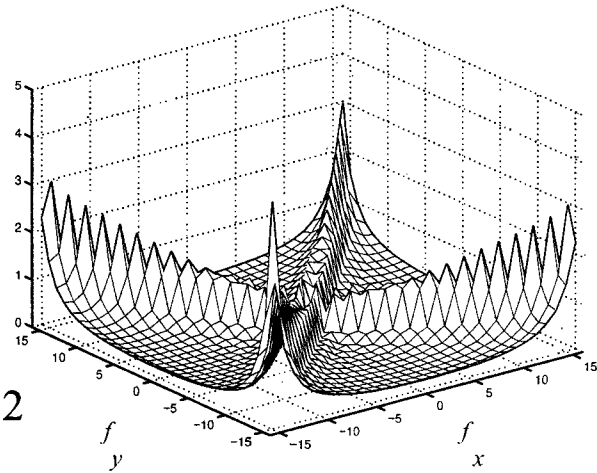
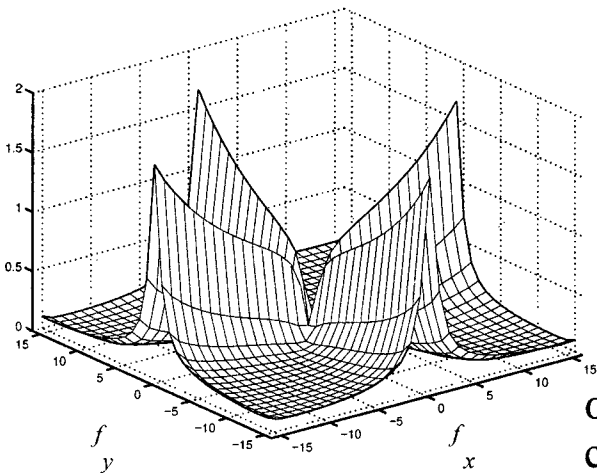
diamond window ('x' bias)



$\alpha = 2$
 $c = N/M$



$\alpha = 2$
 $c = 0$



$\alpha = 2.2$
 $c = 0$

Fig. 1. Mesh plots of the + bias from a square window and the x bias from a diamond window, obtained from image ensembles defined by Eqs. (1) and (2) and for various values of (c, alpha).

Let $A_+(\mathbf{f})$ denote the amplitude spectrum obtained by using the square window, and let $A_\times(\mathbf{f})$ denote the amplitude spectrum obtained by using the diamond window. The subscripts here indicate the orientations at which bias occurs, that is, a $+$ bias or a \times bias (recall Fig. 1). Define the statistic β in terms of the diagonal frequency components of these two spectra:

$$\beta \equiv \sum_{f=f_1}^{f_2} \log_2 \frac{A_\times(f, f)A_\times(f, -f)}{A_+(f, f)A_+(f, -f)}. \quad (3)$$

The statistic β estimates the average bias produced by the diamond window relative to that by the square window and along diagonal frequencies. In these experiments we arbitrarily choose $f_1 = 5$ and $f_2 = M/4$. The reason for ignoring small f is that there is little bias at large scales (see Fig. 1), and it is bias that we are trying to estimate. The reason for ignoring large f is that these frequencies tend to be contaminated by pixel noise.

Finally, we observe that we could have defined β alternatively in terms of the horizontal and vertical frequencies ($f, 0$) and ($0, f$) rather than the diagonal ones (f, f) and ($f, -f$), since the square window yields a positive bias at horizontal frequencies and the diamond window yields zero bias at horizontal frequencies. We choose not to define β in this way to avoid a possible confound. Horizontal and vertical frequencies are known to contain more energy in natural images because of gravitational effects in the environment.^{3,13,14} It is not well understood how the scaling properties of images may differ at these frequencies. By using diagonal frequencies instead, we are able to avoid this confound.

3. EXPERIMENTS

A. van Hateren Ensemble

We use the statistic β to study an ensemble of natural images from the van Hateren database.¹⁵ We partition each of the images numbered 1–400 from this database into 24 mutually disjoint square tiles, each of size $M = 256$, yielding a total of 9600 tiles. For each tile we compute the statistic β . We then partition the entire set of tiles into three equal-size classes based on disjoint intervals of β . We refer to these as the low- β , medium- β , and high- β classes.

Figure 2(a) shows log–log plots of the average power spectrum of the low- β and high- β classes as a function of f and averaged over all orientations. The two curves have similar linear behavior (similar mean values of the best-fitting slope α ; see Table 1) but quite different nonlinear behavior. The average power spectrum of the low- β class is convex overall, and that of the high- β class is concave overall. (The medium- β curve is not shown but falls between the two and is approximately linear.) Thus the main difference between the low- β and high- β classes is primarily in the nonlinear rather than in the linear behavior.

To what extent does this result indicate a failure of $f^{-\alpha}$ scaling in single-image tiles, and to what extent is it merely an artifact of the method that we used? To address this question, we carry out two control experiments

in which we repeat the experiment by using two synthetic image ensembles that are known to scale as $f^{-\alpha}$.

B. Gaussian Ensemble

In the first control experiment, each image was synthesized in the frequency domain as follows. The Fourier co-

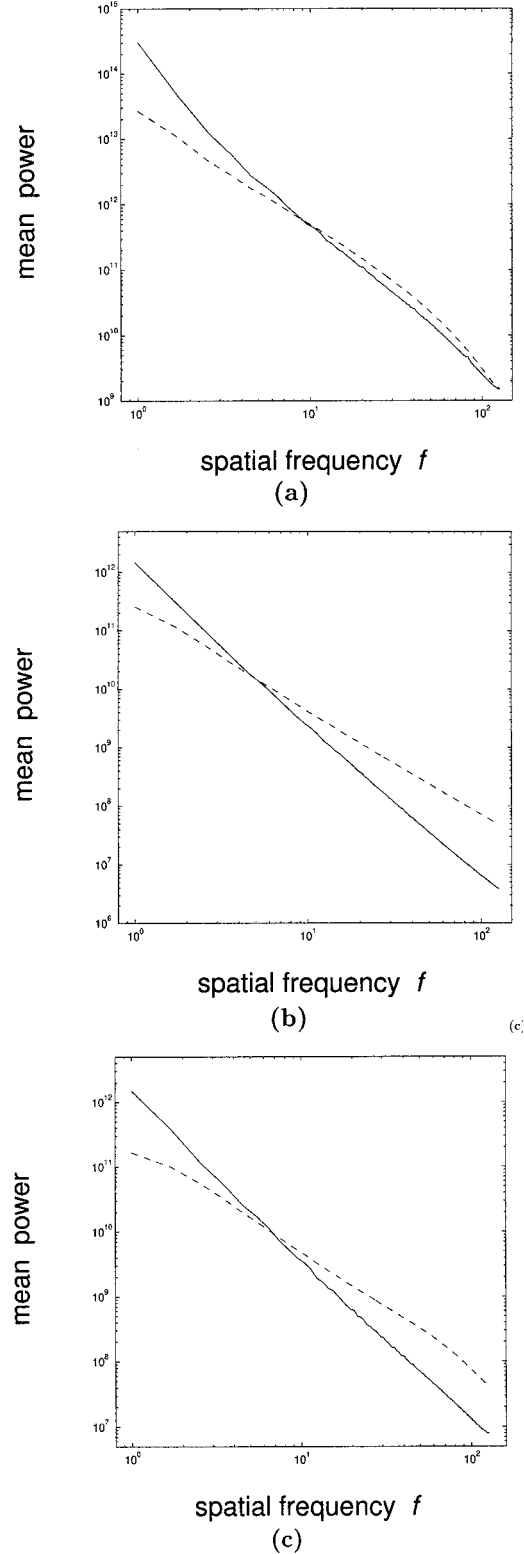


Fig. 2. Mean power spectrum of the low- β (dashed curves) and high- β (solid curves) classes for the (a) van Hateren, (b) Gaussian, and (c) Ruderman ensembles.

Table 1. Mean Values of α for the Images in the Low- β and High- β Classes and for Each of the Three Ensembles Tested

Ensemble	Low β	High β	Overall ^a
van Hateren	2.15	2.36	2.26 ± 0.57
Gaussian	1.78	2.59	2.25 ± 0.56
Ruderman	1.79	2.46	2.24 ± 0.52

^aOverall mean and standard deviation of the α 's for each ensemble. By design, these are roughly identical between ensembles.

efficients were chosen to be statistically independent Gaussian random variables with mean zero and variance $f^{-\alpha}$. The power-law coefficient α was chosen randomly for each synthetic image from the interval (1.4, 3). This range of α 's was used so that the mean and the variance of the α 's were similar to those of the van Hateren ensemble (see column 4 of Table 1). Ninety such images were generated, each of size $N = 1024$. Each image was partitioned into 16 disjoint tiles of size $M = 256$, and the statistic β was computed for each tile. The set of tiles was then partitioned into three classes based on β , as was done for the van Hateren images.

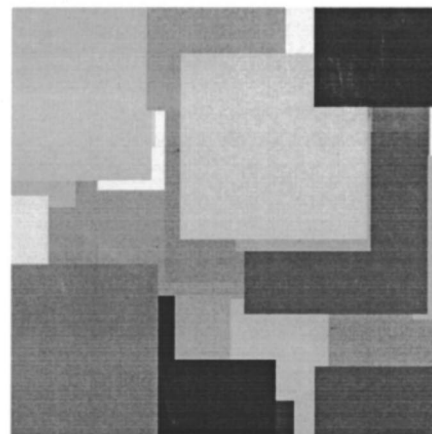
The average power spectra of the low- β and high- β classes for this Gaussian ensemble are shown in Fig. 2(b). The two curves are each roughly linear but differ enormously in slope (see Table 1). There is a slight nonlinearity at the fundamental frequency for the low- β class, which is presumably an artifact of the method, but this effect is small compared with the linear effect. That is, the difference between the mean α 's of the low- β and high- β classes is much greater than what was found for the van Hateren images. We conclude that for the Gaussian $f^{-\alpha}$ ensemble, partitioning image tiles based on β yields large differences in linear behavior of the power spectrum of single tiles and only small differences in nonlinear behavior. This suggests that the relatively large nonlinear behavior observed for the van Hateren images was unlikely to be a mere artifact of the method.

One concern with using Gaussian $f^{-\alpha}$ images as a control, however, is that they have much less spatial structure than natural images. To put the results from the van Hateren ensemble to a further test, we carry out a second control experiment by using an ensemble that contains more coherent spatial structure.

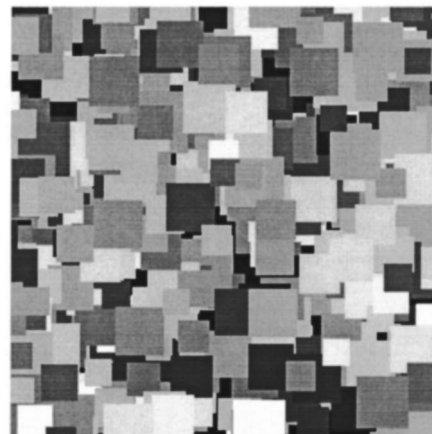
C. Ruderman Ensemble

The second control experiment was based on the synthetic image ensemble introduced by Ruderman.² Each toroidal image ($N = 1024$) was composed of overlapping opaque squares. The squares had a random gray-level intensity and were distributed uniformly in position. Square size was distributed according to a power law, with smaller squares being more likely than larger squares. We used a power law on square sizes $s^{-\gamma}$, where γ was chosen randomly for each image from the interval (2, 3). Each such γ yielded a roughly $f^{-\alpha}$ scaling behavior in the frequency domain, with the α depending on the γ . (For the overall ensemble, there was a slight failure of $f^{-\alpha}$ scaling at small scales, but this was due to a 2-pixel lower bound on the size of squares.) As in the

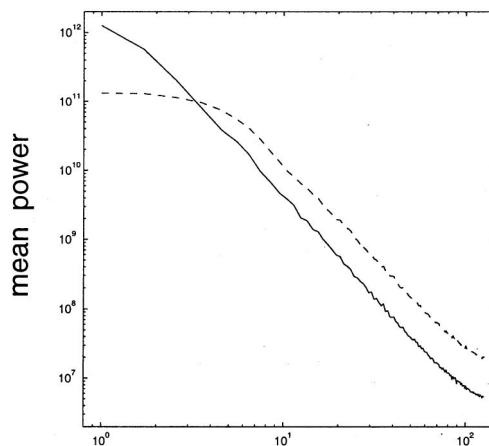
Gaussian ensemble, the range of γ 's was chosen so that the mean and variance of the α 's matched those of the van Hateren images (see Table 1). Two hundred such toroi-



(a)



(b)



(c)

Fig. 3. Two windowed images are shown ($M = 256$). These are composed of overlapping squares of widths in the ranges (a) 64–128 pixels and (b) 16–32 pixels. (c) The mean power spectra of 100 such images are shown for the larger square images in (a) (solid curve) and the smaller square images in (b) (dashed curve).

dal images were generated and partitioned into tiles of disjoint tiles of size $M = 256$, β was computed for each tile, etc.

The results are shown in Fig. 2(c). As with the Gaussian ensemble, the average power spectra of the low- β and high- β classes are both roughly linear. There is a small nonlinearity at the fundamental frequency, again presumably an artifact of the method, but the dominant effect is in the linear behavior, as was the case in the Gaussian ensemble (see Table 1). Thus, again we see that for an ensemble that is known to scale as $f^{-\alpha}$, partitioning the image tiles by using the statistic β yields large differences in the linear scaling behavior and only small differences in the nonlinear behavior. We conclude that the nonlinear effects found in the van Hateren ensemble were not merely an artifact of the method but rather demonstrated large-scale failures of $f^{-\alpha}$ scaling in single images in the ensemble.

4. DISCUSSION

What causes the failures of $f^{-\alpha}$ scaling in single images that we observed in the van Hateren ensemble? Some insight into this question may be gained from Figs. 3(a) and 3(b), which show two examples of synthetic images. Each image is composed of overlapping squares whose widths are distributed over a small range of sizes, with the range differing between the two images. Each of the two images has relatively little large-scale structure, since the maximum square size is less than the width of the image window. Plots of the power spectrum of 100 such images of each type yield $f^{-\alpha}$ scaling at small scales but not at large scales [see Fig. 3(c)]. For the images composed of larger squares, the failure occurs near the fundamental of the image window. For the images composed of smaller squares, the failure occurs at relatively higher spatial frequencies.

Such convex failures of $f^{-\alpha}$ scaling will occur for any image that is composed of objects of a bounded range of sizes such that the upper bound is smaller than the width of the window. Examples are abundant in nature: An image of grass would fail to scale as $f^{-\alpha}$ at scales larger than a single blade. Observe that if the viewing position were to vary from one image to the next, so would the projected size of each object in the image and thus so would the scale f at which the failures of $f^{-\alpha}$ scaling occur. If the images obtained from different viewing distances were grouped into an ensemble, the overall result would be an average ensemble power spectrum that is slightly convex.

What about concave failures of $f^{-\alpha}$ scaling? Such failures can occur when an image is dominated by large-scale structure; for example, an image of a smooth intensity gradient across part of an overcast sky or of a penumbra on the ground. Although it may be tempting to exclude such images from studies of natural image spectra be-

cause they do not contain rich and interesting structure, such images are common in nature and an account of them should be given as well. It may be that large-scale failures of $f^{-\alpha}$ scaling were not noted by previous investigators because the convex and concave failures of $f^{-\alpha}$ scaling, when they are averaged together, serve to cancel each other out.

ACKNOWLEDGMENTS

I am grateful to the reviewers and to J. Oliensis, A. Jepson, D. Ruderman, and W. Bialek for their comments and suggestions.

Address correspondence to the author at Max-Planck Institute for Biological Cybernetics, Spemannstrasse 38, Tübingen 72076 Germany; phone, 49-7071-601-610; fax, 49-7071-601-616; e-mail, michael.langer@tuebingen.mpg.de.

REFERENCES

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. L. Ruderman, "Origins of scaling in natural images," *Vision Res.* **37**, 3385–3398 (1997).
3. A. van der Schaaf and J. H. van Hateren, "Modelling the power spectra of natural images: statistics and information," *Vision Res.* **36**, 2759–2770 (1996).
4. D. L. Ruderman and W. Bialek, "Statistics of natural images: scaling in the woods," *Phys. Rev. Lett.* **73**, 814–817 (1994).
5. G. J. Burton and I. R. Moorhead, "Color and spatial structure in natural scenes," *Appl. Opt.* **26**, 157–170 (1987).
6. J. H. van Hateren, "Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation," *J. Comp. Physiol. A* **171**, 157–170 (1992).
7. D. J. Tolhurst, Y. Tadmor, and T. Chao, "Amplitude spectra of natural images," *Ophthalmic Physiol. Opt.* **12**, 229–232 (1992).
8. D. J. Field and N. Brady, "Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes," *Vision Res.* **37**, 3367–3383 (1997).
9. C. A. Parrago, G. Brelstaff, and T. Troscianko, "Color and luminance information in natural scenes," *J. Opt. Soc. Am. A* **15**, 563–569 (1998).
10. J. N. Lythgoe, *The Ecology of Vision* (Clarendon, Oxford, UK, 1979).
11. M. V. Srinivasan, S. B. Laughlin, and A. Dubs, "Predictive coding: a fresh view of inhibition in the retina," *Proc. R. Soc. London Ser. B* **216**, 427–459 (1982).
12. J. J. Atick and A. N. Redlich, "What does the retina know about natural scenes?" *Neural Comput.* **4**, 196–210 (1992).
13. E. Switkes, M. J. Mayer, and J. A. Sloan, "Spatial frequency analysis of the visual environment: anisotropy and the carpentered environment hypothesis," *Vision Res.* **18**, 1393–1399 (1978).
14. D. L. Ruderman, "The statistics of natural images," *Network* **5**, 517–548 (1994).
15. 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. London Ser. B* **265**, 359–366 (1998).