Scaling and Power Spectra of Natural Images

R. P. Millane, S. Alzaidi and W. H. Hsiao
Department of Electrical and Computer Engineering
University of Canterbury
Private Bag 4800, Christchurch, New Zealand
rick@elec.canterbury.ac.nz

Abstract

The effects of occlusion, scaling and edges in images on the dependence of the power spectrum on spatial frequency is studied using a simple model. The results are compared to the power spectrum behaviour that has been observed for natural images: images of the natural environment that are processed by the human visual system.

Keywords: Natural images, scaling, power spectrum, visual processing

1   Introduction

The human (and other vertebrates) visual system extracts information from the environment. The early visual system samples the visual field rather coarsely relative to the full information content, so that an efficient image-coding scheme is an important part of visual processing [1]. An efficient coding scheme depends on the statistics of the input and the information required in the output. It has been argued that the visual system has developed in such a way that the coding scheme is optimised for the properties of the visual world, as a way of removing image redundancy [1]. Images from the natural environment, that we refer to as *natural images*, are not random patterns, but show a number of consistent statistical properties. The study of such properties is therefore of relevance to the detailed makeup of the visual system, i.e. why visual neurons behave the way they do. Since modern techniques of image display are based in part on image models as well as models of visual perception, such properties of images are of technological interest also.

A property of natural images that has received considerable attention by many authors is the characteristic that the rotationally averaged spatial power spectrum of image ensembles $S(f)$ behaves approximately as $f^{-\gamma}$ where f is spatial frequency and $\gamma \approx 2$ (typically $1.8 < \gamma < 2.3$) [1, 2, 3, 4]. Individual images also exhibit similar behaviour although there is more irregularity in the linear dependence of the log power spectrum versus $\log f$, and the slopes can deviate more from 2 [2]. An example for a single image is shown in Fig. 1. This property is often referred to as *scaling*, although we will reserve this term here for properties of images in the spatial, rather than spatial frequency, domain. This characteristic behaviour of the power spectrum has been discussed in terms of the response properties of visual cortical cells [1].

Relevant questions are then: (1) What are the salient statistical characteristics of natural images, and (2) What is it about these statistics that produces the characteristic dependence of the power spectrum on spatial frequency?

The first question, although complex, can probably be answered in large part as follows. A visual scene is generally made up of opaque objects, the objects have a wide range of sizes and shapes, and the luminosity of individual objects is relatively constant. In other words, partially transparent objects and objects with continuously graded luminosity tend to be quite uncommon in the natural environment. Since the objects are opaque, they *occlude* other objects, or parts of objects, behind them. In this sense then, a simple model of the visual environment is a collage of occluding, constant intensity objects with a wide range of sizes. There is evidence that the distribution of the object sizes is self-similar [3, 5], which we refer to here as *scaling*. An important consequence of this observation is that such a scene will contain many sharp *amplitude edges*, with a characteristic range of spacings between them, at the boundaries between the occluding constant intensity regions. This model of the natural visual environment has been discussed by a number of authors [3, 6].

The second question then becomes whether, and how, occlusion, edges and/or scaling are the source of the characteristic dependence of the power spectrum on spatial frequency. This is the topic of this paper. This question has been investigated by a number of authors although a clear answer is still lacking. Issues revolve around the relative importance of scaling and edges, and the range of spatial frequencies over which power law behaviour of the power spectrum is present [3, 6, 7, 8]. We shed further light on this question by simulation using a simple model of natural images.

(a)
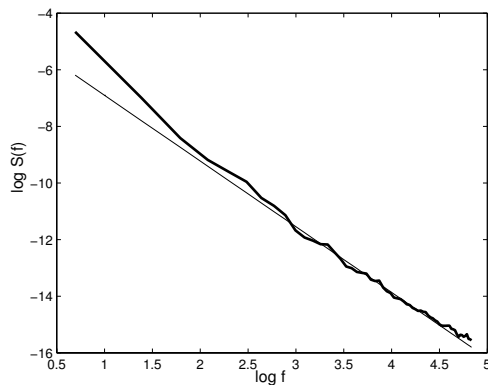
(b)

Figure 1: (a) A natural image (256×256 pixels), and (b) its circularly averaged power spectrum (thick line) and a linear fit to the high frequency portion (thin line). The slope in (b) is 2.3.

2 Methods

We use an image model that consists of disks placed at random in an image frame of $N \times N$ pixels. Although an actual scene consists of objects of a variety of shapes, this approximate model should be sufficient since we are concerned with the circularly averaged power spectrum. Our primary model consists of occluding disks, each of constant amplitude that is uniformly distributed in the range 0 to 255, and the image is self-similar so that the disk radii r_0 follow the power law probability density

$$P(r_0) = A r_0^{-\alpha}, \tag{1}$$

where we refer to α as the *scaling exponent*. Disks are added sequentially to the image such that an added disk occludes the pixels that it occupies, i.e. occludes the parts of disks added previously that occupy the same pixels. The sequence in which disks are added is also governed by the density $P(r)$.

In an image of unbounded extent there is a lower bound on the scaling exponent α, firstly to ensure that Eq. (1) represents a probability density ($\alpha > 1$), and secondly to prevent an over predominance of large disks (that will occlude most of the smaller disks). Note that $\alpha > 3$ is required to give $P(r)$ a finite second moment. However, all real images are of finite extent and it makes sense to impose an upper bound on the disk radii, r_{max}, that is somewhat less than half the extent of the image. It also makes sense to impose a lower bound on the disk radii, r_{min}, that is somewhat greater than the pixel size. For a finite range of disk radii there is no restriction on the exponent α, however the visual world generally contains more smaller features than larger features, i.e. $\alpha > 0$. Large values of α lead to a sharply peaked distribution and a narrow range of radii which is not an appropriate representation of the visual world. We used the range $0.5 < \alpha < 3$ in the simulations.

This image model is consistent with the general characteristics of natural images as described above, and possesses the characteristics of occlusion, edges and self-similarity. We also generated different kinds of images in order to study the effect of each of these characteristics.

In order to study the effect of occlusion, we generated *nonoccluding* images in which the addition of each disk corresponds to *addition* of the amplitude of the disk to the current amplitudes of the pixels that the disk occupies. This corresponds to the superposition of partially transparent objects. The amplitude of nonoccluding images is linearly rescaled to the range 0-255.

To study the effect of edges (or absence of edges) we generated images that do not have edges. The images are made up of nonoccluding disks whose amplitude $l(r)$ is a function of radius r from the center of the disk and is given by

$$
\begin{aligned}
l(r) &= L, & 0 < r < r_0 - a \\
&= \frac{L}{2}\left(1 + \cos\left(\frac{\pi(r - r_0 + a)}{a}\right)\right), & r_0 - a \leq r < r_0
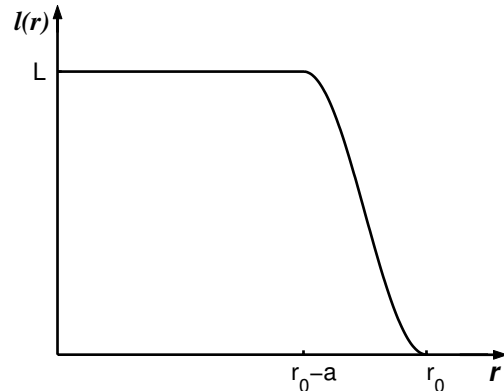\end{aligned}
\tag{2}
$$

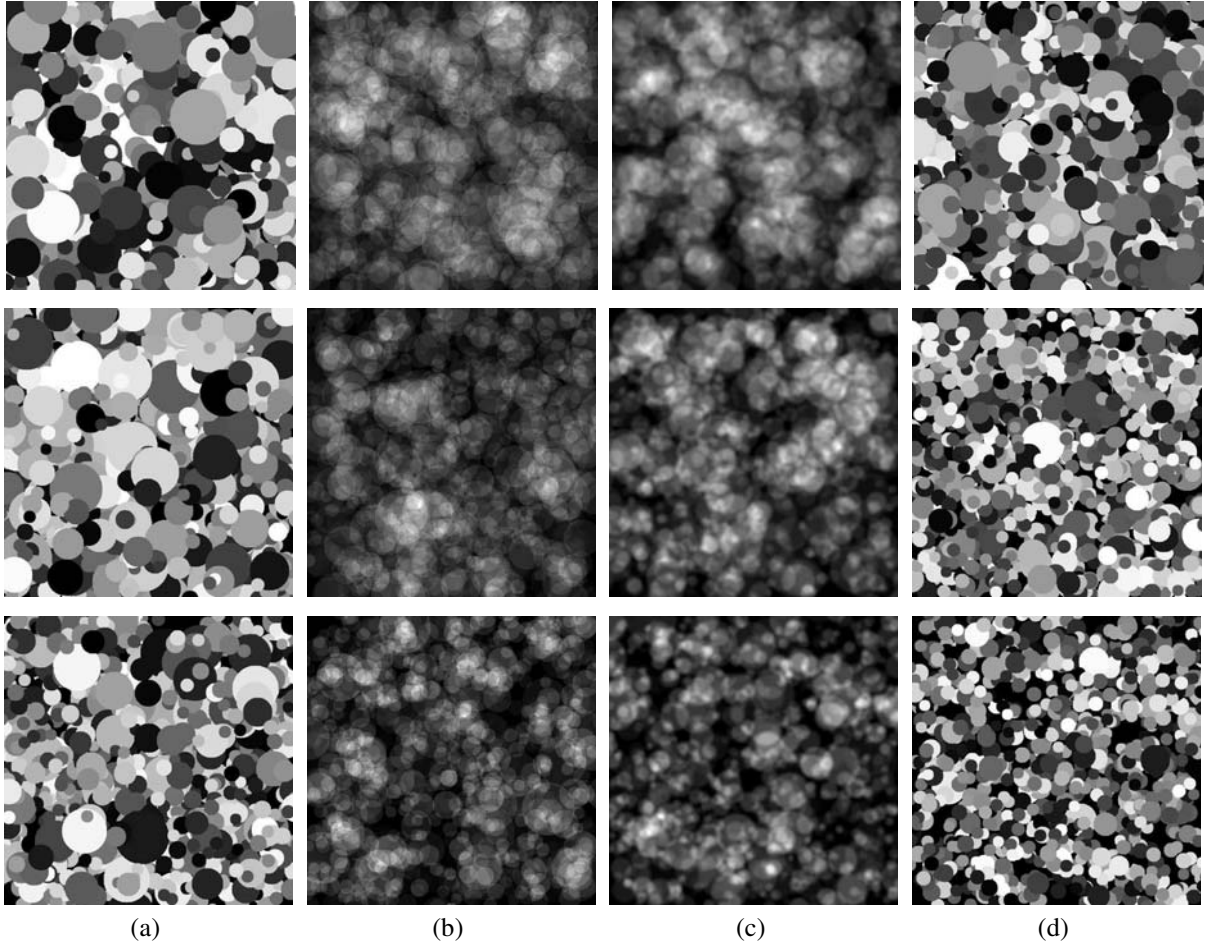Figure 2: Amplitude function for objects in images without edges.

Figure 3: Model self-similar (a) occluding (b) nonoccluding images with edges, and (c) images without edges, for $\alpha = 1$ (top row), $\alpha = 2$ (middle row), and $\alpha = 3$ (bottom row). (d) Non self-similar images with an exponential size distribution for $d = 10$ (top), $d = 5$ (middle) and $d = 3$ (bottom).

where L is the amplitude at the center of the disk, and a is a constant for which we use $a = r_{min}$. The function $l(r)$ is shown in Fig. 2. The form of Eq. (2) is chosen so that its high frequency spectral content is approximately independent of r_0. Images generated with nonoccluding disks with amplitude $l(r)$ do not have edges. Note that L is uniformly distributed and that r_0 is distributed as Eq. (1).

To study the effect of self-similarity we also generated images that consist of occluding, constant amplitude disks, but whose radii are distributed as

$$P(r_0) = Ae^{r_0/d} \qquad (3)$$

where d is a correlation length. Images generated using the distribution Eq. (3) are not self-similar. The ensemble averaged, circularly averaged power spectrum $S(f)$ is defined by

$$S(f) = \langle S_i(f) \rangle_i = \langle \langle |F_i(f, \psi)|^2 \rangle_\psi \rangle_i \qquad (4)$$

where $F_i(f, \psi)$ is the Fourier transform (spectrum) of the i-th image in the ensemble, (f, ψ) are cylindrical polar coordinates in Fourier (spatial frequency) space, and $\langle \rangle_\psi$ and $\langle \rangle_i$ denote averaging over ψ and i, respectively. Individual circularly averaged power spectra were calculated using the FFT and averaging over shells in Fourier space of thickness equal to two sample spacings.
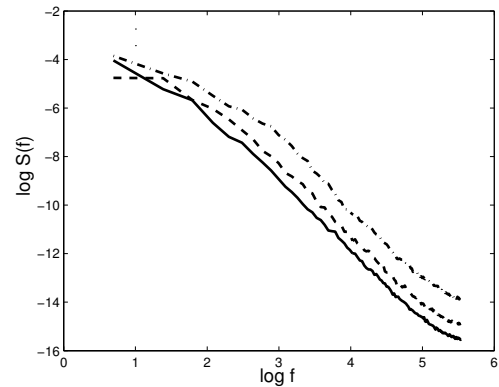

Figure 4: Circularly averaged power spectra for individual model occluding images for $\alpha = 1$ ($-$), $\alpha = 2$ ($--$) and $\alpha = 3$ ($-\cdot-\cdot$).
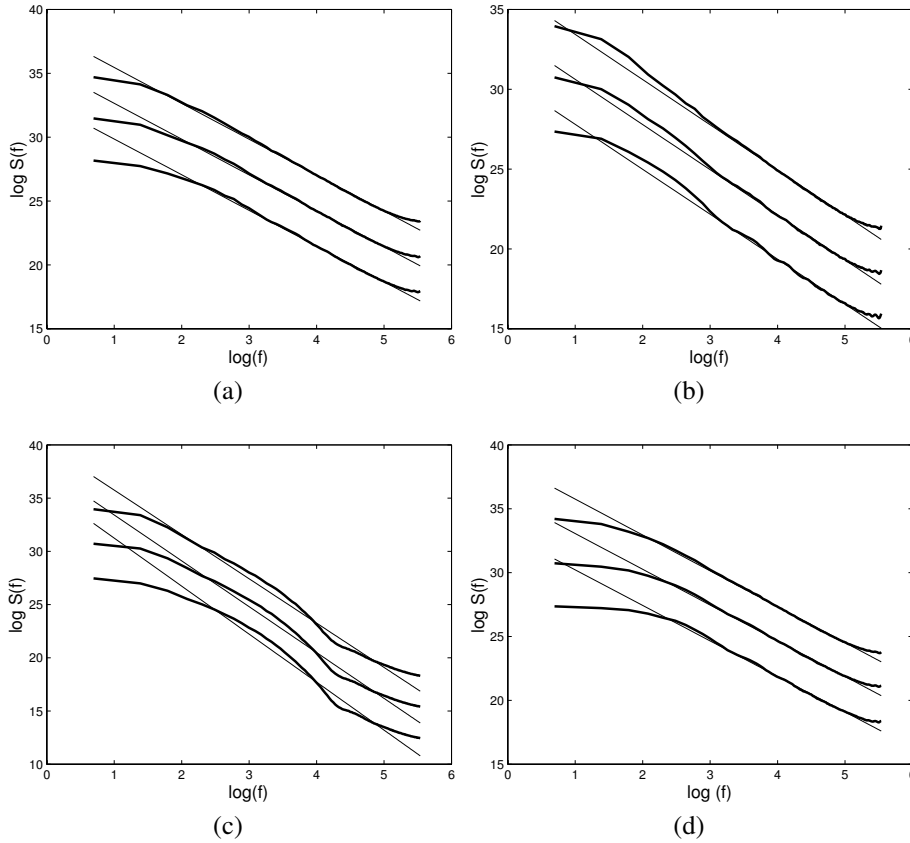
Figure 5: Ensemble averaged power spectra (thick lines) for (a) occluding and (b) nonoccluding self-similar images with edges, and (c) self-similar images without edges for $\alpha = 1$ (top line), $\alpha = 2$ (middle line) and $\alpha = 3$ (bottom line). Ensemble averaged power spectra for images with an exponential size distribution are shown in (d) for $d = 10$ (top line), $d = 5$ (middle line) and $d = 3$ (bottom line). Plots have been shifted vertically to eliminate overlap. Linear fits to the high frequency portions are shown by the thin lines.

The power spectra were fitted to the power law dependence

$$S(f) = Bf^{-\gamma} \qquad (5)$$

by fitting a regression line to the log-log power spectrum data in the range $0.1f_{max} < f < 0.7f_{max}$, where f_{max} is the maximum spatial frequency for which the spectrum is calculated. The exponent γ was calculated from the slope of the regression line.

3 Results

Model occluding images were generated as described above with $N = 512$, $r_{min} = 10$ and $r_{max} = 50$ pixels, 1000 disks per image, and values of α spaced by 0.5. Example images are shown in Fig. 3(a), in which the effects of occlusion and the scaling exponent can be clearly seen. Circularly averaged power spectra of individual images, $S_i(f)$, were calculated as described above and examples plotted on a log-log scale are shown in Fig. 4 for occluding images with different values of α. An immediate observation is that the circularly averaged power spectra are approximately

linear over much of the spatial frequency range, and are rather similar for different values of α although there are some differences at small spatial frequencies. The individual power spectra were averaged over 20 images and the exponent γ calculated as described above. The ensemble averaged power spectra and linear fits to the high frequency portions are shown in Fig. 5(a). Inspection of the figure shows linear behaviour over most of the spatial frequency range, but some deviation from linearity at low frequencies for larger values of α. There is evidence of some oscillations at high spatial frequencies for larger values of α which is due to the distribution of disk radii being quite sharply peaked and the spectrum approaching that of a single disk. The variation of γ versus α is shown in Fig. 6. The most obvious feature is that there is little variation of γ with α.

To examine the effect of occlusion, nonoccluding images were generated as described above and example images are shown in Fig. 3(b). Ensemble averaged power spectra $S(f)$ were calculated and are shown in Fig. 5(b). The power spectra are similar to those for occluding images but linearity is evident over a smaller
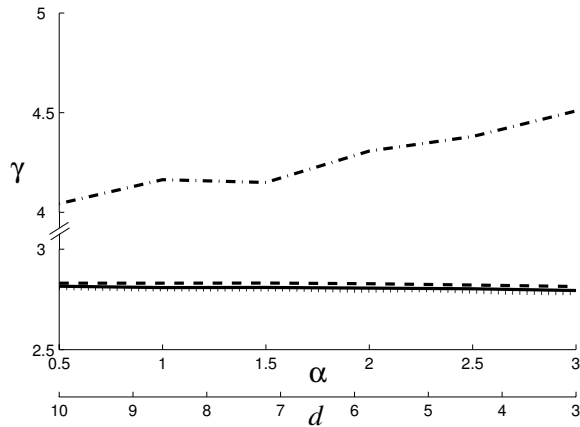
Figure 6: Variation of the exponent γ with the scaling exponent α for occluding ($-$) and nonoccluding ($--$) images with edges, and images without edges ($-\cdot-\cdot$), and with the correlation length d for occluding images with edges ($\cdots$).

range of high spatial frequencies. The high frequency slopes are plotted in Fig. 6 where they are seen to be independent of α and practically identical to those for occluding images.

To examine the effect of edges, images without edges were generated using nonoccluding disks with an amplitude profile give by Eq. (2) as described in Section 2. Examples of such images for three different values of α are shown in Fig. 3(c). Ensemble averaged spectra were calculated and are shown in Fig. 5(c). The spectra are seen to be not particularly linear over any range of spatial frequencies. The average slopes over the high spatial frequency range are much larger than for the images with edges, being in the range $4.0 - 4.6$. The slopes versus α are shown in Fig. 6.

To examine the effect of scaling, images that are not self-similar were generated consisting of occluding, constant amplitude disks with radii following the exponential distribution Eq. (3) with correlation length d varying between 3 and 10 pixels. Example images are shown in Fig. 3(d). Ensemble averaged power spectra were calculated and are shown in Fig. 5(d). The spectra are approximately linear but over a slightly smaller range of spatial frequencies than for self-similar images. The range of spatial frequencies over which the spectrum is linear somewhat reduced for smaller correlation length and there is more evidence of the spectrum of a single disk. The slopes of the linear portions are plotted versus d in Fig. 6 and are seen to be relatively independent of d and of similar value to the case of self-similar images.

4 Conclusions

A detailed study of the effects of various image characteristics on the dependence of the power spectrum on spatial frequency has been presented using a simple model of natural images. For the reference model of self-similar images containing constant amplitude occluding objects, i.e. with edges, the log power spectra are linear over a wide range of spatial frequencies with a slope of ~ 2.8. There is more deviation from linearity for a small range of object sizes (large α) since the transform approaches that of a single object. If the image consists of nonoccluding objects, there is a deviation from linearity for smaller spatial frequencies although the effect is not large. The reasons for this are not clear. Nonoccluding objects lead to more edges of smaller amplitude. This effect will increase as more objects (disks) are included, and for a very large number of objects the image will be more uniform and the effect of edges is expected to be supressed. Images that do not contain edges give log power spectra that do not vary linear with spatial frequency, however this could be due to the objects we used not being not self-similar (although their sizes are self-similar). The significant effect of the absence of edges is that the power spectra falls off more rapidly with spatial frequency with $\gamma \approx 4$. This is due to the spectra of the individual objects falling off more rapidly with spatial frequency. The log power spectra of images containing objects with an exponential size distribution are approximately linear with spatial frequency, although not over quite as wide a range as for those containing objects with a power law size distribution.

There has been some controversy concerning the relative importance of occlusion, edges and self-similarity as contributing factors to the ensemble average power spectra of natural images exhibit power law behaviour over a wide range of spatial frequencies. The results of the detailed study presented here indicate that the presence of edges is necessary for the observed behaviour, and that self-similarity, although not essential, extends the range of spatial frequencies over which power law behaviour is present. Occlusion generally tends to accentuate the presence of edges.

The value of 2.8 we obtain for γ is somewhat larger than is typically seen in natural images although this is probably due to the particular objects (disks) that we have used. We note that the approximate analysis of Ruderman [3] suggests a rather strong dependence of γ on α, although the results presented here do not bear this out for this image model.

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(12):2379–2394, December 1987.

[2] D. J. Tolhurst, Y. Tadmor, and T. Chao. Amplitude spectra of natural images. *Ophthalm. Physiol. Opt.*, 12(2):229–232, April 1992.

[3] D. L. Ruderman. Origins of scaling in natural images. *Vision Res.*, 37(23):3385–3398, December 1997.

[4] D. J. Field and N. Brady. Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Res.*, 37(23):3367–3383, December 1997.

[5] D. L. Ruderman and W. Bialek. Statistics of natural images: scaling in the woods. *Phys. Rev. Lett.*, 73(6):814–817, August 1994.

[6] R. M. Balboa, C. W. Tyles, and N. M. Grzywacz. Occlusions contribute to scaling in natural images. *Vision Res.*, 41(7):955–964, March 2001.

[7] D. L. Ruderman. Letter to the editor. *Vision Res.*, 42(25):2799–2801, November 2002.

[8] N. M. Grzywacz, R. M. Balboa, and C. W. Tyler. Response to letter to the edditor. *Vision Res.*, 42(25):2803–2805, November 2002.