PERGAMON

Vision Research 42 (2002) 2799–2801

Vision Research

www.elsevier.com/locate/visres

Letter to the editor

Dear Editor,

In a recent paper Balboa, Tyler, and Grzywacz (2001) report findings which contradict elements of my work on the origins of scaling in natural images (Ruderman, 1997). Here I refute their claim and in addition show that the theoretical approximations presented in their Appendix are not appropriate for ensembles of natural images.

In a self-similar image ensemble any ensemble averaged statistic has a spatial dependence which is pure power-law. A change in scale alters the coefficient in front of this power-law term but does not change its functional form. One such statistic is the ensemble power spectrum which has been shown to obey a power-law for over 2.5 decades in spatial frequency (Ruderman & Bialek, 1994; van der Schaaf & van Hateren, 1996). Statistics of higher-order, such as entire histograms of image properties, have also been shown to scale (Lee, Huang, & Mumford, 2001; Ruderman & Bialek, 1994). In my original article I proposed occluding objects with sharp luminance edges whose size distribution is power-law as the source of self-similarity in natural image ensembles. I stated that such an image ensemble must have a correlation function whose spatial dependence is a power law (and may contain an added constant): $C(x) \sim x^{-\eta}$. As a counter-example I showed an image of occluding squares whose size distribution gives an exponential ensemble correlation function, $C(x) \sim e^{-x/a}$, and is thus not a self-similar ensemble.

Balboa et al. disagree that power-law spectra can arise only from power-law correlations and conclude that almost any image composed of occluding objects—regardless of their size distribution—will have a power-law spectrum. Central to Balboa et al.'s argument is the demonstration in their first figure that my ensemble of occlusion-based images with exponential correlation has a power-law spectrum, contrary to my previous statement. They show the spectrum of a single image drawn from this distribution. The spectrum in their figure does show a generally power-law trend, though their data are limited only to high frequencies. While their 512×512 pixel image spans 2.5 decades in spatial frequency, they plot the spectrum over only the top 1.7 decades. As Balboa et al. themselves point out, one expects this spectrum on theoretical grounds to behave as $[1 + (2\pi f a)^2]^{-3/2}$, with f the spatial frequency in cycles per pixel and a the correlation length. For a of any reasonable size, like 10 pixels in their example, this spectrum will not be power-law at the lower frequencies.

To demonstrate that this spectrum is poorly approximated by a power-law I created an ensemble of 50 512×512 pixel images using the same object size distribution as their image, although employing circles rather than squares since the circle ensemble's correlation function is more exactly exponential. A sample image is shown in Fig. 1. Fig. 2 contains the measured ensemble power spectrum over the *full* range of spatial frequencies. At the lower half of the logarithmic frequency range the measured spectrum deviates greatly from power-law (dashed line). The solid line in the graph represents the *parameter free* theoretical prediction of the spectrum, which fits the data well and demonstrates that the spectrum is not power-law.

Another difficulty in their treatment is found in their Appendix A.1. Its Eq. (4) breaks the spectrum of 1D cuts through images of luminance edges into the product of two parts: the spectrum of a 1D step edge ($1/\omega^2$) and a factor which depends on the spacing of the edges. Balboa et al. claim that for natural images this second factor is proportional to ω^2 at low frequencies and approximately constant at high frequencies. As an example, a Poisson distribution of edges with average inter-edge distance a makes this factor $(\omega a)^2/[1 + (\omega a)^2]$, which behaves as Balboa et al. describe. They claim that in total the 1D spectrum should be at at low frequencies and fall as $1/\omega^2$ at high frequencies. However, it can be shown theoretically that an image ensemble such as natural images with 2D spectrum $1/\omega^{2-\eta}$ has 1D cuts with spectrum $1/\omega^{1-\eta}$. Thus the approximation presented by Balboa et al. does not match the spectral behavior seen in the natural image ensembles they model. To achieve a power-law spectrum the distribution of object edge intervals must itself be power-law over the relevant frequencies; it cannot have a length scale as does a Poisson process or as in Balboa et al.'s approximation. The $1/\omega^2$ spectrum of edges does not dominate in natural images. Instead the overall spectrum derives from this edge spectrum in combination with the statistics of edge occurrences.

Balboa et al. conclude that "the power spectrum becomes a weak statistical indicator of the structure in the visual world". This statement is based on the spectra of single images limited to their highest frequencies. It is not surprising that individual images of general origin can have spectra with approximately power-law

PII: S0042-6989(02)00357-7

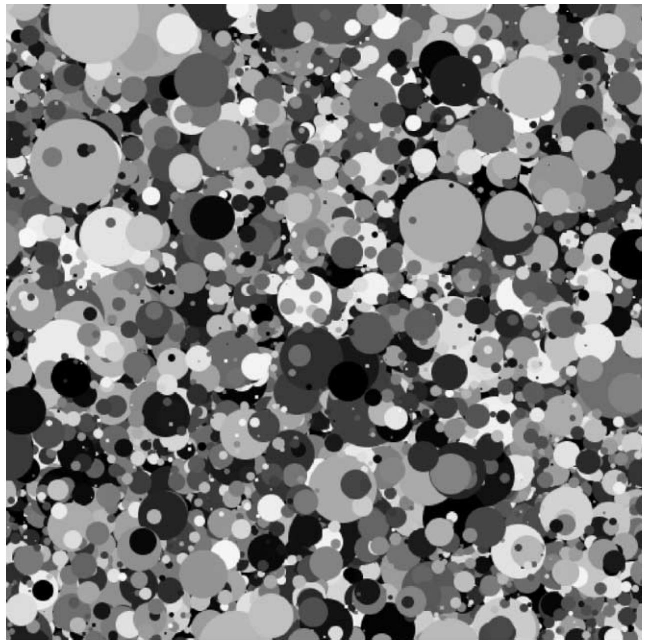

Fig. 1. Example image from an ensemble made of overlapping circles drawn from an exponential size distribution. The ensemble correlation function is approximately exponential with length scale $a = 9.3$.

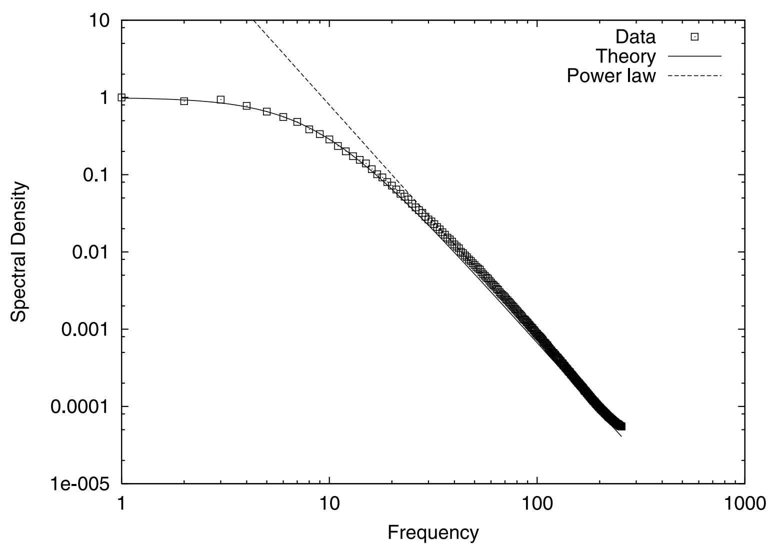

Fig. 2. Power spectrum of image ensemble (squares) showing marked deviation from power-law at low frequency (log–log axes, frequency in cycles/image). Theoretical prediction based on an exponential correlation function matches well (solid line). Dashed line is a fit to the power-law high frequency asymptote, f^{-3}, and deviates significantly from the data for half the frequency range.

behavior over a limited frequency range. But the spectra of natural images which have demonstrated scaling are from *ensembles* of many images and over their entire frequency range, providing a robust set of repeatable measurements. The spectra of these image ensembles restrict which image models can obey them, and they reflect the structure of the natural environment. Discovering which properties give rise to these statistical regularities is of great importance not only for understanding the design of visual systems (Simoncelli & Olshausen, 2001) but for exploring the structure of the visual environment as well.

References

Balboa, R. M., Tyler, C. W., & Grzywacz, N. M. (2001). Occlusions contribute to scaling in natural images. *Vision Research, 41*, 955–964.

Lee, A. B., Huang, J., & Mumford, D. (2001). Random-collage model for natural images. *International Journal of Computer Vision*.

Ruderman, D. L. (1997). Origins of scaling in natural images. *Vision Research, 37*, 3385–3398.

Ruderman, D. L., & Bialek, W. (1994). Statistics of natural images: scaling in the woods. *Physical Review Letters, 73*, 814–817.

van der Schaaf, A., & van Hateren, J. (1996). Modelling the power spectra of natural images: statistics and information. *Vision Research, 28*, 2759–2770.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience, 24*, 1193–1215.

Daniel L. Ruderman[1]
13621 Oak Canyon Avenue
Sherman Oaks, CA 91423, USA
E-mail Address: dan_ruderman@berlex.com

[1] Present address: Berlex Biosciences, 2600 Hilltop Drive, Richmond, CA 94804, USA.