Proceedings of the IEEE Workshop on Statistical and Computational Theories of Vision, Fort Collins, CO, June, 1999. (Published on Web).

An Occlusion Model Generating Scale-Invariant Images

Ann B. Lee[1] and David Mumford[2]

Abstract

We present a model for scale invariance of natural images based on the ideas of images as collages of statistically independent objects. The model takes occlusions into account, and produces images that show translational invariance, and approximate scale invariance under block averaging and median filtering. We compare the statistics of the simulated images with data from natural scenes, and find good agreement for short-range and middle-range statistics. Furthermore, we discuss the implications of the model on a 3D description of the world.

1. Dept. of Physics, Brown Univ., Providence, RI 02912
2. Dept. of Applied Mathematics, Brown Univ., Providence, RI 02912

1 Introduction

One of the most remarkable properties of natural images is an invariance to scale. Scale invariance is interesting because it distinguishes natural scenes from random noise and many man-made images. Scale invariance is also a very robust property of natural images; it depends little on calibration [6], and has been observed in images from very different environments, e.g. scenes from the woods [5], or scenes of mountains, cities, and landscapes [1]. A good model for images which captures the invariance properties of real scenes could be useful for both image compression purposes and for the study of sensory processing in biology.

The most well-known evidence for scale invariance of natural images is perhaps the ensemble power spectrum [2, 5]; it behaves as $1/k^{2-\eta}$ where k is the modulus of the spatial frequency and η is a small constant. The power-law form indicates that the second-order statistics scale, but recently, there has also been evidence of higher-order scaling in natural images. Ruderman [5] and Zhu [8], for example, have shown that the average response histograms of many local filters are the same before and after block averaging images. Such "histogram scaling" indicates that *full scale invariance* $I(\mathbf{x}) \sim I(\sigma\mathbf{x})$ might exist approximately in natural images.

It is not fully understood, however, why natural scenes scale. Most likely it is caused by a combination of (a) objects in the world having different sizes, and (b) objects occurring at arbitrary distances from the viewer. Natural scenes are extremely rich in detail, and it is reasonable to assume that properties (a) and (b) give rise to "projected" objects that span many angular scales.

The standard Gaussian model can produce fully scale invariant images, but the images all look like clouds with no clear objects and borders. There are, however, models for scale invariance which are closer to real images and based on the above notion of images as collages of (independent) "objects". In [4], Mumford proposes a representation of images as sums of elementary objectlets, e.g. wavelets; the objects include independent patches, shadows, textons, etc. This model, however, ignores occlusions. Other models take occlusions into account, for example Chi's "ground plane model" of the 3D world which produces approximately scale invariant images under perspective projections with occlusions [1], or Ruderman's 2D model which shows translational invariance and scaling in the second-order statistics [6].

In this paper, we present a stochastic model for scale invariance based on Ruderman's ideas of randomly placing independent objects in 2D. The model takes occlusions into account, and generate images which show both translational invariance as in [6], and are approximately scale invariant as in [1].

The organization is as follows: In the first part of the paper we lay down the theoretical framework for the model. In the second part of the paper, we generate synthetic images according to the model, and compare the statistics of the simulated images with data from natural images [10]. Finally, in the last section, we discuss the implications of the model on a 3D description of the world.

2 Theoretical Predictions

2.1 Poisson Model with Full Scale Invariance. The $1/r^3$ Law of Sizes.

Our goal is to build a model for scale invariance of natural images. As in [6] and [1], we base the model on the notion that images can be broken down into correlated regions called "objects", with different "objects" being *statistically independent.* We also assume *translational and rotational invariance.* "Objects" can basically occur anywhere in an image, and with any orientation.

The main argument in this section is that *the condition of full scale invariance*

$$P\{I(x, y)\} = P\{I(\sigma x, \sigma y)\} , \tag{1}$$

sets a constraint on the distribution of object sizes. Below we prove this for the general case where opaque objects occlude each other. It is, however, trivial to apply the same reasoning to transparent objects.

For simplicity, we choose circular, uniformly colored objects. Each object is assigned a radius r according to some size distribution $f(r)$ ($r_{\min} < r < r_{\max}$) and a random grey level a from some intensity distribution $p(a)$. Now imagine the situation where the colored discs are randomly "raining down" on a plane. The discs reach the plane in a specific order, and "creates" images with occlusions when viewed from above. We define t as the time a disc hits the plane [1] and let (x, y) denote the position. Note that, if we wait long enough, randomly sampled images belong to a stationary probability distribution $P[I(\mathbf{x})]$.

Mathematically, the above construction (of placing colored objects on a plane) defines a Poisson process

$$\Pi = \{(x, y, r, t, a)\} \tag{2}$$

in 5-dimensional space $\mathbf{R}^2 \times (r_{\min}, r_{\max}) \times (0, \infty) \times (a_{\min}, a_{\max})$. The statistics of Π, and thus the statistics of the generated images $I(\mathbf{x})$, are fully determined by the measure

$$d\mu = \lambda(x, y, r, t, a)\; dx\, dy\, dr\, dt\, da \tag{3}$$

on this space. In our case [2], Eq. 3 reduces to

$$d\mu = f(r)\, p(a)\; dx\, dy\, dr\, dt\, da\ . \tag{4}$$

Now if $I(x, y)$ is a sample from Π, then the rescaled image $I(\sigma x, \sigma y)$ is a sample from a Poisson process

$$\Pi_\sigma = \{(x, y, r, t, a)\} \tag{5}$$

in $\mathbf{R}^2 \times (r_{\min}/\sigma, r_{\max}/\sigma) \times (0, \infty) \times (a_{\min}, a_{\max})$ with measure

$$d\mu_\sigma = \sigma^3 f(\sigma r)\, p(a)\; dx\, dy\, dr\, dt\, da\ . \tag{6}$$

[1] Alternatively, we can tag the objects with a discrete order parameter $k = 1, 2, 3, \ldots$, since the time between two events [object reaching the ground] makes no difference in 2D.

[2] assume translational invariance and separability of the variables x, y, r, t and a

Let us for the time being ignore the short- and long-distance cutoffs on the object sizes. Assume, for example, that $r_{\min} \to 0$ and $r_{\max} \to \infty$. Full scale invariance $P\{I(x,y)\} = P\{I(\sigma x, \sigma y)\}$ then occurs, if and only if

$$d\mu = d\mu_\sigma \tag{7}$$

i.e.

$$f(r) = \sigma^3 f(\sigma r) \Longrightarrow f(r) \propto r^{-3} \, . \tag{8}$$

The $1/r^3$ "law of sizes" has previously been derived for scale-invariant images without occlusions, see for example [1]. Above, we show that the condition for full scale invariance also applies to images *with occlusions* as long as the order of the opaque objects does not depend on the variables x, y, and r.

We would also like to point out that the density parameter C in the density function

$$\lambda(x,y,r,t) = \frac{C \cdot dx \cdot dy \cdot dr \cdot dt}{r^3} \, , \tag{9}$$

has no real meaning in 2D; the value of C does not affect the image statistics. We can absorb the parameter into t, because the images are created by a process which looks at all $t \leq 0$ (assume that $t = 0$ represents the time the image is sampled). Note that placing the objects front to back until the background is filled will give an exact sample from the model (cf. construction of synthetic images in Sec. 3.1).

2.2 Statistics and Scaling for Occlusion Model with $r_{\min}$ and $r_{\max}$ finite

Divergences and the Need for Cut-Offs.
All scale invariant probability distributions have divergences for both short wavelength (UV) and long wavelength (IR) fluctuations [3]. This may be best illustrated by looking at the expected power of stationary, scale-, and rotationally invariant images. These types of images have a covariance of the form $-\log|\mathbf{x}-\mathbf{y}|$ and a power spectrum of the form $1/(\xi^2+\eta^2)$ [4]. Hence, if $A(r_1, r_2)$ is an annulus in the (ξ,η)-plane with inner radius r_1 and the outer radius r_2, then

$$\int\int_{A(r_1,r_2)} E[|\hat{I}(\xi,\eta)|^2] d\xi \, d\eta = \beta \log\left(\frac{r_2}{r_1}\right) \, , \tag{11}$$

where β is a constant. The amount of power in the band depends only on the ratio (r_2/r_1) of the high and low frequencies, not on the frequencies themselves. This indicates that there is no definite angular scale in the images. The problem, however, is that the power

[3] The solution according to Mumford [4] is to consider images $I(x,y)$ not as functions but *distributions modulo constants*. Loosely speaking, we assume that for all test functions ϕ with mean zero, the averages

$$\int\int I(x,y)\,\phi(x,y)\,dx\,dy \tag{10}$$

are well-defined.

[4] For the standard 2D Gaussian model, which is uniquely determined by its mean and covariance, the above form of the covariance or power spectrum is not only a necessary, but also a sufficient, condition for full scale invariance

blows up at both ends, when $r_2 \to \infty$ and $r_1 \to 0$. We call these blow-ups the ultraviolet (UV) and infrared (IR) divergences, respectively.

In the occlusion model described in Sec. 2.1, the IR and UV divergences occur when the object size limits $r_{\min} \to 0$ and $r_{\max} \to \infty$. As $r_{\min} \to 0$, the image is totally covered by microscopic objects. In fact, for each r_0, the proportion of area covered by objects of size $< r_0$ goes to 1. On the other hand, as $r_{\max} \to \infty$, the probability of an image containing only one object tends to 1. Almost all image samples will then have uniform intensities.

Obviously, we need to impose finite bounds $r_{\min}$ and $r_{\max}$ on the object sizes. A natural consequence of finite bounds, however, is that the the population of objects changes under scaling. For example, if we scale an image down by a factor 2, we may introduce new small objects with sizes $r \in [r_{\min}/2, r_{\min}]$. Large objects with sizes $r \in [r_{\max}/2, r_{\max}]$ will also be missing.

Below, we investigate the properties of the occlusion model described in Sec.2.1, for the case when both $r_{\min}$ and $r_{\max}$ are finite. We focus on the following questions:

- What is the two-point correlation function for the images, and how does it depend on the cutoffs $r_{\min}$ and $r_{\max}$?
- How well do the images scale?

Covariance Statistics. Predictions.

We start by deriving an expression for the correlation or covariance between two points $\mathbf{x}_1$ and $\mathbf{x}_2$. For a translational and rotational invariant distribution, it can be written as a function of the separation distance $x = |\mathbf{x}_1 - \mathbf{x}_2|$. Schematically, we write

$$C(x) =< (I(0) - \mu)(I(x) - \mu) > \tag{12}$$

where μ represents the mean of I, and the brackets imply an average over angles, a shift over positions, and an ensemble average over different images $I(\mathbf{x})$.

In our model, the images are composed of *independent, uniformly colored* objects. This means that (a) $C(x) = 0$ for points belonging to different objects, and (b) $C(x)$ is constant, and equal to the variance of the intensity distribution, for points belonging to the same object. From (a) and (b), we obtain a direct relation between the correlation function $C(x)$ and the probability $P_{\text{same}}(x)$ of two points being in the same object. We have

$$C(x) = C_0 \, P_{\text{same}}(x) \, , \tag{13}$$

where C_0 is the constant correlation within objects.

The key step is to connect the probability function $P_{\text{same}}(x)$ above to the density function of the Poisson model. Ruderman has shown [5] that (assume stationarity)

$$P_{\text{same}}(x) = \frac{p_2(x)}{p_1(x) + p_2(x)} \, , \tag{14}$$

where $p_2(x)$ is the probability that the last added object, which is not occluded by any other objects, contains both points $\mathbf{x}_1$ and $\mathbf{x}_2$, and $p_1(x)$ is the probability that the object contains exactly one of the two points. Furthermore, Ruderman has derived an expression for the conditional probability

$$g(x, r) = Pr\{\mathbf{x}_2 \in A \mid \mathbf{x}_1 \in A;\ ||\mathbf{x}_1 - \mathbf{x}_2|| = x\}\ , \tag{15}$$

for a circle A with radius r. In the dimensionless quantity $\xi = x/r$, the function has the form [5]

$$\tilde{g}(0 \leq \xi \leq 2) = \frac{2}{\pi}\left[\arccos\left(\frac{\xi}{2}\right) - \frac{\xi}{2}\sqrt{1 - \left(\frac{\xi}{2}\right)^2}\right]\ . \tag{16}$$

($\tilde{g}(\xi) = 0$, for $\xi > 2$)

In our model, the objects sizes are distributed according to $1/r^3$ where $r_{\min} \leq r \leq r_{\max}$ [5]. We then have

$$\begin{aligned} p_1(x) &= 2\int_{r_{\min}}^{r_{\max}} [1 - g(x, r)]\, p(r)\, dr \\ p_2(x) &= \int_{r_{\min}}^{r_{\max}} g(x, r)\, p(r)\, dr\ , \end{aligned} \tag{17}$$

where

$$p(r)\, dr \approx \frac{dr}{r\, \ln\left(\frac{r_{\max}}{r_{\min}}\right)} \tag{18}$$

is the probability that a given point in the image belongs to an object with a radius in the interval $[r, r + dr]$.

By inserting the above equations into Eq. 13, we get

$$C(x) = \frac{C_0\, B(x)}{2\, \ln\left(\frac{r_{\max}}{r_{\min}}\right) - B(x)}\ , \tag{19}$$

where

$$B(x) = \int_{\frac{r_{\min}}{x}}^{\frac{r_{\max}}{x}} \tilde{g}\left(\frac{1}{u}\right)\, \frac{du}{u}\ . \tag{20}$$

and $\tilde{g}(\xi)$ is given by Eq. 16.

To simplify the integral above, we approximate the function $\tilde{g}(\xi)$ in Eq. 16 with a third-order polynomial. The best fit gives ($0 \leq \xi \leq 2$)

$$\tilde{g}(\xi) \approx \tilde{g}_{\text{poly}}(\xi) = a_3\, \xi^3 + a_2\, \xi^2 + a_1\, \xi + a_0 \tag{21}$$

with coefficients $a_0 \approx 1.0$, $a_1 \approx -0.61$, $a_2 \approx -0.051$, and $a_3 \approx 0.052$. In Fig. 1 we see that the "polynomial approximation" $\tilde{g}_{\text{poly}}(\xi)$ (solid line) fits the "full expression" $\tilde{g}(\xi)$ (diamonds) very well. Inserting Eq. 21 into Eq. 20 leads to an estimation of $B(x)$, and thus

[5] Ruderman's model allows infinite-sized objects, and is only well behaved for power-law size distributions $1/r^{\alpha}$ with $\alpha > 3$. These distributions, however, do not lead to higher-order scaling.

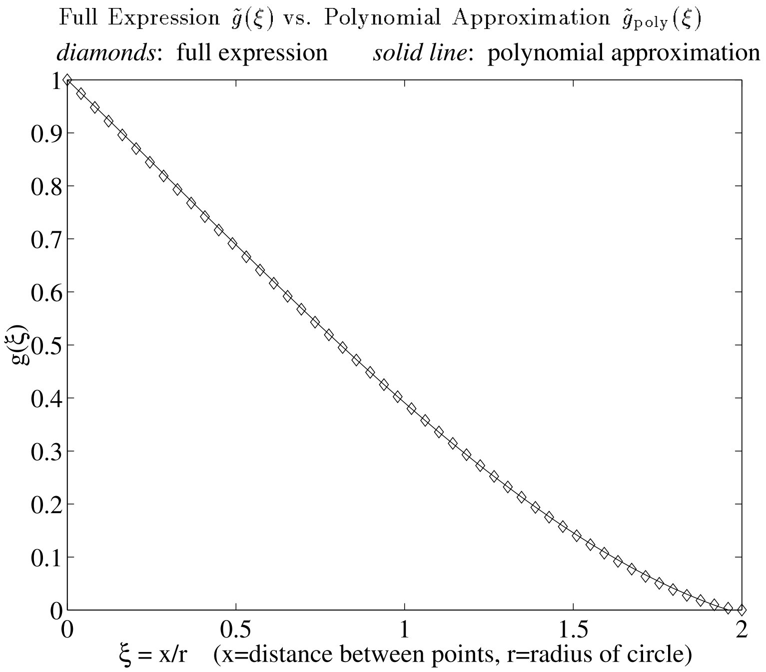

Figure 1: The conditional probability that a point is within a circle with radius r, given that another point a distance x away is within the circle: The diamonds represent values from the "full expression" $\tilde{g}(\xi)$ in Eq. 16 as a function of $\xi = x/r$. The overlapping solid line represents the "polynomial approximation" $\tilde{g}_{\mathrm{poly}}(\xi)$ (Eq. 21).

a numerical expression for the two-point correlation function $C(x)$ according to Eq. 19. For $2r_{\min} \leq x < 2r_{\max}$ [6],

$$B(x) = \frac{a_3}{3}(8 - u^3) + \frac{a_2}{2}(4 - u^2) + a_1(2 - u) + a_0 \ln\left(\frac{2}{u}\right) , \tag{23}$$

where $u = \frac{x}{r_{\max}}$.

Predictions. A Numerical Example.

From Eq. 19 and Eq. 23, we can now predict how the covariance statistics behaves for different values of $r_{\min}$ and $r_{\max}$. Fig. 2 shows $C(x)$ as a function of x for $C_0 = 1$ (i.e. $C(x) = P_{\mathrm{same}}(x)$) and two different choices of cutoffs:

In Fig. 2 (*left*), the solid line represents the correlation function $C(x)$ when $r_{\min} = 1/2$ and $r_{\max} = 2048$. For comparison, we have also fitted $C(x)$ (in the region $2 < x < 64$) to a power-law function of the form $f(x) = -A + B \cdot x^{-\eta}$; In the Fourier domain, this type of function corresponds to a power spectrum of the form $1/k^{2-\eta}$. The best fit with

[6] For $x > 2r_{\max}$, $B = 0$. For $x < 2r_{\min}$,

$$B(x) = \frac{a_3}{3}(s^3 - u^3) + \frac{a_2}{2}(s^2 - u^2) + a_1(s - u) + a_0 \ln\left(\frac{r_{\max}}{r_{\min}}\right) , \tag{22}$$

where $s = \frac{x}{r_{\min}}$ and $u = \frac{x}{r_{\max}}$.

least-square error is obtained for $f_1(x) \approx -0.29 + 1.0 \cdot x^{-0.17}$ (see dotted line). The two curves overlap, except for very small and very large values of x.

In Fig. 2 (*right*), we have $r_{\min} = 10^{-6}$ and $r_{\max} = 10^6$. The solid line represents the predicted correlation function $C(x)$, and the overlapping dotted line shows the power-law fit $f_2(x) \approx -0.59 + 0.91 \cdot x^{-0.033}$. Note that the exponent η in the power-law fit is much smaller here, where $r_{\min}$ is very small and $r_{\max}$ very large, than in the previous case.

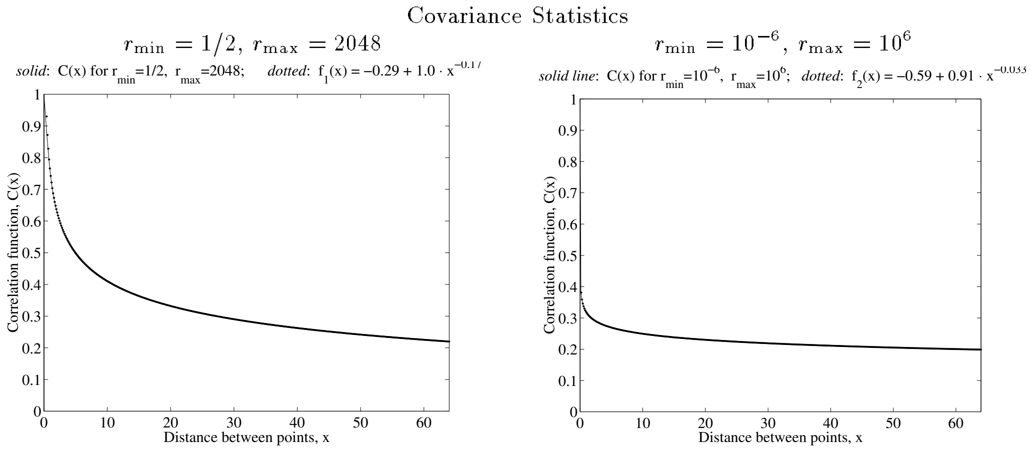

Figure 2: **Left**: The solid line shows the two-point correlation function $C(x)$ (insert Eq. 23 into Eq. 19) as a function of x for $r_{\min} = 1/2$, $r_{\max} = 2048$, and $C_0 = 1$. For comparison, we have fitted a power-law function $f_1(x) \approx -0.29 + 1.0 \cdot x^{-0.17}$ (dotted line) to $C(x)$. The two curves overlap except for very small values of x where the power-law function diverges. **Right**: The solid line shows $C(x)$ for $r_{\min} = 10^{-6}$, $r_{\max} = 10^6$, and $C_0 = 1$. The overlapping dotted line shows the power-law fit $f_2(x) \approx -0.59 + 0.91 \cdot x^{-0.033}$.

For a fully scale invariant system, we expect a power spectrum of the form $1/k^2$, i.e. $\eta = 0$ above, and a covariance with log-behavior. From Eq. 19 and Eq. 23, we see that if $r_{\min} \ll x \ll r_{\max}$, then

$$C(x) \approx \frac{-C_0}{2\,\ln\left(\frac{r_{\max}}{r_{\min}}\right)} \ln\left(\frac{x}{r_{\max}}\right) . \tag{24}$$

In this region, the correlation function is approximately scale invariant up to an additive constant

$$C^{(1)}(x) - C^{(\sigma)}(x) \approx \frac{C_0}{2\,\ln\left(\frac{r_{\max}}{r_{\min}}\right)} \ln(\sigma) . \tag{25}$$

Fig. 3 illustrates the correlation function $C^{(\sigma)}(x) = C(\sigma x)$ for six different scales: $\sigma = 1, 2, 4, 8, 16, 32$. Since the correlation function can only be expected to be scale-invariant modulo constants [4], we have shifted the curves with an additive constant so that $C^{(\sigma)}(x) = C(x)$ at $x = 1$. In the *left* graph, $r_{\min} = 1/2$ and $r_{\max} = 2048$, and in the graph to the *right*, $r_{\min} = 10^{-6}$ and $r_{\max} = 10^6$. As expected, the differences in the shifted $C^{(\sigma)}(x)$-curves are smaller in the latter case.

Before we proceed, we would also like to compare Fig. 3 to the case where the size of the smallest object is determined by the screen resolution. We assume that when an image is down-scaled, all objects with radii less than some *short-distance cutoff* r_0 disappear. Then

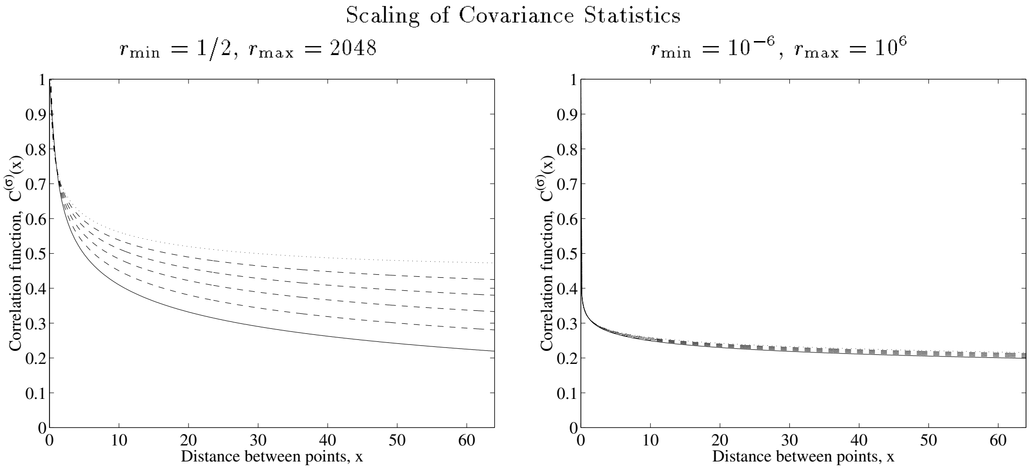

Figure 3: **Left**: The correlation function $C^{(\sigma)}(x)$ as a function of x for $r_{\min} = 1/2$ and $r_{\max} = 2048$. **Right**: $C^{(\sigma)}(x)$ for $r_{\min} = 10^{-6}$ and $r_{\max} = 10^{6}$. In both figures, six different scales are represented: $\sigma = 1$ (solid line), $\sigma = 2$ (dashed line), $\sigma = 4$ (dashed line), $\sigma = 8$ (dashed line), $\sigma = 16$ (dashed line), and $\sigma = 32$ (dotted line). The curves are shifted with an additive constant so that $C^{(\sigma)}(x) = C(x)$ at $x = 1$.

if the original images have objects with sizes $r \in [r_0, r_{max}]$, the corresponding (with a factor σ) down-scaled images will have objects with sizes $r \in [r_0, r_{max}/\sigma]$. Fig. 3 illustrates the correlation function $C^{(\sigma)}(x) = C(\sigma x)$ for six different scales: $\sigma = 1, 2, 4, 8, 16, 32$. As before, the curves are shifted with an additive constant so that $C^{(\sigma)}(x) = C(x)$ at $x = 1$. In the *left* graph, $r_0 = 1/2$ and $r_{\max} = 2048$, and in the graph to the *right*, $r_0 = 10^{-6}$ and $r_{\max} = 10^6$.

Later in Sec. 3.3, we will see that rescaling digitized images by *block averaging* is similar to the procedure in Fig. 3, and that rescaling by taking the *median* of 2×2 blocks is similar to the procedure in Fig. 4.

3 Numerical Simulations

3.1 Construction of Synthetic Images

We generate 1000 images with 256×256 pixels, by successively adding approximately circular objects to a plane. The disc radii r are distributed according to $1/r^3$, where r is between $r_{\min} = 1/2$ pixels and $r_{\max} = 8 \cdot 256 = 2048$ pixels. The construction is as follows:

First, we make an image which is four times as large, i.e. with a size of 1024×1024 pixels. Assume that the "image screen" is defined by $|x| \leq 512$ and $|y| \leq 512$. In each iteration, we pick a random position for the object center, in an "extended screen" with $|x| \leq 512 + 4 \cdot r_{\max}$ and $|y| \leq 512 + 4 \cdot r_{\max}$. The object is assigned a radius r from a $1/r^3$ size distribution with $4r_{\min} \leq r \leq 4r_{\max}$, and a random grey level a according to a double exponential distribution with zero mean and unit variance [7]. We make sure that the generated images are samples from a *stationary* probability distribution by *first*

[7] Let $f(a) = \frac{\lambda}{2} \exp(-\lambda|a|)$, where $\lambda = \sqrt{2}$

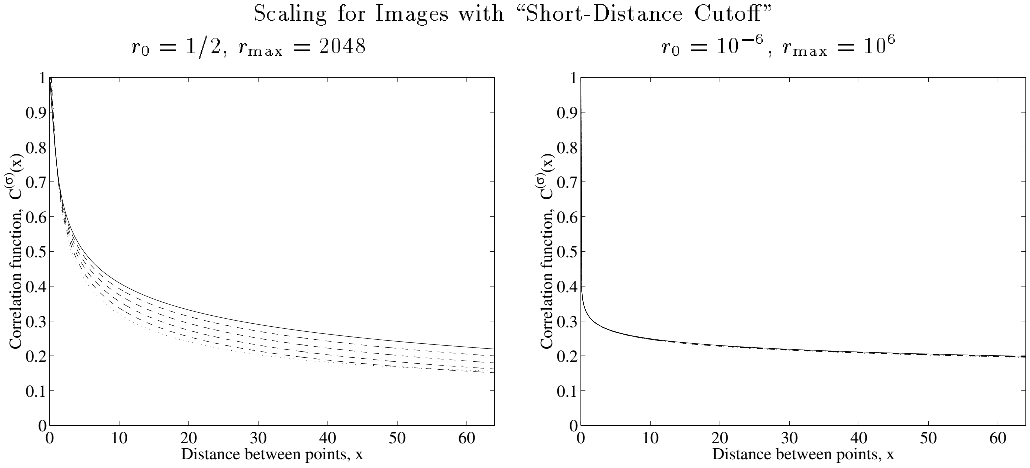

Figure 4: **Left**: The correlation function $C^{(\sigma)}(x)$ as a function of x for $r_{\max} = 2048$ and the short-distance cutoff $r_0 = 1/2$. **Right**: $C^{(\sigma)}(x)$ for $r_{\max} = 10^6$ and the short-distance cutoff $r_0 = 10^{-6}$. In both figures, six different scales are represented: $\sigma = 1$ (solid line), $\sigma = 2$ (dashed line), $\sigma = 4$ (dashed line), $\sigma = 8$ (dashed line), $\sigma = 16$ (dashed line), and $\sigma = 32$ (dotted line). The curves are shifted with an additive constant so that $C^{(\sigma)}(x) = C(x)$ at $x = 1$.

placing the *closest* object on the "screen", and then successively adding objects which are farther away until the whole "screen" is covered. Many of the later added objects are only partly visible in this occlusion-style construction.

In the next step, we scale down the generated images by a factor 4 [8]. This gives image samples that are 256×256 pixels with a subpixel resolution of 1/4 pixel unit. The disc radii are distributed according to $1/r^3$ for $r_{\min} \leq r \leq r_{\max}$, where $r_{\min} = 1/2$ and $r_{\max} = 8 \cdot 256 = 2048$.

Fig. 5 shows four samples from the final image ensemble.

3.2 Statistics of Generated Images. Comparison to Natural Images.

Below we compare the statistics of 1000 256×256 *synthetic images* constructed according to Sec. 3.1, with the statistics of about 4000 1024×1536 *natural images* taken by a digital camera [10, 9].

Single Pixel Statistics

Fig. 6 (*left*) shows the log-histogram of the "log-contrast" for natural images. The "log-contrast" $\tilde{I}$ is defined as

$$\tilde{I}[i,j] = \ln(I_{\text{d.b.}}[i,j]) - < \ln(I_{\text{d.b.}}[i,j]) > \ , \tag{26}$$

where $I_{\text{d.b.}}[i,j]$ are the intensity values provided by the image data base [9]. The average $< \cdot >$ is taken over each image separately. Note that the histogram has a highly non-

[8] Here we have taken the median of 2×2 blocks twice. Block averaging, however, gives similar results.

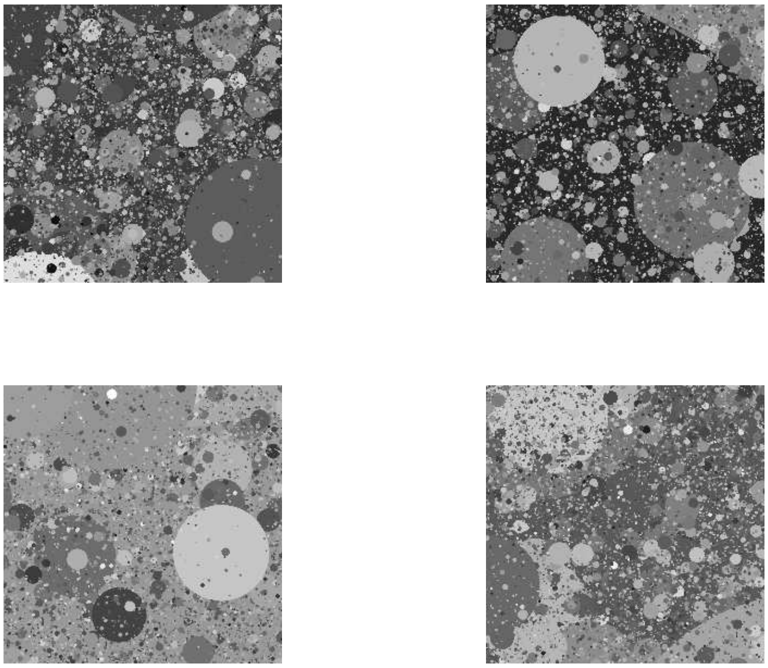

Figure 5: Example of synthetic images from the occlusion model. The images have 256×256 pixels, and are constructed from opaque discs with size distribution $f(r) \sim 1/r^3$, where $1/2 \leq r \leq 8 \cdot 256 = 2048$ pixels.

Gaussian shape with almost linear tails. The mean of the distribution is 0, and the variance is 0.79.

Fig. 6 (*right*) shows the log-histogram of the single pixel intensities $I[i,j]$ in the synthetic images. Both tails are straight, since we in the Poisson model chose the grey levels for the objects from a double-exponential distribution. The sample mean is 0, and the sample variance is 0.83.

Derivative Statistics

We now look at the marginal distribution of horizontal derivatives defined by

$$\nabla_x I[i,j] = I[i,j+1] - I[i,j] \,. \tag{27}$$

In [10], Huang shows, for natural images, that a two-parameter generalized (s, α) Laplace distribution of the form [11, 12]

$$f(x) \propto e^{-|x/s|^{\alpha}} \tag{28}$$

provides a good fit to the log-histogram of derivatives $\nabla \tilde{I}$. Fig. 7 (*left*) displays the log-histogram of $\nabla_x \tilde{I}$ for natural images; it also shows a least-square fit to a Laplacian model with $\alpha = 0.55$. Fig. 7 (*right*) plots the log-histogram of $\nabla_x I$ for images generated with the occlusion model (see dotted line). A double-exponential, i.e. a Laplacian distribution with $\alpha = 1$, gives the best least-square fit (see solid line). Note that the shape of the derivative histogram (large peak at zero, straight tails) follows directly from the assump-

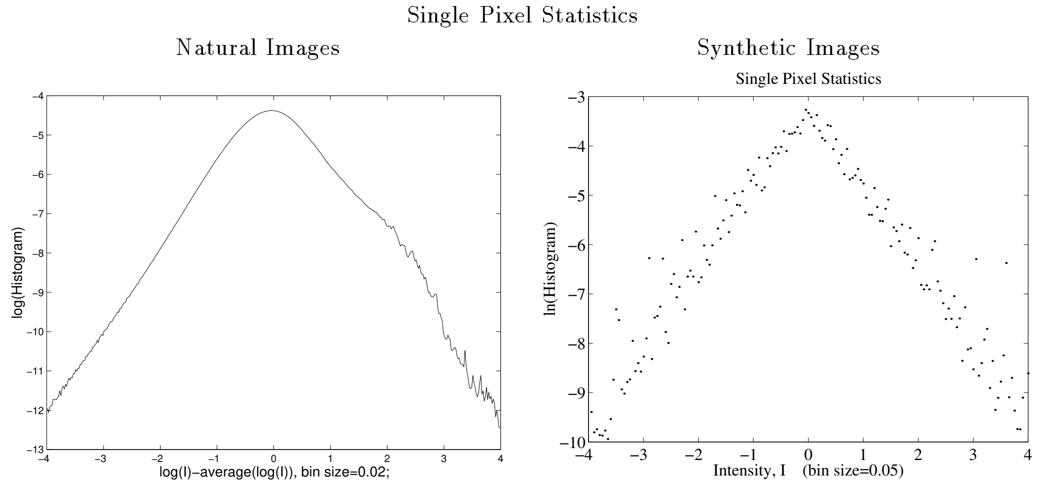

Figure 6: **Left**: Logarithms of normalized histograms of $\tilde{I}$ ("log-contrast") for natural images. The bin size is 0.02. [Courtesy of J. Huang] **Right**: Logarithms of normalized histograms of intensities I for synthetic images from the occlusion model. The bin size is 0.05. (Note: The scales on the vertical axes are different in the two figures).

tion of independent, uniformly colored objects, with a color distribution according to a double-exponential distribution.

It is, however, interesting to note that, if we filter the images by taking the average of $M \times M$ blocks, the tails become more concave. Fig. 8 shows the log-histograms of derivatives for $M = 2, 4, 8, 16$. In all four cases, the histograms fit a Laplacian distribution according to Eq. 28 very well (see solid lines). For $M = 2$ (*top left*), the best fit gives $\alpha = 0.83$. For $M = 4$ (*top right*), $\alpha = 0.78$. For $M = 8$ (*bottom left*), $\alpha = 0.69$, and finally, for $M = 16$ (*bottom right*), we get $\alpha = 0.56$.

The above results seem to indicate that the best fit to natural data is obtained by block averaging [the images generated by the Poisson process] a few times. Alternatively, we can use the above model with suitable subpixel circles, and then make it into a lattice field by taking means (16×16 blocks seems to be the best choice here).

Long-Range Covariances

The most important long-range statistic is probably the correlation between two pixels in an image. We can get an estimate of the two-pixel statistics by calculating the covariance

$$C(x, y) = < I(x, y)\, I(0, 0) > , \tag{29}$$

where $< \cdot >$ denotes an average over all images, or alternatively, by calculating the pixel difference function

$$D(x, y) = < |I(x, y) - I(0, 0)|^2 > \ . \tag{30}$$

The latter formulation is a good choice when images are offset by an unknown constant. The two functions are otherwise equivalent since

$$D(x, y) + 2\, C(x, y) = \text{constant} \ . \tag{31}$$

Derivative Statistics

Natural Images ($\alpha = 0.55$) Synthetic Images ($\alpha = 1.0$)

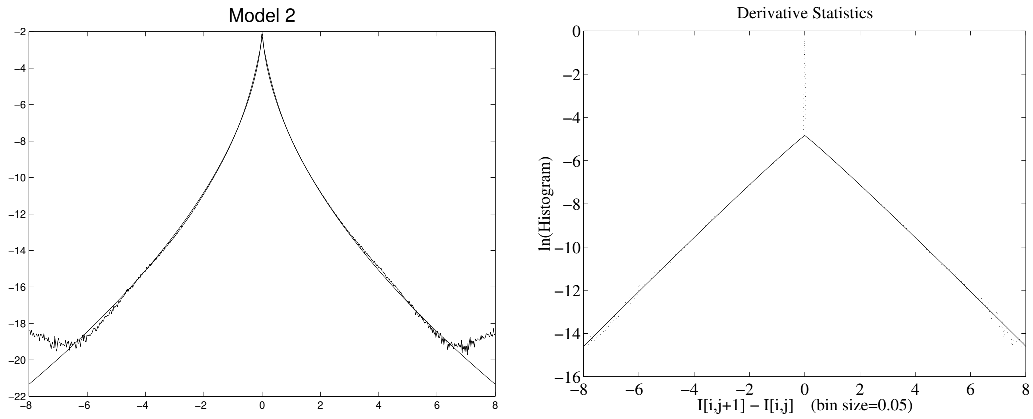

Figure 7: **Left**: Logarithms of normalized histograms of $\nabla_x \tilde{I}$ for natural images; best fit to a Laplacian distribution for $\alpha = 0.55$. The bin size is 0.025. [Courtesy of J. Huang] **Right**: Logarithms of normalized histograms of $\nabla_x I$ for images generated according to the occlusion model (dotted line); a double exponential, i.e. Laplacian distribution with $\alpha = 1$, provides a good fit (solid line). The bin size is 0.05. (Note: The scales on the vertical axes are different in the two figures.)

In [10], Huang shows that the "one-dimensional" difference function for natural images (average $D(x, y)$ over all directions) is best modeled by

$$D(x) = A + B \cdot x^{-\eta} + C \cdot x \ , \tag{32}$$

where x is the separation distance between two pixels. The power-law term dominates the short-range behavior, while the linear term dominates at long distances. In our occlusion model, the linear term above is absent. The difference function for the synthetic images is best modeled by

$$D(x) = A + B \cdot x^{-\eta} \ , \tag{33}$$

which in the Fourier domain corresponds to a power spectrum of the form $1/k^{2-\eta}$.

Fig. 9 (*left*) shows a horizontal cross section of $D(x, y)$ for natural images [10], and Fig. 9 (*right*) shows a log-log plot (base 2) of the derivative of the positive part of the cross section. Note that a power-law behavior according to Eq. 33, would lead to a straight line with slope $(1 + \eta)$ in a log-log plot. However, as pointed out in [10], the curve in Fig. 9 (*right*) is only straight for $2 < \log_2 x < 5$, i.e. for distances between 4 and 32 pixels. In this region the slope is -1.19 (corresponds to $\eta = 0.19$ in Eq. 33), but the curve turns and becomes almost horizontal for large distances.

The solid line in Fig. 10 (*left*) shows the (orientationally averaged) correlation function $C(x)$ for the synthetic images. For comparison, we have also plotted (dashed line) the theoretically predicted two-point correlation function for $r_{\min} = 1/2$, $r_{\max} = 2048$ and $C_0 = 1.06$ (see Eq. 19 and Fig. 2). In Fig. 10 (*right*), we plot $\log_2[-2 \cdot C'(x)] = \log_2[D'(x)]$ as a function of $\log_2 x$. The curve is straight for all distances x. The slope of the line is -1.17, which corresponds to $\eta = 0.17$ in Eq. 33.

Derivative Statistics After Block Averaging Images from the Occlusion Model. Fit to a Generalized Laplacian Distribution.

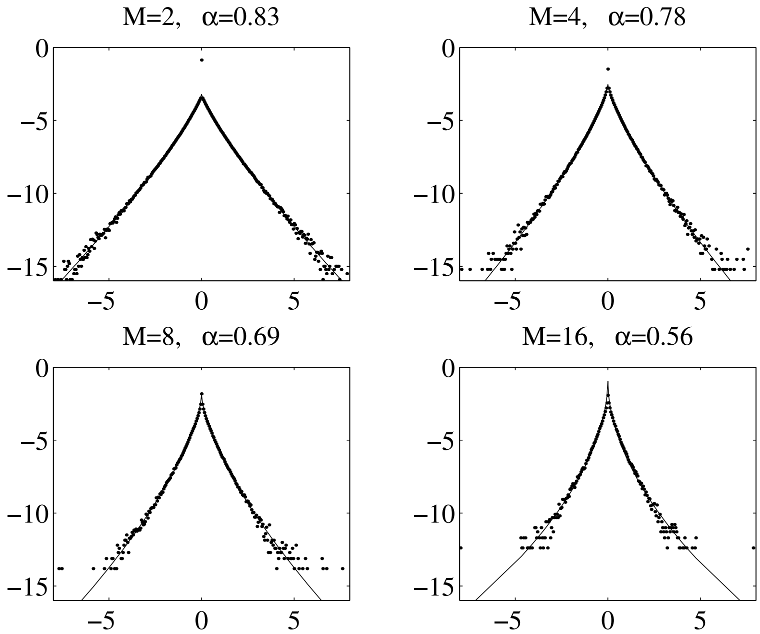

Figure 8: Logarithms of normalized histograms of $\nabla_x I$ after block averaging (see dots). The averages are taken over $M \times M$ blocks. To each histogram, we fit a generalized Laplacian distribution with parameters $\{s, \alpha\}$ (see Eq. 28). **Top left:** M=2. Best fit gives $\alpha = 0.83$. **Top right:** $M = 4$, $\alpha = 0.78$. **Bottom left**: $M = 8$, $\alpha = 0.69$. **Bottom right:** $M = 16$, $\alpha = 0.56$.

3.3 Scaling of Synthetic Images

In this section, we study the scaling properties of images sampled from the occlusion model. Many of the ideas here come from physics: In the theory of critical phenomena, we say that a lattice field is *scale invariant*, or a *fixed point of renormalization*, if it remains invariant under a transformation which involves coarse graining and a change of scale. Block averaging is one possible way of defining renormalization in physics, but other procedures exist.

The idea above, that one can look for probability models which are invariant under many different sorts of transformations, has motivated us to study two different types of scaling procedures for the digitized images:

- *Block averaging (mean filtering).* This is a linear transformation. We scale down an image by a factor of σ by calculating the mean intensity of disjoint $\sigma \times \sigma$ blocks. If the original image is an array $\{I[i,j] \mid 0 \leq i,j \leq N-1\}$, then scaling down by a factor σ gives a rescaled image I^{σ} with

$$I^{(\sigma)}[i,j] = \frac{1}{\sigma^2} \sum_{m=0}^{\sigma-1} \sum_{n=0}^{\sigma-1} I[i\sigma + m, j\sigma + n] \ , \tag{34}$$

where $0 \leq i,j \leq (N/\sigma - 1)$.

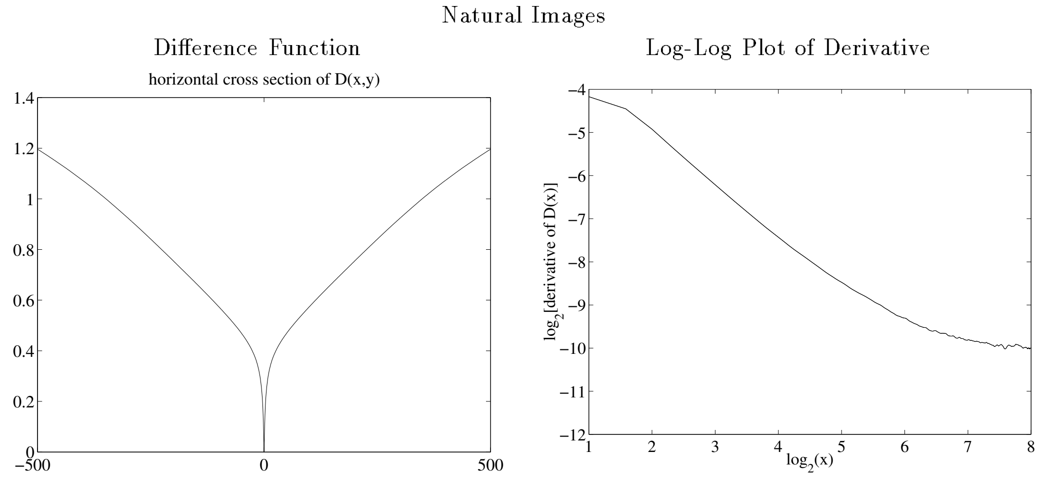

Figure 9: **Left**: Horizontal cross section of the difference function $D(x, y)$ for natural images. [Courtesy of J. Huang] **Right**: Log-log plot with base 2 of the derivative of the cross section (positive part).

- *Median filtering.* This is a non-linear transformation, and has to be done recursively in 2×2 blocks. We scale down an image by a factor $\sigma = 2^k$ by taking the median of disjoint 2×2 blocks k times. If three pixels in the 2×2 block are the same, say (a,a,a,b), we declare the median equal to a, regardless of the value of b. Isolated pixel changes will then have no effect on the median.

Block averaging and median filtering are both coarse graining procedures that take away fluctuations in the system whose scale is smaller than the block size. The two transformations, however, affect the image in different ways. Block averaging *smears out* the small fluctuations. It is as if one looked at the image through an out-of-focus lens. Median filtering, on the other hand, is better described as a *removal* of single pixel fluctuations from a background of larger features.

Scaling of Derivative Statistics

Fig. 11 shows the log-histograms of $\nabla_x I^\sigma$ for six different scales: $\sigma = 1, 2, 4, 8, 16, 32$. As a rule, the variance of the pixel intensities decreases after down-scaling, and the normalized histogram of I^σ becomes narrower. Here, we have divided out the standard deviation from the pixel intensities. For block averaging (*left*), the "renormalization factor" (1/standard deviation) is in the range 1.09-1.10. For median filtering (*right*), the "renormalization factor" is in the range 1.00-1.01. To measure the departure from scale-invariance, we look at *both* the change of the shape of the histograms after rescaling and the deviation of the "renormalization" factor from 1 [9] For both block averaging and median filtering, the histograms remain far from Gaussian for *all* six scales. However, the images can only be said to be fully scale invariant under median filtering – the normalized histograms for different

[9] Note that, according to this definition, white noise images do not scale under block averaging. The standard deviation of the pixel intensities for white noise decreases with a factor of σ when $\sigma \times \sigma$ blocks are averaged; this corresponds to a "renormalization factor" with value 2.

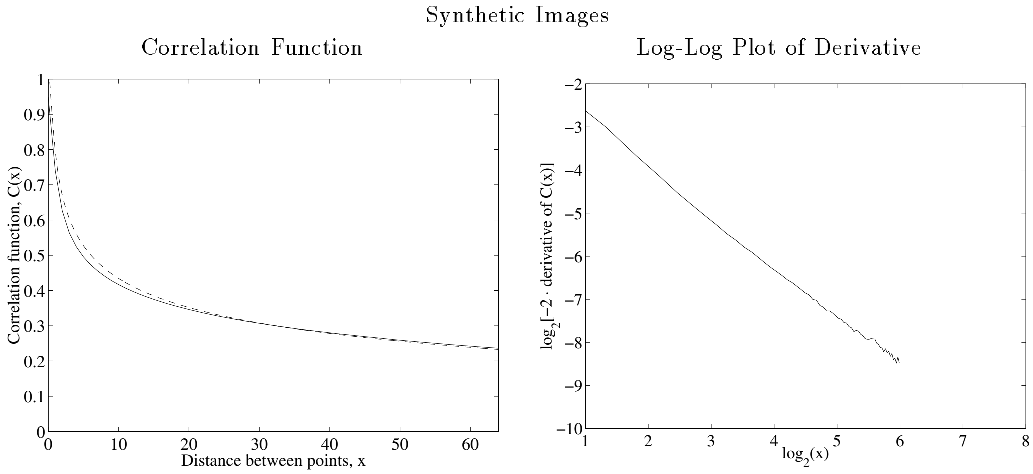

Figure 10: **Left**: The solid line represents the numerically calculated two-pixel correlation function $C(x)$ (averaged over all orientations) for synthetic images. The dashed line represents the theoretically predicted two-point correlation function for the model with $r_{\min} = 1/2$, $r_{\max} = 2048$ and $C_0 = 1.06$ (see Eq. 19 and Fig. 2). **Right**: Log-log plot with base 2 of the derivative $(-2 \cdot C'(x))$.

scales have almost identical shape, and the "renormalization factor" is close to 1.

Scaling of Long-Range Covariances

To study how the long-range statistics of the images scale, we calculate the two-pixel correlation or covariance function $C^\sigma(x)$ for different scales σ. We average over angles, pixel pairs (where two pixels are a distance x apart), and images. The diamond marks in Fig. 12 represent $C^\sigma(x)$ as a function of the distance x between the pixels, for $\sigma = 1, 2, 4, 8, 16, 32$. As mentioned before the covariance can only be excepted to be scale-invariant up to a constant. Note, however, that we have not shifted the curves here as in Sec. 2.2.

In the *left* graph, we scale down by block averaging. The solid lines represent the two-point correlation functions $C(\sigma x)$, defined by Eq. 21 and Eq. 20, for $r_{\min} = 1/2$, $r_{\max} = 2048$, and $C_0 = 1.06$ (cf. Fig. 3). In the *right* graph, we scale down by median filtering. The solid lines represent the two-point correlation functions $C(\sigma x)$ for $r_{\max} = 2048$, a short-distance cutoff $r_0 = 1/2$ (see Sec. 2.2 for definitions), and $C_0 = 1.06$ (cf. Fig. 4).

We see that the numerical results from block averaging agree well with the results from a rescaling $C(x) \to C(2x)$ of the theoretically calculated two-point correlation function. Median filtering, on the other hand, seems similar to a rescaling $C(x) \to C(2x)$ where the short-distance cutoff is kept *fixed* (Let $r_{min}^{(\sigma)} = \sigma \cdot r_{\min}$).

4 Distribution of Objects in 3D

So far we have only considered 2D models where "objects" are directly placed on a plane. These "objects", however, have no real physical meaning; They only make sense when we regard images as perspective projections of the real world, which is three-dimensional,

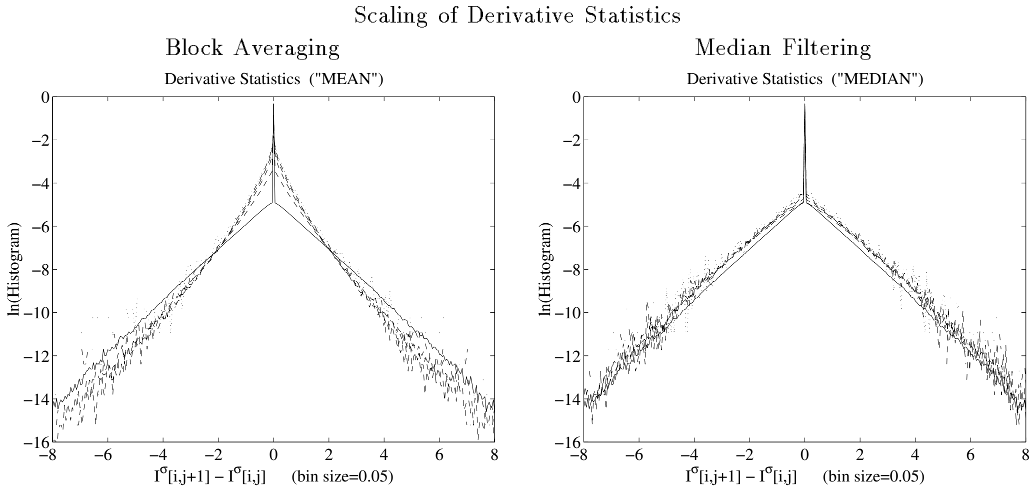

Figure 11: Logarithms of marginal distributions of $\nabla_x I^\sigma$ for scales $\sigma = 1, 2, 4, 8, 16, 32$. **Left**: Scaling by block averaging. **Right**: Scaling by median filtering 2×2 blocks recursively.

onto a planar surface. The *perceived* object sizes in the images are functions of both the *real sizes* of the objects in 3D and the *distances* to the objects.

In Sec. 2.1, we showed that full scale invariance for images requires that the object sizes r in 2D obey a cubic power-law distribution. The question is: What does the $1/r^3$ "law of sizes" in the image imply about the distribution of *real* sizes in the 3D world?

Let us model the world by randomly placing thin, uniformly colored discs in 3D (parallel to the image plane) according to a Poisson process with measure

$$d\mu = g(R)\, dX\, dY\, dZ\, dR \tag{35}$$

on (X, Y, Z, R)-space. The coordinates (X, Y, Z) represent the position of the discs, and R represents the radii of the objects.

We then transform from world coordinates (X, Y, X, R) to screen coordinates (x, y, r, t), where (x, y) represents the positions of the discs on the screen, r the "apparent" sizes of the objects, and t the order or the time an object occurs on the screen (see Sec. 2.1). Perspective viewing with occlusions according to

$$x = \frac{X}{Z}\ , \quad y = \frac{Y}{Z}\ , \quad r = \frac{R}{Z}\ , \quad t = Z\ . \tag{36}$$

leads to the measure

$$d\mu^* = t^3 g(tr)\, dx\, dy\, dr\, dt \tag{37}$$

on "image space". We know from Sec. 2.1 that full scale invariance requires that

$$t^3 g(tr) \propto \frac{1}{r^3}\ , \tag{38}$$

which means that the real objects sizes R has to be distributed according to

$$g(R) \propto \frac{1}{R^3} \tag{39}$$

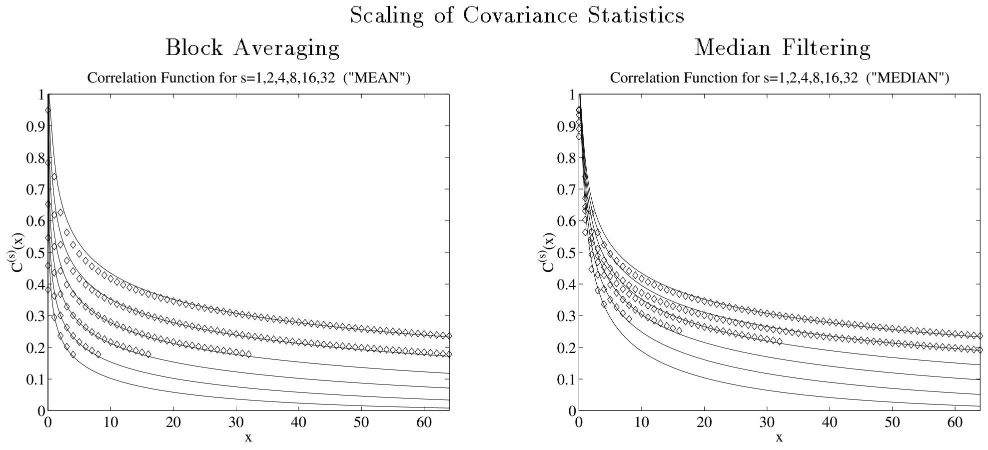

Figure 12: The pixel covariance $C^{\sigma}(x)$ as a function of x for $\sigma = 1, 2, 4, 8, 16, 32$. **Left:** Scaling by block averaging (see diamond marks). For comparison (see solid line), we have plotted the two-point correlation functions $C(\sigma x)$, defined by Eq. 21 and Eq. 20, for $r_{\min} = 1/2$, $r_{\max} = 2048$, and $C_0 = 1.06$ **Right:** Scaling by median filtering (see diamond marks). For comparison (see solid line), we have plotted the two-point correlation functions $C(\sigma x)$, for a *fixed* short-distance cutoff $r_0 = 1/2$ ($r_{\max} = 2048$ and $C_0 = 1.06$).

in the 3D world.

The statement above, that the real sizes of the objects are distributed according to a cubic power-law, is very restrictive, and maybe not so realistic. To get a more plausible model of the world, we may have to relax one of the basic assumptions in the 3D model, such as translational invariance [10], or statistical independence of objects. Note that in the real world, objects often appear in clusters. Single objects also break up into smaller parts. On the ground, for example, we may see groups of trees, and on the trees branches and leaves. A more realistic variant than the Poisson model above may be to generate objects in groups on or near the surface of a parent, as in a random branching process.

5 Summary and Conclusions

We have studied a model for scale invariance of natural images based on the idea that images can be broken down into statistically independent objects. The model takes occlusions into account, and is also translationally invariant. Theoretically speaking, the model is fully scale invariant (with occlusions), if the "apparent" sizes of the objects obey a cubic power-law $1/r^3$, and there are no boundaries on the object sizes. In practice, however, we need to impose a lower and upper limit on the allowed object sizes to prevent the power from blowing up at the UV and IR limits. We have derived an analytical expression for how the covariance statistics of the model depend on the limits on the object sizes. The calculations indicate that the presence of characteristic length scales in the system introduces what in the theory of critical phenomena is called an "anomalous dimension": The

[10] Chi, for example, argues that objects should be modeled as being distributed in a subregion of the 3D space [1]. In his model for the origin of scaling, there is a constant $H > 0$, such that objects are distributed by a homogeneous Poisson law in the region between the earth and the height H.

covariance has the form $C(x) = -A + B \cdot x^{-\eta}$ and a power spectrum of the form $1/k^{2-\eta}$, where η, the "anomalous dimension", is a small positive constant. It is interesting to note that natural images also have a power spectrum of the form $1/k^{2-\eta}$, where η is very small; it is possible that the non-zero constant arises because of a short-distance cut-off present in the images.

We have also systematically compared the statistics of simulated images from our occlusion model with data from natural images. We found that both the single pixel statistics and the derivative statistics agree well with natural data if we block average our images a few times for a suitable subpixel resolution. As mentioned above, our model predicts a power-law form for the correlation function, which agrees well with the short-range and middle-range statistics of natural images. However, our model does not produce a linear tail in the difference function for long distances; the latter has been observed in natural images [10].

Furthermore, we have developed a new approach to "renormalization fixed points" closer to true images and based on the occlusion model. Scaling of the derivative statistics and covariance statistics of the synthetic images show that the images are closer to a fixed point under (recursive) median filtering rather than block averaging. Block averaging and median filtering are both coarse graining procedures which take away short-wavelength fluctuations in the system. The two transformations, however, affect the image in different ways. Block averaging smears out the small fluctuations. Median filtering, on the other hand, is better described as a removal of single pixel fluctuations from a background of larger features; effectively it acts as a short-distance cutoff in the images.

Finally, the condition that the "apparent" sizes r of the objects are distributed according to $1/r^3$ is really equivalent to a 3D picture of the world where the real sizes R of the objects are distributed according to $1/R^3$. This indicates that we may need to relax one of the basic assumptions in the model to make it more plausible. We believe that a more realistic variant of the occlusion model should include dependencies between objects. A next step may be to generate objects in groups on or near the surface of a parent, as in a random branching process.

6 Acknowledgments

We would like to thank Jinggang Huang for data and figures on natural image statistics. We are also grateful to Harel Shouval and Stuart Geman for helpful discussions.

References

[1] Z. Chi. "Probability Models for Complex Systems", Ph.D. Thesis, Division of Applied Mathematics, Brown University, 1998.

[2] D. J. Field. "Relations between the statistics of natural images and the response properties of cortical cells". *Journal of the Optical Society of America* A, 4.2379-2394, Dec 1987

[3] N. Goldenfeld. "Lectures on Phase Transitions and the Renormalization Group". Addison-Wesley, 1992.

[4] D. Mumford, B. Gidas. "Stochastic Models for Generic Images". Submitted to *Quarterly Journal of Mathematics.*

[5] D. L. Ruderman. "The Statistics of Natural Images". *Network*, vol 5, no 4, pp 517-548, 1994

[6] D. L. Ruderman. "Origins of Scaling in Natural Images". *Vision Research*, vol 37, no 23, pp 3385-3395, 1997

[7] Y. G. Sinai. "Self-Similar Probability Distributions". *Theory of Probability and its Applications*, vol XXI, pp 64-80, 1976

[8] S. C. Zhu, and D. Mumford. "Prior Learning and Gibbs Reaction-Diffusion". *IEEE Trans. on PAMI*, vol 19, no 11, pp 1236-1250, 1997.

[9] 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.

[10] J. Huang, and D. Mumford. "Statistics of Natural Images and Models", *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol 1, pp 541-547, Fort Collins, June 1999.

[11] S. G. Mallat. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation". *IEEE Trans. PAMI*, 11:674-693, July 1989.

[12] E. P. Simoncelli and E. H. Adelson. "Noise Removal via Bayesian Wavelet Coring". *Int'l Conf. on Image Processing*, pp 379-383, Lausanne, Sept 1996.