Occlusion Models for Natural Images: A Statistical Study of a Scale-Invariant Dead Leaves Model

ANN B. LEE

Department of Physics, Brown University, Providence, RI 02912, USA

Ann_Lee@brown.edu

DAVID MUMFORD AND JINGGANG HUANG

Division of Applied Mathematics, Brown University, Providence, RI 02912, USA

Abstract. We develop a scale-invariant version of Matheron's "dead leaves model" for the statistics of natural images. The model takes occlusions into account and resembles the image formation process by randomly adding independent elementary shapes, such as disks, in layers. We compare the empirical statistics of two large databases of natural images with the statistics of the occlusion model, and find an excellent qualitative, and good quantitative agreement. At this point, this is the only image model which comes close to duplicating the simplest, elementary statistics of natural images—such as, the scale invariance property of marginal distributions of filter responses, the full co-occurrence statistics of two pixels, and the joint statistics of pairs of Haar wavelet responses.

Keywords: natural images, stochastic image model, non-Gaussian statistics, scaling, dead leaves model, occlusions, clutter

1. Introduction

Recently, there has been a great deal of interest in the statistics of natural images and many investigations of these from both the computational and biological vision perspectives. From the computational side, this has been motivated by many applications including: (i) the need for more effective image compression, e.g., Buccigrossi and Simoncelli (1999), (ii) the search for better deblurring and denoising algorithms, e.g., Freeman and Pasztor (1999), and (iii) the need to estimate the rates of false positives and false negatives in target or face recognition algorithms, e.g., Sullivan et al. (1999).

There are two discoveries which have motivated the search for good models of image statistics. The first is that image statistics are extremely non-Gaussian. For example, a large class of image enhancement algorithms is based on the paradigm of decomposing an observed image I into an enhanced image J and a noise component n and it is standard to assume n is Gaussian noise. These algorithms perform poorly because the 'noise' component which one wants to remove is often really caused by the clutter[1] of small objects or markings, partially resolved by the camera and not important to the understanding of the scene, and the statistics of such clutter are not Gaussian at all. The non-Gaussian nature of image statistics is apparent if one computes histograms of virtually any filter on virtually any database of images: the histogram will, essentially always, have kurtosis greater than 3 (the kurtosis of Gaussian distributions).

The second reason for the interest in image statistics derives from the empirical observation that image statistics *scale*. This means that any local statistic calculated on $n \times n$ images and on block averaged $2n \times 2n$

images should be the same, or that the probability of seeing an image $I(x, y)$ is the same as that of seeing $I(\sigma x, \sigma y)$. Although not exact, this result has been approximately confirmed on all large image databases that we have heard about. The reason this is exciting is that it implies that local image models describing small-scale structures will work as global image models describing the large-scale structures in images. This creates a surprising stability for image statistics and creates a link between the denoising and object recognition problems. In the first case, one wants to eliminate small irrelevant structures, in the second one wants to reject large background objects other than the sought-for object. Since local clutter and generic backgrounds have the same statistics, the same sort of statistical tools can be used. In fact, in automated target recognition (ATR), the term clutter is used to describe the mass of irrelevant foreground and background objects, for example rocks and trees, which occur at all sizes (more smaller ones than larger ones) in natural images, and which significantly degrade the power of object recognition algorithms. The key issue in ATR is whether you have to identify and model every one of the objects in the foreground/background before finding the object of interest, or whether there are common statistics in natural images which enable you to separate the target from the rest of the scene without explicitly describing the clutter in detail.

For all the applications mentioned above, we need new non-Gaussian stochastic models for natural images. We may broadly characterize stochastic models of images into two classes:

– "Descriptive" models. In these models the only random variables are the image pixel values (or the values of filters applied to the image). Typical models of this type are Gibbs models:

$$\Pr(I) = \frac{1}{Z} \, e^{-\sum_{\text{pixels } P} E(I|_{N(P)})}$$

where $N(P)$ is a suitable pixel neighborhood of P, E is a local "energy" function and Z is the normalizing constant. While these models have been able to describe a wide range of simple textures, see e.g., Heeger and Bergen (1995) and Zhu et al. (1998) they lack the concept of structured objects, and usually fail to be scale-invariant or to reproduce the correct long-range dependencies.
– "Generative" models. These models *include hidden variables for the underlying causes of structures in real world imagery*. An image is here a composite

of objects or imagelets with extra random variables such as the location, scale and grey level of particular objects or imagelets. The generative models most common in the literature are (1) models based on templates of specific objects, such as faces, e.g., Hallinan et al. (1999), tanks, machined parts etc., often with extra variables such as non-linear distortions, lighting variables and location of key points on the template, or (2) models which approximate images by a linear superposition of transparent basis images, as in ICA, e.g., Olshausen and Field (1996), PCA and wavelet expansions.

The classical template-based models are useful to model the object of interest in target recognition algorithms, but are much too detailed for the large mass of irrelevant objects in the rest of the scene. Stochastic models based on linear expansions, on the other hand, represent a good first approximation of natural images but fail to capture or realistically model some basic properties of natural images; for example, T-junctions, the presence of regions with almost no contrast variation (in the case of infinitely divisible models with scale invariance, see Mumford and Gidas, 2000), and extended contours or regions broken into pieces due to occlusions.

This paper studies a hybrid class of generative image models which shares some attributes of both classical template-based models and wavelet expansion models. The basic idea behind these so called "dead leaves models" is to assume that the image is formed from a set of template-based elementary objects, whose locations and possibly scale are a sample from a Poisson process, and which partially occlude one another, being laid down in layers. These models were first introduced by Matheron (1968) and Serra (1982) for the morphological analysis of materials, and has only recently been applied to natural image statistics—see, for example, Ruderman (1997) on the origins of power spectrum scaling, and Alvarez et al. (1999) for the analysis and representation of geometric features in natural images.

An important issue is whether such models are only useful as toy models whose statistics can be studied or whether they can be fit to real images with a reasonable amount of computation. For some applications, it may not be necessary to use very complex generator models. In fact, Grenander and Srivastava (2000) have recently showed that a simple generator model made up of the profiles of the same object can already capture the

variability of different types of scenes better than any of the existing clutter models, and that the parameters of the model (which in this case are two) can be directly estimated from the derivative histograms of real images. For applications that require a more accurate description of the complexity of cluttered scenes, there will of course be a larger number of random variables and parameters that have to be estimated.

In a recent paper (Zhu and Guo, 2000), Zhu and Guo have attacked this problem and shown that, at least in some cases, a generative model can be fit to textured scenes with multiple moderately large textons by using sophisticated Monte Carlo techniques. The advances in Monte Carlo methods, including the CONDENSATION algorithm (Isard and Blake, 1998), Data Driven MCMC (Zhu et al., 1999) and adaptive texton segmentation models (Malik et al., 1999), have made the computation of complex hidden variables and complex scene structure seem much more accessible than before. For these reasons, we do think that probability models that are physics-based and include structured primitives are practical.

The goal of this paper is, however, not to develop tools for any of the applications listed above. Our aim is to study whether a simple version of the dead leaves model can reproduce the empirical statistics of natural images better than any of the other classes which have been studied, e.g. Gaussian models, the "scale mixture of Gaussians model" of Wainwright and Simoncelli (2000), the infinitely divisible models by Mumford and Gidas, and Markov Random Fields. Ultimately, we believe that different variations of dead leaves models (possibly with more realistic shapes and dependencies between objects) will be useful for many applications. More specifically, we will study four basic image statistics and statistical properties to see whether simple dead leaves models duplicate the empirical facts drawn from very large databases of natural images. The first is the scaling behavior of images, described above. The second, also mentioned above, is the high kurtosis of filter statistics derived from images, e.g. derivatives (the difference of adjacent pixel values). The third is the highly irregular shape of joint histograms of wavelet coefficients, esp. Haar wavelets. The fourth is the complex behavior of the full two-pixel co-occurrence statistics which, as we will see below, seem to be best modeled as mixtures. This is the most direct confirmation of the accuracy of random-collage models where this mixture property is fundamental.

We wish to explain our emphasis on Haar wavelet filters in this study. Because of the simplicity of Haar filters, any structure in the statistics can be directly related to pixel values (see for example Huang et al., 2000) for an analysis of the local image structure in range images. It also seems that Haar filters show the non-Gaussian structure in local image patches clearer than smooth filters.

There appear to be two types of non-Gaussian behavior found in the statistics of filter responses. The first is the high kurtosis of the distribution of such filters (Field, 1987). Independently, Grenander and Srivastava (2000) and Wainwright and Simoncelli (2000) propose that this is the result of the filter response being a product of an independent "contrast factor" and a "geometry factor" reflecting the geometry of a contrast-normalized image. The kurtosis of such products is the product of the kurtosis of the factors, hence is 9 if the factors were Gaussian.

The second type of non-Gaussian behavior concerns the joint histogram $p(x)dx$ of k filters $\mathbf{R}^k$, $k \leq 2$, and considers the equi-probable contours $p = $ cnst. If these are not ellipsoids, the filters are not jointly Gaussian, not even a "scale mixture of Gaussians", i.e. a product of a scalar contrast factor and a Gaussian distributed vector Wainwright and Simoncelli (2000).[2] It seems that Haar filters produce more non-elliptical contours than smooth filters: compare (Zetzsche et al., 1993; Simoncelli, 1999, esp. Fig. 5) with (Huang and Mumford, 1999, Fig. 8)—hence, we use them as a stronger test for our model. The fact that a dead-leaves model with *circular leaves* and smoothing reproduces these irregular equi-probability contours shows that the contours are not simply an artifact of the interaction of Haar filters with horizontal and vertical edges in natural images.

The paper is organized as follows. In Section 2, we will introduce and analyze an approximately scale-invariant dead leaves model, discussing the need for large and small cut-offs in the sizes of the templates and their scaling behavior. In Section 3, we compare smoothed dead-leaves images with two large databases of natural images. We look first in detail at the derivative statistic and the scaling behavior. A considerable variation in this statistic is encountered in different classes of natural scenes and these will be compared to different versions of the dead leaves model. After this we consider the joint histogram of pairs of Haar wavelet responses. Finally, we look at the two-pixel co-occurrence

statistics and the effects of smoothing in the model.

2.   Dead Leaves Model with Approximate Scale Invariance

2.1.   *Basic Set-up. A 2.1D Sketch of the World*

Our model is based on the notion that the world can be broken down into approximately independent discrete objects of different sizes. When viewed, the 3D world creates a collage of objects that occlude each other (see Fig. 1):

Imagine a simplified 3D picture, where the viewed surfaces of objects are modeled as rigid planar templates parallel to the image plane; and each template is at a random position (x, y, z), where z is the distance to the plane. Mathematically, we write the solid world as

$$\bigcup_i (T_i \oplus s_i), \qquad (1)$$

where s_i are points from a uniform Poisson process $\Pi = \{(x_i, y_i, z_i)\}$ in $\mathbb{R}^3$ (or $\mathbb{R}^2 \times [0, z_{max}]$) of intensity λ; and T_i are closed sets in $\mathbb{R}^2 \times \{0\}$, centered at $\mathbf{0}$ and of random sizes r_i.

Each template T_i is furthermore painted with some albedo a_i. In Section 2.5, we assume a uniform albedo a_i, and disk-like templates; but in principle, a_i can be a function of T_i, and the template can be of any shape,

e.g. a face or a tree (fixed shape), or a polygon with random sides and angles (random shape function).

We assume that the random variables $r_1, r_2, \ldots$ and $a_1, a_2, \ldots$ are independent samples from $f(r)$ and $p(a)$, respectively. The variables r_i and a_i can be regarded as markings of Π, but we can also (according to the marking theorem, Kingman, 1993) consider the set of points

$$\Pi^* = \{(x_i, y_i, z_i, r_i, a_i)\} \qquad (2)$$

as *a Poisson process in the product space* $\mathbb{R}^2 \times [0, z_{max}] \times [r_{min}, r_{max}] \times [a_{min}, a_{max}]$ with measure

$$d\mu^* = \lambda f(r) p(a) \, dx \, dy \, dz \, dr \, da. \qquad (3)$$

Equations (1)–(3) define the 3D world of objects—the physical world. The next issue is how to map the 3D world to 2D images. In a model with transparent objects (see for example the random wavelet expansion in Mumford and Gidas (2000) an image $J(x, y)$ is given by an arithmetic sum:

$$J(x, y) = \sum_i a_i, \qquad (4)$$

where i is such that $(x - x_i, y - y_i) \in T_i$.

In our model, however, objects are *opaque*. We define the image $I(x, y)$ by (orthographic) projections with occlusions, and let $z_{max} \to \infty$ for complete
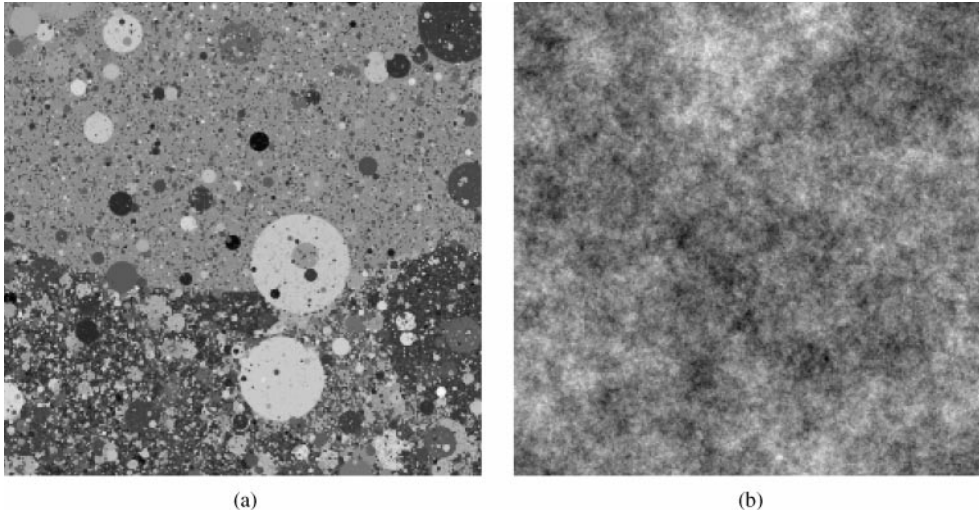

Figure 1.   (a) Computer-simulated sample from a dead leaves model; see Section 2.5 for details. The image is here a collage of discrete objects which partially occlude one another. Compare with (b) a computer-simulated sample from the standard Gaussian model, where there are no clear objects and borders. Both images here are approximately scale invariant.

coverage.[3] This gives

$$I(x, y) = a_{i(x,y)}(x - x_i, y - y_i), \qquad (5)$$

where

$$i(x, y) = \arg\min_{i}\{z_i \mid (x - x_i, y - y_i) \in T_i\}. \quad (6)$$

is the index of the *closest* object in a certain (x, y)-direction.

2.1.1. Remarks on Density Parameter and Sampling. A Perfect Simulation of the Dead Leaves Model.

1. In the images (Eq. (5)) with $z_{max} \to \infty$, the density parameter λ does not affect the amount of "clutter" or other statistics in the image samples—as they do in a transparent world (Eq. (4)) where z_{max} has to be finite: The images are created by a process which looks at all $z \geq 0$, and there is no explicit z-dependency in the model. Thus, rather than sampling in 3D (from the Poisson process Π), we can place random templates in an ordered sequence front to back (the order corresponding to the relative distance to the viewer) until the background is completely covered. The latter construction is used in Section 2.5 for the numerical simulation of dead leaves images.
2. Furthermore, it can be shown (Kendall and Thönnes, 1998) that *a front-to-back simulation until complete coverage* is equivalent to the conventional back-to-front simulation[4] of the dead leaves model until stationarity. In other words, the construction delivers *exact samples from the equilibrium distribution* of a Markov chain $\{I^{(k)}(x, y)\}$ where

$$I^k(x, y) = \begin{cases} I^{(k-1)}(x, y) & \text{if } (x, y) \notin T_k \\ a_k & \text{if } (x, y) \in T_k \end{cases} \quad (7)$$

for $k = 1, 2, \ldots$, and random templates T_k with grey levels a_k.

2.2. Condition for Scale Invariance: Cubic Law of Sizes

As mentioned before, one of the most striking properties of natural images is an invariance to scale. This puts a strong constraint on realistic stochastic models for natural images.

In Ruderman (1997) scaling is related to a power-law size distribution of statistically independent regions, but the discussion is limited to scaling of the

second-order statistics, and dead-leaves models with disks. In this section, we use a different formalism—the full probability measure of a Poisson process—to show that higher-order scaling, which has been observed empirically in the filter responses of natural images (Ruderman, 1994; Zhu and Mumford, 1997), places further restrictions on the images. The objects can also be of general shape.

The argument is that images are fully scale-invariant, i.e.

$$\Pr\{I(x, y)\} = \Pr\{I(\sigma x, \sigma y)\}, \qquad (8)$$

if the Poisson process Π^* (Eq. (2)) is invariant under "2D scaling"

$$x \to \sigma x, \quad y \to \sigma y, \quad r \to \sigma r. \qquad (9)$$

Note that the Poisson process Π^* in $\mathbb{R}^2 \times [0, z_{max}] \times [r_{min}, r_{max}] \times [a_{min}, a_{max}]$ is uniquely determined by the measure $d\mu^*$ (Eq. (3)). Now, scaling according to Eq. (9) leads to a Poisson process

$$\Pi_\sigma^* = \{(x_i, y_i, z_i, r_i, a_i)\} \qquad (10)$$

in $\mathbb{R}^2 \times [0, z_{max}] \times [r_{min}/\sigma, r_{max}/\sigma] \times [a_{min}, a_{max}]$ with measure

$$d\mu_\sigma^* = \lambda \sigma^3 f(\sigma r) p(a) \, dx \, dy \, dz \, dr \, da. \qquad (11)$$

Assume, for the time being, that we can ignore the short- and long-distance cut-offs in object sizes. The model above then scales if and only if

$$d\mu^* = d\mu_\sigma^*. \qquad (12)$$

Furthermore, if Eq. (12) is true, then the images $I(x, y)$ and $I(\sigma x, \sigma y)$—which are projections of samples from Π^* and Π_σ^* (see Eq. (5)), respectively—are statistically equivalent.

Equation (12) leads to a constraint on the sizes r of the objects:

$$f(r) = \sigma^3 f(\sigma r) \Rightarrow f(r) \propto r^{-3}, \qquad (13)$$

We refer to this as the "cubic law of sizes".

2.3. The Need for Cut-Offs in Sizes

It has previously been shown (Mumford and Gidas, 2000) that transparent models (the "random wavelet

expansion") can be fully scale invariant. A possible problem with a model with occlusions is that we need size cut-offs $r_{\min}$ and $r_{\max}$ to obtain non-trivial images. It can be shown that (assume cubic law of sizes): As $r_{\min} \to 0$, the images are totally covered by microscopic objects. 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 (see Appendix A for proof). Almost all image samples will then have uniform intensities.

The finite bounds introduce characteristic length scales in the system; thus, preventing full scaling. Below, we investigate the second-order statistics and scaling of an occlusion model with finite cut-offs.

2.4.   Predictions for Dead Leaves Model with Finite Cut-offs

We assume disk-like templates with random radii from a $1/r^3$ distribution where $r_{\min} \leq r \leq r_{\max}$.

As before, we consider the continuous case—where $I(x, y)$ is a function in $\mathbb{R}^2$ of continuous variables x and y. (The effects of discretization and smoothing are studied empirically in Section 3.)

2.4.1. Two-Point Co-occurrence Function.   Because the images are both translationally and rotationally invariant, the two-point statistics depend only on the *distance* between the points.

Let

$$K(a, b, x) = \Pr\{I(\mathbf{x}_1) = a, I(\mathbf{x}_2)$$
$$= b \mid \| \mathbf{x}_1 - \mathbf{x}_2 \| = x\}, \quad (14)$$

be the co-occurrence function or the joint probability density function (joint pdf) for two points in a random image I; a and b are grey levels, and x is the distance between the points.

In our occlusion model, each object has a *uniform* intensity, and different objects are statistically independent. This gives

$$K(a, b, x) = [1 - P_{\text{same}}(x)] \cdot f(a) \cdot f(b)$$
$$+ P_{\text{same}}(x) \cdot f(b) \cdot \delta(a - b), \quad (15)$$

where $P_{\text{same}}(x)$ is the probability that two points a distance x apart belong to the same object, δ represents the Dirac delta function, and f is the probability density

function for the intensities of the objects. The first term in Eq. (15) represents points on *different objects*, and the second term represents points on the *same object*. This "mixture nature" is fundamental for the model.

In Appendix B, we show that

$$P_{\text{same}}(x) = \frac{B(x)}{2 \log \left(\frac{r_{\max}}{r_{\min}} \right) - B(x)}, \quad (16)$$

where $B(x)$ is defined by Eq. (39), and a numerical estimate is given by Eq. (41).

From $K(a, b, x)$, we can derive *all* statistics of second order; for example, the difference statistics (Section 2.4.2) and the covariance statistics (Section 2.4.3) of two points.

2.4.2. Difference Statistics.   The random variable is here the difference D between two points a fixed distance x apart. From Eq. (15), we have that the probability density function of D is

$$f_D(z) = [1 - P_{\text{same}}(x)] \cdot \int_{-\infty}^{\infty} f(a) \cdot f(a + z) \, da$$
$$+ P_{\text{same}}(x) \cdot \delta(z).$$

Figure 2(a) shows a numerical example for $f(a) = \frac{\lambda}{2} \exp(-\lambda|a|)$ ($\lambda = \sqrt{2}$, $r_{\min} = 1/8$, $r_{\max} = 2048$, $x = 1$). For the "double-exponential" form (which is a first approximation of log-contrast for natural data, see Section 3.1) we get

$$f_D(z) = [1 - P_{\text{same}}(x)] \cdot \frac{\lambda^2}{4} e^{-\lambda|z|} \cdot \left(|z| + \frac{1}{\lambda} \right)$$
$$+ P_{\text{same}} \cdot \delta(z). \quad (17)$$

The peak at $D = 0$ corresponds to regions with no contrast variation ("same object"), and the tails correspond to edges in the images ("different objects"). As we shall see in Section 3.2.2 (Fig. 5(b)), the peak gets shorter and the straight tails become *concave* when images are filtered—but the mixture nature of two distributions (one concentrated at 0, and one with heavy tails) will still remain.

2.4.3. Covariance Statistics.   We write the covariance function schematically as
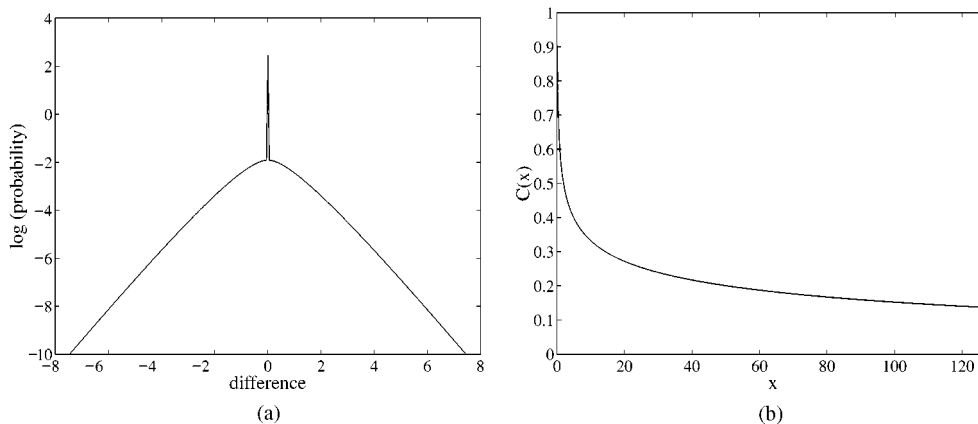
$$C(x) = \langle (I(0)I(x) \rangle \quad (18)$$

Figure 2. Predictions for a continuous dead leaves model with cubic law of sizes and cut-offs $r_{min} = 1/8$ and $r_{max} = 2048$. a: Difference statistics. The figure shows the predicted log (probability) distribution of D for a model with intensities according to a "double-exponential" distribution; D is the grey-level difference for points a distance $\Delta = 1$ apart. The distribution has a sharp peak at $D = 0$, and long straight tails. b: Covariance statistics. The *solid line* shows the predicted covariance function $C(x)$; The overlapping *dotted line* represents a power-law fit $C(x) \approx -0.33 + 0.91 \cdot x^{-0.14}$.

where 0 is an arbitrary origin, x is a location in the image, and the brackets imply an average over angles, a shift over positions, and an ensemble average over different images $I(\mathbf{x})$ (with mean zero). From the co-occurrence function in Eq. (14), we get

$$C(x) = C_0 P_{same}(x), \qquad (19)$$

where $C_0 = \langle I(0)\rangle^2$ is the variance, and $P_{same}(x)$ is given by Eq. (16). This covariance function is approximately a power-law for models with power-law sized objects (see e.g. Ruderman (1997) for disks, and Alvarez et al. (1999) for more general shapes).

Below we use the numerical expression in Eq. (41) to get an estimate of how $C(x)$ depends on the parameters (i.e. the cut-offs) in a model with cubic law of sizes.

Figure 2(b) shows a numerical example for $r_{min} = 1/8$ and $r_{max} = 2048$. The *solid line* represents the predicted covariance function (Eq. (19)), and the *dashed line*) a least-square-error fit (in the region $4 < x < 128$) of this function to

$$C(x) = A + B \cdot x^{-\eta} \qquad (20)$$

The two lines are almost completely overlapping; The best fit occurs for $\eta = 0.14$.

We repeat the power-law fit for different values of r_{min} and r_{max}, and find that the power-law approximation (Eq. (20)) is good for small values of r_{min} and large values of r_{max} (e.g. $r_{min} \lesssim 1$ and $r_{max} \gtrsim 1024$). Figure 3 shows how the number η depends on the ratio r_{max}/r_{min}.

The reason why the figure is interesting is that it gives us an estimate of the deviation from scaling in our model due to the finite cut-offs. Note that *full* scale invariance defined by Eq. (8), implies a power spectrum of the form $1/f^2$, and a covariance function with *log-behavior*—but scaling (with renormalization)

$$\Pr\{I(\mathbf{x})\} = \Pr\{\sigma^\nu I(\sigma\mathbf{x})\}, \qquad (21)$$

leads to a power spectrum $1/f^{2-\eta}$, where $\eta = 2 \cdot \nu$, and a covariance function $C(x)$ with the *power-law form* in Eq. (20). The number η is often known as the "anomalous dimension".

2.5. Numerical Simulation of Dead Leaves Images

So far we have assumed that images $I(x, y)$ are functions of continuous variables x and y. In reality, natural images are given by measurements from a finite array of sensors that average the incident light in some neighborhood. We need to take these things into account in the occlusion model: In Section 3 we analyze a database of 1000 discretized dead leaves images $I[i, j]$ (i and j are the row and column indices, respectively) with subpixel resolution and subpixel objects.

Each image has 256×256 pixels, a subpixel resolution of $1/s$ pixels (length scale), where $s = 16$, and disks as templates. The radii r of the disks are distributed according to $1/r^3$, where r is between $r_{min} = 1/8$ pixels and $r_{max} = 8 \cdot 256 = 2048$ pixels.
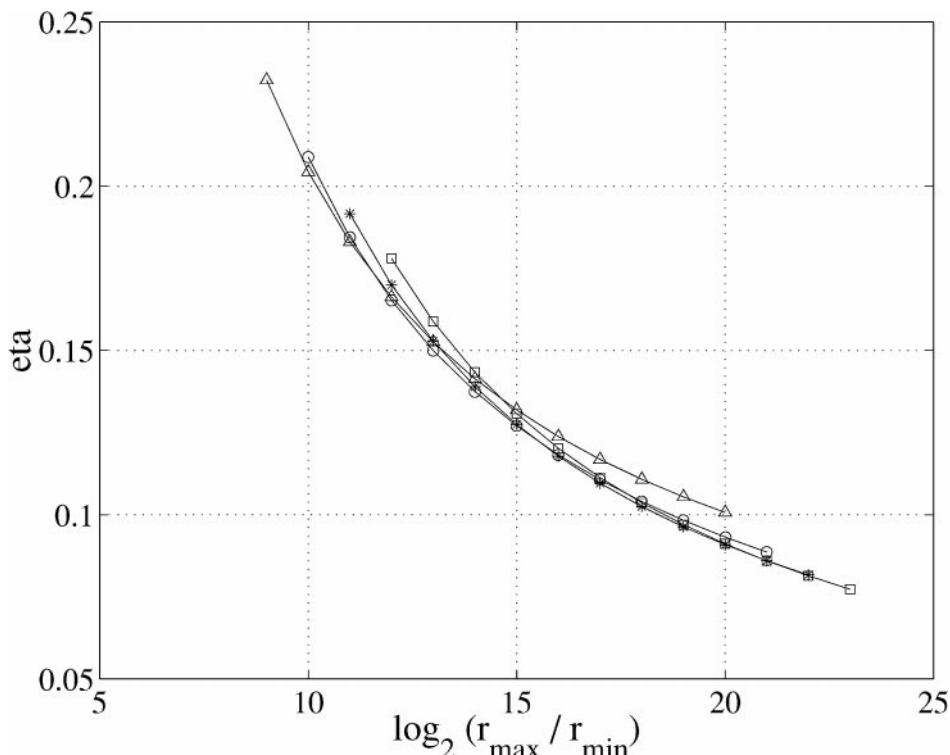
Figure 3. Relation between the number η ("eta") and the cut-offs r_{min} and r_{max} in the dead leaves model with disks and cubic law of sizes. For each curve, r_{max} is fixed and r_{min} is varied. The data points are shown in a plot with η versus $\log_2(r_{max}/r_{min})$. The 4 curves are roughly overlapping, which indicates that the number η is a function of the *ratio* r_{max}/r_{min}. Legend: $\triangle \, r_{max} = 1024$; $\circ \, r_{max} = 2048$; $* \, r_{max} = 4096$; $\diamond \, r_{max} = 8192$.

The exact construction is as follows: First, we make an image which is s times larger than the final image; for $s = 16$, this means 4096×4096 pixels. Assume that the "viewing screen" is defined by $|x| \leq 2048$ and $|y| \leq 2048$. In each iteration, we pick a random position for (the center of) a disk, in an *extended* screen, defined by $|x| \leq 2048 + 16 \cdot r_{max}$ and $|y| \leq 2048 + 16 \cdot r_{max}$—this is to avoid edge effects. The disk is then assigned a radius r from a $1/r^3$ size distribution with $16 \cdot r_{min} \leq r \leq 16 \cdot r_{max}$, and a random grey level a according to the empirical marginal distribution of log-contrast for natural images. We use the 'front-to-back' construction in Section 2.1.1 to make sure that the generated images are samples from a stationary probability distribution: First, we place the *closest* object on the "screen", and then we successively add objects which are farther away until the background is filled.

Finally, we scale down the generated images by averaging pixels in disjoint 16×16 blocks—but other ways of smoothing, such as convolution with a Gaussian filter and subsampling, are also possible.

Figure 1(a) shows an example of a computer-simulated dead leaves image.

3. Statistics of Natural Versus Synthetic Images

Below we compare the empirical statistics of the simulated dead leaves images (see Section 2.5 for details) with natural images from the following two databases:

1. *Database by van Hateren (van Hateren and van der Schaaf, 1998)*. Contains about 4000 1024×1536 calibrated B/W images of mixed urban and rural scenes in Holland.
2. *Database from British Aerospace (courtesy of Andy Wright)*. Consists of 214 calibrated RGB images, with 512×768 pixels, of mixed urban and rural scenes in Bristol, England. Each of these images has been segmented by hand into pixels representing 11 different categories of scenes (see Table 1 and Fig. 4) The segmented database makes possible

Table 1. Different categories and their frequencies.

Category	Frequency	Description
1	10.87	Sky, cloud, mist
2	37.62	Tress, grass, bush, soil, etc.
3	0.20	Road surface marking
4	36.98	Road surface
5	6.59	Road border
6	3.91	Building
7	2.27	Bounding object
8	0.11	Road sign
9	0.28	Telegraph pole
10	0.53	Shadow
11	0.64	Car

the examination of the statistics for each category separately, as well as for the whole ensemble.

Both databases for natural images use calibrated image data, i.e. the images measure light in the world up to an unknown multiplicative constant in each image. To obtain results which are independent of the gain setting, we will work with the *log-contrast* of the

images—defined by

$$I[i, j] = \log(\phi[i, j]) - \langle\log(\phi[i, j])\rangle, \qquad (22)$$

where $\phi[i, j]$ are the calibrated grey-level values and the average $\langle\cdot\rangle$ is taken over each image separately—and use statistics which do not contain the (now additive) constant.

Furthermore, we will always graph the *logarithm of probability*, rather than probability itself, since the log-scale better shows the nature of the tails in the distributions.

3.1. *Single-Pixel Statistics*

The distribution of log-contrast for natural data has a highly non-Gaussian shape with heavy, almost straight tails (see Huang and Mumford, 1999). The single-pixel statistic is however not very informative, as we can strongly modify the histogram of an image without affecting much of its perception. It is also trivial to get a good fit with the dead leaves model—as we always can choose the gray levels of the templates according to the empirical distribution of log-contrast in natural data.
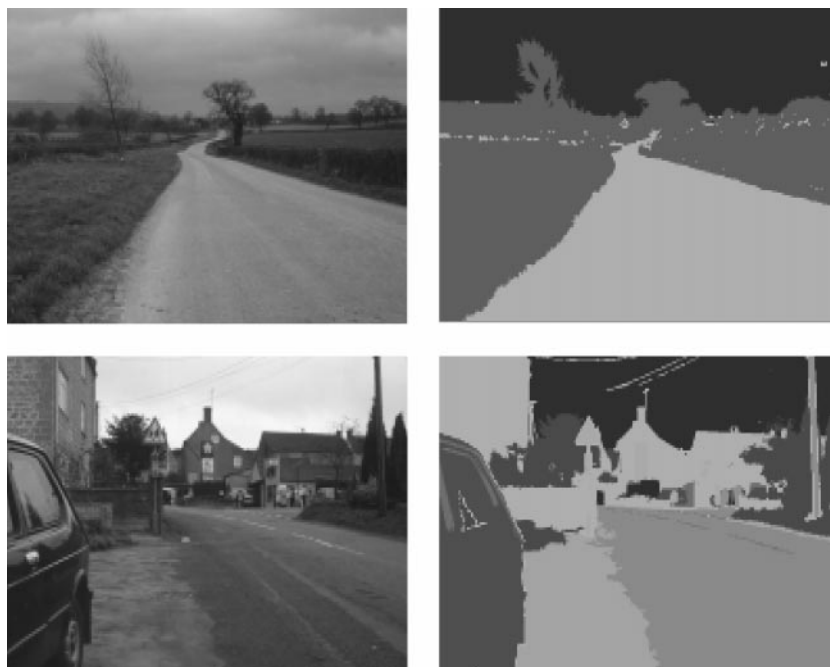

Figure 4. Two sample images from the British Aerospace database and their segmentations into pixels belonging to different categories of scenes (as defined in Table 1).

3.2.   Derivative Statistics and Scaling

In this section, we study the difference of adjacent pixel values, and how well histograms of this difference or derivative statistic scale. More precisely: For each log-contrast image I, we define a scaled-down image $I^{(N)}$ by computing the average of pixel values in disjoint $N \times N$ blocks. The statistic we investigate is the horizontal derivative, which for scale N, is given by

$$D = I^{(N)}[i, j + 1] - I^{(N)}[i, j]. \qquad (23)$$

If natural images are fully scale invariant, $D^{(N)}$ should have the same distribution for different values of N. When we measure departure from scale invariance, we look at *both* the change of the shape of the histograms after rescaling and the change in the standard deviation of D.

3.2.1. Generic Natural Scenes.

First, we note that for *large* databases of natural image, the derivative statistic D is a surprisingly stable statistic—consistent across different datasets. We also get an excellent fit of the probability density function of D to a generalized Laplace distribution[5]

$$f(x) \propto e^{-|x/s|^\alpha}, \qquad (24)$$

where s and α are parameters related to the variance and kurtosis. Figure 5(a) (bottom) shows three curves

which are almost the same. These corresponds to: (1) the log-histogram of D, at scale 2, for van Hateren's database, (2) the corresponding log-histogram for the unsegmented British Aerospace database, and (3) a fit to a generalized Laplace distribution with $\alpha = 0.68$.

The second observation is that the histograms of D for generic images scale almost fully: Fig. 5(a) (top) shows four very similar curves. These correspond to the log-probability distribution of $D^{(N)}$, for $N = 1, 2, 4, 8$ in van Hateren's database (the curves have been shifted vertically for visibility). Except at the first scale (*solid line*), the histograms lie almost completely on top of each other.

3.2.2. Dead Leaves Model for Generic Scenes.

We now compare the marginal distribution of D for natural data, with results from the occlusion model (Fig. 5(b)). We are able to reproduce the *main features* seen in natural data: (1) *the singularity at* 0,[6] which corresponds to large regions with no contrast variation, and (2) the *heavy tails*, which correspond to edges in the images.

The tails are slightly less concave than for natural data; but as before, we get an excellent fit to a generalized Laplace distribution (see the two overlapping curves in Fig. 5(b), bottom): In this case, a Laplace distribution with $\alpha = 0.78$ ($\alpha = 0.68$ for natural data).

Figure 5(b), top part, shows that the synthetic images are close to scale invariant. The *shape* of the log-histograms of D remains the same after scaling,
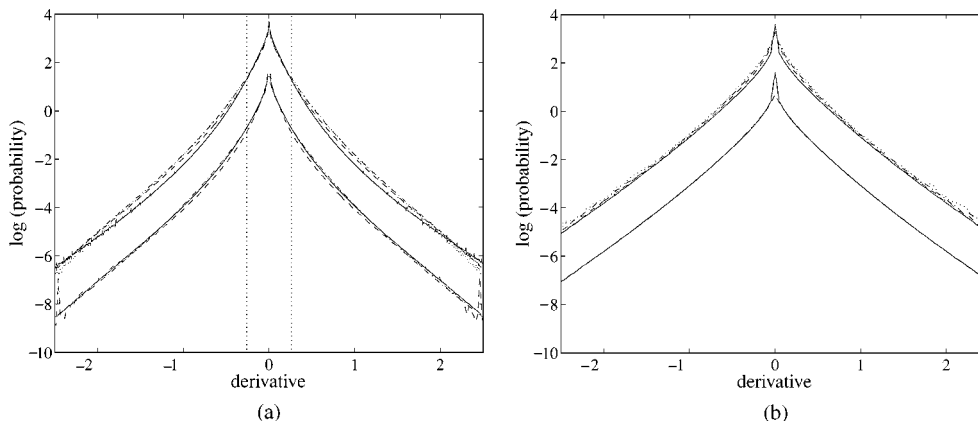

Figure 5. Derivative statistics and scaling. a: Natural images. The difference statistic D between adjacent pixels is amazingly stable, both across different databases and different scales. The bottom three curves, which are overlapping, correspond to the log-histograms of D for van Hateren's database (*solid*), the British Aerospace database (*dashed*), and a fit to a generalized Laplace distribution with parameter $\alpha = 0.68$ (*dotted*). The top four curves (which have been shifted vertically for visibility) correspond to the log-probability of $D^{(N)}$ for scales $N = 1$ (*solid*), 2 (*dashed*), 4 (*dash-dotted*), and 8 (*dotted*). b: Synthetic images. The distribution of D for the model (*solid*, bottom part) can be approximated with a generalized Laplace distribution with parameter $\alpha = 0.68$ (*dashed*, bottom part). After a contrast normalization, the log-histograms of $D^{(N)}$ at scales $N = 1, 2, 4,$ and 8 lie almost on top of each other (see the top four curves which have been shifted for visibility).

but the standard deviation of the derivative decreases somewhat with N (a factor of $N^{0.1}$)—this is because of the anomalous dimension η in the system (see Section 2.4.3). In the figure we have made a contrast renormalization (by dividing out the change in the standard deviation)—the histograms of D for 4 different scales then lie almost completely on top of each other.

3.2.3. Different Types of Natural Scenes.
The segmentation of the British Aerospace database makes it possible to study different types of natural scenes.

To simplify our statements, we merge categories (see Table 1 for definitions and frequencies) that have similar statistics, and get four larger groupings: "sky" (category 1), "vegetation" (categories 2, 7), "road" (category 4), and "man-made" (categories 6, 8, 9, 11).

Figure 6 shows the histograms of the derivative $D^{(N)}$ for these new categories, after a contrast renormalization. For the categories "vegetation", "man-made", and "sky", the *shapes* of the histograms at different scales are about the same (except at the first scale)—this is consistent with the assumption that I scales according to Eq. (21). For the sky category, the *shape* seems to depend on N.

To study the scaling properties more in detail, we plot the logarithm of the standard deviation against the logarithm of N, and perform a linear regression. This is equivalent to fitting a power spectrum of the form $C/f^{(2-\eta)}$: The slope gives us an estimate of half of the "anomalous dimension" η, or if Eq. (21) is valid, the scaling exponent ν.

From Fig. 6 and the linear regression, we conclude that, although natural images as a single ensemble are very nearly scale invariant with a derivative histogram described by a concave-shaped generalized Laplace distribution (see Section 3.2.1), major differences exist between different categories, or different parts of images:
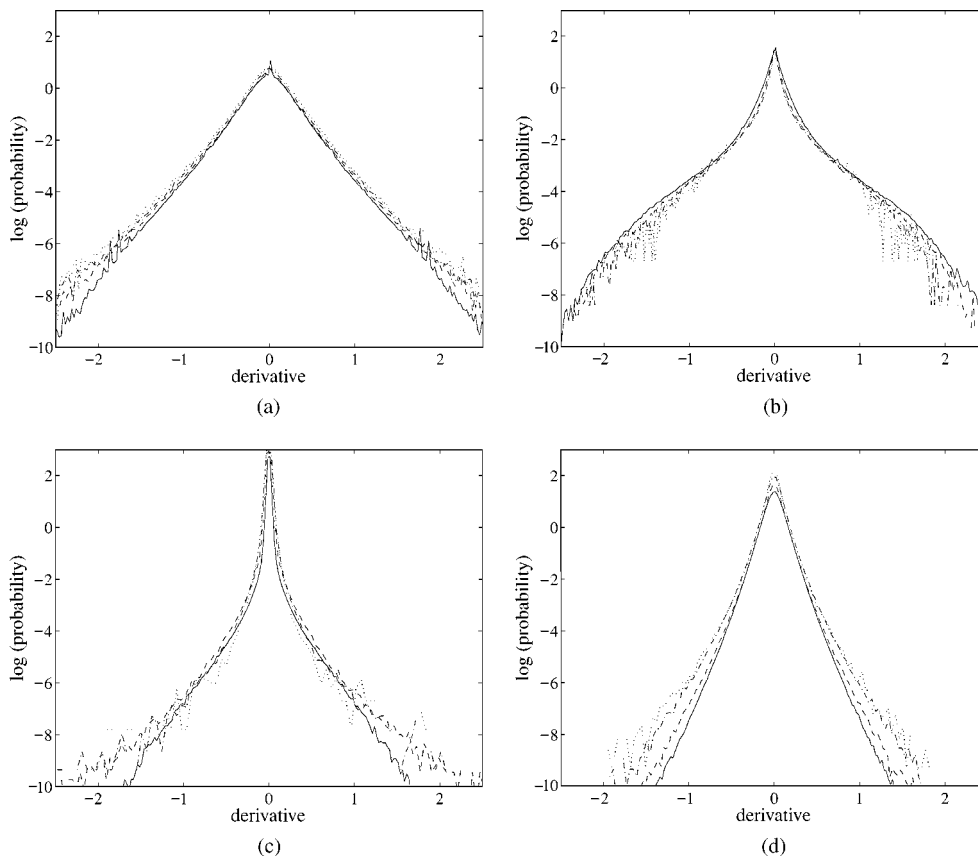

Figure 6. Derivative statistics and scaling for different types of natural scenes after a contrast renormalization. a: Vegetation. b: Man-made. c: Sky. d: Roads. The plots show the log (probability) of $D^{(N)}$ for scales $N = 2$ (*solid*), 4 (*dashed*), 8 (*dash-dotted*), and 16 (*dotted*).

1. For the vegetation category, the log-histograms of D have relatively straight tails. In terms of power spectrum fall-off, we get $C/f^{1.8}$—which is similar to Ruderman's and Bialek's results (Ruderman and Bialek, 1994).

2. For the man-made category, the log-histograms of D have heavy "shoulders" with a convex shape. The power spectrum scales like $C/f^{2.3}$.

3. For the sky category (including clouds), the density of the distribution for D is mainly concentrated around 0. The power spectrum scales like $C/f^{1.0}$, i.e. the category is intermediate between white noise and the standard category with $\eta \approx 0$.

4. For the road category, the log(histograms) are slightly concave. The power spectrum scales roughly like $C/f^{1.4}$.

3.2.4. Different Versions of the Dead Leaves Model.

There are many ways one can vary the dead leaves model so that it fits different types of natural scenes. One can use different kinds of templates, and one can vary the size cut-offs r_{min} and r_{max} to fit the level of clutter and the anomalous dimension in the images. Here we show two different versions of the occlusion model that will be compared to the categories "vegetation" and "man-made" in the previous section:

For *vegetation-like*, we generate 500 relatively cluttered images with 256×256 pixels (see Fig. 7(a)). We use elliptic primitives of random orientation. The length L of the major axis is distributed according to $1/L^3$, where $2 \leq L \leq 1024$ pixels, and the ratio between the lengths of the minor and major axes is a uniform deviate between $1/8$ and $1/2$. We color the
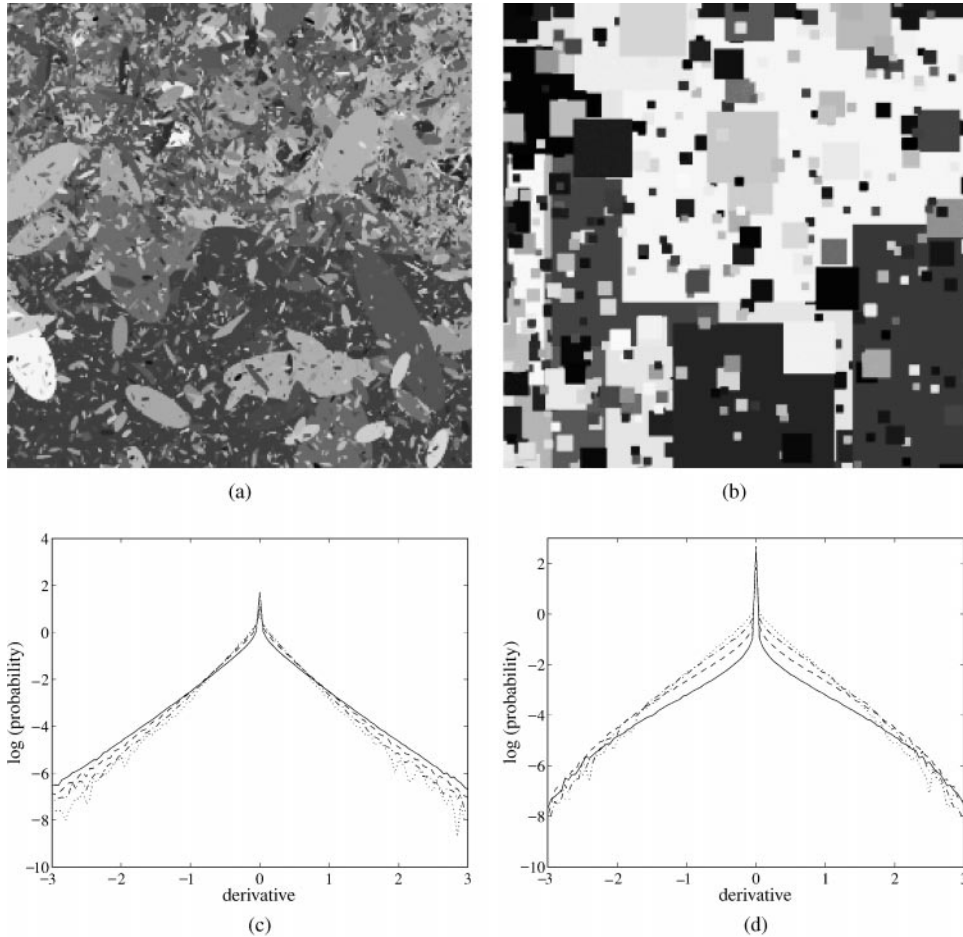

Figure 7. Different versions of the dead leaves model, and their derivative statistics and scaling (no renormalization). a: "Vegetation-like"—Computer-simulated sample with relatively high clutter and elliptic primitives. b: "Man-made-like"—Computer-simulated sample with low clutter and square primitives. c: Log-histogram of $D^{(N)}$ for "vegetation-like" and scales $N = 2$ (*solid*), 4 (*dashed*), 8 (*dash-dotted*), and 16 (*dotted*). d: Corresponding log-histograms for "man-made-like."

ellipses according to a "double-exponential" distribution with mean 0 and variance 1, but add some Gaussian noise for smoothing.

For *man-made-like*, we generate 500 cleaner-looking images with square primitives (see Fig. 7(b)) The length s of the side of the square is distributed according to $1/s^3$, where $2 \leq s \leq 2048$ pixels. We color the squares according to a uniform intensity distribution with mean 0 and variance 1. As before, we add some Gaussian noise for smoothing.

Figures 7(c) and (d) summarize the results. The two plots show the log(probability) of the derivative $D^{(N)}$ at scales $N = 2, 4, 8,$ and 16—for "vegetation-like" and "man-made-like", respectively. Compare the plots to Fig. 6(a) and (b) for the vegetation and man-made categories. Furthermore, linear regression gives the scaling exponent -0.09 for the occlusion model with high clutter and elliptic primitives, and the exponent 0.13 for the model with low clutter and square primitives; Compare this to the "vegetation" and "man-made" categories where the exponents are -0.11 and 0.15, respectively.

3.3. Joint Statistics of Haar Wavelet Responses

In this section, we look at the distribution of 2×2 blocks of pixels. If $a_{i,j} = I[i_0 + i, j_0 + j]$, where $0 \leq i, j \leq 1$, is such a block, then we look at the distribution of $(a_{00}, a_{01}, a_{10}, a_{11}) \in \mathbb{R}^4$. The mean is a relatively un-informative statistic, hence we look at the projection to $\mathbb{R}^3$ given by the 3 Haar wavelet responses:

$$cH = \frac{1}{2}(a_{00} + a_{01} - a_{10} - a_{11})$$

$$cV = \frac{1}{2}(a_{00} - a_{01} + a_{10} - a_{11}) \qquad (25)$$

$$cD = \frac{1}{2}(a_{00} - a_{01} - a_{10} + a_{11})$$

which we call the horizontal, vertical and diagonal filters, respectively. These Haar wavelets show clearly, though only partially, the very specific nature of local statistics (or textons) in natural images.

We use the same definitions as in (Buccigrossi and Simoncelli, 1999), to describe the relative positions of wavelet coefficients: We call the coefficients at adjacent spatial locations in the same subband *brothers* (left, right, upper, or lower—depending on their relative positions; note that the basis functions are disjoint), and we call the coefficients at the same position, but different orientations, *cousins*.

Figure 8 shows the joint wavelet statistics for natural images. We have plotted the contour levels for the joint histograms of some different coefficient pairs (at scale $N = 2$). A common feature for all the histograms is that a cross-section through the origin has a peak at the center and long tails—the shape is very similar to the derivative density function in Section 3.2. More complicated structures also show up in the polyhedra-like contour-level curves: The corners and edges (in the level curves), which are sometimes rounded and sometimes cuspidal, reflect typical local features in the images. For example, in the plot for the horizontal (cH_1)—left brother (cH_2) pair (Fig. 8(c)), the edges along the diagonal $cH_1 = cH_2$ indicate the frequent occurrence of extended horizontal edges in natural scenes. In the plot for the horizontal (cH)—diagonal (cD) pair (Fig. 8(b)), the cusp at $cH = cD$ corresponds to the T-junction $a_{01} = a_{11}$ and $a_{00} \neq a_{10}$; the cusp at $cH = -cD$ corresponds to the T-junction $a_{00} = a_{10}$ and $a_{01} \neq a_{11}$.

In Fig. 9, we have plotted the contour-level curves of the corresponding joint histograms for the synthetic images. We see that the plots are very similar to those in Fig. 8: The corners and edges all appear in the right places, and the shape of the curves are also almost the same as those in Fig. 8. *This is a strong indication that the occlusion model captures much of the local image structure in natural data.* Compare this to the Gaussian model, for example, which totally fails here—all contours in the wavelet domain are elliptic for these images. The random wavelet expansions (Eq. (4)) can reproduce *some* of the polyhedra-like contours for images with *low* levels of clutter (i.e. few objects), but the edges become more rounded, and the contours more elliptic for higher levels of clutter.

In Fig. 9, we also see some smaller differences between natural and computer-simulated images—for example, in the $cH - cV$ plot (Fig. 9(a)) and the cH_1-cH_2 plot (Fig. 9(c)). In natural scenes, there's a strong bias in the horizontal and vertical direction, because of tree trunks, the horizon, buildings etc. The computer-simulated images, on the other hand, have disk-like primitives only, and are also rotationally invariant. This leads to more rounded shapes in Fig. 9(a) and (c).

3.4. Long-Range Covariances

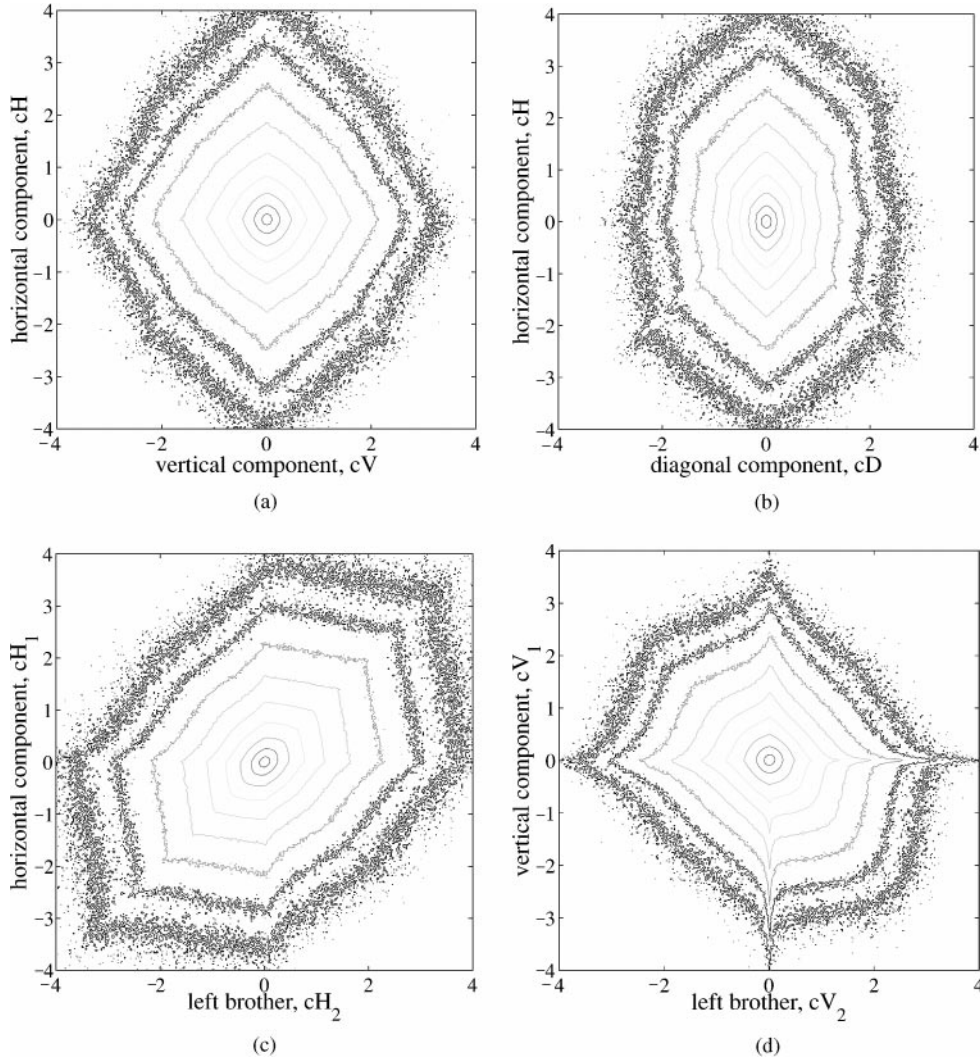So far we have only looked at small-scale statistics, i.e. statistics for single pixels or nearest-neighbor pixels.

Figure 8. Joint statistics in the wavelet domain for natural images. The contour plots show the log(probability) distributions for different waveletcoefficient pairs at scale 1. a: Horizontal (cH) and vertical (cV) components. b: Horizontal (cH) and diagonal (cD) components. c: Horizontal component (cH_1) and its left brother (cH_2). d: Vertical component (cV_1) and its left brother (cV_2).

Here we extend our comparison of natural and simulated images to long-range statistics.

The simplest long-range statistic is probably the correlation between two pixels in, for example, the horizontal direction. We can calculate the covariance function (Eq. (18)) or, alternatively, the variance of the difference of two pixel values, i.e.

$$V(x) = \langle |I(x) - I(0)|^2 \rangle, \qquad (26)$$

where $\langle \cdot \rangle$ denotes an average over all images. The latter formulation is a good choice when images are offset by an unknown constant. The two functions are otherwise equivalent as

$$V(x) + 2C(x) = \text{constant.} \qquad (27)$$

In Huang and Mumford (1999), Huang shows that the "difference function" for natural images (in a fixed direction) is best modeled by

$$V(x) = a_1 + a_2 \cdot x^{-\eta} + a_3 \cdot x, \qquad (28)$$

where x is the separation distance between two pixels, and a_1, a_2, a_3 are constants. The power-law term dominates the short-range behavior, while the linear term dominates at large pixel-distances. The linear term
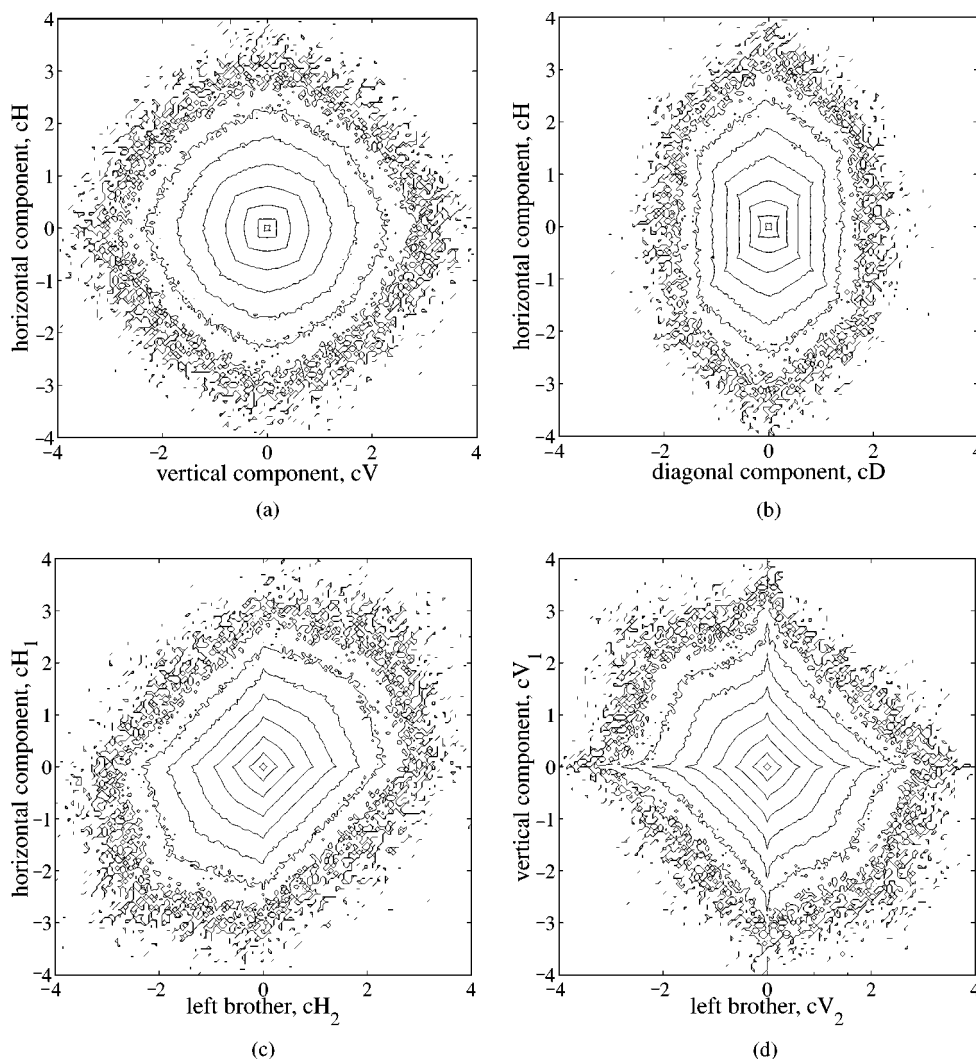
Figure 9. Joint statistics in the wavelet domain for synthetic images. Compare the contour plots with those in Fig. 8 for natural images: The similarities indicate that the dead leaves model captures much of the local image structure of natural data.

indicates that, while the scale-invariance property holds almost exactly locally (i.e. for filters with small supports), there are systematic deviations from scale invariance on a large scale. This may be due to the presence of sky in the images.

In the synthetic images, the linear term is absent. The difference function for the simulated images is best modeled by a power-law

$$V(x) = b_1 + b_2 \cdot x^{-\eta} \qquad (29)$$

for both short and large pixel-distances (see Section 2.4); In the Fourier domain, this corresponds to a power spectrum of the form $1/k^{2-\eta}$.

Figure 10 maybe shows this clearer. Here we have a log-log plot (base 2) of the derivative of $V(x)$ for natural images (*solid line*) and computer-simulated dead leaves images (*dashed line*). A power-law behavior according to Eq. (29), would lead to a straight line with slope $(1 + \eta)$.

The fit between natural images (*solid line*) and synthetic images (*dashed line*) is good in the region where both curves are relatively straight, i.e. for $2 < \log_2 x < 5$, or distances between 2 and 32 pixels. The slopes here are -1.19 ($\eta = 0.19$) for natural data, and -1.16 ($\eta = 0.16$) for synthetic images (cf. $\eta = 0.14$ for the continuous model in Fig. 2(b)). For natural images, however, the curve turns and becomes
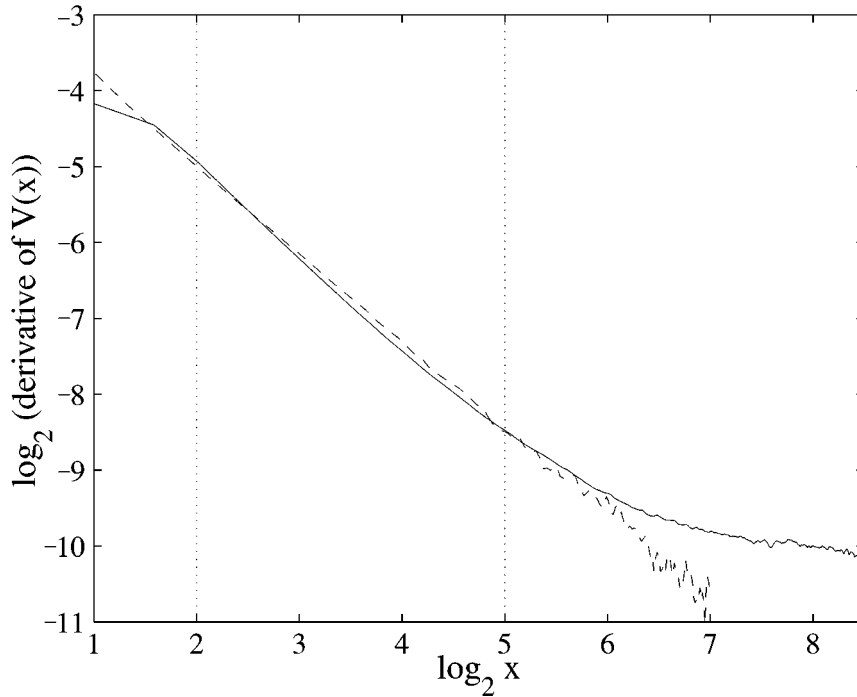
Figure 10. Convariances. The figure shows a log-log plot (base 2) of the derivative of the difference function $V(x)$ for natural images (*solid line*) and synthetic images (*dashed line*).

almost horizontal for large distances (about $1/10$ of the image size); this indicates a linear term in the difference function.

3.5. Two-Pixel Co-occurrence Statistics

In this section, we compare the complex behavior of the *full* co-occurrence statistics in the occlusion model and natural data—previously, we have been comparing the auto-correlation or variance of the difference of two pixel values, and the full histogram of differences in adjacent pixel values. Below, we also test the basic assumption in the occlusion model that natural images can be segmented into two parts: "same object" and "different objects".

In the calculations, we have symmetrized the data so that $\Pr\{I(0) = a\} = \Pr\{I(0) = -a\}$. This is to take away the bias towards high intensity values, caused by the sky in natural images.

3.5.1. Bivariate Fit for Computer-Simulated Images. A Modified Occlusion Model.   As mentioned before, the continuous occlusion model (Eq. (15)) gives good predictions for high-resolution images, but dead leaves

images with subpixel resolution by means really give the best fit to natural data (derivative statistics, scaling, joint statistics of Haar wavelet coefficients, etc.). Thus, one needs to examine how much the formula for the co-occurrence function $K(a, b, x)$ is changed by subpixel averaging.

As an ansatz for the co-occurrence function of two pixels in dead leaves images *with smoothing*, we write (cf. Eq. (15))

$$
\begin{aligned}
\tilde{K}(a, b, x) &= \Pr\{I(0) = a, I(x) = b\} \\
&= [1 - \lambda(x)] \cdot q(a) \cdot q(b) \\
&\quad + 2\lambda(x) \cdot h_x(a + b) \cdot g_x(b - a). \quad (30)
\end{aligned}
$$

As before, we assume that different objects are statistically independent. The first term in Eq. (30) corresponds to pixels on *different* objects, and is equivalent to the product in Eq. (15) of the pdf:s of single-pixel intensities. In the second term—which represents pixels on the *same* object—we replace the previous delta function with a new probability density function g_x that is highly concentrated around 0. We also introduce a new probability density function h_x which is similar to the function f_x for single-pixel intensities.

(The subindex x in h_x and f_x indicates that the functions may depend on the distance x between the pixels.)

For convenience, we make a variable substitution

$$u_x = I(x) + I(0)$$
$$v_x = I(x) - I(0) \tag{31}$$

The joint pdf of the new variables is given by

$$Q(u, v, x) = \frac{1}{2}[1 - \lambda(x)] \cdot q\left(\frac{u+v}{2}\right) \cdot q\left(\frac{u-v}{2}\right)$$
$$+ \lambda(x) \cdot h_x(u) \cdot g_x(v). \tag{32}$$

We now fit the empirical joint pdf $Q^{\text{synth}}(u, v, x)$ of computer-simulated images to the expression in Eq. (32); As a best-fit criterium we minimize the Kullback-Leibler distance. For fixed x, the bivariate fit gives us a value for $\lambda(x)$, and expressions for the 1D functions h_x, g_x and q.

In Fig. 11(a), we see how $\lambda(x)$ varies with the distance x between the pixels. We compare λ-values for $x = 1, 2, 4, 8, 16, 32, 64, 128$ from the bivariate fit (*rings*) to the analytically calculated $P_{\text{same}}(x)$ in the continuous model (*solid line*). The figure shows that $\lambda(x) > P_{\text{same}}(x)$ for fixed x; that is the probability that two points belong to the *same* object is larger for the smoothed dead leaves images than for the continuous images. This is also what we expect of a model where small regions of intensity variations are considered to be parts/textures of larger objects, rather than separate objects.
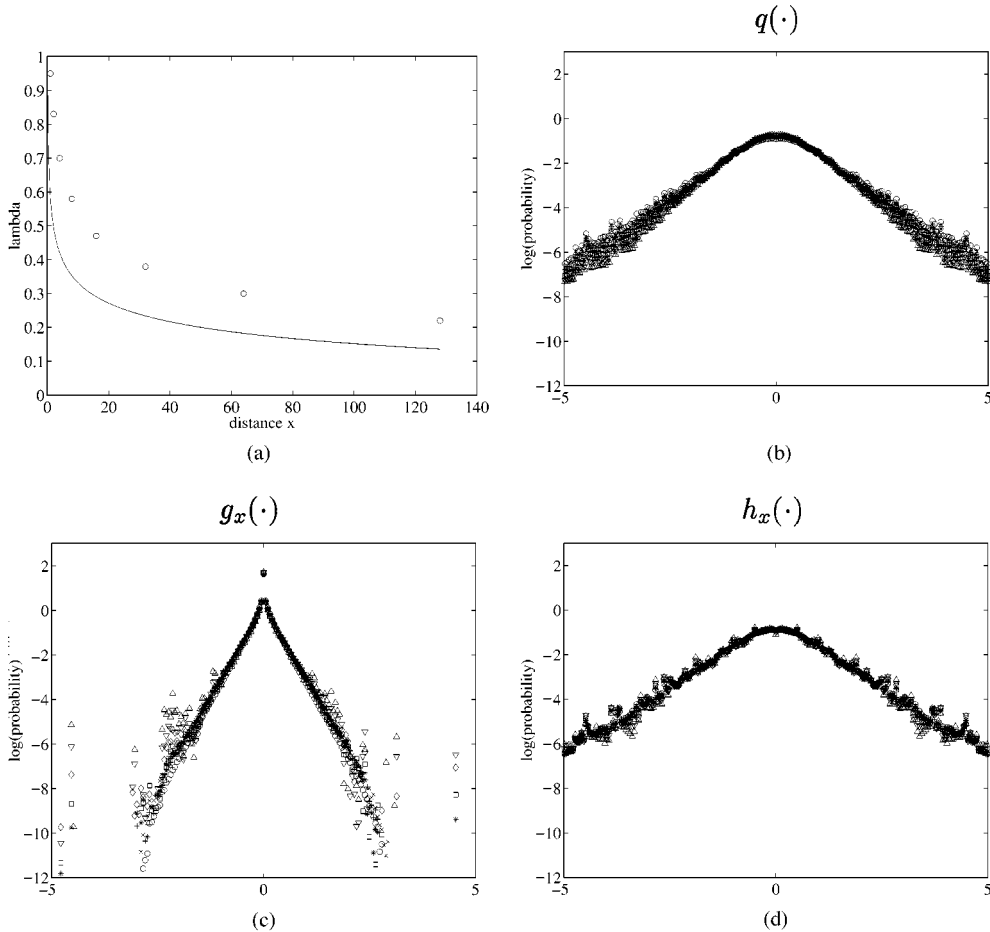

Figure 11. Functions that give the best bivariate fit of computer-simulated images (to Eq. (32)) at different distances x. a: The *rings* represent the λ-values from a bivariate fit at $x = 1, 2, 4, 8, 16, 32, 64, 128$, and the *solid line* represents $P_{\text{same}}(x)$ for a continuous model with uniformly colored objects. b–d: The 3 plots show the 1D functions q, g_x and h_x from the bivariate fit in (a). The functions depend very little on x; h_x and q are also almost the same. Legend: $\circ\, x = 1$; $\times\, x = 2$; $+\, x = 4$; $\star\, x = 8$; $\square\, x = 16$; $\diamond\, x = 32$; $\triangledown\, x = 64$; $\triangle\, x = 128$.

$Q^{\text{synth}}(u, v, x)$ for $x = 2$ Bivariate fit with $\lambda = 0.83$

$Q^{\text{synth}}(u, v, x)$ for $x = 16$ Bivariate fit with $\lambda = 0.47$

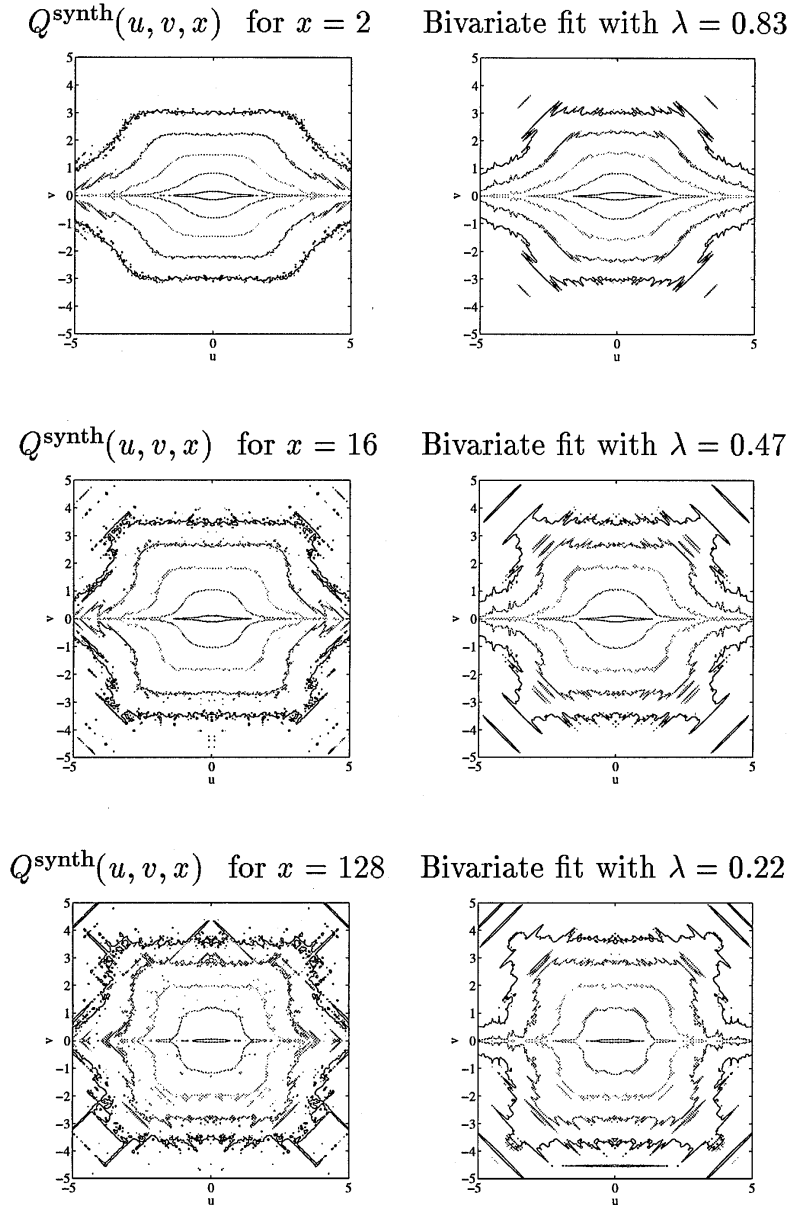$Q^{\text{synth}}(u, v, x)$ for $x = 128$ Bivariate fit with $\lambda = 0.22$

Figure 12. Co-occurrence function $Q^{\text{synth}}(u, v, x)$ and best bivariate fit (to Eq. (32)) for numerically simulated dead leaves images. *Left*: The plots show the contour levels $(-9, -7, -5, -3, -1)$ for the logarithm of the joint pdf of the sum u and difference v for pixels a distance $x = 2, 16$ and 128 pixels apart (horizontal direction). *Right*: The plots show the corresponding contour levels for the best fit to Eq. (32).

The 1D functions q, g_x and h_x that we have used for the bivariate fit at $x = 1, 2, 4, 8, 16, 32, 64, 128$ are plotted in Fig. 11(b)–(d). Note that the functions depend very little on x; h_x and q are also almost the same.

Figure 12 shows the results of a bivariate fit for $x = 2$ (*top*), $x = 16$ (*center*), and $x = 128$ (*bottom*). In the left column, we have the contour levels of Q^{synth} (u, v, x); The right column shows the best fit of the data to Eq. (32). The agreement between the data and the fit is very good, which shows that we can use Eq. (30) or Eq. (32) to accurately describe the two-pixel statistics of dead leaves images with subpixel resolution by means.
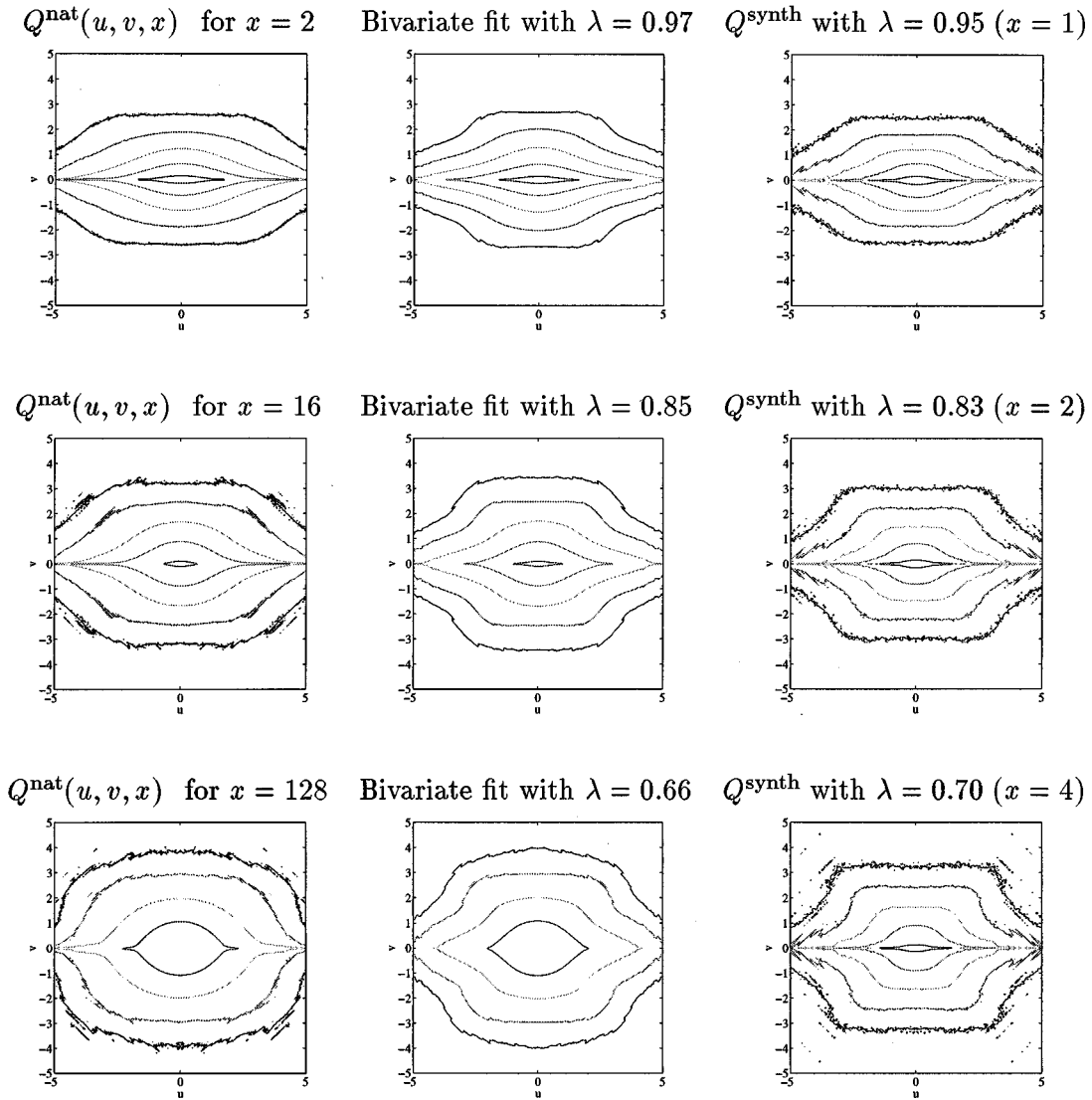
$Q^{\text{nat}}(u, v, x)$ for $x = 2$ Bivariate fit with $\lambda = 0.97$ Q^{synth} with $\lambda = 0.95$ $(x = 1)$

$Q^{\text{nat}}(u, v, x)$ for $x = 16$ Bivariate fit with $\lambda = 0.85$ Q^{synth} with $\lambda = 0.83$ $(x = 2)$

$Q^{\text{nat}}(u, v, x)$ for $x = 128$ Bivariate fit with $\lambda = 0.66$ Q^{synth} with $\lambda = 0.70$ $(x = 4)$

Figure 13. The plots show the contour levels $(-9, -7, -5, -3, -1)$ for the logarithm of the joint pdf of the sum u and difference v for pixels a distance $x = 2, 16$ and 128 pixels. *Left*: $Q^{\text{nat}}(u, v, x)$ for natural images. *Center*: The best fit of $Q^{\text{nat}}(u, v, x)$ to Eq. (32). *Right*: $Q^{\text{synth}}(u, v, x)$ that correspond to similar λ-values.

3.5.2. Bivariate Fit for Natural Images.

Finally, we test how well the modified mixture model fits natural data in van Hateren's database. Figure 13 (left and center columns) hows three examples of a bivariate fit to Eq. (32): The distances between the pixels are $x = 2$ (*top*), $x = 16$ (*bottom*), and $x = 128$ (*bottom*). Although the model is simple, the fit is good—that is, we are able to write the full two-pixel co-occurrence statistic of natural images as a mixture ("same" and "different" objects). *This is a direct confirmation of the accuracy of the dead leaves model.*

In the current version of the dead leaves model, different objects are *independent*. This seems to be a reasonable first approximation. The "objects" defined by such a model, however, become very complex and large, because of the dependencies between different parts of natural scenes. In Fig. 14, we plot the 1D functions h_x, g_x and q that we get from a best fit to Eq. (32) at separation distances $x = 1, 2, 4, 8, 16, 32, 64, 128$. As before, the functions h_x and q are almost the same, and depend little on x. However, the function g_x, which is related to the *difference in intensity* of pixels on the
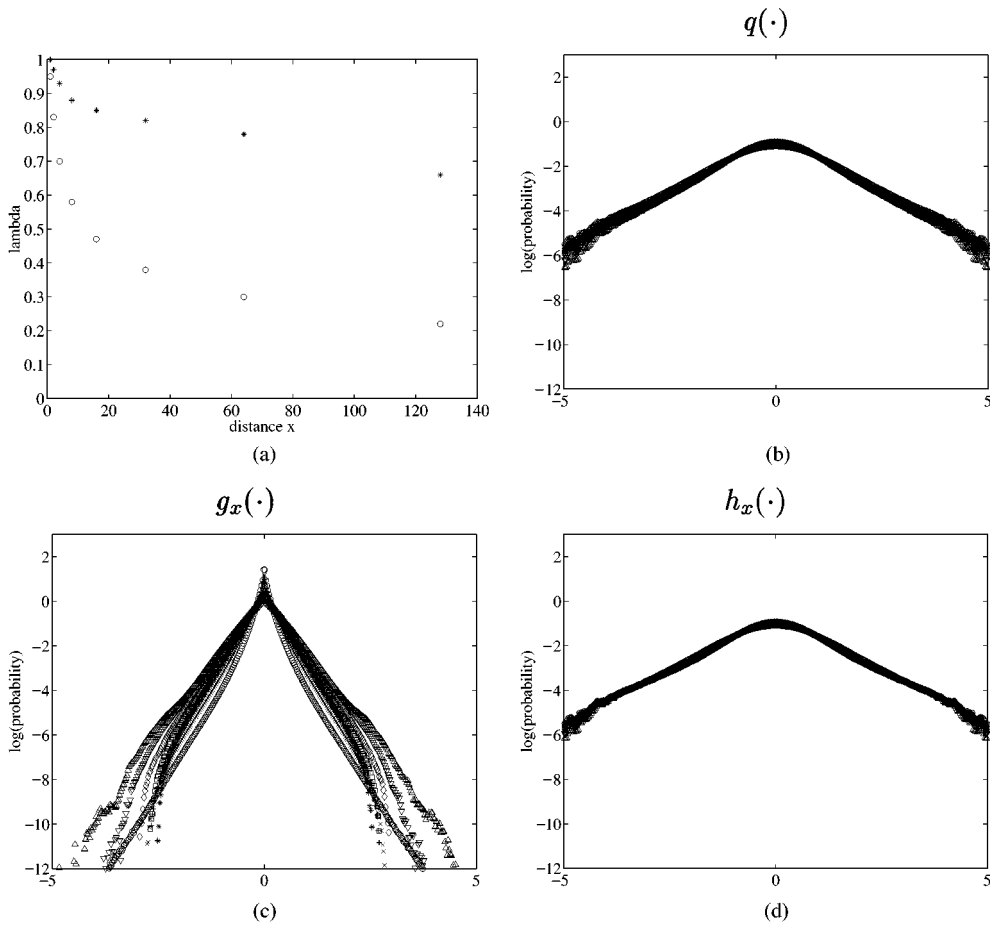
$$q(\cdot)$$

(a)

(b)

$$g_x(\cdot) \qquad\qquad h_x(\cdot)$$

(c)

(d)

Figure 14. Functions that give the best bivariate fit of natural images (to Eq. (32)) at different distances x. a: The *stars* represent the λ-values from a bivariate fit of natural images at $x = 1, 2, 4, 8, 16, 32, 64, 128$; These are compared to the λ-values for synthetic images (*rings*). b–d: The 3 plots show the 1D functions q, g_x and h_x for the bivariate fit of natural images. Note that the function g_x, which is related to the intensity difference of pixel values on the same object, becomes wider for larger values of x. This indicates that the "objects" from the bivariate fit have parts, and parts of parts (see text). Legend: $\circ\, x = 1$; $\times\, x = 2$; $+\, x = 4$; $*\, x = 8$; $\square\, x = 16$; $\diamondsuit\, x = 32$; $\triangledown\, x = 64$; $\triangle\, x = 128$.

same "object", depends strongly on x: The function is relatively concentrated around 0 for small x, but becomes wider for larger values of x. The dependency of g_x on x indicates that the "objects" have *parts*, and *parts of parts*. Assume, for example, that a whole region of a forest is classified as *one* "object". Because the forest divides into trees, and trees, for example, have leaves, we would expect g_x to become successively narrower for smaller values of the pixel distance x. This is also what we found for the bivariate fit of natural images, but not for the bivariate fit of the synthetic images which lack this "object tree structure".

Furthermore, we note that the contour plots for both natural (Fig. 13, left) and synthetic images (Fig. 12, left) become more rectangular for increasing values

of x—as the probability of the pixels being on different "objects" increases. For *fixed* x, $\lambda(x)$ for natural data is considerably larger than $P_{\text{same}}(x)$ for computer-simulated images (see Fig. 14(a); also compare to Fig. 10(a) where, for fixed x, the variance $V(x)$ of the difference of two pixel values is much larger for synthetic than for natural data). Again, this is an indication that a more realistic variant of the occlusion model should include objects with parts. However, if we compare contour plots of $Q^{\text{synth}}(u, v, x)$ and $Q^{\text{nat}}(u, v, x)$ that correspond to *similar* λ-values—compare the left and right columns in Fig. 13—we get an excellent fit between the dead leaves model and natural data, for a range of different pixel separation distances.

3.6. Comparison to Natural Data: Summary of Results

– *Single-Pixel Statistics.* A quantitative fit is here possible. We can color the templates of a dead leaves model according to the observed contrast distribution of natural data.
– *Derivative Statistics.* The smoothed dead leaves images are able to simulate the generalized Laplace distribution, observed in natural data, with a singularity at 0 and long concave tails. (Both Gaussian models and scale-invariant additive models fail here.)
– *Scaling.* The derivative statistic of generic images is almost fully scale invariant. The dead leaves images scale after a contrast renormalization.
– *Joint Statistics of Haar Wavelet Responses.* The contour plots of joint wavelet responses are highly non-ellipsoidal and clearly show the non-Gaussian nature of contrast-normalized images. The model with disks gives a good qualitative fit to natural data, but the detailed structures of the contour plots are different due to the simplifications in the model.
– *Covariance Statistics.* For natural images, a power-law dominates the short-range behavior of the covariance function, while a linear term dominates the long-range covariances. For the dead leaves model with cubic law of sizes, we have an approximate power-law for all distances.
– *Two-Pixel Co-Occurrence Statistics.* For natural images, the bivariate statistics of two pixels approximately fit a parametric form suggested by a dead leaves model with smoothing. The expression implies a mixture nature—"same" vs "different" objects—where the intensities of pixels on different objects are independent, and the sum and difference of intensities are independent for pixels on the same object.

4. Discussion

We have investigated whether a simple version of Matheron's dead leaves model can simulate the statistics of natural images better than Gaussian models and additive stochastic models. In most of the analysis, we assume statistically independent disk-like templates with no contrast variation, and with sizes distributed according to a cubic power law. The building stones of our model are (1) the concept of structured objects (2) approximate scale invariance, and (3) occlusions.

We have compared the statistics of simulated dead leaves images with the empirical statistics of two large databases of natural images. The study includes, for example, the scaling of pixel difference histograms, the full co-occurrence statistics of two pixels, and the joint statistics of pairs of Haar wavelet responses (see Section 3.6 for a summary of the results). In all these cases, we found an excellent qualitative agreement, and good quantitative similarity, between the occlusion model and natural data. At this point, this is the only model which comes close to duplicating the elementary statistics of natural images. The results indicate that this family of models (with structured objects and occlusions) is relevant to natural images—and perhaps also necessary for an understanding of the underlying structures of images. We believe that continued work in this direction—such as a quantitative fit of the dead leaves model to real images; see for example the methods in Zhu and Guo (2000) and Grenander and Srivastava (2000)—will benefit many computer vision application; in particular, clutter removal and object/background segmentation in a natural scene environment of high variability.

We have furthermore found evidence that the current version of the dead leaves model can be further improved by (1) using suitable primitives, (2) adding textures to the primitives, and (3) taking the hierarchical structure of objects into account. For a more realistic variant of the occlusion model, we may also need to generate objects in groups or near the surface of a parent, as in a random branching process.

Appendix

As before, we consider a dead leaves model with disk-like template and a $1/r^3$ size distribution, where the disk radius $r_{\min} \leq r \leq r_{\max}$. The results in Appendix A and B are specific to this choice. (By using the formalism of mathematical morphology (Serra, 1982), it is however possible to derive similar expressions for more general shapes.)

A. Derivation of $P_{same}(x)$ for Dead Leaves Model with Finite Cut-offs in Sizes

Below we follow the calculations in (Ruderman, 1997) of the probability $P_{\text{same}}(x)$ that two given points a distance x apart belong to the same object.

Ruderman has previously shown that at stationarity

$$P_{\text{same}}(x) = \frac{p_2(x)}{p_1(x) + p_2(x)}, \qquad (33)$$

where $p_2(x)$ is the probability that the front-most object, which is not occluded by any other objects, contains both points in the pair, and $p_1(x)$ is the probability that the object contains exactly one of the two points. Furthermore, he has derived an expression for the conditional probability

$$g(x, r) = Pr\{\mathbf{x}_2 \in A \mid \mathbf{x}_1 \in A; \|\mathbf{x}_1 - \mathbf{x}_2\| = x\}, \quad (34)$$

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

$$\tilde{g}(0 \le \xi \le 2) = \frac{2}{\pi}\left[\arccos\left(\frac{\xi}{2}\right) - \frac{\xi}{2}\sqrt{1 - \left(\frac{\xi}{2}\right)^2}\right]. \tag{35}$$

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

In our model, the radii of the disks are distributed according to $1/r^3$ where $r_{\min} \le r \le r_{\max}$.[7] We then have

$$p_1(x) = 2\int_{r_{\min}}^{r_{\max}} [1 - g(x, r)] p(r)\, dr$$
$$\hspace{6cm} (36)$$
$$p_2(x) = \int_{r_{\min}}^{r_{\max}} g(x, r) p(r)\, dr,$$

where

$$p(r)\, dr \approx \frac{dr}{r \ln\left(\frac{r_{\max}}{r_{\min}}\right)} \tag{37}$$

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. (33), we get

$$P_{\text{same}}(x) = \frac{B(x)}{2\ln\left(\frac{r_{\max}}{r_{\min}}\right) - B(x)}, \tag{38}$$

where

$$B(x) = \int_{\frac{r_{\min}}{x}}^{\frac{r_{\max}}{x}} \tilde{g}\left(\frac{1}{u}\right) \frac{du}{u}. \tag{39}$$

and $\tilde{g}(\xi)$ is given by Eq. (35).

To simplify the integral above, we approximate the function $\tilde{g}(\xi)$ in Eq. (35) with a third-order polynomial $\tilde{g}_{\text{poly}}(\xi)$. The best fit gives $(0 \le \xi \le 2)$

$$\tilde{g}(\xi) \approx \tilde{g}_{\text{poly}}(\xi) = a_3\xi^3 + a_2\xi^2 + a_1\xi + a_0 \quad (40)$$

with coefficients $a_0 = 1.0$, $a_1 = -0.61$, $a_2 = -0.051$, and $a_3 = 0.052$.

Inserting Eq. (40) into Eq. (39) leads to a numerical estimate of $B(x)$, and thus a numerical expression for P_{same} in Eq. (38). For $2r_{\min} \le x < 2r_{\max}$,[8]

$$B(x) \approx \frac{a_3}{3}(8 - u^3) + \frac{a_2}{2}(4 - u^2)$$
$$\hspace{1cm} + a_1(2 - u) + a_0 \ln\left(\frac{2}{u}\right), \qquad (41)$$

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

B. Can we let $r_{max} \to \infty$?

In the occlusion model, there is the notion that the larger $r_{\max}$ is, the more likely the image is to be covered by a single object. Below, we show that if we allow infinite-sized objects, the probability of this happening is exactly one.

For simplicity, we assume that the image screen is circular with radius a. As before, we place disks in $\mathbb{R}^2 \times [r_{\min}, r_{\max}]$ according to a Poisson process with rate function

$$\lambda(x, y, r) = \frac{c}{r^3}, \tag{42}$$

where c is a constant and r is the disk radius. A transformation to polar coordinates (ρ, α) gives

$$\lambda^*(\rho, \alpha, r) = \frac{c\rho}{r^3} \tag{43}$$

—that is, the points (ρ, r) form a Poisson process on the product space $(0, \infty) \times (r_{\min}, r_{\max})$ with rate function

$$\lambda^{**}(\rho, r) = \frac{2\pi c\rho}{r^3}. \tag{44}$$

Now consider the front-most object in the image. The object either covers the *whole* image—in which case, the image has a uniform intensity—*or* the object overlaps part of the screen. Below, we calculate the probability that an image has a constant grey level.
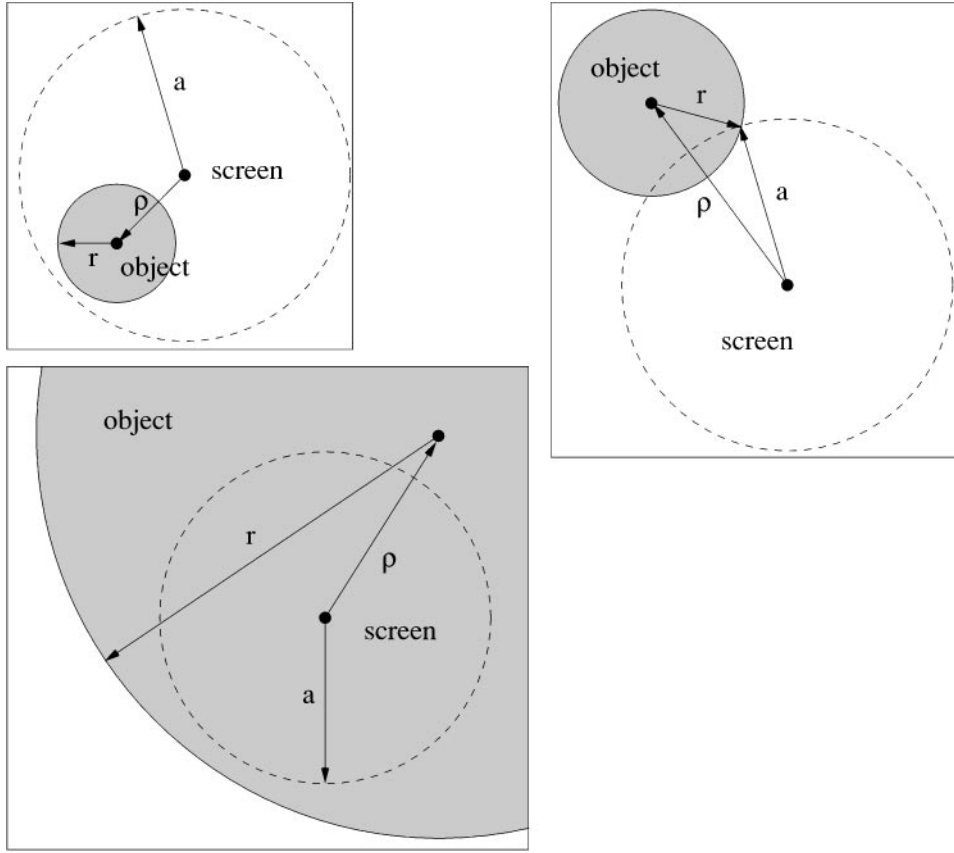
Figure 15. Three examples of objects (*shaded area*) that "overlap" the image screen (*dashed circle*). *Top*: left and right: The object "overlaps" but does *not* "cover" the screen. *Bottom*: The object *both* "overlaps" and "covers" the screen.

This is given by

$$P[I = \text{constant}] = \frac{\mu_{\text{cover}}}{\mu_{\text{overlap}}}, \qquad (45)$$

where μ_{cover} is the measure of the set of "covering" objects in the Poisson process, and μ_{overlap} is the measure of the set of "overlapping" objects.

The definitions for "overlap" and "cover" are straightforward: An object *overlaps* the screen if

$$\rho < a + r, \qquad (46)$$

and an objects *covers* the screen if

$$r \geq \rho + a. \qquad (47)$$

Figure 15 shows three examples of objects that "overlap" the screen. In the top two figures, the objects "overlap" but do *not* "cover" the screen. In the bottom figure, the object *both* "overlaps" and "covers" the screen.

The quantities μ_{cover} and μ_{overlap} can be calculated by integration. Equations (44) and (47) give

$$\mu_{\text{overlap}} = 2\pi c \int_{r=r_{\min}}^{r_{\max}} \left(\int_{\rho=0}^{a+r} \rho \, d\rho \right) \frac{dr}{r^3}$$

$$= \pi c \int_{r_{\min}}^{r_{\max}} \frac{(a+r)^2}{r^3} \, dr$$

$$= \pi c \left[\log\left(\frac{r_{\max}}{r_{\min}}\right) + \frac{2a}{r_{\min}}\left(1 - \frac{r_{\min}}{r_{\max}}\right) \right.$$

$$\left. + \frac{a^2}{2r_{\min}^2}\left(1 - \left(\frac{r_{\min}}{r_{\max}}\right)^2\right) \right].$$

Similarly, Eqs. (44) and (46) give

$$\mu_{\text{cover}} = 2\pi c \int_{\rho=0}^{r_{\max}^{-a}} \left(\int_{r=\rho+a}^{r_{\max}} \frac{dr}{r^3} \right) \rho \, dp$$

$$= \pi c \int_0^{r_{\max}^{-a}} \left(\frac{1}{(\rho + a)^2} - \frac{1}{r_{\max^2}} \right) \rho \; dp$$

$$= \pi c \left[\log \left(\frac{r_{\max}}{a} \right) + \frac{2a}{r_{\max}} - \frac{3}{2} - \frac{a^2}{2r_{\max}^2} \right].$$
(48)

Thus,

$$P(I = \text{constant}) = \frac{\mu_{\text{cover}}}{\mu_{\text{overlap}}} \to 1 \quad \text{as } r_{\max} \to \infty$$
(49)

for $r_{\min}$ finite. Without an upper cutoff on the object sizes, a single object will cover the whole screen.

Acknowledgments

Notes

1. This was pointed out to the senior author by A. Rosenfeld in the 80's.
2. In Wainwright and Simoncelli's analysis, a set of 11 filters supported at 4 adjacent positions is considered. For these a single contrast factor is used, hence the ellipsoidal contours. However, for patches at a distance of 4 or more pixels, they use different contrast factors, hence the scale mixture of Gaussians model produces non-elliptical equi-probability contours for such wavelet coefficient patches (Fig. 4, op. cit.).
3. The probability that a point (x, y) in the image is covered by an object is equal to $1 - e^{-\alpha \lambda z_{\max}}$, where $\alpha = E(\|T\|)$ denotes the expected area of the random templates; Hence, the complete coverage for $z_{\max} \to \infty$.
4. The original name "dead leaves" comes from a picture of leaves falling down on the ground, and an observer viewing them from above.
5. This distribution has previously been suggested for wavelet filter responses; e.g. for application to the entire set of wavelet coefficients (Mallat, 1989) or individual subbands (Simoncelli and Adelson, 1996), and in applications to denoising problems (Moulin and Liu, 1999).
6. Note that scale invariant transparent models (random wavelet expansions) fail to capture the singularity at 0 (Mumford and Gidas, 2000), which seems to be always present in natural data for any zero mean filter response (how pronounced the peak is depends on the amount of texture of "objects" in the images). This is an important clue that *occlusions* play a significant role in images, and that dead leaves models are closer to the truth than models where images are written as sums of "objects".

7. 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.
8. For $x > 2r_{\max}$, $B = 0$. For $x < 2r_{\min}$,

$$B(x) \approx \frac{a_3}{3}(s^3 - u^3) + \frac{a_2}{2}(s^2 - u^2) + a_1(s - u) + a_0 \ln \left(\frac{r_{\max}}{r_{\min}} \right),$$
(50)

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

References

Alvarez, L., Gousseau Y., and Morel J.-M. 1999. The size of objects in natural and artificial images. *Advances in Imaging and Electron Physics*, vol. 111. Academic Press: San Diego, CA, pp. 167–242.

Buccigrossi, R.W. and Simoncelli, E.P. 1999. Image compression via joint statistical characterization in the wavelet domain. *Proc. IEEE Trans. on Image Processing*, 8(12):1688–1701.

Field, D.J. 1987. Relations between the statistics of natural images and the response properties of cortical cells. *J. Optical Society of America*, A4:2379–2394.

Freeman, W.T. and Pasztor, E.C. 1999. Learning low-level vision. In *Proc. IEEE Int. Conf. on Computer Vision*, Corfu, Greece.

Grenander, U. and Srivastava, A. 2000. Probability models for clutter in natural images, *IEEE Trans. PAMI*, in press.

Hallinan P.W., Gordon G.G., Yuille, A.L., Giblin, P., and Mumford, D. 1999. Two- and three-dimensional patterns of the face. A.K. Peters Ltd, Natick, MA.

Heeger, D.J. and Bergen, J.R. 1995. Pyramid based texture analysis/synthesis. In *Computer Graphics Proc.*, pp. 229–238.

Huang, J., Lee, A., and Mumford, D. 2000. Statistics of range images. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 1, Hilton Head Island, SC, pp. 324–331.

Huang, J. and Mumford, D. 1999. Statistics of natural images and models. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Vol. 1*, Fort Collins, CO, pp. 541–547.

Isard, M. and Blake, A. 1998. CONDENSATION—conditional density propagation for visual tracking. *Int. J. Computer Vision*, 291:5–28.

Kendall, W.S. and Thönnes, E. 1998. *Perfect Simulation in Stochastic Geometry*. Preprint 323, Department of Statistics, University of Warwick, UK.

Kingman, J.F.C. 1993. *Poisson Processes*. Oxford Studies in Probability. Clarendon Press, Oxford.

Lee, A.B. and Mumford, D. 1999. An occlusion model generating scale-invariant images. In *Proc. IEEE Workshop on Statistical and Computational Theories of Vision*, Fort Collins, CO.

Malik, J., Belongie, S., Shi, J., and Leung, T. 1999. Textons, contours and regions: Cue integration in image segmentation. In *Proc. IEEE Int. Conf. Computer Vision*, Corfu, Greece.

Mallat, S.G. 1989. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. PAMI*, 11:674–693.

Matheron, G. 1968. Modèle séquentiel de partition aléatoire. Technical report, CMM, 1968.

Matheron, G. 1975. *Random Sets and Integral Geometry*. John Wiley and Sons: New York.

Moulin, P. and Liu, J. 1999. Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors. *IEEE Trans. on Information Theory*, 453:909–919.

Mumford, D. and Gidas, B. 2000. Stochastic models for generic images. *Quarterly Journal of Applied Mathematics* Volume LIX, Number 1, March 2001, pp. 85–111.

Olshausen, B.A. and Field, D.J. 1996. Natural image statistics and efficient coding. *Network*, 7:333–339.

Ruderman, D.L. 1994. The statistics of natural images. *Network*, 5(4):517–548.

Ruderman, D.L. 1997. Origins of scaling in natural images. *Vision Research*, 37(23):3385–3395.

Ruderman, D.L. and Bialek, W. 1994. Statistics of natural images: scaling in the woods. *Physical Review Letters*, 73(6):814–817.

Serra, J.P. 1982. *Image Analysis and Mathematical Morphology*. Academic Press: London.

Simoncelli, E.P. 1999. Modeling the joint statistics of images in the wavelet domain. In *Proc. SPIE 44th Annual Meeting, Vol. 3813*, Denver, Colorado.

Simoncelli, E.P. and Adelson, E.H. 1996. Noise removal via bayesian wavelet coring. In *Third Int'l Conf. on Image Processing*, Lausanne, Switzerland, pp. 379–383.

Sullivan, J., Blake, A., Isard, M., and MacCormick, J. 1999. Object localization by Bayesian correlation. In *Proc. Int. Conf. Computer Vision*, pp. 1068–1075.

van Hateren, J.H. and van der Schaaf, A. 1998. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. R. Soc. Lond.*, B265:359–366.

Wainwright, M.J. and Simoncelli, E.P. 2000. Scale mixtures of gaussians and the statistics of natural images. In *Advances in Neural Information Processing Systems* 12, Solla, Leen, and Müller (Eds.), MIT Press: Cambridge, MA, pp. 885–861.

Zetzsche, B., Wegmann, B., and Barth, E. 1993. Nonlinear aspects of primary vision: Entropy reduction beyond decorrelation. In *Int'l Symp. Soc. for Info. Display, Vol. 24*, pp. 933–936.

Zhu, S.C. and Guo, C. 2000. Mathematical modeling of clutter: Descriptive vs. generative models. In *Proc. of the SPIE AeroSense Conf. on Automatic Target Recognition*, Orlando, FL.

Zhu, S.C., Luo, Q., and Zhang, R. 1999. Effective statistical inference by data-driven markov chain Monte-Carlo. Technical Report, Dept. of Computer and Information Sciences, Ohio State University, Columbus, OH.

Zhu, S. and Mumford, D. 1997. Prior learning and Gibbs reaction-diffusion. *IEEE Trans PAMI*, 19(11):1236–1250.

Zhu, S.C., Wu, Y.N., and Mumford, D. 1998. FRAME: Filters, random field and maximum entropy—towards a unified theory for texture modeling. *Int'l J. Computer Vision*, 27(2):1–20.