

Fragmentation in the Vision of Scenes

Jan-Mark Geusebroek

Arnold W. M. Smeulders

Intelligent Sensory Information Systems
Faculty of Science, University of Amsterdam
Kruislaan 403, 1098 SJ Amsterdam, The Netherlands
E-mail: mark@science.uva.nl

Abstract

Natural images are highly structured in their spatial configuration. Where one would expect a different spatial distribution for every image, as each image has a different spatial layout, we show that the spatial statistics of recorded images can be explained by a single process of sequential fragmentation. The observation by a resolution limited sensory system turns out to have a profound influence on the observed statistics of natural images. The power-law and normal distribution represent the extreme cases of sequential fragmentation. Between these two extremes, spatial detail statistics deform from power-law to normal through the Weibull type distribution as receptive field size increases relative to image detail size.

1. Introduction

Vision provides an enormous amount of information about our physical environment. A general purpose vision system is concerned with the processing of visual sensory information for the purpose of acting, reacting and reflecting on a constantly changing environment. The human visual system is an example of such a general vision system very well adapted to its task.

We start with the prerequisite that any general sensory system will adapt itself to the outside world, specifically to the stochastics of the input signals [2]. When that point of departure is accepted, we note that the statistics of the sensory input are dominated by physical laws of image formation [9]. These physical laws are basically domain independent, as they cover the universally applicable laws of light reflectance from materials. In addition, statistics of the sensory input may be shaped by the structure of our environment. For example, parts of an image which deviate from the common structure around us are likely to contain perceptually salient details. Hence, our motivation to study the statistical regularities in natural images as it implies a

better understanding of cognitive vision systems.

Equipped with an infinitely precise sensor we would see details around us at all scales. This view in all its full complexity is useless to the observer as it would swamp the perceptual processing system. To escape the influx of so much information, a large reduction in information is implemented at the retina where the outside world is integrated over discrete sensory receptive fields. The observation by receptors of finite size imposes spatial coherence to the picture while reducing the complexity of the observed scene.

The spatial statistics of large ensembles of natural images are known to be scale invariant [18]. That is, when examining the marginal distribution of derivative filters or gradient magnitude, an inverse power-law distribution is found in the Fourier domain. However, the statistics of individual images may vary across scale. Consequently the statistical properties for individual images are affected by the observation at finite resolution.

The question is which laws govern *observed* natural image statistics? In previous work, we theoretically derived natural image statistics to follow a sequential fragmentation process [8]. In the current paper, we conduct a large scale experiment on a representative collection of every day scenes, containing almost 50,000 images. Statistical analysis of the data results in a significant finding that natural image statistics obey the sequential fragmentation process, of which the Weibull distribution is the general solution. Hence, we put the empirical findings, for a few images, of Mallat [14] and Simoncelli [19] into a much broader perspective. We provide an alternative theory for the dead-leave model proposed by Matheron and elaborated upon by Lee *et al.* [15, 13]. Furthermore, our model includes the class of transparent objects, as recently modelled by Grenander and Srivastava [11, 20]. Our proposed model includes scaling behavior, where the power-law and normal distribution represent the two extreme ends of resolving power. Between these two extremes, spatial detail statistics follow the Weibull type distribution.

2. The Sequential Fragmentation Process

In this section, we follow [8] for introducing the sequential fragmentation theory. Given the theoretical process, we draw conclusions on the possible options for image statistics. In a later section, we will validate our conclusions on large collections of images.

2.1. Sequential Fragmentation Theory

As a direct implication of causality, we consider that small details are occurring more often in an image than large structures [12]. Diffusion of numerous small structures will result in fewer large structures. Inversely, increasing magnification at large structures will resolve many smaller structures. One may rephrase the statement in that, when resolving power increases, large structures will break-up into new structures, of which some of them are relatively large, but most of them will be small details. An edge or bar filter will respond strongly to large contrasting details, and yield smaller values elsewhere. Hence, the histogram of contrasts for one structure typically shows a power-law distribution,

$$f(x) = \left(\frac{x}{\beta}\right)^{\gamma-1}. \quad (1)$$

When more objects are added to the scene, the image will be fragmented into various patches, each giving rise to an edge of varying contrast. The histogram over the various edges is the results of integrating over the various power-laws caused by every edge,

$$n(x) = c \int_x^\infty n(x')f(x) dx' \quad (2)$$

where $n(x)$ indicates the number of pixels with response magnitude between x and $x + dx$, contributed by all edges with contrast $x' > x$. The integration over a sufficient number of power-laws yields a Weibull distribution,

$$n(x) = \frac{1}{\beta} \left(\frac{x}{\beta}\right)^{\gamma-1} e^{-\frac{1}{\gamma}\left(\frac{x}{\beta}\right)^\gamma}. \quad (3)$$

In integral form, the Weibull distribution given by

$$N(>x) = \int_x^\infty n(x) dx = e^{-\frac{1}{\gamma}\left|\frac{x}{\beta}\right|^\gamma} \quad (4)$$

indicates the relative amount of edges of (positive or negative) contrast larger than x .

In our view, a picture is composed of many details of larger and smaller size, which in turn are composed of even smaller details. We consider small image patches, or textures, like edge parts, blobs, and corners, spots to be the

details an image consists of. In the projection of the details onto the receptive field in the retina, some of the detail is larger than the scale of resolution, whereas other details are smaller and effectively integrated in the response of one receptive field. The size distribution of the details may be inferred from the contours of the details and their cast shadows. The projection of the contours is a linear transform of the three-dimensional detail shape. Hence, the intensity differences in a view are indicative for the size distribution of the projected details in the scene [16].

A similar reasoning is known in the sequential fragmentation of particles by milling [4, 5], which shows much resemblance with the present theory. Brown and Wohletz [5] theoretically derived the power-law process to describe the particle size distribution for the crushing of particles in a mill, providing a solid physical basis for the distributions in Eq. (3) and Eq. (4) with the shape parameter γ related to fractal dimension.

2.2. Consequences for Natural Image Statistics

As a consequence of the sequential fragmentation theory, spatial image statistics are limited to conform to one out of five options:

Power-law: When resolution is extremely fine compared to detail size, spatial layout follows a power-law distribution, being the result of a single fragmentation event. This is the case when we examine a single object against a highly contrasting background. When contrast is reduced by local normalization, a power-law is no longer observed.

Normal: When resolution is too coarse to resolve the details, spatial layout becomes normal distributed. This is the case when we look at sand, or at hairs. The normal distribution is the result of independent details accumulating to the filters response, which will only be true for a small class of textures. On closer inspection with higher resolution, we may resolve the details and spatial statistics convert to Weibull again.

Weibull: In general, with the fine but limited resolution used for the vast majority of scenes we encounter, views we observe are fragmented and their details therefore Weibull distributed. Spatial detail statistics deform from power-law to normal through the Weibull type distribution as resolving power decreases, while the field of view enlarges.

Composition: When the scene is composed of a few parts, the Weibull distribution will not appear. This is often the case for a scene composed of two or three objects, or scenes with sharply distinct distances. In this case, individual parts of the scene may conform to the Weibull distribution with varying parameters. However, the composition of the scene results in the addition of a few power-laws, not yet resembling a Weibull distribution. As the Weibull distribution describes sequential fragmentation, a requirement

for the Weibull distribution is that the composition of the scene is sufficiently complex.

Regular: For repetitive patterns, the visual responses interfere with the repetition in the sensing field, in which case any distribution may describe spatial statistics. Hence, the proposed fragmentation theory breaks for regular textures.

Any image formation process involves the sequential fragmentation of structures to refine a scene. Measurement yields the sieving of the scene to sort out the structures present. Since both processes are dual, one can make no distinction between fragmentation or sieving from the final result. Hence, sieving an image with arbitrary mesh size will result in a Weibull distribution for the local statistics. Scale-space filtering [12] is considered to be the sieving process dual to resolving power. The choice of the mesh size, hence the filter scale, will not affect the statistical result, except for a reparameterization. The dimensionality of the measurement will be absorbed in the exponent γ of the Weibull distribution.

3. Experimental Setup

We experimentally investigated the statistics of spatial detail on three databases. First, we tested if the sequential fragmentation process could be found in the van Hateren outdoor image collection [22]. The collection consist of 4,167 calibrated and uncompressed images of 12-bit grey-value information, image size 1,536 by 1,024.

As edge information is of crucial importance in the coding of images, we expect the sequential fragmentation process so dominantly present in image statistics that it will be preserved by any compression method. Hence, as a second database to show the sequential fragmentation process to be dominant, we tested the Corel photo collection [1]. The collection consists of 46,695 images covering a broad class of general pictures. The large and diverse collection is used to provide evidence for the sequential fragmentation theory to be present in a broad imaging domain, much broader than only dead-leaves occlusion processes, as in [13], or only transparent (additive) image formation, as in [11, 20]. The collection is originally compressed by a wavelet compression technique. For processing purposes, we converted the whole collection to JPEG compressed images, compression factor 0.7.

A detailed study was performed on views of material textures in the Curet collection [6]. The image collection is calibrated and uncompressed, image size 768×576 . The collection consists of 61 materials, each taken under various illumination and viewing directions. The sample contains a diverse collection of materials, including plaster, styrofoam, straw, corduroy, paper, brick, fur, and so on, effectively covering a range of Lambertian reflection, polarized reflection with highlights, to the mirror reflection of Aluminum foil.

We will use this database mainly for illustration purposes.

Edge strength is accessed by Gaussian derivative filters measured in 72 directions by steering the x, y -derivative filters. The effective resolution of the system is given by the spatial width of the filter, here set to correspond to a standard deviation of 3 pixels in all experiments. Note that changing the width to another constant will not change the major results of this paper (data not shown). Further note that applying any alternative zero-average filter will not affect the major results (data not shown). Responses per image were accumulated into histograms, and three distributions were fitted to each histogram. They are a power-law distribution, a Weibull distribution, and a normal distribution. As the histogram contains both positive and negative edge responses, we used the symmetrical versions of the Weibull and power-law distribution.

For filter responses, high values indicate strong correlation between image content and filter shape. Hence the tails of the distribution are much more important in terms of image content [3] than low values, representing uniform areas and noise. Where most hypothesis tests, including the Kullback-Leibler divergence, assign more weight to the often occurring values, we put our emphasis in the tails of the distribution. As the Anderson-Darling test is sensitive to the tails, goodness-of-fit was evaluated by this hypothesis test [7]. Note that this test is commonly used in statistics when accessing goodness-of-fit for both power-law and Weibull distributions. A second important characteristic of the Anderson-Darling test is that critical values are tuned to the distribution at hand, including the free parameters. Models of different complexity –with different degrees of freedom– may be compared at similar confidence levels. Hence, the Anderson-Darling test allows fair comparison between power-law, Gaussian, and Weibull distributed portion of a database, despite the different number of parameters of these models.

The Weibull distribution symmetric integral form is given by

$$f(x) = C \exp\left(-\frac{1}{\gamma} \left|\frac{x - \mu}{\beta}\right|^\gamma\right) \quad (5)$$

the parameters μ , β , and γ representing the center, width, and shape of the distribution, and C being a normalization constant. The shape parameter, γ , ranges from 0 to 2 [10]. For $\gamma = 2$ the Weibull distribution is equivalent to the normal distribution, and for $\gamma = 1$ it is a double exponential. The distribution is also known as the generalized Laplacian [14, 19]. Brown [4] showed the close connection of this integral form to the original distribution proposed by Weibull [23]. For our experiments, the values of the Weibull parameters were estimated using the maximum likelihood method. Goodness-of-fit was evaluated at a significance level $\alpha = 0.05$ ($A^2 < 0.757$) [7] for all cases.

Furthermore, we rejected the Weibull distribution for an estimation of $\gamma > 2.2$, resulting in a value of γ too far out of range to yield a stable statistical process [10].

We consider the symmetric form of the power-law,

$$f(x) = \frac{1}{2}\delta|x|^{-\delta-1} . \quad (6)$$

The parameter δ was estimated by the maximum likelihood method, and goodness-of-fit was evaluated by the Anderson-Darling statistic [17] at a significance level of $\alpha = 0.05$ ($A^2 < 1.341$) [21].

The parameters for the normal distribution were obtained by maximum likelihood estimation, and goodness-of-fit was evaluated by the Anderson-Darling statistic at a significance level of $\alpha = 0.05$ ($A^2 < 0.787$) [21].

4. Results

For the Corel general photo stock, the Anderson-Darling test indicated 48% of the pictures to be Weibull distributed. This is a remarkable result given the variety of generating processes for the pictures in the collection, and the compression of the images. The Weibull distribution apparently describes the spatial statistics for outdoor scenes, indoor scenes, close-ups, and materials of various kinds. Approximately 1% of the Corel collection is normally distributed. Note that these pictures are included in the Weibull distributed part of the collection, the normal distribution being an extreme case with $\gamma = 2$. An additional 9% of the Corel collection is distributed according to a power-law, while being rejected as a Weibull distribution. A portion of 4% is accepted as both power-law distributed and Weibull distributed, the fraction being included in the reported 48%.

For the uncompressed images in the van Hateren collection, similar results are obtained. For this collection, a Weibull distribution is present in 54% of the images. A neglectable amount of images is normally distributed, which can be explained by the high resolution at which these outdoor images are taken. Furthermore, 28% of the images is power-law distributed, indicating that much of the images consist of a high contrasting object against a more or less uniform background. A portion of 5% is accepted as both power-law distributed and Weibull distributed, the fraction being included in the 54% Weibull distributed.

To understand this widespread presence of the Weibull distribution, a detailed study was performed on views of material textures in the Curet collection [6]. It appeared that 54 materials out of 61 (88%) consistently render a Weibull distribution. The estimated Weibull parameter values varied with illumination and viewing direction, consistently rendering a Weibull distribution over all imaging conditions. Note that in the continuing report of our result, examples of the Curet database do not provide anecdotal evidence for

our theory, rather yields detailed illustrations of the different stages of the fragmentation process.

As an explanation for a normal distribution of filter responses consider an infinitely precise sensor. We would see the details around us at infinitely many scales. Increasing the size of the sensor to a finite extent imposes spatial coherence and a limited local scale of detail. When the sensor resolution is much larger than the common size of the random details in the field of view, each sensor response is an average over many impulses. From the central limit theorem, local intensity differences will be normally distributed as is observed in the rabbit fur of the Curet collection (Figure 1a), with individual hairs as the random details, each much smaller than the resolution of observation.

For the power-law distributed images, visual inspection showed much of these pictures to contain an exhibited item, or to display land-sky, thereby fragmenting the scene into a foreground and a background region, while details of intermediate size are missing. When the shape of the foreground figure is sufficiently fractal, the distribution of intensity differences will follow a power-law. See the orange peel in the Curet collection (Figure 1b) as an example.

Between these two extremes, the Weibull distribution occurs. It arises when the scene is progressively fragmented by the addition of objects or detail. Such a process of sequential fragmentation results in a Weibull size distribution [4, 5], where the power-law describes a single fragmentation event. Material textures, with small details extending over a limited depth range, as well as everyday scenes and even mountain views extending over considerable depth, all follow a Weibull sequential fragmentation process. When the size of the details underlying the texture is such that a receptive field typically covers part of a structure, the observed distribution follows Weibull, illustrated by the aluminum foil in the Curet collection (Figure 1c). The power-law process gives the extreme case for a single fragmentation, segmenting the field of view into foreground and background. On the other end of the spectra of size distributions, we have the normal distribution, representing the extreme of fragmentation beyond the visual resolution.

When the spatial detail is not randomly distributed but regular, such that there is repetition between the responses in an image, a Weibull distribution will no longer be found. For the Curet collection, ribbed paper, straw, corduroy, and corn husk break the distribution in the direction of orientation (Figure 1d). In these examples, however, a Weibull distribution reappears when measuring the receptive field response in the perpendicular direction. For the rug and the painted spheres in the Curet collection, the specular reflectance of the material causes a regular pattern of highlights, breaking the standard distribution in any direction.

The distribution of the fragmentation exponent γ as given in Figure 2 indicates the relative importance of the

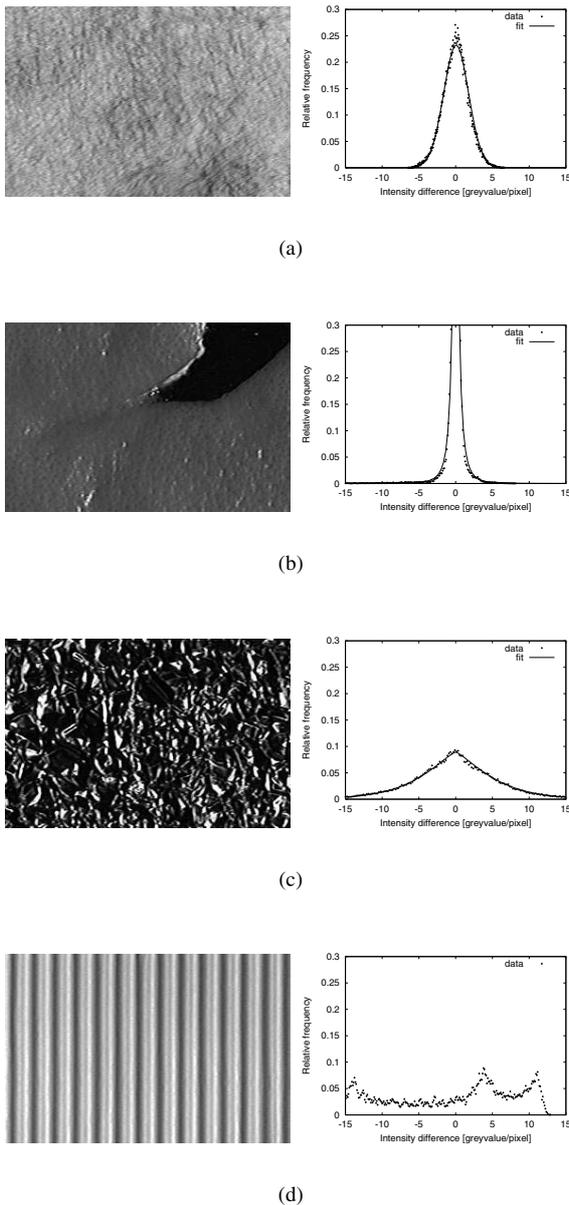


Figure 1: Histograms of intensity differences in the x -direction for images from [6]. The rabbit fur (a) demonstrates the resolution limiting case, for which the Weibull distribution with $\gamma_x = 1.94$ approaches the normal distribution. Orange peel (b) is an example of a single object fragmentation, for which the histogram follows a power-law distribution ($\delta_x = 2.54$). The aluminum foil (c) shows a Weibull distribution with $\gamma_x = 1.27$. Ribbed paper (d) with its regular structure has a non-Weibull distribution in the x -direction, but shows a Weibull distribution for the y -direction.

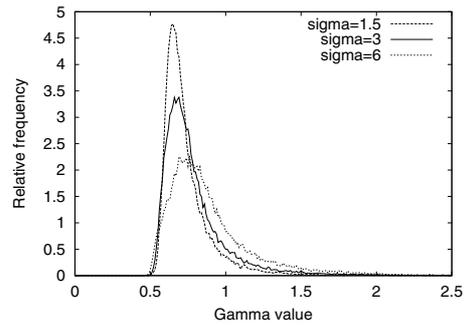


Figure 2: Distribution of the values of γ for the Weibull distribution as estimated for the Corel collection at spatial filter sizes $\sigma = \{1.5, 3, 6\}$.

power-law process and the resolution limited case of the normal distribution. Note that the γ values inherently depend on the observation resolution, hence on the pixel resolution and filter size. For high resolution images observed at small scale, power-law will be dominant. For a low resolution collection or a large observation scale, the distribution of γ -values will shift toward the normal extreme. This scaling behavior is illustrated in Figure 2. For a general vision system, observing at a variety of scales, spatial statistics will cover the complete spectrum of γ -values.

For the gradient magnitude, $f_w = \sqrt{f_x^2 + f_y^2}$, which represents a rotationally invariant filter, 32% of the Corel collection conforms to the Weibull distribution. In this case, we tested for the true Weibull distribution, as magnitude is a strictly positive entity. Furthermore, a photometric invariant $f_n = f_x/f$ is tested. In that case, 45% of the collection is Weibull distributed, whereas the power-law is no longer present. Hence, non-linear sensory combinations still result in a Weibull distribution of observed spatial detail.

The Weibull or power-law distribution is not observed when the image exhibits large uniform regions. Visual inspection of some of the Corel images not conforming to the three distributions revealed these images to show a composition of a few objects. Typical examples are objects exposed against a uniform background, and landscape images under a uniform sky. In such cases, the histogram of the filter responses consists of an addition of two Weibull distributions, with large differences between the parameters. Note that such compositions are not observed in the Corel database, which contains homogeneous textured materials.

In summary, for the Corel general photo stock [1] consisting of 46,695 images, the Anderson-Darling test indicated 57% of the pictures of that collection to conform to the theory of sequential fragmentation. For the van Hateren [22] high resolution and calibrated outdoor image collection, 82% of the images conform to the theory of sequen-

tial fragmentation. This is a remarkable result given the variety of generating processes for the pictures in the collections. The extreme cases of power-law and normal distribution explain just a small part of all scenes. In general, with the fine but limited resolution used for the vast majority of scenes we encounter, views we observe are sequentially fragmented and their details therefore Weibull distributed, seen in 47% of the Corel stock and in 54% of the van Hateren collection. A large portion of the remainder of the collection is a composition of fragmentation processes with different parameters. Unfortunately, no confidence levels are available –or easily obtainable– to access the Anderson-Darling statistic for mixture of power-laws. The fragmentation process breaks for repetitive patterns, causing interference between the receptive field responses.

5. Conclusions

The projection of the fractal world around us on a discrete sensory system, with finite observation resolution, causes scale invariant physical laws to deform to the Weibull statistic. We have found that many images and many local regions therein are dominated by that standard distribution. The significance of our results show the fragmentation process to describe the statistical structure of individual natural images, thereby confirming earlier findings of Mallat [14] and Simoncelli [19]. The model of sequential fragmentation holds for indoor and outdoor scenes, slices through materials like found in biological preparations and at material surfaces, drawings like modern art paintings, astronomy pictures, opaque and transparent compositions of materials, all included in the reported photo stocks.

The fragmentation process results in five options for spatial statistics: power-law, Weibull, normal, composition, or the result of a regular spatial arrangement. Spatial statistics deform from power-law to normal through the Weibull type distribution as the complexity of the scene increases.

Acknowledgments

We acknowledge the discussions with Jan Koenderink and Sylvia Pont. The work is supported by the ICES-KIS Multimedia Information Analysis project.

References

- [1] Corel gallery. *see www.corel.com*.
- [2] H. B. Barlow. Possible principles underlying the transformation of sensory messages. In W. Rosenblith, editor, *Sensory Communication*, page 217. MIT Press, 1961.
- [3] A. J. Bell and T. J. Sejnowski. The independent components of natural scenes are edge filters. *Vision Res.*, 37:3327–3338, 1997.
- [4] W. K. Brown. A theory of sequential fragmentation and its astronomical applications. *J. Astrophys. Astr.*, 10:89–112, 1989.
- [5] W. K. Brown and K. H. Wohletz. Derivation of the weibull distribution based on physical principles and its connection to the rosin-rammler and lognormal distributions. *J. Appl. Phys.*, 78:2758–2763, 1995.
- [6] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink. Reflectance and texture of real world surfaces. *ACM Trans Graphics*, 18:1–34, 1999.
- [7] J. J. Filliben et al. *NIST/SEMTECH Engineering Statistics Handbook*. www.itl.nist.gov/div898/handbook, NIST, Gaithersburg, 2002.
- [8] J. M. Geusebroek and A. W. M. Smeulders. A physical explanation for natural image statistics. In M. Chantler, editor, *Proceedings of the 2nd International Workshop on Texture Analysis and Synthesis (Texture 2002)*, pages 47–52. Heriot-Watt University, 2002.
- [9] J. M. Geusebroek, R. van den Boomgaard, A. W. M. Smeulders, and H. Geerts. Color invariance. *IEEE Trans. Pattern Anal. Machine Intell.*, 23(12):1338–1350, 2001.
- [10] B. V. Gnedenko and A. N. Kolmogorov. *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, Boston, 1968.
- [11] U. Grenander and A. Srivastava. Probability models for clutter in natural images. *IEEE Trans. Pattern Anal. Machine Intell.*, 23(4):424–429, 2001.
- [12] J. J. Koenderink. The structure of images. *Biol. Cybern.*, 50:363–370, 1984.
- [13] A. B. Lee, D. Mumford, and J. Huang. Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model. *Int. J. Comput. Vision*, 41:35–59, 2001.
- [14] S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Machine Intell.*, 11:674–693, 1989.
- [15] G. Matheron. *Random Sets and Integral Geometry*. John Wiley and Sons, New York, 1975.
- [16] A. P. Pentland. Fractal-based description of natural scenes. *IEEE Trans. Pattern Anal. Machine Intell.*, 6:661–674, 1984.
- [17] S. E. Rigdon. Testing goodness-of-fit for the power law process. *Commun. Statist.-Theory Meth.*, 18:4665–4676, 1989.
- [18] D. L. Ruderman and W. Bialek. Statistics of natural images: Scaling in the woods. *Phys. Rev. Lett.*, 73:814–817, 1994.
- [19] E. P. Simoncelli. Modeling the joint statistics of images in the wavelet domain. In *Proc. SPIE*, volume 3813, pages 188–195. SPIE, 1999.
- [20] A. Srivastava, X. Liu, and U. Grenander. Universal analytical forms for modeling image probabilities. *IEEE Trans. Pattern Anal. Machine Intell.*, 24(9):1200–1214, 2002.
- [21] M. A. Stephens. EDF statistics for goodness of fit and some comparisons. *J. Am. Statist. Assoc.*, 69:730–737, 1974.
- [22] 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.
- [23] W. Weibull. A statistical distribution function of wide applicability. *J. Appl. Mech.*, 18:293–297, 1951.