
Journal of Mathematical Imaging and Vision 18: 17–33, 2003

On Advances in Statistical Modeling of Natural Images

A. SRIVASTAVA*
Department of Statistics, Florida State University, Tallahassee, FL 32306, USA
anuj@stat.fsu.edu

A.B. LEE
Division of Applied Mathematics, Brown University, Providence, RI 02912, USA

E.P. SIMONCELLI
Courant Institute for Mathematical Sciences, New York University, New York, NY 10003, USA

S.-C. ZHU
Department of Computer Science, Ohio State University, Columbus, OH 43210, USA

Abstract. Statistical analysis of images reveals two interesting properties: (i) invariance of image statistics to scaling of images, and (ii) non-Gaussian behavior of image statistics, i.e. high kurtosis, heavy tails, and sharp central cusps. In this paper we review some recent results in statistical modeling of natural images that attempt to explain these patterns. Two categories of results are considered: (i) studies of probability models of images or image decompositions (such as Fourier or wavelet decompositions), and (ii) discoveries of underlying image manifolds while restricting to natural images. Applications of these models in areas such as texture analysis, image classification, compression, and denoising are also considered.

Keywords: natural image statistics, non-Gaussian models, scale invariance, statistical image analysis, image manifold, generalized Laplacian, Bessel K form

1. Introduction

Applications dealing with static or dynamic images (movies) have become increasingly important and popular in recent years. Tools for image analysis, compression, denoising, transmission, and understanding have become widely involved in many scientific and commercial endeavors. For instance, compression of images is important in transmission or storage, and automatic recognition of people from their camera images has become important for security purposes. Development of tools for imaging applications starts with mathematical representations of images. Many of these applications need to "learn" some image patterns or tendencies before specific data can be analyzed. The area of natural image statistics has resulted from efforts to observe, isolate, and explain patterns exhibited by natural images. Lately, there has been a greater emphasis on explicit probability models for images. Perhaps one reason for that is the growing appreciation for the variability exhibited by the images, and the realization that exact mathematical/physical models may not be practical, and a statistical approach needs to be adopted. Any statistical approach will need probability models that capture "essential" image variability and yet are

*To whom correspondence should be addressed.

computationally tractable. Another motivation for pursuing image statistics has been to understand sensory coding as a strategy for information storage/processing in animal vision. An understanding of image statistics may provide clues to the architecture of animal visual system [79].

There are several paths to understanding image statistics. Even though images are expressed as elements of a large vector space (e.g. the space of rectangular arrays of positive numbers), referred to as the *image space* here, the subset of interesting images is rather small and restricted. So one path is to isolate this subset, called an *image manifold* here, and impose a simplistic probability model, such as a uniform or a Gaussian-like model, on it. Given the image manifolds and probability distributions on them, statistical tools for imaging applications follow. The other path is to derive probability models that are defined on the larger image space but put any significant probability only on the image manifold. We have divided this paper along these two categories.

In search for statistical descriptions of images, mathematical and physical ideas are not abandoned but are intimately involved. For example, harmonic analysis of image intensities is used to decompose images into individual components that better lend to the model building than the original images. Since image spaces are rather high dimensional, and common density estimation techniques apply mostly to small sizes, one usually starts by decomposing images into their components, and then estimating lower order statistics of these components individually. Also, imaging is a physical process and physical considerations often contribute to the search for statistical descriptors in some form. Physical models have been the motivation of many studies that have led to interesting statistical characterizations of images.

This paper is laid out as follows: In Section 2 we start with a historical perspective on spectral analysis of images and some discoveries that followed. An important achievement was the discovery of non-Gaussianity in image statistics. In next two sections we present some recent results categorized into two sets: Section 3 studies probability models for images decomposed into their spectral components, while Section 4 looks at methods for discovering/approximating image manifolds for natural images. A few applications of these methods are outlined in Section 5, and some open issues are discussed in Section 6.

2. Image Decompositions and Their Statistical Properties

For the purpose of statistical modeling, an image is treated as a realization of a spatial stochastic process defined on some domain in $\mathbb{R}^2$. The domain is assumed to be either a (continuous) rectangular region or a finite, uniform grid. A common assumption in image modeling is that the underlying image process is stationary, i.e. image probabilities are invariant to translations in the image plane.

2.1. Classical Image Analysis

In case the image process is modeled as a second order spatial process, the spectral analysis becomes a natural tool. Defining the covariance function, for two arbitrary pixel values, as $C(x)$ where x is the difference between two pixel locations, one can define the power spectrum as $P(w) = \int_{\mathbb{R}^2} C(x)e^{-jwx}\,dx$, and where w denotes the 2D spatial frequency. Since a Fourier basis is also an eigen basis for circulant matrices, which is the case for $C(x)$ under the stationarity assumption, Fourier analysis also coincides with the popular principal component analysis (discussed later in Section 4). Also, the Fourier representation guarantees uncorrelated coefficients. Images are then represented by their Fourier coefficients, image statistics are studied via coefficient statistics, and one can inherit a stochastic model on images by imposing a random structure on the Fourier coefficients. Specification of means and covariances of the coefficients completely specifies a second order image process. Early studies of spatial power spectra indicated that the power $P(w)$ decays as $\frac{A}{|w|^{2-\eta}}$ where $|w|$ is the magnitude of the spatial frequency. This property, called the power law spectrum for images, was first observed by television engineers in the 50's [25, 49] and discovered for natural images in late 80s by Field [30] and Burton and Moorhead [17]. As summarized in [61], the value of η changes with the image types but is usually a small number.

Although Fourier analysis is central to classical methods, other bases have found popularity for a variety of reasons. For instance, in order to capture the locality of objects in images, decomposition of images using a wavelet basis has become an attractive tool. In particular, it is common to use Gabor wavelets [31] for decomposing the observed images simultaneously

in space and frequency. In addition, Marr [56] suggested using the Laplacian of Gaussian filter to model early vision. If one considers images as realizations on a finite, uniform grid in $\mathbb{R}^2$, the image space becomes finite-dimensional, and one can linearly project images into low-dimensional subspaces that are optimal under different criteria. Some of these linear projections are covered in Section 4. Once again image statistics are studied via the statistics of the projected coefficients.

2.2. *Scale Invariance of Image Statistics*

A discovery closely related to the power law is the invariance of image statistics when the images are scaled up or down [16, 30]. In other words, the marginal distributions of statistics of natural images remain unchanged if the images are scaled. The power law is a manifestation of the fractal or scale invariant nature of images. By studying the histograms of the pixel contrasts ($\log(I(x)/I_0)$) at many scales, Ruderman and Bialek [71] showed its invariance to scaling. Independently, Zhu and Mumford [98] showed a broader invariance by studying the histograms of wavelet decompositions of images. Ruderman [69, 70] also provided evidence of scale invariance in natural images and proposed a physical model for explaining them. Turiel and Parga [84] investigated the multi-fractal structure of natural images and related it to the scale invariance. Scaling of different types of scenes was studied by Huang [43]. In [85], Turiel et al. showed that hyperspectral images (color images) also demonstrate multiscaling properties and the statistics are invariant to scale. Scaling of order statistics of pixel values in small windows was studied by Geman and Koloydenko [32]. In addition to pixel statistics, scaling of topological statistics obtained from morphological operations on images was demonstrated by Alvarez et al. [1].

It must be emphasized that only the statistics of large ensembles of images are scale invariant; statistics of individual images vary across scales. Theoretical models that seek probabilistic description of image ensembles aim for scale invariance, while application driven models that work with individual images aim to capture individual image variability.

2.3. *Non-Gaussianity of Marginal Statistics*

Classical methods assume that images are second order processes but the observations do not support this assumption. Higher order statistics of natural images were found to exhibit interesting patterns and the researchers focused next on these higher order moments. One implication is that image statistics do not follow Gaussian distribution and require higher order statistics. For example, a popular mechanism for decomposing images locally—in space and frequency—using wavelet transforms leads to coefficients that are quite non-Gaussian, i.e. the histograms of wavelet coefficients display heavy tails, sharp cusps at the median and large correlations across different scales. To our knowledge, Field [30] was the earliest to highlight the highly kurtotic shapes of wavelet filter responses. Mallat [55] pointed out that coefficients of multiscale, orthonormal wavelet decompositions of images could be described by generalized Laplacian density (given later in Section 3.2). This non-Gaussian behavior of images has also been studied and modeled by Ruderman [69], Simoncelli and Adelson [76], Moulin and Liu [60], and Wainwright and Simoncelli [88]. Recent work of Thomson [82] studies the statistics of natural images using phase-only second spectrum, a fourth order statistic, and demonstrates both the power law and the scale invariance for this statistic. In fact, projection onto any localized zero-mean linear kernel seems to produce kurtotic responses [95]. Huang [43] showed that images, when filtered by 8×8 random mean-0 filters, have distributions with high kurtosis, a sharp cusp at zero and long exponential tails. This suggests a role for linear decompositions that maximize the kurtosis or some another measure of non-Gaussianity. Such efforts [6, 62, 86] have resulted in bases that are spatially oriented with (spatial) frequency bandwidths being roughly one octave, similar to many multiscale decompositions. Similar results were obtained by minimizing the independence of coefficients, under linear decompositions, leading to independent component analysis [18, 23, 44]. These observations justify widespread use of orthonormal wavelets in general image analysis applications. Use of Gabor wavelets is also motivated by the fact that the receptive fields of simple cells in the visual cortex of animals have been found to resemble Gabor functions [58]. Note, however, that most wavelet decompositions of images are based on separable application of one-dimension filters, which leads to non-oriented (mixed diagonal) subbands. Alternate representations that provide better orientation decomposition, and thus higher kurtosis response, include [24, 27, 78, 90].

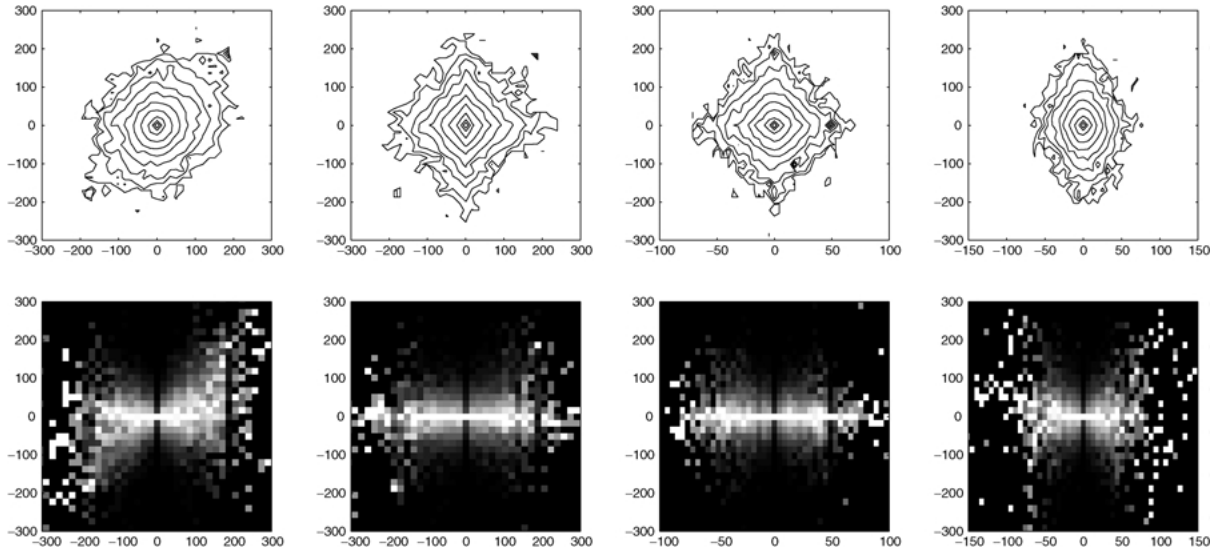

Figure 1. Bivariate histograms of wavelet coefficients associated with different basis functions. Top row shows contour plots, with lines at equal intervals of log probability. Left two cases are for different spatial offsets (same scale and orientation), third is for different orientation (same scale and nearly same position), and the rightmost corresponds to a pair at adjacent scales (same orientation and nearly same position). Bottom row shows some conditional distributions: brightness corresponds to larger frequency.

2.4. Non-Gaussianity of Joint Statistics

In addition to the non-Gaussian behavior of marginal statistics, a number of authors have studied joint statistics of filter responses. In particular, the local structures present in most images lead to dependencies in the responses of local linear operators. A study of joint histograms of wavelet coefficients shows dependency across scales, orientations and positions. Zetzsche and colleagues noted that the joint density of projections onto even- and odd-symmetric Gabor filters have circular symmetry. If the density were Gaussian, this would correspond to independent (uncorrelated) marginals, but the highly kurtotic marginals imply that the responses are strongly dependent [91]. Shapiro developed a heuristic method, for taking advantage of joint dependencies between wavelet coefficients, that revolutionized the field of image compression [73]. Simoncelli and colleagues studied and modeled the dependency between responses to pairs of bandpass filters and found that the amplitudes are strongly correlated, even when the signed responses are not [14, 74]. This is illustrated in Fig. 1, which shows conditional histograms for several pairs of coefficients. Note that unlike second-order correlations, these dependencies can not be eliminated with a linear transform.

Observed statistics of an image patch can be described by joint histograms of coefficients under several filters. For jointly Gaussian coefficients equiprobable contours in joint histograms would all be ellipsoidal, but in the case of natural images the 2D and 3D contour surfaces display striking polyhedra-like shapes. Huang [43] showed that the peaks and cusps in the contours correspond to occurrences of simple geometries in images with partially constant intensity regions and sharp discontinuities. These results points to the ubiquity of "object-like" structures in natural images and underline the importance of object-based models. Grenander and Srivastava [36] also has attributed non-Gaussianity to the presence of objects in images and used that idea for model building. Lee et al. have shown that images synthesized from occlusion models (Section 3.1) show irregular polyhedra-like shapes in contour surfaces of histograms.

3. Emerging Statistical Models on Image Space

Earliest, and still widely used, probability models for images were based on Markov random field models (MRFs) [93]. An image field is as a collection of random variables, denoting pixel values on a uniformly spaced grid in the image plane. In MRFs, the conditional probability of a pixel value given the remaining image is reduced to a conditional probability given a neighborhood of that pixel. Efficiency results if the neighborhood of a pixel is small, and furthermore,

stationarity holds if the same conditional density is used at all pixels. Ising and Potts model are the simplest examples of this family. Besag [8, 9] expressed the joint density of image pixels as a product of conditional densities, and ignored the normalizer to obtain a pseudo-likelihood formulation. Clifford-Hammersely theorem, see for example [93], states that full conditionals completely specify the joint density function (under a positivity assumption) and enabled the analysis of images using a Gibbs sampler. Geman and Geman [33] utilized the equivalence of MRFs and Gibbs distributions to sample from these distributions. Kersten [47] worked on computing the conditional entropies of the pixel values, given the neighboring pixels in an MRF framework. Zhu and Mumford [98] used a Gibbs model on images and estimated model parameters using a minimax entropy criterion. Let $H(I)$ be a concatenation of the histograms of image coefficients, obtained using several wavelet bases. Then the maximum entropy probability takes the form: $P(I \mid \lambda) \propto e^{-\langle \lambda, H(I) \rangle}$. The vector λ is estimated by setting the mean histograms to equal the observed histograms, i.e. $\int H(I)P(I \mid \lambda)\, dI = H_{obs}$, and reduces to the maximum likelihood estimation of λ under $P(I_{obs} \mid \lambda)$.

3.1. Models Motivated by Physics

A number of researchers have studied image statistics from a physical viewpoint, trying to capture the phenomena that generates images in the first place. A common theme to all these models is random placements of planar shapes (lines, templates, objects, etc.) in images according a Poisson process. Different models differ in their choice of shapes (e.g. primitive versus advanced), and their interaction (e.g. some favor occlusion while others favor superposition). Here we summarize a few of these models:

3.1.1. Superposition Models. In order to capture scale invariance and non-Gaussianity of images, Mumford and Gidas [61] utilized a family of infinitely divisible distributions. They showed that such distributions arise when images are modeled as superpositions of random placements of objects. To achieve self-similarity, it was proposed that the sizes of objects present in images be distributed according to density function Zr^{-3} over a subset of $\mathbb{R}_+$ (Z is the normalizing constant). Using this model, Chi [19] described a Poisson placement of objects with sizes sampled according to the $1/r^3$-law. Additionally, he assumed a surface process that models the texture variation on the 2D surfaces of the objects. Bitouk et al. [11] and Grenander and Srivastava [36] have also assumed that images are made up of 2D appearances of the objects g_is placed at homogeneous Poisson points z_is. The objects are chosen randomly from an object space (arbitrary sizes, shapes or texture). A weighted superposition, with weights (or contrasts) given by independent standard normals a_is, is used to form an image; it allows for expressing the linearly filtered images as a superposition of filtered objects. This formulation leads to an analytical (parametric) probability *for individual images* with the final form given in Section 3.2.

3.1.2. Occlusion Models. In more sophisticated models, objects are placed randomly as earlier but now the objects in the front occlude those in the back. An example is the dead leaves (also called a random collage) model which assumes that images are collages of approximately independent objects that occlude each other. One generates images from the model by placing an ordered set of elementary 2D shapes g_is in layers. The locations z_is with sizes r_is are sampled from a Poisson process. For gray level images, the random sets or "leaves" $T_i = g_i(z - z_i)$ are furthermore associated with intensities a_i. The leaves are typically placed front-to-back[1] according to

$$I^{(i)}(z) = \begin{cases} a_i & \text{if } z \notin T_j \; z \in T_i, \forall j < i \\ I^{(i-1)}(z) & \text{otherwise} \end{cases}, \quad \text{where } i = 1, 2, \ldots.$$

Left two panels in Fig. 2 show two samples of a dead leaves model with elliptical and square random shapes, respectively. The final model is defined in the limit $\lim_{i\to\infty} I^{(i)}(z)$, or equivalently, when the finite image domain $\Omega \subset \mathbb{R}^2$ is completely covered. The dead leaves model dates back to Matheron [57] and Serra [72] in mathematical morphology. They showed that the probability of any compact set $K \subset \Omega$ belonging to the same leaf and not being occluded by other leaves equals the ratio $\frac{E[\nu(X_0 \ominus \check{K})]}{E[\nu(X_0 \oplus \check{K})]}$. Here, X_0 is the random shape used in the model, $E[\nu(\cdot)]$ is the expected Lebesque measure, and $X_0 \ominus \check{K} = \{x : x + K \in X_0\}$ and $X_0 \oplus \check{K} = \{x : X_0 \cap (x + K) \neq \emptyset\}$ represent the *erosion* and the *dilation*, respectively, of X_0 by K. Early applications of this model include studies of inhomogeneous materials and the size distribution of grains in powder [45]. More recently, researchers have applied different versions of the dead leaves model to natural

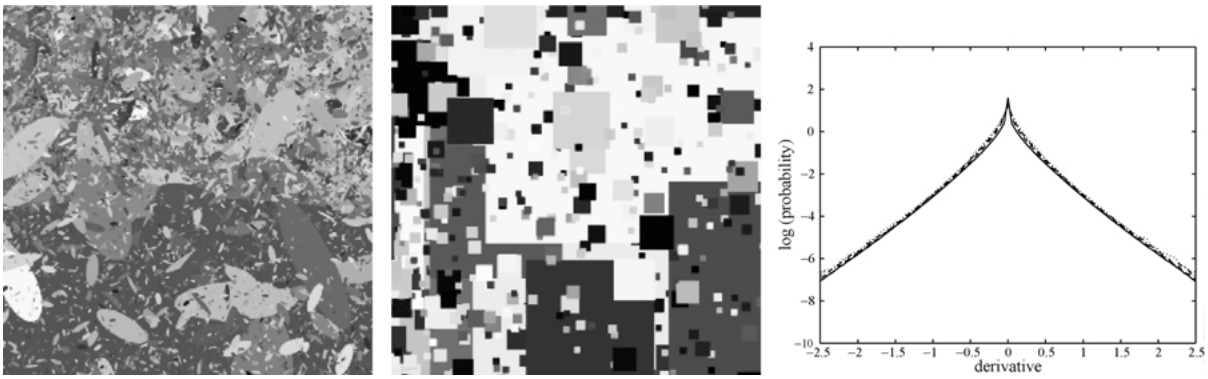

Figure 2. Left two pictures: Two different versions of the dead leaves model with elliptical (left) and square (middle) random shapes, respectively. Right panel: Derivative statistic and scaling of a dead leaves model with disks distributed according to a cubic law of sizes. The curves correspond to the marginal distribution of the differences between adjacent pixels after contrast normalization at four different scales ($N = 1, 2, 4, 8$).

scenes. In [70], Ruderman proposed that a random collage description with statistically independent 'objects' can explain scaling of correlations in images. Using an occluding object model and the difference function of hand-segmented images, he infers a power law for the probability of two points belonging to the same object. This result is consistent with [1] that uses level sets and the original dead leaves formalism of Matheron and Serra. Here, Alvarez et al. analyze the morphological structure of natural images—in particular, the effect of occlusions on the area, perimeter and intercept lengths of homogeneous, connected image regions.

Lee and Mumford [50] have used a dead leaves model with statistically independent objects and the measure of a marked Poisson process to show that $1/r^3$ power law in object sizes also leads to approximate full scale invariance when occlusions are taken into account. Provided that synthesized images are smoothened by, for example by block averaging, their statistics match those observed in large data sets of calibrated natural images. Scale invariance of histograms of filter responses, full co-occurrence statistics of two pixels, and joint statistics of Haar wavelet coefficients are studied in this paper. Figure 2 (right panel) shows a histogram of the derivative statistic, under scaling, in a dead leaves model with smoothing.

In addition, there have been some analogies drawn from other sciences to explain self-similarity of natural images. For example, Ruderman et al. [69, 71] utilized the statistics of turbulent flows. In [83], Turiel et al. formally connected the statistical modeling of turbulent flows to that of images. In turbulent flows, if the second moment of the energy dissipation is proportional to r^τ for some τ, where r is the box size, then self-similarity holds. Using a measure based on edges to compute energy dissipation in images, the authors showed self-similarity of edge statistics.

3.2. *Analytical Densities: Univariate Case*

An important requisite of statistical models is that they be computationally efficient for real-time, or near real-time, applications, and one obvious way to accomplish that is via parametric densities. We want to represent statistical nature of images by means of parametric densities, using only limited parameters. In this section, we discuss families of parametric densities that seems to capture the variability in low-dimensional representations of images. Rather than explaining scale-invariance of a large collection of images, the goal here is to *capture variability of individual images*. For an image I and a linear filter F, this section deals with characterizing the marginal density of the pixel values in $I * F$.

1. *Generalized Laplacian Model*: Marginal densities of image coefficients are well modeled by a generalized Laplacian density (also called generalized Gaussian or stretched exponential) $f_1(x; c, p) = \frac{e^{-|x/c|^p}}{Z_1(p,c)}$, [55, 60, 76], where the normalizer is $Z_1(p, c) = 2\frac{c}{p}\Gamma(\frac{1}{p})$. The parameters, $\{p, c\}$, may be estimated for the subbands of specific images using maximum likelihood or the method of moments. Another way of estimating them is via the (linear) regression of $\log(\log(h(x) + h(-x)) - 2\log(h(0)))$ versus $\log(|x|)$, where $h(x)$ is the histogram value at x and x is the variable for bin centers. Values for the exponent p are typically within the range [0.5, 0.8],

and the width parameter c varies monotonically with the size of the basis functions, producing higher variance for coarser-scale components [76].

Under Grenander's formulation of a superposition model (item 1, Section 3.1), if the random variable $u(z) \equiv \sum_{i=1}^{n}((g_i * F)(z - z_i))^2$ is assumed to have a scaled-Gamma density, then the univariate density of the filtered pixel has been shown in [36] to be: for $p > 0$, $c > 0$, $f_2(x; p, c) = \frac{1}{Z_2(p,c)}|x|^{p-0.5}K_{(p-0.5)}(\sqrt{\frac{2}{c}}|x|)$, where K is the modified Bessel function of third kind, and Z_2 is the normalizing constant. (Its characteristic function is of the form $1/(1 + 0.5c\omega^2)^p$ which for $p = 1$ becomes the characteristic function of a Laplacian density.) This parametric family of densities has been called *Bessel K forms* with (p, c) referred to as *Bessel parameters*. Earlier, Wainwright et al. [89] also investigated the application of the Gaussian scale mixture family to image modeling and referred to f_2 as the *symmetrized Gamma* density (without reaching its analytic, parametric form). As described in [36], p and c are easily estimated from the observed data, with $\hat{p} = \frac{3}{SK(I^{(j)})-3}$ and $\hat{c} = \frac{SV(I^{(j)})}{\hat{p}}$, where SK is the sample kurtosis and SV is the sample variance of the filtered image. A distinct advantage of this model is that parameters can be related mathematically to the physical characteristics of the objects present in the image. If a filter F is applied to an image I to extract some specific feature—vertical edges, say—then, the resulting p has been shown in [80] to depend on two factors: (i) distinctness and (ii) frequency of occurrence of that feature in I. Objects with sharper, distinct edges have low p values, while scenes with many objects have large p values.

2. *Gaussian Scale Mixtures*: Observed statistics of filtered images have pointed towards a rich family of univariate densities that are highly kurtotic and heavy tailed. In statistics literature, there is a wide usage of the variables defined as *normal variance-mean mixture*; X is called a normal variance-mean mixture if the conditional density function of X given u is normal with mean $\mu + u\beta$ and variance $u\Delta$, and u is called the mixing variable. Generalized hyperbolic distributions were introduced by Barndorff-Nielsen [3] as specific normal variance-mean mixtures that result when the mixing variable u is of certain class. For $\mu = \beta = 0$ and $\Delta = 1$, the resulting family is also called Gaussian scale mixture [2] and has seen applications in financial mathematics [12] and speech processing [13]. Furthermore, if u is a scaled Gamma density then Bessel K density results.

Shown in Fig. 3 are some examples of estimating this density function: natural images from van Hateren databse are filtered using arbitrary Gabor filters (not shown), and the resulting pixel values are used to form the histogram $h(x)$. Bottom panels of this figure show the plots of $\log(h(x))$, $\log(f_1(x))$, and $\log(f_2(x))$ with parameters estimated from the corresponding images.

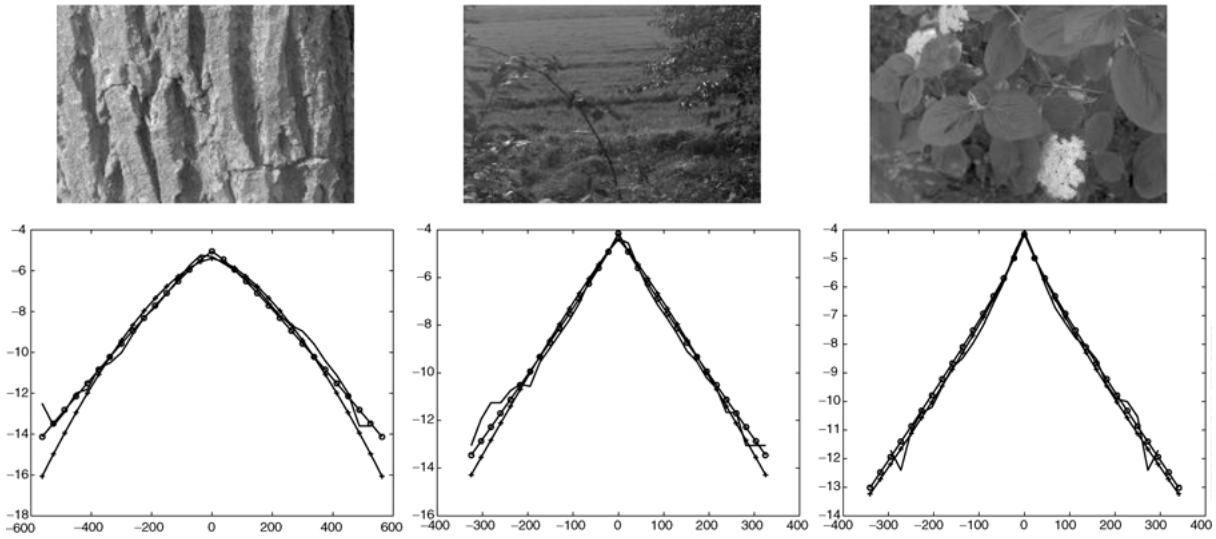

Figure 3. Estimated Bessel K and generalized Laplacian densities compared to the observed histograms for the images in top. Plain lines are histograms, lines with large beads are Bessel K, and lines with small beads are generalized Laplacian. These densities are plotted on a log scale.

There are several ways to judge the performance of any proposed probability model. The simplest idea is to compare the observed frequencies to the frequencies predicted by the model, using any metric on the space of probability distributions. Two commonly used metrics are the Kolmogorov-Smirnov distance: $\max_{x\in\mathbb{R}} |\int_{-\infty}^{x}(f_1(y) - f_2(y))\,dy|$ and the Kullback-Leibler divergence: $\int_{\mathbb{R}} f_1(x)\log(\frac{f_1(x)}{f_2(x)})\,dx$. Both the generalized Laplacian and the Bessel K form perform well in capturing univariate image statistics under these metrics. In general, for small values of p (sharper features and less number of objects in image) the generalized Laplacian performs better while for larger p the Bessel K density better matches the observed densities.

3.3. Bivariate Probability Models

So far we have discussed only univariate models but the complexity of observed images points to statistical interactions of high orders. As a first extension, we look at the models for capturing pairwise interactions between image representations, i.e. the bivariate probability densities. For instance, these models may involve a pair of wavelet coefficients at different scales, positions, or orientations.

3.3.1. Variance Dependency and Gaussian Scale Mixtures. As described in Section 2.4, pairs of wavelet coefficients corresponding to basis functions at nearby positions, orientations, or scales are not independent, even when the responses are (second-order) decorrelated. Specifically, the conditional variance of any given coefficient depends quite strongly on the values of surrounding coefficients [74]. As one considers coefficients that are more distant (either in spatial position, orientation, or scale), the dependency becomes weaker. This dependency appears to be ubiquitous, and may be observed across a wide variety of image types.

What sort of probability model can serve to explain the observations of variance dependency? One candidate is the Gaussian scale mixture model described earlier for the univariate case. In this scheme, wavelet coefficients are modeled as the product of a Gaussian random variable and a hidden "multiplier" random variable (same as mixing variable u earlier) that controls the variance. To explain pairwise statistics shown in Fig. 1, one can prescribe a relationship between the hidden multiplier variables of neighboring coefficients. That is, the hidden variables now depend on one another, generating the variance scaling seen in the data. Modeling of this dependency remains an ongoing topic of research. One possibility is to link these variables together in a Markov tree, in which each multiplier is independent of all others when conditioned on its parent and children in the tree [89].

3.3.2. Bivariate Extension of Bessel K Forms. The univariate Bessel K form has been extended by Grenander to specify bivariate densities of the filtered images. The basic idea is to model all linear combinations of the filtered pixels by Bessel K forms and then invoke the Cramer-Wold device, see [10] for a definition, which states that specification of densities of all linear combinations of a number of random variables specifies uniquely their joint density. The use of the Cramer-Wold device here assumes that there exists a 2D distribution whose Radon-like marginals (for half spaces, not on lines) behave according to Bessel K densities. An imposition of Bessel K forms on marginals seems to agree well with the data qualitatively but its performance remains to quantified over large datasets.

Let $I_1 = I * F^{(1)}$ and $I_2 = I * F^{(2)}$ for two wavelet filters $F^{(1)}$ and $F^{(2)}$, and for $a_1, a_2 \in \mathbb{R}$, let $J(a_1, a_2, z) \equiv a_1 I_1(z) + a_2 I_2(z)$. The Cramer-Wold idea is to compute the characteristic functions of $J(a_1, a_2, z)$, for all pairs (a_1, a_2), using Bessel parameter estimation, and then take an inverse Fourier transform to obtain an estimate of the joint density $f(I_1(z), I_2(z))$. This estimate is parametrized by eight joint moments: $\mu_{0,2}, \mu_{1,1}, \mu_{2,0}, \mu_{4,0}, \mu_{3,1}, \mu_{2,2}, \mu_{1,3}, \mu_{0,4}$, where $\mu_{i,j} = \int\int I_1^i I_2^j f(I_1, I_2)\,dI_1\,dI_2$. Shown in Fig. 4 is an example of this bivariate density estimation. For the image shown in left, consider the bivariate density of its two filtered versions under two Gabor filters at the same scale but orientations 20 degrees apart. Top two panels show the mesh and the contour plots of the estimated density while the bottom panels show the observed bivariate histogram. The densities are all plotted on a log scale.

4. Discovering Image Manifolds

It has been well highlighted that in the space of rectangular arrays of positive numbers, only a small subset has images of natural scenes. One seeks to isolate and characterize this subset for use in image analysis applications. The main idea is to identify this set as a low-dimensional, differentiable manifold and use its geometry to characterize images. Having defined this manifold, a simplistic probability model can help capture the image variability. We now present a summary

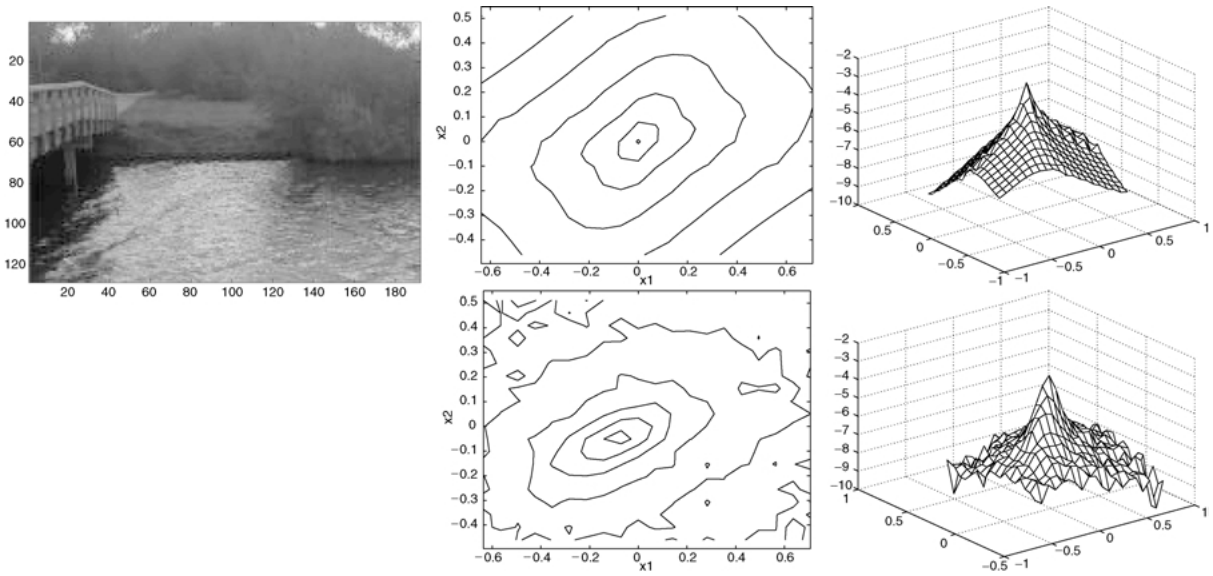

Figure 4. Bivariate extension of Bessel K form: For the image shown in left panel, the estimated (top) and observed (bottom) bivariate densities plotted as meshes and contours, on a log scale.

of some commonly used methods for estimating image manifolds.

4.1. Approximating Image Manifolds by Linear/Locally-Linear Subspaces

Perhaps the easiest technique to approximate the image manifold is to fit a linear subspace through the observations. The fitting criterion will specify the precise subspace that is selected. For example, if one intends to minimize the cumulative residual error (Euclidean distance between the observation points and the subspace), then the optimal subspace is the dominant (or principal) subspace of the data matrix, and is easily computed using eigen decomposition or singular value decomposition. A common application of PCA is in recognition of people from their facial images [48] or study of natural images [39]. Instead, if the goal is to minimize the statistical correlation between the projected components, or to make them as independent as possible, then the independent component basis results [6, 23, 86]. Other criteria lead to similar formulations of the subspace basis such as sparsity [63], Fisher discrimination [5], and non-negative factorization [52]. The use of sparseness is often motivated by the scale invariance of natural images.

Approximating the image manifold by a flat subspace is clearly limiting in general situations. In [96], the authors argue that linear processing of images leaves substantial dependencies between the components, and a nonlinear technique is required. One extension is to seek a "local linear embedding" approximation of the image manifold by fitting neighboring images by low dimensional subspaces [68, 81]. Definition of a neighborhood is through Euclidean metric but that deserves further study. Another idea is to combine local basis elements, such as wavelet bases, into higher level structures that provide a better representation of the image manifold. For instance, Zhu et al. [97] have combined placements of transformed basis elements to form structures called *textons* in order to better characterize images and their manifolds. Shown in the left panel of Fig. 5 is an example of building a "star" texton to match the given image.

4.2. Deformable Template Characterization of Image Manifold

Grenander's pattern theoretic framework for image understanding is motivated by physical models and leads to probabilistic, algebraic representations of scenes. Objects appearing in the scenes are represented by 3D models of shape, texture, and reflectivity, and their occurrences in scenes are captured by group transformations on the typical occurrences or templates. A strong feature of this approach is the logical separation between the actual, physical scenes and the resulting images. Variability of scenes is better modeled in a

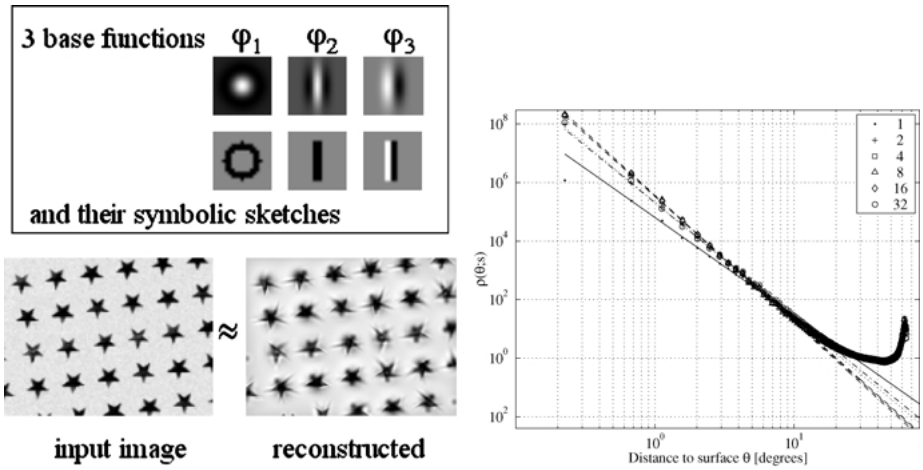

Figure 5. Left panel: To model an image of stars, some wavelet bases are combined to form a star texton. Right Panel: Natural images have an infinite probability density at a manifold in state space that corresponds to blurred edges. The figure shows the empirical probability density as a function of the distance to the manifold of Gaussian scale-space edges in jet space, with curves corresponding to jet representations of natural images at scales $s = 1, 2, 4, 8, 16, 32$.

3D coordinate system as it is guided by the physical principles, rather than the image space where the Euclidean representations do not work well. Since the two, 3D scenes and 2D images, are related by a projection (orthographic or perspective) map, manifolds formed by images can be generated by projecting the manifolds formed using 3D representations. As described in [34, 59], occurrences of physical objects in 3D system are modeled as the orbits of groups acting on the configurations of objects in the scenes. These orbits are projected into the image space to form manifolds on which the images lie. Let C^α be a CAD model of a 3D object labeled by α and let S be the group of transformations (3D rotation, translation, deformation, etc.) that change the object occurrences in a scene. Let, for $s \in S$, sC^α denote the action of s on the template. Then, $\{sC^\alpha : s \in S\}$ is a 3D orbit associated with object α. Further, if T is the imaging projection from a 3D scene to the image plane, then $\{T(sC^\alpha) : s \in S\}$ is the image manifold generated by this object, and dimension of this image manifold equals $\dim(S)$. Geometry of this manifold, such as the tangent spaces or the exponential maps, can also be obtained by projecting their counterparts in the bigger space.

4.3. *Empirical Evidence of Manifolds for Image Primitives*

In contrast to a physical specification, as advocated by the deformable template theory, one can also study image manifolds directly using the observed images. The image based search [51, 64] is also inspired by Marr's idea [56] of representing early vision by converting an array of raw intensities into a symbolic representation—a so called "primal sketch"—with primitives such as edges, bars, blobs and terminations as basic elements. An important issue in computer and human vision is: How are Marr's primitives represented geometrically and statistically in the state space of image data? The geometrical model that arises from Marr's hypothesis is a set of continuous manifolds of the general form $M(s) = [F^{(1)}(\cdot) * sC^\alpha, F^{(2)}(\cdot) * sC^\alpha, \ldots]$, $F^{(j)}(j = 1, \ldots, n)$ form a bank of filters, and sC^α is an image of a primitive α parameterized by $s \in S$. Dimension of resulting image manifold $M(s)$ is determined by an intrinsic dimensionality of the primitives; it is usually a small number (2 for contrast-normalized edges, 3 for bars and circular curves etc.). Furthermore, the manifolds generated by planar primitives form a hierarchical structure: The 2D manifold of straight edges, for example, is a subset of both the 3D manifold of bars and the 3D manifold of circular edges.

In [51], Lee et al. found that the state space of 3×3 natural image patches is extremely sparse with the patches densely concentrated around a non-linear manifold of blurred edges. For the top 20% highest contrast patches in natural images, half of the data lie within a neighborhood of the edge manifold that occupies only 9% of the total state space. Estimated probability density, as a function of the distance to the manifold, takes the form $f(\text{dist}) \sim 1/\text{dist}^\beta$ with $\beta = 2.5$; it has an

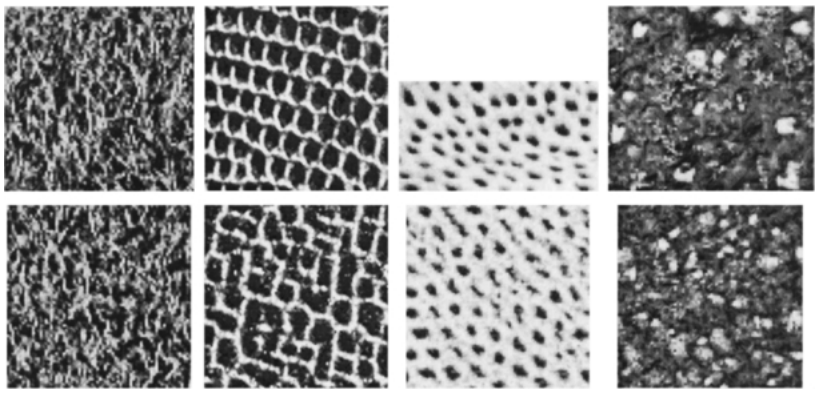

Figure 6. Texture synthesis using Julesz ensemble model (first two columns) and the Gibbs model (last two columns). Top panels: observed real images and bottom panels: synthesized images.

infinite density at the image manifold where dist $= 0$. The paper [64] extends these results by considering filtered patches with the filters being (up to third order) derivatives of Gaussian kernels. Figure 5 right panel shows the estimated probability density as a function of the distance to the edge manifold for multiscale representations of natural images. The density is approximately scale invariant and seems to converge toward the functional form $p(\text{dist}) \sim 1/\text{dist}^{0.7}$.

5. Applications of Statistical Models

Main reason for developing formal statistical models is to apply them in image processing/analysis applications. There is a large array of applications that continually benefit from advances in modeling of image statistics. Here we have selected an important subset.

5.1. Texture Synthesis

Use of newer statistical models has revolutionized the area of texture analysis. In 1980 Faugeras et al. [29] suggested using the marginals of filtered images for texture representations. Bergen and Landy [7], Chubb et al. [21], and Heeger and Bergen [41] also advocated the use of histograms. Zhu et al. [99] showed that marginal distributions of filtered images, obtained using a collection of filters, sufficiently characterize homogeneous textures. Choice of histograms implies that only the frequencies of occurrences are retained and the locations are discarded. Exploiting periodicity of these textures, one extracts features using wavelet decompositions at several scales and orientations, and uses them to represent images. Many schemes have been proposed, two of them are: (i) Julesz ensemble that impose equivalence on all images that lead to the same histograms, and (ii) Gibbs model stated earlier in Section 3. Shown in Fig. 6 are some examples of texture synthesized under these models: the top row shows the real images and the bottom row shows corresponding synthesized textures. It must be noted that this synthesis framework holds well even when the raw histograms are replaced by parametric analytical forms, as shown in [80].

Beyond the use of marginals, Portilla and Simoncelli [65] developed a model based on a characterization of joint statistics of filter responses. Specifically, they measured the correlations of raw coefficients as well as the correlations of their magnitudes, and developed an efficient algorithm for synthesizing random images, subject to these constraints, by iteratively updating the image and forcing the constraints.

5.2. Image Compression

Compression seems to be a natural application for emerging statistical image models. In a typical implementation, the image is decomposed using a linear basis, the coefficients of this representation are quantized to discrete values, and these values are then efficiently encoded by taking advantage of their relative probability of occurrence. In this context, the statistical models considered here provide estimates of these probabilities. Most widely used compression scheme for images is the JPEG standard, which is based on a block-by-block frequency decomposition. In early 80s, it was recognized that multi-scale wavelet-style representations offer more flexibility and better compression

performance [15, 87]. Early coders were based on marginal models of coefficients, but this changed abruptly with the development of the first contextual coder by Shapiro [73]. This coder, and many that followed [20, 67], took heuristic advantage of the joint statistical properties of wavelets. Some subsequent coders have been based more explicitly on such models [14, 54].

5.3. Image Classification

An interesting application is to classify images into some pre-defined categories. If certain lower order statistics of images are found sufficient for this purpose, some level of efficiency can be achieved. We start the discussion with the classical methods for classification.

5.3.1. Multiple Hypothesis Testing. A classical technique for classification, given the probability models associated with the classes, is hypothesis selection [37]. Given an observation, the goal is to select the hypothesis that is most likely to have generated that observation. Let $H_1, H_2, \ldots, H_n$ correspond to the n image classes, and let $P(I \mid H_i)$ be the likelihood of image I belonging to class H_i, then hypothesis selection can be accomplished as a sequence of binary hypothesis tests: $\frac{P(I|H_i)}{P(I|H_j)} \underset{H_j}{\overset{H_i}{\gtrless}} \nu_{ij}$, where ν_{ij} is a threshold value generally taken to be one. Neyman-Pearson lemma provides optimality to hypothesis testing in a certain sense. In case of a Bayesian selection, the threshold ν_{ij} is given by the ratio of priors $\nu_{i,j} = P(H_j)/P(H_i)$. For all these tests, one needs the likelihood function $P(I \mid H_i)$ which will depend upon the choice of statistical models. For example, if a parametric form is chosen to model the image probability, then the classes can be directly related to the parameter values. Typical values ("average values") of the parameter represent a class and the likelihood $P(I \mid H_i)$ is written in the parametric form with the typical parameter values for each class. Deformable templates (Section 4), has been successfully applied to object recognition. The inference is obtained by hypothesis testing for α in presence of nuisance variables $s \in S$. Here, the likelihood is computed via the nuisance integral $P(I \mid H_i) = \int_S P(I, s \mid H_i)\gamma(ds)$. Nuisance variable estimation on a group S and estimation error bounds are derived in [35], while the hypothesis testing for α and recognition error bounds are derived in [37].

5.3.2. Metrics for Image Comparison. A broader goal in image analysis is to quantify differences between two given images. Given such a metric one can perform image clustering, image retrieval, image classification, and even recognition of objects in given images. If the probability models are parametric, one can derive a distance measure on the image space that takes a parametric form. If $f(x \mid p_1, c_1)$ and $f(x \mid p_2, c_2)$ are two univariate density functions parameterized by (p_1, c_1) and (p_2, c_2) respectively, then the distance measure takes the form: $d(p_1, c_1, p_2, c_2) = \tilde{d}(f(x \mid p_1, c_1), f(x \mid p_2, c_2))$, where $\tilde{d}$ is a metric on the space of univariate densities. Several forms have been proposed for $\tilde{d}$ including geodesic length (Riemannian metric), Earth Mover's distance [22], Kullback-Leibler divergence, Renyi's α-divergence [42], Jensen-Renyi divergence [38, 40], χ^2-distance, and the L^p norm for $p = 1, 2, \ldots,$. The choice of metric depends upon the application and the desired computational simplicity. Srivastava et al. [80] have derived a parametric form of L^2 metric between two Bessel K forms. Two densities can be compared directly using their estimated parameters without requiring to compute full histograms.

5.4. Image Denoising

A common approach here is to decompose image into bands of spatial frequency and to threshold the coefficients after some nonlinear transformation, proposed first by Bayer and Powell [4] and later by Donoho [26]. The nonlinear transformation is used essentially to shrink all wavelet coefficients towards zero. This shrinking is based on thresholding which can be implemented as a hard threshold or a soft threshold. For applying statistical approaches, such as Bayesian or MAP techniques, an explicit prior model on the image pixels is required. Simoncelli and Adelson [76] studied image denoising in a Bayesian framework while Moulin and Liu [60] reported the use of generalized Laplacian models in statistical denoising approaches. The fact that statistics of decomposed images are easier to characterize, than that of the original images, has led to many pyramid-based approaches to image denoising. Images are decomposed into multi-scale representations and statistics of coefficients are used for denoising [53, 66, 75] in individual frequency bands. Also, the Gaussian scale mixture model for joint statistics can be used in a Bayesian framework [66], producing results that are significantly better than those achieved with a marginal model.

5.5. *Other Applications*

There are a number of other image analysis/synthesis applications that have benefited from statistical ideas. Yendrikhovskij [94] uses a clustering of color statistics to model the perceived color environment and to compute categories of coloring. A bidirectional radiosity distribution function (BRDF) of an object completely specifies its appearance under arbitrary illumination conditions. Dror et al. [28] have shown that statistics of these illumination maps are similar to the statistics of natural images, and hence the proposed image models can apply there as well [92].

6. Discussion

In this paper we have discussed some recent advances in statistical modeling of natural images. Not only these models provide a better match than the traditional models but also have lead to significant improvements in a number of imaging applications.

Although substantial progress has been made over the past twenty years in understanding complex statistical properties of natural images, we are still quite far from a full probability model. For example, samples drawn from existing models are unable to capture the variety and complexity of real world images, except in the restricted case of homogeneous textures. Beyond univariate and bivariate densities of the image statistics, the computational complexity increases exponentially. An important question is: In the context of a specific imaging application, say face recognition from video images, what order densities are required to ensure a reasonable success? For homogeneous textures, univariate models have been successful but how much more is needed for more general applications. The notion of sufficient statistics needs to be made precise for different application contexts. Even among the proposed models, several issues remain open. For instance, one issue in dead leaves model is how to incorporate texture and dependencies of objects. At this point, there are few analytical results for realistic dead leaves models.

Many of the models described in this paper model statistics of an ensemble of images; their applications for analysis of individual images needs to be clarified.

Beyond applications, such as synthesis and compression of images, an important reason for developing statistical models is image understanding, an area with many outstanding problems. One such problem is: Given an image outside a rectangle, find the most probable interpolation of it inside of the rectangle. Lack of Markovity disallows the classical harmonic analysis, and points to more powerful pattern-theoretic structures underlying the image ensembles.

Acknowledgments

We thank the creators of van Hateren database for making their database available to public. We express gratitude to our collaborators Ulf Grenander, Kim S. Pedersen and David Mumford for their help in this research. Many thanks to Hamid Krim for his invitation and support in writing this paper. AS has been supported in part by the grants NSF DMS-0101429, ARO DAAD19-99-1-0267, and NMA 201-01-2010. EPS was supported in part by NSF CAREER grant MIP-9796040, the Alfred P. Sloan Foundation, and the Howard Hughes Medical Institute.

Note

1. see [46] for a discussion on "forward" and "backward" dead leaves algorithms.

References

1. L. Alvarez, Y. Gousseau, and J.-M. Morel, "The size of objects in natural and artificial images," in *Advances in Imaging and Electron Physics*, P.W.H. et al. (Eds.), Academic Press: New York, 1999.
2. D. Andrews and C. Mallows, "Scale mixtures of normal distributions," *Journal of the Royal Statistical Society*, Vol. 36, pp. 99–102, 1974.
3. O. Barndorff-Nielsen, "Exponentially decreasing distribution for the logarithm of a particle size," *Journal of Royal Statistical Society A*, Vol. 353, pp. 401–419, 1977.
4. B.E. Bayer and P.G. Powell, "A method for the digital enhancement of unsharp, grainy photographic images," *Advances in Computer Vision and Image Processing*, Vol. 2, pp. 31–88, 1986.
5. P.N. Belhumeur, J.P. Hepanha, and D.J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 711–720, 1997.
6. A.J. Bell and T.J. Sejnowski, "The independent components of natural scenes are edge filters," *Vision Research*, Vol. 37, No. 23, pp. 3327–3338, 1997.
7. J.R. Bergen and M.S. Landy, *Computational Modeling of Visual Texture Segregation*, MIT Press: Cambridge, MA, 1991, pp. 253–271.

8. J. Besag, "Spatial interaction and statistical analysis of lattice systems (with discussion)," *J. Royal Statist. Soc.* B, Vol. 36, pp. 192–326, 1974.
9. J. Besag, "On the statistical analysis of dirty pictures," *J. Royal Statistical Society* B, Vol. 48, No. 3, pp. 259–302, 1986.
10. P. Billingsley, *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics, 1995.
11. D. Bitouk, U. Grenander, M.I. Miller, and P. Tyagi, "Fisher information measures for ATR in clutter," in *Proceedings of SPIE, Automatic Target Recognition XI*, 2001, Vol. 4379, pp. 560–572.
12. T. Bollerslev, K. Engle, and D. Nelson, *ARCH Models*, Vol. IV, North-Holland: Amsterdam, 1994, pp. 2959–3038.
13. H. Brehm and W. Stammler, "Description and generation of-spherically invariant speech-model signals," *Signal Processing*, Vol. 12, pp. 119–141, 1987.
14. R.W. Buccigrossi and E.P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Transactions on Image Processing*, Vol. 8, No. 12, pp. 1688–1701, 1999.
15. P.J. Burt and E.H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, Vol. 31, No. 4, pp. 532–540, 1983.
16. G.J. Burton, N.D. Haig, and I.R. Moorhead, "A self-similar stack model for human and mahcine vision," *Biological Cybernetics*, Vol. 53, No. 6, pp. 397–403, 1986.
17. G.J. Burton and I.R. Moorhead, "Color and spatial structures in natural scenes," *Applied Optics*, Vol. 26, No. 1, pp. 157–170, 1987.
18. J.-F. Cardoso, "Source separation using higher order moments," in *Proceedings of ICASSP*, 1989, pp. 2109–2112.
19. Z. Chi, "Probability models for complex systems," Ph.D. Thesis, Division of Applied Mathematics, Brown University, 1998.
20. C. Chrysafis and A. Ortega, "Efficient context-based entropy coding for lossy wavelet image coding," in *Data Compression Conf.*, 1997, Snowbird, Utah.
21. C. Chubb, J. Econopouly, and M.S. Landy, "Histogram contrast analysis and the visual segregation of IID textures," *J. Opt. Soc. Am.* A, Vol. 11, pp. 2350–2374, 1994.
22. S. Cohen and L. Guibas, "The earth mover's distance under transformation sets," in *Proceedings of Seventh IEEE International Conference on Computer Vision*, Vol. 2, pp. 1076–1083, 1999.
23. P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, Special Issue on Higher-Order Statistics, Vol. 36, No. 3, 1994.
24. J. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *Journal of the Optical Society of America* A, Vol. 2, No. 7, pp. 23–26, 1985.
25. N.G. Deriugin, "The power spectrum and the correlation function of the television signal," *Telecommunications*, Vol. 1, pp. 1–12, 1956.
26. D. Donoho, "Denoising by soft-thresholding," *IEEE Trans. Info. Theory*, Vol. 43, pp. 613–627, 1995.
27. D.L. Donoho and A.G. Flesia, "Can recent innovations in harmonic analysis 'Explain' key findings in natural image statistics," *Network: Computation in Neural Systems*, Vol. 12, No. 3, pp. 371–393, 2001.
28. R.O. Dror, T.K. Leung, E.H. Adelson, and A.S. Willsky, "Statistics of real-world illumination," in *Proceedings of 2001 IEEE Conference on CVPR*, 2001, Vol. 2. pp. 164–171.
29. O.D. Faugeras and W.K. Pratt, "Decorrelation methods of texture feature extraction," *IEEE Pat. Anal. Mach. Intell.* Vol. 2, No. 4, pp. 323–332, 1980.
30. D.J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Optical Society of America*, Vol. 4, No. 12, pp. 2379–2394, 1987.
31. D. Gabor, "Theory of Communications," *Journal of IEE (London)*, Vol. 93, pp. 429–457, 1946.
32. D. Geman and A. Koloydenko, "Invariant statistics and coding of natural microimages," in *Proc. of the IEEE Workshop on Statistical and Computational Theories of Vision*, 1999.
33. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, No. 6, pp. 721–741, 1984.
34. U. Grenander, *General Pattern Theory*, Oxford University Press, 1993.
35. U. Grenander, M.I. Miller, and A. Srivastava, "Hilbert-Schmidt lower bounds for estimators on matrix lie groups for ATR," *IEEE Transactions on PAMI*, Vol. 20, No. 8, pp. 790–802, 1998.
36. U. Grenander and A. Srivastava, "Probability models for clutter in natural images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 4, pp. 424–429, 2001.
37. U. Grenander, A. Srivastava, and M.I. Miller, "Asymptotic performance analysis of Bayesian object recognition," *IEEE Transactions of Information Theory*, Vol. 46, No. 4, pp. 1658–1666, 2000.
38. A.B. Hamza, Y. He, and H. Krim, "An information divergence measure for ISAR image registration," in *Proc. of IEEE Workshop on Statistical Signal Processing*, 2001.
39. P.J.B. Hancock, R.J. Baddeley, and L.S. Smith, "The principal components of natural images," *Network*, Vol. 3, pp. 61–70, 1992.
40. Y. He, A.B. Hamza, and H. Krim, "A generalized divergence measure for robust image registration," *IEEE Transactions on Signal Processing*, 2002, to appear.
41. D.J. Heeger and J.R. Bergen, "Pyramid-based texture analysis/synthesis," in *Proceedings of SIGGRAPHS*, 1995, pp. 229–238.
42. A.O. Hero, B. Ma, O. Michel, and J. Gorman, "Alpha-divergence for classification, indexing and retrieval," Communication and Signal Processing Laboratory, Technical Report CSPL-328, U. Mich, 2001.
43. J. Huang, "Statistics of natural images and models," Ph.D. Thesis, Division of Appled Mathematics, Brown University, RI, 2000.
44. A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley and Sons; New York, 2001.
45. D. Jeulin, I. Terol-Villalobos, and A. Dubus, "Morphological analysis of UO2 powder using a dead leaves model," *Microscopy, Microanalysis, Microsctructure*, Vol. 6, pp. 371–384, 1995.
46. W.S. Kendall and E. Thonnes, "Perfect simulation in stochastic geometry," Preprint 323, Department of Statistics, University of Warwick, UK, 1998.
47. D. Kersten, "Predictability and redundancy of natural images," *Journal of Optical Society of America*, Vol. 4, No. 12, pp. 2395–2400, 1987.
48. M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, No. 1, pp. 103–108, 1990.
49. E.R. Kretzmer, "Statistics of television signals," *Bell Syst. Tech. Journal*, Vol. 31, pp. 751–763, 1952.
50. A.B. Lee and D. Mumford, "Occlusion models for natural images: A statistical study of scale-invariant dead leaves model," *International Journal of Computer Vision*, Vol. 41, No. 1/2, 2001.
51. A.B. Lee, K.S. Pedersen, and D. Mumford, "The nonlinear statistics of high-contrast patches in natural images," *Int. J. Computer Vision*, 2002, in press.
52. D.D. Lee and H.S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, Vol. 401, pp. 788–791, 1999.
53. D. Leporini, J.-C. Pesquet, and H. Krim, "Best basis representations with prior statistical models," in *Lecture Notes in Statistics: Bayesian Inferences in Wavelet Based Models*, P. Muller and B. Vidakovic (Eds.), Springer Verlag: Berlin, 1999.
54. S.M. LoPresto, K. Ramchandran, and M.T. Orchard, "Wavelet image coding based on a new generalized gaussian mixture model," in *Data Compression Conf.*, 1997, Snowbird, Utah.
55. S.G. Mallat, "A Theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, pp. 674–693, 1989.
56. D. Marr, *Vision: A computational Investigation into the Human Representation and Processing of Visual Information*, W.H. Freeman and Company: New York, 1982.
57. G. Matheron, *Random Sets and Integral Geometry*, John Wiley and Sons: New York, 1975.
58. K.D. Miller, "A Model for the development of simple-cell receptive fields and the arrangement of orientation columns through activity dependent competetion between on- and off-center inputs," *Journal of NeuroScience*, Vol. 14, pp. 409–441, 1994.
59. M.I. Miller and L. Younes, "Group actions, homeomorphisms, and matching: A general framework," *International Journal of Computer Vision*, Vol. 41, No. 1/2, pp. 61–84, 2002.
60. P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using a generalized Gaussian and complexity priors," *IEEE Trans. Info. Theory*, Vol. 45, pp. 909–919, 1999.
61. D. Mumford and B. Gidas, "Stochastic models for generic images," *Quaterly of Applied Mathematics*, Vol. 59, No. 1, pp. 85–111, 2001.
62. B.A. Olshausen and D.J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, Vol. 381, pp. 607–609, 1996.
63. B.A. Olshausen and D.J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?" *Vision Research*, Vol. 37, No. 23, pp. 3311–3325, 1997.
64. K.S. Pedersen and A.B. Lee, "Toward a full probability model of edges in natural images," in *Proc. of ECCV'02*, 2002.
65. J. Portilla and E.P. Simoncelli, "A parametric texture model based on join statistics of complex wavelet coefficients," *International Journal of Computer Vision*, Vol. 40, No. 1, pp. 49–70, 2000.
66. J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Adaptive Wiener denoising using a Gaussian scale mixture model in the wavelet domain," in *Proc 8th IEEE Int'l Conf on Image Proc*. Thessaloniki, Greece, IEEE Computer Society, 2001, pp. 37–40.
67. R. Rinaldo and G. Calvagno, "Image coding by block prediction of multiresolution subimages," *IEEE Trans Im Proc.*, 1995.
68. S.T. Roweis and L.K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, Vol. 290, pp. 2323–2326, 2000.
69. D.L. Ruderman, "The statistics of natural images," *Network*, Vol. 5, pp. 517–548, 1994.
70. D.L. Ruderman, "Origins of scaling in natural images," *Vision Research*, Vol. 37, No. 23, pp. 3385–3398, 1997.
71. D.L. Ruderman and W. Bialek, "Scaling of natural images: Scaling in the woods," *Physical Review Letters*, Vol. 73, No. 6, pp. 814–817, 1994.
72. J.P. Serra, *Image Analysis and Mathematical Morphology*, Academic Press: London, 1982.
73. J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Sig. Proc.*, Vol. 41, No. 12, pp. 3445–3462, 1993.
74. E.P. Simoncelli, "Statistical models for images: Compression, restoration and synthesis," in *Proc. 31st Asilomar Conf on Signals, Systems and Computers*, Pacific Grove, CA. IEEE Computer Society, 1997, pp. 673–678.
75. E.P. Simoncelli, "Bayesian denoising of visual images in the wavelet domain," in *Bayesian Inference in Wavelet Based Models*, Lecture Notes in Statistics, Vol. 41, Springer Verlag: Berlin, 1999.
76. E.P. Simoncelli and E.H. Adelson, "Noise removal via Bayesian wavelet coring," in *Third Int'l. Conf on Image Proc.*, Lausanne, IEEE Sig Proc Society, 1996, Vol. I, pp. 379–382.
77. E.P. Simoncelli and R.T. Buccigrossi, "Embedded wavelet image compression based on a joint probability model," in *Proceedings of ICIP (1)*, 1997, pp. 640–643.
78. E.P. Simoncelli, W.T. Freeman, E.H. Adelson, and D.J. Heeger, "Shiftable multi-scale transforms," *IEEE Trans Information Theory*, Vol. 38, No. 2, pp. 587–607, 1992.
79. E. Simoncelli and B. Olshausen, "Natural image statistics and neural representation," *Annual Review of Neuroscience*, Vol. 24, pp. 1193–1216, 2001.
80. A. Srivastava, X. Liu, and U. Grenander, "Universal analytical forms for modeling image probability," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 9, 2002.
81. J.B. Tenenbaum, V. Silva, and J.C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, Vol. 290, pp. 2319–2323, 2000.
82. M.G.A. Thomson, "Beats, kurtosis, and visual coding," *Network: Computation in Neural Systems*, Vol. 12, No. 3, pp. 271–287, 2001.
83. A. Turiel, G. Mato, and N. Parga, "The self-similarity properties of natural images resemble those of turbulent flows," *Physical Review Letters*, Vol. 80, No. 5, pp. 1098–1101, 1998.
84. A. Turiel and N. Parga, "The multi-fractal structure of contrast changes in natural images: From sharp edges to textures," *Neural Computation*, Vol. 12, pp. 763–793, 2000.
85. A. Turiel, N. Parga, D.L. Ruderman, and T.W. Cronin, "Multiscaling and information content of natural color images," *Physical Review* E, Vol. 62, No. 1, pp. 1138–1148, 2000.
86. J.A. van Hateren, "Independent component filters of natural images compared with simple cells in primary visual cortex," *Proc. Royal Statistical Society of London* B, Vol. 265, pp. 359–366, 1998.
87. M. Vetterli, "Multidimensional subband coding: Some theory and algorithms," *Signal Processing*, Vol. 6, No. 2, pp. 97–112, 1984.

88. M.J. Wainwright and E.P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," *Advances in Neural Information Processing Systems*, S.A. Solla, T.K. Leen, and K.-R. Muller (Eds.), 2000, pp. 855–861.
89. M.J. Wainwright, E.P. Simoncelli, and A.S. Willsky, "Random cascades on wavelet trees and their use in modeling and analyzing natural imagery," *Applied and Computational Harmonic Analysis*, Vol. 11, No. 1, pp. 89–123, 2001.
90. A.B. Watson, "The cortex transform: Rapid computation of simulated neural images," *Comp. Vis. Graphics Image Proc.*, Vol. 39, pp. 311–327, 1987.
91. B. Wegmann and C. Zetzsche, "Statistical dependence between orientation filter outputs used in a human vision based image code," in *Proceedings of Visual Communcation and Image Processing*, Society of Photo-Optical Intstrumentation Engineers, Vol. 1360, pp. 909–922, 1990.
92. Y. Weiss, "Deriving intrinsic images from image sequences," in *Proceedings Eight IEEE International Conference on Computer Vision*, 2001, Vol. 2, pp. 68–75.
93. G. Winkler, *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Springer: Berlin, 1995.
94. S.N. Yendrikhovskij, "Computing color categoies from statistics of natural images," *Journal of Imaging Science and Technology*, Vol. 45, No. 5, pp. 409–417, 2001.
95. C. Zetzsche, "Polyspectra of natural images," in *Presented at Natural Scene Statistics Meeting*, 1997.
96. C. Zetzsche and F. Rohrbein, "Nonlinear and extra-classical receptive field properties and the statistics of natural images," *Network: Computation in Neural Systems*, Vol. 12, No. 3, pp. 331–350, 2001.
97. S.C. Zhu, C.E. Guo, Y.N. Wu, and Y.Z. Wang, "What are textons?," in *Proc. of Euorpean Conf. on Computer Vision*, 2002.
98. S.C. Zhu and D. Mumford, "Prior learning and Gibbs reaction-diffusion," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, No. 11, pp. 1236–1250, 1997.
99. S.C. Zhu, Y.N. Wu, and D. Mumford, "Minimax entropy principles and its application to texture modeling," *Neural Computation*, Vol. 9, No. 8, pp. 1627–1660, 1997.

Anuj Srivastava is an Assistant Professor in Department of Statistics at Florida State University. He obtained his Ph.D. in electrical engineering from Washington University in 1996 and was a visiting researcher in Division of Applied Mathematics at Brown University during the year 1996–1997. He directs the Laboratory for Computational Vision at FSU that engages in research on image understanding using multiple imaging modalities. His research interests lie in the areas of statistical signal processing and image analysis, statistics on manifolds, and computational statistics. His research has been supported by ARO, NSF, NIMA, and FSU research funds.

Ann Lee received her M.Sc. degree in Engineering Physics at Chalmers University of Technology in Sweden, and her Ph.D. degree in Physics at Brown University. She is currently a visiting research associate in the Division of Applied Mathematics at Brown University and a research instructor in the Mathematics Department at Yale University. Her research is in the areas of machine and biological vision, statistical modeling, and pattern theory.

Eero Simoncelli is an Associate Professor of Neural Science and Mathematics at New York University. He received the B.A. in physics from Harvard University, studied Mathematics for one and a half years at Cambridge University, and then entered the Ph.D. program in electrical engineering and computer science at MIT. He joined the faculty of the Computer and Information Science department at U Pennsylvania in January 1993. In September of 1996, he moved to New York University as part of the Sloan Center for theoretical visual neuroscience. He received an NSF Faculty Early Career Development (CAREER) grant in 1996 for research and teaching in "Visual Information Processing", and a Sloan Research Fellowship in 1998. In August 2000, he became an Associate Investigator of the Howard Hughes Medical Institute, under their new program in Computational Biology.

Song Chun Zhu received his B.S. degree in computer science from the University of Science and Technology of China in 1991. He

received his M.S. and Ph.D. degrees in computer science from Harvard University (Harvard Robotics Lab) in 1994 and 1996 respectively. He was a research associate in the Division of Applied Math. (Pattern Theory Group) at Brown University during 1996–1997, a lecturer in the Computer Science Department at Stanford University during 1997–1998, and an assistant professor in the Department of Computer and Information Sciences at Ohio State University from 1998–2002. He is currently an associate professor in the Departments of Statistics and Computer Science jointly at University of California, Los Angeles. His research is focused on computer vision and learning, statistical modeling, and stochastic computing. He has published over 50 articles, and received various honors including a David Marr prize honorary mention, a NSF Career award, a Sloan Fellow, and a Navy Young Investigator Award.