Published in: Proceedings of IEEE Signal Processing Workshop on Higher-Order Statistics,
Ceasarea, Israel. 14-16 June, 1999.

Higher-order Statistical Models of Visual Images

Eero P. Simoncelli
Center for Neural Science, and
Courant Institute of Mathematical Sciences
New York University
4 Washington Place, Rm 809
New York, NY 10003
eero.simoncelli@nyu.edu

I examine the empirical densities of natural photographic images, and show that although they are highly non-Gaussian, they are quite regular and may be described using fairly simple parameterized density models. I describe two such models, and demonstrate their ability to account for image content.

A probability model for visual images is an important component of many applications in image processing, computer vision, and computer graphics. The power of such a model comes from the fact that the set of visual images is enormous, and yet only a small fraction of these are likely to be encountered in a natural setting. Thus, a statistical prior model, even one that only partially captures these variations in likelihood, can substantially benefit image processing and artificial vision systems. But the problem of inferring a probability density for images is difficult because of their high dimensionality.

A sensible first step is to transform the signal to a basis in which the density is simplified. Choosing a basis that is adapted to statistical properties of an input signal is a classical problem. The traditional solution is principal components analysis (PCA), in which a linear decomposition is chosen to diagonalize the covariance structure of the input. Recently a number of authors have noted that statistics of order greater than two can be utilized in choosing a basis for images. Field [5] noted that the coefficients of frequency subbands of natural scenes have much higher kurtosis than a Gaussian density. Recent work on "independent components analysis" (ICA) has sought linear bases that optimize higher-order statistical measures [e.g., 3, 4]. Several authors have constructed optimal bases for images by optimizing such measurements [7, 1]. The resulting basis functions are spatially oriented and have spatial frequency bandwidths of roughly one octave, similar to the most common multi-scale decompositions.

Our experience is that these observations of highly kurtotic marginals do not depend on the precise choice of basis, beyond the qualitative properties of orientation specificity and octave bandwidth. For our purposes, we utilize a fixed multi-scale orthonormal wavelet basis, in which functions are related by translation, dilation, and frequency modulation. The image is decomposed into a set of *subbands*, each consisting of those coefficients associated with a set of basis functions related by translation (but all having the same orientation and scale). An example is shown in figure 1.

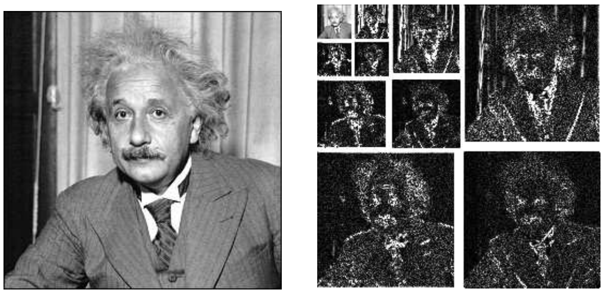
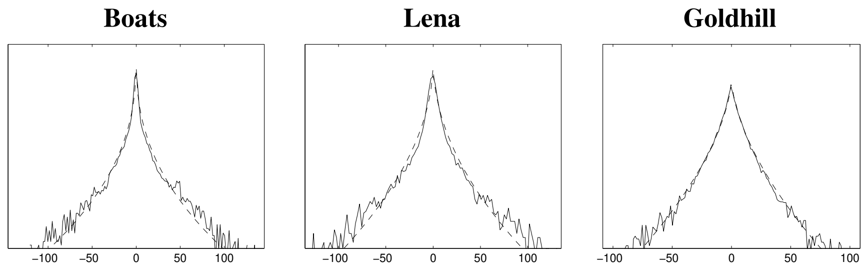


Figure 1. Coefficient magnitudes of a wavelet decomposition. Left: original "Einstein" image. Right: Absolute values of subband coefficients at three scales, and three orientations of a separable wavelet decomposition of this image.

Marginal Model

Empirically, the wavelet coefficients of natural images have significantly non-Gaussian marginal statistics [e.g., 5, 6]. In particular, the histograms are found to have much heavier tails and are more sharply peaked than one would expect from a Gaussian density. As an example, histograms (obtained by gathering the coefficients of a single subband), of separable wavelet decompositions of several images are plotted in figure 2.



$p = 0.62, \frac{\Delta H}{H} = 0.014$ $p = 0.56, \frac{\Delta H}{H} = 0.013$ $p = 0.60, \frac{\Delta H}{H} = 0.0019$

Figure 2. Examples of 256-bin coefficient histograms for a single vertical wavelet subband of three images, plotted in the log domain. All images are size $512x512$. Also shown (dashed lines) are fitted model densities corresponding to equation (1). Below each histogram is the maximum-likelihood value of p used for the fitted model density, and the relative entropy (Kullback-Leibler divergence) of the model and histogram, as a fraction of the total entropy of the histogram.

These marginal densities are well-modeled by a generalized Laplacian (or "stretched exponential") distribution [6, 11]:

$$\mathcal{P}_c(c; s, p) = \frac{e^{-|c/s|^p}}{Z(s, p)}, \qquad (1)$$

where the normalization constant is $Z(s, p) = 2\frac{s}{p}\Gamma(\frac{1}{p})$. Each graph in figure 2 includes a dashed curve corresponding to the best fitting instance of this density function, with the parameters $\{s, p\}$ estimated by maximizing the likelihood of the data under the model. For subbands of images in our collection, values of the exponent p typically lie in the range $[0.5, 0.8]$. The factor s varies monotonically with the size of the basis functions. The density model fits the histograms remarkably well, as indicated by the relative entropy measures given below each plot.

Figure 3 illustrates the entropy gain associated with this model, as compared to a Gaussian density model, and an empirical density model (i.e., assuming a known histogram). The Gaussian model variance was set to match the sample variance. Note that the generalized Laplacian model comes within 0.25 bits/coefficient of the empirical entropy, as compared with the Gaussian density model which often has a relative entropy greater than 1.0 bit/coefficient.

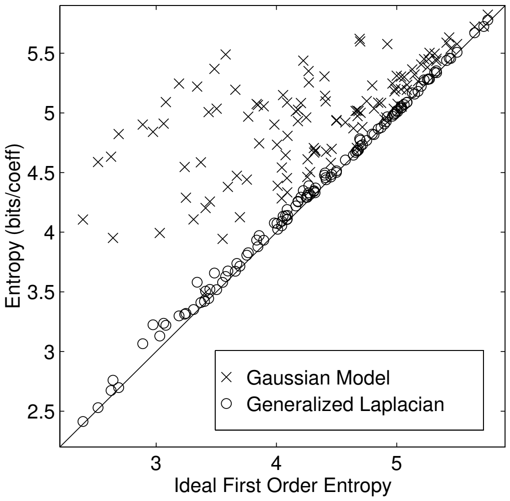


Figure 3. Comparison of encoding costs. Plotted are encoding cost assuming the generalized Laplacian density of equation (1) (O's), and the encoding cost assuming a Gaussian density (X's), versus the empirical ("ideal") encoding cost using a 256-bin histogram. Points are plotted for 9 bands (3 scales, 3 orientations) of 13 images.

Thus, our first model of image statistics is that each wavelet subband consists of i.i.d. random variables drawn from a density of the form given in equation (1). The model is parameterized by the set $\{s_k, p_k\}$, where k indexes the subband.

Joint Model

The marginal model described above assumes that the coefficients of the multi-scale representation are independent. Empirically, the coefficients of orthonormal wavelet decompositions of visual images are found to be fairly well decorrelated (i.e., their covariance is zero). They are *not*, however, independent [2, 9]. Consider two coefficients representing information at adjacent scales, but the same orientation (e.g, horizontal) and spatial location. Figure 4A shows the conditional histogram $\mathcal{H}(c|p)$ of a "child" coefficient conditioned on a coarser-scale "parent" coefficient. As with the marginals, we assume stationarity in order to consider the joint histogram of this pair of coefficients, gathered over the spatial extent of the image, as representative of the underlying density.

The histogram illustrates several important aspects of the relationship between the two coefficients. First, they are second-order decorrelated, since the expected value of c is approximately zero for all values of p. Second, the variance of the conditional histogram of c clearly depends the value of p. Thus, although c and p are uncorrelated, they still ex-

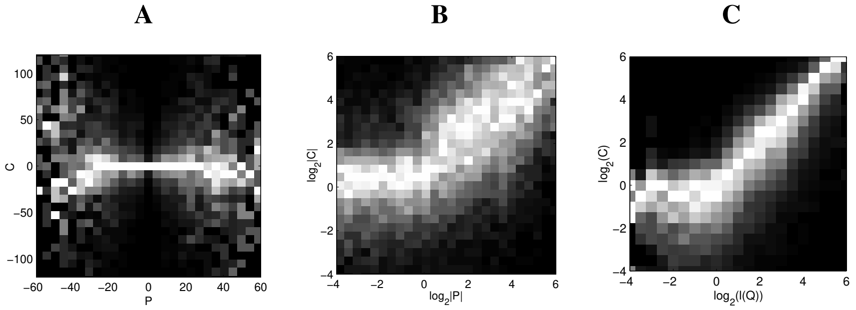


Figure 4. Conditional histograms for a fine scale horizontal coefficient. Brightness corresponds to probability, except that each column has been independently rescaled to fill the full range of display intensities. **A:** Conditioned on the Parent (same location and orientation, coarser scale) coefficient. Data are for the "Boats" image. **B:** Same as **A**, but in the log domain. **C** Conditioned on a linear combination of neighboring coefficient magnitudes. Data are for the same subband of the Boats image as **A**.

hibit statistical dependence. Furthermore, this dependence cannot be eliminated through further linear transformation.

The structure of the relationship between c and p becomes more apparent upon transforming the variables to the log domain. Figure 4B shows the conditional histogram $\mathcal{H}\left(\log_2(c^2)|\log_2(p^2)\right)$ The right side of the distribution is unimodal and concentrated along a unit-slope line, suggesting that the conditional expectation, $\mathbb{E}(c^2|p^2)$, is approximately proportional to p^2. Furthermore, vertical cross sections have approximately the same shape for different values of p^2. Finally, the left side of the distribution is concentrated about a horizontal line, suggesting that c^2 is independent of p^2 in this region. We have modeled the conditional relationship between the two coefficients as:

$$\mathcal{P}\left(c|p\right) \;\propto\; f(c/\sqrt{p^2+\sigma^2}).$$

That is, the distribution of c is described by a density function f, with variance proportional to $p^2+\sigma^2$.

The form of the histograms shown in figure 4 is surprisingly robust across a wide range of images. Furthermore, the qualitative form of these statistical relationships also holds for pairs of coefficients at adjacent spatial locations, adjacent orientations. As one considers coefficients that are more distant (either in spatial position or in scale), the dependency becomes weaker. This suggests that a Markov assumption might be appropriate.

Given the difficulty of characterizing the full density of a coefficient conditioned on a large set of neighbors, we decided to simply extend the variance-scaling relationship described above using a weighted linear predictor for the coefficient variance. This is a somewhat *ad hoc* choice, but it has proven effective in a number of applications of this model. In particular, we assume a Markov model in which the variance of a coefficient, c scales as a linear combination of the squared coefficients in a local neighborhood:

$$\mathcal{P}\left(c \mid \{p_k\}\right) \;\propto\; f(c/\sqrt{\sum_k w_k p_k^2+\sigma^2}). \qquad (2)$$

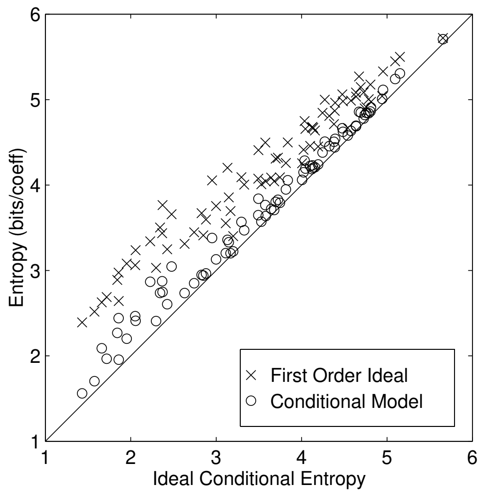


Figure 6. Comparison of encoding cost using the conditional probability model of equation (2) with the encoding cost using the first-order histogram, as a function of the encoding cost using a 256×256-bin joint histogram. Points are plotted for 6 bands (2 scales, 3 orientations) of 13 images.

Unlike traditional Gauss-Markov random fields, in which the conditional mean of a random variable depends linearly on the neighboring variables, here the *variance* of c depends linearly on the *squared* neighbors. If one assumes a particular form for the density f (e.g., Gaussian), the parameters $\{w_k,\sigma\}$ may be determined (numerically) via maximum likelihood estimation.

The model fits the histograms quite well, as can be seen from three examples shown in figure 5 The benefit of the model, in terms of entropy reduction, is summarized in figure 6. The entropy savings are less than those illustrated in figure 3, but still substantial.

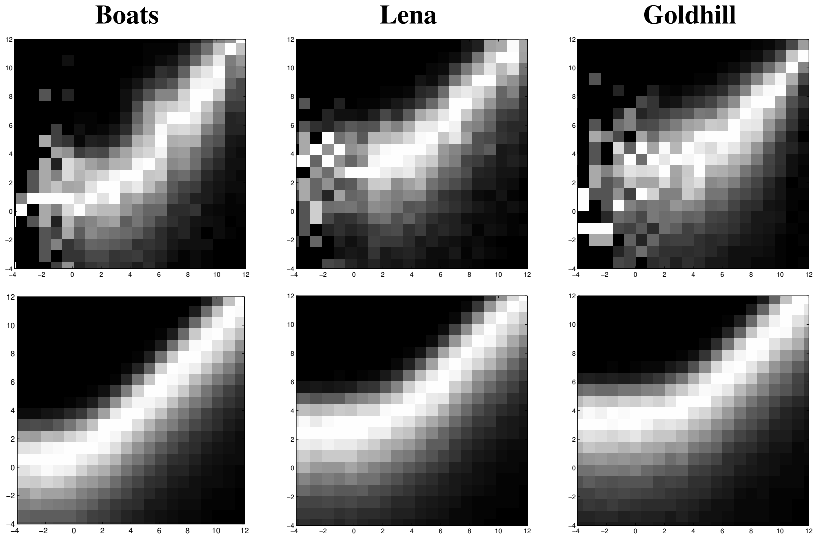


Figure 5. **Top**: Examples of log-domain conditional histograms for the second-level horizontal subband of different images, conditioned on an optimal linear combination of coefficient magnitudes from adjacent spatial positions, orientations, and scales. **Bottom**: Model of equation (2) fitted to the conditional histograms. Intensity corresponds to probability, except that each column has been independently rescaled to fill the full range of intensities.

Discussion

I've described non-Gaussian marginal and joint models for visual images in the wavelet domain. The models are substantially more powerful than traditional Gaussian models. Our previous work has demonstrated that these types of probability model are quite powerful when used in applications of compression [2], denoising [11, 10], and texture synthesis [8].

Although the empirical observations used to motivate the two models are quite striking, there are difficult issues to be resolved, especially in the precise nature of the joint model. A fundamental problem is that the joint model is inconsistent with the marginal model: Marginalizing the joint density does not give a marginal density of the correct form.

Another interesting topic for investigation is the development of sampling algorithms for these density models. Sampling the marginal model is quite simple: one need only fill the coefficients of a wavelet representation with independent samples drawn from generalized Laplacian densities. Sampling from the joint density is more difficult, requiring Monte Carlo or other iterative algorithms.

References

[1] A. J. Bell and T. J. Sejnowski. The 'independent components' of natural scenes are edge filters. *Vision Research*, 37(23):3327–3338, 1997.

[2] R. W. Buccigrossi and E. P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. Technical Report 414, GRASP Laboratory, University of Pennsylvania, May 1997. Accepted (3/99) for publication in IEEE Trans Image Processing.

[3] J. F. Cardoso. Source separation using higer order moments. In *ICASSP*, pages 2109–2112, 1989.

[4] P. Common. Independent component analysis, a new concept? *Signal Process.*, 36:387–314, 1994.

[5] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4(12):2379–2394, 1987.

[6] S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Pat. Anal. Mach. Intell.*, 11:674–693, July 1989.

[7] B. A. Olshausen and D. J. Field. Natural image statistics and efficient coding. *Network: Computation in Neural Systems*, 7:333–339, 1996.

[8] J. Portilla and E. P. Simoncelli. Texture representation and synthesis using correlation of complex wavelet coefficient magnitudes. Technical Report 54, Consejo Superior de Investigaciones Cientificas (CSIC), Madrid, March 1999. Available at ftp://ftp.optica.csic.es/pub/papers/portillaTR54.ps.gz.

[9] E. P. Simoncelli. Statistical models for images: Compression, restoration and synthesis. In *31st Asilomar Conf on Signals, Systems and Computers*, pages 673–678, Pacific Grove, CA, November 1997. IEEE Computer Society.

[10] E. P. Simoncelli. Bayesian denoising of visual images in the wavelet domain. In P. Müller and B. Vidakovic, editors, *Bayesian Inference in Wavelet Based Models*. Springer-Verlag, New York, To appear: Summer 1999.

[11] E. P. Simoncelli and E. H. Adelson. Noise removal via Bayesian wavelet coring. In *Third Int'l Conf on Image Proc*, volume I, pages 379–382, Lausanne, September 1996. IEEE Sig Proc Society.