from *SIAM News*, Volume 35, Number 4

Comparing (Images of) Apples and Oranges

By Anuj Srivastava, Xiuwen Liu, and Ulf Grenander

Automatic recognition of objects (people, license plates, military vehicles, vegetation) from images is an increasing priority, one in which statistical techniques are bound to play an important role. Explicit probability models for images have been attracting the attention of researchers, most likely because of a growing appreciation for the variability of images, coupled with the realization that exact mathematical/physical models may not be feasible. Any statistical approach will rely on probability models that capture "essential" image variability and are computationally tractable.

Because images are very high dimensional, we are faced with the simultaneous tasks of model building and dimension reduction. Tools for dimension reduction include principal components, independent components, wavelet transforms, and Fourier analysis. Recent studies involving empirical distributions of reduced representations have revealed certain interesting patterns. For example, a popular mechanism for decomposing images locally—in space and frequency—via wavelet transforms leads to coefficients that are quite non-Gaussian—that is, the histograms display heavy tails, sharp cusps at the median, and higher correlations at different scales. One such histogram is shown (on a log scale) at the lower left in Figure 1, with the corresponding image directly above it.

Explaining Non-Gaussian Behavior of Images

A recently proposed probability model, built on physical concepts, has been successful in explaining such non-Gaussian behavior. The main idea is as follows: Scenes consisting of objects and images are made up of 2D appearances of the objects. Objects are placed in an image at points $\{z_i \in IR^2 | \ i = 1, \ldots, n\}$ generated according to a random process. The i^{th} object is centered at z_i, and has appearance g_i chosen randomly and weighted by $a_i \sim N(0,1)$. A weighted superposition is used to form an image: $I(z) = \sum_{i=1}^{n} a_i g_i(z - z_i)$.

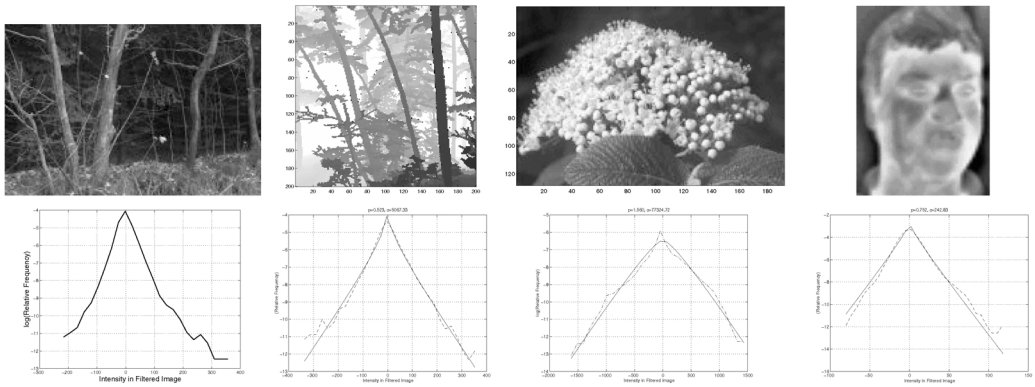

Figure 1. *Marginal densities (bottom row) for images (top row) filtered by arbitrary Gabor filters. Solid lines denote Bessel K densities and broken lines, observed histograms.*

Given an image I and a bank of linear filters $\{F^{(j)}, j = 1, 2, \ldots, J\}$, we compute for each filter $F^{(j)}$ a filtered image $I^{(j)} = I * F^{(j)}$, where * denotes the 2D convolution operation. Possible filters include directional derivatives, Gabor filters, and the Laplacian of the Gaussian. Each filter selects and isolates certain features present in the original image. Under certain assumptions, the density function of the random variable $I^{(j)}(z)$ has been shown [1] to be: for $p > 0$, $c > 0$, $f(x;p,c) = 1/Z(p,c)|x|^{p-0.5} K_{(p-0.5)}((2/c)^{1/2}|x|)$, where K is the modified Bessel function and Z is the normalizing constant. We call these densities *Bessel K forms* and refer to the parameters (p, c) as *Bessel parameters*.

As described in [1], p and c are easily estimated from the observed data,with $\hat{p} = 3/\mathrm{SK}(I^{(j)})-3$ and $\hat{c} = \mathrm{SV}(I^{(j)})/\hat{p}$, where SK is the sample kurtosis and SV is the sample variance of the pixel values in $I^{(j)}$.

Figure 1 shows several examples, with images in different modalities (range, video, infrared) in the top row and the histograms of their filtered versions (with arbitrary filters, not shown) in the bottom row. The overlapping estimated (solid lines) and observed (dotted lines) densities demonstrate the good performance of this model.

If a filter F is applied to an image I to extract some specific feature—vertical edges, say—the resulting p has been shown [2] to depend on two factors: (i) distinctness and (ii) frequency of occurrence of that feature in that image. Objects with sharper, distinct edges have low p values, while scenes with many objects have large p values.

Metrics for Comparing Images

To compare marginal densities that take on Bessel K forms, we can use the L^2-metric between them. A closed-form expression for this metric is given in [2]. For comparing two images, I_1 and I_2, with a filter bank $F^{(1)}, \ldots, F^{(J)}$, we let the parameter values be given by: $(p_1^{(j)}, c_1^{(j)})$ and $(p_2^{(j)}, c_2^{(j)})$, respectively, for $j = 1,2, \ldots, J$. A pseudometric between two images is then defined as:

$$\sqrt{\left(\sum_{j=1}^{J} d\left(p_1^{(j)}, c_1^{(j)}, p_2^{(j)} c_2^{(j)}, \right)^2 \right)},$$

where $d(p_1, c_1, p_2, c_2)$ is the L^2 metric between the corresponding Bessel K forms. For a simple illustration, consider the images of natural scenes (trees, lakes, bushes, leaves) shown in Figure 2. The images are numbered one to ten, from top left to bottom right. Using 27 small-scale Gabor filters, we computed the pairwise distances d_I between them. Shown in the bottom panel is a dendrogram clustering of images based on d_I. We can see that perceptually similar images have been clustered together! Extensive experiments support the claim that d_I provides a powerful tool for image classification, analysis, and understanding.

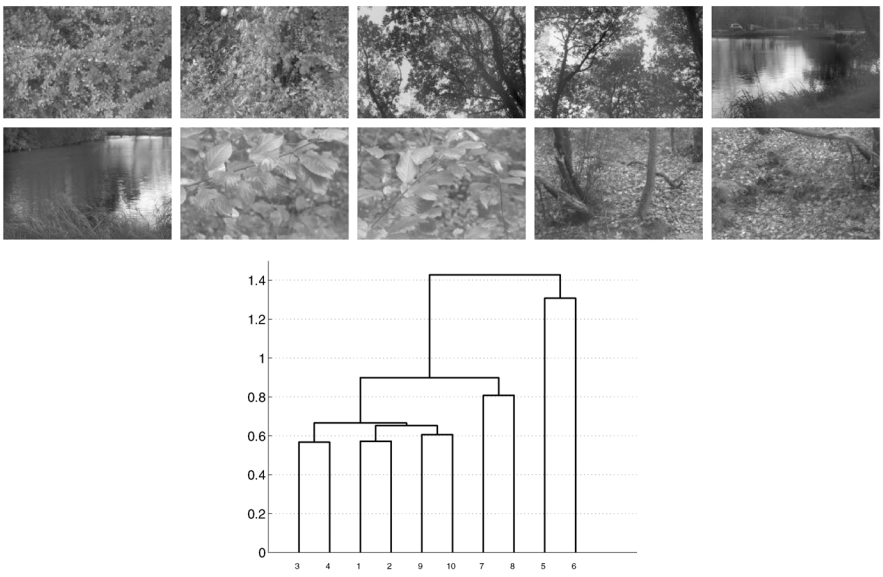

Figure 2. *Dendrogram clustering plot (bottom) for the images in the top row.*

References

[1] U. Grenander and A. Srivastava, *Probability models for clutter in natural images*, IEEE Trans. Pattern Anal. Mach. Intel., 23 (4), April 2001, 424–429.

[2] A. Srivastava, X. Liu, and U. Grenander, *Universal analytical forms for modeling image probability*, IEEE Trans. Pattern Anal. Mach. Intel., to appear, 2002.

Some of the images used in this article are from the van-Hateren natural image database.

Anuj Srivastava and Xiuwen Liu are professors of statistics and computer science, respectively, at Florida State University. Ulf Grenander is a professor in the Division of Applied Mathematics at Brown University.