

# High-pass filters give histograms with positive kurtosis

Eric Clarkson and Harrison H. Barrett

Department of Radiology, University of Arizona, Tucson, Arizona 85724, and  
Optical Sciences Center, University of Arizona, Tucson, Arizona 85721

Received February 8, 2001

If we filter an ensemble of natural images with a high-pass filter, the output is a random variable. The probability density for this variable will, under fairly general assumptions, be a Gaussian scale mixture. The kurtosis for this type of distribution is always positive. © 2001 Optical Society of America  
OCIS codes: 100.2000, 100.2960, 030.6600.

There are many situations in which we are interested in the statistical properties of the outputs of high-pass filters applied to an ensemble of images. Examples of such filters include discrete cosine transforms, wavelet transforms, edge detectors, and the filters that result from analysis of independent components.<sup>1</sup> Examples of image ensembles include natural scenes, satellite images, and medical images. For a set of high-pass filters, the conclusion that the resulting components are statistically independent is often based on the large positive kurtosis of the histogram of each filter output. Indeed, positive kurtosis is often taken as the hallmark of independent components.<sup>2</sup> Large positive kurtosis for the filter-output histograms is also taken as an indication that the representation of an image in terms of these filter outputs is sparse.<sup>3,4</sup> Recently, histograms of the outputs of wavelet filters<sup>5,6</sup> and of discrete cosine filters<sup>7</sup> have been modeled as Gaussian mixtures. We show that, under fairly general assumptions, the output of any high-pass filter applied to an ensemble of natural images can always be at least approximately described by a Gaussian mixture model. We also demonstrate that, if the density of a random variable is a Gaussian mixture, then the kurtosis for this variable must be positive.

The gray-level values of a digital image drawn at random from the image ensemble are given by the function  $f(i, j)$  of integer coordinates  $i$  and  $j$ . The filter coefficients are given by the function  $h(m, n)$  of integers  $m$  and  $n$  from  $-M$  to  $M$ . We assume that the filter is fixed with its center  $(m, n) = (0, 0)$  over location  $(k, l)$  in each image and that the filter support is much smaller than the overall image. The random variable  $z$  that is the output of the filter at this location is given by

$$z = z(k, l) = \sum_{m, n=-M}^M h(m, n)f(k - m, l - n). \quad (1)$$

This is a random variable because  $f(i, j)$  is a random image from the ensemble. We examine some of the statistical properties of this random variable and how they relate to general properties of the image ensemble.

We begin by separating the positive and negative components of the high-pass filter  $h(m, n)$  as follows:

$$h_+(m, n) = \begin{cases} h(m, n) & h(m, n) > 0 \\ 0 & \text{otherwise} \end{cases}, \quad (2)$$

$$h_-(m, n) = \begin{cases} -h(m, n) & h(m, n) < 0 \\ 0 & \text{otherwise} \end{cases}. \quad (3)$$

These filter components in turn give rise to random variables,  $u$  and  $v$ :

$$u = u(k, l) = \sum_{m, n=-M}^M h_+(m, n)f(j - m, k - n), \quad (4)$$

$$v = v(k, l) = \sum_{m, n=-M}^M h_-(m, n)f(j - m, k - n). \quad (5)$$

We then have  $z = u - v$ . We assume that the filter will give zero response in flat regions of an image, which implies that

$$\sum_{m, n=-M}^M h_+(m, n) = \sum_{m, n=-M}^M h_-(m, n). \quad (6)$$

Indeed, this condition is usually satisfied for any high-pass filter.

For image ensembles such as the three mentioned above, i.e., natural scenes, satellite images, and medical images, the ensemble can be divided into subensembles that are characterized by the statistics of the subimage that is covered by the filter. In particular, we can choose the subensembles to have two properties. The first is that the mean value of the pixel gray levels be approximately constant over the filter support when we restrict our attention to images in a particular subensemble. The second property is that the range of the correlation between pixels be short compared with the size of the filter support. Again, this correlation is computed only from the images in a particular subensemble. For example, in medical imaging, subensembles may correspond to images for which the filter is located over bone, soft tissue, lung, and so on. There can also be subensembles that correspond to images for which the filter support contains a bone-soft tissue boundary, a bone-lung boundary, a lung-soft tissue boundary, and so forth. Note that these boundary subensembles can still have the two properties listed above, as long as the boundary is randomly located within the subimage.

Note that the second property does not preclude long-range correlations in the original image ensemble, which certainly exist for the image ensembles

under consideration here.<sup>8-10</sup> To see this, consider two pixel outputs,  $x$  and  $y$ ; let  $\langle * \rangle_{S_n}$  denote the average over the subensemble  $S_n$  and  $\langle * \rangle_n$  denote the average over all subensembles. The covariance of  $x$  and  $y$  is

$$\langle \langle xy \rangle_{S_n} \rangle_n - \langle \langle x \rangle_{S_n} \rangle_n \langle \langle y \rangle_{S_n} \rangle_n = \langle [ \langle xy \rangle_{S_n} - \langle x \rangle_{S_n} \langle y \rangle_{S_n} ] \rangle_n + \langle [ \langle x \rangle_{S_n} \langle y \rangle_{S_n} - \langle x \rangle_{S_n} \langle y \rangle_{S_n} ] \rangle_n. \quad (7)$$

From the second property it follows that the quantity in the first set of brackets in Eq. (7) is negligible when the separation between the two pixels is greater than some correlation length  $L_n$  that is small compared with the filter support. However, we have placed no restrictions on the term in the second set of brackets, which therefore corresponds to long-range correlations in the images.

Given these two properties of the subensembles, we may draw two conclusions about the statistics of the random variables  $u$  and  $v$  when we condition on a particular subensemble. The first is that the conditional mean of  $z = u - v$  be zero. This follows from the fact that the conditional mean value of the pixel gray levels is constant over the filter support. The second conclusion, which follows from the correlation condition and the central-limit theorem, is that the conditional density for the pair  $(u, v)$  be bivariate normal. Therefore, for each subensemble there is a parameter vector  $(a, b, c, s)$  such that the conditional density for  $(u, v)$  is given by

$$p_{UV}(u, v | a, b, c, s) = \frac{1}{2\pi\sqrt{ac - b^2}} \times \exp \left[ -\frac{1}{2} \begin{pmatrix} u - s \\ v - s \end{pmatrix}^T \begin{pmatrix} ab \\ bc \end{pmatrix}^{-1} \begin{pmatrix} u - s \\ v - s \end{pmatrix} \right]. \quad (8)$$

The probability-density function  $p_{ABCS}(a, b, c, s)$  on a four-dimensional region  $\Omega$  describes the statistics of the subensemble parameters  $(a, b, c, s)$  as they occur in the overall ensemble of images. The joint density for  $(u, v)$  is then given by the standard expression

$$p_{UV}(u, v) = \int_0^\infty \int_\Omega p_{UV}(u, v | a, b, c, s) \times p_{ABCS}(a, b, c, s) da db dc ds. \quad (9)$$

Given this density, the probability density for  $z$  is given by

$$p_Z(z) = \int_{-\infty}^\infty p_{UV}(u, u - z) du. \quad (10)$$

We can envisage Eq. (10) as integrating, for each  $z$ , the joint density for  $u$  and  $v$  along the lines  $v = u - z$  in the  $(u, v)$  plane.

After the expression for  $p_{UV}$  is inserted into Eq. (10) and the  $u$  integration has been performed, we have

$$p_Z(z) = \int_0^\infty \int_\Omega p_Z(z | a, b, c, s) \times p_{ABCS}(a, b, c, s) da db dc ds, \quad (11)$$

with the conditional probability density  $p_Z(z | a, b, c, s)$ , given by

$$p_Z(z | a, b, c, s) = \frac{1}{[2\pi(a + c - 2b)]^{1/2}} \times \exp \left[ -\frac{z^2}{2(a + c - 2b)} \right]. \quad (12)$$

If we introduce the random variable  $w$  by

$$w^2 = 1/2(a + c - 2b), \quad (13)$$

then, by standard probability arguments, the density for  $z$  reduces to

$$p_Z(z) = \frac{1}{2\sqrt{\pi}} \int_0^\infty \exp \left( -\frac{z^2}{4w^2} \right) p_W(w) \frac{dw}{w}, \quad (14)$$

where  $p_W(w)$  is the marginal density on  $w$  derived from  $p_{ABCS}(a, b, c, s)$ . We assume that  $p_W(w)$  is not a delta function. In that case, the probability density for  $z$  is called a Gaussian scale mixture density.<sup>11</sup>

This result is represented schematically in Fig. 1. The circles correspond to Gaussian densities for the variables  $(u, v)$  when the images are restricted to various subensembles of images. The discussion above allows for ellipses with arbitrary orientation also, but the end result in that case is similar. Note that the center of each circle is on the line  $u = v$ , because the mean values of  $u$  and  $v$  are equal when we restrict the images to a subensemble. We may think of the radius of each circle as an indication of the width of the two-dimensional Gaussian. The density for  $z$  is simply the

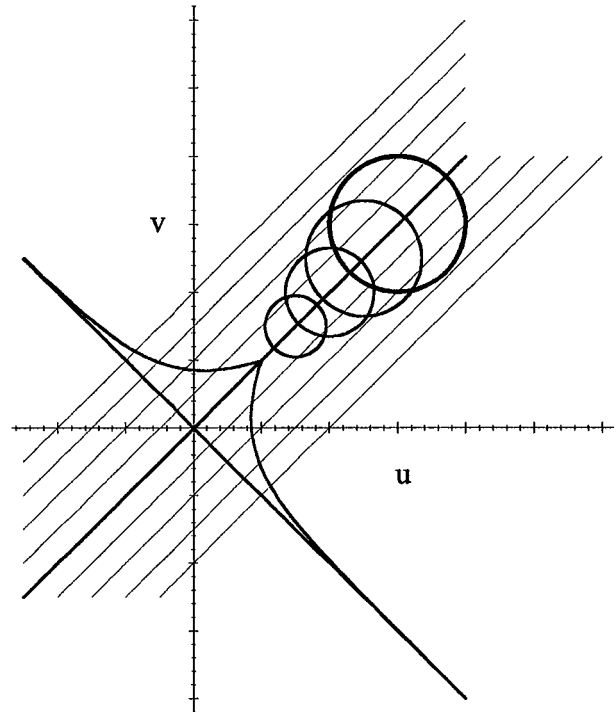


Fig. 1. The probability density for  $z$  is the projection of the joint density for  $(u, v)$  along lines oriented at 45° to the  $u$  axis.

projection of the sum of the Gaussians along the 45° lines shown in the figure. The result is plotted along the angled axis and, intuitively, it would seem that the result must have positive kurtosis.

To verify this intuition we note that the integral in Eq. (14) is the Mellin convolution of  $p_W(w)$  with the Gaussian  $p_G(w) = \exp(-1/4 w^2)$ . This means that the Mellin transform of  $p_Z$  is the product of the Mellin transforms of  $p_W$  and  $p_G$ . In the probability context, the integral that we compute to find the Mellin transform of a density is the  $m$ th moment of that density. Therefore we have a relation between the moments of  $p_Z$  and  $p_W$ :

$$\langle z^m \rangle = \begin{cases} \left[ \frac{2^m \Gamma(m+1/2)}{\sqrt{\pi}} \right] \langle w^m \rangle & m \text{ even} \\ 0 & m \text{ odd} \end{cases}. \quad (15)$$

The expression in brackets in Eq. (15) is the  $m$ th moment of  $p_G$  for even  $m$ . From Eq. (15) we can compute the kurtosis ( $\text{kurt}_Z$ ) of  $p_Z$  in terms of moments of  $p_W$ , and we find that

$$\text{kurt}_Z = \frac{\langle z^4 \rangle - 3\langle z^2 \rangle^2}{\langle z^2 \rangle^2} = 12 \left[ \frac{\langle w^4 \rangle - \langle w^2 \rangle^2}{\langle w^2 \rangle^2} \right] > 0. \quad (16)$$

Inequality (16) follows from the Schwartz inequality. Note that  $\text{kurt}_Z = 0$  if and only if  $w^4 p_W(w) = \lambda p_W(w)$  for some real number  $\lambda$ . This can happen only if  $p_W(w) = \delta(w - w_0)$ , contrary to our assumption about this function. Therefore we can conclude that the kurtosis for  $z$  will be positive. A kurtosis calculation for Gaussian scale mixture densities similar to the one presented here appears in the recent paper of Lam and Goodman on the discrete cosine transform,<sup>7</sup> which is, with the exception of the zero-frequency component, a collection of high-pass filters.

In conclusion, we can state that positive kurtosis is to be expected whenever we compute histograms for natural images filtered through high-pass filters. Models for such kurtotic distributions range from simple Laplacian densities to more-complicated functions, such as  $K$  distributions, which make use of  $K$  Bessel functions. In fact, these are both examples of Gaussian scale mixture distributions, as are some other well-known densities.<sup>6</sup> This is the reason that  $K$  distributions are useful for modeling the random variations in radar<sup>12-14</sup> and in ultrasound signals<sup>15-17</sup> that are due to background clutter. In each of these two imaging modalities the signal strength is roughly proportional to the derivative of the reflectivity in the object, and the derivative operator is the classic example of a high-pass filter. The results above indicate that any family of densities that can reproduce a wide range of densities with positive kurtosis is probably adequate for use in modeling the outputs of high-pass filters. The other conclusion that we may draw is that positive kurtosis alone cannot be used as an indication that filter outputs are statistically independent or form a sparse representation of the image ensemble. This last statement, however, does

not preclude the use of kurtosis as a figure of merit for analysis of independent components, because the goal in this case is to vary a set of high-pass filters until the maximum average kurtosis is achieved for the filter outputs.<sup>2</sup> However, the fact that all the high-pass filters in a given set exhibit positive kurtosis on their outputs does not necessarily imply that any sort of statistical independence has been achieved, because any set of high-pass filters would give the same result.

One can also apply a similar argument to that presented here for the scalar output of one high-pass filter to derive a Gaussian mixture model for the vector output of a collection of such filters. All that is required is that the supports of all of the filters be small compare with the image size but large with respect to typical texture correlations. The resultant model would be a mixture of multivariate Gaussians with different covariance matrices. To reduce this model to that of a Gaussian scale mixture,<sup>6</sup> where these covariance matrices differ by only a scale factor, we would need further assumptions about the filter outputs, which we continue to investigate.

This study was made possible by grants R01 CA52643 and 1 K01 CA87017-01 from the National Cancer Institute. Its content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute. E. Clarkson's e-mail address is clarkson@radiology.arizona.edu.

## References

1. A. Bell and T. J. Sejnowski, *Vision Res.* **37**, 3327 (1997).
2. P. Comon, *Signal Process.* **36**, 287 (1994).
3. B. A. Olshausen and D. Field, *Nature* **381**, 607 (1996).
4. B. A. Olshausen and D. Field, *Network* **7**, 333 (1996).
5. J. J. Heine, S. R. Deans, and L. P. Clarke, *J. Opt. Soc. Am. A* **16**, 6 (1999).
6. M. J. Wainwright and E. P. Simoncelli, in *Advances in Neural Information Processing Systems*, S. A. Solla, T. K. Leen, and K. R. Muller, eds. (MIT, Cambridge, Mass., 2000), Vol. 12, pp. 855-861.
7. E. Y. Lam and J. W. Goodman, *IEEE Trans. Image Process.* **9**, 1661 (2000).
8. M. A. Bernstein, D. M. Thomasson, and W. H. Perman, *Med. Phys.* **16**, 813 (1989).
9. D. L. Ruderman and W. Bialek, *Phys. Rev. Lett.* **73**, 814 (1994).
10. J. J. Heine, S. R. Deans, R. P. Vethuizen, and L. P. Clarke, *Med. Phys.* **26**, 2254 (1999).
11. D. F. Andrews and C. L. Mallows, *J. R. Statist. Soc. Ser. B* **36**, 99 (1974).
12. E. Jakeman and P. N. Pusey, *IEEE Trans. Antennas Propag.* **24**, 806 (1976).
13. E. Jakeman and R. J. A. Tough, *J. Opt. Soc. Am. A* **4**, 1764 (1987).
14. F. A. Pentini, A. Farina, and F. Zirilli, *IEE Proc. F* **139**, 239 (1992).
15. V. M. Narayanan and P. M. Shankar, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **41**, 845 (1994).
16. P. M. Shankar, *Phys. Med. Biol.* **40**, 1633 (1995).
17. P. M. Shankar, R. Molthen, V. M. Narayanan, J. M. Reid, V. Genis, F. Forsberg, C. W. Piccoli, A. E. Lindenmayer, and B. B. Goldberg, *Ultrasound Med. Biol.* **22**, 873 (1996).