Presented at: 3rd IEEE Int'l Conf Image Processing. Lausanne, Switzerland. September 1996.

NOISE REMOVAL VIA BAYESIAN WAVELET CORING

Eero P. Simoncelli

Computer and Information Science Dept.
University of Pennsylvania
Philadelphia, PA 19104

Edward H. Adelson

Brain and Cognitive Science Dept.
Massachusetts Institute of Technology
Cambridge, MA 02139

The classical solution to the noise removal problem is the Wiener filter, which utilizes the second-order statistics of the Fourier decomposition. Subband decompositions of natural images have significantly non-Gaussian higher-order point statistics; these statistics capture image properties that elude Fourier-based techniques. We develop a Bayesian estimator that is a natural extension of the Wiener solution, and that exploits these higher-order statistics. The resulting nonlinear estimator performs a "coring" operation. We provide a simple model for the subband statistics, and use it to develop a semi-blind noise-removal algorithm based on a steerable wavelet pyramid.

A common technique for noise reduction is known as "coring". An image signal is split into two or more bands; the highpass bands are subjected to a threshold non-linearity that suppresses low-amplitude values while retaining high-amplitude values. Use of such techniques is widespread: for example, most consumer VCR's use a simple one-dimensional coring technique.

Many variants of coring have been developed, including two-dimensional coring [1], multi-scale oriented coring [2, 3], pyramid coring [4], and multi-band coring with orthogonal bases [5]. The nonlinear operator is often smoothed to give a "soft" threshold, but the exact choice of function in these techniques has been ad hoc. Similar techniques, based on the statistical concept of "shrinkage", have been recently used with wavelet expansions [6].

The intuition underlying coring is that images typically have spatial structure, consisting of smooth areas interspersed with occasional edges. This notion is evident statistically: the pixels in highpass and bandpass subbands of images have significantly non-Gaussian probability density functions (pdf's) that are sharply peaked at zero with broad tails. Specifically, the coefficient of kurtosis κ (fourth moment divided by squared variance) is typically well above the value of 3 that one expects for a Gaussian pdf.

Field [7] has shown that kurtosis for subbands of natural scenes varies with filter bandwidth, and is maximal at roughly one octave. Significantly wider or narrower bandwidths produce kurtoses near 3 (i.e., Gaussian statistics). Several authors have used Laplacian pdf models (with kurtosis 6) for image subband statistics (e.g., [8, 9]).

Research partially supported by an NSF CAREER grant to EPS.

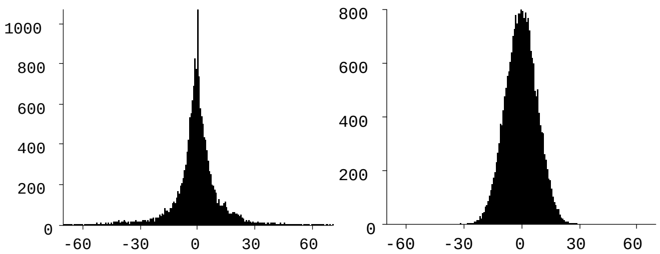

Figure 1: Histograms of a mid-frequency subband in an octave-bandwidth wavelet decomposition for two different images. Left: The"Einstein" image. Right: A white noise image with uniform pdf.

Figure 1 contains an example histogram from a single subband of a wavelet transform built on the "Einstein" image, for which the sample kurtosis is 9.8. For comparison, the histogram of the same subband built on uniform white noise is shown. This histogram is nearly Gaussian, with a sample kurtosis of 2.9. Coring relies on the striking dissimilarity between the point statistics of these two image types.

In the following, we describe a technique for determining the optimal coring function in the Bayesian sense[1], and apply it to a steerable wavelet pyramid.

1. BAYESIAN SIGNAL ESTIMATION

Consider a scalar x contaminated with additive noise n: $y = x + n$. The mean of the posterior distribution provides an unbiased least-squares estimate of the variable x, given measurement y. Bayes' rule allows us to write this in terms of the probability densities of the noise and signal:

$$\hat{x}(y) \quad = \quad \int dx\, \mathcal{P}_{x|y}(x|y)\, x$$

[1]An earlier version of this technique is described in [10], a bachelor's thesis supervised by the authors.

$$= \frac{\int dx\, \mathcal{P}_{y|x}(y|x)\, \mathcal{P}_x(x)\, x}{\int dx\, \mathcal{P}_{y|x}(y|x)\, \mathcal{P}_x(x)}$$

$$= \frac{\int dx\, \mathcal{P}_n(y-x)\, \mathcal{P}_x(x)\, x}{\int dx\, \mathcal{P}_n(y-x)\, \mathcal{P}_x(x)}, \qquad (1)$$

where $\mathcal{P}_n$ indicates the probability density function of the noise, and $\mathcal{P}_x$ the prior probability density function of the signal. The denominator is the pdf of the noisy observation, computed via convolution of the noise and signal pdf's. In order to use this equation to estimate the original signal value x, we must know both of these probability density functions.

Consider a few simple examples. First, let the noise have a zero-mean Gaussian distribution with variance σ_n^2, and let the signal prior be a zero-mean Gaussian with variance σ_s^2. In this case, a well-known closed-form solution exists:

$$\hat{x}(y) = \frac{\sigma_s^2\; y}{\sigma_s^2 + \sigma_n^2}.$$

The solution is a simple linear rescaling of the measurement. When applied to the coefficients of a Fourier transform, this estimator corresponds to the Wiener filter. When applied to subbands of a wavelet transform, the solution is an approximation to the Wiener filter, in which the power spectral density information is averaged over each of the subbands.

Now consider the case in which the noise distribution is Gaussian, but the signal prior is a more sharply peaked distribution, such as that shown in figure 1. In such cases, a closed-form expression for the estimator in equation (1) may not be available, but a numerical solution may be used.[2]. We have computed a numerical approximation of the estimator for the histograms illustrated figure 1. The estimator is illustrated graphically in figure 2. Note that this estimator now corresponds to a nonlinear "coring" operation: large amplitude values are preserved, and small amplitude values are suppressed. This is intuitively sensible: given the substantial signal probability mass at $x = 0$, small values of y are assumed to have arisen from a value of $x = 0$. This curve is similar to the soft-thresholding functions that have been previously devised by more ad hoc methods; the Bayesian derivation thus provides a principled justification for coring systems.

2. PARAMETERIZED MODEL FOR WAVELET COEFFICIENT STATISTICS

The Bayesian estimator discussed above relies on a knowledge of the signal point statistics. In order to utilize it, we need a parameterized model for these pdf's such that: 1) the model provides a good fit to the statistics of natural images, and 2) one can estimate the model parameters from the noisy observation.

[2] One must, in practice, take care to regularize singularities resulting from distribution points with very small probability.

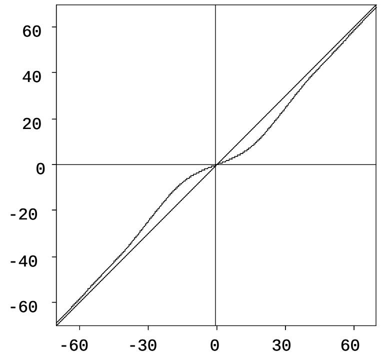

Figure 2: Bayesian estimator (symmetrized) for the signal and noise histograms shown in figure 1. Superimposed on the plot is a straight line indicating the identity function.

For our purposes here, we use a two-parameter generalized Laplacian distribution, also used by Mallat [11]:

$$\mathcal{P}_x(x) \propto e^{-|x/s|^p}. \qquad (2)$$

The distribution is zero-mean and symmetric, and the parameters $\{s, p\}$ are directly related to the second and fourth moments. Specifically (after consultation with an integral table) one obtains:

$$\sigma^2 = \frac{s^2\Gamma(\frac{3}{p})}{\Gamma(\frac{1}{p})}, \qquad \kappa = \frac{\Gamma(\frac{1}{p})\Gamma(\frac{5}{p})}{\Gamma^2(\frac{3}{p})}, \qquad (3)$$

where $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\, dt$, the well known "gamma" function. Given the sample variance and kurtosis of a histogram, we can solve for the two parameters of our model pdf. Typical values for p are in the range $[0.5, 1.0]$. This method of model density estimation is simple and direct, but clearly suboptimal. In the current context, the quality of the estimator should be tested by comparing the noise removal results using the sample (histogram) statistics, and those using the model pdf: such a comparison is given in section 4.

We are also interested in a more realistic "blind" algorithm, in which the parameters are estimated from noisy observations. We note that the second and fourth moments of a generalized Laplacian signal corrupted by additive Gaussian white noise are:

$$\sigma^2 = \sigma_n^2 + \frac{s^2\Gamma(\frac{3}{p})}{\Gamma(\frac{1}{p})}, \quad m_4 = 3\sigma_n^4 + \frac{6\sigma_n^2 s^2\Gamma(\frac{3}{p})}{\Gamma(\frac{1}{p})} + \frac{s^4\Gamma(\frac{5}{p})}{\Gamma(\frac{1}{p})}.$$

Assuming σ_n is known, the measurements of these two moments of the noisy data is sufficient to estimate the model pdf parameters. Results of such an algorithm are given in section 4.

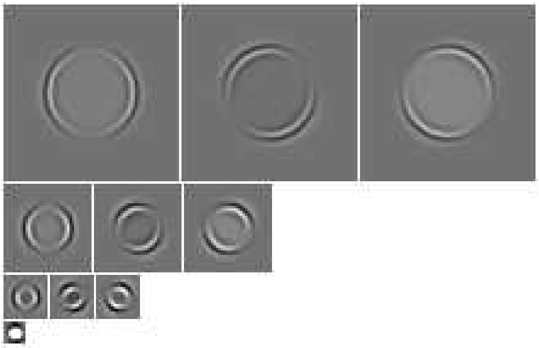

Figure 3: A 3-scale 3-orientation steerable pyramid. Shown are the three oriented bandpass images at each scale, and the residual lowpass image.

3. CHOICE OF WAVELET TRANSFORM

We have implemented a noise reduction scheme based on an oriented multi-scale representation known as a *steerable pyramid* [12]. In this decomposition, the image is subdivided into subbands localized in both scale and orientation. In scale, the subbands have octave bandwidth with a functional form constrained by a recursive system diagram. In orientation, a steerable pyramid can be designed with any number of orientation bands, k. The resulting transform is overcomplete by a factor of $4k/3$. Orientation tuning is constrained by the property of steerability [13].

The transform is "self-inverting" (i.e., the matrix corresponding to the inverse transformation is equal to the transpose of the forward ransformation matrix)[3], and has the additional advantages of being translation-invariant (aliasing-free) and rotation-invariant (steerable). Figure 3 shows an example steerable pyramid decomposition, with three orientation bands.

One disadvantage of the the steerable pyramid for this task is the lack of orthogonality. An orthonormal basis guarantees that the noise component of the transform coefficients will be uncorrelated, assuming that the noise was white in the image domain. For the purposes of this paper, we ignore the off-diagonal terms of the covariance matrix.

4. EXAMPLES

We implemented the "semi-blind" Bayesian de-noising algorithm described previously, in which we assumed a known noise variance. We constructed a 4-scale 4-orientation steerable pyramid from the contaminated image. For each wavelet subband, we estimated the model pdf parameters s and p, and numerically computed an estimator via equation (1). After applying the estimator, we inverted the transform to give the "cleaned" image. To gauge the quality of our model pdf and fitting procedure, we also computed the idealized estimator, based on the *actual* clean signal histograms. Finally, for comparison, we computed a (semi-blind) pyramid-based Wiener filter solution in which we assumed a known noise variance.

[3] In the wavelet literature, such transforms are known as *tight frames*.

Noisy	Bayes ideal	Bayes blind	Wiener
4.78	11.75	11.67	9.79
9.00	13.89	13.82	11.88
13.98	16.49	16.40	14.70

Table 1: SNR values (in dB) for the noise-contaminated image, and images cleaned using each of the three noise removal algorithms. See text.

We applied these three algorithms to the "Einstein" image for three different levels of Gaussian white noise contamination. Table 1 gives signal-to-noise ratios for each case. Note that the semi-blind Bayesian algorithm performs nearly as well in all cases as the ideal Bayesian, indicating that the model pdf is successfully approximating the wavelet coefficient statistics. Note also that the Bayesian algorithm outperforms the Wiener algorithm (as it should, since it is taking advantage of additional information).

Figure 4 shows four images corresponding to the middle row in table 1. The Bayesian image appears to be both sharper (because high-amplitude coefficients are preserved) and less noisy (because low-amplitude coefficients are suppressed). The results can be made more visually appealing by a subsequent sharpening operation, although this reduces the SNR.

5. DISCUSSION

Removal of noise from images relies on differences in the statistical properties of noise and signal. The classic Wiener solution utilizes differences in power spectral density, a second-order property. The Bayesian estimator described above provides a natural extension for incorporating the higher-order statistical regularity present in the point statistics of subband representations. The estimator is based on two factors – a subband representation and a statistical model – both of which can be generalized. Theoretically, one would like a direct link from the properties of the subband pdf to the quality of noise removal, which could then be used to optimize the choice of subband transform. In addition, the statistical model should account for *joint* statistics of wavelet coefficients, both within and between bands. The approach also generalizes to other types of distortion, including blurring and corruption with non-additive noise; one only need have the conditional pdf describing the distortion process. Finally, these types of statistical image model should prove use-

(a) (b)

(c) (d)

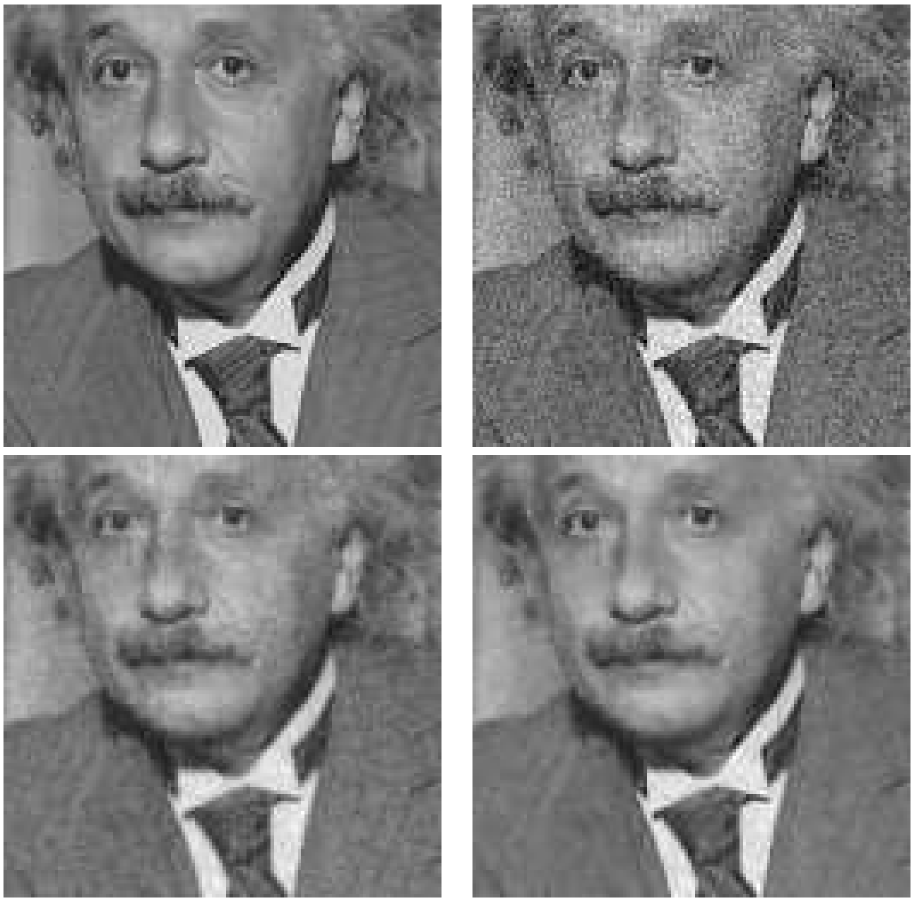

Figure 4: Noise reduction example. (a) Original image (cropped). (b) Image contaminated with additive Gaussian white noise (SNR = 9.00dB). (c) Image restored using (semi-blind) Wiener filter (SNR = 11.88dB). (d) Image restored using (semi-blind) Bayesian estimator (SNR = 13.82dB).

ful in other applications, such as image compression or texture synthesis.

6. REFERENCES

[1] J P Rossi. *JSMPTE*, 87:134–140, 1978.

[2] 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*, 2:31–88, 1986.

[3] E P Simoncelli, W T Freeman, E H Adelson, and D J Heeger. Shiftable multi-scale transforms. *IEEE Trans Information Theory*, 38(2):587–607, March 1992.

[4] C Carlson, E Adelson, and C Anderson. Improved system for coring an image representing signal, 1985. US Patent 4,523,230.

[5] N Ebihara and K Tatsuzawa. Noise reduction system, 1979. US Patent 4,163,258.

[6] D L Donoho. Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. In I. Daubechies, ed, *Proc Symp Appl Math*, vol 47, Providence, RI, 1993.

[7] D J Field. What is the goal of sensory coding? *Neural Computation*, 6:559–601, 1994.

[8] H J Trusell and R P Kruger. Comments on 'nonstationary assumptions for gaussian models in images'. *IEEE Trans SMC*, SMC-8:579–582, 1978.

[9] A N Netravali and B G Haskell. *Digital Pictures.* Plenum, New York, 1988.

[10] A. J. Kalb. Noise reduction in video images using coring on QMF pyramids. MIT, EECS Dept Senior Thesis, May 1991.

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

[12] E P Simoncelli and W T Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *2nd Int'l Conf Image Processing*, Wash, DC, October 1995.

[13] W T Freeman and E H Adelson. The design and use of steerable filters. *IEEE Trans Pat Anal Mach Intell*, 13(9):891–906, 1991.