

# IMAGE DENOISING VIA LOSSY COMPRESSION AND WAVELET THRESHOLDING

S. Grace Chang<sup>1</sup>

Bin Yu<sup>2</sup>

Martin Vetterli<sup>1,3</sup>

<sup>1</sup>Department of Electrical Engineering and Computer Sciences  
University of California, Berkeley, CA 94720, USA

<sup>2</sup>Department of Statistics  
University of California, Berkeley, CA 94720, USA

<sup>3</sup>Département d'Electricité  
Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland  
grchang@eecs.berkeley.edu, binyu@stat.berkeley.edu, vetterli@de.epfl.ch

## ABSTRACT

Some past work has proposed to use lossy compression to remove noise, based on the rationale that a reasonable compression method retains the dominant signal features more than the randomness of the noise. Building on this theme, we explain why compression (via coefficient quantization) is appropriate for filtering noise from signal by making the connection that quantization of transform coefficients approximates the operation of wavelet thresholding for denoising. That is, denoising is mainly due to the zero-zone and that the full precision of the thresholded coefficients is of secondary importance. The method of quantization is facilitated by a criterion similar to Rissanen's minimum description length principle. An important issue is the threshold value of the zero-zone (and of wavelet thresholding). For a natural image, it has been observed that its subband coefficients can be well modeled by a Laplacian distribution. With this assumption, we derive a threshold which is easy to compute and is intuitive. Experiments show that the proposed threshold performs close to optimal thresholding.

## 1. INTRODUCTION

Suppose an image has been corrupted by additive noise, and the goal is to remove the noise. The idea of using a lossy compression algorithm to denoise the signal has been proposed in several works [4, 7]. Continuing on this theme, one main purpose of this paper is to explain why lossy compression can be appropriate for noise filtering. More specifically, we wish to show that quantization (a common step in compression) of transform coefficients achieves denoising by posing quantization as an approximation to an effective denoising method called *wavelet thresholding* [2].

The theoretical formalization of thresholding in the context of removing noise via thresholding wavelet coefficients was pioneered by Donoho and Johnstone [2]. Both the soft- (shrink or kill) and the hard- (keep or kill) thresholding methods compare the input to a given threshold and set it to zero if its magnitude is less than the threshold. The idea is that coefficients insignificant relative to the threshold are likely due to noise, whereas significant coefficients are important signal structures. Thresholding essentially creates a region around zero where the coefficients are considered negligible. Outside of this region, the thresholded coefficients are kept to full precision.

Analogously, in a typical transform domain lossy compression method, negligible coefficients are set to zero, creating what is called a "zero-zone" or "dead-zone", and coefficients outside of this zone are quantized. Our hypothesis is that an appropriate quantization scheme (and hence compression) achieves denoising because it is an approximation to the thresholding operation (see Figure 1). Furthermore, the effectiveness of denoising is mainly due to the zero-zone, and the full precision of the thresholded coefficients is of sec-

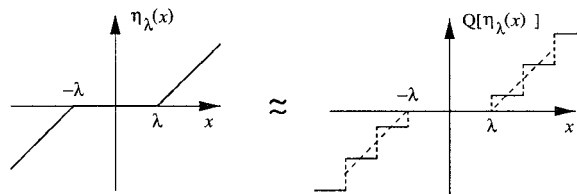


Figure 1. The thresholding function can be approximated by quantization with a zero-zone.

ondary importance. Thus, a comparable level of denoising performance can be achieved by quantizing the coefficients with a zero-zone and a few number of quantization levels outside of the zero-zone. The manner of quantization will be facilitated by a criterion similar to Rissanen's minimum description length (MDL) [5].

One of the most important and frequently asked questions in using wavelet thresholding is "What is the threshold?", or in the compression scenario, "how to choose the zero-zone?" While Donoho and Johnstone [2] have proposed several thresholds such as the universal ( $\sigma\sqrt{2\log(n)}$ ), SURE, and hybrid thresholds, and have demonstrated their asymptotic optimality conditions, these thresholds do not work well in practice. This is particularly true for images, and it is also rather counter-intuitive in signal processing applications to have the threshold values dependent on the sample size  $n$ . There are also many works of thresholding/shrinkage based on standard statistics techniques (e.g. Bayesian, cross-validation), but most of them are not suitable for images and some are rather computationally intensive. Here we propose a threshold for soft-thresholding images which is simple and straightforward to compute.

A large class of natural images has decaying spectrums, which means that the subband energy also follows a certain decay across scales. Within each subband, the coefficients can be well modeled by a generalized Laplacian distribution [3]. Assuming a Laplacian distribution, the proposed threshold is an approximation to the optimal threshold which minimizes the expected squared error among soft-threshold estimators. A different threshold is computed for each subband to adapt to the changing subband characteristics.

## 2. WAVELET THRESHOLDING AND THRESHOLD SELECTION

**Definitions and Notations** For simpler notations, variables are referenced by a single index even though we are working in the 2-D domain.

Suppose the signal  $\mathbf{x} = \{x_i, i = 1, \dots, n\}$  has been corrupted and one observes  $\mathbf{y} = \{y_i = x_i + v_i, i = 1, \dots, n\}$  where  $v_i$  is iid  $N(0, \sigma^2)$  and independent of  $x_i$ . Let  $\mathbf{Y} = \mathcal{W}\mathbf{y}$  denote the vector of wavelet coefficients of  $\mathbf{y}$ , where  $\mathcal{W}$  is

the orthogonal wavelet transform operator, and similarly with  $\mathbf{X}$  and  $\mathbf{V}$ . The readers are referred to [3] for details on wavelet subband decomposition.

Since the transform is orthogonal,  $V_i$  is also *iid* Gaussian  $N(0, \sigma^2)$ . The idea of wavelet thresholding suggests filtering noise from  $\mathbf{y}$  by thresholding its wavelet coefficients (except the coarsest scale coefficients). The soft-thresholding function, defined as  $\eta_\lambda(t) = \text{sgn}(t)(|t| - \lambda)_+$ , where  $\lambda$  is the threshold, is used here because it generally yields visually more pleasing images over hard-thresholding, defined as  $\psi_\lambda(t) = t \cdot 1_{t > \lambda}$ . The denoised estimate  $\hat{\mathbf{x}}$  is then the inverse transform of  $\eta_\lambda(\mathbf{Y})$ ,  $\hat{\mathbf{x}} = \mathbf{W}^{-1} \eta_\lambda(\mathbf{Y})$ .

**Derivation of Threshold** Suppose that we restrict the estimate of  $\mathbf{X}$  to be in the class of soft-threshold estimates,  $\hat{\mathbf{X}} = \eta_\lambda(\mathbf{Y})$ , the goal is to derive a threshold  $\lambda$  which minimizes the averaged squared error,  $1/N \sum_i (\hat{X}_i - X_i)^2$ .

For a large class of natural images, it has been observed that the coefficients in each subband of its wavelet transform (with the exception of the lowest scale) can be well described by a generalized Laplacian distribution [3]. For this work, we assume the Laplacian pdf for simplicity.

Consider now only coefficients from one subband. Let  $X \sim \text{LAP}(\sqrt{2\alpha}) = \frac{1}{2}\sqrt{2\alpha} e^{-\sqrt{2\alpha}|X|}$ . For a large number of coefficients,  $1/N \sum_i (\hat{X}_i - X_i)^2 \approx E(\hat{X} - X)^2$ . Thus, we proceed to minimize  $E(\hat{X} - X)^2$ . It can be shown that the Laplacian pdf is a scaled mixture of normals, and that the denoising problem can be reformulated with the following priors:  $Y|X \sim N(X, \sigma_x^2)$ ,  $X|\sigma_x^2 \sim N(0, \sigma_x^2)$ , and  $\sigma_x^2 \sim \text{EXP}(\alpha) = \alpha e^{-\alpha\sigma_x^2}$ ,  $\sigma_x^2 \geq 0$ .

Breaking the problem into three priors gives implementation advantages because it requires numerical integration of only one variable. That is,

$$\begin{aligned} E(\hat{X} - X)^2 &= \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty (\eta_\lambda(Y) - X)^2 \\ &\quad p(Y|X) p(X|\sigma_x^2) p(\sigma_x^2) dY dX d\sigma_x^2 \\ &= \int_0^\infty p(\sigma_x^2) \sigma_x^2 f\left(\frac{\sigma_x^2}{\sigma^2}, \frac{\lambda}{\sigma}\right) d\sigma_x^2 \triangleq \sigma^2 g(\sigma^2 \alpha, \frac{\lambda}{\sigma}) \end{aligned}$$

where

$$\begin{aligned} f(\sigma_x^2, \lambda) &= \sigma_x^2 + 2(\lambda^2 + 1 - \sigma_x^2) \Phi\left(\frac{\lambda}{\sqrt{1 + \sigma_x^2}}\right) \\ &\quad - 2\lambda(1 + \sigma_x^2) \phi(\lambda, 1 + \sigma_x^2), \end{aligned}$$

$$\phi(x, \beta^2) = \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{x^2}{2\beta^2}\right) \text{ and } \Phi(x) = \int_x^\infty \phi(t, 1) dt.$$

Without loss of generality, assume  $\sigma = 1$ . The optimal threshold for each  $\alpha$  is  $\lambda^*(\alpha) = \arg \min_{\lambda \geq 0} g(\alpha, \lambda)$ . The curve of  $\lambda^*(\alpha)$  is plotted against  $1/\sqrt{\alpha}$  on the  $x$ -axis in Figure 2 (a). This curve is compared with  $\tilde{\lambda}(\alpha) = \sqrt{\alpha}$  which is seen to approximate closely  $\lambda^*(\alpha)$ , and yet is simple and does not need to be calculated through numerical minimization. Their corresponding expected squared errors are shown in Figure 2 (b), and the difference is within 1%.

The choice of the threshold  $\tilde{\lambda}(\alpha) = \sqrt{\alpha}$  also makes intuitive sense. For  $X \sim \text{LAP}(\sqrt{2\alpha})$ , its standard of deviation is  $\text{Std}(X) = 1/\sqrt{\alpha}$ , and  $\tilde{\lambda}(\alpha)$  is inversely proportional to  $\text{Std}(X)$ . Thus, when  $\text{Std}(X)$  is much larger than 1 (recall that  $\sigma = 1$ ), the signal is much stronger than the noise, thus the threshold is chosen to be small to preserve most of the signal and remove some of the noise; vice versa, when  $\text{Std}(X)$  is small relative to 1, the noise dominates and the threshold is large to remove the noise which has

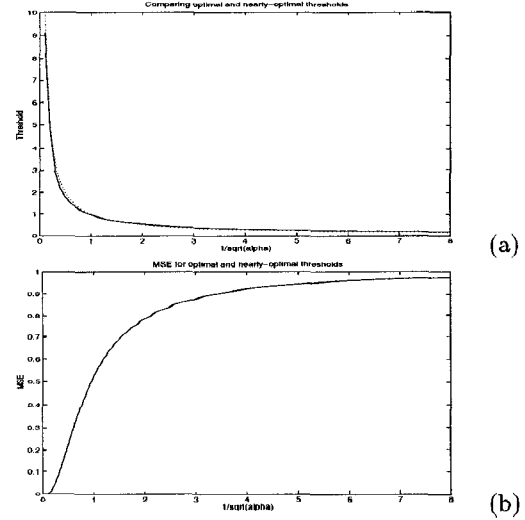


Figure 2. (a) Comparing the optimal threshold  $\lambda^*(\alpha)$  (solid —) and the threshold  $\tilde{\lambda}(\alpha)$  (dotted ···) against  $1/\sqrt{\alpha}$  on the  $x$ -axis. (b) Their corresponding MSEs, with (—) for  $\lambda^*(\alpha)$  and (···) for  $\tilde{\lambda}(\alpha)$ .

overwhelmed the signal. It is also interesting to note that if  $\sigma_x^2$  were treated as deterministic (that is,  $X \sim N(0, \sigma_x^2)$  and  $\sigma_x^2$  is to be estimated), then the threshold  $\lambda(\sigma_x) = 1/\sigma_x$  also approximates the corresponding optimal threshold very well. Now for a general value of  $\sigma$ , the above discussion holds by replacing  $\lambda$ ,  $\alpha$ , and  $\sigma_x$  by  $\lambda/\sigma$ ,  $\sigma^2\alpha$ , and  $\sigma_x/\sigma$ , respectively, and our proposed threshold is

$$\tilde{\lambda}(\alpha) = \sigma^2 \sqrt{\alpha}. \quad (1)$$

A similar threshold to (1) is found independently in [6] for using the same priors with the hard-thresholding function. Soft-thresholding is used here because it is more suitable for images. Furthermore, with these priors, the expected squared error of optimal soft-thresholding is smaller than that of optimal hard-thresholding.

There are two parameters to be estimated: the noise variance  $\sigma^2$  and the hyperparameter,  $\alpha$ , of the Laplacian pdf. The noise variance is estimated by the robust median estimator in the highest subband,  $\hat{\sigma} = \text{Median}(|Y_i|)/.6745$ ,  $Y_i \in$  subband  $\text{HH}_1$ , also used in [2]. Since the signal and noise are assumed to be independent,  $\text{Var}(Y) = 1/\alpha + \sigma^2$ , thus  $\hat{\alpha} = \text{SampleVariance}(\mathbf{Y}) - \hat{\sigma}^2$ . In the rare case that  $\hat{\sigma}^2 > \text{SampleVariance}(\mathbf{Y})$ , the threshold is set to the maximum value of that subband,  $\lambda = \max |\mathbf{Y}|$ ; that is, all coefficients are set to 0.

The above method for estimating  $\hat{\alpha}$  and  $\tilde{\lambda}(\hat{\alpha})$  are done for each subband independently to adapt to the different characteristics.

### 3. QUANTIZATION WITH MDL CRITERION

In the original thresholding scheme, the thresholded coefficients are then inverse transformed back to yield the estimate  $\hat{\mathbf{x}}$ . In this work, to show that quantization approximates thresholding, there is an additional step of quantizing the thresholded coefficients before inverse transform.

Consider again only one particular subband of the wavelet transform, and that the parameters  $\hat{\sigma}$ ,  $\hat{\alpha}$  and the threshold  $\tilde{\lambda}(\hat{\alpha})$  have been calculated. After thresholding, there are the questions of how many quantization levels and what type of

reconstruction values (e.g. uniform, centroid). We propose to use an MDL-like criterion to facilitate this decision.

For a given set of observations, the MDL criterion is useful for choosing a reasonable statistical model which yields the shortest description length [5]. It does this by choosing the model which minimizes the total code-length of a two-part encoding consisting of the data (based on the chosen model) and of the model parameters. The idea is that the chosen model should establish a compromise between fitting the data well and having low complexity (i.e. having a simple representation or a reasonable number of parameters).

More specifically, given the set of noisy transform coefficients  $\mathbf{Y} = \mathbf{X} + \mathbf{V}$ , the framework is to code  $\mathbf{Y}$  given the model  $\mathbf{X}$ . The MDL principle chooses  $\mathbf{X}$  which minimizes the two-part code-length  $L_{\text{total}}(\mathbf{Y}, \mathbf{X}) = L(\mathbf{Y}|\mathbf{X}) + L(\mathbf{X})$ , where  $L(\mathbf{Y}|\mathbf{X})$  is the code-length for  $\mathbf{Y}$  based on  $\mathbf{X}$ , and  $L(\mathbf{X})$  is the code-length for  $\mathbf{X}$ . If it is possible to associate a probability distribution  $p(\cdot)$ , then one can use the idealized code-length  $-\log_2 p$ . In this case where the noise  $V_i$  is iid  $N(0, \sigma^2)$ , the first term becomes

$$\begin{aligned} L(\mathbf{Y}|\mathbf{X}) &= -\log_2 p(\mathbf{Y}|\mathbf{X}) \\ &= \frac{1}{2(\ln 2)\sigma^2} \sum_{i=1}^n (Y_i - X_i)^2 + \frac{1}{2} \log_2 (2\pi\sigma^2n) \end{aligned} \quad (2)$$

The second term in (2) is a constant and irrelevant in the minimization, and thus only the first term is considered.

In Saito's simultaneous compression and denoising method [7] of combining MDL with thresholding, the hard-thresholding function was used and the term  $L(\mathbf{X})$  was taken to be  $(3/2)k \log_2 n$ , where  $k$  is the number of nonzero coefficients:  $\log_2 n$  bits to indicate the location of each nonzero coefficient (assuming a uniform indexing) and  $(1/2)\log_2 n$  bits to represent its value. Although compression has been achieved in the sense that a fewer number of nonzero coefficients are kept, it still does not address the issue that in a practical setting, the coefficients are usually quantized. Thus, our criterion is developed from a coding point of view, and the minimization of  $L_{\text{total}}(\mathbf{Y}, \mathbf{X})$  is restricted to  $\mathbf{X}$  belonging to the set of quantized signals, whose construction will become clear in the following text.

After the zero-zone has been determined (by the threshold), there are  $k$  nonzero coefficients and  $n - k$  zero coefficients to be quantized and coded. For a given threshold, the value  $k$  is fixed and so is the bitrate for coding the locations of zero coefficients (e.g. a naive way is to use  $\log_2 n$  bits to index each of the  $n - k$  nonzero coefficients or, more realistically, to use runlength encoding). Thus, this term is again neglected in the minimization.

For the nonzero coefficients, on each positive and negative side,  $m$  equal-width bins are partitioned between 0 and the maximum absolute coefficient value,  $\max |\mathbf{Y}|$ . The reconstruction values are taken to be the centroid of each bin, assuming a Laplacian distribution,  $\text{LAP}(\sqrt{2}\alpha)$ , with the parameter estimate  $\hat{\alpha}$  as described in Section 2. Let  $\gamma = \sqrt{2}\alpha$ , the closed-form expression for the centroid value of a bin on the positive side with boundaries  $t_i$  and  $t_{i+1}$  is

$$r_i = \frac{\int_{t_i}^{t_{i+1}} xp(x)dx}{\int_{t_i}^{t_{i+1}} p(x)dx} = \frac{1}{\gamma} + \frac{t_i e^{-\gamma t_i} - t_{i+1} e^{-\gamma t_{i+1}}}{e^{-\gamma t_i} - e^{-\gamma t_{i+1}}}.$$

The negative side is similar. The indices of  $t_i$  are  $i \in \{0, \pm 1, \dots, \pm m\}$  and the indices of  $r_i$  are  $i \in \{\pm 1, \dots, \pm m\}$  for the positive and negative sides.

To code the quantized value, first one needs to transmit the range  $\max |\mathbf{Y}|$  and the value of  $m$ . These are fixed overhead and will be ignored in the minimization. Then entropy

encoding is typically used to transmit the bin index of each coefficient. A good estimate of the bitrate for the  $k$  nonzero coefficients is the entropy  $H(m) = -\sum_i k_i \log_2(k_i/k)$ , where  $k_i$ ,  $i \in \{\pm 1, \dots, \pm m\}$ , is the number of coefficients with reconstruction value  $r_i$ .

Let  $Q[x, m]$  denote the operation of quantizing the input  $x$  with  $2m + 1$  levels (including the zero-zone) as described above, and let  $\hat{X}_i^Q = Q[\eta_\lambda(Y_i), m]$  denote the estimate of  $X_i$  obtained by soft-thresholding followed by quantization. We choose the estimate  $\hat{X}^Q$  with the associated  $m$  which minimizes the criterion

$$MDLQ = \frac{1}{2(\ln 2)\sigma^2} \sum_{i=1}^n (Y_i - \hat{X}_i^Q)^2 + H(m) \quad (3)$$

The space-domain estimate is taken to be the inverse transform of  $\hat{X}^Q$ .

Note that the quantized estimate naturally could do worse than the unquantized threshold estimate. However, it is not our goal to achieve better than the unquantized estimate, but rather to establish a connection between compression (via quantization) and thresholding to show that lossy compression can be a good method for noise removal.

This thresholding-quantization scheme is applied to each subband independently. First the noise variance  $\hat{\sigma}^2$  is estimated. Then for each subband, the parameter  $\hat{\alpha}$  and the threshold  $\tilde{\lambda}(\hat{\alpha})$  are calculated, and (3) is minimized to find the desired quantized coefficients. The coarsest component LL is not thresholded but is quantized using (3) with the uniform distribution.

#### 4. EXPERIMENTAL RESULTS

The images "goldhill" and "lena", with various noise strength  $\sigma = 10, 15, 20$ , are used as test data. The chosen wavelet is Daubechies' least asymmetric compactly-supported wavelet with 8 vanishing moments [1], and 4 levels of decomposition is used. Table 1 compares the SNRs of the soft-thresholding denoised results using the oracle threshold,  $\lambda_{\text{orc}}$ , and  $\tilde{\lambda}(\hat{\alpha})$  (adaptive in each subband). For each subband, the oracle threshold is found by  $\lambda_{\text{orc}} = \arg \min_\lambda \sum_i (\eta_\lambda(Y_i) - X_i)^2$  assuming  $X_i$  is known. The SNRs resulting from using  $\tilde{\lambda}(\hat{\alpha})$  are very close to those of  $\lambda_{\text{orc}}$ , indicating that the Laplacian pdf is a good assumption and that the estimation of the parameters is appropriate. Visually, the two sets of images are also very similar (see Figure 3 (b) and (c)).

The last column in Table 1 shows the SNR of the quantized signal using  $\tilde{\lambda}(\hat{\alpha})$  as the zero-zone threshold. The quantized goldhill image with  $\sigma = 15$  is shown in Figure 3(d), where the quantization noise is quite visible. As expected, the quantized signal uses less bits, but at the expense of some degradation. On the average, the quantized signal loses about 1-1.5 dB over the unquantized thresholded signal, although it still has a much higher SNR than the noisy image. This suggests that mainly the zero-zone is responsible for filtering the noise. One thing to note is that the zeroth-order entropy estimate,  $H(m)$ , for the bitrate of the nonzero coefficients is a rather loose estimate. With more sophisticated coding, the same bitrate could yield a higher number of quantization level  $m$ , thus resulting in a better SNR.

Table 2 shows the values of  $m$  chosen for each subband of the goldhill image,  $\sigma = 15$ , averaged over 5 runs. The results show that the MDLQ criterion allocates more levels in the coarser, more important levels, as would be the case in a practical subband coding situation.

SNR	noisy	$\lambda_{orc}$	$\tilde{\lambda}$	$\tilde{\lambda}, Q[.,m]$
goldhill $\sigma=10$	13.84	17.68	17.61	16.12
$\sigma=15$	10.32	15.73	15.63	14.47
$\sigma=20$	7.82	14.46	14.35	13.38
lena $\sigma=10$	13.60	19.10	18.89	17.27
$\sigma=15$	10.08	17.19	16.99	15.57
$\sigma=20$	7.58	15.87	15.71	14.38

Table 1. SNRs (in dB) of (1) the noisy image, (2) oracle soft-thresholding, (3) soft-thresholding with thresholds  $\tilde{\lambda}$ , and (4) quantized signal with zero-zone thresholds  $\tilde{\lambda}$ . Averaged over 5 runs.

Orientation	Scale			
	1 (fine)	2	3	4 (coarse)
HH	0	2.0	3.6	6.2
HL	2.6	4.0	6.0	18.6
LH	2.8	3.6	6.4	12.2
LL	39.0			

Table 2. The value of  $m$  (averaged over 5 runs) for the different subbands of Goldhill, with noise strength  $\sigma = 15$ .

## 5. DISCUSSION AND CONCLUSION

We have addressed two main issues regarding image denoising in the paper. We demonstrated the connection between lossy compression and wavelet thresholding to explain why compression is suitable for denoising. Specifically, it is the zero-zone in coefficient quantization that is the main agent in removing the noise. A suitable threshold for images has also been proposed for wavelet thresholding and for the quantization zero-zone. Results suggest that the proposed method may be appropriate for subband coding, in the decision of bit allocation and the manner of quantization. For future work, it would be interesting to jointly compute the best zero-zone threshold and quantization bins rather than compute them separately.

## REFERENCES

- [1] I. Daubechies, *Ten Lectures on Wavelets*, vol. 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, SIAM, Philadelphia, 1992.
- [2] D.L. Donoho and I.M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol 81, pp. 425-455, 1994.
- [3] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Pat. Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674-693, July 1989.
- [4] B.K. Natarajan, "Filtering random noise from deterministic signals via data compression," *IEEE Trans. on Signal Processing*, vol. 43, no. 11, pp. 2595-2605, November 1995.
- [5] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, World Scientific, 1989.
- [6] F. Ruggeri and B. Vidakovic, "A Bayesian decision theoretic approach to wavelet thresholding," preprint, Duke University, Durham, NC. (<ftp://ftp.isds.duke.edu/pub/Users/brani/papers/Decision.ps>).
- [7] N. Saito, "Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion," in *Wavelets in Geophysics*, Eds. E. Foufoula-Georgiou and P. Kumar, pp. 299-324, Academic Press, 1994.

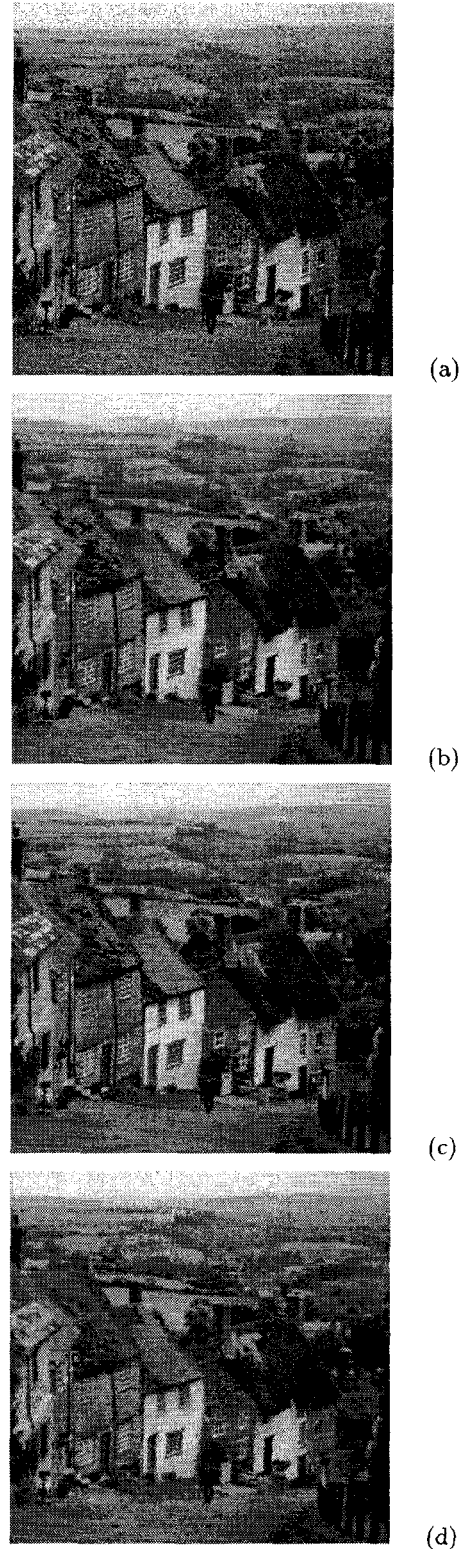


Figure 3. (a) Noisy image,  $\sigma = 15$ . (b) Oracle soft-thresholding. (c) Threshold with  $\hat{\lambda}(\hat{\alpha})$ . (d) Our method of thresholding followed by quantization.