Subband Noise Estimation for Adaptive Wavelet Shrinkage

Xiaohui Yuan and Bill P. Buckles
EECS Dept., Tulane University, New Orleans, LA 70118
{yuanx, buckles}@eecs.tulane.edu

Abstract

In this article, we present an adaptive image denoising method based on subband noise modeling. For wavelet shrinkage, choosing the threshold depends on correctly estimating the noise variance. By modeling the inter-subband noise variance with a parameterized normalized exponential function, the problem becomes identifying the maximum noise variance. Such a maximum exists in the highest decomposition level and can be estimated by locating the extreme of the first derivative of the subband variance function. The experiments demonstrate that our method outperforms peers, especially in the cases of large noise variance.

1. Introduction

In many image processing applications, e.g. image fusion and segmentation, it is beneficial if noise characteristics can be estimated and provided with the denoised image, so that they can be incorporated intelligently in subsequent processings. Difficulties, however, lie in the fact that noise is spatially indistinguishable from the truth image. Much recent research has focused on noise removal using nonlinear wavelet thresholding [4, 5, 10]. This technique, namely wavelet shrinkage, circumvents the confliction of removing noise and preserving details by using a nonlinear mapping in the wavelet domain. It is shown that the wavelet shrinkage scheme is near-optimal in the sense of minimizing the least square error [4, 5, 8]. More approaches can be found in [2] and the references therein.

In the mapping functions, the threshold essentially decides to what degree a coefficient contributes to the signal. The key of determining this threshold is accurate estimation of the noise variance. Donoho et al. [4] proposed the universal threshold, which is given by

$$\lambda = \sigma_n \sqrt{2 \log N}$$

where $N$ is the number of pixels in the image and $\sigma_n$ is the noise standard deviation and assumed known. $\lambda$ tends to be high for large $N$ and attenuates many signal coefficients along with the noise. To solve this problem, a hybrid scheme is developed that uses the universal threshold or SURE threshold depending on the energy of a subband [6]. The SURE threshold is given by

$$\arg \min_{0 \leq \lambda \leq \sqrt{2 \log N}} N - 2 |w| |w| \leq \lambda + \sum \min(|w|, \lambda)^2$$

Chang et al. [2] derived the threshold by minimizing the Bayes risk, i.e. $\sigma_n^2 / \sigma_s^2$, $\sigma_s^2$ is the signal variance. Noise variance is computed using the median value of high-pass subband.

$$\hat{\sigma}_n = \frac{\text{Median}(|Y_{i,j}|)}{0.6745}, \quad Y_{i,j} \in \text{subband } HH_1$$

Since the noise variance estimation tends to be smaller than the actual, the denoised images are usually over-smoothed or sub-optimal in the sense of SNR.

In this article, we describe a method, namely Bayesian shrinkage using noise modeling (BSNM), to estimate the noise variance adaptively for each subband and perform the shrinkage using the general Gaussian density (GGD) prior. A subband of a noise-free (or clean) image is sparse and locally correlated. The histogram of such a subband resembles super Gaussian distribution in that it has high peak and long tails. Figure 1 illustrates examples of wavelet subband histograms. Moreover, noise power decreases along with the wavelet decomposition level, as illustrated in Figure 2. We model the inter-subband noise variance with an exponential function. Hence, noise variance estimation becomes identifying the maximum of the first derivative of variance shrinkage function.

![Figure 1. Histogram of Wavelet subbands.](image-url)
In the next section, we describe an adaptive subband noise estimation approach and incorporate the estimation into GGD-based Bayesian shrinkage. In Section 3, we demonstrate experimental results and compare it with peer approaches. Conclusions are in Section 4.

2. Noise Estimation and Bayesian Shrinkage

2.1 Subband Noise Modeling

Subband coefficients of Gaussian noise retains the Gaussian distribution with an enlarged or suppressed variance depending on the decomposition level. Figure 2(a) illustrates the variance estimations of Gaussian noise subbands. The noise variance attenuates exponentially along with the increasing decomposition level. Yet, within the same level, variances are almost equal and, therefore, can be approximated with an average variance, as illustrated in Figure 2(b). That is, the noise variance is decomposition-level dependent.

![Figure 2. Wavelet subband noise variance.](image)

Let \( l \) denote the wavelet decomposition level. The subband noise variances are modeled with an exponential function as follows.

\[
\sigma_n^2(l; a, b, c) = ae^{-(l+b)c}
\]  

(1)

More than one noise variances, \( \sigma_n(l) \), are needed to fully determine such a function. Obtaining \( \sigma_n \), however, is difficult since in a high-level noisy subband, signal variance is much greater than noise variance, which makes the estimation inaccurate. Let \( a = 1 \) and \( b = 0 \), and normalize the noise variance to [0 1], as illustrated in Figure 3(a). Hence, the model becomes

\[
\bar{\sigma}_n^2(l; c) = \frac{e^{-l\epsilon}}{\max(e^{-l\epsilon})}
\]  

(2)

where \( \bar{\sigma} \) is the normalized noise variance.

It can be proven that when \( c \approx 1.2 \) the model fits the subband noise variance best in the sense of minimizing mean square error. Therefore, the subband noise variances are estimated as follows.

\[
\hat{\sigma}^2_n(l) = \max(\bar{\sigma}_n^2(l; 1.2)) = \sigma_n^2(l = 1)e^{1-l^2}
\]  

(3)

A fitted subband noise model is illustrated in Figure 3(b). The solid line is the subband noise function and the dotted line illustrates the estimated model.

![Figure 3. Wavelet subband noise model and fitted normalized variance.](image)

2.2 Noise Variance Estimation

As illustrated in Figure 1(a), a clean wavelet subband is sparse in that most coefficients are of small magnitude. Intuitively, to estimate the noise variance we can choose those small magnitude coefficients and estimate the noise variance as follows.

\[
\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{m=1}^{n} (I_{o}^* - \bar{I}_o^*)^2
\]  

(4)

where \( I_{o}^* \) consists of only small magnitude coefficients and \( \bar{I}_o^* \) denotes the mean of \( I_{o}^* \). Let \( \tau, \tau = [\bar{I}_o - \epsilon, \bar{I}_o + \epsilon] \), be a neighborhood on the quantization axis (as shown in Figure 4) such that

\[
\forall I(m,n) \in \tau, \quad I(m,n) \in I_{o}^*.
\]  

(5)

Obviously, to optimally estimate the noise variance \( \sigma_n \), two criteria shall be satisfied

1) The noise variance within \( \tau \) is much greater than that of the underlying clean signal, i.e., \( \sigma_n^2 \gg \sigma_s^2 \), and \( \sigma_s^2 \) is the variance of \( I_{o}^* \);

2) The sample size for noise estimation is statistically large, i.e., \( \| I_{o}^* \| \gg 1 \).
Figure 5. Therefore, the projection of each line onto X-axis suggests the maximum magnitude. The noise variance is estimated as illustrated in Figure 5. The projection of each line onto X-axis suggests the maximum magnitude.

Although, in general, \( \sigma_s \) is comparable to \( \sigma_n \), the maximum absolute value of a clean subband is much greater than that of the noise. Let \( \zeta_n \) denote the maximum coefficient magnitude of the subband noise. When the increase of \( \epsilon \) passes \( \zeta_n \), the trend of \( \sigma_n^*(\epsilon) \) matches that of \( \sigma_s^*(\epsilon) \), as illustrated in Figure 5. The projection of each line onto X-axis suggests the maximum magnitude.

\[ \sigma_n^*(\epsilon) = \sigma_s^*(\epsilon) + \sigma_n^2(\epsilon) \approx \sigma_n^2(\epsilon) = \sigma_n^2 \] (6)

This can simply be computed by locating the zero crossing of the second derivative. The noise variance is estimated using Equation 3.

2.3 Bayesian Threshold Selection Using General Gaussian Density Prior

An approximation for the histogram of coefficients at a clean wavelet subband may be achieved by General Gaussian Density (GGD) [3, 7, 9]. The GGD model has been widely used in many image processing applications [1, 3, 7] and is given as follows.

\[ \mathcal{P}(x; \beta, \theta, \sigma) = \frac{\beta}{2\sigma \Gamma(1/\beta)} e^{-((x-\theta)/\sigma)^\beta} \] (8)

where \( \Gamma(\cdot) = \int_0^\infty e^{-tz^{-1}} dt, z > 0 \), is the Gamma function. \( \sigma \) is the standard deviation and determines the width of the distribution, while \( \beta \) is inversely proportional to the decreasing rate of the peak.

In obtaining the threshold, our objective is to minimize the risk function as follows

\[ \lambda = \arg \min_{\lambda} E(\hat{I}_s(\lambda) - I_s)^2 \] (9)

Assume that subband noise is Gaussian with zero mean and variance \( \sigma_n^2 \), the risk function can be written as

\[ \lambda = \arg \min_{\lambda} \int p(I_s)dI_s \int (\hat{I}_s - I_s)^2 p(I_o|I_s)dI_o \]

Although it is difficult to develop an analytical solution, a numerical solution exists [2].

\[ \lambda = \frac{\sigma_n^2}{\sigma_s} \] (10)

where \( \sigma_s \) is obtained by

\[ \sigma_n^2 = \sigma_s^2 + \sigma_n^2 \] (11)

This threshold is used to shrink wavelet coefficients with soft thresholding.

3. Experiments

We constructed a three-level wavelet decomposition. For each level, we estimated the subband noise variance with Equation 3. Table 1 compares our method, BSNM, with BayesShrink, VisuShrink and Wiener filtering using Mean Square Error (MSE). The results were averaged over 10 runs. It is interesting that the Wiener method performs better given large noise standard deviation (Std) than it does given small Std. Similarly, our method gives best performance except when noise Std is 5. This is because when noise Std is small, signal variance dominates the subband.
Table 1. Averaged MSE of denoising results.

<table>
<thead>
<tr>
<th>N.Std</th>
<th>Noisy BSNM</th>
<th>BS</th>
<th>VS</th>
<th>Wiener</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>24.6</td>
<td>20.1</td>
<td>19.8</td>
<td>118.1</td>
</tr>
<tr>
<td>10</td>
<td>97.7</td>
<td>56.3</td>
<td>58.5</td>
<td>242.1</td>
</tr>
<tr>
<td>15</td>
<td>217.0</td>
<td>92.4</td>
<td>102.0</td>
<td>338.5</td>
</tr>
<tr>
<td>20</td>
<td>382.0</td>
<td>144.3</td>
<td>147.9</td>
<td>414.7</td>
</tr>
<tr>
<td>25</td>
<td>591.7</td>
<td>176.2</td>
<td>194.7</td>
<td>476.0</td>
</tr>
<tr>
<td>30</td>
<td>841.3</td>
<td>233.2</td>
<td>238.8</td>
<td>524.0</td>
</tr>
<tr>
<td>35</td>
<td>1124.7</td>
<td>249.4</td>
<td>283.3</td>
<td>564.5</td>
</tr>
</tbody>
</table>

variance function. Hence, the inflection point may not represent the maximum contribution of noise to variance function. Generally, the estimation of $\sigma^2_n(1)$ tends to be greater than the optimal.

Figure 6 illustrates the denoising images using four methods. The noise Std in Figure 6(b) is 20. It is clear that VisuShrink and Wienner method over-smooth or partially over-smooth the image, Figure 6(e) and (f) respectively. For BayesShrink, noise, especially that in the background, still exists, which is successfully attenuated with our method as illustrated in Figure 6(c).

4. Conclusions

In this article, we present a method for adaptive wavelet shrinkage based on exponential noise model. In the wavelet shrinkage, choosing the threshold depends on correctly estimating the noise variance. By modeling the inter-subband noise variance with a parameterized normalized exponential function, the problem becomes identifying the maximum noise variance. Such a maximum exists in the highest decomposition level and can be estimated by locating the extreme of the first derivative of the subband variance function.

References