Image denoising exploiting inter- and intra-scale dependency in complex wavelet domain

Fengxia Yan (严奉霞)[1,2] and Lizhi Cheng (成礼智)[1]

[1]*School of Science, National University of Defense Technology, Changsha 410073*
[2]*Institute of Automation, Chinese Academy of Sciences, Beijing 100080*

Received May 26, 2006

A new locally adaptive image denoising method, which exploits the intra-scale and inter-scale dependency in the dual-tree complex wavelet domain, is presented. Firstly, a recently emerged bivariate shrinkage rule is extended to a complex coefficient and its neighborhood, the corresponding nonlinear threshold functions are derived from the models using Bayesian estimation theory. Secondly, an adaptive weight, which is able to capture the inter-scale dependency of the complex wavelet coefficients, is combined to the obtained bishrink threshold. The experimental results demonstrate an improved denoising performance over related earlier techniques both in peak signal-to-noise ratio (PSNR) and visual effect.

OCIS codes: 100.0100, 110.4280.

An image is often corrupted by noise in its acquisition or transmission and noise elimination is a main concern in computer vision and image processing. In image denoising, a compromise has to be found between noise reduction and preserving significant image details. To achieve a good performance in this respect, a denoising algorithm has to be adapted to image discontinuities. The wavelet representation naturally facilitates the construction of such spatially adaptive algorithms. Smooth image regions are represented by small wavelet coefficients, while edges, ridges, and other singularities are represented by large coefficients. Because of this property, additive noise can be effectively suppressed by simple thresholding[1] of the wavelet coefficients.

Recently, statistical approach has emerged as a new tool for wavelet-based denoising. The basic idea is to model wavelet transform coefficients with prior probability distributions. Statistical models pretended wavelet coefficients are random variables described by some probability distribution. Most models assume that the coefficients are independent and try to characterize them by using Gaussian, Laplacian, generalized Gaussian, or other distributions. Although the wavelet transform nearly decorrelates many images, significant dependencies still exist between wavelet coefficients[2]. Algorithms that exploit the dependency between coefficients can give better results compared with the ones derived by using an independence assumption. For example, Crouse *et al.*[3] developed a new framework to capture the statistical dependencies by using wavelet-domain hidden Markov tree (HMT) models. New bivariate shrinkage[4] functions were proposed by considering the dependencies between the coefficients and their parents. Chen *et al.*[5] introduced an efficient, adaptive threshold denoising algorithm via wavelet soft-thresholding by exploiting the inter-scale persistence. In addition, an information-theoretic analysis of statistical dependencies between wavelet coefficients is described by Liu *et al.*[6].

However, current wavelet-based image denoising algorithms have a tendency to produce denoised images with ringing artifacts around the edges. The main reason is that the traditional critically sampled wavelet transform is not shift-invariant and has not good directional selectivity. The dual-tree complex wavelet transform[7,8] (DT-CWT) is a good way to counteract these problems. Sendur *et al.*[4] applied one bivariate shrinkage function to the magnitude of the DT-CWT coefficients and obtained better denoising performance than using it with the critically sampled wavelet. Wang *et al.*[9] also used this bivariate shrinkage rule with different thresholds in dual-tree complex wavelet domain to reduce noise in their noisy image enhancement method. There are strong dependencies between neighbor coefficients such as between a coefficient and their siblings (adjacent spatial locations), and its parent (adjacent coarser scale locations). But both Sendur and Wang's algorithms only consider the correlations between the coefficients and their parents, i.e., the inter-scale dependencies. In this paper, we address the image denoising problem by exploiting the intra-scale and the inter-scale dependencies simultaneously in the complex wavelet domain. In order to make use of the intra-scale dependency between coefficients, the non-Gaussian bivariate model[4] is extended to a coefficient and its neighborhood, and the corresponding nonlinear threshold function is derived from the model using Bayesian estimation theory. And then, we investigate the coefficients statistics in the dual-tree complex wavelet domain to obtain an adaptive weight to capture the inter-scale dependencies. Lastly, we integrate the adaptive weight to the bishrink threshold to obtain our denoising algorithm.

DT-CWT is a valuable enhancement to the traditional real discrete wavelet transform (DWT), with important additional properties: it is nearly shift invariant and directionally selective in two and higher dimensions. These two properties are very favorable to the image denoising application. The DT-CWT can give a substantial performance boost to DWT-based noise reduction algorithms. Furthermore, denoising algorithms based on statistical models of wavelet coefficients can be more effective for the CWT than for the real DWT because the magnitudes

1671-7694/2007/030156-04

of the coefficients are more dependent in inter-scale and intra-scale neighborhoods[8,10].

The performance gains provided by the DT-CWT come from designing the filters in the two filter banks approximately. The coefficients produced by these filter banks are the real and imaginary parts of a complex coefficient. Assume the sets of coefficients u_i and v_i are produced by these filter banks separately, and the complex coefficients can be represented by $c_i = u_i + iv_i$. The property of near shift invariance means that small signal shifts do not affect the magnitudes of the complex coefficients ($|c_i| = \sqrt{u_i^2 + v_i^2}$). The basis functions have directional selectivity property at $\pm 15°$, $\pm 45°$, $\pm 75°$, which the regular critically sampled transform does not have. And there are six high-frequency directional subbands at each level in the decomposition of two-dimensional (2D) image.

Firstly, we mainly focus on the dependencies between a coefficient and its siblings (neighborhood in the same subband), i.e., the intra-scale dependencies. Consider an image subband and define the neighborhood N_X as the collection of the eight coefficients adjacent to X[6]. To avoid the so-called curse of dimensionality, we would like to assume that the neighborhood $N\_X = \{N\_X_1, \cdots, N\_X_8\}$ provides information to X only through a many-to-one scalar function $T = f(N\_X)$ in the sense that equality of the corresponding mutual information, i.e., $I(X; N\_X) = I(X; T)$, where $I(\cdot)$ represent the mutual information[6]. Simoncelli[11] assumed that the squared magnitude of the current coefficient can be linearly predicted from that of its neighbors at the same scale. His model was a nonlinear Markov model which can be stated as follows. The statistic T is a weighted average of $\left\{|N\_X_i|^2\right\}$, and X is Gaussian conditioned on T. We consider the equal weights $W_i = 1/8$. Here T is an unbiased estimate of the variance of X

$$T = f(N\_X) = \sum_i W_i \left|N\_X_i\right|^2, \tag{1}$$

where N_X is a neighborhood of X (excluding X).

In this paper, the denoising of an image corrupted by additive independent white Gaussian noise with variance σ_n^2 will be considered. Let w_2 represents the neighborhood complex wavelet coefficients of w_1 (w_2 is computed in terms of Eq. (1)). We formulate the problem in the complex wavelet domain as $y_1 = w_1 + n_1$ and $y_2 = w_2 + n_2$ to take into account the statistical dependencies between a coefficient and its neighborhood. y_1, y_2 are noisy observations of w_1, w_2, respectively; and n_1, n_2 are noise sample. We can write $\mathbf{y} = \mathbf{w} + \mathbf{n}$, with $\mathbf{w} = (w_1, w_2)$, $\mathbf{y} = (y_1, y_2)$, $\mathbf{n} = (n_1, n_2)$.

The standard MAP estimator for $\mathbf{w}$ giving the corrupted observation $\mathbf{y}$ is

$$\hat{\mathbf{w}}(\mathbf{y}) = \arg\max_{\mathbf{w}} p_{\mathbf{w}|\mathbf{y}}(\mathbf{w}|\mathbf{y}). \tag{2}$$

According to the property of conditional probability, this equation can be written as

$$\hat{\mathbf{w}}(\mathbf{y}) = \arg\max_{\mathbf{w}} [p_{\mathbf{n}}(\mathbf{y} - \mathbf{w}) \cdot p_{\mathbf{w}}(\mathbf{w})]. \tag{3}$$

We assume the noise is i.i.d. Gaussian, and the noise pdf is

$$p_{\mathbf{n}}(\mathbf{n}) = \frac{1}{2\pi\sigma_n^2} \cdot \exp\left(-\frac{n_1^2 + n_2^2}{2\sigma_n^2}\right). \tag{4}$$

We use the non-Gaussian bivariate model proposed by Sendur[4] to represent the dependency between a coefficient and its neighborhood, i.e.

$$p_{\mathbf{w}}(\mathbf{w}) = \frac{3}{2\pi\sigma_{\mathrm{s}}^2} \cdot \exp\left(-\frac{\sqrt{3}}{\sigma_{\mathrm{s}}}\sqrt{w_1^2 + w_2^2}\right), \tag{5}$$

where σ_{s} represents the standard deviation of the noisy-free wavelet coefficients. Using Eqs. (3)—(5), after some manipulations, the maximum a posterior (MAP) estimator of w_1 is derived to be

$$\hat{w}_1 = \frac{\left(\sqrt{y_1^2 + y_2^2} - \frac{\sqrt{3}\sigma_n^2}{\sigma_{\mathrm{s}}}\right)_+}{\sqrt{y_1^2 + y_2^2}} \cdot y_1. \tag{6}$$

In fact, this is a bivariate (coefficient and its neighborhood) shrink with threshold $T_{\mathrm{bishrink}} = \sqrt{3}\sigma_n^2/\sigma_{\mathrm{s}}$. And in Ref. [9], Wang used bivariate (coefficient and its parent) shrink in dual-tree complex wavelet domain with threshold $T_{\mathrm{bishrink}} = \sigma_n^2/\sigma_{\mathrm{s}}$ for noise reduction.

A typical image usually consists of smooth regions separated by a few singularities. This results in a $1/f$-type spectral behavior, which leads to the variances of the magnitudes of the complex wavelet coefficients tending to decay exponentially across scales. Similar to Chen's method[5], we investigate this property in dual-tree complex wavelet domain. The statistics of the inter-scale dependency of the complex wavelet coefficients is illustrated in Fig. 1, where $\bar{\sigma}_{\mathrm{s}}$ is the average standard deviation of six subbands of complex wavelet coefficients of the DT-CWT. For a number of images it is found that $f(s) = C\beta^s$, $\beta = 2$ shows the best match in the least-squares sense.

In addition, the distribution of standard deviations of the true and noisy image coefficients across scale is shown in Fig. 2. It can be seen that the difference between the plots of noise-free and noisy $\bar{\sigma}_{\mathrm{s}}$ is significant only in the lower scales, which indicates that the higher the scale value s, the higher the signal component on average. Therefore, for the thresholding algorithm of Eq. (6), we give lower weighs to the higher scales and higher

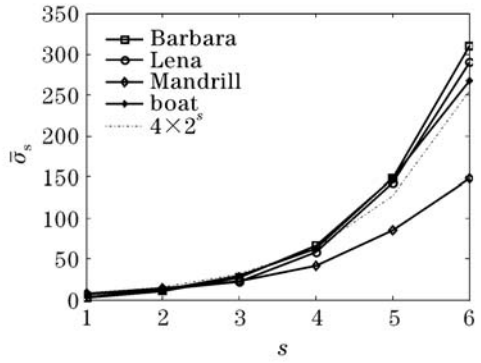

Fig. 1. Average standard deviations $\overline{\sigma_{\mathrm{s}}}$ versus scale of complex wavelet coefficients s for different images.

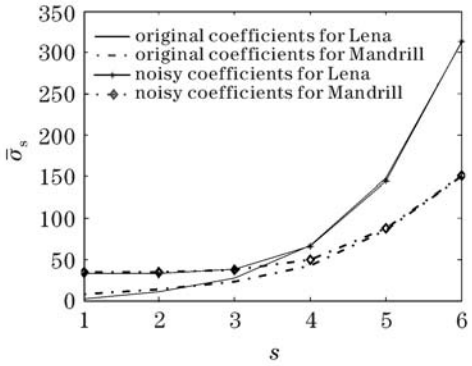

Fig. 2. Average standard deviations of original and noisy images versus scale of complex wavelet coefficients s.

weights to the lower scales. According to the average signal level distribution across scales and the average standard deviation of original and noisy images, we propose a new adaptive threshold as

$$T_{\text{new}} = \frac{1}{\sum_s 2^{-s}} 2^{-(\sigma_{s,j}^2/\sigma_n)\cdot s} \cdot T_{\text{bishrink}}, \tag{7}$$

where $\sigma_{s,j}$ is the standard deviation of the noise-free complex coefficients in subband j for scale s, and σ_n is the standard deviation of the noise, which is estimated from the finest scale wavelet coefficients[1]

$$\hat{\sigma}_n^2 = \text{median}(|y_i|)/0.6475,$$
$$y_i \in \text{subband of } 45^\circ \text{ in the finest scale.} \tag{8}$$

We can see that, of the new thresholding expression in our new bivariate model, the weights before T_{bishrink}, $\frac{1}{\sum_s 2^{-s}} 2^{-(\sigma_{s,j}^2/\sigma_n)\cdot s}$ exploits the inter-scale dependency.

Then, we can summarize our new denoising algorithm as follows. 1) Perform an L-level 2D DT-CWT for a noisy image I to get $6L+2$ noisy complex wavelet coefficients subbands (there are two low-frequency subbands and 6 high-frequency subbands for each scale s), and estimate the noise variance using Eq. (8). 2) Using Eq. (1) to calculate the neighborhood coefficient of the coefficients in each high-frequency subband for each scale s. 3) At each high-frequency subband for scale s, calculate the noise-free signal variance σ_s by an approximate maximum likelihood (ML) estimator: $\hat{\sigma}_s = \sqrt{\max\left(0, \frac{1}{M}\cdot\sum_{i\in N(k)} \text{real}(y[i])^2 - \sigma_n^2\right)}$, where $N(k)$ is the neighborhood of the coefficients considered and M is the number of the coefficients in $N(k)$. 4) Calculate the new adaptive threshold T_{new} using Eq. (7) in each high-frequency subband, and estimate the denoised coefficients in terms of Eq. (6), in which the threshold term $T_{\text{bishrink}} = \sqrt{3}\sigma_n^2/\sigma_s$ is replaced by T_{new}. 5) Perform the inverse DT-CWT to get the reconstructed denoised image $\hat{I}$.

In the experiments, we used three 512×512 standard test grayscale images, namely, Lena, Boat, and Barbara. The proposed algorithm was tested using different noise levels $\sigma_n = 10$, 20, and 30 and compared with adaptive threshold+DWT[5], HMT[3] , Wang's CWT denoising[9] and BiShrink+CWT[4]. Performance analysis is done using the peak signal-to-noise ratio (PSNR) of $10\lg(255^2/\text{mean squared error})$ measure. The quantitative results are listed in Table 1. From it we can see that our method outperforms the other related algorithms, in which Wang's CWT denoising and Bishrink+CWT are also based on the dual-tree complex wavelet. Figure 3 provides a visual comparison of denoised Lena image with different methods. It is clear that, in comparison with other methods, our algorithm got a better compromise in noise reduction and preserving significant image edges and details. In the other hand, our method can efficiently remove the "ringing" of the edge region which is presented in the wavelet-based methods.

The experimental results show that the proposed algorithm achieves much better performance than some other

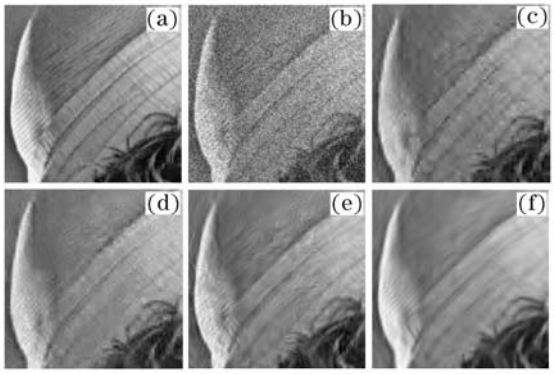

Fig. 3. Comparisons of denoising results on Lena (local 128×128, $\sigma_n = 20$). (a) Original image; (b) noisy image; (c) adaptive threshold+DWT; (d) HMT; (e) Wang's CWT; (f) our method.

Table 1. PSNR [dB] Results for Several Methods

Test Images	σ	Noisy	Adaptive Threshold+DWT[5]	HMT[3]	Wang's CWT[9]	BiShrink+CWT[4]	Our Method
Lena	10	28.165	33.554	33.84	34.88	34.77	35.389
	20	22.136	30.191	30.39	31.70	31.71	32.359
	30	18.627	28.212	28.35	29.56	29.85	30.502
Barbara	10	28.159	30.676	31.36	33.26	31.30	33.677
	20	22.144	26.981	27.80	29.75	27.78	30.007
	30	18.641	24.652	25.11	27.64	25.62	28.041
Boat	10	28.155	31.961	32.28	32.77	32.85	33.19
	20	22.148	28.557	28.84	29.78	29.81	29.911
	30	18.638	26.438	26.83	27.72	27.99	28.163

denoising techniques, both in PSNR and visual effect. We obtained these results by simply exploit the dependencies between coefficients and their 8-neighborhood of the intra-scale (i.e. the local context correlations), and the dependencies across the scales. It is expected that the results can be further improved if the other dependencies between coefficients and its other neighbors, and the more suitable statistical model are exploited.

This work was supported by the National Natural Science Foundation of China under Grant No. 60573027. F. Yan's e-mail address is xialang3@163.com.

References

1. D. L. Donoho and I. M. Johnstone, J. Am. Stat. Assoc. **90,** 1200 (1995).
2. E. P. Simoncelli, in *Bayesian Inference in Wavelet Based Models* P. Muller and B. Vidakovic, (eds.) (Springer-Verlag, New York, 1999) Chap.18, pp.291—308.
3. M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, IEEE Trans. Signal Proc. **46,** 886 (1998).
4. L. Sendur and I. W. Selesnick, IEEE Trans. Signal Proc. **50,** 2744 (2002).
5. Y. Chen and C. Han, Electron. Lett. **41,** 1586 (2005).
6. J. Liu and P. Moulin, IEEE Trans. Image Proc. **10,** 1647 (2001).
7. N. Kingsbury, Appl. Comput. Harmon. Anal. **10,** 234 (2001).
8. I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, IEEE Signal Processing Magazine **22,** 123 (2005).
9. H. Wang, L. Cheng, and Y. Wu, J. Computer-Aided Design & Computer Graphics (in Chinese) **17,** 1911 (2005).
10. J. K. Romberg, H. Choi, and R. G. Baraniuk, in *Proceedings of IEEE Int. Conf. Image Processing* **1,** 614 (2001).
11. E. P. Simoncelli, Proc. SPIE **3813,** 188 (1999).