AN IMAGE DENOISING ALGORITHM WITH AN ADAPTIVE WINDOW

Zhou Dengwen

Department of Computer Science and Technology
North China Electric Power University, Beijing, China
Email: zdw@ncepu.edu.cn

ABSTRACT

Mihcak *et al* proposed a low complexity but powerful image denoising algorithm *LAWML* based on the decimated wavelet transform(DWT). The shortcoming of *LAWML* is to determine the global optimal neighboring window size by experimenting. We improve on *LAWML* using Stein's unbiased risk estimate(SURE). Our method can automatically estimate an optimal neighboring window for every wavelet subband. Its denoising performance also surpasses *LAWML* because the subband adaptive window is superior to the global window. Furthermore, our method on the DWT is extended to on the dual-tree complex wavelet transform (DT-CWT). Experimental results indicate that our method (DT-CWT) delivers the comparable or better performance than some of the already published state-of-the-art denoising algorithms.

Index Terms— Image denoising, wavelet transforms, dual-tree, adaptive method

1. INTRODUCTION

It is a classical signal estimation problem to remove additive white Gaussian noise(AWGN) in images. Wavelet-based methods significantly improve the estimation results since wavelet decompositions simplify the statistical properties of images. A number of researchers have found that wavelet coefficients of images have similar marginal distributions and interscale dependencies. Therefore, it is possible to model the prior distributions of noiseless wavelet coefficients. Recently, many models of noiseless wavelet coefficients have been proposed. The most famous ones include generalized Gaussian distribution(GGD) model[1], Markov random field (MRF)[2] and hidden Markov tree(HMT)[3] models, Gaussian scale mixture (GSM) model[4], bivariate distribution model[5] etc. The estimation results depends heavily on the accurateness of the model. The higher is the accurateness of the model, the better is the effect. In general, the more accurate model suffers the greater computational burden. Mihcak *et al* who were motivated by EQ image compression algorithm proposed a simple doubly stochastic process model and applied it to image denoising[6]. They gave a low complexity but powerful image denoising algorithm *LAWML* in terms of the doubly stochastic process model. The shortcoming of *LAWML* is that it needs to determine a global optimal neighboring window size by experimenting. This process is inefficient. We improve on *LAWML* using Stein's unbiased risk estimate(SURE)[7]. Our method can automatically compute an optimal neighboring window for every wavelet subband. This adaptive method not only avoids determining the window size by experimenting, but also outperforms *LAWML* because the subband adaptive window is superior to the global window. Then our proposed adaptive method is extended to the dual-tree complex wavelet transform(DT-CWT) from the decimated wavelet transform(DWT). The experimental results indicate that our method on the DT-CWT yields the comparable or better performance compared with some of the recent state of art image denoising methods because the DT-CWT has the advantages of shift invariance and direction selectivity over the DWT.

2. *LAWML* APPROACH

We simply introduce *LAWML* algorithm[6] in order to facilitate the discussion of the proposed method. Suppose an image $\{X_{ij}\}$ is contaminated with Gaussian random noise with zero mean and variance σ_n^2:

$$Y_{ij} = X_{ij} + \varepsilon_{ij} \quad i = 1, \cdots, I, j = 1, \cdots, J \tag{1}$$

where $\{\varepsilon_{ij}\}$ is independent and identically Gaussian(normal) distributed (iid) $N(0, \sigma_n^2)$. We apply an orthonormal wavelet transform to Equation (1). Let $\{\theta_{ij}\}$, $\{w_{ij}\}$ and $\{n_{ij}\}$ represent the orthonormal wavelet coefficients of the noiseless image, noisy image and noise, respectively. $\{n_{ij}\}$ is still *iid* $N(0, \sigma_n^2)$ due to orthonormality of the wavelet transform. Given the variance field $\{\sigma_\theta^2(i,j)\}$ of the noiseless coefficients $\{\theta(i,j)\}$, the minimum mean-squared error(MMSE) estimator $\hat{\theta}(i,j)$ of $\theta(i,j)$ is given:

$$\hat{\theta}(i,j) = \frac{\sigma_\theta^2(i,j)}{\sigma_\theta^2(i,j) + \sigma_n^2} w(i,j) \tag{2}$$

In fact, $\{\sigma_\theta^2(i,j)\}$ is unknown and its estimate is a crux. Mihcak *et al* estimated $\sigma_\theta^2(i,j)$ using a local neighborhood $\mathcal{N}(i,j)$ which is a square window and its center is at *w*(*i*,*j*).

$$\begin{aligned}\hat{\sigma}_\theta^2(i,j) &= \arg\max_{\sigma_\theta^2 \ge 0} \prod_{k,l\in\mathcal{N}(i,j)} P(w(k,l)\,|\,\sigma_\theta^2) \\ &= \max\left(0, \frac{1}{M}\sum_{k,l\in\mathcal{N}(i,j)} w^2(k,l) - \sigma_n^2\right) \qquad (3)\end{aligned}$$

where $P(\cdot\,|\,\sigma_\theta^2)$ is the Gaussian distribution with zero mean and variance $\sigma_\theta^2+\sigma_n^2$, and M is the number of coefficients in $\mathcal{N}(k)$. Then, the estimate $\hat{\theta}(i,j)$ of $\theta(i,j)$ is obtained using $\hat{\sigma}_\theta^2(i,j)$ instead of $\sigma_\theta^2(i,j)$ in Equation (2). Mihcak *et al* referred the MMSE-like estimate as *LAWML*. Practically, σ_n is also unknown in Equation (2). A good estimator for σ_n is the median of absolute deviation(MAD) using the highest level wavelet coefficients[8].

$$\hat{\sigma}_n = \frac{\text{median}(|\,w(i,j)\,|)}{0.6745} \quad (w(i,j)\in \text{subband } HH) \qquad (4)$$

3. PROPOSED ALGORITHM

The contribution of our method on the DWT is to determine the optimal neighboring window size of every wavelet subband for *LAWML*. The procedure of the derivation of the optimal neighboring window size for every subband is the following. For ease of notation, let us arrange the N_s noisy wavelet coefficients from subband *s*, $\mathbf{w}_s = \{w_{ij} : i,j \in \text{indices corresponding to subband } s\}$ into the one-dimensional vector $\mathbf{w}_s = \{w_n : n = 1,\cdots,N_s\}$. Let us next combine the N_s unknown noiseless coefficients $\{\theta_{ij} : i,j \in \text{indicies corresponding to subband } s\}$ from subband *s* into the corresponding one-dimensional vector $\boldsymbol{\theta}_s = \{\theta_n : n = 1,\cdots,N_s\}$.

Stein showed[7] that , for almost any fixed estimator $\hat{\boldsymbol{\theta}}_s$ based on the data $\mathbf{w}_s$, the expected loss(i.e. risk) $E\{\|\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s\|_2^2\}$ can be estimated unbiasedly. In general case, the noise standard deviation $\sigma_n = 1$, we have that

$$E\left\{\left\|\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s\right\|_2^2\right\} = N_s + E\left\{\|\mathbf{g}(\mathbf{w}_s)\|_2^2 + 2\nabla\cdot\mathbf{g}(\mathbf{w}_s)\right\} \qquad (5)$$

where $\mathbf{g}(\mathbf{w}_s) = \{g_n\}_{n=1}^{N_s} = \hat{\boldsymbol{\theta}}_s - \mathbf{w}_s$, $\nabla\cdot\mathbf{g} \equiv \sum_n \frac{\partial g_n}{\partial w_n}$.

We have from Equation (2) that for the *n*th wavelet coefficient w_n:

$$g_k(w_k) = \hat{\theta}_k - w_k = \begin{cases} -\dfrac{w_k}{\sigma_\theta^2+1} & (\sigma_\theta > 0) \\ -w_k & (\text{otherwise}) \end{cases} \qquad (6)$$

with

$$\frac{\partial g_k}{\partial w_k} = \begin{cases} \dfrac{1}{\sigma_\theta^2+1}\left(\dfrac{2w_k^2}{M(\sigma_\theta^2+1)} - 1\right) & (\sigma_\theta > 0) \\ -1 & (\text{otherwise}) \end{cases} \qquad (7)$$

$$\|g_k(w_k)\|_2^2 = \begin{cases} \dfrac{w_k^2}{(\sigma_\theta^2+1)^2} & (\sigma_\theta > 0) \\ w_k^2 & (\text{otherwise}) \end{cases} \qquad (8)$$

The quantity

$$SURE(\mathbf{w}_s, L) = N_s + \sum_n \|g_n(w_n)\|_2^2 + 2\sum_n \frac{\partial g_n}{w_n} \qquad (9)$$

is an unbiased estimate of the risk on subband *s* where *L* is the neighborhood window size(*L* is an odd number and greater than 1, for example, 3, 5, 7 etc): $E\{\|\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s\|_2^2\} = E\{SURE(\mathbf{w}_s, L)\}$. Then we choose the neighboring window size L^s on subband *s* which minimize $SURE(\mathbf{w}_s, L)$. Accordingly,

$$L^s = \arg\min_L SURE(\mathbf{w}_s, L) \qquad (10)$$

The above L^s are derived assuming $\sigma_n = 1$. For data with non-unit variance, the coefficients are first standardized by the estimate $\hat{\sigma}_n$ obtained using Equation (4).

4. RESULTS ON THE DECIMATED WAVELET TRANSFORM

We have compared the results of our method with that of *LAWML*. The decimated wavelet transform(DWT) is used with Daubechies' least asymmetric compactly-supported wavelet with eight vanishing moments with four scales. The 512×512 standard test images *Goldhill*, *Barbara* and *Mandrill* are chosen(see Fig. 1). They are contaminated with Gaussian random noise with standard deviations 10, 20, 30, 50, 75 and 100. The optimal neighboring window sizes of *LAWML* are determined by experimenting. The neighboring window sizes of the proposed method are calculated with Equation (10) and the largest window size which is searched for every subband is limited to 25×25. The noise standard deviation is estimated with Equation (4). Our results are measured by the peak signal-to-noise ratio(*PSNR*) in decibels(dB) defined as

$$PSNR = 10*\log_{10}\frac{255^2}{MSE}(dB) \qquad (11)$$

where $MSE = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}(X(i,j)-\hat{X}(i,j))^2$, here , X is the original image, $\hat{X}$ is the estimate of *X*, and *I*J* is the number of pixels. The denoised image is closer to the original when *PSNR* is higher. Table 1 shows the *PSNR* performance of the two denoising methods and the optimal window sizes of *LAWML*. Obviously, the *PSNR*s which the proposed ada-

(a) (b) (c)

Fig.1. The original test images with 512×512 pixels. (a) *Goldhill*; (b) *Barbara*; (c) *Mandrill.*

	LAWML		*Proposed*
Goldhill			
$\sigma = 10$	32.54	5×5	**32.57**
$\sigma = 20$	29.18	7×7	**29.32**
$\sigma = 30$	27.41	9×9	**27.63**
$\sigma = 50$	25.38	13×13	**25.76**
$\sigma = 75$	23.92	17×17	**24.35**
$\sigma = 100$	22.98	25×25	**23.39**
Barbara			
$\sigma = 10$	32.85	5×5	**32.88**
$\sigma = 20$	28.81	7×7	**28.93**
$\sigma = 30$	26.62	7×7	**26.82**
$\sigma = 50$	24.12	9×9	**24.46**
$\sigma = 75$	22.41	13×13	**22.84**
$\sigma = 100$	21.32	15×15	**21.86**
Mandrill			
$\sigma = 10$	**29.50**	3×3	29.43
$\sigma = 20$	**26.00**	5×5	25.97
$\sigma = 30$	24.05	9×9	**24.08**
$\sigma = 50$	21.92	9×9	**22.00**
$\sigma = 75$	20.53	13×13	**20.68**
$\sigma = 100$	19.76	17×17	**19.97**

Table 1. Denoising results(*PSNR*) on the DWT for *Goldhill*, *Barbara* and *Mandrill*.

ptive method yields are higher for all noise levels except for *Mandrill* with noise deviations 10 and 20. The larger neighboring windows are needed when noise levels are increased.

5. EXTENSION TO COMPLEX WAVELET COEFFICIENTS

The dual-tree complex wavelet transform (DT-CWT) [9] is a relatively recent enhancement to the decimated wavelet transform (DWT). It is a slightly redundant transform with a redundancy factor of only 2^d for *d*-dimensional signals and expands an image in terms of a complex wavelet with complementary real and imaginary parts. Its basis functions have directional selectivity property at $\pm 15^\circ$, $\pm 45^\circ$, and $\pm 75^\circ$, which the regular critically sampled transform does not have. The key advantages of the DT-CWT over the DWT are its shift invariance and directional selectivity. It means that DT-CWT-based algorithms will automatically be almost shift invariant, thus reducing many of the artifacts of the critically sampled DWT. The previous adaptive method can be extended to the DT-CWT. The procedure can be described as follows. For the real parts of every subband, we first compute the optimal neighboring window size L^s with Equation (10). In terms of L^s, the real and imaginary parts' variance fields of every subband can be calculated with Equation (3), then the real and imaginary parts of every subband are shrinked separately according to Equation (2). The experimental results show that the proposed method surpasses some of the already published best denoising methods.

	ProbShrink (UWT)	*BiShrink* (DT-CWT)	*Proposed* (DT-CWT)
Goldhill			
$\sigma = 10$	32.50	32.80	**33.04**
$\sigma = 20$	29.65	29.92	**30.03**
$\sigma = 30$	27.99	28.36	**28.46**
$\sigma = 50$	26.15	26.54	**26.62**
$\sigma = 75$	24.79	25.18	**25.27**
$\sigma = 100$	23.85	25.24	**24.33**
Barbara			
$\sigma = 10$	33.27	33.30	**33.68**
$\sigma = 20$	29.32	29.66	**30.03**
$\sigma = 30$	27.14	27.61	**27.96**
$\sigma = 50$	24.44	25.21	**25.58**
$\sigma = 75$	22.78	23.53	**23.83**
$\sigma = 100$	21.94	22.48	**22.73**
Mandrill			
$\sigma = 10$	28.77	29.29	**29.65**
$\sigma = 20$	25.65	26.14	**26.40**
$\sigma = 30$	23.87	24.30	**24.56**
$\sigma = 50$	21.81	22.24	**22.56**
$\sigma = 75$	20.62	20.95	**21.25**
$\sigma = 100$	19.96	20.21	**20.50**

Table 2. Denoising results(*PSNR*) with the UWT or DT-CWT for *Goldhill*, *Barbara* and *Mandrill* for the three denoising methods

We have compared our results with the current state of art schemes *ProbShrink*[10] and *BiShrink*[5]. *ProbShrink* uses the nondecimated wavelet transform(UWT) which employs Daubechies symmlet wavelet with eight vanishing moments with four scales. *BiShrink* and the proposed method use the DT-CWT with six scales. *ProbShrink* and *BiShrink* Matlab implementations are available at the web[11][12]. We still choose the same images with the same noise levels as the previous section. Table 2 shows the *PSNR*s of the three denoising methods.

It can be observed from Table 2 that the proposed method(DT-CWT) consistently yields the highest *PSNR*s for the three test images in all noise levels. The proposed

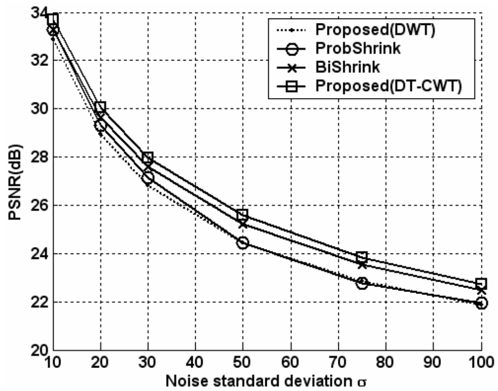

Fig. 2. *PSNR* curves of the proposed method (on the DWT and DT-CWT) and the other two methods for *Barbara.*

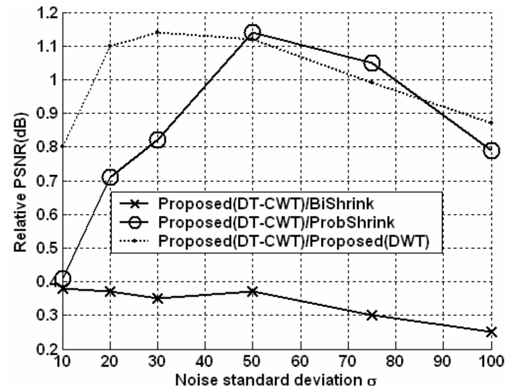

Fig. 3. *PSNR* gain curves of the proposed method(DT-CWT) compared with the other three methods for *Barbara.*

method on the DT-CWT is also significantly superior to that on the DWT. The largest *PSNR* gain of the proposed method on the DT-CWT is 0.94 for *Goldhill*, 1.14 for *Barbara* and 0.57 for *Mandrill* greater than on the DWT. The *PSNR* curves for *Barbara* are shown in Fig. 2 and the *PSNR* gain curves for *Barbara* are shown in Fig. 3. It is obvious that the DWT-based method is inferior to the other methods, but its *PSNR*s are comparable with *ProbShrink* when $\sigma > 45$. We have also compared our results with that of *BLS-GSM* method[4]. The *BLS-GSM* with the 8-orientation steerable pyramid yields slightly better *PSNR* performance than the proposed method on the DT-CWT.

6. CONCLUSION

We have proposed a new adaptive image denoising algorithm. It is an improvement on *LAWML* proposed by Mihcak *et al* on the DWT. Our method can automatically determine the optimal neighboring window size for every wavelet subband while *LAWML* can only determine the global optimal neighboring window size by experimenting. The denoising performance of our method also outperforms *LAWML* because the subband adaptive neighboring window is superior to the global window. Furthermore, our method is extended to the DT-CWT from the DWT. Experimental results indicate that the proposed adaptive denoising method (DT-CWT) yields the comparable or better performance than some of the already published best denoising algorithms.

7. REFERENCES

[1] S. G. Chang, B. Yu, and M. Vetterli, “Adaptive wavelet thresholding for image denoising and compression,” *IEEE Trans. Image Process.*, vol. 9, pp. 1532－1546, Sep. 2000.

[2] M. Malfait and D. Roose, “Wavelet-based image denoising using a Markov random field a priori model,” *IEEE Trans. Image Process.*, vol. 6, pp. 549－565, Apr. 1997.

[3] J. K. Romberg, H. Choi, and R. G. Baraniuk, “Bayesian tree-structured image modeling using wavelet-domain hidden Markov models,” *IEEE Trans. Image Process.*, vol. 10, pp. 1056 – 1068, Jul. 2001.

[4] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, “Image denoising using Gaussian scale mixtures in the wavelet domain,” *IEEE Trans. Image Process.*, vol. 12, pp. 1338 – 1351, Nov. 2003.

[5] L. Sendur and I. W. Selesnick, “Bivariate shrinkage with local variance estimation,” *IEEE Signal Process. Lett.*, vol. 9, pp. 438–441, Dec. 2002.

[6] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, “Low complexity image denoising based on statistical modeling of wavelet coefficients,” *IEEE Signal Process. Lett.*, vol. 6, pp. 300–303, Dec. 1999.

[7] C. Stein, “Estimation of the mean of a multivariate normal distribution,” *Ann. Statist.* vol. 9, pp. 1135–1151, 1981.

[8] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaptation via wavelet shrinkage,” *Biometrika*, vol. 81, pp. 425–455, 1994.

[9] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, “The dual-tree complex wavelet transform,” *IEEE Signal Processing Magazine*. vol. 22, pp. 123–151, Nov., 2005.

[10] A. Pizurica and W. Philips, “Estimating probability of presence of a signal of interest in multiresolution single- and multiband image denoising, ” *IEEE Trans. Image Process.*, vol. 15, pp. 654－665, Mar. 2006.

[11] *ProbShrink* Matlab code, located at the URL: http://telin.ugent.be/~sanja.

[12] DT-CWT and *BiShrink* Matlab code, located at the URL: http://taco.poly.edu/WaveletSoftware/.