Pattern Recognition Letters 29 (2008) 1694–1697

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Image denoising with an optimal threshold and neighbouring window

Zhou Dengwen *, Cheng Wengang

Department of Computer Science and Technology, North China Electric Power University, 2 Beinong Road, Changping District, Beijing 102206, China

ARTICLE INFO

Article history:
Received 7 June 2007
Received in revised form 25 April 2008
Available online 15 May 2008

Communicated by R. Davies

Keywords:
Image denoising
Adaptive
Dual tree
Wavelet transforms
Neighbourhood

ABSTRACT

NeighShrink is an efficient image denoising algorithm based on the decimated wavelet transform (DWT). Its disadvantage is to use a suboptimal universal threshold and identical neighbouring window size in all wavelet subbands. In this paper, an improved method is given, which can determine an optimal threshold and neighbouring window size for every subband by the Stein's unbiased risk estimate (SURE). Its denoising performance is considerably superior to NeighShrink and also outperforms SURE-LET, which is an up-to-date denoising algorithm based on the SURE. It is well known that increasing the redundancy of wavelet transforms can significantly improve the denoising performances. The proposed method is also extended to the redundant dual-tree complex wavelet transform (DT-CWT). Experiments demonstrate that the proposed method on the DT-CWT achieves better results than some of the best denoising algorithms published currently.

1. Introduction

During the last decade, a lot of new methods based on wavelet transforms have emerged for removing Gaussian random noise from images. The denoising process is known as wavelet shrinkage or thresholding. Both VisuShrink and SureShrink are the best-known methods of wavelet shrinkage proposed by Donoho and Johnstone (1994, 1995). For VisuShrink, the wavelet coefficients $\mathbf{w}$ of the noisy signal are obtained first. Then with the universal threshold $T = \sqrt{2\sigma^2 \log(N)}$ (σ is the noise level and N is the length of the noisy signal), the coefficients $w = \{w_i\}_{i=1,2,\ldots,N}$ are shrinked according to the soft-shrinkage rule $\eta_T^s(w_i) = \mathrm{sgn}(w_i) \cdot (|w_i| - T)_+$ and $\eta_T^s(w)$ is used to estimate the noiseless coefficients. Finally, the estimated noiseless signal is reconstructed from the estimated coefficients $\eta_T^s(w)$. VisuShrink is very simple, but its disadvantage is to yield overly smoothed images because the universal threshold T is too large. Just like VisuShrink, SureShrink also applies the soft-shrinkage rule, but it uses independently chosen thresholds for each subband through the minimization of the Stein's unbiased risk estimate (SURE) (Stein, 1981). SureShrink performs better than VisuShrink, producing more detailed images.

After the seminal work of Donoho and Johnstone, many alternative methods have come forth. Cai and Silverman proposed two different shrinkage methods NeighBlock and NeighCoeff for 1-D signals (Cai and Silverman, 2001). They threshold the wavelet coefficients in overlapping blocks rather than individually or term by term as VisuShrink or SureShrink. The basic motivation of block thresholding remains: a coefficient is more likely to contain signal if neighboring coefficients do also. Chen et al. applied NeighCoeff to image denoising and their method is called NeighShrink (Chen et al., 2005). NeighShrink outperforms VisuShrink and SureShrink.

For the other recent shrinkage methods, we cite the following ones:

- *BiShrink (*Sendur and Selesnick, 2002a*)*: This method uses a bivariate shrinkage function which models the statistical dependence between a wavelet coefficient and its parent. It needs to estimate the marginal variance of the coefficient in a local neighbourhood. The neighbouring window size is typically 7×7.
- *ProbShrink (*Pizurica and Philips, 2006a*)*: This method uses a probabilistic shrinkage function. Its core is estimating the probability that a given coefficient contains a significant noise-free component. Then the wavelet coefficient is multiplied with the probability.
- *SURE-LET* (Luisier et al., 2007a,b): This method directly parametrizes the denoising process as a sum of elementary nonlinear processes with unknown weights. It need not hypothesize a statistical model for the noiseless image while it minimizes an estimate of the mean squared error between the noiseless image and the denoised one by the SURE. Consequently, it computes the unknown weights by solving a linear system of equations.

It is well known that increasing the redundancy of wavelet transforms can significantly improve the denoising performances.

* Corresponding author. Present address: 8-4-201, Block 4, Longteng, Huilongguan Residential Area, Changping District, Beijing 102208, China. Tel.: +86 10 51963578; fax: +86 10 51963504.
E-mail addresses: zdw@ncepu.edu.cn (Z. Dengwen), wgcheng@ncepu.edu.cn (C. Wengang).

0167-8655/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2008.04.014

BiShrink, ProbShrink and SURE-LET methods have all been devised for both redundant and nonredundant wavelet transforms. In this paper, we improve NeighShrink using the SURE (Stein, 1981; Fodor and Kamath, 2003). The proposed method can estimate an optimal threshold and neighbouring window size for NeighShrink in every wavelet subband. It is also extended to the redundant dual-tree complex wavelet transform (DT-CWT) from the decimated wavelet transform (DWT). The experimental results demonstrate that the proposed method not only outperforms NeighShrink, but also delivers better results compared with three state-of-the-art methods: i.e. BiShrink, ProbShrink and SURE-LET.

2. Proposed adaptive algorithm

We give a brief introduction to NeighShrink algorithm before we discuss the proposed method. For each noisy wavelet coefficient w_{ij} to be shrinked, it incorporates a square neighbouring window B_{ij} centered at it. The neighbouring window size can be represented as $L \times L$, where L is a positive odd number. Fig. 1 gives a 3×3 neighbouring window centered at the wavelet coefficient to be shrinked.

Suppose $S_{ij}^2 = \sum_{k,l \in B_{ij}} w_{kl}^2$, the NeighShrink shrinkage formula can be represented as

$$\widehat{\theta}_{ij} = w_{ij}\beta_{ij} \quad (1)$$

where $\widehat{\theta}_{ij}$ is the estimator of the unknown noiseless coefficient, $\beta_{ij} = (1 - \lambda^2/S_{ij}^2)_+$ and λ is the universal threshold. $(g)_+$ is defined as $(g)_+ = \max(g, 0)$.

Different wavelet coefficient subbands are shrinked independently, but the threshold λ and neighbouring window size L keep unchanged in all subbands. When S_{ij}^2 summation has pixel indexes out of the wavelet subband range, the corresponding terms in the summation are omitted. The shortcoming of this method is that using the same universal threshold λ and neighbouring window size L in all subbands is suboptimal.

The optimal λ and L of every subband should be data-driven and minimize the mean squared error (MSE) or risk of the corresponding subband. Fortunately, Stein (1981) has stated that the MSE can be estimated unbiasedly from the observed data. We will improve NeighShrink by determining an optimal threshold and neighbouring window size for every wavelet subband using the Stein's unbiased risk estimate (SURE). For ease of notation, we arrange the N_s noisy wavelet coefficients from subband s, $w_s = \{w_{ij}: i,j \in$ indices corresponding to subband$s\}$, into the 1-D vector $w_s = \{w_n: n = 1, \ldots, N_s\}$. Similarly, we combine the N_s unknown noiseless coefficients $\{\theta_{ij}: i,j \in$ indices corresponding to subband$s\}$ from subband s into the corresponding 1-D vector $\theta_s = \{\theta_n: n = 1, \ldots, N_s\}$. Stein shows that, for almost any fixed estimator $\widehat{\theta}_s$ based on the data $\mathbf{w}_s$, the expected loss (i.e. risk) $E\left\{\left\|\widehat{\theta}_s - \theta_s\right\|_2^2\right\}$ can be estimated unbiasedly. Usually, the noise standard deviation σ is set at 1, and then we have that

$$E\left\{\left\|\widehat{\theta}_s - \theta_s\right\|_2^2\right\} = N_s + E\left\{\|g(\mathbf{w}_s)\|_2^2 + 2\nabla \cdot g(\mathbf{w}_s)\right\} \quad (2)$$

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

According to Eq. (1), we have for the nth wavelet coefficient w_n:

$$g_n(w_n) = \widehat{\theta}_n - w_n = \begin{cases} -\frac{\lambda^2}{S_n^2} w_n & (\lambda < S_n) \\ -w_n & (\text{otherwise}) \end{cases} \quad (3)$$

with

$$\frac{\partial g_n}{\partial w_n} = \begin{cases} -\lambda^2 \frac{S_n^2 - 2w_n^2}{S_n^4} & (\lambda < S_n) \\ -1 & (\text{otherwise}) \end{cases} \quad (4)$$

$$\|g_n(w_n)\|_2^2 = \begin{cases} \frac{\lambda^4}{S_n^4} w_n^2 & (\lambda < S_n) \\ w_n^2 & (\text{otherwise}) \end{cases} \quad (5)$$

The quantity

$$\mathrm{SURE}(w_s, \lambda, L) = N_s + \sum_n \|g_n(w_n)\|_2^2 + 2\sum_n \frac{\partial g_n}{w_n} \quad (6)$$

is an unbiased estimate of the risk on subband s where L is the neighbourhood window size (L is an odd number and greater than 1, for example, 3, 5, 7, 9, etc.): $E\left\{\left\|\widehat{\theta}_s - \theta_s\right\|_2^2\right\} = E\{\mathrm{SURE}(w_s, \lambda, L)\}$. Then we choose the threshold λ^s and neighbouring window size L^s on subband s which minimize $\mathrm{SURE}(w_s, \lambda, L)$. Accordingly,

$$(\lambda^s, L^s) = \arg\min_{\lambda, L} \mathrm{SURE}(w_s, \lambda, L) \quad (7)$$

where λ^s and L^s are derived assuming the noise level $\sigma = 1$. For data with nonunit variance, the coefficients are standardized by an appropriate estimator $\widehat{\sigma}$ before calculating the λ^s and L^s with Eq. (7). A good estimator for σ is the median of absolute deviation (MAD) using the highest level wavelet coefficients (Donoho and Johnstone, 1994).

$$\widehat{\sigma} = \frac{\mathrm{median}(|w_s|)}{0.6745} (w_s \in \text{subband } HH) \quad (8)$$

3. Results on the decimated wavelet transform

To verify the validity of the proposed method, we compared its results with those of NeighShrink. In addition, we also compared it with SURE-LET which is the latest method based on the SURE. The DWT was used with Daubechies' least asymmetric compactly-supported wavelet with eight vanishing moments with four scales. SURE-LET Matlab package is available on the web (Luisier et al., 2007b). The 512×512 standard test images, Lena, Barbara and Mandrill, were chosen as the experimental dataset (Fig. 2). They were contaminated with Gaussian random noise with standard deviations 10, 20, 30, 50, 75 and 100. In all wavelet subbands, NeighShrink used the universal threshold $\lambda = \sigma\sqrt{2\log(512)} \approx 3.53\sigma$ and the default neighbouring window size 3×3 which is

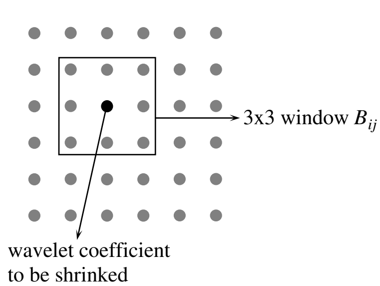

Fig. 1. An illustration of the neighbouring window centered at the wavelet coefficient to be shrinked.

Fig. 2. The original test images with 512×512 pixels: (a) Lena; (b) Barbara; and (c) Mandrill.

recommended by NeighShrink. The threshold and neighbouring window size of the proposed method in every subband were calculated with Eq. (7). We assumed that the noise variances were known in order to focus on the denoising techniques themselves. Generally, noise variances are unknown, but they can be estimated by Eq. (8). We measured the experimental results by the peak signal-to-noise ratio (PSNR) in decibels (dB), which is defined as

$$\mathrm{PSNR} = 10 * \log_{10}\frac{255^2}{\mathrm{MSE}} \ (\mathrm{dB}) \tag{9}$$

where $\mathrm{MSE} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}(X(i,j) - \widehat{X}(i,j))^2$, X is the original image, $\widehat{X}$ is the estimator of X, and $I * J$ is the number of pixels. The denoised image is closer to the original one when PSNR is higher. Table 1 shows the PSNR performance of the three denoising methods.

As expected, the PSNRs that the proposed adaptive method produces are substantially higher than those that NeighShrink does for all noise levels. NeighShrink is also not robust. Its results become inferior as noise levels increase. For Lena image, the largest PSNR gain of the proposed method is approximately 2.8 dB. In most cases, the proposed method performs slightly better than SURE-LET except Barbara image. For Barbara, our method is superior to SURE-LET in all noise levels.

4. Extension to complex wavelet coefficients

The DT-CWT (Selesnick et al., 2005) is a relatively recent enhancement to the 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 ±15°, ±45°, and ±75°, 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 the DT-CWT-based algorithms will automatically be almost shift invariant, thus reducing many of the artifacts of the critically sampled DWT. The aforementioned 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 threshold λ^s and neighbouring window size L^s with Eq. (7). Then the real and imaginary parts of every subband are shrinked separately using λ^s and L^s according to Eq. (1).

The proposed method on the DT-CWT was compared with three state-of-the-art schemes: i.e. BiShrink (Sendur and Selesnick, 2002a), ProbShrink (Pizurica and Philips, 2006a) and SURE-LET (Blu and Luisier, 2007). ProbShrink and SURE-LET used the nondecimated wavelet transform (UWT) with four scales, but the former employed Daubechies symmlet wavelet with eight vanishing moments and the latter did Haar wavelet considering that Blu et al. obtain the best results with Haar wavelet. BiShrink and the proposed method used the DT-CWT with six scales. We still assumed that the noise variances were known just as the last section. ProbShrink and BiShrink Matlab implementations are available on the web (Pizurica and Philips, 2006b; Sendur and Selesnick, 2002b). We also still chose the same images with the same noise levels as the previous section. Table 2 illustrates the PSNRs of the four denoising methods.

It is obvious that the proposed method almost consistently produces the highest PSNRs for the three test images in all noise levels. Our method on the DT-CWT is also significantly superior to that on the DWT. The largest PSNR gain of our method on the DT-CWT is 0.97 dB for Lena, 1.03 dB for Barbara and 0.49 dB for Mandrill greater than that on the DWT (compare Table 2 with Table 1).

Table 1
Denoising results (PSNRs) for Lena, Barbara and Mandrill

σ	NeighShrink	SURE-LET	Proposed
Lena			
10	34.51	34.56	**34.72**
20	31.05	31.37	**31.53**
30	28.90	29.56	**29.70**
50	26.05	27.37	**27.43**
75	23.67	**25.76**	25.62
100	21.87	**24.66**	24.43
Barbara			
10	32.83	32.16	**33.02**
20	28.75	27.96	**29.09**
30	26.48	25.82	**27.01**
50	23.76	23.72	**24.63**
75	21.76	22.54	**22.99**
100	20.34	21.81	**21.94**
Mandrill			
10	29.96	30.13	**30.30**
20	25.80	25.92	**26.20**
30	23.70	23.88	**24.16**
50	21.44	21.86	**22.03**
75	19.92	**20.71**	20.69
100	18.87	**20.09**	19.95

Table 2
Denoising results (PSNRs) with the UWT or DT-CWT for Lena, Barbara and Mandrill for the four denoising methods

σ	ProbShrink (UWT)	BiShrink (DT-CWT)	SURE-LET (UWT)	Proposed (DT-CWT)
Lena				
10	35.06	35.18	35.09	**35.42**
20	31.92	32.24	32.06	**32.39**
30	30.04	30.49	30.31	**30.59**
50	27.68	28.23	28.18	**28.32**
75	25.93	26.42	26.54	**26.59**
100	24.71	25.18	**25.41**	25.39
Barbara				
10	33.45	33.51	32.61	**33.82**
20	29.50	29.86	28.42	**30.12**
30	27.18	27.77	26.21	**28.01**
50	24.45	25.29	24.12	**25.54**
75	22.78	23.54	22.96	**23.81**
100	21.94	22.45	22.24	**22.67**
Mandrill				
10	30.01	30.48	30.22	**30.55**
20	26.02	26.55	26.13	**26.60**
30	23.97	24.55	24.11	**24.63**
50	21.87	22.35	22.10	**22.52**
75	20.62	20.98	20.92	**21.16**
100	19.96	20.21	20.26	**20.38**

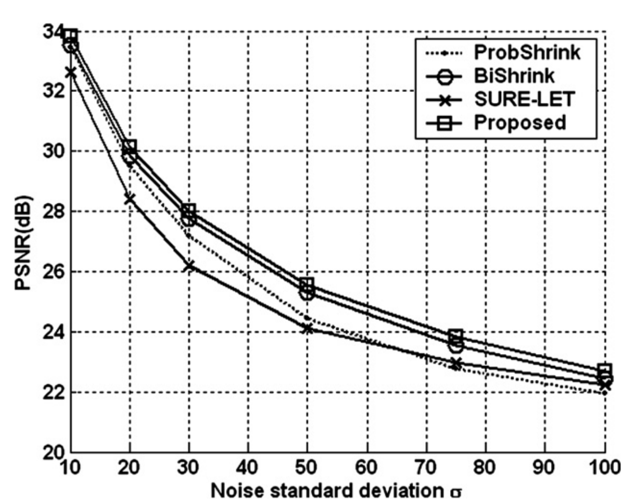

Fig. 3. PSNR curves of the four methods for Barbara.

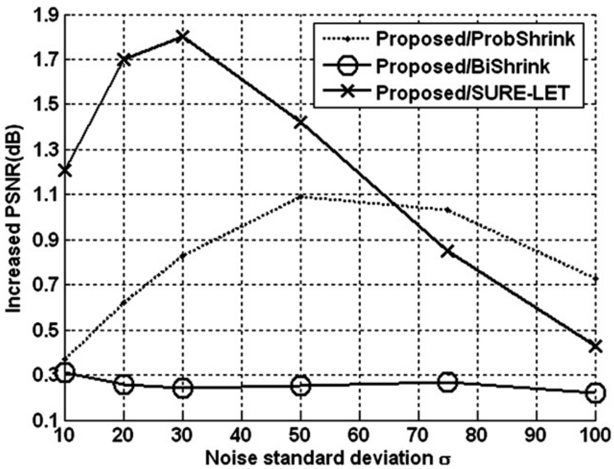

Fig. 4. PSNR gain curves of the proposed method compared with the other three methods for Barbara.

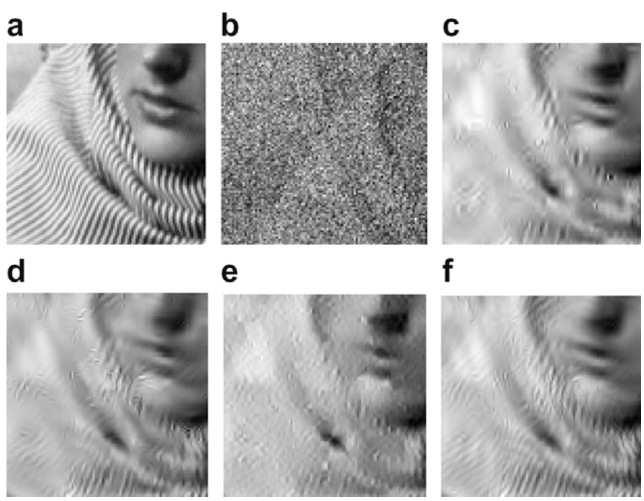

Fig. 5. Denoising results for Barbara: (a) Original image; (b) Noisy image with noise standard deviation $\sigma = 60$, PSNR = 2.57 dB; (c) Denoised image using ProbShrink, PSNR = 23.61 dB; (d) Denoised image using BiShrink, PSNR = 24.48 dB; (e) Denoised image using SURE-LET, PSNR = 23.56 dB; and (f) Denoised image using the proposed method, PSNR = 24.73 dB.

The curves of the PSNRs which the four methods produce for Barbara are shown in Fig. 3. In order to demonstrate more clearly the PSNR advantages of our method over the other three methods, Fig. 4 gives the PSNR gain curves of our method compared with the other three methods for Barbara. It visualizes the PSNR improving trends of our method as the noise level increases. The improvement of the PSNRs which the proposed method yields is large for Barbara compared with ProbShrink and SURE-LET. The largest increments are 1.1 dB and 1.8 dB, respectively.

The visual quality is also better using our method (see Fig. 5f). ProbShrink and SURE-LET suffer substantial blurring. The image that ProbShrink produces is too smooth while the one that SURE-LET does leaves over more noise. BiShrink and our method are considerably superior to ProbShrink and SURE-LET. However, BiShrink produces more disturbing artifacts than our method. Our method also remains more texture details of the Barbara's scarf. Our results also demonstrate the effectiveness of the DT-CWT. As Reviewer 3 has pointed out, the advantages of our method in visual quality and in PSNR are not only due to a different estimation rule but also benefit from the good directional sensitivity of the DT-CWT.

5. Conclusion

In this paper, we improve NeighShrink proposed by Chen et al. using the Stein's unbiased risk estimate (SURE). Compared with NeighShrink, the proposed method can determine an optimal threshold and neighbouring window size for every wavelet subband instead of using the suboptimal universal threshold and same neighbouring window size in all subbands. The proposed method is available on the DT-CWT. Thanks to the shift invariance and directional selectivity of the DT-CWT, the proposed method on the DT-CWT exhibits an excellent performance. Our experimental results indicate that it produces both higher PSNRs and better visual quality than three already published best denoising algorithms.

Acknowledgements

The authors would like to thank the anonymous reviewers and the associate editor for their constructive comments and suggestions to our paper.

References

Blu, T., Luisier, F., 2007. The SURE-LET approach to image denoising. IEEE Trans. Image Process. 16, 2778–2786.

Cai, T.T., Silverman, B.W., 2001. Incorporating information on neighbouring coefficients into wavelet estimation. Sankhya, Ser. B 63, 127–148.

Chen, G.Y., Bui, T.D., Krzyzak, A., 2005. Image denoising with neighbour dependency and customized wavelet and threshold. Pattern Recognition 38, 115–124.

Donoho, D.L., Johnstone, I.M., 1994. Ideal spatial adaptation via wavelet shrinkage. Biometrika 81, 425–455.

Donoho, D.L., Johnstone, I.M., 1995. Adapting to unknown smoothness via wavelet shrinkage. J. Amer. Statist. Assoc. 90, 1200–1224.

Fodor, I.K., Kamath, C., 2003. Denoising through wavelet thresholding: an empirical study. SPIE J. Electron. Imaging 12, 151–160.

Luisier, F., Blu, T., Unser, M., 2007a. A new SURE approach to image denoising: interscale orthonormal wavelet thresholding. IEEE Trans. Image Process. 16, 593–606.

Luisier, F., Blu, T., Unser, M., 2007b. SURE-LET Matlab Codes. Located at the URL: <http://bigwww.epfl.ch/demo/suredenoising/matlab.html>.

Pizurica, A., Philips, W., 2006a. Estimating probability of presence of a signal of interest in multiresolution single- and multiband image denoising. IEEE Trans. Image Process. 15, 654–665.

Pizurica, A., Philips, W., 2006b. ProbShrink Matlab Codes. Located at the URL: <ttp://telin.ugent.be/~sanja>.

Selesnick, I.W., Baraniuk, R.G., Kingsbury, N.G., 2005. The dual-Tree complex wavelet transform. IEEE Signal Process. Mag. 22, 123–151.

Sendur, L., Selesnick, I.W., 2002a. Bivariate shrinkage with local variance estimation. IEEE Signal Process. Lett. 9, 438–441.

Sendur, L., Selesnick, I.W., 2002b. DT-CWT and BiShrink Matlab codes. Located at the URL: <ttp://taco.poly.edu/WaveletSoftware/>.

Stein, C., 1981. Estimation of the mean of a multivariate normal distribution. Ann. Statist. 9, 1135–1151.