Proceedings of the 6th WSEAS International Conference on Multimedia Systems & Signal Processing, Hangzhou, China, April 16-18, 2006 (pp56-61)

Wavelet Based Image Denoising Using Intra Scale Dependency

HYUN-YOUNG GO, JI-BUM LEE, HYUNG-HWA KO
Dept. of Electronics and Communications Engineering
Kwangwoon University
447-1 Wolkye-Dong, Nowon-Gu, Chambit-kwan 813
SEOUL, REPUBLIC OF KOREA
http://ipcl.kw.ac.kr

Abstract:- This paper mainly focuses on the development of using wavelet coefficients' intra scale dependency of natural images. Wavelet transform(WT) coefficients have statistical dependency. WT coefficients have dependency between local coefficients (intra-scale). This paper uses dependency between children's coefficient for estimating corrupted by noise. For this purpose, we derive the bivariate model using correlation between coefficients and shrinkage function for denoising.

Key-Words:- Denoising, Wavelet transform, Wavelet shrinkage, Intra-scale dependency, Inter-scale dependency, MAP, Bivariate model, Gaussian model, Laplacian model

1 Introduction

Multiscale decompositions have shown significant advantages in the representation of signals, and they are used extensively in image compression [1], segmentation [2] and denoising [3, 4, 5] for example. In this thesis, we will deal with the dependency of the wavelet transform coefficients of natural image and its application to the image denoising problem. The denoising of a natural image corrupted by Gaussian noise is a classic problem in signal processing. The wavelet transform has become an important tool for this problem due to its energy compaction property. Crudely, it states that the wavelet transform yields a large number of small coefficients and a small number of large coefficients.

Simple denoising algorithms that use the wavelet tran sform consist of three steps.

1. Calculate the wavelet transform of the noisy signal,
2. Modify the noisy wavelet coefficients according to some rule,
3. Compute the inverse transform using the modified coefficients.

One of the most well-known rules for the second step is soft thresholding analyzed by Donoho [6. Due to its effectiveness and simplicity, it is frequently used in the literature.

Donoho's method of denoising performance is degradation because its method assumed wavelet coefficients are independent. But wavelet coefficients have statistically dependency. So we use the statistical dependency for denoising. The basic idea using the statistical dependency for denoising is to model wavelet transform coefficients with prior probability distributions. Then the problem can be expressed as the estimation of clean coefficients using this priori information with Bayesian estimation techniques, such as the MAP estimator. In this paper, we use correlation between children's coefficients of wavelet coefficients. For this purpose, we show bivariate model and from this, we derive the shrinkage function for denoising.

2 Denoising with Wavelets

In this paper, we will assume the signal is corrupted by additive white Gaussian noise. In denoising, we observe y (noisy signal), and estimate the desired signal x as closely as possible according to some criteria such as mean square error.

$$y = x + m \tag{1}$$

where the m is white Gaussian noise.

If we translate denoising problem in Eq. (1) in wavelet domain, the problem becomes,

$$\hat{w} = w + \eta \tag{2}$$

where $\hat{w}$ is noisy wavelet coefficients, w is desired wavelet coefficients, and η is white Gaussian noise.

To remove noise from a signal in transform domain can be simply explained of three steps.

1. Transform the noisy signal in transform domain.
2. Modify the transform domain coefficients according to some rule.
3. Inverse transform by using modified coefficients into original domain.

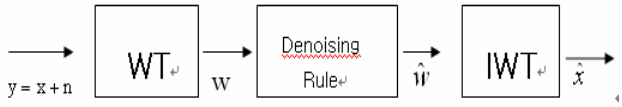


Fig. 1 General block diagram of wavelet domain denoising

The simplest denoising rule is hard thresholding [7] and soft thresholding[6]. Due to its effectiveness and simplicity, it is frequently used in signal processing. MAP based denoising method needs pdf model of wavelet coefficients. The general assumptions are Gaussian distribution and Laplacian distribution. From this assumption, one derive the shrinkage function for denoising. It is as below.

When the pdf is assumed Gaussian distributed, the shrinkage function is as

$$\hat{w}(y) = \frac{\sigma^2}{\sigma^2 + \sigma_n^2} \cdot y \quad (3)$$

where σ^2 , σ_n^2 are signal variance, and noise variance respectively.

When The pdf is assumed Laplacian distributed. the shrinkage function is as

$$\hat{w}(y) = \text{sign}(y)\left(|y| - \frac{\sqrt{2}\sigma_n^2}{\sigma}\right)_+ \quad (4)$$

Here $(g)_+$ is defined as

$$(g)_+ = \begin{cases} 0 & \text{if } g < 0 \\ g & \text{otherwise.} \end{cases}$$

But these all method is derived from the assumption that wavelet coefficients are independent. Due to its false assumation, the denoising performance is degradation. For using the dependency between wavelet coefficients, we show the bivariate model between wavelet coefficients and use it for denoising.

3 Bivariate Model for Denoising

Sendur[8] derive the bivariate model between parent and its children coefficient for denoising using MAP based method. He derived the bivariate model of wavelet coefficients' inter scale dependency between parent-children coefficients. However Wavelet coefficients have intra-scale dependency as well as inter-scale dependency. It means wavelet coefficients have dependency between coefficients in same subband. So we use intra scale dependency for denoising using MAP based method We adopt [8]'s pdf model for modeling the bivariate model between children coefficients. In fact [8]'s pdf model is the joint pdf between parent-children coefficient for inter-scale dependency. But we assume the pdf model between local coefficients (i.e. intra-scale) is same as the pdf model between parent-children coefficient(i.e. inter-scale). Now, we derive the bivariate model using this pdf model.

Let w_2 represent the one neighbor of w_1. Then

$$y_1 = w_1 + n_1$$
$$y_2 = w_2 + n_2$$

where y_1 and y_2 are noisy observations of w_1 and w_2; and n_1 and n_2 are noise samples. We can write

$$\mathbf{y} = \mathbf{w} + \mathbf{n} \quad (5)$$

where $\mathbf{w}=(w_1, w_2)$, $\mathbf{y}=(y_1, y_2)$ and $\mathbf{n}=(n_1, n_2)$.

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

$$\hat{\mathbf{w}}(\mathbf{y}) = \arg\max_{\mathbf{w}} p_{\mathbf{w}|\mathbf{y}}(\mathbf{w} \mid \mathbf{y}) \ . \quad (6)$$

After some manipulations, this equation can be written as

$$\begin{aligned} \hat{\mathbf{w}}(\mathbf{y}) &= \arg\max_{\mathbf{w}} [p_{\mathbf{y}|\mathbf{w}}(\mathbf{y} \mid \mathbf{w}) \cdot p_{\mathbf{w}}(\mathbf{w})] \\ &= \arg\max_{\mathbf{w}} [p_{\mathbf{n}}(\mathbf{y} - \mathbf{w}) \cdot p_{\mathbf{w}}(\mathbf{w})] . \end{aligned} \quad (7)$$

From this equation, Bayes rule allows us to write this estimation in terms of the probability densities of noise and the prior density of the wavelet coefficients. We assume the noise is i.i.d.(identically independent distributed) Gaussian, and we write the noise pdf as

$$p_{\mathbf{n}}(\mathbf{n}) = \frac{1}{2\pi\sigma_n^2} \bullet \exp\left(-\frac{n_1^2 + n_2^2}{2\sigma_n^2}\right) \quad (8)$$

And the joint pdf of signal for children coefficient and its one neighbor is adopted from [3]. It is as

$$p_{\mathbf{w}}(\mathbf{w}) = \frac{3}{2\pi\sigma^2} \cdot \exp\left(-\frac{\sqrt{3}}{\sigma}\sqrt{w_1^2 + w_2^2}\right) \quad (9)$$

Where w_1 means children coefficient, w_2 means its one neighbor. We use this joint pdf for modeling dependency between children coefficient and its one neighbor. This pdf is illustrated in Fig.2. and Fig. 3.

Let's continue on developing the MAP estimator given in (7), which is equivalent to

$$\hat{\mathbf{w}}(\mathbf{y}) = \arg\max_{\mathbf{w}}[\log(p_{\mathbf{n}}(\mathbf{y}-\mathbf{w})) + \log(p_{\mathbf{w}}(\mathbf{w}))] \quad (10)$$

Let's define $f(\mathbf{w}) = \log(p_{\mathbf{w}}(\mathbf{w}))$. By using (8), (10) becomes

$$\hat{\mathbf{w}}(\mathbf{y}) = \arg\max_{\mathbf{w}}[-\frac{(y_1-w_1)^2}{2\sigma_n^2} - \frac{(y_2-w_2)^2}{2\sigma_n^2} + f(\mathbf{w})] \quad (11)$$

This is equivalent to solving the following equations together, if $p_{\mathbf{w}}(\mathbf{w})$ is assumed to be strictly convex and differentiable.

$$\frac{y_1-\hat{w}_1}{\sigma_n^2} + f_1(\hat{\mathbf{w}}) = 0 \quad (12)$$

$$\frac{y_2-\hat{w}_2}{\sigma_n^2} + f_2(\hat{\mathbf{w}}) = 0 \quad (13)$$

where f_1 and f_2 represent the derivative of $f(w)$ with respect to w_1 and w_2 respectively.

Let's find the MAP estimator corresponding to model given in (9). $f(\mathbf{w})$ can be written as

$$f(\mathbf{w}) = \log(\frac{3}{2\pi\sigma^2}) - \frac{\sqrt{3}}{\sigma}\sqrt{w_1^2+w_2^2}\,. \quad (14)$$

From this,

$$f_1(\mathbf{w}) = -\frac{\sqrt{3}w_1}{\sigma\sqrt{w_1^2+w_2^2}}, \quad (15)$$

$$f_2(\mathbf{w}) = -\frac{\sqrt{3}w_2}{\sigma\sqrt{w_1^2+w_2^2}}, \quad (16)$$

Solving (12) and (13) by using (15) and (16), the MAP estimator (or "the joint shrinkage function") can be written as

$$\hat{w}_1 = \frac{\left(\sqrt{y_1^2+y_2^2} - \frac{\sqrt{3}\sigma_n^2}{\sigma})\right)_+}{\sqrt{y_1^2+y_2^2}} \cdot y_1 \quad (17)$$

Fig. 4 shows the plot of this bivariate shrinkage function. As this plot illustrates, there is a circular deadzone (the deadzone is the region where the estimated value is zero), i.e.

$$\text{deadzone} = \left\{(y_1, y_2) : \sqrt{y_1^2+y_2^2} \le \frac{\sqrt{3}\sigma_n^2}{\sigma}\right\}$$

This results clearly show that the estimated value should depend on the neighbor value. The smaller the neighbor value, the greater the shrinkage.

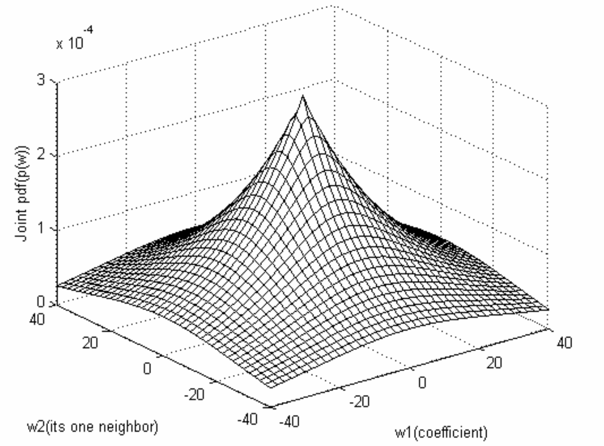


Fig. 2 Bivariate pdf eq. (9) proposed for joint pdf of children coefficient and its one neighbor pairs.

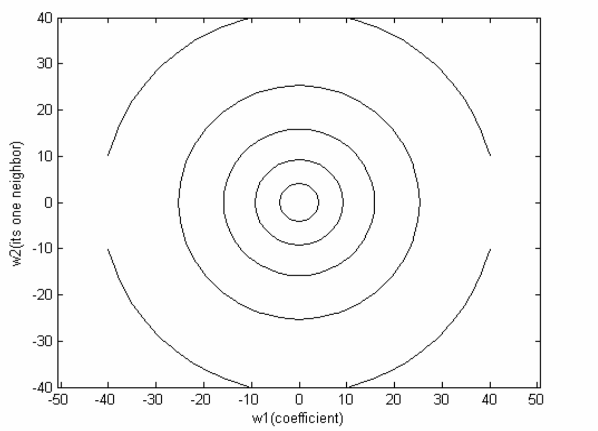


Fig. 3 Contour of bivariate pdf eq. (9)

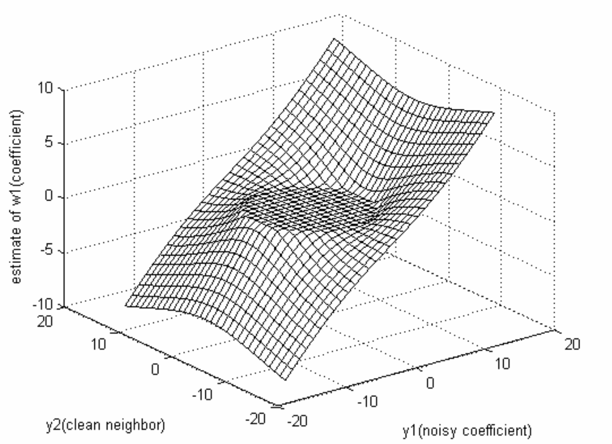


Fig. 4 Bivariate shrinkage function derived from the proposed method (Fig. 2)

4 One Neighbor Selection

We have derived the shirinkage function for denoising using bivariate model between coefficient and its one neighbor coefficient. This method is estimated a coefficient from one neighbor. So, if neighbor coefficient is almost noise free, the estimated result is also good. Now we focus what one neighbor coefficient selects for denoising. For this, we experiment three choosing method.

The choosing method is as below

1. Median Selection
2. Random Selection
3. 1 upper Selection

We use median selection due to one assumption. This assumption is based the noise is zero mean Gaussian noise. As you know, one coefficient has 8 neighbors. The noise is additive Gaussian, and zero mean, so the probability that coefficient is corrupted by additive zero is very high..

Second, we use random selection among 8 neighbors.

Third, we also use 1 upper selection method i.e. selecting 1 upper one of coefficient for estimation. This method is done for fast simulation time. It selects merely 1 upper one. So we also experiment this method how to do good for denoising.

5 Simulation and Result

We simulate the proposed algorithm using the condition –orthogonal wavelet transform-Daubeiches length 8 filter- and 5 level decomposition. We implement the propsed algorithm by using Matlab program. The simulation is done using Pentium4 1.6Ghz computer. The PSNR values of these systems are tabulated in Table 1. The Lena, Barbara 512 by 512 size images are used for comparison purpose with different noise levels, σ_n^2. Sendur's method is using dependency between parent-children coefficient proposed in 2002. Our method is using dependency between coefficient and its one neighbor. It takes about 1.85, 13.30, 3.85, 1.82 seconds for Sendur's, median selection, random selection, 1 upper selection respectively when we check the consuming time for simulation.

As seen in Table 1, we see 1 upper one selection method has best performance among our method. For more comparison we compare with [8] proposed recently. The performance of our proposed method is betten than [8]'s for Barbara image about 0.5dB but has degradation performance for Lena image about maximum 0.59dB.

Denoising example using 512 by 512 Lena image and Barbara image is given in Fig. 5, Fig. 6 respectively.

Table 1 The PSNR values of denoised images for different test images and (a)noise levels σ_n, (b)noisy, (c)hard shirinkage[1], (d)Sendur's, (e)median selection, (f)random selection, (g) 1 upper selection.

unit: dB

	(a)	(b)	(c)	(d)
Lena	10	28.17	30.34	33.94
	15	24.66	28.52	-
	20	22.15	27.24	30.73
	25	20.21	26.34	-
	30	18.60	-	28.94
Bar-bara	10	28.16	27.29	31.13
	15	24.64	25.01	-
	20	22.13	23.65	27.25
	25	20.22	22.83	-
	30	18.62	-	25.21

	(a)	(b)	(e)	(f)	(g)
Lena	10	28.17	33.00	33.20	33.35
	15	24.66	31.15	31.37	31.52
	20	22.15	30.00	30.16	30.32
	25	20.21	28.97	29.20	29.37
	30	18.60	28.24	28.49	28.63
Bar-bara	10	28.16	31.00	31.18	31.26
	15	24.64	28.82	29.20	29.24
	20	22.13	27.09	27.54	27.76
	25	20.22	25.63	26.23	26.55
	30	18.62	24.66	25.12	25.51

(a)

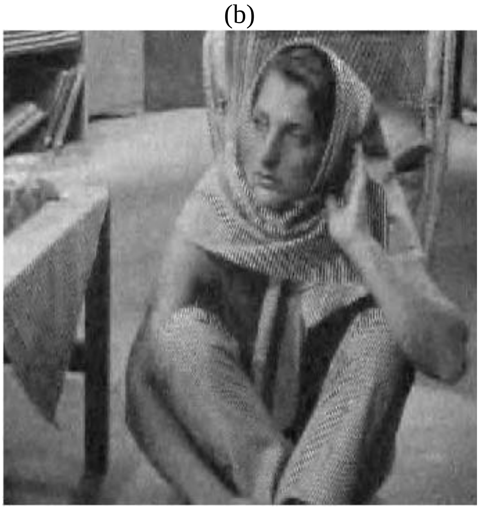
(b)

(c)

Fig. 5 (a) Original Lena image. (b) Noisy image with PSNR= 18.60dB.(σ_n =30) (c) Denoised image using proposed method, PSNR= 28.63dB.

(a)

(b)



(c)

Fig. 6 (a) Original Barbara image. (b) Noisy image with PSNR= 18.62dB.(σ_n =30) (c) Denoised image using proposed method, PSNR= 25.51dB.

6 Conclusion

Wavelet coefficients have intra-scale dependency. So we use this dependency for denoising. First we characterize the dependencies between children and its one neighbor using bivariate model. And we need one noise free coefficient for denoising. For this purpose, we select one neighbor coefficient by selection method among 8 neighbors. We have seen 1 upper selection method has best performance. Performance of our method is better than classical denoising method of hard threshold. And also we compare with recent denoising method[8]. Our method is better than [8] for Barbara image but less for Lena image.

References

[1] C. Christopoulos, A. Skodras, and T. Ebrahimi, The JPEG 2000 still image coding system: an overview, *IEEE Transactions on Consumer Electronics*, 46(4):1103-1127, 1993.

[2] S. G. Chang, B. Yu, and M. Vetterli, Spatially adaptive wavelet thresholding with context modeling for image denoising, *IEEE Trans. on Image Processing*, 9(9):1522-1531, 2000

[3]R.Coifman and D.Donoho, Time-invariant wavelet denoising, *In A. Antoniadis and G..Oppenheim, editors, Wavelet and statistics*, volume 103 of Lecture Notes in Statistics, New York, Springer-Verlag, 1995, pp.125-150

[4] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, Low-complexity image denoising based on statistical modeling of wavelet coefficients, *IEEE Signal Processing Letters*, 6(12):300-303, 1999.

[5] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, Adaptive Wiener denoising using a Gaussian scale mixture model, *Proc. Int. Conf. Image Processing*, 2001.

[6D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans. on Information Theory*, 41(3):613-627, May 1995.

[7]D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, pages 425-455, 1994.

[8]L. Sendur and I.W. Selesnick, Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. *IEEE Transactions on signal Processing*, 30, November 2002.