

# CONTEXT-BASED BIAS REMOVAL OF STATISTICAL MODELS OF WAVELET COEFFICIENTS FOR IMAGE DENOISING

Weisheng Dong<sup>1</sup>, Xiaolin Wu<sup>2</sup>, Guangming Shi<sup>1</sup>, and Lei Zhang<sup>3</sup>

<sup>1</sup>Key Lab. of IPIU of Ministry of Education, Xidian University, Xi'an, China

<sup>2</sup>Dept. of Electrical and Computer Engineering, McMaster University, Canada

<sup>3</sup>Dept. of Computing, Hong Kong Polytechnic University, Hong Kong

## ABSTRACT

Existing wavelet-based image denoising techniques all assume a probability model of wavelet coefficients that has zero mean, such as zero-mean Laplacian, Gaussian, or generalized Gaussian distributions. While such a zero-mean probability model fits a wavelet subband well, in areas of edges and textures the distribution of wavelet coefficients exhibits a significant bias. We propose a context modeling technique to estimate the expectation of each wavelet coefficient conditioned on the local signal structure. The estimated expectation is then used to shift the probability model of wavelet coefficient back to zero. This bias removal technique can significantly improve the performance of existing wavelet-based image denoisers.

*Index Terms*— Context modeling, estimation bias, image denoising, Bayesian shrinkage.

## 1. INTRODUCTION

Image denoising is a classic problem that has as long a history as the discipline of digital image processing, and still remains to be an active research topic. A popular and effective methodology of image denoising is thresholding in the wavelet domain, which was first proposed by Donoho [1]. Since this pioneer work, numerous thresholding-based techniques have been developed. In particular, adaptive shrinkage estimators with Bayesian approach have demonstrated good performance [2][3][4][5][6]. In the design of adaptive shrinkage estimators, a statistical model of the noise-free wavelet coefficients is of central importance.

Based on the fact that the probability density functions (PDFs) of wavelet coefficients in a wavelet subband are sharply peaked at zero and heavy-tailed, current statistical models characterize the wavelet coefficients using parametric distributions, such as zero-mean Gaussian, Laplacian, and generalized Gaussian. To better describe the heavy-tailed PDFs, higher order marginal statistical models, such as Gaussian scale mixture model [3], mixture Laplacian model [4], Gauss-Hermite model [5], and bivariate model [2], are also proposed.

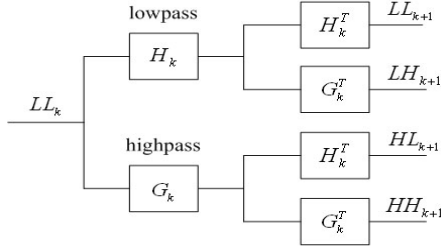
While the set of all wavelet coefficients in a particular subband do have zero mean, in regions of edges and textures the distribution of wavelet coefficients often exhibits a significant bias away from zero. This phenomenon was noticed by researchers of image coding, and exploited in context-based entropy coding of wavelet coefficients [7]. Similarly, in predictive coding the prediction residuals are not zero mean either in different contexts [8]. Inspired by the success of context modeling of coefficients in wavelet-based image coding, we propose to estimate the expectation of each wavelet coefficient conditioned on the local signal structure. The estimated expectation is then used to shift the probability model of the wavelet coefficient back to zero. To estimate the expectation, coefficients of similar contexts, which are not necessarily spatially adjacent, are grouped. The expectation is then estimated from these grouped coefficients. Context modeling was used by Chang et al. to estimate the local signal variances when thresholding wavelet coefficients for denoising [6]. But the authors still assumed zero mean of the context model. In addition to making the context model more general with arbitrary mean, we couple context modeling with overcomplete wavelet transform. This allows us to thoroughly exploit high-order statistics of wavelet coefficients, with the spatial resolution being maximum in all subbands. In order to make context modeling in overcomplete wavelet bases computationally tractable, we choose to form modeling contexts in spatial rather than wavelet domain. Experimental results show that the proposed context modeling technique of bias removal can significantly improve the performance of the existing wavelet-based image denoisers.

The paper is organized as follows. Sec. II presents non-zero mean context modeling of wavelet coefficients. Sec. III presents bias removal of probability models for Bayesian denoising. Experimental results are presented in Sec. IV.

## 2. NON-ZERO MEAN CONTEXT MODELING OF WAVELET COEFFICIENTS

It is well known that the overcomplete wavelet transform (OWT) achieves better denoising performance than critically

sampled wavelet transform [9]. In this paper we adopt OWT. At decomposition level  $k$ , the 2D OWT, as shown in Fig.1, employs low-pass and high-pass filters  $H_k$  and  $G_k$  that are generated by inserting zeros between two adjacent coefficients of corresponding original filters  $H_0$  and  $G_0$ . The restored image is obtained by first downsampling the denoised subbands in horizontal and vertical direction, and then averaging over all possible reconstructed versions of the image signal.



**Fig. 1.** One stage of the 2D overcomplete wavelet transform.

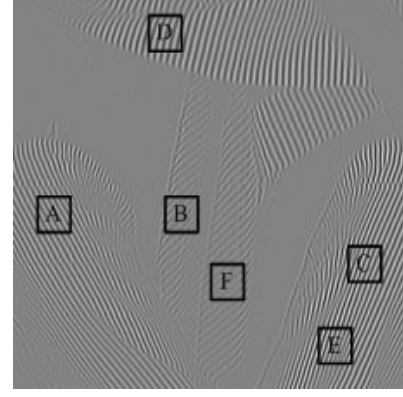
It is a common practice in the fields of image coding and denoising to model wavelet coefficients by heavy-tailed distributions, such as zero-mean Laplacian, Gaussian, Generalized Gaussian. Recent works investigate the higher order marginal statistics of wavelet coefficients, as well as the dependency between wavelet coefficients [2][3][4][5][6]. As pointed out in the introduction, all existing probability models of wavelet coefficients assume zero mean, which is the unfortunate oversight to be rectified by this work. To illustrate such a fact we plot in Fig. 2 a part of the LH wavelet subband of test image Barbara, and in Fig.3 the conditional sample histograms in different contexts as marked in Fig. 2. It can be observed from the figures that sample histograms conditioned on different local patterns of edges and textures can differ significantly from one the other and have means quite far away from zero. This observation exposes a flaw of wavelet-based denoisers that assume a zero-mean probability model.

In context modeling of wavelet coefficients, the problem is to estimate the conditional probability

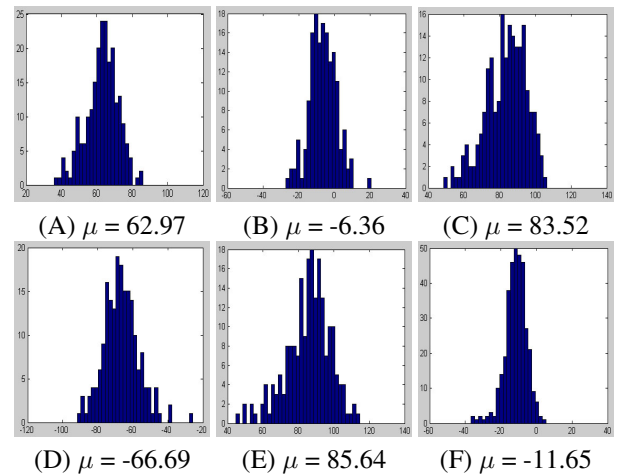
$$P(x_{k,\theta}(i,j)|C_{k,\theta}(i,j)), \quad (1)$$

where  $x_{k,\theta}(i,j)$  is the wavelet coefficient at the position  $(i,j)$  of the subband of scale  $k$  and orientation  $\theta$ , and  $C_{k,\theta}(i,j)$  is the context in which random variable  $x_{k,\theta}(i,j)$  takes its value. To simplify the estimation task we use a parametric model (e.g., Laplacian) for  $P(x_{k,\theta}(i,j)|C_{k,\theta}(i,j))$ . Then the problem is reduced to the estimation of the variance and mean for each context  $C_{k,\theta}(i,j)$ .

Since OWT is used, each subband  $(k,\theta)$  has the same number of coefficients as the number of pixels. To take into consideration all sample dependencies in both spatial and frequency domains, context  $C_{k,\theta}(i,j)$  should consist of



**Fig. 2.** A part of the LH subband in the 1st-level decomposition of image Barbara.



**Fig. 3.** Histograms of OWT coefficients conditioned on contexts  $A, \dots, F$  shown in Fig.2.

wavelet coefficients in a neighborhood of  $(i,j)$  in all subbands  $C_{k,\theta}(i,j)$ ,  $1 \leq k \leq K$ ,  $\theta \in \{LH, HL, HH\}$  where  $K$  being the total number of OWT decomposition levels. But this formation of modeling context faces difficulties. First, the context becomes a vector of very high dimensions. As such context modeling directly in the OWT domain suffers from the problem of context dilution. The second difficulty is high computational complexity involved in computing  $C_{k,\theta}(i,j)$ .

On a second reflection, however, we can circumvent the above difficulties by a simple switch of context space. Since we need to amalgamate the OWT coefficients near spatial position  $(i,j)$  over all subbands as argued in the proceeding paragraph, context  $C_{k,\theta}(i,j)$  can be made the same for all  $k$  and  $\theta$ . Thus we form the context of  $x_{k,\theta}(i,j)$  directly in spatial domain, using pixels in a neighborhood of  $(i,j)$ , denoted by  $C(i,j)$ . Now the context modeling problem becomes to estimate the mean and variance of  $P(x_{k,\theta}(i,j)|C(i,j))$ , noting that each random variable in the set  $\{x_{k,\theta}(i,j)\}_{k,\theta}$  is conditioned on the same spatial context  $C(i,j)$ , regardless which

subband  $(k, \theta)$  is in question. Such a switch of context space from  $C_{k,\theta}(i, j)$  to  $C(i, j)$  does not affect the outcomes of our estimation because the distance between two spatial contexts  $C(i, j)$  and  $C(m, n)$  is proportional to the distance between the corresponding  $C_{k,\theta}(i, j)$  and  $C_{k,\theta}(m, n)$  in the OWT domain. Indeed, for the linear OWT operation  $T_{k,\theta}(C(i, j)) = C_{k,\theta}(i, j)$ , we have

$$\begin{aligned} \|C_{k,\theta}(i, j) - C_{k,\theta}(m, n)\|_2 &= \|T_{k,\theta}(C(i, j) - C(m, n))\|_2 \\ &= (C(i, j) - C(m, n))' T_{k,\theta}' T_{k,\theta} (C(i, j) - C(m, n)) \\ &= \alpha \|C(i, j) - C(m, n)\|_2, \end{aligned} \quad (2)$$

where  $\alpha I = T_{k,\theta}' T_{k,\theta}$  is a constant.

To estimate the mean and variance of  $P(x_{k,\theta}(i, j)|C(i, j))$ , we collect a sample set  $S(i, j)$  of coefficients  $x_{k,\theta}(m, n)$  whose contexts  $C(m, n)$  are similar to  $C(i, j)$ . The mean and variance are then estimated from the sampled coefficients. In practice, the sample set of coefficients  $x_{k,\theta}(m, n)$  is formed such that

$$S(i, j) = \left\{ x_{k,\theta}(m, n) \left\| C(m, n) - C(i, j) \right\|_2 < \tau \right\}, \quad (3)$$

where  $\tau$  is a threshold. Or alternatively, we can include in  $S(i, j)$  those coefficients  $x_{k,\theta}(m, n)$  such that the context  $C(m, n)$  is within the  $\kappa^{\text{th}}$  closest to  $C(i, j)$ .

### 3. CENTERING OF PROBABILITY MODELS FOR BAYESIAN DENOISING

Having estimated the mean of  $P(x_{k,\theta}(i, j)|C(i, j))$ , which is nonzero in general, we need to cancel the bias by centering  $P(x_{k,\theta}(i, j)|C(i, j))$  at the origin. The centered probability model is then used to drive the MAP estimator in the class of wavelet-based Bayesian-type denoisers. Given a prior probability  $p(x)$  of wavelet coefficient  $x$  and assuming an additive zero-mean Gaussian noise of variance  $\sigma_n^2$ , the classic MAP estimate of  $x$  can be obtained by solving the following equation [2]

$$\frac{d}{dx} \left[ -\frac{(y-x)^2}{2\sigma_n^2} + \log(p(x)) \right] = 0. \quad (4)$$

Four examples of the centering of existing probability models for Bayesian denoising are given in this section.

#### 3.1. Bias removal of Gaussian model

Let  $\mu$  and  $\sigma_x^2$  be the expectation and variance of the wavelet coefficient, respectively. Then the bias removed Gaussian density function  $p(x)$  is

$$p(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma_x^2}\right). \quad (5)$$

The corresponding Bayesian estimator is given by

$$\hat{x}(y) = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_n^2} y + \left(1 - \frac{\sigma_x^2}{\sigma_x^2 + \sigma_n^2}\right) \mu \quad (6)$$

#### 3.2. Bias removal of Laplacian model

The bias removed Laplacian density function with expectation  $\mu$  is

$$p(x) = \frac{1}{\sqrt{2}\sigma_x} \exp\left(-\frac{\sqrt{2}|x-\mu|}{\sigma_x}\right) \quad (7)$$

and the modified Bayesian estimator is

$$\hat{x}(y) = \text{sign}(y-\mu) \cdot \max(|y-\mu| - \frac{\sqrt{2}\sigma_n^2}{\sigma_x}, 0) + \mu \quad (8)$$

#### 3.3. Bias removal of bivariate model

Let  $\mu_1$  and  $\mu_2$  be the expectations of the wavelet coefficient and the parent coefficient, respectively. Then the bias removed bivariate model is expressed as

$$p(\mathbf{x}) = \frac{3}{2\pi\sigma_x^2} \exp\left(-\frac{\sqrt{3}}{\sigma_x} \sqrt{(x_1-\mu_1)^2 + (x_2-\mu_2)^2}\right) \quad (9)$$

And the new bivariate estimator is given by:

$$\hat{x}(y) = \frac{\max(0, \sqrt{(y_1-\mu_1)^2 + (y_2-\mu_2)^2} - \frac{\sqrt{3}\sigma_n^2}{\sigma_x})}{(y_1-\mu_1)^{-1} \sqrt{(y_1-\mu_1)^2 + (y_2-\mu_2)^2}} + \mu_1 \quad (10)$$

#### 3.4. Bias removal of Gaussian Scale Mixture model

For the case of more sophisticated Gaussian scale mixture model [3], similarly to the above, we first remove the estimated conditional mean from noisy wavelet coefficients, then apply the BLS-GSM estimator to process the mean-removed noisy coefficients, and finally add the conditional mean back to the denoised wavelet coefficients.

## 4. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed context modeling technique for bias removal in wavelet-based image denoising. Due to the space limit, we present only three test images: Barb, Lena, and Foreman. Gaussian noises of  $\sigma_n = 10, 20, 30$ , are added in our simulation. We report our results for OWT of the well-known 9/7 wavelet with five levels of decomposition in Table I, in which the mean-removed Bayesian estimators are compared to their biased counterparts. In the table the estimators derived from zero-mean Gaussian, Laplacian, Bivariate, Gaussian scale mixture models are labeled as Gauss-0, Lap-0, Bivar-0, and BLS-GSM-0, respectively, while the corresponding mean-removed estimators are labeled by Gauss- $\mu$ , Lap- $\mu$ , Bivar- $\mu$ , and BLS-GSM- $\mu$ . The results show that the mean-removed wavelet denoisers significantly improve the performance of their zero-mean counterparts. In some cases, the gain can be up to 1dB or more. To assess the visual quality we also present sample images in Fig. 4, where one can see that the edges and textures in images denoised by the new methods are much sharper and cleaner than those by the existing methods.

## 5. ACKNOWLEDGEMENT

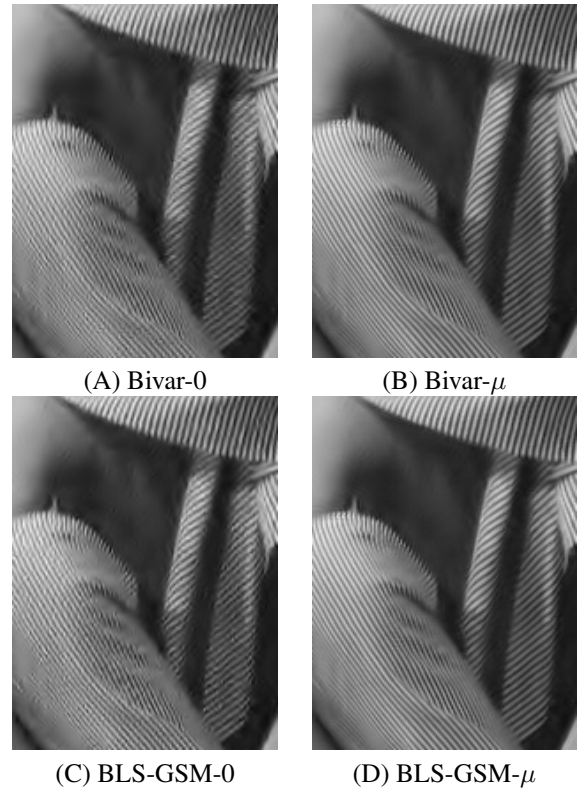
This work is supported in part by National High Technology Research and Development Program of China (NO. 2007AA01Z307), in part by NSFC (NO. 60736043, 60776795, 60672125, 60805012), in part by PCSIRT (NO. IRT0645).

**Table 1.** PSNR results of different methods.

Noise $\sigma_n$	10	20	30
Barbara			
Gauss-0	33.46	29.48	27.17
Gauss- $\mu$	<b>34.27</b>	<b>30.68</b>	<b>28.37</b>
Lap-0	33.35	29.43	27.18
Lap- $\mu$	<b>34.07</b>	<b>30.67</b>	<b>28.49</b>
Bivar-0	33.29	29.43	27.26
Bivar- $\mu$	<b>34.08</b>	<b>30.65</b>	<b>28.57</b>
BLS-GSM-0	33.49	29.49	27.23
BLS-GSM- $\mu$	<b>34.16</b>	<b>30.63</b>	<b>28.56</b>
Foreman			
Gauss-0	34.67	30.86	28.56
Gauss- $\mu$	<b>35.77</b>	<b>31.91</b>	<b>29.39</b>
Lap-0	34.86	31.18	28.99
Lap- $\mu$	<b>35.90</b>	<b>32.26</b>	<b>29.86</b>
Bivar-0	35.05	31.67	29.71
Bivar- $\mu$	<b>35.88</b>	<b>32.54</b>	<b>30.23</b>
BLS-GSM-0	35.28	31.70	29.68
BLS-GSM- $\mu$	<b>35.96</b>	<b>32.51</b>	<b>30.29</b>
Lena			
Gauss-0	35.07	31.57	29.32
Gauss- $\mu$	<b>35.40</b>	<b>32.00</b>	<b>29.76</b>
Lap-0	35.10	31.85	29.81
Lap- $\mu$	<b>35.35</b>	<b>32.23</b>	<b>30.16</b>
Bivar-0	35.20	32.27	30.47
Bivar- $\mu$	<b>35.32</b>	<b>32.38</b>	<b>30.44</b>
BLS-GSM-0	35.35	32.34	30.49
BLS-GSM- $\mu$	<b>35.40</b>	<b>32.38</b>	<b>30.43</b>

## 6. REFERENCES

- [1] D.L. Donoho, "De-noising by soft thresholding," *IEEE Trans. Inform. Theory*, vol. 41, pp. 613–627, May 1995.
- [2] L. Sendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2744–2756, Nov. 2002.
- [3] J. Portilla, V. Strela, M. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of gaussians in the



**Fig. 4.** Visual comparison of different methods for  $\sigma_n = 20$ .

- wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.
- [4] H. Rabbani, M. Vafadust, and S. Gazor, "Image denoising based on a mixture of laplace distributions with local parameters in complex wavelet domain," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2006, pp. 2597–2600.
- [5] S. M. M. Rahman, M. O. Ahmad, and M. N. S. Swamy, "Bayesian wavelet-based image denoising using the gauss-hermite expansion," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1755–1771, Oct. 2008.
- [6] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1522–1531, Sep. 2000.
- [7] X. Wu, "High-order context modeling and embedded conditional entropy coding of wavelet coefficients for image compression," in *Proc. 31st IEEE Asilomar Conf. Signals, System, and Computers*, 1997, pp. 1378–1382.
- [8] X. Wu and N. Memon, "Context-based, adaptive, lossless image coding," *IEEE Trans. Communication*, vol. 45, no. 4, pp. 437–444, Apr. 1997.
- [9] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," in *Wavelets and Statistics*. Springer-Verlag Lecture Notes, 1995.