

# Interscale Image Denoising With Wavelet Context Modeling

Lei Zhang<sup>†</sup>, Paul Bao<sup>\*†</sup> and David Zhang<sup>†</sup>

<sup>†</sup>Computing Dept., The Hong Kong Polytechnic University  
Hung Hum, Kowloon, Hong Kong

<sup>†</sup>Dept. of Information Engineering, The Chinese University of Hong Kong  
Shatin, New Territory, Hong Kong  
Paulbao@ie.cuhk.edu.hk

## ABSTRACT

This paper presents a wavelet-based linear minimum mean square-error estimation (LMMSE) scheme to exploit the strong wavelet interscale dependencies for image denoising. Using overcomplete wavelet expansion (OWE), we group the wavelet coefficients with the same spatial orientation at adjacent scales as a vector. The LMMSE algorithm is then applied to the vector variable. This scheme exploits the correlation information of wavelet scales to improve noise removal. To calculate the statistics of wavelet coefficients more adaptively, we classify them into different clusters by the context modeling technique, which yields a good local discrimination between edge structures and backgrounds. Experiments show that the proposed scheme outperforms some existing denoising methods. And a biorthogonal wavelet, which well characterizes the interscale dependencies, is found very suitable for the scheme.

## 1. INTRODUCTION

A wavelet transform (WT) [4] is a Karhunen-Loève like expansion and it can decorrelate random processes into nearly independent coefficients, which could be statistically modeled more effectively. Many wavelet based schemes for image denoising and coding have been proposed with good results [1, 2, 6, 7, 9, 11].

Although WT decorrelates images well, there still exist strong dependencies between wavelet coefficients in the intrascale or interscale manner. Exploiting the dependency information with proper statistical models will significantly improve the performance of coding and denoising algorithms. As noted by Liu and Moulin [8], the statistical wavelet models can be classified into intrascale models, interscale models and hybrid dependency models. Some intrascale models have been proposed in [1, 2, 10]. Chang *et al* [1] defined each wavelet coefficient as a mixture of GGD variables with unknown slowly spatially varying parameters. The estimation of these parameters is

conditioned on a function of its neighboring coefficients. The morphological coder presented in [10] also exploits the spatial clustering of wavelet coefficients. Mihçak *et al* [2] estimated the two-order local statistics of each coefficient with a centered square window and developed a linear minimum mean squared-error estimation (LMMSE) like method for noise removal.

The image wavelet coefficients are highly correlated at adjacent scales. If a coefficient at a coarser scale has small magnitude, its descendant coefficients at finer scales are also likely to be small. Shapiro exploited this property to develop the now well-known embedded zerotree wavelet coder [3]. Conversely, if a wavelet coefficient produced by a true signal is of large magnitude at a finer scale, its parents at coarser scales will also be large. However for those coefficients produced by noise, the magnitudes will decay rapidly along the scales. Multiplying the adjacent wavelet scales would sharpen the important structures while reducing noise. Such a property has been exploited in denoising [6, 7] and edge detection [12]. The wavelet inter-scale dependencies have also been represented by hidden Markov models [5].

Wavelet intrascale and interscale dependencies can be combined to improve noise removal results. Liu and Moulin [9] employed a composite model in denoising and they analyzed the dependency between wavelet coefficients with a measurement of mutual information [8]. They also compared various wavelet filters with respect to the dependency information capturing ability.

In this paper, an LMMSE-base denoising scheme with an interscale wavelet model is firstly developed. The overcomplete wavelet expansion (OWE), which noticeably outperforms the orthogonal wavelet transform (OWT) in image denoising [11], is employed. We group the wavelet coefficients with the same spatial location across adjacent scales as a vector, and apply the LMMSE to the vector variable to smooth noise. Secondly, to exploit the intrascale dependencies to model the wavelet

---

\* The Corresponding Author.

coefficients adaptively in spatial domain, we cluster the wavelet coefficients using context modeling technique, which gives a local discrimination of image characteristics, such as edge structures and backgrounds. Experiments show that the proposed scheme yields very satisfying denoising results, especially with a biorthogonal wavelet.

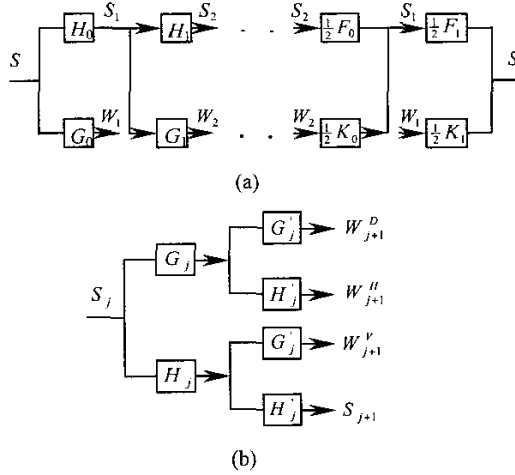


Fig. 1. The illustrations of overcomplete wavelet expansion (OWE). (a) Decomposition and reconstruction of 1-D signals; (b) One-stage decomposition of 2-D images.

## 2. OVERCOMPLETE WAVELET EXPANSION

A wavelet transform (WT) represents a signal  $f$  as a linear combination of elementary atoms or building blocks. A detail description of wavelet theory can be found in [4]. Besides Haar wavelet, there is no other compactly supported orthogonal wavelet that is (anti-)symmetrical, which is an important property in signal processing. Compactly supported biorthogonal wavelets discard the orthogonality to preserve the symmetric property.

(Bi-)orthogonal wavelet transform (OWT) is translation variant due to downsampling in decomposition. This causes some visual artifacts (such as Gibbs phenomena) in denoising applications [11]. The denoising scheme presented in this paper is based on overcomplete wavelet expansion (OWE), whose structure is shown in Fig. 1.  $H_j$  is obtained by put  $(2^{j-1} - 1)$  zeros between each of the coefficients of  $H_0$ , similarly for  $G_j, F_j$  and  $K_j$ . 2-D OWE can be extended from 1-D by separable filtering. The coefficients are in three directions: horizontal, vertical and diagonal. Fig. 1 (b) is the one-stage decomposition structure of 2-D OWE. Filter  $F'$  is the

transpose of  $F$ . At each scale the wavelet coefficients of OWE have the same number of samples as the input image.

## 3. THE DENOISING ALGORITHM

### 3.1. The LMMSE of wavelet coefficients

Suppose  $f$  is corrupted with signal independent white noise:  $g = f + \varepsilon$  with  $\varepsilon \in N(0, \sigma^2)$ . The OWE of  $g$  at scale  $j$  is written as  $Z_j = X_j + V_j$ , where  $X_j$  and  $V_j$  are the OWE of  $f$  and  $\varepsilon$  respectively. Denote by  $\sigma_{X_j}^2$  and  $\sigma_j^2$  the variances of  $X_j$  and  $V_j$ , the linear minimum mean squared-error estimation (LMMSE) of  $x_j$  is

$$\hat{x}_j = \sigma_{X_j}^2 \cdot z_j / (\sigma_{X_j}^2 + \sigma_j^2) \quad (1)$$

The noise standard deviation at scale  $j$  is computed as

$$\sigma_j = \|\psi_j\| \sigma \quad (2)$$

where  $\psi_j(x, y) = 2^{-j} \psi(2^{-j}x, 2^{-j}y)$  is the dilation of 2-D mother wavelet  $\psi$  and  $\|\psi\| = \sqrt{\iint \psi(x, y) dx dy}$ .

$\sigma_{X_j}$  can be estimated as

$$\hat{\sigma}_{X_j}^2 \approx \sigma_{Z_j}^2 - \sigma_j^2 \quad (3)$$

with

$$\sigma_{Z_j}^2 = \frac{1}{M \cdot N} \sum_{m=1}^M \sum_{n=1}^N Z_j^2(m, n) \quad (4)$$

being the variance of  $Z_j$  whose size is  $M \times N$ .

### 3.2. The interscale model-based LMMSE

By OWE each of wavelet subbands has the same number of samples. We assemble the same oriented points at  $J$  scales as a vector

$$\bar{Z}(m, n) = [Z_1(m, n) \dots Z_J(m, n)]^T \quad (5)$$

Then there is  $\bar{Z} = \bar{X} + \bar{V}$ . Obviously  $\bar{V}$  is Gaussian vector noise independent of  $\bar{X}$ . Applying the LMMSE to vector variable  $\bar{Z}$ , we obtain the estimation of  $\bar{X}$ :

$$\hat{\bar{X}} = P(P + R)^{-1} \bar{Z} \quad (6)$$

where  $P = E[\bar{X}\bar{X}^T]$  and  $R = E[\bar{V}\bar{V}^T]$ .

The diagonal element  $E[v_j^2]$  of  $R$  equals to  $\sigma_j^2$ , which can be computed by (2). Noise variables  $v_i$  and  $v_j$ ,  $i \neq j$ , are jointly Gaussian. Their correlation coefficient is

$$\rho_{i,j} = \frac{\iint \psi_i(x, y) \psi_j(x, y) dx dy}{\sqrt{\iint \psi_i^2(x, y) dx dy} \cdot \sqrt{\iint \psi_j^2(x, y) dx dy}} \quad (7)$$

Thus we have

$$R_{i,j} = E[v_i v_j] = \rho_{i,j} \sigma_i \sigma_j \quad (8)$$

The components of matrix  $P$  can be estimated by

$$P_{i,j} = E[x_i x_j] \approx E[z_i z_j] - R_{i,j} \quad (9)$$

where  $E[z_i z_j]$  is calculated as

$$E[z_i z_j] = \frac{1}{M \cdot N} \sum_{m=1}^M \sum_{n=1}^N Z_i(m, n) \cdot Z_j(m, n) \quad (10)$$

Suppose the input image is transformed into  $J$  wavelet scales, and it is found by our experiments that setting  $\bar{z}$  as a  $J$ -D vector won't yield satisfying results. This could be explained in two ways. Firstly, scale  $j$  is strongly correlated with scale  $j+1$  but its correlation with scale  $j+2, j+3, \dots, J$  decreases rapidly. These coarser scales won't convey much information to improve the estimation of scale  $j$ . Secondly a significant structure has larger spatial supports at coarser scales than finer scales. One point at a coarse scale may appear as an edge, but at finer scales the points with the same spatial orientation may be noise predominated. Thus fine scales may impose negative effects on coarse scales.

Based on these considerations, we recover  $X_j$  only by the measurements at scales  $j$  and  $j+1$ , i.e., we set  $\bar{Z}_j(m, n) = [Z_j(m, n) \ Z_{j+1}(m, n)]^T$ . After getting  $\hat{X}_j$ , only the component  $\hat{X}_j$  is preserved. Similarly  $\hat{X}_{j+1}$  is obtained when  $\hat{X}_{j+1}$  is estimated.

In threshold-based denoising schemes [1, 7, 11], wavelet coefficients below a threshold will be set to 0. Those insignificant data are predominantly noise and should be discarded. To merge the merits of thresholding into the LMMSE scheme, we apply LMMSE only to the coefficients above a threshold while shrinking those coefficients below the threshold to 0. In this paper the threshold is set as  $t_j = c\sigma_j$  with  $c \in [2.5, 3.5]$ .

### 3.3. Context modeling of the wavelet coefficients

To statistically model the wavelet coefficients adaptively, we classify them into several clusters. The context modeling technique, which is widely used in coding to differentiate and gather pixels which are similar but not necessarily spatially adjacent, is very appropriate for such clustering. Chang *et al* [1] has proposed the similar clustering in their threshold-based denoising scheme. By computing the context of each wavelet coefficient, they estimated its standard deviation with a collection of pixels whose context values fall into a specified field.

We define the context value of vector coefficient  $\bar{Z}_j(m, n)$  as a function of its neighbors. The weighted average of its adjacent pixels is employed. Denote by

$\bar{z}_{j\{m,n\}}(i)$ ,  $i=1, \dots, 8$  the absolute value of the neighborhood 8 elements of  $\bar{Z}_j(m, n)$ . And define

$$\bar{\mathbf{u}}_j^{m,n} = [\bar{z}_{j\{m,n\}}(1) \ \dots \ \bar{z}_{j\{m,n\}}(8)] \quad (11)$$

The context value of  $\bar{Z}_j(m, n)$  is defined as

$$\bar{C}_j(m, n) = \bar{\mathbf{u}}_j^{m,n} \bar{\mathbf{w}}_j \quad (12)$$

$\bar{\mathbf{w}}_j$  is a  $16 \times 2$  weighted matrix and it can be calculated by the least square estimate:

$$\bar{\mathbf{w}}_j = (\bar{\mathbf{U}}_j^T \bar{\mathbf{U}}_j)^{-1} \bar{\mathbf{U}}_j^T \bar{\mathbf{Y}} \quad (13)$$

$\bar{\mathbf{U}}_j$  is a  $M \cdot N \times 16$  matrix with each row being  $\bar{\mathbf{u}}_j^{m,n}$  and  $\bar{\mathbf{Y}}$  is a  $M \cdot N \times 2$  matrix with each row being  $\bar{Z}_j^T(m, n)$ . The absolute values rather than the original values are used in the context calculating because the absolute values of neighboring coefficients are much more correlated [3]. Thus the average of absolute values would yield more useful information than the original values.

With context modeling, the coefficients with similar natures can be well clustered. Context value  $\bar{C}_j(m, n)$  is a  $1 \times 2$  vector. We group  $\bar{Z}_j(m, n)$  according to  $\bar{C}_j(m, n)$ 's localization in the 2-D Euclidean space. The coordinate plane is divided into several portions and each portion contains the same number  $L$  of context value  $\bar{C}_j(m, n)$ .

Suppose  $\bar{Z}_j(m, n)$  belongs to the  $i$ th group  $\Theta_j^i$ , the covariance matrix of corresponding vector  $\bar{X}_j(m, n)$  is estimated from all the data in  $\Theta_j^i$ .

$$\hat{P}_{j,j+1} \approx \frac{1}{L} \sum_{\bar{Z}_j(m,n) \in \Theta_j^i} Z_j(m, n) \cdot Z_{j+1}(m, n) - R_{j,j+1} \quad (14)$$

The LMMSE scheme in section 3.2 is then applied to each of the data in  $\Theta_j^i$  with the calculated  $\hat{P}_{j,j+1}$ .

## 4. EXPERIMENT RESULTS

Four compactly supported wavelets constructed by Daubechies *et al* [4] were used in the experiments here. The orthogonal wavelet is denoted by  $Dau(N)$ , where  $N$  is the number of vanishing moments of the wavelet whose filter length will be  $2N$ . The biorthogonal wavelet is denoted by  $CDF(N, N')$ , where  $N$  is the vanishing moment number of analytic wavelet filter and  $N'$  is that of synthetic wavelet filter. The four wavelets employed are:  $Dau(2)$ ,  $Dau(4)$ , and  $CDF(1,3)$ ,  $CDF(2,4)$ .

The denoising schemes employed for comparison are the spatially adaptive thresholding of Chang *et al* [1] and the intrascale dependencies exploited LMMSE-like scheme of Mhçak *et al* [2]. For convenience, we denote the two

methods as  $M1$  and  $M2$ . The proposed scheme without context modeling is denoted as  $M3$  and the version with context modeling is denoted as  $M4$ .

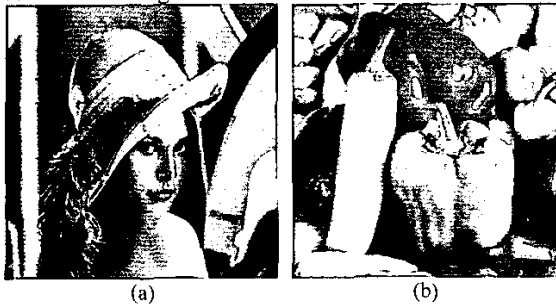


Fig. 2. Images used in the experiments. (a) Lena. (b) Peppers.

Table I. SNR (dB) results of the four methods for *Lena*

<i>Lena</i>		<i>Dau(4)</i>	<i>Dau(2)</i>	<i>CDF(1,3)</i>	<i>CDF(2,4)</i>
$\sigma = 20$	$M1$	25.28	25.61	--	--
	$M2$	25.69	25.75	25.38	25.66
	$M3$	25.69	25.71	25.54	25.32
	$M4$	25.86	25.98	26.12	25.67
$\sigma = 25$	$M1$	24.55	24.59	--	--
	$M2$	24.46	24.55	24.27	24.44
	$M3$	24.54	24.52	24.59	24.22
	$M4$	24.81	24.95	25.07	24.65
$\sigma = 30$	$M1$	23.79	23.83	--	--
	$M2$	23.40	23.55	23.30	23.36
	$M3$	23.64	23.37	23.81	23.29
	$M4$	24.03	24.15	24.29	23.84

Table II. SNR (dB) results of the four methods for *Peppers*.

<i>Peppers</i>		<i>Dau(4)</i>	<i>Dau(2)</i>	<i>CDF(1,3)</i>	<i>CDF(2,4)</i>
$\sigma = 20$	$M1$	24.15	24.33	--	--
	$M2$	24.33	24.49	24.20	24.39
	$M3$	24.52	24.71	24.76	24.27
	$M4$	24.63	24.94	25.21	24.52
$\sigma = 25$	$M1$	23.22	23.40	--	--
	$M2$	23.24	23.42	23.17	23.31
	$M3$	23.53	23.70	23.69	23.24
	$M4$	23.76	24.09	24.32	23.62
$\sigma = 30$	$M1$	22.47	22.65	--	--
	$M2$	22.21	22.44	22.20	22.27
	$M3$	22.63	22.96	22.99	22.31
	$M4$	22.98	23.31	23.56	22.82

Several benchmark images were used for the experiments, and only the results on  $512 \times 512$  images *Lena* and *Peppers* shown in Fig. 2 were reported here. Three levels of Gaussian white noise with  $\sigma = 20, 25, 30$  are added to *Lena* and *Peppers* respectively. The denoising results by the four methods are listed in Table I and Table II.

It can be seen that context modeling significantly improves the denoising performance. Algorithm  $M4$  by wavelet  $CDF(1,3)$  achieves the best results. The highpass analytic filter of  $CDF(1,3)$  has only 1 order vanishing moment and two taps. It is good at characterizing the wavelet interscale dependencies. The vanishing moment of  $CDF(1,3)$ 's synthetic filter is also high enough (3 orders) to smooth noise. As a good wavelet to exploit wavelet interscale dependencies,  $CDF(1,3)$  is very suitable for the proposed LMMSE-based denoising schemes  $M3$  and  $M4$ .

## 5. REFERENCES

- [1] S. G. Chang, B. Yu and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Processing*, vol. 9, pp. 1522-1531, Sep. 2000.
- [2] M. K. Mihçak, I. Kozintsev, K. Ramchandran and Pierre Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters*, vol. 6, pp. 300-303, Dec. 1999.
- [3] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445-3462, Dec. 1993.
- [4] I. Daubechies, *Ten lectures on wavelets*, Philadelphia, PA: SIAM, 1992.
- [5] M. Crouse, R. Nowak and R. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Processing*, vol. 42, pp. 886-902, Apr. 1998.
- [6] Y. Xu *et al*, "Wavelet transform domain filters: a spatially selective noise filtration technique," *IEEE Trans. Image Processing*, vol. 3, pp. 747-758, Nov 1994.
- [7] Q. Pan, L. Zhang *et al*, "Two denoising methods by wavelet transform," *IEEE Trans. Signal Processing*, vol. 47, pp. 3401-3406, Dec. 1999.
- [8] J. Liu and P. Moulin, "Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients," *IEEE Trans. Image Processing*, vol. 10, pp. 1647-1658, Nov. 2001.
- [9] J. Liu and P. Moulin, Image denoising based on scale-space mixture modeling for wavelet coefficients, *Proc. International Conference on Image Processing (ICIP'99)* Kobe, Japan, 1999, pp. I. 386-390.
- [10] S. D. Servetto, K. Ramchandran and M. T. Orchard, "Image coding based on a morphological representation of wavelet data," *IEEE Trans. Image Processing*, vol. 8, pp. 1161-1174, Sep. 1999.
- [11] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," in *Wavelet and Statistics*, A. Antoniadis and G. Oppenheim, Eds. Berlin, Germany: Springer-Verlag, 1995.
- [12] Lei Zhang and Paul Bao, "Edge detection by scale multiplication in wavelet domain," to appear in *Pattern Recognition Letters*, vol. 23, pp. 1771-1784, 2002.