# IMAGE DENOISING BASED ON THE WAVELET CO-OCCURRENCE MATRIX

*Zeyong Shan* and *Selin Aviyente*

Department of Electrical and Computer Engineering
Michigan State University, East Lansing, MI 48824, USA
E-mail: {shanzeyo, aviyente}@egr.msu.edu

## ABSTRACT

Image denoising is a well-known problem in signal processing. Wavelet decomposition based approaches have been applied successfully to the image denoising problem. The majority of the wavelet thresholding methods do not take the spatial correlation between the wavelet coefficients into account. In this paper, a new image denoising approach that incorporates the intra-scale dependencies between the wavelet coefficients into the thresholding algorithm is presented. The co-occurrence matrix of the wavelet coefficients and their neighbors is constructed to represent the spatial dependencies. An information-theoretic criterion, the 2-D joint entropy of the wavelet co-occurrence matrix, is used as the cost function to determine the optimal threshold. Experimental results indicate that the proposed approach yields significant improvement over the universal thresholding both in visual quality and mean square error.

## 1. INTRODUCTION

Image denoising is an important and fundamental task in many image processing applications. Over the past decade, wavelet transform has been applied successfully to the image denoising problem thanks to its energy compaction property. In particular, wavelet-based thresholding algorithms and their extensions (e.g., [1,2]) have been proposed and developed. All of these methods rely on thresholding the wavelet coefficients to minimize a given cost function in order to obtain improved visual quality.

It has been shown that the wavelet coefficients are highly correlated with each other [3-6]. This correlation, mainly caused by features such as lines, edges, and corners, arises between neighboring coefficients in a given subband as well as between coefficients corresponding to different scales and orientations. Although some work [7] has been done to include the inter- and intra-scale dependencies between the wavelet coefficients, in most wavelet thresholding methods the selection of the threshold does not take the spatial correlation between the wavelet coefficients into account. In [5], it is shown that intra-scale models capture most of the dependencies between the wavelet coefficients, and the gains obtained by including the inter-scale dependencies are marginal. In addition, taking inter-scale relationship between the wavelet coefficients into account increases the computational complexity.

In this paper, we focus on the intra-scale dependencies between the wavelet coefficients in the denoising algorithm. The two-dimensional joint probability density function known as the co-occurrence matrix [8] is used to represent the spatial correlation between the wavelet coefficients. The co-occurrence matrix is widely used in texture and image classification and segmentation [9,10]. An information-theoretic criterion, joint entropy, is used as the cost function. The threshold value is determined by applying the maximum entropy sum principle on the wavelet co-occurrence matrix.

Section 2 introduces the background on wavelet thresholding. Section 3 describes the proposed image denoising approach. Experimental results are summarized in Section 4. Finally, Section 5 gives concluding remarks and suggests some future work.

## 2. WAVELET THRESHOLDING AND COST FUNCTION SELECTION

Consider a denoising problem for a $N$ x $N$ image $\boldsymbol{X}$ corrupted by additive white Gaussian noise $\boldsymbol{\varepsilon}$ yielding a noisy image $\boldsymbol{Y}$

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{\varepsilon}. \tag{1}$$

The noise samples $\varepsilon_{ij}$ are independent and identically distributed (*iid*) Gaussian with zero mean and variance $\sigma_n^2$. A 2-D discrete wavelet transform is used to decompose the noisy image $\boldsymbol{Y}$ into the wavelet coefficients $\boldsymbol{W} =$ $\{w_{s,t}^{(k,o)} \mid s,t = 1,\cdots,N/2^k\}$ at different scales $k = 1,\cdots,J$ and orientations $o \in \{HL, LH, HH, LL\}$. The detail wavelet coefficients are then thresholded by a shrinkage operator $T_\tau$ with the pre-selected threshold $\tau$. This is based on the fact that small wavelet coefficients are more likely to be due to noise, and the large ones due to the signal. An inverse wavelet transform is employed to transform the thresholded wavelet coefficients to obtain the denoised estimate of $X$.

Most wavelet thresholding methods are derived based on the marginal distributions of the wavelet coefficients using statistical models such as the generalized Gaussian distribution (*GGD*), and thus do not take the correlation between the wavelet coefficients into account. It is known that the wavelet coefficients in a given subband are spatially correlated. This means that a large wavelet coefficient will probably have large coefficients in its neighborhood, and vice versa. Moreover, the threshold $\tau$ is usually chosen to minimize a cost function such as the mean squared error (*MSE*). The conventional cost functions are based on the second order statistics, which are not optimal for the wavelet coefficients due to their non-Gaussianity.

Therefore, we propose to use an information-theoretic criterion such as the two-dimensional joint entropy, which takes the higher order dependencies between the wavelet coefficients, as the cost function to determine the threshold:

$$H(p) = -\sum_i \sum_j p_{ij} \log p_{ij}. \tag{2}$$

In Equation (2), $p_{ij}$ refers to the joint probability of a pair of wavelet coefficients with values $i$ and $j$ occurring next to each other.

## 3. PROPOSED DENOISING APPROACH

### 3.1. Denoising based on the co-occurrence matrix of the magnitude of wavelet coefficients

In the wavelet domain, it has been observed with some consistency that the wavelet coefficients of natural images have a clustering property. In other words, a wavelet coefficient's magnitude is not independent of its neighbors. In order to take the spatial relationship between the wavelet coefficients into account, we use the co-occurrence matrix.

The elements of the co-occurrence matrix represent the number of occurrences of pairs of wavelet coefficients separated by a certain distance in a given direction. Mathematically, it is defined as:

$$\begin{aligned} C_{\theta,x}(i,j) = card\{((s,t),(u,v)) : I(s,t) = i, I(u,v) = j, \\ d((s,t),(u,v)) = x, \tan^{-1}((s,t)-(u,v)) = \theta\} \end{aligned} \tag{3}$$

where, $card\{\cdot\}$ is the cardinality of a set, $I(\cdot,\cdot)$ denotes an image of size *N* x *N* with *L* gray levels, $d(\cdot,\cdot)$ is a distance measure, $(s,t),(u,v) \in N$ x $N$, and $\theta, x$ are the angle and the displacement between $(s,t)$ and $(u,v)$ respectively. In general, $\theta$ is taken to be $0^0, 45^0, 90^0$, and $135^0$.

In the wavelet denoising scheme, the noisy image is transformed into the wavelet domain $\{w_{s,t}^{(k,o)}\}$. The magnitudes of the wavelet coefficients $\{w_{st}\}$ [1] in each detail subband are then quantized by a uniform scalar quantizer with $L+1$ levels:

$$Q(w_{st}) = round\left(\frac{|w_{st}|}{\Delta}\right), \tag{4}$$

where, $\Delta = \dfrac{\max(|w_{st}|)}{L}$ is the step size. Note that the choice of the value of $L$ is important. If $L$ is too small, then some wavelet coefficients corresponding to the signal and noise are quantized to the same level, which makes it hard to distinguish this part of the signal from noise. On the other hand, if $L$ is too large, then each wavelet coefficient is quantized to a different level. This leads to a sparse co-occurrence matrix, which makes the entropy computation unreliable.

From the quantized wavelet coefficients, the co-occurrence matrix $C_{\theta,x}(i,j)$ is constructed. The entry $p_{ij}$ of the normalized co-occurrence matrix gives us the probability of having two wavelet coefficients at levels $i$ and $j$ located at a distance $x$ and a direction $\theta$:

$$p_{ij} = \frac{C_{\theta,x}(i,j)}{\sum_i \sum_j C_{\theta,x}(i,j)}. \tag{5}$$

If $\tau$, $0 \le \tau \le L$, is the threshold, then $\tau$ partitions the normalized co-occurrence matrix into four quadrants: I, II, III, and IV as shown in Fig. 1. It is known that the magnitudes of the wavelet coefficients corresponding to noise are smaller than those corresponding to the signal.

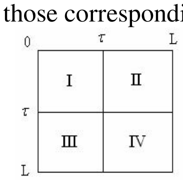


**Fig. 1.** Quadrants of the normalized co-occurrence matrix

The probabilities corresponding to these four quadrants are defined as follows:

$$P_I = \sum_{i=0}^{\tau} \sum_{j=0}^{\tau} p_{ij}, \qquad P_{II} = \sum_{i=0}^{\tau} \sum_{j=\tau+1}^{L} p_{ij}, \tag{6}$$

$$P_{III} = \sum_{i=\tau+1}^{L} \sum_{j=0}^{\tau} p_{ij}, \qquad P_{IV} = \sum_{i=\tau+1}^{L} \sum_{j=\tau+1}^{L} p_{ij}. \tag{7}$$

Normalizing the probabilities within each individual quadrant such that the sum of the probabilities of each

[1] To simplify notation, the superscript $(k,o)$ is dropped and will be used only when necessary for clarity.

quadrant equals one, we get the following cell probabilities for these four different quadrants:

$$p_{ij}^{I} = \frac{p_{ij}}{P_I}, \qquad 0 \le i, j \le \tau \tag{8}$$

$$p_{ij}^{II} = \frac{p_{ij}}{P_{II}}, \qquad 0 \le i \le \tau, \tau+1 \le j \le L \tag{9}$$

$$p_{ij}^{III} = \frac{p_{ij}}{P_{III}}, \qquad \tau+1 \le i \le L, 0 \le j \le \tau \tag{10}$$

and

$$p_{ij}^{IV} = \frac{p_{ij}}{P_{IV}}, \qquad \tau+1 \le i, j \le L. \tag{11}$$

Since small and large wavelet coefficients correspond to noise and the signal respectively, Quadrant I represents noise and Quadrant IV the signal. Quadrants II and III are ignored, because for most natural images the off-diagonal probabilities are negligible.

Therefore, the 2-D joint entropy of noise can be defined as

$$H_I(\tau) = -\sum_{i=0}^{\tau}\sum_{j=0}^{\tau} p_{ij}^{I} \log p_{ij}^{I}, \tag{12}$$

and, the 2-D joint entropy of the signal can be written as

$$H_{IV}(\tau) = -\sum_{i=\tau+1}^{L}\sum_{j=\tau+1}^{L} p_{ij}^{IV} \log p_{ij}^{IV}. \tag{13}$$

The total entropy of the partitioned wavelet coefficients in each detail subband is

$$H_{All}(\tau) = H_I(\tau) + H_{IV}(\tau). \tag{14}$$

The total entropy should be maximized to make the wavelet coefficients in each quadrant homogenous. The homogeneity of the coefficients will ensure a good separation of the signal and noise. Hence, we maximize $H_{All}(\tau)$ with respect to $\tau$ to get the optimal threshold $\tau^*$. Finally, all of the wavelet coefficients in a given detail subband are soft-thresholded by $T_\tau = \Delta \cdot \tau^*$ and inverse transformed to obtain the denoised image.

### 3.2. Threshold selection based on the co-occurrence matrix of wavelet coefficients and their neighbors

In Subsection 3.1, the co-occurrence matrix is constructed based on the magnitude of the wavelet coefficients. However, the magnitude is often not sufficient to represent the spatial correlation between the wavelet coefficients. Therefore, it is common to incorporate additional spatial information in the co-occurrence matrix. There are various approaches for incorporating local information into the co-occurrence matrix such as the mean of the neighboring coefficients.

In a given detail subband, the mean of the $M$ x $M$ neighborhood of each wavelet coefficient is calculated in advance. The magnitudes of the wavelet coefficients and the corresponding neighborhood means are quantized with $L+1$ levels and $S+1$ levels respectively to construct the co-occurrence matrix. The entry of the co-occurrence matrix represents how often different values of a wavelet coefficient and a neighborhood mean occur in a given distance and orientation.

A threshold vector $(\tau, \gamma)$ is determined using the maximum entropy principle. The two orthogonal lines intersecting at $(\tau, \gamma)$ divide the normalized co-occurrence matrix into four quadrants as illustrated in Fig. 2.

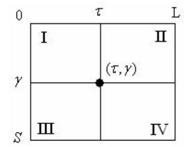


**Fig. 2.** Quadrants produced by a threshold vector

Similar to Subsection 3.1, the threshold vector $(\tau^*, \gamma^*)$ is selected to maximize the sum of the joint entropy of the signal and noise.

## 4. EXPERIMENTAL RESULTS

Two images *Lena* and *Cameraman* of size 256 x 256 are used as test images in order to evaluate the performance of our approach. The proposed approach is compared with *VisuShrink* [1]. *VisuShrink* is a well-known universal soft-thresholding denoising method.

In the experiments, the distance parameter $x$ of the co-occurrence matrix $C_{\theta,x}$ is chosen to be equal to 1. The performance of our denoising approach based on the magnitude of wavelet coefficients is compared for the different directions $\theta$. The *PSNR* results are shown in Table 1 with the best one underlined.

**Table 1.** *PSNR* results for different angles and $\sigma_n$

| $\sigma_n / \theta$ | $0^0$ | $45^0$ | $90^0$ | $135^0$ |
|---|---|---|---|---|
| PSNR | *Lena* | | | |
| 10 | 27.81 | 28.06 | 27.93 | <u>28.24</u> |
| 20 | 25.68 | 25.84 | 25.82 | <u>25.98</u> |
| 30 | 23.42 | 23.46 | 23.43 | <u>23.57</u> |
| | *Cameraman* | | | |
| 10 | <u>24.13</u> | 23.74 | 22.95 | 23.68 |
| 20 | 23.14 | 23.38 | 22.80 | <u>23.48</u> |
| 30 | 21.68 | 21.94 | 21.83 | <u>22.02</u> |

From Table 1, it is seen that the performance of the co-occurrence matrix with four different angles is approximately the same. Although the performance with $\theta = 135^0$ is the best for *Lena* image, this is not a universal result. In practice, the choice of the direction depends on the image.

Next, the performance of our approach is compared to *VisuShrink*. The method based on the magnitude of the wavelet coefficients is named *CocurMagn*, and the one based on the wavelet coefficients and their neighbors is named *CocurNeigh*. The direction $\theta$ is set to $0^0$, and the dimension of the neighborhood is set to 3 by 3. Table 2 gives the experimental results.

**Table 2.** *PSNR* results for different methods

| Methods/$\sigma$ | 10 | 20 | 30 |
|---|---|---|---|
| PSNR | ***Lena*** | | |
| *VisuShrink* | 27.63 | 25.32 | 22.91 |
| *CocurMagn* | 28.23 | 25.69 | 23.58 |
| *CocurNeigh* | 28.74 | 26.13 | 24.05 |
| | ***Cameraman*** | | |
| *VisuShrink* | 23.44 | 21.08 | 19.75 |
| *CocurMagn* | 23.83 | 23.05 | 21.81 |
| *CocurNeigh* | 24.06 | 23.57 | 22.19 |

It is concluded from Table 2, Figs. 3 and 4 that the two methods, *CocurMagn* and *CocurNeigh*, are always superior to the universal thresholding in *MSE* and visual quality with *CocurNeigh* being the best one.

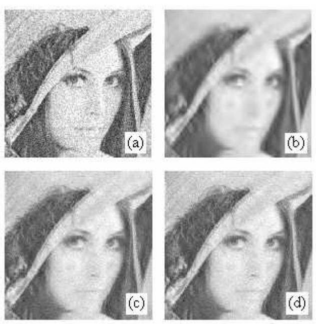


**Fig. 3.** Comparison of denoising results on *Lena* image. (a) Noisy observation (PSNR = 22.07 dB), (b) *VisuShrink* (PSNR = 25.32 dB), (c) *CocurMagn* (PSNR = 25.69 dB), (d) *CocurNeigh* (PSNR = 26.13 dB).


**Fig. 4.** Comparison of denoising results on *Cameraman* image. (a) Noisy observation (PSNR = 22.04 dB), (b) *VisuShrink* (PSNR = 21.08 dB), (c) *CocurMagn* (PSNR = 23.05 dB), (d) *CocurNeigh* (PSNR = 23.57 dB).

## 5. CONCLUSIONS AND FUTURE WORK

In this paper, a new image denoising approach is proposed using the co-occurrence matrix to characterize the intra-scale dependencies between the wavelet coefficients. Since the wavelet coefficients do not follow a Gaussian model, an information-theoretic criterion is employed instead of the conventional cost functions such as *MSE*. The application of the presented approach on test images shows that it is simple to implement, effective, and outperforms the classical soft-thresholding algorithms.

The proposed denoising method can be further improved by exploring the extraction of different spatial features from the wavelet coefficients to construct the co-occurrence matrix, and the different partitions of the co-occurrence matrix such as a threshold line which may provide a better decision boundary between the signal and noise.

## 6. REFERENCES


[1] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, no. 3, pp. 425-455, 1994.

[2] S. G. Chang, B. Yu, and M. Vetterli, “Adaptive wavelet thresholding for image denoising and compression,” *IEEE Trans. on Image Processing*, vol. 9, no. 9, pp. 1532-1546, Sept. 2000.

[3] R. W. Buccigrossi and E. P. Simoncelli, “Image compression via joint statistical characterization in the wavelet domain,” *IEEE Trans. on Image Processing*, vol. 8, no. 12, pp. 1688-1701, Dec. 1999.

[4] E. P. Simoncelli, “Statistical models for images: compression, restoration and synthesis,” *31st Asilomar Conf. on Signals, Systems and Computers*, pp. 673-678, Pacific Grove, CA, Nov. 1997.

[5] J. Liu and P. Moulin, “Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients,” *IEEE Trans. on Image Process*ing, vol. 10, no. 11, pp. 1647-1658, Nov. 2001.

[6] L. Sendur and I. W. Selesnick, “Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency,” *IEEE Trans. on Signal Processing*, vol. 50, no. 11, pp. 2744-2756, Nov. 2002.

[7] S. G. Chang, B. Yu, and M. Vetterli, “Spatially adaptive wavelet thresholding with context modeling for image denoising,” *IEEE Trans. on Image Processing*, vol. 9, no. 9, pp. 1522-1531, Sept. 2000.

[8] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610-621, Nov. 1973.

[9] E. D. Jansing, T. A. Albert, and D. L. Chenoweth, “Two-dimensional entropic segmentation,” *Pattern Recognition Letters*, vol. 20, pp. 329-336, 1999.

[10] A. D. Brink, “Thresholding of digital images using two-dimensional entropies,” *Pattern Recognition*, vol. 25, no. 8, pp. 803-808, 1992.