

Statistical Wavelet-based Image Denoising using Scale Mixture of Normal Distributions with Adaptive Parameter Estimation

M. Saeedzarandi* H. Nezamabadi-pour and S. Saryazdi

Intelligent Data Processing Laboratory (IDPL), Department of Electrical Engineering, Shahid Bahonar University of Kerman, Kerman, Iran.

Received 08 December 2018; Revised 05 January 2020; Accepted 04 February 2020

*Corresponding author: m_saeedzarandi@eng.uk.ac.ir (M. Saeedzarandi).

Abstract

Removing noise from images is a challenging problem in digital image processing. In this paper, we present an image denoising method based on a maximum a posteriori (MAP) density function estimator, which is implemented in the wavelet domain due to its energy compaction property. Performance of the MAP estimator depends on the proposed model for noise-free wavelet coefficients. Thus in the wavelet-based image denoising, selecting a proper model for wavelet coefficients is very important. In this paper, we model wavelet coefficients in each sub-band by heavy-tail distributions that are from scale mixture of the normal distribution family. The parameters of distributions are estimated adaptively to model the correlation between the coefficient amplitudes so the intra-scale dependency of wavelet coefficients is also considered. The denoising results obtained confirm the effectiveness of the proposed method.

Keywords: *Image Denoising, Wavelet Transform, MAP Estimator, Heavy-tail Distributions, Scale Mixture of Normal Distributions.*

1. Introduction

Digital images play an important role in many applications such as satellite television, intelligent traffic monitoring, and signature validation as well as in the areas of research such as geographical information systems and astronomy [1]. Noise can be introduced into the image while capturing or transmitting [2]; therefore, denoising is one of the most important tasks in image processing. Reserving the details of an image and removing the noise as far as possible is the aim of image denoising processes [3-5]. In camera systems, the noise comes from different sources like thermal noise, quantization noise, and photon noise. Thermal noise is due to analog circuits and has a zero-mean Gaussian distribution, and it is the most common noise encountered in different applications. In this work, the additive white Gaussian noise was considered as the noise model. The noise removal can be performed in both the transform and spatial domains [4-8]. The spatial denoising methods such as local and non-local algorithms use the similarity between pixel or patches in an image, and the transform domain-

based methods exploit the similarity of transform coefficients. In a transform domain, the smaller coefficients are related to the high-frequency part of the image containing the noise and image details. By adjusting these small coefficients of the noisy image, the reconstructed image could have less noise.

A wide amount of image processing algorithms are constructed based on wavelet transform due to its energy compaction property [6, 7]. In fact, in the wavelet domain, small coefficients are more likely due to noise, and large coefficients represent important image features (such as edges) [7-9]. Most of the denoising algorithms are based on the thresholding methods, in which each coefficient is compared with a proper threshold; if it is smaller than the threshold, the coefficient is set to zero; otherwise, it is kept (hard thresholding) or modified (soft thresholding). Visueshrink [10], Sureshrink [11, 12], and Bayesshrink [13, 14] are examples of the algorithms that use the thresholding methods for denoising.

The statistical wavelet-based image denoising has been a fundamental tool in image processing analysis over the recent decades, and is still an expanding research area [15-18]. In this work, we modeled the data as the sum of the clean image plus additive white Gaussian noise, and we presented a parametric statistical method in the wavelet domain to estimate clean coefficients from noisy data by a maximum a posteriori (MAP) estimator. The MAP estimator has been widely used in wavelet-based image denoising.

The success of the MAP estimator strongly depends on the employed distribution of clean wavelet coefficients as well as the probability density function of noise [17, 18].

Since natural images contain smooth areas interspersed with occasional sharp transmissions such as edges, the wavelet transform of images is sparse so the marginal distributions of wavelet coefficients are heavy-tail distributions and have a large peak at zero [19, 20]. Various probability distribution functions (PDFs) have been proposed on modeling clean wavelet coefficients. For example, Michak has used a Gaussian PDF [21] that is the simplest probability density, and Rabbani has proposed the Laplacian and Cauchy PDFs for modeling noise-free wavelet coefficients [17, 22]. Still, these distributions do not have enough flexibility to model the wavelet coefficients. Mallat has used generalized Gaussian distribution as a priority for the wavelet coefficients in image processing [23], and the authors in [24] have used α -stable that is a heavy-tail distribution as a priority to capture the sparseness of the wavelet coefficients. Fadili et al. [15] have proposed the Bessel k form density as a prior model and have demonstrated that this distribution is better than the generalized Gaussian and α -stable distributions.

Despite the many advances mentioned above, it is yet a challenging problem to characterize the probability distribution function of wavelet coefficients accurately. The student-t and Slash distributions are heavy-tail distributions that have found many applications in robust statistical data analysis [25, 26].

These distributions, which can be expressed as a scale mixture of normal distributions, have been recently used in machine learning and signal processing problems [27, 28]. Since these distributions have heavy-tails, they are suitable to model the sparseness of wavelet coefficients properly, and we use these distributions as the prior probability density of clean wavelet coefficients in this work.

In practical problems, parameters of the distributions are unknown; in this work, we estimated the parameters adaptively to use the empirically observed correlation between the neighborhood coefficients. The simulation results show the superiority of the proposed methods among the best univariate methods and even some bivariate denoising models.

In this paper, bold-face letters refer to matrices and italic lower-case letters refer to matrix elements. The rest of this paper is organized as follows. In Section 2, after a brief review of the wavelet transform and the basic idea of MAP estimator, we will study the scale mixture of Gaussian distributions, student-t, and Slash as the prior signal distributions. The hyper-parameter estimation is introduced in Section 3. In Section 4, we use our model for wavelet-based denoising of several images corrupted by additive white Gaussian noise in various noise levels, and we compare our results with several published denoising algorithms. Finally, this paper is concluded in Section 5.

2. Statistical wavelet-based image denoising

2.1. Block diagram of denoising process

Similar to many reported works in the field of wavelet-based image denoising, our image denoising model uses three levels: a) decomposing the noisy image into sub-bands at different orientations and scales through wavelet transform, b) denoising each high-pass sub-band, c) applying the inverse wavelet transform to get the estimated noise-free image. The block diagram of the denoising process is depicted in figure 1.

In this work, we considered the case of image denoising corrupted by additive white Gaussian noise:

$$\mathbf{g} = \mathbf{d} + \mathbf{e}, \tag{1}$$

Here, \mathbf{g} , \mathbf{d} , and \mathbf{e} are the noisy image, the original image, and the independent white Gaussian noise, respectively.

Let $\mathbf{y} = \omega(\mathbf{g})$, $\mathbf{x} = \omega(\mathbf{d})$, $\mathbf{n} = \omega(\mathbf{e})$, where ω presents the two-dimensional orthogonal wavelet transform (DWT) operator [29]. By applying the wavelet transform, the image will be decomposed into the sub-bands HH_j , HL_j , and LH_j , $j = 1, 2, \dots, J$ correspond to the detail coefficients in the diagonal, horizontal, and vertical orientations, and LL_j is the low-pass sub-band. It can be concluded from (1) that:

$$y_{m,l}^{\alpha,j} = x_{m,l}^{\alpha,j} + n_{m,l}^{\alpha,j}, \tag{2}$$

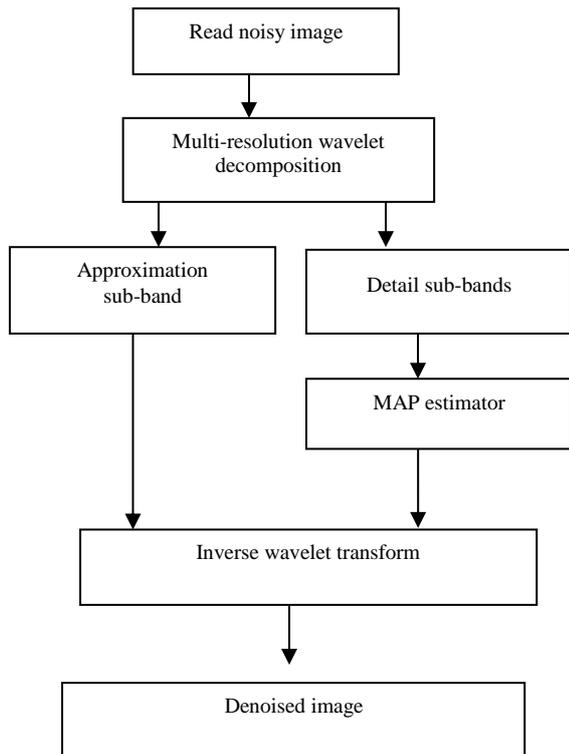


Figure 1. Denoising process.

where, $y_{m,l}^{o,j}$ is the detail coefficient of the noisy image at orientation o , scale j , and location (m, l) , and similarly, for $x_{m,l}^{o,j}$ and $n_{m,l}^{o,j}$ (for original image and noise, respectively). The denoising algorithm was applied to all sub-bands (except the low-pass sub-band) independently.

2.2. MAP estimator

As aforementioned, we consider the case of denoising an image corrupted by white Gaussian noise. Based on (2), in the wavelet domain, we have $y = x + n$, where y is the noisy wavelet coefficient, x is the clean wavelet coefficient that we wish to estimate it as well as possible, and n is the independent white zero-mean Gaussian noise. The standard MAP estimator for x given the corrupted observation y is [30]:

$$\hat{x} = \operatorname{argmax} p(x/y), \quad (3)$$

where, p denotes the density function. Using the Bayesian rule, we have [31]:

$$\hat{x} = \operatorname{argmax} p(x/y) = \operatorname{argmax} \frac{p(y/x)p(x)}{p(y)}, \quad (4)$$

Since $p(y)$ is independent from x , we can omit the denominator of the right-hand side of (4); therefore, we have:

$$\hat{x} = \operatorname{argmax} p(y/x)p(x), \quad (5)$$

Because $y = x + n$, $p(y/x)$ is the PDF of normal distribution, we have:

$$\hat{x} = \operatorname{argmax} p_n(y - x)p(x), \quad (6)$$

As we can see, the estimation depends on the prior probability of signal and the PDF of noise. In figure 2, a 256-bin wavelet coefficient histogram for the horizontal sub-band of three images is plotted. As the figure shows, the histograms have a large peak around zero and with heavy tails. Therefore, in this paper, we used distributions that were heavy-tail to model the sparsity property of wavelet transform.

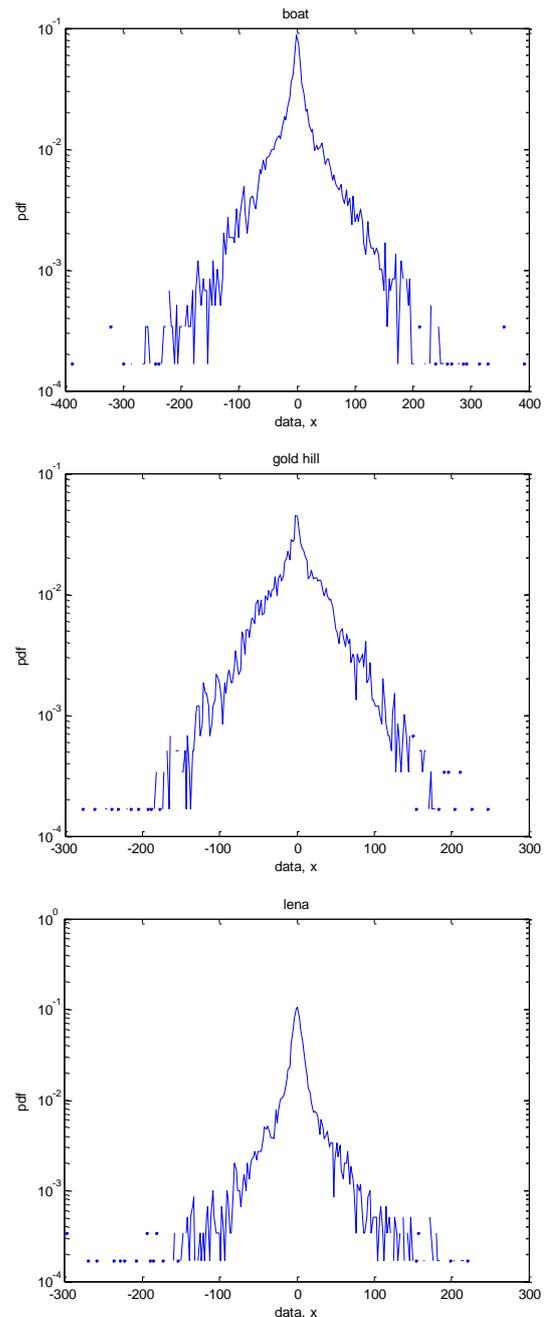


Figure 2. Histograms of horizontal sub-band of wavelet transform (one level decomposition) for three natural images.

2.3. Scale mixture of normal distributions

Suppose that T has a standard normal distribution. Then the distribution of:

$$X = \mu + \sigma U^{-1}T, \quad (7)$$

is referred to as a scale mixture of normal distributions, where μ is the location parameter, σ is the scale parameter, U is a random variable on $(0, \infty)$ called the scale factor with density function $p(0, \nu)$, and ν is a parameter indexing the distribution of U . The names of these classes of distributions become clear when we note the conditional distribution of X , given that $U = u$ is normal. The scale mixture of normal distribution has heavier tails than the normal distribution, and has been used in several areas such as robust inference.

According to the above, the scale mixture of normal distributions can be expressed as [32]:

$$p(x) = \int p(x|u) p(u) du, \quad (8)$$

where, $p(\cdot|u)$ is the conditional density of random variable x given $U = u$ as follows:

$$p(x|u) = \frac{\sqrt{u}}{\sqrt{2\pi}\sigma} e^{-\frac{u}{2\sigma^2}(x-\mu)^2}. \quad (9)$$

It is the normal density with mean μ variance $\sigma^2 u^{-1}$. Different distributions for $p(u)$ result in a rich class of heavy-tail distributions. In this paper, we use heavy-tail distributions such as the student-t and Slash distributions that have a computationally attractive form [32, 33].

Student-t distribution:

In (8), if we consider $u: \text{gamma}(\nu/2, \nu/2)$, where $\text{gamma}(a, b)$ is the *gamma* distribution with mean a/b , the density of x will be:

$$p(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}\sigma} \left(1 + \frac{d}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}, \quad (10)$$

where:

$$d = \frac{(x-\mu)^2}{\sigma^2}.$$

and ν is the degree of freedom parameter, σ is the scale parameter, and μ is the location parameter.

The variance and kurtosis of this distribution are:

$$\text{var}(x) = \sigma^2 \frac{\nu}{\nu-2}, \quad (11)$$

$$\text{kurtosis}(x) = \frac{6}{\nu-4} + 3. \quad (12)$$

Kurtosis is a statistical measure that defines how heavily the tails of a distribution differ from the tails of the normal distribution. Distributions with kurtosis greater than the kurtosis of the normal distribution are said to be leptokurtic, which has tails that asymptotically approach zero more slowly than a Gaussian distribution.

Slash distribution:

In (8), if we consider $p(u) = \nu u^{(\nu-1)}, u \in (0,1)$, the distribution of x will be [34]:

$$p(x) = \int_0^1 \nu u^{(\nu-1)} \phi(x, \mu, u^{-2}\sigma^2) du, \quad (13)$$

where, ϕ is the univariate normal distribution, ν is the shape parameter, σ is the scale parameter, and μ is the location parameter. The variance and kurtosis of Slash distribution are:

$$\text{var}(x) = \sigma^2 \frac{\nu}{\nu-2}, \quad (14)$$

$$\text{kurtosis}(x) = \frac{3(\nu-2)^2}{\nu(\nu-4)}. \quad (15)$$

3. Hyper-parameter estimation

The standard deviation of noise (σ_n) could be estimated using a robust median estimator from *HH* sub-band at the finest scale of wavelet transform, as follows [15]:

$$\hat{\sigma}_n = \frac{\text{MAD}(x^{HH})}{0.6745}, \quad (16)$$

where, *MAD* is the median absolute deviation.

For the student-t and Slash distributions, we should estimate the parameter, degrees of freedom (shape parameter for Slash distribution) that will be estimated in each sub-band, and the scale parameter σ that will be estimated using neighborhood pixels in a window in which the pixel to be denoised is in its center.

The parameter ν of the student-t distribution can be estimated using the second- and fourth-order cumulants (k_2 and k_4) of x as follows (see Appendix I):

$$\text{var}(x) = k_2, \quad (17)$$

$$\text{kurtosis}(x) = \frac{k_4}{k_2^2} + 3. \quad (18)$$

In order to estimate k_2 and k_4 , we use the k unbiased statistic estimator, which has a minimum variance among all the other unbiased estimators [35, 36]:

$$\hat{k}_2 = \frac{n}{n-1} \hat{M}_2, \quad (19)$$

$$\hat{k}_4 = \frac{n^2 \left((n+1)\hat{M}_4 - 3(n-1)\hat{M}_2^2 \right)}{(n-1)(n-2)(n-3)}, \quad (20)$$

where, n is the number of samples, and M_2 and M_4 are the 2nd and 4th sample central moments, respectively. When two or more random variables are statistically independent, the n^{th} -order cumulant of their sum is equal to the sum of their n^{th} -order cumulants. Also the third- and higher-order cumulants of a normal distribution are zero, and it is the only distribution with this property. According to the above and using (2), we have:

$$k_{2,y} = \sigma_y^2 = \text{var}(x) + \sigma_n^2. \quad (21)$$

Finally, using (12), (18), and (21) for the student-t distribution, we will drive the following estimation of v :

$$\frac{\hat{k}_4}{(\hat{k}_2 - \sigma_n^2)^2} + 3 = \frac{6}{\hat{v} - 4} + 3. \quad (22)$$

For estimating the parameter σ , using (11), (21), and (22), we have:

$$\hat{\sigma}^2 = \begin{cases} (\hat{\sigma}_y^2 - \hat{\sigma}_n^2) \frac{\hat{v} - 2}{\hat{v}} & \text{if } \hat{\sigma}_y^2 \geq \hat{\sigma}_n^2 \\ 0 & \text{otherwise} \end{cases}, \quad (23)$$

where, $\hat{\sigma}_y^2$ is computed empirically by:

$$\hat{\sigma}_y^2 = \frac{1}{M} \sum_{y_j \in R} y_j^2, \quad (24)$$

where, M is the size of the neighborhood R .

Estimation of the parameter σ of wavelet coefficients based on a local window will lead to a more robust estimation. In fact, due to the locality property of wavelet transform, if a particular wavelet transform is small/large, then the adjacent spatial coefficients are also very likely to be small/large.

For the Slash distribution, the hyper-parameter estimation method is the same. After estimating the parameters of the proposed distribution, the denoised coefficients will be achieved using (6). In this paper, we use the "fminbnd" Matlab function that uses the golden section and the parabolic interpolation methods [41] to maximize the MAP estimator.

4. Simulation results

In order to show the effectiveness of the proposed model, some experiments were performed, and the results obtained were compared with some well-known denoising algorithms. We evaluated the denoising algorithm on "Lena," "Barbara," "Boat," "Bridge," "Mandrill," and "Goldhill" 512×512

standard images with different noise levels of standard deviations $\sigma_n = 10, 20, 30$. The original images are shown in figure 3.

We used the Daubechies' length eight wavelet transform. The number of scales was set to 3 for low-level noise ($\sigma_n = 10$) and 4 for high-level noise ($\sigma_n = 20, 30$) because, according to our experiments, any further decomposition beyond these levels does not improve the denoising performance significantly.

The performance analysis is performed using the PSNR and MSSIM measures. The PSNR measure is given by:

$$PSNR = 10 \log \left(\frac{255^2}{MSE} \right), \quad (25)$$

where, MSE is the minimum square error given by:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - k(i, j)]^2, \quad (26)$$

where, I and k denote the noise-free and the denoised images, respectively, and m and n are the number of columns and rows of the image.

Let $\mathbf{x} = \{x_i | i = 1, 2, \dots, N\}$ and $\hat{\mathbf{x}} = \{\hat{x}_i | i = 1, 2, \dots, N\}$ be two non-negative image signals. SSIM is defined as follows:

$$SSIM(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x \mu_{\hat{x}} + C_1)(2\sigma_{x\hat{x}} + C_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + C_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2)}, \quad (27)$$

where:

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i, \quad (28)$$

$$\mu_{\hat{x}} = \frac{1}{N} \sum_{i=1}^N \hat{x}_i, \quad (29)$$

$$\sigma_x^2 = \left(\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2 \right)^{1/2}, \quad (30)$$

$$\sigma_{\hat{x}}^2 = \left(\frac{1}{N-1} \sum_{i=1}^N (\hat{x}_i - \mu_{\hat{x}})^2 \right)^{1/2}, \quad (31)$$

$$\sigma_{x\hat{x}} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(\hat{x}_i - \mu_{\hat{x}}), \quad (32)$$

and the constants C_1 and C_2 are included to avoid instability when $\mu_x^2 + \mu_{\hat{x}}^2$ and $\sigma_x^2 + \sigma_{\hat{x}}^2$ are very close to zero. Specifically:

$$C_1 = (k_1 L)^2, \quad (33)$$

$$C_2 = (k_2 L)^2, \quad (34)$$

where, L is the dynamic range of the pixel values (255 for 8-bit grayscale images), and k_1 and k_2 are small constants.

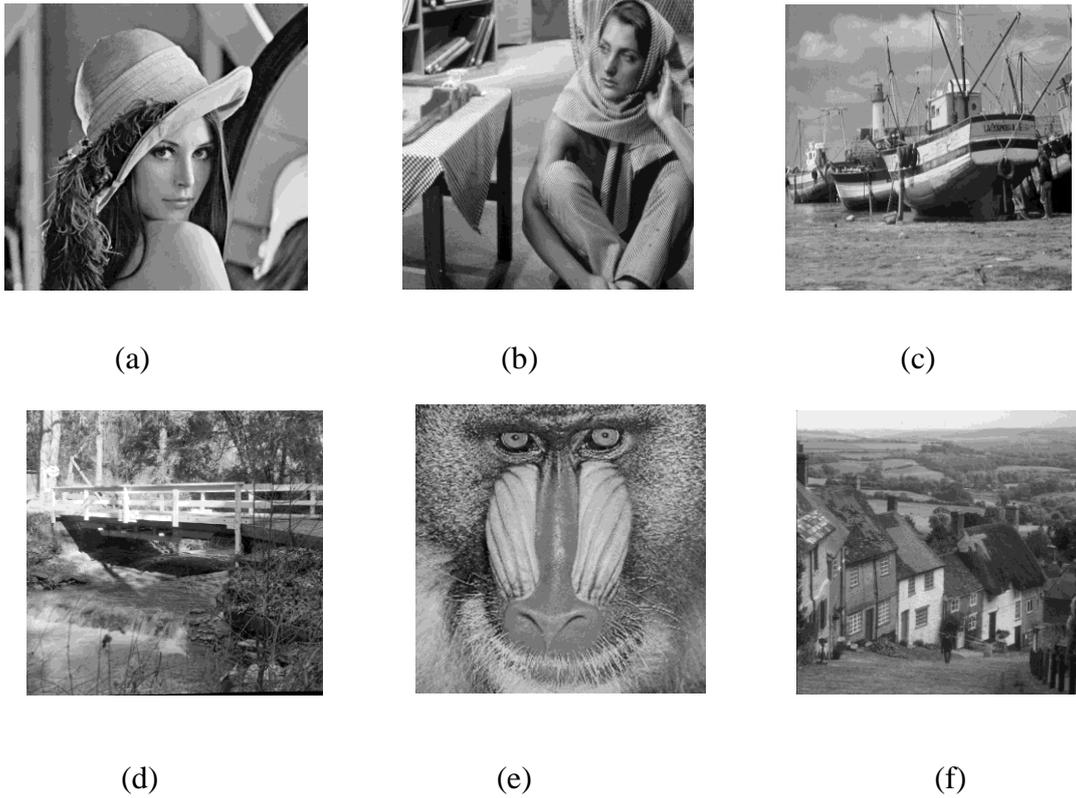


Figure 3. Test images used for denoising (a) Lena, (b) Barbara, (c) Boat, (d) Bridge, (e) Mandrill, (f) Goldhill.

In practice, mean SSIM (MSSIM) is used to measure the entire image quality:

$$MSSIM = \frac{1}{M} \sum_{i=1}^M SSIM(X_i, \hat{X}_i), \quad (35)$$

where, X_i and \hat{X}_i are the image contents at j th local window and M is the number of local windows of image. An MSSIM equal to 1 indicates that X and \hat{X} are the same. It can be inferred from $MSSIM = 0$ that X and \hat{X} are unrelated.

In order to determine the proper window size, we evaluated the denoising method in the five window sizes of 3×3 , 5×5 , 7×7 , 9×9 , and 11×11 . White Gaussian noise with standard deviation $\sigma_n = 10, 20, 30$ was added to the standard images of “Lena,” “Barbara,” and “Boat,” and the denoising model was employed to enhance the noisy images for various window sizes.

The results for the student-t and Slash distributions by averaging over five independent runs are represented in table 1. The results in this table show that a small window size produces a lower PSNR value. On the other hand, increasing the window size without

limitation does not improve the denoising performance.

Also when the noise level increases, the PSNR value is improved by increasing the window size.

We also compared our model with some related denoising algorithms such as Visushrink [10], Sureshshrink [11], Bayesshrink [13], Hidden Markov tree [37], Sendur’s work [38], bivariate Cauchy [17], bivariate shrink [39], mixture of three Laplace PDFs (3-lap-mix) [31], and Michak’s work (LAWMAP) [21] and a mixture of Gaussian PDF [40] that are the best univariate and bivariate statistical wavelet denoising methods reported in the literature.

Table 2 signifies the results obtained. Each PSNR value is averaged over five independent runs of the algorithms. The highest PSNR value is shown with boldface style.

The results obtained show that the proposed methods outperform the competing univariate methods and even some bivariate denoising algorithms in the employed test images.

Table 1. Impact of window size on the denoising performance.

Noise standard deviation	PSNR	σ_n	Student-t distribution					Slash distribution				
			3 × 3	5 × 5	7 × 7	9 × 9	11 × 11	3 × 3	5 × 5	7 × 7	9 × 9	11 × 11
Lena	10		34.36	34.59	34.54	34.50	34.44	34.25	34.60	34.58	34.55	34.47
	20		30.70	31.27	31.32	31.28	31.25	30.33	31.13	31.28	31.29	31.20
	30		28.42	29.26	29.42	29.39	29.37	27.76	29.00	29.28	29.31	29.29
Boat	10		32.59	32.62	32.56	32.50	32.45	32.63	32.72	32.65	32.58	32.51
	20		28.98	29.30	29.33	29.25	29.23	28.83	29.24	29.29	29.27	29.21
	30		26.81	27.34	27.44	27.38	27.23	26.49	27.22	27.38	27.32	27.32
Barbara	10		32.67	32.73	32.65	32.56	32.48	32.76	32.88	32.80	32.67	32.58
	20		28.64	28.84	28.99	28.81	28.74	28.55	28.92	28.95	28.89	28.85
	30		26.29	26.77	26.87	26.76	26.71	26.10	26.76	26.84	26.83	26.81

Table 2. Comparison of performance for image denoising using different algorithms and orthogonal wavelet transform

PSNR	σ_n	Visu	Sure	Bayes	HMT	Sedurs	Bivariate- Cauchy	Bivariate- shrink	3mix-lap	Lawmap	GaussMix- shrinkl	Student-t adaptive	Slash adaptive
		[10]	[11]	[13]	[37]	[38]	[17]	[39]	[31]	[21]	[40]		
Lena	10	28.76	33.42	33.38	33.84	33.94	34.30	34.36	33.63	34.28	34.43	34.59	34.60
	20	26.59	30.22	30.23	30.39	30.73	30.98	31.19	30.24	30.92	30.97	31.32	31.29
	30	25.60	28.47	28.58	28.35	28.91	28.94	29.41	28.75	29.05	28.89	29.42	29.31
Boat	10	26.64	31.63	31.81	32.28	32.25	32.50	32.42	31.99	32.25	32.68	32.62	32.72
	20	24.59	28.55	28.48	28.84	28.93	29.06	29.18	28.63	28.97	29.15	29.33	29.29
	30	23.63	26.73	26.65	26.83	27.11	27.13	27.29	26.84	27.03	27.11	27.44	27.38
Barbara	10	24.81	30.21	30.86	30.86	31.13	32.32	32.25	31.43	31.99	32.60	32.73	32.88
	20	22.81	25.91	27.13	27.80	27.25	28.40	28.36	27.30	27.94	28.60	28.99	28.95
	30	22.00	24.33	25.16	25.11	25.21	26.20	26.28	25.18	25.80	26.43	26.87	26.84

Table 3. PSNR and MSSIM comparison of performance for image denoising in the orthogonal wavelet transform domain.

PSNR/MSSIM	σ_n	Bivariate- shrink [39]	Student-t adaptive	Slash adaptive
Bridge	10	29.51/0.96	29.75/0.96	29.97/0.96
	20	26.01/0.89	26.41/0.90	26.56/0.90
	30	24.32/0.81	24.76/0.83	24.73/0.83
Mandrill	10	31.66/0.94	32.09/0.96	32.15/0.96
	20	27.59/0.89	28.15/0.90	28.13/0.90
	30	25.50/0.82	26.09/0.84	26.06/0.83
Goldhill	10	32.27/0.94	32.48/0.94	32.60/0.94
	20	29.12/0.87	29.40/0.88	29.38/0.88
	30	27.59/0.81	27.71/0.82	27.60/0.82

In table 3, we compared the PSNR and MSSIM measures of the proposed methods with the bivariate shrink method on “Bridge,” “Mandrill,”

and “Goldhill” test images. These simulation results confirm the superiority of the proposed models.

In figure 4, the MAP estimated coefficient as a function of the noisy coefficient is depicted (using the student-t distribution as the prior probability of wavelet coefficients). The left graph shows the influence of the degrees of freedom ν for constant values of scale parameter σ and noise standard deviation. As demonstrated by this plot, the proposed MAP estimator shrinks the small observed wavelet coefficients heavily and the large ones slightly. The amount of shrinkage decreases as ν decreases. This can be understood from the fact that as ν decreases, the tails will be heavier,

and the higher the probability that smaller values are due to the true function d (original image) not due to noise. The graph on the right depicts the influence of the scale parameter σ on the MAP estimator for constant values of degrees of freedom ν and noise standard deviation. As we can see, the amount of shrinkage decreases as σ increases. The reason could be as σ increases SNR, ($SNR = \frac{\sigma^2}{\sigma_n^2} (\frac{\nu}{\nu-2})$) increases, so yielding less shrinkage.

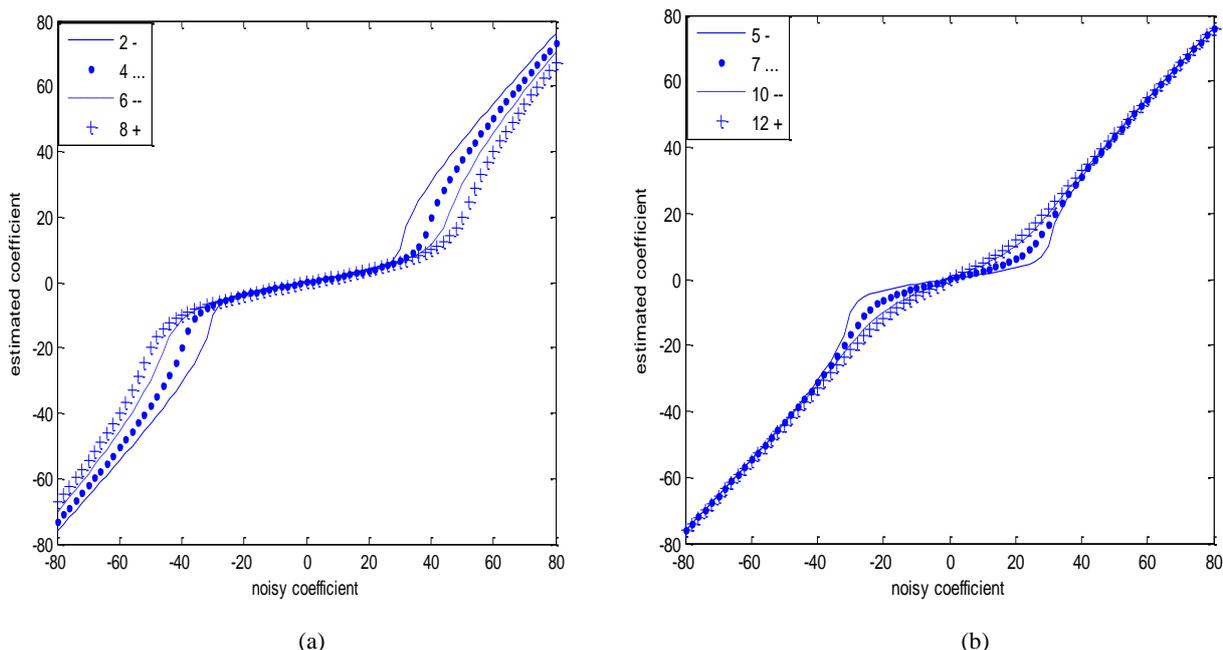


Figure 4. MAP estimator as a function of noisy coefficient when using the student-t distribution as the prior probability of wavelet coefficients a) influence of the degrees of freedom, b) influence of scale parameter σ .

We have also employed the overcomplete representation, namely Dual-Tree-Complex-Wavelet Transform (DTCWT) [42, 43] as the transform domain in which it is nearly shift-invariant, so it is not affected by the pseudo-Gibbs phenomena; it also has a good directional selectivity property. we compared the denoising results with some state of the art denoising algorithms, DTWF [44], AND [45], WF [46], DCTBI [47], GNW [48], NASW [49], LPG-PCA [50], GSM [51], EWD [52], MMSE [53], NSTSD [54], and R-NL [55]. The PSNR values of these methods and the student-t distribution in the DTCWT domain for two of the test images, ‘Barbara’ and ‘Lena’, are depicted in table 4. The results obtained show the superiority of the proposed method. The plots depicted in figure 5 are the observed (red-solid) and estimated (blue-dashed) densities

for one scale of wavelet coefficients for the Lena, Boat, and Barbara images. These images were corrupted by the Gaussian noise ($\sigma_n = 20$). The first line corresponds to the student-t model, and the second line corresponds to the Slash distribution, and the three columns are the HH , HL , and LH orientations, respectively. As we can see, the estimations are in good agreement with those obtained from the noise-free images.

5. Conclusion and future works

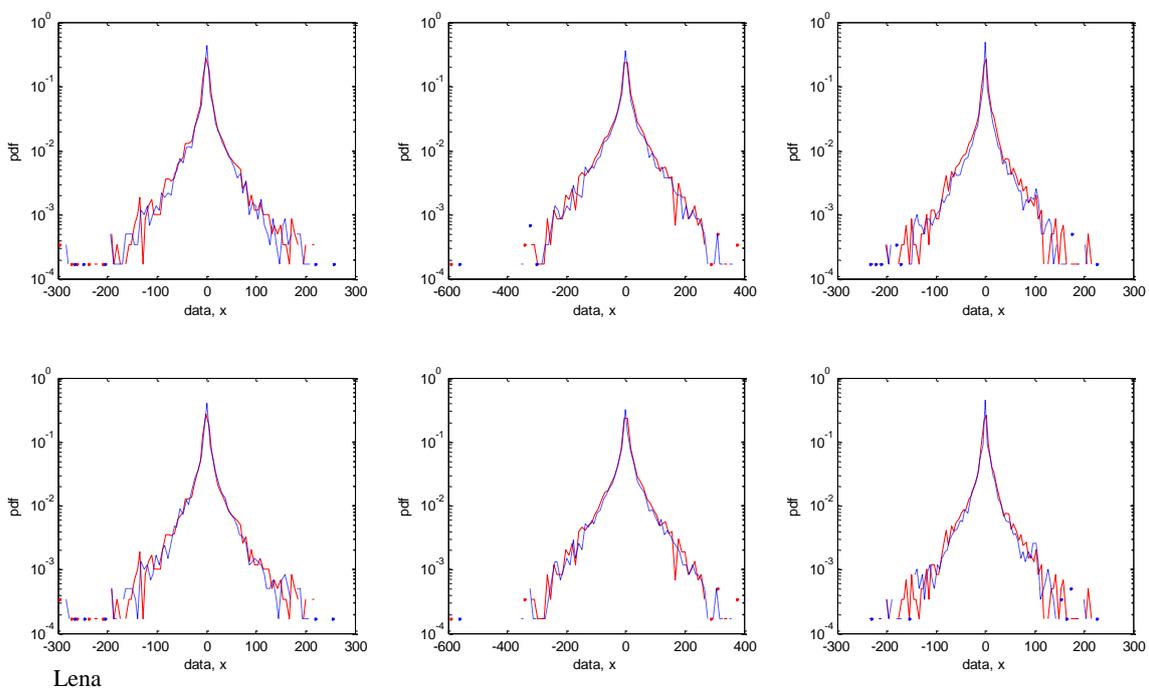
In this paper, we have proposed a statistical model to characterize the wavelet coefficient in each sub-band and developed a MAP estimator using this model for the digital image denoising problem. The student-t and Slash distributions were used as the prior distributions of noise-free coefficients, which can be expressed as the scale mixture of Gaussian distribution and can describe the heavy-tail behavior of the wavelet coefficient. Also by

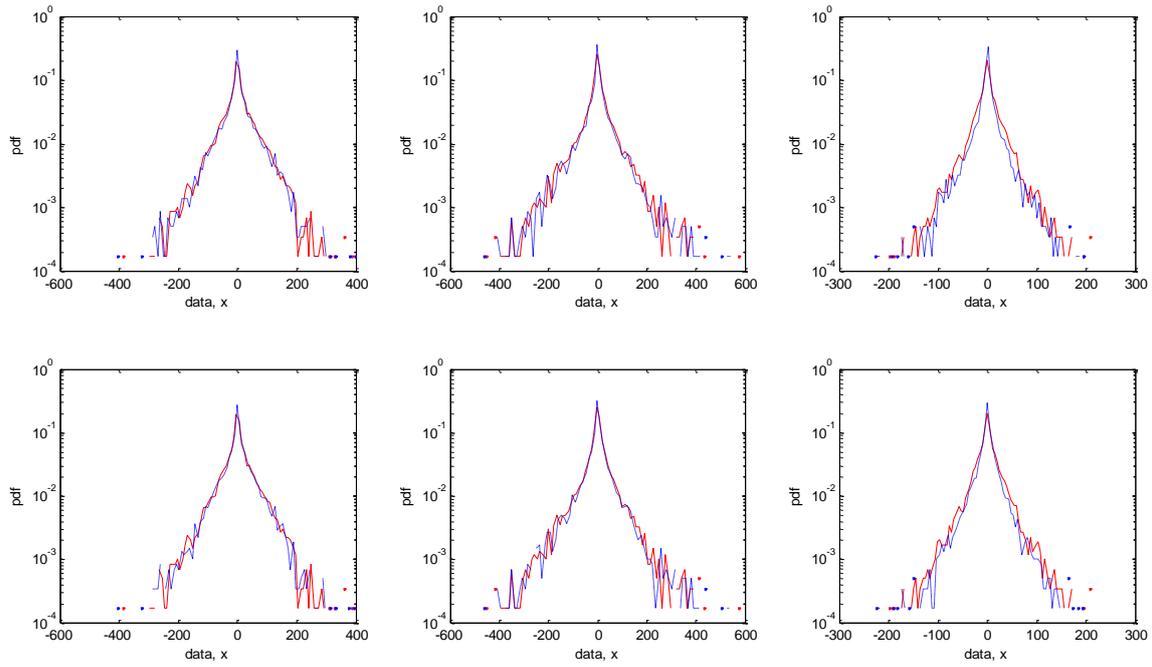
estimating the parameters of distribution adaptively, we could model the correlation between the coefficient amplitudes. We implemented the denoising algorithm in the DTCWT domain and compared our denoising results with some state of the art denoising methods. The simulation results showed the

superiority of the proposed methods. Better denoising results could be achieved by considering groups of wavelet coefficients together using the multivariate statistical models instead of processing each wavelet coefficient individually, which will be investigated in our future works.

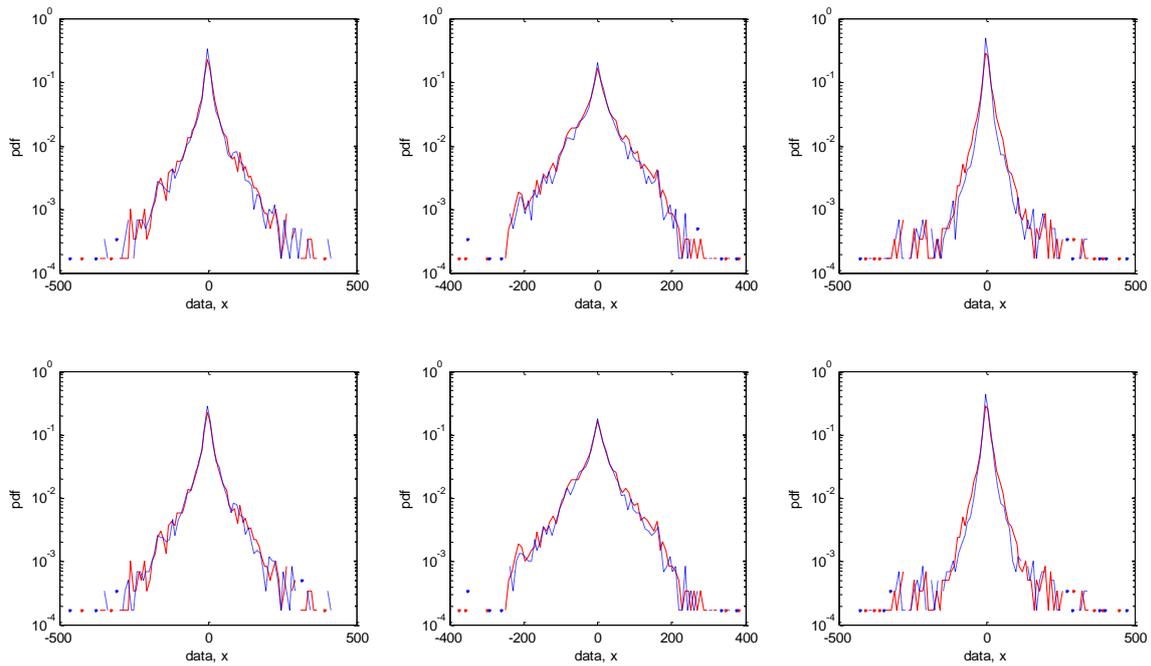
Table 4. PSNR comparison of performance for image denoising using different algorithms.

σ_η	Barbara				Lena			
	10	15	20	25	10	15	20	25
DTWF[44]	31.38	29.01	27.35	26.22	34.08	32.01	30.18	29.52
AND[45]	26.64	26.28	25.28	24.96	32.70	30.56	29.79	28.16
WF [46]	32.06	30.03	28.36	27.03	34.40	32.59	31.30	30.27
DCTBI [47]	-	29.42	28.19	27.20	-	31.29	30.82	29.09
GNW [48]	32.41	-	27.64	-	33.96	-	30.62	-
NASW [49]	31.40	29.03	27.46	26.29	34.09	32.02	30.16	29.54
LPG-PCA [50]	32.50	-	28.50	-	33.70	-	29.70	-
GSM [51]	33.08	30.71	29.05	27.79	35.28	-	31.74	-
EWD [52]	31.13	29.11	27.25	26.06	32.94	31.01	29.73	28.76
MMSE [53]	32.34	-	27.33	-	34.00	-	30.17	-
NSTSD [54]	33.16	-	28.96	-	35.12	-	31.93	-
R-NL [55]	-	-	29.76	-	-	-	31.04	-
Student-t-dtcwt	33.64	31.47	29.86	28.68	35.31	33.43	32.08	30.98





Boat



Barbara

Figure 5. Estimated (blue-dashed) and observed (red-solid) densities (in log scale) of the wavelet coefficient for one scale of the wavelet transform. The first row corresponds to the student t model and the second one corresponds to the slash distribution model. The three columns are HH , HL , and LH orientations.

6. Appendix I

In the probability theory and statistics, kurtosis is a measure of the tailedness of the probability distribution of a random variable, and variance is the expectation of the squared deviation of a random variable from its mean, and are defined as:

$$\text{var}(x) = E(x - \mu)^2 = M_2, \quad (36)$$

$$\text{kurtosis}(x) = \frac{E[(x - \mu)^4]}{E[(x - \mu)^2]^2} = \frac{M_4}{M_2^2}, \quad (37)$$

where, M_4 and M_2 are the 4th and 2nd central moments, respectively.

The cumulants (k_n) of a random variable x are defined via the cumulant generating function ($K(t)$), which is the natural logarithm of the moment generation function:

$$K(t) = \log E[e^{tx}]. \quad (38)$$

The 4th and 2nd cumulants are related to the central moments by the following equations [42]:

$$M_2 = k_2, \quad (39)$$

$$M_4 = k_4 + 3k_2^2. \quad (40)$$

Thus kurtosis and variance in terms of cumulants are:

$$\text{var}(x) = M_2 = k_2, \quad (41)$$

$$\text{Kurtosis}(x) = \frac{M_4}{M_2^2} = \frac{k_4 + 3k_2^2}{k_2^2} = \frac{k_4}{k_2^2} + 3. \quad (42)$$

References

[1] Kaur, S. & Singeh, N. (2014). Image denoising techniques: a review. *International Journal of Innotative Research in Computer and Communication Engineering*, vol. 2, pp. 4578-4583.

[2] Wang, X., Song, R., Song, C., Tao, J. (2018). The NSCT-HMT model of remote sensing image based on Gaussian-Cauchy mixture distribution. *IEEE Access* 6, vol. 6, pp. 66007-66019.

[3] Shao, L., Yan, R & Li, X. (2013). From heuristic optimization to dictionary Learning: a review and comprehensive comparison of Image denoising Algorithms. *IEEE Transaction on Cybernetics*, vol. 44, no. 7, pp. 1-14.

[4] Liu, H., Zhang, X. & Xiong, R. (2016). Content-adaptive low rank regularization for image denoising. *IEEE International Conference on Image Processing*, Arizona, USA, 2016.

[5] Lu, L., Jin, W. & Wang, X. (2015). Non local means image denoising with a soft threshold. *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 833-837.

[6] Moonon, S. A., Jianwen, H. & Shutao, L. (2015). Remote sensing image fusion method based on

nonsampled shearlet transform and sparse representation. *Sensing and Imaging*.

[7] Zeng, W., Fu, X., Hu, C. & Du, Y. (2018). Wavelet denoising with generalized bivariate prior model. *Multimedia Tools and Applications*, vol. 77, no. 16, pp. 20863-20887.

[8] Gwanggil, J. (2016). Denoising in Contrast-Enhanced X-ray Images. *Sensing and Imaging*, doi:10.1007/s11220-016-0140-9.

[9] Qian, W. (2015). Research on image denoising with an improved wavelet threshold algorithm. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 8, no. 9, pp 257-266.

[10] Donoho, D. L. (1995). Denoising by soft-thresholding. *IEEE Transaction Information Theory*, vol. 4, no. 3, pp. 613-627.

[11] Donoho, D. L. & Johnstone, I. M. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, vol. 81, no. 3, pp. 425-455.

[12] Gupta, S. & Minakshi. (2014). A review and comprehensive comparison of image denoising techniques. *IEEE International Conference on Computing for Sustainable Global Development*, New Delhi, India 2014.

[13] Donoho, D. L. & Johnstone, Z. M. (1995). Adapting to unknown smoothness via wavelet Shrinkage. *Journal of American Statistical Association*, vol. 90, no. 432, pp. 1200-1224.

[14] Chang, S. G., Yu, B. & Vetterli, M. (2000). Adaptive wavelet thresholding for image denoising and compression. *IEEE Transaction on Image Processing*, vol. 9, no. 9, pp. 1532-1546.

[15] Fadili, J. & Boubchir, L. (2005). Analytical form for a bayesian wavelet estimator of images using the bessel k forms densities. *IEEE Transaction on Image Processing*, vol. 14, pp. 231-240.

[16] Saeedzarandi, M., Nezamabadi-pour, H., Jamalizadeh, A. (2019). Dual-Tree Complex Wavelet Coefficient Magnitude Modeling Using Scale Mixtures of Rayleigh Distribution for Image Denoising. *Circuits, Systems, and Signal Processing*, pp. 1-26.

[17] Rabbani, H., Vafadust, M., Gazor, S. & Selesnick, I. (2006). Image denoising employing a bivariate cauchy distribution with local variance in complex wavelet domain. *IEEE 12th Digital Signal Processing Workshop*, doi: 10.1109/DSPWS.2006.265407, 2006.

[18] Amini, M., Ahmad, M. O. & Swamy, M. N. S. (2015). A new map estimator for wavelet domain image denoising using vector-based hidden Markov model. *IEEE International Symposium on Circuits and Systems*, Lisbon, Portugal, 2015.

[19] Bhuiyan, M. I., Ahmad, M. O. & Swamy, M. N. S. (2007). On the modeling of the wavelet image coefficients using a symmetric generalized hyperbolic

PDF. 50th Midwest Symposium on Circuits and Systems, Montreal, Canada, 2007.

[20] Naimi, H., Adamou-Mitiche, A. B. H. & Mitiche, L. (2015). Medical image denoising using dual tree complex thresholding wavelet transform and Wiener filter. *Journal of King Saud University- Computer and Information Sciences*, vol. 27, pp. 40-45.

[21] Mihcak, M. K., Kozintsev, I. & Ramchandran, K. (1999). Low complexity image denoising based on statistical modeling of wavelet coefficient. *IEEE signal processing letters*, vol. 6, no. 12, pp. 300-303.

[22] Rabbani, H. & Vafadoost, M. (2006). Noise reduction in wavelet domain based on a local laplace pdf. 14th Iranian Conference on Electrical Engineering, 2006.

[23] Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Transaction on Pattern Analysis and Machine Intelligence*. vol. 11, no. 7, pp. 674–693.

[24] Achim, A., Bezerianos, A. & Tsakalides, P. (2001). Novel bayesian multiscale method for speckle removal in medical ultrasound images. *IEEE Transaction on Medical Imag*, vol. 20, no. 5, pp. 772–783.

[25] Lange, K. L. & Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics*, vol. 2, no. 2, pp.175–198.

[26] Wang, J. & Taafe, M. R. (2015). Multivariate mixtures of normal distributions: properties, random Vector generation, fitting, and as models of market daily changes. *Inform Journal on Computing*. vol. 27, no. 2, pp.193-203.

[27] Dias, J. (2003). Fast GEM wavelet-based deconvolution algorithm. *IEEE International Conference on Image Processing*, Barcelona, Spain, 2003.

[28] Figueiredo, M. A. T. (2003). Adaptive sparseness for supervised learning. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1150-1159.

[29] Kour, p. (2015). Image processing using discrete wavelet transform. *International Journal of Electronic and Communication*, vol. 3, pp. 53-59.

[30] Simoncelli, E. P. (1999). Bayesian denoising of visual images in the wavelet domain, In: Muller, P. & Vidakovic, B. (Eds.), *Bayesian inference in wavelet based models*. Lecture notes in statistics, Springer, New York, pp. 291-308.

[31] Rabbani, H. & Vafadoost, M. (2006). Wavelet based image denoising based on a mixture of Laplace distributions. *Iranian Journal Science & Technology*. vol. 30, no. B6, pp. 711-733.

[32] Basso, R. M., Lachos, V. H., Cabral, C. R. B. & Ghosh, P. (2010). Robust mixture modeling based on scale mixture of skew-normal distribution.

Computational Statistics and Data Analysis, vol. 54, pp. 2927-2941.

[33] Chos, C. & Prates, M. O. (2012). Multivariate mixture modeling using skew-normal independent distributions. *Computational. Statistics & Data Analysis*, vol. 56, pp. 126–142.

[34] Rojas, M. A., Bolfarine, H. & Gomez, H. W. (2014). An extension of the slash-elliptical distribution. *Statistics and Operations Research Transactions*. vol. 38, no. 2, pp. 215-230.

[35] Rose, C. & Smith, M. D. (2002). k-statistics: unbiased estimators of cumulants. In: *Mathematical Statistics with Mathematica*, New York, Springer-Verlag, pp. 256-259.

[36] Fisher, R. A. (1950). Moments and product moments of sampling distributions. Wiley, New York.

[37] Crouse, M. S., Nowak, R. D. & Baraniuk, R. G. (1998). Wavelet-based statistical signal processing using hidden Markov model. *IEEE Transaction on signal processing*, vol. 46, no. 4, pp. 886-902.

[38] Sendur, L. & Selesnick, I. W. (2002). Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. *IEEE Transaction on Signal Processing*, vol. 50, no. 11, pp. 2744-2756.

[39] Sendur, L. & Selesnick, I. W. (2002). Bivariate shrinkage with local variance estimation. *IEEE Signal Processing letters*. vol. 9, no. 12, pp. 438-441.

[40] Rabbani, H & Vafadust, M. (2006). Image denoising with a mixture of Gaussian distribution with local parameters in wavelet domain. 4th Iranian conference on Image Processing, Tehran, Iran, 2006.

[41] Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (2007). *Minimization or Maximization of Functions*, in: *Numerical Recipes, the art of scientific computing*. Cambridge University, New York, pp. 492-496.

[42] Kingsbury, N. G. (1999). Image processing with complex wavelets. *Philosophical Transactions of the Royal Society B Biological Science*, vol. 357, pp. 2543–2560.

[43] Kingsbury, N. G. (2001). Complex wavelets for shift invariant analysis and filtering. *Applied and Computational Harmonic Analysis*, vol. 10, pp. 234–253, (2001).

[44] Khan, J., Bhuiyan, S. M. A., Muurphy, G & Arline, M. (2015). Embedded zero tree wavelet based data denoising and compression for smart grid. *IEEE Transactions on Industry Applications*, vol. 51, pp. 4190-4200.

[45] Mandava, A. K. & Regentova, E. E. Image denoising based on adaptive nonlinear diffusion in wavelet domain. *Journal of Electronic Imaging*, vol. 20, pp. 1-7.

- [46] Wang, J., Wu, J., Wu, Z., Jeong, J. & Jeon, G. (2017). Wiener filter based wavelet domain denoising. *Displays*, vol. 46, pp. 37-41.
- [47] Min, D. Jiuwen, Z. & Yide, M. (2015). Image denoising via bivariate shrinkage function based on a new structure of dual contourlet transform. *Signal Processing*, pp. 25-37.
- [48] Om, H & Biswas, M. (2013). A generalized image denoising method using neighbouring wavelet coefficients. *Signal, Image and Video Processing*, vol. 9, pp. 191–200.
- [49] Gaci, S. (2014). The use of wavelet based denoising to enhance the first arrival picking on seismic traces. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 52, pp. 4558-4563.
- [50] Zhang, L., Dong, W., Zhang, D. & Shi, G. (2010). Two-stage image denoising by principal component analysis with local pixel grouping. *Pattern Recognition*, vol. 43, pp. 1531–1549.
- [51] Portilla, J., Strela, V., Wainwright, M. & Simoncelli, E. (2003). Image denoising using scale mixture of Gaussian in the wavelet domain”, *IEEE transaction on Image Processing*. vol. 12, pp. 1338-1351.
- [52] Parmar, J. & Patil, S. (2013). Performance evaluation and comparison of modified denoising method and the local adaptive wavelet image denoising method. *International Conference on Intelligent Systems and Signal Processing*, Gujarat, India, 2013.
- [53] Om, H. & Biswas, M. (2014). MMSE based map estimation for image denoising. *Optics and laser technology*, vol. 57, pp. 252-264.
- [54] Meiyu, L., Junping, D. & Honggang, L. (2014). Self adaptive spatial image denoising model based on scale correlation and sure-let in the nonsubsampled contourlet transform domain. *Science China Information Sciences*, vol. 57, pp. 1-15.
- [55] Sutour, C., Deledalle, C. A. & Aujol, J. F. (2014). Adaptive regularization of the NL means: application to image and video denoising. *IEEE Transaction on Image Processing*. vol. 23, pp. 3506–3521.

روش آماری کاهش نویز تصاویر در حوزه موجک با استفاده از توزیع‌های مخلوط وزنی توزیع نرمال و تخمین پارامترها به صورت وقتی

منصوره سعیدزرنندی، حسین نظام آبادی پور* و سعید سریزدی

آزمایشگاه پردازش داده هوشمند، دانشکده برق، دانشگاه شهید باهنر کرمان.

ارسال ۲۰۱۸/۱۲/۰۸؛ بازنگری ۲۰۲۰/۰۱/۰۵؛ پذیرش ۲۰۲۰/۰۲/۰۴

چکیده:

کاهش نویز تصاویر، مبحثی مهم در حوزه پردازش تصاویر دیجیتال می‌باشد. در این مقاله، روش کاهش نویز بر اساس روش تخمین با استفاده از ماکزیمم کردن احتمال پسین (MAP) در حوزه موجک ارائه می‌شود. کارایی روش MAP به مدل احتمال پیشین نویز و احتمال پیشین ضرایب موجک بستگی دارد. در این مقاله، ضرایب موجک در هر زیر باند، با استفاده از توزیع‌های دنباله سنگین که از خانواده مخلوط وزنی توزیع نرمال هستند مدل می‌شوند. همچنین جهت استفاده از وابستگی بین ضرایب موجک، پارامترهای توزیع به صورت وقتی، در هر زیرباند، مدل شده‌اند. نتایج شبیه‌سازی، کارایی روش پیشنهادی را تایید می‌کنند.

کلمات کلیدی: کاهش نویز، تبدیل موجک، تخمین گر MAP، توزیع‌های دنباله سنگین، مخلوط وزنی توزیع نرمال.