Image Denoising Employing Two-Sided Gamma Random Vectors with Cycle-Spinning in Wavelet Domain

Pichid Kittisuwan[1], **Sanparith Marukatat**[2],
Thitiporn Chanwimaluang[3], and **Widhyakorn Asdornwised**[4], Non-members

ABSTRACT

In this work, we present new Bayesian estimator for *circularly-contoured Two-Sided Gamma* random vector in additive white Gaussian noise (AWGN). This PDF is used in view of the fact that it is more peaked and the tails are heavier to be incorporated in the probabilistic modeling of the wavelet coefficients. One of the cruxes of the Bayesian image denoising methods is to estimate statistical parameters for a shrinkage function. We employ maximum a posterior (MAP) estimation to calculate local variances with Rayleigh density prior for local observed variances and Gaussian distribution for noisy wavelet coefficients. Several denoising methods (ProbShrink with redundant wavelet transform) using un-decimated wavelet transforms provide good results. The un-decimated wavelet transforms can also be viewed as applying an orthogonal wavelet transform to a set of shifted versions of the signal. This procedure was first suggested by Coifman and Donoho where they termed it *cycle-spinning* method. We apply *cycle-spinning* with orthogonal wavelet transforms in our work. The experimental results show that the proposed method yields good denoising results.

Keywords: *Spherically-Contoured Two-Sided Gamma* Random Vectors, MAP Estimation, *Cycle-Spinning*, and Wavelet Transforms

1. INTRODUCTION

The application of the wavelet transform in image denoising has shown remarkable success over the last decade. There are two major approaches for statistical wavelet-based denoising. The first approach is to design a statistically optimal threshold parameter for some nonlinear thresholding techniques or shrinkage functions. Standard choices for nonlinear thresholding techniques or shrinkage functions are soft or hard thresholding [1], firm-shrinkage [2], and non-negative garrote shrinkage [3]. Hard thresholding technique yields a less biased estimate from observed signal but with a higher variance and discontinuity [3]. Soft thresholding yields moderate variance compared to hard thresholding but with more biased estimate. Firm-shrinkage relying on two threshold parameters outperforms both hard and soft thresholding. The only disadvantage of the firm shrinkage is that it requires two thresholds. The non-negative garrote shrinkage provides a good compromise between the soft and hard shrinkage function. Indeed, this method is based on a continuity, therefore it is more stable than hard thresholding and less bias than soft thresholding for large coefficient. These shrinkage methods are computationally simple, but rely on arbitrarily chosen non-linear function [4]. This drawback could be overcome under Bayesian denoising framework. In the second approach, the shrinkage function could be designed by minimizing Bayesian risk, typically under the maximum a posterior (MAP) or minimum mean square error (MMSE) criteria. Note that, in this case the prior knowledge about the distribution of wavelet coefficients is required. The choice of distribution representing the wavelet coefficients is one of the key to get better performance for Bayesian techniques.

It is widely known that the wavelet coefficients's

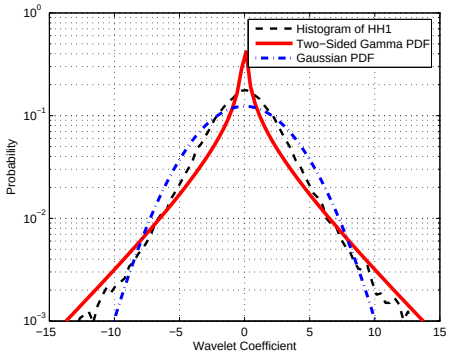

Fig.1: *Histogram of clean wavelet coefficients in the HH subband at scale-1 of Lena image, marginal Two-Sided Gamma distribution (1) and Gaussian distribution.*

Manuscript received on January 16, 2009 ; revised on April 2, 2009.

[1,4] The authors are with Department of Electrical Engineering, Chulalongkorn University, Bangkok, Thailand, E-mail:Pichid.K@Student.chula.ac.th and Widhyakorn.A@chula.ac.th

[2,3] The authors are with DNational Electronics and Computer Technology Center (NECTEC), Pathumthani, Thailand, E-mail:sanparith@nectec.or.th and marukatat@nectec.or.th

amplitude tends to propagate across scales. This parent-child relation is also underlined by the empirical joint histogram between parent and child coefficients as shown in [5]. Various authors have proposed different PDFs to approximate this joint histogram. For example, in [5] and [6], bivariate radial exponential density and Laplacian random vectors are proposed to model this parent-child joint PDF respectively. *Two-Sided Gamma* distribution is one of the PDFs that is successfully used for image denoising. In [7], the authors develop Bayesian estimation in combination with *Two-Sided Gamma* distribution for discrete fourier transform (DFT) coefficients. We recall that a zero-mean of marginal *Two-Sides Gamma* distribution with variance σ^2 has the density

$$f_x(x) = \frac{\sqrt[4]{3}}{2\sqrt{2\pi\sigma}} |x|^{\frac{-1}{2}} \exp\left(\frac{-\sqrt{3}|x|}{2\sigma}\right). \quad (1)$$

This model provides a simple and symmetric distribution. Fig.1 illustrates the histogram of the coefficients in one subband of the wavelet transform of a photographic image in log-scale. While the Gaussian PDF has a smooth peak and lean tails, the wavelet-coefficient histogram has different behavior patterns so it is more peaked and the tails are heavier. These kind of histograms are more kurtotic than the Gaussian distributions [6]. Fig.1 also illustrates PDF plots. The PDF plots illustrate the marginal *Two-Sided Gamma* distribution and Gaussian distribution. The histogram in Fig.1 is very symmetric with zero mean and the histogram of wavelet coefficients are more like marginal *Two-Sided Gamma* distribution, it is more peaked and the tails are heavier, than the Gaussian distribution.

In this paper, we propose the use of *spherically-contoured Two-Sided Gamma* random vectors to model parent-child wavelet coefficients in the expression for the MAP estimation of noise-free wavelet coefficients. These model is chosen because it is more peaked and the tails are heavier to be incorporated in the probabilistic modeling of the wavelet coefficients. *The estimator for statistical parameters of shrinkage function is one of the cruxes of the proposed algorithm.* In [8], the authors put a stochastic prior on the local observed variances and obtained MAP estimator for local observed variances. In this work, we employ MAP estimation to calculate local variances with Rayleigh density prior for local observed variances and Gaussian distribution for noisy wavelet coefficients. Although several denoising methods (ProbShrink with redundant wavelet transform) using un-decimated wavelet transforms provides good results [9]. The un-decimated wavelet transforms can also be viewed as applying an orthogonal wavelet transform to a set of shifted versions of the signal. This procedure was first suggested by Coifman and Donoho where they termed it *cycle-spinning* method [10]. So we applying *cycle-spinning* with orthogonal wavelet transforms in our work. Orthogonal wavelet transform has been exploited to decimated wavelet transform by using the *cycle-spinning* technique. We chose to employ the *cycle-spinning* method because it shifts the wavelet coefficients, and at the same time, the noise has been reduced. Prior to the noise reduction, the shifted wavelet coefficients are averaged (The detailed explanation can be found at Section 4 in the paper). We can simply write an efficient computer program for noise reduction when the wavelet coefficients have been shifted. The program can be written in parallel which significantly decreases the computational time. Hence, the approach is noticeably faster than other un-decimated wavelet transform methods, e.g., dual-tree complex wavelet transform, DT-CWT [5].

The rest of this paper is organized as follows. In Section 2, after a brief review on the basic idea of Bayesian denoising, we obtain a shrinkage function using *Two-Sided Gamma* random vectors with local variance namely, *GammaShrink*. Section 3 describes the MAP estimator for the local variances of wavelet coefficients using Rayleigh marginal prior density with Gaussian distribution, since this estimate is the needed parameter for the shrinkage function in Section 2. Section 4 describes *cycle-spinning* method due to the fact that un-decimated wavelet transform can be viewed as applying an orthogonal transform to a set of shifted versions of the image. In Section 5, the proposed method is applied for wavelet-based denoising of several images corrupted with additive Gaussian noise at various noise levels. Simulation results demonstrate the effectiveness of our proposed algorithms compared with other state-of-the-art methods. The experimental results show that our algorithm achieves better performance visually and in terms of PSNR. Finally the concluding remarks and discussion are given in Section 6.

2. BAYESIAN DENOISING

In this section, the idea of MAP estimator will be explained. Assume that we want to estimate a parameter $\theta = [\theta_1, \theta_2, \ldots, \theta_d]^T$ on the observations $\mathbf{y} = [y_1, y_2, \ldots, y_d]^T$. Lets' $f_{\mathbf{y}|\theta}(\mathbf{y}|\theta)$ is known as the likelihood function. Now assume that a prior distribution $f_\theta(\theta)$ exists. The posterior distribution of θ is $f_{\theta|\mathbf{y}}(\theta|\mathbf{y})$. The method of MAP estimation then estimate θ as the mode of the posterior distribution of the random variable [11]: $\hat{\theta}_{MAP} = \arg\max_\theta \left[f_{\theta|\mathbf{y}}(\theta|\mathbf{y})\right]$. Using Bayes' rule, one gets

$$\hat{\theta}_{MAP} = \arg\max_\theta \left[f_{\mathbf{y}|\theta}(\mathbf{y}|\theta) f_\theta(\theta)\right]. \quad (2)$$

Indeed, the denoising of an image corrupted by AWGN with variance σ_n^2 will be considered. For a wavelet coefficient x_1, let $x_2, x_3, \ldots, x_d$ represent its parent, i.e., $x_2, x_3, \ldots, x_d$ is the wavelet coefficient at the same position as the wavelet coefficient

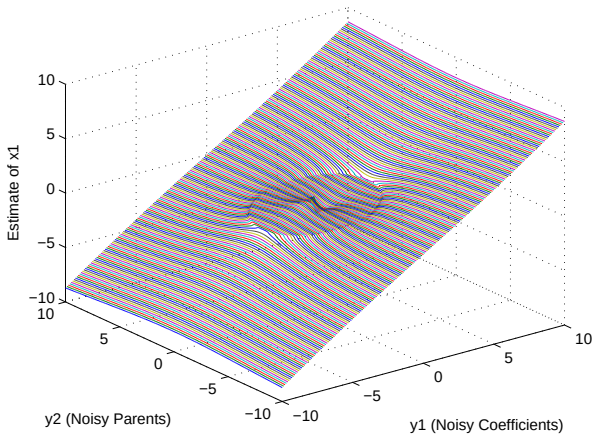

Fig.2: *New shrinkage function produced from (6), GammaShrink with* $\sigma^2 = 4, \sigma_n^2 = 2$

x_1, but at the next coarser scale. We suppose that the coefficients are contaminated by additive noise, that is $y_1 = x_1 + n_1,\ y_2 = x_2 + n_2, \ldots, y_{\mathrm{d}} = x_{\mathrm{d}} + n_{\mathrm{d}}$ where $y_1, y_2, y_3, \ldots, y_d$ are noisy observations of $x_1, x_2, x_3, \ldots, x_d$, and $n_1, n_2, n_3, \ldots, n_d$ are noise samples respectively. To take into account the statistical dependencies between a coefficient and its parent, we combine them into vector form as follow: $\mathbf{y} = \mathbf{x} + \mathbf{n}$. Let us continue on developing the MAP estimator given in (2), where $\theta = \mathbf{x}$, which is equivalent to

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} \left[f_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})\, f_{\mathbf{x}}(\mathbf{x}) \right].$$

After some manipulations, this equation can be written as

$$\hat{\mathbf{x}}(\mathbf{y}) = \arg\max_{\mathbf{x}} [\ln f_{\mathbf{n}}(\mathbf{y} - \mathbf{x}) + \ln f_{\mathbf{x}}(\mathbf{x})]. \quad (3)$$

The proposed model for joint pdf ($f_{\mathbf{x}}(\mathbf{x})$) is important in the above equation. Generally, it is hard to find a model for this PDF. In this paper, we proposed the following *Two-Sided Gamma* random vectors for coefficient and its parent:

$$f_{\mathbf{x}}(\mathbf{x}) = k_1 (\|\mathbf{x}\|)^{\frac{-1}{2}} \exp\left(\frac{-\sqrt{3}\,\|\mathbf{x}\|}{2\sigma} \right) \quad (4)$$

where σ is standard deviation of this model and k_1 is the normalizing constant parameter. We assume that the noise is i.i.d white Gaussian, the noise PDF is given by [5]

$$f_{\mathbf{n}}(\mathbf{n}) = \frac{1}{(2\pi\sigma_n^2)^{\frac{d}{2}}} \exp\left(\frac{-\|\mathbf{n}\|^2}{2\sigma_n^2} \right). \quad (5)$$

Solving (3) using (4) and (5), the MAP estimator for x_1 is derived as a solution of:

$$\hat{x}_1 = \frac{r_+}{\|\mathbf{y}\|} y_1 \quad (6)$$

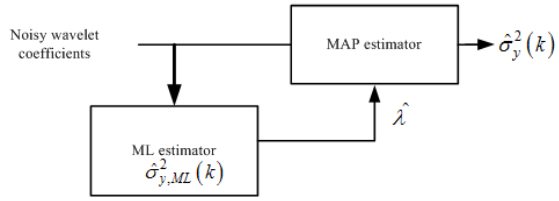

Fig.3: *Block-diagram of the proposed MAP estimator for local observed variance, the parameter* $\hat{\lambda}$ *is calculated from local observed variance using ML estimation with Gaussian assumption* ($\hat{\sigma}_{y,ML}^2(k)$) *by (10).*

where

$$r = \frac{-a + \sqrt{a^2 - 2\sigma_n^2}}{2}, a = \frac{\sqrt{3}\sigma_n^2}{\sigma} - \|\mathbf{y}\|.$$

Here, $(g)_+$ is defined as $g_+ = \max(0, g)$. The derivation can be found in Appendix A. This shrinkage function is called *GammaShrink*. From this point onward, empirically verify GammaShrink just in bivariate case $\left(d = 2, \|\mathbf{y}\| = \sqrt{y_1^2 + y_2^2}\right)$ only. Fig. 2 shows the proposed shrinkage function produced from (6). As we can see in Fig. 2, the shrinkage function of *GammaShrink* is non-linear.

3. PARAMETER ESTIMATION

To apply (6), we need to know the noise variance σ_n^2 and variance of noise-free wavelet coefficients σ^2. To estimate the noise variance from the noisy wavelet coefficients, a robust median estimator is obtained from the HH_1 Subband [12]: $\sigma_n^2 = (\frac{median(|HH_1|)}{0.6745})^2$.

Under the assumption that marginal variances are different for each data point $y(k)$, an estimated local observed variance $\sigma_y^2(k)$ can be found using local neighborhood $N(k)$. We use a square window $N(k)$ centered at $y(k)$. Now, assume that a prior marginal distribution $f_{\sigma_y^2(k)}(\sigma_y^2(k))$ is available. By using MAP estimator in (2), where $\theta = \sigma_y^2(k)$. After some manipulation, we can obtain MAP estimator for $\sigma_y^2(k)$ as

$$\hat{\sigma}_y^2(k) = \arg\max_{\sigma_y^2(k) \geq 0} \left[\ln \left(\left(\prod_{j \in N(k)} f\left(y_j | \sigma_y^2(k)\right) \right) \times f_{\sigma_y^2(k)}\left(\sigma_y^2(k)\right) \right) \right]. \quad (7)$$

In this section, we assume that noisy wavelet coefficients is the Gaussian distribution with zero mean and variance $\sigma_y^2(k)$: $f\left(y_j|\sigma_y^2(k)\right) = \frac{1}{\sqrt{2\pi\sigma_y^2(k)}} \exp\left(\frac{-y_j^2}{2\sigma_y^2(k)}\right)$. Thus, the MAP estimator for $\sigma_y^2(k)$ can be obtained using a Rayleigh density prior $f_{\sigma_y^2(k)}\left(\sigma_y^2(k)\right) = \lambda^2 \sigma_y^2(k) \exp\left(\frac{-\lambda^2 \left(\sigma_y^2(k)\right)^2}{2}\right), \lambda > 0.$

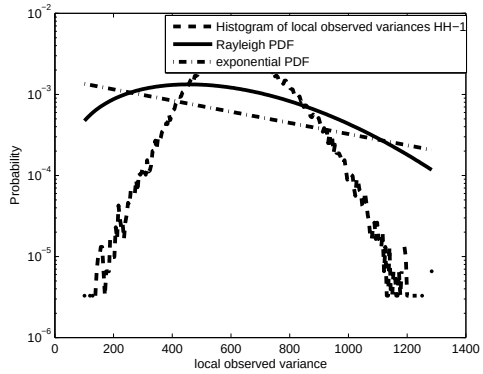

Fig.4: *Histogram of ML estimator for local observed variances* $(\hat{\sigma}^2_{y,ML}(k))$ *by (10) of HH-1 and* $\sigma_n = 25$ *of Lena image, Rayleigh density with* $\hat{\lambda} = 0.002$ *(11), and exponential density.*

As derived in Appendix B, we can write

$$\left(\sigma_y^2(k)\right)^3 + \left(\frac{M-2}{2\lambda^2}\right)\sigma_y^2(k) - \frac{\sum\limits_{j\in N(k)} y_j^2}{2\lambda^2} = 0. \quad (8)$$

giving that λ is parameters of Rayleigh density, and M is number of coefficient in $N(k)$. Coincidently, (8) is also a third order equation. Based on Cardano's method (The description can be found in Appendix C), we can also obtain:

$$\sigma_y^2(k) = \sqrt[3]{C(k)} + \sqrt[3]{D(k)} \quad (9)$$

where

$$C(k) = \frac{\sum\limits_{j\in N(k)} y_j^2}{4\lambda^2} + \sqrt{\frac{\left(\sum\limits_{j\in N(k)} y_j^2\right)^2}{16\lambda^4} + \frac{(M-2)^3}{216\lambda^6}},$$

$$D(k) = \frac{\sum\limits_{j\in N(k)} y_j^2}{4\lambda^2} - \sqrt{\frac{\left(\sum\limits_{j\in N(k)} y_j^2\right)^2}{16\lambda^4} + \frac{(M-2)^3}{216\lambda^6}}.$$

To select the parameter λ, we use the fact that our Rayleigh density prior assumption $\sigma_y^2(k)$, computed over all coefficients, should distribute according to Rayleigh density. First, the parameter λ has to be calculated from the maximum likelihood (ML) estimation with Gaussian distribution zero mean and variance $\sigma^2_{y,ML}(k)$, that is [13]:

$$\hat{\sigma}^2_{y,ML}(k) = \frac{\sum\limits_{j\in N(k)} y_j^2}{M} \quad (10)$$

where [14]

$$\hat{\lambda} = \sqrt{\frac{2N}{\sum\limits_{k=1}^{N}\left(\hat{\sigma}^2_{y,ML}(k)\right)^2}} \quad (11)$$

and N is the number of wavelet coefficients in each subband. The estimation of λ is then substituted in (9) for MAP estimation of local observed variance $(\sigma_y^2(k))$. The block-diagram in Fig.3 describes the MAP estimator for local observed variances.

One of the motivations for using Rayleigh density prior is that Rayleigh density requires the estimation only one additional parameter (λ) per image wavelet subband. It is easy to find λ in the same way as using exponential density prior. In [8], the authors used exponential density as the local observed variances prior. Histogram of ML estimate of local observed variance $(\hat{\sigma}^2_{y,ML}(k))$ of Lena image in HH-1 with $\sigma_n = 25$, using window size 7×7 is illustrated in Fig.4. In addition, Rayleigh density, and exponential density of local observed variance are estimated and illustrated in Fig.4. It can be observed that the histogram of $\hat{\sigma}^2_{y,ML}(k)$ are more like Rayleigh density than exponential density.

We use the fact that the wavelet coefficients and the additive noise are independent, thus we have the following relation between their statistical parameters:

$$\hat{\sigma}^2(k) = \left(\hat{\sigma}_y^2(k) - \hat{\sigma}_n^2\right)_+. \quad (12)$$

Then

$$\hat{\sigma}^2(k) = \sqrt{\left(\left(\sqrt[3]{C(k)} + \sqrt[3]{D(k)}\right) - \hat{\sigma}_n^2\right)_+}. \quad (13)$$

4. CYCLE-SPINNING METHOD

This section introduces *cycle-spinning* method. The shrinkage approach (e.g., BiShrink with DT-CWT, [15] ProbShrink [9], BLS-GSM [16] and Gaussian-Hermite Expansion [4]) using unitary transforms provides good results. Significant improvement is achieved when this technique is implemented using over-complete representations. In most cases, there are implemented methods using either un-decimated wavelet transform or any other sliding windowed transforms (sliding local DCT, sliding local DFT, steerable pyramid or complex wavelet transform, etc.) The un-decimated wavelet transform can be viewed as applying an orthogonal transform to a set of shifted versions of the image. This way, the shrinkage operation is applied to each transformed image independently. This is followed by transforming each modified transform back to the image space, and averaging all the corrected images after shifting them back to their original positions. This is illustrated in Fig. 5. This procedure was first suggested by Coifman and Donoho where they termed it *cycle-spinning* denoising.

5. EXPERIMENTAL RESULTS

This section presents image denoising examples using our new statistical model for wavelet coefficients with *cycle-spinning* (*GammaShrink*) and

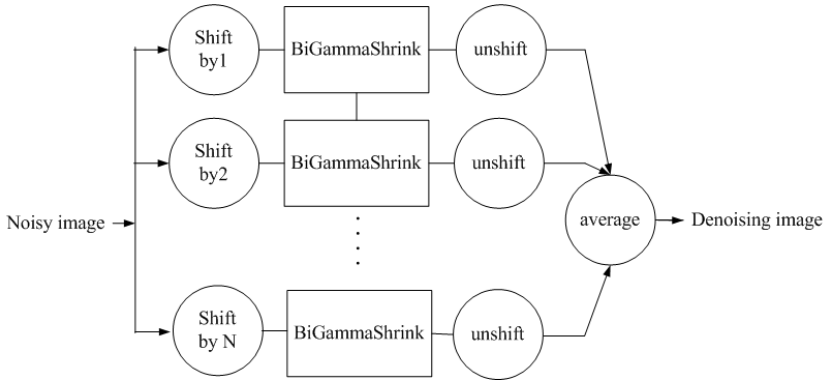

Fig.5: *The pipe-line of the* cycle-spinning *method*

Fig.6: *(Lena) From left to right and clockwise: Noise-free image, Noisy image with* $\sigma_n = 30$*, LAWMAP [8] (PSNR = 28.98), and GammaShrink (PSNR = 30.52).*

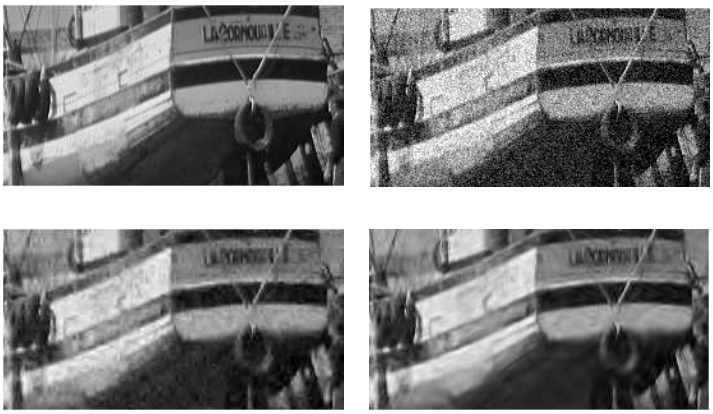

Fig.7: *(Boat) From left to right and clockwise: Noise-free image, Noisy image with* $\sigma_n = 30$*, BayeShrink (DT-CWT) [17] (PSNR = 28.08), and GammaShrink (PSNR = 28.16).*

compare it with MMSE-Laplace [6] LAWMAP [8], Hard−Threshold [1], and BayeShrink with dual-tree complex wavelet transform (DT-CWT) [17], undecimated wavelet transform version of BayeShrink. These algorithms are evaluated with different noise level $\sigma_n = 5, 10, 20, 30, 40$ and 50. Three 512×512 grayscale images and one 256×256 grayscale image, namely Lena, Boat, Barbara, and Cameraman respectively, are used to assess the algorithm's performance. We use peak-signal-to-noise ratio (PSNR) as an objective criterion for image performance evaluation. All test images can be obtained from the same sources as mentioned in [18]. Here, the Daubechies length-10 filter and a 7×7 window size ($N(k)$) with periodic boundary extensions [19] are used. We have also investigated different window sizes. A 9×9 window size can also be a good choice. However, using a 3×3 window sizes resulted in a slight performance loss. In this paper, we have not considered different shapes for $N(k)$. The results can be seen in Table 1-4. Each PSNR values in the table is averaged over five runs. In these Tables, the highest PSNR value is bolded. We test on the contributions of cycle-spinning, *GammaShrink*, and the proposed local variance estimation parts to the performance of the proposed denoising method. Having listed from high to low performance improvement respectively, Method I (using *GammaShrink*, proposed local variance, and cycle-spinning), Method III (using *GammaShrink* and cycle-spinning with ML local variance estimate), Method IV (using BayeShrink, proposed local variance, and cycle-spinning), and Method II (using *GammaShrink* and proposed local variance without cycle-spinning) are shown in Table 5. As we can seen from Table 6, the computational time of our proposed method using 6 co-processors is slightly faster than conventional redundant wavelet transform techniques. The more co-processors are used, the less computational time is required. The essential reason is that several redundant wavelet transforms, e.g. steerable pyramid transform, perform in sequence by performing wavelet transform first and then denoising the transformed coefficients later. This way, the computation of redundant wavelet transform plays an important role and usually prefers not to be implemented in a parallel manner. Fig. 6 illustrates denoising algorithms on Lena image with $\sigma_n = 30$, original image, noisy image, denoising image obtained from LAWMAP, and *GammaShrink*. Fig. 7 shows results from several denoising algorithms on Boat image with $\sigma_n = 30$, original image, noisy image, denoising image obtained from BayeShrink (DT-CWT), and our proposed method.

6. CONCLUSION AND DISCUSSION

This paper presents the use of *spherically-contoured Two-Sided Gamma* random vectors to model parent-child wavelet coefficients in the expression for the MAP estimation of noise-free wavelet coefficients. To estimate statistical parameters for a shrinkage function, we employ MAP estimation to calculate local observed variances. Instead of using overcomplete transform, we apply *cycle-spinning* with orthogonal wavelet transforms. Instead of using this proposed shrinkage function, other nonlinear shrinkage functions can also be used. For example, instead of using MAP estimator we can use the MMSE estimator to obtain the denoising shrinkage function. For better capture the heavy-tailed nature of wavelet coefficients, we can use mixture models. Therefore, if we use a mixture model of *Two-Sided Gamma* random vectors for wavelet, the performance of denoising algorithm will be improved.

APPENDICES

A DERIVATION OF *GAMMASHRINK*

Solving (3) using (4) and (5) gives

$$\ln f_{\mathbf{n}}(\mathbf{y}-\mathbf{x}) + \ln f_{\mathbf{x}}(\mathbf{x}) = \ln\left(\frac{1}{(2\pi\sigma_n^2)^{\frac{d}{2}}}\right) - \frac{\|\mathbf{y}-\mathbf{x}\|^2}{2\sigma_n^2} + \ln(k_1) - \frac{1}{2}\ln(\|\mathbf{x}\|) - \frac{\sqrt{3}\|\mathbf{x}\|}{2\sigma},$$

$$\frac{\partial\left[\ln f_{\mathbf{n}}(\mathbf{y}-\mathbf{x}) + \ln f_{\mathbf{x}}(\mathbf{x})\right]}{\partial x_i} = 0$$

Then,

$$x_i\left(1 + \frac{\sigma_n^2}{2\|\mathbf{x}\|^2} + \frac{\sqrt{3}\sigma_n^2}{2\sigma\|\mathbf{x}\|}\right) = y_i.$$

Setting $r = \|\mathbf{x}\|$

$$x_i = \frac{y_i}{\left(1 + \frac{\sigma_n^2}{2r^2} + \frac{\sqrt{3}\sigma_n^2}{2\sigma r}\right)}. \tag{14}$$

Taking the square root of the sum of the square over $1 \le i \le d$ gives

$$\|\mathbf{x}\| = r = \frac{\|\mathbf{y}\|}{\left(1 + \frac{\sigma_n^2}{2r^2} + \frac{\sqrt{3}\sigma_n^2}{2\sigma r}\right)}.$$

Then,

$$\left(1 + \frac{\sigma_n^2}{2r^2} + \frac{\sqrt{3}\sigma_n^2}{2\sigma r}\right) = \frac{\|\mathbf{y}\|}{r}, \tag{15}$$

$$r^2 + \left(\frac{\sqrt{3}\sigma_n^2}{2\sigma} - \|\mathbf{y}\|\right)r + \frac{\sigma_n^2}{2} = 0.$$

Finding r and Substituting (15) in (14) gives

$$x_i = \frac{r_+}{\|\mathbf{y}\|}y_i,$$

$$r = \frac{-a + \sqrt{a^2 - 2\sigma_n^2}}{2}, a = \frac{\sqrt{3}\sigma_n^2}{\sigma} - \|\mathbf{y}\|.$$

Table 1: *Average PSNR Values of Denoising Image Over Five Runs for Lena Image.*

Standard Deviation Noise	5	10	20	30	40	50
MMSE-Laplace [6]	37.77	34.47	30.95	28.88	27.45	26.26
LAWMAP [8]	37.73	34.45	30.98	28.98	27.67	26.47
Hard−Threshold [1]	33.78	31.10	28.33	26.71	25.61	24.95
BayeShrink (DT-CWT) [17]	37.63	34.89	31.75	29.74	28.38	27.26
GammaShrink	**38.12**	**35.03**	**31.93**	**30.05**	**28.80**	**27.79**

Table 2: *Average PSNR Values of Denoising Image Over Five Runs for Boat Image.*

Standard Deviation Noise	5	10	20	30	40	50
MMSE-Laplace [6]	**35.76**	32.62	29.23	27.21	25.83	24.79
LAWMAP [8]	35.71	32.56	29.20	27.24	25.89	24.93
Hard−Threshold [1]	31.16	28.69	26.04	24.73	23.82	23.11
BayeShrink (DT-CWT) [17]	34.71	32.36	29.41	27.55	26.29	25.32
GammaShrink	**36.26**	**33.22**	**30.01**	**28.11**	**26.82**	**25.90**

Table 3: *Average PSNR Values of Denoising Image Over Five Runs for Barbara Image.*

Standard Deviation Noise	5	10	20	30	40	50
MMSE-Laplace [6]	36.44	32.46	28.59	26.43	25.05	23.97
LAWMAP [8]	36.56	32.60	28.76	26.63	25.24	24.18
Hard−Threshold [1]	30.75	27.37	24.18	22.96	22.40	21.96
BayeShrink (DT-CWT) [17]	35.69	32.54	29.15	**27.21**	**25.83**	**24.81**
GammaShrink	**37.26**	**33.20**	**29.19**	26.89	25.44	24.38

Table 4: *Average PSNR Values of Denoising Image Over Five Runs for Cameraman Image.*

Standard Deviation Noise	5	10	20	30	40	50
MMSE-Laplace [6]	36.92	32.61	28.57	26.44	24.92	23.81
LAWMAP [8]	36.69	32.33	28.29	26.17	24.69	23.70
Hard−Threshold [1]	31.34	28.19	25.10	23.53	22.35	21.59
BayeShrink (DT-CWT) [17]	35.98	32.07	28.44	26.54	25.21	24.18
GammaShrink	**37.34**	**32.91**	**29.01**	**26.94**	**25.54**	**24.57**

Table 5: *Average PSNR Values of Lena Image Over Five Runs using Key Techniques in Our Proposed Algorithms (A = GammaShrink, B = Proposed Local Variance Estimation, C = Cycle Spinning, $\overline{A}$ = BayeShrink, $\overline{B}$ = Local Variance Estimation using ML, $\overline{C}$ = Without Cycle Spinning).*

Image	Techniques			PSNR					
	A	**B**	**C**	5	10	20	30	40	50
Method I	O	O	O	**38.12**	**35.03**	**31.93**	**30.05**	**28.80**	**27.79**
Method II	O	O		37.28	34.05	30.70	28.83	27.63	26.55
Method III	O		O	**38.12**	35.01	31.88	29.99	28.74	27.69
Method IV		O	O	38.11	35.01	31.92	29.97	28.71	27.68

Table 6: *Computational Time (Second)*

	BayeShrink +Local Variance	GammaShrink (1 Processor)	GammaShrink (25 Processors)	GammaShrink (6 Processors)
Lena	15.4100	84.7400	**3.3900**	**14.1250**
Boat	17.3300	85.0000	**3.4000**	**14.1667**
Barbara	17.3900	85.2500	**3.4100**	**14.2083**
Cameraman	6.0200	32.5000	**1.3000**	**5.4167**

B DERIVATION OF THE MAP ESTIMATION FOR $\sigma_Y^2(K)$ USING RAYLEIGH DENSITY PRIOR FOR LOCAL OBSERVED VARIANCES AND GAUSSIAN DISTRIBUTION FOR NOISY WAVELET COEFFICIENTS

Where $f\left(y_j|\sigma_y^2(k)\right) = \frac{1}{\sqrt{2\pi\sigma_y^2(k)}}\exp\left(\frac{-y_j^2}{2\sigma_y^2(k)}\right)$ is the Gaussian distribution with zero mean and variance $\sigma_y^2(k)$ and $f_{\sigma_y^2(k)}\left(\sigma_y^2(k)\right)$ is PDF of Rayleigh density $f_{\sigma_y^2(k)}\left(\sigma_y^2(k)\right) = \lambda^2\sigma_y^2(k)\exp\left(\frac{-\lambda^2\left(\sigma_y^2(k)\right)^2}{2}\right), \lambda > 0$, thus,

$$\left(\prod_{j\in N(k)} f\left(y_j|\sigma_y^2(k)\right)\right) f_{\sigma_y^2(k)}\left(\sigma_y^2(k)\right) =$$

$$\left(\frac{1}{\sqrt{2\pi\sigma_y^2(k)}}\right)^M \exp\left(\frac{-\sum_{j\in N(k)} y_j^2}{2\sigma_y^2(k)}\right) \times\lambda^2\sigma_y^2(k)\exp\left(\frac{-\lambda^2\left(\sigma_y^2(k)\right)^2}{2}\right)$$

$$\ln\left(\left(\prod_{j\in N(k)} f\left(y_j|\sigma_y^2(k)\right)\right) f_{\sigma_y^2(k)}\left(\sigma_y^2(k)\right)\right) =$$

$$\frac{M}{2}\ln\left(\frac{1}{2\pi}\right) - \frac{M}{2}\ln\left(\sigma_y^2(k)\right) - \frac{\sum_{j\in N(k)} y_j^2}{2\sigma_y^2(k)} + \ln\left(\lambda^2\right)$$

$$+\ln(\sigma_y^2(k)) - \frac{\lambda^2(\sigma_y^2(k))^2}{2}.$$

In order to find $\sigma_y^2(k)$, we use (7). Then

$$\frac{\partial\ln\left(\left(\prod_{j\in N(k)} f\left(y_j|\sigma_y^2(k)\right)\right) f_{\sigma_y^2(k)}\left(\sigma_y^2(k)\right)\right)}{\partial\left(\sigma_y^2(k)\right)} = 0,$$

$$\left(\sigma_y^2(k)\right)^3 + \left(\frac{M-2}{2\lambda^2}\right)\sigma_y^2(k) - \frac{\sum_{j\in N(k)} y_j^2}{2\lambda^2} = 0.$$

Finally, we use Cardano's method to find the optimal $\sigma_y^2(k)$.

C CARDANO'S METHOD

In mathematic, a cubic function is a function of the form

$$f(x) = c_4x^3 + c_1x^2 + c_2x + c_3,$$
$$c_1,\ c_2,\ c_3,\ c_4 \in R, c_4 \neq 0.$$

We first normalize this standard equation by dividing the equation with the first coefficient. Thus, we can write,

$$x^3 + ax^2 + bx + c = 0.$$

In summary, if we use cardano's method [20] to solve the above equation,The roots of the cubic equation will be

$$x = \begin{cases} \sqrt[3]{A} + \sqrt[3]{B} - \frac{a}{3} \\ \frac{-1}{2}(\sqrt[3]{A} + \sqrt[3]{B}) \pm \frac{\sqrt{3}}{2}j(\sqrt[3]{A} + \sqrt[3]{B}) - \frac{a}{3} \end{cases}$$

where

$$p = b - \frac{a^2}{3}, \quad q = c - \frac{ab}{3} + \frac{2a^3}{27}$$

$$A = \frac{-q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}, \quad B = \frac{-q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}.$$

Note that only a real root will be used here. In MATLAB language, we use *nthroot* function to find a real root.

ACKNOWLEDGMENT

The authors would like to thank Thailand Graduate Institute of Science and Technology (TGIST) for Research and Development, Dr. I. Selesnick for making available the codes of wavelet denoising.

References

[1] D. L. Donoho, "Denoising by soft-thresholding," *IEEE Transaction Information Theory*, vol. 41, no. 3, pp. 613-627, Mar 1995.

[2] H. Y. Gao and A. G. Bruce, "Waveshrink with firm shrinkage," *Statistis Sinica*, vol. 7, no. 4, pp. 855-874, 1997.

[3] H. Y. Gao, "Wavelet shrinkage denoising using the non-negative garrote," *J Computation Graph Statistic*, vol. 7, no. 4, pp. 469-488, 1998.

[4] S. M. M. Rahman, M. O. Ahmad, and M. N. S. Swamy,"Bayesian Wavelet-Based Image Denoising Using the Gauss-Hermite Expansion," *IEEE Transaction Image Processing*, vol. 17, no. 10, pp 1755-1771, October 2008.

[5] L. Sendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Transaction Signal Processing*," vol. 50, no. 11, pp. 2744-2756, November 2002.

[6] I. W. Selesnick, "Estimation of Laplace Random Vectors in Adaptive White Gaussian Noise," *IEEE Transactions on Signal Processing*, vol 56, no. 8, pp. 3482-3496. August 2008.

[7] R. Martin, "Speech Enhancement Based on Minimum Mean-Square Error Estimation and Supergaussian Priors," *IEEE Transeactions on Speech*

and Audio Processing", vol. 13, No. 5, September 2005.

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

[9] A. Pizurica and W. Philips, "Estimating the probability of the presence of a signal of interest in multi-resolution single- and multiband image denoising," *IEEE Transection Image Processing*, vol. 15, no. 3, pp. 654-665, Mar. 2006.

[10] R. R. Coifman and D. L. Donoho. Translation invariant de-noising. In A. Antoniadis and G. Oppenheim, editors, Wavelets and Statistics, pages 125-150. Springer-Verlag, 1995.

[11] H. W. Sorenson, Parameter Estimation: Principles and Problems. Marcel Dekker: 1980.

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

[13] M. R. Spiegel, Theory and Problems of Probability and Statistics, McGraw-Hill: 1992.

[14] H. Stark and J. W. Woods, Probability and Random Processes with Application. Prentice Hall: 2002.

[15] L. Sendur and I. W. Selesnick "Bivariate shrinkage with local variance estimation," *IEEE Signal. Processing Letters*, vol.9, no.12, pp. 438-441 December 2002.

[16] J. Portilla, V. Strela, M. Wainwright and E. Simoncelli, "Adaptive Wiener denoising using a Gaussian scale mixture model," IEEE Signal Processing, ICIP 2001.

[17] S.G. Chang, B .Yu and M. Vetterli," Adaptive wavelet thresholding forimage denoising and compression," *IEEE Transection, Image Processing*" 9 (9): 1522-1531, 2000.

[18] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Transaction Image Processing*, vol. 16, no. 8, August 2007.

[19] I. Daubechies, Ten Lectures on Wavelets. Singapore: SIAM, 1992.

[20] R.W.D. Nickalls, "A new approach to solving the cubic: Cardan's solution revealed," *The Mathematical Gazette*, vol. 77,1993, pp. 354-359.

Pichid Kittisuwan received the B.Eng. and M.Eng degree from Kasetsart University and Chulalongkorn University, Thailand, in 2005 and in 2008, respectively. His research interests include Image Processing, Image denoising, and Wavelet applications.

Sanparith Marukatat received the License and Maitrise degree from University of French-Computer. He has finished his DEA (a kind of French one-year Master degree) and his doctoral degree at University of Paris 6 in 2000 and 2004 respectively, He is currently a researcher in Information Tecnology laboratory at NECTEC. His current research interests include Image processing.

Thitiporn Chanwimaluang received the B.Eng. degree in electrical engineering from Chulalongkorn University, Bangkok, Thailand, the M.S. degree in Electrical Engineering from the Pennsylvania State University, University Park, in 1996 and 2001, respectively, and the Ph.D. degree from the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater. From 1996 to 1997, she was with Siemens, Thailand. From 1997 to 1998, she was with Thai Telephone and Telecommunications, Thailand. She is currently with the National Electronics and Computer Technology Center (NECTEC), Klong Luang, Thailand. Her current research interests include Image processing, Image registration, Image segmentation, and 3-D reconstruction.

Widhyakorn Asdornwised is Currently an Assistant Profressor at Department of Electrical Engineering, Chulalongkorn University, Thailand. From 2007 to 2010; he has been the Head of the Communication Division, Chulalongkorn University. His research interests include Image denoising, multiple classifier systems, Invariant pattern recognition, and Wavelet applications.