Accurate Prior Modeling in the Locally Adaptive Window-Based Wavelet Denoising

Yun-Xia Liu[1,2,3(✉)], Yang Yang[4], and Ngai-Fong Law[5]

[1] School of Information Science and Engineering, University of Jinan, Jinan, China
ise_liuyx@ujn.edu.cn
[2] Shandong Provincial Key Laboratory of Network Based Intelligent Computing, Jinan, China
[3] School of Control Science and Engineering, Shandong University, Jinan, China
[4] School of Information Science and Engineering, Shandong University, Jinan, China
[5] Centre for Signal Processing, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong, China

Abstract. The locally adaptive window-based (LAW) denoising method has been extensively studied in literature for its simplicity and effectiveness. However, our statistical analysis performed on its prior estimation reveals that the prior is not estimated properly. In this paper, a novel maximum likelihood prior modeling method is proposed for better characterization of the local variance distribution. Goodness of fit results shows that our proposed prior estimation method can improve the model accuracy. A modified LAW denoising algorithm is then proposed based on the new prior. Image denoising experimental results demonstrate that the proposed method can significantly improve the performance in terms of both peak signal-to noise ratio (PSNR) and visual quality, while maintain a low computation.

Keywords: Image denoising · Orthogonal wavelet transform · Adaptive parameter estimation · Maximum likelihood estimation · Visual quality

1 Introduction

Image denoising has drawn huge research interest [1–3]. It is, in fact, an important engineering application as noise is unavoidable during image capture and transmission. Due to effective modeling of both image and noise in wavelet domain, wavelet methods have been widely studied and have proved to provide good performances [4–12].

Accurate modeling of the wavelet coefficients is of vital importance to the success of the denoising tasks. More and more sophisticated coefficients representation models [8, 13] are adopted for better denoising performance. For example, the locally adaptive window-based (LAW) denoising method [4] adopts a doubly stochastic process (DSP) model [8]. Utilizing the MAP estimator and assuming an exponential density for the local variance, LAWMAP achieves good performance at low computational cost.

BiShrink [6] incorporates the inter-scale relationship by modeling the joint parent-child distribution. ProbShrink [7] assumes a generalized Laplacian prior and models the denoising problem as detection of "signal of interest". The SURE-Let [9] method, making use of Stein's unbiased risk estimator (SURE) and linear expansion of threshold (Let), is considered to be the state-of-the-art orthogonal wavelet domain denoising method.

LAWMAP is widely studied due to its simplicity and efficiency. It is now returning to researcher's horizon due to recent rapid developments in photo-response nonuniformity noise based image forensics, where it is adopted as the denoiser [14]. There have been extensive discussion on accurate modeling within the LAW framework. For example, non-parametric approach is adopted in [11] for marginal distribution modeling. A bilateral scheme [12] is proposed for better estimation of the local variance.

An accurate prior modeling algorithm that better characterized the local variance filed is proposed in this paper. Firstly, our statistical analysis reveals that the prior modeling in LAWMAP is improper in that the choice of a key parameter λ that characterizes the exponential prior is not optimal. Secondly, we propose a maximum likelihood (ML) method for estimating λ. Goodness of fit experiments are carried out on abundant images to investigate the modeling accuracy of the proposed prior estimation scheme. Results show that our proposed prior is more efficient in characterizing the global subband statistics of wavelet coefficients than the original approach. Thirdly, as an application to image denoising, the modified LAWMAP algorithm gains significant performance improvement as compared to its original approach, with no extra computation cost. These improvements lead the modified LAWMAP an appealing competitor among orthogonal wavelet domain denoisers, especially for texture regions in images.

2 LAWMAP Denoising

Consider the general denoising scenario that the clean wavelet coefficients x are corrupted by zero mean additive Gaussian white noise n with standard deviation σ_n. Then,

$$y(k) = x(k) + n(k), \tag{1}$$

where y is observed noisy coefficients and k is the position index in wavelet domain. The aim of denoising is to recover x as accurate as possible.

2.1 Doubly Stochastic Process Model

In the wavelet domain doubly stochastic process model, wavelet coefficient x and its neighboring coefficients are assumed to be independently drawn from a zero-mean Gaussian distribution $N(0, \theta)$ given their variances θ. The local variance θ is modeled as independently, identically distributed, spatially highly correlated random variables following a distribution $\pi(\theta)$.

Among all $\pi(\theta)$ candidates which characterize the high-peak and heavy-tail properties of local variance, exponential density

$$\pi(\theta) = \lambda e^{-\lambda\theta}, \quad for \quad \theta \geq 0 \tag{2}$$

is adopted [4] as it results in close-form MAP variance estimator, where λ is known as the rate parameter.

2.2 Two-Step Minimum Mean Square Error Estimator

The LAW denoising method [4] consists of two steps: (1) estimation of the variance field for each wavelet coefficient; and (2) a pointwise local wiener filtering. See Fig. 1 for the block diagram of the algorithm.

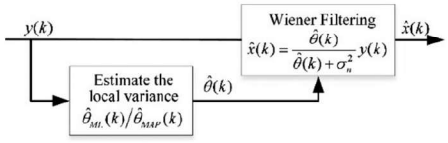

Fig. 1. Block-diagram of the LAW denoising algorithm.

For estimation of the underlying variance field, we can resort to either the ML or the MAP estimation method. Given the current coefficient $y(k)$ and its neighborhoods in a square window $\mathcal{N}(k)$, the ML estimator $\hat{\theta}_{ML}(k)$ is given by

$$\hat{\theta}_{ML}(k) = \max\left(0, \sum_{j\in \mathrm{N}(k)} y^2(j) \Big/ M - \sigma_n^2\right) \tag{3}$$

where M denotes the number of coefficients in $\mathcal{N}(k)$. The approximate MAP estimator $\hat{\theta}_{MAP}(k)$ considers also the prior marginal exponential distribution $\pi(\theta)$ as

$$\hat{\theta}_{MAP}(k) = \max\left(0, \frac{M}{4\lambda}\left[-1 + \sqrt{1 + \frac{8\lambda}{M^2}\sum_{j\in N(k)} y^2(j)}\right] - \sigma_n^2\right) \tag{4}$$

has shown to consistently outperform $\hat{\theta}_{ML}(k)$ [4].

In the second step, LAWMAP adopts the minimum mean square error (MMSE) criteria for denoising. Given $\hat{\theta}_{MAP}(k)$, the denoised wavelet coefficient is obtained as,

$$\hat{x}(k) = \frac{\hat{\theta}(k)}{\hat{\theta}(k) + \sigma_n^2} y(k) \tag{5}$$

3 Accurate Prior Modeling of LAWMAP

From the framework of LAWMAP, we see that λ is a key parameter that characterizes the global distribution of the variance field of the subband, thus has a direct impact on the estimation accuracy.

3.1 Analysis of Prior Modeling in LAWMAP

In LAWMAP [4], the parameter λ is set to be the inverse of the standard deviation of wavelet coefficients that was initially denoised using $\hat{\theta}_{ML}(k)$ in Eqs. (3) and (5).

Assume $\hat{\theta}_{ML}(k)$ provides an accurate estimation of the real variance field, i.e., $\hat{\theta}_{ML}(k)$ is very close to $\theta(k)$. From Eq. (1) we get

$$\mathrm{var}[\hat{x}(k)] = \left(\frac{\hat{\theta}_{ML}(k)}{\hat{\theta}_{ML}(k)+\sigma_n^2}\right)^2 \mathrm{var}[y(k)] \approx \left(\frac{\theta(k)}{\theta(k)+\sigma_n^2}\right)^2 \left(\theta(k)+\sigma_n^2\right) = \frac{\theta^2(k)}{\theta(k)+\sigma_n^2} \quad (6)$$

Thus the prior λ_1 used in [4] can be rewritten as

$$\lambda_1 = \frac{1}{std[\hat{x}(k)]} = \frac{\sqrt{\theta(k)+\sigma_n^2}}{\theta(k)} \quad (7)$$

An exponential density has a mean value of $1/\lambda$. This indicates that $\pi(\theta)$ is fitted with a variance field whose mean value is

$$\frac{1}{\lambda_1} = \frac{\theta(k)}{\sqrt{\theta(k)+\sigma_n^2}} < \sqrt{\theta(k)+\sigma_n^2} \quad (8)$$

which is related to $\sqrt{\theta(k)}$, i.e., the *local standard deviation*. This ***contradicts*** with the initial idea of using an exponential prior for modeling the *variance* field. Hence, this leaves room for an improved prior modeling.

3.2 Proposed Method for Estimating λ

As compared with $\hat{x}(k)$, the ML variance estimators $\hat{\theta}_{ML}(k)$ are more informative for estimating λ that characterize the global subband statistics. We thus propose to use maximum likelihood estimator of $\hat{\theta}_{ML}(k)$ within the same subband for estimating λ.

Assume $\{\theta_1, \theta_2, \ldots, \theta_N\}$ are independent exponentially distributed random variables, the maximum likelihood estimator can be proved to be

$$\hat{\lambda}_{ML} = N \Big/ \sum_{i=1}^{N} \theta_i \quad (9)$$

Notice from Eq. (3) that $\hat{\theta}_{ML}(k)$ can be truncated to zero. This will disturb the ML estimator and thus they are left out. Hence, the proposed prior λ_2 is obtained as,

$$\lambda_2(b) = \frac{N(b)}{\sum_{k=1}^{N(b)} \hat{\theta}_{ML}(k)} \tag{10}$$

where b is the index for bandpass subbands, and $N(b)$ is the number of non-zero $\hat{\theta}_{ML}(k)$ in subband b.

3.3 Goodness of Fit Test for Prior Models λ_1 and λ_2

To investigate the accuracy of the two prior models in representing the global subband statistics, we compare several goodness of fit measures [15] between the empirical histogram H and the two estimated exponential probability density function (pdf)s. As the exponential distribution is continuous, we have to first construct the discrete probability mass functions (pmf) for the estimated $E(\lambda_1)$ and $E(\lambda_2)$. The following four goodness of fit measures are considered:

$$\begin{array}{ll} \text{SS:} & \sum (h_i - e_i)^2 \\ \text{WSS:} & \sqrt{\sum (h_i - e_i)^2 \cdot h_i} \\ \text{KLD:} & \sum h_i \cdot \log(h_i/e_i) \\ \chi^2: & \sum \frac{(h_i - e_i)^2}{e_i} \end{array} \quad i = 1, \ldots, n_b \tag{11}$$

where n_b denotes the total number of bins in the pmf, h_i and e_i are pmf values of the empirical histogram H and the estimated exponential E at bin i respectively. SS is the sum of square errors between empirical histogram and the estimated exponential priors. WSS is a variant of SS, it is enhanced by weighting of the empirical histogram. Kullback-Leibler divergence (KLD) which is a measure originate from information theory depicts extra number of bits needed to code samples from H using codes based on E. χ^2 is the famous Pearson's chi-square statistic. For all of the four measures, smaller values indicate higher modeling quality.

The image set consists of eight 512 × 512 images: Lena, Barbara, Boat, Peppers, Bicycle, Fingerprint, Flinstones, Al and eight 256 × 256 images: Baboon, House, Coco, Mit, Bridge, Cameraman, Parrot, Zoneplate (See Fig. 2). These images are diverse in their contents. Some have rich smooth areas such as portrait images, natural scenes, cartoons, some have sharp edges and some are full of textures, e.g. fingerprints, architectures. Square shaped 5 × 5 window is adopted for spatial adaptivity. Orthogonal wavelet transform with five levels of decomposition and "sym8" filter are used. The two prior modeling schemes are compared for all 15 bandpass subbands.

Table 1 shows four indicators: the minimum, the maximum, the mean and the standard deviation for the image set with n_b = 256. From results in Table 1, we see that the proposed λ_2 prior modeling method produces consistent smaller goodness of fit values as compared with the original approach λ_1. This theoretically justifies its fitting

Fig. 2. Image Sets.

Table 1. Goodness of fit measures for prior models λ_1 and λ_2.

Goodness of fit measures		Min	Max	Mean	Std
SS	λ_1	0.0002	0.6455	0.0829	0.1302
	λ_2	**0.0001**	**0.549**	**0.0594**	**0.0963**
WSS	λ_1	0.0022	0.7369	0.1189	0.1649
	λ_2	**0.002**	**0.6472**	**0.0974**	**0.1334**
KLD	λ_1	0.0132	3.9475	0.8365	0.7301
	λ_2	**0.0127**	**3.1989**	**0.6555**	**0.6075**
χ^2	λ_1	0.0211	37.8918	3.604	6.1336
	λ_2	**0.0197**	**30.25**	**2.1792**	**3.8717**

superiority in modeling. Furthermore, similar results have been observed when we tried different parameters for n_b (varing from 256 to 64, 128 and 512) and noise levels (varing from 10 to 100). This indicates that the proposed λ_2 prior modeling method consistently performs better than λ_1, irrespective of the values of n_b and the noise levels.

Figure 3 depicts the histogram of the estimated variance for a typical subband of the "Barbara" image in log scale, together with the two fitted exponential priors. We can clearly see that λ_2 better describes the subband characteristics than λ_1, especially in the modeling of the heavy tail characteristics of the wavelet coefficients. Thus we can expect performance improvement in denoising images due to this accurate prior modeling.

4 Application in Image Denoising

In this part, we apply the proposed prior modeling method in image denosing, and investigate how the accurate modeling of local variance field could benefit LAWMAP image denoising.

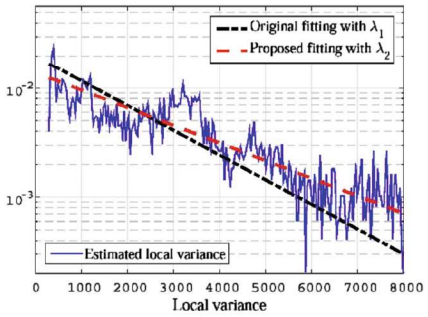

Fig. 3. Comparison of different prior modeling schemes. (Color figure online)

4.1 The Modified LAWMAP Image Denoising Algorithm

In our modified LAWMAP method, the local variances are replaced with those estimated by λ_2 modeling. The algorithm can be summarized as follows:

1. Perform an L-level orthogonal wavelet transform of the noisy image to get the noisy wavelet coefficients y.
2. For each of the wavelet coefficient $y(k)$ within subband b,
   (a) Compute the maximum likelihood estimator $\hat{\theta}_{ML}(k)$ using Eq. (3);
   (b) Estimate exponential prior λ_2 using Eq. (10) and get $\hat{\theta}_{MAP}(k)$ using Eq. (4);
   (c) Estimate the denoised wavelet coefficient $x(k)$ using Eq. (5).
3. Obtain the denoised image via inverse wavelet transform of $\hat{x}$.

4.2 Denoising Performance Comparison with Other Wavelet Algorithms

To compare the proposed denoising method with the state-of-the-art orthogonal wavelet-domain algorithms, we carried out comprehensive experiments on images in Fig. 2. Only results of three images were reported in Table 2 due to the space limitation, where each figure had been averaged over ten noise realizations. Results of other methods were obtained either by programs provided by authors [4, 6, 7, 9, 11] or simulated with parameters suggested in the paper [12]. For the proposed method, orthogonal wavelet transform with five levels of decomposition and "sym8" filter were used. Considering both estimation robustness and spatial adaptivity, window size was set to 5×5.

As is evident in Table 2, the proposed algorithm performs almost the best among the eight methods for different types of images at varying noisy levels. It consistently outperforms LAWML and LAWMAP λ_1 method [4], averaged at 1.66 dB and 0.54 dB, respectively. By accurate modeling, the LAW method with bilateral variance estimation (Law-B) [12] and the proposed method show clear advantage for images with rich textures and details (e.g. "Barbara") over the other competitors. The proposed method is more efficient at noisy circumstances though it reports slight lower PSNR (within 0.02 dB) at low noisy levels. PSNR gain over ProbShrink [7] and the non-parametric

Table 2. PSNR comparison with other orthogonal wavelet domain algorithms (in dB).

Image	Methods	Noise Standard Deviation σ_n								
		10	15	20	25	30	40	50	75	100
Lena	LAWML [4]	34.14	31.99	30.44	29.21	28.20	26.56	25.26	22.77	20.93
	LAWMAP [4]	34.32	32.37	31.01	29.99	29.17	27.89	26.94	25.26	24.12
	BiShrink [6]	34.32	32.47	31.16	30.15	29.34	28.09	27.14	25.44	24.29
	ProbShrink [7]	34.08	32.15	30.81	29.8	28.97	27.68	26.74	25.10	24.04
	SURE-Let [9]	34.56	32.67	31.35	30.32	29.53	28.26	27.33	25.69	**24.61**
	NonPara [11]	34.27	32.09	30.51	29.26	28.23	26.59	25.28	22.77	20.97
	Law-B [12]	34.60	32.48	30.91	29.69	28.68	27.03	25.76	23.35	21.52
	Proposed	**34.63**	**32.75**	**31.43**	**30.41**	**29.59**	**28.32**	**27.37**	**25.70**	24.58
Barbara	LAWML [4]	32.54	30.10	28.44	27.17	26.17	24.62	23.44	21.30	19.75
	LAWMAP [4]	32.58	30.20	28.62	27.43	26.52	25.17	24.2	22.62	21.67
	BiShrink [6]	32.14	29.82	28.23	27.06	26.13	24.81	23.88	22.47	21.65
	ProbShrink [7]	32.34	29.88	28.24	27.02	26.08	24.68	23.69	22.28	21.45
	SURE-Let [9]	32.19	29.67	27.98	26.76	25.84	24.54	23.72	22.52	21.80
	NonPara [11]	32.59	30.24	28.59	27.32	26.3	24.72	23.49	21.30	19.75
	Law-B [12]	**32.95**	**30.61**	28.95	27.70	26.69	25.15	23.95	21.79	20.23
	Proposed	32.94	30.59	**29.01**	**27.84**	**26.91**	**25.52**	**24.52**	**22.84**	**21.91**
Coco	LAWML [4]	35.31	32.92	31.23	29.91	28.79	27.05	25.63	23.05	21.10
	LAWMAP [4]	35.61	33.46	31.96	30.85	29.96	28.57	27.58	25.75	24.49
	BiShrink [6]	36.11	34.09	32.66	31.58	30.71	29.29	28.29	26.47	25.21
	ProbShrink [7]	35.35	33.08	31.47	30.39	29.48	28.20	27.09	25.35	24.20
	SURE-Let [9]	36.22	34.13	32.63	31.55	30.58	29.17	28.07	26.18	24.80
	NonPara [11]	35.33	32.89	31.18	29.75	28.71	26.92	25.44	22.82	20.95
	Law-B [12]	35.87	33.50	31.83	30.54	29.56	27.70	26.17	23.65	21.64
	Proposed	**36.31**	**34.21**	**32.81**	**31.75**	**30.84**	**29.55**	**28.46**	**26.70**	**25.41**

(NonPara) [11] methods are prominent. BiShrink [6] mainly models the parent-child interscale relationship. It performs relatively well for smooth regions (e.g. "Lena", "Coco"), but is inefficient for characterization of high-frequency detailed textures. The proposed algorithm produces significant better PSNR and visual quality in texture regions as compared with the SURE-Let [9] method, which only takes care of the inter-scale relationships. Note that all of these benefit is due to the accurate modeling of local variance field, which maintains the LAW model's simplicity and adds no extra computation cost to the algorithm.

For subjective visual quality comparison, Fig. 4 shows some of the denoised images by different orthogonal wavelet domain denoising methods. As shown in Fig. 4 (i), the texture details in Lena's hat are better preserved and less influenced by artifacts (see parts in the left highlighted ellipses) due to the accurate prior modeling. In Fig. 4 (ii), we can observe that denoised image (h) by the proposed method also demonstrate superiority in denoising smooth regions. Furthermore, from the comparison in Fig. 4 (iii) we can learn that the visual advantage over other methods is more obvious at high

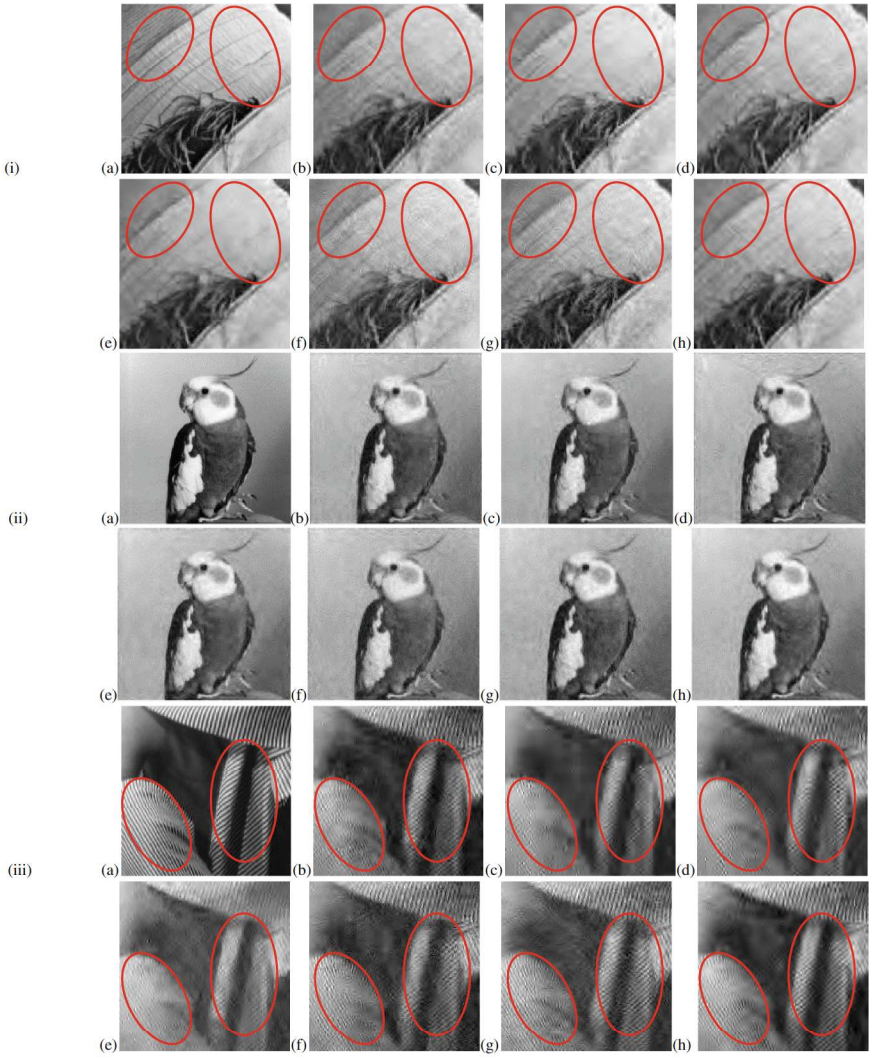

Fig. 4. Subjective quality comparison of various orthogonal wavelet denoising methods for: (i) "Lena" (σ_n = 25), (ii) "Coco" (σ_n = 25) and (iii) "Barbara" (σ_n = 50), where (a) Clean image; (b) LAWMAP [4]; (c) BiShrink [6]; (d) ProbShrink [7]; (e) SURE-Let [9]; (f) NonPara [11]; (g) LawB [12] and (h) the proposed method.

noisy levels. The proposed method recovers the texture details in Barbara's trousers and shawl best as compared with other methods, with least ringing artifacts.

To further test the robustness of the proposed method on different types of images, we have conducted denoising over the image set shown in Fig. 2 at nine different noise levels as in Table 2. Figure 5 depicts the averaged PSNR with respect to the computation time. It is shown that the computational complexities of the proposed λ_2 and the

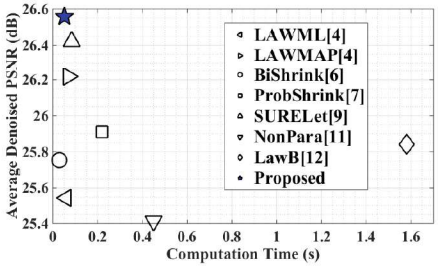

Fig. 5. Denoising performance and computation time comparison for different orthogonal wavelet domain denoising algorithms.

original λ_1 schemes are essentially the same, however, λ_2 improves the average PSNR of about 0.34 dB as compared to λ_1. PSNR advantages over BiShrink [6] and ProbShrink [7] methods are prominent. By accurately modeling of the variance field, the proposed method performs better than the state-of-the-art SURE-Let by 0.14 dB. Computational complexity superiority over NonPara [11] and LawB [12] algorithms are prominent. Hence our proposed method maintains good performance in computational efficiency as well as denoising effectiveness.

5 Conclusions

Accurate modeling of wavelet coefficients is of vital importance in image denoising. Our statistical analysis on the LAWMAP algorithm shows that the prior estimation is not proper in the sense that it cannot model the global distribution of wavelet domain local variance accurately. This motivates us to develop a maximum likelihood estimator for λ. Experimental results show that our proposed method can significantly improve the original algorithm while maintain a low computational complexity. This makes the LAWMAP method a competitive choice for wavelet based denoising. In the future, we will use the parent and child information [9, 10] to capture the inter-scale relationship so that a more accurate estimation of the local variance can be obtained.

Acknowledgement. This work was supported by the National Nature Science Foundation of China (No. 61203269, No. 61305015), Postdoctoral Science Foundation of China (No. 2015M580591). Yun-Xia LIU acknowledges the research scholarships provided by the Centre for Multimedia Signal Processing, Department of Electronic and Information Engineering and the Hong Kong Polytechnic University where the work is partially done.

References

1. Liu, Y., Peng, Y., Qu, H., et al.: Energy-based adaptive orthogonal FRIT and its application in image denoising. Sci. China Inf. Sci. **50**(2), 212–226 (2007)
2. Shen, X., Wang, K., Guo, Q.: Local thresholding with adaptive window shrinkage in the contourlet domain for image denoising. Sci. China Inf. Sci. **56**(9), 61–69 (2013)

3. Liu, Y., Law, N., Siu, W.: Patch based image denoising using the finite ridgelet transform for less artifacts. J. Vis. Commun. Image Represent. **25**(5), 1006–1017 (2014)
4. Mihcak, M., Kozintsev, I., Ramchandran, K., et al.: Low complexity image denoising based on statistical modeling of wavelet coefficients. IEEE Signal Process. Lett. **6**(12), 300–303 (1999)
5. Chang, S., Yu, B., Vetterli, B.: Spatially adaptive wavelet thresholding with context modeling for image denoising. IEEE Trans. Image Process. **9**(9), 1522–1531 (2000)
6. Sendur, L., Selesnick, I.: Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. IEEE Trans. Signal Process. **50**(11), 2744–2756 (2002)
7. Pizurica, A., Philips, W.: Estimating the probability of the presence of a signal of interest in multiresolution single and multiband image denoising. IEEE Trans. Image Process. **15**(3), 645–665 (2006)
8. Tan, S., Jiao, L.: Multivariate statistical models for image denoising in the wavelet domain. Int. J. Comput. Vision **75**(2), 209–230 (2007)
9. Luisier, F., Blu, T., Unser, M.: A new SURE approach to image denoising: interscale orthonormal wavelet thresholding. IEEE Trans. Image Process. **16**(3), 593–606 (2007)
10. Boubchir L., Naitali A., Petit E.: Multivariate statistical modeling of images in sparse multiscale transforms domain. In: 17th IEEE International Conference on Image Processing, Hong Kong, pp. 1877–1880 (2010)
11. Tian, J., Chen, L., Ma, L.: A wavelet-domain non-parametric statistical approach for image denoising. IEICE Electron. Express **7**(18), 1409–1415 (2010)
12. Shi, J., Liu, Z., Tian, J.: Bilateral signal variance estimation for wavelet-domain image denoising. Sci. China Inf. Sci. **56**(6), 83–88 (2013)
13. Liu, J., Moulin, P.: Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients. IEEE Trans. Image Process. **10**(11), 1647–1658 (2001)
14. Chen, M., Fridrich, J., Goljan, M., et al.: Determining Image Origin and Integrity Using Sensor Noise. IEEE Trans. Inf. Forensics Secur. **3**(1), 74–90 (2008)
15. Pi, M., Tong, C., Choy, S., et al.: A fast and effective model for wavelet subband histograms and its application in texture image retrieval. IEEE Trans. Image Process. **15**(10), 3078–3088 (2006)