Image Denoising Using Bivariate α-Stable Distributions in the Complex Wavelet Domain

Alin Achim and Ercan E Kuruo$\breve{g}$lu

Signals and Images Laboratory
Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"
Area della Ricerca CNR di Pisa
Via G. Moruzzi 1
56124 Pisa, ITALY
{*Alin.Achim, Ercan.Kuruoglu*}*@isti.cnr.it*

Abstract

Recently, the dual-tree complex wavelet transform has been proposed as a novel analysis tool featuring near shift-invariance and improved directional selectivity compared to the standard wavelet transform. Within this framework, we describe a novel technique for removing noise from digital images. We design a bivariate *maximum a posteriori* (MAP) estimator, which relies on the family of isotropic α-stable distributions. Using this relatively new statistical model we are able to better capture the heavy-tailed nature of the data as well as the interscale dependencies of wavelet coefficients. We test our algorithm for the Cauchy case, in comparison with several recently published methods. The simulation results show that our proposed technique achieves state-of-the-art performance in terms of root mean squared error.

Index Terms

Wavelet transform, alpha-stable distributions, bivariate models, MAP estimation, Monte-Carlo methods.

I. INTRODUCTION

In the last decade, there has been considerable interest in using multiscale decompositions as a framework to develop algorithms aiming at recovering signals from noisy data. Many of these algorithms have been developed based on Donoho's pioneering work [1] on soft thresholding. However, wavelet shrinkage techniques have come a long way in recent years due to the better statistical models adopted for the wavelet coefficients. For example, in a number of recent publications [2], [3], [4], it has been shown that alpha-stable distributions, a family of heavy-tailed densities, are sufficiently flexible and rich to appropriately model wavelet coefficients of images in various applications. On the other hand, algorithms that exploit dependencies between coefficients could achieve better results compared with the ones based on the independence assumption [5], [6], [7], [8]. In this context, Sendur and Selesnick [7] have developed a *maximum a posteriori* (MAP) estimator based on a circular-symmetric Laplacian model for a coefficient and its parent.

Our approach for image enhancement is similar to the method reported in [7] in that we implement a wavelet-domain Bayesian denoising processor that takes into account the correlation of wavelet coefficients in two adjacent scales. The innovative aspect of the present work is twofold: First, we propose a new representation for modeling the interscale dependencies of wavelet coefficients in visual image, which is based on the family of isotropic alpha-stable distributions. Second, we propose a maximum-likelihood method for estimating the parameters of the alpha-stable distribution from noisy observations by making use of Monte Carlo integration techniques.

The paper is organized as follows: In Section II, we provide some preliminaries on alpha-stable processes with a special emphasis on bivariate models. The design of a bivariate Bayesian estimator that exploits the signal heavy-tailed nature and interscale correlations is described in Section III. Section IV compares the performance of our proposed algorithm with the performance of other current wavelet-based denoising methods applied on two test images. Finally, Section V draws conclusions and describes future work directions.

II. THE ALPHA-STABLE FAMILY OF DISTRIBUTIONS

This section provides a brief, necessary overview of the alpha-stable statistical model used to characterize wavelet coefficients of natural images. Since our interest is in modeling wavelet coefficients, which are symmetric in nature, we restrict our exposition to the case of symmetric alpha-stable ($S\alpha S$) distributions. For detailed accounts of the properties of the general stable family, we refer the reader to [9] and [10].

A. Univariate Symmetric Alpha-Stable Distributions

The appeal of $S\alpha S$ distributions as a statistical model for signals derives from two main theoretical reasons. First, stable random variables satisfy the stability property which states that linear combinations of jointly stable variables are indeed stable. Second, stable processes arise as limiting processes of sums of independent identically distributed (i.i.d.) random variables via the generalized central limit theorem. Actually, the *only* possible non-trivial limit of normalized sums of i.i.d. terms is stable.

The $S\alpha S$ distribution lacks a compact analytical expression for its probability density function (pdf). Consequently, it is most conveniently represented by its characteristic function

$$\varphi(\omega) = \exp(\jmath\delta\omega - \gamma|\omega|^{\alpha}) \tag{1}$$

where α is the *characteristic exponent*, taking values $0 < \alpha \leq 2$, δ ($-\infty < \delta < \infty$) is the *location parameter*, and γ ($\gamma > 0$) is the *dispersion* of the distribution. For values of α in the interval $(1, 2]$, the location parameter δ corresponds to the mean of the $S\alpha S$ distribution,

while for $0 < \alpha \leq 1$, δ corresponds to its median. The dispersion parameter γ determines the spread of the distribution around its location parameter δ, similar to the variance of the Gaussian distribution.

The characteristic exponent α is the most important parameter of the $S\alpha S$ distribution and it determines the shape of the distribution. The smaller the characteristic exponent α is, the heavier the tails of the $S\alpha S$ density. This implies that random variables following $S\alpha S$ distributions with small characteristic exponents are highly impulsive. One consequence of heavy tails is that only moments of order less than α exist for the non-Gaussian alpha-stable family members. As a result, stable laws have infinite variance. Gaussian processes are stable processes with $\alpha = 2$ while Cauchy processes result when $\alpha = 1$.

B. Bivariate Stable Distributions

Much like univariate stable distributions, bivariate stable distributions are characterized by the stability property and the generalized central limit theorem [10]. The characteristic function of a bivariate stable distribution has the form

$$\varphi(\boldsymbol{\omega}) = \begin{cases} \exp\left(j\boldsymbol{\omega}^T\boldsymbol{\delta} - \boldsymbol{\omega}^T A\boldsymbol{\omega}\right) & for\ \alpha = 2 \\ \exp\left(j\boldsymbol{\omega}^T\boldsymbol{\delta} - \int_S \left|\boldsymbol{\omega}^T\mathbf{s}\right|^\alpha \mu(d\mathbf{s}) + j\beta_\alpha(\boldsymbol{\omega})\right) & for\ 0 < \alpha < 2 \end{cases} \tag{2}$$

where

$$\beta_\alpha(\boldsymbol{\omega}) = \begin{cases} \tan\frac{\alpha\pi}{2}\int_S \left|\boldsymbol{\omega}^T\mathbf{s}\right|^\alpha \operatorname{sign}\left|\boldsymbol{\omega}^T\mathbf{s}\right| \mu(d\mathbf{s}) & for\ \alpha \neq 1,\ 0 < \alpha < 2 \\ \int_S \boldsymbol{\omega}^T\mathbf{s} \log\left|\boldsymbol{\omega}^T\mathbf{s}\right| \mu(d\mathbf{s}) & for\ \alpha = 1 \end{cases} \tag{3}$$

and where $\boldsymbol{\omega} = (\omega_1, \omega_2)$, $|\boldsymbol{\omega}| = \sqrt{\omega_1^2 + \omega_2^2}$, and $\boldsymbol{\delta} = (\delta_1, \delta_2)$. S is the unit circle, the measure $\mu(\cdot)$ is called the *spectral measure* of the α-stable random vector and A is a positive semidefinite symmetric matrix.

Unlike univariate stable distribution, bivariate stable distributions form a nonparametric set being thus much more difficult to describe. An exception is the family of multidimensional isotropic stable distributions who's characteristic function has the form

$$\varphi(\omega_1, \omega_2) = \exp(\jmath(\delta_1\omega_1 + \delta_2\omega_2) - \gamma|\boldsymbol{\omega}|^\alpha) \tag{4}$$

The distribution is isotropic with respect to the location point (δ_1, δ_2). The two marginal distributions of the isotropic stable distribution are $S\alpha S$ with parameters $(\delta_1, \gamma, \alpha)$ and $(\delta_2, \gamma, \alpha)$. Since our further developments are in the framework of wavelet analysis, in the following we will assume that $(\delta_1, \delta_2) = (0, 0)$. The bivariate *isotropic Cauchy* and *Gaussian* distributions are special cases for $\alpha = 1$ and $\alpha = 2$, respectively. The bivariate pdf in these two cases can be written as:

$$P_{\alpha,\gamma}(x_1, x_2) = \begin{cases} \frac{\gamma}{2\pi(x_1^2 + x_2^2 + \gamma^2)^{3/2}} & \text{for } \alpha = 1 \\ \frac{1}{4\pi\gamma}\exp\left[-\frac{x_1^2 + x_2^2}{4\gamma}\right] & \text{for } \alpha = 2. \end{cases} \tag{5}$$

Figure 1 shows an example of a two-dimensional Cauchy surface. The important observation regarding this plot is that although the bivariate Cauchy density behaves approximately like a Gaussian density near the origin, its tails decay at a lower rate compared with a Gaussian density.

As in the case of the univariate $S\alpha S$ density function, when $\alpha \neq 1$ and $\alpha \neq 2$, no closed form expressions exist for the density function of the bivariate stable random variable. Kuruoglu et al. [11] have shown that approximate densities can be obtained by expressing the 2-D $S\alpha S$ pdf as Gaussian mixtures with an appropriate number of components. A direct numerical calculation algorithm has been developed by Nolan [12].

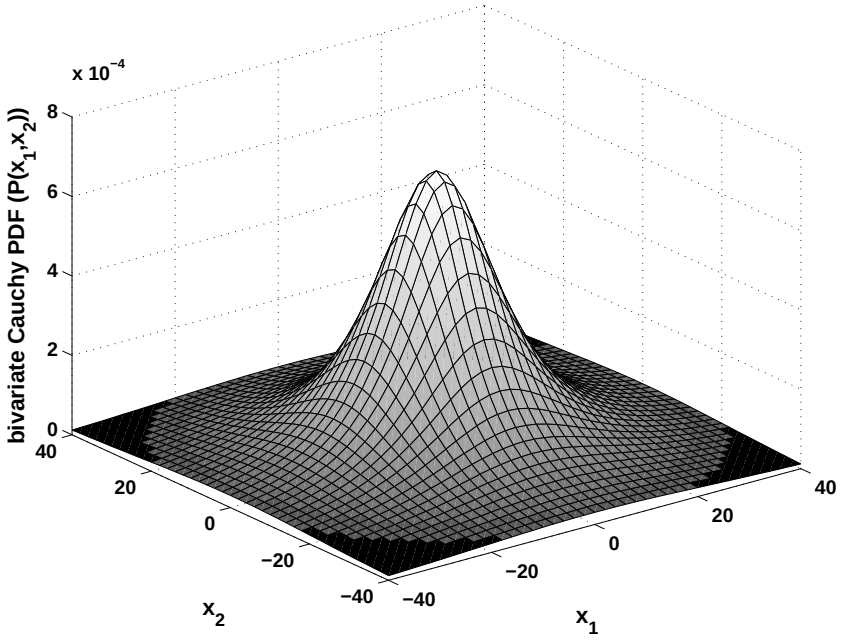

Fig. 1. Two dimensional Cauchy density surface ($\alpha = 1$, $\gamma = 15$).

III. The proposed algorithm

Our approach for image denoising uses the classical 3-steps technique generally employed in wavelet-based methods: i) analysis of raw data by means of wavelet transform, ii) noisy coefficients shrinkage using an appropriately designed algorithm, and iii) synthesis of the denoised image from the processed wavelet coefficients through the inverse wavelet transform. One should note at this point that in our implementation we use an overcomplete representation, namely the dual-tree complex wavelet transform (DTCWT) proposed by Kingsbury [13], [14]. The advantage of doing so is twofold: First, the DTCWT is nearly shift-invariant, so it keeps us away from unwanted effects like pseudo-Gibbs phenomena. Also, it offers a good directional selectivity property that the classical separable wavelet transform does not possesses. The cost to be paid is only a small increase in signal redundancy and computational load.

A. Bivariate MAP Processor for Noise Removal

The wavelet transform is a linear operation. Consequently, after decomposing an image we get, in each of the six oriented subbands and for every two adjacent levels, sets of noisy wavelet coefficients represented as the sum of the transformations of the signal and of the noise:

$$\begin{aligned} y_j &= x_j + n_j \\ y_{j+1} &= x_{j+1} + n_{j+1} \end{aligned} \tag{6}$$

where $1 < j < J$ refers to the decomposition level. The above set of equations can be written in vectorial form as:

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \tag{7}$$

where $\mathbf{y} = (y_1, y_2)$, $\mathbf{x} = (x_1, x_2)$, $\mathbf{n} = (n_1, n_2)$. The MAP estimator of $\mathbf{x}$ given the noisy observation $\mathbf{y}$ can be easily derived as being:

$$\hat{\mathbf{x}}(\mathbf{y}) = \arg\max_{\hat{\mathbf{x}}} P_{\mathbf{x}|\mathbf{y}}(\mathbf{x}\,|\mathbf{y}) \tag{8}$$

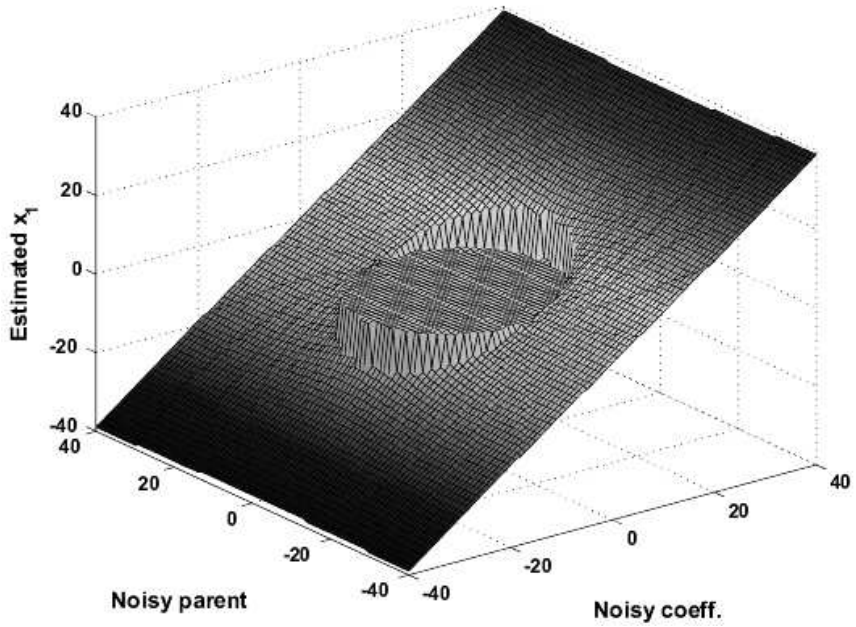

Fig. 2. Numerically computed MAP surface from Eq. (10) for bivariate Cauchy signal ($\alpha_x = 1$) and Gaussian noise ($\alpha_n = 2$) prior distributions.

Bayes' theorem gives the *a posteriori* PDF of **x** based on the measured data:

$$P_{\mathbf{x}|\mathbf{y}}(\mathbf{x}\,|\mathbf{y}) = \frac{P_{\mathbf{y}|\mathbf{x}}(\mathbf{y}\,|\mathbf{x})\; P_{\mathbf{x}}(\mathbf{x})}{P_{\mathbf{y}}(\mathbf{y})}, \tag{9}$$

where $P_{\mathbf{x}}(\mathbf{x})$ is the *prior* PDF of the signal component of the measurements and $P_{\mathbf{y}|\mathbf{x}}(\mathbf{y}\,|\mathbf{x})$ is the *likelihood* function. Substituting (9) in (8), we get:

$$\begin{aligned}\hat{\mathbf{x}}(\mathbf{y}) &= \arg\max_{\hat{\mathbf{x}}} P_{\mathbf{y}|\mathbf{x}}(\mathbf{y}\,|\mathbf{x})\; P_{\mathbf{x}}(\mathbf{x}) = \arg\max_{\hat{\mathbf{x}}} P_{\mathbf{n}}(\mathbf{y}\text{-}\mathbf{x}) P_{\mathbf{x}}(\mathbf{x}) \\ &= \arg\max_{\hat{\mathbf{x}}} P_{\mathbf{n}}(\mathbf{n}) P_{\mathbf{x}}(\mathbf{x})\end{aligned} \tag{10}$$

In the above equation we use a bivariate Cauchy model for the signal component while we use a zero-mean bivariate Gaussian model for the noise component of the wavelet coefficients. In other words, the observed signal is a mixture of Cauchy signal and Gaussian noise. Figure 2 illustrates the numerically computed MAP surface using these two prior density functions with (10). The behavior of this processor incorporates the main property of the ones proposed in [3], [4], namely that large-amplitude observations are essentially preserved while small-amplitude values are suppressed. In addition, the shrinkage of a coefficient is also conditioned on the value of the corresponding coefficient at the next decomposition level (parent value): The smaller the parent value, the greater the shrinkage [7].

B. $S\alpha S$ Parameters Estimation from Noisy Observations

Naturally, in order for the processor in Eq. (10) to be of any practical use, one should be able to estimate the parameters γ_x for the signal and σ_n of the noise from the observed data. As proposed in [1], a robust estimate of the noise standard deviation, σ_n, is obtained using the median absolute deviation (MAD) of coefficients at the first level of an wavelet decomposition

$$\hat{\sigma}_n = \frac{MAD(y_1)}{0.6745}, \tag{11}$$

For the purpose of $S\alpha S$ parameters estimation from noisy observations, methods based on empirical characteristic functions have been previously proposed by Ilow et. al [15] and by Achim et al. [4]. Naturally, those methods can easily be adapted for the particular case of Cauchy distributions. Specifically, since the PDF of the measured coefficients $(\mathbf{y})$ is the convolution between the PDFs of the signal $(\mathbf{x})$ and noise components $(\mathbf{n})$, the associated characteristic function of the measurements is given by the product of the characteristic functions of the signal and noise:

$$\varphi_y(\omega) = \exp\left(-\gamma_x|\omega|\right) \cdot \exp\left(-\frac{\sigma_n^2}{2}|\omega|^2\right) \tag{12}$$

After some straightforward manipulations one gets:

$$\gamma_x = -\frac{\log \varphi_y^2(\omega) - \sigma_n^2 \, |\omega|^2}{2|\omega|} \tag{13}$$

In principle, γ_x could be thus estimated using any nonzero value of ω. However, as suggested in [15] in order to reduce the overall variance of the estimate, we chose to average the results from the estimates corresponding to many possible choices of ω.

This method is computationally efficient and provides consistent estimates, at least in large data samples. However, if one is concerned with the implementation of local algorithms in which the parameters are to be estimated in small size windows, then a more direct maximum-likelihood (ML) approach should be employed. For estimation problems involving α-stable distributions, ML methods are generally difficult to implement due to the lack of close-form expressions for their pdf. Even in the case of Cauchy distribution, if the parameter γ needs to be estimated from noisy observations, this leads to solving integrals which are intractable analytically and computationally extensive when solved by classical numerical methods. For this reason, we propose a method based on Monte Carlo (MC) integration coupled with an sample-importance re-sampling (SIR) step to increase efficiency. We will refer to the new method as Monte Carlo maximum likelihood estimation of $S\alpha S$ parameters (SαSMCMLE).

The ML estimator of parameters $\{\alpha, \gamma\}$ of a mixture of $S\alpha S$ signal and Gaussian noise can be written as:

$$\{\hat{\alpha}, \hat{\gamma}\} = \arg\max_{\hat{\alpha},\hat{\gamma}} \sum_{k=1}^{N} \log\left(\int_{-\infty}^{\infty} P_{\alpha,\gamma}\left(y_k - x\right) \mathrm{e}^{-\frac{x^2}{2\sigma^2}} dx\right) \tag{14}$$

where $P_{\alpha,\gamma}$ is the $S\alpha S$ pdf and $P_\sigma(x) = e^{-\frac{x^2}{\sigma^2}}$ is the Gaussian kernel corresponding to the noise. The integral in the above expression can be evaluated using a simple MC approximation [16], so that (14) becomes

$$\{\hat{\alpha}, \hat{\gamma}\} = \arg\max_{\hat{\alpha},\hat{\gamma}} \sum_{k=1}^{N} \log\left(\frac{1}{M}\sum_{m=1}^{M} P_{\alpha,\gamma}\left(y_k - x^m\right)\right) \tag{15}$$

where x^m represents a sample drawn from $P_\sigma \sim N(0, \sigma^2)$.

Furthermore, in order to use importance sampling we make a change of variables, rewriting the integral in (14) as

$$\int_{-\infty}^{\infty} P_{\alpha,\gamma}\left(y_k - x\right) P_\sigma(x) dx = \int_{-\infty}^{\infty} \frac{P_{\alpha,\gamma}\left(y_k - x\right) P_\sigma(x)}{f\left(x\right)} f\left(x\right) dx \tag{16}$$

If $f(x)$ is a normalized density who's support includes the support of $P_\sigma(x)$, and from which we are able to generate random draws, then the above integral can be estimated using L draws

$x^1, ..., x^L$ from $f(x)$ by the expression $\frac{1}{L}\sum_{l=1}^{L} P_{\alpha,\gamma}\left(y_k - x^l\right) w\left(x^l\right)$. The factors

$$w\left(x^l\right) = \frac{P_\sigma(x^l)}{f\left(x^l\right)}$$

are called *importance weights*, and we use a Student-t distribution with 3 degrees of freedom for the kernel f. A Cauchy distribution could have been an equally appropriate choice. Consequently, the SαSMCMLE in (15) becomes

$$\{\hat{\alpha},\hat{\gamma}\} = \arg\max_{\hat{\alpha},\hat{\gamma}} \sum_{k=1}^{N} \log\left(\frac{1}{L}\sum_{l=1}^{L} P_{\alpha,\gamma}\left(y_k - x^l\right) w\left(x^l\right)\right) \tag{17}$$

Let us note that we perform the SIR step above without replacement, which means that, according to their weights, we select only a number L of draws out of a total number of $M(L < M)$ initially generated samples. As a consequence, by expression (17) we are able to obtain an estimate as consistent as the one in (15) but using a smaller number of the generated samples, being thus computationally more efficient.

The above SαSMCMLE method is applicable for mixtures of $S\alpha S$ and Gaussian signals in general. Nevertheless, in the results presented in the next section we use it by replacing $P_{\alpha,\gamma}$ with the Cauchy pdf and performing the optimization only in respect with the parameter γ.

IV. SIMULATION RESULTS

We compared our proposed algorithm to other effective techniques from the recent literature using the standard "lena" and "boats" images. To quantify the denoising performance of each algorithm we employed the peak signal to noise ratio (PSNR) defined as:

$$PSNR = 20\log_{10}(256/\sqrt{\frac{1}{N^2}\sum_i (\hat{s}_i - s_i)^2}) \tag{18}$$

where s and $\hat{s}$ denote the noise-free and the denoised images respectively, and N^2 is the total number of pixels.

In a first set of simulations, we tested our subband adaptive (SA) method in comparison with systems that model only the marginal distribution of wavelet coefficients or that takes into account only their interscale dependency. From the first category we selected a hard thresholding (HT) method and the alpha-stable based Bayesian processor (WIN-SAR) in [4] from which we have of course taken out the LOG and EXP modules; both approaches were re-implemented using the DTCWT. From the second category we chose to include in the comparison the hidden Markov tree (HMT) model in [5] and the bivariate shrinkage (BiShrink) method [7]. The results of these experiments are shown in Table I for 5 different levels of computer-generated white Gaussian noise.

As a second set of experiments, we considered different local adaptive (LA) algorithms, in which the parameters are estimated within square-shaped neighborhoods around the reference coefficient. Apart from our system, in Table II we show results obtained with an adaptive Wiener filter implemented in the CWT domain (WienerCWT) and the local adaptive algorithm from [17] (LBishrink). For all three methods we considered windows of size 7×7 within each subband. The results refer to the "boats" image for 5 noise levels.

From the tables it can be seen that our processor achieves the best results in most situations, in both its subband adaptive and locally adaptive implementations. Although conceptually similar with Sendur's technique, we attribute the better performance of our algorithm to the better ability of the Cauchy model in fitting the tails of the underlying distribution of the signal wavelet coefficients.

TABLE I

PSNR VALUES (DB) OBTAINED BY THE 5 DENOISING METHODS APPLIED ON THE 2 TEST IMAGES.

Image	Lena					Boats				
σ	10	15	20	25	30	10	15	20	25	30
Method										
Noisy	28.18	24.65	22.14	20.17	18.62	28.16	24.65	22.15	20.15	18.62
HT	34.23	32.60	31.30	30.23	29.40	32.40	30.50	29.27	28.23	27.44
WIN-SAR	34.33	32.74	31.52	30.57	29.89	32.85	30.87	29.65	28.64	27.92
HMT	33.84	31.76	30.39	29.24	28.35	32.28	30.31	28.84	27.68	26.83
BiShrink	**34.77**	32.78	31.71	30.63	29.85	32.85	31.01	29.81	28.77	27.99
proposed-SA	34.75	**33.03**	**31.87**	**30.89**	**30.18**	**33.26**	**31.22**	**30.00**	**28.95**	**28.22**

TABLE II

PSNR VALUES (DB) OBTAINED BY 3 LOCAL ADAPTIVE DENOISING METHODS APPLIED ON THE BOATS IMAGE.

σ	Noisy	WienerCWT	LBiShrink	proposed-LA
10	28.16	**33.51**	33.10	33.09
15	24.65	31.43	31.36	**31.44**
20	22.15	29.92	30.08	**30.19**
25	20.15	28.74	29.06	**29.21**
30	18.62	27.73	28.31	**28.51**

Remember however, that in image processing applications, we are interested in suppressing noise while at the same time preserving the edges of the original image that often constitute features of interest for further image analysis tasks. Having in mind this observation, in Figure 3 we show the results from the processing of the "boats" image. As it can be seen, our proposed algorithms have mitigated noise retaining as well most of the features that are clearly distinguishable in the noisy data.

V. CONCLUSIONS

We proposed an effective wavelet-domain MAP processor that makes use of bivariate α-stable distributions to account for the interscale dependencies of natural image subbands. We tested our algorithm for the Cauchy case and we compared it with several other results reported in the recent literature. We conclude that our proposed method achieves state-of-the-art performance, being competitive with the best wavelet-based denoising systems.

We are currently exploring several ways to extend the work presented here. An interesting direction is the further improvement of the statistical image model, by considering the case of general bivariate α-stable distributions. The implementation of the method employing isotropic stable densities, with $0 < \alpha \leq 2$ in general, is straightforward, provided that we are equipped with an efficient program for computing bivariate $S\alpha S$ densities. Instead, considering general α-stable densities would be a much more challenging task, presupposing accurate estimation of the spectral measure, μ (see Eq. 2), from noisy observations. Finally, capturing intrascale dependencies of wavelet coefficients by means of α-stable Markov random fields promises to significantly improve the overall performance of a denoising scheme.

Fig. 3. Results of various denoising methods: (a) Original image; (b) Noisy image ($\sigma = 25$); (c) Subband adaptive algorithm ($PSNR = 28.95\ dB$); (d) Local adaptive algorithm ($PSNR = 29.21\ dB$)

REFERENCES

[1] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, Aug. 1994.

[2] P. Tsakalides, P. Reveliotis, and C. L. Nikias, "Scalar quantization of heavy-tailed signals," *IEE Proc. - Vision, Imag. Sign. Proc.*, vol. 147, pp. 475–484, Oct. 2000.

[3] A. Achim, A. Bezerianos, and P. Tsakalides, "Novel Bayesian multiscale method for speckle removal in medical ultrasound images," *IEEE Trans. Med. Imag.*, vol. 20, pp. 772–783, Aug. 2001.

[4] A. Achim, P. Tsakalides, and A. Bezerianos, "SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling," *IEEE Trans. Geosc. and Remote Sensing*, vol. 41, pp. 1773–1784, Aug. 2003.

[5] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based signal processing using hidden Markov models," *IEEE. Tran. Sig. Proc.*, vol. 46, pp. 886–902, April 1998.

[6] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising," *IEEE Trans. Image Processing*, vol. 11, pp. 545–557, May 2002.

[7] L. Sendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE. Tran. Sig. Proc.*, vol. 50, pp. 2744–2756, Nov. 2002.

[8] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE. Tran. Image. Proc.*, vol. 12, pp. 1338–1351, Nov. 2003.

[9] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. New York: Chapman and Hall, 1994.

[10] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*. New York: John Wiley and Sons, 1995.

[11] E. E. Kuruoglu, C. Molina, and W. J. Fitzgerald, "Approximation of α-stable probability densities using finite Gaussian mixtures," *Proc. EUSIPCO'98*, Sep. 1998.

[12] J. P. Nolan, "Multivariate stable distributions: approximation, estimation, simulation and identification," in *A Practical Guide to Heavy Tails* (R. J. Adler, R. E. Feldman, and M. S. Taqqu, eds.), Boston: Birkhauser, 1998.

[13] N. G. Kingsbury, "Image processing with complex wavelets," *Phil. Trans. R. Soc. London A*, vol. 357, pp. 2543–2560, Sept. 1999.

[14] N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering," *Appl. Compt. Harmon. Anal.*, vol. 10, pp. 234–253, May 2001.

[15] J. Ilow and D. Hatzinakos, "Applications of the empirical characteristic function to estimation and detection problems," *Sig. Proc.*, vol. 65, pp. 199–219, March 1998.

[16] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, "Simple Monte Carlo integration," in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.*, pp. 295–299, Cambridge, England: Cambridge University Press, 1992.

[17] L. Sendur and I. W. Selesnick, "Bivariate shrinkage with local variance estimation," *IEEE. Sig. Proc. Letters*, vol. 9, pp. 438–441, Dec. 2002.