An Efficient Approach to Wavelet Image Denoising

Alaa A. Hefnawy, Heba A. Elnemr

Computer & Systems Dept., Electronics Research Institute, Cairo, Egypt
alaahouse@yahoo.com, heba_elnemr@yahoo.com

ABSTRACT

This paper proposed an efficient approach to orthonormal wavelet image denoising, based on minimizing the mean square error (MSE) between the clean image and the denoised one. The key point of our approach is to use the accurate, statistically unbiased, MSE estimate—Stein's unbiased risk estimate (SURE). One of the major advantages of this method is that; we don't have to deal with the noiseless image model.Since the estimate here is quadratic in the unknown weights, the problem of findingthresholding function is downgraded to solve a linear system of equations, which is obviously fast and attractive especially for large images. Experimental results on several test images are compared with the standard denoising techniqueBayesShrink, and to benchmark against the best possible performance of soft-threshold estimate, the comparison also include Oracleshrink. Results show that the proposed technique yieldssignificantly superior image quality.

KEYWORDS

Image denoising; Orthonormal wavelet transform; Wavelet image denoising

1. INTRODUCTION

Very often, image acquisition systems are not perfectand images are, thus, corrupted by noise during their digitization.Besides, communication channels are not ideal and imagesare further degraded during their transmission. Hence, denoisingis a crucial step before image analysis.The main objective of image denoising is to reduce noise as much as possible while preserving image features. The most popular approaches of image denoisingare the transform-domainone, where the noisy images are first transformed using linear ormultiscaletransformation.Then nonlinearly processing the resulted coefficients, and finally retrieving the image by applying the inverse linear transformation. To thisrespect, the wavelet transform (WT) has emerged as the premier tool for image denoising, due to the statistically useful properties of wavelet coefficients of natural images. The sparseness property of wavelet coefficients and tendency of wavelets bases to diagonalise images allows us to break the problem into modelling a small number of 'neighbouring' coefficients (in space and scale) to reduce the dimensionality and improve the tractability of the problem. Indeed, the WT mainly concentrates the energies of many signals of interest in a few coefficients, whereas the power of the noise often spreads out over all the coefficients [1].

The simplest way to distinguish the data from the noise in the WT domain is bythresholding the wavelet coefficients.Its principle consists ofsetting to zero all the coefficients below a certainthreshold value, while either keeping the remaining ones unchanged(*hard-thresholding*) or shrinking them by the thresholdvalue (*soft-thresholding*, which was originally theorized byDonoho andJohnstone[2]).It has been shown that the shrinkage rule is near-optimal in the minimaxsense,giving an expression of the optimal thresholdvalue T (called the “universal threshold”) as a function ofthe noise power $\sigma^2$when the number of samples$N$ is large: $T = \sqrt{2\sigma^2 \log N}$. The *VisuShrink* method is known as the one where the universal threshold is used for denoisingimages [3].Although, minimax is theoreticallyattractive, themean-squared error (MSE) still considered a better different measure of error. In literature, a lot of workhas been done to findbetter alternative thresholdingmethodologies in terms of MSE than *VisuShrink*[4]–[9]. Donohoand Johnstone themselves acknowledged this flaw andhave subsequently proposedto choose the optimal threshold value *T*by minimizingStein’s unbiased risk estimator (SURE) [10].Thismethod has been called*SureShrink*by their authors [2].Manyextensions of this seminal work have been since realized (see,for example, [11]–[14]).

The Bayesian is another popular statistical approach.In this framework the unknown signal is viewed asa realization of a random field with a given prior probabilitydistribution and the objective is to make a realistic choice ofthis distribution that will yield an efficient denoising procedure.Without challenging the soft-thresholding strategy,one ofthe most popular threshold value selection techniques was proposed by Chang *et al.* [15].Assuming that the wavelet coefficientsdistribution is Gaussian, a spatiallyadaptive thresholding is performed in a Bayesian framework. This solution is known as *BayesShrink* and has a better MSE performance than *SureShrink*.

Amongthe recent many denoising algorithms, Luisier*et al.* [16]have showed that a quite competitive results compared to the best state-of-the-art denoising algorithms ([17]–[22]), could be obtained without involving sophisticated statistical image model. They used the SURE estimator—that is based onthe noisy data alone—to minimize the MSEbetween noisy and clean image.

In this paper, we will perform the denoising process in the transformation domain using orthonormal wavelet transform. Assuming white Gaussian noise, we will use Stein’sunbiased risk estimate (*SURE*).SURE can be considered as a very good estimate of MSEbetween the noiseless image and the resulteddenoisedone. The proposed thresholding function, that willminimize the SURE, is expressed as alinear expansion of thresholds, to get benefitfrom the *quadratic* form of theSURE. This solution is considered, computationally very efficient, especially for the practice large images.

2. THEORETICAL BACKGROUND

2.1. Problem Formulation

In the standard wavelet denoising problem: given noisy data $y_n = x_n + b_n$, for n=1 …. N,where x is the noiseless data(clean image),$b$is the noise, and N is the number of samples.Then by defining the wavelet and its inverse linear transformations *D*—decomposition—and*R*—reconstruction, such that *RD* =Identity. As long as the size of theinput and output data are the same, these linear operators can be characterizedby matrices;$\mathtt{D} = (d_{i,j})_{(i,j)\in[1;L]\times[1;N]}$, and $\mathtt{R} = (r_{i,j})_{(i,j)\in[1;N]\times[1;L]}$,which satisfy the *Ideal reconstruction*property$\mathtt{RD} = \mathtt{Id}$. Then, the whole denoisingprocess can be summarized in the following steps

1) Calculate the transformed noisy coefficients by applying$\mathtt{D}$ to the noisy signal y = x + b $\mathtt{w} = \mathtt{Dy} = (w_i)_{i\in[1;L]}$
2) Apply a *pointwise*thresholding function $\Theta(\mathtt{w}) = (\theta_i(w_i))_{i\in[1;L]}$

3) Revert to the original domain to calculate the denoised estimate $\hat{x} = R\Theta(w)$,through applying R to the thresholdedcoefficients $\Theta(w)$

This above algorithm can be written as a function of the noisyinput coefficients

$$\hat{x} = F(y) = R\Theta(Dy) \tag{1}$$

Our goal is, thus, to find a function ofthe noisy data alone $F(y) = (f_n(y))_{n=1,2,....N} = \hat{x}$ which willminimize the MSE defined by

$$\text{MSE} = \frac{1}{N}\|\hat{x} - x\|^2 = \frac{1}{N}\sum_{n=1}^{N}\left|\hat{x}_n - x_n{}^2\right| \tag{2}$$

To develop our denoising method, which allow usto apply deferentthresholding functionin every highpasssubband,we will follow two important assumptions:

- We will consideronly additive Gaussian noise that defined by zero mean and a σ^2 variance; i.e., $b \sim \mathcal{N}(0, \sigma^2)$, and can be accurately estimatedfrom the first decomposition level diagonal subband HH_1 by the robust and accurate median estimator [2].

$$\sigma^2 = \left[\frac{median(|w|)}{0.6745}\right]^2, w \in coefficients \rightarrow HH_1 \text{ sub band} \tag{3}$$

- We will consider only*orthonormal* wavelet transform; and hence:
  - MSE (in the space domain) = weighted sum of the MSE of each individual subband
  - In the wavelet domain, the noise stays Gaussian while keeping the same statistics

This means that our solution is"subband-adaptive"as the most of the successful wavelet denoisingmethods.

2.2. Stein's Unbiased MSE Estimate (SURE)

The peak signal-to-noise ratio (PSNR); is the most common used measure in the denoising applications, which can beexpressed as:

$$\text{PSNR} = 10\log_{10}\left(\frac{\max(x^2)}{\text{MSE}}\right) \tag{4}$$

Where, usually, for 8-bit images $\max(x^2) = 255^2$. As the noise is a random process, an expectationoperator $\varepsilon\{\ \}$, were used to estimate the potential results of the processed noisy data y. Here, the noiseless data x isnot a random process; thus $\varepsilon\{x\} = x$. Generally, the target of image denoisingapproaches is to maximize thePSNR and, which imply minimizing the MSE defined in (2).

Since we do not have access to the original signal x, wecannot compute $\|\hat{x} - x\|^2 / N$ —the *Oracle* MSE. However,regardless any restrictions on the noiseless data, we will prove that; this quantitycan be replaced by an unbiased estimatewhich is a function of y only.
In the following version of Stein's lemma[8], one can easily notice that, it is possible to replace any unknown x data expression by another one (having the same expectation),but containing the only known y data.

Lemma 1: Let F(y) be an N-dimensional vector functionsuch that $\varepsilon\{|\partial f_n(y)/\partial y_n|\} < \infty$ for n=1, …. N . The expressions $F(y)^T x$ and $F(y)^T y - \sigma^2 \mathrm{div}\{F(y)\}$ have the same expectation, (assuming white Gaussian noise).

$$\varepsilon\left\{\sum_{n=1}^{N} f_n(y)x_n\right\} = \varepsilon\left\{\sum_{n=1}^{N} f_n(y)y_n\right\} - \sigma^2 \varepsilon\left\{\sum_{n=1}^{N} \frac{\partial f_n(y)}{\partial y_n}\right\} \tag{5}$$

We can get the MSE estimate (Stein's unbiased risk "SURE"), by applying Lemma 1 to Eq. (2).
Theorem 1: For the same assumptions as Lemma 1, thefollowing random variable

$$\varsigma = \frac{1}{N}\|F(y) - y\|^2 + \frac{2\sigma^2}{N}\mathrm{div}\{F(y)\} - \sigma^2 \tag{6}$$

is an unbiased estimator of the MSE, i.e.,

$$\varepsilon\{\varsigma\} = \frac{1}{N}\varepsilon\left\{\|F(y) - x\|^2\right\}$$

Proof: We getthe following formula, by expanding the expectation of the MSE, and applying Lemma 1

$$\varepsilon\left\{\|F(y) - x\|^2\right\} = \varepsilon\left\{\|F(y)\|^2\right\} - 2\varepsilon\left\{F(y)^T x\right\} + \varepsilon\left\{\|x\|^2\right\}$$

$$= \varepsilon\left\{\|F(y)\|^2\right\} - 2\varepsilon\left\{F(y)^T y\right\} + 2\sigma^2\varepsilon\{\mathrm{div}\{F(y)\}\} + \varepsilon\left\{\|x\|^2\right\}$$

Since the noise b has zeromean, we can replace $\varepsilon\left\{\|x\|^2\right\}$ by $\varepsilon\left\{\|y\|^2\right\} - N\sigma^2$. A rearrangementof the y terms then provides the result of Theorem1.If we take into consideration the fact that in image processing applications the number of samples is usually large, hence, the estimate ς has a small variance (typically $\propto 1/N$).Thus the estimate is very close to its expectation, which isthe true MSE of the denoising process.

2.3."SURE"- based Image Denoising

Luisier*et al.* [16], assumed that theminimization of the MSE over a certain rangeof functions θ, could be obtained through minimizing ς over the samedenoising functions, up to a small random error.Now, if θ(the well-known *soft-thresholding*function) can be defined as follows

$$\theta(w) = sign(w)\max(0, |w| - T) \quad (7)$$

Then, by applyingTheorem 1, we could search for the optimal value T by minimizing the following expression over T

$$\tilde{\varsigma}(T) = \left\langle (2\sigma^2 + T^2 - w^2).\max(0, |w| - T) \right\rangle \quad (8)$$

Eq. (8) has its minimum for the same Tasfollows

$$SURE(T; w) = \sigma^2 - \frac{1}{N} * (2\sigma^2.\#\left\{ i : |w_i| \le T - \sum_{i=1}^{N} \min(|w_i|, T)^2 \right\} \quad (9)$$

as shown in [2], (the operator #{A} returns the cardinality of the set A).The estimated optimal threshold value is then: $\tilde{T}_{opt} = \arg\min_T (SURE(T; w)) = \arg\min_T (\tilde{\varsigma}(T))$. The *soft-thresholding* function exhibits two maindrawbacks: 1)It depends on a single parameter T,thus, its shape can'tbe flexible; 2) this dependency is nonlinear. Consequently,the *soft-thresholding*function is very sensitive to the valueof T,which requires a nonlinear algorithmto find the optimal threshold.

To mitigate this issue, Blu and Luisier [23] proposed a general form of denoisingfunction that depends *linearly* on a set of parameters (involves several degreesof freedom):thelinear transformation, a number $K$of linear parameters, andthe thresholding functions Θ_k, as follows:

$$\mathrm{F}(\mathrm{y}) = \sum_{k=1}^{K} a_k \mathrm{F}_k(\mathrm{y}) = \sum_{k=1}^{K} a_k \mathrm{R}\Theta_k(\mathrm{Dy}) \quad (10)$$

Here, the unknown weights a_k can be obtained by minimizing theestimatorçin Eq. (6).As the MSE estimate has a quadratic form (very similar to the real MSE), The linear minimization process is simple.The coefficients a_k are, thus, the solution ofthe following linear system

$$\sum_{l=1}^{K} \underbrace{\mathrm{F}_k(\mathrm{y})^{\mathrm{T}} \mathrm{F}_l(\mathrm{y}) a_l}_{[\mathrm{M}]_{k,l}} = \underbrace{\mathrm{F}_k(\mathrm{y})^{\mathrm{T}} \mathrm{y} - \sigma^2 \mathrm{div}\{\mathrm{F}_k(\mathrm{y})\}}_{[\mathrm{c}]_k}, \qquad \text{for } k = 1, 2, \ldots K$$

The above system of equations can be rewritten in matrix form as: $\mathrm{Ma} = \mathrm{c}$ (11)

3. THE PROPOSED EFFICIENT THRESHOLDING FUNCTION

We will consider only a *nonredundant* wavelet transform, i.e.; D & R are full rank matrices of size $N \times N$ .Further, as we already assumed orthonormal transform, i.e.; $\mathrm{R} = \mathrm{D}^{\mathrm{T}}$, then the Eq. (6) can be rewritten as

$$\varsigma = \frac{1}{N} \sum_{i=1}^{N} ((\theta_i(w_i) - w_i)^2 + 2\sigma^2 \theta_i'(w_i)) - \sigma^2 \quad (12)$$

Where w_i is the *i*th component of Dy ; i.e., it is a sum of the specific MSE estimates for each transformed coefficient w_i . Thus, the optimization process can be donecompletelyin the transform domain. Hence, we will express our thresholding function as a linear expans'ionof thresholds

$$\Theta(\mathrm{w}) = \sum_{k=1}^{K} a_k T_k(\mathrm{w}) \qquad (13)$$

If we introduce (13) into the estimate of the MSE given in (12)and perform differentiations overthe a_k, the exactminimization could be easily obtained by solving the linear system of equations for the unknown weights a_k as in (11) $\mathrm{a} = \mathrm{M}^{-1}\mathrm{c}$, where a and c are vectors of size $K \times 1$ and M is a matrix of size $K \times K$ as follows

$$\mathrm{c} = [c_k] = T_k(\mathrm{w})^{\mathrm{T}}\mathrm{w} - \sigma^2 T_k'(\mathrm{w}), \qquad k = 1, 2, \ldots\ldots K$$

$$\mathrm{M} = [M_{k,l}] = T_k(\mathrm{w})^{\mathrm{T}} T_l(\mathrm{w}), \qquad 1 \le k, l \le K \qquad (14)$$

As it is preferable to reduce the number of degreesof freedom (parameters *K*) in order for the estimate ς tokeep a small variance, we will consider only two parameters (*K* = 2). The clue now is to picka suitablebasis functions T_k which will define the shape of the proposedthresholdingfunction.Now, athresholding function is considered to be efficient if ithas the following minimal properties:

- *differentiability*: required to apply Theorem 1;
- *anti-symmetry*: (for the sign of the wavelet coefficients);
- *lineardealingwith large coefficients*;as for largecoefficient the noise corruptionis negligible, hencewe can let it without any modifications.

Thus, a good choice has been experimentally found to be of the form

$$\theta_i(w) = a_{i,1} t_1(w) + a_{i,2} t_2(w), \quad \text{in each band } i,$$

$$\text{Where } t_1(w) = w \text{ and } t_2(w) = w\left(1 - e^{-\frac{w^4}{4\sigma^2}}\right) \qquad (15)$$

4. RESULTS

The experiments are conducted on several standard gray scale test images like Lena&Boat of size 512 × 512, and House & Peppers of size 256 × 256 at different noise levelsσ =10, 20, 30, 50. The wavelet transform employs*generalized orthonormal Daubechies*.To test our proposed denoisingthresholdingfunction (15),a comparison with other various threshold denoising methods has been done. Ourresults, in terms of PSNR, have been compared with both the standard denoising method *"BaysShrink"*, and the theoretically best possible results that can be obtained by *soft-threshold* with an optimal threshold choice *"OracleShrink*". The results in Table 1 show thatthe proposedthresholding technique out performs the other denoisingmethods, and gives better PSNRs than the optimal*soft-threshold*. Further, the visual quality test in Figures 1 &2has provedthe significant capabilities of our approach forprocessing the artifacts andconservationof image edges.

Table 1: PSNR results for different test images and σ values, of 1) BayesShrink, 2) OracleShrink, 3) Our Proposed Threshold.

	BaysShrink	**OracleShrink**	**Proposed Method**
Lena 512 × 512			
σ = 10	33.4213	33.6221	33.8156
σ = 20	30.2556	30.3546	30.5899
σ = 30	28.5110	28.7895	28.9118
σ = 50	26.7987	26.9354	27.2147
Boat 512 × 512			
σ = 10	31.8544	32.1215	32.5044
σ = 20	28.3981	28.5947	28.7921
σ = 30	26.6121	26.9221	27.3201
σ = 50	24.5239	24.8455	25.0298
House 256 × 256			
σ = 10	31.8955	32.2418	32.3928
σ = 20	29.5628	29.7124	29.8854
σ = 30	27.7882	27.9132	28.2011
σ = 50	25.4877	25.6281	25.7209
Peppers 256 × 256			
σ = 10	29.8956	32.2113	32.5661
σ = 20	27.8678	28.1548	28.6124
σ = 30	25.7582	25.9776	26.4257
σ = 50	23.1938	23.4648	23.7254

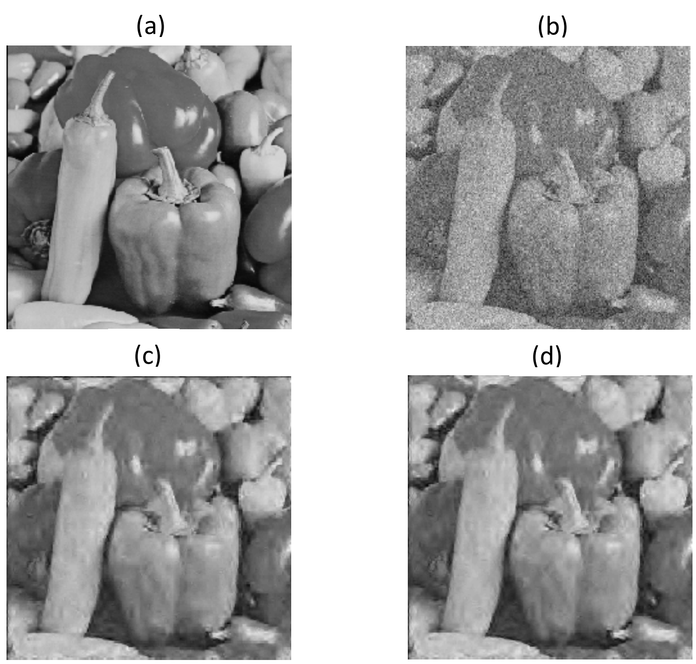

Figure 1. (a) Noise-free "Peppers" image 256×256. (b) A noisy version with (σ = 40): PSNR = 16.09 dB. (c) *BayesShrink*denoise: PSNR = 25.11 dB. (d) Proposed denoise: PSNR = 25.62 dB

Figure 2. (a) Noise-free "Lena" image 512×512. (b) A noisy version with (σ = 30): PSNR=18.59 dB. (c) *BayesShrink*denoise: PSNR=28.82 dB. (d) Proposed denoise: PSNR=29.56 dB

5. CONCLUSIONS

We have presented in this paper, an efficient and computationally attractive approach for image denoising. Using a modified Stein's unbiased risk estimate (SURE), that is only based on the noisy image, we obtained an accurate estimate of the MSEbetween noisy and clean image. Hence we not need any prior sophisticated statistical modelizationof the wavelet coefficients, and the optimal solution of MSE could be directly estimated by minimizingthe proposed thresholding function.

Experiments have been conducted to assess the performanceof our denoising techniquein comparison with the standard one*BayesShrink*, and the best soft-threshold method*OracleShrink*. The results show that the proposed techniquegave the best output PSNRs for the tested images.Furthermore, the visualassessmentshows that our denoisingmethod resulted images outperforms the other methods.

REFERENCES

[1] D. L. Donoho, “Unconditional bases are optimal bases for data compression and for statistical estimation,” J. Appl. Comput. Harmon. Anal., vol. 1, no. 1, pp. 100–115, Dec. 1993.

[2] D. L. Donoho and I. M. Johnstone, “Adapting to unknown smoothness via wavelet shrinkage,” J. Amer. Statist. Assoc., vol. 90, no. 432, pp. 1200–1224, Dec. 1995.

[3] ——, “Ideal spatial adaptation via wavelet shrinkage,” Biometrika, vol. 81, pp. 425–455, 1994.

[4] L. Breiman, “Better subset regression using the non-negative garrote,” Technometrics, vol. 37,no. 4, pp. 373–384, Nov. 1995.

[5] N. G. Kingsbury, “Image processing with complex wavelets,” Phil. Trans. Roy. Soc. A., Sep. 1999.

[6] H.-Y. Gao and A. G. Bruce, “Waveshrink with firm shrinkage,” Statist. Sin., vol. 7, pp. 855–874, 1997.

[7] ——, “Wavelet shrinkage denoising using the non-negative garrote,” J. Comput. Graph.Statist., vol. 7, no. 4, pp. 469–488, 1998.

[8] S. Kalavathy, and R. M. Suresh, “Analysis of Image Denoising using Wavelet Coefficient and Adaptive SubbandThresholding Technique,” IJCSI International Journal of Computer Science Issues, vol. 8, no. 1, pp. 166–172, 2011.

[9] S. D. Ruikar, and D. D. Doye, “Wavelet Based Image Denoising Technique,” (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 2, no.3, pp. 49–53, 2011.

[10] C. Stein, “Estimation of the mean of a multivariate normal distribution,” Ann. Statist., vol. 9, pp. 1135–1151, 1981.

[11] G. P. Nason, “Wavelet shrinkage using cross-validation,” J. Roy. Stat. Soc. B, vol. 58, no. 58, pp. 463–479, 1996.

[12] Y. Wang, “Function estimation via wavelet shrinkage for long-memory data,” Ann. Stat., vol. 9, pp. 466–484, 1996.

[13] N.Weyrich and G. T.Warhola, “Wavelet shrinkage and generalized cross validation for image denoising,” IEEE Trans. Image Process., vol. 7, no. 1, pp. 82–90, Jan. 1998.

[14] T. T. Cai and B. W. Silverman, “Incorporating information on neighboring coefficients into wavelet estimation,” Sankhya, ser. B, vol. 63, pp. 127–148, 2001.

[15] S. G. Chang, B. Yu, and M. Vetterli, “Adaptive wavelet thresholding for image denoising and compression,” IEEE Trans. Image Process., vol. 9, no. 9, pp. 1135–1151, Sep. 2000.

[16] F. Luisier, T. Blu, and M. Unser, “A new SURE approach to image denoising: Inter-scale orthonormal wavelet thresholding,” IEEE Trans. Image Process., vol. 16, no. 3, pp. 593–606, Mar. 2007.

[17] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, “Image denoising using scale mixtures of Gaussians in the wavelet domain,” IEEE Trans. Image Process., vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[18] J. A. Guerrero-Colón, L. Mancera, and J. Portilla, “Image restoration using space-variant Gaussian scale mixtures in overcomplete pyramids,” IEEE Trans. Image Process., vol. 17, no. 1, pp. 27–41, Jan.,2008.

[19] A. Piˇzurica and W. Philips, “Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising,” IEEE Trans. Image Process., vol. 15, no. 3, pp. 654–665, Mar. 2006.

[20] B. Goossens, A. Piˇzurica, and W. Philips, “Image Denoising Using Mixtures of ProjectedGaussian Scale Mixtures,” IEEE Trans. Image Process., vol. 18, no. 8, pp. 1689–1702, Aug., 2009.

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

[22] L. Sendur and I. W. Selesnick, “Bivariate shrinkage with local variance estimation,” IEEE Signal Process. Lett., vol. 9, no. 12, pp. 438–441, Dec. 2002.

[23] T. Blu and F. Luisier, “The SURE-LET approach to image denoising,” IEEE Trans. Image Process., vol. 16, no. 11, pp. 2778–2786, Nov. 2007.