

An Extension of the Interscale SURE-LET Approach for Image Denoising

Regular Paper

Lihong Cui¹, Zhan Wang¹, Yigang Cen², Xuguang Li¹ and Jianjun Sun^{3,*}

¹ Department of Mathematics and Computer Science, Beijing University of Chemical Technology, Beijing, China

² Institute of Information Science, Beijing Jiaotong University, Beijing, China

³ College of Chemical Engineering, Beijing University of Chemical Technology, Beijing, China

* Corresponding author E-mail: [sunjj@mail.buct.edu.cn](mailto:sunj@buct.edu.cn)

Received 05 Jul 2012; Accepted 21 Oct 2013

DOI: 10.5772/57294

© 2014 The Author(s). Licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract In this paper, an extension of the interscale SURE-LET approach exploiting the interscale and intrascale dependencies of wavelet coefficients is proposed to improve denoising performance. This method incorporates information on neighbouring coefficients into the linear expansion of thresholds (LETs) without additional parameters to capture the texture characteristics of this image. The resulting interscale-intrascale wavelet estimator consists of a linear expansion of multivariate thresholding functions, whose parameters are optimized thanks to a multivariate Stein's unbiased risk estimate (SURE). Some experimental results are given to demonstrate the strength of the proposed method.

Keywords Orthonormal Wavelet, Multivariate SURE-LET, Interscale and Intrascale Dependence, Multivariate Stein's Unbiased Risk Estimate, Image Denoising

1. Introduction

During the last decade, image denoising has undergone dramatic improvement; lots of new methods based on wavelet transforms have emerged for removing Gaussian

noise. A standard methodology proceeds by wavelet transforming the image, operating on the transform coefficients with nonlinear estimation functions, and then inverting the wavelet transform to obtain the denoised image. The choice of estimation function is an essential part of the denoising problem. Estimation functions generally take the form of "shrinkage" operators that are applied independently to each transform coefficient (e.g., [1-4]), or are applied to the neighbourhoods of coefficients at adjacent spatial positions and/or from other sub-bands (e.g., [5-16]). As demonstrated by several algorithms presented in the above literature, the performance of image-denoising algorithms can be improved significantly by taking into account the statistical dependence between interscale and intrascale coefficients. Figure 1 illustrates the statistical dependence between interscale and intrascale coefficients.

Image denoising can be accomplished by many different approaches. For example: using a prior model for the transform coefficients or using a parametric form for the estimation function. Generally, the generalized Gaussians [7, 11], scale mixtures models [13], Bessel K densities [17], and symmetric alpha stable densities [18], have been used

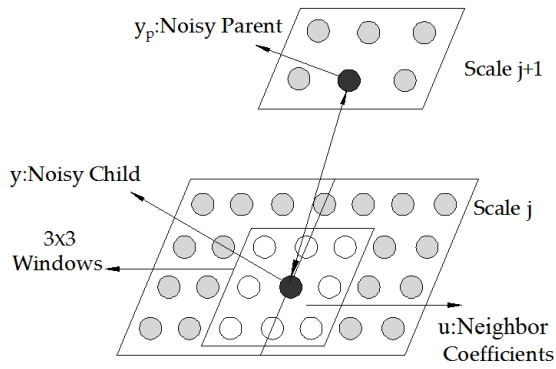


Figure 1. Illustration of the Parent-Child Relation and 3×3 Neighborhood Window

as prior models for the transform coefficients, which may be used to drive a Bayes-optimal estimator such as a MAP or MMSE estimator. Alternatively, one may directly assume a parametric form estimation function, such as [6, 8, 9, 19, 20], and select parameters by optimizing performance under certain conditions. In [8, 9], Luisier et al. introduced a new SURE [21] approach to image denoising - interscale orthonormal wavelet thresholding - which parameterized the denoising process as a sum of elementary nonlinear processes - LETs - with unknown weights instead of postulating a statistical prior model for the wavelet coefficients, and then adaptively optimized the parametric estimator by minimizing SURE, which provides an approximation of the mean squared error (MSE) as a function of the observed noisy data. Risk minimization and unknown weights' estimations ultimately come down to solving a linear system of equations. However, they only took into account interscale dependency using an interscale prediction model group delay compensation (GDC), and they disposed of intrascale dependency. Their experimental result demonstrated that, for most of the images, the interscale SURE-based approach is competitive in relation to the best techniques available that consider orthonormal wavelet transforms. However, it should be noted that this approach did not obtain good performance for images with substantial textures, such as the Barbara image. The main reason for this is that some local information (especially the texture of Barbara's trousers) is completely lost at coarser scales. Interscale correlations may be too weak for this image, which indicates that an efficient denoising process may require intrascale information as well. For other denoising methods, the reader is referred to [22, 23] and the references cited there.

In this paper, we propose a multivariate SURE-LET approach to orthonormal wavelet image denoising as an extension of Luisier's bivariate approach. This method incorporates information on neighbouring coefficients into the LET without additional parameters to capture the

texture characteristics of this image. The resulting interscale-intrascale wavelet estimator consists of a linear expansion of multivariate thresholding functions, whose parameters are optimized thanks to multivariate SURE. This paper is organized as follows: In Section 2, we explain the multivariate SURE theory for a neighbourhood vector and generalize the corresponding linear parameterization strategy. In Section 3, the competitive results with the best up-to-date algorithms will be shown. The conclusion can be found in Section 4.

2. Multivariate SURE-LET

Let $g_k, k \in \mathbb{K}$ be equally-spaced samples of a real-valued image, where \mathbb{K} is a set of spatial indexes ($\mathbb{K} \subset \mathbb{Z}^2$). Consider the standard nonparametric regression setting:

$$f_k = g_k + n_k \quad (2.1)$$

where n_k are i.i.d. normal random variables with a mean zero and variance σ^2 independent of g_k . Let \mathbf{f} , \mathbf{g} and \mathbf{n} denote the matrix representation of the corresponding samples. Let $\mathbf{Y} = W\mathbf{f}$, $\mathbf{X} = W\mathbf{g}$ and $\mathbf{B} = W\mathbf{n}$, where W is the two-dimensional dyadic orthonormal wavelet transform (DWT) operator [24]. Let y_k^{oj} be the detail coefficient of the noisy image \mathbf{f} at location \mathbf{k} , scale j and orientation o , and similarly for x_k^{oj} and b_k^{oj} . It follows from (2.1) that:

$$y_k^{oj} = x_k^{oj} + b_k^{oj} \quad (2.2)$$

For facilitation, we will drop scale j and the orientation o indexes, change the spatial indexes \mathbf{k} to 1-D index n , and consider the standard simplified denoising problem in each orientation sub-band at a given scale: given noisy data $y_n = x_n + b_n$, for $n = 1 \dots N$, where b_n is an independent white Gaussian noise of variance σ^2 , and due to the orthogonality of the basis we observe $\mathbf{y} = \{y_n\}, n = 1, 2 \dots N$ and seek to estimate the desired $\mathbf{x} = \{x_n\}, n = 1, 2 \dots N$ as accurately as possible according to various criteria. This is a classical problem in estimation theory. Our aim is to find an estimation function θ of the observed noisy coefficient neighbourhood alone, such that:

$$\hat{\mathbf{x}}_n = \theta(\mathbf{u}_n), n = 1, 2, \dots, N \quad (2.3)$$

which will minimize the MSE defined by:

$$\text{MSE} = \frac{1}{N} \sum_{n=1}^N |\hat{x}_n - x_n|^2 = \frac{1}{N} \|\hat{\mathbf{x}} - \mathbf{x}\|^2 \quad (2.4)$$

where $\mathbf{u} = \{\mathbf{u}_n\}_{n=1,2,\dots,N}$ is an observation neighbourhood sequence, and \mathbf{u}_n of a d -dimensional real-valued ($d \in \mathbb{N}, d > 1$) vector is spatial neighbourhood vector of y_n . Specifically, \mathbf{u}_n is defined as all those coefficients within a square-shaped window that is centred at the n th coefficient, as illustrated in Figure 1. Without loss of generality, we can assume that $\mathbf{u}_n = [y_n, \tilde{\mathbf{y}}_n^\top]^\top$, where $\tilde{\mathbf{y}}_n^\top$ is the last $d-1$ components of \mathbf{u}_n . So far, we have introduced an explicit dependence between x_n and $\tilde{\mathbf{y}}_n$. As we know from [8, 9], the MSE in the space domain is a weighted sum of the MSE of each individual sub-band, which allows us to apply the denoising function independently in every high-pass sub-band.

2.1 Unbiased Estimate of the MSE

It seems impossible to compute the $\|\hat{\mathbf{x}} - \mathbf{x}\|^2 / N$ since we do not have access to the signal \mathbf{x} . However, in the case of Gaussian noise, it is possible to apply an extension of Stein's principle [21] for deriving an explicit expression. The following lemma 1 shows how it is possible to replace an expression that contains the unknown coefficient \mathbf{x} by another one with the same expectation, yet containing the known noise coefficient \mathbf{y} only. Lemma 1 and Theorem 1 essentially recap the derivation of SURE, which can be found in [8].

Lemma 1. Let $\theta : \mathbb{R}^d \rightarrow \mathbb{R}$ be a continuous and almost everywhere differentiable function, such that:

$$\forall \xi \in \mathbb{R}^d, \lim_{\|\mathbf{t}\| \rightarrow +\infty} \theta(\mathbf{t}) \exp\left(-\frac{(\mathbf{t}-\xi)^\top \Gamma^{-1}(\mathbf{t}-\xi)}{2}\right) = 0 \quad (2.5)$$

$$\mathbf{E}[\|\theta(\mathbf{u}_n)\|^2] < +\infty \text{ and } \mathbf{E}\left[\left\|\frac{\partial \theta(\mathbf{u}_n)}{\partial y_n}\right\|\right] < +\infty \quad (2.6)$$

Then, under the additive white Gaussian noise assumption:

$$\mathbf{E}[\theta(\mathbf{u}_n)x_n] = \mathbf{E}[\theta(\mathbf{u}_n)y_n] - \sigma^2 \mathbf{E}\left[\frac{\partial \theta(\mathbf{u}_n)}{\partial y_n}\right] \quad (2.7)$$

where $\Gamma = \sigma^2 \mathbf{I}_d$, \mathbf{I}_d is a unit matrix and $\mathbf{E}[\cdot]$ stands for the mathematical expectation operator.

Proof. Let $\mathbf{T} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be a continuous and almost everywhere differentiable function, such that:

$$\forall \xi \in \mathbb{R}^d, \lim_{\|\mathbf{t}\| \rightarrow +\infty} \mathbf{T}(\mathbf{t}) \exp\left(-\frac{(\mathbf{t}-\xi)^\top \Gamma^{-1}(\mathbf{t}-\xi)}{2}\right) = 0 \quad (2.8)$$

$$\mathbf{E}[\|\mathbf{T}(\mathbf{u}_n)\|^2] < +\infty \text{ and } \mathbf{E}\left[\left\|\frac{\partial \mathbf{T}(\mathbf{u}_n)}{\partial \mathbf{u}_n}\right\|_F\right] < +\infty \quad (2.9)$$

where $\|\cdot\|_F$ is the Frobenius norm. In this multivariate context, Stein's principle [21] can be expressed as:

$$\mathbf{E}[\mathbf{T}(\mathbf{u}_n)\mathbf{w}_n^\top] = \mathbf{E}[\mathbf{T}(\mathbf{u}_n)\mathbf{u}_n^\top] - \mathbf{E}\left[\frac{\partial \mathbf{T}(\mathbf{u}_n)}{\partial \mathbf{u}_n}\right] \Gamma \quad (2.10)$$

where \mathbf{w}_n is, according the spatial neighbourhood vector of x_n , similar to \mathbf{u}_n formally. Equation (7) follows by choosing $\mathbf{T} : \mathbf{t} \mapsto [\theta(\mathbf{t}), 0, \dots, 0]^\top$ and focusing on the top-left element of matrix $\mathbf{E}[\mathbf{T}(\mathbf{u}_n)\mathbf{w}_n^\top]$.

Theorem 1. Under the same hypotheses as Lemma 1, the random variable:

$$\varepsilon = \frac{1}{N} \|\theta(\mathbf{u}) - \mathbf{y}\|^2 + \frac{2\sigma^2}{N} \mathbf{div}\{\theta(\mathbf{u})\} - \sigma^2 \quad (2.11)$$

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

$$\mathbf{E}[\varepsilon] = \frac{1}{N} \mathbf{E}[\|\theta(\mathbf{u}) - \mathbf{x}\|^2] \quad (2.12)$$

where $\mathbf{div}\{\theta(\mathbf{u})\} = \sum_{n=1}^N \frac{\partial \theta(\mathbf{u}_n)}{\partial y_n}$.

Proof. By expanding the expectation of the MSE, we have:

$$\begin{aligned} \mathbf{E}[\|\theta(\mathbf{u}) - \mathbf{x}\|^2] &= \mathbf{E}[\|\theta(\mathbf{u})\|^2] - 2\mathbf{E}[\theta(\mathbf{u})^\top \mathbf{x}] + \mathbf{E}[\|\mathbf{x}\|^2] \\ &= \mathbf{E}[\|\theta(\mathbf{u})\|^2] - 2\mathbf{E}[\theta(\mathbf{u})^\top \mathbf{y}] + 2\sigma^2 \mathbf{E}[\mathbf{div}\{\theta(\mathbf{u})\}] + \mathbf{E}[\|\mathbf{x}\|^2] \end{aligned}$$

Since the noise \mathbf{b} has zero mean, we can replace $\mathbf{E}[\|\mathbf{x}\|^2]$ by $\mathbf{E}[\|\mathbf{y}\|^2] - N\sigma^2$. A rearrangement of the \mathbf{y} terms then provides the result of Theorem 1.

The expression in equation (2.12) may be evaluated on a single observation \mathbf{y} (\mathbf{u} can be assembled by overlapping \mathbf{y}) to produce an unbiased estimate of the MSE. Although the derivation of this expression is relatively simple, it leads us to the somewhat counterintuitive conclusion that the estimator may be optimized without explicit knowledge of the clean coefficients \mathbf{x} . It must be emphasized that this estimate is close to its expectation, which is the MSE of the denoising procedure, because the standard deviation of ε is small by the law of large numbers.

2.2 The Multivariate SURE-LET Approach

Similar to the LET of [5, 8, 9], we build a linearly parameterized multivariate estimation function

incorporating information on neighbouring coefficients of the form:

$$\theta(\mathbf{u}_n) = \sum_{k=1}^K a_k \varphi_k(\mathbf{u}_n) = \underbrace{[a_1, a_2, \dots, a_K]}_{\mathbf{a}^\top} \times \underbrace{\begin{bmatrix} \varphi_1(\mathbf{u}_n) \\ \varphi_2(\mathbf{u}_n) \\ \vdots \\ \varphi_K(\mathbf{u}_n) \end{bmatrix}}_{\Phi(\mathbf{u}_n)} \quad (2.13)$$

Here, $\Phi(\mathbf{u}_n)$ is a $K \times 1$ vector function, the a_k is an unknown weight specified by minimizing the SURE given by (2.11), and \mathbf{a} is a $K \times 1$ vector. It should be noted that the new multivariate estimation function does not introduce more parameters, compared with LET in [8-9], which means that this improvement still maintains efficiency of calculation. In this formalism, $\frac{\partial \theta(\mathbf{u}_n)}{\partial y_n}$ can be expressed as:

$$\frac{\partial \theta(\mathbf{u}_n)}{\partial y_n} = \mathbf{a}^\top \nabla_{y_n} \Phi(\mathbf{u}_n)$$

where we have denoted by $\nabla_{y_n} \Phi(\mathbf{u}_n)$ the vector containing the partial derivatives of the components y_n , i.e., $\nabla_{y_n} \Phi(\mathbf{u}_n) = \left[\frac{\partial \varphi_1(\mathbf{u}_n)}{\partial y_n}, \frac{\partial \varphi_2(\mathbf{u}_n)}{\partial y_n}, \dots, \frac{\partial \varphi_K(\mathbf{u}_n)}{\partial y_n} \right]^\top$.

The MSE estimate \mathcal{E} is quadratic in \mathbf{a} , as follows:

$$\begin{aligned} \mathcal{E} &= \frac{1}{N} \sum_{n=1}^N \left| \mathbf{a}^\top \Phi(\mathbf{u}_n) - y_n \right|^2 + \frac{2\sigma^2}{N} \sum_{n=1}^N \mathbf{a}^\top \nabla_{y_n} \Phi(\mathbf{u}_n) - \sigma^2 \\ &= \frac{1}{N} \sum_{n=1}^N \left(\mathbf{a}^\top \Phi(\mathbf{u}_n) \Phi(\mathbf{u}_n)^\top \mathbf{a} - 2y_n \mathbf{a}^\top \Phi(\mathbf{u}_n) + y_n^2 \right) \\ &\quad + \frac{2\sigma^2}{N} \sum_{n=1}^N \mathbf{a}^\top \nabla_{y_n} \Phi(\mathbf{u}_n) - \sigma^2 \\ &= \mathbf{a}^\top \mathbf{M} \mathbf{a} - 2\mathbf{a}^\top \mathbf{c} + \frac{1}{N} \|\mathbf{y}\|^2 - \sigma^2 \end{aligned} \quad (2.14)$$

where we have defined:

$$\mathbf{M} = \frac{1}{N} \sum_{n=1}^N \Phi(\mathbf{u}_n) \Phi(\mathbf{u}_n)^\top \quad (2.15)$$

$$\mathbf{c} = \frac{1}{N} \sum_{n=1}^N \left(y_n \Phi(\mathbf{u}_n) - \sigma^2 \nabla_{y_n} \Phi(\mathbf{u}_n) \right) \quad (2.16)$$

Finally, the minimization of (2.14) with respect to \mathbf{a} boils down to the following linear system of equations:

$$\mathbf{a} = \mathbf{M}^{-1} \mathbf{c}. \quad (2.17)$$

Note that since the minimum of \mathcal{E} always exists, it is ensured that there will always be a solution to this system. When $\text{rank}(\mathbf{M}) < K$, we can simply take its pseudo-inverse to choose any one among the admissible

solutions. Of course, it is desirable to keep the number of degrees of freedom K as low as possible in order for the estimate \mathcal{E} to maintain a small variance.

2.3 The New Inter- and Intracscale Thresholding Function

To compensate for feature misalignment between child coefficients and parent coefficients, we will also use the GDC scheme [8, 9], which builds an interscale predictor out of the low-pass sub-band at the same scale. Let y_{p_n} denote the value of the GDC output, which can be interpreted as a discriminator between high SNR wavelet coefficients and low SNR wavelet coefficients, corresponding to the noisy coefficient y_n . A Gaussian smoother function proposed in [8, 9] is chosen, namely the decision function:

$$f(y_{p_n}) = e^{-\frac{y_{p_n}^2}{2T^2}}. \quad (2.18)$$

where $T = \sqrt{6}\sigma$ is the universal threshold.

In order to incorporate information on neighbouring coefficients into the LET without additional parameters, we propose the following pointwise radial exponential function:

$$\varphi_k(\mathbf{u}_n) = y_n e^{-\frac{(k-1)\|\mathbf{u}_n\|^2}{2dT^2}}, k=1, \dots, K. \quad (2.19)$$

Here, d is the dimension of vector \mathbf{u}_n and the radial profile of this pointwise function is exponential in $\|\mathbf{u}_n\|$.

By joining the interscale predictor and multivariate SURE-LET approach, we lead to the following general inter- and intracscale thresholding function:

$$\tilde{\theta}(\mathbf{u}_n, y_{p_n}) = f(y_{p_n}) \sum_{k=1}^K a_k \varphi_k(\mathbf{u}_n) + (1 - f(y_{p_n})) \sum_{k=1}^K a_{k+K} \varphi_k(\mathbf{u}_n) =$$

$$\underbrace{[a_1, \dots, a_K, a_{K+1}, \dots, a_{2K}]}_{\mathbf{a}^\top} \times \underbrace{\begin{bmatrix} f(y_{p_n}) \varphi_1(\mathbf{u}_n) \\ \vdots \\ f(y_{p_n}) \varphi_K(\mathbf{u}_n) \\ (1 - f(y_{p_n})) \varphi_1(\mathbf{u}_n) \\ \vdots \\ (1 - f(y_{p_n})) \varphi_K(\mathbf{u}_n) \end{bmatrix}}_{\Phi(\mathbf{u}_n, y_{p_n})} \quad (2.20)$$

Here, $\Phi(\mathbf{u}_n, y_{p_n})$ is a $2K \times 1$ vector function and \mathbf{a} is $2K \times 1$ vector. It is essential to notice that, because of the statistical independence between sub-bands of different iteration depths, \mathbf{u}_n and y_{p_n} will also be statistically independent. Therefore, the partial derivatives of $\tilde{\theta}(\mathbf{u}_n, y_{p_n})$ with respect to the component y_n are

uncorrelated with y_{p_n} - Theorem 1 remains true - and then the linear parameter vector \mathbf{a} is solved by minimizing the MSE estimate \mathcal{E} defined in Theorem 1, i.e., \mathbf{a} can be obtained by (2.17).

We can summarize our denoising algorithm as follows:

- 1) Perform a J level DWT to the noisy image \mathbf{f} , i.e. $\mathbf{Y} = \mathbf{W}\mathbf{f}$.
- 2) For each sub-band (except the low-pass residual), compute the interscale predictor $\{y_{p_n}\}_{1 \leq n \leq N}$ using the GDC approach [8, 9].
- 3) Compute the $\{\|\mathbf{u}_n\|^2\}_{1 \leq n \leq N}$ under a given neighbourhood window size, and obtain $\Phi(\mathbf{u}_n, y_{p_n})$ according to (2.19) and (2.20).
- 4) Determine \mathbf{M} and \mathbf{c} using (2.15) and (2.16), and then solve the linear system (2.17) to obtain \mathbf{a} .
- 5) Sub-band adaptive image denoising using (2.20).
- 6) Reconstruct the denoised image from the processed sub-bands and the low-pass residual.

3. Numerical Experiments

In what follows, we carried out all the experiments on 8-bit greyscale test images of sizes 512×512 and 256×256 , as presented in Figure 2. The test images were obtained from the same sources, as mentioned in [8, 9, 11]. We applied our multivariate SURE-LET (abbreviated as MuSURE-LET) algorithm according to the expression (2.20) with $K = 3$, after four or five decomposition levels (depending on the size of the image) of an orthonormal wavelet transform (OWT) using the standard Daubechies symlets with eight vanishing moments (sym8 in MatLab). A good estimator for σ is the median of absolute deviation (MAD) using the highest level wavelet coefficients [2], as follows:

$$\hat{\sigma} = \frac{\text{median}(|w_s|)}{0.6745} (w_s \in \text{subbandHH}). \quad (3.1)$$

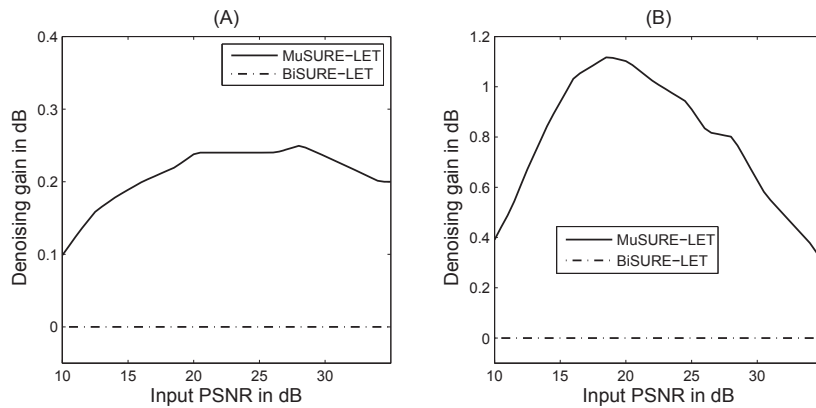


Figure 3. PSNR improvements brought by our multivariate SURE-LET strategy compared to bivariate SURE-LET: (A) Lena image; (B) Barbara image



Figure 2. The test images used in the experiments, referred to as 'Lena', 'Barbara', 'Boat', 'Mandrill', 'Fingerprint' and 'Bridge' (numbered from left to right and top to bottom)

Here, sub-band HH is the finest scale wavelet sub-band in the diagonal direction. The denoising performances are measured in terms of a peak signal-to-noise ratio (PSNR), defined as:

$$PSNR = 10 \log_{10} \frac{255^2}{\frac{1}{N} \sum_{n=1}^N w_n^2} \quad (3.2)$$

where N is the total number of pixels and $w_n = \hat{x}_n - x_n$.

The window size is dependent on the abundance of the textures of the example images. In our experiments, the window size 7×7 yields the best results for those images with substantial textures, while the window size 3×3 yields the best results for those images with less detailed textures. Table 1 shows the error variances of the denoised images, expressed as the PSNR defined in (3.2), at eight different power levels $\sigma \in [10, 15, 20, 25, 30, 50, 75, 100]$. Note that, for all the images, there is very little improvement at the lowest noise level. This makes sense, since the "clean" images in fact include quantization errors and have an implicit PSNR of 58.9 dB.

3.1 Comparisons with the Interscale SURE-LET Approach

In order to understand the relative contribution of our method, we first want to evaluate the improvements brought by the integration of neighbouring coefficients' dependencies. Compared with the bivariate SURE-LET approach defined in [8,9], we can evaluate the improvements brought by our multivariate SURE-LET function (2.20) (see Figure 3). As can be observed, the integration of neighbouring coefficients' dependencies improves the denoising performance considerably. For those images that have substantial textures, such as the Barbara image, the denoising gains are up to 0.8–1.1 dB when the range of the PSNR values of input noisy images are in [15, 30], and for ones that have less detailed textures, such as the Lena image, the denoising gains are up to 0.2–0.3 dB when the range of the PSNR values of input noisy images are in [15, 30]. Figure 4 provides a visual comparison of an example image (Barbara) between the above-mentioned two methods. Our method is seen to provide fewer artefacts - for example, in parts of the forehead and hair of the woman - which means that our method can better suppress noise in the uniform areas.

In order to understand the relative contribution of our method, we first want to evaluate the improvements brought by the integration of neighbouring coefficients' dependencies. In Figure 3, we compare our multivariate SURE-LET function (2.20) with the bivariate SURE-LET (abbreviated as BiSURE-LET) defined in [8]. As can be observed, the integration of neighbouring coefficients' dependencies improves the denoising performance considerably. For those images that have substantial textures, such as the Barbara image, the denoising gains are up to 0.8–1.1 dB when the range of the PSNR values of input noisy images are in [15, 30], and for ones that have less detailed textures, such as the Lena image, the denoising gains are up to 0.2–0.3 dB when the range of the PSNR values of input noisy images are in [15, 30]. Figure 4 provides a visual comparison of an example image (Barbara) with the above-mentioned two methods. Our method is seen to provide fewer artefacts - for example, in parts of the forehead and hair of the woman - which means that our method can better suppress noise in the uniform areas.

3.2 Comparisons with State of the Art Denoising Schemes

Compared with state-of-the-art denoising algorithms, for which the code is freely distributed by the authors: Bishrink (7×7) [15,16], ProbShrink (3×3) [12], BLS-GSM (3×3) [13], Block-matching and 3D filtering (BM3D)

[25], Non-local Means (NL-Means) [26] and Field of Experts (FoE) [27, 28]. Since the versions of the noise standard are not on a unit level, we have averaged the output PSNRs over eight noise realizations so as to apply the same noise realizations to different algorithms.

Table 1 reports the PSNR results we obtained with the various denoising methods, the best results being shown in boldface. As we can see, our algorithm (Multivariate SURE-LET) matches or overmatches the other methods' results for most of the images. Noisy ($\sigma=60$) and denoised fingerprint and mandrill images are shown in Figures 5 and 6, respectively.

When looking more closely at the results, we observe the following.

Our method gives better results than Sendur's Bishrink 7×7 , which integrates both the inter- and the intrascale dependencies (an average gain of +0.8 dB).

Our method gives better results than Pižurica's ProbShrink 3×3 , which integrates the intrascale dependencies (an average gain of +0.6 dB).

Our method outperforms the Portilla's BLS-GSM 3×3 or NL-Means by more than 0.2–0.3 dB on average.

Our method improves the PSNR by about 0.6 dB on average in comparison with FoE.

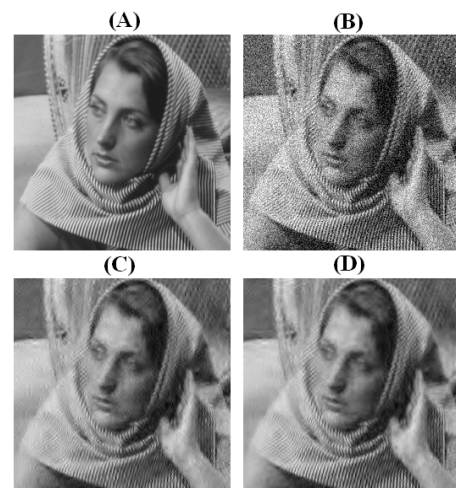


Figure 4. Comparison of the denoising results on the Barbara image (cropped to 256×256 to show the artefacts): (A) Part of the noise-free Barbara image; (B) Part of the noisy Barbara image: $\sigma = 30$, $PSNR = 18.59$ dB; (C) Result of the BiSURE-LET: $PSNR = 25.82$ dB; (D) Result of the MuSURE-LET: $PSNR = 26.94$ dB.

σ	BiShrink	ProbShrink	BLS-GSM	BiSURE-LET	MuSURE-LET	BM3D	NL-Means	FoE
Lena								
10	34.47	34.30	34.74	34.56	34.81	35.91	34.31	35.04
15	32.63	32.41	32.90	32.68	32.92	34.26	32.07	33.26
20	31.30	31.05	31.59	31.37	31.61	33.04	31.55	31.84
25	30.30	30.02	30.57	30.36	30.60	32.08	30.46	30.82
30	29.49	29.25	29.74	29.56	29.78	31.28	29.49	29.81
50	27.16	27.22	27.44	27.37	27.55	28.85	27.39	26.49
75	25.47	25.56	25.72	25.76	25.87	26.99	25.31	24.13
100	24.31	24.30	24.54	24.66	24.72	25.55	23.75	21.87
Barbara								
10	32.52	32.48	32.89	32.16	32.96	34.93	33.16	32.86
15	30.14	30.04	30.54	29.65	30.59	33.05	30.79	30.22
20	28.51	28.40	28.93	27.96	28.98	31.71	30.21	28.30
25	27.29	27.20	27.72	26.74	27.84	30.64	28.92	27.05
30	26.33	26.27	26.76	25.82	26.94	29.73	28.03	26.01
50	23.91	23.86	24.25	23.72	24.58	27.15	25.72	23.15
75	22.49	22.50	22.72	22.54	22.98	25.13	23.35	21.36
100	21.67	21.68	21.53	21.81	22.01	23.57	21.86	19.77
Boat								
10	32.46	32.53	32.89	32.91	33.09	33.90	32.90	33.05
15	30.47	30.50	30.89	30.86	31.06	32.11	30.69	31.23
20	29.08	29.11	29.49	29.47	29.67	30.85	29.69	29.82
25	28.03	28.05	28.43	28.44	28.63	29.86	28.63	28.72
30	27.20	27.22	27.58	27.63	27.81	29.06	27.65	27.86
50	25.05	25.12	25.34	25.52	25.66	26.64	25.21	24.53
75	23.67	23.82	23.97	24.04	24.14	24.84	23.35	22.48
100	22.66	22.69	22.64	23.09	23.16	23.64	22.10	20.80
Mandrill								
10	30.05	29.78	30.17	30.20	30.39	30.70	30.32	30.19
15	27.48	27.27	27.66	27.65	27.92	28.31	27.93	27.69
20	25.84	25.65	26.02	26.02	26.33	26.75	26.41	25.93
25	24.65	24.48	24.85	24.88	25.20	25.62	25.19	24.23
30	23.77	23.61	23.98	24.03	24.34	24.75	24.31	23.92
50	21.65	21.93	21.91	22.07	22.30	22.42	21.90	21.75
75	20.54	20.81	20.70	20.96	21.07	21.14	20.57	20.38
100	19.95	20.12	20.04	20.36	20.40	20.39	19.91	19.51
Fingerpt								
10	30.93	31.62	31.65	31.70	31.76	32.45	31.02	32.03
15	28.67	29.29	29.36	29.47	29.54	30.28	28.69	29.42
20	27.18	27.80	27.82	27.95	28.03	28.80	27.20	27.34
25	26.04	26.55	26.65	26.79	26.89	27.70	26.15	25.05
30	25.10	25.68	25.70	25.87	25.98	26.82	25.24	23.60
50	22.57	23.12	23.21	23.39	23.56	24.32	22.97	22.68
75	20.70	21.33	21.43	21.49	21.76	22.63	21.04	20.29
100	19.47	20.14	20.23	20.18	20.55	21.30	19.55	18.75
Bridge								
10	29.08	29.61	30.02	30.19	30.32	30.71	30.43	30.92
15	26.96	27.20	27.52	27.80	27.90	28.28	27.96	28.48
20	25.62	25.74	26.02	26.31	26.40	26.76	26.48	26.78
25	24.69	24.73	25.03	25.27	25.35	25.75	25.37	25.70
30	23.99	23.97	24.29	24.48	24.56	25.02	24.55	24.81
50	22.28	22.11	22.48	22.52	22.59	23.10	22.16	22.50
75	21.05	21.13	21.12	21.15	21.21	21.86	20.60	20.83
100	20.22	19.97	20.19	20.28	20.38	20.91	19.63	19.50

Table 1. Comparison of Some of the Most Efficient Denoising Methods (sym8)

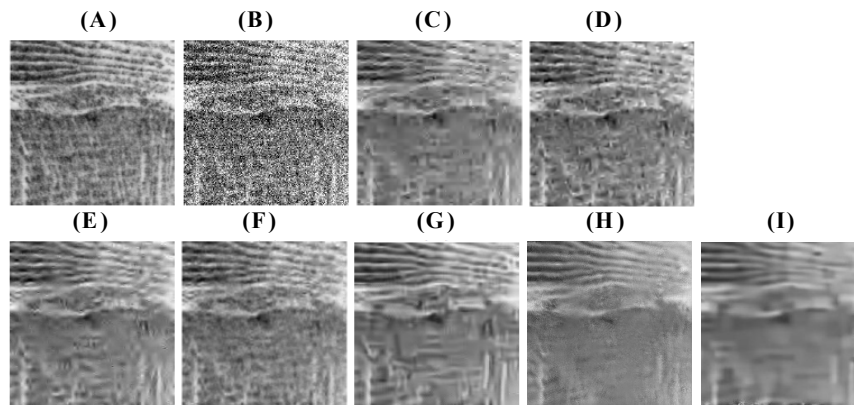


Figure 5. Comparison of the denoising results on the Fingerprint image (cropped to 200×200 to show the artefacts): (A) Part of the noise-free Fingerprint image; (B) Part of the noisy Fingerprint image: $\sigma = 60$, $PSNR = 12.57$ dB; (C) Result of the BiShrink: $PSNR = 21.71$ dB; (D) Result of the ProbShrink (3×3): $PSNR = 22.60$ dB; (E) Result of the BLS-GSM (3×3): $PSNR = 22.39$ dB; (F) Result of the MuSURE-LET: $PSNR = 22.73$ dB; (G) Result of the BM3D: $PSNR = 23.55$; (H) Result of the NL-Means: $PSNR = 22.00$; (I) Result of the FoE: $PSNR = 21.60$.

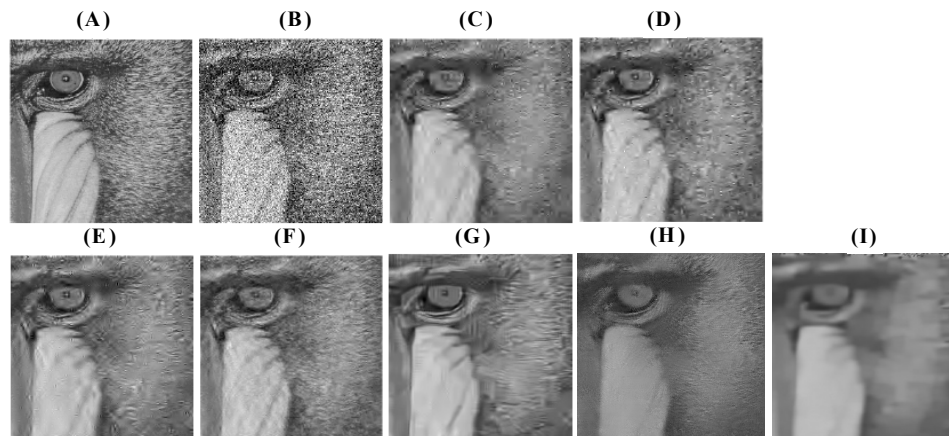


Figure 6. Comparison of the denoising results on the Mandrill image (cropped to 256×256 to show the artefacts.): (A) Part of the noise-free Mandrill image; (B) Part of the noisy Mandrill image: $\sigma = 60$, $PSNR = 12.57$ dB; (C) Result of the BiShrink: $PSNR = 21.12$ dB; (D) Result of the ProbShrink (3×3): $PSNR = 21.48$ dB; (E) Result of the BLS-GSM (3×3): $PSNR = 21.32$ dB; (F) Result of the MuSURE-LET: $PSNR = 21.71$ dB; (G) Result of the BM3D: $PSNR = 21.79$; (H) Result of the NL-Means: $PSNR = 21.22$; (I) Result of the FoE: $PSNR = 21.10$.

Although the PSNR of our method is less than approximately 0.9 dB on average when compared to BM3D, the denoised images of our method are very similar to the original and qualitatively superior to BM3D (comparing (F) with (G) in Figure 5 and Figure 6).

In particular, our algorithm obtains better results for those images with substantial textures, for which the bivariate SURE-LET is not very effective [8, 9], such as the Barbara and Mandrill images.

From a visual point of view, our algorithm can be seen to provide fewer artefacts as well as a better preservation of edges and other details. These observations are clearly illustrated in Figure 5 and Figure 6.

4. Conclusions

The paper integrates the intrascale dependencies within the SURE-LET approach successfully such that, as an extension of the interscale and bivariate SURE-LET approaches in [8], the results are interesting.

The comparison of the denoising results obtained with our algorithm and with the best state-of-the-art non-redundant techniques (that integrate both inter- and intrascale dependencies) demonstrate the efficiency of our multivariate SURE-LET approach, which gave superior output PSNRs for most of the images. The visual quality of our denoised images is characterized more by fewer artefacts and a better preservation of edges and other details than the other methods.

5. Acknowledgements

The authors would like to thank the support of the National Natural Science Foundation of China (grant no. 61272028), the National Undergraduate Training Programmes for Innovation and Entrepreneurship (no. 201210010077) and the Science and Technology Innovation Foundation for College Students (no. pt2012064).

The authors also appreciate the help of Dr. F. Luisier, Dr. I. W. Selesnick, Dr. A. Pizurica, Dr. J. Portilla, Dr. K. Dabov, Dr. A. Buades and Dr. S. Roth for making available the codes of the methods OWT SURE-LET, BiShrink, ProbShrink, BLS-GSM, BM3D, NL-Means and FoE, respectively, on their websites.

6. References

- [1] D. L. Donoho, "Denoising by soft-thresholding," *IEEE Trans. Information Theory*, Vol. 41, No. 3, pp. 613–627, 1995.
- [2] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, Vol. 81, No. 3, pp. 425–455, 1994.
- [3] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, Vol. 90, No. 432, pp. 1200–1224, 1995.
- [4] M. Figueiredo and R. Nowak, "Wavelet-based image estimation: An empirical Bayes approach using Jeffrey's noninformative prior," *IEEE Trans. Image Process*, Vol. 10, No. 9, pp. 1322–1331, 2001.
- [5] T. Blu and F. Luisier, "The sure-let approach to image denoising," *IEEE Trans. Image Process*, Vol. 16, No. 11, pp. 2778–2786, 2007.
- [6] T. T. Cai and B. W. Silverman, "Incorporating information on neighboring coefficients into wavelet estimation," *The Indian Journal of Statistics, Series B*, Vol. 63, No. 2, pp. 127–148, 2001.
- [7] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Processing*, Vol. 9, No. 9, pp. 1532–1546, 2000.
- [8] F. Luisier, T. Blu, and M. Unser, "A new sure approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Trans. Image Process*, Vol. 16, No. 3, pp. 593–606, 2007.
- [9] F. Luisier, T. Blu, and M. Unser, "Sure-let multichannel image denoising: interscale orthonormal wavelet thresholding," *IEEE Trans. Image Process*, Vol. 17, No. 4, pp. 482–492, 2008.
- [10] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low complexity image denoising based on statistical modeling of wavelet coefficients," *Signal Processing Letters*, Vol. 6, No. 12, pp. 300–303, 1999.
- [11] A. Pizurica and W. Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single and multiband image denoising," *IEEE Trans. Image Processing*, Vol. 15, No. 3, pp. 654–665, 2006.
- [12] 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 Process*, Vol. 11, No. 5, pp. 545–557, 2002.
- [13] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using a scale mixture of Gaussians in the wavelet domain," *IEEE Trans. Image Process*, Vol. 12, No. 11, pp. 1338–1351, 2003.
- [14] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *IEEE Trans. Image Processing*, Vol. 10, No. 7, pp. 1056–1068, 2001.
- [15] L. Sendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. Image Process*, Vol. 50, No. 11, pp. 2744–2756, 2002.
- [16] L. Sendur and I. W. Selesnick, "Bivariate shrinkage with local variance estimation," *IEEE Trans. Signal Processing Letters*, Vol. 9, No. 12, pp. 438–441, 2002.
- [17] J. M. Fadili and L. Boubchir, "Analytical form for a Bayesian wavelet estimator of images using the Bessel k form densities," *IEEE Trans. Image Processing*, Vol. 14, No. 2, pp. 231–240, 2005.
- [18] L. Boubchir and J. M. Fadili, "A closed-form nonparametric Bayesian estimator in the wavelet domain of images using an approximate alpha-stable prior," *Pattern Recognition Letters*, Vol. 27, No. 12, pp. 1370–1382, 2006.
- [19] C. Chaux, L. Duval, A. Benazza-Benyahia, and J. C. Pesquet, "A nonlinear Stein-based estimator for multichannel image denoising," *IEEE Trans. Image Process*, Vol. 56, No. 8, pp. 3855–3870, 2008.
- [20] Y. Hel-Or and D. Shaked, "A discriminative approach for wavelet denoising," *IEEE Trans. Image Process*, Vol. 56, No. 8, pp. 443–457, 2008.
- [21] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *The Annals of Statistics*, Vol. 9, No. 6, pp. 1135–1151, 1981.
- [22] P. Qiu and P. S. Mukherjee, "Edge structure preserving image denoising," *Signal Processing*, Vol. 90, No. 10, pp. 2851–2862, 2010.
- [23] I. Gijbels, A. Lambert, and P. Qiu, "Edge-preserving image denoising and estimation of discontinuous surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 7, pp. 1075–1087, 2006.
- [24] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA, 1992.
- [25] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3D filtering," *Proceedings of SPIE*, Vol. 6064, pp. 354–365, 2006.

- [26] A. Buades, B. Coll, and J.M. Morel, "Non-local means denoising," *Image Processing On Line*, Vol. 2011, 2011.
- [27] S. Roth and M.J. Black, "Fields of experts: A framework for learning image priors," *Computer Vision and Pattern Recognition*, IEEE Computer Society Conference on, Vol. 2, pp. 860-867, 2005.
- [28] S. Roth and M.J. Black, "Fields of experts," *International Journal of Computer Vision*, Vol. 82, No. 2, pp. 205-229, 2009.