IMAGE DENOISING BY POINTWISE THRESHOLDING OF THE UNDECIMATED WAVELET COEFFICIENTS: A GLOBAL SURE OPTIMUM

Florian Luisier and Thierry Blu

Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

ABSTRACT

We devise a new undecimated wavelet thresholding for denoising images corrupted by additive Gaussian white noise. The first key point of our approach is the use of a linearly parameterized pointwise thresholding function. The second key point consists in optimizing the parameters *globally* by minimizing Stein's unbiased MSE estimate (SURE) directly in the image-domain, and not separately in the wavelet subbands.

Amazingly, our method gives similar results to the best state-of-the-art algorithms, despite using only a simple pointwise thresholding function; we demonstrate it in simulations over a wide range of noise levels for a representative set of standard grayscale images.

Index Terms— Image denoising, undecimated wavelet thresholding, SURE minimization

1. INTRODUCTION

The multi-resolution analysis performed by the wavelet transform has been proved to be particularly efficient in image denoising. Since the early use of the classical orthonormal wavelet transform for removing additive Gaussian white noise through thresholding [1], a lot of work has been done [2, 3, 4, 5], leading to two main observations:

- better performances can be achieved with shift-invariant transformations such as the undecimated wavelet transform [6];
- further improvements can be obtained with more sophisticated thresholding functions which incorporates inter- and intra-scale dependencies.

In this paper, we will only concentrate on the first of these two observations, without forgetting the pioneer work of Donoho, to work out a new undecimated wavelet thresholding algorithm based on the minimization of Stein's unbiased MSE estimate (SURE) [7]. Up to our knowledge, the SURE has only been properly used in the framework of orthonormal wavelet transform. Indeed, its MSE preservation property allows a wavelet-domain subband-independent SURE minimization. Unfortunately, this classical approach is sub-optimal with the undecimated wavelet representation, because the MSE is not preserved anymore in the reconstruction, despite the fact that the undecimated wavelet transform (UWT) is a tight frame. In this paper, we show how to overcome this difficulty and efficiently apply the SURE to undecimated wavelet thresholding. The results show that, despite not using any form of inter- or intra-scale dependencies, our simple thresholding strategy rivals the best state-of-the-art algorithms in the undecimated wavelet transform setting[1].

The paper is organized as follows: first, we recall the matrix formulation of the undecimated wavelet transform and describe the main details of our denoising strategy; then, we expose our new SURE-based undecimated wavelet thresholding; finally, we demonstrate its efficiency.

2. CONTEXT

2.1. The undecimated wavelet transform

In this paper, we propose to use the classical undecimated wavelet transform whose dyadic analysis/synthesis filterbank for 1D signal is presented in Figure 1. The redundancy factor of this overcomplete wavelet expansion is $J+1$ where J stands for the number of decomposition levels. For 2D signals like images, we simply extend this representation in a separable way, leading to an overall redundancy factor of $3J+1$. Thanks to this redundancy, the UWT is a completely shift-invariant transformation.

The linearity of the UWT allows the following matrix formulation:

$$\mathbf{w} = \mathbf{D}\mathbf{x} \tag{1}$$

where: $\mathbf{D} = [\mathbf{D}_1 \dots \mathbf{D}_j \dots \mathbf{D}_{J+1}]^{\mathrm{T}}$ is a matrix of size $(J+1)N \times N$, where each $\mathbf{D}_j$ is a square matrix of size $N \times N$ and $\mathbf{x} \in \mathbb{R}^N$ is an input vector. The columns of the sub-matrices $\mathbf{D}_j$ are circularly shifted versions of the impulse response of the corresponding wavelet filter (or scaling filter for $\mathbf{D}_{J+1}$). Since the wavelet filters G and the scaling filter H are normalized, $\mathrm{diag}(\mathbf{D}_j^{\mathrm{T}}\mathbf{D}_j) = \mathbf{1} \in \mathbb{R}^N, \forall j \in [1; J+1]$.

This work was supported by the Center for Biomedical Imaging (CIBM) of the Geneva - Lausanne Universities and the EPFL, the foundations Leenaards and Louis-Jeantet, as well as by the Swiss National Science Foundation under grant 200020-109415.

[1]The best UWT for these methods turn out to be the *Haar* UWT for the vast majority of test images.

Perfect reconstruction $\mathbf{RD} = \mathbf{I}$ is then achieved as follows:

$$\mathbf{x} = \mathbf{R}\mathbf{w} \tag{2}$$

where: $\mathbf{R} = [\frac{1}{2}\mathbf{D}_1 \dots \frac{1}{2^j}\mathbf{D}_j \dots \frac{1}{2^J}\mathbf{D}_J \ \frac{1}{2^J}\mathbf{D}_{J+1}] = \mathbf{D}^{\mathrm{T}}\Lambda$.

$\Lambda = \begin{bmatrix} \frac{1}{2}\mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \frac{1}{2^J}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \frac{1}{2^J}\mathbf{I} \end{bmatrix}$ is a diagonal matrix of size $(J+1)N \times (J+1)N$ which compensates for the redundancy of the transformation.

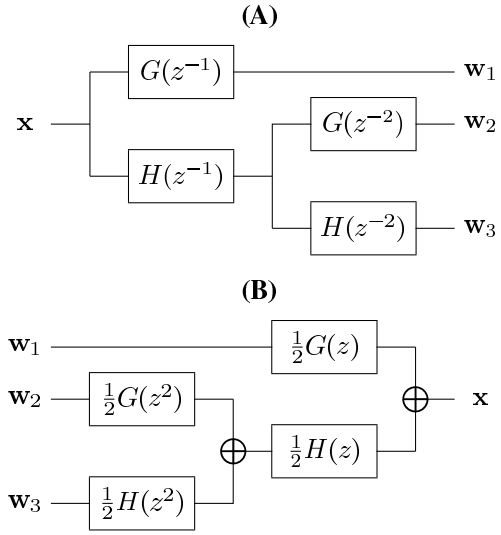

Fig. 1. The classical undecimated wavelet filterbank for 1D signal. **(A)** Analysis. **(B)** Synthesis.

2.2. Our denoising strategy

The aim of the denoising algorithm is to find the best estimate $\hat{\mathbf{x}} \in \mathbb{R}^N$ of an unknown signal $\mathbf{x} \in \mathbb{R}^N$ from a noisy observation of it $\mathbf{y} = \mathbf{x} + \mathbf{b} \in \mathbb{R}^N$. Every element of the noise $\mathbf{b}$ is assumed to follow an independent normal law specified by a zero-mean and a known variance σ^2.

In this context, our denoising strategy consists of the following steps:

1. Apply the UWT to the noisy signal $\mathbf{y}$ to get the transformed noisy coefficients $\mathbf{w} = \mathbf{D}\mathbf{y} = (\mathbf{w}_j)_{j\in[1;J+1]}$, where $\mathbf{w}_j \in \mathbb{R}^N$;
2. Apply a *pointwise* thresholding function Γ to the transformed noisy coefficients to get:

   $$\hat{\mathbf{w}} = \Gamma(\mathbf{w}),$$

   a vector of $(J+1) \times N$ components $\hat{w}_{j,n} = \gamma_j(w_{j,n})$, where γ_j, $j = 1 \cdots (J+1)$, are nonlinear functions;
3. Come back to the original domain by applying the inverse UWT to the thresholded coefficients.

The whole denoising process can thus be fully described by a function of the noisy input coefficients:

$$\hat{\mathbf{x}} = \mathrm{F}(\mathbf{y}) = \mathbf{R}\Gamma(\mathbf{D}\mathbf{y}) \tag{3}$$

The first key point of our approach is to build F as a *linear* combination of simple —possibly nonlinear— functions $\mathrm{F}_{j,k}$, that is:

$$\mathrm{F}(\mathbf{y}) = \sum_{j=1}^{J+1}\sum_{k=1}^{K} a_{j,k} \underbrace{\mathbf{R}\Gamma_{j,k}(\mathbf{D}\mathbf{y})}_{\mathrm{F}_{j,k}(\mathbf{y})}, \tag{4}$$

where: $\Gamma_{j,k}(\mathbf{w}) = [\mathbf{0}^{\mathrm{T}} \cdots \mathbf{0}^{\mathrm{T}}\ \Gamma_k(\mathbf{w}_j)^{\mathrm{T}}\ \mathbf{0}^{\mathrm{T}} \cdots]^{\mathrm{T}}$, i.e., the simple denoising function $\mathrm{F}_{j,k}$ modifies only subband j.

Thanks to this linear parametrization, the search for the optimal —in the minimum mean square error (MMSE) sense— parameters a_k boils down to the resolution of a *linear system of equations* (see section 3.3), because the MSE, defined as

$$\mathrm{MSE} = \frac{1}{N}\|\hat{\mathbf{x}} - \mathbf{x}\|^2 = \frac{1}{N}\|\mathrm{F}(\mathbf{y}) - \mathbf{x}\|^2, \tag{5}$$

is a quadratic function in $\mathrm{F}(\mathbf{y})$ and thus, in the $a_{j,k}$'s.

Of course, since we do not have access to the noise-free signal $\mathbf{x}$, we can not compute the actual MSE. To overcome this difficulty, we will use Stein's unbiased MSE estimate (SURE) [7] (see section 3.1).

The second key point of our approach is that the parameters optimization is performed in the *image domain*, contrary to what is usually done in other UWT-based denoising algorithms. This image domain optimization is unavoidable when using the UWT, because the MSE is not preserved in the transformed domain due to the non-orthonormality of the transformation.

3. SURE-BASED ALGORITHM FOR UNDECIMATED WAVELET THRESHOLDING

3.1. MSE estimation

It is possible to have an unbiased estimate of the MSE using only the noisy observation $\mathbf{y}$, not the original noise-free signal $\mathbf{x}$. There is even no need to put any statistical assumption on this unknown underlying signal. This unbiased MSE estimate, known as SURE, takes the following expression:

$$\begin{aligned} \epsilon \ = \ & \frac{1}{N}\Big(\|\mathrm{F}(\mathbf{y})\|^2 - 2\mathrm{F}(\mathbf{y})^{\mathrm{T}}\mathbf{y} + 2\sigma^2 \mathrm{div}\{\mathrm{F}(\mathbf{y})\}\Big) \\ & + \frac{1}{N}\|\mathbf{y}\|^2 - \sigma^2 \end{aligned} \tag{6}$$

In general, it is quite painful to express the divergence term which appears in equation (6). Fortunately, in the framework of the undecimated wavelet transform presented in section 2, this computation becomes simple, leading to the following form of the MSE estimate:

Theorem 1 *After a pointwise undecimated wavelet processing, the MSE can be unbiasedly estimated by the following random variable:*

$$\epsilon = \frac{1}{N}\Big(\|\mathrm{F}(\mathbf{y})\|^2 - 2\mathrm{F}(\mathbf{y})^{\mathrm{T}}\mathbf{y} + 2\sigma^2\mathrm{diag}(\Lambda)^{\mathrm{T}}\Gamma'(\mathbf{D}\mathbf{y})\Big) + \frac{1}{N}\|\mathbf{y}\|^2 - \sigma^2 \quad (7)$$

i.e.:

$$\mathscr{E}\{\epsilon\} = \frac{1}{N}\mathscr{E}\{\|\mathrm{F}(\mathbf{y}) - \mathbf{x}\|^2\}$$

where $\mathscr{E}\{\cdot\}$ stands for the expectation operator and $\Gamma'(\mathbf{w}) = [\Gamma'_1(\mathbf{w}_1)^{\mathrm{T}} \cdots \Gamma'_{J+1}(\mathbf{w}_{j+1})^{\mathrm{T}}]^{\mathrm{T}}$.

Since the variance of the MSE estimate given by Theorem 1 mainly depends on the number of samples N, ϵ is really close to the actual MSE in image denoising applications.

3.2. Parametric pointwise denoising function

To apply Theorem 1, we need a differentiable pointwise thresholding function; for an easier optimization, it is really appropriate to use a linear parametrization. Both of these constraints are satisfied by the following function

$$\gamma(w) = \sum_{k=1}^{K} a_k \underbrace{we^{-(k-1)\frac{w^2}{3K^2\sigma^2}}}_{\theta_k(w)} \quad (8)$$

which is simply a weighted sum of first derivatives of Gaussians. For example, for $K = 2$ terms, we obtain:

$$\gamma(w) = a_1 w + a_2 we^{-\frac{w^2}{12\sigma^2}}, \quad (9)$$

which has been already shown in [8] to be more efficient than the classical *soft-thresholding*.

Furthermore, we propose to link the number of parameters K to the level of decomposition j as follows:

$$K_j = \min(J - j + 2, 4) \quad (10)$$

to speed up the algorithm by assigning less parameters to the coarsest scales.

3.3. Parameters optimization

Our aim is to find the remaining linear parameters $a_{j,k}$ which will minimize the MSE estimate ϵ. Starting from equation (7), these are solved for in the following way:
$\forall (k,j) \in [1;K] \times [1;J+1]$, we have:

$$\frac{\partial \epsilon}{\partial a_{j,k}} = 0 \Longleftrightarrow \mathbf{a} = \mathbf{M^{-1}c} \quad (11)$$

where: $\mathbf{a}$, $\mathbf{c}$ and $\mathbf{M}$ are vector/matrix straightforwardly built using, respectively, the coefficients $a_{j,k}$, the SURE expressions $\Big(\mathrm{F}_{j,k}(\mathbf{y})^{\mathrm{T}}\mathbf{y} - \sigma^2\mathrm{diag}(\Lambda)^{\mathrm{T}}\Gamma'_{j,k}(\mathbf{D}\mathbf{y})\Big)$ and the scalar products $\mathrm{F}_{j,k}(\mathbf{y})^{\mathrm{T}}\mathrm{F}_{j',k'}(\mathbf{y})$.

Because of this linear feature, our algorithm is fast and simple to implement.

4. EXPERIMENTS

4.1. Optimization method assessment

Before comparing our algorithm with the best state-of-the-art denoising schemes, we have performed several experiments on simulated additive Gaussian white noise to validate our image-domain optimization method. We have noticed an overall gain of up to 1 dB compared to an independent undecimated wavelet in-band MSEs minimization. This large gap is caused by the high correlation of the UWT wavelet subbands.

4.2. Comparison with state-of-the-art denoising schemes

We propose to compare our pointwise undecimated wavelet thresholding with two other wavelet denoising procedures: the *BiShrink* [5] and the *BLS-GSM* [3] which gives, up to our knowledge, the best results known in the literature. Contrary to our approach, both of these methods integrate the parent —coefficient at the next coarser scale— and a local neighborhood to better estimate the current noisy coefficient. We have run these algorithms under the same conditions as ours using the respective Matlab codes of the authors, kindly provided on their respective webpages. Their parameters have been chosen according to what was suggested in their papers [5, 3].

We have applied five decomposition stages of an undecimated wavelet transform for each of the various denoising procedures. Surprisingly, the best results were obtained for all methods with the *Haar* filters for over a vast majority of test images. To perform a reliable PSNR[2] comparison, we have then averaged the results over eight noise realizations in an input PSNR range of $8.13 - 34.15$ dB.

In Table 1, we have reported the output PSNRs of the various algorithms and, in Figure 2, we show a visual result. As it can be observed, our method is really competitive: we achieve the same average quality as the *BLS-GSM* and an average gain of 0.4 dB against the *BiShrink*. These observations have been confirmed over most standard test images, to the notable exception of *Barbara*, where we obtain similar results as the *BiShrink*, but experience an average loss of 0.5 dB compared to the *BLS-GSM*. This may suggest that with textured images, the integration of local neighborhood information may become important.

From a practical point of view, the computation time of our method is approximately 5s for 256×256 images and about 35s for 512×512 images, whereas the *BLS-GSM* required respectively 12s and 60s under the same conditions.

5. CONCLUSION

We have presented a new pointwise SURE-based undecimated wavelet thresholding algorithm. The parameters involved in

[2] defined as $\mathrm{PSNR} = 10\log 10\frac{255^2}{\mathrm{MSE}}$

Table 1. Comparison of some of the most efficient denoising methods (UWT *Haar*).

σ	5	10	20	30	50	100
Input PSNR	34.15	28.13	22.11	18.59	14.15	8.13
Method	**House** 256×256					
BiShrink	38.40	34.94	31.90	30.04	27.62	24.14
BLS-GSM	38.35	35.29	32.39	30.53	28.01	24.59
Our method	**38.78**	**35.49**	**32.44**	**30.59**	**28.15**	**24.79**
Method	**Lena** 512×512					
BiShrink	37.98	34.73	31.71	29.93	27.66	24.56
BLS-GSM	38.26	**35.19**	**32.10**	**30.25**	27.93	25.05
Our method	**38.29**	35.08	32.00	30.21	**28.03**	**25.28**
Method	**Boat** 512×512					
BiShrink	36.65	33.12	29.97	28.17	25.96	23.17
BLS-GSM	36.82	33.51	**30.31**	**28.44**	26.17	23.44
Our method	**37.18**	**33.57**	30.26	28.42	**26.21**	**23.59**
Method	**Al** 512×512					
BiShrink	38.20	34.79	31.85	30.20	28.00	24.72
BLS-GSM	38.89	35.51	32.42	30.60	28.30	25.26
Our method	**39.05**	**35.60**	**32.46**	**30.72**	**28.55**	**25.59**

Note: output PSNRs have been averaged over eight noise realizations.

the thresholding function are linear and therefore, can be solved for elegantly by minimizing Stein's unbiased MSE estimate (SURE). Due to the non-orthonormality of the undecimated wavelet transform, we have proposed a rigorous image-domain optimization which finally makes our pointwise algorithm competitive with the best state-of-the-art denoising procedures.

6. REFERENCES

[1] David L. Donoho and Iain M. Johnstone, "Adapting to Unknown Smoothness via Wavelet Shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, December 1995.

[2] S. Grace Chang, Bin Yu, and Martin Vetterli, "Adaptive Wavelet Thresholding for Image Denoising and Compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, September 2000.

[3] Javier Portilla, Vasily Strela, Martin J. Wainwright, and Eero P. Simoncelli, "Image Denoising using Scale Mixtures of Gaussians in the Wavelet Domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, November 2003.

[4] Aleksandra Pižurica and Wilfried Philips, "Estimating the Probability of the Presence of a Signal of Interest in Multiresolution Single- and Multiband Image Denoising," *IEEE Transactions on Image Processing*, vol. 15, no. 3, pp. 654–665, March 2006.

[5] Levent Sendur and Ivan W. Selesnick, "Bivariate Shrinkage With Local Variance Estimation," *IEEE Signal Processing Letters*, vol. 9, no. 12, pp. 438–441, December 2002.

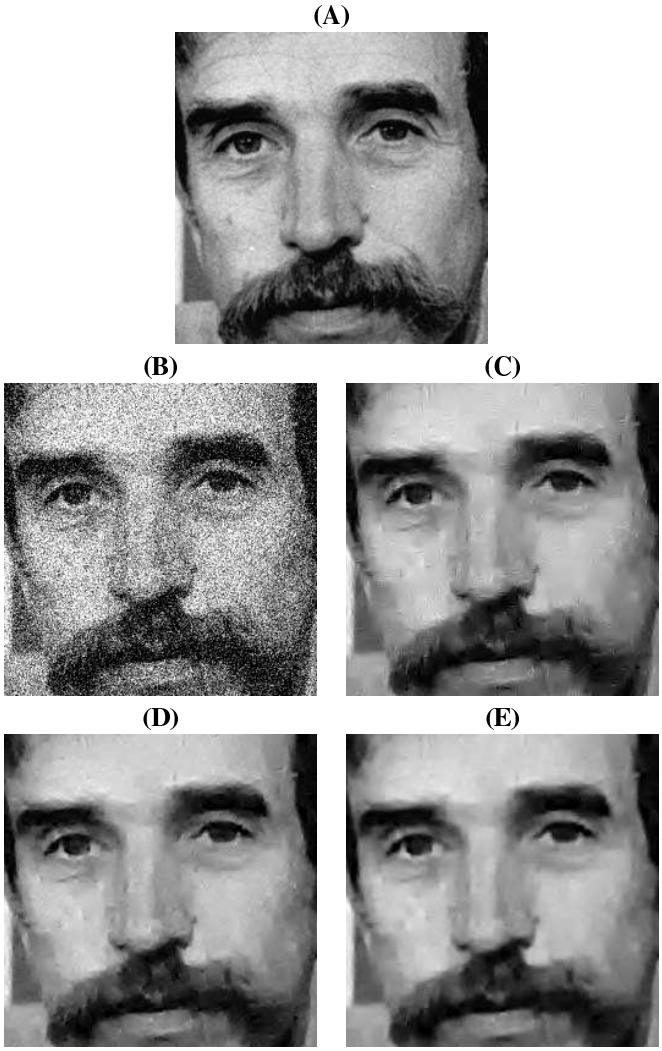

Fig. 2. **(A)** Part of the noise-free *Al* image. **(B)** Noisy image: PSNR $= 14.15$ dB. **(C)** *BiShrink*: PSNR $= 28.05$ dB. **(D)** *BLS-GSM*: PSNR $= 28.34$ dB. **(E)** Our method: PSNR $= 28.62$ dB.

[6] S. Mallat, "A wavelet Tour of Signal Processing," *Academic Press*, 1998.

[7] C. Stein, "Estimation of the Mean of a Multivariate Normal Distribution," *The Annals of Statistics*, vol. 9, pp. 1135–1151, 1981.

[8] Florian Luisier, Thierry Blu, and Michael Unser, "A New SURE Approach to Image Denoising: Inter-scale Orthonormal Wavelet Thresholding," *To appear in IEEE Transactions on Image Processing*, 2007.