SURE-BASED WAVELET THRESHOLDING INTEGRATING INTER-SCALE DEPENDENCIES

Florian Luisier, Thierry Blu and Michael Unser

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

ABSTRACT

We propose here a new pointwise wavelet thresholding function that incorporates inter-scale dependencies. This non-linear function depends on a set of four linear parameters per subband which are set by minimizing Stein's unbiased MSE estimate (SURE). Our approach assumes additive Gaussian white noise.

In order for the inter-scale dependencies to be faithfully taken into account, we also develop a rigorous feature alignment processing, that is adapted to arbitrary wavelet filters (e.g. non-symmetric filters).

Finally, we demonstrate the efficiency of our denoising approach in simulations over a wide range of noise levels for a representative set of standard images.

Index Terms— Orthonormal wavelet transform, inter-scale dependencies, MSE estimate, denoising

1. INTRODUCTION

Over the past fifteen years, the wavelet transform (whether redundant or not) has been shown to be particularly efficient in image denoising, because of its energy compaction and subband decorrelation properties [2, 4, 5, 6]. Here, we will only consider critically-sampled orthonormal transforms for three main reasons:

- computation time;
- minimal memory occupation (a critical point when dealing with volumes and higher dimensional signals);
- MSE preservation (justifies independent processing of the subbands).

Previous studies [5, 6] have demonstrated that the denoising quality can be slightly improved by taking into account the scale dependencies between the wavelet coefficients (known as the parent-child relationship). Although within-scale dependencies seem to bring substantially larger denoising gain, this is done at the expense of a much heavier computation load [6], which is why we will stick to a simpler pointwise approach in this paper. Indeed, we will show that an adequate choice of the threshold function brings denoising results that are competitive with the best state-of-the-art non-redundant techniques.

This paper is organized in three parts: first, we will work out a new processing algorithm for aligning parent and child subbands. Then, we will present our new threshold function and show how to optimize for its free parameters using Stein's unbiased MSE estimate (SURE) principle. Finally, we demonstrate the superior performance of our approach.

This work was supported in part by the Swiss National Science Foundation under grant 200020-109415.

2. INTER-SCALE DEPENDENCIES

2.1. Features alignment between adjacent scales

In a critically sampled orthonormal wavelet decomposition, the parent subband is half-size the child subband (see Figure 1). The usual way of putting the two subbands in correspondence is simply to expand the parent by a factor two.

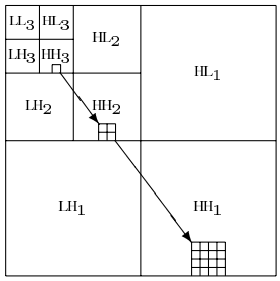

Fig. 1. Three stages of a fully decimated orthogonal wavelet transform and the so-called parent-child relationship.

Here, we propose a more sophisticated solution, which ensures the image features alignment between the child and its parent. Our idea is to filter the low-pass subband at scale j by a suitable filter W along the corresponding high-pass filtered rows or columns in order to obtain images of the same size as HL_j, LH_j and HH_j, as depicted in Figure 2. Considering the filterbank of Figure 3, we say that the prediction map $z(k)$ is aligned with $y_h(k)$ if and only if:

- $H(z^{-1})W(z^2) = G(z^{-1})Q(z)$, where $Q(z)$ is a symmetric (zero phase) filter;

- the alignment filter, of impulse response $\{w_n\}_{n\in\mathbb{Z}}$, preserves the signal energy, i.e.:

$$\sum_n w_n^2 = 1$$

In addition, the filter has to be high-pass.

We can prove that the general form of such a filter is given by:

Theorem 1 *In order for the output of the dyadic filterbank of Figure 3 to be aligned, it is necessary and sufficient that:*

$$W(z^2) \quad = \quad G(z^{-1})G(-z^{-1})(\epsilon + z^{-2})R(z^2) \qquad (1)$$

where $\epsilon = \pm 1$ *and* $R(z) = R(z^{-1})$ *is arbitrary.*

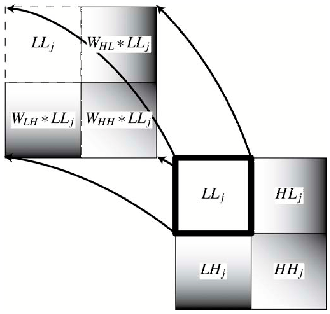

Fig. 2. Obtaining the whole parent information out of the low-pass subband.

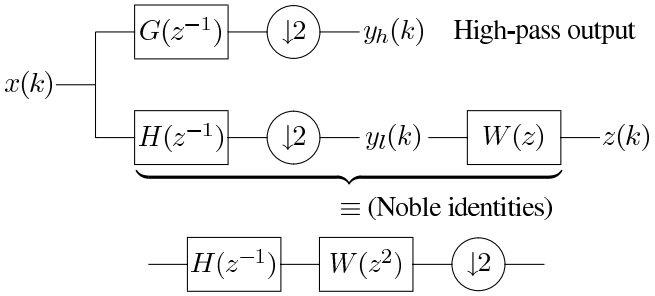

Fig. 3. One iteration of a dyadic orthonormal filterbank, where $G(z) = -z^{-1}H(-z^{-1})$.

From a practical point of view, we aim at using the shortest possible W. It is easy to see that in the case of symmetric or nearly symmetric —such as the Daubechies *symlets* [3]— filters, the shortest W satisfying the perfect alignment condition is in fact the simple gradient filter: $W(z) = 1 - z$. If the symmetry or the quasi-symmetry is not centered at the origin but at a position n_0, then $W(z) = z^{-n_0}(1 - z)$. When the low-pass filter is asymmetric, we can simply take $R(z^2) = 1$ in (1).

Additionally, in order to increase the homogeneity inside regions of similar magnitude coefficients, we apply a 2D-smoo-thing filter by a normalized Gaussian kernel.

Figure 4 shows the whole construction of the final interscale prediction, to which we will refer from now on as Y.

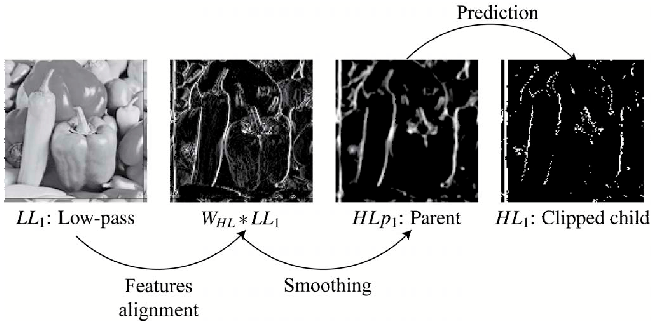

Fig. 4. Building an efficient interscale predictor, illustrated with a particular subband (HL_1) of the noise-free *Peppers* image.

3. A NEW DATA-DRIVEN THRESHOLDING FUNCTION

Let us denote by $y = (y_n)_{n\in[1;N]}$ the noisy image:

$$y = \underbrace{x}_{\text{noise-free image}} + \underbrace{b}_{\text{Gaussian white noise}} \qquad (2)$$

The noise samples b_n are assumed to follow a normal law $\mathcal{N}(0, \sigma^2)$ with known σ.

Due to orthonormality, this noise model carries over to the wavelet domain, i.e. $y^j = x^j + b^j$, with $b_n^j \sim \mathcal{N}(0, \sigma^2)$. Thanks to orthonormality again, the MSE can be written as a weighted sum of the MSE in every subband j:

$$\begin{aligned} \text{MSE} &= < |x - y|^2 > \\ &= \sum_{j=1}^{J} \frac{N^j}{N} \underbrace{< |x^j - y^j|^2 >}_{\text{MSE}^j} \end{aligned} \qquad (3)$$

where we have introduced the notation:

$$< u >= \frac{1}{N} \sum_{n=1}^{N} u_n \qquad (4)$$

The denoising process can therefore be applied independently in each wavelet subband j. For the next part, we will thus drop any reference to j.

3.1. Stein's unbiased MSE estimate

The main difficulty of the image denoising problem is that we have to minimize a quantity —the MSE— that we cannot compute, because in practice we do not have access to the

noise-free image. Fortunately, this difficulty has been overcome by Stein in [1] who has proposed an unbiased estimate of the MSE —or "risk", in the statistical literature. Our aim is to find the best estimate $\tilde{x} = \theta(y, Y)$ of the noise-free data x using the noisy data y and the noisy parent information Y. For this purpose, we can apply the following theorem which is a consequence of [1, Theorem 1]:

Theorem 2 *The following random variable:*

$$\begin{aligned}\epsilon &= <\theta(y,Y)^2 - 2y\theta(y,Y) + 2\sigma^2\theta'(y,Y)> \\ &\quad + <y^2> - \sigma^2 \\ &= \frac{1}{N}\sum_{n=1}^{N}\Big(\theta^2(y_n,Y_n) - 2y_n\theta(y_n,Y_n) + 2\sigma^2\theta'(y_n,Y_n)\Big) \\ &\quad + \frac{1}{N}\sum_{n=1}^{N} y_n^2 \;-\; \sigma^2 \end{aligned} \tag{5}$$

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

$$\mathscr{E}\{\epsilon\} = \mathscr{E}\{<|\theta(y,Y) - x|^2>\}$$

where $\mathscr{E}\{\cdot\}$ stands for the expectation operator.

The result given by Theorem 2 is particularly interesting in image denoising applications because the number of samples is large. In that case the standard deviation of ϵ is small; i.e., close to its expectation which is the MSE of the denoising procedure. As a result, we can use ϵ as if it were the true MSE.

3.2. Construction of a parametric thresholding function

We propose now the construction of an efficient parameterized wavelet denoising function θ. For this purpose, we want it to be differentiable (necessary condition to apply Theorem 2) and anti-symmetric (the wavelet coefficients are not supposed to exhibit a preference of sign). Moreover, we want a non-linear transition between low-magnitude and high-magnitude coefficients. Without consideration of the inter-scale predictor Y, these requirements are satisfied by the following parametric thresholding function:

$$\theta_0(y;\mathbf{u},T) = \Big(u_1 + u_2 e^{-\frac{y^2}{T^2}}\Big)y \tag{6}$$

where: $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ and T are parameters.

The role of the exponential is crucial, because it rules the non-linear transition between low-SNR and high-SNR wavelet coefficients. The inter-scale predictor developed in section 2 can be integrated into this denoising function by using two different θ_0 depending on whether the parent is large or not:

$$\begin{aligned}\theta(y,Y;\mathbf{u},\mathbf{v},T) &= e^{-\frac{Y^2}{T^2}}\;\theta_0(y;\mathbf{u},T) \\ &\quad + \Big(1 - e^{-\frac{Y^2}{T^2}}\Big)\;\theta_0(y;\mathbf{v},T)\end{aligned} \tag{7}$$

where: $\mathbf{u}$, $\mathbf{v}$ and T are some parameters that are determined by minimizing (5) as shown in the next section.

$\theta(y,Y;\mathbf{u},\mathbf{v},T)$	$\|Y\| >> T$	$\|Y\| << T$
$\|y\| >> T$	$v_1 y$	$u_1 y$
$\|y\| << T$	$(v_1+v_2)y$	$(u_1+u_2)y$

Table 1. Limit behavior of the pointwise thresholding function (7).

3.3. Parameters optimization

The parameter T of our function (6) and (7) can be viewed as a threshold, in the same way as in the well-known *soft-thresholding* function developed by Donoho *et al.* in [2]. But in our case, its sensitivity is really small and thus there is no need to optimize it in each wavelet subband. We have experimentally found that it can be directly linked to the noise standard deviation σ according to $T = 3.5\sigma$.

In order to find the optimal parameters $\mathbf{u}$ and $\mathbf{v}$, we can simply apply Theorem 2 by introducing our pointwise inter-scale dependent thresholding function $\theta(y,Y;\mathbf{u},\mathbf{v},T)$ into (5). Since ϵ is quadratic in u_1, u_2, v_1 and v_2, minimizing it leads to a linear system of equations:

$$\underbrace{\begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{12} & m_{22} & m_{23} & m_{24} \\ m_{13} & m_{23} & m_{33} & m_{34} \\ m_{14} & m_{24} & m_{34} & m_{44} \end{bmatrix}}_{\mathbf{M}} \underbrace{\begin{bmatrix} u_1 \\ u_2 \\ v_1 \\ v_2 \end{bmatrix}}_{\mathbf{p}} = \underbrace{\begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix}}_{\mathbf{c}}$$

$$\mathbf{p} \;=\; \mathbf{M}^{-1}\mathbf{c} \tag{8}$$

where:

$m_{11} =< f_p^2 y^2 >$ $\quad$ $m_{33} =< \bar{f}_p^2 y^2 >$
$m_{12} =< f_p^2 f y^2 >$ $\quad$ $m_{34} =< \bar{f}_p^2 f y^2 >$
$m_{13} =< f_p \bar{f}_p y^2 >$ $\quad$ $m_{44} =< \bar{f}_p^2 f^2 y^2 >$
$m_{14} =< f_p \bar{f}_p f y^2 >$ $\quad$ $c_1 =< f_p y^2 - \sigma^2 f_p >$
$m_{22} =< f_p^2 f^2 y^2 >$ $\quad$ $c_2 =< f_p f y^2 - \sigma^2 f_p (f + f'y) >$
$m_{23} =< f_p \bar{f}_p f y^2 >$ $\quad$ $c_3 =< \bar{f}_p y^2 - \sigma^2 \bar{f}_p >$
$m_{24} =< f_p \bar{f}_p f^2 y^2 >$ $\quad$ $c_4 =< \bar{f}_p f y^2 - \sigma^2 \bar{f}_p (f + f'y) >$

where: $f = e^{-\frac{y^2}{T^2}}$; $\bar{f} = 1 - f$; $f_p = e^{-\frac{Y^2}{T^2}}$; $\bar{f}_p = 1 - f_p$.

The optimal set of parameters $\mathbf{p}$ is thus the solution of a linear system of equations, which makes our approach very simple to implement.

4. METHOD EVALUATION

We propose now to compare our pointwise inter-scale dependent thresholding function (7) with some of the best state-of-the-art denoising procedures. The first one, by Sendur *et al.* [5], is known as the *BiShrink*. The second one, by Portilla *et al.* [6], is known as the *BLS-GSM* and gives the best state-of-the-art results. Contrary to our approach, these two methods require an explicit prior to model the noise-free coefficients. Moreover, besides integrating the parent-child dependency, they both take into account the local neighborhood to better estimate the child coefficient, which is not yet the case of our approach. We have also added in our comparison the popular *BayesShrink* [4] which is a Bayesian soft-thresholding rule.

We have applied four decomposition stages of a critically-sampled orthonormal (*symlet* with eight vanishing moments) filterbank and denoise all the high-pass wavelet subbands using the respective Matlab codes of the authors, kindly provided on their respective webpages. Their parameters have been chosen accordingly to what was suggested in their referred papers. To perform a reliable PSNR[1] comparison, we have then averaged the results over ten noise realizations in an input PSNR range of $8.13-34.15$ dB. In Figure 5, we have reported the PSNR differences between our method (reference) and the other ones and, in Figure 6, we show a visual result. As one can observe, our method is really competitive (average gain of: 0.2 dB towards the *BLS-GSM*; 0.6 dB towards the *BiShrink* and up to 1 dB towards the *BayesShrink*). This lead has been confirmed over most standard test images, to the notable exception of *Barbara*; this may suggest that with textured images, the integration of local neighborhood information is probably important.

The computation time of our method is below 1s for 256×256 images and about 3s for 512×512 images, which is approximately four times faster than the *BLS-GSM*.

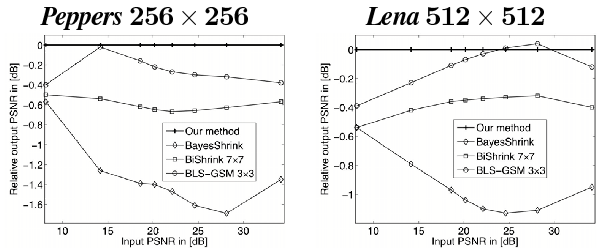

Fig. 5. PSNR comparison between some of the best state-of-the-art methods and our pointwise inter-scale dependent thresholding function (7).

5. CONCLUSION

We have proposed a new parametric thresholding function, whose parameters can be solved for by minimizing an unbiased estimate of the MSE. In order to take inter-scale dependencies into account, we have devised a new way of building the parent subband used to predict its child. By appropriately integrating this inter-scale predictor into the thresholding function, we have demonstrated that our approach favorably compares to the other non-redundant state-of-the-art methods.

6. REFERENCES

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

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

Fig. 6. **(A)** Part of the noise-free *Lena* image. **(B)** Noisy image: PSNR = 20.17 dB. **(C)** BayesShrink: PSNR = 29.31 dB. **(D)** Our method: PSNR = 30.36 dB.

[2] 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.

[3] I. Daubechies, "Ten Lectures on Wavelets," *CBMS-NSF Regional Conference series in Applied Mathematics*, vol. 61 of Proc., March 1992.

[4] 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, September 2000.

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

[6] 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, November 2003.