A New SURE Approach to Image Denoising: Interscale Orthonormal Wavelet Thresholding

Florian Luisier, Thierry Blu, *Senior Member, IEEE*, and Michael Unser, *Fellow, IEEE*

***Abstract*—This paper introduces a new approach to orthonormal wavelet image denoising. Instead of postulating a statistical model for the wavelet coefficients, we directly parametrize the denoising process as a sum of elementary nonlinear processes with unknown weights. We then minimize an estimate of the mean square error between the clean image and the denoised one. The key point is that we have at our disposal a very accurate, statistically unbiased, MSE estimate—Stein's unbiased risk estimate—that depends on the noisy image alone, not on the clean one. Like the MSE, this estimate is quadratic in the unknown weights, and its minimization amounts to solving a linear system of equations. The existence of this *a priori* estimate makes it unnecessary to devise a specific statistical model for the wavelet coefficients. Instead, and contrary to the custom in the literature, these coefficients are not considered random anymore. We describe an interscale orthonormal wavelet thresholding algorithm based on this new approach and show its near-optimal performance—both regarding quality and CPU requirement—by comparing it with the results of three state-of-the-art nonredundant denoising algorithms on a large set of test images. An interesting fallout of this study is the development of a new, group-delay-based, parent–child prediction in a wavelet dyadic tree.**

***Index Terms*—Image denoising, interscale dependencies, orthonormal wavelet transform, Stein's unbiased risk estimate (SURE) minimization.**

I. INTRODUCTION

DURING acquisition and transmission, images are often corrupted by additive noise that can be modeled as Gaussian most of the time. The main aim of an image denoising algorithm is then to reduce the noise level, while preserving the image features. The multiresolution analysis performed by the wavelet transform has been shown to be a powerful tool to achieve these goals. Indeed, in the wavelet domain, the noise is uniformly spread throughout the coefficients, while most of the image information is concentrated in the few largest ones (sparsity of the wavelet representation).

The most straightforward way of distinguishing information from noise in the wavelet domain consists of thresholding the wavelet coefficients. Of the various thresholding strategies, *soft-thresholding* is the most popular and has been theoretically justified by Donoho and Johnstone [1]. These authors have shown that the shrinkage rule is near-optimal in the minimax sense and provided the expression of the optimal threshold value T—called the "universal threshold"—as a function of the noise power σ^2 when the number of samples N is large: $T = \sqrt{2\sigma^2 \log N}$. The use of the universal threshold to denoise images in the wavelet domain is known as *VisuShrink* [2].

Yet, despite its theoretical appeal, minimax is different from mean-squared error (MSE) as a measure of error. A lot of work has been done to propose alternative thresholding strategies that behave better in terms of MSE than *VisuShrink* [3]–[6]. Donoho and Johnstone themselves acknowledged this flaw and suggested to choose the optimal threshold value T by minimizing Stein's unbiased risk estimator (SURE) [7] when the data fail to be sparse enough for the minimax theory to be valid. This hybrid approach has been coined *SureShrink* by their authors [1]. Without challenging the soft-thresholding strategy, alternative threshold value selections have been proposed as well. One of the most popular was proposed by Chang *et al.*, who derived their threshold in a Bayesian framework, assuming a generalized Gaussian distribution for the wavelet coefficients. This solution to the wavelet denoising problem is known as *BayesShrink* [8] and has a better MSE performance than *SureShrink*.

Beyond the pointwise approach, more recent investigations have shown that substantially larger denoising gains can be obtained by considering the intra- and interscale correlations of the wavelet coefficients. In addition, increasing the redundancy of the wavelet transform is strongly beneficial to the denoising performances, a point to which we will come back later. We have selected three such techniques reflecting the state-of-the-art in wavelet denoising, against which we will compare our results.

- *Portilla et al.* [9]:[1] Their main idea is to model the neighborhoods of coefficients at adjacent positions and scales as a Gaussian scale mixture (GSM); the wavelet estimator is then a Bayes least squares (BLS). Their denoising method, consequently called *BLS-GSM*, is the most efficient up-to-date approach.
- *Pižurica et al.* [10]:[2] Assuming a generalized Laplacian prior for the noise-free data, their approach called *ProbShrink* is driven by the estimation of the probability that a given coefficient contains significant information—notion of "signal of interest."
- *Sendur et al.* [11], [12]:[3] Their method, called *BiShrink*, is based on new non-Gaussian bivariate distributions

---

Manuscript received February 28, 2006; revised September 14, 2006. This work was supported in part by the Center for Biomedical Imaging (CIBM) of the Geneva–Lausanne Universities and the EPFL, in part by the foundations Leenaards and Louis-Jeantet, and in part by the Swiss National Science Foundation under Grant 200020-109415. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Srdjan Stankovic.

The authors are with the Biomedical Imaging Group (BIG), Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland (e-mail: florian.luisier@epfl.ch; thierry.blu@epfl.ch; michael.unser@epfl.ch).

Digital Object Identifier 10.1109/TIP.2007.891064

[1] Available at http://www.decsai.ugr.es/~javier/denoise/software/index.htm, with a 3 × 3 neighborhood as suggested by the authors.

[2] Available at http://www.telin.ugent.be/~sanja/, with a 3 × 3 neighborhood and a threshold value $T = \sigma$ as suggested by the authors.

[3] Available at http://www.taco.poly.edu/WaveletSoftware/denoise2.html, with a 7 × 7 neighborhood as suggested by the authors.

1057-7149/$25.00 © 2007 IEEE

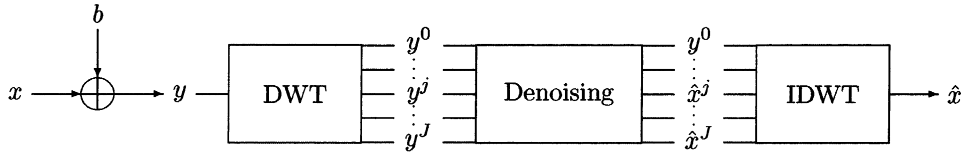

Fig. 1. Principle of wavelet denoising.

to model interscale dependencies. A nonlinear bivariate shrinkage function using the maximum *a posteriori* (MAP) estimator is then derived. In a second paper, these authors have extended their approach by taking into account the intrascale variability of wavelet coefficients.

These techniques have been devised for both redundant and nonredundant transforms.

Despite reports on the superior denoising performances of redundant transforms [13], [14], we will only consider critically sampled wavelet transforms in this paper. The rationale behind our choice is that, since there is no *added* information—only *repeated* information—in redundant transforms, we believe that, eventually, a nonredundant transform may match the performance of redundant ones. This would potentially be very promising since the major drawback of redundant transforms are their memory and CPU time requirements which limits their routine use for very large images and, above all, usual volumes of data.

More than a specific denoising algorithm, this paper is about a powerful new method for optimizing *beforehand*—unaware of the clean image—the performance of a denoising method. Here, we want in particular to promote Stein's unbiased risk estimate (SURE) which is nothing less than an *a priori* estimation of the MSE resulting from an arbitrary processing of noisy data. This estimate turns out to be more accurate as more data are available, which is the case of images. Wavelet denoising methods routinely involve a statistical description of the coefficient distribution [15], an estimation of the—always nonlinear—statistical parameters and then, a search for the best denoising algorithm for this type of statistics. In contrast, by taking advantage of Stein's MSE estimate, our method goes directly to the last step, without caring for the statistical description: in short, we do not make any explicit hypotheses on the clean image. In fact, we do not consider it as a random process at all; the randomness in our formulation follows from the Gaussian white noise alone.

Our approach consists, thus, in parametrizing the denoising method and choosing the parameters that minimize this MSE estimate. Previous techniques using the SURE required the minimization of complicated expressions for few nonlinear parameters [16], [17] or the use of parallel block iterative convex programming [18]. What makes our approach more tractable and efficient, is precisely the parametrizing method: a *linear combination* of nonlinear denoising functions—thresholding functions. Because of this "linear" choice, the minimization of the MSE estimate merely amounts to solving a linear system of equations, whose size is the number of weights in the linear combination. Obviously, the number of parameters, or degrees of freedom, is not a challenge and highly complicated thresholding behaviors can be obtained this way. In the context of image denoising, a univariate linear parametrization combined with an implicit SURE minimization was already evoked in [19] (*sigmoidal filtering*).

Because of the particular simplicity of Stein's estimate for pointwise denoising functions, we will not exploit the full potential of the theory in this paper and will only consider interscale pointwise thresholding in the orthonormal wavelet transform. This excludes any intrascale considerations. Yet, we will show that our denoising method performs better than the nonredundant versions of the state-of-the-art methods [9], [10], [12] on almost all tested images, to the noteworthy exception of *Barbara*, which may require intrascale processing. Without any optimization attempts in our implementation, the comparison of computation times already show how economical our method is.

The paper is organized as follows. In Section II, we expose the SURE theory for functions of one or several statistically independent variables, and sketch the principles of our parametrization strategy. In Section III, we show how these principles can be exploited to build an efficient pointwise thresholding function that outperforms all known pointwise techniques. In Section IV, we extend the approach to a thresholding function that involves coarser scale parents as well. On this occasion, we develop a new formula to build a parent coefficient out of parent subbands, and, finally, we compare our denoising method to the best available nonredundant techniques (Section V). Both the competitiveness and robustness of our method validate our new approach as an attractive solution for image denoising.

II. THEORETICAL ELEMENTS

A. Problem Setting

Wavelet denoising consists of three main stages (see Fig. 1).

i) Perform a discrete wavelet transform (DWT) to the noisy data $y = (y_n)_{n\in[1,N]}$ which is the sum of the noise-free data $x = (x_n)_{n\in[1,N]}$ and the noise $b = (b_n)_{n\in[1,N]}$.

ii) Denoise J noisy wavelet subimages $y^j = x^j + b^j = (y_n^j)_{n\in[1,N_j]}$, $j \in [1, J]$ by computing J estimates $\hat{x}^j$ of the noise-free highpass subbands x^j.

iii) Reconstruct the denoised image by applying the inverse discrete wavelet transform (IDWT) on the processed highpass wavelet subimages $\hat{x}^j$ to obtain an estimate $\hat{x}$ of the noise-free data x.

One can make two important remarks that set the context in which we will develop our denoising method.

- We will only consider additive Gaussian white noise following a normal law defined by a zero mean and a known[4] σ^2 variance; i.e., $b \sim \mathcal{N}(0, \sigma^2)$.

---

[4] In practice, the noise standard deviation can be accurately estimated using a robust median estimator [1].

- We will only consider *orthonormal* wavelet transform; the consequences are as follows.
  - —The mean-square error (MSE) in the space domain is a weighted sum of the MSE of each individual subband

    $$\underbrace{\left\langle|\hat{x}-x|^2\right\rangle}_{\text{MSE}}=\sum_{j=0}^{J}\frac{N_j}{N}\underbrace{\left\langle|\hat{x}^j-x^j|^2\right\rangle}_{\text{MSE}^j} \tag{1}$$

    where we have introduced the notation

    $$\langle u\rangle=\frac{1}{N}\sum_{n=1}^{N}u_n \tag{2}$$

    for the statistical mean estimate.
  - —The noise remains white and Gaussian with same statistics in the orthonormal wavelet domain, i.e., $b^j\sim\mathcal{N}(0,\sigma^2)$.

This allows us to apply a new denoising function independently in every highpass subband, which means that our solution is subband-adaptive like most of the successful wavelet denoising approaches.

B. Stein's Unbiased MSE Estimate (SURE)

In denoising applications, the performance is often measured in terms of peak signal-to-noise ratio (PSNR), which can be defined as follows:[5]

$$\text{PSNR}=10\log_{10}\left(\frac{\max(x^2)}{\langle|\hat{x}-x|^2\rangle}\right). \tag{3}$$

Since the noise is a random process, we introduce an expectation operator $\mathcal{E}\{\}$ to guess the potential results obtained after processing the noisy data y. Note that the noise-free data x is not modeled as a random process; thus, $\mathcal{E}\{x\}=x$.

The aim of image denoising is naturally to maximize the PSNR and, thus, to minimize the MSE defined in (1). In this paper, we choose to estimate each x^j by a pointwise function of y^j

$$\left(\hat{x}_n^j\right)_{n\in[1,N_j]}=\left(\theta^j\left(y_n^j\right)\right)_{n\in[1,N_j]}.$$

From now on, we will drop the subband index j since a new denoising function is independently applied in each individual subband. Our goal is to find a function θ that minimizes

$$\text{MSE}=\left\langle|\theta(y)-x|^2\right\rangle=\langle\theta(y)^2\rangle-2\langle x\theta(y)\rangle+\langle x^2\rangle. \tag{4}$$

In practice, we only have access to the noisy signal $y=x+b$, and not to the original signal x. In (4), we, thus, need to remove the explicit dependence on x. Note that, since $\langle x^2\rangle$ has no influence in the minimization process, we do not need to estimate it. The remaining problematic term is only $\langle x\theta(y)\rangle$. However, the following theorem, a version of which was proposed by Stein in [7], allows us to overcome this difficulty.

Theorem 1: Let $\theta:\mathbb{R}\to\mathbb{R}$ be a (weakly) differentiable function that does not explode at infinity.[6] Then, the following random variable:

$$\begin{aligned}\epsilon&=\left\langle\theta(y)^2-2y\theta(y)+2\sigma^2\theta'(y)\right\rangle+\langle x^2\rangle\\&=\underbrace{\frac{1}{N}\sum_{n=1}^{N}\left(\theta^2(y_n)-2y_n\theta(y_n)+2\sigma^2\theta'(y_n)\right)}_{\tilde{\epsilon}}+\langle x^2\rangle\end{aligned} \tag{5}$$

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

$$\mathcal{E}\{\epsilon\}=\mathcal{E}\left\{\left\langle|\theta(y)-x|^2\right\rangle\right\}.$$

Proof: We can develop the square error between x_n and its estimate $\theta(y_n)$ as

$$\begin{aligned}\mathcal{E}\left\{|\theta(y_n)-x_n|^2\right\}=&\mathcal{E}\left\{\theta^2(y_n)\right\}-2\mathcal{E}\left\{x_n\theta(y_n)\right\}+\mathcal{E}\left\{x_n^2\right\}\\=&\mathcal{E}\left\{\theta^2(y_n)\right\}-2\mathcal{E}\left\{y_n\theta(y_n)\right\}\\&+2\mathcal{E}\left\{b_n\theta(y_n)\right\}+x_n^2\end{aligned}$$

where each term is well-defined thanks to the hypothesis on θ.

We then use the fact that the Gaussian probability density $q(b_n)$ satisfies $b_nq(b_n)=-\sigma^2q'(b_n)$ to evaluate $\mathcal{E}\{b_n\theta(y_n)\}$

$$\begin{aligned}\mathcal{E}\left\{b_n\theta(y_n)\right\}=&\int\theta(x_n+b_n)b_nq(b_n)db_n\\=&-\sigma^2\int\theta(x_n+b_n)q'(b_n)db_n\\=&\sigma^2\int\theta'(x_n+b_n)q(b_n)db_n\quad\text{(by parts)}\\=&\sigma^2\mathcal{E}\left\{\theta'(y_n)\right\}.\end{aligned} \tag{6}$$

Note that the integrated part $[\sigma^2\theta(x_n+b_n)q(b_n)]_{-\infty}^{+\infty}$ vanishes by hypothesis. This is known as Stein's Lemma [7] and leads to

$$\begin{aligned}\mathcal{E}\left\{|\theta(y_n)-x_n|^2\right\}=\mathcal{E}\left\{\theta^2(y_n)\right\}-2\mathcal{E}\left\{y_n\theta(y_n)\right\}\\+2\sigma^2\mathcal{E}\left\{\theta'(y_n)\right\}+x_n^2.\end{aligned}$$

Since the expectation of a sum is equal to the sum of the expectations, we can deduce that

$$\begin{aligned}\mathcal{E}\left\{\left\langle|\theta(y)-x|^2\right\rangle\right\}=\mathcal{E}\left\{\langle\theta^2(y)\rangle\right\}-2\mathcal{E}\left\{\langle y\theta(y)\rangle\right\}\\+2\sigma^2\mathcal{E}\left\{\langle\theta'(y)\rangle\right\}+\langle x^2\rangle.\end{aligned}$$

■

As said before, there is no need to estimate $\langle x^2\rangle$, since this term will disappear in the minimization. So, in practice, we will consider $\tilde{\epsilon}$ which is the only part of the MSE estimate that depends on the choice of the denoising function θ.

Note that Theorem 1 is still valid if $\theta(y)$ is replaced by a two-variable denoising function $\theta(y,z)$ where y is random, but independent[7] of y. In particular, in an orthonormal wavelet transform—which transforms Gaussian white noise into Gaussian white noise—z can be any wavelet coefficients other than y itself.

---

[5] For 8-bit images, usually $\max(x^2)=255^2$.

[6] Typically, such that $|\theta(z)|\leq\text{Const}\cdot\exp(az^2)$ for $a<(1/2\sigma^2)$.

[7] We recall that the randomness of $y=x+b$ only results from the Gaussian white noise b, because no statistical model is assumed on the noise-free data x.

The result given by Theorem 1 becomes particularly interesting in image denoising applications, where the number of samples is large. Indeed, by the law of large numbers, the standard deviation of ϵ is small; i.e., the estimate ϵ is 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. The next section shows how to use Theorem 1 efficiently.

C. SURE Approach to Image Denoising

Our denoising approach amounts to minimizing ϵ over a range of reasonable denoising functions θ. We claim that this will result in the minimization of the MSE over the same range of functions, up to a small random error inversely proportional to the square root of the number of samples. Before defining more precisely which denoising functions we consider reasonable, we can illustrate the search for the optimal value T by applying Theorem 1 when θ is the well-known *soft-thresholding* function defined by

$$\theta(y) = \mathrm{sign}(y)\,(|y| - T)_+ \tag{7}$$

where $(x)_+ = \max(x, 0)$.

By Theorem 1, the following expression has to be minimized over T

$$\tilde{\epsilon}(T) = \left\langle (2\sigma^2 + T^2 - y^2)\,(|y| - T)_+^0 \right\rangle. \tag{8}$$

The last expression has its minimum exactly for the same T as the following formula:

$$\mathrm{SURE}(T; y) = \sigma^2 - \frac{1}{N} \times \left(2\sigma^2 \cdot \#\{n : |y_n| \leq T\} - \sum_{n=1}^{N} \min(|y_n|, T)^2 \right) \tag{9}$$

which appears in [1].

The estimated optimal threshold value is then: $\tilde{T}_{\mathrm{opt}} = \arg\min_T(\mathrm{SURE}(T; y)) = \arg\min_T(\tilde{\epsilon}(T))$.

We must notice here that the so-called *SureShrink* procedure developed by Donoho and Johnstone in [1] uses, in fact, a hybrid scheme between the SURE theory and the universal threshold (asymptotically optimal when the data exhibit a high level of sparsity). Their minimization of $\mathrm{SURE}(T; y)$ is, thus, restricted to $T \in [0; T_{\mathrm{univ}}]$, where $T_{\mathrm{univ}} = \sqrt{2\sigma^2 \log N}$ is the universal threshold. Our opinion, however, is that this restriction is unnecessary—and often suboptimal—in image denoising applications where quality is measured by a mean-square criterium. This is because, even though natural images have small wavelet coefficients, these are not vanishing as required by the strict sparsity results. It may even be argued that these small coefficients convey important texture information and should not, thus, be set to zero.

As we can verify in Fig. 3, the estimate of the theorem is statistically very reliable and robust, making it completely suitable for an accurate estimation of the optimal threshold.

The *soft-thresholding* function (see Fig. 2) exhibits two main drawbacks. First, it only depends on a single parameter T, and, thus, its shape is not very flexible; second, this dependency is not linear. The consequence of these two remarks is that the sensitivity of the *soft-thresholding* function with respect to the value of T is high, and that finding the optimal threshold requires a nonlinear search algorithm.

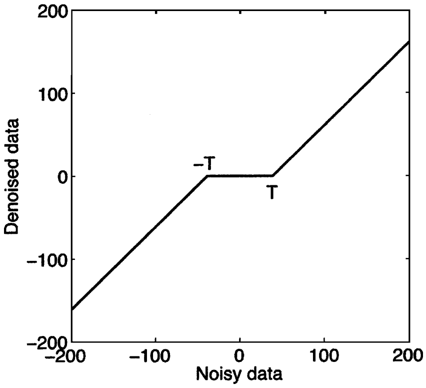

Fig. 2. *Soft-thresholding* function.

In order to mitigate this issue, we choose to build a denoising function that depends *linearly* on a set of parameters—degrees of freedom—which we will determine exactly by minimizing ϵ. The exact minimization is especially simple (linear) because the MSE estimate ϵ has a quadratic form, much like the true MSE. The key idea is, thus, to build a linearly parameterized denoising function of the form

$$\theta(y) = \sum_{k=1}^{K} a_k \varphi_k(y) \tag{10}$$

where K is the number of parameters.

If we introduce (10) into the estimate of the MSE given in Theorem 1 and perform differentiations over the a_k, we obtain for all $k \in [1; K]$

$$0 = \frac{1}{2}\frac{\partial \epsilon}{\partial a_k} = \left\langle \theta(y)\varphi_k(y) - y\varphi_k(y) + \sigma^2 \varphi_k'(y) \right\rangle$$

$$\Updownarrow$$

$$\sum_{l=1}^{K} \underbrace{\left\langle \varphi_k(y)\varphi_l(y) \right\rangle}_{M_{k,l}} a_l - \underbrace{\left\langle y\varphi_k(y) - \sigma^2 \varphi_k'(y) \right\rangle}_{c_k} = 0.$$

These equations can be summarized in matrix form as $\mathbf{Ma} = \mathbf{c}$, where $\mathbf{a} = [a_1 \ldots a_K]^T$ and $\mathbf{c} = [c_1 \ldots c_K]^T$ are vectors of size $K \times 1$, and $\mathbf{M} = [M_{k,l}]_{1 \leq k, l \leq K}$ is a matrix of size $K \times K$. This linear system is solved for $\mathbf{a}$ by

$$\mathbf{a} = \mathbf{M}^{-1}\mathbf{c} \tag{11}$$

which makes our approach very simple to implement. Note that, since we are only interested in the minimum of ϵ, we are ensured that there will always be a solution. When several solutions are admissible (e.g., when $\mathrm{rank}(\mathbf{M}) < K$) any one of them will be acceptable—in particular, the one provided by the pseudoinverse of $\mathbf{M}$. When this degeneracy occurs, we will conclude that the parameters a_k belong to some linear subspace and, thus, that some of them are useless (the function is "over-parameterized"). Of course, it is desirable to keep the number of degrees of freedom K as low as possible in order for the estimate ϵ to keep a small variance.

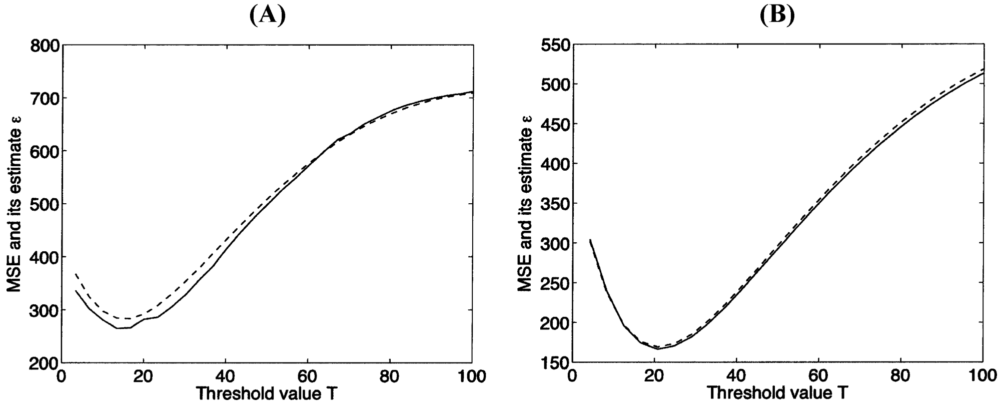

Fig. 3. Statistical accuracy of Theorem 1 illustrated with the *soft-threshold*: true MSE is in dashed lines, while its estimate ϵ is in solid line. (a) $N = 32 \times 32$ samples and $\sigma = 20$. (b) $N = 256 \times 256$ samples and $\sigma = 20$. The variance of the estimator decreases when the number of samples N increases, making Theorem 1 statistically reliable for image denoising applications.

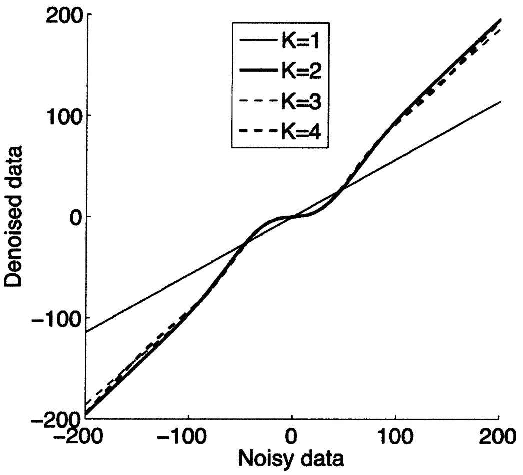

Fig. 4. Shape of our denoising function (12) in a particular subband, for various K and optimized a_k's and T.

III. EFFICIENT SURE-BASED POINTWISE THRESHOLDING

In the previous section, we have proposed a general form of denoising functions (10). The difficulty is now to choose suitable basis functions φ_k that will determine the shape of our denoising function. Therefore, we want the denoising function θ to satisfy the following properties:

- differentiability: required to apply Theorem 1;
- anti-symmetry: the wavelet coefficients are not expected to exhibit a sign preference;
- linear behavior for large coefficients: because $\theta(y)$ should asymptotically tend to y.

After trying several types of φ_k, we have found that all of them give quite similar results, when the above conditions are satisfied. We have, thus, decided to retain the following pointwise denoising function:

$$\theta(y) = \sum_{k=1}^{K} a_k y e^{-(k-1)\frac{y^2}{2T^2}}. \tag{12}$$

We choose derivatives of Gaussians (DOG) because they decay quite fast, which ensures a linear behavior close to the identity for large coefficients (see Fig. 4).

In addition to the linear coefficients, our denoising function contains two nonlinear dependencies: the number of terms K and the parameter T. We will see later that they can be fixed independently of the image.

If we consider only one parameter ($K = 1$), our denoising function simply becomes $\theta(y) = a_1 y$, which is the simplest linear pointwise denoising function. The direct minimization of the estimate ϵ provides

$$a_1 = 1 - \frac{\sigma^2}{\langle y^2 \rangle} \tag{13}$$

which is known as the James–Stein estimator [20].

Practical tests (with optimization over the parameter T, independently in each subband) on various images and with various noise levels have shown that, as soon as $K \geq 2$, the results become quite similar. It, thus, appears that it is sufficient to keep as few as $K = 2$ terms in (12). This is confirmed in Fig. 4, which shows that the shape of our denoising function is nearly insensitive to the variation of $K \geq 2$.

Moreover, the optimal value of the parameter T is closely linked to the standard deviation σ of the noise and in a lesser way to the number of parameters K. Its interpretation is quite similar as in the case of the *soft-threshold*: It manages the transition between low SNR to high SNR coefficients. In our case though, the variations of the minimal ϵ (over a_k) when T changes are quite small (see Fig. 5), because our denoising function is much more flexible than the *soft-threshold*. This sensitivity becomes even smaller as the number of parameters K increases. In fact, this indicates that some parameters are, in that case, useless.

To summarize, we have shown that both the number of terms K and the parameter T have only a minor influence on the quality of the denoising process. This indicates that these two parameters do not have to be optimized; instead, they can be fixed once for all, independently of the type of image. From a practical point of view, we suggest to use $K = 2$ terms and $T = \sqrt{6}\sigma$ (see Fig. 5), leading to the following pointwise thresholding function:

$$\theta_0(y; \mathbf{a}) = \left(a_1 + a_2 e^{-\frac{y^2}{12\sigma^2}}\right) y. \tag{14}$$

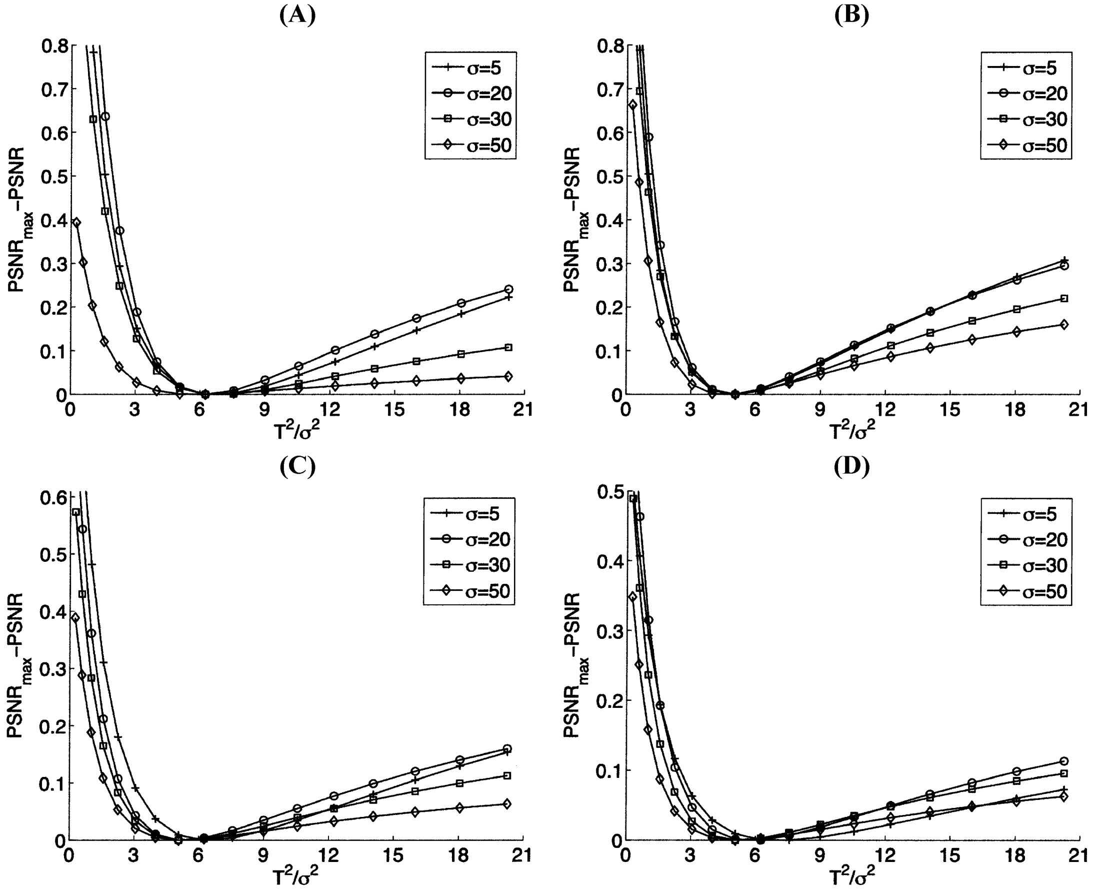

Fig. 5. Sensitivity of our denoising function (14) with respect to variations of T. (a) *Peppers* 256 × 256. (b) *MIT* 256 × 256. (c) *Lena* 512 × 512. (d) *Boat* 512 × 512. We can notice that for all images and for the whole range of input PSNR the maximum of the PSNR is reached for $(T^2/\sigma^2) \simeq 6$.

TABLE I

COMPARISON OF OUR SUM OF DOG (14) WITH THE ORACLE *SOFT-THRESHOLD* (NONREDUNDANT *SYM8*, FOUR ITERATIONS)

σ	5	10	20	30	50	5	10	20	30	50
Method	**Boat** 512 × 512					**Goldhill** 512 × 512				
OracleShrink	36.09	32.11	28.64	26.81	24.79	35.99	31.97	28.75	27.18	25.45
Sum of DOG (*Oracle*)	**36.35**	**32.37**	**28.85**	**27.03**	**25.01**	**36.21**	**32.25**	**28.99**	**27.42**	**25.67**
Sum of DOG (SURE)	**36.35**	**32.37**	**28.85**	27.02	25.00	**36.21**	**32.25**	**28.99**	27.41	25.66
Method	**Peppers** 256 × 256					**Bridge** 256 × 256				
OracleShrink	36.38	32.06	28.03	25.84	23.34	34.83	29.81	25.77	23.93	22.06
Sum of DOG (*Oracle*)	**36.67**	**32.36**	**28.28**	**25.97**	**23.47**	**34.89**	**30.00**	**26.10**	**24.29**	**22.40**
Sum of DOG (SURE)	**36.67**	32.35	28.67	25.95	23.45	**34.89**	**30.00**	26.09	24.28	22.39

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

Now, it is interesting to evaluate the efficiency of our denoising function (14) and the accuracy of our minimization process based on an estimate ϵ of the MSE. We propose to compare our results with the best results that can be reached by the popular *soft-threshold* with an optimal threshold choice (*OracleShrink*). Two main observations naturally come out of Table I.

i) SURE is a reliable estimate of the MSE, since the resulting average loss in PSNR is within 0.02 dB for all images.

ii) Our sum of DOG (14) gives better PSNRs than the optimal *soft-threshold*.

IV. EFFICIENT SURE-BASED INTERSCALE THRESHOLDING

The integration of interscale information has been shown to improve the denoising quality, both visually and in terms of PSNR [9], [11], [21]. However, the gain brought is often modest, especially considering the additional complications involved by this processing [9]. In this section, we reformulate the problem by first building a loose prediction y_p of wavelet coefficients y out of a suitably filtered version of the lowpass subband at the same scale, and then by including this predictor in an explicit pointwise denoising function. Apart from the

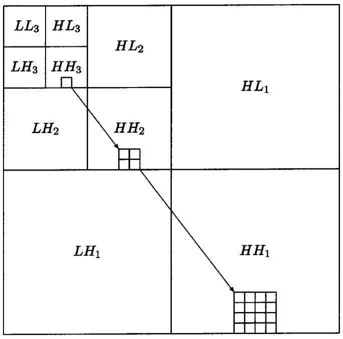

Fig. 6. Three stages of a fully decimated orthogonal wavelet transform and the so-called parent–child relationship.

specific denoising problem addressed in this paper, we believe more generally that other applications (e.g., compression, detection, segmentation) could benefit as well from the theory that leads to this predictor.

A. Building the Interscale Prediction

The wavelet coefficients that lie on the same dyadic tree (see Fig. 6) are well known to be large together in the neighborhood of image discontinuities. What can, thus, be predicted with reasonably good accuracy are the position of large wavelet coefficients out of parents at lower resolutions. However, getting the actual values of the finer resolution scale coefficients seem somewhat out of reach. This suggests that the best we can get out of between-scale correlations is a segmentation between regions of large and small coefficients. This comes back to the idea of signal of interest proposed by Pižurica *et al.* in [10].

In a critically sampled orthonormal wavelet decomposition, the parent subband is half the size of the child subband. The usual way of putting the two subbands in correspondence is simply to expand the parent by a factor two. Unfortunately, this approach does not take into account the potential—noninteger— shift caused by the filters of the DWT. We, thus, propose a more sophisticated solution, which addresses this issue and ensures the alignment of image features between the child and its parent.

Our idea comes from the following observation: Let LH_j and LL_j be, respectively, bandpass and lowpass outputs at iteration j of the filterbank. Then, if the *group delay*[8] between the bandpass and the lowpass filters are *equal*, no shift between the features of LH_j and LL_j will occur. Of course, depending on the amplitude response of the filters, some features may be attenuated, blurred, or enhanced, but their location will remain unchanged. When the group delays differ—which is the general case—we, thus, propose to filter the lowpass subband LL_j in order to *compensate for the group delay difference* with LH_j. This operation is depicted in Fig. 7(a): LL_j is filtered in the three bandpass "directions" by adequately designed filters W_{HL}, W_{HH}, and W_{LH}, providing aligned—i.e., group delay compensated—subbands with HL_j, HH_j, and LH_j.

Because the filters considered in this paper are separable, we only have to consider 1-D group delay compensation (GDC).

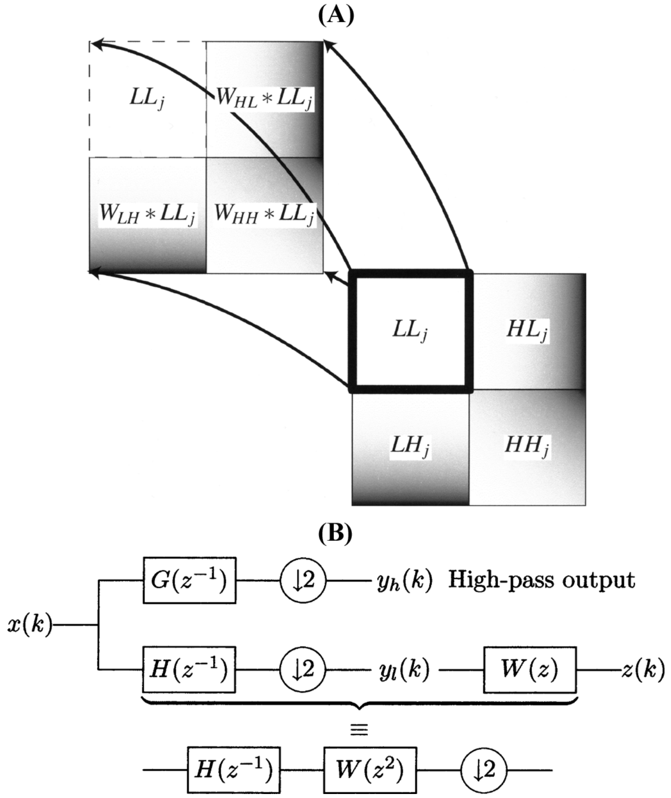

Fig. 7. One way of obtaining the whole parent information out of the lowpass subband: (a) 2-D illustration; (b) 1-D filterbank illustration.

Definition 1: We say that two filters $H(z)$ and $G(z)$ are group delay compensated if and only if the group delay of the quotient filter $H(z)/G(z)$ is zero identically; i.e., if and only if there exists a (anti-) symmetric filter $R(z) = \pm R(z^{-1})$ such that $H(z) = G(z)R(z)$.

The following result shows how to choose a GDC filter in a standard orthonormal filterbank.

Theorem 2: For the output of the dyadic orthonormal filterbank of Fig. 7(b) to be group delay compensated, it is necessary and sufficient that

$$W(z^2) = G(z^{-1})G(-z^{-1})(1 + \epsilon z^{-2})R(z^2) \qquad (15)$$

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

Proof: Group delay compensation between the two filterbank branches is equivalent to [see Fig. 7(b)]

$$H(z^{-1})W(z^2) = G(z^{-1})R_1(z) \qquad (16)$$

where $R_1(z) = \epsilon R_1(z^{-1})$ is an arbitrary symmetric ($\epsilon = 1$) or anti-symmetric ($\epsilon = -1$) filter.

Because the filters H and G are orthonormal, we have $H(z^{-1}) = zG(-z)$, and, thus, (16) can be rearranged as

$$W(z^2) = \frac{G(z^{-1})R_1(z)}{zG(-z)} = G(z^{-1})G(-z^{-1})\underbrace{\frac{z^{-1}R_1(z)}{G(-z)G(-z^{-1})}}_{R_2(z)}. \qquad (17)$$

---

[8] For example, the frequency gradient of the phase response, with a minus sign.

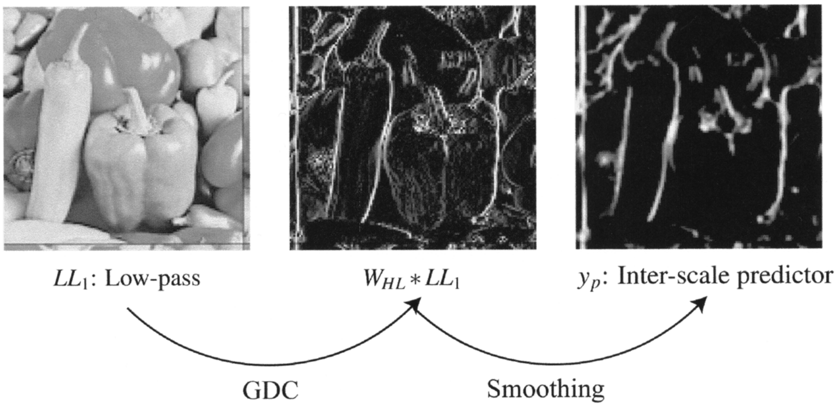

Fig. 8. Building an efficient interscale predictor, illustrated with a particular subband (HL_1) of the noise-free *Peppers* image.

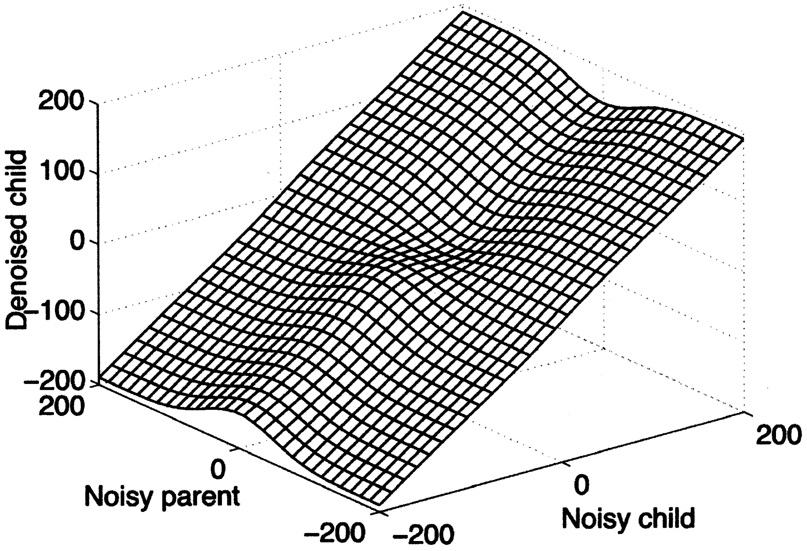

Fig. 9. Three-dimensional surface plot of a possible realization of our interscale thresholding function (21).

We observe that $R_2(z)$ is an even polynomial because both $G(z^{-1})G(-z^{-1})$ and $W(z^2)$ are. If we denote $R_2(z) = (1 + \epsilon z^{-2})R(z^2)$, then the symmetry of $R_1(z)$ implies that

$$\begin{aligned} R(z^{-2}) &= \frac{zR_1(z^{-1})}{(1+\epsilon z^2)G(-z)G(-z^{-1})} \\ &= \frac{\epsilon z R_1(z)}{(1+\epsilon z^2)G(-z)G(-z^{-1})} \\ &= \frac{z^{-1}R_1(z)}{(1+\epsilon z^{-2})G(-z)G(-z^{-1})} \\ &= R(z^2) \end{aligned}$$

i.e., $R(z)$ is and arbitrary zero-phase filter.

After substitution in (17), this finally leads us to the formulation (15), as an equivalent characterization of the group delay compensation in the filterbank of Fig. 7(b). ■

In addition to (15), the GDC filter $W(z)$ has to satisfy a few constraints:

- *energy preservation*, i.e., $\sum_{n\in\mathbb{Z}} w_n^2 = 1$, in order for the amplitude of the two outputs to be comparable;
- *highpass behavior*, in order for the filtered lowpass image to "look like" the bandpass target;
- *shortest possible response*, in order to minimize the enlargement of image features.

We can give a simple GDC filter in the case of symmetric filters. The shortest highpass $W(z)$ satisfying the GDC condition is in fact the simple gradient filter: $W(z) = z - 1$. If the symmetry is not centered at the origin but at a position n_0, then $W(z) = z^{-n_0}(z-1)$. This type of solution is still adequate for near-symmetric filters such as the Daubechies *symlets* [22]. When the lowpass filter is not symmetric, we can simply take $R(z^2) = 1$ in (15).

Finally, in order to increase the homogeneity inside regions of similar magnitude coefficients, we apply a 2-D smoothing filter—a normalized Gaussian kernel $G(x) = (1/\sqrt{2\pi})e^{-(x^2/2)}$—onto the absolute value of the GDC output. In the rest of the paper, we will refer to the so-built interscale predictor by y_p.

B. Integrating the Interscale Predictor

Now that we have built the interscale predictor y_p, we have to suitably integrate it into our pointwise denoising function. As mentioned before, this interscale predictor does not tell us much about the actual value of its corresponding child wavelet

TABLE II
DENOISING PERFORMANCE IMPROVEMENT BROUGHT BY OUR INTERSCALE STRATEGY (NONREDUNDANT *SYM8*, FOUR ITERATIONS)

σ	5	10	20	30	50	100	5	10	20	30	50	100
Method	**Peppers** 256 × 256						**House** 256 × 256					
Expansion by 2	36.76	32.49	28.46	26.21	23.62	20.92	37.50	33.59	30.03	28.07	25.78	22.92
Proposed	**37.17**	**33.18**	**29.33**	**27.13**	**24.43**	**21.32**	**37.88**	**34.29**	**30.93**	**28.98**	**26.58**	**23.51**

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

TABLE III
COMPARISON OF SOME OF THE MOST EFFICIENT DENOISING METHODS (NONREDUNDANT *SYM8*, FOUR ITERATIONS)

σ	5	10	15	20	25	30	50	100	5	10	15	20	25	30	50	100
Input PSNR	34.15	28.13	24.61	22.11	20.17	18.59	14.15	8.13	34.15	28.13	24.61	22.11	20.17	18.59	14.15	8.13
Method	**Peppers** 256 × 256								**House** 256 × 256							
BayesShrink	35.83	31.49	29.30	27.85	26.72	25.73	23.17	20.73	36.91	32.92	30.81	29.42	28.44	27.66	25.49	22.87
BiShrink 7 × 7	36.61	32.55	30.25	28.66	27.47	26.51	23.89	20.80	37.54	33.60	31.56	30.16	29.07	28.20	25.83	22.84
ProbShrink 3 × 3	36.72	32.68	30.41	28.85	27.67	26.70	23.85	20.85	37.59	33.84	31.74	30.29	29.20	28.35	25.99	23.17
BLS-GSM 3 × 3	36.80	32.86	30.62	29.07	27.90	26.97	24.40	20.88	**38.01**	34.26	32.23	30.79	29.65	28.72	26.15	22.97
Our bivariate (21)	**37.17**	**33.18**	**30.91**	**29.33**	**28.12**	**27.13**	**24.43**	**21.32**	37.88	**34.29**	**32.32**	**30.93**	**29.86**	**28.98**	**26.58**	**23.51**
Best redundant	*37.09*	*33.33*	*31.26*	*29.84*	*28.74*	*27.84*	*25.30*	*21.98*	*38.43*	*35.04*	*33.23*	*31.91*	*30.87*	*30.01*	*27.62*	*24.53*
Method	**Al** 512 × 512								**Bridge** 256 × 256							
BayesShrink	37.77	34.17	32.10	30.67	29.63	28.84	26.67	23.84	34.81	29.80	27.30	25.75	24.69	23.90	22.04	19.99
BiShrink 7 × 7	38.01	34.50	32.57	31.23	30.21	29.39	27.09	24.01	34.94	29.93	27.38	25.81	24.75	23.97	22.11	19.97
ProbShrink 3 × 3	38.11	34.58	32.64	31.28	30.08	29.32	27.18	24.24	34.59	29.61	27.20	25.74	24.73	23.97	22.10	20.08
BLS-GSM 3 × 3	38.38	34.83	32.93	31.58	30.53	29.68	27.35	24.20	34.98	29.98	27.50	26.02	25.01	24.25	22.34	20.00
Our bivariate (21)	**38.43**	**34.90**	**32.97**	**31.64**	**30.64**	**29.84**	**27.61**	**24.56**	**35.06**	**30.22**	**27.84**	**26.36**	**25.33**	**24.56**	**22.60**	**20.35**
Best redundant	*38.90*	*35.46*	*33.66*	*32.42*	*31.46*	*30.67*	*28.46*	*25.51*	*35.23*	*30.46*	*28.07*	*26.60*	*25.58*	*24.83*	*22.98*	*20.78*
Method	**Barbara** 512 × 512								**Boat** 512 × 512							
BayesShrink	35.78	31.25	28.86	27.32	26.22	25.34	23.14	21.36	35.99	31.98	29.94	28.55	27.52	26.71	24.74	22.44
BiShrink 7 × 7	36.76	32.52	30.14	28.51	27.29	26.33	23.91	21.47	36.18	32.46	30.47	29.08	28.03	27.20	25.05	22.52
ProbShrink 3 × 3	36.75	32.48	30.04	28.40	27.20	26.27	23.86	21.58	36.20	32.53	30.50	29.11	28.05	27.22	25.12	22.69
BLS-GSM 3 × 3	**37.05**	**32.89**	**30.54**	**28.93**	**27.72**	**26.76**	**24.25**	21.53	36.46	32.89	**30.89**	**29.49**	28.43	27.58	25.34	22.64
Our bivariate (21)	36.71	32.18	29.66	27.98	26.76	25.83	23.70	**21.76**	**36.70**	**32.90**	30.85	29.47	**28.44**	**27.63**	**25.50**	**22.97**
Best redundant	*37.69*	*33.90*	*31.71*	*30.16*	*28.96*	*27.99*	*25.32*	*22.47*	*36.94*	*33.53*	*31.64*	*30.32*	*29.30*	*28.48*	*26.28*	*23.65*
Method	**Crowd** 512 × 512								**Goldhill** 512 × 512							
BayesShrink	34.60	29.31	26.53	24.73	23.45	22.47	20.07	17.46	35.93	31.94	29.96	28.69	27.79	27.13	25.41	23.32
BiShrink 7 × 7	34.71	29.48	26.70	24.88	23.57	22.57	20.13	17.40	36.17	32.27	30.32	29.07	28.15	27.44	25.57	23.26
ProbShrink 3 × 3	34.42	29.29	26.59	24.83	23.56	22.58	20.15	17.43	36.07	32.30	30.35	29.07	28.13	27.43	25.62	23.47
BLS-GSM 3 × 3	34.79	29.63	26.91	25.12	23.84	22.85	20.39	17.51	36.37	32.61	30.68	29.41	28.47	27.73	25.73	23.30
Our bivariate (21)	**34.86**	**29.77**	**27.11**	**25.38**	**24.13**	**23.17**	**20.75**	**17.97**	**36.53**	**32.69**	**30.76**	**29.52**	**28.60**	**27.89**	**26.06**	**23.82**
Best redundant	*34.96*	*30.05*	*27.49*	*25.83*	*24.63*	*23.69*	*21.24*	*18.33*	*36.88*	*33.24*	*31.37*	*30.13*	*29.22*	*28.50*	*26.60*	*24.30*

Note: output PSNRs have been averaged over ten noise realizations. The best redundant results are obtained using the *BLS-GSM* 3 × 3 with an 8-orientations full steerable pyramid; results slightly differ from the ones published in [9], because no boundary extension has been applied here.

coefficients. It only gives an indication on its expected magnitude. Here, we, thus, propose to use the parent y_p as a discriminator between high SNR wavelet coefficients and low SNR wavelet coefficients, leading to the following general pointwise denoising function:

$$\theta(y, y_p) = f(y_p) \sum_{k=1}^{K} a_k \varphi_k(y) + (1 - f(y_p)) \sum_{k=1}^{K} b_k \varphi_k(y). \tag{18}$$

The linear parameters a_k and b_k are then solved for by minimizing the MSE estimate ϵ defined in Theorem 1, for the linear parameters a_k and b_k. The optimal coefficients are obtained in the same way as in Section II-C and involve a solution similar to (11).

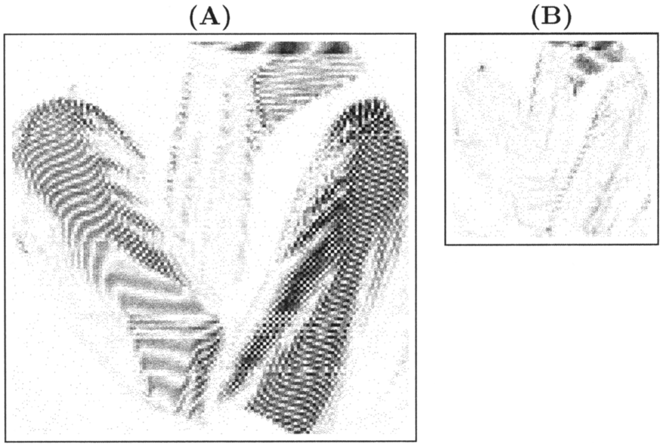

Fig. 10. (a) Zoom at Barbara's trousers at the finest scale of an orthonormal wavelet transform: the stripes are clearly visible. (b) Zoom at Barbara's trousers at the next coarser scale: the stripes are not visible anymore.

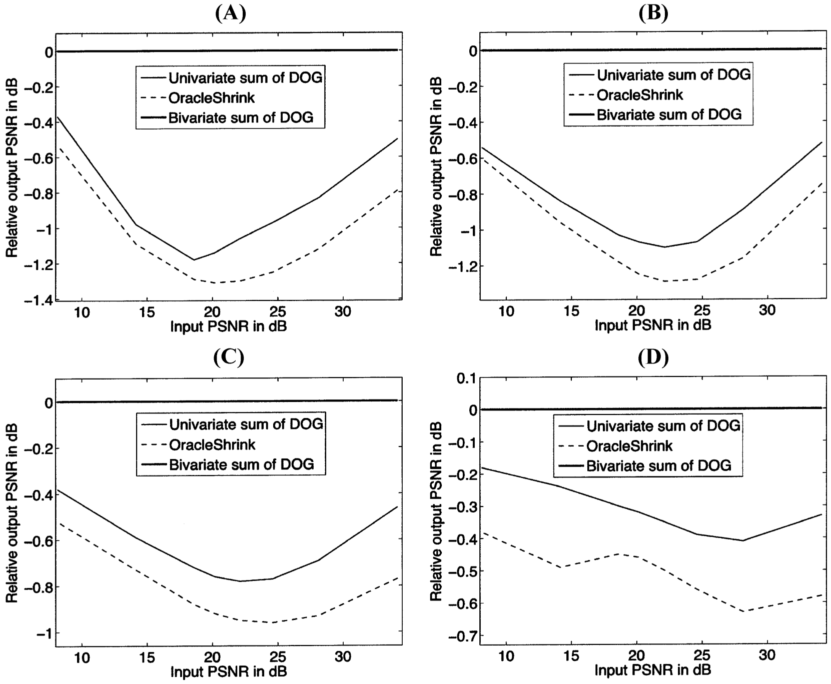

Fig. 11. Comparison of our interscale dependent thresholding function (21) with the best possible soft-threshold *OracleShrink* and with our simple univariate denoising function (14). (a) *Peppers* 256 × 256. (b) *House* 256 × 256. (c) *Lena* 512 × 512. (d) *Barbara* 512 × 512.

A first thought choice for the function f in (18) is simply the Heaviside function

$$H(y_p) = \begin{cases} 1, & \text{if } |y_p| \geq T \\ 0, & \text{if } |y_p| < T \end{cases} \tag{19}$$

where T can be interpreted as a decision factor. However, since the classification will not be perfect (i.e., some small parent coefficients may correspond to high-magnitude child coefficients, and *vice versa*), it is more appropriate to use a smoother decision function. Instead, we, thus, propose to use

$$f(y_p) = e^{-\frac{y_p^2}{2T^2}}. \tag{20}$$

As in the univariate case (Section III), we suggest to use a sum of DOG with $K = 2$ terms for each class of wavelet coefficients and[9] $T = \sqrt{6}\sigma$, leading to the following bivariate denoising function:

$$\begin{aligned} \theta(y, y_p; \mathbf{a}, \mathbf{b}) &= e^{-\frac{y_p^2}{12\sigma^2}} \theta_0(y; \mathbf{a}) + \left(1 - e^{-\frac{y_p^2}{12\sigma^2}}\right) \theta_0(y; \mathbf{b}) \\ &= e^{-\frac{y_p^2}{12\sigma^2}} \left(a_1 + a_2 e^{-\frac{y^2}{12\sigma^2}}\right) y \\ &\quad + \left(1 - e^{-\frac{y_p^2}{12\sigma^2}}\right) \left(b_1 + b_2 e^{-\frac{y^2}{12\sigma^2}}\right) y. \end{aligned} \tag{21}$$

Table II quantifies the improvement introduced by this new way of integrating the interscale information, as compared to the usual expansion of the parent subband.

V. Experimental Results

In this section, we compare our interscale dependent thresholding function (21) with some of the best state-of-the-art techniques: Sendur's *et al.* bivariate MAP estimator with local variance estimation, Portilla's *BLS-GSM* and Pižurica's *ProbShrink.*

In all comparisons, we use a critically sampled orthonormal wavelet basis with eight vanishing moments (*sym8*) over four decomposition stages.

A. PSNR Comparisons

We have tested the various denoising methods for a representative set of standard 8-bit grayscale images such as *Al*, *Barbara*, *Boat*, *Crowd*, *Goldhill* (size 512 × 512) and *Peppers*, *House*, *Bridge* (size 256 × 256), corrupted by simulated additive Gaussian white noise at eight different power levels $\sigma \in [5, 10, 15, 20, 25, 30, 50, 100]$, which corresponds to PSNR decibel values [34.15, 28.13, 24.61, 22.11, 20.17, 18.59, 14.15, 8.13]. The denoising process has been performed over ten different noise realizations for each standard deviation and the resulting PSNRs averaged over these ten runs. The parameters of each method have been set according to the values given by their respective authors in the corresponding referred papers. Variations in output PSNRs are, thus, only due to the denoising techniques themselves. This reliable comparison was only possible thanks to the kindness of the various authors who have provided their respective Matlab codes on their personal websites.

Table III summarizes the results obtained. To the noteworthy exception of *Barbara*, our results are already competitive with the best techniques available that consider nonredundant orthonormal transforms. We stress again that our processing consists of a simple pointwise threshold, driven by interscale infor-

---

[9] Side investigations have shown that the T needed in (20) and the one optimized in Section III can be chosen identical for optimal performances and equal to $\sqrt{6}\sigma$.

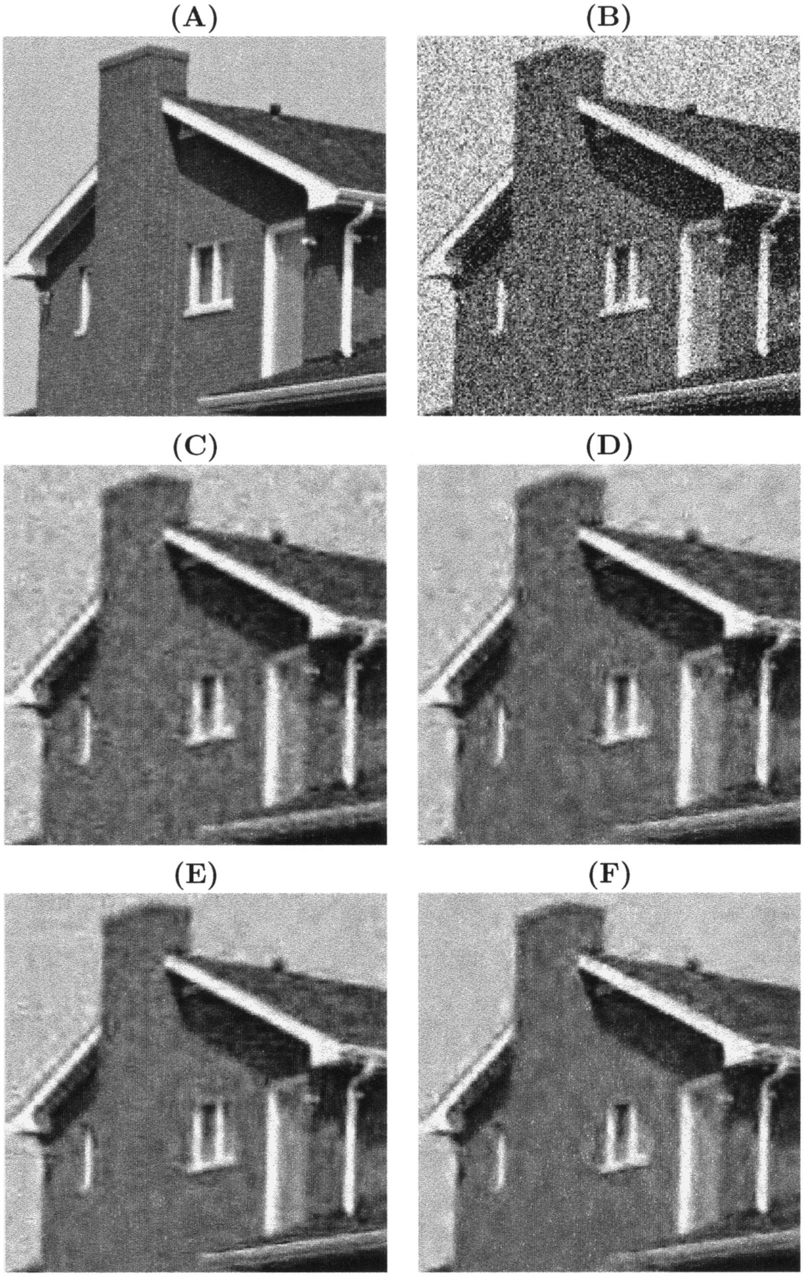

Fig. 12. (a) Part of the noise-free 256 × 256 *House* image. (b) Noisy version of it: PSNR $= 18.59$ dB. (c) Denoised result using the *BayesShrink*: PSNR $= 27.57$ dB. (d) Denoised result using the *BiShrink* 7 × 7: PSNR $= 28.19$ dB. (e) Denoised result using the *BLS-GSM* 3 × 3: PSNR $= 28.73$ dB. (f) Denoised result using our interscale dependent thresholding function (21): PSNR $= 28.96$ dB.

mation; i.e., without taking intrascale dependencies into consideration, contrary to the best performing methods (*ProbShrink*, *BiShrink* and *BLS-GSM*).

When looking closer at the results, we observe the following.

- Our method outperforms the classical *BayesShrink* by more than +1 dB on average.
- Our method gives better results than Sendur's *BiShrink* 7× 7 which integrates both the inter- and the intrascale dependencies (average gain of +0.6 dB).
- Our method gives better results than Pižurica's *ProbShrink* 3 × 3 which integrates the intrascale dependencies (average gain of +0.4 dB).
- We obtain similar or sometimes even better results than Portilla's *BLS-GSM* 3 × 3 for most of the images.
- For the *Barbara* image, our method is among the worst performers together with the pointwise *BayesShrink*. Our explanation for this is that some local information (especially the texture in Barbara's trousers) is completely lost at coarser scales (see Fig. 10). Interscale correlations may be too weak for this image, which indicates that an efficient denoising process may require intrascale information as well.
- The gap between our nonredundant SURE-based approach and the best up-to-date redundant results lies in the range of 0.5–1 dB for most images.

It is instructive to compare the results (see Fig. 11) obtained with our interscale dependent thresholding function (21), with the ones obtained with our simple univariate denoising function

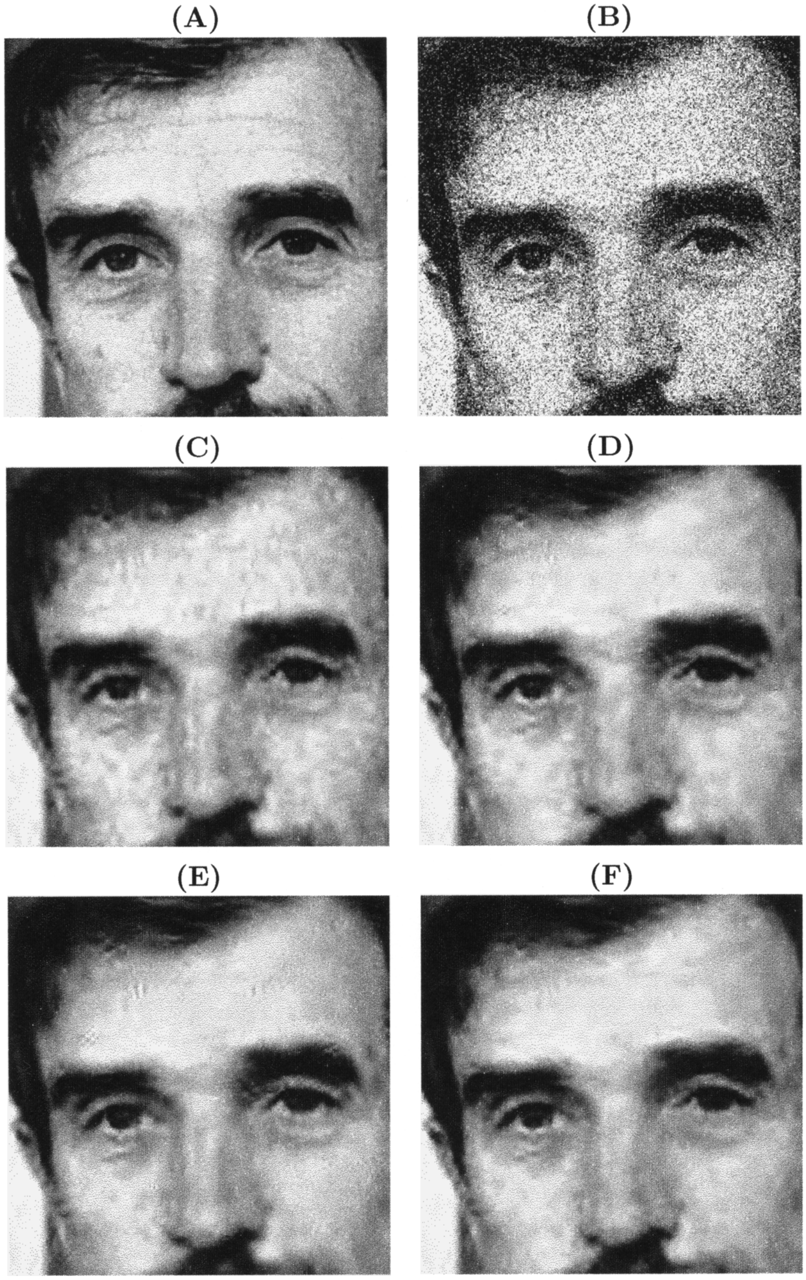

Fig. 13. (a) Part of the noise-free 512×512 *Al* image. (b) Noisy version of it: PSNR $= 14.15$ dB. (c) Denoised result using the *BayesShrink*: PSNR $= 26.71$ dB. (d) Denoised result using the *BiShrink* 7×7: PSNR $= 27.12$ dB. (e) Denoised result using the *BLS-GSM* 3×3: PSNR $= 27.34$ dB. (f) Denoised result using our interscale dependent thresholding function (21): PSNR $= 27.66$ dB.

(14). The improvement (often more than $+1$ dB) is quite significant for most standard images (see Fig. 11). Yet, for images that have substantial high-frequency contents, the integration of interscale dependencies does not lead to such an impressive gain. On the same graphs, we have also included the results obtained with the *OracleShrink*, showing a systematic underperformance with regards to even our simple univariate denoising function.

B. Visual Quality

Although there is no consensual objective way to judge the visual quality of a denoised image, two important criteria are widely used: the visibility of processing artifacts and the conservation of image edges. Processing artifacts usually result from a modification of the spatial correlation between wavelet coefficients (often caused by the zeroing of small neighboring coefficients) and are likely to be reduced by taking into account intrascale dependencies. Instead, image edge distortions usually arise from modifications of the interscale coefficient correlations. The amplitude of these modifications is likely to be reduced by a careful consideration of interscale dependencies in the denoising function.

Since our algorithm only includes interscale considerations, we expect it to be specifically robust to noise with regards to

TABLE IV
RELATIVE COMPUTATION TIME OF VARIOUS DENOISING TECHNIQUES

Method	Unit of time [U]	
	256×256 **images**	512×512 **images**
BayesShrink	1.0	3.9
BiShrink 7×7	1.4	5.4
ProbShrink 3×3	2.8	6.6
BLS-GSM 3×3	7.8	30.0
Univariate sum of DOG (14)	1.2	4.5
Bivariate sum of DOG (21)	2.0	7.9
Redundant *BLS-GSM* 3×3	*81.5*	*311.8*

Note: The computation times have been averagde over twenty runs.

edge preservation. Additionally, we would like to stress that our method exhibits the fewest number of artifacts, which we attribute to the fact that we are never forcing any wavelet coefficients to zero. These observations are illustrated in Figs. 12 and 13.

C. Computation Time

It is also interesting to evaluate the various denoising methods from a practical point of view: the computation time. Indeed, the results achieved by overcomplete representation are admittedly superior than the ones obtained by critically sampled wavelet transforms, but their weakness is the time they require (nearly 27 s on a Power Mac G5 workstation with 1.8-GHz PowerPC 970 CPU for 256×256 images to obtain the redundant results reported in Table III). With our simple univariate method (14), the whole denoising process (including four iterations of an orthonormal wavelet transform) lasts approximately 0.4 s for 256×256 images (1.6 s for 512×512 images), using a similar workstation. With our interscale dependent thresholding function (21), the whole denoising task takes between 0.6–0.7 s for 256×256 images and about 2.7 s for 512×512 images. To compare with, Portilla's BLS-GSM with a 3×3 window size lasts approximately 10 s for 512×512 images, using the same orthonormal transform. Besides giving competitive results, our method is, thus, also much faster.

Table IV summarizes the relative computation time of the various methods considered in this paper. Note that the main part of the *ProbShrink* is contained in a precompiled file, making its execution time a bit faster than the other algorithms which are fully implemented in Matlab.

VI. CONCLUSION

We have presented a new approach to orthonormal wavelet image denoising that does not need any prior statistical modelization of the wavelet coefficients. This approach is made possible thanks to the existence of an efficient estimate of the MSE between noisy and clean image—the SURE—that is based on the noisy data alone. Its minimization over a set of denoising processes automatically provides a near-optimal solution in the sense of the *a posteriori* MSE. For efficiency reasons, we have chosen this set to be a linear span of basic nonlinear mappings. Using this approach, we have designed an image denoising algorithm that takes into account interscale dependencies, but discards intrascale correlations. In order to compensate for features misalignment, we have developed a rigorous procedure based on the relative group delay between the scaling and wavelet filters—*group delay compensation*. The information brought by this new interscale predictor is used to classify smoothly between high- and low-SNR wavelet coefficients.

The comparison of the denoising results obtained with our algorithm, and with the best state-of-the-art nonredundant techniques (that integrate both inter- and intrascale dependencies), demonstrate the efficiency of our SURE-based approach which gave the best output PSNRs for most of the images. The visual quality of our denoised images is moreover characterized by fewer artifacts than the other methods.

We are currently working on an efficient integration of the intrascale correlations within the SURE-based approach. Our goal is to show that the consideration of inter- and intrascale dependencies brings denoising gains that rival the quality of the best redundant techniques such as *BLS-GSM*.

REFERENCES

[1] 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.

[2] ——, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

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

[4] N. G. Kingsbury, "Image processing with complex wavelets," *Phil. Trans. Roy. Soc. A.*, Sep. 1999.

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

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

[7] C. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 9, pp. 1135–1151, 1981.

[8] 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.

[9] 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.

[10] A. Pižurica 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. 645–665, Mar. 2006.

[11] 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.

[12] ——, "Bivariate shrinkage with local variance estimation," *IEEE Signal Process. Lett.*, vol. 9, no. 12, pp. 438–441, Dec. 2002.

[13] N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *J. Appl. Comput. Harmon. Anal.*, vol. 10, no. 3, pp. 234–253, May 2001.

[14] J.-L. Starck, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 670–684, Jun. 2002.

[15] M. S. Crouse, R. D. Nowak, and R. G. Baraniuki, "Wavelet-based signal processing using Hidden Markov models," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 886–902, Apr. 1998.

[16] X.-P. Zhang and M. D. Desai, "Adaptive denoising based on SURE risk," *IEEE Signal Process. Lett.*, vol. 5, no. 10, pp. 265–267, Oct. 1998.

[17] A. Benazza-Benyahia and J.-C. Pesquet, "Building robust wavelet estimators for multicomponent images using Stein's principle," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1814–1830, Nov. 2005.

[18] P. L. Combettes and J.-C. Pesquet, "Wavelet-constrained image restoration," *Int. J. Wavelets, Multires. Inf. Process.*, vol. 2, no. 4, pp. 371–389, Dec. 2004.

[19] J.-C. Pesquet and D. Leporini, "A new wavelet estimator for image denoising," in *Proc. 6th Int. Conf. Image Processing and its Applications*, Jul. 14–17, 1997, vol. 1, pp. 249–253.

[20] W. James and C. Stein, "Estimation with quadratic loss," in *Proc. 4th Berkeley Symp. Math. Statist. Probab.*, 1961, vol. 1, pp. 361–379.

[21] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1522–1531, Sep. 2000.

[22] I. Daubechies, "Ten lectures on wavelets," presented at the CBMS-NSF Regional Conf. Ser. Applied Mathematics, Mar. 1992.

[23] M. K. Mihçak, Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Process. Lett.*, vol. 6, no. 12, pp. 300–303, Dec. 1999.

[24] F. Abramovitch, T. Sapatinas, and B. W. Silverman, "Wavelet thresholding via a Bayesian approach," *J. Roy Statist. Soc. B*, vol. 60, no. 4, pp. 725–749, 1998.

[25] J. S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Pattern Anal. Mach. Intell.*, vol. PAMI-2, no. 3, pp. 165–168, Mar. 1980.

[26] E. P. Simoncelli, *Bayesian Interference in Wavelet Based Models*, ser. Lecture Notes in Statistics. New York: Springer-Verlag, Mar. 1999, vol. 141, ch. 18, pp. 291–308.

[27] B. Vidakovic, *Statistical Modeling by Wavelets*. New York: Wiley-Interscience, Apr. 1999.

Florian Luisier was born in Switzerland in 1981. In 2005, he received the M.S. degree in microengineering from the Swiss Federal Institute of Technology (EPFL), Lausanne, Swizerland. He is currently pursuing the Ph.D. degree within the Biomedical Imaging Group (BIG), EPFL.

His research interests mainly include multiresolution analysis and the restoration of biomedical images.

Thierry Blu (M'96–SM'06) was born in Orléans, France, in 1964. He received the "Diplôme d'ingénieur" from the École Polytechnique, France, in 1986 and from Télécom Paris (ENST), France, in 1988, and the Ph.D. degree in electrical engineering from ENST in 1996 for a study on iterated rational filterbanks, applied to wideband audio coding.

He is with the Biomedical Imaging Group, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, on leave from the France Télécom National Center for Telecommunications Studies (CNET), Issy-les-Moulineaux, France. At EPFL, he teaches the theory of Signals and Systems for Microengineering and Life Science students. His research interests include (multi)wavelets, multiresolution analysis, multirate filterbanks, interpolation, approximation and sampling theory, image denoising, psychoacoustics, optics, wave propagation, etc.

Dr. Blu is the recipient of the 2003 best paper award (SP Theory and Methods) from the IEEE Signal Processing Society. He is currently an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING and, since 2006, for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Michael Unser (M'89–SM'94–F'99) received the M.S. (summa cum laude) and Ph.D. degrees in electrical engineering from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1981 and 1984, respectively.

From 1985 to 1997, he was a Scientist with the National Institutes of Health, Bethesda, MD. He is now a Professor and Director of the Biomedical Imaging Group, EPFL. His main research topics are biomedical image processing, splines, and wavelets. He is the author of over 120 published journal papers in these areas.

Dr. Unser has been actively involved with the IEEE TRANSACTIONS ON MEDICAL IMAGING, holding the positions of Associate Editor (1999–2002 and 2006–present), member of steering committee, and Associate Editor-in-Chief (2003–2005). He has acted as an Associate Editor or member of the editorial board for eight more international journals, including the *IEEE Signal Processing Magazine*, the IEEE TRANSACTIONS ON IMAGE PROCESSING (1992–1995), and the IEEE SIGNAL PROCESSING LETTERS (1994–1998). He organized the first IEEE International Symposium on Biomedical Imaging (ISBI 2002). He currently chairs the technical committee of the IEEE-SP Society on Bio Imaging and Signal Processing (BISP), and well as the ISBI steering committee. He is a recipient of three Best Paper Awards from the IEEE Signal Processing Society.