The SURE-LET Approach to Image Denoising

Thierry Blu, *Senior Member, IEEE*, and Florian Luisier

***Abstract*—We propose a new approach to image denoising, based on the *image-domain minimization* of an estimate of the mean squared error—*Stein's unbiased risk estimate* (SURE). Unlike most existing denoising algorithms, using the SURE makes it needless to hypothesize a statistical model for the noiseless image. A key point of our approach is that, although the (nonlinear) processing is performed in a transformed domain—typically, an undecimated discrete wavelet transform, but we also address nonorthonormal transforms—this minimization is performed in the image domain. Indeed, we demonstrate that, when the transform is a "tight" frame (an undecimated wavelet transform using orthonormal filters), separate subband minimization yields substantially worse results. In order for our approach to be viable, we add another principle, that the denoising process can be expressed as a linear combination of elementary denoising processes—*linear expansion of thresholds* (LET). Armed with the SURE and LET principles, we show that a denoising algorithm merely amounts to solving a *linear system of equations* which is obviously fast and efficient. Quite remarkably, the very competitive results obtained by performing a simple threshold (image-domain SURE optimized) on the undecimated Haar wavelet coefficients show that the SURE-LET principle has a huge potential.**

I. Introduction

DURING acquisition and transmission, images are often corrupted by additive noise. The main aim of an image denoising algorithm is then to reduce the noise level, while preserving the image features.

Transform domain image denoising—the most popular approaches to process noisy images consist in first applying some linear—often multiscale—transformation, then performing a usually nonlinear—and sometimes multivariate—operation on the transformed coefficients, and finally reverting to the image domain by applying an inverse linear transformation. Among the many denoising algorithms to date, we would like to cite the following ones.

- *Portilla et al.* [1]:[1] The authors' 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). The resulting denoising method, consequently called *BLS-GSM*, is the most efficient up-to-date approach in terms of peak signal-to-noise ratio (PSNR).
- *Pižurica et al.* [2]:[2] Assuming a generalized Laplacian prior for the noise-free data, the authors' 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.* [3], [4]:[3] The authors' method, called *BiShrink*, is based on new non-Gaussian bivariate distributions 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.

While the choice of the transformation is easily justified by well-accepted general considerations—e.g., closeness to the Karhunen–Loève transformation, "sparsity" of the transformed coefficients, "steerability" of the transformation—the nonlinear operation that follows is more frequently based on *ad hoc* statistical hypotheses on the transformed coefficients that are specific to each author. The final performance of the algorithms—typically, PSNR results—is, thus, inconclusively related to the accuracy of this modelization.

SURE-LET denoising—In this paper, we want to promote quite a different point of view, which avoids any *a priori* hypotheses on the noiseless image—in particular, *no* random process modelization—but for the usual white Gaussian noise assumption. This approach is made possible by the existence of an excellent unbiased estimate of the mean squared error (MSE) between the noiseless image and its denoised version—Stein's unbiased risk estimate (*SURE*). If we evaluate denoising performances by comparing PSNRs, then this MSE is precisely the quantity that we want to minimize. Similar to the MSE, the SURE takes the form of a quadratic expression in terms of the denoised image (see Theorem1).

Our approach, thus, consists in reformulating the denoising problem as the search for the denoising process that will minimize the SURE—in the image domain. In practice, the process is completely characterized by a set of parameters. Now, to take full advantage of the *quadratic* nature of the SURE, we choose to consider only denoising processes that can be expressed as a *linear combination* of "elementary" denoising processes—linear expansion of thresholds (*LET*). This "SURE-LET" stategy is computationally very efficient because minimizing the SURE for the unknown weights gives rise to a mere linear system of equations, which in turn allows to

Manuscript received March 20, 2007; revised July 13, 2007. 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. Pier Luigi Dragotti.

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

Digital Object Identifier 10.1109/TIP.2007.906002

[1]Available at http://www.io.csic.es/PagsPers/JPortilla/denoise/software/index.htm.

[2]Available at http://telin.ugent.be/~sanja/.

[3]Available at http://taco.poly.edu/WaveletSoftware/denoise2.html.

1057-7149/$25.00 © 2007 IEEE

consider processes described by quite a few parameters. There is, however, a tradeoff between the sharpness of the description of the process which increases with the number of parameters, and the predictability of the MSE estimate, which is inversely related to the number of parameters. We have already applied our approach within a nonredundant, orthonormal wavelet framework, and showed that a simple thresholding function that takes interscale dependences into account is very efficient, both in terms of computation time and image denoising quality[4] [5].

SURE-related literature—Despite its simple MSE justification (a mere integration by parts), the SURE does not belong to the toolbox of the standard signal processing practitioner—although it is, of course, much better established among statisticians. The best-known use of the SURE in image denoising is Donoho's SureShrink algorithm [6] in which a soft-threshold is applied to the *orthonormal* wavelet coefficients, and where the threshold parameter is optimized separately in each subband through the minimization of the SURE. Otherwise, the approach that is most closely related to SURE-LET—but for a multichannel image denoising application—is the contribution by Pesquet and his collaborators [7]–[9] which perform *separate in-band* minimization of the SURE applied to a denoising process that contains both nonlinear and linear parameters.

Yet, the specificity of SURE-LET for redundant or nonorthonormal transforms lies in the fact that this minimization is performed *in the image domain*. While it is true that, due to some Parseval-like MSE conservation, image domain MSE/SURE minimization is equivalent to separate in-band MSE/SURE minimization whenever the analysis transformation is—nonredundant—orthonormal [5], this is grossly wrong as soon as the transformation is, either *redundant* (even when it is a "tight frame") or *nonorthonormal*. This is actually the observation made by those who apply soft-thresholding to an undecimated wavelet transform: the SureShrink threshold determination yields substantially worse results than an empirical choice (see Fig. 3). Unfortunately, this may lead practitioners to wrongly conclude that the SURE approach is unsuitable for redundant transforms, whereas a correct diagnosis should be that it is the independent subband approach that is flawed.

Organization of the paper—In Section II, we expose the multivariate SURE theory for vector functions, and sketch the principles of our linear parametrization strategy; we also address practical issues like how the SURE is modified depending on the choice for boundary conditions, and provide explicit SURE formulæ for pointwise thresholding. In Section III, because we want to exemplify the power of the SURE-LET approach, we restrict the processing to simple pointwise thresholds in the transformed domain and show that, by using an undecimated Haar wavelet transform, a SURE image-domain minimization yields very competitive results with the best up-to-date algorithms [1], [2], [4] (Section IV-C). In comparison, without any optimization attempts in our implementation, the SURE-LET method is quite CPU-time friendly. Yet, a huge margin of improvement can be envisioned if intrascale and interscale dependencies are taken into account. Both the competitiveness and robustness of our method validate our new approach as an attractive solution for image denoising.

[4]See our demo http://bigwww.epfl.ch/demo/suredenoising/.

II. Theoretical Background

We consider the standard simplified denoising problem: given noisy data $y_n = x_n + b_n$, for $n = 1 \dots N$ where b_n is a white Gaussian noise of variance σ^2, find a reasonably good estimate $\hat{\mathbf{x}}$ of $\mathbf{x} = \{x_n\}_{n=1,2,\dots N}$. Our goal is, thus, to find a function of the noisy data alone $\mathbf{F}(\mathbf{y}) = (f_n(\mathbf{y}))_{n=1,2,\dots N} = \hat{\mathbf{x}}$ which will minimize the MSE defined by

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

A. Unbiased Estimate of the MSE

Since we do not have access to the original signal $\mathbf{x}$, we cannot compute $\|\hat{\mathbf{x}} - \mathbf{x}\|^2/N$—the *Oracle* MSE. However, without any assumptions on the noise-free data, we will see that it is possible to replace this quantity by an unbiased estimate which is a function of $\mathbf{y}$ only. This has an important consequence: contrary to what is frequently done in the literature, the noise-free signal is not modeled as a random process in our framework—we do not even require $\mathbf{x}$ to belong to a specific class of signals. Thus, the observed randomness of the noisy data originates only from the Gaussian white noise $\mathbf{b}$.

The following lemma which states a version of Stein's lemma [10], shows how it is possible to replace an expression that contains the unknown data $\mathbf{x}$ by another one with the same expectation, but containing the known data $\mathbf{y}$ only.

Lemma 1: Let $\mathbf{F}(\mathbf{y})$ be an N-dimensional vector function such that $\mathcal{E}\{|\partial f_n(\mathbf{y})/\partial y_n|\} < \infty$ for $n = 1, \dots, N$. Then, under the additive white Gaussian noise assumption, the expressions $\mathbf{F}(\mathbf{y})^{\mathrm{T}}\mathbf{x}$ and $\mathbf{F}(\mathbf{y})^{\mathrm{T}}\mathbf{y} - \sigma^2\mathrm{div}\{\mathbf{F}(\mathbf{y})\}$ have the same expectation

$$\mathcal{E}\left\{\sum_{n=1}^{N} f_n(\mathbf{y})x_n\right\} = \mathcal{E}\left\{\sum_{n=1}^{N} f_n(\mathbf{y})y_n\right\} - \sigma^2\mathcal{E}\Bigg\{\underbrace{\sum_{n=1}^{N}\frac{\partial f_n(\mathbf{y})}{\partial y_n}}_{\mathrm{div}\{\mathbf{F}(\mathbf{y})\}}\Bigg\} \tag{2}$$

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

Proof: We use the fact that a Gaussian white probability density $q(b_n)$ satisfies $b_n q(b_n) = -\sigma^2 q'(b_n)$. Thus, denoting by $\mathcal{E}_{b_n}\{\cdot\}$ the partial expectation over the nth component of the noise, we have the following sequence of equalities:[5]

$$\begin{aligned}
\mathcal{E}_{b_n}\{f_n(\mathbf{y})x_n\} &= \mathcal{E}_{b_n}\{f_n(\mathbf{y})y_n\} - \mathcal{E}_{b_n}\{f_n(\mathbf{y})b_n\} \\
&= \mathcal{E}_{b_n}\{f_n(\mathbf{y})y_n\} - \int f_n(\mathbf{y})b_n q(b_n)db_n \\
&= \mathcal{E}_{b_n}\{f_n(\mathbf{y})y_n\} + \sigma^2\int f_n(\mathbf{y})q'(b_n)db_n \\
&= \mathcal{E}_{b_n}\{f_n(\mathbf{y})y_n\} - \sigma^2\int \frac{\partial f_n(\mathbf{y})}{\partial y_n}q(b_n)db_n \\
&\quad \text{(by parts)} \\
&= \mathcal{E}_{b_n}\{f_n(\mathbf{y})y_n\} - \sigma^2\mathcal{E}_{b_n}\left\{\frac{\partial f_n(\mathbf{y})}{\partial y_n}\right\}.
\end{aligned}$$

[5]To be fully rigorous, we need to assume that $f_n(\mathbf{y})q(y_n - x_n)$ tends to zero with $|y_n|$, which is very broadly ensured whenever $f_n(\mathbf{y})$ is bounded by some *fastly increasing* function, like $\exp(\|\mathbf{y}\|^2/2\sigma'^2)$ where $\sigma' > \sigma$.

TABLE I
COMPARISON OF SOME OF THE MOST EFFICIENT DENOISING METHODS

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							
BiShrink [4]	37.18	33.38	31.28	29.80	28.67	27.76	25.28	22.11	38.35	34.71	32.89	31.63	30.64	29.83	27.54	24.51
ProbShrink [2]	37.45	33.75	31.71	30.25	29.15	28.24	25.72	22.48	38.51	35.15	33.43	32.19	31.21	30.38	27.98	24.76
BLS-GSM [1]	37.32	33.77	31.74	30.31	29.21	28.33	25.90	**22.67**	38.67	35.34	33.60	32.35	31.35	30.52	28.21	25.09
UWT SURE-LET	**37.63**	**34.00**	**31.97**	**30.53**	**29.40**	**28.48**	**25.94**	22.60	**38.71**	**35.52**	**33.81**	**32.60**	**31.66**	**30.89**	**28.58**	**25.25**
UWT Oracle	*37.64*	*34.01*	*31.98*	*30.55*	*29.43*	*28.51*	*26.00*	*22.72*	*38.71*	*35.53*	*33.83*	*32.63*	*31.70*	*30.94*	*28.68*	*25.44*
Method	**Al** 512×512								**Bridge** 256×256							
BiShrink [4]	38.72	35.34	33.51	32.24	31.26	30.45	28.15	24.99	35.18	30.47	28.11	26.62	25.56	24.77	22.88	20.73
ProbShrink [2]	38.67	35.42	33.68	32.45	31.51	30.66	28.46	25.40	35.08	30.36	27.97	26.52	25.52	24.79	23.03	20.93
BLS-GSM [1]	**38.98**	**35.57**	**33.81**	**32.60**	**31.67**	**30.91**	**28.73**	**25.75**	**35.26**	30.49	28.11	26.65	25.66	24.92	23.11	20.98
UWT SURE-LET	38.88	35.43	33.60	32.36	31.42	30.66	28.57	25.69	35.23	**30.54**	**28.24**	**26.82**	**25.83**	**25.10**	**23.27**	**21.09**
UWT Oracle	*38.88*	*35.43*	*33.61*	*32.37*	*31.43*	*30.67*	*28.59*	*25.76*	*35.23*	*30.55*	*28.25*	*26.83*	*25.84*	*25.11*	*23.30*	*21.16*
Method	**Barbara** 512×512								**Boat** 512×512							
BiShrink [4]	37.35	33.51	31.37	29.87	28.72	27.79	25.30	22.46	36.72	33.17	31.30	29.98	28.96	28.14	25.97	23.31
ProbShrink [2]	37.39	33.49	31.24	29.60	28.33	27.30	24.54	22.02	36.69	33.29	31.34	29.97	28.91	28.06	25.83	23.17
BLS-GSM [1]	**37.79**	**34.02**	**31.84**	**30.29**	**29.10**	**28.12**	**25.44**	**22.59**	36.98	**33.58**	**31.70**	**30.37**	**29.35**	**28.54**	**26.35**	**23.70**
UWT SURE-LET	36.98	32.65	30.16	28.45	27.18	26.23	24.13	22.26	**37.13**	33.53	31.57	30.22	29.20	28.39	26.20	23.61
UWT Oracle	*36.98*	*32.65*	*30.16*	*28.45*	*27.19*	*26.24*	*24.14*	*22.29*	*37.13*	*33.54*	*31.58*	*30.23*	*29.21*	*28.40*	*26.22*	*23.65*
Method	**Crowd** 512×512								**Goldhill** 512×512							
BiShrink [4]	34.86	29.85	27.28	25.61	24.40	23.47	21.05	18.18	36.78	33.11	31.23	29.99	29.08	28.37	26.52	24.19
ProbShrink [2]	34.79	30.00	27.47	25.80	24.57	23.61	21.14	18.24	36.76	33.20	31.33	30.12	29.22	28.53	26.71	24.51
BLS-GSM [1]	34.97	30.07	27.52	25.87	24.67	23.73	21.29	18.37	**36.98**	**33.36**	**31.50**	**30.28**	**29.39**	**28.69**	**26.85**	24.61
UWT SURE-LET	**35.10**	**30.20**	**27.64**	**25.96**	**24.74**	**23.78**	**21.32**	**18.43**	36.85	33.20	31.37	30.17	29.30	28.61	26.83	**24.69**
UWT Oracle	*35.10*	*30.20*	*27.64*	*25.96*	*24.74*	*23.78*	*21.33*	*18.45*	*36.85*	*33.21*	*31.37*	*30.18*	*29.30*	*28.62*	*26.85*	*24.75*

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

Now, taking the expectation over the remaining components of the noise, we get

$$\mathcal{E}\{f_n(\mathbf{y})x_n\} = \mathcal{E}\{f_n(\mathbf{y})y_n\} - \sigma^2\mathcal{E}\left\{\frac{\partial f_n(\mathbf{y})}{\partial y_n}\right\}.$$

Since the expectation is a linear operator, (2) follows directly.■

By applying Lemma 1 to the expression of the MSE, we then get Stein's unbiased risk—or MSE—estimate (SURE).

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

$$\epsilon = \frac{1}{N}\|\mathbf{F}(\mathbf{y}) - \mathbf{y}\|^2 + \frac{2\sigma^2}{N}\mathrm{div}\{\mathbf{F}(\mathbf{y})\} - \sigma^2 \tag{3}$$

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

$$\mathcal{E}\{\epsilon\} = \frac{1}{N}\mathcal{E}\left\{\|\mathbf{F}(\mathbf{y}) - \mathbf{x}\|^2\right\}.$$

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

$$\begin{aligned}\mathcal{E}\left\{\|\mathbf{F}(\mathbf{y}) - \mathbf{x}\|^2\right\} &= \mathcal{E}\left\{\|\mathbf{F}(\mathbf{y})\|^2\right\} - 2\mathcal{E}\left\{\mathbf{F}(\mathbf{y})^{\mathrm{T}}\mathbf{x}\right\} \\ &\quad + \mathcal{E}\left\{\|\mathbf{x}\|^2\right\} \\ &= \mathcal{E}\left\{\|\mathbf{F}(\mathbf{y})\|^2\right\} - 2\mathcal{E}\left\{\mathbf{F}(\mathbf{y})^{\mathrm{T}}\mathbf{y}\right\} \\ &\quad + 2\sigma^2\mathcal{E}\{\mathrm{div}\{\mathbf{F}(\mathbf{y})\}\} + \mathcal{E}\left\{\|\mathbf{x}\|^2\right\}\end{aligned}$$

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

We want to emphasize here the fact that in image denoising applications the number of samples is usually large—typically 256^2—and, thus, the estimate ϵ has a small variance—typically $\propto 1/N$. This estimate is, thus, close to its expectation, which is indeed the true MSE of the denoising process.

B. SURE-LET Approach

Our general denoising strategy consists in expressing the denoising process, $\mathbf{F}(\mathbf{y})$, as a linear combination (LET: linear expansion of thresholds) of given elementary processes, $\mathbf{F}_k(\mathbf{y})$

$$\mathbf{F}(\mathbf{y}) = \sum_{k=1}^{K} a_k\mathbf{F}_k(\mathbf{y}). \tag{4}$$

Here, the unknown weights a_k are specified by minimizing the SURE given by (3). It is also possible, in order to evaluate the performance of the algorithm, to compare the result with what the minimization of the MSE would provide—i.e., the *Oracle* optimization (see Table I). A limitation of the LET approach is that the elementary denoising functions $\mathbf{F}_k$ have to fulfill the hypothesis of Lemma 1 (differentiability); moreover, the number of parameters K must not be "too large" compared to the number of pixels (typically, less than 100 for usual 256×256 images), in order for the variance of the SURE to remain small.

The linearity of the expansion (4) is a crucial advantage for solving the minimization problem, because the SURE is quadratic in $\mathbf{F}(\mathbf{y})$. The coefficients a_k are, thus, the solution of a linear system of equations

$$\begin{aligned}\sum_{l=1}^{K} \underbrace{\mathbf{F}_k(\mathbf{y})^{\mathrm{T}}\mathbf{F}_l(\mathbf{y})}_{[\mathbf{M}]_{k,l}} a_l &= \underbrace{\mathbf{F}_k(\mathbf{y})^{\mathrm{T}}\mathbf{y} - \sigma^2\mathrm{div}\{\mathbf{F}_k(\mathbf{y})\}}_{[\mathbf{c}]_k} \\ &\quad \text{for } k = 1, 2, \ldots K \\ &\Updownarrow \\ \mathbf{Ma} &= \mathbf{c}.\end{aligned} \tag{5}$$

Note that, since the minimum of ϵ always exists, we are ensured that there will always be a solution to this system. When

$\text{rank}(\mathbf{M}) < K$, the function $\mathbf{F}$ is *over-parameterized* and consequently, several sets of parameters a_k yield equivalent results; in that case, we may simply consider the solution provided by the pseudoinverse of $\mathbf{M}$. Of course, it is also possible to reduce the parametrization order K so as to make the matrix $\mathbf{M}$ full rank—at no quality loss.

What this approach suggests is that the practitioner may choose at will (restricted only by the differentiability constraint of Theorem 1) a set of K different denoising algorithms—ideally with complementary denoising behaviors—and optimize a weighting of these algorithms to get the best of them at once.

Among the potentially interesting algorithms are those that work in a transformed domain such as:

- the nonredundant wavelet transforms, either orthogonal or bi-orthogonal [11];
- the classical undecimated wavelet transform [12];
- the *curvelet* [13] transform;
- the *contourlet* [14] transform;
- the steerable pyramids [15], [16];

as well as the discrete cosine transform (*DCT*) or its overcomplete variant: the block discrete cosine transform (*B-DCT*). In the remainder of this paper, we will consider only pointwise thresholding in such transform domains.

C. Pointwise SURE-LET Transform Denoising

Transform domain denoising consists in first defining a couple of linear transformations $\mathcal{D}$—decomposition—and $\mathcal{R}$—reconstruction—such that $\mathcal{R}\mathcal{D} = \text{Identity}$: typically, $\mathcal{D}$ is a bank of decimated or undecimated filters. Once the size of the input and output data are frozen, these linear operators are characterized by matrices, respectively $\mathbf{D} = (d_{i,j})_{(i,j)\in[1;L]\times[1;N]}$ and $\mathbf{R} = (r_{i,j})_{(i,j)\in[1;N]\times[1;L]}$ that satisfy the *perfect reconstruction* property $\mathbf{RD} = \mathbf{Id}$. Then, the whole denoising process boils down to the following steps.

1) Apply $\mathbf{D}$ to the noisy signal $\mathbf{y} = \mathbf{x} + \mathbf{b}$ to get the transformed noisy coefficients $\mathbf{w} = \mathbf{Dy} = (w_i)_{i\in[1;L]}$.
2) Apply a *pointwise* thresholding function $\Theta(\mathbf{w}) = (\theta_i(w_i))_{i\in[1;L]}$.
3) Revert to the original domain by applying $\mathbf{R}$ to the thresholded coefficients $\Theta(\mathbf{w})$, yielding the denoised estimate $\hat{\mathbf{x}} = \mathbf{R}\Theta(\mathbf{w})$.

This algorithm can be summarized as a function of the noisy input coefficients

$$\hat{\mathbf{x}} = \mathbf{F}(\mathbf{y}) = \mathbf{R}\Theta(\mathbf{Dy}). \tag{6}$$

The SURE-LET approach suggests to express $\mathbf{F}$ as a *linear expansion* of denoising algorithms $\mathbf{F}_k$, according to

$$\mathbf{F}(\mathbf{y}) = \sum_{k=1}^{K} a_k \underbrace{\mathbf{R}\Theta_k(\mathbf{Dy})}_{\mathbf{F}_k(\mathbf{y})} \tag{7}$$

where $\Theta_k(\cdot)$ are elementary pointwise thresholding functions.

As we have noticed in the previous subsection [see (5)], retrieving the parameters a_k boils down to the resolution of a linear system of equations. Note that this linear parametrization does not imply a linear denoising; indeed, the thresholding functions Θ_k can be chosen nonlinear.

In the SURE-LET framework, Theorem 1 can be reformulated in the following way.

Corollary 1: Let $\mathbf{F}$ be defined according to (6) where Θ denotes *pointwise* thresholding. Then the MSE between the original and the denoised signal can be *unbiasedly* estimated by the following random variable:

$$\epsilon = \frac{1}{N}\|\mathbf{F}(\mathbf{y}) - \mathbf{y}\|^2 + \frac{2\sigma^2}{N}\boldsymbol{\alpha}^{\mathrm{T}}\Theta'(\mathbf{Dy}) - \sigma^2 \tag{8}$$

where

- $\boldsymbol{\alpha} = \text{diag}\{\mathbf{DR}\} = \{[\mathbf{DR}]_{1,1}, [\mathbf{DR}]_{2,2}, \ldots, [\mathbf{DR}]_{L,L}\}$ is a vector made of the diagonal elements of the matrix $\mathbf{DR}$;
- $\Theta'(\mathbf{Dy}) = \Theta'(\mathbf{w}) = (\theta_i'(w_i))_{i\in[1;L]}$.

In particular, when $\mathbf{D} = [\mathbf{D}_1^{\mathrm{T}}, \mathbf{D}_2^{\mathrm{T}}, \ldots, \mathbf{D}_J^{\mathrm{T}}]^{\mathrm{T}}$ and $\mathbf{R} = [\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_J]$ where $\mathbf{D}_i, \mathbf{R}_i$ are $N_i \times N$ and $N \times N_i$ matrices, then $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^{\mathrm{T}}, \boldsymbol{\alpha}_2^{\mathrm{T}}, \ldots, \boldsymbol{\alpha}_J^{\mathrm{T}}]^{\mathrm{T}}$ where $\boldsymbol{\alpha}_i = \text{diag}\{\mathbf{D}_i\mathbf{R}_i\}$.

Proof: By applying Theorem 1, we only have to prove that in the SURE-LET framework

$$\text{div}\{\mathbf{F}(\mathbf{y})\} = \boldsymbol{\alpha}^{\mathrm{T}}\Theta'(\mathbf{Dy}). \tag{9}$$

By using the reconstruction formula $\mathbf{F}(\mathbf{y}) = \mathbf{R}\Theta(\mathbf{w})$, i.e., $f_n(\mathbf{y}) = \sum_{l=1}^{L} r_{n,l}\theta_l(w_l)$, and the decomposition formula $\mathbf{w} = \mathbf{Dy}$, i.e., $w_l = \sum_{k=1}^{N} d_{l,k}y_k$, we can successively write the following equalities:

$$\begin{aligned}
\text{div}\{\mathbf{F}(\mathbf{y})\} &= \sum_{n=1}^{N} \frac{\partial f_n(\mathbf{y})}{\partial y_n} \\
&= \sum_{n=1}^{N}\sum_{l=1}^{L} r_{n,l}\theta_l'(w_l)\frac{\partial w_l}{\partial y_n} \\
&= \sum_{n=1}^{N}\sum_{l=1}^{L} r_{n,l}\theta_l'(w_l)d_{l,n} \\
&= \sum_{l=1}^{L} \theta_l'(w_l) \underbrace{\sum_{n=1}^{N} d_{l,n}r_{n,l}}_{[\mathbf{DR}]_{l,l}}
\end{aligned} \tag{10}$$

and, finally, conclude that $\text{div}\{\mathbf{F}(\mathbf{y})\} = \text{diag}\{\mathbf{DR}\}^{\mathrm{T}}\Theta'(\mathbf{Dy})$. ■

As it appears in this corollary, the computation of the divergence term—i.e., of $\text{diag}\{\mathbf{DR}\}$—is a crucial point.

1) Evaluation of the Divergence Term—$\boldsymbol{\alpha}$: In the general case where $\mathbf{D}$ and $\mathbf{R}$ are known only by their action on vectors, and not explicitly by their matrix coefficients—typically, when only $\mathcal{D}$ and $\mathcal{R}$ are specified—the analytical expression for $\boldsymbol{\alpha}$ is quite painful to compute: in order to build $\boldsymbol{\alpha}$, for each $l = 1, 2, \ldots L$ it is necessary to compute the reconstruction $\mathbf{f}_l = \mathbf{R}\mathbf{e}_l$ (where $[\mathbf{e}_l]_n = \delta_{n-l}$ is the canonical basis of $\mathbb{R}^L$), then the decomposition $\mathbf{Df}_l$ and keep the lth component. Given that L is of the order of 256^2—and even much more in the case of redundant transforms—this process may be extremely costly, even considering that it has to be done only once. Fortunately, we can always compute a very good approximation of it using the following numerical algorithm.

For $i = 1 \cdots I$

1) Generate a normalized Gaussian white noise $\mathbf{b}_i \in \mathbb{R}^L$.

2) Apply the reconstruction matrix to $\mathbf{b}_i$ to get the vector $\mathbf{r}_i = \mathbf{R}\mathbf{b}_i$ of size $N \times 1$.
3) Apply the decomposition matrix to $\mathbf{r}_i$ to get the vector $\mathbf{b}'_i = \mathbf{D}\mathbf{R}\mathbf{b}_i$ of size $L \times 1$.
4) Compute the element-by-element product of $\mathbf{b}'_i$ with $\mathbf{b}_i$ to get a vector of L coefficients $\mathbf{v}_i = \mathrm{diag}\{\mathbf{b}'_i\mathbf{b}_i^{\mathrm{T}}\}$, which can be viewed as a realization of the random vector $\mathbf{v} = \mathrm{diag}\{\mathbf{D}\mathbf{R}\mathbf{b}\mathbf{b}^{\mathrm{T}}\}$.

end

An approximate value $\hat{\boldsymbol{\alpha}}$ for $\mathrm{diag}\{\mathbf{DR}\}$ is finally obtained by averaging the realizations $\mathbf{v}_i$ over I runs (typically, $I = 1000$ provides great accuracy)

$$\hat{\boldsymbol{\alpha}} = \frac{1}{I}\sum_{i=1}^{I}\mathbf{v}_i. \tag{11}$$

The above algorithm is justified by the following lemma.

Lemma 2: Let $\mathbf{b}$ be a normalized Gaussian white noise with L components. Then, we have the following equality:

$$\mathcal{E}\left\{\mathrm{diag}\{\mathbf{D}\mathbf{R}\mathbf{b}\mathbf{b}^{\mathrm{T}}\}\right\} = \mathrm{diag}\{\mathbf{DR}\}. \tag{12}$$

Proof:

$$\begin{aligned}\mathcal{E}\left\{\mathrm{diag}\{\mathbf{D}\mathbf{R}\mathbf{b}\mathbf{b}^{\mathrm{T}}\}\right\} &= \mathrm{diag}\Big\{\mathbf{DR}\underbrace{\mathcal{E}\{\mathbf{b}\mathbf{b}^{\mathrm{T}}\}}_{\mathbf{Id}}\Big\}\\ &= \mathrm{diag}\{\mathbf{DR}\}.\end{aligned}$$

■

The numerical computation of $\mathrm{diag}\{\mathbf{DR}\}$ can be performed offline for various image sizes, since it does not depend specifically on the image—but for its size—nor on the noise level.

2) Influence of the Boundary Extensions: One of the main drawbacks of any transform-domain denoising algorithm is the potential generation of boundary artifacts by the transform itself. Decreasing these effects is routinely done by performing boundary extensions, the most popular choice being symmetric extension and periodic extension. Thus, the effect of these extensions boils down to replacing the transformation $\mathbf{D}$ by another transformation, $\mathbf{D}'$.

Indeed, usual boundary extensions are linear preprocessing applied to the available data $\mathbf{y}$ and can, therefore, be expressed in a matrix form. In particular, for a given boundary extension of length E, i.e., characterized by an $E \times N$ matrix $\mathbf{H}$, the denoising process becomes

$$\begin{aligned}\mathbf{F}(\mathbf{y}) &= [\mathbf{Id}_N\ \mathbf{0}_{N\times E}]\mathbf{R}_{N+E}\Theta\left(\mathbf{D}_{N+E}\begin{bmatrix}\mathbf{y}\\ \mathbf{H}\mathbf{y}\end{bmatrix}\right)\\ &= \mathbf{R}'\Theta(\mathbf{D}'\mathbf{y})\end{aligned}$$

where $\mathbf{D}_{N+E}$ (resp., $\mathbf{R}_{N+E}$) is the matrix corresponding to the linear transformation $\mathcal{D}$ (resp., $\mathcal{R}$) when the input signal is of size $N + E$. Any boundary handling can, therefore, be seen as a modification of the decomposition matrix $\mathbf{D}$ that must be taken into account when computing the divergence term, namely $\mathrm{diag}\{\mathbf{D}'\mathbf{R}'\}$. This is where Lemma 2 is particularly useful: although the implementation of the transformations $\mathcal{D}$ and $\mathcal{R}$ with the adequate boundary extensions may be straightforward, the explicit computation of the coefficients of the matrices $\mathbf{R}'$ and $\mathbf{D}'$ is tedious—and Lemma 2 avoids this computation.

3) Applications to Standard Linear Transforms: In some particular cases of linear transforms, it is possible to easily compute $\mathrm{diag}\{\mathbf{DR}\}$ analytically, as shown in the following.

a) Nonredundant transforms: Here, we assume that the number of samples is preserved in the transform domain, and more precisely:

- $\mathbf{D}$ is a full rank matrix of size $N \times N$;
- $\mathbf{R}$ is also a full rank matrix of size $N \times N$.

Then, the following lemma shows how to compute the divergence term of Corollary 1.

Lemma 3: When $\mathbf{D}$ is nonredundant, the divergence term $\boldsymbol{\alpha}$ in (8) is given by

$$\boldsymbol{\alpha} = \underbrace{[1, 1, \ldots, 1]}_{L\text{ times}}^{\mathrm{T}}. \tag{13}$$

Proof: Because $\mathbf{DR} = \mathbf{RD} = \mathbf{Id}$, we have $\boldsymbol{\alpha} = \mathrm{diag}\{\mathbf{DR}\} = \mathrm{diag}\{\mathbf{Id}\}$. ■

Note that, when additionally the transformation is *orthonormal*, the reconstruction matrix is simply the transpose of the decomposition matrix, i.e., $\mathbf{R} = \mathbf{D}^{\mathrm{T}}$. Consequently, in corollary 1, the SURE becomes

$$\begin{aligned}\epsilon &= \frac{1}{N}\left\|\mathbf{F}(\mathbf{y}) - \mathbf{y}\right\|^2 + \frac{2\sigma^2}{N}\boldsymbol{\alpha}^{\mathrm{T}}\Theta'(\mathbf{D}\mathbf{y}) - \sigma^2\\ &= \frac{1}{N}\left\|\Theta(\mathbf{D}\mathbf{y}) - \mathbf{D}\mathbf{y}\right\|^2 + \frac{2\sigma^2}{N}\boldsymbol{\alpha}^{\mathrm{T}}\Theta'(\mathbf{D}\mathbf{y}) - \sigma^2\\ &\quad (\text{orthogonality of } \mathbf{R})\\ &= \frac{1}{N}\sum_{i=1}^{N}\Big((\theta_i(w_i) - w_i)^2 + 2\sigma^2\theta'_i(w_i)\Big) - \sigma^2\end{aligned} \tag{14}$$

where w_i is the ith component of $\mathbf{D}\mathbf{y}$; i.e., it is a sum of the specific MSE estimates for each transformed coefficient w_i. The optimization procedure can, thus, be performed *separately* in the transform domain [5]. This is specific to orthonormal transforms: nonredundant biorthogonal transforms do not enjoy this property; i.e., the optimization in the transform domain is *not* equivalent to the optimization in the image domain. Yet, Lemma 3 still applies and is actually particularly useful for applying our SURE minimization strategy.

b) Undecimated filterbank transforms: Here, we will consider linear redundant transforms characterized by J analysis filters $\tilde{G}_i(z) = \sum_n \tilde{g}_i[n]z^{-n}$ and J synthesis filters $G_i(z) = \sum_n g_i[n]z^{-n}$ as shown in Fig. 1.

A periodic boundary extension implementation of this structure gives rise to decomposition and reconstruction matrices $\mathbf{D}$ and $\mathbf{R}$ made of J circulant submatrices—i.e., diagonalized with an N-point DFT matrix—$\mathbf{D}_i$ and $\mathbf{R}_i$ of size $N \times N$ each, with coefficients

$$\begin{aligned}[\mathbf{D}_i]_{k,l} &= \sum_n \tilde{g}_i[l - k + nN]\\ [\mathbf{R}_i]_{k,l} &= \sum_n g_i[k - l + nN].\end{aligned}$$

We then have the following lemma to be used in Corollary 1:

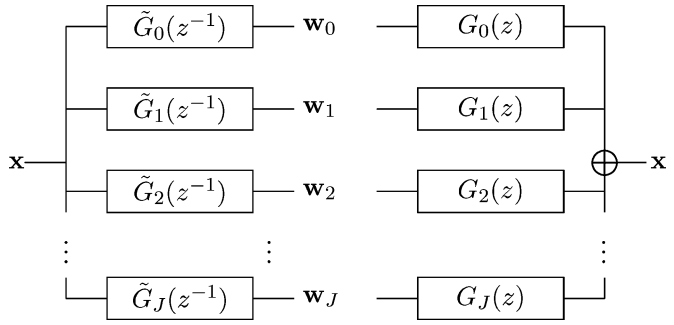

Fig. 1. Undecimated J-band analysis-synthesis filterbank.

Lemma 4: When $\mathbf{D}$ and $\mathbf{R}$ are periodically extended implementations of the analysis-synthesis filterbank of Fig. 1, the divergence term $\boldsymbol{\alpha}$ in (8) is given by $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^{\mathrm{T}}, \boldsymbol{\alpha}_2^{\mathrm{T}}, \dots \boldsymbol{\alpha}_J^{\mathrm{T}}]^{\mathrm{T}}$ where

$$\boldsymbol{\alpha}_i = \left(\sum_n \gamma_i[nN] \right) \underbrace{[1, 1, \dots, 1]^{\mathrm{T}}}_{N\text{times}}. \tag{15}$$

and where $\gamma_i[n]$ is the nth coefficient of the filter $\tilde{G}_i(z^{-1})G_i(z)$. The extension to filterbanks in higher dimensions is straightforward—the summation in (15) running over a multidimensional index n.

Proof: According to Corollary 1, we have to compute $\boldsymbol{\alpha}_i = \mathrm{diag}\{\mathbf{D}_i\mathbf{R}_i\}$. Since $\mathbf{D}_i$ and $\mathbf{R}_i$ are circulant matrices the product $\mathbf{D}_i\mathbf{R}_i$ is also circulant and is built using the N-periodized coefficients of the filter $\tilde{G}_i(z^{-1})G_i(z)$, i.e.,

$$[\mathbf{D}_i\mathbf{R}_i]_{k,l} = \sum_n \gamma_i[k - l + nN]$$

the diagonal of which yields (15). ∎

It is often assumed that $\tilde{G}_i$ and G_i satisfy the biorthogonality condition

$$\sum_{k=0}^{M-1} \underbrace{\tilde{G}_i(z^{-1}e^{-j2\pi k/M_i})\tilde{G}_i(ze^{j2\pi k/M_i})}_{\Gamma_i(ze^{j2\pi k/M_i})} = \lambda_i \tag{16}$$

where M_i is a divisor of N, because undecimated filterbanks are usually obtained from critically sampled filterbanks—for which (16) holds with $\lambda_i = M_i$. In this case, since (16) actually specifies the coefficients $\gamma_i[nM_i]$, we find that $\boldsymbol{\alpha}_i = \lambda_i/M_i[1, 1, \dots, 1]^{\mathrm{T}}$.

An example of such a transform is the standard undecimated wavelet transform (UWT) which uses $J + 1$ ($3J + 1$ in two dimensions) orthonormal filters (see Fig. 2). In that case, the equivalent filters are given by

$$\begin{aligned}\tilde{G}_i(z) &= 2^i G_i(z) \\ &= H(z)H(z^2)\dots H\left(z^{2^{i-2}}\right) G\left(z^{2^{i-1}}\right) \\ &\quad \text{for } i = 1, 2, \dots, J \\ \tilde{G}_{J+1}(z) &= 2^J G_{J+1}(z) = H(z)H(z^2)\dots H\left(z^{2^{J-1}}\right).\end{aligned}$$

They satisfy (16) for $\lambda_i = 1$. This shows that $\boldsymbol{\alpha}_i = 2^{-i}[1, 1, \dots, 1]^{\mathrm{T}}$ for all $i = 1, 2, \dots J$ and $\boldsymbol{\alpha}_{J+1} = 2^{-J}[1, 1, \dots, 1]^{\mathrm{T}}$. In a 2-D separable framework, these values are extended straightforwardly, taking into account that the 2-D filters still satisfy (16) for $\lambda_i = 1$: the general result is, thus, that $\boldsymbol{\alpha}_i$ is given by the (2-D) downsampling factor $1/M_i$.

III. Example of a SURE-LET Denoising Algorithm

In Section II-C, we have proposed a general form of denoising function (7), which involves several degrees of freedom: the linear transformation, the number K of linear parameters, and the thresholding functions Θ_k. This section studies a possible choice. The denoising performance of the resulting algorithm will be evaluated in the next section.

First, we will restrict ourselves to the undecimated wavelet transform,[6] although other linear transforms may in some cases be more advisable—e.g., the undecimated DCT, the curvelet transform, etc…

A. Choosing an Efficient Thresholding Function

A pointwise thresholding function is likely to be efficient if it satisfies the following minimal properties:

- differentiability: required to apply Theorem 1—this rules out pure hard-thresholds;
- anti-symmetry: we assume that the coefficients are not expected to exhibit a sign preference;
- linear behavior for large coefficients: because when a coefficient is large, it can be kept unmodified—noise corruption is negligible.

A good choice has been experimentally found to be of the form

$$\begin{aligned}\theta_i(w) &= a_{i,1}t_1(w) + a_{i,2}t_2(w) \\ \text{where} \quad t_1(w) &= w \text{ and } t_2(w) = w\left(1 - e^{-\left(\frac{w}{3\sigma}\right)^8}\right)\end{aligned} \tag{17}$$

in each band i. The nonlinear term, $t_2(w)$, can be seen as a regular approximation of a *Hard-threshold*.

Similarly to what was observed empirically in other settings [5], [17], adding more thresholding functions only bring marginal (∼0.1–0.2 dB) improvement to the overall denoising quality.

B. Solving for the Linear Parameters

Finding the parameters $a_{i,k}$ that minimize the MSE estimate ϵ amounts to solving the linear system of (5) in which it is necessary to replace $\mathbf{F}(\mathbf{y})$ by

$$\mathbf{F}(\mathbf{y}) = \sum_{i=1}^{J}\sum_{k=1}^{2} a_{i,k}\mathbf{F}_{i,k}(\mathbf{y}) + \text{lowpass}$$

where $\mathbf{F}_{i,k}$ is the image obtained by zeroing all the bands $i' \neq i$ and processing the subband i with the thresholding function $t_k(w)$. Note that, as usual in denoising algorithms, the $(J+1)$th band, lowpass, is not processed.

As shown in Section II-C3b, the divergence term in (5) has an exact expression, namely $\mathrm{div}\{\mathbf{F}_{i,k}(\mathbf{y})\} = 4^{-i}\sum_{w\in\mathrm{band}i} t'_k(w)$. Alternatively, in particular, in the case of nonperiodic boundary image extensions, it is possible to use the approximate algorithm presented in Section II-C2.

C. Summary of the Algorithm

1) Perform a boundary extension on the noisy image.

[6] In our tests, the best performer was the Haar wavelet.

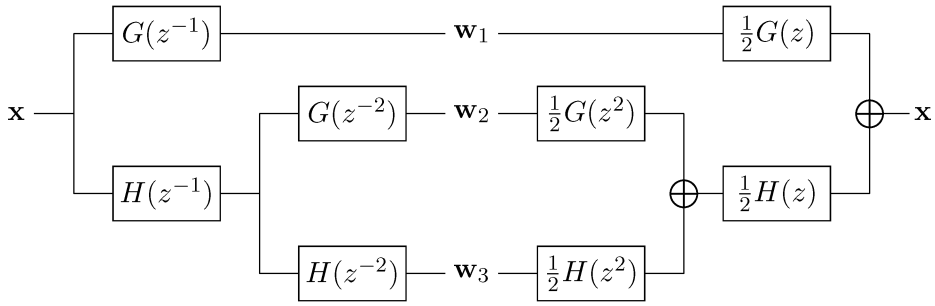

Fig. 2. Classical undecimated wavelet filterbank for 1-D signal.

2) Perform an UWT on the extended noisy image.
3) **For** $i = 1 \ldots J$ (number of bandpass subbands), **For** $k = 1, 2$:
   a) Apply the *pointwise* thresholding functions t_k defined in (17) to the current subband $\mathbf{w}_i$.
   b) Reconstruct the processed subband by setting all the other subbands to zero to obtain $\mathbf{F}_{i,k}(\mathbf{y})$.
   c) Compute the first derivative of t_k for each coefficient of the current subband $\mathbf{w}_i$ and build the corresponding coordinate of $\mathbf{c}$ as exemplified by (5).

   end

   end
4) Compute the matrix $\mathbf{M}$ and deduce the optimal—in the minimum SURE sense—linear parameters $a_{i,k}$'s using the matrix formulation of (5).
5) The noise-free image $\hat{\mathbf{x}}$ is finally estimated by the sum of each $\mathbf{F}_{i,k}$ weighted by its corresponding SURE-optimized $a_{i,k}$.

IV. Results

A. Wavelet-Domain Versus Image-Domain Optimization

Before comparing our SURE-LET approach with the best state-of-the-art algorithms, we demonstrate here that, in order to optimize the denoising process, it is essential to perform the minimization in the *image-domain*. Instead, an *independent wavelet subband* processing is suboptimal, often by a significant margin, even in a "tight" frame representation. This is because we usually do not have energy preservation between the denoised "tight" frame coefficients $\hat{\mathbf{w}}$ and the reconstructed image $\hat{\mathbf{x}} = \mathbf{R}\hat{\mathbf{w}} : \|\mathbf{R}\hat{\mathbf{w}}\| \neq \|\hat{\mathbf{w}}\|$. This is not in contradiction with the well-known energy conservation between the "tight" frame coefficients $\mathbf{w} = \mathbf{D}\mathbf{y}$ and the noisy image $\mathbf{y} : \|\mathbf{D}\mathbf{y}\| = \|\mathbf{y}\|$.

In Fig. 3, we compare a classical wavelet domain SURE-based optimization of our thresholding function (17) with the image domain optimization based on Lemma 4 in the framework of the undecimated *Haar* wavelet transform. We notice that the rigorous image domain optimization provides large improvements—up to $+1$ dB—over the independent in-band optimization. A closer examination of the "optimal" thresholds in both cases indicates that this difference may be related to the difference between the slopes of these functions around zero: the image-domain solution is actually much flatter, making it able to suppress small coefficients almost exactly.

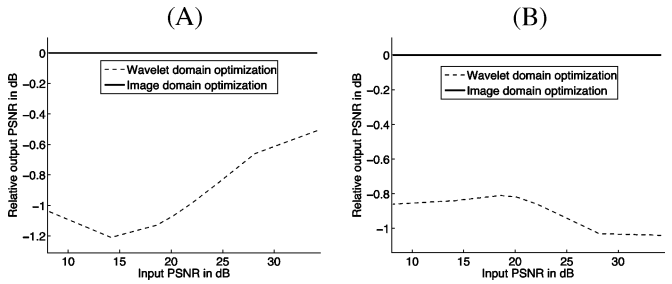

Fig. 3. Comparison of the proposed *SURE-LET* denoising procedure with a SURE-based denoising algorithm optimized in the wavelet domain when using the undecimated wavelet (*Haar*) transform: (a) *House*; (b) *Al*.

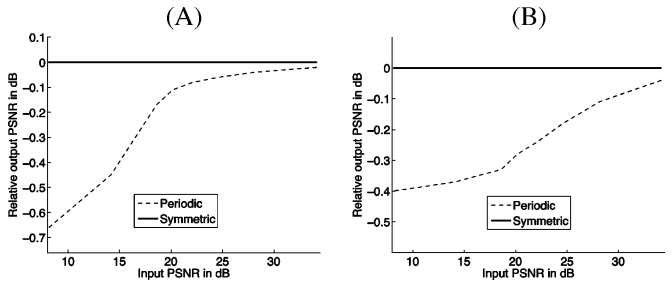

Fig. 4. Influence of the boundary extensions when using the undecimated wavelet (*Haar*) transform: (a) *Peppers*; (b) *House*.

B. Periodic Versus Symmetric Boundary Extensions

It is also worth quantifying the effects of particular boundary extensions. In Fig. 4, we compare symmetric boundary extensions (rigorous SURE computation, as described in Section II-C-2) with the periodic ones. As it can be observed, the symmetric boundary extension can lead to up to $+0.5$ dB of PSNR improvements over the periodic one.

C. Comparison With State-of-the-Art Denoising Schemes

We have compared our Haar wavelet SURE-LET denoising algorithm with some of the best state-of-the-art techniques for which the code is freely distributed by the authors: *BiShrink* [4] (dual tree complex wavelet decomposition), *ProbShrink* [2] (undecimated Daubechies *symlets*) and *BLS-GSM* [1] (full steerable—eight orientations per scale—pyramidal decomposition). Depending on the size of the images, 256×256 or 512×512, we have performed 4 or 5 decomposition levels.

For a reliable comparison, we have run all the algorithms[7] on a comprehensive set of standard grayscale[8] images of size

[7]We have used the same parameters as those suggested by the authors in their respective papers and softwares.

[8]8-bit images with pixels values between 0 and 255.

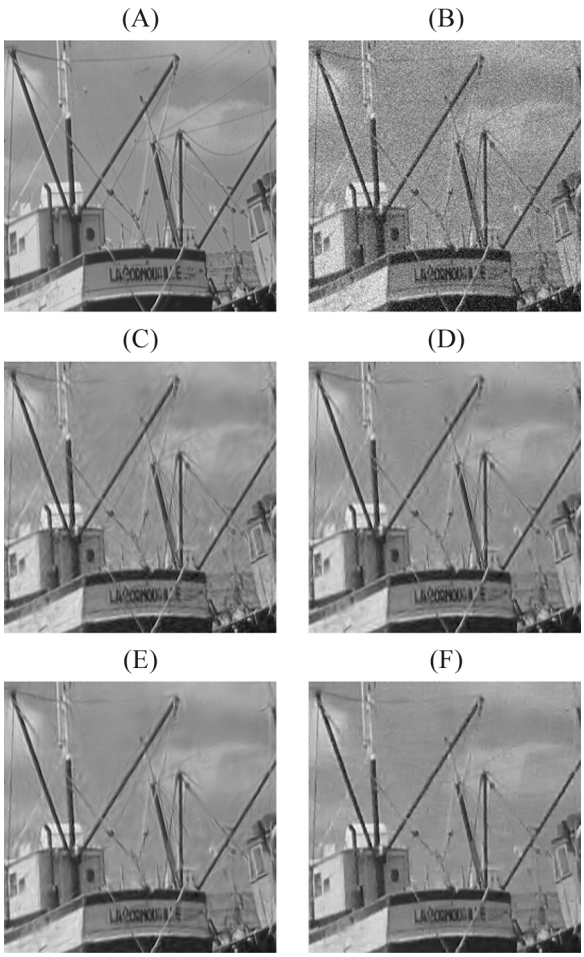

Fig. 5. (a) Part of the noise-free *Boat* image. (b) A noisy version of it: PSNR = 22.11 dB. (c) *BiShrink* denoising result: PSNR = 29.99 dB. (d) *ProbShrink* denoising result: PSNR = 29.97 dB. (e) *BLS-GSM* denoising result: PSNR = 30.36 dB. (f) *UWT SURE-LET* denoising result: PSNR = 30.24 dB.

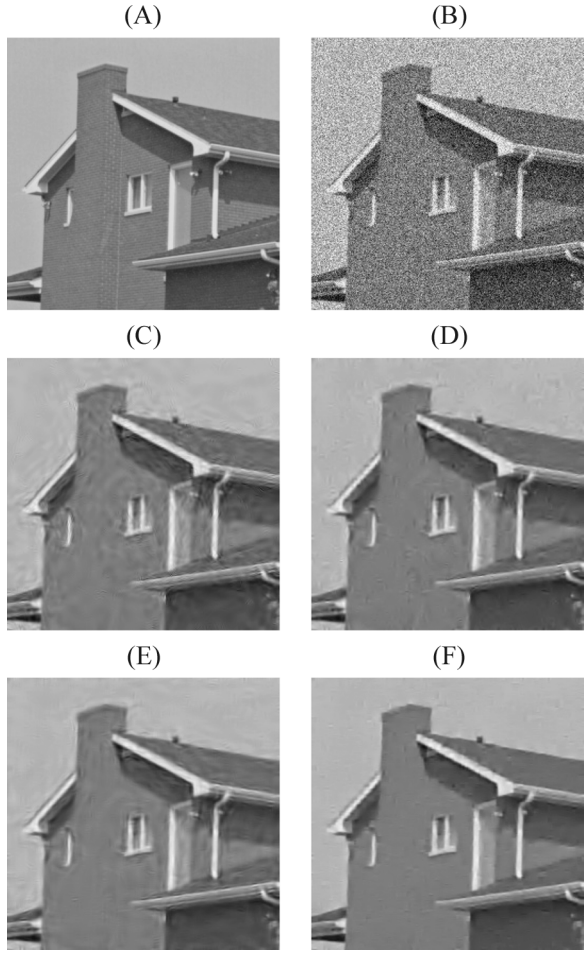

Fig. 6. (a) Noise-free *House* image. (b) A noisy version of it: PSNR = 18.59 dB. (c) *BiShrink* denoising result: PSNR = 29.77 dB. (d)*ProbShrink* denoising result: PSNR = 30.33 dB. (e) *BLS-GSM* denoising result: PSNR = 30.50 dB. (f) *UWT SURE-LET* denoising result: PSNR = 30.90 dB.

256 × 256 (*Peppers*, *House*, *Bridge*) and of size 512 × 512 (*Al*, *Barbara*, *Boat*, *Crowd*, *Goldhill*), each one corrupted with 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]. We have then averaged the output PSNRs over eight noise realizations (the different algorithms are applied to the same noise realizations).

Table I reports the PSNR results we have obtained with the various denoising methods, the best results being shown in boldface. As we can notice, our algorithm (*UWT SURE-LET*) matches the best state-of-the-art results for most of the images, except for Barbara where it may be argued that, either a finer subband decomposition, or a more sophisticated, multivariate, thresholding function should be used in order to capture the texture information that characterizes this image. Note also how the SURE minimization is close to the MSE one (*Oracle* in Table I), which is an evidence of the robustness of the SURE-LET approach.

We want to stress that the denoising algorithm we propose in this section is limited to a pointwise thresholding, contrary to the above mentioned algorithms which involve some kind of multivariate thresholding. Because it simply boils down to solving a linear system of equations, our algoritm is quite fast compared to *BLS-GSM* which has the best denoising results. More precisely, the execution of our current un-optimized Matlab implementation of the whole denoising task lasts on average 3.5 s for 256 × 256 images and about 26 s for 512 × 512 on a Power Mac G5 with CPU speed of 1.8 GHz and 1 GB of memory, whereas Portilla *et al. BLS-GSM* lasts, respectively, 25 and 100 s on the same workstation. Note that the main part of our computational time is dedicated to the independent reconstruction of all the subbands.

Other preliminary tests indicate that if, for images like Barbara, we choose transforms that have more subbands (such as the undecimated DCT), our simple pointwise thresholding strategy may provide slightly better results than *BLS-GSM* (typically, +0.2 dB); moreover, it is possible to select a transform or the other based only on the SURE values. We may also envision that thresholding schemes that involve inter and intrascale dependences substantially improve the denoising performance, as this is the case with orthonormal wavelet transforms [5].

We can finally notice in Figs. 5 and 6 that our *SURE-LET* denoising procedure gives quite a decent visual quality compared to the best state-of-the-art spatially adaptive method.

V. Conclusion

We have presented a new approach to image denoising that is especially useful when redundant or nonorthonormal transforms are involved. In this paper, we have emphasized the theoretical part of our approach and its implementation aspects, in order to make the SURE-LET principle easily applicable for others. Accordingly, we did not try to take advantage of all the degrees of freedom (multivariate thresholding, increased number of parameters, more sophisticated transforms) to make our example of algorithm optimal. And yet, the obtained results are quite competitive with the best state-of-the-art denoising algorithms—which require involved statistical image models. This indicates that there is a substantial margin of improvement of SURE-LET type algorithms.

Acknowledgment

The authors would like to thank Prof. M. Unser for useful discussions, and the anonymous reviewers for their suggestions.

References

[1] 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.

[2] 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. 654–665, Mar. 2006.

[3] 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.

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

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

[6] 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.

[7] 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.

[8] 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.

[9] C. Chaux, L. Duval, A. Benazza-Benyahia, and J.-C. Pequet, "A nonlinear Stein based estimator for multichannel image denoising," *IEEE Trans. Signal Process.*, to be published.

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

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

[12] G. Nason and B. W. Silverman, *The Stationary Wavelet Transform and Some Statistical Applications*. New York: Springer-Verlag, 1995, vol. 103.

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

[14] M. N. Do and M. Vetterli, "The Contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, Dec. 2005.

[15] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sep. 1991.

[16] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multi-scale transforms," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 587–607, Mar. 1992.

[17] M. Raphan and E. P. Simoncelli, "Learning to be Bayesian without supervision," presented at the NIPS Conf., Dec. 2006.

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

[19] 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.

[20] 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. 1532–1546, Sep. 2000.

[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] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[23] 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.

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

[25] 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.

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

[27] E. P. Simoncelli, *Bayesian Denoising of Visual Images in the Wavelet Domain*, ser. Lecture Notes in Statistics. New York: Springer-Verlag, Mar. 1999, vol. 141.

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

[29] 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.

Thierry Blu (M'96–SM'06) was born in Orléans, France, in 1964. He received the "Diplôme d'ingénieur" from 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 at the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, on leave from the France Telecom R&D center in Issy-les-Moulineaux. 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, and wave propagation.

Dr. Blu was the recipient of two best paper awards from the IEEE Signal Processing Society (2003 and 2006). From 2002 and 2006, he was an Associate Editor for the IEEE Transactions on Image Processing and, since 2006, he has been an Associate Editor for the IEEE Transactions on Signal Processing.

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

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