SURE-LET Multichannel Image Denoising: Interscale Orthonormal Wavelet Thresholding

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

***Abstract*—We propose a vector/matrix extension of our denoising algorithm initially developed for grayscale images, in order to efficiently process multichannel (e.g., color) images. This work follows our recently published SURE-LET approach where the denoising algorithm is parameterized as a *linear expansion of thresholds* (LET) and optimized using *Stein's unbiased risk estimate* (SURE). The proposed wavelet thresholding function is pointwise and depends on the coefficients of same location in the other channels, as well as on their parents in the coarser wavelet subband. A *nonredundant*, orthonormal, wavelet transform is first applied to the noisy data, followed by the (subband-dependent) vector-valued thresholding of individual *multichannel* wavelet coefficients which are finally brought back to the image domain by inverse wavelet transform. Extensive comparisons with the state-of-the-art multiresolution image denoising algorithms indicate that despite being *nonredundant*, our algorithm matches the quality of the best *redundant* approaches, while maintaining a high computational efficiency and a low CPU/memory consumption. An online Java demo illustrates these assertions.**

***Index Terms*—Color image denoising, interscale dependencies, multichannel image denoising, nonredundant transforms, orthonormal wavelet transforms (OWTs), Stein's unbiased risk estimate (SURE).**

I. Introduction

DENOISING has become an essential step in image analysis. Indeed, due to sensors imperfections, transmission channels defects, as well as physical constraints, noise deteriorates the quality of almost every acquired images. A considerable breakthrough has been achieved thanks to the development of new multiresolution tools such as the wavelet transform [1]–[3]. Its energy-compaction property has been shown to be particularly suitable to bring out the key signal informations from the noise.

Initially, multiresolution denoising algorithms were based on pointwise *wavelet thresholding*: its principle consists of setting to zero all the wavelet coefficients below a certain threshold value, while either keeping the remaining ones unchanged (*hard-thresholding*) or shrinking them by the threshold value (*soft-thresholding*, which was originally theorized by Donoho *et al.* [4]). Since then, a lot of work has been carried out to improve this simple, yet quite successful, pointwise approach. Intra- and interscale correlations have been successively introduced in more sophisticated estimators often derived in a Bayesian framework (initiated by [5]–[7]), assuming non-Gaussian [8] or generalized Laplacian priors [9], as well as scale mixtures of Gaussians [10], in order to model the statistics of the underlying noise-free signal. In this paper, we will indistinctly use the term *thresholding* to characterize all these wavelet-based denoisers.

In conjunction with the expansion of new wavelet estimators, some researchers have worked on improving the wavelet transform itself. Since the early—nonredundant—*orthonormal* wavelet transform (OWT), substantial improvements have been reached by using shift-invariant transformations with better directional selectivity [8]–[11]. Recently, we have proposed a general methodology, the "SURE-LET" paradigm, for building (using a *linear expansion of thresholds*: "LET" parameterization) and optimizing (using *Stein's unbiased risk estimate*: SURE principle) denoising algorithms adapted to any kind of linear transforms [12].

The new properties resulting from the use of often highly redundant transformations have been obtained at the expense of the loss of orthogonality, a substantially more intensive memory usage and a higher computational cost than that of the original OWT. The latter point becomes a major concern in image volume denoising and more generally in multichannel image denoising, in particular when the number of channels is large. For instance, even though the usual color image representations require no more than 3–4 channels (RGB, HSV, YUV, or CMYK descriptions), the computational cost is already quite large when shift-invariant (i.e., undecimated) transforms are involved. This is why, in this paper, we will only consider orthonormal (i.e., nonredundant) wavelet transforms for multichannel image denoising.

The easiest approach to denoising multichannel images is simply to apply an existing denoiser separately in each channel. However, this solution is far from being optimal, due to the presence of potentially strong common information between the various channels. There are then two main conceivable strategies: the first one consists of "decorrelating" the data based on some *a priori* model for the relation between the noiseless channels, and then separately apply a standard denoiser; the second one is to devise specific nonseparable multichannel denoising algorithms.

The first strategy has been exploited in color image denoising, where the data are denoised in more appropriate color representations than the standard red-green-blue (RGB) color space. Some satisfactory results have been obtained directly in the image domain using partial differential equations (PDE)/variational-based algorithms [13]–[15] in a chromaticity-brightness (CB) decomposition, or by promoting good continuation of hue

Manuscript received July 10, 2007; revised January 19, 2008. 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.

F. Luisier is with the Biomedical Imaging Group (BIG), Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland (e-mail: florian.luisier@epfl.ch).

T. Blu is with the Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong (e-mail: thierry.blu@m4x.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2008.919370

1057-7149/$25.00 © 2008 IEEE

transitions in the hue-saturation-value (HSV) representation of colors [16]. For wavelet thresholding algorithms, the color image denoising in the luminance-chrominance (YUV) color space has been shown to bring an often significant improvement over the standard RGB denoising [17], [18]. Quite recently, Lian *et al.* have even proposed a data-adaptive procedure for an optimal luminance/color-difference projection [17]. Combined with an efficient wavelet thresholding algorithm [19], their solution stands among the best state-of-the-art color image denoisers.

The second strategy has been quite early chosen by Sapiro *et al.*, who have adapted the anisotropic diffusion for multivalued images [20]. Later on, Blomgren *et al.* have extended the total variation (TV) method to vector-valued functions, and applied it to the restoration of color images [21]. In the wavelet community, specific multichannel algorithms have only quite recently been designed [22]–[24]. In particular, Scheunders has exploited the interchannel dependencies by summing products of wavelet coefficients from different channels to better isolate the noise and thus to derive an analytical noise probability density function (pdf), from which a thresholding value can be determined [25]. More recently, he proposed a multicomponent adaptation of Portilla *et al.*'s *BLS-GSM* by applying the Gaussian scale mixture model to each vector of multicomponent coefficients [26]; yet, his adaptation does not take into account neighboring wavelet coefficients, nor their interscale relations. Pižurica *et al.* have also adapted their original *ProbShrink* by including the interchannel dependencies in the definition of their local spatial activity indicator [9]. Finally, in the context of multiband satellite image denoising, Benazza-Benyahia *et al.* have proposed robust wavelet estimators based on the assumption that the noise-free wavelet coefficients follow a Bernoulli–Gaussian distribution [27]; the resulting estimator is then derived in a Bayesian framework as the *a posteriori* conditional mean. The parameters involved in their estimator are then optimized using Stein's unbiased risk estimate [28] (SURE). A recently submitted manuscript from Chaux with the same authors [11] goes a step further by generalizing the parameterization of the denoiser.

Here, we extend our original monochannel denoiser [29] to multichannel image denoising in an orthonormal wavelet representation. This requires a novel vector/matrix formulation of our theory which is shown to bring substantial denoising improvements by efficiently exploiting interchannel correlations. Indeed, we observe that, besides outperforming other nonredundant techniques, our algorithm even reaches the quality obtained by current multiresolution-based *redundant* methods. Moreover, the gain in CPU/memory consumption made it possible to work out a java version of the proposed algorithm (see our online demo [30]).

The paper is organized as follows. In the next section, we give the explicit expression of a multichannel MSE estimate for a vector-valued function of two statistically independent random vectors. In Section III, we present the generalization of the linear parametrization introduced in [29], as well as the corresponding SURE minimization for vector-valued functions, thus abiding by the SURE-LET methodology. We then recall the key steps which lead to our interscale predictor, and finally conclude the section by giving the expression of our interscale-interchannel wavelet thresholding. In Section IV, we demonstrate the competitiveness of our solution for color image denoising and more generally for multiband (*LandSat*) image denoising, through comprehensive experimentations on various test images for a wide range of input noise levels. In addition, we show that our algorithm is insensitive to the color representation (RGB or YUV), which is not the case of most other algorithms.

II. Multichannel SURE

In this paper, we consider N-pixel images with C channels—typically, $C = 3$ color channels for RGB images, but for biological images (fluorescence) or multiband satellite images, C might be much larger. We denote these multichannel images by a $C \times N$ matrix whose columns are the channel values of each pixel:

$$\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N] \quad \text{where} \quad \mathbf{x}_n = \begin{bmatrix} x_{n,1} \\ x_{n,2} \\ \vdots \\ x_{n,C} \end{bmatrix}.$$

These images are corrupted by an additive channel-wise Gaussian white noise[1] $\mathbf{b} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_N]$ of known $C \times C$ interchannel covariance matrix $\mathbf{R}$, i.e.,

$$\mathcal{E}\left\{\mathbf{b}_n \mathbf{b}_{n'}^{\mathrm{T}}\right\} = \mathbf{R}\,\delta_{n-n'}.$$

We denote the resulting noisy image by $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N]$ and we have

$$\mathbf{y} = \mathbf{x} + \mathbf{b}. \tag{1}$$

We want to stress that, in this paper, the original image, $\mathbf{x}$, will *never* be considered as a realization of some random process. The only source of randomness in our setting is the noise $\mathbf{b}$. As a consequence, the noisy image $\mathbf{y}$ is also random; yet, all the statistical considerations (typically, on the independence between wavelet subbands) will only be a consequence of the statistical properties of the noise itself.

Denoising the image $\mathbf{y}$ boils down to finding an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ as a function of $\mathbf{y}$ alone. We will evaluate the quality of the denoising by a classic *mean-squared error* (MSE) which can be expressed using Frobenius matrix norm

$$\begin{aligned} \mathrm{MSE} &= \frac{1}{CN}\|\hat{\mathbf{x}} - \mathbf{x}\|_F^2 \\ &= \frac{1}{CN}\mathrm{Tr}\{(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^{\mathrm{T}}\} \\ &= \frac{1}{CN}\sum_{n=1}^{N}\|\hat{\mathbf{x}}_n - \mathbf{x}_n\|^2. \end{aligned}$$

Retaining the strategy that was used in [29], we choose an estimate that involves a *critically sampled orthonormal wavelet transform* (OWT) applied to each channel (see Fig. 1 for the discrete wavelet decomposition of a RGB color image). Denoting the resulting wavelet images at scale $j \in [1, J]$ by a superscript, we thus have

$$\mathbf{y}^j = \mathbf{x}^j + \mathbf{b}^j. \tag{2}$$

[1]"Channel-wise" means here that the noise is Gaussian and white inside each channel.

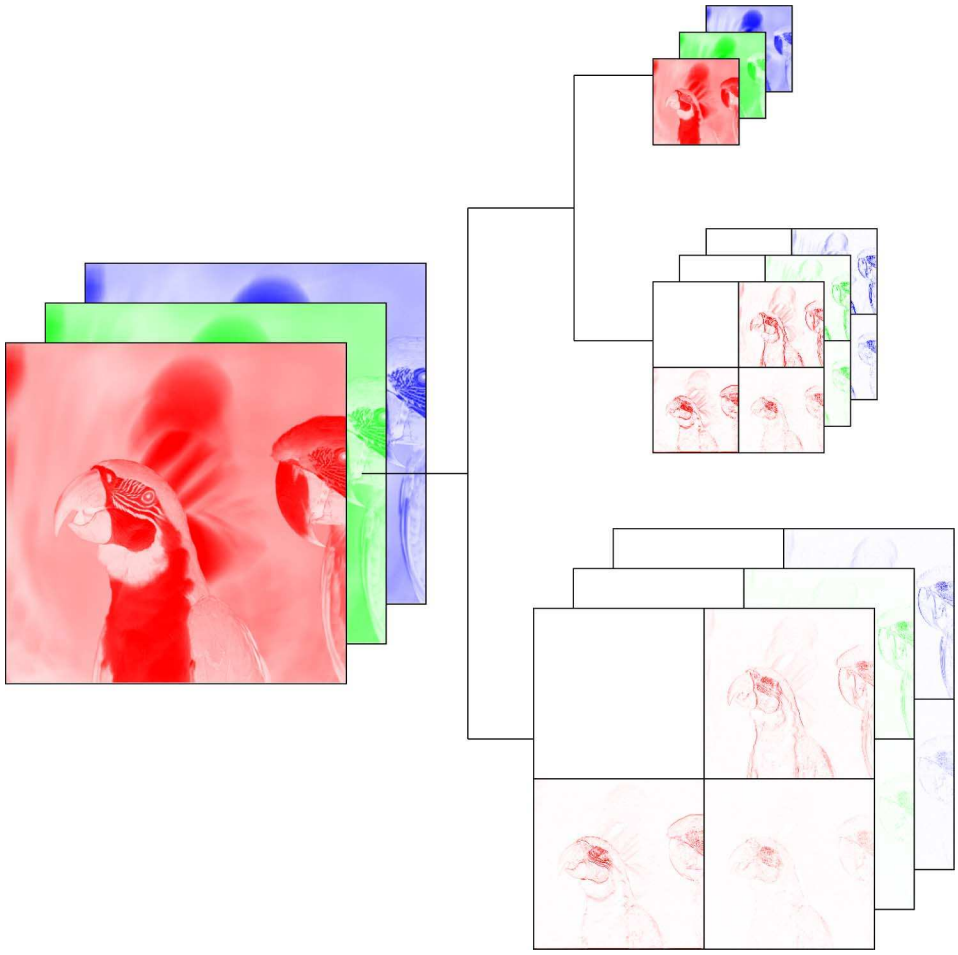

Fig. 1. Two iterations of a 2-D orthonormal wavelet transform applied to a RGB image.

Thanks to its linearity and orthonormality, the OWT enjoys two important conservation properties.

- *Noise statistics*—Given that the noise is white and Gaussian in the image domain, its wavelet coefficients are Gaussian as well, and are independent within and between the subbands. Moreover, the interchannel covariance matrix remains unchanged

$$\mathcal{E}\left\{\mathbf{b}_n^j \mathbf{b}_{n'}^{j'\,\mathrm{T}}\right\} = \mathbf{R}\,\delta_{n-n'}\,\delta_{j-j'}.$$

- *Mean-Squared Error*—The image domain MSE and the subband MSEs are related through

$$\underbrace{\|\hat{\mathbf{x}}-\mathbf{x}\|_F^2}_{CN\times\mathrm{MSE}} = \sum_{j=1}^{J} \underbrace{\|\hat{\mathbf{x}}^j-\mathbf{x}^j\|_F^2}_{CN^j\times\mathrm{MSE}^j} \tag{3}$$

where N^j is the number of samples in subband j.

These two key properties make it particularly attractive to perform independent processing $\boldsymbol{\theta}^j$ in each individual noisy wavelet subbands $\mathbf{y}^j$. To take advantage of both the interchannels similarities and the interscale consistencies that may be intrinsic to $\mathbf{x}$, the thresholding function $\boldsymbol{\theta}^j$ will also involve an interscale predictor $\mathbf{p}^j$ built using subbands $j' \geq j+1$ as detailed in Section III.A. We will, however, remain "pointwise" in the sense that the estimate $\hat{\mathbf{x}}_n^j$ of the nth pixel of subband j will depend only on $\mathbf{x}_n^j$ and $\mathbf{p}_n^j$, without taking their neighbours into account. It is essential to notice that, because of the statistical independence between subbands of different iteration depth, $\mathbf{y}^j$ and $\mathbf{p}^j$ will also be statistically independent.

From now on, we will drop the subband exponent j when no ambiguity is likely to arise. More abstractly, we are thus going to consider the denoising of a multichannel (subband) image $\mathbf{y}=\mathbf{x}+\mathbf{b}$, given an independent prediction (parent) $\mathbf{p}$, by using a $\mathbb{R}^C \times \mathbb{R}^C \to \mathbb{R}^C$ function $\boldsymbol{\theta}$ relating the coefficients of $\mathbf{y}$ and $\mathbf{p}$ to the coefficients of the estimate $\hat{\mathbf{x}}$ through

$$\hat{\mathbf{x}}_n = \boldsymbol{\theta}(\mathbf{y}_n, \mathbf{p}_n), \quad \text{for } n = 1, 2, \ldots, N. \tag{4}$$

Ideally, our aim would be to choose $\boldsymbol{\theta}$ so that it minimizes the MSE defined in (3). A difficulty is that we do not have access to the original noise-free image $\mathbf{x}$. Fortunately, we can rely on an adapted version of Stein's unbiased risk estimate (SURE) [28] to accurately estimate this actual MSE, as shown in the following theorem.

Theorem 1: Assume that $\boldsymbol{\theta}(\cdot,\cdot)$ is (weakly) differentiable w.r.t. its first variable and that it does not explode at infinity; typically, such that $\|\boldsymbol{\theta}(\mathbf{u},\mathbf{v})\| \leq \mathrm{const}(\mathbf{v}) \times \exp(\|\mathbf{u}\|^2/(2s^2))$ where $s > \sigma$. Then, if the estimate $\hat{\mathbf{x}}$ is built according to (4), the following random variable:

$$\epsilon = \frac{1}{CN}\sum_{n=1}^{N} \|\boldsymbol{\theta}(\mathbf{y}_n,\mathbf{p}_n) - \mathbf{y}_n\|^2 + \frac{2}{CN}\sum_{n=1}^{N} \mathrm{Tr}\{\mathbf{R}^{\mathrm{T}}\nabla_1\boldsymbol{\theta}(\mathbf{y}_n,\mathbf{p}_n)\} - \frac{1}{C}\mathrm{Tr}\{\mathbf{R}\} \tag{5}$$

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

$$\mathcal{E}\{\epsilon\} = \frac{1}{CN}\mathcal{E}\left\{\|\hat{\mathbf{x}} - \mathbf{x}\|_F^2\right\}.$$

Here, we have denoted by $\nabla_1\boldsymbol{\theta}$ the matrix containing the partial derivatives of the components of $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_C]^{\mathrm{T}}$ with respect to its *first* variable

$$\nabla_1\boldsymbol{\theta}(\mathbf{u},\mathbf{v}) = \begin{bmatrix} \frac{\partial\theta_1(\mathbf{u},\mathbf{v})}{\partial u_1} & \frac{\partial\theta_2(\mathbf{u},\mathbf{v})}{\partial u_1} & \cdots & \frac{\partial\theta_C(\mathbf{u},\mathbf{v})}{\partial u_1} \\ \frac{\partial\theta_1(\mathbf{u},\mathbf{v})}{\partial u_2} & \frac{\partial\theta_2(\mathbf{u},\mathbf{v})}{\partial u_2} & \cdots & \frac{\partial\theta_C(\mathbf{u},\mathbf{v})}{\partial u_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial\theta_1(\mathbf{u},\mathbf{v})}{\partial u_C} & \frac{\partial\theta_2(\mathbf{u},\mathbf{v})}{\partial u_C} & \cdots & \frac{\partial\theta_C(\mathbf{u},\mathbf{v})}{\partial u_C} \end{bmatrix}.$$

Proof: Note that, because $\mathbf{p}_n$ and $\mathbf{y}_n$ are independent, we may simply prove the result without considering $\mathbf{p}_n$ to be random. We can then develop the squared error between $\mathbf{x}_n = \mathbf{y}_n - \mathbf{b}_n$ and its estimate $\hat{\mathbf{x}}_n = \boldsymbol{\theta}(\mathbf{y}_n, \mathbf{p}_n)$ as

$$\begin{aligned} &\mathcal{E}\{\|\boldsymbol{\theta}(\mathbf{y}_n,\mathbf{p}_n) - \mathbf{x}_n\|^2\} \\ &\quad = \mathcal{E}\{\|\boldsymbol{\theta}(\mathbf{y}_n,\mathbf{p}_n)\|^2\} \\ &\qquad - 2\mathcal{E}\{\boldsymbol{\theta}(\mathbf{y}_n,\mathbf{p}_n)^{\mathrm{T}}(\mathbf{y}_n - \mathbf{b}_n)\} + \|\mathbf{x}_n\|^2 \\ &\quad = \mathcal{E}\{\|\boldsymbol{\theta}(\mathbf{y}_n,\mathbf{p}_n) - \mathbf{y}_n\|^2\} \\ &\qquad + 2\mathcal{E}\{\boldsymbol{\theta}(\mathbf{y}_n,\mathbf{p}_n)^{\mathrm{T}}\mathbf{b}_n\} + \|\mathbf{x}_n\|^2 - \mathcal{E}\{\|\mathbf{y}_n\|^2\}. \end{aligned} \quad (6)$$

Now we use the fact that a zero-mean multivariate Gaussian probability density function $q(\mathbf{b}_n)$ with covariance matrix $\mathbf{R}$ satisfies $\mathbf{b}_n q(\mathbf{b}_n) = -\mathbf{R}\nabla q(\mathbf{b}_n)$ to evaluate $\mathcal{E}\left\{\boldsymbol{\theta}(\mathbf{y}_n,\mathbf{p}_n)^{\mathrm{T}}\mathbf{b}_n\right\}$

$$\begin{aligned} &\mathcal{E}\left\{\boldsymbol{\theta}(\mathbf{y}_n,\mathbf{p}_n)^{\mathrm{T}}\mathbf{b}_n\right\} \\ &\quad = \int_{\mathbb{R}^C} \boldsymbol{\theta}(\mathbf{x}_n + \mathbf{b}_n, \mathbf{p}_n)^{\mathrm{T}}\mathbf{b}_n q(\mathbf{b}_n)\,\mathrm{d}^C\mathbf{b}_n \\ &\quad = -\int_{\mathbb{R}^C} \boldsymbol{\theta}(\mathbf{x}_n + \mathbf{b}_n, \mathbf{p}_n)^{\mathrm{T}}\mathbf{R}\,\nabla q(\mathbf{b}_n)\,\mathrm{d}^C\mathbf{b}_n \\ &\quad = \int_{\mathbb{R}^C} \mathrm{div}_{\mathbf{b}_n}\{\mathbf{R}^{\mathrm{T}}\boldsymbol{\theta}(\mathbf{x}_n + \mathbf{b}_n, \mathbf{p}_n)\}\, q(\mathbf{b}_n)\,\mathrm{d}^C\mathbf{b}_n \quad \text{(by parts)} \\ &\quad = \int_{\mathbb{R}^C} \mathrm{Tr}\{\mathbf{R}^{\mathrm{T}}\nabla_1\boldsymbol{\theta}(\mathbf{x}_n + \mathbf{b}_n, \mathbf{p}_n)\} q(\mathbf{b}_n)\,\mathrm{d}^C\mathbf{b}_n \\ &\quad = \mathcal{E}\{\mathrm{Tr}\{\mathbf{R}^{\mathrm{T}}\nabla_1\boldsymbol{\theta}(\mathbf{y}_n,\mathbf{p}_n)\}\}. \end{aligned}$$

Using the above relation, as well as the standard result $\mathcal{E}\{\|\mathbf{y}_n\|^2\} = \|\mathbf{x}_n\|^2 + \mathrm{Tr}\{\mathbf{R}\}$ into (6), we get the desired result. ∎

The variance of the above MSE estimate ϵ depends on the number of C-channel pixels N. Since in multichannel image denoising the data are usually quite huge (typically, $256\times256\times C$), ϵ can be reliably used as the actual MSE. In particular, its minimization will closely tend to the minimization of the actual mean squared error between the processed image and the—unknown—noise-free image.

III. Algorithm

In this section, we show how to adapt the monochannel SURE-based denoiser presented in [29] to multichannel image denoising. The two fundamental ingredients of our original approach remain the same.

1) The denoising function $\boldsymbol{\theta}$ will be built as a *linear expansion* of simple—possibly nonlinear—thresholding functions $\boldsymbol{\theta}_k$

$$\begin{aligned} \boldsymbol{\theta}(\mathbf{y}_n,\mathbf{p}_n) &= \sum_{k=1}^{K} \mathbf{a}_k^{\mathrm{T}}\boldsymbol{\theta}_k(\mathbf{y}_n,\mathbf{p}_n) \\ &= \underbrace{[\mathbf{a}_1^{\mathrm{T}}, \mathbf{a}_2^{\mathrm{T}}, \ldots, \mathbf{a}_K^{\mathrm{T}}]}_{\mathbf{A}^{\mathrm{T}}} \\ &\quad \times \underbrace{\begin{bmatrix} \boldsymbol{\theta}_1(\mathbf{y}_n,\mathbf{p}_n) \\ \boldsymbol{\theta}_2(\mathbf{y}_n,\mathbf{p}_n) \\ \vdots \\ \boldsymbol{\theta}_K(\mathbf{y}_n,\mathbf{p}_n) \end{bmatrix}}_{\boldsymbol{\Theta}(\mathbf{y}_n,\mathbf{p}_n)}. \end{aligned} \quad (7)$$

Here, $\boldsymbol{\Theta}(\mathbf{y}_n,\mathbf{p}_n)$ is a $KC\times1$ vector, the $\mathbf{a}_k$ are $C\times C$ matrices and $\mathbf{A}$ is a $KC\times C$ matrix. In this formalism, the gradient of $\boldsymbol{\theta}(\mathbf{y}_n,\mathbf{p}_n)$ with respect to the first variable can be expressed as

$$\nabla_1\boldsymbol{\theta}(\mathbf{y}_n,\mathbf{p}_n) = \nabla_1\boldsymbol{\Theta}(\mathbf{y}_n,\mathbf{p}_n)\,\mathbf{A}.$$

2) The MSE estimate ϵ is *quadratic* in $\mathbf{A}$, as follows:

$$\begin{aligned} \epsilon &= \frac{1}{CN}\sum_{n=1}^{N}\|\mathbf{A}^{\mathrm{T}}\boldsymbol{\Theta}(\mathbf{y}_n,\mathbf{p}_n) - \mathbf{y}_n\|^2 \\ &\quad + \frac{2}{CN}\sum_{n=1}^{N}\mathrm{Tr}\{\mathbf{R}^{\mathrm{T}}\nabla_1\boldsymbol{\Theta}(\mathbf{y}_n,\mathbf{p}_n)\,\mathbf{A}\} \\ &\quad - \frac{1}{C}\mathrm{Tr}\{\mathbf{R}\} \\ &= \frac{1}{CN}\sum_{n=1}^{N}\mathrm{Tr}\{(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Theta}(\mathbf{y}_n,\mathbf{p}_n) - \mathbf{y}_n) \\ &\qquad \times (\mathbf{A}^{\mathrm{T}}\boldsymbol{\Theta}(\mathbf{y}_n,\mathbf{p}_n) - \mathbf{y}_n)^{\mathrm{T}}\} \\ &\quad + \frac{2}{CN}\sum_{n=1}^{N}\mathrm{Tr}\{\mathbf{R}^{\mathrm{T}}\nabla_1\boldsymbol{\Theta}(\mathbf{y}_n,\mathbf{p}_n)\,\mathbf{A}\} \\ &\quad - \frac{1}{C}\mathrm{Tr}\{\mathbf{R}\} \\ &= \frac{1}{CN}\mathrm{Tr}\{\mathbf{A}^{\mathrm{T}}\mathbf{M}\mathbf{A} - 2\mathbf{B}^{\mathrm{T}}\mathbf{A}\} \\ &\quad + \frac{1}{CN}\sum_{n=1}^{N}\mathrm{Tr}\left\{\mathbf{y}_n\mathbf{y}_n^{\mathrm{T}}\right\} - \frac{1}{C}\mathrm{Tr}\{\mathbf{R}\} \end{aligned} \quad (8)$$

where we have defined

$$\begin{aligned} \mathbf{M} &= \sum_{n=1}^{N}\boldsymbol{\Theta}(\mathbf{y}_n,\mathbf{p}_n)\boldsymbol{\Theta}(\mathbf{y}_n,\mathbf{p}_n)^{\mathrm{T}} \\ \mathbf{B} &= \sum_{n=1}^{N}\left(\boldsymbol{\Theta}(\mathbf{y}_n,\mathbf{p}_n)\mathbf{y}_n^{\mathrm{T}} - (\nabla_1\boldsymbol{\Theta}(\mathbf{y}_n,\mathbf{p}_n))^{\mathrm{T}}\mathbf{R}\right). \end{aligned}$$

Finally, the minimization of (8) with respect to $\mathbf{A}$ boils down to the following linear system of equations:

$$\mathbf{A}_{\mathrm{opt}} = \mathbf{M}^{-1}\mathbf{B}. \quad (9)$$

Notice that if $\mathbf{M}$ is not a full rank matrix, we can simply take its pseudo-inverse to choose among the admissible solutions. In fact, rank deficiency indicates that $\boldsymbol{\Theta}$ is *over-parame-*

terized, and, consequently, that fewer parameters yield the same minimal MSE estimate. Obviously, it is advisable to have the smallest possible number of parameters, in order for $\mathbf{M}$ to be nonsingular and in order to make the standard deviation of ϵ as small as possible, in such a way that any of its realizations is close to the actual MSE.

It may also be interesting to restrict the number of degrees of freedom of the coefficient matrices $\mathbf{a}_k$ and, in exchange, increase the actual number K of these coefficients: typically, one may choose $\mathbf{a}_k$ to be of the form $\mathbf{u}_k\mathbf{v}_k^{\mathrm{T}}$ where $\mathbf{v}_k$ is some *known* $C \times 1$ vector, while $\mathbf{u}_k$ is an *unknown* $C \times 1$ vector. This means that the $KC \times C$ matrix $\mathbf{A}$ lives in some linear subspace of dimension $D < KC^2$ spanned by, say, a basis of $KC \times C$ matrices $\{\mathbf{E}_d\}_{d=1,2,\ldots,D}$. Once again, minimizing (8) with respect to all the degrees of freedom of $\mathbf{A}$ leads to a linear system of equations

$$\mathrm{Tr}\left\{\mathbf{E}_d^{\mathrm{T}}(\mathbf{M}\mathbf{A}_{\mathrm{opt}} - \mathbf{B})\right\} = 0 \quad \text{for } d = 1, 2, \ldots D$$

from which the D (linear) degrees of freedom of $\mathbf{A}_{\mathrm{opt}}$ can be computed.

A. Interscale Predictor

In [29], we proposed to build an interscale predictor out of the low-pass subband at the same scale, contrary to the custom in the literature which consists of expanding the parent subband—i.e., the subband at the next coarser scale—by a factor of two. We showed that our new construction ensures a perfect feature alignment between the current subband and its interscale predictor. For this reason, we are going to use this approach to build the parent estimate $\mathbf{p}_n$ to compute the estimate $\hat{\mathbf{x}}_n$ according to (4).

In the monochannel case of [29], the interscale predictor of the wavelet subband at scale j is obtained by applying a suitable *group delay compensation* filter (GDC) to the low-pass subband at the same scale. We showed that group delay compensation ensures the alignment between the wavelet subband and the computed parent subband. In one dimension, the GDC filter takes the following general form:

$$W(z^2) = G(z^{-1})G(-z^{-1})(1 + \lambda z^{-2})R(z^2) \qquad (10)$$

where $\lambda = \pm 1$; $R(z) = R(z^{-1})$ is an arbitrary zero-phase filter; $G(z) = -z^{-1}H(-z^{-1})$ is the wavelet filter given that $H(z)$ is the—orthonormal—scaling filter. The absolute value of each resulting subband is then smoothed by a normalized Gaussian kernel (see Fig. 2).

In the case of symmetric or nearly symmetric (around n_0) scaling filters, the shortest normalized GDC filter is the discrete gradient operator

$$W(z) = \frac{1}{\sqrt{2}} z^{-n_0}(1 - z). \qquad (11)$$

For multichannels images, we simply apply the same procedure separately in every channel, which yields the multichannel interscale predictor $\mathbf{p}_n$.

B. New Interscale-Interchannel Thresholding Function

We propose now a natural vectorization of the thresholding function presented in [29] by taking into account the strong similarities that may occur between the various channels. More specifically, we build this thresholding function according to the expression (7) with $K = 4$ in which each $\boldsymbol{\theta}_k$ denoises a particular zone of the wavelet subband, characterized by large or small values of the parents/wavelet coefficients. This zone selection makes use of a "trigger" function $\gamma(x)$ which is essentially unity for small values of x, and vanishes for large values. We have chosen the following expression:

$$\gamma(x) = \exp\left(-\frac{|x|}{12C}\right). \qquad (12)$$

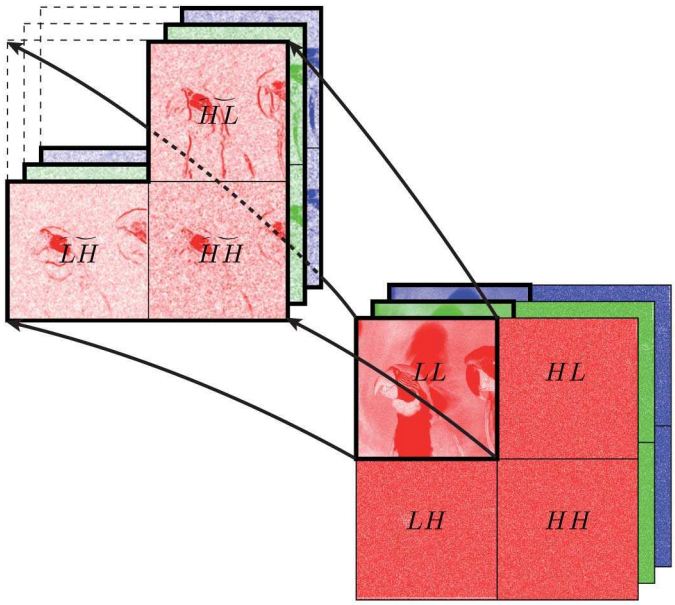

Fig. 2. Interscale predictor for a 2-D OWT. At each level of decomposition, the low-pass subband LL is used to build predictors $\widetilde{HL}$, $\widetilde{LH}$, and $\widetilde{HH}$ of each of the three high-pass subbands HL, LH, and HH.

Fig. 3. Test images used in the experiments, referred to as *Image 1* to *Image 8* (numbered from left to right and top to bottom).

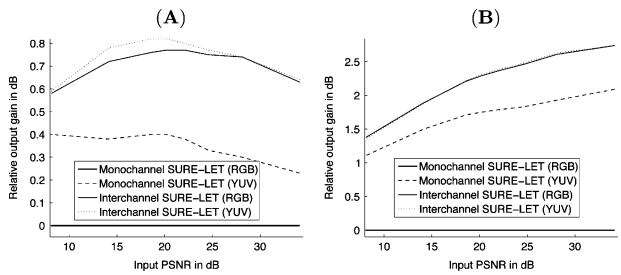

Fig. 4. PSNR improvements brought by our interchannel strategy, compared to the worst case (Monochannel SURE-LET in RGB). (A) *Image 1*. (B) *Image 7*.

TABLE I
COMPARISON OF COLOR DENOISING ALGORITHMS (SAME NOISE LEVEL IN EACH RGB CHANNEL)

$\sigma_R = \sigma_G = \sigma_B$	5	10	15	20	25	30	50	100	5	10	15	20	25	30	50	100
Input PSNR [dB]	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	*Image 1* 256 × 256								*Image 2* 256 × 256							
OWT ProbShrink-MB [9]	36.65	33.18	31.32	29.98	28.94	28.07	25.61	22.69	34.37	30.01	28.17	26.93	26.00	25.25	23.24	20.72
OWT SURE-LET	**37.91**	**34.46**	**32.60**	31.29	30.25	29.39	26.95	23.73	**35.40**	**31.22**	**29.24**	**27.98**	**27.07**	**26.34**	**24.38**	**21.76**
UWT ProbShrink-MB [9]	37.69	34.22	32.30	30.94	29.96	29.06	26.55	23.72	35.31	31.21	29.22	27.92	26.99	26.21	24.17	21.65
BLS-GSM [10]	37.57	34.20	32.52	**31.31**	**30.34**	**29.52**	**27.21**	**24.12**	35.35	31.01	29.09	27.91	27.04	26.34	24.37	21.74
Method	*Image 3* 512 × 512								*Image 4* 512 × 512							
OWT ProbShrink-MB [9]	36.37	33.45	31.78	30.59	29.67	28.92	26.88	24.28	35.48	32.49	31.02	29.98	29.14	28.41	26.20	23.55
OWT SURE-LET	**37.80**	**34.64**	**33.02**	**31.90**	**31.04**	**30.33**	**28.35**	**25.66**	**36.62**	**33.35**	31.79	30.72	29.89	29.19	27.16	24.48
UWT ProbShrink-MB [9]	37.46	34.42	32.69	31.47	30.61	29.83	27.76	25.03	36.33	**33.35**	31.81	30.74	29.96	29.20	26.85	24.28
BLS-GSM [10]	37.29	34.45	32.90	31.78	30.89	30.15	28.09	25.40	36.34	33.26	**31.89**	**30.92**	**30.13**	**29.46**	**27.47**	**24.73**
Method	*Image 5* 512 × 512								*Image 6* 512 × 512							
OWT ProbShrink-MB [9]	33.86	28.90	26.44	24.87	23.73	22.89	20.94	19.31	37.58	34.03	32.01	30.64	29.29	28.24	25.90	23.30
OWT SURE-LET	**35.12**	**30.49**	**28.15**	**26.64**	**25.55**	**24.71**	**22.59**	**20.37**	**39.11**	**35.70**	**33.71**	**32.29**	**31.19**	**30.29**	**27.77**	**24.77**
UWT ProbShrink-MB [9]	34.83	30.15	27.72	26.17	25.04	24.16	21.98	19.81	38.78	35.23	33.20	31.80	30.77	29.81	26.87	23.97
BLS-GSM [10]	35.01	30.13	27.66	26.08	24.95	24.07	21.92	19.89	38.40	35.01	33.09	31.74	30.69	29.84	27.47	24.45
Method	*Image 7* 512 × 512								*Image 8* 512 × 512							
OWT ProbShrink-MB [9]	35.47	31.25	29.15	27.76	26.70	25.87	23.77	21.44	39.47	35.97	33.87	32.05	30.85	29.98	27.76	25.00
OWT SURE-LET	**38.69**	**34.24**	**31.87**	**30.29**	**29.10**	**28.15**	**25.63**	**22.72**	**41.05**	**37.56**	**35.49**	**34.00**	**32.84**	**31.88**	**29.26**	26.11
UWT ProbShrink-MB [9]	37.05	32.64	30.36	28.88	27.80	26.87	24.58	21.91	40.49	36.92	34.82	33.36	32.25	31.24	28.61	25.55
BLS-GSM [10]	36.36	32.17	30.14	28.86	27.92	27.18	25.13	22.50	40.16	37.03	35.11	33.71	32.62	31.72	29.24	**26.18**

Notes: 1. *UWT ProbShrink-MB* and *BLS-GSM* are applied in a *highly redundant* representation, whereas *OWT ProbShrink-MB* and *OWT SURE-LET* only use a *non-redundant* (orthonormal) transform.
2. Output PSNRs have been averaged over ten noise realizations.

The interscale predictor $\mathbf{p}$ will then be used in order to smoothly discriminate between high-SNR and low-SNR wavelet coefficients, which finally leads to the following interscale-interchannel thresholding function:

$$\begin{aligned}\boldsymbol{\theta}(\mathbf{y}_n, \mathbf{p}_n) &= \underbrace{\gamma\left(\mathbf{p}_n^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{p}_n\right)\gamma\left(\mathbf{y}_n^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}_n\right)}_{\text{small parents and small coefficients}}\ \mathbf{a}_1^{\mathrm{T}}\mathbf{y}_n \\ &+ \underbrace{\left(1-\gamma\left(\mathbf{p}_n^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{p}_n\right)\right)\gamma\left(\mathbf{y}_n^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}_n\right)}_{\text{large parents and small coefficients}}\ \mathbf{a}_2^{\mathrm{T}}\mathbf{y}_n \\ &+ \underbrace{\gamma\left(\mathbf{p}_n^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{p}_n\right)\left(1-\gamma\left(\mathbf{y}_n^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}_n\right)\right)}_{\text{small parents and large coefficients}}\ \mathbf{a}_3^{\mathrm{T}}\mathbf{y}_n \\ &+ \underbrace{\left(1-\gamma\left(\mathbf{p}_n^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{p}_n\right)\right)\left(1-\gamma\left(\mathbf{y}_n^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}_n\right)\right)}_{\text{large parents and large coefficients}}\ \mathbf{a}_4^{\mathrm{T}}\mathbf{y}_n \end{aligned} \tag{13}$$

where $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, and $\mathbf{a}_4$ are $C \times C$ matrices, leading to an overall number of $4C^2$ parameters.

In the following tests, we have retained this expression because of its simplicity. However, we have observed that by increasing K from 4 to $4C$ (by increasing the number of zones, e.g., by distinguishing between parents in the same channel from parents in other channels) and decreasing the number of degrees of freedom of the coefficients $\mathbf{a}_k$ from $C \times C$ full-rank matrices to $C \times C$ matrices having nonzero elements in a single column (the overall number of parameters thus remains $4C^2$) yields often better denoising results that may in some cases reach up to +0.3 dB.

IV. EXPERIMENTS

A. Color Image Denoising

Color spaces usually consist of $C = 3$ channels and we mostly consider red-green-blue (RGB) representations here. In order to demonstrate the performance of our approach, we assume that the interchannel noise covariance matrix is given by

$$\mathbf{R} = \begin{bmatrix} \sigma_R^2 & 0 & 0 \\ 0 & \sigma_G^2 & 0 \\ 0 & 0 & \sigma_B^2 \end{bmatrix}.$$

This assumption implies that, in other color spaces, there will usually be noise correlations between the color channels. As an illustration, suppose that we want to perform the denoising in the luminance-chrominance space YUV. An image $\tilde{\mathbf{y}}$ in YUV is obtained form an original RGB image $\mathbf{y}$ through the following linear transformation:

$$\tilde{\mathbf{y}} = \underbrace{\begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.1 \end{bmatrix}}_{\mathbf{S}} \mathbf{y}. \tag{14}$$

Then, the noise covariance matrix in the YUV color space becomes $\tilde{\mathbf{R}} = \mathbf{S}\mathbf{R}\mathbf{S}^{\mathrm{T}}$, and the MSE estimate in the YUV color space is finally obtained by replacing $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{R}$ by, respectively, $\tilde{\mathbf{x}} = \mathbf{S}\mathbf{x}$, $\tilde{\mathbf{y}} = \mathbf{S}\mathbf{y}$, and $\tilde{\mathbf{R}}$ in the expression of the SURE (5).

All the experiments of this section have been carried out on $N = 256 \times 256$ and $N = 512 \times 512$ RGB test images from the set presented in Fig. 3. We have applied our interscale-interchannel thresholding algorithm after four or five decomposition levels (depending on the size of the image: 256^2 or 512^2) of an orthonormal wavelet transform (OWT) using the standard Daubechies symlets with eight vanishing moments (*sym8*

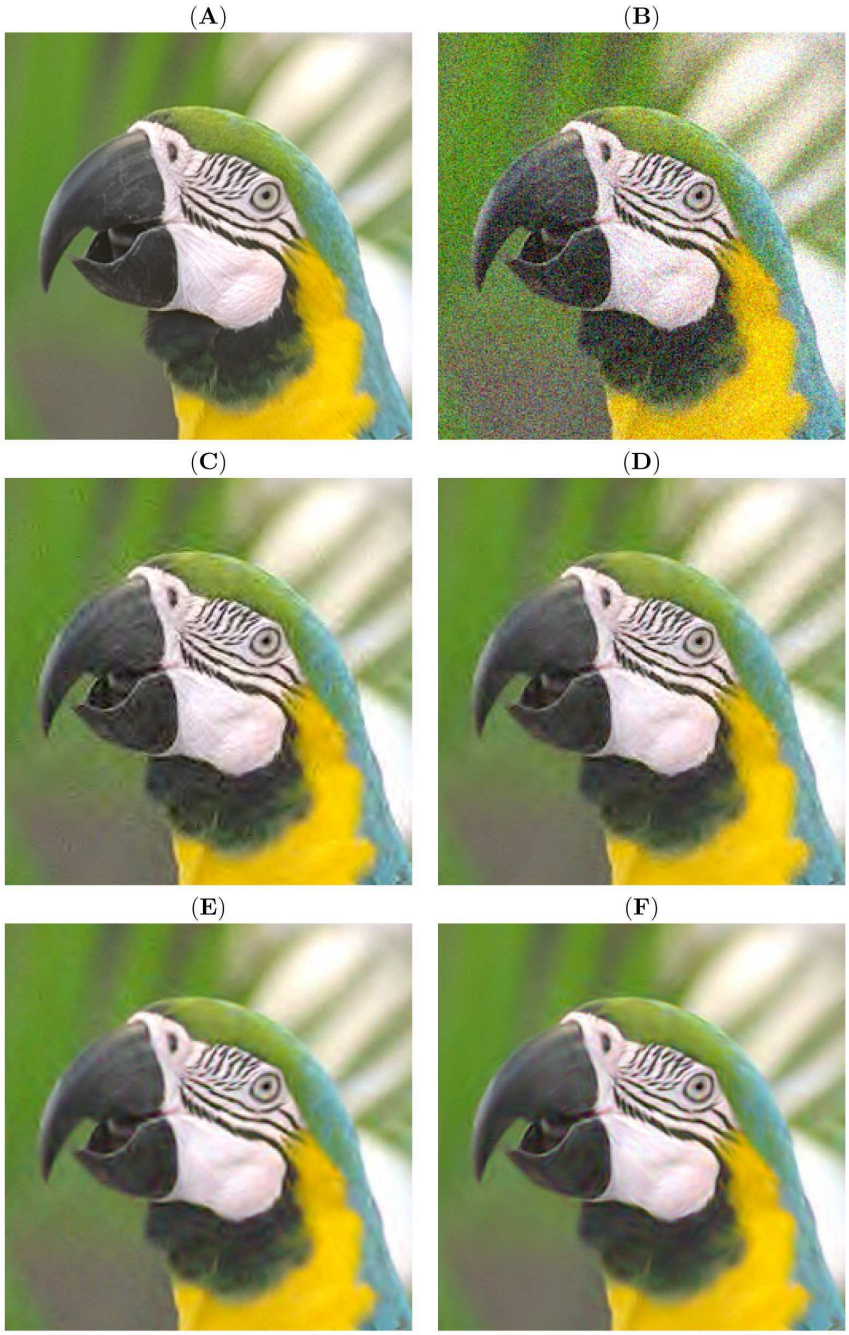

Fig. 5. (A) Part of the noise-free *Image 8*. (B) Part of the noisy *Image 8*: PSNR = 20.17 dB. (C) Result of the *ProbShrink-MB* (OWT *sym8*): PSNR = 30.88 dB. (D) Result of our *SURE-LET* (OWT *sym8*): PSNR = 32.83 dB. (E) Result of the *ProbShrink-MB* (UWT *sym8*): PSNR = 32.22 dB. (F) Result of the *BLS-GSM* (full steerable pyramid): PSNR = 32.60 dB.

in Matlab). The denoising performances are measured in terms of peak signal-to-noise ratio (PSNR) defined as

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{255^2}{\mathrm{MSE}}\right)\ \mathrm{dB}. \tag{15}$$

1) Interchannel Versus Independent Monochannel Thresholding: Before comparing our results with some of the state-of-the-art denoising procedures, we first want to evaluate the improvements brought by the integration of interchannel dependencies. In Fig. 4, we compare our interscale-interchannel thresholding function (13) with the interscale thresholding defined in [29] applied separately in each channels, both in the standard RGB color space and in the luminance-chrominance space YUV.

As can be observed, the integration of interchannel dependencies improves the denoising performance considerably, both in the RGB color space (more than +1 dB) and in the more "decorrelated" YUV space (around +0.5 dB). Note that these improvements become even more pronounced (around +1.5–2 dB) when the power of the noise is different in each channels.

Remarkably and contrary to the other algorithms that have been published previously, our results are quite insensitive to the color representation (variations of ±0.1 dB). Indeed, the pa-

TABLE II
COMPARISON OF COLOR DENOISING ALGORITHMS (DIFFERENT NOISE LEVEL IN EACH RGB CHANNEL)

$\sigma_R = 38.25, \sigma_G = 25.50, \sigma_B = 12.75$								
Input PSNR: 19.33 [dB]								
Method	***Image 1***	***Image 2***	***Image 3***	***Image 4***	***Image 5***	***Image 6***	***Image 7***	***Image 8***
OWT uHMT-OCP [17]	N/A	N/A	N/A	N/A	N/A	N/A	29.16	31.46
OWT SURE-LET	**30.63**	**27.19**	**31.41**	29.93	**26.12**	**31.34**	**30.09**	**33.17**
UWT ProbShrink-YUV [18]	29.53	26.55	30.47	29.37	24.93	29.95	28.25	32.03
BLS-GSM [10]	30.47	27.18	30.91	**30.13**	25.10	30.45	27.83	32.45

Notes: 1. *UWT ProbShrink-YUV* and *BLS-GSM* are applied in a *highly redundant* representation, whereas the *uHMT-OCP* and our *SURE-LET* only use a *non-redundant* (orthonormal) transform.
2. Output PSNRs have been averaged over ten noise realizations.

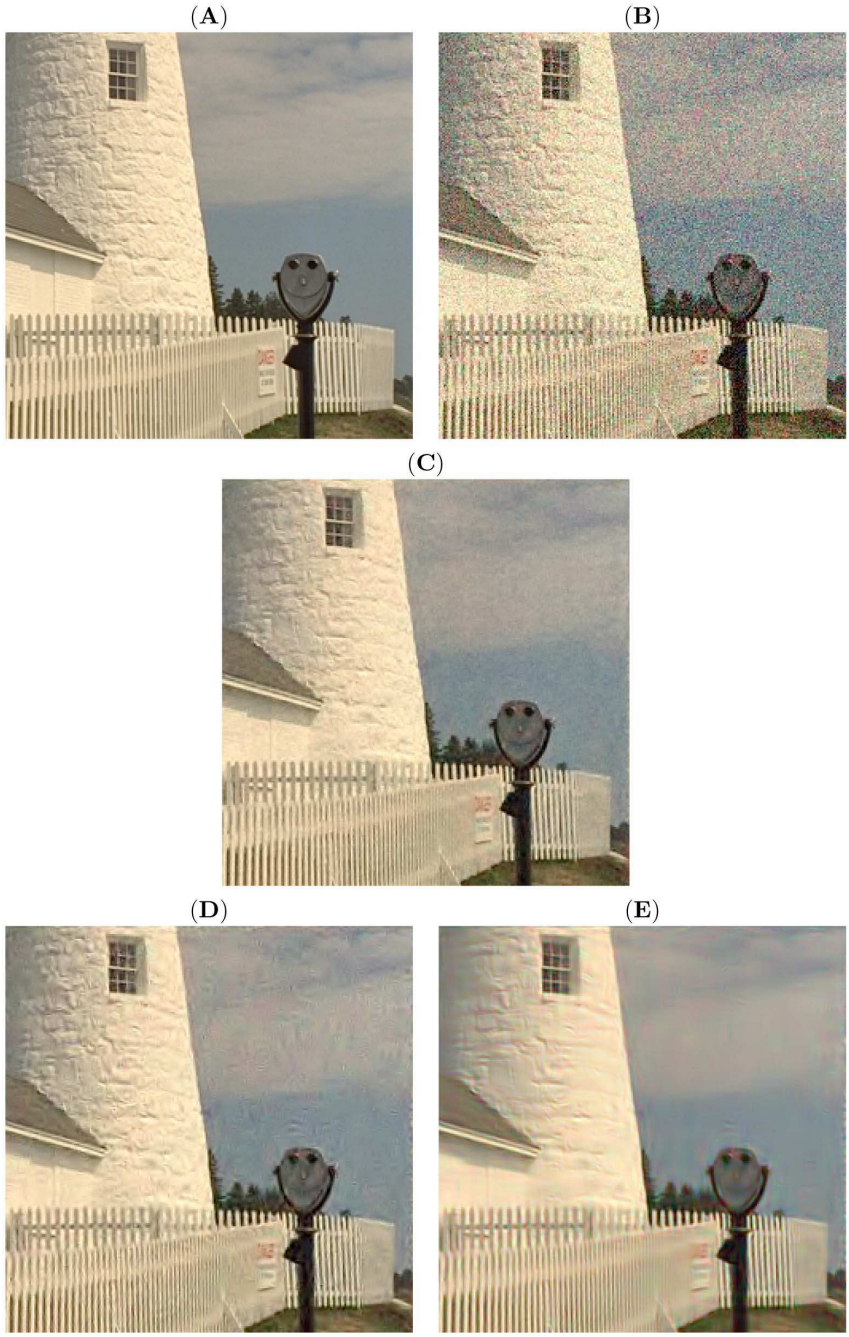

Fig. 6. (A) Part of the noise-free *Image 7*. (B) Part of the noisy *Image 7*: PSNR $= 19.33$ dB ($\sigma_R = 38.25$, $\sigma_G = 25.50$ and $\sigma_B = 12.75$). (C) Result of our *SURE-LET* (OWT *sym8*): PSNR $= 30.05$ dB. (D) Result of the *ProbShrink-YUV* (UWT *sym8*): PSNR $= 28.27$ dB. (E) Result of the *BLS-GSM* (full steerable pyramid): PSNR $= 27.83$ dB.

rameters $\mathbf{a}_k$ of (13) can be understood as statistically optimized[2] linear color space transformations in each wavelet zone. From now on, we will thus apply our algorithm in the RGB color space only.

2) Comparisons: We have chosen to compare our method with two state-of-the-art multiresolution-based denoising algorithms.

- Pižurica *et al.*'s *ProbShrink-MB* [9], which is a multiband extension of the original grayscale denoiser of the same authors. For color image denoising, it has to be applied in the standard RGB representation, and for equal noise variance in each channels. We have applied this algorithm with a nonredundant orthonormal wavelet transform, as well as with the—highly redundant—*undecimated* wavelet tranform (UWT) using the code of the authors[3] with their suggested parameters; we have considered the same number of decomposition levels and the same wavelet (*sym8*) as with our method. Since this algorithm has been shown in [9], [18] to favorably compare with the multiband wavelet thresholding described in [25], as well as with the vector-based linear minimum mean squared error estimator proposed in [31], it constitutes a good reference for evaluating our solution.
- Portilla *et al.*'s *BLS-GSM* [10]: although this algorithm has not been designed for multichannel denoising, this is currently the most efficient multiresolution-based grayscale denoiser we know of. It consists of a multivariate thresholding—integrating both inter- and intrascale dependencies—performed in a highly redundant (eight orientations per scale) full steerable pyramid. We have used the code and the settings provided by the authors,[4] except that periodic boundary conditions have been considered as in the other methods. For color image denoising, we have simply applied the *BLS-GSM* independently in each RGB channels.

  Note that, in all likelihood, a more complete (i.e., including the modeling of local neighborhoods and parents) multichannel extension of the *BLS-GSM* than the one recently initiated by Scheunders *et al.* in [26], would certainly give substantially better results than the independent application of the original *BLS-GSM* that we propose to use here.

In the first experiment, we have corrupted the test images with additive (synthetic) Gaussian white noise having the same variance in each RGB channel. The PSNR results are displayed in Table I. Using the same orthonormal wavelet transform, our *SURE-LET* algorithm clearly outperforms (often by more than $+1$ dB) the *ProbShrink-MB*. Despite being performed in a *nonredundant* wavelet representation, our solution gives even better (average gain of nearly $+0.5$ dB) output PSNRs than the *ProbShrink-MB* applied in the *undecimated* wavelet representation, and similar results to *BLS-GSM* for all the tested images, as well as for the whole range of input noise levels. From a visual point of view, our algorithm holds its own against the best redundant approaches based on multiresolution (see Fig. 5).

In a second experiment, the test images have been corrupted with additive (synthetic) Gaussian white noise having a different power in each RGB channel. As a comparison, we have used another version of Pižurica *et al.*'s *ProbShrink* [18] which is an application of their original grayscale denoiser in the luminance-chrominance color space in the undecimated wavelet transform, consequently referred to as the *UWT ProbShrink-YUV*. The PSNR results are displayed in Table II. We have also reported in this table the results published in [17]. Their algorithm is developed in an orthonormal wavelet transform framework and combines the universal hidden Markov tree (uHMT), a statistical approach devised in [19], with an optimal luminance/color-difference space projection (OCP); it will, therefore, be referred to as the *OWT uHMT-OCP*. As can be observed, our *SURE-LET* approach outperforms these two algorithms in terms of PSNR (almost $+1$ dB); it even gives better results than the *BLS-GSM* for most images. In Fig. 6, we show the visual quality of the various algorithms: ours exhibits very few color artifacts, and preserves most of the image details.

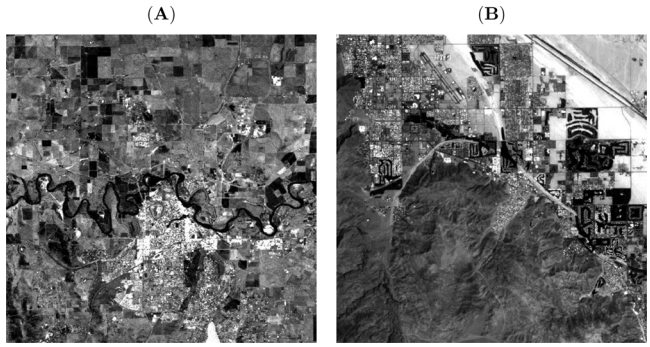

Fig. 7. (A) First band of a *Landsat* image of Wagga Wagga. (B) First band of a *Landsat* image showing a part of Southern California.

Finally, we must emphasize that the execution of the unoptimized Matlab implementation of our algorithm only lasts around 6 s for 512×512 color images on a Power Mac G5 workstation with 1.8-Hz CPU. To compare with, the best *ProbShrink* requires approximately 19 s under the same conditions, whereas the *BLS-GSM* requires about 260 s. Besides achieving very competitive denoising results, the proposed solution is also faster than most state-of-the-art algorithms: the interested reader may wish to check these claims with our online demo [30]. Not only is it faster, but it is also much more memory effective because it makes use of a nonredundant transformation, an approach that could prove even more valuable for the processing of higher dimensional data—in particular, tridimensional and moving pictures.

B. Multiband Image Denoising

Our *SURE-LET* algorithm is particularly well-suited to the denoising of multiband images, such as satellite images, and more generally, any stack of images with significant common content (e.g., consecutive moving images or consecutive volume slices). Indeed, thanks to the SURE-based optimization of the linear parameters, the potentially strong similarities between the various channels are efficiently—and automatically—taken into account. There is thus no need to decorrelate the bands beforehand.

For the experiments, we have used two different seven-band *Landsat* images.[5]

[2]In the minimum SURE sense.

[3]Available at: http://telin.rug.ac.be/~sanja/

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

[5]Data by courtesy of the following website: http://ceos.cnes.fr:8100/cdrom-00b2/ceos1/datasets.htm

TABLE III
COMPARISON OF MULTIBAND DENOISING ALGORITHMS (SAME NOISE LEVEL IN EACH CHANNEL)

σ_i, $i \in [1,6]$	5	10	15	20	25	30	50	100	5	10	15	20	25	30	50	100
Input PSNR [dB]	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	***Wagga Wagga*** 512×512								***Southern California*** 512×512							
BLS-GSM [10]	35.06	30.21	27.70	26.09	24.92	24.02	21.76	19.26	35.52	30.83	28.33	26.67	25.47	24.54	22.22	19.71
OWT SURE-LET	**35.37**	**30.88**	**28.59**	**27.09**	**26.00**	**25.14**	**22.95**	**20.39**	**36.57**	**32.19**	**29.78**	**28.15**	**26.96**	**26.03**	**23.64**	**20.92**

Notes: 1. *BLS-GSM* is applied in a *highly redundant* representation (a full steerable pyramid), whereas our *SURE-LET* only uses a *non-redundant* (orthonormal) transform.
2. Output PSNRs have been averaged over ten noise realizations.

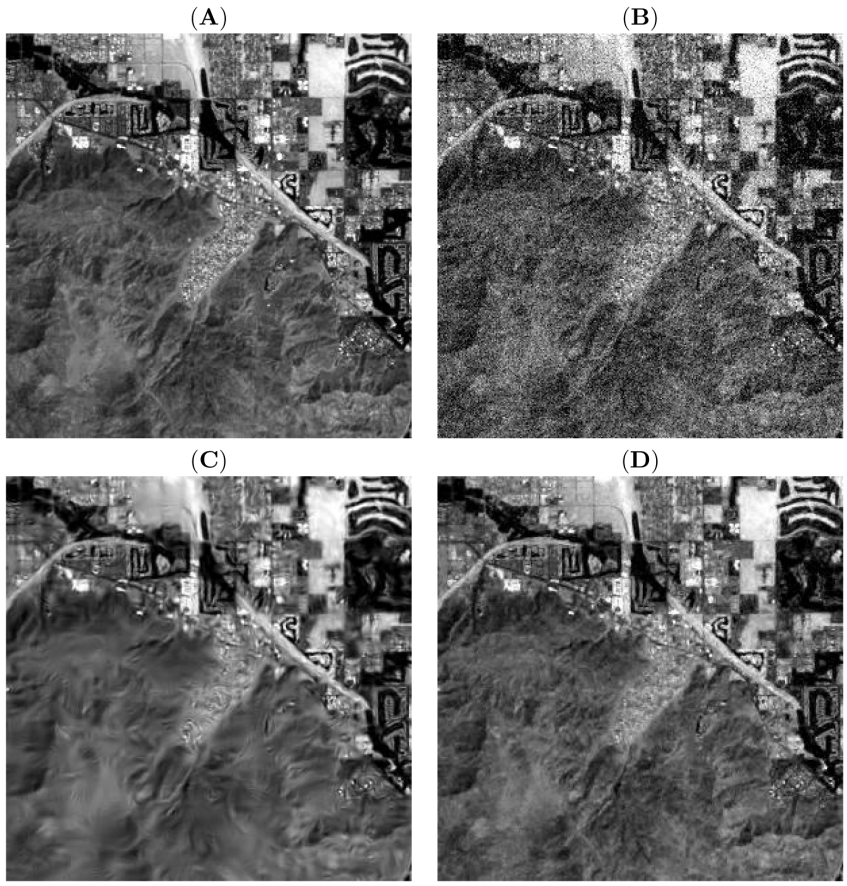

Fig. 8. (A) Part of the first band of the noise-free *Southern California* image. (B) Noisy version of it: PSNR $= 18.59$ [dB]. (C) Result of the *BLS-GSM* (full steerable pyramid): PSNR $= 24.54$ [dB]. (D) Result of our *IS-IC SURE-LET* (OWT *sym8*): PSNR $= 26.03$ [dB].

- The first one covers the inland city of Wagga Wagga in Australia. The coverage area shown in Fig. 7(A) is approximately 15×15 km with a resolution of 30 m (image size of $N = 512 \times 512 \times 7$).
- The second one shows a part of a scene taken over Southern California, encompassing the region from Long Beach to San Diego. The coverage area shown in Fig. 7(B) is also approximately 15×15 km with a resolution of 30 m (image size of $N = 512 \times 512 \times 7$).

For the denoising experiments, we have disregarded band 6 of both *Landsat* images, since it is very different from the others (a thermal infrared channel at lower resolution); our test data are, therefore, of size $N = 512 \times 512 \times 6$. Unfortunately, we were unable to compare our results with the ones obtained by other algorithms specifically devised to handle more than three bands because we could neither get the test data used in their experiments nor find the corresponding implementations. However, to have a point of comparison we show the results obtained by the *BLS-GSM* applied separately in each bands.

As it can be observed in Table III, our *SURE-LET* clearly outperforms (often by more than $+1$ dB) the *BLS-GSM*, although it is applied in an orthonormal wavelet representation. A visual comparison is also shown in Fig. 8 for one particular band. From a computational time point of view, there is an obvious interest in considering nonredundant transformation: the denoising of the six bands of a 512×512 *Landsat* image lasts 22 s with our algorithm, whereas it takes 520 s—more than 8 min—with *BLS-GSM*.

V. CONCLUSION

We have proposed an extension of our previous monochannel denoiser to properly handle multichannel images. The resulting interscale-interchannel wavelet estimator consists of a linear expansion of thresholding functions, whose parameters are solved for by minimizing an unbiased estimate of the expected mean squared error between the noise-free signal and the denoised one. This linear parametrization has two main benefits: first, thanks to the quadratic form of the MSE estimate, the

parameters optimization amounts to resolve a linear system of equations, which makes our approach computationally *light* and *fast*; second, the optimized linear parameters act as an optimal—in the minimum SURE sense—transformation of the data. For color image denoising, the consequence is that the denoising performances are nearly insensitive to the color representation, contrary to most existing approaches. Compared to even the best multiresolution algorithms (which involve highly redundant transforms), the results confirm the efficiency of our approach, both from a computational and from a quality point of view. The gains may even become quite significant when the number of channels increases.

Preliminary investigations using an interchannel adaptation of the SURE-LET approach [12] within an undecimated wavelet transform indicate that substantial additional denoising gains can be achieved (+1 dB and often more). Of course, these are obtained at the expense of a higher computational cost and a worse memory management.

REFERENCES

[1] I. Daubechies, "Orthonormal Bases of Compactly Supported Wavelets," *Commun. Pure Appl. Math.*, vol. 41, pp. 909–996, 1988.

[2] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

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

[4] 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.

[5] H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch, "Adaptive Bayesian wavelet shrinkage," *J. Roy. Statist. Soc.*, vol. 92, no. 440, pp. 1413–1421, Dec. 1997.

[6] M. Clyde, G. Parmigiani, and B. Vidakovic, "Multiple shrinkage and subset selection in wavelets," *Biometrika*, vol. 85, no. 2, pp. 391–401, Jun. 1998.

[7] 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.

[8] L. Şendur 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.

[9] 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.

[10] 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.

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

[12] T. Blu and F. Luisier, "The SURE-LET approach to image denoising," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2778–2786, Nov. 2007.

[13] T. F. Chan, S. H. Kang, and J. Shen, "Total variation denoising and enhancement of color images; Based on the CB and HSV color representation," *J. Vis. Commun. Image Represent.*, vol. 12, no. 4, pp. 422–435, Jun. 2001.

[14] B. Tang, G. Sapiro, and V. Caselles, "Color image enhancement via chromaticity diffusion," *IEEE Trans. Image Process.*, vol. 10, no. 5, pp. 701–707, May 2001.

[15] S. Kim, "PDE-based image restoration: A hybrid model and color image denoising," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1163–1170, May 2006.

[16] O. Ben-Shahar and S. W. Zucker, "Hue fields and color curvatures: A perceptual organization approach to color image denoising," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Jun. 2003, vol. 2, pp. II-713–II-720.

[17] N.-X. Lian, V. Zagorodnov, and Y.-P. Tan, "Edge-preserving image denoising via optimal color space projection," *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2575–2587, Sep. 2006.

[18] A. Pižurica, W. Philips, and P. Scheunders, "Wavelet domain denoising of single-band and multiband images adapted to the probability of the presence of features of interest," *Proc. SPIE*, vol. 5914, no. 591411, Sep. 2005.

[19] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1056–1068, Jul. 2001.

[20] G. Sapiro and D. L. Ringach, "Anisotropic diffusion of multivalued images with applications to color filtering," *IEEE Trans. Signal Process.*, vol. 5, no. 11, pp. 1582–1586, Nov. 1996.

[21] P. Blomgren and T. F. Chan, "Color TV: Total variation methods for restoration of vector-valued images," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 304–309, Mar. 1998.

[22] S. G. Chang, B. Yu, and M. Vetterli, "Wavelet thresholding for multiple noisy image copies," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1631–1635, Sep. 2000.

[23] J.-L. Starck and P. Querre, "Multispectral image restoration by the wavelet-Karhunen–Loeve transform," *Signal Process.*, vol. 81, no. 12, pp. 2449–2459, Dec. 2001.

[24] A. K. Fletcher, V. K. Goyal, and K. Ramchandran, "On multivariate estimation by thresholding," in *Proc. Int. Conf. Image Process.*, Sep. 14–17, 2003, vol. 1, pp. 61–64.

[25] P. Scheunders, "Wavelet thresholding of multivalued images," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 475–483, Apr. 2004.

[26] P. Scheunders and S. D. Backer, "Wavelet denoising of multicomponent images, using a Gaussian scale mixture model," in *Proc. 18th Int. Conf. Pattern Recognition*, 2006, vol. 3, pp. 754–757.

[27] 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.

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

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

[30] *SURE-LET Color Denoising Demo*, [Online]. Available: http://bigwww.epfl.ch/demo/suredenoising-color/index.html

[31] P. Scheunders and J. Driesen, "Least-squares interband denoising of color and multispectral images," in *Proc. IEEE Int. Conf. Image Processing*, Oct. 2004, vol. 2, pp. 985–988.

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 in 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 École Polytechnique, France, in 1986, and from Télécom Paris (ENST), Paris, France, in 1988, and the Ph.D. degree in electrical engineering from ENST for a study on iterated rational filterbanks, applied to wideband audio coding in 1996.

Between 1998 and 2007, he was with the Biomedical Imaging Group, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland. He is currently a Professor in the Department of Electronic Engineering, The Chinese University of Hong Kong. His research interests include (multi)wavelets, multiresolution analysis, multirate filterbanks, interpolation, approximation and sampling theory, image denoising, psychoacoustics, optics, wave propagation, etc.

Dr. Blu was the recipient of two best paper awards from the IEEE Signal Processing Society (2003 and 2006). Between 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.