This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

Image Denoising in Mixed Poisson-Gaussian Noise

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

Abstract

We propose a general methodology (PURE-LET) to design and optimize a wide class of transform-domain thresholding algorithms for denoising images corrupted by mixed Poisson-Gaussian noise. We express the denoising process as a linear expansion of thresholds (LET) that we optimize by relying on a purely data-adaptive unbiased estimate of the mean-squared error (MSE), derived in a non-Bayesian framework (PURE: Poisson-Gaussian Unbiased Risk Estimate). We provide a practical approximation of this theoretical MSE estimate for the tractable optimization of arbitrary transform-domain thresholding. We then propose a pointwise estimator for undecimated filterbank transforms, which consists of subband-adaptive thresholding functions with signal-dependent thresholds that are globally optimized in the image domain. We finally demonstrate the potential of the proposed approach through extensive comparisons with state-of-the-art techniques that are specifically tailored to the estimation of Poisson intensities. We also present denoising results obtained on real images of low-count fluorescence microscopy.

EDICS: TEC-RST, TEC-MRS, TEC-BIP

Copyright (c) 2010 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org

Florian Luisier and Michael Unser are with the Biomedical Imaging Group (BIG), École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland (e-mail: florian.luisier@a3.epfl.ch; michael.unser@epfl.ch).

Thierry 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).

This work was supported by the Center for Biomedical Imaging (CIBM) of the Geneva–Lausanne Universities and the EPFL, the Leenaards and Louis-Jeantet foundations, the Swiss National Science Foundation under grant 200020-109415, as well as by the General Research Fund CUHK410209 from the Hong Kong Research Grant Council.

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

I. Introduction

The two predominant sources of noise in digital image acquisition are:

1) The stochastic nature of the photon-counting process at the detectors.
2) The intrinsic thermal and electronic fluctuations of the acquisition devices.

Under standard illumination conditions, the second source of noise, which is signal-independent, is stronger than the first one. This motivates the usual additive-white-Gaussian-noise (AWGN) assumption. However, in many applications such as fluorescence microscopy or astronomy, only few photons are collected by the photosensors, due to various physical constraints (low-power light source, short exposure time, photo-toxicity). Under these imaging conditions, the major source of noise is strongly signal-dependent. Consequently, it is more reasonable to model the output of the detectors as a Poisson-distributed random vector. The problem is then to estimate the underlying intensities of Poisson random variables, potentially further degraded by independent AWGN. In this paper, we propose fast and high-quality nonlinear algorithms for denoising digital images corrupted by mixed Poisson-Gaussian noise.

Among various image-denoising strategies, the transform-domain approaches in general, and in particular the multiscale ones, are very efficient for AWGN reduction (e.g. [1]–[3]). As many natural images can be represented by few significant coefficients in a suitable basis/frame, the associated transform-domain processing amounts to a (possibly multivariate) *thresholding* of the transformed coefficients, which results in a fast denoising procedure. Since the present work lies within this scope of transform-domain thresholding strategies, we discuss hereafter the main multiscale techniques that have been considered for *Poisson* intensity estimation. Note that there are also non-multiscale methods for Poisson denoising, e.g. [4]–[6].

A. Related Work

Since the Poisson statistics are generally more difficult to track in a transformed domain than the traditional Gaussian ones, a natural solution consists in "Gaussianizing" the Poisson measurements. This is usually performed by applying a nonlinear mapping (*e.g.*, a square root) to the raw data, an approach

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

that has been theorized by Anscombe in [7] and first exploited in denoising applications by Donoho in [8]. The so-called *Anscombe* variance-stabilizing transform (VST) has been later generalized by Murtagh *et al.* to stabilize the variance of a Poisson random variable corrupted by AWGN [9]. After stabilization, any high-quality AWGN denoiser can be applied (*e.g.*, [1]).

Several works take advantage of the fact that the unnormalized Haar wavelet transform[1] has the remarkable property of preserving Poisson statistics in its lowpass channel. Fryzlewicz *et al.* have proposed a VST based on the observation that the scaling coefficients at a given scale are good local estimates of the noise variances of the same-scale wavelet coefficients [10]. Their approach gave state-of-the-art denoising results (in the minimum mean-squared error sense) at the time of its publication (2004). Another interesting property of the unnormalized Haar transform applied to Poisson data is that the statistical relation between a scaling coefficient (parent) and its child (scaling coefficient at the next finer scale) is very simple; the distribution of a child conditioned on its parent is binomial. These properties have been exploited in a Bayesian framework in [11]–[14], as well as in a user-calibrated hypothesis testing [15]. Hirakawa *et al.* have taken advantage of the Skellam distribution of the unnormalized Haar *wavelet* coefficients to derive a so-called *SkellamShrink* [16], [17], which can be viewed as a Poisson variant of Donoho's *et al. SUREshrink* [18]. Recently, we have proposed a non-Bayesian framework to estimate Poisson intensities in the unnormalized Haar wavelet domain (*PURE-LET* [19], see Fig. 1 for an illustration of its principle). The qualitative and computational efficiency of this approach results from the combination of the following two key ingredients:

1) A *prior-free* unbiased estimate of the expected mean-squared error (MSE) between the unknown original image and the denoised one. Under an AWGN hypothesis, this estimator is known as Stein's unbiased risk estimate (SURE) [20], while, for Poisson data, we called it PURE, which stands for Poisson's unbiased risk estimate. Note that, contrary to the Bayesian approach, the noise-free image is not considered as random in [19].

[1]See Fig. 1 for a filterbank implementation of the unnormalized Haar wavelet transform.

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

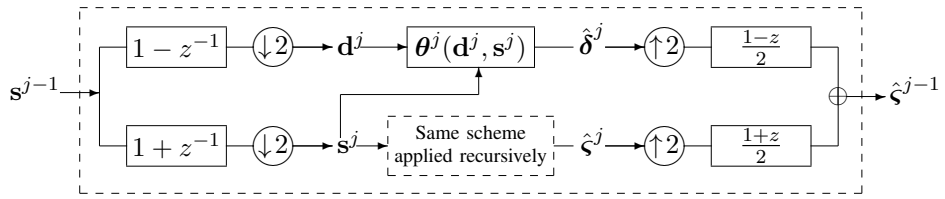

Fig. 1. Filterbank implementation of the unnormalized discrete Haar wavelet transform and principle of the class of denoising algorithms described in [19]. Notations: the superscript $j = 1 \dots J$ indicates the level of decomposition; $\mathbf{s}$ is the vector of noisy scaling coefficients ($\mathbf{s}^0$ is thus the noisy input); $\mathbf{d}$ is the vector of noisy wavelet coefficients; $\boldsymbol{\theta}$ is the subband-dependent thresholding function; $\hat{\boldsymbol{\delta}}$ (resp. $\hat{\boldsymbol{\varsigma}}$) is the vector of the estimated noise-free wavelet (resp. scaling) coefficients.

2) A *linear* parameterization of the denoising process, through a linear expansion of thresholds (LET). The optimal parameters of this expansion are then the solution of a system of linear equations, resulting from the minimization of the subband-dependent quadratic unbiased estimate of the MSE.

In the more standard case of Gaussian noise reduction, better results have been usually obtained with wavelets that are smoother than Haar, and/or with shift-invariant transformations (e.g. the undecimated wavelet transform). This has motivated researchers to devise Poisson-intensity estimators applicable to arbitrary multiscale transforms. Kolaczyk has developed (a pair of) soft/hard-thresholds for arbitrary wavelet shrinkage of "burst-like" Poisson intensities [21]. This pair of thresholds can be seen as an adapted version of Donoho's *et al.* universal threshold that was designed for AWGN [22]. This approach was generalized to arbitrary kinds of Poisson-distributed data by Charles and Rasson [23]. Based on the statistical method of cross-validation, Nowak *et al.* derived a wavelet shrinkage, whose threshold is locally adapted to the estimated noise variance [24]. The modulation estimator devised by Antoniadis and Spatinas [25], which is based on cross-validation as well, covers all univariate natural exponential families with quadratic variance functions, of which Gaussian and Poisson distributions are two particular cases. Sardy *et al.* proposed a generalization of Donoho and Johnstone's wavelet shrinkage for a broad class of exponential noise distributions, including the Poisson case [26]. Their estimator is the solution of a log-likelihood problem, regularized by the addition of a wavelet-domain ℓ_1-penalty. Using the concept of multiscale likelihood factorizations, Kolaczyk and Nowak introduced complexity-penalized estimators that

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

can also handle a wide class of distributions (Gaussian, Poisson, and multinomial) [27]. This methodology was further exploited by Willett *et al.* who proposed a platelet-based penalized likelihood estimator that has been demonstrated to be particularly efficient for denoising piecewise-smooth signals [28]. Recently, the VST-based approach has been revitalized thanks to the contributions of Jansen, who combined VST with multiscale Bayesian models [29], and Zhang *et al.*, who proposed a multiscale VST that can better stabilize very low intensity signals, and showed how it can be efficiently used with the latest multiresolution transformations (*e.g.*, curvelets) [30]. This latter VST solution can also stabilize Poisson data embedded in AWGN [31].

Most of the denoising algorithms discussed in this section have been carefully evaluated by Besbeas *et al.* [32] using 1D data. Most of them are specifically designed for pure Poisson noise. To the best of our knowledge, there are very few denoising algorithms that can properly handle mixed Poisson-Gaussian noise.

B. Contributions

In this work, we extend the PURE-LET approach in three main directions. First, we lift the restricted use of the unnormalized Haar wavelet transform by generalizing to arbitrary (redundant) transform-domain (nonlinear) processing. Second, we consider a more realistic noise model: a Poisson random vector degraded by AWGN, for which we derive a new theoretically unbiased MSE estimate; this new estimate, for which we keep the name PURE, combines both SURE and PURE. Third, we show that PURE can be used to globally optimize a LET spanning several (redundant) bases.

Because we are dealing with transformations that are not necessarily orthonormal anymore, the denoising process has to be optimized in the image domain, to ensure a global MSE minimization [3], [33]. To make tractable the optimization of arbitrary transform-domain processing, we also need a practical approximation of the PURE. All these extensions allow us to get the best (in the minimum PURE sense) linear combination of the processed subbands coming from complementary transforms, such as the undecimated wavelet transform (appropriate for piecewise-smooth images) and the overcomplete

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

block-discrete cosine transform (effective in sparsely representing textured images).

The paper is organized as follows: In the next section, we introduce the theoretical basis of this work. We first derive an unbiased estimate of the MSE for an arbitrary processing of Poisson data degraded by AWGN, and then propose a practical approximation of this general result. A *pointwise* estimator for undecimated filterbank transforms is presented in Section III and compared with some state-of-the-art approaches in Section IV. Results on real fluorescence-microscopy data are finally shown in Section V.

II. THEORY

A. Definitions and Notations

- A random variable z follows a Poisson distribution of intensity $x \in \mathbb{R}_+$ if and only if its conditional probability density function is $p(z = k|x) = \dfrac{x^k}{k!}e^{-x}$. We use the standard notation $z \sim \mathcal{P}(x)$.
- A random variable b follows a Gaussian distribution with zero-mean and variance σ^2 if and only if its probability density function is $q(b) = \dfrac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{b^2}{2\sigma^2}}$. We use the standard notation $b \sim \mathcal{N}(0, \sigma^2)$.

Throughout this paper, for a given vector $\mathbf{v} \in \mathbb{R}^N$, we use the notation v_n to refer to its nth component, where $n \in \{1, \ldots, N\}$. The variable $\mathbf{z}$ denotes a vector of N independent Poisson random variables z_n of underlying intensities x_n, with $z_n \sim \mathcal{P}(x_n)$. A realization of $\mathbf{z}$ can be thought of as a noisy measurement of the intensity signal $\mathbf{x}$. Note that, in contrast with Bayesian approaches, $\mathbf{x}$ is considered to be *deterministic* in the present work.

We further assume that the noisy measurements z_n are degraded by i.i.d. AWGNs b_n of known variance σ^2, so that $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. The final observation vector $\mathbf{y}$ is therefore given by

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

Our goal is then to find an estimate $\hat{\mathbf{x}} = \mathbf{f}(\mathbf{y})$ that is the closest possible to the original signal in the minimum MSE sense. That is, we want to minimize

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

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

B. Useful Properties of the Poisson and Gaussian Random Variables

Property 1 (Stein's Lemma)**.** *Let* $\mathbf{y} = \mathbf{z} + \mathbf{b}$*, where* $\mathbf{z}$ *is independent of* $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ *and let* $\mathbf{f} : \mathbb{R}^N \rightarrow \mathbb{R}^N$ *be a (weakly) differentiable real-valued vector function such that* $\forall n$*,* $\mathscr{E}_{\mathbf{b}}\left\{|\partial f_n(\mathbf{y})/\partial y_n|\right\} < \infty$*. Then*

$$\mathscr{E}_{\mathbf{b}}\left\{\mathbf{b}^{\mathrm{T}}\mathbf{f}(\mathbf{y})\right\} = \sigma^2\mathscr{E}_{\mathbf{b}}\left\{\mathrm{div}\left\{\mathbf{f}(\mathbf{y})\right\}\right\}$$

where $\mathrm{div}\left\{\mathbf{f}(\mathbf{y})\right\} = \sum_{n=1}^{N} \dfrac{\partial f_n(\mathbf{y})}{\partial y_n}$ *is the divergence of the function* $\mathbf{f}$ *and* $\mathscr{E}_{\mathbf{b}}\left\{\cdot\right\}$ *stands for the mathematical expectation taken over all realizations of the random variable* $\mathbf{b}$*.*

This is a standard result in the statistical literature, known as "Stein's lemma", which was first established by Charles Stein [20]. An alternative proof can also be found in [3].

Property 2. *Let* $\mathbf{z} \sim \mathcal{P}(\mathbf{x})$ *and let* $\mathbf{f} : \mathbb{R}^N \rightarrow \mathbb{R}^N$ *be a real-valued vector function and let the family of vectors* $(\mathbf{e}_n)_{n=1\ldots N}$ *be the canonical basis of* $\mathbb{R}^N$*. Then,*

$$\mathscr{E}_{\mathbf{z}}\left\{\mathbf{x}^{\mathrm{T}}\mathbf{f}(\mathbf{z})\right\} = \mathscr{E}_{\mathbf{z}}\left\{\mathbf{z}^{\mathrm{T}}\mathbf{f}^{-}(\mathbf{z})\right\}$$

where $\mathbf{f}^{-}(\mathbf{z}) = [f_n(\mathbf{z} - \mathbf{e}_n)]_{n=1\ldots N}$*.*

This property can be thought of as the "Poisson's equivalent" of Stein's lemma. A proof of a similar result can be found for instance in [19], [34], [35].

C. PURE: An Unbiased MSE Estimate for Poisson Data Degraded by AWGN

In practice, we obviously do not have access to the original noise-free signal $\mathbf{x}$. Therefore, we cannot compute the actual MSE and minimize it. However, we can rely on its unbiased estimate given in the following theorem:

Theorem 1. *Let* $\mathbf{y}$ *be the random variable defined in* (1) *and let* $\mathbf{f}(\mathbf{y}) = [f_n(\mathbf{y})]_{n=1\ldots N}$ *be an* N*-dimensional (weakly) differentiable real-valued vector function such that* $\forall n = 1 \ldots N$*:* $\mathscr{E}_{\mathbf{y}}\left\{|\partial f_n(\mathbf{y})/\partial y_n|\right\} < \infty$*. Then, the random variable*

$$\epsilon = \frac{1}{N}\left(\|\mathbf{f}(\mathbf{y})\|^2 - 2\mathbf{y}^{\mathrm{T}}\mathbf{f}^{-}(\mathbf{y}) + 2\sigma^2\mathrm{div}\left\{\mathbf{f}^{-}(\mathbf{y})\right\}\right) + \frac{1}{N}\left(\|\mathbf{y}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{y}\right) - \sigma^2 \tag{2}$$

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

is an unbiased estimate of the expected MSE, that is, $\mathscr{E}_{\mathbf{y}}\{\epsilon\} = \frac{1}{N}\mathscr{E}_{\mathbf{y}}\left\{\|\mathbf{f}(\mathbf{y}) - \mathbf{x}\|^2\right\}$.

Proof: By expanding the expectation of the squared error between $\mathbf{x}$ and its estimate $\mathbf{f}(\mathbf{y})$, we obtain

$$\mathscr{E}_{\mathbf{y}}\left\{\|\mathbf{f}(\mathbf{y}) - \mathbf{x}\|^2\right\} = \mathscr{E}_{\mathbf{y}}\left\{\|\mathbf{f}(\mathbf{y})\|^2\right\} - 2\underbrace{\mathscr{E}_{\mathbf{y}}\left\{\mathbf{x}^{\mathrm{T}}\mathbf{f}(\mathbf{y})\right\}}_{\text{(I)}} + \underbrace{\|\mathbf{x}\|^2}_{\text{(II)}}. \tag{3}$$

We can now evaluate the two expressions (I,II) which involve the unknown data $\mathbf{x}$.

(I) Since $\mathbf{b}$ is independent of $\mathbf{x}$ and $\mathbf{z}$, we can write the following sequence of equalities:

$$\begin{aligned}
\mathscr{E}_{\mathbf{y}}\left\{\mathbf{x}^{\mathrm{T}}\mathbf{f}(\mathbf{y})\right\} &= \mathscr{E}_{\mathbf{b}}\left\{\mathscr{E}_{\mathbf{z}}\left\{\mathbf{x}^{\mathrm{T}}\mathbf{f}(\mathbf{z}+\mathbf{b})\right\}\right\} \\
&\overset{\text{Prop. 2}}{=} \mathscr{E}_{\mathbf{b}}\left\{\mathscr{E}_{\mathbf{z}}\left\{\mathbf{z}^{\mathrm{T}}\mathbf{f}^{-}(\mathbf{z}+\mathbf{b})\right\}\right\} \\
&\overset{(1)}{=} \mathscr{E}_{\mathbf{y}}\left\{\mathbf{y}^{\mathrm{T}}\mathbf{f}^{-}(\mathbf{y})\right\} - \mathscr{E}_{\mathbf{z}}\left\{\mathscr{E}_{\mathbf{b}}\left\{\mathbf{b}^{\mathrm{T}}\mathbf{f}^{-}(\mathbf{y})\right\}\right\} \\
&\overset{\text{Prop. 1}}{=} \mathscr{E}_{\mathbf{y}}\left\{\mathbf{y}^{\mathrm{T}}\mathbf{f}^{-}(\mathbf{y})\right\} - \sigma^2\mathscr{E}_{\mathbf{y}}\left\{\operatorname{div}\left\{\mathbf{f}^{-}(\mathbf{y})\right\}\right\}.
\end{aligned} \tag{4}$$

(II) We notice that:

$$\begin{aligned}
\|\mathbf{x}\|^2 &= \mathscr{E}_{\mathbf{y}}\left\{\mathbf{x}^{\mathrm{T}}\mathbf{y}\right\} \\
&\overset{(4)}{=} \mathscr{E}_{\mathbf{y}}\left\{\|\mathbf{y}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{y}\right\} - N\sigma^2.
\end{aligned} \tag{5}$$

Putting back (4) and (5) into (3) finally demonstrates Theorem 1. ■

In the remainder of this paper, the unbiased MSE estimate for mixed Poisson-Gaussian noise defined in Equ. (2) will be referred to as PURE (Poisson-Gaussian Unbiased Risk Estimate). It is a generalization of SURE (Stein's unbiased risk estimate) [20], which was derived for AWGN only, and the PURE we recently exposed in [19], which was devised for pure Poisson noise reduction in the unnormalized Haar wavelet domain. Note that the result of Theorem 1 was given in [36] without proof, where it was applied to linear image deconvolution.

The variance of the PURE estimator notably depends on the number of samples y_n, as well as on the number of parameters (degrees of freedom) involved in $\mathbf{f}$. Therefore, provided that the solution is sufficiently constrained, PURE remains very close to the actual MSE in image processing, due to the high number of available samples.

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

D. Taylor-Series Approximation of PURE

Unfortunately, the presence of the term $\mathbf{f}^{-}(\mathbf{y})$ in Theorem 1 makes PURE impractical to evaluate for an arbitrary nonlinear processing. Indeed, in order to compute just one component $f_n(\mathbf{y}-\mathbf{e}_n)$ of $\mathbf{f}^{-}(\mathbf{y})$, one would need to apply the whole denoising process to a slightly modified version of the noisy input. This operation would have to be repeated N times to get the full vector $\mathbf{f}^{-}(\mathbf{y})$. Such a "brute force" approach is thus prohibitive in practice, considering that a typical image contains $N = 256 \times 256$ pixels. Instead, we propose to use the first-order Taylor-series approximation of $\mathbf{f}^{-}(\mathbf{y})$ given by

$$f_n(\mathbf{y}-\mathbf{e}_n) \;\simeq\; f_n(\mathbf{y}) - \frac{\partial f_n(\mathbf{y})}{\partial y_n}$$

for all $n = 1 \dots N$. Consequently, provided that the above approximation is reasonable (*i.e.*, that f_n varies slowly), the function $\mathbf{f}^{-}(\mathbf{y})$ is well approximated by $\mathbf{f}(\mathbf{y}) - \partial\mathbf{f}(\mathbf{y})$, where $\partial\mathbf{f}(\mathbf{y}) = [\frac{\partial f_n(\mathbf{y})}{\partial y_n}]_{n=1\dots N}$ is the $N \times 1$ vector made of the first derivative of each function f_n with respect to y_n. The PURE unbiased MSE estimate defined in (2) is, in turn, well approximated by[2]

$$\begin{aligned}\hat{\epsilon} \;=\; & \frac{1}{N}\left(\|\mathbf{f}(\mathbf{y})\|^2 - 2\mathbf{y}^{\mathrm{T}}(\mathbf{f}(\mathbf{y}) - \partial\mathbf{f}(\mathbf{y})) + 2\sigma^2 \mathrm{div}\left\{\mathbf{f}(\mathbf{y}) - \partial\mathbf{f}(\mathbf{y})\right\}\right) + \\ & \frac{1}{N}(\|\mathbf{y}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{y}) - \sigma^2. \end{aligned} \tag{6}$$

Note that, if the processing $\mathbf{f}$ is *linear*, then the two MSE estimates (2) and (6) are identical.

E. PURE for Transform-Domain Denoising

In this section, we derive an expression for the MSE estimate given in (6), in the particular case of a *transform-domain pointwise* processing $\boldsymbol{\Theta}$. The principle of transform-domain denoising is illustrated in Fig. 2: once the size of the input and output data are frozen, the linear decomposition and reconstruction operators are characterized by the matrices $\mathbf{D} = [d_{i,j}]_{(i,j)\in[1\dots L]\times[1\dots N]}$ and $\mathbf{R} = [r_{i,j}]_{(i,j)\in[1\dots N]\times[1\dots L]}$, respectively, that satisfy the perfect reconstruction property $\mathbf{RD} = \mathbf{I}$.

Dealing with a signal-dependent noise makes it generally difficult to express the observation model in the transformed domain. Therefore, we assume here that a (coarse) "map" of the signal-dependent noise

[2]To be fully rigorous, we must further assume that each f_n is twice differentiable with respect to y_n.

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

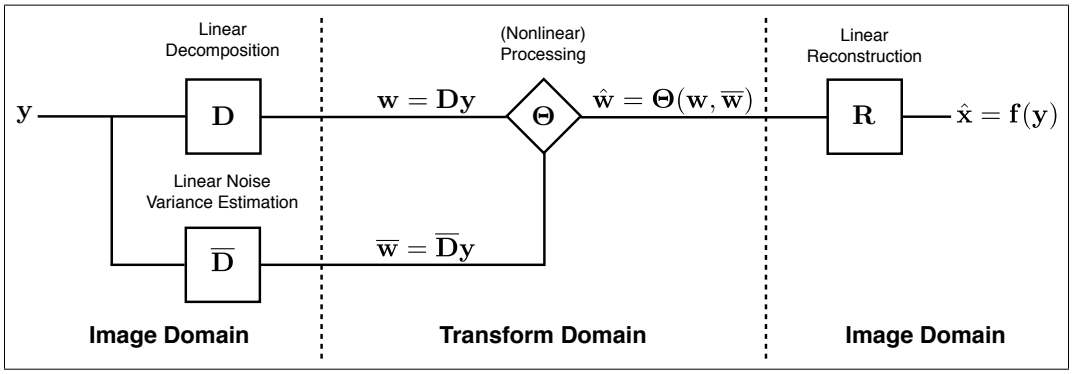

Fig. 2. Principle of transform-domain denoising for signal-dependent noise.

variance can be obtained in the transformed domain, by applying a linear transformation (typically, a smoothing) $\overline{\mathbf{D}} = [\overline{d}_{i,j}]_{(i,j)\in[1\ldots L]\times[1\ldots N]}$ to the noisy data $\mathbf{y}$, as shown in Fig. 2. The denoised estimate $\hat{\mathbf{x}}$ can be thus finally expressed as a function $\mathbf{f}$ of the noisy input signal $\mathbf{y}$ as

$$\hat{\mathbf{x}} = \mathbf{f}(\mathbf{y}) = \mathbf{R}\boldsymbol{\Theta}(\underbrace{\mathbf{D}\mathbf{y}}_{\mathbf{w}}, \underbrace{\overline{\mathbf{D}}\mathbf{y}}_{\overline{\mathbf{w}}}) \tag{7}$$

where $\boldsymbol{\Theta}(\mathbf{w}, \overline{\mathbf{w}}) = [\theta_l(w_l, \overline{w}_l)]_{l\in[1\ldots L]}$ is a *pointwise* (nonlinear) processing.

Note that the *PUREshrink* and the PURE-LET algorithms introduced in [19] belong to the general class of processing defined in (7). In that case, $\mathbf{D}$ and $\mathbf{R}$ implement the (unnormalized) Haar wavelet decomposition and reconstruction, respectively (see Fig. 1). $\mathbf{w}$ and $\overline{\mathbf{w}}$ then represent the Haar wavelet $\mathbf{d}$ and scaling coefficients $\mathbf{s}$, respectively.

Corollary 1. *For the transform-domain pointwise processing defined in* (7)*, the MSE estimate given in* (6) *can be reformulated as*

$$\begin{aligned}\hat{\epsilon} \;=\; & \frac{1}{N}\|\mathbf{f}(\mathbf{y}) - \mathbf{y}\|^2 + \frac{2}{N}\left(\boldsymbol{\Theta}_1(\mathbf{w}, \overline{\mathbf{w}})^{\mathrm{T}}(\mathbf{D} \bullet \mathbf{R}^{\mathrm{T}})\mathbf{y} + \boldsymbol{\Theta}_2(\mathbf{w}, \overline{\mathbf{w}})^{\mathrm{T}}(\overline{\mathbf{D}} \bullet \mathbf{R}^{\mathrm{T}})\mathbf{y}\right) + \\ & \frac{2\sigma^2}{N}\left(\operatorname{diag}\left\{\mathbf{D}\mathbf{R}\right\}^{\mathrm{T}}\boldsymbol{\Theta}_1(\mathbf{w}, \overline{\mathbf{w}}) + \operatorname{diag}\left\{\overline{\mathbf{D}}\mathbf{R}\right\}^{\mathrm{T}}\boldsymbol{\Theta}_2(\mathbf{w}, \overline{\mathbf{w}})\right) - \\ & \frac{2\sigma^2}{N}\Big(\operatorname{diag}\left\{(\mathbf{D} \bullet \mathbf{D})\mathbf{R}\right\}^{\mathrm{T}}\boldsymbol{\Theta}_{11}(\mathbf{w}, \overline{\mathbf{w}}) - \operatorname{diag}\left\{(\overline{\mathbf{D}} \bullet \overline{\mathbf{D}})\mathbf{R}\right\}^{\mathrm{T}}\boldsymbol{\Theta}_{22}(\mathbf{w}, \overline{\mathbf{w}}) - \\ & 2\operatorname{diag}\left\{(\mathbf{D} \bullet \overline{\mathbf{D}})\mathbf{R}\right\}^{\mathrm{T}}\boldsymbol{\Theta}_{12}(\mathbf{w}, \overline{\mathbf{w}})\Big) - \frac{1}{N}\mathbf{1}^{\mathrm{T}}\mathbf{y} - \sigma^2\end{aligned} \tag{8}$$

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

where

- ☐ *The* $L \times 1$ *vector* $\boldsymbol{\Theta}_1(\mathbf{w}, \overline{\mathbf{w}}) = \left[\frac{\partial \theta_l(w_l, \overline{w}_l)}{\partial w_l}\right]_{l \in [1...L]}$ *is made of the first derivative, with respect to its* first *variable, of each function* θ_l.
- ☐ *The* $L \times 1$ *vector* $\boldsymbol{\Theta}_2(\mathbf{w}, \overline{\mathbf{w}}) = \left[\frac{\partial \theta_l(w_l, \overline{w}_l)}{\partial \overline{w}_l}\right]_{l \in [1...L]}$ *is made of the first derivative, with respect to its* second *variable, of each function* θ_l.
- ☐ *The* $L \times 1$ *vector* $\boldsymbol{\Theta}_{12}(\mathbf{w}, \overline{\mathbf{w}}) = \left[\frac{\partial^2 \theta_l(w_l, \overline{w}_l)}{\partial w_l \partial \overline{w}_l}\right]_{l \in [1...L]}$ *is made of the first derivative, with respect to its first variable and second variable, of each function* θ_l.
- ☐ *The* $L \times 1$ *vector* $\boldsymbol{\Theta}_{11}(\mathbf{w}, \overline{\mathbf{w}}) = \left[\frac{\partial^2 \theta_l(w_l, \overline{w}_l)}{\partial w_l^2}\right]_{l \in [1...L]}$ *is made of the second derivative, with respect to its* first *variable, of each function* θ_l.
- ☐ *The* $L \times 1$ *vector* $\boldsymbol{\Theta}_{22}(\mathbf{w}, \overline{\mathbf{w}}) = \left[\frac{\partial^2 \theta_l(w_l, \overline{w}_l)}{\partial \overline{w}_l^2}\right]_{l \in [1...L]}$ *is made of the second derivative, with respect to its* second *variable, of each function* θ_l.
- ☐ *The operator "*$\bullet$*" denotes the Hadamard (element-by-element) product between two matrices.*

Proof: Using the result given in (6) and the fact that $\boldsymbol{\Theta}(\mathbf{w}, \overline{\mathbf{w}}) = [\theta_l(w_l, \overline{w}_l)]_{l \in [1...L]}$, we can further develop $\partial \mathbf{f}(\mathbf{y}) = \left[\frac{\partial f_n(\mathbf{y})}{\partial y_n}\right]_{1 \leq n \leq N}$ as

$$\frac{\partial f_n \mathbf{y})}{\partial y_n} = \sum_{l=1}^{L} r_{n,l} \Big(\underbrace{\frac{\partial \theta_l(w_l, \overline{w}_l)}{\partial w_l}}_{[\boldsymbol{\Theta}_1(\mathbf{w}, \overline{\mathbf{w}})]_l} d_{l,n} + \underbrace{\frac{\partial \theta_l(w_l, \overline{w}_l)}{\partial \overline{w}_l}}_{[\boldsymbol{\Theta}_2(\mathbf{w}, \overline{\mathbf{w}})]_l} \overline{d}_{l,n} \Big). \tag{9}$$

Similarly, we have that

$$\frac{\partial^2 f_n(\mathbf{y})}{\partial y_n^2} = \sum_{l=1}^{L} r_{n,l} \Big(\underbrace{\frac{\partial^2 \theta_l(w_l, \overline{w}_l)}{\partial w_l^2}}_{[\boldsymbol{\Theta}_{11}(\mathbf{w}, \overline{\mathbf{w}})]_l} d_{l,n}^2 + \underbrace{\frac{\partial^2 \theta_l(w_l, \overline{w}_l)}{\partial \overline{w}_l^2}}_{[\boldsymbol{\Theta}_{22}(\mathbf{w}, \overline{\mathbf{w}})]_l} \overline{d}_{l,n}^2 + 2 \underbrace{\frac{\partial^2 \theta_l(w_l, \overline{w}_l)}{\partial w_l \partial \overline{w}_l}}_{[\boldsymbol{\Theta}_{12}(\mathbf{w}, \overline{\mathbf{w}})]_l} d_{l,n} \overline{d}_{l,n} \Big), \tag{10}$$

and consequently that

$$\begin{aligned} \mathbf{y}^{\mathrm{T}} \partial \mathbf{f}(\mathbf{y}) &= \sum_{l=1}^{L} [\boldsymbol{\Theta}_1(\mathbf{w}, \overline{\mathbf{w}})]_l \sum_{n=1}^{N} \underbrace{d_{l,n} r_{n,l}}_{[\mathbf{D} \bullet \mathbf{R}^{\mathrm{T}}]_{l,n}} y_n + \sum_{l=1}^{L} [\boldsymbol{\Theta}_2(\mathbf{w}, \overline{\mathbf{w}})]_l \sum_{n=1}^{N} \underbrace{\overline{d}_{l,n} r_{n,l}}_{[\overline{\mathbf{D}} \bullet \mathbf{R}^{\mathrm{T}}]_{l,n}} y_n \\ &= \boldsymbol{\Theta}_1(\mathbf{w}, \overline{\mathbf{w}})^{\mathrm{T}} (\mathbf{D} \bullet \mathbf{R}^{\mathrm{T}}) \mathbf{y} + \boldsymbol{\Theta}_2(\mathbf{w}, \overline{\mathbf{w}})^{\mathrm{T}} (\overline{\mathbf{D}} \bullet \mathbf{R}^{\mathrm{T}}) \mathbf{y} \end{aligned} \tag{11}$$

$$\begin{aligned} \operatorname{div}\{\mathbf{f}(\mathbf{y})\} &= \sum_{l=1}^{L} [\boldsymbol{\Theta}_1(\mathbf{w}, \overline{\mathbf{w}})]_l \sum_{n=1}^{N} \underbrace{d_{l,n} r_{n,l}}_{[\mathbf{D}\mathbf{R}]_{l,l}} + \sum_{l=1}^{L} [\boldsymbol{\Theta}_2(\mathbf{w}, \overline{\mathbf{w}})]_l \sum_{n=1}^{N} \underbrace{\overline{d}_{l,n} r_{n,l}}_{[\overline{\mathbf{D}}\mathbf{R}]_{l,l}} \\ &= \operatorname{diag}\{\mathbf{D}\mathbf{R}\}^{\mathrm{T}} \boldsymbol{\Theta}_1(\mathbf{w}, \overline{\mathbf{w}}) + \operatorname{diag}\{\overline{\mathbf{D}}\mathbf{R}\}^{\mathrm{T}} \boldsymbol{\Theta}_2(\mathbf{w}, \overline{\mathbf{w}}) \end{aligned} \tag{12}$$

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

$$\begin{aligned}\operatorname{div}\{\partial\mathbf{f}(\mathbf{y})\} &= \sum_{l=1}^{L}[\boldsymbol{\Theta}_{11}(\mathbf{w},\overline{\mathbf{w}})]_l \sum_{n=1}^{N} \underbrace{d_{l,n}^2 r_{n,l}}_{[(\mathbf{D}\bullet\mathbf{D})\mathbf{R}]_{l,l}} + \sum_{l=1}^{L}[\boldsymbol{\Theta}_{22}(\mathbf{w},\overline{\mathbf{w}})]_l \sum_{n=1}^{N} \underbrace{\overline{d}_{l,n}^2 r_{n,l}}_{[(\overline{\mathbf{D}}\bullet\overline{\mathbf{D}})\mathbf{R}]_{l,l}} + \\ &\quad 2\sum_{l=1}^{L}[\boldsymbol{\Theta}_{12}(\mathbf{w},\overline{\mathbf{w}})]_l \sum_{n=1}^{N} \underbrace{d_{l,n}\overline{d}_{l,n} r_{n,l}}_{[(\mathbf{D}\bullet\overline{\mathbf{D}})\mathbf{R}]_{l,l}} \\ &= \operatorname{diag}\{(\mathbf{D}\bullet\mathbf{D})\mathbf{R}\}^{\mathrm{T}}\boldsymbol{\Theta}_{11}(\mathbf{w},\overline{\mathbf{w}}) + \operatorname{diag}\left\{(\overline{\mathbf{D}}\bullet\overline{\mathbf{D}})\mathbf{R}\right\}^{\mathrm{T}}\boldsymbol{\Theta}_{22}(\mathbf{w},\overline{\mathbf{w}}) + \\ &\quad 2\operatorname{diag}\left\{(\mathbf{D}\bullet\overline{\mathbf{D}})\mathbf{R}\right\}^{\mathrm{T}}\boldsymbol{\Theta}_{12}(\mathbf{w},\overline{\mathbf{w}}). \end{aligned} \tag{13}$$

Putting back (11), (12) and (13) into (6) finally leads to the desired result (8). ■

F. The PURE-LET Strategy

Similarly to what has been proposed for SURE-based denoising [2], [3], [33], [37], we describe the denoising function $\mathbf{f}$ as the *linear expansion of thresholds* (LET) defined as

$$\mathbf{f}(\mathbf{y}) = \sum_{k=1}^{K} a_k \underbrace{\mathbf{R}\boldsymbol{\Theta}_k(\mathbf{w},\overline{\mathbf{w}})}_{\mathbf{f}_k(\mathbf{y})}. \tag{14}$$

Thanks to this linear parameterization, PURE becomes quadratic in the a_k's. Therefore, the search for the optimal (in the minimum PURE sense) vector of parameters $\mathbf{a} = [a_1\ a_2\ \ldots a_k]^{\mathrm{T}}$ boils down to the solution of the following system of K *linear* equations

$$\begin{aligned}\sum_{l=1}^{K} \underbrace{\mathbf{f}_k(\mathbf{y})^{\mathrm{T}}\mathbf{f}_l(\mathbf{y})}_{[\mathbf{M}]_{k,l}} a_l &= \underbrace{\mathbf{y}^{\mathrm{T}}\mathbf{f}_k^{-}(\mathbf{y}) - \sigma^2\operatorname{div}\left\{\mathbf{f}_k^{-}(\mathbf{y})\right\}}_{[\mathbf{c}]_k} \quad \text{for } k = 1\ldots K \\ &\Updownarrow \\ \mathbf{M}\mathbf{a} &= \mathbf{c}. \end{aligned} \tag{15}$$

When using the first-order Taylor-series approximation of PURE defined in Section II-D, we obtain a similar system of K linear equations that involves the same matrix $\mathbf{M}$ and an approximated $\mathbf{c}$ given by

$$\hat{\mathbf{c}} = \left[\mathbf{y}^{\mathrm{T}}\left(\mathbf{f}_k(\mathbf{y}) - \partial\mathbf{f}_k(\mathbf{y})\right) - \sigma^2\operatorname{div}\left\{\mathbf{f}_k(\mathbf{y}) - \partial\mathbf{f}_k(\mathbf{y})\right\}\right]_{k\in[1\ldots K]}.$$

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

III. Pointwise Estimator for Undecimated Filterbank Transforms

In this section, we show how to use the result of Corollary 1 to globally optimize a subband-dependent pointwise thresholding applied to the coefficients of the J-band undecimated filterbank depicted in Fig. 3. In this case, the decomposition and reconstruction matrices are made of J $(N \times N)$ circulant submatrices $\mathbf{D}_j$ and $\mathbf{R}_j$ given by

$$\begin{cases} [\mathbf{D}_j]_{k,l} = \sum_n \tilde{g}_j[l-k+nN] \\ [\mathbf{R}_j]_{k,l} = \sum_n g_j[k-l+nN]. \end{cases} \tag{16}$$

Although all the results presented in this section are derived for 1D signals, they can be straightforwardly extended to higher dimensional signals by considering separable filters.

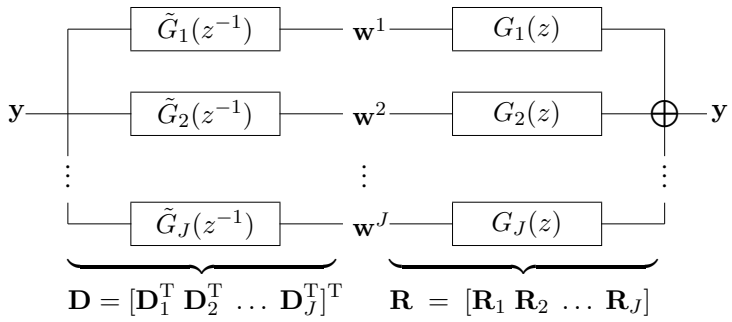

Fig. 3. Undecimated J-band analysis–synthesis filterbank.

A. Choice of $\overline{\mathbf{D}}$—Group-Delay Compensation (GDC)

In an undecimated wavelet transform (UWT), the scaling coefficients of the lowpass residual at a given scale j could be used as a reliable estimation of the signal-dependent noise variance for each bandpass subband from the same scale j, up to the scale-dependent factor $\beta_j = 2^{-j/2}$. However, except for the Haar wavelet filters, there is a group-delay difference between the output of the lowpass and bandpass channels that needs to be compensated for. Contrary to the *highpass* group delay compensation (GDC) filter derived for interscale prediction in critically sampled orthonormal filterbank proposed in [2], we are looking here for an *allpass* filter $Q(z)$ with real coefficients such that $Q(z)^2 = Q(z)/Q(z^{-1})$ and

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

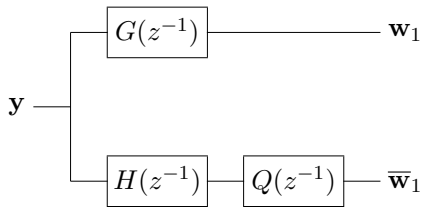

Fig. 4. Group-delay compensation of the lowpass channel of an undecimated filterbank.

(see Fig. 4)

$$H(z^{-1})Q(z^{-1}) = G(z^{-1})R_1(z), \tag{17}$$

where $H(z)$ and $G(z)$ are the orthonormal scaling and wavelet filters, respectively, and $R_1(z) = \varepsilon R_1(z^{-1})$ is an arbitrary symmetric ($\varepsilon = 1$) or anti-symmetric ($\varepsilon = -1$) filter.

In the case of symmetric filters characterized by $H(z^{-1}) = z^{n_0}H(z)$, the shortest-support GDC filter is simply given by $Q(z) = z^{n_0-1}$. Note that this result can also be used for nearly symmetric filters, such as the Daubechies *symlets*. From a theoretical point of view, considering arbitrary wavelet filters would be interesting as well, but would probably be less useful here because, in our observations, the best denoising performances are achieved by symmetric (*Haar*) or nearly symmetric (*sym8*) filters.

The $L \times N$ matrix $\overline{\mathbf{D}}$ is then defined as $\overline{\mathbf{D}} = [\overline{\mathbf{D}}_1^{\mathrm{T}}\ \overline{\mathbf{D}}_2^{\mathrm{T}} \dots \overline{\mathbf{D}}_{J+1}^{\mathrm{T}}]^{\mathrm{T}}$, with

$$\left[\overline{\mathbf{D}}_j\right]_{k,l} = \sum_{n\in\mathbb{Z}} \overline{h}_j[l-k+nN]\ ,$$

where $\overline{h}_j[n]$ is the nth coefficient of the filter

$$\begin{cases} \overline{H}_j(z) &= 2^j Q(z^{2^{j-1}})H_j(z) = Q(z^{2^{j-1}})H(z)H(z^2)\dots H(z^{2^{j-1}}), \text{ for } j = 1\dots J \\ \overline{H}_{J+1}(z) &= 2^J H_J(z) = H(z)H(z^2)\dots H(z^{2^{J-1}}). \end{cases}$$

In an overcomplete M-block discrete cosine transform (BDCT) representation, the lowpass residual band can directly serve as a coarse estimate of the noise variance[3] for each bandpass subband, since the filters of the BDCT all have the same group delay.

[3]Up to the scaling factor $\beta_j = M^{-1/2}$.

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

B. Computation of Transform-Dependent Terms

To compute the MSE estimate (8), we need to evaluate several terms that depend only on the choice of transformation.

1) Computation of $\mathrm{diag}\left\{\mathbf{DR}\right\}$:

   In the case of periodic boundary conditions, we have that

$$\mathrm{diag}\left\{\mathbf{D}_j\mathbf{R}_j\right\} = \begin{cases} \frac{1}{M_j}[\underbrace{1\ 1\ldots 1}_{N \text{ times}}]^{\mathrm{T}} & \text{, for } j = 1\ldots J \\ \frac{1}{M_J}[\underbrace{1\ 1\ldots 1}_{N \text{ times}}]^{\mathrm{T}} & \text{, for } j = J+1 \end{cases}$$

   where, for multiscale filterbanks, $M_j = 2^j$ is the downsampling factor. For an overcomplete BDCT, $M_j = M$, where M is the size of the considered blocks.

2) Computation of $\mathrm{diag}\left\{\overline{\mathbf{D}}\mathbf{R}\right\}$, $\mathrm{diag}\left\{(\mathbf{D}\bullet\mathbf{D})\mathbf{R}\right\}$, $\mathrm{diag}\left\{(\overline{\mathbf{D}}\bullet\overline{\mathbf{D}})\mathbf{R}\right\}$, and $\mathrm{diag}\left\{(\mathbf{D}\bullet\overline{\mathbf{D}})\mathbf{R}\right\}$:

   Contrary to $\mathrm{diag}\left\{\mathbf{DR}\right\}$, the computation of the remaining terms does not generally lead to simple expressions that are independent from the coefficients of the underlying filters. However, all diagonal terms presented in this section can still be easily computed offline, using the numerical Monte Carlo-like algorithm described in [3]. In particular, this numerical scheme applies when the various matrices are not explicitly given, or when non-periodic boundary extensions are performed.

Note that the vectors $(\mathbf{D}\bullet\mathbf{D})\mathbf{y}$, $(\overline{\mathbf{D}}\bullet\overline{\mathbf{D}})\mathbf{y}$, $(\mathbf{D}\bullet\mathbf{R}^{\mathrm{T}})\mathbf{y}$, and $(\overline{\mathbf{D}}\bullet\mathbf{R}^{\mathrm{T}})\mathbf{y}$ are obtained by analyzing the noisy data $\mathbf{y}$ with the considered undecimated filterbank using modified analysis filters with coefficients $\tilde{g}_j^2[k]$, $\overline{h}_j^2[k]$, $\tilde{g}_j[k]\cdot g_j[-k]$, and $\overline{h}_j[k]\cdot g_j[-k]$, respectively.

C. Thresholding Function

We propose to use a subband-dependent pointwise thresholding function defined by

$$\mathbf{\Theta}(\mathbf{w},\overline{\mathbf{w}}) = \left[\theta_j(w_{j,n},\overline{w}_{j,n})\right]_{(j-1)N+n},\ (j,n)\in[1\ldots J]\times[1\ldots N] \tag{18}$$

For the sake of conciseness, we drop the subband subscripts (n,j) in this section and denote by w (resp. $\overline{w}$) any of the $w_{j,n}$'s (resp. $\overline{w}_{j,n}$'s).

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

In the case of Poisson data, we need a signal-dependent transform-domain threshold to take into account the non-stationarity of the noise. If we consider unit-norm filters, the variance σ^2 of the AWGN is preserved in the transformed domain. An estimation of the variance of the Poisson-noise component is given by the magnitude of $|\overline{w}|$ (built as described in Section III-A), up to the scale-dependent factors $\beta_j = 2^{-j/2}$ and $\beta_j = M^{-1/2}$ for a multiscale transform and for an overcomplete BDCT, respectively. Therefore, we propose the signal-dependent threshold

$$t_j(\overline{w}) = \sqrt{\beta_j |\overline{w}| + \sigma^2} \tag{19}$$

which is then embedded in a subband-dependent thresholding function, similar to the one proposed in [3] for AWGN reduction in redundant representations

$$\theta_j(w, \overline{w}) = a_{j,1} \cdot \underbrace{w}_{\theta_{j,1}(w,\overline{w})} + a_{j,2} \cdot \underbrace{w \exp\left(-\left(\frac{w}{3t_j(\overline{w})}\right)^8\right)}_{\theta_{j,2}(w,\overline{w})}. \tag{20}$$

To compute the MSE estimate given in (8), a differentiable (at least up to the second order) approximation of the absolute-value function is required. In practice, we suggest to use $\tanh(k\overline{w})\overline{w}$ ($k = 100$, typically) instead of $|\overline{w}|$ in the threshold defined in (19). As observed in Fig. 5, the proposed thresholding function can be perceived as a smooth hard-thresholding. Thanks to the use of a signal-dependent threshold, each transformed coefficient is adaptively thresholded according to its estimated amount of noise.

D. Implementation

In the generalized PURE-LET framework for J-band undecimated filterbank, the whole transform-domain thresholding is rewritten as

$$\mathbf{f}(\mathbf{y}) = \sum_{j=1}^{J} \sum_{k=1}^{2} a_{j,k} \underbrace{\mathbf{R}_j \mathbf{\Theta}_{j,k}(\mathbf{D}_j \mathbf{y}, \overline{\mathbf{D}}_j \mathbf{y})}_{\mathbf{f}_{j,k}(\mathbf{y})} + \underbrace{\mathbf{R}_{J+1} \mathbf{D}_{J+1} \mathbf{y}}_{\text{lowpass}}. \tag{21}$$

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

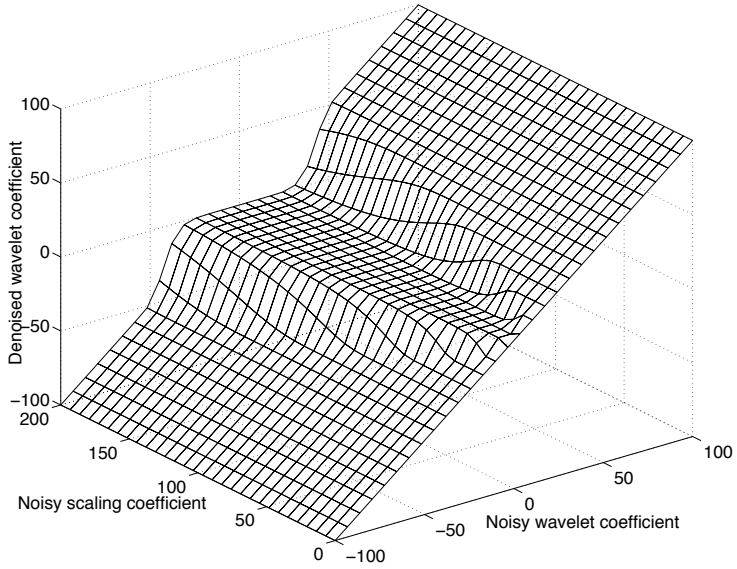

Fig. 5. A possible realization of the proposed thresholding function (20) in a particular subband of an undecimated Haar wavelet representation.

The parameters $a_{j,k}$ that minimize the MSE estimate $\hat{\epsilon}$ given in (6) are then the solution of the system of $2J$ linear equations $\mathbf{Ma} = \hat{\mathbf{c}}$, where, for $\mathbf{k}, \mathbf{l} \in [1 \dots J] \times [1\ 2]$,

$$\begin{cases} \mathbf{M} &= \left[\mathbf{f_k}(\mathbf{y})^{\mathrm{T}} \mathbf{f_l}(\mathbf{y})\right]_{2(k_1-1)+k_2, 2(l_1-1)+l_2} \\ \hat{\mathbf{c}} &= \Big[\mathbf{f_k}(\mathbf{y})^{\mathrm{T}} (\mathbf{I} - \mathbf{R}_{J+1}\mathbf{D}_{J+1})\mathbf{y} - \\ & \quad \left(\mathbf{y}^{\mathrm{T}} \partial \mathbf{f_k}(\mathbf{y}) + \sigma^2 \mathrm{div}\left\{\mathbf{f_k}(\mathbf{y}) - \partial \mathbf{f_k}(\mathbf{y})\right\}\right) \Big]_{2(k_1-1)+k_2}. \end{cases} \tag{22}$$

For very low intensity signals (typically such that $x_n < 5$, $\forall n$), the first-order Taylor-series approximation of some nonlinear functions $\mathbf{f}_{j,2}^{-}$ might be inaccurate, leading (22) to be an unreliable approximation of the $(j,2)$th component of the PURE vector $\mathbf{c}$ given in (15). To illustrate this scenario in the case of an undecimated Haar wavelet representation, we have plotted in the first row of Fig. 6 the $\mathrm{SNR} = 10 \log_{10} \frac{c_{j,2}^2}{(\hat{c}_{j,2} - c_{j,2})^2}$ between $c_{j,2}$ (for $j = 1$ and $j = 3$) and its approximation $\hat{c}_{j,2}$, as a function of the average noise-free signal energy $E_{\mathrm{mean}} = \|\mathbf{x}\|^2/N = \mathscr{E}_{\mathbf{y}}\left\{\|\mathbf{y}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{y}\right\}/N - \sigma^2$. Any nonlinearly processed subband $\mathbf{f}_{j,2}(\mathbf{y})$ for which the SNR of the approximation is below a given threshold (around

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

40 dB, typically), should be disregarded from the LET (*i.e.*, its LET coefficient $a_{j,2}$ set to zero). In practice, we need to identify the badly approximated functions without knowing the SNR of the approximation. For this identification, we propose an empirical approach that is based on the following observation: We notice in the first row of Fig. 6 that the SNR of the approximation is an increasing function of the average signal energy. We further observe that, for a given average signal energy, the quality of the approximation increases as the iteration depth j increases. Consequently, we suggest to keep only those nonlinearly processed subbands $\mathbf{f}_{j,2}(\mathbf{y})$ for which $T_j = M_j E_{\text{mean}}$ is above a given threshold T. We experimentally found that any value $T \in [5, 15]$ was a reasonable choice[4] (see the second row of Fig. 6). Thanks to this restriction, the PURE-based optimization of the parameters gives similar PSNR results to the MSE-based optimization, even when the latter considers all the nonlinearly processed subbands $\mathbf{f}_{j,2}(\mathbf{y})$, as shown in Fig. 7.

E. Denoising in Mixed Bases

In order to get the best out of several transforms, we propose to make the LET span several transformed domains with complementary properties (*e.g.*, UWT and overcomplete BDCT)

$$\mathbf{f}(\mathbf{y}) = \sum_{k=1}^{K_1} a_k \mathbf{f}_k^{\text{UWT}}(\mathbf{y}) + \sum_{k=1}^{K_2} b_k \mathbf{f}_k^{\text{BDCT}}(\mathbf{y}) + \dots \tag{23}$$

PURE is then used to jointly optimize the weights of this linear combination of processed subbands. In this case, the union of several transforms can be interpreted as an overcomplete dictionary of bases which sparsely represents a wide class of natural images. The idea of combining several complementary transforms was exploited in the context of AWGN reduction by Starck *et al.* in [38] and Fadili *et al.* in [39]. The use of an overcomplete dictionary, either fixed in advance (as in our case) or trained, is at the core of the *K-SVD*-based denoising algorithm of Elad *et al.* [40].

[4]For all the experiments presented in this paper, we have used $T = 10$.

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

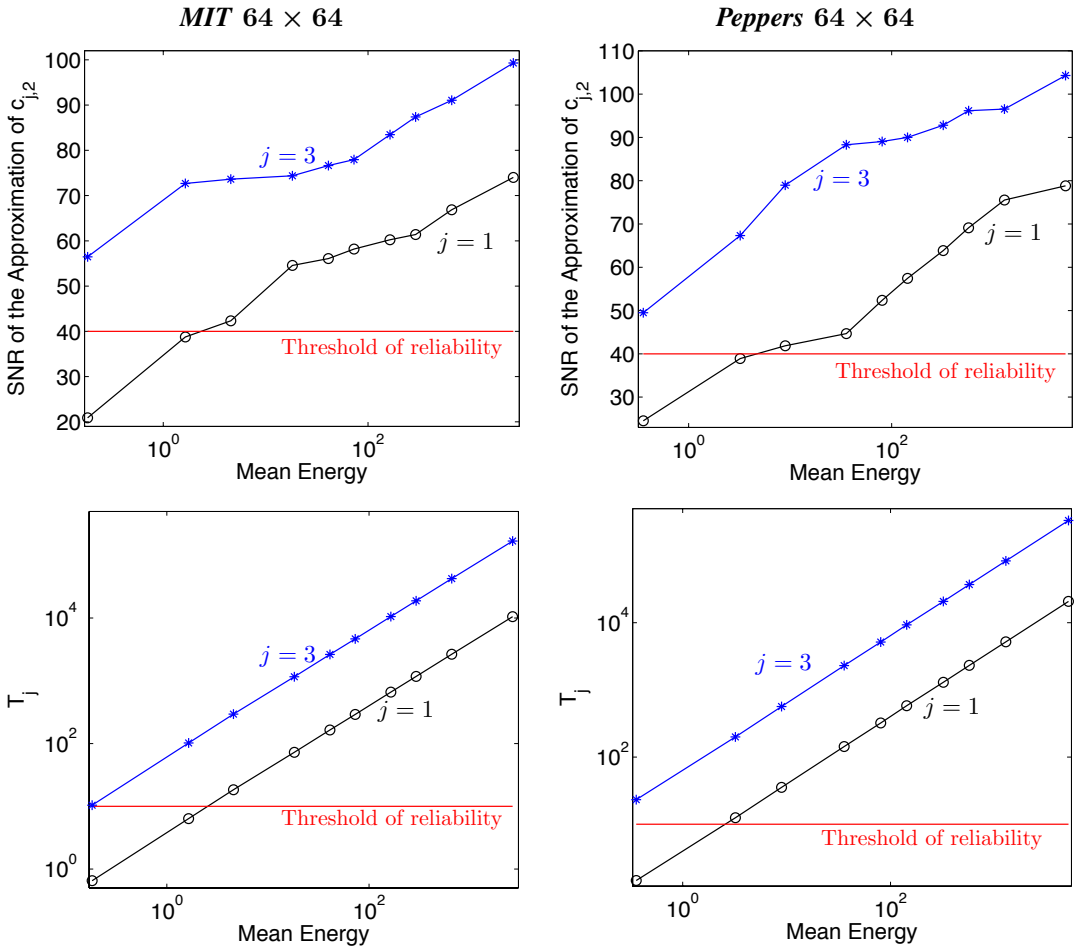

Fig. 6. Validity of the first-order Taylor-series approximation of some nonlinear functions $\mathbf{f}_{j,2}^{-}$. The first row shows the SNR of the approximation of the component $c_{j,2}$ of the PURE vector $\mathbf{c}$. The values of $T_j = M_j E_{\text{mean}}$ are plotted in the second row.

IV. Simulations

In this section, we propose to compare our PURE-LET approach with three multiscale-based methods in simulated experiments.

- A variant of the Haar-Fisz algorithm described in [10]: The Haar-Fisz variance-stabilizing transform (VST) is followed by Donoho's *SUREshrink* [18] applied to each bandpass subband of an undecimated Haar wavelet transform. Since the Haar-Fisz transform is not shift-invariant, 20 "external"

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

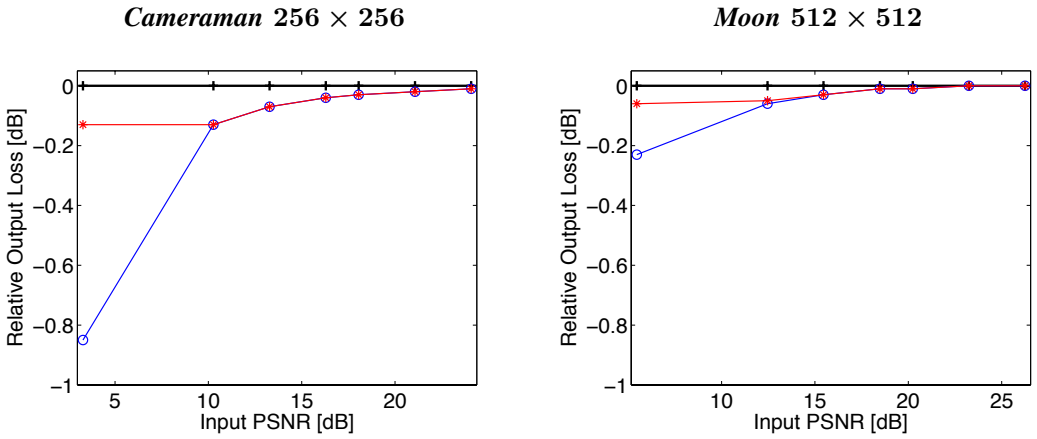

Fig. 7. Reliability of the PURE-based optimization of the LET parameters. The blue curve ("$\circ$" markers) represents the output PSNRs obtained by a PURE-based optimization of the LET parameters without restriction, taking all nonlinear terms into consideration. The red curve ("$*$" markers) displays the output PSNRs obtained by a restricted PURE-based optimization of the LET parameters, where all the nonlinearly processed subband $\mathbf{f}_{j,2}(\mathbf{y})$ with $T_j < 10$ have been disregarded. The output PSNRs obtained by the MSE-based optimization of the LET parameters (without restriction) is the baseline ("$+$" markers).

cycle-spins (CS) are also applied to the whole algorithm (Haar-Fisz VST + UWT-*SUREshrink* + Haar-Fisz inverse VST), as suggested in [10].

- A standard variance-stabilizing transform followed by a high-quality AWGN denoiser: As VST, we have retained the generalized Anscombe transform (GAT) [9], which can also stabilize Poisson random variables degraded by AWGN. For the denoising part, we have considered Portilla *et. al. BLS-GSM* [1], which consists of a multivariate estimator resulting from Bayesian least-squares (BLS) optimization, assuming Gaussian scale mixtures (GSM) as a prior for neighborhoods of coefficients at adjacent positions and scales[5]. Applied in a full steerable pyramid, this algorithm stands among the best state-of-the-art multiresolution-based methods for AWGN reduction.
- A state-of-the-art denoising algorithm specifically designed for Poisson intensity estimation: We have retained Willett and Nowak's *Platelet* approach [28]. Their Poisson intensity estimation consists in optimizing (through a penalized likelihood) the coefficients of polynomials fitted to a recursive

[5]Matlab code available at: `http://www.io.csic.es/PagsPers/JPortilla/denoise/software/index.htm`

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

dyadic partition of the support of the Poisson intensity[6]. The near shift-invariance of their approach is achieved by averaging the denoising results of several cycle-spins (20, in our experiments).

The tuning parameters of these three methods have been set according to the values given in the corresponding papers/Matlab codes.

We have tested our PURE-LET approach with three undecimated transformations: the Haar UWT (suitable for piecewise-smooth images), the 16×16 overcomplete BDCT (efficient for textured images), and a dictionary containing the basis functions of both transforms. This latter option allows to take the best (in the minimum PURE sense) combination of the two complementary transformations. As a measure of quality, we use the peak signal-to-noise ratio (PSNR), defined as

$$\text{PSNR} = 10 \log_{10} \left(\frac{I_{\max}^2}{\text{MSE}} \right),$$

where $I_{\max}$ is the maximum intensity of the noise-free image. Various input PSNRs have been obtained by rescaling the original images between $I_{\max} = 1$ and $I_{\max} = 120$.

A. Poisson Noise Reduction

In Table I, we compare our method with the state-of-the-art multiresolution-based denoising algorithms for Poisson-intensity estimation (no further AWGN degradation). When available, we have also reported the results recently obtained by Lefkimmiatis *et al.* with their multiscale Poisson-Haar hidden Markov tree (PH-HMT) algorithm [14]. As observed, we obtain, on average, the best PSNR results. The Haar-Fisz algorithm is consistently outperformed by our method, and usually by the two algorithms specifically designed for Poisson-intensity estimation. We can notice that the *Anscombe+BLS-GSM* solution becomes more competitive as the peak intensities increase, whereas the other approaches are not restricted to high-intensity images. We obtain similar or better results than the *Platelet* and the *PH-HMT* methods, which are also competitive with the recent approach described in [30]. However, the major drawback of these two recent approaches is their computation time. Denoising the 256×256 grayscale *Cameraman* image

[6]Matlab code available at: `http://nislab.ee.duke.edu/NISLab/Platelets.html`

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

at peak intensity 20 requires around 1300s (65s per cycle-spin) for the *Platelet* and 92s for the *PH-HMT*. By contrast, the execution of 20 cycle-spins of the Haar-Fisz algorithm lasts ~ 5.7s, the *Anscombe+BLS-GSM* takes ~ 7.7s, whereas our PURE-LET algorithm lasts 1.3s using the Haar UWT, 12.2s in the 16×16 BDCT, and 13.5s considering the UWT/BDCT dictionary, respectively. In most cases, the simple Haar UWT gives optimal or nearly optimal results, except for textured images at relatively high input PSNR.

In Fig. 8, we show a visual result of the various algorithms applied on *Moon* image. We can notice that the PURE-LET denoised image exhibits very few artifacts, without over-smoothing, contrary to the other methods.

B. Mixed Poisson-Gaussian Noise Reduction

We show in Table II the PSNR results obtained by *GAT+BLS-GSM* and by the proposed algorithm, when estimating Poisson intensity degraded by AWGN. Here again, the best results are obtained by our PURE-LET approach, on average. In particular, at low intensities, the GAT fails to stabilize the noise variance and thus, huge gains ($>>$ 1 dB) can be obtained with a direct handling of Poisson statistics. The visual comparison of both algorithms is given in Fig. 9 for *Barbara* image. The various textures present in this particular image are better restored with the proposed PURE-LET algorithm applied in a BDCT representation than with the competing methods.

V. Application to Real Fluorescence Microscopy Images

We propose now to apply our denoising algorithm to real fluorescence microscopy images (Fig. 10A), acquired on a Leica SP5 confocal laser scanning microscope at the Department of Biology of the Chinese University of Hong Kong. The dataset contains 100 instances of 512×512 images of fixed tobacco cells, labeled with two fluorescent dyes (*GFP* in the green channel and *Alexa568* in the red channel). The average of the 100 images provides a "ground truth" for visual comparison (Fig. 10B). We refer to [41] for a general introduction to fluorescence microscopy and to [42] for a detailed investigation on the various sources of noise in confocal microscopy.

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

TABLE I

COMPARISON OF MULTISCALE-BASED POISSON NOISE REMOVAL ALGORITHMS

Peak Intensity	120	60	30	20	10	5	2	1
Image	***Cameraman*** 256×256							
Input PSNR	24.08	21.07	18.05	16.29	13.28	10.27	6.29	3.28
Haar-Fisz [10]	29.73	27.98	26.35	25.45	23.70	22.55	20.77	19.77
Anscombe [7] + BLS-GSM [1]	30.85	29.13	27.54	26.56	24.63	22.50	19.07	14.44
Platelet [28]	30.54	29.10	27.66	26.80	25.14	23.56	**21.72**	**20.57**
PH-HMT [14]	N/A	N/A	N/A	26.61	24.97	23.37	21.41	20.03
UWT PURE-LET	31.03	29.29	27.67	26.72	25.10	23.50	21.67	20.44
BDCT PURE-LET	30.72	28.92	27.25	26.30	24.77	23.32	21.57	20.40
UWT/BDCT PURE-LET	**31.35**	**29.58**	**27.91**	**26.93**	**25.29**	**23.65**	**21.72**	20.48
Image	***Barbara*** 512×512							
Input PSNR	24.00	20.99	17.98	16.22	13.21	10.20	6.22	3.21
Haar-Fisz [10]	28.36	26.59	25.06	24.28	23.09	22.18	21.24	20.47
Anscombe [7] + BLS-GSM [1]	31.24	29.40	27.55	26.46	24.65	22.61	19.90	14.97
Platelet [28]	28.62	26.53	24.59	23.74	22.96	22.33	21.47	20.65
PH-HMT [14]	N/A	N/A	N/A	24.92	23.45	22.33	21.27	20.48
UWT PURE-LET	29.55	27.52	25.69	24.80	23.56	22.61	21.50	20.74
BDCT PURE-LET	31.75	29.96	28.23	27.23	25.52	23.87	21.83	20.74
UWT/BDCT PURE-LET	**31.86**	**30.08**	**28.34**	**27.33**	**25.59**	**23.89**	**21.87**	**20.79**
Image	***Fluorescent Cells*** 512×512							
Input PSNR	28.25	25.23	22.22	20.46	17.45	14.44	10.47	7.45
Haar-Fisz [10]	34.15	32.61	31.28	30.43	29.03	27.48	25.96	24.58
Anscombe [7] + BLS-GSM [1]	**35.24**	**33.54**	31.67	30.57	28.30	25.51	19.38	13.25
Platelet [28]	34.04	32.50	31.05	30.30	28.86	27.77	26.19	25.11
UWT PURE-LET	34.98	33.29	31.72	30.84	29.41	28.10	26.43	25.22
BDCT PURE-LET	35.17	33.44	31.82	30.92	29.47	28.11	26.39	25.19
UWT/BDCT PURE-LET	35.23	33.52	**31.91**	**31.00**	**29.52**	**28.18**	**26.47**	**25.25**
Image	***Moon*** 512×512							
Input PSNR	26.26	23.24	20.25	18.48	15.46	12.46	8.47	5.46
Haar-Fisz [10]	29.70	28.04	26.68	25.44	24.48	24.13	23.71	23.46
Anscombe [7] + BLS-GSM [1]	29.78	27.99	26.27	25.43	24.17	23.17	19.70	13.98
Platelet [28]	28.80	27.40	26.20	25.58	24.74	24.14	23.49	23.29
UWT PURE-LET	29.98	28.35	26.95	26.24	25.21	24.47	23.84	23.47
BDCT PURE-LET	29.78	28.16	26.81	26.14	25.21	**24.52**	**23.85**	23.41
UWT/BDCT PURE-LET	**30.10**	**28.46**	**27.05**	**26.33**	**25.30**	**24.52**	**23.85**	**23.49**

Output PSNRs have been averaged over 10 noise realizations, except for the *Platelet* algorithm.

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

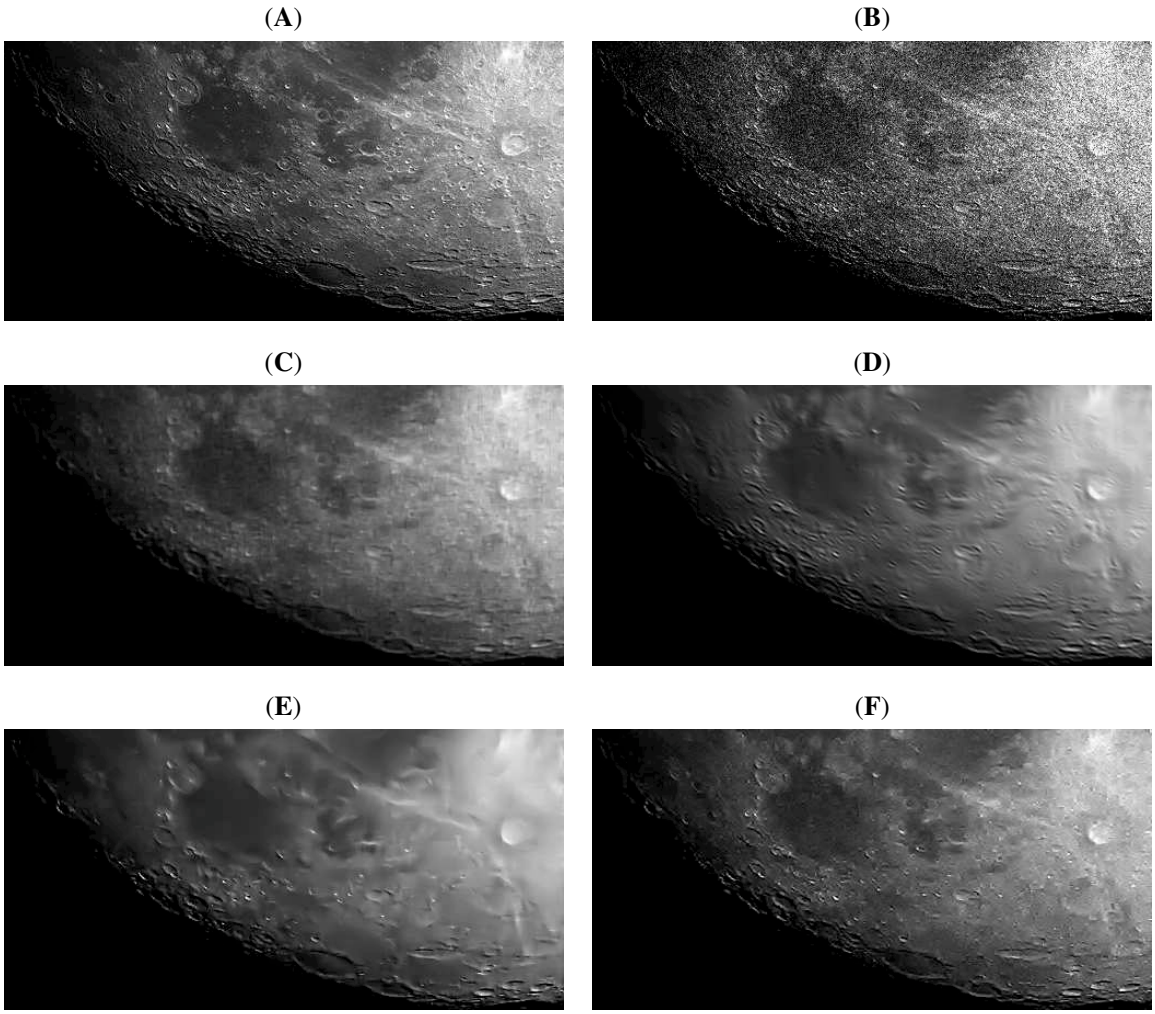

Fig. 8. (**A**) Part of the original *Moon* image at peak intensity 20. (**B**) Noisy version, PSNR $=$ 18.48 dB. (**C**) Denoised with a translation-invariant *Haar-Fisz* [10] algorithm, PSNR $=$ 25.44 dB in 28s. (**D**) Denoised with *Anscombe* [7] + *BLS-GSM* [1], PSNR $=$ 25.45 dB in 38s. (**E**) Denoised with translation-invariant *Platelet* [28], PSNR $=$ 25.66 dB in 2234s. (**F**) Denoised with the proposed UWT PURE-LET, PSNR $=$ 26.23 dB in 7.5s.

In order for our data to fit the noise model described in (1), we need to first subtract the offset value δ of the CCD detectors, and then, divide by their amplification factor α. These two parameters, as well as the variance σ^2 of the AWGN, are determined from the data. In practice, a robust linear regression is first performed on a collection of local estimates of the sample mean and sample variance [43], [44]; then,

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

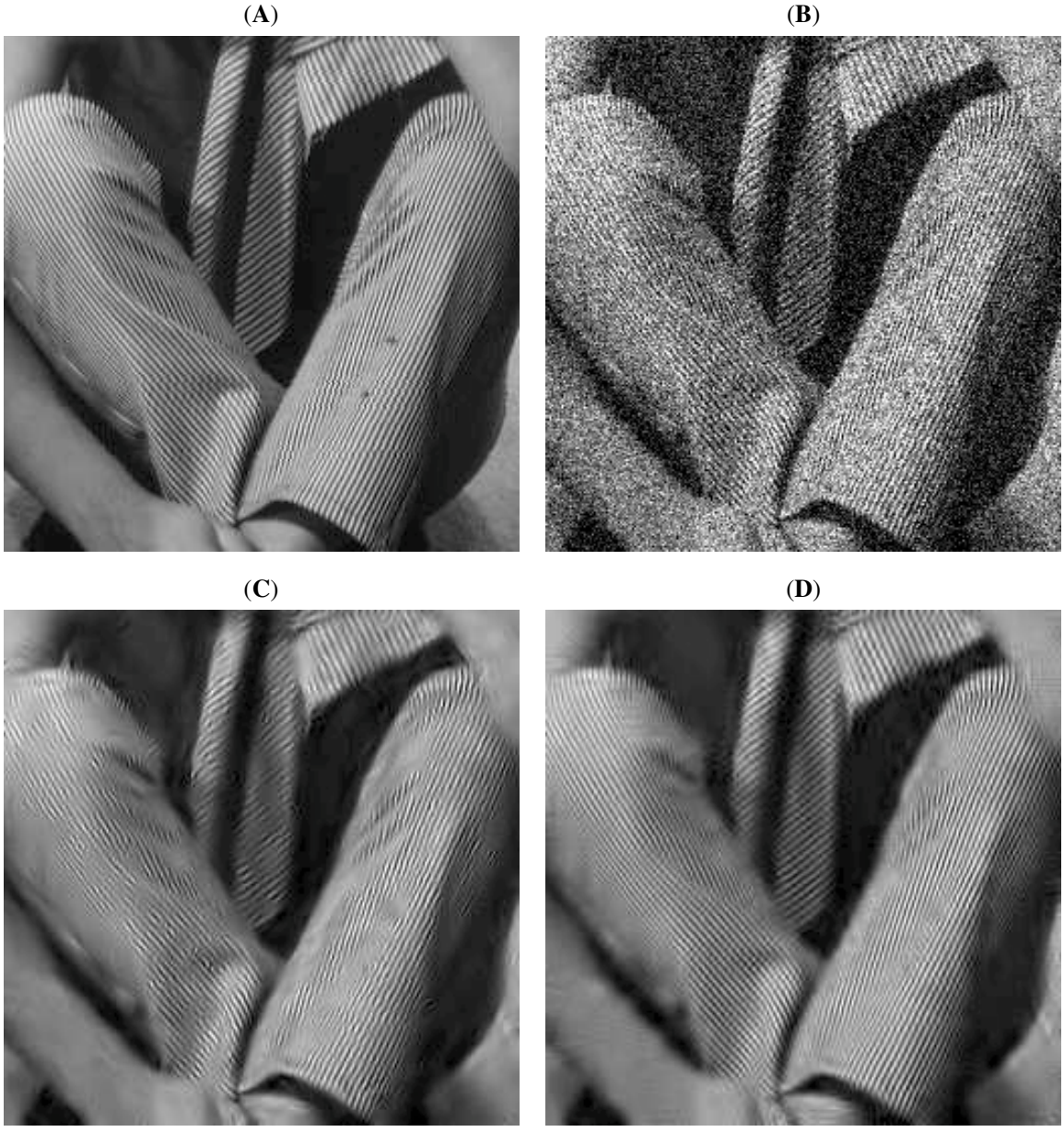

Fig. 9. (**A**) Part of the original *Barbara* image at peak intensity 30. (**B**) Noisy version, PSNR $= 15.87$ dB. (**C**) Denoised with GAT [9] + *BLS-GSM* [1], PSNR $= 26.19$ dB in 32s. (**D**) Denoised with the proposed BDCT PURE-LET, PSNR $= 27.00$ dB in 47s.

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

TABLE II

COMPARISON OF MULTISCALE-BASED MIXED POISSON-GAUSSIAN NOISE REMOVAL ALGORITHMS

Peak Intensity	120	60	30	20	10	5	2	1
$\boldsymbol{\sigma}$	12	6	3	2	1	0.5	0.2	0.1
Image	***Cameraman*** 256×256							
Input PSNR	18.56	17.48	15.90	14.75	12.45	9.83	6.11	3.19
GAT [9] + BLS-GSM [1]	27.56	27.02	26.19	25.59	24.43	22.58	19.39	14.63
UWT PURE-LET	27.68	27.14	26.32	25.74	24.56	23.22	21.57	20.42
BDCT PURE-LET	27.17	26.64	25.87	25.35	24.24	23.05	21.47	20.36
UWT/BDCT PURE-LET	**27.92**	**27.37**	**26.53**	**25.95**	**24.74**	**23.36**	**21.61**	**20.44**
Image	***Fluorescent Cells*** 512×512							
Input PSNR	19.39	18.86	17.96	17.21	15.53	13.38	10.01	7.22
GAT [9] + BLS-GSM [1]	**30.85**	30.47	29.67	29.12	28.31	25.67	20.07	13.49
UWT PURE-LET	30.68	30.37	29.86	29.46	28.63	27.67	26.23	25.07
BDCT PURE-LET	30.64	30.33	29.83	29.44	28.64	27.67	26.21	25.09
UWT/BDCT PURE-LET	30.82	**30.50**	**29.98**	**29.56**	**28.71**	**27.73**	**26.28**	**25.15**

Output PSNRs have been averaged over 10 noise realizations.

the slope of the fitted line gives the amplification factor α. The parameters δ and σ^2 can be estimated independently in signal-free regions of the image.

We have applied our UWT PURE-LET algorithm independently on each channel of a single image from the dataset. As observed in Fig. 10D, the proposed algorithm considerably reduces the level of noise, and is still able to recover most of the activated red spots. By contrast, we also display in Fig. 10C the result obtained after 20 CS of the *Platelet* algorithm, where it can be seen that the noise has been effectively reduced but fewer activated red spots have been preserved.

VI. CONCLUSION

In this paper, we have provided an unbiased estimate of the MSE for the estimation of Poisson intensities degraded by AWGN, a practical measure of quality that we called PURE. We have then defined a generic PURE-LET framework for designing and jointly optimizing a broad class of (redundant) transform-domain nonlinear processing. To obtain a computationally fast and efficient algorithm for undecimated filterbank

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

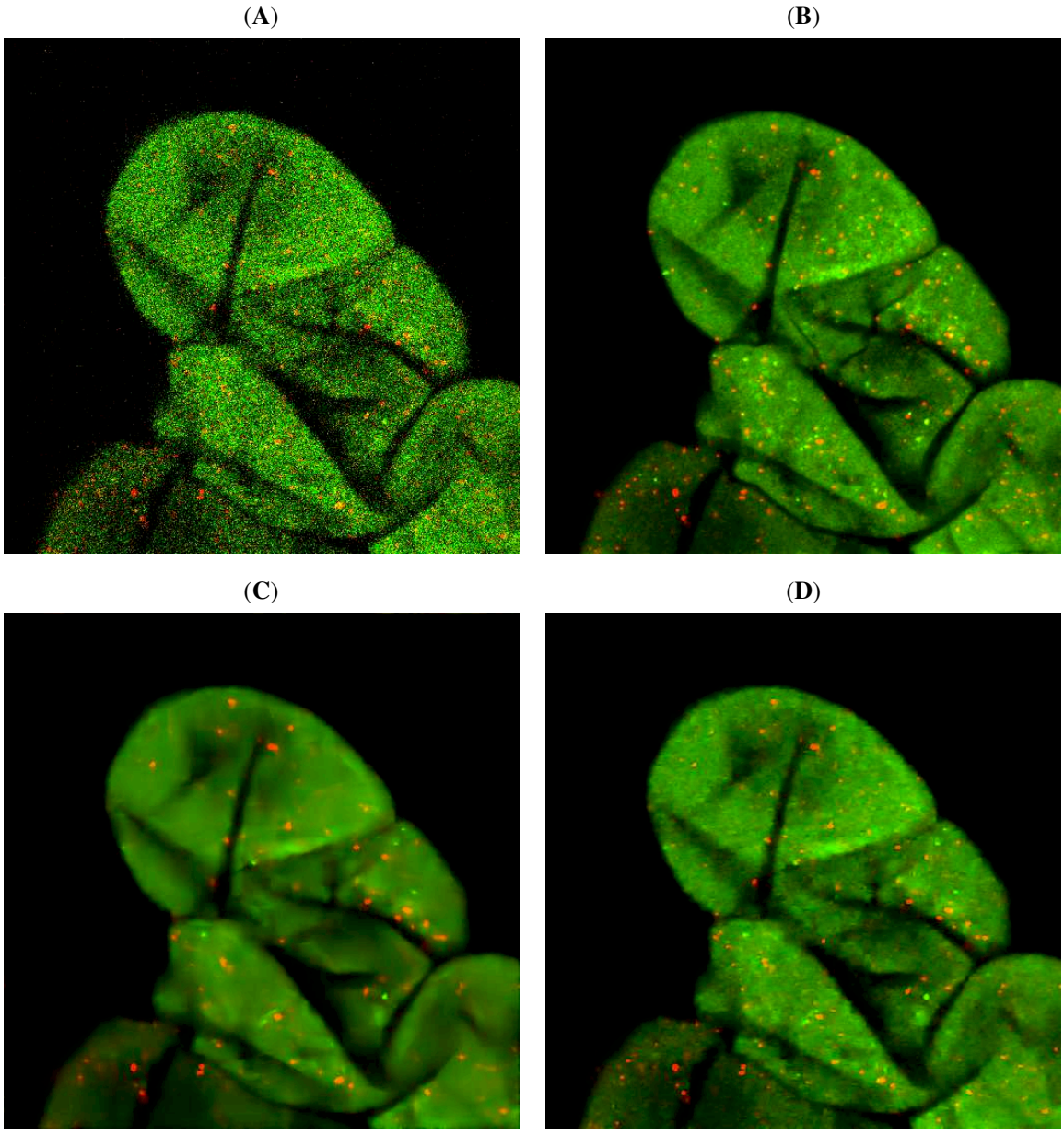

Fig. 10. (**A**) Observed data: single acquisition (two-color image). (**B**) "Ground truth" (average of 100 acquisitions). (**C**) Single acquisition denoised with *Platelets* [28] in 1240s. (**D**) Single acquisition denoised with the proposed UWT PURE-LET in 15s.

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

transforms, we have proposed a first-order Taylor-series approximation of PURE. For each nonlinearly processed subband, the reliability of this approximation can be controlled, ensuring near optimal MSE performances for the considered class of algorithms.

As an illustrative algorithm, we have devised a pointwise subband-dependent thresholding. In each bandpass subband, the amount of shrinkage is related to the signal-dependent noise variance, estimated from suitably chosen lowpass coefficients. The resulting denoising algorithm is faster than other state-of-the-art approaches that make use of redundant transforms. It gives better results, both quantitatively (lower MSE) and qualitatively, than the standard VST-based algorithms. The proposed solution also favorably compares with some of the most recent multiscale methods specifically devised for Poisson data.

Finally, we have shown that our PURE-LET strategy constitutes a competitive solution for fast and high-quality denoising of real fluorescence microscopy data.

VII. Acknowledgments

The authors are especially grateful to Prof. Liwen Jiang and Dr. Yansong Miao from the Department of Biology of the Chinese University of Hong Kong for the preparation of the biological samples and for their help on acquiring the fluorescence microscopy images.

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 Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, November 2003.

[2] F. Luisier, T. Blu, and M. Unser, "A New SURE Approach to Image Denoising: Interscale Orthonormal Wavelet Thresholding," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 593–606, March 2007.

[3] T. Blu and F. Luisier, "The SURE-LET Approach to Image Denoising," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2778–2786, November 2007.

[4] T. Le, R. Chartrand, and T. J. Asaki, "A Variational Approach to Reconstructing Images Corrupted by Poisson Noise," *Journal of Mathematical Imaging and Vision*, vol. 27, no. 3, pp. 257–263, April 2007.

[5] R. H. Chan and K. Chen, "Multilevel Algorithm for a Poisson Noise Removal Model with Total-Variation Regularization," *International Journal of Computer Mathematics*, vol. 00, no. 5, pp. 1–18, May 2007.

[6] A. Sawatzky, C. Brune, J. Müller, and M. Burger, *Total Variation Processing of Images with Poisson Statistics*. Springer Berlin / Heidelberg, 2009, ch. Computer Analysis of Images and Patterns, pp. 533–540.

[7] F. J. Anscombe, "The Transformation of Poisson, Binomial and Negative-Binomial Data," *Biometrika*, vol. 35, no. 3/4, pp. 246–254, December 1948.

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

[8] D. L. Donoho, "Nonlinear Wavelet Methods for Recovery of Signals, Densities, and Spectra From Indirect and Noisy Data," in *Proceedings of Symposia in Applied Mathematics*. American Mathematical Society, 1993, pp. 173–205.

[9] F. Murtagh, J.-L. Starck, and A. Bijaoui, "Image Restoration with Noise Suppression using a Multiresolution support," *Astronomy & Astrophysics Supplement Series*, vol. 112, pp. 179–189, 1995.

[10] P. Fryzlewicz and G. P. Nason, "A Haar-Fisz Algorithm for Poisson Intensity Estimation," *Journal of Computational and Graphical Statistics*, vol. 13, no. 3, pp. 621–638, September 2004.

[11] K. E. Timmermann and R. D. Nowak, "Multiscale Modeling and Estimation of Poisson Processes with Application to Photon-Limited imaging," *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 846–862, April 1999.

[12] E. D. Kolaczyk, "Bayesian Multi-Scale Models for Poisson Processes," *Journal of the American Statistical Association*, vol. 94, no. 447, pp. 920–933, September 1999.

[13] H. Lu, Y. Kim, and J. M. M. Anderson, "Improved Poisson Intensity Estimation: Denoising Application Using Poisson Data," *IEEE Transactions on Image Processing*, vol. 13, no. 8, pp. 1128–1135, August 2004.

[14] S. Lefkimmiatis, P. Maragos, and G. Papandreou, "Bayesian Inference on Multiscale Models for Poisson Intensity Estimation: Applications to Photon-Limited Image Denoising," *IEEE Transactions on Image Processing*, vol. 18, no. 8, pp. 1724–1741, Aug. 2009.

[15] E. D. Kolaczyk, "Nonparametric Estimation of Intensity Maps Using Haar Wavelets and Poisson Noise Characteristics," *The Astrophysical Journal*, vol. 534, pp. 490–505, 2000.

[16] K. Hirakawa and P. Wolfe, "SkellamShrink: Poisson Intensity Estimation for Vector-Valued Data," in *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'09)*, April 2009, pp. 3441–3444.

[17] K. Hirakawa and P. J. Wolfe, "Efficient multivariate skellam shrinkage for denoising photon-limited image data: An empirical bayes approach," in *Proceedings of the 2009 IEEE International Conference on Image Processing (ICIP'09)*, Cairo, Egypt, November 7-11, 2009, pp. 2961–2964.

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

[19] F. Luisier, C. Vonesch, T. Blu, and M. Unser, "Fast Interscale Wavelet Denoising of Poisson-Corrupted Images," *Signal Processing*, vol. 90, no. 2, pp. 415–427, February 2010.

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

[21] E. D. Kolaczyk, "Wavelet Shrinkage Estimation of Certain Poisson Intensity Signals Using Corrected Thresholds," *Statistica Sinica*, vol. 9, no. 1, pp. 119–135, January 1999.

[22] D. L. Donoho and I. M. Johnstone, "Ideal Spatial Adaptation via Wavelet Shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[23] C. Charles and J. Rasson, "Wavelet Denoising of Poisson-Distributed Data and Applications," *Elsevier Computational Statistics and Data Analysis*, vol. 43, no. 2, pp. 139–148, 2003.

[24] R. D. Nowak and R. G. Baraniuk, "Wavelet-Domain Filtering for Photon Imaging Systems," *IEEE Transactions on Image Processing*, vol. 8, no. 5, pp. 666–678, May 1999.

[25] A. Antoniadis and T. Sapatinas, "Wavelet Shrinkage for Natural Exponential Families with Quadratic Variance Functions," *Biometrika*, vol. 88, no. 3, pp. 805–820, 2001.

[26] S. Sardy, A. Antoniadis, and P. Tseng, "Automatic Smoothing With Wavelets for a Wide Class of Distributions," *Journal of Computational & Graphical Statistics*, vol. 13, no. 2, pp. 399–423, June 2004.

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

[27] E. D. Kolaczyk and R. D. Nowak, "Multiscale Likelihood Analysis and Complexity Penalized Estimation," *Annals of statistics*, vol. 32, no. 2, pp. 500–527, 2004.

[28] R. M. Willett and R. D. Nowak, "Multiscale Poisson Intensity and Density Estimation," *IEEE Transactions on Information Theory*, vol. 53, no. 9, pp. 3171–3187, September 2007.

[29] M. Jansen, "Multiscale Poisson Data Smoothing," *Journal of the Royal Statistical Society B*, vol. 68, no. 1, pp. 27–48, 2006.

[30] B. Zhang, J. M. Fadili, and J.-L. Starck, "Wavelets, Ridgelets, and Curvelets for Poisson Noise Removal," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1093–1108, July 2008.

[31] B. Zhang, J. Fadili, J.-L. Starck, and J.-C. Olivo-Marin, "Multiscale Variance-Stabilizing Transform for Mixed-Poisson-Gaussian Processes and its Applications in Bioimaging," in *Proceedings of the 2009 IEEE International Conference on Image Processing (ICIP'07)*, vol. 6, 16 2007-Oct. 19 2007, pp. VI –233–VI –236.

[32] P. Besbeas, I. D. Feis, and T. Sapatinas, "A Comparative Simulation Study of Wavelet Shrinkage Estimators for Poisson Counts," *International Statistical Review*, vol. 72, no. 2, pp. 209–237, August 2004.

[33] M. Raphan and E. P. Simoncelli, "Optimal Denoising in Redundant Representations," *IEEE Transactions on Image Processing*, vol. 17, no. 8, pp. 1342–1352, Aug. 2008.

[34] Hudson, "A Natural Identity for Exponential Families with Applications in Multiparameter Estimation," *The Annals of Statistics*, vol. 6, no. 3, pp. 473–484, May 1978.

[35] K.-W. Tsui and S. J. Press, "Simultaneous Estimation of Several Poisson Parameters Under K-Normalized Squared Error Loss," *The Annals of Statistics*, vol. 10, no. 1, pp. 93–100, March 1982.

[36] S. Ramani, C. Vonesch, and M. Unser, "Deconvolution of 3D Fluorescence Micrographs with Automatic Risk Minimization," in *Proceedings of the Fifth IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'08)*, Paris, France, May 14-17, 2008, pp. 732–735.

[37] M. Raphan and E. P. Simoncelli, "Learning to be Bayesian without Supervision," in *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. Platt, and T. Hoffman, Eds. Cambridge, MA: MIT Press, 2007, pp. 1145–1152.

[38] J.-L. Starck, D. L. Donoho, and E. J. Candes, "Very High Quality Image Restoration by Combining Wavelets and Curvelets," in *Wavelets: Applications in Signal and Image Processing IX*, A. F. Laine, M. A. Unser, and A. Aldroubi, Eds., vol. 4478, no. 1. SPIE, 2001, pp. 9–19.

[39] M. Fadili, J.-L. Starck, and L. Boubchir, "Morphological Diversity and Sparse Image Denoising," in *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'07)*, vol. 1, april 2007, pp. I–589 –I–592.

[40] M. Elad and M. Aharon, "Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[41] C. Vonesch, F. Aguet, J.-L. Vonesch, and M. Unser, "The Colored Revolution of Bioimaging," *IEEE Signal Processing Magazine*, vol. 23, no. 3, pp. 20–31, May 2006.

[42] J. Pawley, *Three-Dimensional Confocal Microscopy: Volume Investigation of Biological Specimens*. Academic Press, 1994, ch. Sources of noise in three-dimensional microscopical data sets, pp. 47–94.

[43] S. Delpretti, F. Luisier, S. Ramani, T. Blu, and M. Unser, "Multiframe SURE-LET Denoising of Timelapse Fluorescence Microscopy Images," in *Proceedings of the Fifth IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'08)*, Paris, France, May 14-17, 2008, pp. 149–152.

[44] J. Boulanger, C. Kervrann, J. Salamero, J.-B. Sibarita, P. Elbau, and P. Bouthemy, "Patch-Based Non-Local Functional

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

for Denoising Fluorescence Microscopy Image Sequences," *IEEE Transactions on Medical Imaging*, vol. 29, no. 2, pp. 442–454, 2010.

Florian Luisier (M'10) was born in Bagnes, Switzerland, in 1981. In 2005, he received his Master degree in Microengineering from the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. In January 2010, he obtained his Ph.D. in Computer, Communication, and Information Sciences from the same institution. From 2005 to March 2010, he was with the EPFL's Biomedical Imaging Group. He is now a postdoctoral fellow in the Statistics and Information Sciences Laboratory at Harvard University. His research interests include statistical signal processing, multiresolution representations, and the restoration of multidimensional biomedical data.

Dr. Luisier received a 2009 Young Author Best Paper Award from the IEEE Signal Processing Society for a paper coauthored with Prof. Thierry Blu and Prof. Michael Unser.

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. In 1996, he obtained a Ph.D in electrical engineering from ENST for a study on iterated rational filterbanks, applied to wideband audio coding.

Between 1998 and 2007, he was with the Biomedical Imaging Group at the Swiss Federal Institute of Technology (EPFL) in Lausanne, Switzerland. He is now a Professor in the Department of Electronic Engineering, The Chinese University of Hong Kong.

Dr. Blu was the recipient of two best paper awards from the IEEE Signal Processing Society (2003 and 2006). He is also coauthor of a paper that received a Young Author best paper award (2009) from the same society. Between 2002 and 2006, he has been an Associate Editor for the IEEE Transactions on Image Processing and since 2006, for the IEEE Transactions on Signal Processing. He is also Associate Editor of Elsevier Signal Processing and Eurasip J. on Image and Video Processing.

Research interests: (multi)wavelets, multiresolution analysis, multirate filterbanks, interpolation, approximation and sampling theory, sparse sampling, image denoising, psychoacoustics, biomedical imaging, optics, wave propagation

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

Michael Unser (M'89–SM'94–F'99) received the M.S. (summa cum laude) and Ph.D. degrees in Electrical Engineering in 1981 and 1984, respectively, from the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. From 1985 to 1997, he worked as a scientist with the National Institutes of Health, Bethesda USA. He is now full professor and Director of the Biomedical Imaging Group at the EPFL.

His main research area is biomedical image processing. He has a strong interest in sampling theories, multiresolution algorithms, wavelets, and the use of splines for image processing. He has published over 150 journal papers on those topics, and is one of ISI's Highly Cited authors in Engineering (http://isihighlycited.com).

Dr. Unser has held the position of associate Editor-in-Chief (2003-2005) for the IEEE Transactions on Medical Imaging and has served as Associate Editor for the same journal (1999-2002; 2006-2007), the IEEE Transactions on Image Processing (1992-1995), and the IEEE Signal Processing Letters (1994-1998). He is currently member of the editorial boards of Foundations and Trends in Signal Processing, and Sampling Theory in Signal and Image Processing. He co-organized the first IEEE International Symposium on Biomedical Imaging (ISBI'2002) and was the founding chair of the technical committee of the IEEE-SP Society on Bio Imaging and Signal Processing (BISP).

Dr. Unser received the 1995 and 2003 Best Paper Awards, the 2000 Magazine Award, and the 2008 Technical Achievement Award from the IEEE Signal Processing Society. He is an EURASIP Fellow and a member of the Swiss Academy of Engineering Sciences.

Copyright (c) 2010 IEEE. Personal use is permitted. For any other purposes, Permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.