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.

Index Terms—Filterbank, Gaussian noise, image denoising, MSE estimation, Poisson noise, thresholding, unbiased risk estimate.

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 a few photons are collected by the photosensors, due to various physical constraints (low-power light source, short exposure time, phototoxicity). 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-denosing 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 non-linear mapping (e.g., a square root) to the raw data, an approach 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 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 \cite{16}, \cite{17}, which can be viewed as a Poisson variant of Donoho’s et al. SUREshrink \cite{18}. Recently, we proposed a non-Bayesian framework to estimate Poisson intensities in the unnormalized Haar wavelet domain (PURE-LET \cite{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) \cite{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 \cite{19}.

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 \cite{21}. This pair of thresholds can be seen as an adapted version of Donoho’s et al. universal threshold that was designed for AWGN \cite{22}. This approach was generalized to arbitrary kinds of Poisson-distributed data by Charles and Rasson \cite{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 \cite{24}. The modulation estimator devised by Antoniadis and Spatias \cite{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 \cite{26}. Their estimator is the solution of a log-likelihood problem, regularized by the addition of a wavelet-domain $\ell_1$-penalty. Using the concept of multiscale likelihood factorizations, Kolaczyk and Nowak introduced complexity-penalized estimators that can also handle a wide class of distributions (Gaussian, Poisson, and multinomial) \cite{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 \cite{28}. Recently, the VST-based approach has been revitalized thanks to the contributions of Jansen, who combined VST with multiscale Bayesian models \cite{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) \cite{30}. This latter VST solution can also stabilize Poisson data embedded in AWGN \cite{31}.

Most of the denoising algorithms discussed in this section have been carefully evaluated by Besbeas et al. \cite{32} using 1-D 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 \cite{3}. \cite{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 block-discrete cosine transform (effective in sparsely representing textured images).

This paper is organized as follows. In Section II, we introduce the theoretical basis of this work. We first derive an unbiased estimate of the MSE for an arbitrary processing of Poisson

![Image](Fig. 1. Filterbank implementation of the unnormalized discrete Haar wavelet transform and principle of the class of denoising algorithms described in \cite{19}. The superscript $j = 1, \ldots, J$ indicates the level of decomposition; $s$ is the vector of noisy scaling coefficients ($s^0$ is thus the noisy input); $d$ is the vector of noisy wavelet coefficients; $\theta$ is the subband-dependent thresholding function; $\delta$ (resp. $\xi$) is the vector of the estimated noise-free wavelet (resp. scaling) coefficients.)
data degraded by AWGN, and then propose a practical approximation of this general result. A pointwise estimator for under-
imated filterbank transforms is presented in Section III and compared with some state-of-the-art approaches in Section IV. Re-
results on real fluorescence-microscopy data are finally shown in Section V.

II. THEORY

Definitions and Notations

- A random variable \( z \) follows a Poisson distribution of in-
tensity \( x \in \mathbb{R}_+ \) if and only if its conditional probability density function is \( p(z = k|x) = \frac{x^k}{k!}e^{-x} \). We use the standard notation \( z \sim \mathcal{P}(x) \).

- A random variable \( b \) follows a Gaussian distribution with zero-
mean and variance \( \sigma^2 \) if and only if its probability density function is \( q(b) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-b^2/2\sigma^2} \). 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 \( n \)th component, where \( n \in \{1, \ldots, N\} \). The variable \( z \) 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 \( z \) can be thought of as a noisy measurement of the intensity signal \( x \). Note that, in contrast with Bayesian approaches, \( x \) 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 \( \epsilon_n \) of known variance \( \sigma^2 \), so that \( b \sim \mathcal{N}(0, \sigma^2) \). The final observation vector \( \mathbf{y} \) is therefore given by

\[
\mathbf{y} = \mathbf{z} + \mathbf{b}.
\]

(1)

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

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

A. Useful Properties of the Poisson and Gaussian Random Variables

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

\[
\mathbb{E}_b \{ \mathbf{b}^T \mathbf{f}(y) \} = \sigma^2 \mathbb{E}_b \{ \text{div} \{ \mathbf{f}(y) \} \}
\]

where \( \text{div} \{ \mathbf{f}(y) \} = \sum_{n=1}^{N} \partial f_n(y)/\partial \epsilon_n \) is the divergence of the function \( \mathbf{f} \) and \( \mathbb{E}_b \{ \cdot \} \) 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].

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

\[
\mathbb{E}_z \{ \mathbf{z}^T \mathbf{f}(\mathbf{z}) \} = \mathbb{E}_z \{ \mathbf{z}^T \mathbf{f}(\mathbf{z}) \}
\]

where \( \mathbf{f}(\mathbf{z}) = [f_n(z - \epsilon_n)]_{n=1}^{N} \).

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], and [35].

B. 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 \( x \). 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 \( \mathbf{f}(\mathbf{y}) = [f_n(y)]_{n=1}^{N} \) be an \( N \)-dimensional (weakly) differentiable real-valued vector function such that \( \forall n, \mathbb{E}_y \{ \partial f_n(y)/\partial \epsilon_n \} < \infty \). Then, the random variable

\[
\epsilon = \frac{1}{N} \left( ||\mathbf{f}(y)||^2 - 2y^T \mathbf{f}(\mathbf{y}) + 2\sigma^2 \text{div} \{ \mathbf{f}(\mathbf{y}) \} \right)
\]

\[
+ \frac{1}{N} \left( ||y||^2 - 1^T y \right) - \sigma^2
\]

(2)

is an unbiased estimate of the expected MSE, that is,

\[
\mathbb{E}_y \{ \epsilon \} = \frac{1}{N} \mathbb{E}_y \{ ||\mathbf{f}(y) - \mathbf{x}||^2 \}.
\]

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

\[
\mathbb{E}_y \{ ||\mathbf{f}(y) - \mathbf{x}||^2 \} = \mathbb{E}_y \{ ||\mathbf{f}(y)||^2 \} - 2 \mathbb{E}_y \{ \mathbf{x}^T \mathbf{f}(\mathbf{y}) \} + ||\mathbf{x}||^2.
\]

(3)

We can now evaluate the two expressions which involve the unknown data \( x \) and are given here.

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

\[
\mathbb{E}_y \{ \mathbf{x}^T \mathbf{f}(\mathbf{y}) \} = \mathbb{E}_b \{ \mathbb{E}_x \{ \mathbf{x}^T \mathbf{f}(\mathbf{z} + \mathbf{b}) \} \}
\]

\[
= \mathbb{E}_b \{ \mathbb{E}_x \{ \mathbf{z}^T \mathbf{f}(\mathbf{z} + \mathbf{b}) \} \}
\]

\[
= \mathbb{E}_y \{ \mathbf{y}^T \mathbf{f}(\mathbf{y}) \} - \mathbb{E}_b \{ \mathbb{E}_x \{ \mathbf{b}^T \mathbf{f}(\mathbf{y}) \} \}
\]

\[
= \mathbb{E}_y \{ \mathbf{y}^T \mathbf{f}(\mathbf{y}) \} - \sigma^2 \mathbb{E}_y \{ \text{div} \{ \mathbf{f}(\mathbf{y}) \} \}.
\]

(4)

2) We notice that

\[
||\mathbf{x}||^2 = \mathbb{E}_y \{ \mathbf{x}^T \mathbf{y} \}
\]

\[
= \mathbb{E}_y \{ ||\mathbf{y}||^2 - 1^T \mathbf{y} \} - N\sigma^2.
\]

(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 (2) will be referred to as PURE. It is a generalization of SURE [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 \( f \). 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.

C. Taylor-Series Approximation of PURE

Unfortunately, the presence of the term \( f^{-1}(y) \) in Theorem 1 makes PURE impractical to evaluate for an arbitrary nonlinear processing. Indeed, in order to compute just one component \( f_n(y - e_n) \) of \( f^{-1}(y) \), one would need to apply the whole deconvolution process to a slightly modified version of the noisy input. This operation would have to be repeated \( N \) times to get the full vector \( f^{-1}(y) \). 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 \( f^{-1}(y) \) given by

\[
f_n(y - e_n) \approx f_n(y) - \frac{\partial f_n(y)}{\partial y} e_n
\]

for all \( n = 1 \ldots N \). Consequently, provided that the above approximation is reasonable (i.e., that \( f_n \) varies slowly), the function \( f^{-1}(y) \) is well approximated by \( f(y) - \partial f(y) \), where \( \partial f(y) = [\partial f_n(y)/\partial y]_{n=1 \ldots N} \) 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:

\[
\hat{\varepsilon} = \frac{1}{N} (||f(y)||^2 - 2y^T (f(y) - \partial f(y))) + \frac{1}{N} (2\sigma^2 \text{div} \{f(y) - \partial f(y)\}) + ||y||^2 - 1^T y - \sigma^2.
\]

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

D. PURE for Transform-Domain Denoising

Here, we derive an expression for the MSE estimate given in (6), in the particular case of a transform-domain pointwise processing \( \Theta \). 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 \( D = [d_{ij}]_{(i,j) \in \{1 \ldots L\} \times \{1 \ldots N\}} \) and \( R = [r_{ij}]_{(i,j) \in \{1 \ldots N\} \times \{1 \ldots L\}} \), respectively, that satisfy the perfect reconstruction property \( RD = I \).

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

\[
\hat{x} = f(y) = R\Theta(Dy, \tilde{D}y)
\]

where \( \Theta(w, \tilde{w}) = [\theta_j(w_j, \tilde{w}_j)]_{j \in \{1 \ldots L\}} \) 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, \( D \) and \( R \) implement the (unnormalized) Haar wavelet decomposition and reconstruction, respectively (see Fig. 1). \( w \) and \( \tilde{w} \) then represent the Haar wavelet coefficients, respectively.

Corollary 1: For the transform-domain pointwise processing defined in (7), the MSE estimate given in (6) can be reformulated as (8), shown at the bottom of the page, with the following definitions.

- The \( L \times 1 \) vector \( \Theta_1(w, \tilde{w}) = [\partial \theta_j(w_j, \tilde{w}_j)/\partial w_j]_{j \in \{1 \ldots L\}} \) is made of the first derivative, with respect to its first variable, of each function \( \theta_j \).
- The \( L \times 1 \) vector \( \Theta_2(w, \tilde{w}) = [\partial \theta_j(w_j, \tilde{w}_j)/\partial \tilde{w}_j]_{j \in \{1 \ldots L\}} \) is made of the first derivative, with respect to its second variable, of each function \( \theta_j \).
- The \( L \times 1 \) vector \( \Theta_{12}(w, \tilde{w}) = [\partial^2 \theta_j(w_j, \tilde{w}_j)/\partial w_j \partial \tilde{w}_j]_{j \in \{1 \ldots L\}} \) is made of the first derivative, with

\[
\hat{\varepsilon} = \frac{1}{N} ||f(y) - y||^2 - \frac{1}{N} y^T y - \sigma^2 + \frac{2}{N} (\Theta_1(w, \tilde{w})^T (D \cdot R^T) y + \Theta_2(w, \tilde{w})^T (D \cdot R^T) y)
\]

\[
+ \frac{2\sigma^2}{N} \left( \text{diag} \{D R\} \Theta_1(w, \tilde{w}) + \text{diag} \{D \tilde{R}\} \Theta_2(w, \tilde{w}) \right)
\]

\[
- \frac{2\sigma^2}{N} \left( \text{diag} \{(D \cdot D) R\} \Theta_{11}(w, \tilde{w}) + \text{diag} \{(D \cdot \tilde{D}) R\} \Theta_{22}(w, \tilde{w}) + 2 \text{diag} \{(D \cdot \tilde{D}) R\} \Theta_{12}(w, \tilde{w}) \right)
\]
respect to its first variable and second variable, of each function \( \theta_l \).

- The \( L \times 1 \) vector \( \Theta_{11}(\mathbf{w}, \mathbf{\bar{w}}) = \left[ \partial \theta_l(u_l, \mathbf{\bar{w}}_l) / \partial \mathbf{\bar{w}}_l \right]_{l=1}^{L} \) is made of the second derivative, with respect to its first variable, of each function \( \theta_l \).

- The \( L \times 1 \) vector \( \Theta_{22}(\mathbf{w}, \mathbf{\bar{w}}) = \left[ \partial \theta_l(u_l, \mathbf{\bar{w}}_l) / \partial \mathbf{\bar{w}}_l \right]_{l=1}^{L} \) is made of the second derivative, with respect to its second variable, of each function \( \theta_l \).

- The operator “\( \bullet \)” denotes the Hadamard (element-by-element) product between two matrices.

Proof: Using the result given in (6) and the fact that \( \Theta(\mathbf{w}, \mathbf{\bar{w}}) = \left[ \theta_l(u_l, \mathbf{\bar{w}}_l) \right]_{l=1}^{L} \), we can further develop \( \partial \mathbf{f}(\mathbf{y}) = \left[ \partial f_n(\mathbf{y}) / \partial y_m \right] \) as

\[
\frac{\partial f_n(\mathbf{y})}{\partial y_m} = \sum_{l=1}^{L} \sum_{n=1}^{N} \left[ \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial y_m} \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial \mathbf{\bar{w}}_l} \frac{\partial y_n}{\partial \mathbf{\bar{w}}_l} \right]_{l=1}^{L} + \left[ \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial \mathbf{\bar{w}}_l} \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial \mathbf{\bar{w}}_l} \frac{\partial y_n}{\partial \mathbf{\bar{w}}_l} \right]_{l=1}^{L}.
\]

Similarly, we have that

\[
\frac{\partial^2 f_n(\mathbf{y})}{\partial y_m^2} = \sum_{l=1}^{L} \sum_{n=1}^{N} \left[ \frac{\partial^2 \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial y_m^2} \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial \mathbf{\bar{w}}_l} \frac{\partial y_n}{\partial \mathbf{\bar{w}}_l} \right]_{l=1}^{L} + \left[ \frac{\partial^2 \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial y_m^2} \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial \mathbf{\bar{w}}_l} \frac{\partial y_n}{\partial \mathbf{\bar{w}}_l} \right]_{l=1}^{L} + 2 \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial y_m} \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial \mathbf{\bar{w}}_l} \frac{\partial y_n}{\partial \mathbf{\bar{w}}_l}.
\]

and, consequently, that

\[
y^T \partial \mathbf{f}(\mathbf{y}) = \sum_{l=1}^{L} \sum_{n=1}^{N} \left[ \Theta(\mathbf{w}, \mathbf{\bar{w}}) \right]_{l=1}^{L} \left[ \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial y_m} \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial \mathbf{\bar{w}}_l} \frac{\partial y_n}{\partial \mathbf{\bar{w}}_l} \right]_{l=1}^{L} + \left[ \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial \mathbf{\bar{w}}_l} \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial \mathbf{\bar{w}}_l} \frac{\partial y_n}{\partial \mathbf{\bar{w}}_l} \right]_{l=1}^{L} + 2 \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial y_m} \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial \mathbf{\bar{w}}_l} \frac{\partial y_n}{\partial \mathbf{\bar{w}}_l}.
\]

\[
\text{div} \{ \mathbf{f}(\mathbf{y}) \} = \sum_{l=1}^{L} \sum_{n=1}^{N} \left[ \Theta(\mathbf{w}, \mathbf{\bar{w}}) \right]_{l=1}^{L} \left[ \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial y_m} \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial \mathbf{\bar{w}}_l} \frac{\partial y_n}{\partial \mathbf{\bar{w}}_l} \right]_{l=1}^{L} + \left[ \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial \mathbf{\bar{w}}_l} \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial \mathbf{\bar{w}}_l} \frac{\partial y_n}{\partial \mathbf{\bar{w}}_l} \right]_{l=1}^{L} + 2 \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial y_m} \frac{\partial \theta_l(u_l, \mathbf{\bar{w}}_l)}{\partial \mathbf{\bar{w}}_l} \frac{\partial y_n}{\partial \mathbf{\bar{w}}_l}.
\]

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

E. 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

\[
\mathbf{f}(\mathbf{y}) = \sum_{k=1}^{K} a_k \mathbf{R}_k(\mathbf{w}, \mathbf{\bar{w}}).
\]

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 \( \mathbf{a} = [a_1 \ a_2 \ ... \ a_K]^{T} \) boils down to the solution of the following system of \( K \) linear equations: for \( k = 1 \ldots K \), we have

\[
\sum_{l=1}^{L} \left[ \mathbf{f}_k(\mathbf{y}) \mathbf{f}_l(\mathbf{y}) - \sigma^2 \text{div} \{ \mathbf{f}_k(\mathbf{y}) \} \right]_{l=1}^{L} = \Theta(\mathbf{w}, \mathbf{\bar{w}})^T(D \cdot \mathbf{R}^T) \mathbf{y} + \Theta(\mathbf{w}, \mathbf{\bar{w}})^T(D \cdot \mathbf{R}^T) \mathbf{y}.
\]

III. POINTWISE ESTIMATOR FOR UNDECIMATED FILTERBANK TRANSFORMS

Here, 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

\[
\mathbf{D}_j = \sum_{n=0}^{N-1} \mathbf{y}_j[k - nN], \quad \mathbf{R}_j = \sum_{n=0}^{N-1} \mathbf{y}_j[k - nN].
\]
Although all of the results presented in this section are derived for 1-D signals, they can be straightforwardly extended to higher dimensional signals by considering separable filters.

A. Choice of $\hat{D}$—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

$$Q(z)^2 = Q(z)/Q(z^{-1})$$

(see Fig. 4)

$$H(z^{-1})Q(z^{-1}) = G(z^{-1})R_{11}(z)$$

(17)

where $H(z)$ and $G(z)$ are the orthonormal scaling and wavelet filters, respectively, and $R_{11}(z) = \varepsilon R_{11}(z^{-1})$ is an arbitrary symmetric ($\varepsilon = 1$) or anti-symmetric ($\varepsilon = -1$) filter.

In the case of symmetric filters characterized by $H(z^{-1}) = z^{\tau_{0}}H(z)$, the shortest-support GDC filter is simply given by $Q(z) = z^{\tau_0 - 1}$. 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 $\hat{D}$ is then defined as

$$\hat{D} = [D_{1}^{T} D_{2}^{T} \ldots D_{J+1}^{T}]^{T},$$

where $\tilde{h}_{j}[n]$ is the $nt_{1}$ coefficient of the filter

$$\tilde{H}_{j}(z) = 2^{j}Q(z^{2^{j-1}})H_{j}(z)$$

$$= Q(z^{2^{j-1}})H(z)H(z^{2}) \cdots H(z^{2^{J-1}}),$$

for $j = 1 \ldots J$

$$\tilde{H}_{j+1}(z) = 2^{j}H_{j}(z)$$

$$= H(z)H(z^{2}) \cdots H(z^{2^{J-1}}).$$

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

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 $\text{diag}(\{D \cdot R \})$:

In the case of periodic boundary conditions, we have that

$$\text{diag}(\{D \cdot R_j \}) = \begin{cases} \frac{1}{M} \{1, \ldots, 1\}^{T}, & \text{for } j = 1 \ldots J \vspace{1em} \\ \frac{1}{M} \{1, \ldots, 1\}^{T}, & \text{for } j = J + 1 \end{cases}$$

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 $\text{diag}(\{D \cdot D \cdot R \})$, $\text{diag}(\{D \cdot D \cdot D \cdot R \})$, and $\text{diag}(\{D \cdot D \cdot R \})$:

Contrary to $\text{diag}(\{D \cdot R \})$, 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 nonperiodic boundary extensions are performed.

Note that the vectors $(D \cdot D)y$, $(D \cdot D \cdot y)$, and $(D \cdot R \cdot y)$ are obtained by analyzing the noisy data $y$ with the considered undecimated filterbank using modified analysis filters with coefficients $g_{j}^{(2)}[k]$, $\tilde{g}_{j}^{(2)}[k]$, $g_{j}[k]$, $\tilde{g}_{j}[k]$, and $\tilde{g}_{j}[k]$, respectively.

C. Thresholding Function

We propose to use a subband-dependent pointwise thresholding function defined for $j \in [1 \ldots J] \times [1 \ldots N]$ by

$$\Theta(w, \pl) = \{\theta_{j}(w_{j,n}, \pl_{jn})\}_{(j-1)N+n \cdot}$$

(18)

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

In the case of Poisson data, we need a signal-dependent transform-domain threshold to take into account the nonstationarity

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

of the noise. If we consider unit-norm filters, the variance $\sigma^2$ 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 $\hat{\sigma}(\mathbf{f})$ (built as described in Section III-A), up to the scale-dependent factors $\beta_j = 2^{-j/2}$ and $\hat{\sigma}_j = M^{-1/2}$ for a multiscale transform and for an overcomplete BDCT, respectively. Therefore, we propose the signal-dependent threshold

$$t_j(\mathbf{f}) = \sqrt{\beta_j |\mathbf{f}| + \sigma^2}$$

(19)

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

$$\theta_j(w, \mathbf{f}) = \alpha_jw \cdot \left( \frac{w}{\theta_{j1}(w, \mathbf{f})} + \alpha_jw \exp\left( -\left( \frac{w}{\theta_{j2}(w, \mathbf{f})} \right)^8 \right) \right).$$

(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 $\tanh(k\mathbf{f})\mathbf{f}$ ($k = 100$, typically) instead of $|\mathbf{f}|$ 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

$$\mathbf{f}(\mathbf{y}) = \sum_{j=1}^{J} \sum_{k=1}^{2} \alpha_{j,k} \frac{\mathbf{R}_j \mathbf{\Theta}_{j,k}(\mathbf{D}_j \mathbf{y})}{\mathbf{f}_{j,k}(\mathbf{y})} \mathbf{D}_j \mathbf{y} + \frac{\mathbf{R}_{j+1} \mathbf{D}_{j+1} \mathbf{y}}{\mathbf{f}_{j,k}(\mathbf{y})}$$

(21)

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

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

Fig. 7. Reliability of the PURE-based optimization of the LET parameters. The blue (online version) curve (‘+’ 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 (online version) curve (‘*’ markers) displays the output PSNRs obtained by a restricted PURE-based optimization of the LET parameters, where all the nonlinearly processed subband $r_{j,k}(\mathbf{y})$ with $T_j < 10$ have been disregarded. The output PSNRs obtained by the MLE-based optimization of the LET parameters (without restriction) is the baseline (‘*’ markers).

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

$$\begin{align}
\mathbf{M} &= \left[ \hat{\mathbf{f}}_k(\mathbf{y})^T \mathbf{f}_j(\mathbf{y}) \right]_{2(k_j-1)+k_2,2(k_j-1)+k_2} \\
\hat{\mathbf{c}} &= \left[ \hat{\mathbf{f}}_k(\mathbf{y})^T (\mathbf{I} - \mathbf{R}_{j+1} \mathbf{D}_{j+1}) \mathbf{y} \right] \\
&- (\mathbf{y}^T \partial \hat{\mathbf{f}}_k(\mathbf{y}) + \sigma^2 \text{div}(\hat{\mathbf{f}}_k(\mathbf{y}) - \partial \hat{\mathbf{f}}_k(\mathbf{y}))) \\
&\quad \text{div}(\hat{\mathbf{f}}_k(\mathbf{y}) - \partial \hat{\mathbf{f}}_k(\mathbf{y}))) \right]_{2(k_j-1)+k_2},
\end{align}$$

(22)

For very low intensity signals (typically such that $x_n < 5$, $\forall n$), the first-order Taylor-series approximation of some nonlinear functions $f_j$ might be inaccurate, leading (22) to be an unreliable approximation of the $(j, 2)$th component of the PURE vector $c$ 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 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 $E_{\text{mean}} = |x|^2/N = E_N \left\{ |y|^2 - 1^T x \right\}/N = \sigma^2$. Any nonlinearly processed subband $f_{j,2}(y)$ for which the SNR of the approximation is below a given threshold (around 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 $T$ increases. Consequently, we suggest to keep only those nonlinearly processed subbands $f_{j,2}(y)$ for which the SNR of the approximation is above a given threshold ($T_j = M_j E_{\text{mean}}$ is above a given threshold $T$). We experimentally found that any value $T = 10$ was a reasonable choice (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 $f_{j,2}(y)$, as shown in Fig. 7.

TABLE I

<table>
<thead>
<tr>
<th>Peak Intensity</th>
<th>120</th>
<th>60</th>
<th>30</th>
<th>20</th>
<th>10</th>
<th>5</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Input PSNR</td>
<td>24.08</td>
<td>21.07</td>
<td>18.05</td>
<td>16.29</td>
<td>13.28</td>
<td>10.27</td>
<td>7.29</td>
<td>3.28</td>
</tr>
<tr>
<td>Haar-Fisz [10]</td>
<td>29.73</td>
<td>27.98</td>
<td>26.35</td>
<td>25.45</td>
<td>23.70</td>
<td>22.55</td>
<td>20.77</td>
<td>19.77</td>
</tr>
<tr>
<td>Platelet [28]</td>
<td>30.54</td>
<td>29.10</td>
<td>27.66</td>
<td>26.80</td>
<td>25.14</td>
<td>23.56</td>
<td>21.72</td>
<td>20.87</td>
</tr>
<tr>
<td>PH-HMT [14]</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>26.61</td>
<td>24.97</td>
<td>23.37</td>
<td>21.41</td>
</tr>
<tr>
<td>UWT PURE-LET</td>
<td>31.03</td>
<td>29.29</td>
<td>27.67</td>
<td>26.72</td>
<td>25.10</td>
<td>23.50</td>
<td>21.67</td>
<td>20.44</td>
</tr>
<tr>
<td>UWT/BDCT PURE-LET</td>
<td>31.35</td>
<td>29.58</td>
<td>27.91</td>
<td>26.93</td>
<td>25.29</td>
<td>23.65</td>
<td>21.72</td>
<td>20.48</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Image</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Input PSNR</td>
<td>24.00</td>
<td>20.99</td>
<td>17.98</td>
<td>16.22</td>
<td>13.21</td>
<td>10.20</td>
<td>7.22</td>
<td>3.21</td>
</tr>
<tr>
<td>PH-HMT [14]</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>24.92</td>
<td>23.45</td>
<td>23.23</td>
<td>21.27</td>
</tr>
<tr>
<td>UWT PURE-LET</td>
<td>29.75</td>
<td>27.52</td>
<td>25.69</td>
<td>24.80</td>
<td>23.56</td>
<td>22.61</td>
<td>21.50</td>
<td>20.74</td>
</tr>
<tr>
<td>BDCT PURE-LET</td>
<td>31.75</td>
<td>29.96</td>
<td>28.23</td>
<td>27.23</td>
<td>25.52</td>
<td>23.87</td>
<td>21.83</td>
<td>20.74</td>
</tr>
<tr>
<td>UWT/BDCT PURE-LET</td>
<td>31.86</td>
<td>30.08</td>
<td>28.34</td>
<td>27.33</td>
<td>25.59</td>
<td>23.89</td>
<td>21.87</td>
<td>20.79</td>
</tr>
</tbody>
</table>

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

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)

$$f(y) = \sum_{k=1}^{K_1} a_k f_{k}^{\text{UWT}}(y) + \sum_{k=1}^{K_2} b_k f_{k}^{\text{BDCT}}(y) + \cdots$$  \hspace{0.5cm} (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. [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].
IV. SIMULATIONS

Here, 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 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” 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. 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 dyadic partition of the support of the Poisson intensity. 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-


Matlab code available at: http://nislab.ee.duke.edu/NISLab/Platelets.html

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 28 s. (d) Denoised with Anscombe [7] + BLS-GSM[1], PSNR = 25.45 dB in 38 s. (e) Denoised with translation-invariant Platelet [28], PSNR = 25.66 dB in 2234 s. (f) Denoised with the proposed UWT PURE-LET, PSNR = 26.23 dB in 7.5 s.
TABLE II
COMPARISON OF MULTISCALE-BASED MIXED POISSON–GAUSSIAN NOISE REMOVAL ALGORITHMS

<table>
<thead>
<tr>
<th>Peak Intensity</th>
<th>120</th>
<th>60</th>
<th>30</th>
<th>20</th>
<th>10</th>
<th>5</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image</td>
<td>Cameraman 256 × 256</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Input PSNR</td>
<td>18.56</td>
<td>17.48</td>
<td>15.90</td>
<td>14.75</td>
<td>12.45</td>
<td>9.85</td>
<td>6.11</td>
<td>3.19</td>
</tr>
<tr>
<td>UWT PURE-LET</td>
<td>27.68</td>
<td>27.14</td>
<td>26.32</td>
<td>25.74</td>
<td>24.56</td>
<td>23.22</td>
<td>21.57</td>
<td>20.42</td>
</tr>
<tr>
<td>BDCT PURE-LET</td>
<td>27.17</td>
<td>26.64</td>
<td>25.87</td>
<td>25.35</td>
<td>24.24</td>
<td>23.05</td>
<td>21.47</td>
<td>20.36</td>
</tr>
<tr>
<td>UWT/BDCT PURE-LET</td>
<td>27.92</td>
<td>27.37</td>
<td>26.53</td>
<td>25.95</td>
<td>24.74</td>
<td>23.36</td>
<td>21.61</td>
<td>20.44</td>
</tr>
<tr>
<td>Image</td>
<td>Fluorescent Cells 512 × 512</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Input PSNR</td>
<td>19.39</td>
<td>18.86</td>
<td>17.96</td>
<td>17.21</td>
<td>15.53</td>
<td>13.38</td>
<td>10.01</td>
<td>7.22</td>
</tr>
<tr>
<td>UWT PURE-LET</td>
<td>30.68</td>
<td>30.37</td>
<td>29.86</td>
<td>29.46</td>
<td>28.63</td>
<td>27.67</td>
<td>26.23</td>
<td>25.07</td>
</tr>
<tr>
<td>BDCT PURE-LET</td>
<td>30.64</td>
<td>30.33</td>
<td>29.83</td>
<td>29.44</td>
<td>28.64</td>
<td>27.67</td>
<td>26.21</td>
<td>25.09</td>
</tr>
<tr>
<td>UWT/BDCT PURE-LET</td>
<td>30.82</td>
<td>30.50</td>
<td>29.98</td>
<td>29.56</td>
<td>28.71</td>
<td>27.73</td>
<td>26.28</td>
<td>25.15</td>
</tr>
</tbody>
</table>

Output PSNRs have been averaged over 10 noise realizations.

![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 32 s. (d) Denoised with the proposed BDCT PURE-LET, PSNR = 27.00 dB in 47 s.](image)

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

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

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

A. Poisson Noise Reduction

In Table I, we compare our method with the state-of-the-art multisresolution-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 $256 \times 256$ grayscale Cameraman image at peak intensity $20$ requires around $1300$ s (65 s per cycle-spin) for the Platelet and $92$ s for the PH-HMT. By contrast, the execution of 20 cycle-spins of the Haar-Fisz algorithm lasts $\sim 5.7$ s, the Anscombe+BLS-GSM takes $\sim 7.7$ s, whereas our PURE-LET algorithm lasts $1.3$ s using the Haar UWT, $12.2$ s in the $16 \times 16$ BDCT, and $13.5$ s 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 ($\gg 1$ 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. 10(a)], 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 $512 \times 512$ 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. 10(b)]. 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.

In order for our data to fit the noise model described in (1), we need to first subtract the offset value $\delta$ of the CCD detectors, and then, divide by their amplification factor $\alpha$. These two parameters, as well as the variance $\sigma^2$ 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, the slope of the fitted line gives the amplification factor $\alpha$. The parameters $\delta$ and $\sigma^2$ 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. 10(d), 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. 10(c) 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 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.

ACKNOWLEDGMENT

The authors are especially grateful to Prof. L. Jiang and Dr. Y. 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


