Signal Processing 90 (2010) 415–427

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Fast interscale wavelet denoising of Poisson-corrupted images

Florian Luisier [a], Cédric Vonesch [a,*], Thierry Blu [b], Michael Unser [a]

[a] Biomedical Imaging Group (BIG), Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
[b] Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

ARTICLE INFO

Article history:
Received 12 March 2009
Received in revised form
24 June 2009
Accepted 3 July 2009
Available online 17 July 2009

Keywords:
Poisson
Interscale
Denoising
Wavelets
Risk estimation
Linear expansion of thresholds
Fluorescence microscopy

ABSTRACT

We present a fast algorithm for image restoration in the presence of Poisson noise. Our approach is based on (1) the minimization of an unbiased estimate of the MSE for Poisson noise, (2) a linear parametrization of the denoising process and (3) the preservation of Poisson statistics across scales within the Haar DWT. The minimization of the MSE estimate is performed independently in each wavelet subband, but this is equivalent to a global image-domain MSE minimization, thanks to the orthogonality of Haar wavelets. This is an important difference with standard Poisson noise-removal methods, in particular those that rely on a non-linear preprocessing of the data to stabilize the variance.

Our non-redundant interscale wavelet thresholding outperforms standard variance-stabilizing schemes, even when the latter are applied in a translation-invariant setting (cycle-spinning). It also achieves a quality similar to a state-of-the-art multiscale method that was specially developed for Poisson data. Considering that the computational complexity of our method is orders of magnitude lower, it is a very competitive alternative.

The proposed approach is particularly promising in the context of low signal intensities and/or large data sets. This is illustrated experimentally with the denoising of low-count fluorescence micrographs of a biological sample.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

1.1. Motivation and objectives

"Additive white Gaussian noise" is a ubiquitous model in the context of statistical image restoration. In many applications, however, the current trend towards quantitative imaging calls for less generic models that better account for the physical acquisition process. The need for such models is particularly stringent in biomicroscopy, where live samples are often observed at very low light levels, due to acquisition-time and phototoxicity constraints [1]. In this regime, the performance of the imaging device is typically shot-noise limited, i.e., the measurement noise is strongly signal dependent. Thus, opting for a non-additive, non-Gaussian model can yield significant restoration-quality improvements in such applications.

In this paper we present a novel algorithmic solution for denoising low-count images. Motivated by practical applications in biomedical imaging, we aim at a tractable compromise between restoration quality, computational complexity and automation. Specifically, our goals are the following. First, we want a method that is designed for Poissonian noise, which is the most common model for low-intensity imaging. Second, the method should satisfy strict constraints in terms of computational cost and memory requirements, so as to be able to process large data sets. Finally, we want the algorithm to require as less user input as possible in order to facilitate its application and to enhance the reproducibility of its results.

* Corresponding author.
E-mail address: cedric.vonesch@epfl.ch (C. Vonesch).

0165-1684/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2009.07.009

1.2. The PURE-LET approach

To address the aforementioned challenges, our method is based on the following concepts.

The fundamental tool is a statistical estimate of the mean square error (MSE), or "risk", between the (unknown) noiseless image and the processed noisy image. Owing to the Poisson noise hypothesis, we refer to this result as the Poisson unbiased risk estimate (PURE); this is the equivalent of Stein's unbiased risk estimate (SURE) which holds for Gaussian statistics. In particular, we develop an interscale PURE that we believe to be novel. We then *minimize* this MSE estimate over a collection of "acceptable" denoising processes to find the best one, in the sense of the signal-to-noise ratio (SNR), which is a widespread measure of restoration quality [2]. To our knowledge, this is actually the first reported use of an (unbiased) MSE estimate in the Poisson-noise case for image processing.

The efficiency of our method stems from the use of a simple unnormalized Haar-wavelet transform and from the concept of linear expansion of thresholds (LET): the "acceptable" denoising processes are expressed as a linear combination of elementary denoising processes, from which only the weights are unknown. It is these weights that are then computed by minimizing the PURE, through the resolution of a simple linear system of equations. This means that all the parameters of the algorithm are adjusted completely automatically, without requiring user input.

For each subband, our restoration functions involve several parameters, which provides more flexibility than standard single-parameter thresholding functions. Importantly, the thresholds are adapted to local estimates of the (signal-dependent) noise variance; this is a fundamental difference with our previous work [3]. These estimates are derived from the corresponding low-pass coefficients at the same scale; the latter are also used to incorporate interscale relationships into the denoising functions. The resulting procedure can be easily integrated into the wavelet decomposition, which is *non-redundant*. The MSE estimate is optimized independently for each subband by exploiting the orthogonality of the Haar wavelet basis.

As a result, our algorithm has low computational complexity and modest memory requirements. These are valuable features for denoising large data sets, such as those typically produced in biomedical applications. Importantly, this computational efficiency is not traded for quality. On the contrary, the algorithm yields improved results compared to traditional Gaussian-inspired approaches, and it performs competitively with a state-of-the-art multiscale method that was specially developed for Poisson data.

1.3. Related work

In the context of image denoising, several works are based on the fact that Poisson statistics are preserved across scales in the low-pass channels of an unnormalized Haar wavelet transform. This fundamental property was for example used by Timmermann and Nowak [4] to construct a multiscale Bayesian model of the image; an extension for estimating all parameters of the model using the expectation–maximization algorithm was derived in [5]. A similar model was proposed independently by Kolaczyk for 1D signals [6], using the concept of recursive dyadic partition. The aforementioned property was also used within a user-calibrated hypothesis-testing approach for astrophysical imaging [7].

Notice that the Bayesian framework has also been used in conjunction with more involved multiscale transformations than the Haar transform. For example, Sardy et al. [8] considered a general ℓ_1-penalized-likelihood framework for arbitrary wavelet bases and noise models, including the Poisson case. More recently, Willett and Nowak have proposed a platelet-based penalized-likelihood estimator that was shown to be very efficient for denoising piecewise-smooth images [9].

A widespread alternative to the direct handling of Poisson statistics is to apply variance-stabilizing transforms (VSTs)—with the underlying idea of exploiting the broad class of denoising methods that are based on a Gaussian noise model [10]. Since the seminal work of Anscombe [11], more involved VSTs have been proposed, such as the Haar–Fisz transform [12]. Such approaches belong to the state-of-the-art for 1D wavelet-based Poisson noise removal [2,12]. They have been combined with various other methodologies, e.g., Bayesian multiscale likelihood models that can be applied to arbitrary wavelet transforms [13]. Very recently, a hybrid approach that combines VSTs, hypothesis testing, ℓ_1-penalized reconstruction and advanced redundant multiscale representations has been proposed by Zhang et al. [14].

With the exception of cross-validation methods [8,15,16], however, the potential of purely data-driven techniques seems to have remained under-exploited for the wavelet-based restoration of images corrupted by Poisson noise. What we claim to be original in the present paper is the combination of a risk estimate for Poisson statistics with the Haar multiresolution, which leads to a fast denoising procedure. In particular, up to our knowledge, the use of a Poisson MSE estimate is unreported, at least in the signal/image processing literature, not to mention the more sophisticated interscale extension that we have worked out.

While this paper was under review, we became aware of the independent work of Hirakawa and Wolfe [31], which is related to the results presented here.

1.4. Organization of the paper

The paper is organized as follows. In Section 2, we derive an unbiased risk estimate for a broad class of Poisson denoising algorithms formulated in the Haar-wavelet domain. In Section 3, we specify several such algorithms based on our previous work and experimental insights. In Section 4, we compare our approach to typical variance-stabilizing methods, as well as a recent algorithm specifically designed for Poisson statistics. Finally we present results obtained with real fluorescence-microscopy data in Section 5.

2. Theory

Recall that m is a Poisson random variable of intensity $\mu \in \mathbb{R}^+$ if and only if

$$P(m=k) = e^{-\mu}\frac{\mu^k}{k!} \tag{1}$$

for every $k \in \mathbb{N}$; we use the standard notation $m \sim \mathscr{P}(\mu)$. 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 $n \in \{1,\ldots,N\}$. $\mathbf{m}$ will denote a vector of N independent Poisson random variables m_n of underlying intensities μ_n, i.e., $m_n \sim \mathscr{P}(\mu_n)$. A realization of m can be thought of as a noisy measurement of the intensity signal $\boldsymbol{\mu}$. Note that in contrast with Bayesian approaches, $\boldsymbol{\mu}$ is considered to be *deterministic* in the present work. Based on the measurement $\mathbf{m}$, our goal is to find an estimate $\hat{\boldsymbol{\mu}}$ that is the closest possible to the original signal in the minimum *mean squared error* (MSE) sense; that is, we want to minimize

$$\mathrm{MSE} = \frac{1}{N}\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2 = \frac{1}{N}\sum_{n=1}^{N}(\hat{\mu}_n - \mu_n)^2.$$

The MSE is directly related to the peak-signal-to-noise ratio (PSNR), which is the most common measure of restoration quality in the denoising literature:

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

where $I_{\max}$ is the maximum intensity of the noise-free signal. In addition, the MSE has nice mathematical properties that greatly facilitate its manipulation and optimization in comparison with other quality measures.

2.1. Some useful properties of the Poisson distribution

The Poisson distribution enjoys the following useful properties.

Property 1. *The sum of independent Poisson random variables is also a Poisson random variable, whose intensity is the sum of the original intensities.*

For example, $m_1 + m_2 \sim \mathscr{P}(\mu_1 + \mu_2)$.

Property 2. *If $m \sim \mathscr{P}(\mu)$ and $\theta : \mathbb{R} \to \mathbb{R}$ is a real function such that $\mathscr{E}\{|\theta(m)|\} < \infty$, then*

$$\mathscr{E}\{\mu\theta(m)\} = \mathscr{E}\{m\theta(m-1)\},$$

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

Proof.

$$\begin{aligned}\mathscr{E}\{\mu\theta(m)\} &= \sum_{k\in\mathbb{N}} \mu\theta(k)\frac{\mu^k}{k!}e^{-\mu} \\ &= \sum_{k\in\mathbb{N}\setminus\{0\}} \theta(k-1)\frac{\mu^k}{(k-1)!}e^{-\mu} \\ &= \sum_{k\in\mathbb{N}} k\theta(k-1)\frac{\mu^k}{k!}e^{-\mu} \\ &= \mathscr{E}\{m\theta(m-1)\}.\end{aligned}$$

Note that Property 2 is the Poisson equivalent of Stein's lemma for Gaussian statistics [17]; similar results can be found e.g., in [18,19]. As a straightforward application of this property, we note that we have

$$\mu^2 = \mathscr{E}\{m(m-1)\}.$$

Finally we state a result dating back to [20] that will be useful for the experimental part of this work (Section 5). □

Property 3 (*Binomial selection*). *Let $m \sim \mathscr{B}(\ell,\eta)$ be a binomial random variable, where $\eta \in [0,1]$ represents the probability of success. If the number of trials $\ell \in \mathbb{N}$ is random and follows a Poisson distribution with mean λ, then m is itself Poisson distributed with mean $\mu = \eta\lambda$.*

2.2. The unnormalized Haar discrete wavelet transform

The unnormalized Haar discrete wavelet transform (DWT) can be seen as a standard two-channel filterbank (see Fig. 1). The analysis pair of lowpass/highpass filters is given in the z-transform domain by

$$\begin{cases} H_a(z) = 1 + z^{-1}, \\ G_a(z) = 1 - z^{-1}. \end{cases}$$

The corresponding synthesis pair is

$$\begin{cases} H_s(z) = \frac{1}{2}H_a(z^{-1}), \\ G_s(z) = \frac{1}{2}G_a(z^{-1}). \end{cases}$$

In this work, the unnormalized Haar scaling coefficients of the measurement $\mathbf{m}$ at scales $j = 1,\ldots,J$ are denoted by $\mathbf{s}^j \in \mathbb{R}^{N_j}$, where $N_j = N/2^j$, and $d^j \in \mathbb{R}^{N_j}$ stands for the associated wavelet coefficients (we assume that the signal dimension is divisible by 2^J). Setting $\mathbf{s}^0 = \mathbf{m}$, these coefficients are obtained from the following sums and

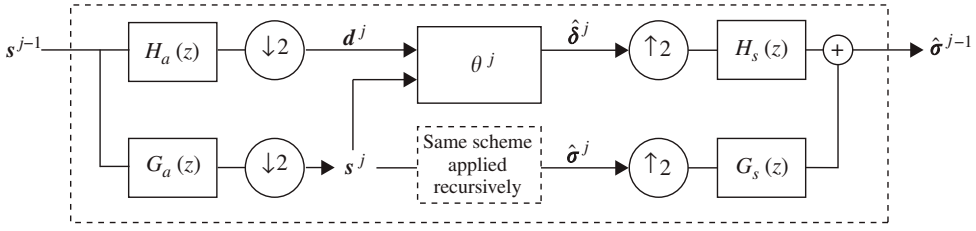

Fig. 1. Filter bank implementation of the unnormalized discrete Haar wavelet transform and principle of the class of denoising algorithms described by (2). The scheme is applied recursively on the low-pass channel output.

differences:

$$\begin{cases} s_n^j = s_{2n}^{j-1} + s_{2n-1}^{j-1} \\ d_n^j = s_{2n}^{j-1} - s_{2n-1}^{j-1} \end{cases} \quad \text{for } j = 1,\ldots,J.$$

The original sequence $\mathbf{m} = \mathbf{s}^0$ is simply recovered by computing

$$\begin{cases} s_{2n}^{j-1} = (s_n^j + d_n^j)/2 \\ s_{2n-1}^{j-1} = (s_n^j - d_n^j)/2 \end{cases} \quad \text{for } j = J,\ldots,1.$$

Similarly, we denote by $\boldsymbol{\sigma}^j$ and $\boldsymbol{\delta}^j$ the scaling and wavelet coefficients of the original signal $\boldsymbol{\mu}$ at a given scale j. Note that, by linearity of the wavelet transform, we have $\mathscr{E}\{d_n^j\} = \delta_n^j$ and $\mathscr{E}\{s_n^j\} = \sigma_n^j$.

The key properties of the unnormalized Haar DWT are the following.

1. It is an *orthogonal* transform. In particular, we can split the MSE into subband-specific error terms:

$$\text{MSE} = \frac{2^{-J}}{N}\left\|\hat{\boldsymbol{\sigma}}^J - \boldsymbol{\sigma}^J\right\|^2 + \sum_{j=1}^{J}\frac{2^{-j}}{N}\|\hat{\boldsymbol{\delta}}^j - \boldsymbol{\delta}^j\|^2.$$

   This implies that we can minimize the MSE for each subband independently, while ensuring a global signal-domain MSE minimization.
2. At a given scale j, the scaling coefficients of an input vector of independent Poisson random variables are also independent Poisson random variables, thanks to Property 1 of the Poisson distribution.

2.3. PURE: a Haar-wavelet-domain Poisson unbiased risk estimate

In principle, the estimate $\hat{\boldsymbol{\delta}}^j$ may be constructed using all available subbands of the measurement $\boldsymbol{\mu}$; in practice, however, standard thresholding techniques only use the corresponding wavelet subband of the measurement, $\mathbf{d}^j$. In the sequel, we will consider denoising algorithms where $\hat{\boldsymbol{\delta}}^j$ also depends on the scaling coefficients at the same scale $\mathbf{s}^j$. This means that we have the following functional relationship:

$$\hat{\boldsymbol{\delta}}^j = \boldsymbol{\theta}^j(\mathbf{d}^j, \mathbf{s}^j). \tag{2}$$

As usual, the lowpass residual is not processed, i.e., $\hat{\boldsymbol{\sigma}}^J = \mathbf{s}^J$. Our algorithmic framework is illustrated in Fig. 1.

The above choice is advantageous from a computational standpoint because such a restoration procedure can be implemented in parallel with the wavelet decomposition, which yields the scaling coefficients $\mathbf{s}^j$ as a by-product. Furthermore, this framework comprises advanced denoising schemes that exploit interscale dependencies via the scaling coefficients $\mathbf{s}^j$ (see Section 3.2.3). Finally, it allows us to minimize $\text{MSE}_j = \|\boldsymbol{\theta}^j(\mathbf{d}^j, \mathbf{s}^j) - \boldsymbol{\delta}^j\|^2/N_j$ independently for each wavelet subband. We will thus focus on a fixed scale and drop the superscript j to simplify the notations.

The MSE is a very popular measure of restoration quality in phantom experiments, where the ground-truth signal is known. However, it is not accessible in real situations where the original noise-free signal is unknown. Our approach is thus to construct a statistical estimate of this quantity that solely depends on the measurement. This type of estimators is very accurate in the context of image processing, because it can rely on a large number of samples. In the sequel, we borrow the term risk from the statistics literature, to refer to the expectation of the MSE.

In the Gaussian case, the scaling coefficients $\mathbf{s}^j$ are statistically independent of the wavelet coefficients $\mathbf{d}^j$ (in an orthogonal wavelet domain). This is in contrast with the Poisson case, for which $\mathbf{d}^j$ and $\mathbf{s}^j$ are statistically dependent and even correlated. This makes the derivation of a bivariate (i.e., involving both $\mathbf{d}^j$ and $\mathbf{s}^j$) SURE-like MSE estimate less obvious. In the following theorem, we give an expression of an unbiased estimate of the MSE in a given subband j of the unnormalized Haar DWT defined in Section 2.2. This novel result serves as a data-dependent quantitative measure to be minimized for Poisson intensity estimation. The family of vectors $(\mathbf{e}_n)_{n=1,\ldots,N_j}$ denotes the canonical basis of $\mathbb{R}^{N_j}$, i.e., all components of $\mathbf{e}_n$ are zero, except for the n-th component, which is equal to one.

Theorem 1. *Let $\boldsymbol{\theta}(\mathbf{d},\mathbf{s}) = \boldsymbol{\theta}^j(\mathbf{d}^j,\mathbf{s}^j)$ be an estimate of the noise-free wavelet coefficients $\boldsymbol{\delta} = \boldsymbol{\delta}^j$. Define $\boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})$ and $\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s})$ by*

$$\begin{cases} \theta_n^+(\mathbf{d},\mathbf{s}) = \theta_n(\mathbf{d}+\mathbf{e}_n, \mathbf{s}-\mathbf{e}_n), \\ \theta_n^-(\mathbf{d},\mathbf{s}) = \theta_n(\mathbf{d}-\mathbf{e}_n, \mathbf{s}-\mathbf{e}_n). \end{cases} \tag{3}$$

Then the random variable

$$\varepsilon_j = \frac{1}{N_j}(\|\boldsymbol{\theta}(\mathbf{d},\mathbf{s})\|^2 + \|\mathbf{d}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{s} - \mathbf{d}^{\mathrm{T}}(\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s}) + \boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})) - \mathbf{s}^{\mathrm{T}}(\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s}) - \boldsymbol{\theta}^+(\mathbf{d},\mathbf{s}))) \tag{4}$$

is an unbiased estimate of the MSE for the subband under consideration, i.e., $\mathscr{E}\{\varepsilon_j\} = \mathscr{E}\{\text{MSE}_j\}$.

Proof. The proof relies centrally on the fact that, within a given scale the scaling coefficients are independent Poisson random variables. We consider the case where $j = 1$, so that we can use $\mathbf{m} = \mathbf{s}^{j-1}$ and $\boldsymbol{\mu} = \boldsymbol{\sigma}^{j-1}$ to avoid superscripts.

We first develop the squared error between $\boldsymbol{\delta}$ and its estimate $\boldsymbol{\theta}(\mathbf{d},\mathbf{s})$, using the fact that $\boldsymbol{\delta}$ is a deterministic quantity:

$$\mathscr{E}\{\|\boldsymbol{\theta}(\mathbf{d},\mathbf{s}) - \boldsymbol{\delta}\|^2\} = \mathscr{E}\{\|\boldsymbol{\theta}(\mathbf{d},\mathbf{s})\|^2\} + \underbrace{\|\boldsymbol{\delta}\|^2}_{\text{(I)}} - 2\underbrace{\mathscr{E}\{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\theta}(\mathbf{d},\mathbf{s})\}}_{\text{(II)}}. \tag{5}$$

Now, we can evaluate the two expressions (I,II) that involve the unknown data $\boldsymbol{\delta}$.

(I) $\|\boldsymbol{\delta}\|^2 = \sum_{n=1}^{N_j}(\delta_n)^2$:
We notice that

$$(\delta_n)^2 = \mathscr{E}\{\mu_{2n}(m_{2n} - m_{2n-1})\} + \mathscr{E}\{\mu_{2n-1}(m_{2n-1} - m_{2n})\}.$$

By applying Property 2 for $\theta(m) = m - m_{2n-1}$ and for $\theta(m) = m - m_{2n}$, by using the fact that m_{2n} and m_{2n-1} are statistically independent, we get

$$(\delta_n)^2 = \mathscr{E}\{(m_{2n} - m_{2n-1})^2 - (m_{2n} + m_{2n-1})\} = \mathscr{E}\{(d_n)^2 - s_n\}.$$

Therefore $\|\boldsymbol{\delta}\|^2 = \mathscr{E}\{\|\mathbf{d}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{s}\}$.

(II) $\mathscr{E}\{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\theta}(\mathbf{d},\mathbf{s})\} = \sum_{n=1}^{N_j} \mathscr{E}\{\delta_n\theta_n(\mathbf{d},\mathbf{s})\}$:
We can successively write

$$\begin{aligned}
&\mathscr{E}\{\delta_n\theta_n(\mathbf{d},\mathbf{s})\}\\
&= \mathscr{E}\{\mu_{2n}\theta_n(\mathbf{d},\mathbf{s})\} - \mathscr{E}\{\mu_{2n-1}\theta_n(\mathbf{d},\mathbf{s})\}\\
&\overset{\text{Prop.2}}{=} \mathscr{E}\{m_{2n}\theta_n(\mathbf{d}-\mathbf{e}_n,\mathbf{s}-\mathbf{e}_n)\}\\
&\quad - \mathscr{E}\{m_{2n-1}\theta_n(\mathbf{d}+\mathbf{e}_n,\mathbf{s}-\mathbf{e}_n)\}\\
&= \mathscr{E}\left\{\frac{m_{2n}-m_{2n-1}}{2}(\theta_n^-(\mathbf{d},\mathbf{s}) + \theta_n^+(\mathbf{d},\mathbf{s}))\right\}\\
&\quad + \mathscr{E}\left\{\frac{m_{2n}+m_{2n-1}}{2}(\theta_n^-(\mathbf{d},\mathbf{s}) - \theta_n^+(\mathbf{d},\mathbf{s}))\right\}\\
&= \frac{1}{2}\mathscr{E}\{d_n(\theta_n^-(\mathbf{d},\mathbf{s}) + \theta_n^+(\mathbf{d},\mathbf{s}))\}\\
&\quad + \frac{1}{2}\mathscr{E}\{s_n(\theta_n^-(\mathbf{d},\mathbf{s}) - \theta_n^+(\mathbf{d},\mathbf{s}))\}.
\end{aligned}$$

Thus $2\mathscr{E}\{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\theta}(\mathbf{d},\mathbf{s})\} = \mathscr{E}\{\mathbf{d}^{\mathrm{T}}(\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s}) + \boldsymbol{\theta}^+(\mathbf{d},\mathbf{s}))\} + \mathscr{E}\{\mathbf{s}^{\mathrm{T}}(\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s}) - \boldsymbol{\theta}^+(\mathbf{d},\mathbf{s}))\}$.

Putting these results back into (5) gives the desired equality. For $j \geq 2$, the proof is based on the same idea. □

We will refer to (4) as the *Poisson unbiased risk estimate* (PURE). This estimate involves finite differences of the restoration function (instead of derivatives in the Gaussian case). It can be used to evaluate the restoration quality of any algorithm of the form (2) in terms of MSE. In the next section, we will consider algorithms that depend on a set of parameters, and we will minimize the PURE in order to obtain their optimal values.

2.4. Extension to multidimensional signal denoising

While the above result (4) is stated in a 1D setting, it can easily be extended to arbitrary dimensions when using the separable Haar DWT. Indeed, Theorem 1 essentially relies on the fact that the wavelet coefficients are differences of Poisson random variables. In higher dimensions, it suffices to observe that the wavelet coefficients are differences of *sums of* scaling coefficients; these sums are still Poisson distributed, according to Property 1.

3. PURE-based wavelet thresholding

3.1. PUREshrink

As a benchmark for illustrating our approach, we propose a wavelet-domain estimator which consists in applying the popular *soft-threshold* with a PURE-optimized threshold. Our *PUREshrink* estimator can be viewed as the transposition of Donoho and Johnstone's *Gaussian SUREshrink* [21] to Poisson noise removal. An important difference is that the method described in [21] forces the threshold T to be smaller than the universal threshold (otherwise it is set to the value of the universal threshold); this is known to be suboptimal for image-denoising applications [22]. Our threshold optimization totally relies on the minimization of the PURE (without restrictions).

Contrary to the Gaussian case, where the noise is stationary and completely described by its variance,[1] for Poisson data, the amount of noise directly depends on the intensity we want to estimate. Thus, for deciding on the amount of shrinkage, we use a threshold T that is proportional to the square root of the scaling coefficient at the same location and scale. This quantity is an estimate of the local noise standard deviation, so that it is a good reference for assessing the significance of a wavelet coefficient. Indeed, each wavelet coefficient of the unnormalized Haar transform follows a Skellam distribution [23], whose variance is equal to the sum of the two underlying Poisson intensities, i.e., approximately the corresponding scaling coefficient. Note that in the *Haar–Fisz* transform [12], the scaling coefficients are also considered as an estimate of the local noise variance and thus used to stabilize it. The *PUREshrink* estimator is therefore defined as

$$\theta_n^{\text{PUREshrink}}(\mathbf{d},\mathbf{s};a) = \text{sign}(d_n)\max(|d_n| - a\sqrt{|s_n|}, 0), \tag{6}$$

where, for each wavelet subband, the parameter a is set to the value that minimizes the PURE (4) with $\boldsymbol{\theta}(\mathbf{d},\mathbf{s}) = \boldsymbol{\theta}^{\text{PUREshrink}}(\mathbf{d},\mathbf{s};a)$.

3.2. PURE-LET

We will now introduce a family of more elaborate denoising functions and compare them with the basic PUREshrink procedure. Following our recently devised *SURE-LET* strategy [3], we propose to consider a wavelet estimator that is formulated as a *linear expansion of thresholds* (LET), i.e.,

$$\boldsymbol{\theta}^{\text{LET}}(\mathbf{d},\mathbf{s};\mathbf{a}) = \sum_{k=1}^{K} a_k\boldsymbol{\theta}_k(\mathbf{d},\mathbf{s}),$$

where the $\boldsymbol{\theta}_k$'s are generic estimators that can be chosen arbitrarily. In this section, we will gradually specify more elaborate estimators $\boldsymbol{\theta}_k$ (e.g., by taking into account redundant information across scales), and we will show how these choices yield substantially improved denoising quality.

Thanks to this linear parameterization, the unbiased estimate of the MSE in (4) is *quadratic* with respect to the parameters $\mathbf{a} \in \mathbb{R}^K$. Therefore, its minimization boils down to the resolution of a *linear* system of equations with *small* dimension K:

$$\mathbf{a} = \mathbf{M}^{-1}\mathbf{c}, \tag{7}$$

[1] The noise is usually assumed to be zero-mean.

where for $1 \le k, \ell \le K$,

$$\begin{cases} c_k = [\mathbf{d}^{\mathrm{T}}(\boldsymbol{\theta}_k^-(\mathbf{d},\mathbf{s}) + \boldsymbol{\theta}_k^+(\mathbf{d},\mathbf{s})) + \mathbf{s}^{\mathrm{T}}(\boldsymbol{\theta}_k^-(\mathbf{d},\mathbf{s}) - \boldsymbol{\theta}_k^+(\mathbf{d},\mathbf{s}))]/2, \\ M_{k,\ell} = \boldsymbol{\theta}_k(\mathbf{d},\mathbf{s})^{\mathrm{T}}\boldsymbol{\theta}_\ell(\mathbf{d},\mathbf{s}). \end{cases} \tag{8}$$

The definition of $\boldsymbol{\theta}_k^+(\mathbf{d},\mathbf{s})$ and $\boldsymbol{\theta}_k^-(\mathbf{d},\mathbf{s})$ is similar to (3).

3.2.1. Basic thresholding function

Similarly to [22], we propose a linearly parameterized thresholding function with $K = 2$ parameters (a_1 and a_2), whose n-th component is defined by

$$\theta_n^{\mathrm{LET0}}(\mathbf{d},\mathbf{s};[a_1\ a_2]^{\mathrm{T}}) = a_1 d_n + a_2\left(1 - \exp\left(-\frac{d_n^2}{2T^2}\right)\right)d_n. \tag{9}$$

In this expression, the linear parameters a_1 and a_2 define a compromise between two regimes: either the wavelet coefficient d_n is kept as is (signal preservation) or it is shrinked towards zero (noise suppression). The exponential function has the advantage of being smooth, which reduces the variance of the estimator.

As in the *PUREshrink* estimator, the threshold T is directly linked to the local noise variance, estimated from the magnitude of the corresponding scaling coefficient $|s_n|$. However, thanks to the degrees of freedom provided by the two *linear* parameters a_1 and a_2, the value of this threshold does not need to be optimized. By running several simulations, we found that $T^2 = 6|s_n|$ constituted a good choice, inducing no significant loss compared to a subband-optimized threshold. Our experiments (see Fig. 3) show that the above thresholding function is already more efficient (approx. +0.25 dB) than the previously presented *PUREshrink* (6).

3.2.2. Interscale sign dependencies

The integration of interscale dependencies has already been shown to bring a substantial improvement in the context of additive Gaussian white noise removal [22]. Therefore, we propose here an analogous interscale wavelet thresholding, but for Poisson intensity estimation. The idea is to exploit the scaling coefficients $\mathbf{s}$ to "predict" and reinforce the significant wavelet coefficients of $\mathbf{d}$ at the same scale. Indeed, the scaling coefficients offer improved SNR because they arise from Poisson random variables with summed intensities. They also contain all the information about coarser-scale wavelet coefficients.

To construct an interscale predictor of the wavelet coefficient d_n, we simply compute the difference between the two scaling coefficients that surround s_n:

$$\tilde{d}_n = s_{n-1} - s_{n+1}.$$

The whole procedure boils down to applying a centered gradient filter on the low-pass subband, as illustrated in the diagram of Fig. 2.

By taking a closer look at Fig. 2, it can be observed that the signs of the interscale predictor coefficients are consistent with those of the corresponding highpass subband. This suggests adding a term proportional to this interscale predictor into the simple thresholding function

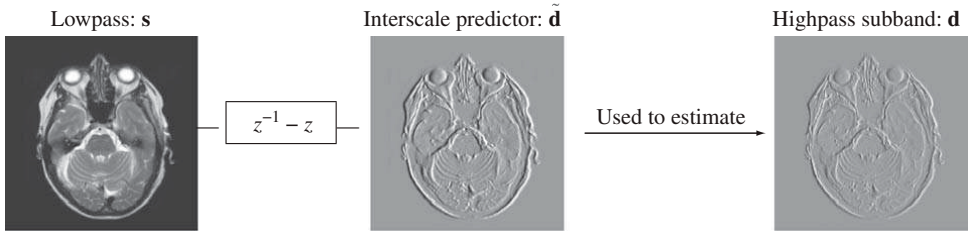

Fig. 2. The interscale predictor of subband HL_j (resp. LH_j, resp. HH_j) is obtained by horizontally (resp. vertically, resp. horizontally and vertically) filtering the same-scale lowpass subband LL_j with the centered gradient filter, whose z-transform is $z^{-1} - z$.

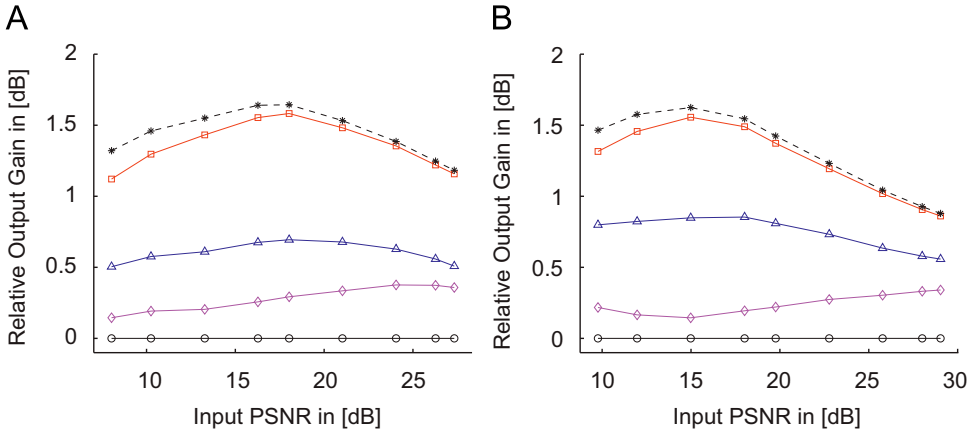

Fig. 3. PSNR improvements brought by the *PURE-LET* strategy over *PUREshrink* (Formula (6): "○" markers), for two standard grayscale images. Formula (9): "⋄" markers. Formula (10): "△" markers. Formula (11): "□" markers). Oracle results for Formula (11) are shown with "∗" markers: (A) Cameraman; (B) MIT.

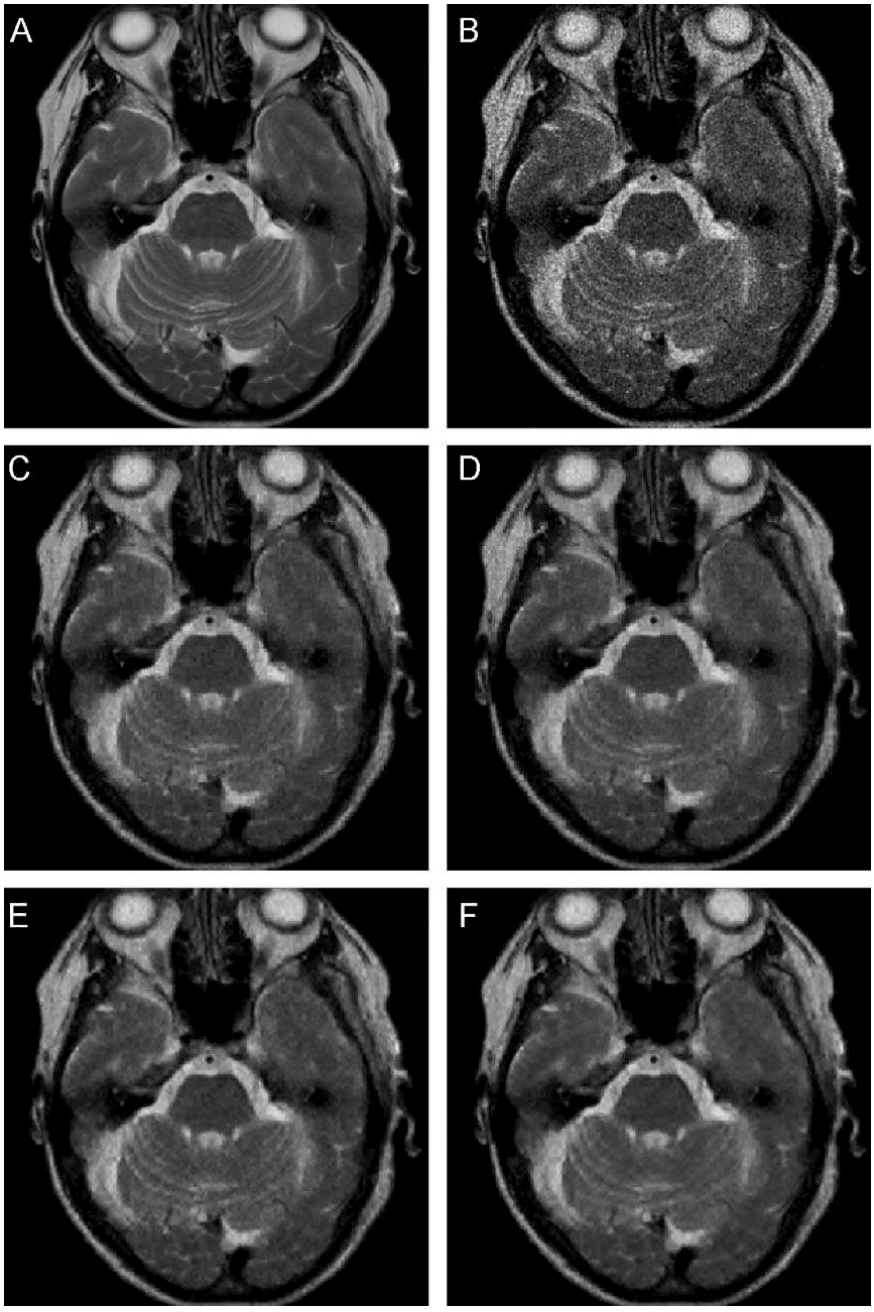

Fig. 4. Visual quality of the various estimators presented in Section 3. (A) Part of the original *MRI* slice ($I_{\max} = 20$). (B) Noisy realization of it: PSNR = 22.31 dB. (C) Denoised by *PUREshrink* (6): PSNR = 29.53 dB. (D) Denoised by *PURE-LET*0 (9): PSNR = 29.81 dB. (E) Denoised by *PURE-LET*1 (10): PSNR = 31.19 dB. (F) Denoised by *PURE-LET*2 (11): PSNR = 31.43 dB.

(9), leading to

$$\theta_n^{\mathrm{LET1}}(\mathbf{d},\mathbf{s};[a_1\ a_2\ a_3]^{\mathrm{T}}) = \theta_n^{\mathrm{LET0}}(\mathbf{d},\mathbf{s};[a_1\ a_2]^{\mathrm{T}}) + a_3 \underbrace{(s_{n-1}-s_{n+1})}_{\tilde{d}_n}. \tag{10}$$

This simple strategy brings significant improvements (approximately +0.5 dB). This was confirmed by multiple experiments on standard grayscale images; some of these simulations are reported in Fig. 3.

3.2.3. *Interscale amplitude-sign dependencies*

Further improvements can be obtained by grouping together wavelet coefficients of similar magnitudes [22]. To increase the robustness towards noise, this grouping is based on the magnitude of a smoothed version $\mathbf{p}$ of the previously defined interscale predictor $\tilde{\mathbf{d}}$. The smoothed version of the interscale predictor is simply obtained by applying a normalized Gaussian kernel on the absolute value of $\tilde{\mathbf{d}}$, i.e., $p_n = \sum_k |\tilde{d}_k| f_{n-k}$, where $f_k = e^{-k^2/2}/\sqrt{2\pi}$. The proposed interscale wavelet thresholding is thus

Table 1
Comparison of multiscale Poisson noise removal algorithms.

Images	*Peppers* (256 × 256)							*Cameraman* (256 × 256)						
I_{max}	120	60	30	20	10	5	1	120	60	30	20	10	5	1
Input PSNR	23.92	20.92	17.91	16.17	13.14	10.12	3.14	24.05	21.03	18.03	16.27	13.27	10.25	3.28
Non-redundant														
Anscombe + *SUREshrink* (OWT *sym*8)	29.29	27.27	24.85	23.56	21.47	19.77	15.32	28.53	26.37	24.26	23.29	21.81	19.61	15.36
Haar−Fisz + *SUREshrink* (OWT*sym*8)	29.27	27.30	24.54	23.82	21.98	21.03	18.69	28.49	26.40	24.36	23.69	22.39	20.93	18.57
Platelet	29.07	27.44	25.73	24.92	23.23	21.57	18.17	28.29	26.79	25.44	24.60	23.24	21.49	18.70
Interscale *PURE-LET* (OWT *Haar*)	30.28	**28.51**	**26.72**	**25.70**	**23.81**	**21.99**	**18.92**	**30.07**	**28.28**	**26.54**	**25.55**	**23.94**	**22.42**	**19.18**
Redundant														
Haar−Fisz + *SUREshrink* (25 cyclic shifts of OWT *sym*8)	30.29	28.37	25.45	24.49	22.43	21.53	18.90	29.35	27.30	25.09	24.46	23.02	21.46	18.71
Platelet (2 cyclic shifts)	30.61	28.52	27.08	26.11	**24.34**	22.13	18.88	29.34	27.84	26.21	25.32	24.05	22.35	19.34
Interscale *PURE-LET* (2 cyclic shifts of OWT *Haar*)	**30.79**	**29.07**	**27.27**	**26.18**	24.29	**22.52**	**19.33**	**30.36**	**28.56**	**26.87**	**25.89**	**24.32**	**22.76**	**19.67**

Images	*MIT* (256 × 256)							*Moon* (512 × 512)						
I_{max}	120	60	30	20	10	5	1	120	60	30	20	10	5	1
Input PSNR	25.78	22.78	19.76	18.02	14.98	11.97	4.99	26.27	23.25	20.23	18.48	15.47	12.46	5.47
Non-redundant														
Anscombe + *SUREshrink* (OWT *sym*8)	28.80	26.70	24.10	22.58	21.21	18.14	13.93	29.00	26.92	24.83	24.09	22.99	21.38	18.37
Haar−Fisz + *SUREshrink* (OWT *sym*8)	28.78	26.74	24.25	22.88	21.78	20.00	16.03	29.03	26.89	25.01	24.33	23.55	22.78	21.98
Platelet	28.25	26.44	24.70	23.47	21.65	20.18	16.24	27.16	26.01	25.05	24.60	23.96	23.63	22.97
Interscale *PURE-LET* (OWT *Haar*)	**30.03**	**27.92**	**25.96**	**24.85**	**23.01**	**21.18**	**17.43**	**29.62**	**27.97**	**26.56**	**25.87**	**24.92**	**24.23**	**23.16**
Redundant														
Haar−Fisz + *SUREshrink* (25 cyclic shifts of OWT *sym*8)	29.67	27.67	25.21	23.77	22.64	20.89	16.59	29.61	27.39	25.40	24.68	23.83	23.06	22.09
Platelet (2 cyclic shifts)	29.23	27.66	25.64	24.63	22.99	21.27	17.27	27.99	26.76	25.60	25.09	24.28	23.74	22.69
Interscale *PURE-LET* (2 cyclic shifts of OWT *Haar*)	**30.47**	**28.41**	**26.55**	**25.34**	**23.49**	**21.63**	**17.82**	**29.77**	**28.09**	**26.70**	**25.97**	**24.99**	**24.28**	**23.19**

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

Table 2
Cycle-spinning: PSNR improvement and computation time.

Algorithms	Non-redundant		2 cyclic shifts		25 cyclic shifts	
	PSNR (dB)	Time (s)	PSNR (dB)	Time (s)	PSNR (dB)	Time (s)
Cameraman at 17.25 dB						
Haar−Fisz + *SUREshrink*	24.01	0.06	24.49	0.11	24.83	1.3
Interscale *PURE-LET*	26.09	0.19	26.45	0.37	26.75	4.6
Platelet	25.01	44.5	25.83	89	27.17	1112
MIT at 18.98 dB						
Haar−Fisz + *SUREshrink*	23.47	0.06	24.10	0.11	24.64	1.3
Interscale *PURE-LET*	25.48	0.19	26.02	0.37	26.43	4.6
Platelet	24.10	36	25.33	72	26.80	891

Note: Output PSNRs and computation times have been averaged over 10 noise realizations, except for the *Platelet* approach.

finally defined as

$$\theta_n^{\mathrm{LET2}}(\mathbf{d},\mathbf{s};\mathbf{a},\mathbf{b}) = \exp\left(-\frac{p_n^2}{12|s_n|}\right)\theta_n^{\mathrm{LET1}}(\mathbf{d},\mathbf{s};\mathbf{a}) + \left(1-\exp\left(-\frac{p_n^2}{12|s_n|}\right)\right)\theta_n^{\mathrm{LET1}}(\mathbf{d},\mathbf{s};\mathbf{b}). \quad (11)$$

The above thresholding function has mainly two regimes: when the squared magnitude of the predictor p_n is small with respect to the local noise variance, LET2 essentially behaves like LET1 with parameter **a**; when p_n^2 is large with respect to the local noise variance, LET2 essentially behaves like LET1 with parameter **b**. This "classification" based on the predictor coefficients increases the adaptivity of the denoising process.

In Figs. 3 and 4, it is seen that this latter interscale wavelet estimator clearly gives the best results, both quantitatively and visually, among all estimators presented here. Note that the PURE-based adjustment of the parameters **a** and **b** gives a SNR gain that is very close to

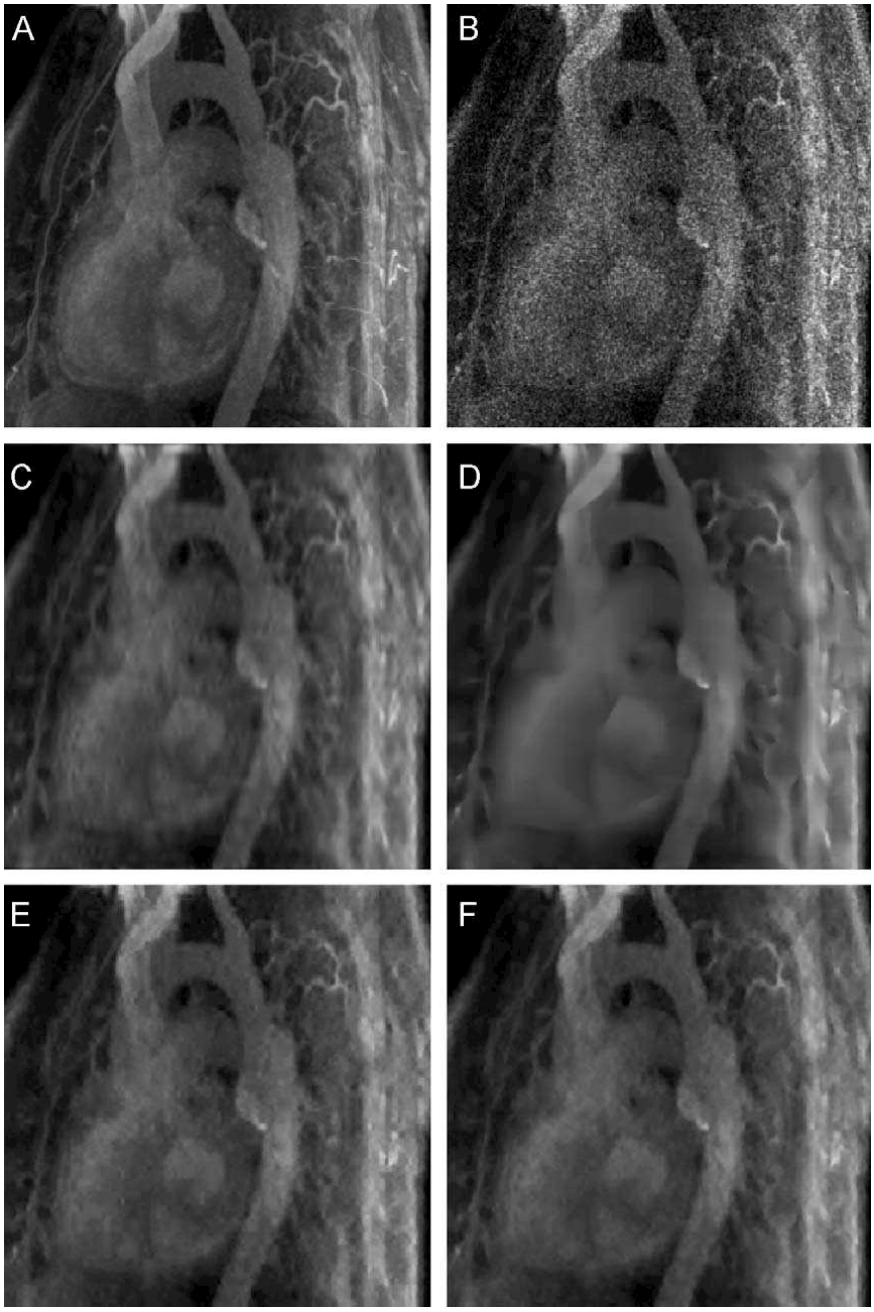

Fig. 5. (A) The original *Pneumatix* image ($I_{max} = 30$), by courtesy of the following website: http://pubimage.hcuge.ch:8080/. (B) Noisy realization of it: PSNR = 20.27 dB. (C) Denoised with *Haar−Fisz + SUREshrink* (25 cyclic shifts of OWT *sym*8): PSNR = 29.16 dB in 1.3 s. (D) Denoised with *Platelet* (25 cyclic shifts): PSNR = 29.52 dB in 960 s. (E) Denoised with our non-redundant *PURE-LET* (OWT *Haar*): PSNR = 29.21 dB in 0.2 s. (F) Denoised with our redundant *PURE-LET* (25 cyclic shifts of OWT *Haar*): PSNR = 29.73 dB in 4.4 s.

the optimum (which is obtained from an *oracle* adjustment of these parameters using the knowledge of the original image).

4. Results on simulated data

We now propose to compare our PURE-based approach with the following multiscale methods in simulated experiments:

- A standard procedure, which consists of three steps:
  1. Apply a variance-stabilizing transformation (VST) on the Poisson-distributed data. In the experiments, we have tried the Anscombe [11] and the Haar–Fisz [12] transforms.
  2. Use any available denoiser designed for additive Gaussian white noise removal: we have chosen Donoho and Johnstone's (subband-dependent) *SURE-shrink* [21].

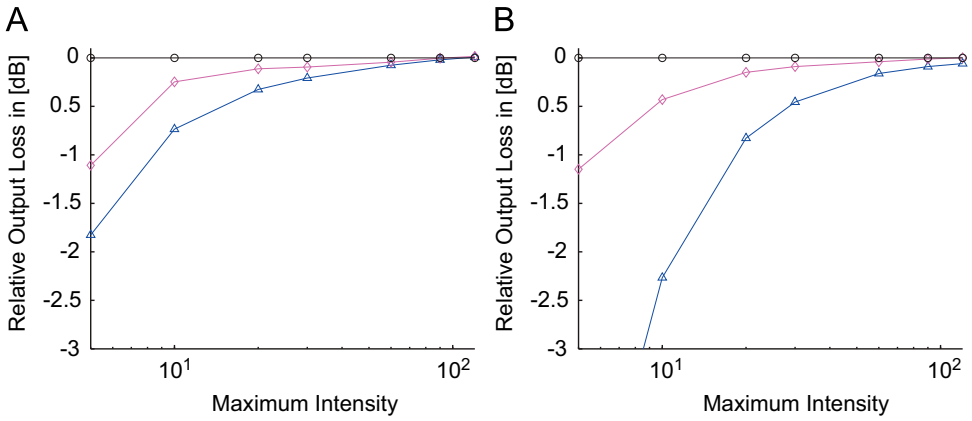

Fig. 6. PSNR improvements brought by the direct handling of Poisson statistics (*PURE-LET* (OWT Haar): "○" markers) over VST + *SURE-LET* schemes (*Anscombe* + *SURE-LET* (OWT Haar): "△" markers; *Haar*−*Fisz* + *SURE-LET* (OWT Haar): "⋄" markers): (A) Peppers; (B) MIT.

3. Apply the corresponding inverse variance-stabilizing transformation to the denoised data to finally get an estimation of the underlying Poisson intensities.

This type of approach has the advantage of giving relatively good results at low computational cost; moreover it is easily reproducible.

- A more involved algorithm: we have retained Willett and Nowak's *Platelet* approach,[2] [24,9] which stands among the state-of-the-art algorithms for Poisson intensity estimation [14].

For all wavelet-based methods presented in this section, we use the same number of decomposition levels, i.e., 4 (resp. 5) for 256×256 (resp. 512×512) images. The input peak-signal-to-noise ratios (PSNR) are adjusted by rescaling the original test images,[3] from a maximum intensity of 120 to a minimum of 1.

Table 1 summarizes the PSNRs obtained by the various algorithms, both in a non-redundant and in a redundant framework (using cycle spinning [25]). It can be observed that the PURE-based approach clearly outperforms (around +1.5 dB, on average) the standard VST-based wavelet denoisier applied in an *orthonormal* wavelet basis. Note that the retained basis (Daubechies' *symlets* with eight vanishing moments [26]) is smoother—and thus more suitable for image-restoration tasks—than the basic *Haar* wavelets that we use. Our solution also gives significantly better PSNRs (∼0.5−1 dB) than the non-redundant version of the *Platelet* approach. In most cases, the proposed algorithm gives even better results (between 0.5 and 1.5 dB, on average) than the VST-based thresholding applied in a *redundant* cycle-spinning framework, and results similar to two cyclic shifts of *Platelets*.

Table 2 gives more insights concerning the tradeoff between the degree of redundancy and the computation time of the various algorithms. The *Platelet* procedure achieves the best PSNRs when considering a high number (25) of cyclic shifts. However, these results are obtained at a prohibitive computational cost. Cyclic shifts of our PURE-based approach also brings some gains (around 1 dB), despite the fact that an independent "shift-wise" PURE minimization is sub-optimal (as shown in [3,27] for the Gaussian case). There is probably room for improvement by deriving a rigorous unbiased estimate of the MSE for redundant processing of Poisson data; but this is outside the scope of the present paper.

As shown in Fig. 5, our interscale *PURE-LET* algorithm removes most of the Poisson noise, without over-smoothing the underlying intensities. Moreover, from a computational point of view, it takes only ∼0.5 s to denoise a 512×512 image with the current Matlab implementation of our algorithm; this corresponds to the optimization of 90 parameters. Under the same conditions, the execution of the cycle-spinning *SUREshrink* (25 cyclic shifts) combined with the *Haar–Fisz* variance-stabilizing transform requires ∼5.5 s, while a single shift of the *Platelets* lasts 150 s, on average.

We also compared the proposed *PURE-LET* algorithm with our previously described interscale *SURE-LET* strategy, specifically devised for Gaussian noise statistics [22]. For a fair comparison, we used an adapted implementation of the *SURE-LET* algorithm that involved the same number of parameters as the present *PURE-LET* method. We also considered the same wavelet transformation, i.e., OWT Haar, for both techniques. As can be seen in Fig. 6, applying the *SURE-LET* strategy in the VST-domain is less efficient for small intensities (over 0.5 dB loss for intensities lower than 10). This can be attributed to the rigorous minimization of an estimate of the actual MSE that is performed by the proposed *PURE-LET* algorithm.

To conclude this series of synthetic experiments, let us highlight the two main aspects of our approach:

1. We use sophisticated denoising functions with multiple linear parameters per subband (LET representation) that are able to process, among others, interscale dependencies; these typically outperform a simple

[2] Matlab code downloadable at: http://www.ee.duke.edu/∼willett/Research/platelets.html.

[3] Available at: http://bigwww.epfl.ch/luisier/Test-Images.zip.

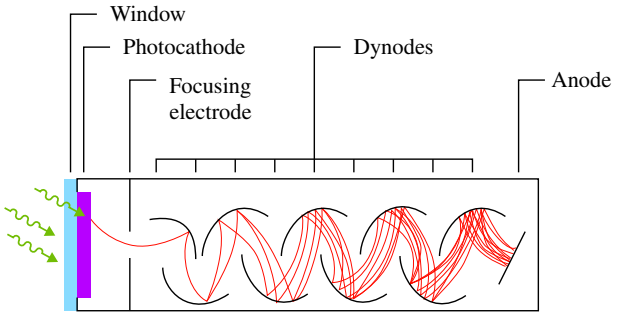

Fig. 7. Schematic representation of a photomultiplier tube (PMT).

thresholding with one parameter per subband (see Fig. 3).

2. The parameters are adjusted using a *noisy data*-based estimate of the MSE that is statistically unbiased; this typically yields better results than alternative adjustment methods (see Fig. 6).

It is the combination of these two facts that makes our method competitive with the state of the art, as shown in Table 1.

5. Application to real biological data

In this last section, we describe the application of our denoising algorithm to real fluorescence-microscopy images of biological samples. The images were acquired on a Leica TCS SP5 confocal microscope at the Imaging Center of the IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France). This microscope is equipped with a water-cooled scan-head and low-noise PMTs[4] (Hamamatsu R 9624). We refer to [1] for a general introduction to fluorescence microscopy.

5.1. Brief description of the acquisition process

The measurement process is the same for every scan position (pixel); it is illustrated in Fig. 7 and can be summarized as follows (see e.g., [28] for a more detailed description). The number of photons arriving at the PMT during the integration time follows a Poisson distribution of mean λ determined by the source intensity. Each photon may traverse the protection window, penetrate the photocathode and be converted to an electron with a certain probability η; this probability is known as the quantum efficiency of the PMT and is on the order of 30% for the best models. The conversion process can be seen as a binomial selection [20,29] and according to Proposition 3, the number of electrons at the output of the photocathode (= photoelectrons) follows a Poisson distribution of mean $\mu = \eta\lambda$. The number of photoelectrons represents a shot-noise-corrupted measurement m of the intensity signal μ in our framework.

The electrons are then multiplied (via several amplification stages) and converted to an electric current that is integrated and quantized. The recorded signal is essentially proportional to the number of photoelectrons; although the amplification factor may fluctuate in practice, recent work [30] suggests that the newest PMTs with high first-stage gain have the ability to discriminate between multi-photoelectron events, at least for low count numbers. Under this assumption the measurement process is purely shot-noise limited and one can estimate the number of photoelectrons by adequately renormalizing the data.

5.2. Data preprocessing and verification of the statistical model

Our normalization procedure is based on the characteristic property of a Poisson random variable that its mean is equal to its variance. Thus, by dividing the data by the amplification factor (gain), we should approximately retrieve this behavior.

To validate our approach, we acquired a set of 100 images of the same object (a *Caenorhabditis elegans* embryo) under low illumination intensity. We could thus compute estimates of the mean and variance for every pixel. The amplification factor (gain) was determined by fitting a linear function to these mean–variance measurements. After dividing by the gain (and subtracting a constant corresponding to the offset of the detector), the frequency distribution of the pixel values was found to be in good agreement with Poisson statistics. Fig. 8 shows the histogram of the normalized pixel values for those pixels whose mean was equal to a given value μ ($\pm 5\%$).

In practice one cannot use multiple realizations of the same image to obtain the amplification gain. We found that estimates of the local mean and variance based on spatial averaging can yield a good estimate of the gain. This approach has been used for the results presented in the next subsection.

5.3. Denoising of color fluorescence micrographs

Fig. 9(A) shows the confocal image of a *C. elegans* embryo labeled with three fluorescent dyes. We reproduced very noisy experimental conditions by reducing the laser power and using short exposure times. The three color channels were processed separately.

The result of the SUREShrink algorithm with a Haar–Fisz transform and 25 cyclic shifts is shown in Fig. 9(C). The result of our algorithm is sharper and shows less artifacts (Fig. 9(D)). Besides, it is less noisy than the image shown in Fig. 9(B), which corresponds to a 4 times longer exposure time.

6. Conclusion

The above results suggest that our PURE-based approach is a promising alternative for denoising Poisson-corrupted images. Although our PURE-LET method is based on a maximally decimated Haar wavelet transform, it yields results that are comparable or superior to standard translation-invariant approaches. At the same

[4] PMT stands for photomultiplier tube.

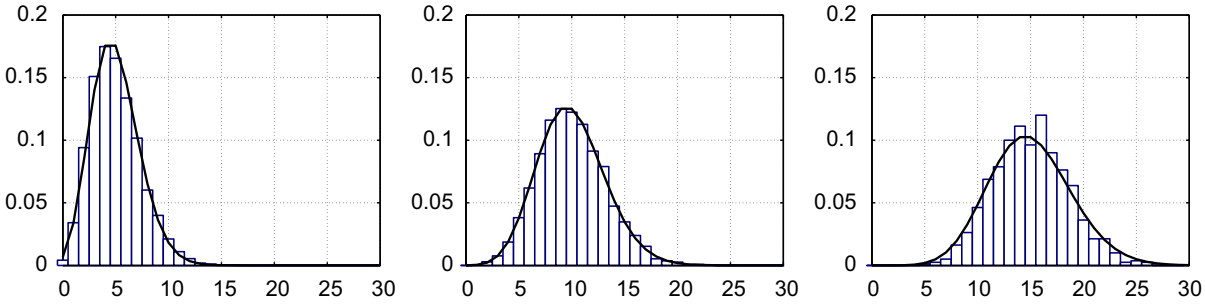

Fig. 8. Comparison of the Poisson distribution with the histogram of the normalized data for $\mu = 5, 10, 15$.

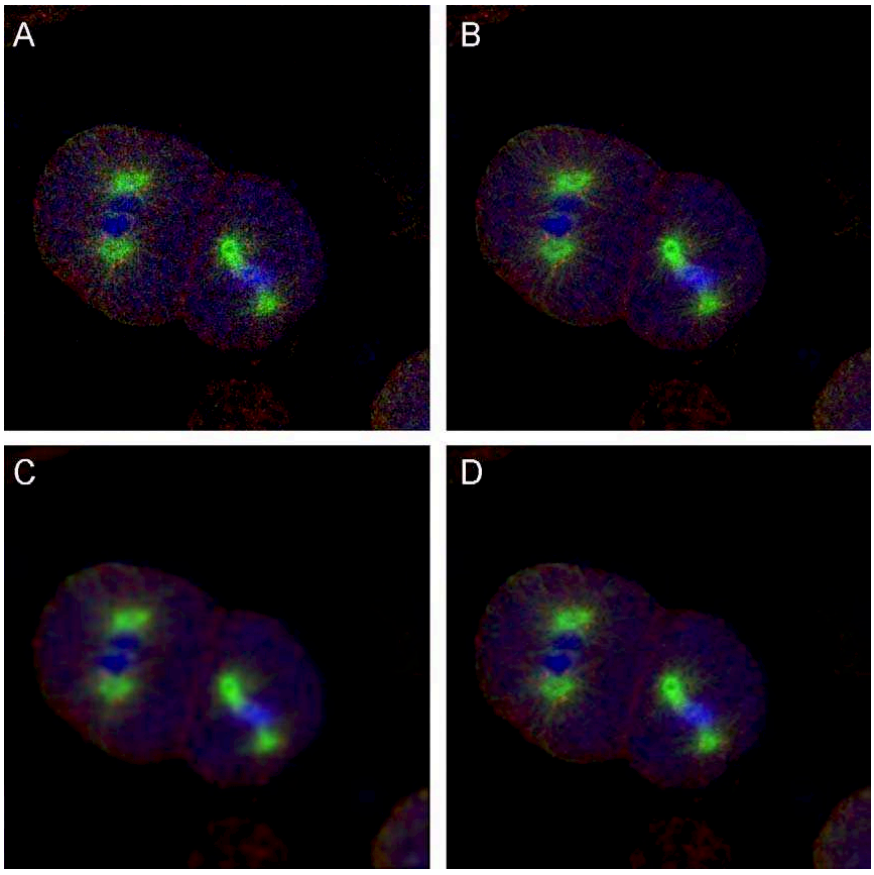

Fig. 9. (A) Input image for denoising (scan frequency of 1000 Hz). (B) Reference image (average of four images at 1000 Hz). (C) Input image denoised with *Haar*−*Fisz* + *SUREshrink* (25 cyclic shifts of OWT *sym*8). (D) Input image denoised with our interscale *PURE-LET* (OWT *Haar*).

time, our algorithm has substantially lower computational complexity and smaller memory requirements than the latter.

These are appealing features for applications in dynamic fluorescence microscopy, where biologists must often acquire large time-lapse image series under very low light conditions. Our method potentially allows for a reduction in exposure time and/or laser intensity, which are critical parameters when imaging fast-moving live samples.

The methodology presented here paves the way for further research. We are currently working on an extension of our method that is applicable to a more general noise model, suitable for imaging situations that are not necessarily shot-noise-limited. We are also trying to adapt the method to more general (possibly redundant) wavelet transforms, or even arbitrary linear transforms. This could allow to overcome one of the limitations of wavelet representations, which is that they are mostly suitable for piecewise smooth signals, and less well for textured images. These extensions will be the subject of a future report.

Acknowledgments

The authors would like to thank the team of the IGBMC Imaging Center for providing the equipment and the

assistance for acquiring the data sets presented in Section 5. They would also like to thank Prof. Pierre Gönczy from the Swiss Institute for Experimental Cancer Research (ISREC, EPFL) for providing the biological sample. This work was supported by the Hasler foundation and the Center for Biomedical Imaging (CIBM) of the Geneva - Lausanne Universities and the EPFL, as well as by Direct Grant #2050420 from the Hong Kong Research Grants Council.

References

[1] C. Vonesch, F. Aguet, J.-L. Vonesch, M. Unser, The colored revolution of bioimaging, IEEE Signal Processing Magazine 23 (3) (2006) 20–31.
[2] P. Besbeas, I.D. Feis, T. Sapatinas, A comparative simulation study of wavelet shrinkage estimators for Poisson counts, International Statistical Review 72 (2) (2004) 209–237.
[3] T. Blu, F. Luisier, The SURE-LET approach to image denoising, IEEE Transactions on Image Processing 16 (11) (2007) 2778–2786.
[4] K.E. Timmermann, R.D. Nowak, Multiscale modeling and estimation of Poisson processes with application to photon-limited imaging, IEEE Transactions on Information Theory 45 (3) (1999) 846–862.
[5] H. Lu, Y. Kim, J.M.M. Anderson, Improved Poisson intensity estimation: denoising application using Poisson data, IEEE Transactions on Image Processing 13 (8) (2004) 1128–1135.
[6] E.D. Kolaczyk, Bayesian multi-scale models for Poisson processes, Journal of the American Statistical Association 94 (447) (1999) 920–933.
[7] E.D. Kolaczyk, Nonparametric estimation of intensity maps using Haar wavelets and Poisson noise characteristics, The Astrophysical Journal 534 (2000) 490–505.
[8] S. Sardy, A. Antoniadis, P. Tseng, Automatic smoothing with wavelets for a wide class of distributions, Journal of Computational & Graphical Statistics 13 (2) (2004) 399–423.
[9] R.M. Willett, R.D. Nowak, Multiscale Poisson intensity and density estimation, IEEE Transactions on Information Theory 53 (9) (2007) 3171–3187.
[10] D.L. Donoho, Nonlinear wavelet methods for recovery of signals densities and spectra from indirect and noisy data, in: I. Daubechies (Ed.), Different Perspectives on Wavelets, Proceedings of Symposia in Applied Mathematics, vol. 47, American Mathematical Society, Providence, RI, USA, 1993, pp. 173–205.
[11] F.J. Anscombe, The transformation of Poisson, binomial and negative binomial data, Biometrika 35 (3/4) (1948) 246–254.
[12] P. Fryzlewicz, G.P. Nason, A Haar–Fisz algorithm for Poisson intensity estimation, Journal of Computational and Graphical Statistics 13 (3) (2004) 621–638.
[13] M. Jansen, Multiscale Poisson data smoothing, Journal of the Royal Statistical Society B 68 (1) (2006) 27–48.
[14] B. Zhang, J.M. Fadili, J.L. Starck, Wavelets, ridgelets and curvelets for Poisson noise removal, IEEE Transactions on Image Processing 17 (7) (2008) 1093–1108.
[15] R.D. Nowak, R.G. Baraniuk, Wavelet-domain filtering for photon imaging systems, IEEE Transactions on Image Processing 8 (5) (1999) 666–678.
[16] A. Antoniadis, T. Sapatinas, Wavelet shrinkage for natural exponential families with quadratic variance functions, Biometrika 88 (3) (2001) 805–820.
[17] C. Stein, Estimation of the mean of a multivariate normal distribution, The Annals of Statistics 9 (1981) 1135–1151.
[18] H.M. Hudson, A natural identity for exponential families with applications in multiparameter estimation, The Annals of Statistics 6 (3) (1978) 473–484.
[19] K.W. Tsui, S.J. Press, Simultaneous estimation of several Poisson parameters under K-normalized squared error loss, The Annals of Statistics 10 (1) (1982) 93–100.
[20] L. Takács, Über die Wahrscheinlichkeitstheoretische Behandlung der Anodenstromschwankungen von Elektronenröhren, Acta Physica Academiae Scientiarum Hungaricae 7 (1957) 25–50.
[21] D.L. Donoho, I.M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, Journal of the American Statistical Association 90 (432) (1995) 1200–1224.
[22] F. Luisier, T. Blu, M. Unser, A new SURE approach to image denoising: interscale orthonormal wavelet thresholding, IEEE Transactions on Image Processing 16 (3) (2007) 593–606.
[23] J.G. Skellam, The frequency distribution of the difference between two Poisson variates belonging to different populations, Journal of the Royal Statistical Society 109 (3) (1946) 296.
[24] R.M. Willett, R.D. Nowak, Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging, IEEE Transactions on Medical Imaging 22 (3) (2003) 332–350.
[25] R.R. Coifman, D.L. Donoho, Translation Invariant De-noising, in: Lecture Notes in Statistics: Wavelets and Statistics, vol. 103, Springer, New York, 1995, pp. 125–150.
[26] I. Daubechies, Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 61, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992.
[27] M. Raphan, E.P. Simoncelli, Optimal denoising in redundant representations, IEEE Transactions on Image Processing 17 (8) (2008) 1342–1352.
[28] J. Pawley, Sources of noise in three-dimensional microscopical data sets, Three-dimensional Confocal Microscopy: Volume Investigation of Biological Specimens, Academic Press, New York, 1994, pp. 47–94.
[29] H.H. Barrett, K.J. Myers, Foundations of Image Science, Wiley, New York, 2003.
[30] A.G. Wright, The statistics of multi-photoelectron pulse-height distributions, Nuclear Instruments and Methods in Physics Research A 579 (2007) 967–972.
[31] K. Hirakawa, P.J. Wolfe. Skellam shrinkage: Wavelet-based intensity estimation for inhomogeneous Poisson data. 2009, Submitted.