UNDECIMATED HAAR THRESHOLDING FOR POISSON INTENSITY ESTIMATION

Florian Luisier[1], *Thierry Blu*[2] *and Michael Unser*[1]

[1]Biomedical Imaging Group, Ecole Polytechnique Fédérale de Lausanne, Switzerland
[2]Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong

ABSTRACT

We propose a novel algorithm for denoising Poisson-corrupted images, that performs a signal-adaptive thresholding of the undecimated Haar wavelet coefficients. A Poisson's unbiased MSE estimate is devised and adapted to arbitrary transform-domain pointwise processing. This prior-free quadratic measure of quality is then used to globally optimize a linearly parameterized subband-adaptive thresholding, which accounts for the signal-dependent noise variance. We demonstrate the qualitative and computational competitiveness of the resulting denoising algorithm through comprehensive comparisons with some state-of-the-art multiscale techniques specifically designed for Poisson intensity estimation. We also show promising denoising results obtained on low-count fluorescence microscopy images.

Index Terms— Image denoising, Poisson noise, Haar wavelet, MSE estimation, fluorescence microscopy

1. INTRODUCTION

Digital image acquisition involves several optical and electronic devices, whose imperfections (e.g. thermal instabilities) generate noise in the acquired image. Besides these mostly *signal-independent* degradations, the intrinsic randomness of photon emission and detection also contributes to decrease the image quality. When the measured light intensities are sufficiently high, the common additive white Gaussian noise (AWGN) assumption remains relevant. In this work, we focus on the low illumination conditions, where the photon-counting noise becomes dominant. In this scenario, the measurements are more accurately modeled as independent Poisson variables. These low-count imaging conditions are typically encountered in biomedical imaging (e.g. in fluorescence microscopy) and in astronomy. Denoising of Poisson-corrupted images then becomes an essential step prior to visualization and extraction of meaningful information.

There are two main approaches for Poisson intensity estimation:

- The first consists of a three-step procedure: application of a variance-stabilizing transform (VST) (e.g. Anscombe [1] or Haar-Fisz [2]), processing of the stabilized data with an existing AWGN denoiser, and inversion of the VST. The advantage of this approach is to take advantage of the huge amount of work dedicated to AWGN reduction (e.g. [3,4]). Yet, most VSTs fail to stabilize the noise variance for very low-intensity images, making the AWGN assumption inaccurate.
- The second is the direct handling of Poisson statistics often in a Bayesian [5–7] or in a penalized likelihood [8] framework. Statistical priors or penalty terms are commonly formulated in a multiscale decomposition, where the Poisson intensities are sparsely represented. In particular, several denoising algorithms are constructed around the non-redundant (unnormalized) Haar wavelet transform, taking advantage of its multiscale preservation of Poisson statistics [5, 9, 10]. The algorithms specifically tailored to Poisson data do not lead to a drop of denoising performances at low-count regime, but they require the development of new estimators.

Note that some recent algorithms combine VST with multiscale Bayesian models [11] and multiscale hypothesis testing [12].

In a recent paper [10], we proposed a fast algorithm for estimating Poisson intensities that was restricted to the use of the *non-redundant* (unnormalized) Haar wavelet transform. In this work, we lift this restriction by devising a novel *undecimated* Haar wavelet thresholding. Our main contribution is the development of an image-domain *prior-free* Poisson's unbiased risk estimate (PURE). In particular, we propose a tractable approximation of PURE for an arbitrary transform-domain pointwise processing. We then use PURE to optimize an undecimated Haar wavelet-domain linear expansion of thresholds (LET), adapted to the signal-dependent noise variance. Although the thresholding is subband-adaptive, its parameters optimization is performed in the *image-domain* to ensure a *global* MSE optimality [4, 13]. The proposed PURE-LET strategy can thus be seen as the Poisson extension of the SURE-LET paradigm [4].

2. THEORY

We assume that the discrete imaging of an original natural scene yields a Poisson-corrupted observation $\mathbf{y} = [y_1\ y_2 \dots y_N]^{\mathrm{T}}$, i.e. N independent Poisson random variables y_n of underlying intensities x_n. Each vector of the canonical basis of $\mathbb{R}^N$ will be denoted by $\mathbf{e}_n = [\delta_{n-k}]_{1 \le k \le N}$.

2.1. PURE: a Poisson's Unbiased Risk Estimate

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

$$\mathrm{MSE} = \frac{1}{N}\|\hat{\mathbf{x}} - \mathbf{x}\|^2 = \frac{1}{N}\sum_{n=1}^{N}(f_n(\mathbf{y}) - x_n)^2 \qquad (1)$$

The next theorem gives us a practical estimate of the MSE that does not require any knowledge about the unknown intensities $\mathbf{x}$. In contrast with Bayesian approaches, $\mathbf{x}$ is even considered to be *deterministic* here.

This work was supported by the Center for Biomedical Imaging (CIBM) of the Geneva - Lausanne Universities and the EPFL, the foundations Leenaards and Louis-Jeantet, the Swiss National Science Foundation under grant 200020-109415, as well as by a Direct Grant #2050420 from the Hong Kong Research Grants Council.

Theorem 1 (PURE). *Let* $\mathbf{F}(\mathbf{y}) = [f_n(\mathbf{y})]_{1\leq n\leq N}$ *be a real-valued vector function such that* $\forall n = 1\dots N$, $\mathscr{E}\{|f_n(\mathbf{y})|\} < \infty$ *and define* $\mathbf{F}^-(\mathbf{y}) = [f_n(\mathbf{y}-\mathbf{e}_n)]_{1\leq n\leq N}$. *Then the random variable*

$$\epsilon = \frac{1}{N}\left(\|\mathbf{F}(\mathbf{y})\|^2 - 2\mathbf{y}^{\mathrm{T}}\mathbf{F}^-(\mathbf{y}) + \|\mathbf{y}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{y}\right) \qquad (2)$$

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

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

The proof of Theorem 1 is based on Hudson's identity [14], and will appear elsewhere.

The variance of PURE (2) decreases as the number N of measurements y_n increases. Because of the large number of pixels in images, PURE can be considered as a very reliable MSE estimate.

2.2. Tractable Approximation of PURE

The computationally tedious part of PURE (2) lies in the evaluation of the term $\mathbf{F}^-(\mathbf{y})$ for an arbitrary nonlinear processing. Indeed, to compute a single component $f_n(\mathbf{y}-\mathbf{e}_n)$ of $\mathbf{F}^-(\mathbf{y})$, one needs to apply the whole denoising process to a slightly perturbed version of the noisy input. This operation has to be repeated N times to get the full vector $\mathbf{F}^-(\mathbf{y})$. Such a "brute force" approach cannot be envisaged in practice, since a typical image contains $N = 256^2$ pixels. Instead, we propose to use a first order Taylor series approximation of $\mathbf{F}^-(\mathbf{y})$, i.e. for all $n = 1\dots N$:

$$\mathbf{F}^-(\mathbf{y}) \quad \simeq \quad \mathbf{F}(\mathbf{y}) - \partial\mathbf{F}(\mathbf{y})$$

where $\partial\mathbf{F}(\mathbf{y}) = \left[\frac{\partial f_n(\mathbf{y})}{\partial y_n}\right]$, assuming that $f_n \in C^1(\mathbb{R}^N)$, $\forall n$.

Consequently, provided that the above approximation is reasonable (i.e. $|y_n| >> 1$), PURE is well approximated by:

$$\tilde{\epsilon} \quad = \quad \frac{1}{N}\left(\|\mathbf{F}(\mathbf{y}) - \mathbf{y}\|^2 + 2\mathbf{y}^{\mathrm{T}}\partial\mathbf{F}(\mathbf{y}) - \mathbf{1}^{\mathrm{T}}\mathbf{y}\right) \qquad (3)$$

2.3. PURE for Transform-Domain Pointwise Processing

We propose now to derive a tractable PURE for the particular case of a transform-domain pointwise processing,

$$\mathbf{F}(\mathbf{y}) = \mathbf{R}\boldsymbol{\Theta}(\underbrace{\mathbf{D}\mathbf{y}}_{\mathbf{w}}, \underbrace{\widetilde{\mathbf{D}}\mathbf{y}}_{\widetilde{\mathbf{w}}}) \qquad (4)$$

where:

- $\mathbf{D}$ is a $L \times N$ matrix which implements a forward transformation (e.g. a redundant wavelet decomposition).
- $\mathbf{R}$ is a $N \times L$ matrix which implements the inverse transformation associated with $\mathbf{D}$, such that $\mathbf{RD} = \mathbf{Id}$.
- $\widetilde{\mathbf{D}}$ is a $L \times N$ smoothing matrix which yields a coarse estimation $\widetilde{\mathbf{w}} = \widetilde{\mathbf{D}}\mathbf{y}$ of the transform-domain signal-dependent noise variance. A particular choice of such matrix will be given in Section 3.1.
- $\boldsymbol{\Theta}(\mathbf{w}, \widetilde{\mathbf{w}}) = [\theta_l(w_l, \widetilde{w}_l)]_{l\in[1;L]}$ is an arbitrary *pointwise* processing that will be specified in Section 3.1.

Corollary 1. *For a transform-domain pointwise processing as defined in* (4)*, PURE* (3) *can be expressed as:*

$$\tilde{\epsilon} \;=\; \frac{1}{N}\left(\|\mathbf{F}(\mathbf{y}) - \mathbf{y}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{y}\right) + \qquad (5)$$
$$\frac{2}{N}\left(\partial_1\boldsymbol{\Theta}(\mathbf{w}, \widetilde{\mathbf{w}})^{\mathrm{T}}(\mathbf{D} \bullet \mathbf{R}^{\mathrm{T}})\mathbf{y} + \partial_2\boldsymbol{\Theta}(\mathbf{w}, \widetilde{\mathbf{w}})^{\mathrm{T}}(\widetilde{\mathbf{D}} \bullet \mathbf{R}^{\mathrm{T}})\mathbf{y}\right)$$

where:

- $\partial_1\boldsymbol{\Theta}(\mathbf{w}, \widetilde{\mathbf{w}}) = \left[\frac{\partial\theta_l(w_l, \widetilde{w}_l)}{\partial w_l}\right]_{l\in[1;L]}$ *is the* $L \times 1$ *vector made of the first derivative, with respect to its first variable, of each function* θ_l.
- $\partial_2\boldsymbol{\Theta}(\mathbf{w}, \widetilde{\mathbf{w}}) = \left[\frac{\partial\theta_l(w_l, \widetilde{w}_l)}{\partial \widetilde{w}_l}\right]_{l\in[1;L]}$ *is the* $L \times 1$ *vector made of the first derivative, with respect to its second variable, of each function* θ_l.
- *"*$\bullet$*" denotes the Hadamard (element-by-element) product between two matrices.*

3. PURE-LET THRESHOLDING IN UNDECIMATED HAAR REPRESENTATION

In this section, we propose to use the practical measure of quality given in Corollary 1 to optimize a subband-adaptive pointwise thresholding applied to the coefficients of the undecimated Haar wavelet transform.

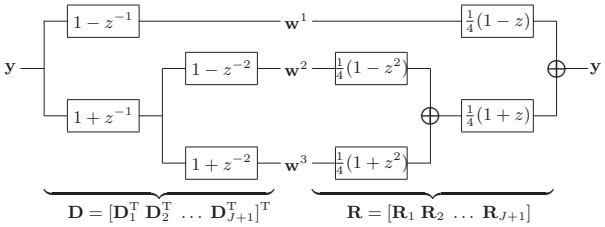

Fig. 1. The undecimated (unnormalized) Haar filterbank for $J = 2$ levels of decomposition.

3.1. LET Thresholding

Analyzing a vector $\mathbf{y}$ of independent Poisson variables through the (unnormalized) undecimated Haar filterbank depicted in Fig. 1, yields nothing but sums (scaling coefficients) and differences (wavelet coefficients) of independent Poisson random variables. The variance of each *wavelet* coefficient is thus the sum of the two underlying Poisson intensities, i.e. the corresponding noise-free *scaling* coefficient. Each noisy scaling coefficient can thus be used as a coarse estimate of the signal-dependent noise variance of the same-scale wavelet coefficient. In this case, $\widetilde{\mathbf{D}}$ simply implements each scale ($j = 1\dots J$) of the lowpass channel of the filterbank depicted in Fig. 1 and $\widetilde{\mathbf{w}} = \widetilde{\mathbf{D}}\mathbf{y}$ is the vector made of the resulting scaling coefficients.

Similarly to what has been recently proposed for fast denoising [4, 10, 13], we propose to build our subband-adaptive thresholding function $\theta_j(w, \widetilde{w})$ as a *linear expansion of thresholds* (LET):

$$\theta_j(w, \widetilde{w}) = a_{j,1}\underbrace{w}_{\theta_{j,1}(w,\widetilde{w})} + a_{j,2}\underbrace{w\exp\left(-\left(\frac{w}{t(\widetilde{w})}\right)^8\right)}_{\theta_{j,2}(w,\widetilde{w})} \qquad (6)$$

The above thresholding function is very similar to the one we introduced in [4] for AWGN reduction in redundant representation, except that it involves a *signal-dependent* threshold $t(\widetilde{w})$ that is tuned by the estimated noise variance $\widetilde{w}$. We experimentally found that $t(\widetilde{w}) = 3\sqrt{|\widetilde{w}|}$ yields, on average, the minimum MSE results.

3.2. LET Optimization by PURE Minimization

In the proposed undecimated Haar wavelet-domain LET framework, the whole denoising process can be finally expressed as:

$$\mathbf{F}(\mathbf{y}) \;=\; \sum_{j=1}^{J}\sum_{k=1}^{2} a_{j,k}\, \underbrace{\mathbf{R}_j \mathbf{\Theta}_{j,k}(\mathbf{D}_j\mathbf{y}, \widetilde{\mathbf{D}}_j\mathbf{y})}_{\mathbf{F}_{j,k}(\mathbf{y})} + \underbrace{\mathbf{R}_{J+1}\mathbf{D}_{J+1}\mathbf{y}}_{\text{lowpass}}$$

The parameters $a_{j,k}$ that minimize the approximate PURE given in Equ. (3) and (5) are then the solution of the *linear* system of $2J$ equations $\mathbf{M}\mathbf{a} = \tilde{\mathbf{c}}$, where for $\mathbf{k}, \mathbf{l} \in [1; J] \times [1; 2]$,

$$\begin{cases} \mathbf{M} = \Big[\mathbf{F_k}(\mathbf{y})^{\mathrm{T}}\mathbf{F_l}(\mathbf{y})\Big]_{2(k_1-1)+k_2, 2(l_1-1)+l_2} \\ \tilde{\mathbf{c}} \;=\; \Big[\mathbf{F_k}(\mathbf{y})^{\mathrm{T}}(\mathbf{Id} - \mathbf{R}_{J+1}\mathbf{D}_{J+1})\mathbf{y} - \partial\mathbf{F_k}(\mathbf{y})^{\mathrm{T}}\mathbf{y}\Big]_{2(k_1-1)+k_2} \end{cases}$$

The first order Taylor series approximation of some nonlinear functions $\mathbf{F}^{-}_{j,2}(\mathbf{y})$ might be inaccurate for very low-intensity signals. As the intensities of the Poisson processes linearly increase with the scale j, we propose to only keep those nonlinearly processed subband $\mathbf{F}_{j,2}(\mathbf{y})$ for which $2^j E_{\mathrm{mean}}$ is above a given threshold T. The averaged signal energy E_{mean} is unbiasedly estimated as: $E_{\mathrm{mean}} = \frac{1}{N}(\|\mathbf{y}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{y})$. We empirically found that any values of $T \in [5, 15]$ ensure that the PURE-based parameters optimization gives near minimum MSE performances (see Table 1).

4. EXPERIMENTS

We now evaluate the performance of the proposed PURE-LET undecimated Haar wavelet thresholding for denoising Poisson-corrupted images. We have chosen $J = 5$ levels of decomposition for the undecimated Haar wavelet transform and $T = 10$ for the threshold of reliability discussed in Section 3.2.

4.1. On Simulated Data

We have compared our algorithm with three state-of-the-art methods in simulated experiments: the Anscombe VST [1] followed by the *BLS-GSM* denoiser applied in a full steerable pyramid [3], the *Platelet* approach exposed in [8] and the Poisson-Haar hidden Markov tree (*PH-HMT*) introduced in [7]. The near shift-invariance of these last two algorithms is achieved by averaging the denoising results obtained on several shifted versions of the input image (cycle-spinning strategy).

The performance of the various methods has been quantified in term of peak signal-to-noise ratio (PSNR), computed as: $\mathrm{PSNR} = 10\log_{10}\left(\frac{I^2_{\max}}{\mathrm{MSE}}\right)$. Various input noise levels have been obtained by rescaling the original images between peak intensities $I_{\max} = 20$ and $I_{\max} = 1$.

As observed in Table 1, the VST-based method is usually outperformed by the other algorithms at low intensities. The proposed solution gives sometimes similar and often higher PSNRs than the state-of-the-art approaches specifically designed for Poisson intensity estimation. Note that better results could be obtained with a more sophisticated (e.g. multivariate) thresholding, but this extension is out of the scope of the present paper. As expected, our PURE-based parameters optimization consistently remains within 0.2 dB from the highest PSNR performances achievable by the pointwise thresholding proposed in (6). This demonstrates the reliability of PURE as an unbiased estimate of the actual MSE.

A visual comparison between the *Platelet* algorithm and the proposed approach is presented in Fig. 2 for the *Fingerprint* image. As observed, our denoised image exhibits no oversmoothing.

From a practical point of view, it must be stressed that the proposed algorithm requires only $\sim$ 1s to denoise a 256×256 image. To compare with, VST+*BLS-GSM* lasts around 8s on the same workstation, whereas the execution of 20 cycle-spins of *Platelet* requires nearly 1300s and the reported computation time of *PH-HMT* is 92s.

Table 1. Comparison of multiscale Poisson denoising algorithms.

Peak Intensities	20	10	5	3	2	1
Image	***Cameraman* 256 × 256**					
Input PSNR	16.29	13.28	10.27	8.07	6.29	3.28
Anscombe [1] + BLS-GSM [3]	26.56	24.63	22.50	20.75	19.07	14.44
Platelet [8]	**26.80**	**25.14**	**23.56**	**22.56**	**21.72**	**20.57**
PH-HMT [7]	26.61	24.97	23.37	22.31	21.41	20.03
PURE-LET	**26.72**	**25.10**	**23.50**	22.39	**21.67**	**20.48**
MSE-LET	*26.76*	*25.17*	*23.63*	*22.57*	*21.81*	*20.61*
Image	***Boat* 512 × 512**					
Input PSNR	15.95	12.94	9.92	7.70	5.94	2.93
Anscombe [1] + BLS-GSM [3]	**27.39**	**25.95**	**24.32**	22.71	20.92	15.57
Platelet [8]	26.92	25.50	**24.33**	23.39	**22.88**	21.75
PH-HMT [7]	26.96	25.57	**24.31**	**23.45**	22.77	21.76
PURE-LET	27.23	25.81	**24.39**	**23.53**	**22.88**	**21.92**
MSE-LET	*27.25*	*25.87*	*24.54*	*23.66*	*23.01*	*21.98*
Image	***Fingerprint* 512 × 512**					
Input PSNR	15.57	12.56	9.54	7.32	5.57	2.56
Anscombe [1] + BLS-GSM [3]	**24.94**	**23.36**	**21.89**	**20.55**	18.88	13.96
Platelet [8]	23.72	21.84	20.01	18.65	17.83	16.52
PH-HMT [7]	23.46	21.91	20.42	19.36	18.55	17.39
PURE-LET	**24.84**	23.09	21.38	20.19	**19.31**	**18.05**
MSE-LET	*24.85*	*23.12*	*21.45*	*20.27*	*19.37*	*18.08*

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

4.2. On Real Fluorescence Microscopy Data

We finally show the potential of the proposed approach for denoising low-intensities fluorescence microscopy data. The noisy image shown in Fig. 3 A has been acquired on a confocal microscope at the Imaging Center of the IGBMC, Illkirch, France[1]. It shows the cellular division of a *C. elegans* embryo labeled with three fluorescent dyes. Each fluorescence channel has been denoised independently using the proposed algorithm and the denoised image is shown in Fig. 3 B. As observed, the Poisson noise has been effectively removed, revealing some fine structures.

5. CONCLUSION

We have devised a novel undecimated Haar thresholding for robust and fast estimation of Poisson intensities. The use of a purely data-driven unbiased estimate of the MSE (PURE) allows the accurate monitoring of the denoising quality, while the linear parametrization of the denoising process (LET) leads to a straightforward parameters optimization. The proposed solution turns out to be qualitatively and computationally very competitive with some state-of-the-art multiscale methods and its application to real low-intensity fluorescence microscopy images yields promising denoising results.

[1]We are grateful to the IGBMC's team for providing the equipment and the assistance for acquiring the data and to Prof. Pierre Gönczy from the ISREC at EPFL for providing the biological sample.

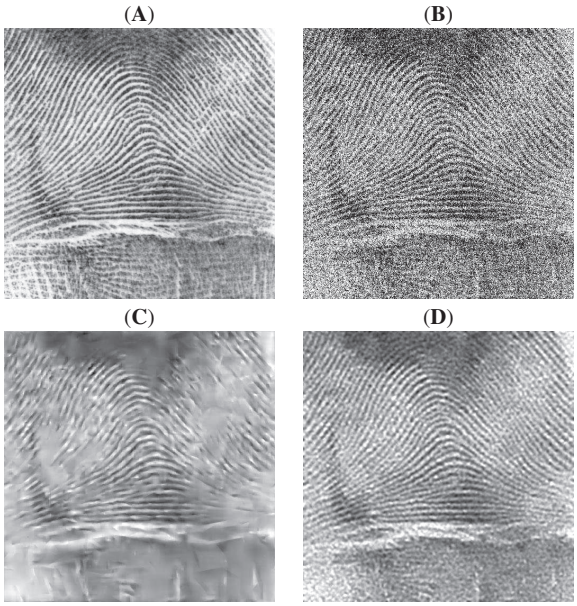

Fig. 2. (**A**) Original *Fingerprint* image. (**B**) Noisy version of it: PSNR $=$ 6.52 dB. (**C**) Denoised with *Platelet* [8]: PSNR $=$ 18.34 dB. (**D**) Denoised with the proposed algorithm: PSNR $=$ 19.79 dB.

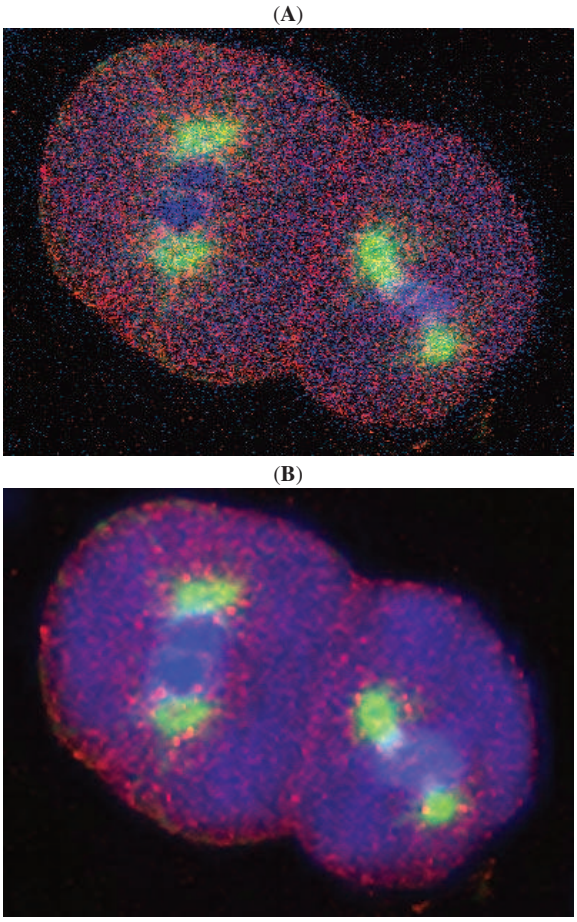

Fig. 3. (**A**) Observed data (three-color image). (**B**) Denoised with the proposed algorithm.

6. REFERENCES

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

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

[3] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image Denoising using Scale Mixtures of Gaussians in the Wavelet Domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, November 2003.

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

[5] K. E. Timmermann and R. D. Nowak, "Multiscale modeling and estimation of poisson processes with application to photon-limited imaging," *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 846–862, April 1999.

[6] E. D. Kolaczyk, "Bayesian multi-scale models for Poisson processes," *Journal of the American Statistical Association*, vol. 94, no. 447, pp. 920–933, September 1999.

[7] S. Lefkimmiatis, P. Maragos, and G. Papandreou, "Bayesian inference on multiscale models for poisson intensity estimation: Applications to photon-limited image denoising," *Image Processing, IEEE Transactions on*, vol. 18, no. 8, pp. 1724–1741, Aug. 2009.

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

[9] K. Hirakawa and P. Wolfe, "Skellamshrink: Poisson intensity estimation for vector-valued data," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, April 2009, pp. 3441–3444.

[10] F. Luiser, C. Vonesch, T. Blu, and M. Unser, "Fast interscale wavelet denoising of poisson-corrupted images," *Signal Processing*, vol. 90, no. 2, pp. 415–427, February 2010.

[11] M. Jansen, "Multiscale poisson data smoothing," *Journal of the Royal Statistical Society B*, vol. 68, no. 1, pp. 27–48, 2006.

[12] B. Zhang, J. M. Fadili, and J.-L. Starck, "Wavelets, ridgelets, and curvelets for poisson noise removal," *Image Processing, IEEE Transactions on*, vol. 17, no. 7, pp. 1093–1108, July 2008.

[13] M. Raphan and E. P. Simoncelli, "Optimal denoising in redundant representations," *Image Processing, IEEE Transactions on*, vol. 17, no. 8, pp. 1342–1352, Aug. 2008.

[14] H. M. Hudson, "A natural identity for exponential families with applications in multiparameter estimation," *The Annals of Statistics*, vol. 6, no. 3, pp. 473–484, May 1978.