OPTIMAL PARAMETER SELECTION FOR BILATERAL FILTERS USING POISSON UNBIASED RISK ESTIMATE

Harini Kishan and Chandra Sekhar Seelamantula

Department of Electrical Engineering
Indian Institute of Science
Bangalore - 560 012, India
Emails: harini@ee.iisc.ernet.in, chandra.sekhar@ieee.org

ABSTRACT

Bilateral filters perform edge-preserving smoothing and are widely used for image denoising. The denoising performance is sensitive to the choice of the bilateral filter parameters. We propose an optimal parameter selection for bilateral filtering of images corrupted with Poisson noise. We employ the Poisson's Unbiased Risk Estimate (PURE), which is an unbiased estimate of the Mean Squared Error (MSE). It does not require a priori knowledge of the ground truth and is useful in practical scenarios where there is no access to the original image. Experimental results show that quality of denoising obtained with PURE-optimal bilateral filters is almost indistinguishable with that of the Oracle-MSE-optimal bilateral filters.

***Index Terms*—** Poisson noise, Denoising, PURE, Bilateral filter, Raised-cosine kernel

1. INTRODUCTION

The bilateral filter was proposed by Tomasi and Manduchi [1]. It has been used for various applications such as denoising [2], demosaicking [3], and optical-flow estimation. Spatial filtering does not preserve edges because it employs shift-invariant kernels. On the other hand, the bilateral filter employs a specific shift-variant kernel, which helps preserve edges. The bilateral filter ϕ is a combination of the domain filter and range filter, that is, $\phi_{k,m}(y_k, y_m) = w_{k-m}r(y_k - y_m)$, wherein the domain filter w_{k-m} is based on the geometric proximity between the pixel of interest at spatial coordinate k and a nearby pixel at coordinate m. The desirable properties of an appropriate domain kernel are symmetry, non-negativity, and localization. Pixels located closer to the pixel of interest are assigned higher weights than the farther ones, localizing the averaging operation to a neighbourhood $\mathcal{N}$ of k. The range filter $r(y_k - y_m)$ measures the photometric similarity between the pixel of interest y_k and its neighbour y_m. It is symmetric, non-negative, and assigns higher weights to pixels that are photometrically similar to the pixel of interest than the dissimilar ones. The general expression for the bilateral filter output corresponding to an image $\mathbf{y}$ is given by

$$f_k(\mathbf{y}, \theta) = \frac{\sum\limits_{m \in \mathcal{N}_k} w_{k-m} r(y_k - y_m) y_m}{\sum\limits_{m \in \mathcal{N}_k} w_{k-m} r(y_k - y_m)}, \tag{1}$$

where θ denotes the parameters of the bilateral filter.

The nonlinearity of filtering and lack of shift-invariance do not facilitate FFT-based implementations of (1). This problem was addressed by Porikli [4], who proposed a constant-time O(1) implementation of the bilateral filter (for arbitrary spatial kernels) using polynomial range kernels. Recently, Chaudhury et al. [5] proposed constant-time bilateral filtering using a raised-cosine range kernel.

Bilateral filtering is often employed as a preprocessing step for noise suppression in images. Digital image quality degradation can be attributed to signal-independent and signal-dependent noise sources. Imperfections in the optical and electronic devices of an image acquisition system are major sources of signal-independent noise. Thermal noise and noise in the electronic circuits of the acquisition system can be modeled as additive white Gaussian noise (AWGN). Under the AWGN assumption, a methodology to select the Gaussian bilateral filter parameters optimally was proposed by Peng and Rao [6]. Their methodology is based on the Stein's unbiased risk estimate (SURE). Van De Ville and Kocher developed a SURE approach for non-local means [7]. We developed SURE-based optimal parameter selection techniques for the raised-cosine-kernel-based fast bilateral filter in [8]. The randomness of photon emission is a source of signal-dependent noise. It becomes significant under low illumination conditions and degrades the image quality. In this paper, the noisy measurements are modeled as independent Poisson random variables.

One of the approaches to Poisson intensity estimation involves a three-step procedure: application of a variance stabilizing transform (VST) such as the Anscombe transform [9] to make the variance of the transformed data independent of its mean, processing the transformed data with a Gaussian-noise-optimal denoising function, and inversion of the VST. Although this approach can potentially take advantage of the vast literature on denoising of images in Gaussian noise conditions, VSTs fail to stabilize the noise variance under low illumination conditions when the photon counting noise is dominant.

Recently, Luisier et al. [10–12] introduced the Poisson unbiased risk estimator (PURE) for the MSE and employed it for wavelet-domain denoising of Poisson-noise images [10, 11]. To make an analogy, the PURE is to Poisson noise what the SURE is to Gaussian noise. In this paper, we propose to use PURE for optimal parameter selection of bilateral filters. We assume that the ground truth image is deterministic. The key advantage of PURE is that it is independent of the original image, making it suitable and relevant for practical applications when ground truth is not available.

The paper is organized as follows. We provide the problem statement in Section 2. The theory related to PURE is given in Section 3. We provide the PURE calculations for the standard bilateral filter, the Gaussian bilateral filter, and the raised-cosine-kernel-based fast bilateral filter in Section 4. The accuracy of PURE derived for the Gaussian and fast bilateral filters is validated in Section 5 by comparing the PURE-optimal denoising versus Oracle-MSE-optimal denoising. Concluding remarks are given in Section 6.

2. PROBLEM FORMULATION

We consider a deterministic digital image $\mathbf{x} = [x_1, x_2, \cdots, x_N]$ (vectorized notation where N denotes the total number of pixels in the image) corrupted by Poisson noise. The noisy observation be $\mathbf{y} = [y_1, y_2, \cdots, y_N]$. y_k follows the Poisson distribution: $P(y_k = n) = \frac{\exp(-x_k)x_k^n}{n!}$. The Poisson noise model assumed here is neither an additive nor a multiplicative one. It is based on the assumption that the intensity of pixels in the noisy image are i.i.d. Poisson random variables with mean as the intensity of the corresponding pixel in the original image. Given the noisy image, our aim is to compute the optimal parameter $\hat{\theta}$ of the bilateral filter such that the bilaterally filtered image $\mathbf{f}(\mathbf{y}, \hat{\theta})$ is the best estimate of the ground truth in the sense of minimizing the MSE.

The MSE between the original and denoised image is defined as $\text{MSE}(\mathbf{f}(\mathbf{y}, \theta)) = \mathcal{E}\{\|\mathbf{f}(\mathbf{y}, \theta) - \mathbf{x}\|^2\}$. Our goal is to select the bilateral filter domain and range kernel parameters such that the bilateral filter output $\mathbf{f}(\mathbf{y}, \theta) = [f_k(\mathbf{y}, \theta)]_{1 \le k \le N}$ is closest to the original image in the minimum MSE sense. In other words, we are interested in computing $\mathbf{f}(\mathbf{y}, \hat{\theta})$ such that $\hat{\theta} = \arg\min_{\theta} \text{MSE}(\mathbf{f}(\mathbf{y}, \theta))$. In practice, the Oracle MSE cannot be computed as we do not have access to the original image $\mathbf{x}$. PURE is an unbiased estimator of the MSE and can be computed without the knowledge of the ground truth. The practical alternative to minimizing the Oracle MSE is to obtain $\mathbf{f}(\mathbf{y}, \hat{\theta})$ such that $\hat{\theta} = \arg\min_{\theta} \text{PURE}(\mathbf{f}(\mathbf{y}, \theta))$.

3. THEORETICAL BACKGROUND FOR PURE

We recall the following theorem from [10–12], which is based on Hudson's identity [13].

Theorem: Let $\mathbf{f}(\mathbf{y}, \theta) = [f_k(\mathbf{y}, \theta)]$ be an N-dimensional real-valued vector function such that $\mathcal{E}\{|f_k(\mathbf{y}, \theta)|\} < \infty$, $k = 1, 2, 3, \cdots, N$. Define $\mathbf{f}^-(\mathbf{y}, \theta) = [f_k(\mathbf{y} - \mathbf{e}_k, \theta)]_{1 \le k \le N}$, where $\mathbf{e}_k = [\delta_{k-n}]_{1 \le n \le N}$ denotes each vector of the canonical basis of $\mathbb{R}^N$. Then, the random variable

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

is an unbiased estimate of MSE and is referred to as PURE, that is, $\mathcal{E}\{\text{PURE}\} = \mathcal{E}\{\|\mathbf{f}(\mathbf{y}, \theta) - \mathbf{x}\|^2\}$.

Computing PURE involves applying the whole denoising process to a slightly perturbed noisy input to compute a single component $f_k(\mathbf{y} - \mathbf{e}_k, \theta)$ of $\mathbf{f}^-(\mathbf{y}, \theta)$. This process has to be repeated N times to obtain the full vector $\mathbf{f}^-(\mathbf{y}, \theta)$. Thus, computing PURE directly as stated in (2) is tedious. Luisier et al [12] assume that f has continuous first-order partial derivatives $\left(f_k \in \mathcal{C}^1(\mathbb{R}^N), \forall k\right)$, and use the first-order Taylor series approximation of $\mathbf{f}^-(\mathbf{y}, \theta)$:

$$\mathbf{f}^-(\mathbf{y}, \theta) \simeq \mathbf{f}(\mathbf{y}, \theta) - \partial\mathbf{f}(\mathbf{y}, \theta) \text{ ,where } \partial\mathbf{f}(\mathbf{y}, \theta) = \left[\frac{\partial f_k(\mathbf{y}, \theta)}{\partial y_k}\right]_{1 \le k \le N},$$

which is valid if $|y_k| \gg 1$. Then, PURE is well approximated as

$$\text{PURE}(\mathbf{f}(\mathbf{y}, \theta)) = \frac{1}{N}\left(\|\mathbf{f}(\mathbf{y}, \theta) - \mathbf{y}\|^2 + 2\mathbf{y}^T\partial\mathbf{f}(\mathbf{y}, \theta) - 1^T\mathbf{y}\right). \tag{3}$$

4. PURE-OPTIMAL BILATERAL FILTERS

4.1. PURE for the general bilateral filter

The derivative of the bilateral filter output with respect to the noisy image is given by

$$\frac{\partial f_k(\mathbf{y}, \theta)}{\partial y_k} = \frac{1}{W_k}\Bigg(\sum_{m \in \mathcal{N}_k} \frac{\partial \phi_{k,m}(y_k, y_m)}{\partial y_k} y_m + \phi_{k,k}(y_k, y_k) - f_k(\mathbf{y}, \theta)\sum_{m \in \mathcal{N}_k} \frac{\partial \phi_{k,m}(y_k, y_m)}{\partial y_k}\Bigg), \tag{4}$$

where $W_k = \sum_{m \in \mathcal{N}_k} \phi_{k,m}(y_k, y_m)$ and $\phi_{k,k}(y_k, y_k) = 1$. The corresponding PURE is computed by substituting (1) and (4) in (3):

$$\text{PURE}(\mathbf{f}(\mathbf{y}, \theta)) = \frac{\|\mathbf{f}(\mathbf{y}, \theta) - \mathbf{y}\|^2 - 1^T\mathbf{y}}{N} + \frac{2}{N}\sum_{k=1}^{N}\frac{y_k}{W_k} \left(\sum_{m \in \mathcal{N}_k} \frac{\partial \phi_{k,m}(y_k, y_m)}{\partial y_k} y_m + 1 - f_k(\mathbf{y}, \theta)\sum_{m \in \mathcal{N}_k} \frac{\partial \phi_{k,m}(y_k, y_m)}{\partial y_k}\right), \tag{5}$$

where $\|\mathbf{f}(\mathbf{y}, \theta) - \mathbf{y}\|^2$ denotes the square error between the bilateral filter output and the noisy image. The optimum parameters of the bilateral filter are computed as $\hat{\theta} = \arg\min_{\theta} \text{PURE}(\mathbf{f}(\mathbf{y}, \theta))$.

4.2. PURE for the Gaussian bilateral filter

The classical Gaussian bilateral filter consists of Gaussian range and domain kernels and is given by

$$\phi_{k,m}(y_k, y_m) = \exp\left(\frac{-\|k - m\|^2}{2\sigma_d^2}\right)\exp\left(\frac{-|y_k - y_m|^2}{2\sigma_r^2}\right). \tag{6}$$

The parameters of the Gaussian bilateral filter are $\theta = \{\sigma_d, \sigma_r\}$.

Taking the derivative of (6) with respect to y_k, we get that

$$\frac{\partial \phi_{k,m}(y_k, y_m)}{\partial y_k} = \phi_{k,m}(y_k, y_m)\Big(\frac{y_m - y_k}{\sigma_r^2}\Big). \tag{7}$$

On substituting (7) in (5), we get that

$$\text{PURE}(\mathbf{f}(\mathbf{y}, \sigma_d, \sigma_r)) = \frac{\|\mathbf{f}(\mathbf{y}, \sigma_d, \sigma_r) - \mathbf{y}\|^2 - 1^T\mathbf{y}}{N} + \frac{2}{N}\sum_{k=1}^{N}\frac{y_k}{W_k} \left(1 + \sum_{m \in \mathcal{N}_k}\phi_{k,m}(y_k, y_m)\Big(\frac{y_m - y_k}{\sigma_r^2}\Big)(y_m - f_k(\mathbf{y}, \sigma_d, \sigma_r))\right). \tag{8}$$

The optimum parameters of the Gaussian bilateral filter are computed as $\hat{\sigma}_d, \hat{\sigma}_r = \arg\min_{\sigma_d, \sigma_r} \text{PURE}(\mathbf{f}(\mathbf{y}, \sigma_d, \sigma_r))$. Direct implementation of the Gaussian bilateral filter and computation of the differential term in (7) are both computationally intensive, which makes it tedious to compute PURE in (8) and to optimize for the Gaussian bilateral filter parameters.

4.3. PURE for the fast bilateral filter

The raised-cosine (RC) kernel closely approximates the Gaussian function [5]. It is given by $r(y_k - y_m) = \left[\cos\left(\frac{\gamma}{\rho\sqrt{L}}(y_k - y_m)\right)\right]^L$. The RC kernel is non-negative when its argument takes values between $-\pi/2$ and $\pi/2$. Normalizing the argument by $\gamma = \pi/2T$ ensures non-negativity of the kernel. $[0, T]$ is the dynamic range of the image $\mathbf{y}$. The parameters L and ρ of the RC kernel depend on the standard deviation σ_r of the Gaussian to be approximated. ρ is given by $\rho = \gamma\sigma_r$. If $\sigma_r > \gamma^{-2}$, a large value is chosen for L, otherwise $L = \rho^{-2}$. Writing $\cos\theta = (e^{\mathrm{j}\theta} + e^{-\mathrm{j}\theta})/2$ and applying the binomial theorem, the RC kernel is expressed as

$$r(y_k - y_m) = \sum_{l=0}^{L} 2^{-L}\binom{L}{l}\exp\left(\frac{\mathrm{j}\gamma}{\rho\sqrt{L}}(2l - L)(y_k - y_m)\right). \tag{9}$$

The domain filter used in [5] is

$$w_{k-m} = \exp\left(\frac{-\|k - m\|^2}{2\sigma_d^2}\right). \tag{10}$$

The RC-kernel-based bilateral filter is parameterized by σ_d of its domain kernel and σ_r of the Gaussian range kernel it approximates, that is, $\theta = \{\sigma_d, \sigma_r\}$. The bilateral filter output is obtained by substituting (9) and (10) in (1), that is,

$$\begin{aligned}
f_k(\mathbf{y}, \sigma_d, \sigma_r) &= \frac{\sum_{l=0}^{L} d_k(l)\bar{g}_k(l)}{\sum_{l=0}^{L} d_k(l)\bar{h}_k(l)}, & (11)\\
\text{where } d_k(l) &= 2^{-L}\binom{L}{l}\exp\left(\frac{\mathrm{j}\gamma}{\rho\sqrt{L}}(2l - L)y_k\right), & (12)\\
h_k(l) &= \exp\left(-\frac{\mathrm{j}\gamma}{\rho\sqrt{L}}(2l - L)y_k\right), \text{ and} & (13)\\
g_k(l) &= h_k(l)y_k.
\end{aligned}$$

In (11), $\bar{g}_k(l)$ and $\bar{h}_k(l)$ are obtained by filtering $g_k(l)$ and $h_k(l)$, respectively with a Gaussian kernel of mean zero and variance σ_d^2 using FFT-based implementation. From the bilateral filter definition, (9), and (10), the derivative of ϕ with respect to y is computed as

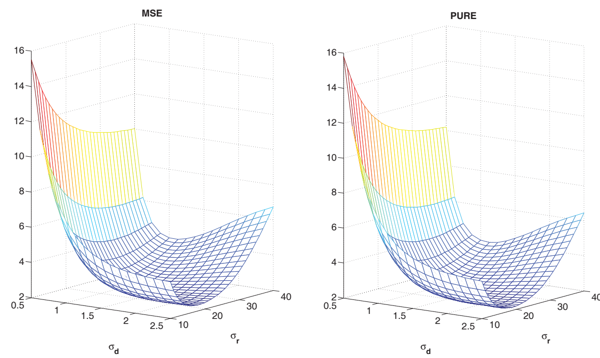
$$\frac{\partial\phi_{k,m}(y_k, y_m)}{\partial y_k} = w_{k-m}\sum_{l=0}^{L} c_k(l)\exp\left(-\frac{\mathrm{j}\gamma}{\rho\sqrt{L}}(2l - L)y_m\right), \tag{14}$$

where $c_k(l) = \frac{\mathrm{j}\gamma}{\rho\sqrt{L}}(2l - L)d_k(l)$. Substituting (14) in (5), we obtain the expression for PURE as

$$\begin{aligned}
\mathrm{PURE}(\mathbf{f}(\mathbf{y}, \sigma_d, \sigma_r)) =& \frac{\|\mathbf{f}(\mathbf{y}, \sigma_d, \sigma_r) - \mathbf{y}\|^2 - 1^T\mathbf{y}}{N}\\
&+ \frac{2}{N}\sum_{k=1}^{N} y_k\frac{\partial f_k(\mathbf{y}, \sigma_d, \sigma_r)}{\partial y_k}, \text{ where} \qquad (15)
\end{aligned}$$

$$\frac{\partial f_k(\mathbf{y}, \sigma_d, \sigma_r)}{\partial y_k} = \frac{\sum_{l=0}^{L} c_k(l)\bar{g}_k(l) + 1 - f_k(\mathbf{y}, \sigma_d, \sigma_r)\sum_{l=0}^{L} c_k(l)\bar{h}_k(l)}{\sum_{l=0}^{L} d_k(l)\bar{h}_k(l)}. \tag{16}$$

The optimal parameters of the fast bilateral filter are computed as $\hat{\sigma}_d, \hat{\sigma}_r = \arg\min_{\sigma_d, \sigma_r} \mathrm{PURE}(\mathbf{f}(\mathbf{y}, \sigma_d, \sigma_r))$. In [5], it has been shown that (11) can be implemented in constant time. Computing the bilateral filter output in (11) and the differential term in (16) using FFT-based implementations accelerates the computation of PURE for different parameters of the filter. Parallel computation of the $2(L + 1)$ different filtered outputs in (11) and (16) can further accelerate the speed of computation of PURE. Therefore, the optimal parameter computation for the fast bilateral filter is not as computationally intensive as that of the Gaussian bilateral filter.



(a) (b)

Fig. 1: Comparison of (a) MSE and (b) PURE for the fast bilateral filter.

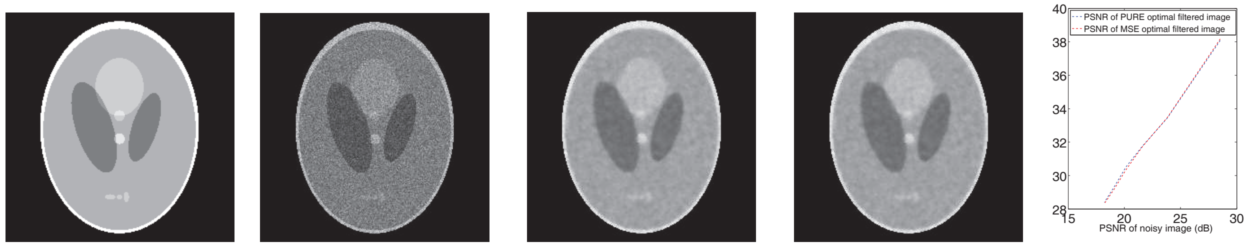
5. EXPERIMENTS

We next validate the accuracy of the PURE derivations for the Gaussian and RC-kernel-based fast bilateral filters.

A 22.66 dB noisy realization of Shepp-Logan phantom was denoised with a RC-kernel-based fast bilateral filter for different parameter settings. σ_d was varied from 0.5 to 2.5 in steps of 0.125 and σ_r was varied from 10 to 40 in increments of 1. The results are shown in Figure 1. We observe that PURE approximates the MSE accurately. The cost functions seem to be well behaved to enable parameter search by optimization techniques. We have used gradient-descent optimization technique to compute optimal parameters. We found the PURE-optimal parameters to be $\hat{\sigma}_d = 1.8, \hat{\sigma}_r = 20$, whereas the Oracle-MSE-optimal parameters turned out to be $\hat{\sigma}_d = 1.8, \hat{\sigma}_r = 19$. As the MSE and PURE-optimal parameters and PSNR values of the MSE and PURE-optimally denoised images are close enough, we infer that the MSE-optimal and PURE-optimal RC-kernel-based fast bilateral filters offer similar performance and the quality of denoising is good. From Figure 2(e), we observe that the MSE- and PURE-optimal values for different noisy realizations match well. The experimental results for 256×256 Pepper image using Gaussian bilateral filter are shown in Figure 2. A fingerprint image of size 256×256 was denoised using MSE- and PURE-optimal RC-kernel-based fast bilateral filters. The results are shown in Figure 3. The PSNR values of the MSE and PURE-optimally filtered images are identical.

6. CONCLUSION

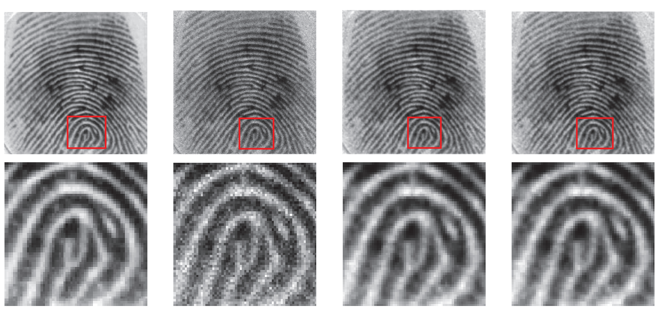
We addressed the problem of choosing optimal bilateral filter parameters to denoise Poisson-noise corrupted images. We proposed a technique based on minimizing the Poisson unbiased risk estimator of the MSE. We derived PURE expressions for the Gaus-



(a) Ground truth Shepp-Logan phantom | (b) Noisy image PSNR = 22.66 dB | (c) Oracle MSE; fast BF PSNR = 32.58 dB | (d) PURE; fast BF PSNR = 32.58 dB | (e) Output PSNR (dB) vs. input PSNR (dB)

(a) Ground truth Pepper | (b) Noisy image PSNR = 24.18 dB | (c) Oracle MSE; Gaussian BF PSNR = 30.74 dB | (d) PURE; Gaussian BF PSNR = 30.74 dB | (e) Output PSNR (dB) vs. input PSNR (dB)

Fig. 2: Comparison of denoising performance using Oracle-MSE- and PURE-optimal fast/Gaussian bilateral filters (BF).



(a) (b) (c) (d)

Fig. 3: (a) Original image, (b) Noisy image (PSNR=25.84 dB), (c) Denoised image obtained with MSE-optimal RC-kernel-based fast bilateral filter (PSNR=31.94 dB), (d) Denoised image obtained with PURE-optimal RC-kernel-based fast bilateral filter (PSNR=31.94 dB). The bottom row corresponds to the zoomed portions of the parts highlighted in the images in the top row.

sian and raised-cosine-based fast bilateral filters. Experimental results showed that the quality of denoising obtained with the PURE-optimal denoising methodology is almost indistinguishable from the Oracle-MSE-optimal denoising.

7. REFERENCES

[1] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," *IEEE Intl. Conf. Comp. Vis.*, pp. 839–846, 1998.

[2] E. P. Bennett, J. L. Mason, and L. McMillan, "Multispectral bilateral video fusion," *IEEE Trans. Image Process.*, vol. 16, pp. 1185–1194, 2007.

[3] R. Ramanath and W. E. Snyder, "Adaptive demosaicking," *Jl. of El. Imaging*, vol. 12, pp. 633–642, 2003.

[4] F. M. Porikli, "Constant-time O(1) bilateral filtering," *IEEE Conf. on Comp. Vis. and Patt. Recog.*, pp. 1–8, 2008.

[5] K. N. Chaudhury, D. Sage, and M. Unser, "Fast O(1) bilateral filtering using trigonometric range kernels," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3376–3382, 2011.

[6] H. Peng and R. Rao, "Bilateral kernel parameter optimization by risk minimization," *Proc. IEEE Intl. Conf. Image Proc.*, pp. 3293–3296, 2010.

[7] D. Van De Ville and M. Kocher, "SURE-based non-local means," *IEEE Signal Process. Lett.*, vol. 16, no. 11, pp. 973–976, Nov. 2009.

[8] H. Kishan and C. S. Seelamantula, "SURE-fast bilateral filters," *Proc. IEEE Intl. Conf. Acoust., Speech, Sig. Proc.*, pp. 1129–1132, 2012.

[9] F. J. Anscombe, "The transformation of Poisson, Binomial and negative-Binomial data," *Biometrika*, vol. 35, pp. 246–254, 1948.

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

[11] F. Luisier, T. Blu, and M. Unser, "Image denoising in mixed Poisson-Gaussian noise," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 696–708, March 2011.

[12] F. Luisier, T. Blu, and M. Unser, "Undecimated Haar thresholding for Poisson intensity estimation," *IEEE Intl. Conf. Image Proc.*, pp. 1697–1700, 2010.

[13] H. M. Hudson, "A natural identity for exponential families with applications in multiparameter estimation," *The Annals of Stat.*, vol. 6, pp. 473–484, 1978.