Approximate Bayesian Computation, stochastic algorithms and non-local means for complex noise models

Charles Kervrann, Philippe Roudot, François Waharte

APPROXIMATE BAYESIAN COMPUTATION, STOCHASTIC ALGORITHMS AND NON-LOCAL MEANS FOR COMPLEX NOISE MODELS

Charles Kervrann, Philippe Roudot

Inria Rennes - Bretagne Atlantique, Serpico Team
Campus universitaire de Beaulieu
35 042 Rennes Cedex, France

Francois Waharte

UMR 144, PICT IBiSA, CNRS-Institut Curie
12, rue Lhomond
75 248 Paris Cedex 05, France

ABSTRACT

In this paper, we present a stochastic NL-means-based denoising algorithm for generalized non-parametric noise models. First, we provide a statistical interpretation to current patch-based neighborhood filters and justify the Bayesian inference that needs to explicitly accounts for discrepancies between the model and the data. Furthermore, we investigate the Approximate Bayesian Computation (ABC) rejection method combined with density learning techniques for handling situations where the posterior is intractable or too prohibitive to calculate. We demonstrate our stochastic Gamma NL-means (SGNL) on real images corrupted by non-Gaussian noise.

Index Terms— Denoising, Approximate Bayesian Computation, density learning, NL-means, fluorescence imaging.

1. INTRODUCTION

A digital imaging system consists of an optical system followed by a photodetector and associated electrical filters. The photodetector converts the incident optical intensity (i.e photons) to electrons. During the process, the true signals are damaged by many different sources of noise. For instance, in optical microscopy imaging, signals are known to be corrupted by intensity dependent Poisson noise but also by additional sources of electron noise [1, 2]. More generally, we have to deal with images corrupted with heterogeneous mixed noises, which requires to adapt the usual denoising approaches including the non local(NL)-means algorithm [3].

In the case of images damaged by additive white Gaussian noise, the NL-means [3] has been shown to be efficient to reduce additive noise. This method exploits image redundancy captured by patches to restore information. In order to optimally perform in the case of Poisson noise, the NL-means has been combined with variance stabilizing transforms such as Anscombe [1] and Fisz transform [4]. Other authors proposed to combine Principal Component Analysis [5] and dictionary [6] to patch-based representation to reduce Poisson noise. In [7], the authors proposed an extension of the NL-means based on probabilistic similarities to compare noisy patches and pre-estimated patches; this framework has been used to denoise SAR images [8] and is appropriate for general parametric noise models [9].

In this paper, a new Bayesian motivation for the NL-means is given for generalized non-parametric noise models. Our approach inspired from the Approximate Bayesian Computation (ABC) framework [10, 11], is especially adapted to complex situations where the posterior cannot be easily derived or computed. Furthermore, we exploit empirical noise statistics and propose a distance learning framework to adapt to different conditions. This is particularly relevant for images contaminated by heterogeneous sources of noise [1, 2]. A major difference with previous methods [8, 7] is that we directly handle the structure of the noise, without precise parametric modeling of the noise.

The organization of the paper is as follows: in Section 2, we describe the NL-means in the Bayesian framework [12]. In Section 3 we describe the ABC method which serves to compute the data-driven posterior distribution. In Section 4 we demonstrate the flexibility of our stochastic Gamma NL-means (SGNL-means) by showing how it can be adapted to tackle the noise in frequency domain fluorescence lifetime imaging microscopy (FD-FLIM) and cryo-electron tomography. We illustrate the potential of our approach on two examples of ultra-sound and SAR imaging.

2. BAYESIAN INTERPRETATION OF NL-MEANS

Consider a gray-scale image $v = (v(x))_{x \in \Omega}$ defined over a bounded domain $\Omega \subset \mathbb{R}^2$ and $v(x) \in \mathbb{R}_+$ is the noisy observed intensity at pixel $x \in \Omega$. Assume that the image v is a noisy version of an unknown image u, that is $v = u + noise$. Define a $\sqrt{n} \times \sqrt{n}$ observed patch $S_v(x)$ at pixel x as: $S_v(x)(t) \overset{\triangle}{=} v(x+t)$, $\forall t \in [-\frac{\sqrt{n}-1}{2}, \cdots, \frac{\sqrt{n}-1}{2}]^2$. The NL-means at pixel x is a weighted average of all gray values in the entire image. Formally, we have [3]:

$$u_{NL}(x) = \frac{1}{Z(x)} \sum_{y \in \Omega} \exp\left(-\frac{\|S_v(x) - S_v(y)\|_2^2}{h^2}\right) v(y) \quad (1)$$

and $Z(x)$ is a normalizing factor. The weights express the amount of similarity ($\| \cdot \|_2$ denotes the Euclidean distance)

between the n-dimensional image patches $S_v(x)$ and $S_v(y)$ of each pair of pixels x and y involved in the computation. For the sake of simplicity, we omitted in (1) the choice of a weighted Euclidean norm over the patches as described in [3]. The decay parameter $h^2 \approx 10\sigma^2$ acts as a filtering parameter and σ^2 denotes the variance of the noise assumed to be Gaussian. The range of the search space can be as large as the whole image but, in practice, it is necessary to restrict the computation of weights to 21×21 pixel neighborhoods.

In the line of work of [3, 13], we describe a more recent interpretation of NL-means in the Bayesian setting [12]. Define a prior on patches $z \in \mathbb{R}^n$ from the noisy image v. A simple histogram is given by $p(z) = \frac{1}{|\Omega|} \sum_{y \in \Omega} \mathbf{1}[z = S_v(y)]$ where $\mathbf{1}[\cdot]$ denotes the indicator function. The Bayesian estimate $\widehat{S_u}(x)$ of a patch $S_u(x)$ with L_2-risk and prior p on patches is given by the posterior expectation $\widehat{S_u}(x) = \mathbb{E}_p[S_U(x)|S_v(x)]$, i.e.

$$\widehat{S_u}(x) = \int \overbrace{p(z|S_v(x))}^{posterior} z\, dz = \int \overbrace{p(z)}^{prior} \overbrace{p(S_v(x)|z)}^{likelihood} z\, dz.$$

In the case of zero-mean white Gaussian noise (variance σ^2),

$$\widehat{S_u}(x) = \frac{1}{Z(x)} \int p(z) e^{-\frac{\|S_v(x)-z\|_2^2}{2\sigma^2}} z\, dz. \quad (2)$$

Given $p(z)$ as defined above and switching the sums, we get

$$\widehat{S_u}(x) = \frac{1}{Z'(x)} \int \sum_{z \in \mathbb{R}^n} \mathbf{1}[z = S_v(y)]\, e^{-\frac{\|S_v(x)-z\|_2^2}{2\sigma^2}} z\, dz. \quad (3)$$

For a given y, a patch z yielding a non-zero term can only be $z = S_v(y)$. Finally, by selecting the central pixel [14], we get $u_{NL}(x)$ (see (1)). NL-means is therefore a posterior expectation and the prior model is based on the empirical histogram of patches taken in the input noisy image. In the next section, we consider more general likelihood models and priors.

3. APPROXIMATE BAYESIAN COMPUTATION NL-MEANS FILTERING

3.1. Principles of ABC rejection method

One of the basic problem in Bayesian statistics is the computation of the posterior for general forms of noise distributions. If the posterior density cannot be computed explicitly or is time consuming, we usually resort to stochastic simulation to generate samples for the posterior. The commonly-used approach is the rejection method but, more recently, Beaumont et al. [10] described a generalization of the usual rejection method in the domain of genetics.

Formally, assume data $\mathcal{D}$ generated from a model determined by λ whose prior is denoted $p(\lambda)$. The so-called ABC method is as follows [10, 11]:

1. Generate λ from $p(\cdot)$;
2. Simulate $\mathcal{D}'$ from the model with parameter λ;
3. Calculate a distance $\rho(\mathcal{D}, \mathcal{D}')$ between $\mathcal{D}'$ and $\mathcal{D}$, accept λ if $\rho(\mathcal{D}, \mathcal{D}') \leq \delta$, and return to 1.

As $\delta \to \infty$, accepted observations come from the prior. When $\delta \to 0$, this rejection algorithm is exact and accepted observations are independent and identically distributed from the posterior distribution $p(\lambda|\mathcal{D})$. Nevertheless, most samples are rejected if we set $\delta = 0$. Then, this approach requires the setting of δ and the selection of a metric $\rho(\cdot, \cdot)$ (e.g. L_2 distance).

The next step is to calculate expectations of the form $\mathbb{E}(\lambda|D) = \int p(\lambda|D)\, \lambda\, d\lambda$ where the expectation is taken with respect to the posterior distribution of λ. The simplest way to approach this is to draw samples $\{\lambda_i\}_{i=1,\cdots,N}$, from $p(\lambda|D)$ using the previous algorithm and then approximate using the sum $N^{-1} \sum \lambda_i$. However, a more stable estimate can be obtained by weighting the λ values with the posterior. Consequently, all values of λ are included in the the sum and there is no rejection step. This is a direct extension of the estimate given in Beaumont et al. [11] which used Epanechnikov kernels to weight each value of λ.

3.2. Patch-based ABC method

The interpretation of ABC given above allows us to revisit previous analyzes of the NL-means in the Bayesian setting. The objective is to restore the pixel x given an observed patch $S_v(x)$. Denote λ the unknown scalar intensity value at a given pixel whose prior $p(\lambda)$ is assumed to be uniform in the range $[0, \cdots, \lambda_M]$ where λ_M is the maximum intensity value. Consider the ABC procedure following the previous guidelines in the case of zero-mean white Gaussian noise:

1. Generate $\lambda \sim U[0, \cdots, \lambda_M]$;
2. Find a pixel y in the entire image such that $y = \arg\min_{y' \in \Omega} |v(y') - \lambda|$ (not unique) and select the patch $S_v(y)$ whose center is y;
3. Calculate the error $\epsilon(x, y) = \frac{\|S_v(y)-S_v(x)\|_2^2}{2\sigma^2}$ between the "simulated" patch $S_v(y)$ and the "observed" patch $S_v(x)$ at pixel x, compute $\pi(\epsilon(x,y))$ and return to 1.

Here $\pi(\cdot)$ denotes the unknown probability density function of the error term. Instead of uniformly drawing independent samples in the n-dimensional space, Step 2 is expected to generate more plausible "artificial data" in the sense of ABC, "closer" to the observed data. Note that Step 2 amounts to uniformly drawing a patch in the entire image domain.

3.3. Data-driven density learning

In the Gaussian case and non-overlapping patches, the errors are not centered at 0 but are expected to follow a chi-square distribution χ_n^2 with n degrees of freedom. Yet, for overlapping patches, it is established that the error is the sum of three independent χ_n^2 variables: $\epsilon(x, y) = z_1 + 2z_2 + 3z_3$ such as $z_1 \sim \chi_{n-p}^2, z_2 \sim \chi_{2p-n}^2$ and $z_3 \sim \chi_{n-p}^2$ where $\frac{n}{2} < p < n$ controls the rate of overlapping. This is explicitly true for sliding windows in one-dimensional signals and we have $\mathbb{E}[\epsilon(x,y)] = 2n$ and $\text{Var}[\epsilon(x,y)] = (12n - 4p)$.
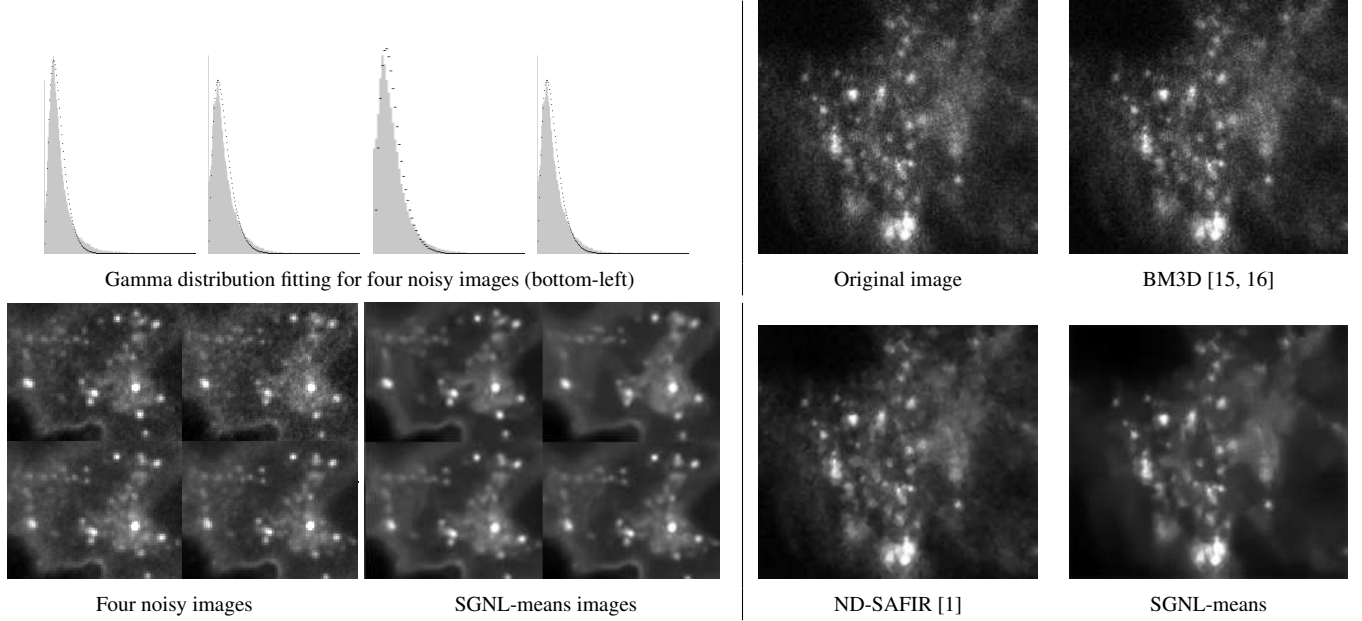
Gamma distribution fitting for four noisy images (bottom-left)

Original image

BM3D [15, 16]

Four noisy images

SGNL-means images

ND-SAFIR [1]

SGNL-means

Fig. 1. Experiments in FD-FLIM microscopy. Left: FNAR1 tagged with Green Fluoresence Protein (GFP) observed in a epithelial cell with mCHerry-tagged Tyk2 (confocal microscopy with spinning disk set-up, UMR 144 CNRS Institut Curie, Paris, France); Gamma distribution fitting and SGNL-means denoising on four subsequent images with temporally varying signal-to-noise ratios. Right: views of denoised images with state-of-the-art methods and comparisons.

In the case of variable and multiple overlappings in 2D, a general form for the distribution of the error cannot be easily obtained (e.g. see [18]). Nevertheless, it is established in [17] that the sum of weighted chi-squares variables can be approximated by a Gamma distribution controlled essentially by two parameters k and β. Consequently, we have experimentally investigated this idea of approximating the empirical density $\pi(\cdot)$ of errors by fitting Gamma distributions using the moment method, yielding the following algorithm.

3.4. Stochastic Gamma NL-means (SGNL-means)

The proposed ABC-based Gamma NL-means is based on the following stochastic two-step procedure (one iteration):

Step 1: Data-driven density learning
1. Draw uniformly with replacement $|B_r||\Omega|$ pairs of patches $(S_v(x), S_v(y))$ in the noisy image v such as $\|x - y\|_2 \leq r$ and B_r is a ball of radius $r > 0$. Compute the empirical density $\pi(\cdot)$ of errors $\epsilon(x, y) = \|S_v(y) - S_v(x)\|_2^2/(2\sigma^2)$;
2. Estimate the Gamma distribution parameters k and β by fitting the empirical mode = $(k - 1)\beta$ and variance = $k\beta^2$ of the density.

Step 2: ABC-based denoising

$$u_{SGNL}(x; k, \beta) = \frac{1}{C(x)} \sum_{y \sim U_r(x)} (\epsilon(x,y))^{k-1} e^{-\frac{\epsilon(x,y)}{\beta}} v(y) \quad (4)$$

where the set of N variables $y \sim U_r(x)$ in the sum are uniformly drawn from a ball B_r of radius r (same value as in Step 1) centered at pixel x and $C(x)$ is normalization constant.

In the implementation, we adopted a blockwise approach with patch overlapping [3, 13, 19]. Due to the overlap of patches, the restored value at a pixel x is finally obtained by uniform averaging the different estimators available at that position. Finally, we denote $u_{SGNL} = D_{SGNL}v$ the "filtered" image using the proposed SGNL-means.

3.5. Bias reduction and noise adaptation

Instead of repeating applications of the SGNL-means to reduce noise progressively, it is possible to get better denoising results by exploiting the residual image $v - u_{SGNL}$, as first suggested in [20]. The so-called "twicing" approach has been recently described in [21, 22] and is known to improve the estimator bias. Formally, define the bias of the estimator as:

$$\begin{aligned} \text{Bias}[u_{SGNL}] &= \mathbb{E}[u_{SGNL}] - u = -\mathbb{E}[v - u_{SGNL}] \quad (5) \\ &\approx -D_{SGNL}(v - u_{SGNL}). \end{aligned}$$

Given this approximation of the bias, we correct u_{SGNL} (using the same estimated parameters k and β) as [21, 22]:

$$\tilde{u}_{SGNL} = u_{SGNL} + D_{SGNL}(v - u_{SGNL}) \quad (6)$$

Reducing the bias is done at the cost of increasing the variance but the second iteration better preserves structural details. Additional iterations do not improve the results.

In the statistical framework, we have assumed that images are corrupted by white Gaussian noise. An approach to potentially adapt the SGNL-means is to substitute a dedicated noise variance model to σ^2 in (4), for instance as explained in [2] in FD-FLIM imaging. Such a noise variance model can be derived in closed-form for other image modalities [13, 19].
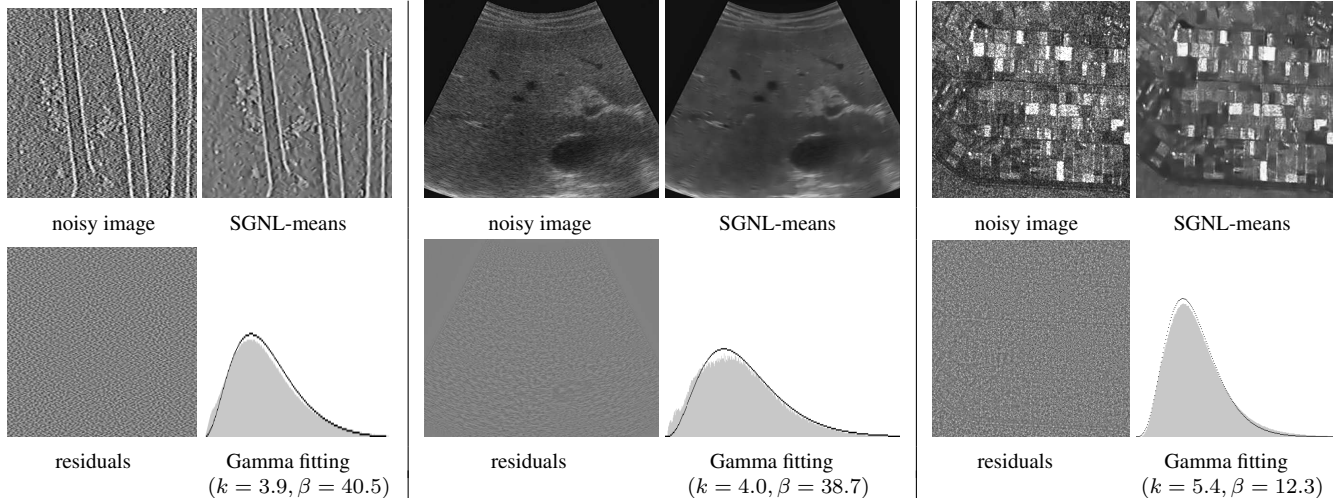
Fig. 2. Gamma distribution fitting, denoising of images with SGNL-means and residual images ($v - \tilde{u}_{SGNL}$). Left: Xenopus Microtubule Associated proteins and taxoal stablized microtubule interaction in cryo-electron tomography (Technai 200KV LaB_6, Ultrascan 1000 Gatan CCD Camera, IGDR UMR6290, Rennes, France). Middle: ultra-sound 2D image (liver) from a 3D volume (see [19] for details). Right: SAR *Lelystadt* image (Netherlands) ©ESA (see [8]).

4. EXPERIMENTAL RESULTS

All our results use the two following control parameters: $n = 7 \times 7$ and $r = 7$. We considered $N = |B_r|$ draws to compute (4) and the SGNL-means provided better results with 2 iterations (twicing). As illustrated in Figs. 1-2 for several image modalities, the parameters k and β are robustly estimated given a large set of distances computed from pairs of noisy patches. We automatically discard too large distances and we select the most homogeneous patches in the input image by using global and local noise variances estimated as in [23]. The computational time of the SGNL-means (including twicing) is of about 80s on a 512×512 image and C++ implementation on an Intel Core i7 64-bit CPU 2.4GHz.

First, on images corrupted artificially by white Gaussian noise, SGNL-means produced better results[1] than NL-means (IPOL / www.ipol.im). Performances of SGNL-means are especially demonstrated on real images in confocal imaging combined with frequency domain fluorescence lifetime imaging (FD-FLIM) (see Fig. 1). In [24], it is established that the ICCD response $v(x)$ is of the following form: $v(x) = g_{\text{INT}}(x)g_{\text{CCD}}\aleph(x) + \xi(x)$ where $\aleph(x)$ is the incident photon number assumed to follow a Poisson distribution, ξ is the CCD read-out Gaussian noise such as $\xi \sim \mathcal{N}(m_\xi, \sigma_\xi^2)$, g_{CCD} is the constant gain of the CCD sensor and $g_{\text{INT}}(x)$ is the spatially varying gain of the intensifier assumed to be random. Noise variance $\sigma^2(x)$ is spatially varying [2] but the likelihood function cannot be explicitly derived. Consequently, we have resort to stochastic simulation to generate samples from the posterior distribution. The denoising results on four subsequent images with temporally varying signal-to-noise ratios are depicted in Fig. 1 (left bottom). Assuming a mixed Poisson-Gaussian noise, we compared the performance of the ND-SAFIR algorithm [1] and BM3D [15] combined with variance stabilization [16] to adaptive variance SGNL-means. BM3D has not a significant impact on the resulting image and ND-SAFIR tends to create piecewise constant areas contrary to SGNL-means combined to adaptive noise variance [2].

In the second experiment, we addressed the problem of denoising in cryo-electron microscopy as already investigated in [25, 26]. Specimen are known to be very sensitive to electron radiation involving the spreading out of the electron dose on the whole tilt series, causing noise in the background. A closed-form for the likelihood function cannot be easily derived, and we have applied SGNL-means to improve signal-to-noise ratio on 2D images from a 3D stack (see Fig. 2 left). Finally we tested the algorithm on images damaged by multiplicative noise in ultra-sound (Fig. 2 middle) and SAR (Fig. 2 right) imaging. The procedure preserves sharp discontinuities and details while reducing noise with fewer processing artifact. In these experiments, we considered the noise variance to be constant. Results can be improved by considering more dedicated signal-dependent noise variance models [19, 8].

5. CONCLUSION

An extension of the NL means has been proposed for images damaged by different sources of noises. It is based on similarity metric learning to compare noisy patches. A Bayesian estimator for NL means, based on the idea of ABC, has been proposed. With the ABC framework, it is also possible to capture information using a number of marginal statistics [13, 19].Finally, visual results support the efficiency of the proposed method. More satisfying results will be obtained if a noise variance model is specialized in each study case.

---

[1]e.g. Lena image ($\sigma_{noise} = 20$): 32.56 (SGNL) vs 31.52 (NL-means); Barbara image ($\sigma_{noise} = 20$): 31.05 (SGNL) vs 30.21 (NL-means)

6. REFERENCES

[1] J. Boulanger, C. Kervrann, P. Bouthemy, P. Elbau, J.-B. Sibarita, and J. Salamero, "Patch-based non-local functional for denoising fluorescence microscopy image sequences," *IEEE Trans. Medical Imaging*, vol. 29, no. 2, pp. 442–454, 2010.

[2] P. Roudot, C. Kervrann, J. Boulanger, and F. Waharte, "Noise modeling for intensified camera in fluorescence imaging: application to image denoising," in *IEEE Int. Symp. Biomedical Imaging (ISBI 2013)*, San-Francisco, CA, 2013, pp. 600–603.

[3] A. Buades, B. Coll, and J.M. Morel, "A review of image denoising methods, with a new one.," *SIAM J. Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[4] A. de Decker, J.A. Lee, and M. Verlysen, "Variance stabilizing transformations in patch-based bilateral filters for poisson noise image denoising," in *IEEE Engineering in Medicine and Biology Society (EMBS 2009)*, 2009, pp. 3673–3676.

[5] J. Salmon, C.A. Deledalle, R. Willett, and Z.T. Harmany, "Poisson noise reduction with non-local PCA," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP 2012)*, Kyoto, Japan, 2012, pp. 1109–1112.

[6] R. Giryes and M. Elad, "Sparsity based poisson denoising," in *Electrical Electronics Engineers in Israel (IEEEI 2012)*, Nov 2012, pp. 1–5.

[7] C.-A. Deledalle, F. Tupin, and L. Denis, "Poisson NL-means: Unsupervised non local means for poisson noise," in *IEEE Int. Conf. Image Processing (ICIP 2010)*, Hong Kong, China, 2010, pp. 801–804.

[8] C.A. Deledalle, L. Denis, and F. Tupin, "Iterative weighted maximum likelihood denoising with probabilistic patch-based weights," *IEEE Trans. Image Processing*, vol. 18, no. 12, pp. 2661–2672, 2009.

[9] C.-A. Deledalle, L. Denis, and F. Tupin, "How to compare noisy patches? patch similarity beyond Gaussian noise," *Int. J. Computer Vision*, vol. 99, pp. 86–102, 2012.

[10] M. A. Beaumont, W. Zhang, and D. J. Balding, "Approximate Bayesian computation in population genetics," *Genetics*, vol. 162, pp. 2025–2035, 2002.

[11] P. Marjoram, P. Molitor, V. Plagnol, and S. Tavaré, "Markov chain Monte Carlo without likelihoods," *Proceedings of the National Academy of Sciences*, vol. 100, no. 26, pp. 15324–15328, 2003.

[12] C. Louchet, *Variational and Bayesian models for image denoising : from total variation towards non-local means*, Ph.D. thesis, University of Paris V, 2008.

[13] C. Kervrann, J. Boulanger, and P. Coupe, "Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal," in *Scale Space and Variational Methods (SSVM'07)*, Ischia, Italy, 2007, pp. 520–532.

[14] J. Salmon and Y. Strozecki, "Patch reprojections for Non-Local methods," *Signal Processing*, vol. 92, no. 2, pp. 477–489, 2012.

[15] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[16] M. Makitalo and A. Foi, "Optimal inversion of the generalized Anscombe transformation for Poisson-Gaussian noise," *IEEE Trans. Image Processing*, vol. 2, no. 1, pp. 91–103, 2013.

[17] A. H. Feiveson and F.C. Delaney, "The distribution and properties of a weighted sum of chi squares," Tech. Rep. NASA TN D-4575, National Aeronautics and Space Administration, Houston, Texas, May 1968.

[18] Y. Wu, B. Tracey, P. Natarajan, and J.P. Noonan, "Probabilistic non-local means," *IEEE Signal Process. Lett.*, vol. 20, no. 8, pp. 763–766, 2013.

[19] P. Coupé, P. Hellier, C. Kervrann, and C. Barillot, "Nonlocal means-based speckle filtering for ultrasound images," *IEEE Trans. Image Processing*, vol. 18, no. 10, pp. 2221–2229, 2009.

[20] J.W. Tukey, *Exploratory Data Analysis*, Addison-Wesley, 1977.

[21] P. Milanfar, "A tour of modern image filtering: New insights and methods, both practical and theoretical," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 106–128, 2013.

[22] P.-A. Cornillon, N. Hengartner, Jégou N., and E. Matzner-Lober, "Iterative bias reduction: a comparative study," *Statistics and Computing*, vol. 23, pp. 777–791, 2013.

[23] C. Kervrann, "An adaptive window approach for image smoothing and structures preserving," in *Proc. Eur. Conf. Computer Vision (ECCV'04)*, Prague, Czech Republic, May 2004, pp. 132–144.

[24] F.R. Boddeke, *Quantitative fluorescence microscopy*, Ph.D. thesis, Delft University Press, 1999.

[25] J. Darbon, A. Cunha, T. F. Chan, S. Osher, and G.J. Jensen, "Fast nonlocal filtering applied to electron cryomicroscopy," in *IEEE Int. Symp. Biomedical Imaging (ISBI 2008)*, Paris, France, 2008, pp. 1331–1334.

[26] D.Y. Wei and C.C. Yin, "An optimized locally adaptive non-local means denoising filter for cryo-electron microscopy data," *J. Structural Biology*, , no. 172, pp. 211–218, 2010.