POISSON NL MEANS: UNSUPERVISED NON LOCAL MEANS FOR POISSON NOISE

Charles-Alban Deledalle, Florence Tupin

Institut Telecom, Telecom ParisTech
CNRS LTCI
Paris, France

Loïc Denis

Observatoire de Lyon, CNRS CRAL
UCBL, ENS de Lyon, Université de Lyon
Lyon, France

ABSTRACT

An extension of the non local (NL) means is proposed for images damaged by Poisson noise. The proposed method is guided by the noisy image and a pre-filtered image and is adapted to the statistics of Poisson noise. The influence of both images can be tuned using two filtering parameters. We propose an automatic setting to select these parameters based on the minimization of the estimated risk (mean square error). This selection uses an estimator of the MSE for NL means with Poisson noise and Newton's method to find the optimal parameters in few iterations.

Index Terms— Non local means, Poisson noise, mean square error, SURE, PURE, Newton's method

1. INTRODUCTION

Poisson noise appears in low-light conditions when the number of collected photons is small, such as in optical microscopy or astronomy. Poisson noise is signal-dependent and, thus, needs to adapt the usual denoising approaches.

NL means have been proposed by Buades *et al.* in [2] to denoise images damaged by additive white Gaussian noise. While local filters lead to biases and resolution loss, NL techniques are known to efficiently reduce noise and preserve structures. Instead of combining neighboring pixels, the NL means average similar pixels. Let k_s be the observed noisy value at site s and λ_s its underlying noise-free value. NL means define the estimate $\hat{\lambda}_s$ as a weighted average:

$$\hat{\lambda}_s = \frac{\sum_t w_{s,t} k_t}{\sum_t w_{s,t}} \tag{1}$$

where t is a pixel index and $w_{s,t}$ is a data-driven weight depending on the similarity between pixels with indexes s and t. In practice, the pixels t are located in a search window centered on s. For robustness reasons, pixel similarity is evaluated by comparing surrounding patches around s and t. Patch-similarity is classically defined by the Euclidean

distance, leading to the following weight expression:

$$w_{s,t} = \exp\left(-\frac{\sum_b (k_{s+b} - k_{t+b})^2}{\alpha}\right) \tag{2}$$

where $s+b$ and $t+b$ denote the b-th pixels in the patches $\mathcal{B}_s$ and $\mathcal{B}_t$ centered on s and t, and α is a filtering parameter.

In case of low signal-to-noise ratio images, it has been shown that the performances of the NL means can be improved by refining the weights using a pre-estimate $\hat{\theta}$ of the noise-free image [3, 4, 5, 6]. A general expression of refined NL means is:

$$\hat{\lambda}_s = \frac{\sum_t w_{s,t} k_t}{\sum_t w_{s,t}} \tag{3}$$

$$\text{with} \quad w_{s,t} = \exp\left(-\frac{F_{s,t}}{\alpha} - \frac{G_{s,t}}{\beta}\right),$$

$$F_{s,t} = \sum_b f(k_{s+b}, k_{t+b})$$

$$\text{and} \quad G_{s,t} = \sum_b g\left(\hat{\theta}_{s+b}, \hat{\theta}_{t+b}\right)$$

where α and β are filtering parameters, and f and g are two similarity criteria suitable respectively to compare noisy data and pre-estimated data. A typical choice for f and g is the squared difference: $f(x,y) = g(x,y) = (x-y)^2$. The choice of the filtering parameters for α and/or β is a critical task already explored in [3, 6] and [7]. According to our knowledge, there is no known method to jointly set α and β.

2. PATCH-SIMILARITIES UNDER POISSON NOISE

Let k be a noisy observation following a Poisson distribution with parameters described by a noise-free value λ:

$$p(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}. \tag{4}$$

We take inspiration from the probabilistic approach of [6, 8] to extend the refined NL means in (3) to the Poisson noise degradation model. The squared difference classically used

for f can be replaced by:

$$f_L(k_1, k_2) = -\log \frac{\max_\lambda p(k_1|\lambda_1 = \lambda) p(k_2|\lambda_2 = \lambda)}{\max_\lambda p(k_1|\lambda_1 = \lambda) \max_\lambda p(k_2|\lambda_2 = \lambda)}$$

$$= k_1 \log k_1 + k_2 \log k_2 - (k_1 + k_2) \log \left(\frac{k_1 + k_2}{2} \right). \quad (5)$$

This similarity evaluates the likelihood ratio corresponding to the hypothesis that k_1 and k_2 share an identical λ parameter against the hypothesis that their λ parameters are independent. The squared difference generally used for g can be replaced by the symmetric Kullback-Leibler divergence:

$$g_{KL}\left(\hat{\theta}_1, \hat{\theta}_2\right) = D_{KL}\left(\hat{\theta}_1 \| \hat{\theta}_2\right) = \left(\hat{\theta}_1 - \hat{\theta}_2\right) \log \frac{\hat{\theta}_1}{\hat{\theta}_2}. \quad (6)$$

This criterion is a good candidate to define similarities between estimated values since it can be considered as a statistical test of the hypothesis $\lambda_1 = \lambda_2$.

The setting of the parameters α and β in the case of Poisson noise is maybe a more critical problem than in other denoising tasks. In [2, 3] and [6], the authors propose to define the parameters according to the variance or the quantiles of the similarity criteria. Unfortunately, in the case of Poisson noise, these quantities depend on the unknown image λ since the noise is signal-dependent. Van De Ville *et al.* propose a risk minimization approach for Gaussian noise [7]. Their method selects the parameters minimizing the risk (without any specific assumption on the underlying image λ). This kind of approach seems relevant in the case of Poisson noise.

3. AUTOMATIC SETTING OF PARAMETERS BASED ON RISK MINIMIZATION

The parameters of the denoising technique can be selected as those that minimize the expected MSE:

$$\mathsf{E}\left[\frac{1}{N}\|\lambda - \hat{\lambda}\|^2\right] = \frac{1}{N} \sum_s \left(\lambda_s^2 + \mathsf{E}\left[\hat{\lambda}_s^2\right] - 2\mathsf{E}\left[\lambda_s \hat{\lambda}_s\right]\right) \quad (7)$$

with N the image size and $\mathsf{E}[.]$ the expectation operator. The MSE requires the knowledge of the noise-free image λ but can still be estimated from the noisy image k alone. Since the first term λ_s^2 in (7) is independent of $\hat{\lambda}_s$, it can be omitted when minimizing the MSE with respect to the denoising parameters. The Stein's unbiased risk estimator (SURE) is an estimator of the MSE under Gaussian noise [9]. It is based on an estimator of $\mathsf{E}[\lambda_s \hat{\lambda}_s]$ which does not require λ. SURE has already been used successfully on images damaged by additive white Gaussian noise for wavelet filtering [10] and NL means filtering [7]. The main result in [7] is that SURE for NL means can be obtained in closed form. For Poisson noise, we can use the result of Chen [1] to follow the same approach:

$$\mathsf{E}\left[\lambda_s h(k)_s\right] = \mathsf{E}\left[k_s h(\overline{k})_s\right] \quad (8)$$

$$\text{with} \quad \overline{k}_t = \begin{cases} k_t & \text{if } t \neq s \\ k_t - 1 & \text{otherwise} \end{cases}$$

where $h(.)_s$ denotes the estimated value on site s obtained by the application of the estimator h on the given noisy image. By injecting (8) in (7), we obtain the Poisson unbiased risk estimator (PURE) as introduced in [11] for wavelet denoising. Considering now that the estimator h is the refined NL means given in eq. (3), PURE can be expressed as follows:

$$R(\hat{\lambda}) = \frac{1}{N} \sum_s \left(\lambda_s^2 + \hat{\lambda}_s^2 - 2k_s \overline{\lambda}_s\right). \quad (9)$$

The value $\overline{\lambda}_s$ refers to the denoised value obtained by the application of the NL means on the noisy image $\overline{k}$, i.e:

$$\overline{\lambda}_s = \frac{\sum_t \overline{w}_{s,t} \overline{k}_t}{\sum_t \overline{w}_{s,t}} \quad (10)$$

$$\text{with} \quad \overline{w}_{s,t} = \exp\left(-\frac{\overline{F}_{s,t}}{\alpha} - \frac{G_{s,t}}{\beta}\right)$$

$$\text{and} \quad \overline{F}_{s,t} = \sum_b f(\overline{k}_{s+b}, \overline{k}_{t+b}).$$

According to (7), (8) and (9), it is straightforward to show that $R(\hat{\lambda})$ is unbiased. However, note that (10) holds by assuming that $G_{s,t}$ (i.e the pre-estimate $\hat{\theta}$) does not depend on the noise component of k. To satisfy this assumption, the noise variance in $\hat{\theta}$ has to be reduced significantly.

In terms of time complexity, we note as in [7] that the computation time is unchanged since the computation of PURE can be incorporated within the core of the NL means. Moreover, the scan of the patches of $\overline{k}$ can be avoided thanks to the following relation:

$$\overline{F}_{s,t} = F_{s,t} +$$
$$\begin{cases} f(\overline{k}_s, \overline{k}_s) - f(k_s, k_s), & \text{if } s = t, \\ f(\overline{k}_s, k_t) - f(k_s, k_t) \\ \quad + f(k_{2s-t}, \overline{k}_s) - f(k_{2s-t}, k_s), & \text{if } s \in \mathcal{B}_t, \\ f(\overline{k}_s, k_t) - f(k_s, k_t), & \text{otherwise.} \end{cases}$$

Selecting parameters that minimize PURE gives parameters close to that minimizing the MSE. In the case of the classical NL means, the authors of [7, 11] compute the optimal parameters by exhaustive search. Optimization techniques can be applied to reach the optimal parameters in few iterations. We suggest to optimize $R(\hat{\lambda})$ for the refined NL means by using Newton's method on the joint filtering parameters α and β. Newton's method iteratively refines α and β with the update:

$$\begin{pmatrix} \alpha^{(n+1)} \\ \beta^{(n+1)} \end{pmatrix} = \begin{pmatrix} \alpha^{(n)} \\ \beta^{(n)} \end{pmatrix} - H^{-1}\nabla \quad (11)$$

$$\text{with } H^{-1}\nabla = \begin{pmatrix} \frac{\partial^2 R(\hat{\lambda}^{(n)})}{\partial \alpha^2} & \frac{\partial^2 R(\hat{\lambda}^{(n)})}{\partial \alpha \partial \beta} \\ \frac{\partial^2 R(\hat{\lambda}^{(n)})}{\partial \beta \partial \alpha} & \frac{\partial^2 R(\hat{\lambda}^{(n)})}{\partial \beta^2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial R(\hat{\lambda}^{(n)})}{\partial \alpha} \\ \frac{\partial R(\hat{\lambda}^{(n)})}{\partial \beta} \end{pmatrix}.$$

where n is the current iteration index. To perform the optimization procedure in (11), the first and second order differ-
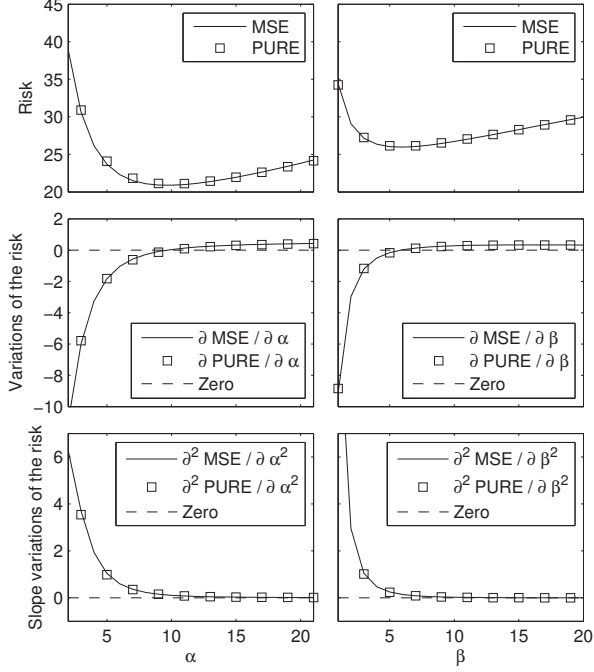
Fig. 1. The risk (MSE and PURE) and their two first order variations (from top to bottom) with respect to the parameters α (left) and β (right).

entials are required. Their expressions are given by substituting x and y by α or β in the following equations:

$$\frac{\partial R(\hat{\lambda})}{\partial x} = \frac{2}{N} \sum_s \hat{\lambda}_s \frac{\partial \hat{\lambda}_s}{\partial x} - \frac{2}{N} \sum_s k_s \frac{\partial \overline{\lambda}_s}{\partial x},$$

$$\frac{\partial^2 R(\hat{\lambda})}{\partial x \partial y} = \frac{2}{N} \sum_s \hat{\lambda}_s \frac{\partial^2 \hat{\lambda}_s}{\partial x \partial y} + \frac{2}{N} \sum_s \left(\frac{\partial \hat{\lambda}_s}{\partial x} \right) \left(\frac{\partial \hat{\lambda}_s}{\partial y} \right)$$
$$- \frac{2}{N} \sum_s k_s \frac{\partial^2 \overline{\lambda}_s}{\partial x \partial y},$$

$$\frac{\partial \hat{\lambda}_s}{\partial x} = \frac{\sum X_{s,t} w_{s,t} (k_t - \hat{\lambda}_s)}{x^2 \sum w_{s,t}},$$

$$\frac{\partial^2 \hat{\lambda}_s}{\partial x^2} = \frac{\sum X_{s,t}^2 w_{s,t} (k_t - \hat{\lambda}_s)}{x^4 \sum w_{s,t}} - 2 \frac{\partial \hat{\lambda}_s}{\partial x} \frac{\sum (X_{s,t} + x) w_{s,t}}{x^2 \sum w_{s,t}},$$

$$\frac{\partial^2 \hat{\lambda}_s}{\partial x \partial y} = \frac{\sum X_{s,t} Y_{s,t} w_{s,t} (k_t - \hat{\lambda}_s)}{x^2 y^2 \sum w_{s,t}}$$
$$- \frac{\partial \hat{\lambda}_s}{\partial x} \frac{\sum Y_{s,t} w_{s,t}}{y^2 \sum w_{s,t}} - \frac{\partial \hat{\lambda}_s}{\partial y} \frac{\sum X_{s,t} w_{s,t}}{x^2 \sum w_{s,t}},$$

where $X = F$ (resp. $Y = F$) when $x = \alpha$ (resp. $y = \alpha$) and $X = G$ (resp. $Y = G$) when $x = \beta$ (resp. $y = \beta$). The differentials for $\overline{\lambda}$ are the same with respect to $\overline{k}$, $\overline{w}$ and $\overline{F}$.

Newton's method finds in few iterations the best tradeoff between the information brought by the noisy image and the pre-estimated image to define the weights. For instance,

Peppers (256×256)				
Noisy	3.14	13.14	17.91	23.92
MA filter	19.20	20.93	21.11	21.16
PURE-LET [11]	19.33	24.29	27.27	30.79
NL means [2]	18.12	23.33	26.98	30.64
Refined NL means	19.84	24.48	27.56	30.93
Poisson NL means	**19.90**	**25.32**	**28.07**	**31.06**
α_{opt}	(209)	(13.6)	(10.05)	(9.21)
β_{opt}	(0.72)	(1.31)	(2.76)	(7.64)
$\#iterations$	(13.5)	(8.02)	(7.03)	(6.90)
Cameraman (256×256)				
Noisy	3.28	13.27	18.03	24.05
MA filter	18.71	20.15	20.29	20.33
PURE-LET [11]	19.67	24.32	26.87	**30.36**
NL means [2]	18.17	23.53	26.77	29.39
Refined NL means	19.80	24.53	27.18	29.44
Poisson NL means	**19.89**	**25.07**	**27.42**	29.47
α_{opt}	(62.1)	(9.48)	(8.81)	(7.34)
β_{opt}	(0.51)	(1.19)	(3.57)	(16.19)
$\#iterations$	(11.0)	(6.80)	(7.60)	(11.3)

Table 1. PSNR values averaged over ten realisations using different methods on images damaged by Poisson noise with different levels of degradation. The averaged optimal parameters and the averaged number of iterations of the proposed Poisson NL means are given.

β will get a high value when the pre-estimated image has a poor quality, resulting in weights determined only from the noisy image. Reciprocally, α will get a high value when the pre-estimated image has a high quality: the weights will be determined only from the well pre-estimated image.

4. EXPERIMENTS AND RESULTS

The proposed extension of the NL means (Poisson NL means) is applied with a search window of size 21×21 and patches of size 7×7. Newton's method is performed until PURE does not change between two successive iterations. The pre-estimated image is obtained by a moving average (MA) filter with a 13×13 disk kernel. Using the optimization of [12], the computational time is of about 10s per iteration on a 256×256 image and C implementation on an Intel Core 2 Duo 64-bit CPU 3.00GHz.

Figure 1 shows the risk and its two first order differentials with respect to α and β. These curves have been computed by applying the proposed method on a 150×150 noisy image for different values of the parameters. The MSE and its differentials have been computed from the noise-free image and finite differences. PURE and its differentials have been evaluated using the expressions given in Section 3.

Table 1 gives the peak signal-to-noise ratio (PSNR) values obtained by different denoising methods averaged over
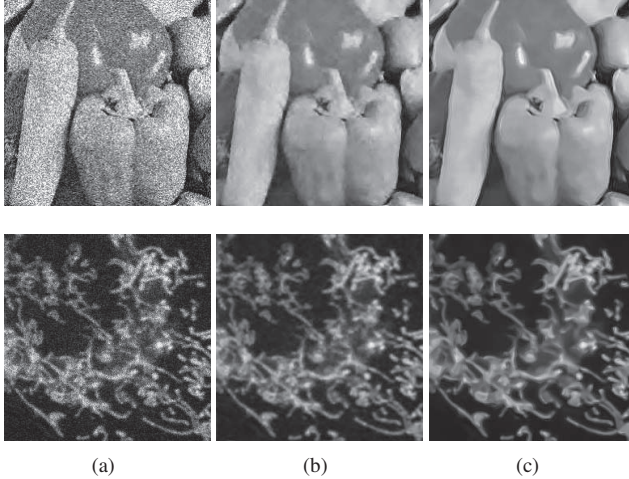
Fig. 2. (a) Original images damaged by Poisson noise, denoised images obtained by (b) PURE-LET [11] and (c) the proposed Poisson NL means.

5. CONCLUSION

Inspired by the methodology of [6], an extension of the NL means has been proposed for images damaged by Poisson noise. It is based on probabilistic similarities to compare noisy patches and patches of a pre-estimated image. A risk estimator for NL means, based on the idea of [7], has been derived for Poisson noise. This risk estimator is used in an optimization method to automatically select the filtering parameters in few iterations. Numerical results as well as visual results support the efficiency of the proposed method.

ten realizations of two 256×256 reference images damaged by synthetic Poisson noise with different degradation levels. The MA filter is applied with a 13×13 disk kernel. We compare our method with the PURE based wavelet approach (PURE-LET) proposed in [11]. Three versions of NL means are applied. NL means denotes the classical one i.e $f(x, y) = (x - y)^2$ and $\beta = \infty$. We call the refined NL means when $f(x, y) = g(x, y) = (x - y)^2$. Poisson NL means denotes our proposed method with f and g defined as in Section 2. For all NL means versions, the optimal parameters are obtained by PURE minimization using Newton's method. The table gives the optimal parameters α_{opt} and β_{opt} and the number of iterations for the Poisson NL means. The refined and the Poisson NL means both use the pre-estimated image obtained by the MA filter. Poisson NL means provides generally better performance with on average 6 to 14 iterations. The parameters behave as predicted with respect to the relative qualities of the noisy image and the pre-estimated image.

Figure 2 presents visual results[1] obtained by the PURE-LET approach[2] and Poisson NL means on two images. The first image is an image degraded by synthetic Poisson noise and the second one is an image[3] of a mitochondrion sensed in low-light conditions by confocal fluorescence microscopy [13]. Even if both filters seem to preserve well the resolution while reducing the noise, the Poisson NL means seem to provide a more regular result with fewer processing artifacts.

---

[1]More results available at http://perso.telecom-paristech.fr/~deledall/poisson_nlmeans.php

[2]we are grateful to F. Luisier for providing the results of PURE-LET

[3]image courtesy of Y. Tourneur

6. REFERENCES

[1] L.H.Y. Chen, "Poisson approximation for dependent trials," *The Annals of Probability*, vol. 3, no. 3, pp. 534–545, 1975.

[2] A. Buades, B. Coll, and J.M. Morel, "A Non-Local Algorithm for Image Denoising," in *Proc. IEEE Computer Society Conf. CVPR*, 2005, vol. 2, pp. 60–65.

[3] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 2866–2878, 2006.

[4] T. Brox, O. Kleinschmidt, and D. Cremers, "Efficient Nonlocal Means for Denoising of Textural Patterns," *IEEE Transactions on Image Processing*, 2007.

[5] B. Goossens, H. Luong, A. Pižurica, and W. Philips, "An improved non-local denoising algorithm," in *Proc. Int. Workshop on LNLA in Image Processing, Lausanne, Switzerland*, 2008.

[6] C.A. Deledalle, L. Denis, and F. Tupin, "Iterative Weighted Maximum Likelihood Denoising with Probabilistic Patch-Based Weights," *IEEE Transactions on Image Processing*, vol. 18, no. 12, pp. 2661–2672, Dec. 2009.

[7] D. Van De Ville and M. Kocher, "SURE-Based Non-Local Means," *IEEE Signal Processing Letters*, vol. 16, no. 11, pp. 973–976, November 2009.

[8] F. Alter, Y. Matsushita, and X. Tang, "An intensity similarity measure in low-light conditions," *Lecture Notes in Computer Science*, vol. 3954, pp. 267, 2006.

[9] C.M. Stein, "Estimation of the mean of a multivariate normal distribution," *The Annals of Statistics*, pp. 1135–1151, 1981.

[10] D.L. Donoho and I.M. Johnstone, "Adapting to Unknown Smoothness Via Wavelet Shrinkage.," *Journal of the american statistical association*, vol. 90, no. 432, 1995.

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

[12] J. Darbon, A. Cunha, T.F. Chan, S. Osher, and G.J. Jensen, "Fast nonlocal filtering applied to electron cryomicroscopy," in *Proc. IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro*, 2008, pp. 1331–1334.

[13] S. Pelloux, J. Robillard, R. Ferrera, A. Bilbaut, C. Ojeda, V. Saks, M. Ovize, and Y. Tourneur, "Non-beating HL-1 cells for confocal microscopy: application to mitochondrial functions during cardiac preconditioning," *Progress in biophysics and molecular biology*, vol. 90, no. 1-3, pp. 270–298, 2006.