

AN ADAPTIVE WINDOW APPROACH FOR POISSON NOISE REDUCTION AND STRUCTURE PRESERVING IN CONFOCAL MICROSCOPY

Charles Kervrann

Alain Trubuil

IRISA-INRIA Rennes / INRA - MIA
Campus de Beaulieu
35042 Rennes Cedex, France

INRA - MIA
Domaine de Vilvert
78352 Jouy-en-Josas, France

ABSTRACT

In domains like confocal microscopy, the imaging process is based on detection of photons. It is established the additive Gaussian noise model is a poor description of the actual photon-limited image recording, compared with that of a Poisson process. This motivates the use of restoration methods optimized for Poisson noise distorted images. In this paper, we propose a novel restoration approach for Poisson noise reduction and discontinuities preservation in images. The method is based on a locally piecewise constant modeling of the image with an adaptive choice of a window around each pixel. The restoration technique associates with each pixel the weighted sum of data points within the window. It is worth noting the proposed technique applied to confocal microscopy is data-driven and does not require the hand tuning of parameters.

1. INTRODUCTION

Confocal systems offer the chance to image thick biological tissue in 2D or 3D dimensions. At each point of the image, the emitted fluorescence for the object is focused on the detector. This light is converted by a photomultiplier tube (PMT) into an electrical signal and represented by a discrete value after an A/D conversion [10]. Finally, the PMT essentially behaves as a photon counter and the distortions caused by the quantum nature of the photon detection process are better described by Poisson statistics. Operations that invert these distortions of the microscope are then necessary to improve the quantitative analysis of images. Generally restoration methods yield an estimate of the original image given an imaging model, a noise model and additional criteria. In previous work [12, 6, 7], the Richardson-Lucy algorithm, which is a maximum likelihood estimator for the intensity of a Poisson process, incorporates a non-negativity constraint in the algorithm. This algorithm is generally capable of partially reducing the distortions found in confocal 3D images. However, it is sensitive to noise [6] and additional methods (Gaussian pre-filtering) are necessary to

produce better restoration results [7].

In this paper, we also address the adaptive image restoration problem and present a nonparametric estimation method that smooths homogeneous regions and inhibits smoothing in the neighborhood of discontinuities. The observed data are imperfect and in the form of Poisson process. Since we do not address the image formation of the confocal fluorescence microscope, ideally modeled as a convolution of the object function with the point spread function [10], the proposed method can be seen also as a sophisticated pre-filtering method before starting the more complex *deconvolution* process using the Lucy-Richardson algorithm [7]. The proposed *adaptive window approach* is conceptually very simple being based on the key idea of estimating a locally regression function with an adaptive choice of the window size (neighborhood) for which the applied model fits the data well [11]. At each pixel, we estimate the regression function by iteratively growing a window and adaptively weighting input data to achieve an optimal compromise between the bias and variance [8, 5, 9]. The proposed algorithm complexity is actually controlled by simply restricting the size of the larger window and setting the window growing factor. In contrast to most digital diffusion-based filtering processes for which the input noisy image is “abandoned” after the first iteration [2], the adaptive window approach recycles at each step the original data.

2. A NONPARAMETRIC APPROACH

In photon-limited imaging, the major source of errors is Poisson noise due to the discrete nature of photon detection. Unlike Gaussian noise, Poisson noise is signal dependent, which makes separating signal from noise a very difficult task. However, by applying the Anscombe transform [13] defined as $Y_i = 2\sqrt{f_i + \frac{3}{8}}$, the Poisson data $f_i \sim \text{Poiss}(\nu_i)$ of intensity ν_i are transformed to data with a Gaussian distribution with variance $\sigma^2 = 1$. Such an assertion is asymptotically correct as $\nu_i \rightarrow \infty$. This transformation allows one to use well-studied methods for Gaussian noise on data

corrupted with the much trickier Poisson noise. Then we applied a method developed for Gaussian noise. After denoising, the inverse Anscombe transform is applied.

2.1. Image model and basic idea

We observe the regression function u with some additive errors ξ_i : $Y_i = u(x_i) + \xi_i$, $i = 1, \dots, n$, where $x_i \in \mathbb{R}^d$, $d = 2, 3$, represents the spatial coordinates of the discrete image domain S of n pixels and $Y_i \in \mathbb{R}$ is the observed intensity at location x_i and defined as previously. We suppose the errors ξ_i to be independent identically distributed zero-mean random variables with unknown variances, i.e., $\text{var}(\xi_i) = \sigma^2$. However, to cope with more complex degradations, the noise variance is also an unknown parameter of the method.

A classical nonparametric estimation approach is based on the structural assumption that regression function $u(x)$ is constant in the vicinity of a point x . An important question under such an approach is first how to determine for each pixel the size and shape of the neighborhood under concern from image data. The regression function $u(x)$ can be then estimated from the observations lying in the estimated neighborhood of x by a local maximum likelihood (ML) method.

The proposed procedure is iterative and mostly realizes this idea. First, suppose we are given a local window $W_i^{(0)}$ containing the point of estimation x_i . We can calculate an initial ML estimate $\hat{u}_i^{(0)}$ at point x_i (and its variance $\hat{\vartheta}_i^{(0)}$) by averaging observations over a small neighborhood $W_i^{(0)}$ of x_i as

$$\hat{u}_i^{(0)} = \frac{1}{|W_i^{(0)}|} \sum_{x_j \in W_i^{(0)}} Y_j \quad \text{and} \quad \hat{\vartheta}_i^{(0)} = \frac{\hat{\sigma}^2}{|W_i^{(0)}|} \quad (1)$$

where $\hat{\sigma}^2$ is the unknown estimate of the noise variance and $|W_i^{(0)}|$ denotes the number of points $x_j \in W_i^{(0)}$. At the next iteration, a larger neighborhood $W_i^{(1)}$ with $W_i^{(0)} \subset W_i^{(1)}$ centered at x_i is considered, and every point x_j from $W_i^{(1)}$ gets a weight $w_{ij}^{(1)}$ which is defined by comparing the estimates $\hat{u}_i^{(0)}$ and $\hat{u}_j^{(0)}$ obtained at the first iteration. Then we recalculate the estimate $\hat{u}_i^{(1)}$ as the weighted average of data points lying in the neighborhood $W_i^{(1)}$. We continue this way, increasing with k the considered neighborhood $W_i^{(k)}$; for each $k \geq 1$, the ML estimate $\hat{u}_i^{(k)}$ and its variance are given by

$$\hat{u}_i^{(k)} = \sum_{x_j \in W_i^{(k)}} w_{ij}^{(k)} Y_j \quad \text{and} \quad \hat{\vartheta}_i^{(k)} = \hat{\sigma}^2 \sum_{x_j \in W_i^{(k)}} [w_{ij}^{(k)}]^2 \quad (2)$$

where weights $w_{ij}^{(k)}$ are continuous variables ($0 \leq w_{ij}^{(k)} \leq 1$), computed by comparison of the preceding estimates $\hat{u}_i^{(k-1)}$

and $\hat{u}_j^{(k-1)}$. In the next section, statistical arguments for calculating weights $w_{ij}^{(k)}$ are given. We also add a pointwise control step for selecting the optimal window size for each pixel, by comparing the new estimate $\hat{u}_i^{(k)}$ with the estimate $\hat{u}_i^{(k-1)}$ obtained at the preceding iteration [11]. In Equation (2), the weight function $w_{ij}^{(k)}$ does not directly depend on input data but are only calculated from neighboring local estimates, which contributes to the regularization effect.

2.2. Adaptive weights

In our approach, we may decide on the basis of the estimates $\hat{u}_i^{(k-1)}$ and $\hat{u}_j^{(k-1)}$, whether points x_i and $x_j \in W_i^{(k)}$ are in the same region or not and then prevent from significant discontinuities oversmoothing [11]. In the local Gaussian case, significance is measured using a contrast $|\hat{u}_i^{(k-1)} - \hat{u}_j^{(k-1)}|$.

If this contrast is high compared to $\sqrt{\hat{\vartheta}_i^{(k-1)}}$, then x_j should not participate to the estimation of $\hat{u}_i^{(k)}$ and $w_{ij}^{(k)} \rightarrow 0$. Hence, motivated by the robustness and smoothing properties of the Huber M-estimator in the probabilistic approach of image denoising [1], we introduce the following related weight function (but other weight functions are possible [2]):

$$w_{ij}^{(k)} = \frac{g_{ij}^{(k)}}{\sum_{x_j \in W_i^{(k)}} g_{ij}^{(k)}}, \quad (3)$$

$$g_{ij}^{(k)} = \begin{cases} 1 & \text{if } |\hat{u}_i^{(k-1)} - \hat{u}_j^{(k-1)}| \leq 3\sqrt{\hat{\vartheta}_i^{(k-1)}} \\ \frac{3\sqrt{\hat{\vartheta}_i^{(k-1)}}}{|\hat{u}_i^{(k-1)} - \hat{u}_j^{(k-1)}|} & \text{otherwise} \end{cases}$$

Here $3\sqrt{\hat{\vartheta}_i^{(k-1)}}$ is related to the spatially varying fraction of contamination of the Gaussian distribution (“rule of 3 sigma” for Gaussian distributions.): for the majority of points $x_j \in W_i$, the differences $\hat{u}_i^{(k-1)} - \hat{u}_j^{(k-1)}$ can be approximately modeled as being constant (zero) with random Gaussian noise. Large differences are assumed to be outliers which should not have a large effect on the estimator.

2.3. Optimal window selection

The classical measure of the closeness of the estimator \hat{u} obtained in the window W_i to its target value u is the mean squared error (MSE) which is decomposed into the sum of the squared bias and variance. As explained before, we should choose a window that achieves an optimal compromise between the squared bias and variance. Accordingly, we make the reasonable assumption that the squared bias is an increasing function of the neighborhood size and the variance is a decreasing function of the neighborhood size.

Then, in order to minimize the pointwise MSE we search for the window where the square bias and the variance of the estimate are equal for each point x_i .

Now, let us introduce a finite set of k_M windows $\{U_i^{(0)}, \dots, U_i^{(k_M)}\}$ centered at x_i , with $W_i^{(k)} \subset W_i^{(k+1)}$, starting with a small $U_i^{(0)}$ and the corresponding estimates $\hat{u}_i^{(0)}$ of the true image $u(x_i)$. Denote by $U_i^{(\hat{k}_i)}$ the ideal window size corresponding to the minimum value of the MSE. Then the optimal window $U_i^{(\hat{k}_i)}$ can be obtained according to the following statistical rule [5, 8, 9]:

$$\hat{k}_i = \max \left\{ k : \forall k' < k : \left| \hat{u}_i^{(k)} - \hat{u}_i^{(k')} \right|^2 \leq 8\hat{\vartheta}_i^{(k')} \right\}. \quad (4)$$

In other words, as long as successive estimates $\hat{u}_i^{(k)}$ stay close to each other, we decide that the bias is small and the size of the estimation window can be increased to improve the estimation of the constant model (and to decrease the variance of the estimate $\hat{u}_i^{(k)}$). If an estimated point $\hat{u}_i^{(k')}$ appears far from the previous ones, we interpret this as the dominance of the bias over the variance term. For each pixel, the detection of this transition enables to determine the critical window size that balances bias and variance.

2.4. Convergence

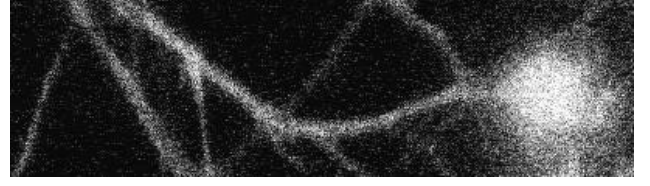
In this section, we propose a stopping rule to save computation time if two successive solutions are very close and prevents from an useless setting of the larger window size. In our approach, we adopt the so-called Csiszár's I-divergence [3, 7] to detect global convergence defined as:

$$I(\hat{u}^{(k)}, \hat{u}^{(k+1)}) = \sum_{i=1}^n \left[\hat{u}_i^{(k)} \log \frac{\hat{u}_i^{(k)}}{\hat{u}_i^{(k+1)}} - \hat{u}_i^{(k)} + \hat{u}_i^{(k+1)} \right]. \quad (5)$$

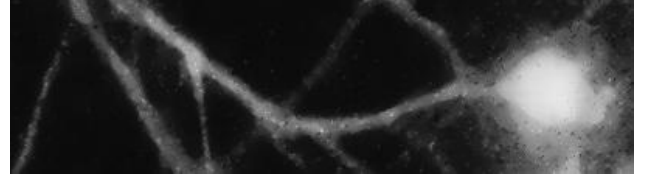
In practice, the I-divergence is normalized with its maximal occurring value at iteration $k = 0$. When $I(\hat{u}^{(k)}, \hat{u}^{(k+1)})$ sinks under a threshold (of the order 10^{-3} for typical images) that sufficiently accounts for convergence, the algorithm is stopped at the final iteration $k_c = k$, with $k_c \leq k_M$, where k_M is the maximal value for k set by the user. Finally, the window size increases at each iteration k if the global convergence criterion is not met (or $k \leq k_M$) without changing the estimate $\hat{u}_i^{(k)}$ if the pointwise rule (4) has already been violated at x_i , i.e. $\hat{u}_i^{(k)} = \hat{u}_i^{(\hat{k}_i)}$ if $\hat{k}_i < k$. If the rule (4) has not been violated at x_i , we have $k_i = k$ where k is the current iteration of the algorithm.

3. THE ALGORITHM

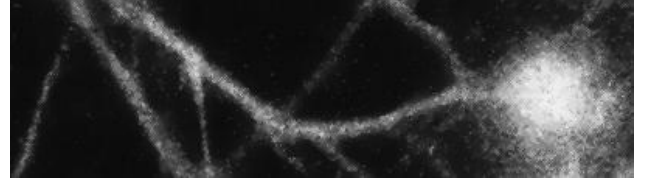
The key ingredient of the proposed algorithm is an increasing sequence of neighborhoods $W_i^{(k)} \subset W_i^{(k+1)}$ centered at each image pixel x_i . In what follows, $|W_i^{(k)}|$ denotes



(a) confocal image



(b) denoised image using Poisson statistics



(c) denoised image using Gaussian statistics

Fig. 1. Denoising of a confocal image showing an individual neuronal cell.

the number of points x_j in $W_i^{(k)}$. We arbitrarily choose $|W^{(k)}| = (2k+1) \times (2k+1)$ pixels with $k = 0, 1, \dots, k_M$. For each point x_i , we naturally choose $|W^{(0)}| = 1$ and calculate initial estimates $\hat{u}_i^{(0)}$ and $\hat{\vartheta}_i^{(0)}$ using Equation (1) and set $k = 1$. Here $\hat{\sigma}^2$ is the noise variance robustly estimated from data as it is explained in [4, 2]. The algorithm is then as follows:

Estimation For all x_j in $W_i^{(k)}$, we compute weights $w_{ij}^{(k)}$ using Equation (3) and new estimates $\hat{u}_i^{(k)}$ and $\hat{\vartheta}_i^{(k)}$ using Equation (2).

Pointwise control After the estimate $\hat{u}_i^{(k)}$ has been computed, we compare it to the previous estimates $\hat{u}_i^{(k')}$ at the same point x_i for all $k' < k$. From the rule (4), if there is at least one index $k' < k$ such that $\left| \hat{u}_i^{(k)} - \hat{u}_i^{(k')} \right| > 2\sqrt{2} \sqrt{\hat{\vartheta}_i^{(k')}}$ then we do not accept $\hat{u}_i^{(k)}$ and keep the estimates $\hat{u}_i^{(k-1)}$ from the preceding iteration as the final estimate at location x_i [5, 9]. This estimate is unchanged if $k > \hat{k}_i$.

Convergence We stop the procedure if $k = k_M$ or if $I(\hat{u}^{(k)}, \hat{u}^{(k+1)}) < 10^{-3}$, otherwise increase k by 1 and continue with the estimation step. Here we use k_M to bound the numerical complexity. This parameter is set to 15 overestimates the number of necessary iterations.

4. EXPERIMENTS

We have tested the algorithm (also implemented for processing 3D data) on confocal fluorescence microscopy images that contain complex structures. The potential of the adaptive window method is first shown on a 2D confocal microscopy 115×512 image (Fig. 1, courtesy of INSERM 413 IFRMP $n^{\circ}23$, Rouen, France) depicting a neurite in cultured cerebellar granule cells. Fig. 1a shows the highly noisy image where high gray-level values correspond to elevated calcium concentration. Figures 1b and 1c contain the denoised images using the adaptive window method combined with respectively Poisson statistics and Gaussian statistics; the image denoised using Poisson statistics contains larger homogeneous regions. The adaptive window method using Poisson statistics has been also applied to 3D data: a typical 2D image taken from a 3D stack of 20 images depicting membranes of about fifty cultured human cells is shown in Fig. 2a (courtesy of INRA - UFDNH, Nantes, France). The image is denoised (Fig. 2b) using the set of parameters used in the previous experiments. Finally, the performance of the restoration procedure is demonstrated for a 3D fluorescence microscopic ($179 \times 144 \times 16$) stack. Figure 3a shows a typical 2D image taken from the 3D stack of 16 images depicting moving chromosomes (with dark values) standing over a spatially varying background, during mitosis (courtesy of Institut Curie, Paris, France). Figure 3b shows the image denoised using Poisson statistics, where the location of chromosomes (dark spots) are well preserved. Figure 3c shows also the locations and sizes of optimal estimation; we have coded small windows with black and large windows with white. As expected, small windows are in the neighborhood of chromosomes.

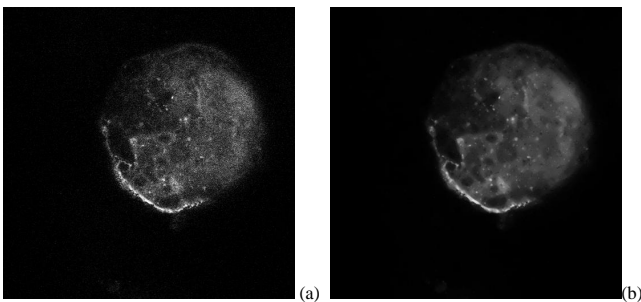


Fig. 2. Denoising of a 3D confocal stack depicting membranes of cultured human cells.

5. CONCLUSION

We have described a novel feature-preserving adaptive algorithm that reduces Poisson noise with a controllable computational complexity. The processing of the 256×256 image required typically 3 seconds on a PC (2.6 Ghz, Pentium IV) using a standard C++ implementation of the al-

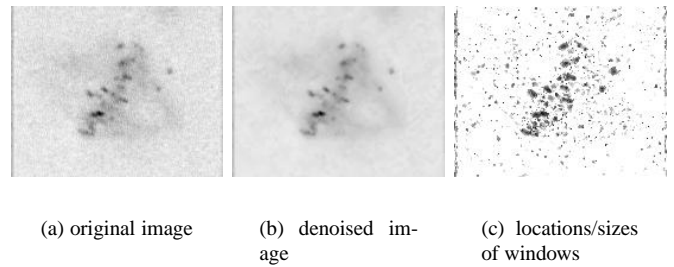


Fig. 3. Denoising of a fluorescence microscopic image showing human chromosomes (centromeres).

gorithm. The proposed scheme can be seen also as an alternative method to the anisotropic diffusion [2] and energy minimization methods. An advantage of the method is that no hidden parameters need to be precisely adjusted as in many other methods and the algorithm can be parallelized. Experimental results show its effectiveness and demonstrate its potential for 2D and 3D confocal microscopy.

6. REFERENCES

- [1] A. Ben Hamza, H. Krim, "A variational approach to maximum a posteriori estimation for image denoising", in *Proc. EMMCVPR'01*, LNCS 2134, Sophia-Antipolis, France, 2001, pp. 19-34.
- [2] M.J. Black, G. Sapiro, D. Marimont, D. Heeger, "Robust anisotropic diffusion", *IEEE Trans. Image Proc.*, vol. 7, no. 3, pp. 421-432, 1998.
- [3] I. Csiszár, "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems", *Ann. Statist.*, vol. 19, pp. 2032-2066, 1991.
- [4] T. Gasser, L. Sroka, C. Jennen Steinmetz, "Residual variance and residual pattern in nonlinear regression", *Biometrika*, vol. 73, pp. 625-633, 1986.
- [5] A. Juditsky, "Wavelet estimators: adapting to unknown smoothness", *Math. Methods of Statistics*, no. 1, pp. 1-20, 1997.
- [6] G.M.P. van Kempen, L.J. Van Vliet, P.J. Verveer, H.T.M. van der Voort, "A quantitative comparison of image restoration methods for confocal microscopy", *J. Microscopy*, vol. 185, no. 3, pp. 354-365, 1997.
- [7] G.M.P. van Kempen, H.T.M. van der Voort, J.G.J. Bauman, K.C. Strasters, "Comparing maximum likelihood estimation and constrained Tikhonov-Miller restoration", *IEEE. Eng. Med. Biol.*, vol. 15, pp. 76-83, 1996.
- [8] O. Lepski, "Asymptotically minimax adaptive estimation 1: upper bounds", *SIAM J. Th. Prob. Appl.*, vol. 36, no. 4, pp. 654-659, 1991.
- [9] M. Maurizot, P. Bouthemy, B. Delyon, A. Iouditski, J.-M. Odobez, "Determination of singular points in 2D deformable flow fields", in *IEEE Int. Conf. Image Processing*, Washington DC, 1995.
- [10] J.B. Pawley, "Fundamental limits in confocal microscopy", *Handbook of Biological Confocal Microscopy* (ed. by J.B. Pawley), pp. 15-26, Plenum Press, New-York, 1990.
- [11] J. Polzehl, V. Spokoiny, "Adaptive weights smoothing with application to image restoration", *J. R. Stat. Soc. B.*, vol. 62, no. 2, pp. 335-354, 2000.
- [12] D.L. Snyder, M.I. Miller, *Random Point Process in Time and Space*, Springer Verlag, Berlin, 1991.
- [13] J.L. Starck, F. Murtagh, A. Bijaoui, *Image Processing and Data Analysis, the Multiscale Approach*, Cambridge University Press, 2000.