Poisson Image Denoising Using Best Linear Prediction: A Post-processing Framework

Milad Niknejad, Mário A.T. Figueiredo

Instituto de Telecomunicações, and

Instituto Superior Técnico, Universidade de Lisboa, Portugal

Abstract—In this paper, we address the problem of denoising images degraded by Poisson noise. We propose a new patch-based approach based on best linear prediction to estimate the underlying clean image. A simplified prediction formula is derived for Poisson observations, which requires the covariance matrix of the underlying clean patch. We use the assumption that similar patches in a neighborhood share the same covariance matrix and we use off-the-shelf Poisson denoising methods in order to obtain an initial estimate of these covariance matrices. Our method can be seen as a post-processing step for other Poisson denoising methods and the results show that it improves upon them by relevant margins.

I. INTRODUCTION

Image denoising under Poissonian observations is an important problem, appearing in several application areas such as astronomy and medical imaging, where photon-limited images are very common.

In Poisson denoising, each pixel in the noisy image is a realization of a Poisson random variable with expected value equal to the true underlying pixel to be estimated. Some Poisson image denoising methods apply a nonlinear *variance stabilization transform* (VST), such as the Anscombe transform [1], to the noisy image, in order to approximately transform the noise into Gaussian-distributed. The resulting image is then processed using a denoising method for Gaussian noise; finally, an inverse of the transform is applied to the output of the denoising method, in order to obtain the final image estimate. Although some methods have been proposed to improve the accuracy of this inverse transform [2], VST-based methods are less accurate in regimes of low *signal-to-noise ratios* (SNR), which has led to the development of methods that directly tackle the Poissonian image. Two recent examples of such denoising methods are *Poisson non-local means* (PNLM) [3] and *non-local PCA* (NL-PCA) [4] (more references to earlier work can be found in [3] and [4]). Approaches that directly deal with the Poissonian nature of the observations have also been proposed for image deconvolution under Poisson noise (see [5], [6], and references therein).

The underlying assumption in many image restoration methods is that the clean patches lie in a low dimensional subspace. Whereas in the Gaussian noise case, this assumption can be used efficiently [7], [8], in the Possionian case, applying this assumption as a prior leads to an intractable optimization problem due to the non-quadratic log-likelihood. To circumvent this difficulty, some methods, such as NL-PCA and the sparsity-based method in [9], assume the patches are represented as the exponential of some data lying in a low-dimensional subspace. However, the exponential of low-dimensional data is not guaranteed to be low-dimensional. Recently, approaches based on sampling have been proposed to directly approximate the *minimum mean squared error* (MMSE) estimate [10]. Although these approaches are generally computationally complex for generic images, it has been shown that it can be effectively applied in a class-specific setting [11].

In this paper, we propose a new method based on the *best linear prediction* (BLP) principle [12], in order to estimate the clean patches as an affine transformation of the noisy ones. This estimator is independent of the noise distribution and it requires merely the covariance matrix of the underlying clean vector. We assume that similar patches in a neighborhood of a reference patch share the same covariance matrix. In order to estimate this covariance matrix, we use an off-the-shelf Poisson denoising algorithm. Thus, our method can be seen as a post-processing step for Poisson denoising methods. The experimental results reported in this paper show that our method improves over state-of-the-art denoising methods by relevant margins.

In the following sections, we first review the BLP principle; then, the proposed method is described. Finally, experimental results are reported and conclusions are drawn.

II. BEST LINEAR PREDICTION – BLP

In this paper, we use uppercase normal font to indicate random vectors and uppercase bold font for deterministic matrices. Consider that a random vector $X \in \mathbf{R}^n$ with unknown probability density function is to be estimated from an observed random vector $Y \in \mathbf{R}^m$ as an affine function with the form

$$\widehat{X} = \mathbf{B}Y + \mathbf{a}, \tag{1}$$

where $\mathbf{B} \in \mathbf{R}^{n \times m}$ and $\mathbf{a} \in \mathbf{R}^n$ are a fixed matrix and vector, respectively. Let $\boldsymbol{\mu}_y$ and $\boldsymbol{\mu}_x$ be the mean of the random variables Y and X, respectively. Based on the BLP theorem [12], the optimal choice for $\mathbf{B}$ and $\mathbf{a}$ in the MMSE sense is

$$\mathbf{B} = \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}, \ \mathbf{a} = \boldsymbol{\mu}_x - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\mu}_y \tag{2}$$

where $\boldsymbol{\Sigma}_{xy}$ and $\boldsymbol{\Sigma}_{yy}$ are the cross-covariance matrix between X and Y, and the covariance matrix of Y, respectively. A

remarkable feature of BLP is that it is independent of the distribution of the random variables involved, depending only on the aforementioned matrices and mean vectors. In the case of Gaussian noise, it can easily be shown that BLP is equivalent to using a multivariate Gaussian prior with covariance $\boldsymbol{\Sigma}_{xx}$, which is often used in image denoising [8], [13]; in that case, (2) is not only the best *linear* predictor, but the best predictor overall in the MMSE sense [14].

In this paper, we propose to use BLP (1)–(2) in the case where Y is a Poisson random vector with the underlying mean X ($\mathbb{E}[Y|X] = X$), that is, for $m = n$ and

$$\mathbb{P}(Y = \mathbf{y}|X = \mathbf{x}) = \prod_{j=1}^{n} \frac{e^{-x_j} x_j^{y_j}}{y_j!}, \qquad (3)$$

where $\mathbf{y} \in \mathbb{N}_0^n$ and $\mathbf{x} \in \mathbb{R}_+^m$. The main challenge in this approach will be to obtain the aforementioned covariance matrices and mean vectors, as discussed in the next section.

III. THE PROPOSED METHOD

A. BLP from Poisson Observations

The Poisson conditional distribution of Y in (3) has an important implication in the cross-covariance matrix $\boldsymbol{\Sigma}_{xy}$. Recall that the cross-covariance matrix is given by

$$\boldsymbol{\Sigma}_{xy} = \mathbb{E}_{X,Y}[XY^T] - \boldsymbol{\mu}_x \boldsymbol{\mu}_y^T, \qquad (4)$$

where $\mathbb{E}_{X,Y}$ indicates the expectation with respect to the joint distribution of X and Y. Using the so-called *law of iterated expectation*, the above can be written as

$$\begin{aligned} \boldsymbol{\Sigma}_{xy} &= \mathbb{E}_X \big[\mathbb{E}_{Y|X}[XY^T] \big] - \boldsymbol{\mu}_x \boldsymbol{\mu}_y^T \\ &= \mathbb{E}_X (XX^T) - \boldsymbol{\mu}_x \boldsymbol{\mu}_x^T \\ &= \boldsymbol{\Sigma}_{xx}, \end{aligned} \qquad (5)$$

because $\boldsymbol{\mu}_y = \boldsymbol{\mu}_x$ and $\mathbb{E}_{Y|X}[XY^T] = X\mathbb{E}_{Y|X}[Y^T] = XX^T$. In conclusion (as is also true for any zero-mean additive independent noise), in the case of Poissonian observations (which is not additive noise), we have that $\boldsymbol{\Sigma}_{xy} = \boldsymbol{\Sigma}_{xx}$.

A particular property of the Poisson distribution (namely that its mean and variance are equal) underlies a simple relationship between $\boldsymbol{\Sigma}_{yy}$ and $\boldsymbol{\Sigma}_{xx}$. A similar relationship exists in the case of additive independent zero-mean noise with known covariance[1], but for the Poisson model (3), no further information is needed. Using iterated expectation again,

$$\begin{aligned} \boldsymbol{\Sigma}_{yy} &= \mathbb{E}_Y (YY^T) - \boldsymbol{\mu}_y \boldsymbol{\mu}_y^T \\ &= \mathbb{E}_X \big[\mathbb{E}_{Y|X}[YY^T] \big] - \boldsymbol{\mu}_x \boldsymbol{\mu}_x^T. \end{aligned} \qquad (6)$$

Conditioned on X, the components of Y are independent, thus the off-diagonal elements of $\mathbb{E}_{Y|X}[YY^T]$ are simply

$$(i \neq j) \;\Rightarrow\; \big(\mathbb{E}_{Y|X}[YY^T] \big)_{i,j} = \mathbb{E}[Y_i|X]\, \mathbb{E}[Y_j|X] = X_i \, X_j.$$

Concerning the diagonal elements,

$$\big(\mathbb{E}_{Y|X}[YY^T] \big)_{i,i} = \mathbb{E}_{Y|X}[Y_i^2] = X_i + X_i^2,$$

[1]For example, for additive zero-mean white noise of known variance σ^2, we have the well-known relationship $\boldsymbol{\Sigma}_{yy} = \boldsymbol{\Sigma}_{xx} + \sigma^2 \mathbf{I}$.

because the mean and variance of a Poisson random variable are identical. Consequently,

$$\mathbb{E}_{Y|X}[YY^T] = XX^T + \mathrm{diag}(X_1, ..., X_n)$$

and

$$\begin{aligned} \mathbb{E}_X \big[\mathbb{E}_{Y|X}[YY^T] \big] &= \mathbb{E}_X[XX^T] + \mathbb{E}_X[\mathrm{diag}(X_1, ..., X_n)] \\ &= \boldsymbol{\Sigma}_{xx} + \boldsymbol{\mu}_x \boldsymbol{\mu}_x^T + \mathrm{diag}(\boldsymbol{\mu}_x), \end{aligned} \qquad (7)$$

where $\mathrm{diag}(\boldsymbol{\mu}_x)$ denotes a diagonal matrix with the components of $\boldsymbol{\mu}_x$. Finally, plugging (7) into (6), yields

$$\boldsymbol{\Sigma}_{yy} = \boldsymbol{\Sigma}_{xx} + \mathrm{diag}(\boldsymbol{\mu}_x), \qquad (8)$$

which is a natural result, since Poisson noise can be seen as additive independent noise with variance equal to the underlying clean value.

Finally, the BLP (1) of X from Poisson observations Y modeled by (3) is given by

$$\widehat{X} = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xx}(\mathrm{diag}(\boldsymbol{\mu}_x) + \boldsymbol{\Sigma}_{xx})^{-1}(Y - \boldsymbol{\mu}_y). \qquad (9)$$

B. Application to Patch-Based Poisson Denoising

As in classical patch-based methods, the image is divided into overlapping $\sqrt{n} \times \sqrt{n}$ patches [3], [15], [16]. The patches are denoised in collaboration with similar patches and then returned to the original position in the image, with the multiple estimates of each pixel (due to the overlapping nature of the patches) averaged to reconstruct the final clean image estimate. Denoting the (vectorized) i-th noisy patch as $\mathbf{y}_i$, we estimate the corresponding clean patch $\mathbf{x}_i$ by BLP, *i.e.*, using (9),

$$\widehat{\mathbf{x}}_i = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xx}(\mathrm{diag}(\boldsymbol{\mu}_x) + \boldsymbol{\Sigma}_{xx})^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_y). \qquad (10)$$

where $\boldsymbol{\mu}_x$ and $\boldsymbol{\Sigma}_{xx}$ are estimated as describe in the next paragraphs.

Some methods for Gaussian denoising use the same underlying covariance matrix for similar patches in the image [8], [13]. These approaches can all be seen as inspired by the collaborative filtering idea pioneered in the famous BM3D denoising method [16]. In this paper, we follow [8], by assuming the same covariance matrix $\boldsymbol{\Sigma}_{xx}$ for patches that are similar to a reference patch $\mathbf{y}_r$ in a neighborhood thereof.

Estimating $\boldsymbol{\mu}_x$ and $\boldsymbol{\Sigma}_{xx}$ from a set of clean patches would simply amount to computing the sample mean and sample covariance. In the presence of additive zero-mean white Gaussian noise of known variance σ^2, it would still be possible to obtain estimates $\boldsymbol{\mu}_x$ and $\boldsymbol{\Sigma}_{xx}$ from noisy patches [17]. However, this is not so simple in the case of Poissonian observations, because the variance of each observation depends on the underlying clean value, which is of course unknown. Here, we propose to use an off-the-shelf Poisson denoising methods to obtain an initial estimate of the clean image, from which $\boldsymbol{\mu}_x$ and $\boldsymbol{\Sigma}_{xx}$ can be estimated; these estimates are then plugged into (10) to obtain the final patch estimator.

C. Proposed Algorithm

Let **y** denote the whole Poisson noisy image. The first step is to use an off-the-shelf Poisson denoising method to obtain a so-called *pilot estimate* $\tilde{\mathbf{x}}$. In the experiments reported below, we will use NL-PC [4], VST+BM3D, and the recent state-of-the-art method in [18].

The remaining steps of the algorithm follow closely the macro structure of BM3D [16], [19]. Denoting a reference patch of the pilot estimate as $\tilde{\mathbf{x}}_r$, this image is searched within a window of size $N \times N$, centered at $\tilde{\mathbf{x}}_r$, for the k-nearest patches in Euclidean distance; let $\mathcal{X}_r$ denote the set of patches thus obtained. From this set of patches, their sample mean and sample covariance are obtained and denote as $\boldsymbol{\mu}_x$ and $\boldsymbol{\Sigma}_{xx}$.

Let $\mathcal{Y}_r$ denote the set of patches in the noisy image at the same locations as the patches in $\mathcal{X}_r$. Using $\boldsymbol{\mu}_x$ and $\boldsymbol{\Sigma}_{xx}$ obtained as explained in the previous paragraph, all the patches in $\mathcal{Y}_r$ are denoised by BLP, according to (10). These denoised patches are then returned to their locations, and averaged wherever they overlap.

The algorithm may be repeated L times, using the obtained denoised image as the next pilot estimate.

It is clear from the description that, rather than a full self-contained denoising method, our algorithm can be seen as a post-processing step for other Poisson denoising methods. As shown below in the experiments, the proposed approach improves over several state-of-the-art Poisson denoising algorithms by a non-negligible margin.

IV. RESULTS

In this section, we evaluate the performance of our algorithm using different methods to obtain the *pilot estimate*: NL-PCA [4], VST+BM3D [2], and a recent state-of-the-art method in [18]. The parameters used in the proposed method are as follows. Reference patches are selected with step-size 4 along both the row and the columns of the image. The patches have size 8×8 and the search window is 40×40. The number of selected similar patches to each reference patch is set to $k = 30$. Finally, we use $L = 2$ iterations of the algorithm.

Table I, II, and III show the results of the proposed method for some benchmark images, in comparison to the methods used to obtain the corresponding pilot estimates. It can be seen that, in all the examples, the BLP improves the pilot methods for every tested images. The improvements often exceed 0.5 dB. The average improvement of our method is also noticeable. In these tables, mainly low SNR values, with peak values from 1 to 10 are considered. Although the peak value of 10 in images is sometimes not considered as a low SNR, it is included to show the ability of our method to improve the higher SNR's even when VST methods are quite accurate in these ranges.

Some examples of the proposed denoising methods using different initialization are shown and compared to the corresponding initialized images in Figures 1, 2, and 3. It can be seen that the proposed technique is able to reduce some of the artifacts produced by the underlying initialization method, yielding visually more pleasing results.

Fig. 1. Example of denoising results of a part of the image Cameraman (a) Noisy image (Peak=5), (b) VST+BM3D (PSNR=23.14) (c) Proposed method initialized by VST+BM3D (PSNR=23.51)

Fig. 2. Example of denoising of a part of the image Barbara, (a) Noisy image (Peak=15), (b) VST+BM3D (PSNR=26.29), (c) VST+BM3D+BLP (PSNR=26.61)

TABLE I
PSNR/ SSIM RESULTS OF THE PROPOSED METHOD IN COMPARISON WITH NLPCA [4]

	peak=2		peak=5		peak=10	
	NLPCA	NLPCA+BLP	NLPCA	NLPCA+BLP	NLPCA	NLPCA+BLP
Cameraman	20.84 / .9984	**21.21 / .9986**	20.69 / .9921	**21.56 / .9941**	20.82 / .9795	**22.02 / .9861**
House	22.64 / .9984	**23.40 / .9987**	23.42 / .9960	**24.29 / .9965**	23.97 / .9931	**25.14 / .9948**
Barbara	21.47 / .9986	**21.80 / .9988**	21.83 / .9954	**22.19 / .9960**	22.02 / .9885	**22.52 / .9900**
Lena	23.64 / .9988	**24.30 / .9989**	24.70 / .9968	**25.35 / .9973**	24.99 / .9935	**25.81 / .9947**
Average	22.14 / .9985	**22.68 / .9987**	22.66 / .9950	**23.35 / .9959**	22.95 / .9886	**23.87 / .9914**

TABLE II
PSNR/SSIM RESULTS OF THE PROPOSED METHOD IN COMPARISON WITH VST+BM3D [2]

	peak=2		peak=5		peak=10	
	VST+BM3D	VST+BM3D+BLP	VST+BM3D	VST+BM3D+BLP	VST+BM3D	VST+BM3D+BLP
Cameraman	22.04 / .9987	**22.32 / .9989**	24.40 / .9976	**24.79 / .9979**	26.09 / .9957	**26.51 / .9959**
House	24.04 / .9988	**24.81 / .9990**	26.86 / .9983	**27.43 / .9986**	28.63 / .9977	**29.15 / .9979**
Barbara	22.09 / .9988	**22.36 / .9989**	24.81 / .9977	**25.19/ .9979**	26.60 / .9960	**27.02 / .9964**
Lena	24.37 / .9990	**24.99 / .9991**	26.76 / .9983	**27.33 / .9985**	28.60 / .9973	**29.08 / .9976**
Average	23.13 / .9988	**23.62 / .9990**	25.70 / .9980	**26.18 / .9982**	27.48 / .9967	**27.94 / .9970**

TABLE III
PSNR/SSIM RESULTS OF THE PROPOSED METHOD IN COMPARISON WITH THE METHOD IN [18]

	peak=1		peak=2		peak=5		peak=10	
	[18]	[18]+BLP	[18]	[18]+BLP	[18]	[18]+BLP	[18]+BLP	[18]+BLP
Cameraman	21.17 / .9966	**21.29 / .9966**	22.34 / .9990	**22.65 / .9991**	24.53 / .9977	**25.19 / .9982**	26.38 / .9961	**26.88 / .9966**
House	23.56 / .9996	**23.74 / .9997**	24.74 / .9992	**25.15 / .9993**	26.78 / .9984	**27.74 / .9988**	28.76 / .9977	**29.80 / .9982**
Barbara	21.37 / .9996	**21.56 / .9996**	22.33 / .9990	**22.51 / .9991**	24.57 / .9977	**25.12 / .9981**	26.52 / .9960	**27.31 / .9967**
Lena	24.01 / .9997	**24.32 / .9997**	25.27 / .9992	**25.77 / .9993**	27.07 / .9983	**28.01 / .9987**	28.54 / .9974	**29.59 / .9979**
Average	22.53 / .9989	**22.73 / .9989**	23.67 / .9991	**24.02/ .9992**	25.73 / .9980	**26.51 / .9985**	27.55 / .9968	**28.39 / .9974**
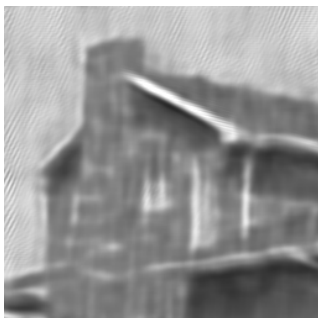
Fig. 3. Example of denoising results of the House image, (a) Noisy image (peak=2), (b) NLPCA (PSNR=24.04), (c) NLPCA+BLP (PSNR=24.84)

V. CONCLUSION

In this paper, we have proposed a new approach based on *best linear prediction* (BLP) for denoising images contaminated with Poisson noise. BLP is independent of the noise distribution and we showed that, in the Poisson noise case, the only required information is the mean and covariance matrix of the underlying vector being estimated. To implement the BLP approach, our method relies on an off-the-shelf Poisson denoising method in order to obtain a pilot estimate, from which the required means and covariances are estimated. Consequently, the proposed method can be seen as a post-processing step for Poisson denoising methods. The experiments reported show that our method is able to obtain a noticable improvement over the denoised images obtained by several state-of-the-art methods.

REFERENCES

[1] F. Anscombe, "The transformation of Poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 3/4, pp. 246–254, 1948.

[2] M. Mäkitalo and A. Foi, "Optimal inversion of the Anscombe transformation in low-count Poisson image denoising," *IEEE Transactions on Image Processing*, vol. 20, no. 1, pp. 99–109, 2011.

[3] C. Deledalle, F. Tupin, and L. Denis, "Poisson nl means: Unsupervised non local means for poisson noise," in *Image processing (ICIP), 2010 17th IEEE int. conf. on.* IEEE, 2010, pp. 801–804.

[4] J. Salmon, Z. Harmany, C. Deledalle, and R. Willett, "Poisson noise reduction with non-local PCA," *Journal of Mathematical Imaging and Vision*, vol. 48, no. 2, pp. 279–294, 2014.

[5] M.A.T. Figueiredo and J.M. Bioucas-Dias, "Restoration of Poissonian images using alternating direction optimization," *IEEE Transactions on Image Processing*, vol. 19, pp. 3133 –3145, 2010.

[6] Z. Harmany, R. Marcia, and R. Willett, "This is SPIRAL-TAP: Sparse Poisson intensity reconstruction algorithmstheory and practice," *IEEE Transactions on Image Processing*, vol. 21, pp. 1084–1096, 2012.

[7] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2862–2869.

[8] M. Niknejad, H. Rabbani, and M. Babaie-Zadeh, "Image restoration using Gaussian mixture models with spatially constrained patch clustering," *IEEE Transactions on Image Processing*, vol. 24, pp. 3624–3636, 2015.

[9] A. Rond, R. Giryes, and M. Elad, "Poisson inverse problems by the plug-and-play scheme," *Journal of Visual Communication and Image Representation*, vol. 41, pp. 96–108, 2016.

[10] S. Pyatykh and J. Hesser, "MMSE estimation for Poisson noise removal in images," *arXiv:1512.00717*, 2015.

[11] M. Niknejad, J. Bioucas-Dias, and M. A. T. Figueiredo, "Class-specific Poisson denoising by patch-based importance sampling," in *IEEE Intern. Conf. Image Processing (ICIP)*, 2017.

[12] S. Searle, G. Casella, and C. McCulloch, *Variance components*, John Wiley & Sons, 2009.

[13] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From Gaussian mixture models to structured sparsity," *IEEE Transactions on Image Processing*, vol. 21, pp. 2481–2499, 2012.

[14] L. Scharf and C. Demeure, *Statistical signal processing: detection, estimation, and time series analysis*, Addison-Wesley, 1991.

[15] Antoni Buades, Bartomeu Coll, and J-M Morel, "A non-local algorithm for image denoising," in *2005 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2005, vol. 2, pp. 60–65.

[16] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, pp. 2080–2095, 2007.

[17] A. Teodoro, M. Almeida, and M. Figueiredo, "Single-frame image denoising and inpainting using Gaussian mixtures," in *4th International Conference on Pattern Recognition Applications and Methods*, 2015.

[18] L. Azzari and A. Foi, "Variance stabilization for noisy+estimate combination in iterative Poisson denoising," *IEEE Signal Processing Petters*, vol. 23, pp. 1086–1090, 2016.

[19] M. Lebrun, "An analysis and implementation of the BM3D image denoising method," *Image Processing Online*, vol. 2, 2012.