

Optimal inversion of the Anscombe transformation in low-count Poisson image denoising

Markku Mäkitalo and Alessandro Foi

Abstract—The removal of Poisson noise is often performed through the following three-step procedure. First, the noise variance is stabilized by applying the Anscombe root transformation to the data, producing a signal in which the noise can be treated as additive Gaussian with unitary variance. Second, the noise is removed using a conventional denoising algorithm for additive white Gaussian noise. Third, an inverse transformation is applied to the denoised signal, obtaining the estimate of the signal of interest.

The choice of the proper inverse transformation is crucial in order to minimize the bias error which arises when the nonlinear forward transformation is applied. We introduce optimal inverses for the Anscombe transformation, in particular the exact unbiased inverse, a maximum likelihood (ML) inverse, and a more sophisticated minimum mean square error (MMSE) inverse. We then present an experimental analysis using a few state-of-the-art denoising algorithms and show that the estimation can be consistently improved by applying the exact unbiased inverse, particularly at the low-count regime. This results in a very efficient filtering solution that is competitive with some of the best existing methods for Poisson image denoising.

Index Terms—denoising, photon-limited imaging, Poisson noise, variance stabilization.

I. INTRODUCTION

Poisson noise is characteristic of many image acquisition modalities, and its removal is of fundamental importance for many applications and particularly in astronomy and medical imaging. As the noise variance equals the expected value of the underlying true signal, Poisson noise is signal dependent, which makes the premise for Poisson denoising very different from the case of additive white Gaussian noise with constant variance typically assumed by signal processing filters.

Although denoising algorithms specifically designed for Poisson noise have been proposed (e.g., [1], [2], [3], [4], [5], [6], [7]), often the removal of Poisson noise is performed through the following three-step procedure. First, the noise variance is stabilized by applying the Anscombe root transformation [8] $f : z \mapsto 2\sqrt{z + \frac{3}{8}}$ to the data. This produces a signal in which the noise can be treated as additive Gaussian with unitary variance. Second, the noise is removed using a conventional denoising algorithm for additive white Gaussian noise. Third, an inverse transformation is applied to the denoised signal, obtaining the estimate of the signal of interest.

This paper focuses on this last step and aims at identifying and emphasizing the role that the inversion plays in ensuring the success of the whole procedure.

This work was supported by the Academy of Finland (project no. 213462, Finnish Programme for Centres of Excellence in Research 2006-2011, project no. 118312, Finland Distinguished Professor Programme 2007-2010, and project no. 129118, Postdoctoral Researcher's Project 2009-2011).

In the recent years, variance stabilization has often been questioned as a viable method for Poisson noise removal because of the poor numerical results achieved at the low-count regime, i.e. for low-intensity signals, which corresponds to the case of low signal-to-noise ratio (SNR). We show that this disappointing performance, reported in many earlier works (e.g., [1], [2], [3]), is not due to the stabilization itself (i.e. to the forward transformation), but rather to the inverse transformation.

The choice of the proper inverse transformation is crucial in order to minimize the bias error which arises when the nonlinear forward transformation is applied. Both the algebraic inverse and the asymptotically unbiased inverse proposed by Anscombe [8] lead to a significant bias at low counts. In particular, the latter inverse provides unbiasedness only asymptotically for large counts while at low counts it leads to a larger bias than the former one.

This work extends our preliminary paper [9] by considering more general optimal inverses for the Anscombe transformation. First we introduce the exact unbiased inverse and show that it coincides with a form of maximum likelihood (ML) inverse, and then we consider a more sophisticated minimum mean square error (MMSE) inverse. After that, we present an extensive experimental analysis using a few state-of-the-art denoising algorithms and show that the results can be consistently improved by applying the exact unbiased inverse. In particular, the combination of BM3D [10] and the exact unbiased inverse outperforms some of the best existing algorithms specifically targeted at Poisson noise removal, while maintaining low computational complexity.

The rest of the paper is organized as follows: Section II introduces some preliminaries about Poisson noise, variance stabilization and the conventional inverses of the Anscombe transformation. In Section III we consider optimal inverse transformations: first we propose the exact unbiased inverse, which can be interpreted as a maximum likelihood inverse, and finally we discuss the minimum mean square error inverse. Section IV consists of various experiments, followed by discussion and conclusions in Section V.

II. PRELIMINARIES

A. Poisson noise

Let $z_i, i = 1, \dots, N$, be the observed pixel values obtained through an image acquisition device. We consider each z_i to be an independent random Poisson variable whose mean $y_i \geq 0$ is the underlying intensity value to be estimated. Explicitly,

the discrete Poisson probability of each z_i is

$$P(z_i | y_i) = \frac{y_i^{z_i} e^{-y_i}}{z_i!}. \quad (1)$$

In addition to being the mean of the Poisson variable z_i , the parameter y_i is also its variance:

$$E\{z_i | y_i\} = y_i = \text{var}\{z_i | y_i\}. \quad (2)$$

Poisson noise can be formally defined as

$$\eta_i = z_i - E\{z_i | y_i\}, \quad (3)$$

thus, we trivially have $E\{\eta_i | y_i\} = 0$ and $\text{var}\{\eta_i | y_i\} = \text{var}\{z_i | y_i\} = y_i$. Since the noise variance depends on the true intensity value, Poisson noise is signal dependent. More specifically, the standard deviation of the noise η_i equals $\sqrt{y_i}$. Due to this, the effect of Poisson noise increases (i.e. the signal-to-noise ratio decreases) as the intensity value decreases.

B. Variance stabilization and the Anscombe transformation

The rationale behind applying a variance-stabilizing transformation is to remove the data-dependence of the noise variance, so that it becomes constant throughout the whole data z_i , $i = 1, \dots, N$. Moreover, if the transformation is also normalizing (i.e. it results in a Gaussian noise distribution), we can estimate the intensity values y_i with a conventional denoising method designed for additive white Gaussian noise. Neither exact stabilization nor exact normalization are possible [11], [12], therefore, in practice, approximate or asymptotical results are employed.

One of the most popular variance-stabilizing transformations is the Anscombe transformation [8]

$$f(z) = 2\sqrt{z + \frac{3}{8}}. \quad (4)$$

Applying (4) to Poisson distributed data gives a signal whose noise is asymptotically additive standard normal.

The denoising of $f(z)$ produces a signal D that can be considered as an estimate of $E\{f(z) | y\}$. We need to apply an inverse transformation to D in order to obtain the desired estimate of y . The direct algebraic inverse of (4) is

$$\mathcal{I}_A(D) = f^{-1}(D) = \left(\frac{D}{2}\right)^2 - \frac{3}{8}, \quad (5)$$

but the resulting estimate of y is biased, because the nonlinearity of the transformation f means we generally have

$$E\{f(z) | y\} \neq f(E\{z | y\}), \quad (6)$$

and, thus,

$$f^{-1}(E\{f(z) | y\}) \neq E\{z | y\}. \quad (7)$$

Another possibility is to use the adjusted inverse [8]

$$\mathcal{I}_B(D) = \left(\frac{D}{2}\right)^2 - \frac{1}{8}, \quad (8)$$

which provides asymptotical unbiasedness for large counts. This is the inverse typically used in applications.

III. OPTIMAL INVERSE TRANSFORMATIONS

While the asymptotically unbiased inverse (8) provides good results for high-count data, applying it to low-count data leads

to a biased estimate, as can be seen, e.g., in [1]. Here we consider three types of optimal inverses.

A. Exact unbiased inverse

Provided a successful denoising, i.e. D is treated as $E\{f(z) | y\}$, the exact unbiased inverse of the Anscombe transformation f is an inverse transformation \mathcal{I}_C that maps the values $E\{f(z) | y\}$ to the desired values $E\{z | y\}$:

$$\mathcal{I}_C : E\{f(z) | y\} \mapsto E\{z | y\}. \quad (9)$$

Since $E\{z | y\} = y$ for any given y , the problem of finding the inverse \mathcal{I}_C reduces to computing the values $E\{f(z) | y\}$, which is done by numerical evaluation of the integral corresponding to the expectation operator E :

$$E\{f(z) | y\} = \int_{-\infty}^{+\infty} f(z)p(z | y) dz, \quad (10)$$

where $p(z | y)$ is the generalized probability density function of z conditioned on y . In our case we have discrete Poisson probabilities $P(z | y)$, so we can replace the integral by summation:

$$E\{f(z) | y\} = \sum_{z=0}^{+\infty} f(z)P(z | y). \quad (11)$$

Further, since here $f(z)$ is the forward Anscombe transformation (4), we can write (11) as

$$E\{f(z) | y\} = 2 \sum_{z=0}^{+\infty} \left(\sqrt{z + \frac{3}{8}} \cdot \frac{y^z e^{-y}}{z!} \right). \quad (12)$$

Figure 1 shows the plots of the inverse transformations \mathcal{I}_A , \mathcal{I}_B and \mathcal{I}_C . Since \mathcal{I}_C is unbiased, we see that at low counts the asymptotically unbiased inverse actually leads to a larger bias than the algebraic inverse.

Let us remark that if the exact unbiased inverse (9) is applied to the denoised data D with some errors (in the sense that $D \neq E\{f(z) | y\}$), then the estimation error in $\hat{y} = \mathcal{I}_C(D)$ can include variance as well as bias components. In general, the unbiasedness of \mathcal{I}_C holds only provided that $D = E\{f(z) | y\}$ exactly, as it is assumed when defining (9).

B. ML inverse

In the previous Section III-A we assumed that the denoising is successful (i.e. we can treat the denoised signal D as $E\{f(z) | y\}$), which then lead us to the concept of the exact unbiased inverse. Now we consider a more general scenario, where this assumption does not necessarily hold: instead of the strict equality, we assume that the pointwise mean square error of D as an estimate of $E\{f(z) | y\}$ is

$$\varepsilon^2 = E\left\{(D - E\{f(z) | y\})^2\right\}. \quad (13)$$

In practice the distribution of D is unknown. For simplicity, we assume that D is normally distributed around $E\{f(z) | y\}$ with variance ε^2 :

$$D \sim \mathcal{N}\left(E\{f(z) | y\}, \varepsilon^2\right). \quad (14)$$

While formally (14) implies that D is an unbiased estimate of $E\{f(z) | y\}$, in fact also unknown estimation-bias errors can be considered as contributors of ε^2 , with the symmetry of the

distribution about $E\{f(z) | y\}$ reflecting our uncertainty about the sign of the bias.

By treating D as the data, the maximum likelihood (ML) inverse is defined as

$$\mathcal{I}_{\text{ML}}(D) = \arg \max_y p(D | y), \quad (15)$$

where, according to (14),

$$p(D | y) = \frac{1}{\sqrt{2\pi} \varepsilon^2} e^{-\frac{1}{2\varepsilon^2}(D - E\{f(z)|y\})^2}. \quad (16)$$

Under the above assumptions, this equals to (see Appendix for details)

$$\mathcal{I}_{\text{ML}}(D) = \begin{cases} \mathcal{I}_C(D), & \text{if } D \geq 2\sqrt{3/8} \\ 0, & \text{if } D < 2\sqrt{3/8}. \end{cases} \quad (17)$$

Thus, the exact unbiased inverse coincides with this form of ML inverse. Note also that $\mathcal{I}_{\text{ML}}(D)$ is independent of ε . The obtained result (17) holds for any unimodal distribution whose mode is $E\{f(z) | y\}$.

C. MMSE inverse

Under the same hypotheses of Section III-B, we define the minimum mean square error (MMSE) inverse, which is parametrized by ε , as

$$\begin{aligned} \mathcal{I}_{\text{MMSE}}(D, \varepsilon) &= \arg \min_{\hat{y}} E\{(y - \hat{y})^2 | D\} \\ &= \arg \min_{\hat{y}} \int_{-\infty}^{+\infty} p(y | D) (y - \hat{y})^2 dy. \end{aligned} \quad (18)$$

It is worth reminding that we assume D to be normally distributed according to (14); if this assumption does not hold, the obtained inverse is not necessarily the true minimum MSE inverse.

Additionally assuming that the true signal y is uniformly distributed, solving for \hat{y} (see Appendix for details) produces the following formula for computing the inverse:

$$\hat{y} = \frac{\int_0^{+\infty} p(D | y) y dy}{\int_0^{+\infty} p(D | y) dy}, \quad (19)$$

where $p(D | y)$ is given by (16). Note that the exact unbiased inverse can be considered a limit case of the MMSE inverse, obtained when $\varepsilon = 0$, because $p(D | y)$ becomes a Dirac impulse centered at that particular value of y such that $E\{f(z) | y\} = D$. In other words,

$$\mathcal{I}_{\text{MMSE}}(D, 0) = \mathcal{I}_C(D) = \mathcal{I}_{\text{ML}}(D). \quad (20)$$

Figure 2 shows the MMSE inverse transformations for some values of ε , including the case $\varepsilon = 0$.

IV. EXPERIMENTS

All of our experiments consist of the same three-step denoising procedure: First we apply the forward Anscombe transformation (4) to a noisy image. Then we denoise the transformed image (assuming additive white Gaussian noise of unit variance) with either BM3D [10], SAFIR [13] or BLS-GSM [14], and finally we apply an inverse transformation in order to get the final estimate. We do not provide comparisons against the Haar-Fisz algorithm [4] or platelets [6], [7], as both the MS-VST [1] and PURE-LET [3] algorithms have been

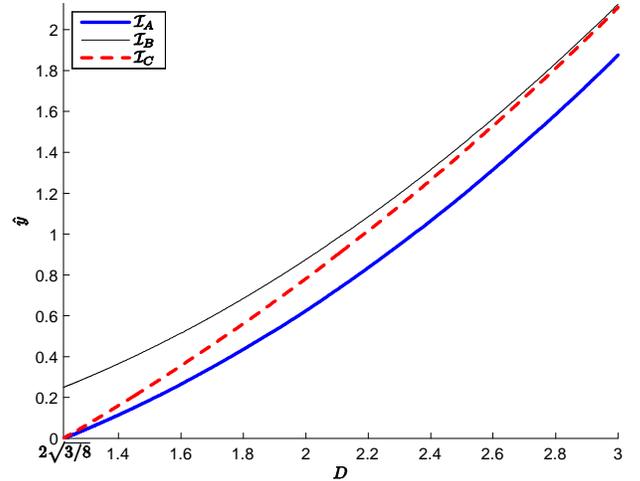


Fig. 1. Inverse Anscombe transformations \mathcal{I}_A (algebraic), \mathcal{I}_B (asymptotically unbiased) and \mathcal{I}_C (exact unbiased). For the exact unbiased inverse, D coincides with $E\{f(z) | y\}$, hence its bias is zero.

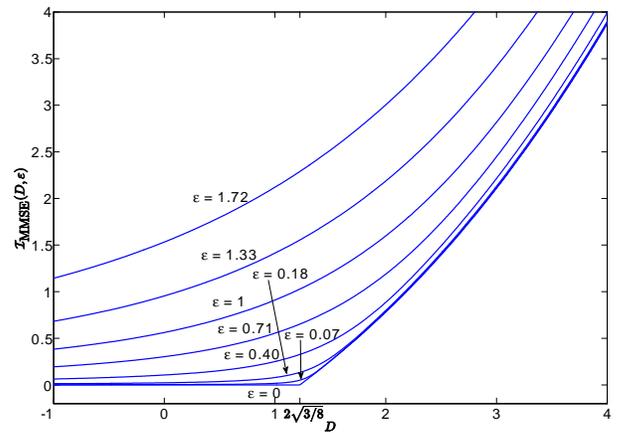


Fig. 2. MMSE inverse transformations (18) for some values of ε . The case $\varepsilon = 0$ corresponds to the exact unbiased inverse transformation (9) and the ML inverse transformation (17).

shown to outperform them, and we further compare these two algorithms against ours.

To implement the exact unbiased inverse \mathcal{I}_C in practice, it is sufficient to compute (12) for a limited set of values y ; for arbitrary values of y we then use linear interpolation based on these computed values of (12), and for large values¹ of y we approximate \mathcal{I}_C by \mathcal{I}_B . In similar fashion, the MMSE inverse can be obtained based on numerical evaluation of the two integrals in (19).

Matlab functions implementing these two optimal inverse transformations are available online at <http://www.cs.tut.fi/~foi/invansc>.

We evaluate the performance either by normalized mean integrated square error (NMISE) or by peak signal-to-noise

¹In our implementation, we consider y to be large if $y > 2500$.

ratio (PSNR). The NMISE is calculated using the formula

$$\frac{1}{\bar{N}} \sum_{i: y_i > 0} \left((\hat{y}_i - y_i)^2 / y_i \right), \quad (21)$$

where \hat{y}_i are the estimated intensities, y_i the respective true values, and the sum is computed over the \bar{N} pixels in the image for which $y_i > 0$. The PSNR is calculated using the formula

$$10 \log_{10} \left(\frac{\max_i (y_i)^2}{\left(\sum_i (\hat{y}_i - y_i)^2 / N \right)} \right), \quad (22)$$

where N is the total number of pixels in the image.

In Section IV-A we consider the exact unbiased inverse, and Section IV-B consists of experiments with the MMSE inverse. Section IV-C addresses the computational complexity of the inverse transformations and the denoising algorithms.

A. Exact unbiased inverse

We consider three sets of experiments in order to compare against the three recent works [1], [2] and [3], each of which proposes an algorithm specifically designed for Poisson noise removal (MS-VST, PH-HMT and Interscale PURE-LET, respectively).

1) *NMISE comparison against MS-VST [1] and PH-HMT [2]*: For the first set we proceed in the same way as in [1] in order to produce comparable results: The above-mentioned three-step denoising procedure is performed five times for each image, each time with a different realization of the random noise. We evaluate the performance by using NMISE, and the obtained NMISE values are finally averaged over these five replications. This metric was chosen because of the available results for comparison in [1] and [2]. The authors of [1] also kindly provided us with their set of test images (all of them 256×256 in size), shown in Figure 3.

The denoising is done with either BM3D, SAFIR or BLS-GSM, and for the inversion of the denoised signal we use the exact unbiased inverse. The same experiments are also done for the asymptotically unbiased inverse (8), whose results serve as a point of comparison.

The numerical results of our experiments are presented in Table I, where we also compare them to the state-of-the-art results obtained with the PH-HMT and MS-VST algorithms proposed in [2] and [1], respectively. In addition, we have included the results obtained in [1] with the asymptotically unbiased inverse Anscombe transformation combined with various undecimated wavelet transforms (here collectively denoted as WT). Table I shows not only that the exact unbiased inverse produces significantly better results at low counts than the asymptotically unbiased inverse, but also that the method is competitive with both PH-HMT and MS-VST. In particular, the combination of BM3D and the exact unbiased inverse outperforms both of them in terms of NMISE.

Figures 4–5 illustrate the improvement that is achieved (especially at low counts) by applying the exact unbiased inverse instead of the asymptotically unbiased inverse, while Figure 6 compares the different algorithms for the denoising of the Cells image (the exact unbiased inverse combined with

BM3D, SAFIR and BLS-GSM, and the best MS-VST result from [1]). In addition, we present a chosen cross-section (i.e. one row) of some of the test images in Figure 7. These plots also clearly demonstrate that at low counts the exact unbiased inverse provides a significant improvement over the asymptotically unbiased inverse, whereas at high counts the difference is expectedly negligible.

For additional figures we refer to our preliminary paper [9].

2) *PSNR comparison against PH-HMT [2]*: In the second part of the experiments we use the test images shown in Figure 8 and evaluate the performance in terms of PSNR, thus enabling us to compare against the PH-HMT results in [2]. This time we scale each image to seven different peak intensity levels (1, 2, 3, 4, 5, 10 and 20), and for each of them we perform the denoising procedure ten times, with ten different realizations of the random noise.

As above, we use either BM3D, SAFIR or BLS-GSM for the denoising, and the inversion is done with either the exact unbiased inverse or the asymptotically unbiased inverse.

The results, which are averages of ten PSNR values, are reported in Table II. We see again that at low peak intensities we get a substantial improvement by applying the exact unbiased inverse instead of the asymptotically unbiased inverse, regardless of the used denoising algorithm. Indeed, the best results of the table mainly correspond to algorithms combined with the exact unbiased inverse. In particular, the best overall performance is obtained with BM3D, although both SAFIR and PH-HMT provide competitive results especially at the lowest peak intensities. The different performance at low counts is possibly explained by SAFIR exploiting adaptive window sizes, as opposed to BM3D, which uses fixed-size blocks.

Note that the average PSNR values of the noisy images in Table II have minor differences to those reported in [2] due to different realizations of the random noise.

3) *PSNR comparison against PURE-LET [3]*: The third part of our experiments is very similar to the second one, with the following differences: Now we use the four test images shown in Figure 9 and compare our results against the best results obtained with the Interscale PURE-LET (with two cyclic shifts) [3]. Each image is scaled to the peak intensity levels 1, 5, 10, 20, 30, 60 and 120, so we do not focus at low counts as much as earlier.

As in [3], the denoising performance is evaluated in terms of PSNR. Table III presents the obtained results (averages of ten values), which are consistent with the results in Table II: the overall performance of BM3D is strong, but it is often outperformed by SAFIR at the lower peak intensity levels. It is interesting to note that for an image like Moon, which presents a large black background area that is completely flat, the performance gap between multiscale (BLS-GSM and PURE-LET) and patch-based (BM3D and SAFIR) methods is reduced, in as much as in a few cases the former methods are producing slightly better numerical results than the latter ones.

4) *Summary*: All three sets of experiments produce consistent results, showing that at low intensities we obtain significantly better results by applying the exact unbiased inverse instead of the asymptotically unbiased inverse, whereas at high

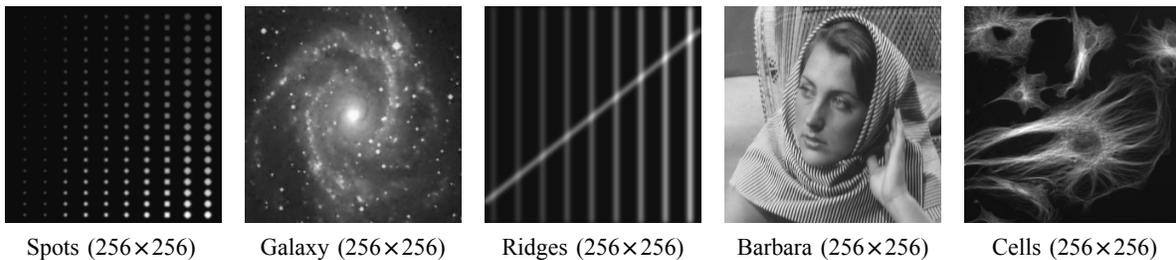


Fig. 3. The five test images used in the experiments of Sections IV-A1 and IV-B.

TABLE I

AVERAGE NMISE VALUES FOR THE ASYMPTOTICALLY UNBIASED INVERSE AND THE EXACT UNBIASED INVERSE, AND A COMPARISON TO THE RESULTS OBTAINED IN [2] AND [1] WITH ALGORITHMS SPECIFICALLY DESIGNED FOR POISSON NOISE REMOVAL. THE INTENSITY RANGE OF EACH IMAGE IS INDICATED IN BRACKETS.

	Asymptotically unbiased inverse				Exact unbiased inverse			Other algorithms	
	WT [1]	BM3D	SAFIR	BLS-GSM	BM3D	SAFIR	BLS-GSM	PH-HMT [2]	MS-VST [1]
Spots [0.03, 5.02]	2.34	1.7424	1.7495	2.0370	0.0365	0.0384	0.1871	0.048	0.069
Galaxy [0, 5]	0.15	0.1026	0.1110	0.1253	0.0299	0.0301	0.0385	0.030	0.035
Ridges [0.05, 0.85]	0.83	0.7025	0.7252	0.7694	0.0128	0.0173	0.0331	-	0.017
Barbara [0.93, 15.73]	0.26	0.0881	0.1178	0.1122	0.0881	0.1178	0.1123	0.159	0.17
Cells [0.53, 16.93]	0.095	0.0660	0.0683	0.0718	0.0649	0.0671	0.0707	0.082	0.078

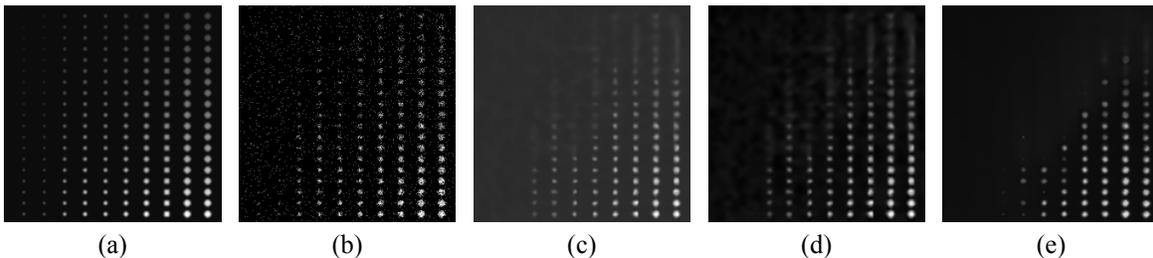


Fig. 4. (a) Original Spots image (intensity range [0.03, 5.02]), (b) Poisson-count image, (c) image denoised with BM3D and the asymptotically unbiased inverse (average NMISE = 1.7395), (d) image denoised with BM3D and the exact unbiased inverse (average NMISE = 0.0365), (e) image denoised with SAFIR and the exact unbiased inverse (average NMISE = 0.0384). The images shown here are gamma-corrected ($\gamma = 0.6$) for improved visibility of the darker areas. A cross-section of images (a), (c) and (d) is shown in Figure 7(a).

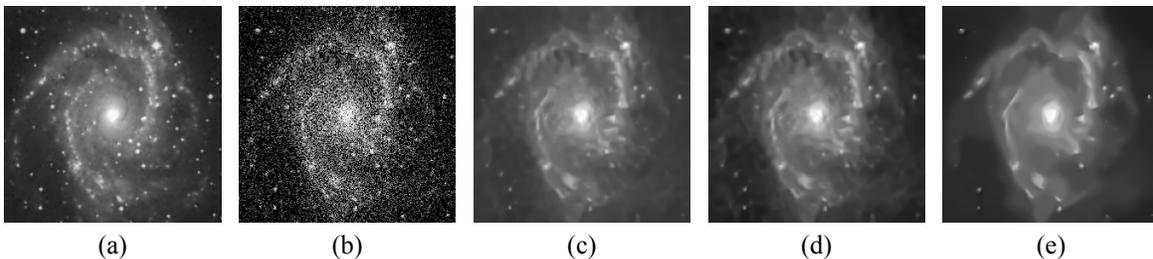


Fig. 5. (a) Original Galaxy image (intensity range [0, 5]), (b) Poisson-count image, (c) image denoised with BM3D and the asymptotically unbiased inverse (average NMISE = 0.1025), (d) image denoised with BM3D and the exact unbiased inverse (average NMISE = 0.0299), (e) image denoised with SAFIR and the exact unbiased inverse (average NMISE = 0.0301). The images shown here are gamma-corrected ($\gamma = 0.6$) for improved visibility of the darker areas. A cross-section of images (a), (c) and (d) is shown in Figure 7(b).

intensities there is expectedly no significant improvement.

The results also show that combined with a state-of-the-art Gaussian denoising algorithm, the exact unbiased inverse is competitive with some of the best algorithms targeted at Poisson noise removal.

B. MMSE inverse

Assuming that (14) is valid, the use of the MMSE inverse $\mathcal{I}_{\text{MMSE}}$ (18) requires knowledge of the pointwise mean square

error (13) for the estimate D produced by the denoising algorithm. In other words, for each pixel, a pair (D, ε) is used as an argument for $\mathcal{I}_{\text{MMSE}}$. First, in order to illustrate the full potential of this inverse, we show results obtained by employing an oracle estimate of the MSE computed by Monte-Carlo simulations. Second, as an example of the actual performance that can be achieved in practice, we compute an estimate of the MSE using Stein's unbiased risk estimate (SURE) [15].

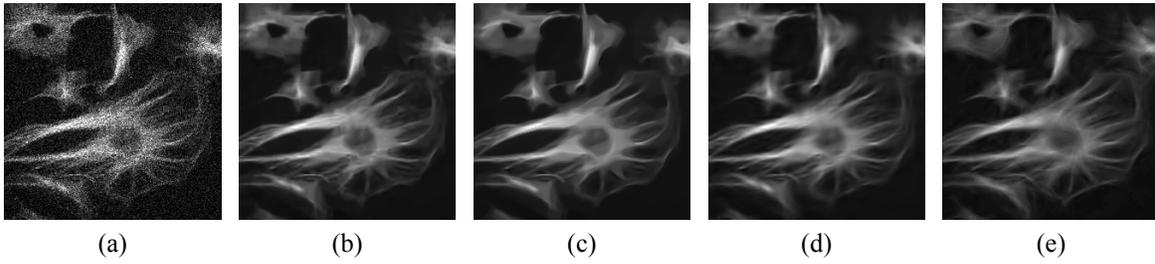


Fig. 6. (a) Noisy Cells image, denoised with (b) BM3D and the exact unbiased inverse (average NMISE = 0.0649), (c) SAFIR and the exact unbiased inverse (average NMISE = 0.0671), (d) BLS-GSM and the exact unbiased inverse (average NMISE = 0.0707), (e) MS-VST + curvelets (average NMISE = 0.078) [1]. The original image is shown in Figure 3.

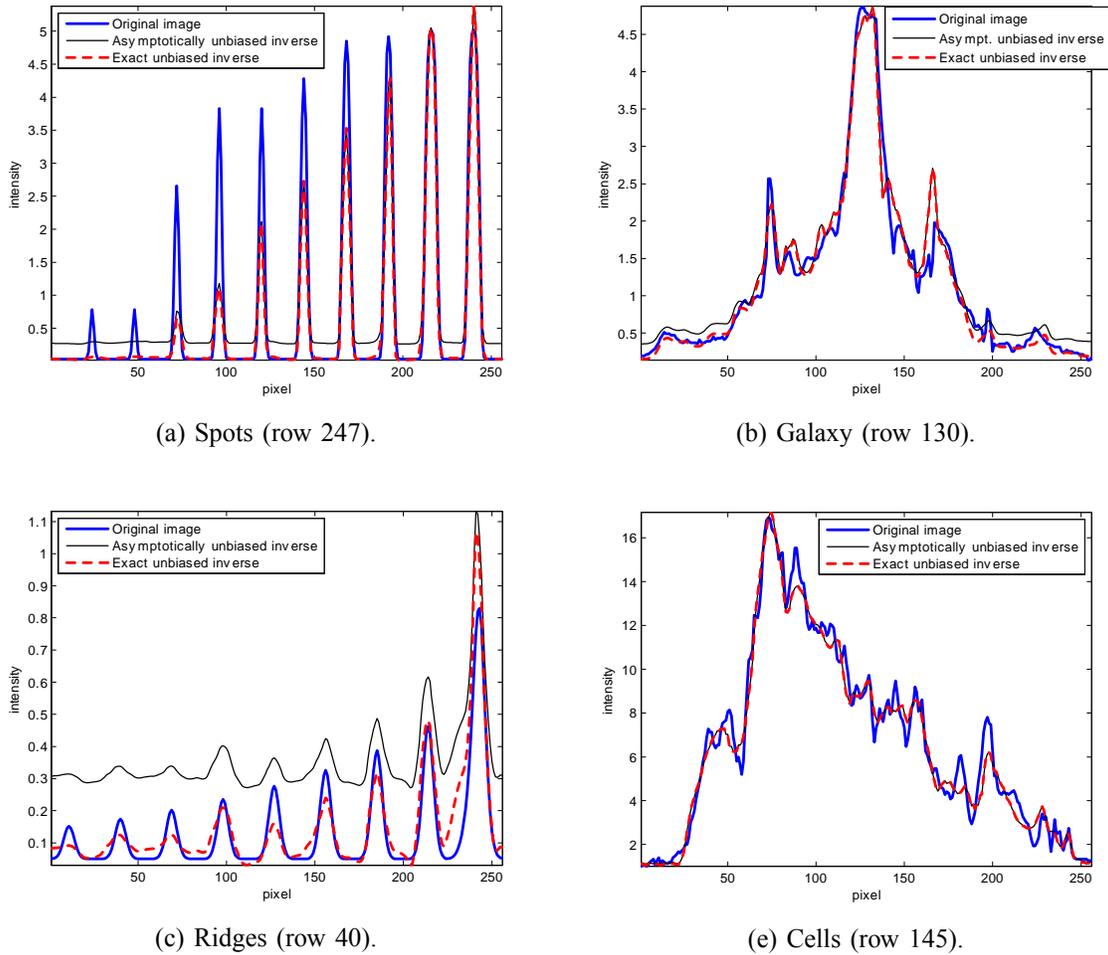


Fig. 7. Cross-sections of some of the images denoised with BM3D. For Cells the intensities are large enough for the two inverses to practically coincide.

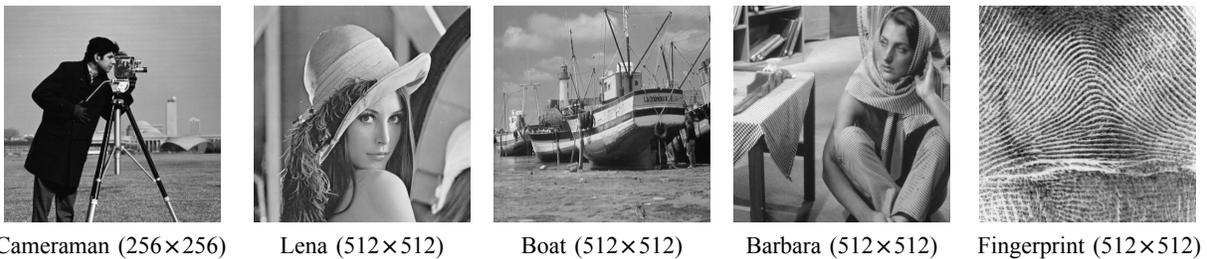


Fig. 8. The five test images used in the experiments of Section IV-A2.

TABLE II

AVERAGE PSNR VALUES (dB) FOR VARIOUS PEAK INTENSITIES FOR THE ASYMPTOTICALLY UNBIASED INVERSE AND THE EXACT UNBIASED INVERSE, AND A COMPARISON TO THE RESULTS OBTAINED IN [2] WITH PH-HMT, AN ALGORITHM SPECIFICALLY DESIGNED FOR POISSON NOISE REMOVAL.

Image	Peak	Noisy	Asymptotically unbiased inverse			Exact unbiased inverse			Other algorithms
			BM3D	SAFIR	BLS-GSM	BM3D	SAFIR	BLS-GSM	PH-HMT [2]
Cameraman	1	3.27	14.90	14.97	14.37	19.89	20.37	18.44	20.03
	2	6.26	20.49	20.21	18.94	22.10	21.88	20.11	21.41
	3	8.05	22.30	22.12	20.69	23.07	22.87	21.22	22.31
	4	9.28	23.40	23.25	21.67	23.86	23.71	21.98	22.90
	5	10.27	24.12	24.01	22.36	24.42	24.31	22.57	23.37
	10	13.26	26.03	25.83	24.52	26.08	25.89	24.57	24.97
	20	16.29	27.65	27.31	26.49	27.65	27.31	26.49	26.61
Lena	1	2.96	16.13	16.26	15.79	22.22	23.41	21.52	22.66
	2	5.99	22.88	23.00	21.98	24.07	24.77	23.42	23.91
	3	7.75	24.91	25.28	24.24	25.23	25.74	24.67	24.69
	4	9.00	25.96	26.32	25.43	26.06	26.45	25.59	25.29
	5	9.96	26.53	26.88	26.15	26.56	26.91	26.21	25.78
	10	12.97	28.31	28.51	28.03	28.31	28.51	28.03	27.21
	20	15.98	29.99	30.03	29.60	29.99	30.03	29.59	28.66
Boat	1	2.93	15.88	15.87	15.56	20.97	21.41	20.49	21.76
	2	5.95	21.80	21.57	20.90	22.74	22.76	21.98	22.77
	3	7.71	23.36	23.28	22.69	23.67	23.62	23.04	23.45
	4	8.96	24.15	24.07	23.63	24.28	24.20	23.79	23.90
	5	9.92	24.71	24.63	24.33	24.77	24.70	24.42	24.31
	10	12.94	26.27	26.06	25.93	26.28	26.07	25.94	25.57
	20	15.94	27.83	27.44	27.41	27.83	27.44	27.41	26.96
Barbara	1	3.20	15.28	15.33	14.95	20.43	20.78	19.64	20.48
	2	6.23	20.88	20.50	19.87	21.91	21.67	20.93	21.27
	3	7.99	22.74	21.88	21.31	23.07	22.18	21.62	21.72
	4	9.24	23.71	22.38	21.90	23.84	22.48	22.01	22.07
	5	10.21	24.42	22.68	22.57	24.48	22.72	22.62	22.33
	10	13.21	26.35	24.22	24.67	26.35	24.22	24.67	23.45
	20	16.22	28.18	26.91	26.50	28.18	26.91	26.50	24.92
Fingerprint	1	2.55	14.61	14.28	13.89	17.12	16.61	15.98	17.39
	2	5.57	19.55	19.23	18.87	19.86	19.59	19.36	18.55
	3	7.34	20.93	20.80	20.52	20.98	20.85	20.64	19.36
	4	8.58	21.68	21.54	21.34	21.69	21.54	21.39	19.94
	5	9.54	22.22	22.03	21.89	22.22	22.03	21.91	20.42
	10	12.56	23.80	23.35	23.38	23.80	23.35	23.38	21.91
	20	15.57	25.37	24.62	24.94	25.37	24.62	24.94	23.46



Peppers (256×256)



Cameraman (256×256)



MIT (256×256)



Moon (512×512)

Fig. 9. The four test images used in the experiments of Section IV-A3.

For both cases, due to space limitation, we present only the results corresponding to Table I produced using the BM3D algorithm.

1) *Oracle Monte-Carlo MSE*: Here we compute the mean square error (13) from 50 independent replications of the denoising experiment. This estimate, denoted as ε_{MC}^2 , is an oracle estimate which cannot obviously be produced if y is unknown. Note that very fine structures of the image are visible in ε_{MC} , as shown in the leftmost images in Figures 10–11. The corresponding MMSE estimate of y is obtained as $\mathcal{I}_{MMSE}(D, \varepsilon_{MC})$. The average NMISE and PSNR results over 5 independent replications of z are reported in Table IV.

For some images, the improvement is dramatic (up to almost 2.7 dB PSNR difference for Spots). However, for those images

where already the exact unbiased inverse did not provide substantial improvement over the asymptotically unbiased inverse (see Table I), the differences are much smaller. In the case of Barbara the results are even slightly worse. This can be attributed to the failure of the normal model (14) in describing the actual estimation errors for this particular image dominated by repeated texture.

2) *Empirical SURE estimate*: Stein's unbiased risk estimate (SURE) [15] can be used to provide a surrogate for the mean square error (13) for an arbitrary denoising algorithm without needing y to be known and without resorting to multiple realizations of the noise. Assuming that the noise corrupting $f(z)$ is zero-mean Gaussian with diagonal covariance matrix

TABLE III

AVERAGE PSNR VALUES (dB) FOR VARIOUS PEAK INTENSITIES FOR THE ASYMPTOTICALLY UNBIASED INVERSE AND THE EXACT UNBIASED INVERSE, AND A COMPARISON TO THE BEST RESULTS OBTAINED IN [3] WITH THE INTERSCALE PURE-LET, AN ALGORITHM SPECIFICALLY DESIGNED FOR POISSON NOISE REMOVAL.

Image	Peak	Noisy	Asymptotically unbiased inverse			Exact unbiased inverse			Other algorithms
			BM3D	SAFIR	BLS-GSM	BM3D	SAFIR	BLS-GSM	PURE-LET [3]
Peppers	1	3.16	15.20	15.30	14.67	19.98	20.34	18.44	19.33
	5	10.13	24.59	24.89	23.47	24.70	25.03	23.57	22.52
	10	13.15	26.41	26.49	25.39	26.43	26.51	25.41	24.29
	20	16.15	28.05	28.02	27.00	28.05	28.03	27.00	26.18
	30	17.92	29.05	28.93	27.97	29.05	28.93	27.97	27.27
	60	20.92	30.75	30.48	29.62	30.75	30.48	29.62	29.07
	120	23.94	32.47	32.10	31.36	32.47	32.10	31.36	30.79
Cameraman	1	3.27	14.90	14.97	14.37	19.89	20.37	18.44	19.67
	5	10.26	24.07	23.99	22.39	24.36	24.28	22.59	22.76
	10	13.26	26.05	25.87	24.52	26.11	25.93	24.57	24.32
	20	16.26	27.65	27.30	26.49	27.65	27.30	26.49	25.89
	30	18.06	28.56	28.14	27.51	28.56	28.14	27.50	26.87
	60	21.05	30.04	29.41	29.10	30.04	29.41	29.10	28.56
	120	24.07	31.66	30.76	30.78	31.66	30.76	30.78	30.36
MIT	1	5.00	13.41	13.14	12.28	19.17	17.90	15.67	17.82
	5	11.97	23.81	24.19	21.19	24.43	24.88	21.52	21.63
	10	14.98	25.99	26.20	23.32	26.16	26.37	23.41	23.49
	20	18.01	27.88	27.76	25.62	27.93	27.80	25.64	25.34
	30	19.78	28.96	28.64	26.84	28.98	28.65	26.84	26.55
	60	22.78	30.81	30.18	28.80	30.81	30.18	28.80	28.41
	120	25.77	32.63	31.89	30.71	32.63	31.89	30.71	30.47
Moon	1	5.46	13.96	14.03	13.98	22.64	23.34	23.08	23.19
	5	12.46	23.66	23.45	23.17	24.33	24.11	23.79	24.28
	10	15.48	24.85	24.58	24.21	25.05	24.76	24.38	24.99
	20	18.47	25.86	25.56	25.50	25.92	25.62	25.56	25.97
	30	20.24	26.50	26.18	26.35	26.53	26.21	26.38	26.70
	60	23.25	27.80	27.36	28.00	27.81	27.37	28.01	28.09
	120	26.25	29.45	28.82	29.80	29.45	28.83	29.80	29.77

TABLE IV

AVERAGE NMISE AND PSNR (dB) VALUES FOR THE EXACT UNBIASED INVERSE, MMSE INVERSE WITH THE ORACLE MONTE-CARLO ESTIMATE ϵ_{MC} , AND MMSE INVERSE WITH THE SURE ESTIMATE ϵ_{SURE} . THE INTENSITY RANGE OF EACH IMAGE IS INDICATED IN BRACKETS.

	NMISE			PSNR (dB)		
	Exact unbiased inverse	MMSE ϵ_{MC}	MMSE ϵ_{SURE}	Exact unbiased inverse	MMSE ϵ_{MC}	MMSE ϵ_{SURE}
Spots [0.03, 5.02]	0.0365	0.0324	0.0362	31.96	34.65	32.81
Galaxy [0, 5]	0.0299	0.0241	0.0308	28.05	30.06	28.16
Ridges [0.05, 0.85]	0.0128	0.0129	0.0129	25.89	25.98	25.94
Barbara [0.93, 15.73]	0.0881	0.0971	0.0911	25.92	25.80	25.91
Cells [0.53, 16.93]	0.0649	0.0686	0.0668	30.19	30.23	30.18

$\sigma^2 \mathbf{1}$, the SURE for D is

$$\text{SURE}(D_i) = (D_i - f(z_i))^2 + \sigma^2 \left(2 \frac{\partial D_i}{\partial f(z_i)} - 1 \right), \quad (23)$$

where $i = 1, \dots, N$. The Anscombe variance-stabilizing transformation ensures that these assumptions approximately hold with $\sigma^2 = \text{var}\{f(z)|y\} \simeq 1$. We compute the partial derivative $\frac{\partial D_i}{\partial f(z_i)}$ in (23) as the finite difference $(D_i^\delta - D_i)/\delta$, where D_i^δ is the denoised output obtained after perturbing $f(z_i)$ with a finite increment $\delta = 1.5$ (this value is chosen so that the perturbation can compete with the noise). Thus, calculating (23) requires N individual denoising procedures. Although these can be first accelerated by processing only a neighborhood of the perturbed sample and then parallelized, the computational cost of this SURE approach remains obviously very high. Depending on the particular denoising filter, more sophisticated approaches to empirically estimate (23) exist (see, e.g., [16]).

As this risk estimator is unbiased, it is reasonable to smooth $\text{SURE}(D)$ in order to approximate its expectation. Also for this

smoothing we use the BM3D filter. Further, since $\text{var}\{f(z)|y\}$ is constant only approximately, we replace the factor σ^2 in (23) by $\text{var}\{f(z)|\mathcal{I}_C(D)\}$ (this conditional variance of $f(z)$ can be computed numerically as it is done for its conditional expectation). We denote the obtained MSE estimate as ϵ_{SURE}^2 . In Figures 10–11 we compare ϵ_{SURE} estimates with oracle Monte-Carlo estimates ϵ_{MC} . The average results obtained from the MMSE estimate $\mathcal{I}_{MMSE}(D, \epsilon_{SURE})$ are given in Table IV. On a very simple image, such as Spots, ϵ_{SURE}^2 can provide a reasonable approximation of the mean square error, and thus a noticeable improvement in the PSNR. However, for all other images this approximation is too coarse and the results do not differ on average from the results of the exact unbiased inverse. Figures 12–13 compare the results of the three inverses $\mathcal{I}_C(D) = \mathcal{I}_{ML}(D)$, $\mathcal{I}_{MMSE}(D, \epsilon_{MC})$ and $\mathcal{I}_{MMSE}(D, \epsilon_{SURE})$ for Spots and Barbara, respectively. While visually the three estimates of Barbara are virtually indistinguishable, one can observe that for Spots the MMSE inverses, particularly $\mathcal{I}_{MMSE}(D, \epsilon_{MC})$, provide sharper details

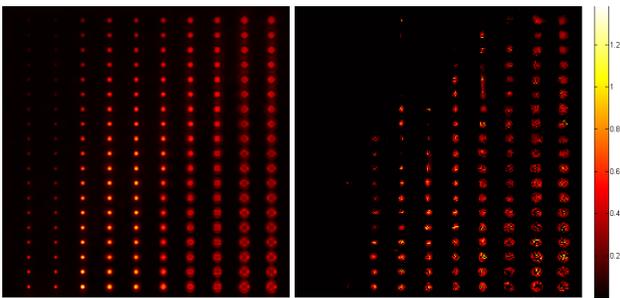


Fig. 10. Oracle Monte-Carlo estimate ε_{MC} for Spots (left) compared with the respective SURE estimate ε_{SURE} (right).

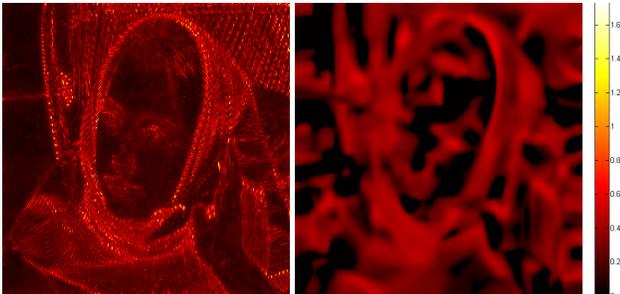


Fig. 11. Oracle Monte-Carlo estimate ε_{MC} for Barbara (left) compared with the respective SURE estimate ε_{SURE} (right).

than the exact unbiased inverse.

C. Computational complexity

As our Matlab implementation of the exact unbiased inverse takes advantage of precomputed values of $E\{f(z) | y\}$ and $E\{z | y\}$, the inverse transformation can be executed quickly. Thus, the computation time of the whole denoising procedure mainly depends on the execution time of the chosen denoising algorithm. Table V shows average computation times of the denoising of Lena (512×512 , peak 10) and Cameraman (256×256 , peak 10) with the exact unbiased inverse, for two different CPUs. It is worth noting that for SAFIR we do not use the default parameters, but the ones that should give the best results (no subsampling, patch radius=3, iterations=8, lambda=66, eta=3.7). This significantly increases the computation time by a factor of about 580, but provides an improvement of about 0.6 dB (see [17] for details about the complexity/performance scaling of the algorithm). Also for BLS-GSM we use its full steerable pyramid implementation.

Note that the faster execution times for the dual core T8300 are rather explained by the fact that the CPU is much newer than the Pentium 4, as at least the Matlab implementation of BM3D does not take advantage of more than one CPU core.

In comparison, the authors of [1] report that the MS-VST + curvelet denoising of Cells (see Figure 6(e)) required 1287 seconds on a 1.1 GHz PC. The authors of [2] do not specify their hardware, but they report the PH-HMT denoising of Cameraman (peak 20) taking 92 seconds with unoptimized Matlab code. Finally, the authors of [3] report the Interscale PURE-LET denoising of Cameraman at 17.25 dB taking only 0.37 seconds with two cyclic shifts and 4.6 seconds with 25

cyclic shifts (hardware not specified). Note that in Table III we compare our results against the PURE-LET with two cyclic shifts, as similar results for 25 cyclic shifts are not presented in [3].

Regarding our MMSE inverse, even with exploiting some acceleration, the time needed for computing the empirical MSE estimate ε_{SURE}^2 is hundreds of times higher than that of a single denoising run. Therefore, based on the minor improvements over $\mathcal{I}_C = \mathcal{I}_{ML}$ shown in Table IV, the practical use of the MMSE inverse appears extremely limited.

V. DISCUSSION AND CONCLUSIONS

In this paper we showed that the three-step procedure of first stabilizing the noise variance by applying the Anscombe transformation, then denoising with an algorithm designed for Gaussian noise, and finally applying an inverse transformation, can still be considered a viable approach for Poisson noise removal. In particular, the poor performance of the asymptotically unbiased inverse at low counts can be overcome by replacing it with the exact unbiased inverse. The excellent performance achieved through the exact unbiased inverse is justified by the fact that this inverse can be interpreted as a maximum likelihood inverse under rather generic hypotheses.

Further, when combined with a state-of-the-art Gaussian denoising algorithm, this method is competitive with some of the best Poisson noise removal algorithms, such as PH-HMT [2] and MS-VST [1]. While most of the improvement is due to the exact unbiased inverse, the choice of the denoising algorithm does also matter, and of the methods considered here BM3D seems to be the best choice due to its overall strong results combined with low complexity.

We have also proposed an MMSE inverse parametrized by the pointwise MSE of the denoised stabilized data. While this inverse is not suitable for practical applications with generic filters, we argue that it can be relevant within specific implementations where more knowledge about the statistics of the estimates is available.

In connection with our contributions, we would like to mention the work of Neyman and Scott [18] on the unbiased inversion of transformed and stabilized variables, and highlight the aspects that make their results different than ours. First, in [18] it is assumed that the transformed variables are exactly stabilized and normalized, which (as the authors also point out) is not possible for Poisson distributed variables stabilized by a root transformation. In our paper we instead always provide a fully accurate statistical modelling of the distribution of the stabilized Poisson variables $f(z)$. Second, in the case when the estimation is inaccurate, even though we and they both assume a normal distribution of the estimate D , the inverse sought by them is an unbiased one (minimum variance unbiased estimate) whereas we address this case by the ML and MMSE inverses.

In this paper we contemplated only imaging, but it is worth noting that the same procedure can be applied to data of any dimension, including 1-D signals and volumetric data.

Let us also remark that even though our focus is on the Anscombe transformation, there exist a variety of other variance-stabilizing transformations for Poisson data, such as

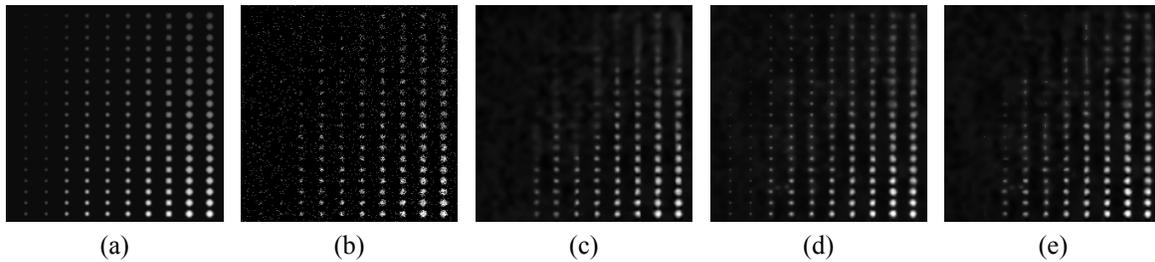


Fig. 12. (a) Original Spots image (intensity range $[0.03, 5.02]$), (b) Poisson-count image, (c) image denoised with BM3D and the exact unbiased inverse (average NMISE = 0.0365 and PSNR = 31.96 dB), (d) image denoised with BM3D and the MMSE inverse with the oracle Monte-Carlo estimate ϵ_{MC} (average NMISE = 0.0324 and PSNR = 34.65 dB), (e) image denoised with BM3D and the MMSE inverse with the SURE estimate ϵ_{SURE} (average NMISE = 0.0362 and PSNR = 32.81 dB). The images shown here are gamma-corrected ($\gamma = 0.6$) for improved visibility of the darker areas.



Fig. 13. (a) Original Barbara image (intensity range $[0.93, 15.73]$), (b) Poisson-count image, (c) image denoised with BM3D and the exact unbiased inverse (average NMISE = 0.0881 and PSNR = 25.92 dB), (d) image denoised with BM3D and the MMSE inverse with the oracle Monte-Carlo estimate ϵ_{MC} (average NMISE = 0.0971 and PSNR = 25.80 dB), (e) image denoised with BM3D and the MMSE inverse with the SURE estimate ϵ_{SURE} (average NMISE = 0.0911 and PSNR = 25.91 dB).

TABLE V

AVERAGE COMPUTATION TIMES OF THE DENOISING PROCEDURES AND TRANSFORMATIONS FOR THE LENA (512×512) AND CAMERAMAN (256×256) IMAGES, FOR TWO DIFFERENT CPUs. FOR SAFIR WE USE THE PARAMETERS GIVING THE BEST DENOISING RESULTS, WHICH SIGNIFICANTLY INCREASES THE COMPUTATION TIME, AND FOR BLS-GSM WE USE ITS FULL STEERABLE PYRAMID IMPLEMENTATION.

CPU	Image	BM3D	SAFIR	BLS-GSM	Forward Anscombe (transformation only)	Exact unbiased inverse (transformation only)
Intel Pentium 4 HT 560 (single core) @ 3.6 GHz	Lena	6.3 s	60 min	68 s	0.07 s	0.22 s
	Cameraman	1.6 s	14 min	17 s	0.014 s	0.063 s
Intel Mobile Core 2 Duo T8300 @ 2.4 GHz	Lena	4.1 s	43 min	-	0.025 s	0.15 s
	Cameraman	1.0 s	9.5 min	-	0.006 s	0.044 s

the Freeman-Tukey transformation [19], or more optimized ones discussed in [20] and [21]. However, the emphasis of this paper is on the improvement gained through applying a suitable inverse, rather than on the improvement gained through optimized variance stabilization. We chose to use the Anscombe transformation because it is in wide use, but the proposed method is in no way limited to this particular transformation.

Finally, we wish to note that the concept of optimal inverse transformations is of course not restricted to the Poisson distribution. As a notable example, in our recent paper [22], we have studied the stabilization and the exact unbiased inverse transformation for the raw data from digital imaging sensors, modelling these data by a doubly censored heteroskedastic normal distribution.

VI. ACKNOWLEDGMENT

We would like to thank Bo Zhang and Jalal Fadili for providing us with the images shown in Figure 3, and Charles

Kervrann and Jérôme Boulanger for providing us with their denoising software (SAFIR) and suitable parameters for it.

We would also like to thank the reviewers, whose valuable feedback helped in improving the paper.

REFERENCES

- [1] Zhang B., J.M. Fadili, and J-L. Starck, "Wavelets, ridgelets, and curvelets for Poisson noise removal", *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1093–1108, July 2008.
- [2] Lefkimmatis, S., P. Maragos, and G. Papandreou, "Bayesian inference on multiscale models for Poisson intensity estimation: Applications to photon-limited image denoising", *IEEE Trans. Image Process.*, vol. 18, no. 8, pp. 1724–1741, Aug. 2009.
- [3] Luisier, F., C. Vonesch, T. Blu, and M. Unser, "Fast interscale wavelet denoising of Poisson-corrupted images", *Signal Processing*, vol. 90, no. 2, pp. 415–427, Feb. 2010.
- [4] Fryzlewicz, P., and G.P. Nason, "A Haar-Fisz Algorithm for Poisson Intensity Estimation", *Journal of Computational and Graphical Statistics*, vol. 13, no. 3, pp. 621–638, 2004.
- [5] Kolaczyk, E.D., and D.D. Dixon, "Nonparametric estimation of intensity maps using Haar wavelets and Poisson noise characteristics", *The Astrophysical Journal*, vol. 534, no. 1, pp. 490–505, 2000.
- [6] Willett, R.M., and R.D. Nowak, "Platelets: A Multiscale Approach for Recovering Edges and Surfaces in Photon-Limited Medical Imaging", *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 332–350, March 2003.

- [7] Willett, R.M., “Multiscale Analysis of Photon-Limited Astronomical Images”, *Statistical Challenges in Modern Astronomy (SCMA) IV*, 2006.
- [8] Anscombe, F.J., “The transformation of Poisson, binomial and negative-binomial data”, *Biometrika*, vol. 35, no. 3/4, pp. 246–254, Dec. 1948.
- [9] Mäkitalo, M., and A. Foi, “On the inversion of the Anscombe transformation in low-count Poisson image denoising”, *Proc. Int. Workshop on Local and Non-Local Approx. in Image Process., LNLA 2009*, Tuusula, Finland, pp. 26–32, Aug. 2009.
- [10] Dabov, K., A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3D transform-domain collaborative filtering”, *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
- [11] Curtiss, J.H., “On transformations used in the analysis of variance”, *The Annals of Mathematical Statistics*, vol. 14, no. 2, pp. 107–122, June 1943.
- [12] Efron, B., “Transformation theory: How normal is a family of distributions?”, *The Annals of Statistics*, vol. 10, no. 2, pp. 323–339, 1982.
- [13] Boulanger, J., J.B. Sibarita, C. Kervrann, and P. Bouthemy, “Non-parametric regression for patch-based fluorescence microscopy image sequence denoising”, *Proc. IEEE Int. Symp. on Biomedical Imaging (ISBI’08)*, pp. 748–751, May 2008.
- [14] Portilla, J., V. Strela, M.J. Wainwright, and E.P. Simoncelli, “Image denoising using scale mixtures of Gaussians in the wavelet domain”, *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.
- [15] Stein, C.M., “Estimation of the mean of a multivariate normal distribution”, *The Annals of Statistics*, vol. 9, no. 6, pp. 1135–1151, Nov. 1981.
- [16] Ramani, S., T. Blu, and M. Unser, “Monte-Carlo SURE: A Black-Box Optimization of Regularization Parameters for General Denoising Algorithms”, *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1540–1554, Sept. 2008.
- [17] Kervrann, C., and J. Boulanger, “Local adaptivity to variable smoothness for exemplar-based image denoising and representation”, *International Journal of Computer Vision*, vol. 79, no. 1, pp. 45–69, Aug. 2008.
- [18] Neyman, J., and E.L. Scott, “Correction for Bias Introduced by a Transformation of Variables”, *The Annals of Mathematical Statistics*, vol. 31, no. 3, pp. 643–655, 1960.
- [19] Freeman, M., and J. Tukey, “Transformations related to the angular and the square root”, *The Annals of Mathematical Statistics*, vol. 21, no. 4, pp. 607–611, 1950.
- [20] Foi, A., “Optimization of variance-stabilizing transformations”, submitted 2009, preprint available at <http://www.cs.tut.fi/~foi/>.
- [21] Foi, A., “Direct optimization of nonparametric variance-stabilizing transformations”, presented at *8èmes Rencontres de Statistiques Mathématiques, CIRM, Luminy*, Dec. 2008.
- [22] Foi, A., “Clipped noisy images: heteroskedastic modeling and practical denoising”, *Signal Processing*, vol. 89, no. 12, pp. 2609–2629, Dec. 2009. doi:10.1016/j.sigpro.2009.04.035

APPENDIX

Here we present the derivations of the maximum likelihood inverse (17) and the MMSE inverse (19).

A. Derivation of the ML inverse

To prove (17), let us consider two cases. First, if $D \geq 2\sqrt{3}/8$, (15) can be maximized by choosing y in such a way that the maximum of the probability density function (PDF) of D given y (16) coincides with D , i.e. $E\{f(z) | y\} = D$. Thus, from (9) we obtain the first half of (17). Second, if $D < 2\sqrt{3}/8$, it is not possible for the maximum of the PDF (16) to coincide with D , since $y \geq 0$; however, because this PDF is monotonically increasing between $-\infty$ and $E\{f(z) | y\}$, and variations of y correspond to translations of the PDF, $\arg \max_y p(D | y)$ is achieved with the smallest possible y , i.e. $y = 0$.

Note that this proof relies only on the fact that the distribution of D is unimodal with mode at $E\{f(z) | y\}$.

B. Derivation of the MMSE inverse

Equation (19) in Section III-C is derived as follows: We define our MMSE inverse transformation by

$$\mathcal{I}_{\text{MMSE}}(D, \varepsilon) = \arg \min_{\hat{y}} \int_{-\infty}^{+\infty} p(y | D)(y - \hat{y})^2 dy. \quad (24)$$

According to Bayes’ theorem,

$$p(y | D) = \frac{p(D | y)p(y)}{p(D)}. \quad (25)$$

Thus, (24) is equivalent to

$$\begin{aligned} \mathcal{I}_{\text{MMSE}}(D, \varepsilon) &= \arg \min_{\hat{y}} \int_{-\infty}^{+\infty} \frac{p(y)}{p(D)} p(D | y)(y - \hat{y})^2 dy. \\ &= \arg \min_{\hat{y}} \int_{-\infty}^{+\infty} p(y)p(D | y)(y - \hat{y})^2 dy. \end{aligned} \quad (26)$$

Assuming y has an improper uniform distribution over \mathbb{R}^+ (uninformative prior), we can further write

$$\mathcal{I}_{\text{MMSE}}(D, \varepsilon) = \arg \min_{\hat{y}} \int_0^{+\infty} p(D | y)(y - \hat{y})^2 dy. \quad (27)$$

The \hat{y} minimizing the integral in (27) is found by differentiation, giving us

$$\hat{y} = \frac{\int_0^{+\infty} p(D | y)y dy}{\int_0^{+\infty} p(D | y) dy}. \quad (28)$$

In practice, the same result can be obtained by equivalently assuming y to be uniformly distributed between 0 and M , where $M > 0$ is a constant much larger than any of our observations z_i , because

$$\int_0^{+\infty} p(D | y)(y - \hat{y})^2 dy \approx \int_0^M p(D | y)(y - \hat{y})^2 dy.$$

The difference between the two integrals is negligible due to our choice of M and the exponential decay of $p(D | y)$.