

# Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data

Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian, *Senior member, IEEE*

**Abstract**—We present a simple and usable noise model for the raw-data of digital imaging sensors. This signal-dependent noise model, which gives the pointwise standard-deviation of the noise as a function of the expectation of the pixel raw-data output, is composed of a Poissonian part, modeling the photon sensing, and Gaussian part, for the remaining stationary disturbances in the output data. We further explicitly take into account the clipping of the data (over- and under-exposure), faithfully reproducing the nonlinear response of the sensor. We propose an algorithm for the fully automatic estimation of the model parameters given a single noisy image. Experiments with synthetic images as well as with real raw-data from various sensors prove the practical applicability of the method and the accuracy of the proposed model.

**Index Terms**—clipping, digital imaging sensors, noise estimation, noise modeling, overexposure, Poisson noise, raw-data.

## I. INTRODUCTION

Progress in hardware design and manufacturing has introduced digital imaging sensors having a dramatically increased resolution. This is mainly achieved by an increase of the pixel density. Despite the electrical and thermal characteristics of the sensors have noticeably improved in the last decade [18], [15], with the size of each pixel becoming smaller and smaller the sensor output signal's susceptibility to photon noise has become greater and greater. As of now, this source of noise appears as the most significant contributor of the overall noise in a digital imaging sensor [1]. This makes the noise component of the raw-data output of the sensor markedly signal-dependent, thus far from the conventional additive white Gaussian noise modeling so widely used in image processing. Further, with the intention of making full use of the rather limited dynamic range of digital sensors, pictures are usually taken with some areas purposely overexposed or *clipped*, i.e. accumulating charge beyond the full-well capacity of the individual pixels. These pixels obviously present highly nonlinear noise characteristics, which are completely different than those of normally exposed pixels.

The raw-data which comes from sensor always undergoes various processing stages (e.g., denoising, demosaicing, deblurring, compression) before the final “cooked” image reaches the user. In order to process the data and/or to attenuate

the noise in the most efficient and effective way, it is vital that a proper modeling of the noise is considered during the various stages of digital image processing. However, the technical datasheets of the devices usually provide vague and inadequate figures for the noise that are of a global nature (i.e., “average” values which are meant to be valid for the whole sensor) [17]. Consequently, raw-data filtering algorithms either assume independent stationary noise models or, if a signal-dependent model is assumed, the correct parameters for the noise are often not specified. Such rough noise estimates are inadequate for the high-quality image processing filters which are rapidly becoming an integral part of the imaging chain.

Two are the contributions in this paper. First, we present a simple noise model which can accurately be used for the raw-data. Based on the above considerations, it is a signal-dependent noise model based on a *Poissonian* part, modeling the photon sensing, and *Gaussian* part, for the remaining stationary disturbances in the output data. We explicitly take into account the problem of *clipping* (over- and under-exposure), faithfully reproducing the nonlinear response of the sensor. Only two parameters are sufficient to fully describe the model. These parameters are explained in relation to the sensor's hardware characteristics (quantum efficiency, pedestal, gain). As a second and most important contribution, we propose an algorithm for the fully automatic estimation of the model parameters given a *single* noisy image.

The paper is organized as follows. In Section II we present the model in its basic form, which ignores the clipping. The parameter estimation algorithm is then presented in Section III. The general model with clipping requires more involved mathematics, and it is given in Section IV, followed by the modified estimation algorithm in Sections V and VI. Throughout these sections, we demonstrate the accuracy of the algorithm with synthetic test images, for which the exact noise parameters are known. Experiments with real raw-data are presented in Section VII; these experiments prove the *practical* applicability of the method and confirm that the raw-data noise can indeed be accurately modeled as a clipped Poissonian-Gaussian process. Further comments and details on the algorithm and its implementation are given in Section VIII.

## II. POISSONIAN-GAUSSIAN MODELING

Let us consider the generic signal-dependent noise observation model of the form

$$z(x) = y(x) + \sigma(y(x)) \xi(x) \quad (1)$$

where  $x \in X$  is the pixel position in the domain  $X$ ,  $z : X \rightarrow \mathbb{R}$  is the observed (recorded) signal,  $y : X \rightarrow \mathbb{R}$  is the original

A. Foi, V. Katkovnik, and K. Egiazarian are with the Department of Signal Processing, Tampere University of Technology, P.O. Box 553, 33101 Tampere, Finland. E-mail: firstname.lastname@tut.fi

M. Trimeche is with the Multimedia Technologies Laboratory, Nokia Research Center, Tampere, Finland. E-mail: firstname.lastname@nokia.com

This work was supported by the Finnish Funding Agency for Technology and Innovation (Tekes), AVIPA/AVIPA2 projects, and by the Academy of Finland, project No. 213462 (Finnish Centre of Excellence program 2006 - 2011).

(unknown) signal,  $\xi : X \rightarrow \mathbb{R}$  is zero-mean independent random noise with standard deviation equal to 1, and  $\sigma : \mathbb{R} \rightarrow \mathbb{R}^+$  is a function of  $y$  that gives the standard deviation of the overall noise component. Throughout the paper, we denote the expected value (or mathematical expectation) of a random variable as  $E\{\cdot\}$ , its variance as  $\text{var}\{\cdot\}$ , and its standard deviation as  $\text{std}\{\cdot\} = \sqrt{\text{var}\{\cdot\}}$ ; when any of these operators is applied to a sequence (resp. matrix) of random variables, its output is defined as the sequence (resp. matrix) of the operator's outputs for the individual random variables. The symbol  $\sigma$  is used exclusively to denote this function of the model (1). From  $E\{\xi(x)\} = 0$  follows that  $E\{z(x)\} = y(x)$ , i.e. the original signal can be defined as the expected value of the noisy observations. Consequently, we have that  $\text{std}\{z(x)\} = \sigma(E\{z(x)\})$ , i.e. the standard deviation of the noise is a function, namely  $\sigma$ , of the expectation of the noisy signal.

In our modeling, we assume that the noise term is composed of two mutually independent parts, a *Poissonian* signal-dependent component  $\eta_p$  and a *Gaussian* signal-independent component  $\eta_g$ :

$$\sigma(y(x))\xi(x) = \eta_p(y(x)) + \eta_g(x). \quad (2)$$

In terms of distributions, these two components are characterized as follows,

$\chi(y(x) + \eta_p(y(x))) \sim \mathcal{P}(\chi y(x))$ ,  $\eta_g(x) \sim \mathcal{N}(0, b)$ , where  $\chi > 0$  and  $b \geq 0$  are real scalar parameters and  $\mathcal{P}$  and  $\mathcal{N}$  denote the Poisson and normal (i.e., Gaussian) distributions. From the elementary properties of the Poisson distribution, we obtain the following equation for the mean and variance

$$\begin{aligned} E\{\chi(y(x) + \eta_p(y(x)))\} &= \text{var}\{\chi(y(x) + \eta_p(y(x)))\} \\ &= \chi y(x). \end{aligned}$$

Since  $E\{\chi(y(x) + \eta_p(y(x)))\} = \chi y(x) + \chi E\{\eta_p(y(x))\}$  and  $\chi^2 \text{var}\{\eta_p(y(x))\} = \chi y(x)$ , it follows that

$$E\{\eta_p(y(x))\} = 0 \quad \text{and} \quad \text{var}\{\eta_p(y(x))\} = y(x)/\chi.$$

Thus, the Poissonian  $\eta_p$  has varying variance that depends on the value of  $y(x)$ ,  $\text{var}\{\eta_p(y(x))\} = ay(x)$ , where  $a = \chi^{-1}$ . The Gaussian component  $\eta_g$  has instead constant variance equal to  $b$ .

Consequently, the overall variance of  $z$  in (1) has the affine form

$$\sigma^2(y(x)) = ay(x) + b, \quad (3)$$

which gives the standard deviation  $\sigma$  as the square root

$$\sigma(y(x)) = \sqrt{ay(x) + b}, \quad (4)$$

and, in particular,  $\sigma(0) = \sqrt{b}$  and  $\sigma(1) = \sqrt{a+b}$ .

Some examples of standard-deviation functions  $\sigma$  for different combinations of the constants  $a$  and  $b$  are shown, as an illustration, in Figure 1 (solid lines).

Figure 2 presents a simple piecewise smooth image which is degraded by Poissonian and Gaussian noise with parameters  $\chi = 100$  ( $a = 0.01$ ) and  $b = 0.04^2$ . As illustrated in Figure 1, these parameters imply that the noise standard-deviation in the brightest parts of the image is more than twice as large as in the darker ones.

### A. Raw-data modeling

The Poissonian-Gaussian model (1-2) is naturally suited for the raw-data of digital imaging sensors. The Poissonian component  $\eta_p$  models the signal-dependent part of the errors, which is essentially due to the photon-counting process, while the Gaussian  $\eta_g$  accounts for the signal-independent errors such as electric and thermal noise. We briefly mention how the above model parameters relate to elementary aspects of the digital sensor's hardware.

1) *Quantum efficiency*: The parameter  $\chi$  of  $\eta_p$  is related to the quantum efficiency of the sensor: the larger the number of photons necessary to produce a response of the sensor (generation of an electron), the smaller the  $\chi$ .

2) *Pedestal parameter*: In digital imaging sensors, the collected charge is always added to some base “pedestal” level  $p_0 \in \mathbb{R}^+$ . This constitutes an offset-from-zero of the output data and it can be rewritten as a shift in the argument of the signal-dependent part of the noise:

$$\begin{aligned} z(x) &= y(x) + \sigma(y(x) - p_0)\xi(x) = \\ &= y(x) + \eta_p(y(x) - p_0) + \eta_g(x). \end{aligned}$$

3) *Analog gain*: We model the analog gain as an amplification of the collected charge. Let us denote the variables before amplification by the circle superscript  $\circ$ ,

$$\hat{z}(x) = \hat{y}(x) + \hat{\eta}_p(\hat{y}(x) - p_0) + \hat{\eta}_g(x).$$

We formalize the amplification  $\Theta$  of  $\hat{z}$  as the multiplication of the noise-free signal, of the Poissonian noise, and of a part of the Gaussian noise, by a scaling constant  $\theta > 1$ ,

$$z(x) = \Theta(\hat{z}(x)) = \theta(\hat{y}(x) + \hat{\eta}_p(\hat{y}(x) - p_0) + \hat{\eta}'_g(x)) + \hat{\eta}''_g(x).$$

Here, the Gaussian noise term  $\hat{\eta}_g$  has been split in two components  $\hat{\eta}'_g$  and  $\hat{\eta}''_g$ ,  $\hat{\eta}'_g + \hat{\eta}''_g = \hat{\eta}_g$ , where  $\hat{\eta}''_g$  represents the portion of the noise that is introduced after the amplification and thus not affected by the factor  $\theta$ . The expectation and variance for  $z$  are

$$\begin{aligned} E\{z(x)\} &= y(x) = \theta\hat{y}(x), & \text{var}\{z(x)\} = \\ &= \theta^2\chi^{-1}(\hat{y}(x) - p_0) + \theta^2\text{var}\{\hat{\eta}'_g(x)\} + \text{var}\{\hat{\eta}''_g(x)\}. \end{aligned}$$

Hence, we come again to a model of the form (3)-(4) with  $a = \chi^{-1}\theta$ ,  $b = \theta^2\text{var}\{\hat{\eta}'_g(x)\} + \text{var}\{\hat{\eta}''_g(x)\} - \theta^2\chi^{-1}p_0$ .

Note that now this  $b$  can be negative, provided a large pedestal  $p_0$  and a small variance of  $\hat{\eta}_g$ . This does not mean that there is a “negative” variance. Indeed, because of the pedestal,  $y \geq \theta p_0$  and therefore  $ay + b \geq 0$ .

In digital cameras, the analog gain (i.e.,  $\theta$ ) is usually controlled by the choice of the ISO sensitivity setting. This can be done manually by the user, or automatically by the camera (“auto mode”). Large ISO numbers (e.g., 800 or 1600) correspond to large  $\theta$ , and thus worse signal-to-noise ratio (SNR). Lower values (e.g., ISO 50) yield a better SNR but at the same time produce darker images, unless these are taken with a longer exposure time (which corresponds to having larger values of  $\hat{y}$  before the multiplication by  $\theta$ ).

Figure 3 shows few examples of the standard-deviation functions  $\sigma$  which can typically be found for the raw data. Two of these examples have  $b < 0$ , which corresponds to a pedestal  $p_0 > 0$ .

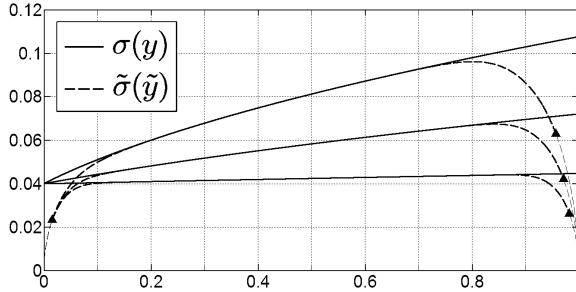


Fig. 1. Some examples of the standard-deviation functions  $\sigma$  (solid lines) from the model (1) for different combinations of the constants  $a$  and  $b$  of Equation (4): (left)  $a = 0.02^2, 0.06^2, 0.10^2$ ,  $b = 0.04^2$  and (right)  $a = 0.4^2, b = 0.02^2, 0.06^2, 0.10^2$ . The dashed lines show the corresponding functions  $\tilde{\sigma}$  of the clipped observation model (30), as functions of the clipped  $\tilde{y}$  (see Section IV). The small black triangles indicate the points  $(\tilde{y}, \tilde{\sigma}(\tilde{y}))$  which correspond to  $y = 0$  and  $y = 1$ .

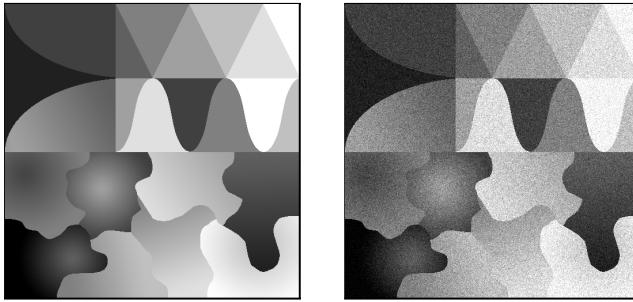


Fig. 2. A piecewise smooth test image of size  $512 \times 512$ : original  $y$  and observation  $z$  degraded by Poissonian and Gaussian noise with parameters  $\chi = 100$  ( $a = 0.01$ ) and  $b = 0.04^2$ .

### B. Heteroskedastic normal approximation

Throughout the following sections, we need to derive a few results and relations which depend not only on the mean and variance, but also on the particular distribution of the processed samples. For the sake of simplification, we exploit the usual normal approximation of the Poisson distribution, which gives

$$\mathcal{P}(\lambda) \approx \mathcal{N}(\lambda, \lambda). \quad (5)$$

The accuracy of this approximation increases with the parameter  $\lambda$  and in practice, for large enough<sup>1</sup>  $\lambda$ , a Poissonian process can be treated as a special heteroskedastic Gaussian one. We thus obtain the following normal approximations of the errors

$$\sigma(y(x))\xi(x) = \sqrt{ay(x) + b}\xi(x) \simeq \eta_h(y(x)), \quad (6)$$

where  $\eta_h(x) \sim \mathcal{N}(0, ay(x) + b)$ .

## III. THE ALGORITHM

Our goal is to estimate the function  $\sigma : \mathbb{R} \rightarrow \mathbb{R}^+$  of the observation model (1) from a noisy image  $z$ . The proposed algorithm is divided in two main stages: *local estimation* of multiple expectation/standard-deviation pairs and *global parametric model fitting* to these local estimates. An initial

<sup>1</sup>How large  $\lambda$  is enough really depends on the considered application and desired accuracy. The fact that the Poisson distribution is discrete is a secondary aspect, because quantization of the digital data makes anyway discrete even errors due to continuous distributions.

For the considered standard-deviation estimation problem, we found experimentally that already with  $\lambda = 10$  (corresponding to  $\chi = 20$  for the middle intensity  $y = 0.5$ ) there is virtually no difference between the estimation accuracy of a truly Poissonian variable and that of its Gaussian approximation.

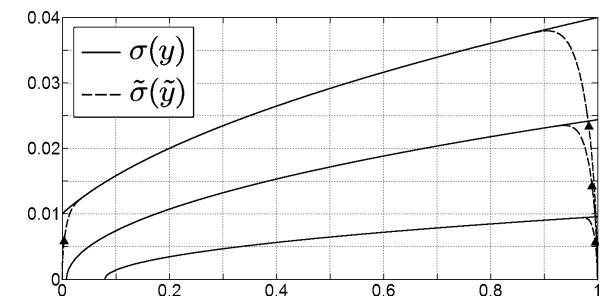
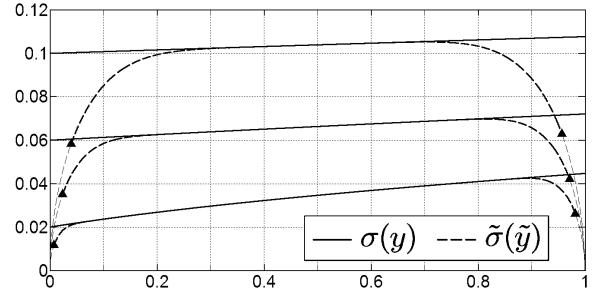


Fig. 3. Some examples of the standard-deviation functions  $\sigma$  (solid lines) which are often found for the raw data. In these three examples the parameters  $(a, b)$  from Equation (4) are  $(1.5 \cdot 10^{-3}, 10^{-4})$ ,  $(6 \cdot 10^{-4}, -5 \cdot 10^{-6})$ , and  $(10^{-4}, -8 \cdot 10^{-6})$ . The dashed lines show the corresponding functions  $\tilde{\sigma}$  of the clipped observation model (30), as functions of the clipped  $\tilde{y}$  (see Section IV). The small black triangles indicate the points  $(\tilde{y}, \tilde{\sigma}(\tilde{y}))$  which correspond to  $y = 0$  and  $y = 1$ .

preprocessing stage, in which the data is transformed to the wavelet domain and then segmented into non-overlapping level sets where the data is smooth, precedes the estimation.

### A. Wavelet domain analysis

Similar to [4], we facilitate the noise analysis by considering wavelet detail coefficients  $z^{\text{wdet}}$  defined as the downsampled convolution

$$z^{\text{wdet}} = \downarrow_2(z \otimes \psi),$$

where  $\psi$  is a 2-D wavelet function with zero mean and unity  $\ell^2$ -norm,  $\sum \psi = 0$ ,  $\|\psi\|_2 = 1$ , and  $\downarrow_2$  denotes the decimation operator that discards every second row and every second column. Analogously, we define the normalized approximation coefficients as

$$z^{\text{wapp}} = \downarrow_2(z \otimes \varphi),$$

where  $\varphi$  is the corresponding 2-D wavelet scaling function, which we specially normalize so that  $\sum \varphi = 1$ .

For noisy images, the detail coefficients  $z^{\text{wdet}}$  contain mostly noise and, due to the normalizations of the convolution kernels, we have

$$\begin{aligned} \text{std}\{z^{\text{wdet}}\} &= \downarrow_2(\text{std}\{z \otimes \psi\}) = \downarrow_2\left(\sqrt{\text{var}\{z\} \otimes \psi^2}\right) \simeq (7) \\ &\simeq \downarrow_2(\text{std}\{z\} \|\psi\|_2) = \downarrow_2(\text{std}\{z\}) = \\ &= \downarrow_2(\sigma(y)) = \sigma(\downarrow_2 y) = \sigma(\downarrow_2(y \sum \varphi)) \simeq \\ &\simeq \sigma(\downarrow_2(y \otimes \varphi)) = \sigma(E\{z^{\text{wapp}}\}), \end{aligned} \quad (8)$$

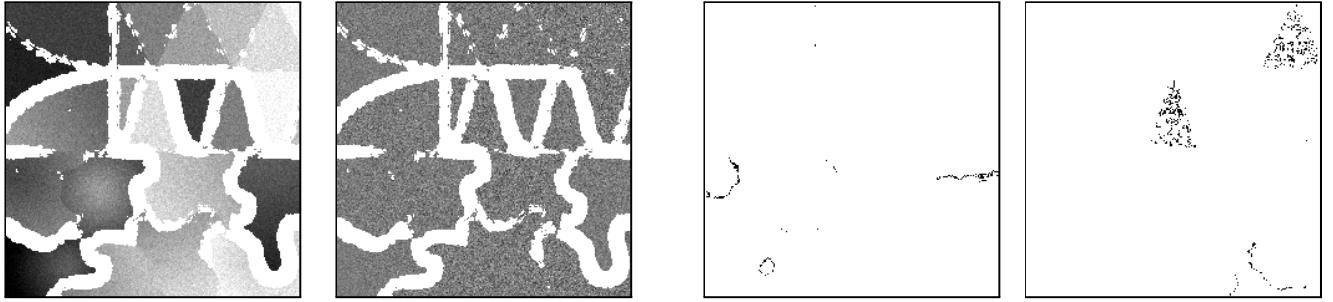


Fig. 4. From left to right: wavelet approximation and detail coefficients  $z^{\text{wapp}}$  and  $z^{\text{wdet}}$ , restricted on the set of smoothness  $X^{\text{smo}}$ , and two level-sets  $S_i$  (13) computed for  $\Delta_i = \Delta = 1/300$ . The scale of this figure is half that of Fig. 2.

with the approximate equalities  $\simeq$  becoming accurate at points in regions where  $y$  (and hence  $\text{std}\{z\}$ ) is uniform, as we can assume that the distribution of  $z$  does not change over the small support of the wavelets. Thus, in particular, at a point  $x$  in such uniform regions, we can assume that

$$z^{\text{wdet}}(x) \sim \mathcal{N}(0, \sigma(E\{z^{\text{wapp}}(x)\})), \quad (9)$$

and, because of decimation and orthogonality properties of wavelet functions, that the noise degrading  $z^{\text{wdet}}$ , as well as the noise degrading  $z^{\text{wapp}}$ , are independent ones.

Note that, always,  $\|\varphi\|_2 \neq 1$ . Therefore, when considering  $\text{std}\{z^{\text{wapp}}\}$ , the above equations can be repeated, replacing  $\psi$  with  $\varphi$ , only provided that the factor  $\|\varphi\|_2$  is kept. Thus, we come to

$$\text{std}\{z^{\text{wapp}}\} \simeq \|\varphi\|_2 \sigma(z^{\text{wapp}}).$$

In our implementation, we use separable kernels  $\psi = \psi_1 \circledast \varphi_1^T$  and  $\varphi = \varphi_1 \circledast \varphi_1^T$  where  $\psi_1$  and  $\varphi_1$  are 1-D Daubechies wavelet and scaling functions

$$\psi_1 = [0.035 \ 0.085 \ -0.135 \ -0.460 \ 0.807 \ -0.333], \quad (10)$$

$$\varphi_1 = [0.025 \ -0.060 \ -0.095 \ 0.325 \ 0.571 \ 0.235].$$

### B. Segmentation

Like in our previous work [6], we segment the data into level sets, in each of which the image can be reasonably assumed to be uniformly close to a certain value. Having nothing but a noisy image at our disposal, we shall employ spatial smoothing (as opposed to temporal smoothing, used in [6]) in order to attenuate the noise and an edge-detector in order to stay clear from edges when analyzing the data, thus enabling the conditions (7)-(8).

There exist a myriad of different methods which can be used for smoothing or for edge detection. However, for our purposes, the following simple and *non-adaptive* methods proved adequate for all considered experimental cases.

1) *Smoothed approximation*: From  $z^{\text{wapp}}$ , we compute a smoothed (low-pass) image  $z^{\text{smo}}$ ,

$$z^{\text{smo}} = z^{\text{wapp}} \circledast \varpi, \quad (11)$$

where  $\varpi$  is positive smoothing kernel,  $\varpi \geq 0$  and  $\|\varpi\|_1 = 1$ . The smoothing action of the kernel should be especially strong, so to effectively suppress most of the noise. In our implementation, we use a uniform  $7 \times 7$  kernel for  $\varpi$ .

In the corresponding regions where  $y$  itself is smooth,  $z^{\text{smo}}$  is approximately equal to  $E\{z^{\text{wapp}}\}$ , and thus to  $\downarrow_2 y$ . This is

a reasonable assumption provided that the support of  $\varpi$  does not intersect edges during the calculation of the convolution (11).

2) *Edges and set of smoothness*: To detect edges, we use the conventional approach where some smoothed derivatives of the image are thresholded against an estimate of the local standard deviation. Exploiting the fact that the mean of the absolute deviations of  $\mathcal{N}(0, 1)$  is equal to  $\sqrt{2/\pi}$  [7], we can define a rough estimate of the local standard-deviations of  $z^{\text{wdet}}$  as the map

$$s = \sqrt{\frac{\pi}{2}} |z^{\text{wdet}}| \circledast \varpi.$$

We define the *set of smoothness*  $X^{\text{smo}}$  as

$$X^{\text{smo}} = \{x \in \downarrow_2 X : \\ : |\nabla(\Lambda(z^{\text{wapp}}))(x)| + |\Lambda(z^{\text{wapp}})(x)| < \tau \cdot s(x)\}, \quad (12)$$

$$\Lambda(z^{\text{wapp}}) = \nabla^2 \text{medfilt}(z^{\text{wapp}}),$$

where  $\nabla$  and  $\nabla^2$  are, respectively, gradient and Laplacian operators, medfilt denotes a  $3 \times 3$  median filter,  $\downarrow_2 X$  is the decimated domain of the wavelet coefficients  $z^{\text{wapp}}$ , and  $\tau > 0$  is positive threshold constant. We realize both the Laplacian operator  $\nabla^2$  and the gradient operator  $\nabla$  as convolutions against  $9 \times 9$  kernels. Thresholding the sum of the moduli of the Laplacian and of its gradient is a heuristic way to obtain “thickened” edges.

In Figure 4, we show the wavelet approximation and detail coefficients  $z^{\text{wapp}}$  and  $z^{\text{wdet}}$ , restricted on the set of smoothness  $X^{\text{smo}}$  (whose complement thus appears as white in the figure), calculated for the test image  $z$  of Figure 2. Note that some of the weakest edges have not been detected as such.

3) *Level sets (segments)*: In the set of smoothness  $X^{\text{smo}}$ , we can assume that edges of the image did not interfere with the smoothing (11), hence, that the conditions (7)-(8) hold and that, for  $x \in X^{\text{smo}}$ ,

$$z^{\text{smo}}(x) = E\{z^{\text{wapp}}(x)\} = E\{(\downarrow_2 z)(x)\} = (\downarrow_2 y)(x), \\ \text{std}\{z^{\text{wdet}}(x)\} = \text{std}\{(\downarrow_2 z)(x)\} = (\downarrow_2 (\sigma(y)))(x).$$

We identify in the smoothness set  $X^{\text{smo}}$  a collection of  $N$  non-overlapping level sets (segments)  $S_i \subset X^{\text{smo}}$ ,  $i = 1, \dots, N$  of the smoothed image  $z^{\text{smo}}$ . Each level set, characterized by its centre value  $u_i$  and allowed deviation  $\Delta_i > 0$ , is defined as

$$S_i = \{x \in X^{\text{smo}} : z^{\text{smo}}(x) \in [u_i - \Delta_i/2, u_i + \Delta_i/2]\}. \quad (13)$$

By non-overlapping we mean that  $S_i \cap S_j = \emptyset$  if  $i \neq j$ . In practice, assuming a signal in the range  $[0, 1]$ , one can take fixed  $\Delta_i \equiv \Delta$  and equispaced  $u_i \in \{\Delta j, j = 1, \dots, \bar{N} = \lceil \Delta^{-1} \rceil\}$ , where the  $\lfloor \cdot \rfloor$  brackets indicate the rounding to the nearest larger or equal integer. Further, we require that the level sets are non-trivial, in the sense that each set  $S_i$  must contain at least two samples<sup>2</sup>; thus,  $N \leq \bar{N}$  and  $\bigcup_{i=1}^N S_i \subseteq X^{\text{smo}} \subseteq \downarrow_2 X$ . Figure 4 shows two of the level sets computed for the example in Figure 2 for  $\Delta = 1/300$ . Observe that these sets are meager and quite fragmented.

### C. Local estimation of expectation/standard-deviation pairs

For each level set  $S_i$ , we define the (unknown) variable

$$y_i = \frac{1}{n_i} \sum_{j=1}^{n_i} E\{z^{\text{wapp}}(x_j)\}, \quad \{x_j\}_{j=1}^{n_i} = S_i. \quad (14)$$

Note that  $y_i$  and  $u_i$  might not coincide. The level set  $S_i$  is used as a one domain for the computation of a pair of estimates  $(\hat{y}_i, \hat{\sigma}_i)$ , where  $\hat{y}_i$  is an estimate of  $y_i$  and  $\hat{\sigma}_i$  is an estimate of  $\sigma(y_i)$ . In what follows, although we shall refer explicitly to  $y_i$ , this variable is always used implicitly and, in the final estimation of the function  $y \mapsto \sigma(y)$ , the many  $y_i, i, \dots, N$ , remain “hidden” variables which are modeled as unknown. Similarly, the smoothed data  $z^{\text{smo}}$  and the values  $u_i$  and  $\Delta_i$  used for the construction of  $S_i$  do not appear in the following estimation, where only  $z^{\text{wdet}}$ ,  $z^{\text{wapp}}$  and  $S_i$  are used in order to compute the estimates  $\hat{y}_i$  and  $\hat{\sigma}_i$ .

1) *Estimation of  $y_i$ :* We estimate  $y_i$  as the sample mean of the approximation coefficients  $z^{\text{wapp}}$  on  $S_i$

$$\hat{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} z^{\text{wapp}}(x_j), \quad \{x_j\}_{j=1}^{n_i} = S_i. \quad (15)$$

2) *Estimation of  $\sigma(y_i)$ :* The estimate  $\hat{\sigma}_i$  is calculated as the unbiased sample standard-deviation of the detail coefficients  $z^{\text{wdet}}$  on  $S_i$

$$\hat{\sigma}_i = \frac{1}{\kappa_{n_i}} \sqrt{\frac{\sum_{j=1}^{n_i} (z^{\text{wdet}}(x_j) - \bar{z}_i^{\text{wdet}})^2}{n_i - 1}}, \quad (16)$$

where  $\bar{z}_i^{\text{wdet}} = \frac{1}{n_i} \sum_{j=1}^{n_i} z^{\text{wdet}}(x_j)$  and the factor  $\kappa_{n_i}^{-1}$  is defined [7]

$$\kappa_n = \sqrt{\frac{2}{n-1} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}} = 1 - \frac{1}{4n} - \frac{7}{32n^2} + O\left(\frac{1}{n^3}\right). \quad (17)$$

This factor, which comes from the mean of the chi-distribution with  $n - 1$  degrees of freedom, makes the estimate unbiased for normally and identically independently distributed (i.i.d.)  $z^{\text{wdet}}(x_j)$ .

3) *Unbiasedness:* Clearly from the definition (14),  $\hat{y}_i$  is an unbiased estimator of  $y_i$ .

The unbiasedness of  $\hat{\sigma}_i$  as an estimator of  $\sigma(y_i)$  is a more complex issue. As observed above,  $\hat{\sigma}_i$  is an unbiased estimator of  $\sigma(y_i)$  provided that  $z^{\text{wdet}}$  is normally i.i.d. on the level set  $S_i$ . However, we cannot claim, in general, that  $z^{\text{wdet}}$  is identically distributed on  $S_i$ . We remark that

<sup>2</sup>The smoothness threshold  $\tau$  (12) can be automatically increased in the rare event of  $N < 2$ , i.e. when there are not enough non-trivial level sets for the estimation. Note that  $X^{\text{smo}}$  is monotonically enlarging to  $\downarrow_2 X$  with  $\tau \xrightarrow[\tau \rightarrow \infty]{} \downarrow_2 X$ .

the assumed validity of (7)-(8) concerns individual points. It does not mean that  $\text{std}\{z^{\text{wdet}}\}$  is constant over  $S_i$ . As a matter of fact, especially for large  $\Delta_i$ ,  $E\{z^{\text{wapp}}(x)\}$  is not constant for  $x \in S_i$ , which implies that the standard deviations of the wavelet detail coefficients (8) are not constant over  $S_i$ . Lacking any particular hypothesis on the image  $y$ , it is nevertheless reasonable to assume that  $\{E\{z^{\text{wapp}}(x)\}, x \in S_i\}$  has a symmetric (discrete) distribution centred at  $y_i$  (with diameter bounded by  $\Delta_i$ ). Because of (3) and (9), we have that  $\{\text{var}\{z^{\text{wdet}}(x)\}, x \in S_i\}$  has also a symmetrical distribution, which is centred at  $\text{var}\{y_i\}$ . This makes  $\kappa_{n_i}^2 \hat{\sigma}_i^2$  an unbiased<sup>3</sup> estimator of  $\text{var}\{y_i\}$  and, since  $\kappa_n \xrightarrow[n \rightarrow \infty]{} 1$ ,  $\hat{\sigma}_i$  is an asymptotically unbiased estimator of  $\sigma(y_i)$ . This asymptotic unbiasedness is relevant in the practice, since a large  $\Delta_i$  corresponds to large  $n_i$ .

We further note that, despite the segmentation and removal of edges, the presence of sharp image features, singularities, or even texture in the segment  $S_i$  is not completely ruled out. This can be effectively compensated by means of non-linear robust estimators of the standard deviation, such as the well-known median of absolute deviations (MAD) [12]. For the sake of expository simplicity, in the current and in the next section we restrict ourselves to the basic estimator (16) and postpone considerations on robust estimation of the standard-deviation to Section VI.

4) *Variance of the estimates:* The variance of the estimates  $\hat{y}_i$  and  $\hat{\sigma}_i$  depends directly on the variances of the samples used for the estimation, which are degraded by independent noise. With arguments similar to Section III-C.3, the variances of the estimates can be expressed as

$$\text{var}\{\hat{y}_i\} = \sigma^2(y_i) c_i, \quad \text{var}\{\hat{\sigma}_i\} = \sigma^2(y_i) d_i, \quad (18)$$

$$c_i = \frac{\|\varphi\|_2^2}{n_i}, \quad d_i = \frac{1 - \kappa_{n_i}^2}{\kappa_{n_i}^2} = \frac{1}{2n_i} + \frac{5}{8n_i^2} + O\left(\frac{1}{n_i^3}\right), \quad (19)$$

where these expressions coincide with those for the perfect case when  $\text{var}\{z^{\text{wdet}}\}$  and  $\text{var}\{z^{\text{wapp}}\}$  are constant on  $S_i$  [7].

5) *Distribution of the estimates:* The estimates  $\hat{y}_i$  and  $\hat{\sigma}_i$  are distributed, respectively, following a normal distribution and a scaled non-central chi-distribution, which can also be approximated, very accurately for large  $n_i$ , as a normal distribution [7]. Thus, in what follows, we treat both  $\hat{y}_i$  and  $\hat{\sigma}_i$  as normally distributed random variables and, in particular, as

$$\hat{y}_i \sim \mathcal{N}(y_i, \sigma^2(y_i) c_i), \quad \hat{\sigma}_i \sim \mathcal{N}(\sigma(y_i), \sigma^2(y_i) d_i), \quad (20)$$

where  $c_i$  and  $d_i$  are defined as in (19).

### D. Maximum-likelihood fitting of a global parametric model

The maximum-likelihood (ML) approach is used to fit a global parametric model of the function  $\sigma$  on the estimates  $\{\hat{y}_i, \hat{\sigma}_i\}_{i=1}^N$ . Depending on the parameters  $a$  and  $b$ , we have  $\sigma^2(y) = ay + b$ . For reasons of numerical consistency (note that formally this  $\sigma^2(y)$  may be zero or negative), for the fitting we define a simple regularized variance-function  $\sigma_{\text{reg}}^2$  as

$$\sigma_{\text{reg}}^2(y) = \max(\varepsilon_{\text{reg}}^2, \sigma^2(y)) \quad (21)$$

<sup>3</sup>This can be proved easily since, for  $x \in X^{\text{smo}}$ , we can treat  $E\{z^{\text{wdet}}(x)\}$  as zero.

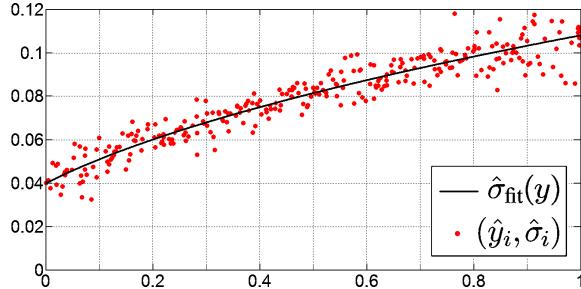


Fig. 5. Each dot of the scatter plot corresponds to a pair  $(\hat{y}_i, \hat{\sigma}_i)$  of estimates of  $y_i$  and  $\sigma(y_i)$ . The solid line shows the maximum-likelihood estimate  $\hat{\sigma}_{\text{fit}}$  of the true standard-deviation function  $\sigma$ . The plot of  $\hat{\sigma}_{\text{fit}}$  overlaps perfectly with that of the true  $\sigma$  (shown in Figure 1). The estimated parameters are  $\hat{a} = 0.01008$  ( $\hat{\chi} = 99.20$ ) and  $\hat{b} = 0.001583$  ( $\sqrt{\hat{b}} = 0.03979$ ). The initialization parameters, found as the least-squares solution (26), were  $\hat{a}_0 = 0.00994$  ( $\hat{\chi}_0 = 100.62$ ) and  $\hat{b}_0 = 0.001649$  ( $\sqrt{\hat{b}_0} = 0.04061$ ).

where  $\varepsilon_{\text{reg}} > 0$  is a small regularization parameter. Hence, the regularized standard-deviation  $\sigma_{\text{reg}} = \sqrt{\sigma_{\text{reg}}^2(y)}$  is always well defined, for any choice of  $a$ ,  $b$ , and  $y$ .

As discussed in Section III-C.5, we can assume normality and unbiasedness for both  $\hat{y}_i$  and  $\hat{\sigma}_i$ . Thus, the conditional probability densities of  $\hat{y}_i$  and  $\hat{\sigma}_i$  given  $y_i = y$  are, respectively,

$$\begin{aligned}\wp(\hat{y}_i|y_i=y) &= \frac{1}{\sqrt{2\pi\sigma_{\text{reg}}^2(y)c_i}} e^{-\frac{1}{2\sigma_{\text{reg}}^2(y)c_i}(\hat{y}_i-y)^2}, \\ \wp(\hat{\sigma}_i|y_i=y) &= \frac{1}{\sqrt{2\pi\sigma_{\text{reg}}^2(y)d_i}} e^{-\frac{1}{2\sigma_{\text{reg}}^2(y)d_i}(\hat{\sigma}_i-\sigma_{\text{reg}}(y))^2}.\end{aligned}$$

Further, we observe that, because of the orthogonality of the wavelets,  $\hat{y}_i$  and  $\hat{\sigma}_i$  are mutually independent<sup>4</sup>. Hence,

$$\begin{aligned}\wp((\hat{y}_i, \hat{\sigma}_i)|y_i=y) &= \wp(\hat{y}_i|y_i=y) \wp(\hat{\sigma}_i|y_i=y) = \\ &= \frac{1}{2\pi\sqrt{c_id_i}} \frac{1}{\sigma_{\text{reg}}^2(y)} e^{-\frac{1}{2\sigma_{\text{reg}}^2(y)}\left(\frac{(\hat{y}_i-y)^2}{c_i} + \frac{(\hat{\sigma}_i-\sigma_{\text{reg}}(y))^2}{d_i}\right)}.\end{aligned}\quad (22)$$

The posterior likelihood  $L$  is obtained by considering all measurements  $\{(\hat{y}_i, \hat{\sigma}_i)\}_{i=1}^N$  and by integrating the densities  $\wp((\hat{y}_i, \hat{\sigma}_i)|y_i=y)$  with respect to a prior probability density  $\wp_0(y)$  of  $y$ ,

$$L(a, b) = \prod_{i=1}^N \int_{-\infty}^{\infty} \wp((\hat{y}_i, \hat{\sigma}_i)|y_i=y) \wp_0(y) dy. \quad (23)$$

The integration copes with the fact that  $y_i$  and  $y$  are unknown. For images in the range  $[0, 1]$ , the simplest and most obvious choice is  $\wp_0$  to be uniform on  $[0, 1]$ , which implies that (23) becomes  $L(a, b) = \prod_{i=1}^N \int_0^1 \wp((\hat{y}_i, \hat{\sigma}_i)|y_i=y) dy$ . In our experiments with synthetic images we use this prior. However, we wish to note that other prior statistics have been shown to be more representative of the histograms of natural images [10]. Let us observe that  $(\hat{y}_i, \hat{\sigma}_i)$  and  $(\hat{y}_j, \hat{\sigma}_j)$ ,  $i \neq j$ , are mutually independent because the corresponding level sets  $S_i$  and  $S_j$  are non-overlapping.

<sup>4</sup>This independence is a general property of the sample mean and sample standard-deviation, which property holds also when the estimates are computed from the very same samples [7]. However, by sampling two independent sets of wavelet coefficients, we have that the two estimates are necessarily independent, regardless of the particular mean and standard-deviation estimators used, a fact that comes useful for the forthcoming sections.

From (23), the sought parameter estimates  $\hat{a}$  and  $\hat{b}$  are defined as the solution of

$$\begin{aligned}(\hat{a}, \hat{b}) &= \underset{a,b}{\operatorname{argmax}} L(a, b) = \underset{a,b}{\operatorname{argmin}} -\ln L(a, b) = \\ &= \underset{a,b}{\operatorname{argmin}} -\sum_{i=1}^N \ln \int \wp((\hat{y}_i, \hat{\sigma}_i) | y_i = y) \wp_0(y) dy.\end{aligned}\quad (24)$$

Hence, our final estimate of the function  $\sigma$  is

$$\hat{\sigma}_{\text{fit}}(y) = \sqrt{\max(0, \hat{a}y + \hat{b})}. \quad (25)$$

Figure 5 shows the result of the above optimization for the test example shown in Figure 2. It can be seen that the procedure estimates the parameters of the noise with great accuracy.

1) *Iterative solution and initialization:* In our implementation, we solve the problem (24) numerically, using the Nelder-Mead iterative downhill simplex method [13] and evaluating the integrals as finite sums. As initial parameters  $\hat{a}_0, \hat{b}_0$  for this iterative optimization we take the least-squares solution

$$\begin{aligned}[\hat{a}_0, \hat{b}_0] &= \underset{a,b}{\operatorname{argmin}} ([a \ b] \Phi^T - \hat{\mathbf{v}}) ([a \ b] \Phi^T - \hat{\mathbf{v}})^T = \\ &= \hat{\mathbf{v}} \Phi (\Phi^T \Phi)^{-1},\end{aligned}\quad (26)$$

where

$$\Phi = \begin{bmatrix} \hat{y}_1 & 1 \\ \hat{y}_2 & 1 \\ \vdots & \vdots \end{bmatrix}, \quad \hat{\mathbf{v}} = [\kappa_{n_1}^2 \hat{\sigma}_1^2 \quad \kappa_{n_2}^2 \hat{\sigma}_2^2 \quad \dots], \quad (28)$$

with the factors  $\kappa_n$  defined as  $\kappa_n = \kappa_n$ . The linear problem (26) allows a simple direct solution by means of the normal equations (27). While in (24) we aim at fitting the standard-deviation curve  $\sigma_{\text{reg}}$  to the estimates  $\{(\hat{y}_i, \hat{\sigma}_i)\}_{i=1}^N$ , Equation (26) minimizes the residuals with respect to the variances  $\hat{\sigma}^2$ , treated as a linear function of the parameters  $a$  and  $b$ . Here, the factor  $\kappa_{n_i}^2$  makes  $\kappa_{n_i}^2 \hat{\sigma}_i^2$  an unbiased estimate of the variance (contrary to  $\hat{\sigma}_i$  (16), which is an unbiased estimate of the standard deviation).

## IV. CLIPPING (CENSORING)

### A. Clipped observations model

In practice, the data range, or *dynamic range*, of acquisition, transmission, and storage systems is always limited. Without loss of generality, we consider data given on the normalized range  $[0, 1]$ , where the extremes correspond to the maximum and minimum pixel values for the considered noisy image (e.g., raw data) format. Even if the noise-free image  $y$  is within the  $[0, 1]$  range, the noise can cause  $z$  to exceed these bounds. We shall assume that values exceeding these bounds are replaced by the bounds themselves, as this corresponds to the behavior of digital imaging sensors in the case of over- or underexposure. Thus, we define the *clipped* (or censored<sup>5</sup>)

<sup>5</sup>Strictly speaking, the form of the so-called censored samples [2] is really  $\tilde{z} = z$  if  $0 \leq z \leq 1$  and no sample (i.e., censoring) if  $z < 0$  or  $z > 1$ . Usually, the amount of censored samples below and above the extrema are assumed as known (Type-1 censoring). Thus, clipped (29) and censored observations can, in a sense, be considered as equivalent. However, the formulas and estimators for censored variables which can be found in the literature cannot be used directly in the case of the clipped observations (29).

observations  $\tilde{z}$  as

$$\tilde{z}(x) = \max(0, \min(z(x), 1)), \quad x \in X, \quad (29)$$

where  $z$  is given by the signal-dependent noise model (1). Let  $\tilde{y}(x) = E\{\tilde{z}(x)\}$ . The corresponding noise model for the clipped observations (29) is

$$\tilde{z}(x) = \tilde{y}(x) + \tilde{\sigma}(\tilde{y}(x))\tilde{\xi}(x), \quad (30)$$

where again  $E\{\tilde{\xi}(x)\} = 0$ ,  $\text{var}\{\tilde{\xi}(x)\} = 1$ , and the function  $\tilde{\sigma} : [0, 1] \rightarrow \mathbb{R}^+$  is defined as  $\tilde{\sigma}(\tilde{y}(x)) = \text{std}\{\tilde{z}(x)\}$ .

In general,  $\tilde{y}(x) = E\{\tilde{z}(x)\} \neq E\{z(x)\} = y(x)$ ,  $\tilde{\sigma}(\tilde{y}(x)) = \text{std}\{\tilde{z}(x)\} \neq \text{std}\{z(x)\} = \sigma(y(x))$ , and, even though  $\text{var}\{\tilde{\xi}(x)\} = \text{var}\{\xi(x)\} = 1$ , the distributions of  $\xi$  and  $\tilde{\xi}$  are different.

In Figure 1 one can compare the standard-deviation functions  $\tilde{\sigma}$  (dashed line) and  $\sigma$  (solid line) for different combinations of the constants  $a$  and  $b$  in (4).

In the next sections, we rely on the heteroskedastic normal approximation (6) and hence treat  $z$  as a purely Gaussian variable. Consequently, we model  $\tilde{z}$  as a clipped (censored) normal variable. We note that this normal approximation is especially relevant for values of  $y = E\{z\}$  close to 0 or 1, where the clipping effects may be dominant. For  $y$  close to 1, we have that (5) holds with the largest values of  $\lambda$ , hence is for this values of  $y$  that the Gaussianization of the Poissonian component  $\eta_p(y(x))$  is most accurate. For  $y$  close to 0, although  $\lambda$  in (5) might not be large, the approximation holds because the variance  $ay(x)$  of  $\eta_p(y(x))$  becomes negligible compared to the variance  $b$  of the Gaussian part  $\eta_g(x)$ . This is true provided that  $b \neq 0$ . However, if  $b = 0$  the noise has only the Poissonian component  $\eta_p$ , which is always positive. It means that  $z \geq 0$  and, thus,  $\tilde{z} = \min(1, z)$ . Therefore, if  $b = 0$ , for our purposes it is sufficient to consider only the normal approximation for  $y$  close to 1, as no clipping happens at 0.

### B. Expectations, standard deviations, and their transformations

To simplify the calculations, we shall assume that the two clippings, the one from below ( $z < 0$ ,  $\tilde{z} = 0$ ) and the one from above ( $z > 1$ ,  $\tilde{z} = 1$ ), are not mixed by the randomness of the noise, and can thus be computed independently<sup>6</sup>. In other words, this means that, *in practice*, if  $y$  is close enough to 0 so that it is possible that  $z < 0$ , then it is impossible that  $z > 1$ ; similarly, if  $y$  is close enough to 1, so that  $z$  can be larger than 1, then  $z$  cannot be smaller than 0; for intermediate values of  $y$ , with  $0 \ll y \ll 1$ , we have that  $0 < z < 1$ , i.e. clipping is not happening.

In what follows, we therefore treat separately the two cases:

<sup>6</sup>Formally, this corresponds to assuming that, for a given  $y(x)$ , the product probability  $P(z(x) > 1) \cdot P(z(x) < 0)$  is negligibly small. This condition is satisfied provided, e.g., the stronger condition that  $P(z(x) < 0 | y(x) > 1 - \epsilon)$  and  $P(z(x) > 1 | y(x) < \epsilon)$  are both negligibly small for  $\epsilon = 0.5$ . These conditions are all surely met in the practical cases, since there the standard deviation  $\sigma(y(x))$  of  $z(x)$  is always much smaller (in fact, several orders smaller) than 0.5 and its distribution does not have heavy tails (note that  $[y - \gamma\sigma(y), y + \gamma\sigma(y)]$  with  $\gamma \geq 4$  can be a rather “safe” confidence interval for  $z$ , with higher than 99.99% confidence for Gaussian distributions).

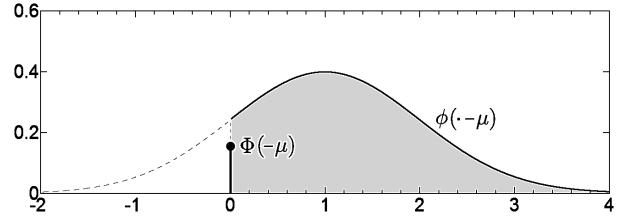


Fig. 6. The probability density function of  $\tilde{\nu} = \max(0, \nu)$ , as defined by Equation (31). In this illustration  $\mu = 1$ . The height of the impulse at 0 is equal to the area under the bell curve between  $-\infty$  and 0.

- clipping from below (left single censoring):  $y$  and  $z$  are near 0 and  $z < 1$ , thus,  $\tilde{z} = \max(0, z)$ ;
- clipping from above (right single censoring):  $y$  and  $z$  are near 1 and  $z > 0$ , thus,  $\tilde{z} = \min(z, 1)$ .

Further, we combine the results for the two cases, so to obtain formulas which are valid for the case  $\tilde{z} = \max(0, \min(z, 1))$ , where clipping can happen from above or below (double censoring).

1) *Clipping from below (left single censoring)*: Since  $\tilde{z} = \max(0, z)$ , we have that  $E\{\tilde{z}\} \geq E\{z\} = y$ .

Let  $\nu \sim \mathcal{N}(\mu, 1)$  be a normally distributed random variable with mean  $E\{\nu\} = \mu$  and unitary variance and  $\tilde{\nu} = \max(0, \nu)$ .

The probability density  $f_{\tilde{\nu}}$  of  $\tilde{\nu}$  is a generalized function defined as follows

$$f_{\tilde{\nu}}(t) = \begin{cases} \phi(t - \mu) + \Phi(-\mu) \delta_0(t) & t \geq 0, \\ 0 & t < 0, \end{cases} \quad (31)$$

where  $\phi$  and  $\Phi$  are the probability density and cumulative distribution functions (p.d.f. and c.d.f.) of the standard normal  $\mathcal{N}(0, 1)$  and  $\delta_0$  is the Dirac delta impulse. This function is illustrated in Figure 6.

Tedious but simple calculations (see, e.g., [7], or [8], Chapter 20) show that the expectation  $E\{\tilde{\nu}\}$  and the variance  $\text{var}\{\tilde{\nu}\}$  of the clipped  $\tilde{\nu}$  are

$$E\{\tilde{\nu}\} = \Phi(\mu)\mu + \phi(\mu), \quad (32)$$

$$\text{var}\{\tilde{\nu}\} = \Phi(\mu) + \phi(\mu)\mu - \phi^2(\mu) + \Phi(\mu)\mu(\mu - \Phi(\mu)\mu - 2\phi(\mu)). \quad (33)$$

The plots of the expectation  $E\{\tilde{\nu}\}$  and of the standard deviation  $\text{std}\{\tilde{\nu}\} = \sqrt{\text{var}\{\tilde{\nu}\}}$  as functions  $\mathcal{E}_m$  and  $\mathcal{S}_m$  of  $\mu = E\{\nu\}$  are shown in Figure 7. Observe that  $E\{\tilde{\nu}\}$  is strictly positive even for negative values of  $\mu$ ; it is convex and increasing,  $E\{\tilde{\nu}\} \rightarrow 0$  as  $\mu \rightarrow -\infty$  and  $E\{\tilde{\nu}\}$  is asymptotic to  $\mu = E\{\nu\}$  as  $\mu \rightarrow +\infty$ . The standard deviation  $\text{std}\{\tilde{\nu}\}$  approaches  $1 = \text{std}\{\nu\}$  as  $\mu \rightarrow +\infty$  and goes to zero as  $\mu$  decreases.

The normal approximation (6) gives that  $z \sim \mathcal{N}(y, \sigma^2(y))$ . “Standardization” of the noise is obtained dividing the variables by  $\sigma(y)$ , which gives  $\frac{\tilde{z}}{\sigma(y)} \sim \mathcal{N}\left(\frac{y}{\sigma(y)}, 1\right)$ . It means that, by taking  $\mu = \frac{y}{\sigma(y)}$ , we can write  $z = \sigma(y)\nu$ ,  $\tilde{z} = \sigma(y)\tilde{\nu}$ . It follows that  $\tilde{y}(x) = E\{\tilde{z}(x)\} = \sigma(y)E\{\tilde{\nu}\}$  and  $\tilde{\sigma}(\tilde{y}) = \text{std}\{\tilde{z}\} = \sigma(y)\text{std}\{\tilde{\nu}\}$ . Exploiting this standardization, we can formulate the *direct* and *inverse* transformations which link  $\sigma$  and  $y$  to  $\tilde{y}$  and  $\tilde{\sigma}$ .

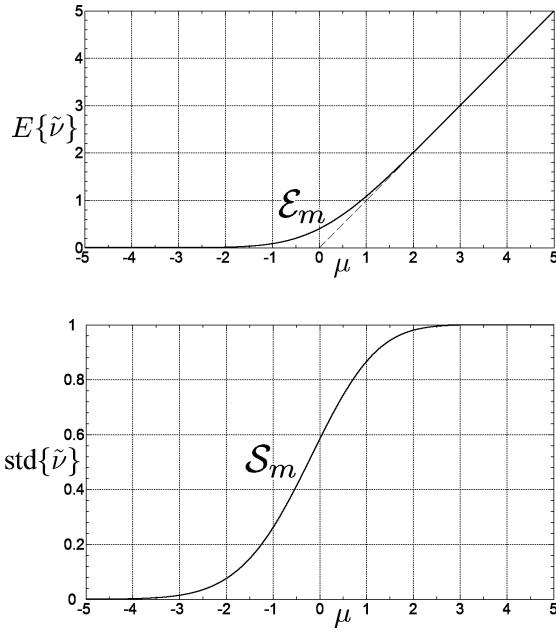


Fig. 7. Expectation  $E\{\tilde{\nu}\}$  and standard deviation  $\text{std}\{\tilde{\nu}\}$  of the clipped  $\tilde{\nu} = \max(0, \nu)$  as functions  $\mathcal{E}_m$  and  $\mathcal{S}_m$  of  $\mu$ , where  $\mu = E\{\nu\}$  and  $\nu \sim \mathcal{N}(\mu, 1)$ .

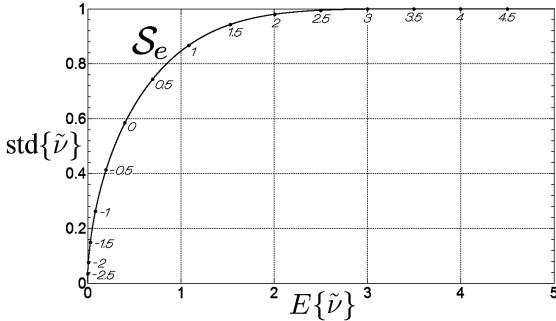


Fig. 8. Standard deviation  $\text{std}\{\tilde{\nu}\}$  of the clipped  $\tilde{\nu} = \max(0, \nu)$  as function  $S_e$  of its expectation  $E\{\tilde{\nu}\}$ . The numbers in italic indicate the corresponding value of  $\mu$ , where  $\mu = E\{\nu\}$  and  $\nu \sim \mathcal{N}(\mu, 1)$ .

a) *Direct transformation ( $\tilde{y}$  and  $\tilde{\sigma}$  from  $y$  and  $\sigma$ ):* From the above formulas we obtain

$$\tilde{y} = \sigma(y) \mathcal{E}_m \left( \frac{y}{\sigma(y)} \right), \quad (34)$$

$$\tilde{\sigma}(\tilde{y}) = \sigma(y) \mathcal{S}_m \left( \frac{y}{\sigma(y)} \right), \quad (35)$$

which give an explicit expression for the clipped observation model (30), provided that  $\sigma(y)$  from the basic model (1) is known. In particular, (34) and (35) define the transformations that bring the standard deviation curve  $(y, \sigma(y))$  to its clipped counterpart  $(\tilde{y}, \tilde{\sigma}(\tilde{y}))$ .

The two plots in Figure 7 can be unified, plotting  $\text{std}\{\tilde{\nu}\}$  as a function of  $E\{\tilde{\nu}\}$ . This is shown by the function  $\mathcal{S}_e$  in Figure 8. Naturally, between  $\mathcal{S}_m$  and  $\mathcal{S}_e$  there is only a change of the independent variables,  $\mu \longleftrightarrow E\{\tilde{\nu}\}$ , hence, from (35) follows that

$$\tilde{\sigma}(\tilde{y}) = \sigma(y) \mathcal{S}_e \left( \frac{\tilde{y}}{\sigma(y)} \right), \quad (36)$$

where  $\tilde{y}$  can be obtained from (34).

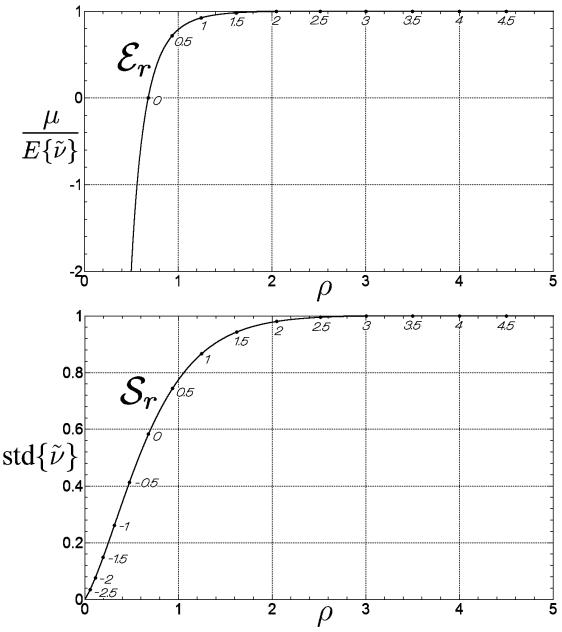


Fig. 9. Expectation  $E\{\tilde{\nu}\}$  and standard deviation  $\text{std}\{\tilde{\nu}\}$  of the clipped  $\tilde{\nu} = \max(0, \nu)$  as functions  $\mathcal{E}_r$  and  $\mathcal{S}_r$  of  $\rho = \frac{E\{\tilde{\nu}\}}{\text{std}\{\tilde{\nu}\}}$ . The numbers in italic indicate the corresponding value of  $\mu$ .

b) *Inverse transformation ( $\sigma$  and  $y$  from  $\tilde{\sigma}$  and  $\tilde{y}$ ):* As clearly seen in Figure 7, the plot of  $\mathcal{S}_e$  is strictly convex, which implies that the (incremental) ratio  $\rho = \frac{E\{\tilde{\nu}\}}{\text{std}\{\tilde{\nu}\}} = \frac{E\{\tilde{\nu}\}-0}{\text{std}\{\tilde{\nu}\}-0}$  is in bijection with  $\mu$ . This means that  $\mu$  can be univocally determined given  $\rho$ . Note that this ratio is scale-invariant and that, in particular,  $\rho = \frac{E\{\tilde{\nu}\}}{\text{std}\{\tilde{\nu}\}} = \frac{\tilde{y}\sigma^{-1}(y)}{\tilde{\sigma}(\tilde{y})\sigma^{-1}(y)} = \frac{\tilde{y}}{\tilde{\sigma}(\tilde{y})}$ . Therefore, given both  $\tilde{y}$  and  $\tilde{\sigma}(\tilde{y})$ , we can obtain  $\mu$  and hence also  $E\{\tilde{\nu}\}$  and  $\text{std}\{\tilde{\nu}\}$ .

In Figure 9 we show the plots of  $\mathcal{E}_r$  and  $\mathcal{S}_r$  which represent  $\frac{\mu}{E\{\tilde{\nu}\}} = \frac{E\{\nu\}}{E\{\tilde{\nu}\}}$  and  $\text{std}\{\tilde{\nu}\}$  as functions of  $\rho$ , respectively.

From the definition of  $\mathcal{E}_r$  follows that  $y = E\{z\} = \sigma(y) E\{\nu\} = \sigma(y) \mu = \sigma(y) E\{\tilde{\nu}\} \mathcal{E}_r(\rho)$ . Substituting  $\tilde{y} = E\{\tilde{z}\} = \sigma(y) E\{\tilde{\nu}\}$  in the previous equation (observe that, at this stage,  $\sigma(y)$  is considered as unknown) we obtain

$$y = \tilde{y} \mathcal{E}_r(\rho) = \tilde{y} \mathcal{E}_r \left( \frac{\tilde{y}}{\tilde{\sigma}(\tilde{y})} \right). \quad (37)$$

Analogously for the standard deviation,  $\sigma(y) = \text{std}\{z\} = \sigma(y) \text{std}\{\nu\} = \sigma(y) \frac{\text{std}\{\tilde{\nu}\}}{\mathcal{S}_r(\rho)}$ . Substituting  $\tilde{\sigma}(\tilde{y}) = \text{std}\{\tilde{z}\} = \sigma(y) \text{std}\{\tilde{\nu}\}$  we have

$$\sigma(y) = \text{std}\{z\} = \frac{\tilde{\sigma}(\tilde{y})}{\mathcal{S}_r(\rho)} = \frac{\tilde{\sigma}(\tilde{y})}{\mathcal{S}_r \left( \frac{\tilde{y}}{\tilde{\sigma}(\tilde{y})} \right)}. \quad (38)$$

The Equations (37) and (38) define the transformation that brings the clipped standard deviation curve  $(\tilde{y}, \tilde{\sigma}(\tilde{y}))$  to its non-clipped counterpart  $(y, \sigma(y))$ .

2) *Clipping from above (right single censoring):* The case of clipping from above,  $\tilde{z} = \min(1, z)$ , can be treated exactly as the clipping from below, provided simple manipulations and the following obvious change of variables:

$$\begin{aligned} y &\longleftrightarrow 1-y, & z &\longleftrightarrow 1-z, \\ \tilde{y} &\longleftrightarrow 1-\tilde{y}, & \tilde{z} &\longleftrightarrow 1-\tilde{z}. \end{aligned}$$

a) Direct transformation ( $\tilde{y}$  and  $\tilde{\sigma}$  from  $y$  and  $\sigma$ ):

$$\tilde{y} = 1 - \sigma(y) \mathcal{E}_m \left( \frac{1-y}{\sigma(y)} \right), \quad (39)$$

$$\tilde{\sigma}(\tilde{y}) = \sigma(y) \mathcal{S}_m \left( \frac{1-y}{\sigma(y)} \right), \quad (40)$$

$$\tilde{\sigma}(\tilde{y}) = \sigma(y) \mathcal{S}_e \left( \frac{1-\tilde{y}}{\sigma(y)} \right). \quad (41)$$

b) Inverse transformation ( $\sigma$  and  $y$  from  $\tilde{\sigma}$  and  $\tilde{y}$ ):

$$y = 1 - (1 - \tilde{y}) \mathcal{E}_r \left( \frac{1-\tilde{y}}{\tilde{\sigma}(\tilde{y})} \right), \quad (42)$$

$$\sigma(y) = \text{std}\{z\} = \frac{\tilde{\sigma}(\tilde{y})}{\mathcal{S}_r \left( \frac{1-\tilde{y}}{\tilde{\sigma}(\tilde{y})} \right)}, \quad (43)$$

3) Combined clipping from above and below (double censoring): The formulas for the two separate clippings, the one from below and the one from above, can be combined into “universal” formulas which can be applied to data which is clipped in any of the two ways. Here, we undertake the assumption, discussed in Section IV-B, that the product probability of  $z$  being clipped both from above and from below is negligibly small.

a) Direct transformation ( $\tilde{y}$  and  $\tilde{\sigma}$  from  $y$  and  $\sigma$ ): Since only one kind of clipping can happen for a given  $y$ , it means that either (34) or (39) is equal to  $y$ . Therefore, Equations (37) and (42) can be combined by summing the two right-hand sides and subtracting  $y$ ,

$$\tilde{y} = \sigma(y) \mathcal{E}_m \left( \frac{y}{\sigma(y)} \right) - y + 1 - \sigma(y) \mathcal{E}_m \left( \frac{1-y}{\sigma(y)} \right). \quad (44)$$

Similarly, (35) and (40) cannot be simultaneously different than  $\sigma(y)$ . So do (36) and (41). It means that their combinations are simply the products of the respective factors in the right-hand sides:

$$\tilde{\sigma}(\tilde{y}) = \sigma(y) \mathcal{S}_m \left( \frac{y}{\sigma(y)} \right) \mathcal{S}_m \left( \frac{1-y}{\sigma(y)} \right), \quad (45)$$

$$\tilde{\sigma}(\tilde{y}) = \sigma(y) \mathcal{S}_e \left( \frac{\tilde{y}}{\sigma(y)} \right) \mathcal{S}_e \left( \frac{1-\tilde{y}}{\sigma(y)} \right). \quad (46)$$

b) Inverse transformation ( $\sigma$  and  $y$  from  $\tilde{\sigma}$  and  $\tilde{y}$ ):

Analogous considerations hold also for combining of Equation (37) with (42) and Equation (38) with (43). Consequently, we have

$$y = \tilde{y} \mathcal{E}_r \left( \frac{\tilde{y}}{\tilde{\sigma}(\tilde{y})} \right) - \tilde{y} + 1 - (1 - \tilde{y}) \mathcal{E}_r \left( \frac{1-\tilde{y}}{\tilde{\sigma}(\tilde{y})} \right), \quad (47)$$

$$\sigma(y) = \text{std}\{z\} = \frac{\tilde{\sigma}(\tilde{y})}{\mathcal{S}_r \left( \frac{\tilde{y}}{\tilde{\sigma}(\tilde{y})} \right) \mathcal{S}_r \left( \frac{1-\tilde{y}}{\tilde{\sigma}(\tilde{y})} \right)}. \quad (48)$$

### C. Expectation and standard deviation in the wavelet domain

All the above results are valid also in the more general case where the mean and the standard deviation are not calculated for  $\tilde{\nu}$ , but rather from the corresponding detail or approximation wavelet coefficients, respectively. More precisely,

$$E\{\tilde{\nu}\} = E\{\tilde{\nu} \otimes \varphi\}, \quad \text{std}\{\tilde{\nu}\} = \text{std}\{\tilde{\nu} \otimes \varphi\},$$

since these equalities follow from the independence of  $\tilde{\nu}$  and on the normalizations  $\sum \varphi = 1$  and  $\|\psi\|_2 = 1$ . Therefore, in the next section, we consider the wavelet coefficients

calculated from the clipped observations:

$$\tilde{z}^{\text{wdet}} = \downarrow_2 (\tilde{z} \otimes \psi), \quad \tilde{z}^{\text{wapp}} = \downarrow_2 (\tilde{z} \otimes \varphi).$$

### V. ALGORITHM: CLIPPED CASE

Our goal is to estimate the functions  $\tilde{\sigma}$  and  $\sigma$  which correspond to the clipped observation model (30) from the clipped image  $\tilde{z}$ .

Pragmatically, we approach the problem using the estimators  $\hat{y}_i$  (15) and  $\hat{\sigma}_i$  (16) of mean and standard-deviation, without any particular modification. Because of clipping, these are no longer unbiased estimators of  $y_i$  and  $\sigma(y_i)$ . However, as discussed below, they can be treated as unbiased estimators of the unknown variable  $\tilde{y}_i$ , defined analogously to (14) as

$$\tilde{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} E\{\tilde{z}^{\text{wapp}}(x_j)\}, \quad \{x_j\}_{j=1}^{n_i} = S_i,$$

and of its associated standard deviation  $\tilde{\sigma}(\tilde{y}_i)$ .

Exploiting the transformations defined in the previous section and by modeling the statistics of the estimates computed from the wavelet coefficients of clipped variables, we modify the likelihood function (22) and the least-squares normal equations (Section III-D.1). Thus, we come to the desired estimates  $\hat{\sigma}_{\text{fit}}$  of  $\sigma$  and  $\hat{\tilde{\sigma}}_{\text{fit}}$  of  $\tilde{\sigma}$ .

#### A. Local estimation of expectation/standard-deviation pairs

1) Estimate of  $\tilde{\sigma}(\tilde{y}_i)$ : The standard-deviation estimator (16) is an asymptotically (for large samples) unbiased estimator of the standard deviation regardless of their particular distribution. However, for finite samples, we can guarantee unbiasedness only when the samples are normally distributed.

On this respect, applying the estimator on the wavelet detail coefficients  $\tilde{z}^{\text{wdet}}$  (rather than directly on  $\tilde{z}$ ) has the important beneficial effect of “Gaussianizing” the analyzed data, essentially by the central-limit theorem. In practice, the larger is the support of the filter  $\psi$ , the closer to a normal is the distribution of  $\tilde{z}^{\text{wdet}}$ . To make the issue transparent, let us consider the example of a constant  $y(x) \equiv y, \forall x \in X$ , and restrict our attention to the clipping from below (single left censoring). According to the models (1) and (30),  $\text{var}\{z(x)\} = \sigma^2(y)$  and  $\text{var}\{\tilde{z}(x)\} = \tilde{\sigma}^2(\tilde{y})$  are also obviously constant. Then, provided that  $\psi$  has zero mean and  $\|\psi\|_2 = 1$ , we have that  $\tilde{z}^{\text{wdet}}$  has a distribution that approaches, for an enlarging support of  $\psi$ , the normal distribution  $\mathcal{N}(0, \tilde{\sigma}^2)$ . Indeed, the probability density  $f_{\tilde{z}^{\text{wdet}}}$  of  $\tilde{z}^{\text{wdet}}$  can be calculated as the generalized cascaded convolutions of the densities  $f_{\psi(j)\tilde{z}}$ ,  $j = 1, \dots, n_\psi$  of  $n_\psi$  clipped normal distributions, where  $n_\psi$  is the number of non-zero elements  $\psi(\cdot)$  of the wavelet  $\psi$ . We remark that all these densities are generalized functions with a scaled Dirac impulse at 0. From (31), we have that the impulse in  $f_{\psi(j)\tilde{z}}$  is  $\Phi(-y/\sigma(y)) \delta_0$  (note that the scale of the impulse does not depend on  $\psi$ ). Because of the independence of  $\tilde{z}$ , the probability  $P(\tilde{z}^{\text{wdet}} = 0)$  is the product probability  $\prod_j P(\psi(j)\tilde{z} = 0) = \Phi(-y/\sigma(y))^{n_\psi}$ , thus the impulse in  $f_{\tilde{z}^{\text{wdet}}}$  is  $\Phi(-y/\sigma(y))^{n_\psi} \delta_0$ , showing that the discrete part of the distribution vanishes at exponential rate with  $n_\psi$ . The convergence to a normal distribution is rather fast, and even for small wavelet kernels such as  $\psi = \psi_1 \otimes \psi_1^T$  with  $\psi_1$  defined

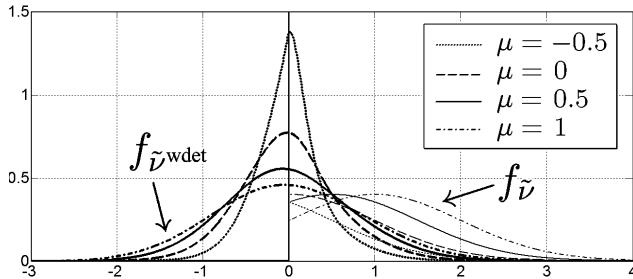


Fig. 10. Probability densities of the clipped (from below)  $\tilde{\nu} = \tilde{z}/\sigma(y)$  (thin lines) and of its wavelet detail coefficients  $\tilde{\nu}^{\text{wdet}} = \tilde{z}^{\text{wdet}}/\sigma(y) = \downarrow_2(\tilde{z} \otimes \psi)/\sigma(y)$  (thick lines) for different values of  $\mu = y/\sigma(y)$ , when  $\psi$  is the 2-D Daubechies wavelet. The “Dirac peaks” at 0, characteristic of these densities, appear here as vertical asymptotes at 0 and cannot thus be seen in the drawing.

as (10), for which  $n_\psi = 36$ , the distribution of  $\tilde{z}^{\text{wdet}}$  is very similar to a normal for values  $\mu = \frac{y}{\sigma(y)}$  as low as 0 (observe that the larger is  $\mu$ , the closer is the normal approximation), as shown in Figure 10<sup>7</sup>. Note that for  $\mu = -0.5, 0, 0.5, 1$  and  $n_\psi = 36$ , the amplitudes of the step discontinuity at 0 in the distribution of  $\tilde{z}^{\text{wdet}}$  are  $\Phi(\mu)^{n_\psi} \simeq 1.7 \cdot 10^{-6}, 1.5 \cdot 10^{-11}, 4.1 \cdot 10^{-19}, 1.6 \cdot 10^{-29}$ , respectively, thus all these distributions are practically continuous and therefore the plots in the figure are a faithful illustration of the generalized probability densities  $f_{\tilde{z}^{\text{wdet}}}$ .

The described “Gaussianization” is important, because it ensures that the bias due to finite samples is not significant, allowing to use the same constant  $\kappa_n$  (17) as in the non-clipped case. As a rough quantitative figure of the error which may come from this simplification, in Table I we give the values of the expectation<sup>8</sup>  $E\{\hat{\sigma}_i\}/\sigma(y_i)$  for different combinations of  $\mu = y/\sigma(y)$  and  $n_i$ . The cases “ $n_i = \infty$ ” correspond to the true values of the standard deviation  $\tilde{\sigma}(\tilde{y})/\sigma(y) = \text{std}\{\tilde{z}\}/\sigma(y) = \text{std}\{\tilde{\nu}\} = \mathcal{S}_m(\mu) = \mathcal{S}_m(y/\sigma(y))$  of the clipped data, calculated from (33) and plotted in Figure 7. From the table one can see that a handful of samples are sufficient for the finite-sample estimation bias  $E\{\hat{\sigma}_i\} - \text{std}\{\tilde{\nu}\}$  to be negligible.

2) *Estimate of  $\hat{y}_i$ :* Let us now consider the estimates of the mean. Clearly, being a sample average,  $\hat{y}_i$  is an unbiased estimate of  $\tilde{y}_i$ , regardless of the number of samples  $n_i$  or of the distribution of  $\tilde{z}$ . The central-limit theorem and similar arguments as above show that  $\tilde{z}^{\text{wapp}}$  and  $\hat{y}_i$  are both normally distributed with mean  $\tilde{y}_i$ .

3) *Variance of the estimates:* Ignoring the possible dependence of the noise in the wavelet coefficients (due to non-

<sup>7</sup>In [5], we consider the case where the wavelet  $\psi$  is replaced by the basis elements of the 2-D discrete cosine transform (DCT) and show analogous figures for  $\mu = 0$  and transforms of size  $n \times n$ ,  $n = 2, 3, 4$ . The “Gaussianization” is observable there as well, particularly for the AC terms.

<sup>8</sup>The finite-sample numbers in Tables I and III are obtained by Monte Carlo simulations. The simulations were computed with enough replications to have a sample standard-deviation of the averages lower than 0.0001. Thus, the numbers given in the tables can be considered as precise for all shown digits. The taken samples  $z^{\text{wdet}}$  were contiguous in the set  $S_i$ , therefore some dependence was present (exact independence is found only for samples farther than the diameter of the support of  $\psi$ , because in the considered case the distribution of  $z$  is not normal), however, as  $n_i$  grows the dependence becomes negligible, since  $z^{\text{wdet}}$  can be split in  $n_\psi/4 = 9$  subsets, each with a growing number of fully independent samples.

Gaussianity of the clipped variables), simple estimates of the variances of  $\hat{y}_i$  and  $\hat{\sigma}_i$  can be obtained from the variances  $\|\varphi\|_2^2 \tilde{\sigma}^2(\tilde{y}_i)$  and  $\tilde{\sigma}^2(\tilde{y}_i)$  of the wavelet coefficients  $\tilde{z}^{\text{wapp}}$  and  $\tilde{z}^{\text{wdet}}$ , respectively, as in Section III-C.4.

4) *Distribution of the estimates:* In conclusion, similar to Section III-C.5, we model the distributions of the estimates  $\hat{y}_i$  and  $\hat{\sigma}_i$  as the normal

$$\hat{y}_i \sim \mathcal{N}(\tilde{y}_i, \tilde{\sigma}^2(\tilde{y}_i) c_i), \quad \hat{\sigma}_i \sim \mathcal{N}(\tilde{\sigma}(\tilde{y}_i), \tilde{\sigma}^2(\tilde{y}_i) d_i), \quad (49)$$

where the factors  $c_i$  and  $d_i$  are defined as in (19).

### B. Maximum-likelihood fitting of the clipped model

It is straightforward to exploit the above analysis for the estimation of the functions  $\tilde{\sigma}$  (30) and  $\sigma$  (1) from the clipped data  $\tilde{z} = \max(0, \min(z, 1))$ . In fact, for the ML solution (24), it suffices to introduce the functions  $\mathcal{E}_m$  and  $\mathcal{S}_m$  into the definition of the function to be fitted to the measured data, which are pairs  $(\hat{y}_i, \hat{\sigma}_i)$  centered—according to (49)—at  $(\tilde{y}_i, \tilde{\sigma}(\tilde{y}_i))$ . From (45) follows that we can define  $\tilde{\sigma}_{\text{reg}}(\tilde{y})$  as

$$\tilde{\sigma}_{\text{reg}}(\tilde{y}) = \sigma_{\text{reg}}(y) \mathcal{S}_m\left(\frac{y}{\sigma_{\text{reg}}(y)}\right) \mathcal{S}_m\left(\frac{1-y}{\sigma_{\text{reg}}(y)}\right), \quad (50)$$

where the argument  $\tilde{y}$  is, according to (44),

$$\tilde{y} = \sigma_{\text{reg}}(y) \mathcal{E}_m\left(\frac{y}{\sigma_{\text{reg}}(y)}\right) - y + 1 - \sigma_{\text{reg}}(y) \mathcal{E}_m\left(\frac{1-y}{\sigma_{\text{reg}}(y)}\right). \quad (51)$$

The conditional probability density (22) is thus modified into

$$\wp((\hat{y}_i, \hat{\sigma}_i) | \tilde{y}_i = \tilde{y}) = \wp(\hat{y}_i | \tilde{y}_i = \tilde{y}) \wp(\hat{\sigma}_i | \tilde{y}_i = \tilde{y}) = \quad (52)$$

$$= \frac{1}{2\pi\sqrt{c_i d_i}} \frac{1}{\tilde{\sigma}_{\text{reg}}^2(\tilde{y})} e^{-\frac{1}{2\tilde{\sigma}_{\text{reg}}^2(\tilde{y})} \left( \frac{(\hat{y}_i - \tilde{y})^2}{c_i} + \frac{(\hat{\sigma}_i - \tilde{\sigma}_{\text{reg}}(\tilde{y}))^2}{d_i} \right)}.$$

Analogously to (23), the posterior likelihood  $L$  is obtained by considering all measurements  $\{(\hat{y}_i, \hat{\sigma}_i)\}_{i=1}^N$  and by integrating the densities  $\wp((\hat{y}_i, \hat{\sigma}_i) | \tilde{y}_i = \tilde{y})$  with respect to a prior  $\wp_0(y)$  as

$$\tilde{L}(a, b) = \prod_{i=1}^N \int \wp((\hat{y}_i, \hat{\sigma}_i) | \tilde{y}_i = \tilde{y}) \wp_0(y) dy. \quad (53)$$

Note that the integration in (53) is still with respect to  $y$  and that  $\wp((\hat{y}_i, \hat{\sigma}_i) | \tilde{y}_i = \tilde{y})$  is itself an explicit function of  $y$ , as it is clear from Equations (50-52). Therefore, (53) allows for direct calculation, and by solving (24) with the likelihood  $\tilde{L}(a, b)$  in place of  $L(a, b)$  (23) we obtain the parameters  $\hat{a}$  and  $\hat{b}$ , which define both the ML estimate  $\hat{\sigma}_{\text{fit}}$  of  $\sigma$ , exactly as in (25), and the ML estimate  $\hat{\tilde{\sigma}}_{\text{fit}}$  of  $\tilde{\sigma}$ , which can be obtained from  $\hat{\sigma}_{\text{fit}}$  by application of the transformations (44) and (45). Note that for the clipped raw-data it is unnatural to assume that  $\wp_0$  is uniform on  $[0, 1]$ , because in the case of overexposure the true signal could be much larger than 1. Therefore, for the clipped raw-data, we assume that all positive values of  $y$  are equiprobable and we maximize<sup>9</sup>

$$\tilde{L}(a, b) = \prod_{i=1}^N \int_0^{+\infty} \wp((\hat{y}_i, \hat{\sigma}_i) | \tilde{y}_i = \tilde{y}) dy.$$

1) *Least-squares initialization:* Similar to the non-clipped case, we use a simple least-squares solution as the initial condition for the iterative maximization of the likelihood

<sup>9</sup>Equivalently, we maximize

$$\tilde{L}(a, b) = \lim_{j \rightarrow +\infty} \prod_{i=1}^N \int_0^{1+j} \wp((\hat{y}_i, \hat{\sigma}_i) | \tilde{y}_i = \tilde{y}) \wp_j(y) (1+j) dy,$$

where  $\wp_j$  is a uniform density on  $[0, 1+j]$  and the normalization factor  $(1+j)$  enables the convergence of the sequence of integrals.

$n_i$	$E\{\hat{\sigma}_i\}/\sigma(y_i)$ for $n < \infty$ ,					$\tilde{\sigma}(\tilde{y})/\sigma(y) = \text{std}\{\tilde{z}\}/\sigma(y) = S_m(y/\sigma(y))$ for $n = \infty$				
	$\mu = -1$	$\mu = -0.5$	$\mu = 0$	$\mu = 0.5$	$\mu = 1$	$\mu = 1.5$	$\mu = 2$	$\mu = 2.5$	$\mu = 3$	$\mu = 5$
2	0.226	0.389	0.572	0.741	0.868	0.946	0.982	0.995	0.999	1.000
10	0.247	0.404	0.580	0.743	0.867	0.944	0.981	0.995	0.999	1.000
50	0.258	0.411	0.583	0.744	0.867	0.943	0.980	0.994	0.999	1.000
100	0.260	0.412	0.583	0.744	0.867	0.943	0.980	0.994	0.999	1.000
500	0.261	0.413	0.584	0.744	0.867	0.943	0.980	0.994	0.999	1.000
$\infty$	0.262	0.413	0.584	0.744	0.867	0.943	0.980	0.994	0.999	1.000

TABLE I

EXPECTATION  $E\{\hat{\sigma}_i\}/\sigma(y_i)$  FOR DIFFERENT COMBINATIONS OF  $\mu = y_i/\sigma(y_i)$  AND  $n_i$ . THE CASES “ $n_i = \infty$ ” CORRESPOND TO THE TRUE VALUES OF THE STANDARD DEVIATION  $\tilde{\sigma}(\tilde{y})/\sigma(y) = \text{std}\{\tilde{z}\}/\sigma(y) = \text{std}\{\tilde{\nu}\} = S_m(\mu) = S_m(y/\sigma(y))$  OF THE CLIPPED DATA, CALCULATED FROM (33) AND PLOTTED IN FIGURE 7.

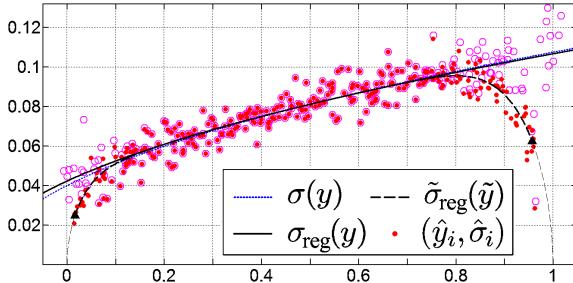


Fig. 11. Estimation with clipped observations  $\tilde{z}$  (Fig. 2): Least-squares initialization. Each dot of the scatter plot corresponds to a pair  $(\hat{y}_i, \hat{\sigma}_i)$  of estimates of  $\tilde{y}_i$  and  $\tilde{\sigma}(\tilde{y}_i)$ . The circles indicate these pairs of estimates after inverse-transformation (see Eqs. (55) and (56)). The solid line shows the square root of the least-squares estimate of the variance function  $\sigma^2$  (see Section V-B.1),  $\hat{a}_0 = 0.00945$  ( $\hat{\chi}_0 = 105.82$ ),  $\hat{b}_0 = 0.001822$  ( $\sqrt{\hat{b}_0} = 0.04268$ ). The dotted line is the true  $\sigma$ , while the dashed-line is the function  $\tilde{\sigma}_{\text{reg}}$  with parameters  $\hat{a}_0, \hat{b}_0$  used as initial condition for the iterative maximization of the likelihood (53).

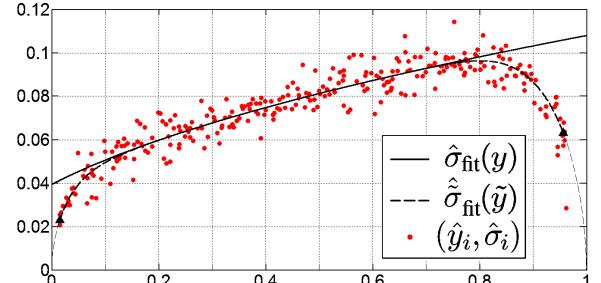


Fig. 12. Estimation with clipped observations  $\tilde{z}$  (Fig. 2): ML solution. Each dot of the scatter plot corresponds to a pair  $(\hat{y}_i, \hat{\sigma}_i)$  of estimates of  $\tilde{y}_i$  and  $\tilde{\sigma}(\tilde{y}_i)$ . The solid line and dashed line show the maximum-likelihood estimates  $\hat{\sigma}_{\text{fit}}$  and  $\hat{\sigma}_{\text{fit}}$  of the standard-deviation functions  $\sigma$  and  $\tilde{\sigma}$ , respectively.  $\hat{a} = 0.00995$  ( $\hat{\chi} = 100.52$ ),  $\hat{b} = 0.001552$  ( $\sqrt{\hat{b}} = 0.03940$ ). The plot of  $\hat{\sigma}_{\text{fit}}$  overlaps perfectly with that of the true  $\sigma$ .

function. We exploit the inverse transformations (47)-(48) from Section IV-B.3.b to attain a fit of  $\sigma^2$  with respect to the non-clipped variables. Hence, the initial parameters are given as  $[\hat{a}_0, \hat{b}_0] = \hat{\mathbf{v}}\Phi(\Phi^T\Phi)^{-1}$ , with the dependent and independent variables transformed as

$$\rho_{i,0}^{\text{reg}} = \sqrt{\left(\frac{\hat{y}_i}{\hat{\sigma}_i}\right)^2 + \varepsilon_{\text{reg}}^2}, \quad \rho_{i,1}^{\text{reg}} = \sqrt{\left(\frac{1 - \hat{y}_i}{\hat{\sigma}_i}\right)^2 + \varepsilon_{\text{reg}}^2}, \quad (54)$$

$$\Phi = \begin{bmatrix} \hat{y}_1 \mathcal{E}_r(\rho_{1,0}^{\text{reg}}) - \hat{y}_1 + 1 - (1 - \hat{y}_1) \mathcal{E}_r(\rho_{1,1}^{\text{reg}}) & 1 \\ \hat{y}_2 \mathcal{E}_r(\rho_{2,0}^{\text{reg}}) - \hat{y}_2 + 1 - (1 - \hat{y}_2) \mathcal{E}_r(\rho_{2,1}^{\text{reg}}) & 1 \\ \vdots & \vdots \end{bmatrix}, \quad (55)$$

$$\hat{\mathbf{v}} = \begin{bmatrix} \frac{\kappa_{n_1}^2 \hat{\sigma}_1^2}{(\mathcal{S}_r(\rho_{1,0}^{\text{reg}}) \mathcal{S}_r(\rho_{1,1}^{\text{reg}}))^2} & \frac{\kappa_{n_2}^2 \hat{\sigma}_2^2}{(\mathcal{S}_r(\rho_{2,0}^{\text{reg}}) \mathcal{S}_r(\rho_{2,1}^{\text{reg}}))^2} & \dots \end{bmatrix}. \quad (56)$$

Figures 11 and 12 respectively show the initial  $\tilde{\sigma}_{\text{reg}}$ , which corresponds to the parameters  $\hat{a}_0, \hat{b}_0$ , and the ML estimates  $\hat{\sigma}_{\text{fit}}$  and  $\hat{\sigma}_{\text{fit}}$  found using  $\tilde{\sigma}_{\text{reg}}$  as initialization in the iterative maximization of the likelihood. In Figure 11 we can see that the inverse transformations (47)-(48) used in (55) and (56) effectively move the clipped estimates pairs near to their respective “non-clipped” positions. Note also the increased accuracy of the ML estimates compared to that of the least-squares ones.

## VI. ROBUST ESTIMATES

Despite the removal of edges from  $X^{\text{smo}}$ , small singularities or fine textures and edges of the image can still be present in  $z^{\text{wdet}}$ , within  $S_i$ . The accuracy of the sample standard-deviation estimator (16) is consequently degraded, since  $z^{\text{wdet}}$  would contain wild errors of large amplitude, which can cause the distribution of  $z^{\text{wdet}}$  to become heavy-tailed. This typically leads to an over-estimate of the standard-deviation. It is well-known that robust estimators based on order-statistics can effectively deal with these situations.

### A. Robust standard-deviation estimates

To reduce the influence of these wild errors, we replace the sample standard-deviation estimator (16) with the robust estimator based on the median of the absolute deviations (MAD) [12], [9]<sup>10</sup>

$$\hat{\sigma}_i^{\text{mad}} = \frac{1}{\kappa_{n_i}^{\text{mad}}} \text{median}_{x_j \in S_i} \{|z^{\text{wdet}}(x_j)|\}, \quad (57)$$

where  $\kappa_{n_i}^{\text{mad}}$  is again a scaling factor to make the estimator unbiased. It is well known that, for large normally i.i.d. samples,  $\kappa_n^{\text{mad}}$  approaches  $\Phi^{-1}(3/4) = 0.6745$ , where  $\Phi^{-1}$  is the inverse c.d.f. of the standard normal. For small finite

<sup>10</sup>In its general form, this estimator is defined as  $\frac{1}{\kappa_{n_i}^{\text{mad}}} \text{median}_{x_j \in S_i} |z^{\text{wdet}}(x_j) - \text{median}_{x_j \in S_i} (z^{\text{wdet}}(x_j))|$ . However, when used on wavelet detail coefficients, the subtraction of the median in the deviation is often discarded, since its expected value for these coefficients is typically zero.

$n$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$\kappa_n^{\text{mad}}$	0.798	0.732	0.712	0.702	0.696	0.693	0.690	0.688	0.686	0.685										

TABLE II

BIAS FACTOR  $\kappa_n^{\text{mad}}$  FOR THE MAD ESTIMATOR (57) FOR SMALL FINITE SAMPLES OF  $n$  INDEPENDENT NORMAL VARIABLES;  $\kappa_{2n}^{\text{mad}} = \kappa_{2n-1}^{\text{mad}}$ .

$n_i$	$E\{\hat{\sigma}_i^{\text{mad}}\}/\sigma(y_i)$									
	$\mu = -1$	$\mu = -0.5$	$\mu = 0$	$\mu = 0.5$	$\mu = 1$	$\mu = 1.5$	$\mu = 2$	$\mu = 2.5$	$\mu = 3$	$\mu = 5$
2	0.215	0.380	0.566	0.739	0.869	0.946	0.983	0.995	0.999	1.000
10	0.168	0.341	0.543	0.733	0.872	0.952	0.985	0.997	0.999	1.000
50	0.158	0.332	0.538	0.731	0.873	0.953	0.986	0.997	0.999	1.000
100	0.156	0.330	0.537	0.731	0.873	0.953	0.986	0.997	0.999	1.000
500	0.155	0.329	0.536	0.731	0.873	0.953	0.986	0.997	0.999	1.000
500000	0.154	0.329	0.536	0.731	0.873	0.953	0.986	0.997	0.999	1.000

TABLE III

EXPECTATION  $E\{\hat{\sigma}_i^{\text{mad}}\}/\sigma(y_i)$  FOR DIFFERENT COMBINATIONS OF  $\mu = y_i/\sigma(y_i)$  AND  $n_i$ .

samples, the values of  $\kappa_n^{\text{mad}}$  are larger and up to  $\sqrt{2/\pi} = 0.7979$  (mean of absolute deviations of  $\mathcal{N}(0, 1)$ ); in Table II we give the values of  $\kappa_n^{\text{mad}}$  for  $n = 1, \dots, 20$ . For larger  $n$ , we can approximate  $\kappa_n^{\text{mad}}$  as  $\kappa_n^{\text{mad}} \simeq \frac{1}{5n} + \Phi^{-1}(3/4)$ . Note that  $\kappa_{2n}^{\text{mad}} = \kappa_{2n-1}^{\text{mad}}$ ; this is because in a set of  $2n$  independent random variables, any individual variable has probability 0.5 of belonging to the subset of  $n$  variables smaller (or larger) than the median value. Tables similar to Table II can be found in [14] for a few other estimators of the standard deviation.

1) *Variance and distribution of the standard-deviation estimates:* The variance of the robust estimates (57) can be approximated<sup>11</sup> as

$$\text{var}\{\hat{\sigma}_i^{\text{mad}}\} = \sigma^2(y_i) d_i^{\text{mad}}, \quad d_i^{\text{mad}} \simeq \frac{1.35}{n_i + 1.5}. \quad (58)$$

Thus, we pay the increased robustness with respect to outliers with a more than twice as large variance of the estimates, in comparison to (19) (this larger variance can be seen clearly by visual comparison of Figures 12 and 14). However, in practice, when working with many-megapixels images, the variance (58) is often quite low, due to the large number of samples  $n_i$ . Hence, the use of the robust estimator is ordinarily recommendable.

Like the sample standard-deviation estimates, also the MAD estimates (57) have a distribution which can be approximated by a normal<sup>12</sup>. In particular (and analogous to (20)),

$$\hat{\sigma}_i^{\text{mad}} \sim \mathcal{N}(\sigma(y_i), \sigma^2(y_i) d_i^{\text{mad}}).$$

2) *Estimates of the variance:* An unbiased robust estimate of the variance (as used by the least-squares initialization (26)) can be obtained from the squared  $(\hat{\sigma}_i^{\text{mad}})^2$  (57), provided multiplication with a bias correction factor. Using the same symbols of Section III-D.1, we denote this estimate of the variance as  $(\check{\kappa}_{n_i}^{\text{mad}} \hat{\sigma}_i^{\text{mad}})^2$ , where the factor  $(\check{\kappa}_n^{\text{mad}})^2$  can be

<sup>11</sup>The approximation of  $d_i^{\text{mad}}$  in (58) can be obtained by Monte Carlo simulations. A table with few of these values is found also in [14].

<sup>12</sup>The normal approximation can be easily verified by numerical simulations. Despite all the necessary ingredients for an analytical proof can be found in [7], it seems that that this result is not explicitly reported in the literature.

approximated as  $(\check{\kappa}_n^{\text{mad}})^2 \simeq 1 + \frac{1}{5n}$ .

### B. Maximum-likelihood fitting (non-clipped)

The ML solution is found exactly as in Section III-D, provided that the estimates and factors  $\hat{\sigma}_i, d_i, \check{\kappa}_{n_i}$  are replaced by their respective “mad” counterparts  $\hat{\sigma}_i^{\text{mad}}, d_i^{\text{mad}}, \check{\kappa}_{n_i}^{\text{mad}}$  in Equations (22-24) and (28).

### C. Clipped observations

Let us now apply the MAD estimator (57) to the wavelet coefficients  $\tilde{z}^{\text{wdet}} = \downarrow_2(\tilde{z} \circledast \psi)$  of the clipped observations  $\tilde{z}$ ,

$$\hat{\sigma}_i^{\text{mad}} = \frac{1}{\kappa_{n_i}^{\text{mad}}} \text{median}_{x_j \in S_i} \{|\tilde{z}^{\text{wdet}}(x_j)|\}. \quad (59)$$

Although robust with respect to outliers of large amplitude, the MAD estimator is sensitive to the asymmetry in the distribution of the samples [14] and even the limiting value  $\Phi^{-1}(3/4)$  is, as one can expect from the presence of the inverse c.d.f. of the normal distribution, essentially correct for normally distributed samples only. Thus, contrary to the sample standard-deviation, the MAD estimator is not asymptotically unbiased:

$$\lim_{n_i \rightarrow \infty} E\{\hat{\sigma}_i^{\text{mad}}\} \neq \text{std}\{\tilde{z}^{\text{wdet}}\} = \lim_{n_i \rightarrow \infty} E\{\hat{\sigma}_i\}. \quad (60)$$

Let us investigate this estimation bias for large as well as for small finite samples. As in Section V-A, we restrict ourselves to the case of clipping from below for a constant  $y(x) \equiv y, \forall x \in X$ , and apply the estimator (57) to the corresponding  $\tilde{z}^{\text{wdet}}$ . In Table III, we give the expectations  $E\{\hat{\sigma}_i^{\text{mad}}\}/\sigma(y_i)$  for different combinations of  $\mu = y_i/\sigma(y_i)$  and  $n_i$ . The same considerations which we made commenting Table I can be repeated also for Table III.

We use the expectations of large-sample estimates (values with  $n_i = 500000$  in Table III) as a numerical definition of  $\lim_{n_i \rightarrow \infty} E\{\hat{\sigma}_i^{\text{mad}}\}$ . In this way, we define the function  $S_m^{\text{mad}}$  which gives  $\lim_{n_i \rightarrow \infty} E\{\hat{\sigma}_i^{\text{mad}}\}$  as a function of  $E\{\tilde{v}\}$ . Hence, are also defined  $S_e^{\text{mad}}, S_r^{\text{mad}}$ , and  $E_r^{\text{mad}}$  and the corresponding analogs of the direct and inverse transformation

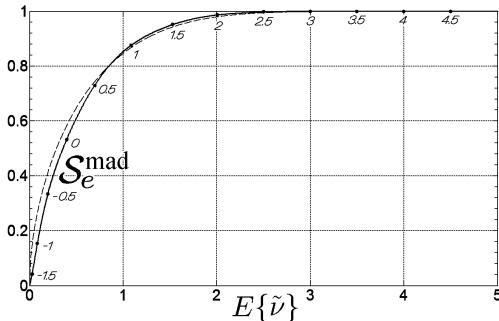


Fig. 13. Large-sample asymptotic expectation  $\lim_{n_i \rightarrow \infty} E\{\hat{\sigma}_i^{\text{mad}}\}$  of the median of the absolute deviations (57) as a function  $S_e^{\text{mad}}$  of the expectation  $E\{\tilde{\nu}\} = \tilde{y}_i/\sigma(y_i)$  of the clipped variables. The dashed line corresponds to the function  $S_e$  from Fig. (8).

formulas (45)-(48). In particular, we define  $\tilde{\sigma}^{\text{mad}}$  by

$$\tilde{\sigma}^{\text{mad}}(\tilde{y}) = \sigma(y) \mathcal{S}_m^{\text{mad}}\left(\frac{y}{\sigma(y)}\right) \mathcal{S}_m^{\text{mad}}\left(\frac{1-y}{\sigma(y)}\right).$$

In Figure 13, we show the plot of  $\mathcal{S}_e^{\text{mad}}$  superimposed on the plot of  $\mathcal{S}_e$  (dashed line). The vertical difference between the two plots in the Figure is the bias<sup>13</sup> of (57) as an estimator of  $\text{std}\{\hat{\nu}\} = \text{std}\{\hat{z}^{\text{wdet}}\}/\sigma(y)$ . These differences can also be seen (as a function of  $\mu$ ) by comparing the last rows of the Tables I and III, for the cases  $n_i = \infty$  and  $n_i = 500000$ , respectively.

For the case of the MAD estimator, formula (49) needs therefore to be modified as follows:

$$\hat{\sigma}_i^{\text{mad}} \sim \mathcal{N} \left( \tilde{\sigma}^{\text{mad}}(\tilde{y}_i), \left( \tilde{\sigma}^{\text{mad}}(\tilde{y}_i) \right)^2 d_i^{\text{mad}} \right).$$

The specific  $S_m^{\text{mad}}$  allows us to take into account of the difference  $\tilde{\sigma}^{\text{mad}}(\tilde{y}_i) - \tilde{\sigma}(\tilde{y}_i)$  (60) here and in the following ML estimation of the functions  $\sigma$  and  $\tilde{\sigma}$ .

1) *Maximum-likelihood fitting (clipped)*: The ML solution is found exactly as in Section V-B, provided that the functions  $\mathcal{S}_m^{\text{mad}}$ ,  $\mathcal{S}_r^{\text{mad}}$ , and  $\mathcal{E}_r^{\text{mad}}$  defined above and estimates and factors  $\hat{\sigma}_i^{\text{mad}}$ ,  $d_i^{\text{mad}}$ ,  $\tilde{\kappa}_{n_i}^{\text{mad}}$  are used, in place of their respective “non-robust” counterparts, in the Equations (50), (52), (53), (54), (55), and (56). The found parameters  $\hat{a}$  and  $\hat{b}$  define simultaneously three functions: from (25) we obtain  $\hat{\sigma}_{\text{fit}}$ , a ML estimate of  $\sigma$ ;  $\hat{\tilde{\sigma}}_{\text{fit}}$ , a ML estimate of  $\tilde{\sigma}$ ; and  $\hat{\hat{\sigma}}_{\text{fit}}^{\text{mad}}$ , a ML estimate of  $\tilde{\sigma}_{\text{mad}}$ .

$$\tilde{\sigma}^{\text{mad}}(\tilde{y}) = \sigma(y) \mathcal{S}_m^{\text{mad}}\left(\frac{y}{\sigma(y)}\right) \mathcal{S}_m^{\text{mad}}\left(\frac{1-y}{\sigma(y)}\right),$$

around which are scattered the estimates  $(\hat{y}_i, \hat{\sigma}_i^{\text{mad}})$ .

In Figure 14 we show the ML estimates  $\hat{\sigma}_{\text{fit}}$  and  $\hat{\sigma}_{\text{fit}}^{\text{mad}}$  obtained for the clipped  $\tilde{z}$  from Figure 2 using the MAD. We can see that, despite the larger variance of the estimates  $\hat{\sigma}_i^{\text{mad}}$  (as compared to  $\hat{\sigma}_i$  in Figure 12), the final estimated parameters and the corresponding  $\hat{\sigma}_{\text{fit}}$  are essentially the same as those obtained using the sample standard-deviation. In the Figure, note the slightly different shape of the plot of  $\hat{\sigma}_{\text{fit}}^{\text{mad}}$  compared to  $\hat{\sigma}_{\text{fit}}$ .

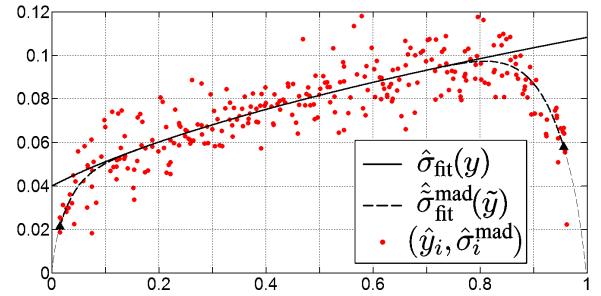


Fig. 14. Estimation with clipped observations  $\tilde{z}$  and MAD estimator (59) (Fig. 2): Maximum-likelihood solution. Each dot of the scatter plot corresponds to a pair  $(\hat{y}_i, \hat{\sigma}_i)$  of estimates of  $\tilde{y}_i$  and  $\tilde{\sigma}(\tilde{y}_i)$ . The solid line and dashed line show the maximum-likelihood estimates  $\hat{\sigma}_{fit}$  and  $\hat{\sigma}_{fit}^{mad}$  of the standard-deviation functions  $\sigma$  and  $\tilde{\sigma}$ , respectively. The plot of  $\hat{\sigma}_{fit}$  overlaps perfectly with that of the true  $\sigma$  (shown in Figure 1). The estimated parameters are  $\hat{a} = 0.01000$  ( $\hat{x} = 100.04$ ) and  $\hat{b} = 0.001594$  ( $\sqrt{\hat{b}} = 0.03992$ ).

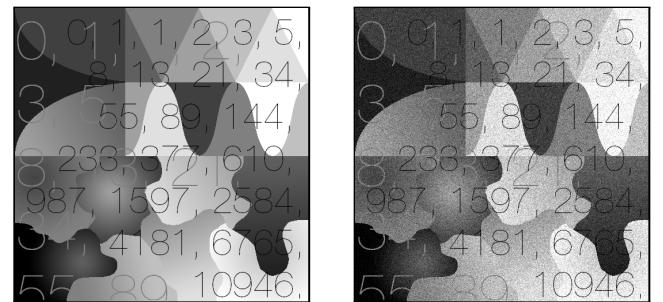


Fig. 15. The piecewise smooth test image of Fig. 2 with thin text superimposed: original  $y$  and observation  $z$  degraded by Poissonian and Gaussian noise with parameters  $\chi = 100$  ( $a = 0.01$ ) and  $b = 0.04^2$ .

#### D. Another example

To demonstrate a situation where the robust estimates are remarkably more accurate than the non-robust ones, we introduce a number of thin and sharp discontinuities in the test image, as shown in Figure 15. At many places, due to low contrast (and also due to the simplicity of our edge-detector), these discontinuities cannot be detected properly and are thus eventually incorporated in the smoothness set  $X^{\text{smo}}$ . In Figure 16, we show the estimates  $\hat{\sigma}_{\text{fit}}$  and  $\hat{\hat{\sigma}}_{\text{fit}}$  obtained using the robust and the non-robust estimator. As easily expected, the estimates  $\hat{\sigma}_i$  are inaccurate and typically biased in favour of larger standard-deviation values. As a result, the  $\hat{\sigma}_{\text{fit}}$  curve does not match with the true  $\sigma$ . The result obtained from the robust estimates  $\hat{\sigma}_i^{\text{mad}}$  is essentially better, with only a mild overestimation of the signal-dependent component of the noise.

## VII. EXPERIMENTS WITH RAW DATA

We performed extensive experiments with raw data<sup>14</sup> of various digital imaging sensors under different acquisition parameters. The devices included three CMOS sensors from

<sup>13</sup>Although it is not insignificant, this asymptotic bias is as not large as it would be if applying the MAD (57) directly on  $\tilde{z}$  instead of  $z^{\text{wdet}}$ . In fact, it is easy to realize that  $\text{median}\{\tilde{\nu}\} = 0$  for  $\mu \leq 0$ . Since obviously  $\text{median}\{|\tilde{\nu}|\} = \text{median}\{\tilde{\nu}\}$ , we have that  $\text{median}\{|\tilde{\nu}|\} = \text{median}\{|\tilde{\nu} - \text{median}\{\tilde{\nu}\}|\} = 0$  for  $\mu < 0$ .

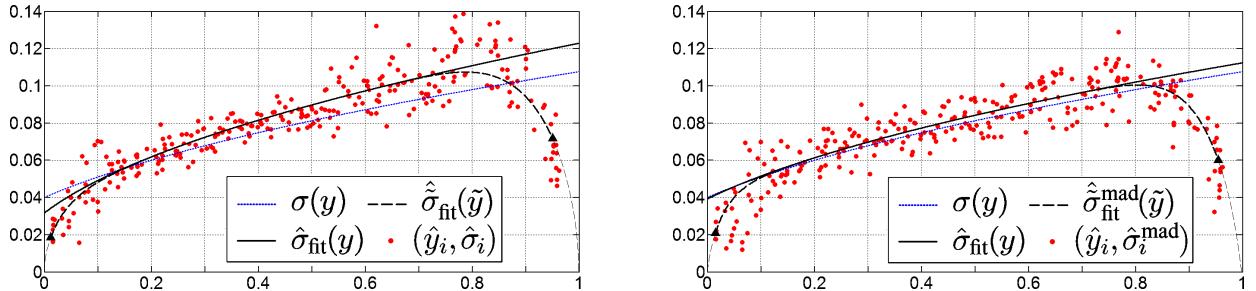


Fig. 16. Robust (right) vs. non-robust (left) estimation with clipped observations  $\tilde{z}$  (Fig. 15): ML solutions. Each dot of the scatter plot corresponds to a pair  $(\hat{y}_i, \hat{\sigma}_i)$  of estimates of  $\hat{y}_i$  and  $\hat{\sigma}(\hat{y}_i)$ . The solid line and dashed line show the maximum-likelihood estimates  $\hat{\sigma}_{\text{fit}}$  and  $\hat{\sigma}_{\text{fit}}^{\text{mad}}$  of the standard-deviation functions  $\sigma$  and  $\hat{\sigma}$ , respectively. The dotted line is the true  $\sigma$ . The parameters estimated by the two methods are  $\hat{a} = 0.01415$  ( $\hat{\chi} = 70.68$ ),  $\hat{b} = 0.000951$  ( $\sqrt{\hat{b}} = 0.03084$ ) and  $\hat{a} = 0.01108$  ( $\hat{\chi} = 90.29$ ),  $\hat{b} = 0.001524$  ( $\sqrt{\hat{b}} = 0.03904$ ), respectively.

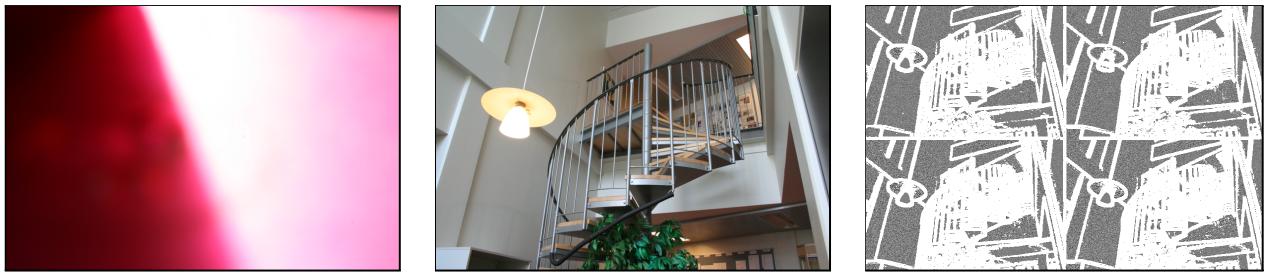


Fig. 17. From left to right: out-of-focus image with under- and overexposure (Canon EOS 350D, ISO 100), a natural image (Canon EOS 350D, ISO 1600) and its wavelet detail coefficients  $z^{\text{wdet}}$  restricted on the set of smoothness  $X^{\text{simo}}$  (the four subimages are arranged as  $[R, B; G_1, G_2]$ ).

Nokia cameraphones (300 Kpixel, 1.3 Mpixel, 5 Mpixel), Super CCD HR sensors [16] from Fujifilm FinePix S5600 (5 Mpixel) and S9600 (9.1 Mpixel) cameras, and two CMOS sensors from Canon EOS 350D and 400D SLR cameras (8 Mpixel, 10 Mpixel). In all experiments we found near-perfect fit of our proposed clipped Poissonian-Gaussian model to the data. We also compared the parametric curves  $\hat{\sigma}_{\text{fit}}$ , estimated from a single image by the proposed algorithm, with the nonparametric curves estimated by the algorithm [6] using 50 images; we found the agreement to be very good, with minor differences due to the fact that the present algorithm includes the fixed-pattern noise (FPN) in the noise estimate, whereas [6], being a pixelwise procedure, estimates only the temporal noise. Because of length limitation, we present here only few most significant examples of the obtained results.

First, we show the estimated curves for the raw-data of Canon EOS 350D with ISO 100 and 1600 (lowest and highest user-selectable analog-gain options). An out-of-focus, hence smooth, target (shown in Figure 17(left)) was used, with under- and over-exposed parts, thus providing a complete and reliable coverage of the dynamic range and beyond. Besides the excellent match between the fitted parametric curve  $\hat{\sigma}_{\text{fit}}^{\text{mad}}$  and the local estimates  $(\hat{y}_i, \hat{\sigma}_i^{\text{mad}})$ , one should observe that the curve accurately follows the estimates as these approach  $(1, 0)$ , in agreement with our clipped data modeling. Nearly identical curves are found when the smooth out-of-focus target is replaced by one, shown in Figure 17(center), which presents various complex structures that may potentially impair the estimation. The wavelet coefficients  $z^{\text{wdet}}$  are shown in Figure 17(right). The estimated curves are shown in Figure 19, where one can also observe the wider dispersion of the estimates

$(\hat{y}_i, \hat{\sigma}_i^{\text{mad}})$  (due to the much smaller number  $n_i$  of usable samples in the level sets  $S_i$ ) and that the  $\hat{y}_i$  are not distributed over the full data-range. In Figure 20 we show a remarkable example of clipping from above and from below within the same frame, as it can be found with the Fujifilm S5600 using ISO 1600. The plot on the left is estimated from the raw-data of an evenly exposed out-of-focus-target, shown in Figure 21(left); observe that the fit of the model to the data is again nearly perfect. The plot on the right is estimated from the raw-data of the dark image shown in Figure 21(center). The curves estimated in the two cases coincide. A further example of estimation from a dark image, showing the accuracy of the proposed model, is given in Figure 22 for the Canon EOS 350D using ISO 1600. Even though in this case the estimates  $(\hat{y}_i, \hat{\sigma}_i^{\text{mad}})$  are concentrated at one side of the diagram, this plot and those shown in Figures 18(right) and 19(right) are nearly identical. Finally, a comparison with the nonparametric estimate  $\hat{\sigma}_{\text{np}}$  obtained by the method [6] is given in Figure 23. The curve  $\hat{\sigma}_{\text{np}}$  was computed analyzing 50 shots of the same target, whereas only one of these 50 images has been used to estimate the function  $\hat{\sigma}_{\text{fit}}$  with the proposed algorithm. The shots were acquired by a 1.3-Mpixel CMOS sensor of a cameraphone, with an analog gain of 10dB. We note that the nonparametric method provides an estimate of  $\sigma(y)$  only for the range of values  $y$  covered by the used images. Moreover, it produces erroneous results approaching the extrema of this range (about 0.07 and 0.41), due to lack of samples. Within these extrema (i.e.,  $0.07 < y < 0.41$ ) the two plots are however very close, with minor differences due to the lack of FPN contribution to the  $\hat{\sigma}_{\text{np}}$  estimate.

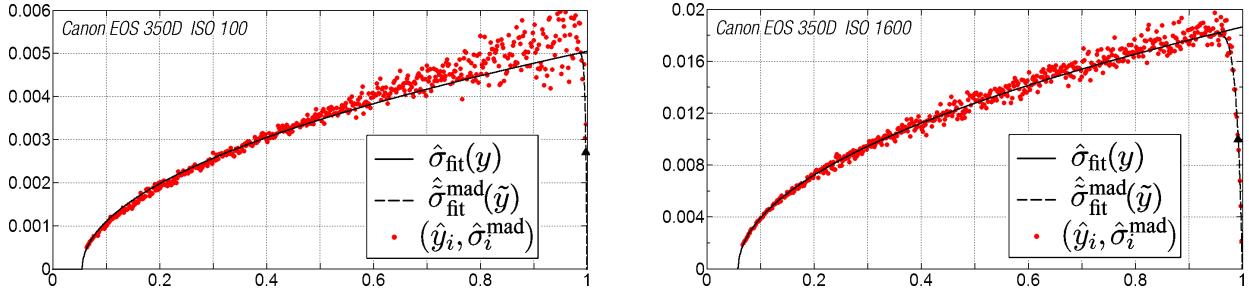


Fig. 18. Estimation of the standard-deviation function  $\sigma$  from the raw-data of the out-of-focus image (Fig. 17, left).

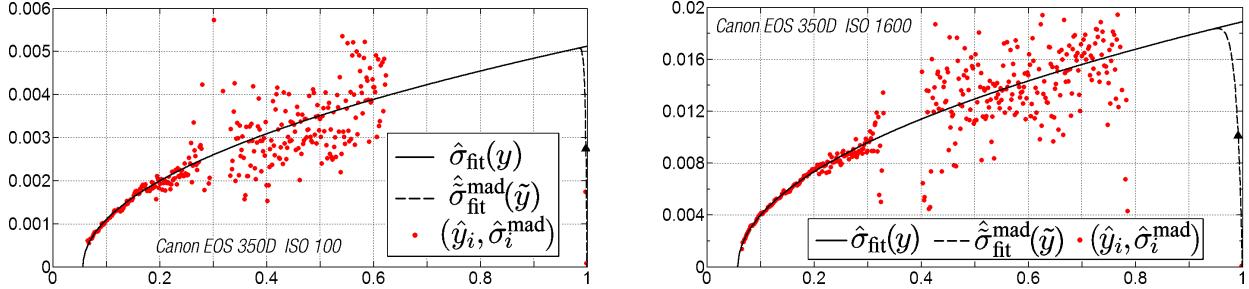


Fig. 19. Estimation of the function  $\sigma$  from the raw-data of the natural image (Fig. 17, center). Compare with the corresponding plots in Fig. 18.

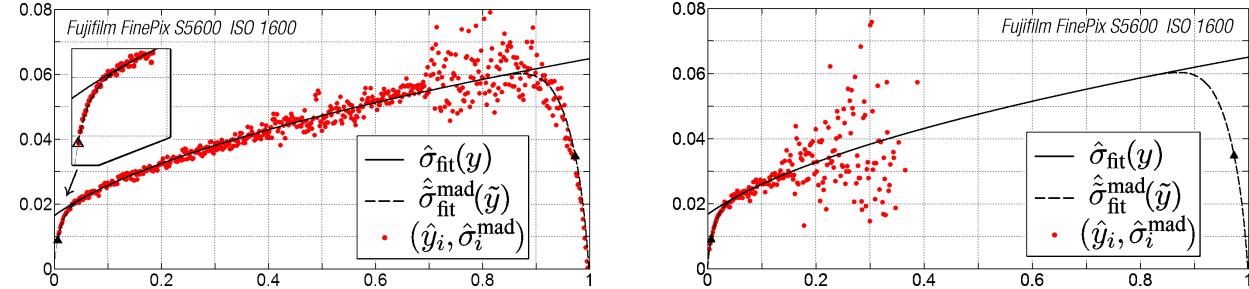


Fig. 20. Estimation from raw-data which exhibits both clipping from above (overexposure) and from below (underexposure), as with the Fujifilm FinePix S5600 camera at ISO 1600. The plot on the left is estimated from raw-data of an evenly exposed out-of-focus image shown in Figure 21(left), the one on the right is estimated from the raw-data of the dark image shown in Figure 21(center). The curves estimated in the two cases coincide.



Fig. 21. From left to right: out-of-focus image with under- and over-exposure (Fujifilm FinePix S5600, ISO 1600), and two largely underexposed dark shots (Fujifilm FinePix S5600, ISO 1600, and Canon EOS 350D, ISO 1600).

## VIII. COMMENTS

### A. Different parametric models for the $\sigma$ function

We remark that the proposed algorithm is not restricted to the particular model (4). In fact, the parameters of any other parametric model can be estimated in the same way. It is sufficient to modify the expression of the function  $\sigma_{\text{reg}}$  in the likelihood (23) according to the assumed parametric model. Therefore, our algorithm has a broader scope of application than shown in this paper and can be applied for parameter

estimation of other signal-dependent noise models, which can be approximated as heteroskedastic normal. Heuristic models for  $\sigma$ , such as those found using the principal component analysis in [11], can also be exploited in our estimation framework.

### B. Multiple images

If two or more independent realizations of the image  $z$  are available, they can be easily exploited in a fashion similar to

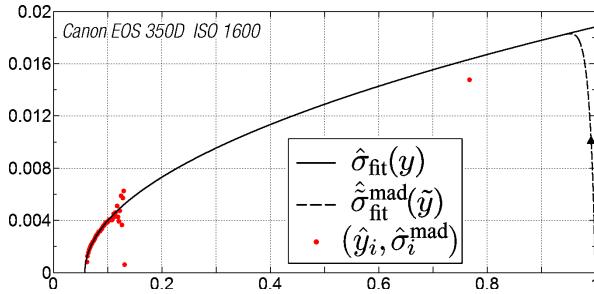


Fig. 22. Estimation from the raw-data of the dark image shown in Figure 21(right) (Canon EOS 350D, ISO 1600); compare with plots in Figures 18(right) and 19(right).

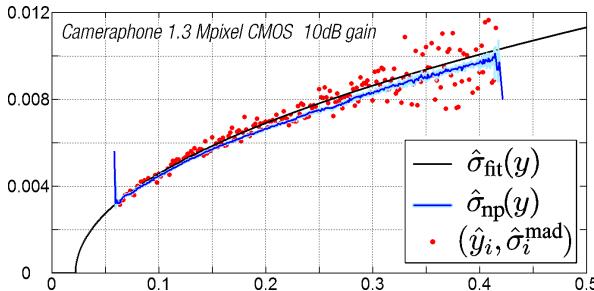


Fig. 23. Comparison between the parametric  $\hat{\sigma}_{\text{fit}}$ , estimated from a single image, and the nonparametric curve  $\hat{\sigma}_{\text{np}}$  [6] computed from 50 images.

[6]. Let us denote the different realizations as  $z_1, \dots, z_J$ . From (1), we have

$z_j(x) = y(x) + \sigma(y(x))\xi_j(x) \quad \forall x \in X, \quad j = 1, \dots, J$ , where  $\xi_1, \dots, \xi_J$  are mutually independent and, for a fixed  $x$ ,  $\xi_1(x), \dots, \xi_J(x)$  are i.i.d. random variables. Thus, by averaging we obtain

$$z^{\text{ave}}(x) = \sum_{j=1}^J \frac{z_j(x)}{J} = y(x) + \frac{\sigma(y(x))}{\sqrt{J}}\xi(x), \quad \forall x \in X, \quad (61)$$

where  $\xi(x)$  has the same distribution as any  $\xi_j(x)$ . Applying the proposed estimation procedure on  $z^{\text{ave}}$  permits to estimate the function  $J^{-\frac{1}{2}}\sigma$  and hence  $\sigma$ . In principle, the advantage of the averaging (61) lies in the lower variance of the observation  $z^{\text{ave}}$ , which allows for better edge-removal and results in estimates  $(\hat{y}_i, \hat{\sigma}_i)$  with lower variance. However, in practice,  $J$  cannot be taken arbitrarily large because a very large  $J$  would render the noise-to-signal ratio of  $z^{\text{ave}}$  too low for the noise  $J^{-\frac{1}{2}}\sigma\xi$  to be measured accurately. Hence, the averaging (61) is valuable only provided that the true  $y$  is sufficiently smooth and that the computational precision is high.

### C. Denoising clipped signals

A generic denoising procedure can be modeled as an operator whose output is an estimate of the expectation of the noisy input. It means that when we denoise  $\tilde{z}$ , as the output we do not get an estimate of  $y$ , but rather an estimate of  $\tilde{y}$ . However, by applying (47) on the output, we can transform it to an estimate of  $y$ . In the same way, we can “take advantage of the noise” to obtain an image with a higher dynamic range, since the range of  $\tilde{z}$  and  $\tilde{y}$  is always smaller than that of  $y$ . The interested reader can refer to the recent work [5], where

a detailed discussion about the denoising of clipped noisy images is given.

### D. Interpolation of the functions $\mathcal{S}_m$ , $\mathcal{S}_e$ , $\mathcal{S}_r$ , $\mathcal{E}_r$

In our current implementation of the algorithm, we use interpolated values for the functions  $\mathcal{S}_m$ ,  $\mathcal{S}_e$ ,  $\mathcal{S}_r$ ,  $\mathcal{E}_r$ ,  $\mathcal{S}_m^{\text{mad}}$ ,  $\mathcal{S}_e^{\text{mad}}$ ,  $\mathcal{S}_r^{\text{mad}}$ , and  $\mathcal{E}_r^{\text{mad}}$  as no closed form expression is available. For practicality, we resort to indirect (nonlinear) polynomial interpolation with exponential or logarithmic functions. The particular expressions of the used interpolant are as follows,

$$\begin{aligned} \mathcal{S}_m(\mu) &\approx \frac{1+\tanh(p(\mu))}{2}, & \mathcal{S}_e(\xi) &\approx 1 - e^{p(\sqrt{\xi})}, \\ \mathcal{E}_r(\rho) &\approx 1 - e^{p(\log(\rho))}, & \mathcal{S}_r(\rho) &\approx 1 - e^{p(\rho)}, \end{aligned} \quad (62)$$

where  $p(t) = \sum_k p_k t^k$  is a polynomial with coefficients  $p_k$  as given in Table IV. For the MAD estimates  $\hat{\sigma}_i^{\text{mad}}$  (57) and related functions  $\mathcal{S}_m^{\text{mad}}$ ,  $\mathcal{S}_e^{\text{mad}}$ ,  $\mathcal{S}_r^{\text{mad}}$ , and  $\mathcal{E}_r^{\text{mad}}$ , we use the same interpolant expressions as in (62) but with different polynomial coefficients, which are given in Table V.

We note that, with (62) and the coefficients in Tables IV and V, the interpolation achieved for  $\mathcal{E}_r^{\text{mad}}$  and  $\mathcal{S}_r$  is diverging at  $\rho \approx 0.5$  and  $\rho \approx 11.5$ , respectively. However, the interpolation is accurate for  $0.65 \leq \rho \leq 11$ . Therefore, in our experiments we constrain  $\rho$  within these bounds. Since  $\rho$ ,  $\mathcal{E}_r^{\text{mad}}$ , or  $\mathcal{S}_r$  are used only for the weighted least-squares problem (26) and not for the likelihood equation (23), the restriction on  $\rho$  does not affect the final estimation of the noise model parameters.

The interpolants and tables presented in this paper complement and extend similar (although not equivalent) numerical data found in the literature [2] (and references therein), [3]. To the best of the authors’ knowledge, no other studies of indirect (e.g., in the wavelet domain) and robust (e.g., median-based) estimators of clipped samples have appeared to date and, although limited, the results in Tables III and V are therefore valuable on their own. Further, we wish to emphasize that the various estimators proposed in the cited publications are developed for censored Gaussian processes with *fixed* mean and variance, and are thus not applicable to the more general estimation problem considered by us.

### E. Matlab software

A MATLAB implementation of the algorithm is available at <http://www.cs.tut.fi/~foi/sensornoise.html>

## IX. ACKNOWLEDGEMENT

The first author would like to thank Giacomo Boracchi for his strenuous and accurate review of the very early drafts and for the many constructive comments and suggestions which eventually led to a substantial improvement of this paper.

## X. CONCLUSION

We presented and analyzed a Poissonian-Gaussian noise model for clipped (and non clipped) raw-data. An algorithm for the estimation of the model parameters from a single noisy image is proposed. The algorithm utilizes a special ML fitting of the parametric model on a collection of local wavelet-domain estimates of mean and standard-deviation.

	$\mathcal{S}_m$	$\mathcal{S}_e$	$\mathcal{E}_r$	$\mathcal{S}_r$
$p_{10}$	—	$-5.7155077 \cdot 10^{-3}$	—	$2.3348876 \cdot 10^{-7}$
$p_9$	$1.4308530 \cdot 10^{-6}$	$9.9630886 \cdot 10^{-2}$	—	$-8.7619300 \cdot 10^{-6}$
$p_8$	$3.2172868 \cdot 10^{-7}$	$-7.3358563 \cdot 10^{-1}$	—	$1.2534052 \cdot 10^{-4}$
$p_7$	$-2.6295693 \cdot 10^{-5}$	$2.9464712 \cdot 10^0$	$-5.5320701 \cdot 10^{-3}$	$-7.0907114 \cdot 10^{-4}$
$p_6$	$-8.5123452 \cdot 10^{-5}$	$-6.9596693 \cdot 10^0$	$-5.5542026 \cdot 10^{-2}$	$-1.4943642 \cdot 10^{-3}$
$p_5$	$-1.7851033 \cdot 10^{-5}$	$9.7330082 \cdot 10^0$	$-2.0363415 \cdot 10^{-1}$	$4.5981260 \cdot 10^{-2}$
$p_4$	$2.0282884 \cdot 10^{-3}$	$-8.3760888 \cdot 10^0$	$-3.4651219 \cdot 10^{-1}$	$-2.8638716 \cdot 10^{-1}$
$p_3$	$2.4377832 \cdot 10^{-2}$	$3.9881199 \cdot 10^0$	$-4.1222715 \cdot 10^{-1}$	$8.5412513 \cdot 10^{-1}$
$p_2$	$3.7234715 \cdot 10^{-2}$	$-1.8598057 \cdot 10^0$	$-9.1504182 \cdot 10^{-1}$	$-1.5725702 \cdot 10^0$
$p_1$	$7.0309281 \cdot 10^{-1}$	$-7.0210049 \cdot 10^{-1}$	$-4.3779025 \cdot 10^0$	$-5.2653050 \cdot 10^{-1}$
$p_0$	$1.6923658 \cdot 10^{-1}$	$-3.1389654 \cdot 10^{-6}$	$-1.5498697 \cdot 10^0$	$-1.2319839 \cdot 10^{-10}$

TABLE IV

COEFFICIENTS OF THE POLYNOMIAL  $p(t) = \sum_k p_k t^k$  USED FOR INDIRECT INTERPOLATION OF THE FUNCTIONS  $\mathcal{S}_m, \mathcal{S}_e, \mathcal{S}_r, \mathcal{E}_r$  AS IN (62).

	$\mathcal{S}_m^{\text{mad}}$	$\mathcal{S}_e^{\text{mad}}$	$\mathcal{E}_r^{\text{mad}}$	$\mathcal{S}_r^{\text{mad}}$
$p_{10}$	—	$-2.0882655 \cdot 10^1$	—	$-6.8695634 \cdot 10^{-1}$
$p_9$	$6.8722511 \cdot 10^{-4}$	$1.5778478 \cdot 10^2$	$-2.4852997 \cdot 10^1$	$1.0659232 \cdot 10^1$
$p_8$	$-3.3132811 \cdot 10^{-3}$	$-4.9232870 \cdot 10^2$	$6.2356790 \cdot 10^1$	$-7.1938586 \cdot 10^1$
$p_7$	$4.6401970 \cdot 10^{-4}$	$8.1719776 \cdot 10^2$	$-3.4588152 \cdot 10^1$	$2.7680210 \cdot 10^2$
$p_6$	$1.4193996 \cdot 10^{-2}$	$-7.7148949 \cdot 10^2$	$-2.3467375 \cdot 10^1$	$-6.6844401 \cdot 10^2$
$p_5$	$-3.3370736 \cdot 10^{-3}$	$4.0631819 \cdot 10^2$	$2.2066307 \cdot 10^1$	$1.0501179 \cdot 10^3$
$p_4$	$-4.0537889 \cdot 10^{-2}$	$-1.0564855 \cdot 10^2$	$1.7531850 \cdot 10^0$	$-1.0745079 \cdot 10^3$
$p_3$	$7.8410754 \cdot 10^{-2}$	$8.7374360 \cdot 10^0$	$-5.0041080 \cdot 10^0$	$6.9444029 \cdot 10^2$
$p_2$	$1.6003810 \cdot 10^{-2}$	$-1.6521034 \cdot 10^0$	$1.3236318 \cdot 10^{-1}$	$-2.6249852 \cdot 10^2$
$p_1$	$8.3418294 \cdot 10^{-1}$	$5.2647043 \cdot 10^{-2}$	$-4.7818191 \cdot 10^0$	$4.5940232 \cdot 10^1$
$p_0$	$7.0493620 \cdot 10^{-2}$	$-1.3640079 \cdot 10^{-6}$	$-1.4931770 \cdot 10^0$	$-1.3097339 \cdot 10^0$

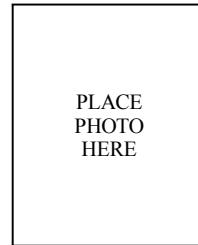
TABLE V

COEFFICIENTS OF THE POLYNOMIAL  $p(t) = \sum_k p_k t^k$  USED FOR INDIRECT INTERPOLATION OF THE FUNCTIONS  $\mathcal{S}_m^{\text{mad}}, \mathcal{S}_e^{\text{mad}}, \mathcal{S}_r^{\text{mad}}, \mathcal{E}_r^{\text{mad}}$  AS IN (62).

Experiments with synthetic images and real raw-data from camera sensors demonstrate the effectiveness and accuracy of the algorithm in estimating the model parameters and confirm the validity of the proposed model.

## REFERENCES

- [1] Blanksby, A.J., M.J. Loinaz, D.A. Inglis, and B.D. Ackland, “Noise performance of a color CMOS photogate image sensor”, *IEEE Int. Electron Devices Meeting 97 Tech. Dig.*, pp. 205-208, 1997.
- [2] Cohen, A.C., *Truncated and Censored Samples*, CRC Press, 1991.
- [3] Khasawneh, M.T., S.R. Bowling, S. Kaewkuekool, B.R. Cho, “Tables of a Truncated Standard Normal Distribution: A Singly Truncated Case”, *Quality Engineering*, vol. 17, no. 1, pp. 33-50, 2005.
- [4] Donoho, D.L., and I.M. Johnstone, “Ideal spatial adaptation via wavelet shrinkage”, *Biometrika*, n. 81, pp. 425-455, 1994.
- [5] Foi, A., “Practical denoising of clipped or overexposed noisy images”, *Proc. 16th European Signal Process. Conf. EUSIPCO 2008*, Lausanne, Switzerland, August 2008.
- [6] Foi, A., S. Alenius, V. Katkovnik, and K. Egiazarian, “Noise measurement for raw-data of digital imaging sensors by automatic segmentation of non-uniform targets”, *IEEE Sensors Journal*, vol. 7, no. 10, pp. 1456-1461, Oct. 2007.
- [7] Johnson, N.L., S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, 2<sup>nd</sup> ed., Wiley Series in Probability and Statistics, 1994.
- [8] Greene, W.H., *Econometric Analysis*, 4<sup>th</sup> ed., Prentice Hall, 2000.
- [9] Hampel, F.R., “The influence curve and its role in robust estimation”, *Journal of American Statistical Association*, 62, pp. 1179-1186, 1974.
- [10] Huang, J., and D. Mumford, “Statistics of natural images and models”, *Proc. IEEE Conf. Computer Vision and Pattern Recognition, CVPR 1999*, vol. 1., pp. 541-547, June 1999.
- [11] Liu, C., W.T. Freeman, R. Szeliski, and S.B. Kang, “Noise estimation from a single image”, *Proc. IEEE Conf. Computer Vision and Pattern Recognition, CVPR 2006*, pp. 901-908, 2006.
- [12] Mosteller, F., and J.W. Tukey, *Data Analysis and Regression: A Second Course in Statistics*, Addison Wesley, 1997.
- [13] Nelder, J.A., and R. Mead, “A simplex method for function minimization”, *The Computer Journal*, vol. 7, pp. 308-313, 1965.
- [14] Rousseeuw, P.J., and C. Croux, “Alternatives to the Median Absolute Deviation”, *J. American Statistical Assoc.*, vol. 88, no. 424, pp. 1273-1283, December, 1993.
- [15] Rhodes, H., G. Agranov, C. Hong, U. Boettiger, R. Mauritzson, J. Ladd, I. Karasev, J. McKee, E. Jenkins, W. Quinlin, I. Patrick, J. Li, X. Fan, R. Panicacci, S. Smith, C. Mouli, J. Bruce, “CMOS imager technology shrinks and image performance”, *Proc. 2004 IEEE Workshop on Microelectronics and Electron Devices, WMED*, pp. 7-18, April 2004.
- [16] Tamayama, H., K. Ito, and T. Nishimura, “Technology trends of high-definition digital still camera systems”, *Sym. VLSI Circuits, Digest of Technical Papers*, pp. 100-105, June 2002.
- [17] Theuwissen, A., *Course on camera system*, Lecture notes, CEI-Europe, 2005.
- [18] Wong, H.-S., D.J. Frank, P. Solomon, C. Wann, and J. Wesler, “Nanoscale CMOS”, *Proc. of the IEEE*, vol. 87, no. 4, pp. 537-570, April 1999.



Alessandro Foi received the M.Sc. degree in Mathematics from the Università degli Studi di Milano, Italy, in 2001, the Ph.D. degree in Mathematics from the Politecnico di Milano in 2005, and the D.Sc.Tech. degree in Signal Processing from Tampere University of Technology, Finland, in 2007. His research interests include mathematical and statistical methods for signal processing, functional analysis, and harmonic analysis. Currently, he is a senior researcher at the Department of Signal Processing, Tampere University of Technology. His work focuses on spatially adaptive algorithms for digital image filtering and on noise modeling for digital imaging sensors.

PLACE  
PHOTO  
HERE

**M**ejdi Trimeche received the B.Sc. degree in electrical engineering from Bilkent University, Turkey, in 1998. In 2000, he received the M.Sc. degree with distinction from Tampere University of Technology, Finland. Later in 2006, he received his Ph.D. from the same university. Since 2000, he is with Nokia Research Center, Tampere, Finland. His research interests include image/video processing and computer vision. In particular, he has been working on super-resolution, image restoration, video coding, visual content indexing and retrieval.

PLACE  
PHOTO  
HERE

**V**ladimir Katkovnik received the M.Sc., Ph.D., and D.Sc. degrees in technical cybernetics from the Leningrad Polytechnic Institute, Leningrad, Russia, in 1960, 1964, and 1974, respectively. From 1964 to 1991, he held the positions of Associate Professor and Professor at the Department of Mechanics and Control Processes, Leningrad Polytechnic Institute. From 1991 to 1999, he was a Professor of statistics with the Department of the University of South Africa, Pretoria. From 2001 to 2003, he was a Professor of mechatronics with the Kwangju Institute of Science and Technology, Korea. From 2000 to 2001 and since 2003 he is a Research Professor with the Institute of Signal Processing, Tampere University of Technology, Tampere, Finland. He has published seven books and more than 200 papers. His research interests include stochastic signal processing, linear and nonlinear filtering, nonparametric estimation, imaging, nonstationary systems, and time-frequency analysis.

PLACE  
PHOTO  
HERE

**K**aren Egiazarian (SM'96) was born in Yerevan, Armenia, in 1959. He received the M.Sc. degree in mathematics from Yerevan State University in 1981, the Ph.D. degree in physics and mathematics from Moscow State University, Moscow, Russia, in 1986, and the D.Tech. degree from the Tampere University of Technology (TUT), Tampere, Finland, in 1994. He has been Senior Researcher with the Department of Digital Signal Processing, Institute of Information Problems and Automation, National Academy of Sciences of Armenia. Since 1996, he has been an Assistant Professor with the Institute of Signal Processing, TUT, where he is currently a Professor, leading the Transforms and Spectral Methods Group. His research interests are in the areas of applied mathematics, signal processing, and digital logic.