A Multiscale Denoising Framework Using Detection Theory with Application to Images from CMOS/CCD Sensors

Khuram Naveed [1,*] iD **, Shoaib Ehsan** [2]**, Klaus D. McDonald-Maier** [2] **and Naveed ur Rehman** [1]

[1] Department of Electrical and Computer Engineering, COMSATS University, Park Road, Islamabad 45550, Pakistan; naveed.rehman@comsats.edu.pk
[2] School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, UK; sehsan@essex.ac.uk (S.E.); kdm@essex.ac.uk (K.D.M.-M.)
* Correspondence: khurram.naveed@comsats.edu.pk; Tel.: +92-51-9049-223

Abstract: Output from imaging sensors based on CMOS and CCD devices is prone to noise due to inherent electronic fluctuations and low photon count. The resulting noise in the acquired image could be effectively modelled as signal-dependent Poisson noise or as a mixture of Poisson and Gaussian noise. To that end, we propose a generalized framework based on detection theory and hypothesis testing coupled with the variance stability transformation (VST) for Poisson or Poisson–Gaussian denoising. VST transforms signal-dependent Poisson noise to a signal independent Gaussian noise with stable variance. Subsequently, multiscale transforms are employed on the noisy image to segregate signal and noise into separate coefficients. That facilitates the application of local binary hypothesis testing on multiple scales using empirical distribution function (EDF) for the purpose of detection and removal of noise. We demonstrate the effectiveness of the proposed framework with different multiscale transforms and on a wide variety of input datasets.

1. Introduction

Digital images acquired using complementary metal oxide semiconductors (CMOS) or charged coupled devices (CCD) image sensors are subject to noise from two notable sources, i.e., electronic instruments and the photo-sensing devices [1,2]. This noise is typically modelled using a mixture of Poisson and Gaussian distributions, namely, Poisson–Gaussian distribution [3,4]. In cases where the Poisson component of the noise is dominant, the Gaussian component is ignored and noise is modelled using the Poisson distribution [5,6].

For the purpose of denoising, estimation of noise parameters of the signal-dependent noise from CMOS/CCD sensors is a problem of interest [7,8]. A mixed Poisson–Gaussian distribution was used to model the practical sensor noise which was subsequently used for denoising [9]. In [10], Poisson statistics in combination with maximum likelihood estimation are used to restore images from optic acquisition systems. A Bayesian framework is developed for denoising and deconvolution of Poisson–Gaussian noise [11]. In addition, a post processing technique for Poisson denoising using best linear prediction on the local image patches is introduced in [12].

Similar to the Stein's unbiased risk estimator (SURE) which is an estimate of the mean squared error (MSE) for Gaussian noise [13], a Poisson unbiased risk estimator (PURE) is estimated and used with the linear expansion techniques (LET) to formulate a state-of-the-art Poisson denoising method known as *PureLet* [14]. This technique was also extended for Poisson–Gaussian denoising, whereby

Poisson–Gaussian risk estimate (PGURE) was minimized [15]. A Purelet deconvolution technique has also emerged for restoring images in the presence Poisson–Gaussian noise [16]. Furthermore, *PURE* estimate has also been employed for extension of the non-local mean (NLM) filtering to Poisson denoising, which we call *Poiss-NLM* [17]. A non-local application of principle component analysis (NL-PCA) for Poisson denoising has been achieved by patch-ization of the noisy image [18].

An alternate approach to the problem involves 'Gaussianization' of Poisson noise through variance stability transformation (VST), followed by traditional noise filtering methods [19]. In this class of methods, first, VST is applied on a low (photon) count image, resulting in transformation of Poisson noise to approximately Gaussian noise which is not only signal independent but also has constant variance. In this way, a Poisson denoising problem can be modeled as a Gaussian denoising problem post VST. As a result, Gaussian denoising methods can be used for suppressing noise, followed by the inverse VST (IVST) to obtain the Poisson denoised image.

Poisson denoising using Anscombe variance stability transformation (AT) [20] has been studied in [21–23], which involves state-of-the-art Gaussian denoising methods. However, inverse Anscombe transformation does not perform the exact reconstruction due to a bias at low photon count, which has been empirically improved in [22]. Similarly, VST has also been used for removing/reducing Poisson–Gaussian noise in [24] where generalized Anscombe transformation (GAT) [25] is used for approximate Gaussian transformation of Poisson–Gaussian noise. This is followed by the use of Gaussian denoising methods such as BLSGSM (introduced in [26]). Finally, inverse GAT transformation is performed to yield the denoised image.

A multiscale extension of VST (MS-VST) is employed for removing mixture of the Poisson–Gaussian noise from medical images in [27]. A similar approach using MS-VST is also devised for Poisson denoising in [28]. Other multiscale approaches for Poisson denoising employ hypothesis testing using multiscale Haar transform [29,30].

In this work, we propose a framework for denoising data corrupted with Poisson or mixed Poisson–Gaussian noise, which is found in images obtained from CMOS or CCD image sensors. The proposed framework uses hypothesis testing framework, typically used in signal detection theory, at multiple data scales and employs test statistics based on empirical distribution function (EDF), e.g., AD and CVM statistics. While our previous work used goodness-of-fit (GoF) test based hypothesis testing for Gaussian denoising of 1D (time series) [31–33] and 2D data (images) [34,35], the novel contributions in this work are as follows: (i) the proposed framework caters for both Poisson and mixed Poisson–Gaussian noise from CMOS/CCD sensors; (ii) the proposed framework is generalized in a sense that it has been implemented by using linear and non-linear multiscale transform domain methods; (iii) complete theoretical and mathematical framework of the proposed methodology is presented in the context of detection theory; (iv) extensive results on images corrupted with non-Gaussian noise have been included with special emphasis on images obtained from CMOS/CCD sensors.

The paper is arranged as follows: Section 2 gives an insight into CMOS imaging and noise models while statement of the problem is given in Section 3. A review of the signal detection theory is presented in Section 4. Section 5 presents the proposed non-Gaussian denoising framework while the discussion on experimental results is presented in Section 6. Section 7 concludes the paper with a discussion on avenues for future work.

2. CMOS Imaging and Sources of Noise

Recent progress in the design of CMOS imaging technology has led to extensive use of image acquisition systems in real world applications, i.e., installation of cameras in mobile phones, tablets, personal computers etc. A CMOS imaging system is essentially composed of two dimensional array of CMOS sensors each of which produces a pixel value [36,37]. The architecture of the CMOS image acquisition system is given in Figure 1 whereby a pipeline of modules required to capture an image is shown. To begin with, a micro-lens array is installed to increase the concentration of light incident

on the photo-detector. This is followed by a color filter array which generates only one of the red, green or blue signal at each pixel. Next, an array CMOS imaging sensors capture the analog signals and converts them into electrical signals which are subsequently digitized using the analog to digital converters to generate pixel values. Finally, post processing operation is performed to adjust white balancing and perform color correction [38].
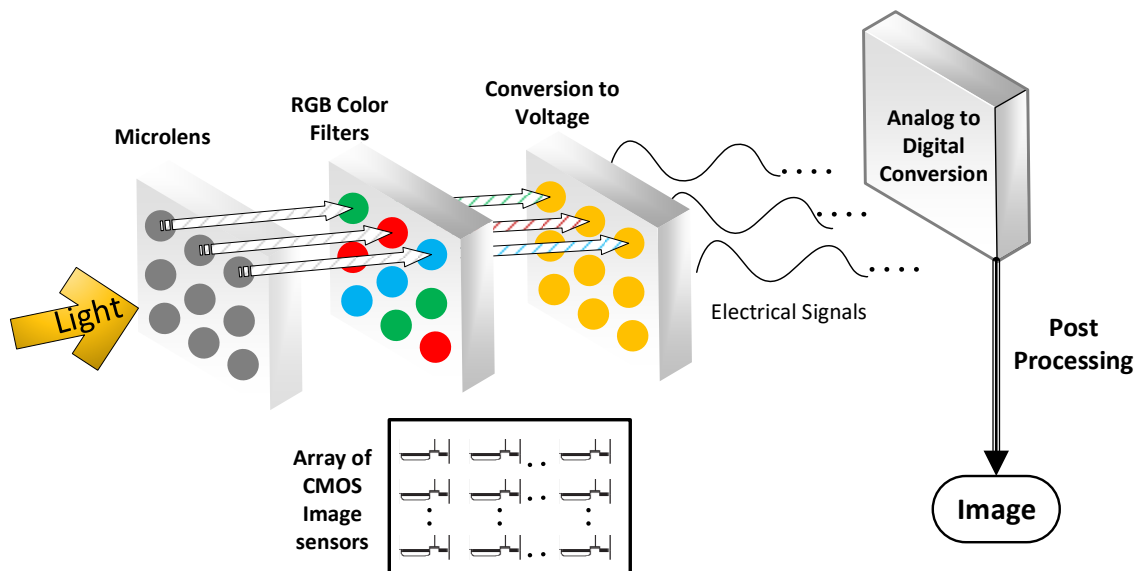

Figure 1. Pipeline of CMOS imaging acquisition system (adapted from [38]).

While CMOS image sensing technology is revolutionizing the digital imaging by shrinking the pixel pitch [39], one of the major challenges includes reducing image noise at the time of acquisition [40]. In the case of CMOS imaging sensors, main sources of noise include sensor electronics and photon starvation. For precise modeling of noise in the CMOS sensors, it is important to look into the pixel sensing architecture shown in Figure 2. The pixel circuit is composed of a photodiode and a switching transistor. The imaging principle works as follows: During exposure, photons fall on the reverse biased photodiode leading to the decrease of the reverse voltage across the diode. Subsequently, voltage across photodiode is measured or read at the end of the exposure and photodiode is reset for another exposure [41]. This is known as passive pixel model while the active CMOS pixel model includes the amplification of the read signal which not only increases the sensitivity of the CMOS sensor but also helps reduce the noise [41]. Figure 3 gives an account of various types of noise corrupting the CMOS sensed image.
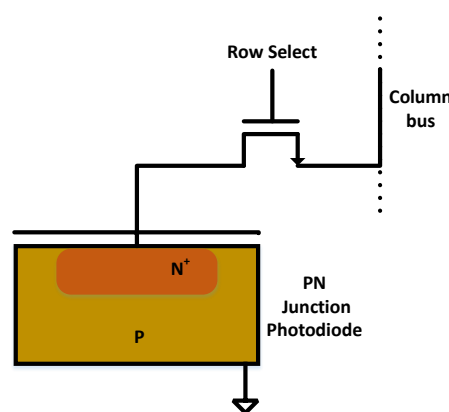

Figure 2. CMOS photo-sensing pixel architecture (adapted from [41]).

The quantum nature of light dictates that the amount of photons incident on the photodiode is never a certainty. This fluctuation in photon count results in shot noise in the acquired image [41]. Similarly, dark current non-uniformity is an exposure dependent fixed pattern noise which also has a temporal noise associated with itself, known as dark current shot noise. Noise due to photon fluctuations are signal-dependent and are modelled using Poisson distribution [42], where the mean and the variance of the Poisson process equals the signal strength.

Moreover, charge to voltage conversion and subsequent amplification of the electrical signal are also noisy owing to electronic fluctuations. These processes notably introduce flicker or $1/f$ noise and thermal noise, which are modelled using the additive white Gaussian noise (AWGN) [42]. In addition, residual error due to quantization is also modelled using the independent additive white Gaussian noise [15].

Based on the above discussion, noise due to CMOS image sensors is typically modelled using the mixture of Poisson and Gaussian distributions also termed as Poisson–Gaussian distribution [15,16,42]. However, in situations of poor illumination or low light conditions noise may be dominantly Poisson distributed [14,20] since the effect of AWGN can be neglected due to low photon count in such cases.
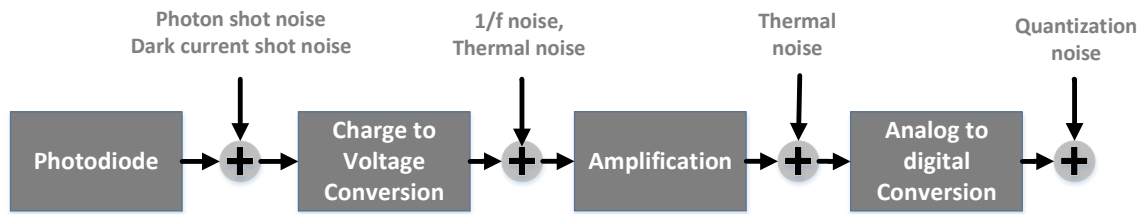
Figure 3. Noise model of CMOS image sensor (adapted from [42]).

3. Statement of Problem

Let $z_{\mathbf{i}}$ denotes the pixels of the acquired noisy image **Z** using a CMOS sensor, which may be mathematically modelled as

$$z_{\mathbf{i}} = s_{\mathbf{i}} + \eta_{\mathbf{i}}^{p}(s_{\mathbf{i}}) + \eta_{\mathbf{i}}^{g}, \tag{1}$$

where $s_{\mathbf{i}}$ denotes expected pixels of true image **S**, $\eta_{\mathbf{i}}^{g} \sim \mathcal{N}(0,\sigma^2)$ denote the additive white Gaussian noise (AWGN) with zero mean and arbitrary variance σ^2 and $\eta_{\mathbf{i}}^{p}(s_{\mathbf{i}}) \sim P(z_{\mathbf{i}}|s_{\mathbf{i}})$ denotes the signal-dependent Poisson noise. Here, Poisson distribution $P(z_{\mathbf{i}}|s_{\mathbf{i}})$ is given as follows

$$P(z_{\mathbf{i}}|s_{\mathbf{i}}) = \frac{s_{\mathbf{i}}^{z_{\mathbf{i}}} e^{-s_{\mathbf{i}}}}{z_{\mathbf{i}}!}, \tag{2}$$

where ! denotes the factorial operation. Note that the vector **i** denotes pixel location, i.e., row and column indexes.

Under various physical limitations, i.e., low light or short exposure time, the effect of Gaussian noise may be neglected due to relative strength of the signal-dependent noise. In this case, the acquired pixel model (1) reduces to

$$z_{\mathbf{i}} = s_{\mathbf{i}} + \eta_{\mathbf{i}}^{p}(s_{\mathbf{i}}). \tag{3}$$

In this case, $z_{\mathbf{i}}$ will also be distributed by $P(z_{\mathbf{i}}|s_{\mathbf{i}})$ with a non zero mean $E[z_{\mathbf{i}}|s_{\mathbf{i}}] = s_{\mathbf{i}}$. Hence, it can be concluded that the mean and variance of $\eta_{\mathbf{i}}^{p}(s_{\mathbf{i}})$ are given as follows

$$\begin{aligned} E[\eta_{\mathbf{i}}^{p}(s_{\mathbf{i}})] &= 0, \\ var[\eta_{\mathbf{i}}^{p}(s_{\mathbf{i}})] &= s_{\mathbf{i}}. \end{aligned} \tag{4}$$

This means that the variance of each noise coefficient $\eta_{\mathbf{i}}^{p}(s_{\mathbf{i}})$ is dependent on the corresponding true signal value $s_{\mathbf{i}}$ and is parametrized by the peak value of the signal.

4. Preliminaries: Introduction to Hypothesis Testing in Detection Theory

Classical detection theory [43] based on hypothesis testing assumes prior distribution models for signal and noise where the aim is to detect signal (with or without an additive noise part) while avoiding noise, e.g., communication channel, radar signal processing etc. A detection problem using hypothesis testing comprises of (i) null hypothesis $\mathcal{H}_0$ of *noise only* case and (ii) alternate hypothesis $\mathcal{H}_1$ concerning with *signal plus noise* detection. A classical example may be the detection of signal in a transmission medium where the noise is assumed to be distributed by zero mean Gaussian distribution $\mathcal{N}(0,\sigma^2)$ and *signal plus noise* is modeled by a non-zero mean Gaussian distribution $\mathcal{N}(\mu,\sigma^2)$ as signal values when added to zero mean noise, contribute a mean μ to the distribution. Mathematically, this detection problem may be modeled as

$$\begin{aligned} \mathcal{H}_0 &: x \in \mathcal{N}(0,\sigma), \\ \mathcal{H}_1 &: x \in \mathcal{N}(\mu,\sigma), \end{aligned} \tag{5}$$

where $\mathcal{H}_0$ and $\mathcal{H}_1$ denote the null and the alternate hypothesis, respectively, while x denotes an arbitrary value from the noisy signal **x**. Figure 4a plots probability distribution functions of each hypothesis, i.e., $p(x|\mathcal{H}_0) = \mathcal{N}(0,\sigma^2)$ and $p(x|\mathcal{H}_1) = \mathcal{N}(\mu,\sigma^2)$. The problem here is to differentiate between the following hypothesis

$$\begin{aligned} \mathcal{H}_0 &: x = \eta, \\ \mathcal{H}_1 &: x = s + \eta, \end{aligned} \tag{6}$$

where s denotes an arbitrary value from the true signal **s** and η denotes an arbitrary value from $\boldsymbol{\eta}$ distributed by $\mathcal{N}(0,\sigma^2)$.

Here, the decision in Equation (6) can be taken by comparing the observations x against a threshold λ. This is elaborated graphically in Figure 4a, where the threshold is plotted as a dotted line. Note that the values greater than the threshold λ_k are more likely to be distributed according to $p(x|\mathcal{H}_1)$ while the values less then the threshold λ_k are more likely to be distributed in accordance with $p(x|\mathcal{H}_0)$. Therefore, the above hypothesis testing problem can be given as

$$\begin{aligned} \mathcal{H}_0 &: x \leq \lambda, \\ \mathcal{H}_1 &: x > \lambda. \end{aligned} \tag{7}$$

A trivial choice of threshold may be the point of intersection of $p(x|\mathcal{H}_0)$ and $p(x|\mathcal{H}_1)$ as shown in Figure 4a while a different choice of threshold is depicted in Figure 4b.

However, this kind of detector makes two types of errors: (type I) detecting $\mathcal{H}_0$ when $\mathcal{H}_1$ is given; and (type II) detecting $\mathcal{H}_1$ when $\mathcal{H}_0$ is given $p(\mathcal{H}_1|\mathcal{H}_0)$. Figure 4 shows probability regions $p(\mathcal{H}_0|\mathcal{H}_1)$ and $p(\mathcal{H}_1|\mathcal{H}_0)$ corresponding to type I and II errors respectively, in the context of a simple detection problem. The type II error is also known as *false alarm* and the probability $p(\mathcal{H}_1|\mathcal{H}_0)$ is termed as the *probability of false alarm* (P_{fa}). Minimizing both errors simultaneously is not possible as decreasing one increases the other, however, these errors can be traded off for each other by adjusting the value of threshold λ, (as depicted by different choices of threshold in Figure 4). Typically, it is required to keep P_{fa} very low in order to avoid the severe consequences of noise detected as signal. Hence, P_{fa} is fixed to a very small value α to estimate a suitable threshold using the following relation

$$P_{fa} = p(\mathcal{H}_1|\mathcal{H}_0) = Prob(x > \lambda \,|\mathcal{H}_0) = \int_{\{x \,|\, x>\lambda\}} p(x|\mathcal{H}_0)\mathrm{d}x = \alpha, \tag{8}$$

where the range $\{x \mid x > \lambda\}$ denote the values x detected as signal (i.e., $\mathcal{H}_1$) and $Prob(\cdot)$ denotes the probability of the given event and α is of the order of 10^{-3}–10^{-6}.
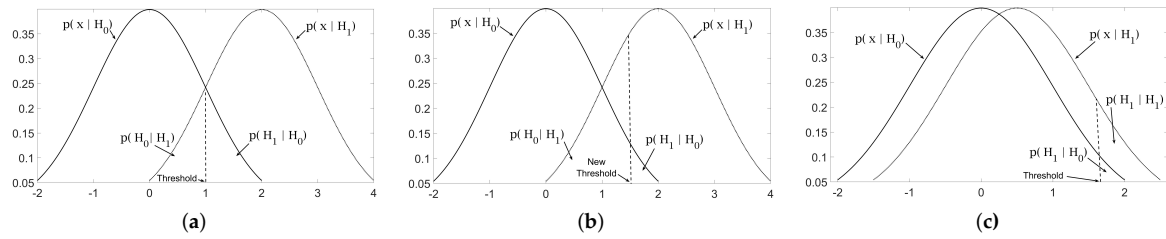
Figure 4. Depiction of a simple detection problem where the probability of null hypothesis $p(x|\mathcal{H}_0) = \mathcal{N}(0, \sigma^2)$ and probability of alternate hypothesis $p(x|\mathcal{H}_1) = \mathcal{N}(\mu, \sigma^2)$ are plotted together. Here, (**a**) show the relationship between error probability regions $p(\mathcal{H}_0|\mathcal{H}_1)$, $p(\mathcal{H}_1|\mathcal{H}_0)$ and the detection probability region $p(\mathcal{H}_1|\mathcal{H}_1)$; (**b**,**c**) highlight the trade off between P_{fa} and P_d with an increase in threshold value λ.

On the other hand, the probability of the first type of error $p(\mathcal{H}_0|\mathcal{H}_1)$ is minimized which in turn maximizes the probability of signal detection when signal is present, i.e., $p(\mathcal{H}_1|\mathcal{H}_1) = 1 - p(\mathcal{H}_0|\mathcal{H}_1)$, see Figure 4c. The probability of true signal detections $p(\mathcal{H}_1|\mathcal{H}_1)$, also termed as probability of detection (P_d), is mathematically given as

$$P_d = p(\mathcal{H}_1|\mathcal{H}_1) = Prob(x > \lambda \mid \mathcal{H}_1) = \int_{\{x \mid x > \lambda\}} p(x|\mathcal{H}_1)\mathrm{d}x. \tag{9}$$

In detection theory, P_d is required to be maximized with the minimization of P_{fa}.

The binary hypothesis defined in Equation (7) directly compares the data x against a threshold. More generally, some metric $\mathcal{S}(x)$ on x is used for this purpose, e.g., Neyman Pearson optimal detector [44] uses a statistic $\mathcal{S}(x) = \frac{p(x|\mathcal{H}_1)}{p(x|\mathcal{H}_0)}$ to compare against the threshold λ for hypothesis testing as follows

$$\begin{aligned}\mathcal{H}_0 &: \mathcal{S}(x) \leq \lambda, \quad i.e., \quad x \in \eta, \\ \mathcal{H}_1 &: \mathcal{S}(x) > \lambda, \quad i.e., \quad x \in \mathbf{s} + \eta,\end{aligned} \tag{10}$$

where the distributions of null hypothesis $p(x|\mathcal{H}_0)$ and alternate hypothesis $p(x|\mathcal{H}_1)$ must be known a priori. Consequently, definition of P_{fa} in the presence of test statistic $\mathcal{S}(x)$ changes to the following

$$P_{fa} = Prob(\mathcal{S}(x) > \lambda \mid \mathcal{H}_0) = \int_{\{x \mid \mathcal{S}(x) > \lambda\}} p(x|\mathcal{H}_0)\mathrm{d}x, \tag{11}$$

where $\{x \mid \mathcal{S}(x) > \lambda\}$ is composed of the values x for which $\mathcal{S}(x) > \lambda$ and as a consequence, threshold λ may be estimated by fixing the $P_{fa} = \alpha$. Similarly, P_d changes to the following

$$P_d = Prob(\mathcal{S}(x) > \lambda \mid \mathcal{H}_1) = \int_{\{x \mid \mathcal{S}(x) > \lambda\}} p(x|\mathcal{H}_1)\mathrm{d}x. \tag{12}$$

A popular approach in detection theory is based on goodness-of-fit (GoF) test in which the test statistic $\mathcal{S}(x)$ is based on the information regarding the empirical distribution model of the data at hand. This approach also avoids the need to assume a prior distribution model for alternate hypothesis, since prior knowledge of the null distribution is adequate for binary hypothesis testing in Equation (7). Under such conditions, the test statistic $\mathcal{S}(x)$ estimates the distance between the empirical distribution function (EDF) $\mathcal{F}(t) = \sum_t \mathbf{1}.(x > t)$ of noisy observations $\mathbf{x}$, and the null cumulative distribution function (CDF) $\mathcal{F}_0(t) = \int_t p(x|\mathcal{H}_0)\mathrm{d}x$, where t is the support vector. There are a number of test statistics/measures used as detectors within the framework of the GoF test, but Anderson

Darling (AD) statistic [45], Cramer Von Mises (CVM) statistic [46] are frequently used in detection problems [47,48], and are given, respectively, as follows

$$\mathcal{S}_{AD}(x) = \int_{-\infty}^{\infty} (\mathcal{F}_0(t) - \mathcal{F}(t))^2 \psi(\mathcal{F}_0(t)) \mathrm{d}(\mathcal{F}_0(t)). \tag{13}$$

$$\mathcal{S}_{CVM}(x) = \int_{-\infty}^{\infty} (\mathcal{F}_0(t) - \mathcal{F}(t))^2 \mathrm{d}(\mathcal{F}_0(t)). \tag{14}$$

5. Proposed Denoising Framework Using Detection Theory

In this section, we propose a denoising framework to remove Poisson and Poisson–Gaussian distributed arising due to the CMOS/CCD image sensors. For this purpose, we first employ variance stability transformation (VST) to 'Gaussianize' the noise present in the CMOS/CCD images. Following that, image denoising is formulated as a detection problem whereby local hypothesis testing based on empirical distribution function (EDF) is employed.

Since, detection theory is employed on time series data [47,48], formulation of detection problem based on local EDF statistics for spatio-temporal data (images) in our case requires following notable adjustments.

1. To ensure the preservation of spatio-temporal characteristics of multiscale coefficients of noisy image, two dimensional (2D) windows of size $l \times l$ are considered around the coefficient for local hypothesis testing.
2. Two dimensional EDFs are not unique and are computationally expensive [49], therefore, their use for GoF testing on 2D data is not suitable. Consequently, in our work, we list the coefficients in the windows as 1D vectors followed by the computation of their unique (1D) EDF. Note that listing of 2D segments as 1D vectors is a common practice in image denoising methods whereby multivariate statistical distributions are used to model multiscale dependencies [50].

The block diagram of the proposed method is shown in Figure 5. The method involves VS transformation followed by multiscale hypothesis testing of data at local level using EDF statistics. The following subsections illustrate the main steps of the proposed framework.
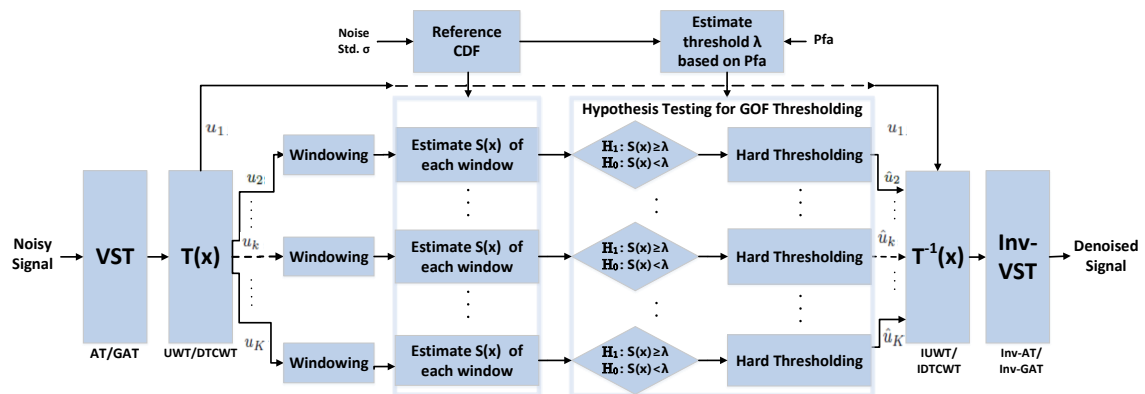
Figure 5. Block diagram of the proposed framework for Poisson and Poisson–Gaussian denoising using detection theory.

5.1. Variance Stability Transform (VST)

We propose to use VST as a preprocessing step in the case of denoising images corrupted by Poisson and Poisson–Gaussian noise. Following the preprocessing step, the noise is effectively transformed into an independent Gaussian noise with constant variance which can be handled through

a Gaussian denoising framework. For Poisson image pixels $z_\mathbf{i} \in \mathbf{Z}$, the AT [20] could be used for variance stabilization of signal-dependent noise as follows,

$$x_\mathbf{i} = 2\sqrt{z_\mathbf{i} + 3/8}. \tag{15}$$

For variance stabilization of mixed Poisson–Gaussian noisy image $\mathbf{Z}$, generalized Anscombe transformation (GAT) [25] is used

$$x_\mathbf{i} = \begin{cases} \frac{2}{\alpha}\sqrt{\alpha z_\mathbf{i} + \frac{3}{8}\alpha^2 + \sigma^2}, & z_\mathbf{i} > -\frac{3}{8}\alpha - \frac{\sigma^2}{\alpha}, \\ 0 & z_\mathbf{i} \leq -\frac{3}{8}\alpha - \frac{\sigma^2}{\alpha}, \end{cases} \tag{16}$$

where the problem (1) is now transformed to Gaussian denoising problem since $x_\mathbf{i} \in \mathbf{X}$ in Equations (15) and (16) are pixels of the variance stabilized image corrupted with approximate Gaussian noise. Note that, GAT is a generalization of AT, as for scaling factor $\alpha = 1$ and $\sigma = 0$ (i.e., absence of Gaussian noise), Equation (16) reduces to Equation (15).

5.2. Multiscale Local Hypothesis Testing Based on EDF

Typically, in detection problems, we are mainly interested in the detection of signal at a particular time (with or without additive noise part). In signal denoising, on the other hand, we are interested in separating signal from noise so that the effect of noise could be cancelled from the output data. That requires a modification in the classical binary hypothesis testing framework to be applied for denoising applications. In the denoising problem, the alternate hypothesis must correspond to the *signal only* case whereas the null hypothesis corresponds to the *noise only* case as before. To achieve that, we propose to apply the modified hypothesis testing within the framework of goodness-of-fit (GoF) test at multiple scales obtained via a multiscale transform.

Let $\mathcal{T}(\cdot)$ denote a multiscale transform which decomposes a noisy signal $\mathbf{x}$ into multiscale coefficients $u_k^{(i)}$ at scale k and location i, as given below

$$u_k^{(\mathbf{i})} = \mathcal{T}(\mathbf{x}). \tag{17}$$

For multiscale coefficients $u_k^{(\mathbf{i})}$ to either correspond to the true signal (only) or the noise (only), the $\mathcal{T}(\cdot)$ must fulfill the following conditions:

1. $\mathcal{T}(\cdot)$ must decompose a signal across multiple scales.
2. Across each scale, signal and noise must be distributed among separate coefficients/values.

Set of transform domain methods fulfilling the above conditions may include DWT and its redundant variants like DDDWT [51], DT-CWT [52] and UWT [53] etc.

Given that $\mathcal{T}(\cdot)$ fulfills aforementioned conditions, we propose to formulate the *denoising* problem as a transformed hypothesis testing problem as follows

$$\begin{aligned} \tilde{\mathcal{H}}_0 &: u_k^{(\mathbf{i})} \in \mathcal{T}(\boldsymbol{\eta}) \quad \text{(noise)} \\ \tilde{\mathcal{H}}_1 &: u_k^{(\mathbf{i})} \in \mathcal{T}(\mathbf{s}) \quad \text{(signal)} \end{aligned} \tag{18}$$

where $\tilde{\mathcal{H}}_0$ and $\tilde{\mathcal{H}}_1$, respectively, denote the transformed null and alternate hypothesis while $\mathcal{T}(\mathbf{s})$ denotes multiscale true signal coefficients or the multiscale version of *signal only* case and $\mathcal{T}(\boldsymbol{\eta})$ are multiscale noise (only) coefficients.

Based on proposed hypothesis testing problem for multiscale denoising in Equation (18), foundations of the *multiscale detection theory* can be built. To this end, a scale adaptive threshold λ_k may be obtained by fixing the probability of a false alarm at the kth scale, i.e., $P_{fa}^{(k)} = \alpha^{(k)}$. Now a test statistic $\mathcal{S}(u_k^{(\mathbf{i})})$ may be employed to compute the statistical distance between the multiscale

coefficients $u_k^{(i)}$ from the distribution of noise at multiple scales, i.e., distribution of $\mathcal{T}(\boldsymbol{\eta})$. Henceforth, the transformed hypothesis testing problem in Equation (18) can be re-written as follows

$$
\begin{aligned}
\tilde{\mathcal{H}}_0 &: \mathcal{S}(u_k^{(\mathbf{i})}) \leq \lambda_k; \quad i.e., \quad u_k^{(\mathbf{i})} \in \mathcal{T}(\boldsymbol{\eta}) \\
\tilde{\mathcal{H}}_1 &: \mathcal{S}(u_k^{(\mathbf{i})}) > \lambda_k; \quad i.e., \quad u_k^{(\mathbf{i})} \in \mathcal{T}(\mathbf{s}).
\end{aligned}
\tag{19}
$$

Remark 1. *Null and alternate hypotheses in the proposed approach correspond to the noise only and signal only detections at multiple scales (i.e., $\tilde{\mathcal{H}}_0 : \mathcal{T}(\boldsymbol{\eta})$ & $\tilde{\mathcal{H}}_1 : \mathcal{T}(\mathbf{s})$ respectively), whereas the null and alternate hypothesis in traditional detection problem correspond to noise only and signal plus noise detections at original signal (image) scale (i.e., detection of $\mathcal{H}_0 : \boldsymbol{\eta}$ & $\mathcal{H}_1 : \mathbf{S} + \boldsymbol{\eta}$ respectively).*

5.3. Estimation of Threshold λ_k

As a consequence of the modified hypothesis testing problem in Equation (19), the definition of $P_{fa}^{(k)}$ is also modified accordingly, which directly follows from Equation (11) as

$$
P_{fa}^{(k)} = Prob(\mathcal{S}(u_k^{(\mathbf{i})}) > \lambda_k \mid \tilde{\mathcal{H}}_0) = \int_{\{u_k^{(\mathbf{i})} \mid \mathcal{S}(u_k^{(\mathbf{i})}) > \lambda_k\}} p(u_k^{(\mathbf{i})} \mid \tilde{\mathcal{H}}_0) \mathrm{d}u_k^{(\mathbf{i})},
\tag{20}
$$

where $\{u_k^{(\mathbf{i})} \mid \mathcal{S}(u_k^{(\mathbf{i})}) > \lambda_k\}$ is the set of multiscale noise coefficients which are falsely detected as signal, i.e., the set of coefficients yielding false alarms.

In the proposed framework, the threshold λ_k for each scale k is estimated using Equation (20) for a given probability of false alarm $P_{fa}^{(k)}$ at scale k. For that purpose, probability distribution function of noise coefficients at multiple scales $p(u_k^{(\mathbf{i})} \mid \tilde{\mathcal{H}}_0)$ is required. The challenge is that the $\mathcal{T}(\cdot)$ might change input noise distribution, e.g., UWT does not retain the Gaussianity at multiple scales. As a result, the probability density function $p(u_k^{(\mathbf{i})} \mid \tilde{\mathcal{H}}_0)$ at scale k is obtained by taking the derivative of the empirically estimated null (cumulative) distribution function $\mathcal{F}_0^{(k)}(t)$ at kth scale as follows

$$
p(u_k^{(\mathbf{i})} \mid \tilde{\mathcal{H}}_0) = \frac{\mathrm{d}}{\mathrm{d}t} \mathcal{F}_0^{(k)}(t),
\tag{21}
$$

where $\frac{\mathrm{d}}{\mathrm{d}t}$ denotes the first order difference in the discrete case. The empirical estimation of the null EDF $\mathcal{F}_0^{(k)}(t)$ for a non-linear transform is discussed in the next section.

Similarly, the definition of P_d changes to the following

$$
P_d^{(k)} = Prob(\mathcal{S}(u_k^{(\mathbf{i})}) > \lambda_k \mid \tilde{\mathcal{H}}_1) = \int_{\{u_k^{(\mathbf{i})} \mid \mathcal{S}(u_k^{(\mathbf{i})}) > \lambda_k\}} p(u_k^{(\mathbf{i})} \mid \tilde{\mathcal{H}}_1) \mathrm{d}u_k^{(\mathbf{i})},
\tag{22}
$$

where $p(u_k^{(\mathbf{i})} \mid \tilde{\mathcal{H}}_0)$ denotes the probability distribution function of multiscale noise $\mathcal{T}(\boldsymbol{\eta})$ and $p(u_k^{(\mathbf{i})} \mid \tilde{\mathcal{H}}_1)$ denotes the probability distribution function of multiscale true signal $\mathcal{T}(\mathbf{s})$.

In order to estimate $P_d^{(k)}$ from Equation (22), distribution model $p(u_k^{(\mathbf{i})} \mid \tilde{\mathcal{H}}_1)$ must be known a priori. One limitation of the extension of detection theory to multiscale denoising lies in the non-availability of a prior distribution model for multiscale signal coefficients. One exception to this could be the multiscale coefficients obtained from the DWT, which have been shown to follow heavy tailed exponential distributions [26]. However, a prior assumption for $p(u_k^{(\mathbf{i})} \mid \tilde{\mathcal{H}}_1)$ may not be possible for other transforms, which means that P_d could not be easily computed for a general multiscale transform $\mathcal{T}(\cdot)$. Consequently, the tradeoff between P_{fa} and P_d is fixed experimentally.

Remark 2. *For a given noise distribution, the threshold estimation is performed only once.*

5.4. Multiscale GoF Statistics Estimation

In order to perform local hypothesis testing based on EDF, GoF test statistic $\mathcal{S}(u_k^{(\mathbf{i})})$ must be estimated for each window of local coefficients $u_k^{(\mathbf{i})}$ at multiple scales. The statistic $\mathcal{S}(u_k^{(\mathbf{i})})$ computes the distance between the EDF $\mathcal{F}^{\mathbf{i}}(t)$ of the window centered around the coefficient $u_k^{(\mathbf{i})}$ from the null distribution function $\mathcal{F}_0^{(k)}$ at scale k. The computationally convenient expressions of Anderson Darling (AD) and Cramer Von Mises (CVM) statistics used within the proposed framework are given as follows

$$\mathcal{S}_{AD}(u_k^{(\mathbf{i})}) = -L - \sum_{j=1}^{L} \frac{(2j-1)}{L} \left(\ln(\mathcal{F}_0^{(k)}(u_k^{(\mathbf{i})})) - \ln(\mathcal{F}_0^{(k)}(u_k^{(L+1-j)})) \right). \tag{23}$$

$$\mathcal{S}_{CVM}(u_k^{(\mathbf{i})}) = \frac{1}{12L} + \sum_{j=1}^{L} \left((\mathcal{F}_0^{(k)}(u_k^{(j)}) - \frac{(2j-1)}{L} \right). \tag{24}$$

For generality we denote both AD and CVM statistic by $\mathcal{S}(u_k^{(\mathbf{i})})$.

As discussed above, the non-linear transforms change the distribution of noise at multiple scale, hence, to compute $\mathcal{S}(u_k^{(\mathbf{i})})$ using Equations (23) and (24), $\mathcal{F}_0^{(k)}(t)$ must be known a priori for each scale k for a given non-linear transform $\mathcal{T}(\cdot)$. To this end, we estimate $\mathcal{F}_0^{(k)}(t)$ empirically by assuming a large sized AWGN $\boldsymbol{\eta}$ which is subsequently decomposed using multiscale transform as follows $\underline{u}_k^{(\mathbf{i})} = \mathcal{T}(\boldsymbol{\eta})$. Next, multiscale noise coefficients $\underline{u}_k^{(\mathbf{i})}$ at each scale k are divided into M windows of local coefficients centered at the spatial location $\mathbf{i}$. Subsequently, EDF $\mathcal{F}_\eta^{\mathbf{i}}(t)$ of all windows is computed followed by the ensemble average of these EDFs resulting in the reference or null distribution $\mathcal{F}_0^{(k)}(t) = Avg\{\mathcal{F}_\eta^{\mathbf{i}}(t)\} \ \forall \ \mathbf{i}$.

To give an insight into the GoF based hypothesis testing, Figure 6 plots the null distribution $\mathcal{F}_0^{(k)}(t)$ of a DTCWT scale (solid line) along with the EDF $\mathcal{F}_s^{\mathbf{i}}(t)$ of signal coefficients (dashed line) as well as the EDF $\mathcal{F}_\eta^{\mathbf{i}}(t)$ of noise coefficients (dotted line). As $\mathcal{F}_0^{(k)}(t)$ and $\mathcal{F}_s^{\mathbf{i}}(t)$ are far apart, the statistic $\mathcal{S}(u_k^{(\mathbf{i})})$ is expected to be quite large in this case and consequently suggests alternate hypothesis (detection of signal). Contrarily, $\mathcal{F}_0^{(k)}(t)$ and $\mathcal{F}_\eta^{\mathbf{i}}(t)$ are very close suggesting that $\mathcal{S}(u_k^{(\mathbf{i})})$ corresponds to noise.
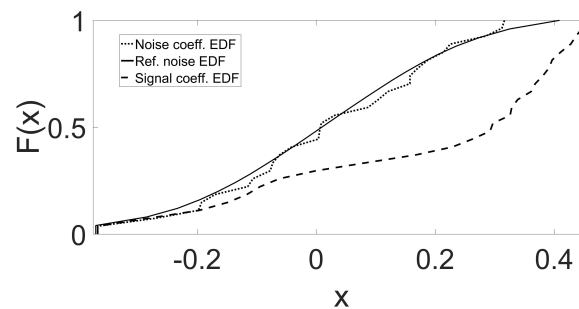

Figure 6. Hypothesis testing based signal and noise detection: $\mathcal{F}_0^{(k)}(t)$ is reference Gaussian CDF plotted along with of $\mathcal{F}_s^{\mathbf{i}}(t)$ of signal coefficients (dashed line) and the EDF $\mathcal{F}_\eta^{\mathbf{i}}(t)$ of noise coefficients (dotted line).

5.5. Multiscale Thresholding Based on Hypothesis Testing

For a threshold λ_k obtained for given $P_{fa} = \alpha$ in Equation (20), the following hard thresholding function is employed based on the proposed hypothesis testing in Equation (19) for all the windows of multiscale coefficients corresponding to noisy signal

$$\hat{u}_k^{(i)} = \begin{cases} 0 & if \ \mathcal{S}(u_k^{(\mathbf{i})}) \leq \lambda_k, \\ u_k^{(\mathbf{i})} & if \ \mathcal{S}(u_k^{(\mathbf{i})}) > \lambda_k, \end{cases} \tag{25}$$

where the central coefficient $u_k^{(\mathbf{i})}$ is replaced by zero if null hypothesis is fulfilled, i.e., $\mathcal{S}(u_k^{(\mathbf{i})}) \leq \lambda_k$; otherwise $u_k^{(\mathbf{i})}$ is retained yielding the thresholded coefficients $\hat{u}_k^{(\mathbf{i})}$.

In order to obtain the denoised image, the thresholded multiscale coefficients from Equation (25) are reconstructed by employing the transform $\mathcal{T}^{-1}(\cdot)$ as follows

$$\hat{\mathbf{S}} = \mathcal{T}^{-1}(\hat{u}_k^{(\mathbf{i})}), \tag{26}$$

where $\hat{s}$ an estimate of the true signal (or image) **s**, or simply stated the denoised signal or image. Implementation of various forward and inverse wavelet transforms is reported in [54–57].

5.6. Inverse VST

In the case of Poisson denoing, exact unbiased inverse of Anscombe transform (Inv-AT) [20] of the $\hat{\mathbf{S}}$ is performed to obtain denoised image $\hat{\hat{\mathbf{S}}}$. Similarly for Poisson–Gaussian denoising, inverse generalized Anscombe transformation (Inv-GAT) performed on the Gaussian denoised $\hat{\mathbf{S}}$ to obtain the Poisson–Gaussian denoised image $\hat{\hat{\mathbf{S}}}'$.

6. Experimental Results

This section presents image denoising results of proposed methods and gives comparison with respect to the state-of-the-art. To that end, we choose to employ undecimated wavelet transform (UWT) and dual tree complex wavelet transform (DTCWT) as transform domain methods within the proposed framework. The test statistics include AD and CVM for hypothesis testing within the framework of the GoF test. The input images used in this study include standard test images 'Lena', 'Plane', 'Peppers' and 'Boat' while two images, respectively, capturing aerial views of 'Padma River' and 'Ogden Valley' are obtained from on-line repository https://earthobservatory.nasa.gov/, which freely allows use, publication and redistribution of images. Figure 7 shows 'Lena' (https://en.wikipedia.org/wiki/File: Lenna_(test_image).png), 'Padma River' and 'Ogden Valley' while standard images such as 'Boat', 'Plane' and 'Peppers' are provided as supporting material with the manuscript, see the Figure S1 in the Supplementary Materials. We report quantitative denoising performance using the peak signal to noise ratio (PSNR), where each reported PSNR value is the average of $J = 20$ iterations. Furthermore, a representative denoising example of a CMOS image corrupted with real noise is demonstrated where the real noisy image is obtained from the *RENOIR* dataset introduced in [58].
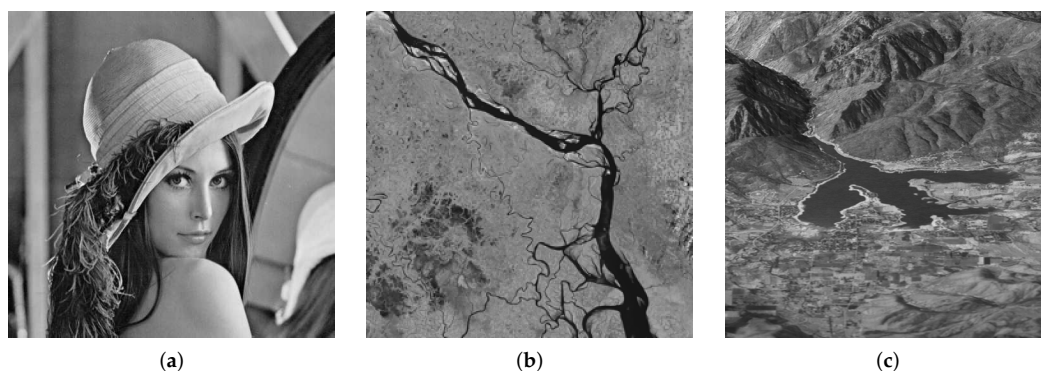

|(a)|(b)|(c)|

Figure 7. Input (2D) signals or images used for experimentation in this work including (**a**) 'Lena' image, (**b**) 'Padma River' image and (**c**) 'Ogden Valley' image.

6.1. Poisson Denoising

In this section, we discuss the performance of the proposed Poisson denoising method against the state-of-the-art. For Poisson denoising using the proposed framework, UWT and DTCWT were employed as transform domain methods while AD statistics was used as test statistic within the GoF framework. We compare the proposed Poisson denoising methods against *MSVST* [28], *NLPCA* [18],

PureLet [14] and *Poiss-NLM* [17]. The set of input test images is composed of standard 'Lena', 'Plane', 'Peppers' and 'Boat' images along with two images capturing the aerial view of 'Padma River' and 'Ogden Valley'. These images were corrupted by signal-dependent Poisson noise arising from CMOS/CCD sensors where Poisson noise with varying intensities was added to the input images to simulate sensor noise. Since noise here is signal-dependent, increasing the peak amplitude of signal, increases the peak signal to noise ratio (PSNR) of the noisy image.

Each input image was corrupted by Poisson noise at varying signal peaks, i.e., 1–100 and the resulting input PSNR values are listed in Table 1, along with output PSNR values obtained by denoising these images using the state-of-the-art and the proposed Poisson denoising methods. Results in Table 1 show that the proposed *AT-AD-DTCWT* method yielded highest output PSNR values on most instances when compared to the other methods. The proposed *AT-AD-UWT* method also demonstrated comparable performance by consistently yielding second or third highest output PSNR values while at times it also managed to outperform all of the comparative methods. *PureLet*, which is considered as the gold standard method in Poisson denoising, remained competitive against the proposed methods and managed to beat them at a few input noise levels. *Poiss-NLM* also showed comparable denoising results but it mostly remained behind the *PureLet* and *AT-AD-DTCWT* methods in terms of the output PSNR values. *MSVST* failed to match the performance of the best methods but showed good denoising performance. *NLPCA* shows competitive performance on higher noise levels but as the noise level was reduced in the signal, *NLPCA* failed to improve its performance for all images.

Table 1. Poisson denoising: Output PSNR value obtained from the proposed method and state-of-the-art comparative methods for input images at varying input PSNR levels (bold denotes highest PSNR).

Peaks	Inp. PSNR	MS-VST	NL-PCA	Pure-Let	Pois-NLM	AT-AD UWT	AT-AD DTCWT	Inp. PSNR	MS VST	NL PCA	Pure-Let	Pois-NLM	AT-AD UWT	AT-AD DTCWT
					Lena							Plane		
1	2.93	17.19	20.67	22.08	19.73	21.65	**22.22**	1.11	14.62	19.06	20.53	17.33	19.80	**20.45**
2	5.96	17.70	20.87	**23.31**	21.40	23.02	23.26	4.15	14.61	19.09	21.68	19.34	21.34	**21.70**
3	7.73	17.86	20.74	**23.98**	22.64	23.85	23.87	5.90	14.86	19.18	22.56	20.71	22.41	**22.77**
4	8.95	18.04	20.77	24.51	23.57	24.73	**24.86**	7.16	15.43	19.19	23.07	21.65	23.02	**23.32**
5	9.95	18.44	20.78	24.89	24.28	**25.67**	25.43	8.13	15.94	19.15	23.54	22.40	23.58	**23.96**
10	12.97	20.23	20.49	26.27	26.39	26.79	**27.18**	11.13	18.09	19.07	24.90	24.44	25.21	**25.58**
20	15.97	22.39	20.32	27.75	27.54	28.47	**28.84**	14.13	20.48	18.84	26.42	25.59	26.86	**27.29**
50	19.91	25.45	17.98	29.58	30.15	30.48	**30.75**	18.12	23.65	18.44	28.48	28.20	29.09	**29.49**
100	22.95	27.78	19.53	31.28	32.23	32.07	**32.32**	21.13	26.04	18.73	30.14	30.42	30.74	**31.23**
					Peppers							Boat		
1	2.73	16.61	18.82	**21.45**	19.28	20.96	21.44	2.96	16.83	20.17	21.28	19.06	20.72	**21.36**
2	5.76	17.00	18.94	22.80	21.12	22.58	**22.80**	5.96	17.43	20.22	22.22	20.48	21.89	**22.24**
3	7.54	17.25	18.81	**23.55**	22.40	23.26	23.39	7.72	17.47	20.29	**22.77**	21.56	22.63	22.74
4	8.80	17.58	18.87	24.17	23.36	**24.22**	24.18	8.94	17.59	20.31	23.27	22.39	23.11	**23.44**
5	9.72	17.96	18.74	24.49	24.04	24.85	**24.89**	9.93	17.92	20.20	23.69	22.98	23.62	**23.86**
10	12.75	19.83	18.36	25.75	26.01	26.34	**26.38**	12.94	19.72	20.21	24.90	24.77	24.88	**25.23**
20	15.79	22.05	18.44	27.15	27.00	27.73	**27.81**	15.94	21.79	20.09	26.19	25.80	26.28	**26.72**
50	19.72	24.81	16.67	29.01	29.07	**29.37**	29.51	19.90	24.71	19.99	28.06	28.09	28.14	**28.67**
100	22.76	26.87	16.63	**30.61**	30.35	30.55	30.55	22.96	26.85	19.81	29.78	30.07	29.66	**30.22**
					Padma River							Ogden Valley		
1	3.27	16.71	**19.63**	19.91	18.66	19.48	19.58	4.01	18.12	20.81	**21.78**	20.13	21.54	21.68
2	6.26	17.53	19.77	20.66	19.90	20.44	**20.70**	7.05	18.75	20.83	**22.66**	21.18	22.54	22.42
3	7.97	17.53	19.77	21.15	20.75	21.07	**21.17**	8.82	18.88	20.79	23.09	22.01	**23.16**	22.97
4	9.29	17.50	19.80	21.58	21.43	21.47	**21.61**	10.06	18.92	20.90	23.60	22.64	**23.61**	23.44
5	10.25	17.75	19.81	21.86	21.80	21.74	**21.90**	11.01	19.18	20.87	23.91	23.10	**23.96**	23.81
10	13.29	19.11	19.82	22.91	**23.11**	22.73	22.97	14.02	20.42	20.90	25.04	24.50	25.06	**25.15**
20	16.30	20.87	19.71	24.06	23.81	23.79	**24.10**	17.00	22.40	20.82	26.26	25.33	26.28	**26.58**
50	20.24	23.42	19.44	**25.92**	25.35	25.13	25.54	20.99	25.12	20.75	28.10	27.43	28.07	**28.63**
100	23.29	25.43	19.61	**27.56**	26.78	26.21	26.73	24.01	27.34	20.60	29.70	29.66	29.62	**30.28**

We also display denoised images of 'Lena' and 'Padma River', respectively, in Figures 8 and 9. In Figure 8, noisy 'Lena' image at signal peak = 20 is displayed in Figure 8a along with the denoised 'Lena' images by *MSVST*, *Poiss. NLM*, *PureLet*, *AT-AD-UWT* and *AT-AD-DTCWT*, respectively, in Figure 8b–f. Observe from Figure 8d that *PureLet* yielded blurry artifacts which were spread all around

the denoised 'Lena' image. Figure 8c shows the denoised signal by the *Poiss-NLM* which seems to be devoid of the artifacts but at the cost of the loss of image details, due to over-smoothing of the denoised signal. The denoised signals by the proposed methods *AT-AD-UWT* and *AT-AD-DTCWT*, shown in Figure 8e,f respectively, showed lesser artifacts as compared to the *PureLet* and the *MSVST* while also extracting more signal details. When compared to *Poiss-NLM*, denoised images by the proposed *AT-AD-DTCWT* method extracted more details but at the expense of slight artifacts. The *AT-AD-UWT* also extracted higher signal details compared to the *Poiss-NLM* but with some visible artifacts. *PureLet* and *MSVST* also changed the brightness of the denoised images whereas the proposed methods did not alter signal brightness even at such higher noise level.

Figure 9 compares the denoising performance of the proposed *AT-AD-DTCWT* method against the *PureLet* and the *NLPCA* at very high noise level, i.e., signal peak = 5, on 'Padma River' image. Original and noisy versions of 'Padma River' image are shown in Figure 9a,b while the denoised images from *NLPCA*, *PureLet* and the proposed *AT-AD-DTCWT* are shown in Figure 9c–e respectively. *NLPCA* not only over-smoothed the recovered image details but also blurred it, see Figure 9c. *PureLet* and the proposed *AT-AD-DTCWT* managed to recover significant image details even at such a high level of signal-dependent additive noise, see Figure 9d,e. Note that *Purelet* showed significant blurring artifacts in Figure 9d while denoised image by the proposed *AT-AD-DTCWT* recovered image details effectively with very little artifacts even at such a high input noise level, see Figure 9e.
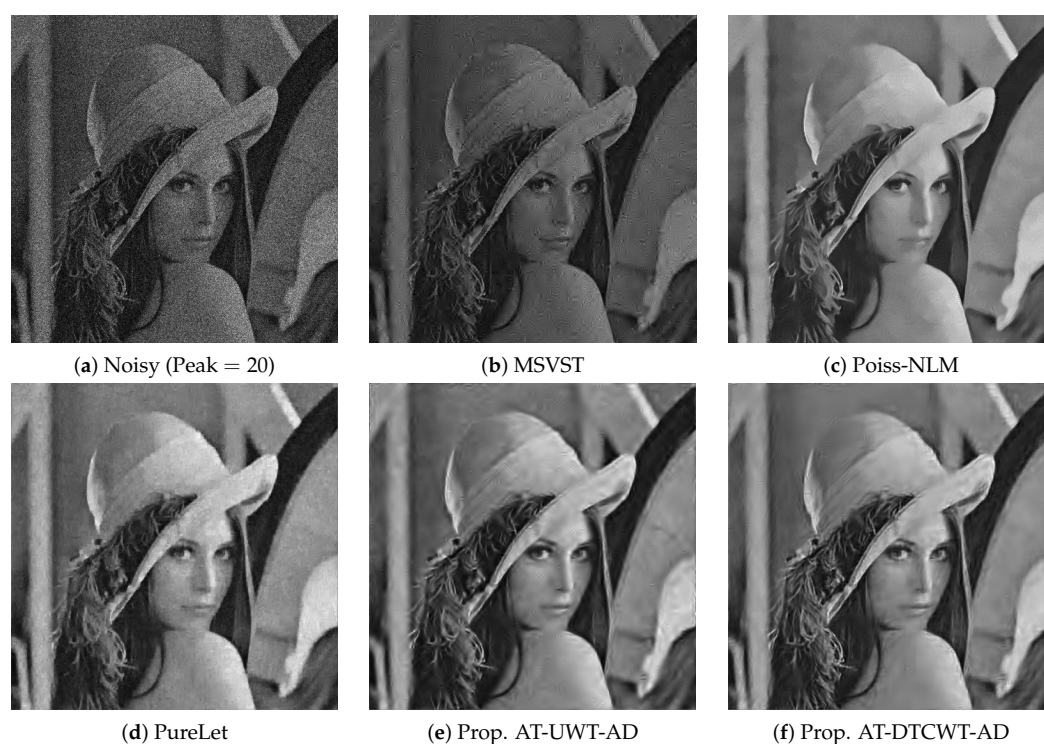

(**a**) Noisy (Peak = 20) (**b**) MSVST (**c**) Poiss-NLM

(**d**) PureLet (**e**) Prop. AT-UWT-AD (**f**) Prop. AT-DTCWT-AD

Figure 8. Poisson denoising results on the 'Lena' image by various methods at signal peak = 20.

(**a**) Original (**b**) Noisy (Peak = 5) (**c**) NLPCA (**d**) PureLet (**e**) AT-AD-DTCWT
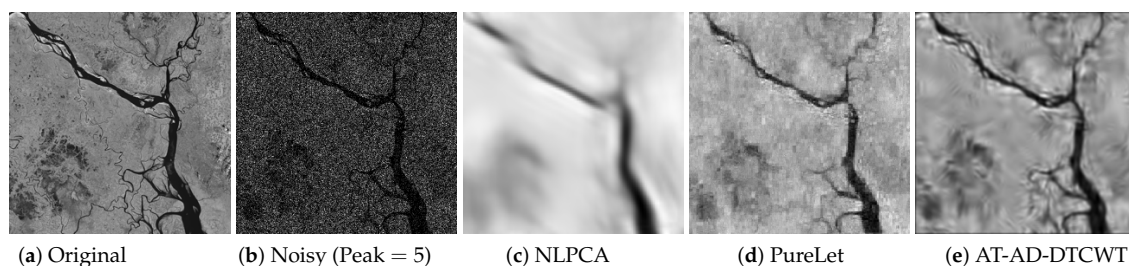
Figure 9. Poisson denoising results on the 'Padma River' image by various methods at signal peak = 5.

6.2. Poisson–Gaussian denoising

We now provide comparative results of the proposed framework for Poisson–Gaussian denoising against the existing denoising methods. We use DTCWT as a transform domain method in the proposed methodology while *AD* is used as test statistic for the GoF based hypothesis testing. We name the proposed method *GAT-AD-DTCWT* and compare it against the *PGureLet* [15], *GAT-BLSGSM* [24] and *MSVST-MPG* [27]. We report the denoising results on all input images used in previous two sections, namely, 'Lena', 'Plane', 'Peppers', 'Boat', 'Padma River' and 'Ogden Valley'. These images were corrupted with input Poisson–Gaussian noise of varying noise levels to simulate CMOS/CCD sensor noise where the strength of the Poisson noise was defined by signal peak = 1, 2 ,3, 4, 5, and 10 and standard deviation σ of Gaussian noise was selected as $\sigma = \text{peak}/10$.

Input PSNR values corresponding to these parameters of Poisson–Gaussian noise are reported in Table 2, along with the output PSNR values of the denoised images by the comparative methods. Note that denoised images from the proposed method had highest output PSNR values on most input noise levels while the *GAT-BLSGSM* yielded competitive results. It was observed that *PGureLet* showed competitive results on higher noise levels while the *GAT-BLSGSM* yielded competitive results on lower noise levels. However, the proposed *GAT-AD-DTCWT* showed consistently improved performance at all noise levels. *MSVST-MPG* yielded lowest output PSNR values among all the methods.

Figure 10 shows original, noisy and denoised 'Ogden Valley' images obtained from the comparative methods. Noisy images displayed in Figure 10b,f were, respectively, corrupted by Poisson–Gaussian noise at signal peak = 5 and 10 and AWGN standard deviation $\sigma = 0.5$ & 1. These noisy images were denoised using *PGureLet*, *GAT-BLSGSM* and proposed *GAT-AD-DTCWT* method which are, respectively, displayed in second, third and fourth columns of Figure 10. Observe that the *PGureLet* blurred the recovered images and distorted the information, see Figure 10c,g. *GAT-BLSGSM* recovered the image details well (see Figure 10d,h) as compared to the denoised images from *PGureLet*. However, denoised images by the *GAT-BLSGSM* yielded spike-like artifacts which are more evident in denoising at higher input noise level, as shown in Figure 10h. Denoised images by the proposed *GAT-AD-DTCWT* method are displayed in Figure 10e,i where not only image details have been preserved well but also the contrast and bright information is intact.

Table 2. Poisson–Gaussian denoising: Output PSNR value obtained from the proposed method and state-of-the-art comparative methods for input images at varying input PSNR levels (bold denotes highest PSNR).

Peaks	Noise std. σ	Inp. PSNR	MSVST MPG	GAT-BL SGSM	PGure-Let	GAT-AD DTCWT	Inp. PSNR	MSVST MPG	GAT-BL SGSM	PGure-Let	GAT-AD DTCWT	Inp. PSNR	MSVST MPG	GAT-BL SGSM	PGure-Let	GAT-AD DTCWT
				Lena					Plane					Lena		
1	0.1	2.87	17.06	18.63	21.98	**22.09**	1.07	14.28	19.98	20.10	**20.63**	2.69	16.47	17.27	**21.45**	21.30
2	0.2	5.79	16.70	23.10	23.12	**23.30**	4.03	13.90	21.69	21.18	**21.71**	5.63	16.23	21.16	22.59	**22.82**
3	0.3	7.48	16.41	24.57	23.65	**24.00**	5.71	13.97	22.43	21.69	**22.64**	7.30	16.04	23.26	23.12	**23.53**
4	0.4	8.63	16.64	**24.96**	23.97	24.64	6.92	14.66	23.17	21.95	**23.23**	8.48	16.48	24.21	23.39	**24.25**
5	0.5	9.53	17.17	**25.16**	24.13	25.08	7.48	15.36	23.54	22.11	**23.68**	9.37	16.98	**24.91**	23.56	24.86
10	1	12.17	19.21	26.51	24.52	**26.58**	10.6	17.46	25.43	22.45	**25.51**	12.00	18.84	**26.53**	23.83	26.02
				Boat					Padma River					Ogden Valley		
1	0.1	2.84	16.64	18.32	21.10	**21.31**	3.21	16.80	18.86	**19.90**	19.71	3.90	17.96	19.16	**21.84**	21.62
2	0.2	5.78	16.45	21.22	21.93	**22.27**	6.09	16.36	20.57	20.44	**20.69**	6.79	17.82	22.41	22.54	**22.56**
3	0.3	7.47	16.14	22.58	22.27	**22.86**	7.79	16.00	**21.28**	20.72	20.97	8.48	17.29	**23.06**	23.01	22.86
4	0.4	8.63	16.38	23.24	22.46	**23.29**	8.93	16.20	21.27	20.84	**21.39**	9.62	17.26	23.21	23.13	**23.23**
5	0.5	9.52	16.91	23.61	22.60	**23.69**	9.83	16.64	21.85	20.91	**21.93**	10.48	17.61	**23.65**	23.35	23.54
10	1	12.15	18.69	**25.12**	22.86	24.88	12.43	18.14	**23.02**	21.04	22.59	13.04	19.31	24.52	23.62	**24.64**

(**a**) Original

(**b**) Noisy (Peak = 10, $\sigma = 1$)   (**c**) PGureLet   (**d**) GAT-BLSGSM   (**e**) GAT-AD-DTCWT

(**f**) Noisy (Peak = 5, $\sigma = 0.5$)   (**g**) PGureLet   (**h**) GAT-BLSGSM   (**i**) GAT-AD-DTCWT
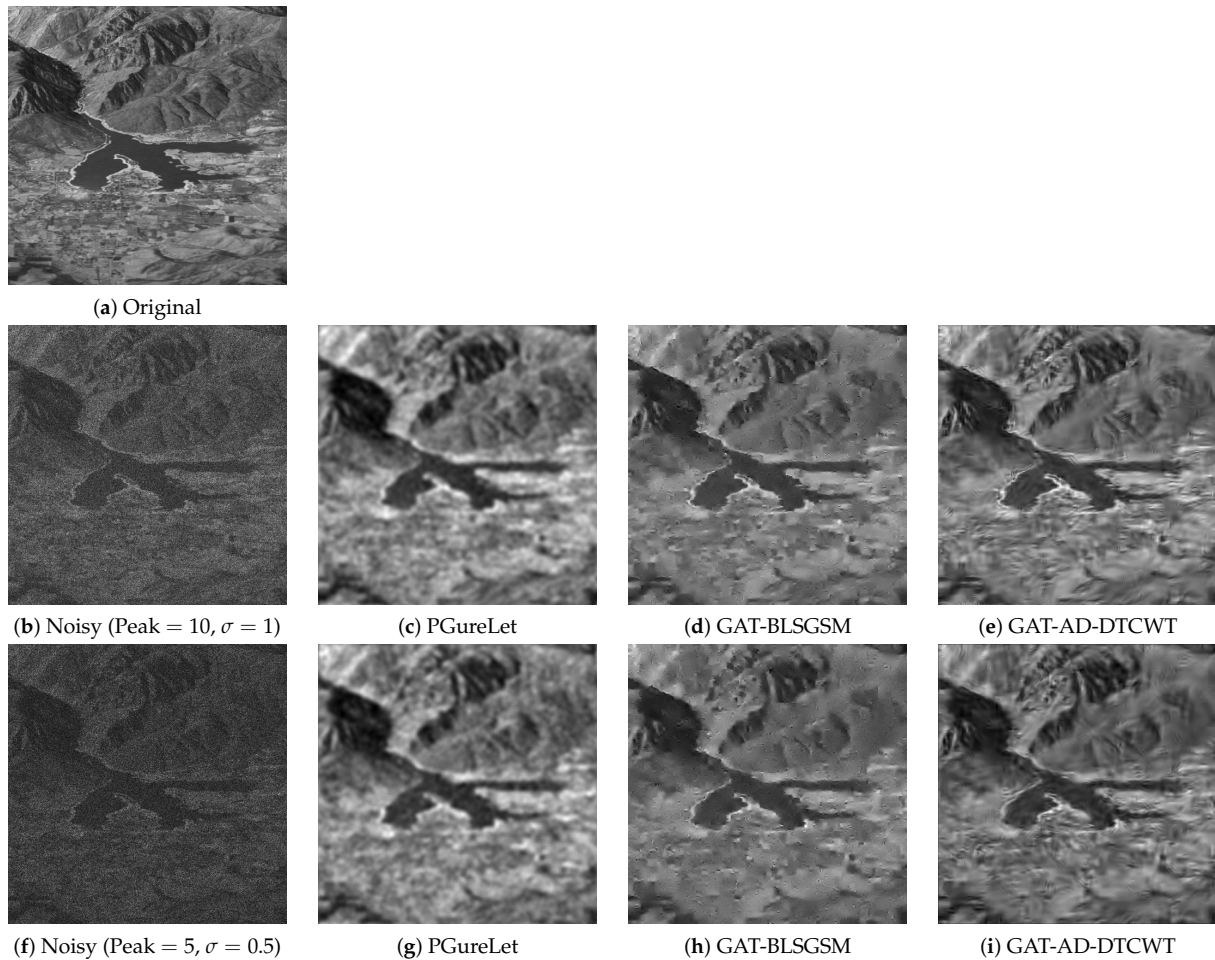
Figure 10. Poisson–Gaussian denoising results on the 'Ogden Valley' image by various methods where noisy image is corrupted by Poisson noise at signal peak = 10 & $\sigma = 1$ (second row) and at signal peak = 5 & $\sigma = 0.5$ (third row).

6.3. A Denoising Example of an Image Obtained from CMOS Sensor

In this section, we employ proposed *GAT-AD-DTCWT* to suppress real CMOS sensor noise from an image obtained through a CMOS camera installed in Xiomi Mi3 mobile. This image is made freely available as part of the RENOIR dataset [58] containing CMOS images corrupted by sensor noise. The study in [58] not only offers a noisy dataset but also compares the performance of the state-of-the-art Poisson and Poisson–Gaussian denoising methods for removing the sensor noise.

Figure 11 displays the noisy CMOS image along with the denoised image using the *GAT-AD-DTCWT* method. Noisy image (top row) and the zoomed in view of the highlighted part (lower row) are, respectively, shown in Figure 11a, while the denoised image using the *GAT-AD-DTCWT* (top row) and the zoomed in view of the highlighted region (lower row) in Figure 11b. As can be observed from the Figure 11 (top row) that the proposed method successfully suppresses majority of the real CMOS sensor noise. The zoomed-in view of the highly detailed region in the noisy and the denoised image are shown in Figure 11 (lower row), where granular noise pattern or shot noise spikes are visible in the noisy image patch. However, in the denoised image patch, these granular patterns have been successfully suppressed by the proposed method.
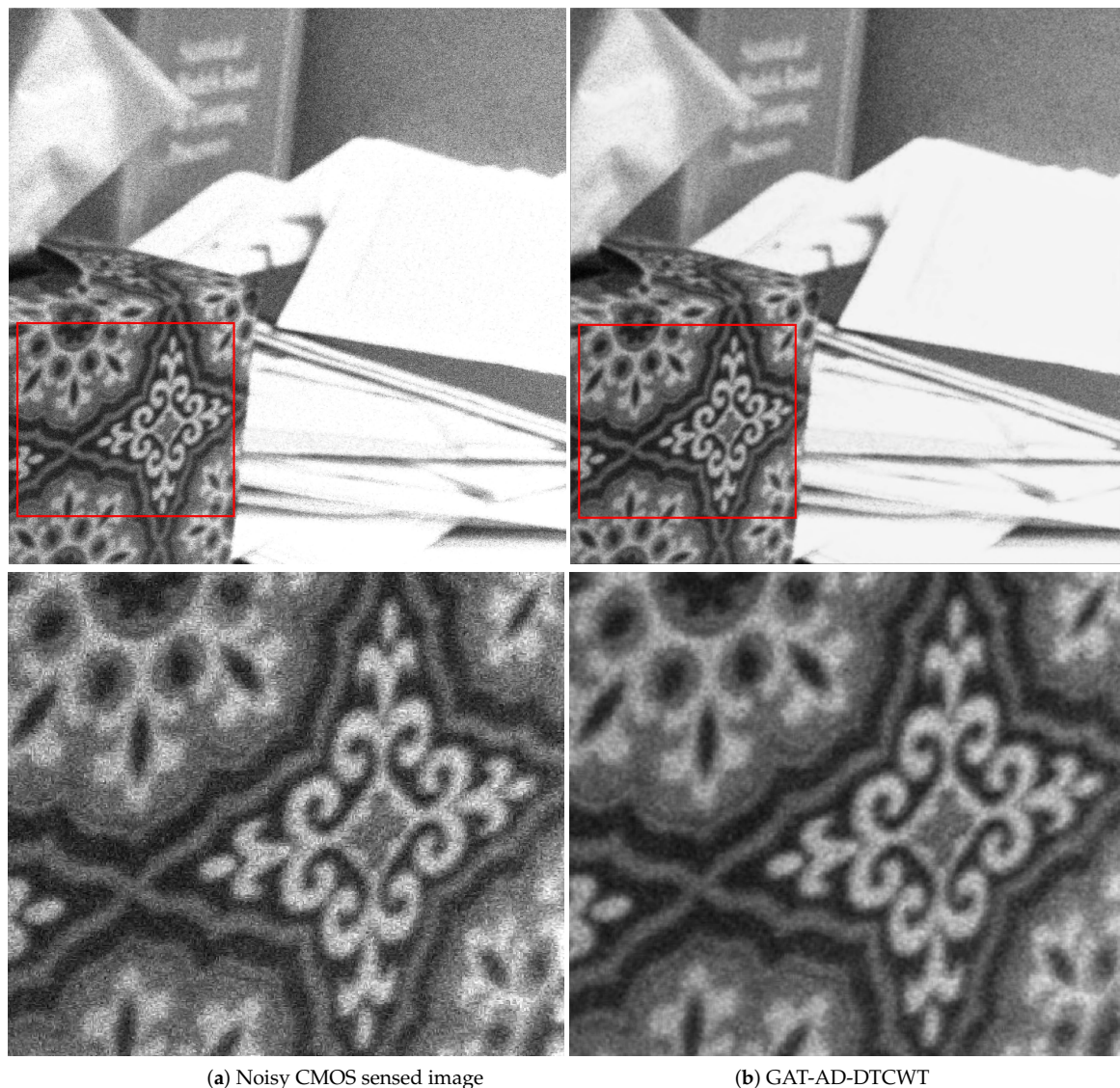
(**a**) Noisy CMOS sensed image　　　　　　　　　(**b**) GAT-AD-DTCWT

Figure 11. Performance analysis of the proposed *GAT-AD-DTCWT* on a noisy image obtained from RENOIR dataset [58] which contains noisy images from CMOS sensors corrupted by real sensor noise.

7. Discussion and Conclusions

This article proposes a generalized denoising framework based on detection theory and applies it to remove Poisson and Poisson–Gaussian noise from CMOS/CCD image sensors. To this end, variance stability transformation (VST) has been combined with the proposed binary hypothesis testing framework to enable the detection and removal of Poisson and Poisson–Gaussian noise at multiple scales. For local hypothesis testing, statistical measures of goodness-of-fit test based on empirical distribution function (EDF), i.e., Anderson Darling (AD), Cramer Von Mises (CVM) have been employed in our work. Furthermore, different 2D transform domain methods have been tested within the proposed framework.

The proposed methodology has been shown to outperform the comparative state-of-the-art methods in Poisson and Poisson–Gaussian denoising. This could be attributed to the effectiveness of the proposed framework to handle the non-standard noise distributions due to its data driven nature. To further stress this point, an example of denoising a CMOS image corrupted with real sensor noise using the proposed *GAT-AD-DTCWT* is also presented which demonstrates the efficacy of the proposed framework for suppressing the noise due to CMOS/CCD sensors.

Computational complexity of the proposed framework can be minimized by offline estimation of thresholds versus P_{fa} table which is only required to be computed only once. Similarly, the estimation of the reference CDF (in case of nonlinear transformation) may be also be performed offline. Other computationally intensive step involved in the proposed method is the local estimation of the EDF which requires computations of the order of $O(L\ logL)$ while rest of the steps in the proposed methodology require computations equivalent to standard multiscale denoising methods.

The scope of this work is limited to the algorithm design for denoising CMOS/CCD images. However, in the following, we discuss few important aspects related to the hardware implementation of the proposed method. The proposed algorithm has a potential to be implemented in six stage-pipelined architecture, which enables parallel computations and increases the throughput of the system in real-time. The first-pipelined stage applies VST as a pre-processing step on the input noisy signal. The computationally expensive part of this stage is to evaluate square-root, CORDIC algorithms are used to compute square-root in hardware platforms (e.g., microcontrollers, processors and FPGAs). The second-pipelined stage computes the transform (UWT/DTCWT) operation on the pre-processed signal of the first stage. Real-time implementations of different transforms are reported in the literature [54–57]. Addition and average mathematical operations are needed to be executed in the third-pipelined stage of the proposed algorithm. In addition, these operations are executed in windowing fashion, therefore different blocks can be executed in parallel. In the fourth-pipelined stage, hard thresholding is done using GOF thresholding technique, which can be done by comparator implementation. The fifth-pipelined stage computed inverse transform operation, which has same computational complexity as the forward transform. Sixth-pipelined stage computed the exact inverse VST which requires pre-computed tables to empirically remove the bias at low photon count as stated in [20,25]. These tables can be stored and used in look-up table to decrease the computational complexity of this stage.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1: Standard input images used for performance analysis of various denoising methods in this study including (a) 'Peppers' image, (b) 'Boat' image, (c) 'Plane' image.

Author Contributions: Conceptualization, K.N. and N.u.R.; Data curation, K.N.; Formal analysis, K.N. and N.u.R.; Funding acquisition, S.E. and K.D.M.-M.; Investigation, K.N., S.E., K.D.M.-M. and N.u.R.; Methodology, K.N. and N.u.R.; Project administration, K.D.M.-M. and N.u.R.; Resources, K.N.; Software, K.N.; Supervision, N.u.R.; Validation, K.N. and N.u.R.; Visualization, K.N., S.E., K.D.M.-M. and N.u.R.; Writing—original draft, K.N.; Writing—review and editing, K.N., S.E., K.D.M.-M. and N.u.R.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kawai, N.; Kawahito, S. Noise analysis of high-gain, low-noise column readout circuits for CMOS image sensors. *IEEE Trans. Electron Devices* **2004**, *51*, 185–194. [CrossRef]
2. Cho, M.; Javidi, B. Three-dimensional photon counting imaging with axially distributed sensing. *Sensors* **2016**, *16*, 1184. [CrossRef] [PubMed]
3. Lundberg, K.H. Noise Sources in Bulk CMOS. Unpublished Paper. 2002. Available online: www.mit.edu/~klund/papers/UNP_noise.pdf (accessed on 8 December 2018).
4. Foi, A.; Trimeche, M.; Katkovnik, V.; Egiazarian, K. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE Trans. Image Process.* **2008**, *17*, 1737–1754. [CrossRef] [PubMed]
5. Lee, S.W.; Maik, V.; Jang, J.; Shin, J.; Paik, J. Noise-adaptive spatio-temporal filter for real-time noise removal in low light level images. *IEEE Trans. Consum. Electron.* **2005**, *51*, 648–653.
6. Kolaczyk, E.D. Nonparametric estimation of intensity maps using Haar wavelets and Poisson noise characteristics. *Astrophys. J.* **2000**, *534*, 490–505. [CrossRef]

7.	Zhang, Y.; Guangyi, W.; Xu, J. Parameter estimation of signal-dependent random noise in CMOS/CCD image sensor based on numerical characteristic of mixed Poisson noise samples. *Sensors* **2018**, *18*, 2276. [CrossRef] [PubMed]

8.	Lee, S.; Lee, M.S.; Kang, M.G. Poisson–Gaussian Noise Analysis and Estimation for Low-Dose X-ray Images in the NSCT Domain. *Sensors* **2018**, *18*, 1019.

9.	Zhang, J.; Hirakawa, K. Improved denoising via Poisson mixture modeling of image sensor noise. *IEEE Trans. Image Process.* **2017**, *26*, 1565–1578. [CrossRef]

10.	Li, D.; Sun, C.; Yang, J.; Liu, H.; Peng, J.; Zhang, L. Robust multi-frame adaptive optics image restoration algorithm using maximum likelihood estimation with poisson statistics. *Sensors* **2017**, *17*, 785. [CrossRef]

11.	Gao, Q.; Eck, S.; Matthias, J.; Chung, I.; Engelhardt, J.; Rippe, K.; Rohr, K. Bayesian joint super-resolution, deconvolution, and denoising of images with Poisson–Gaussian noise. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 938–942.

12.	Niknejad, M.; Figueiredo, M.A. Poisson Image Denoising Using Best Linear Prediction: A Post-Processing Framework. *arXiv* **2018**, arXiv:1803.00389.

13.	Donoho, D.L.; Johnstone, I.M. Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Stat. Assoc.* **1995**, *90*, 1200–1224. [CrossRef]

14.	Luisier, F.; Vonesch, C.; Blu, T.; Unser, M. Fast interscale wavelet denoising of Poisson-corrupted images. *Signal Process.* **2010**, *90*, 415–427. [CrossRef]

15.	Luisier, F.; Blu, T.; Unser, M. Image denoising in mixed Poisson–Gaussian noise. *IEEE Trans. Image Process.* **2011**, *20*, 696–708. [CrossRef] [PubMed]

16.	Li, J.; Luisier, F.; Blu, T. PURE-LET image deconvolution. *IEEE Trans. Image Process.* **2018**, *27*, 92–105. [CrossRef] [PubMed]

17.	Deledalle, C.A.; Tupin, F.; Denis, L. Poisson NL means: Unsupervised non local means for Poisson noise. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Hong Kong, China, 26–29 September 2010; pp. 801–804.

18.	Salmon, J.; Harmany, Z.; C.-Deledalle, A.; Willett, R. Poisson noise reduction with non-local PCA. *J. Math. Imaging Vis.* **2014**, *48*, 279–294. [CrossRef]

19.	Donoho, D.L. Nonlinear wavelet methods for recovery of signals, densities and spectra from indirect and noisy data. In *Proceedings of Symposia in Applied Mathematics: Different Perspectives on Wavelets*; American Mathematical Society: Providence, RI, USA, 1993; Volume 47, pp. 173–205.

20.	Anscombe, F.J. The transformation of Poisson, binomial and negative- binomial data. *Biometrika* **1948**, *35*, 246–254. [CrossRef]

21.	Makitalo, M.; Foi, A. On the inversion of the Anscombe transformation in low-count Poisson image denoising. In Proceedings of the IEEE International Workshop on Local and Non-Local Approximation in Image Processing, Tuusula, Finland, 19–21 August 2009; pp. 26–32.

22.	Makitalo, M.; Foi, A. Optimal inversion of the Anscombe transformation in low-count Poisson image denoising. *IEEE Trans. Image Process.* **2011**, *20*, 99–109. [CrossRef] [PubMed]

23.	Azzari, L.; Foi, A. Variance Stabilization for Noisy+ Estimate Combination in Iterative Poisson Denoising. *IEEE Signal Process. Lett.* **2016**, *23*, 1086–1090. [CrossRef]

24.	Makitalo, M.; Foi, A. Optimal inversion of the generalized Anscombe transformation for Poisson–Gaussian noise. *IEEE Trans. Image Process.* **2013**, *22*, 91–103. [CrossRef]

25.	Starck, J.L.; Murtagh, F.; Bijaoui, A. *Image Processing and Data Analysis*; Cambridge University Press: Cambridge, UK, 1998.

26.	Portilla, J.; Strela, V.; Wainwright, M.J.; Simoncelli, E.P. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Trans. Image Process.* **2003**, *12*, 1338–1351. [CrossRef]

27.	Zhang, B.; Fadili, M.J.; Starck, J.L.; Olivo-Marin, J.C. Multiscale variance-stabilizing transform for mixed-Poisson–Gaussian processes and its applications in bioimaging. In Proceedings of the IEEE International Conference on Image Processing (ICIP), San Antonio, TX, USA, 16 September–19 October 2007; Volume 6, pp. VI–233.

28.	Zhang, B.; Fadili, J.M.; Starck, J.L. Wavelets, ridgelets, and curvelets for Poisson noise removal. *IEEE Trans. Image Process.* **2008**, *17*, 1093–1108. [CrossRef]

29. Zhang, B.; Fadili, J.M.; Starck, J.L.; Digel, S.W. Fast Poisson noise removal by biorthogonal Haar domain hypothesis testing. *Stat. Methodol.* **2008**, *5*, 387–396. [CrossRef]

30. Charles, C.; Rasson, J.P. Wavelet denoising of Poisson-distributed data and applications. *Comput. Stat. Data Anal.* **2003**, *43*, 139–148. [CrossRef]

31. ur Rehman, N.; Abbas, S.Z.; Asif, A.; Javed, A.; Naveed, K.; Mandic, D.P. Translation invariant multi-scale signal denoising based on goodness-of-fit tests. *Signal Process.* **2017**, *131*, 220–234. [CrossRef]

32. Naveed, K.; Shaukat, B.; ur Rehman, N. Dual tree complex wavelet transform-based signal denoising method exploiting neighbourhood dependencies and goodness-of-fit test. *R. Soc. Open Sci.* **2017**, *5*, 180436. [CrossRef]

33. Naveed, K.; Shaukat, B.; ur Rehman, N. Signal denoising based on dual tree complex wavelet transform and goodness of fit test. In Proceedings of the IEEE International Conference on Digital Signal Processing (DSP), London, UK, 23–25 August 2017; pp. 1–5.

34. ur Rehman, N.; Naveed, K.; Ehsan, S.; McDonald-Maier, K. Multi-scale image denoising based on goodness-of-fit tests. In Proceedings of the European Signal Processing Conference (EUSIPCO), Budapest, Hungary, 29 August–2 September 2016; pp. 1548–1552.

35. Naveed, K.; Shaukat, B.; Ehsan, S.; McDonald-Maier, K.; ur Rehman, N. Multiscale image denoising using goodness-of-fit test based on EDF statistics. *PLOS ONE* **2018**, submitted.

36. Nakamura, J. *Image Sensors and Signal Processing for Digital Still Cameras*; CRC Press: Boca Raton, FL, USA, 2016.

37. Chiabrando, F.; Chiabrando, R.; Piatti, D.; Rinaudo, F. Sensors for 3D imaging: Metric evaluation and calibration of a CCD/CMOS time-of-flight camera. *Sensors* **2009**, *9*, 10080–10096. [CrossRef] [PubMed]

38. Gamal, A.E.; Eltoukhy, H. CMOS image sensors. *IEEE Circuits Devices Mag.* **2005**, *21*, 6–20. [CrossRef]

39. Rizzolo, S.; Goiffon, V.; Estribeau, M.; Paillet, P.; Marcandella, C.; Durnez, C.; Magnan, P. Total-Ionizing Dose Effects on Charge Transfer Efficiency and Image Lag in Pinned Photodiode CMOS Image Sensors. *IEEE Trans. Nucl. Sci.* **2018**, *65*, 84–91. [CrossRef]

40. Yeh, S.F.; Chou, K.Y.; Tu, H.Y.; Chao, C.Y.P.; Hsueh, F.L. A 0.66 e rms-Temporal-Readout-Noise 3-D-Stacked CMOS Image Sensor With Conditional Correlated Multiple Sampling Technique. *IEEE J. Solid-State Circuits* **2018**, *53*, 527–537. [CrossRef]

41. Theuwissen, A.J. CMOS image sensors: State-of-the-art. *Solid-State Electron.* **2008**, *52*, 1401–1406. [CrossRef]

42. Gow, R.D.; Renshaw, D.; Findlater, K.; Grant, L.; McLeod, S.J.; Hart, J.; Nicol, R.L. A comprehensive tool for modeling CMOS image-sensor-noise performance. *IEEE Trans. Electron Devices* **2007**, *54*, 1321–1329. [CrossRef]

43. Kay, S.M. *Fundamentals of Statistical Signal Processing: Detection Thery*; Prentice Hall PTR: Upper Saddle River, NJ, USA, 1993; Volume II.

44. Reich, E.; Swerling, P. Detection of a sine wave in Gaussian noise. *J. Appl. Phys* **1953**, *24*, 289–296. [CrossRef]

45. Anderson, T.W.; Darling, D.A. A Test of Goodness-of-Fit. *J. Am. Stat. Assoc.* **1954**, *49*, 765–769. [CrossRef]

46. Cramr, H. On the composition of elementary errors. *Scand. Actuar. J.* **1928**, *11*, 141–180. [CrossRef]

47. Lei, S.; Wang, H.; Shen, L. Spectrum sensing based on goodness-of-fit tests. In Proceedings of the IEEE International Conference on Electronics, Communications and Control, Ningbo, China, 9–11 September 2011; pp. 485–489.

48. Wang, H.; Yang, E.; Zhao, Z.; Zhang, W. Spectrum Sensing in Cognitive Radio Using goodness-of-fit Testing. *IEEE Trans. Wirel. Commun.* **2009**, *8*, 5427–5430. [CrossRef]

49. Peacock, J.A. Two-dimensional goodness-of-fit testing in astronomy. *Mon. Not. R. Astron. Soc.* **1983**, *202*, 615–627. [CrossRef]

50. Cho, D.; Bui, T.D. Multivariate statistical modeling for image denoising using wavelet transforms. *Signal Process. Image Commun.* **2005**, *20*, 77–89. [CrossRef]

51. Selesnick, I.W. The double density DWT. In *Wavelets in Signal and Image Analysis*; Springer: Dordrecht, The Netherlands, 2001; pp. 39–66.

52. Selesnick, I.W.; Ivan, W.; Baraniuk, R.G.; Kingsbury, N.C. The dual-tree complex wavelet transform. *IEEE Signal Process. Mag.* **2005**, *22*, 123–151. [CrossRef]

53. Starck, J.L.; Fadili, J.; Murtagh, F. The undecimated wavelet decomposition and its reconstruction. *IEEE Trans. Image Process.* **2007**, *16*, 297–309. [CrossRef]

54. Singh, R.; Mohanty, S.R.; Kishor, N.; Thakur, A. Real-Time Implementation of Signal Processing Techniques for Disturbances Detection. *IEEE Trans. Ind. Electron.* **2018**. [CrossRef]

55. Kumar, A.; Rastogi, P.; Srivastava, P. Design and FPGA Implementation of DWT, Image Text Extraction Technique. *Procedia Comput. Sci.* **2015**, *57*, 1015–1025. [CrossRef]

56. Zhang, J.; Chow, P.; Liu, H. FPGA implementation of low-power and high-PSNR DCT/IDCT architecture based on adaptive recoding CORDIC. In Proceedings of the IEEE International Conference on Field Programmable Technology (FPT), Queenstown, New Zealand, 7–9 December 2015; pp. 128–135.

57. Siddiqui, M.F.; Reza, A.W.; Kanesan, J.; Ramiah, H. Investigation of a novel common subexpression elimination method for low power and area efficient DCT architecture. *Sci. World J.* **2014**, *2014*, 620868. [CrossRef] [PubMed]

58. Anaya, J.; Barbu, A. RENOIR–A dataset for real low-light image noise reduction. *J. Vis. Commun. Image Represent.* **2018**, *51*, 144–154. [CrossRef]