Minimum Risk Wavelet Shrinkage Operator For Poisson Image Denoising

Wu Cheng, Student Member, IEEE, and Keigo Hirakawa, Senior Member, IEEE

Abstract—The pixel values of images taken by an image sensor are said to be corrupted by Poisson noise. To date, multi-scale Poisson image denoising techniques have processed Haar frame and wavelet coefficients—the modeling of coefficients is enabled by the Skellam distribution analysis. We extend these results by solving for shrinkage operators for Skellam that minimizes the risk functional in the multi-scale Poisson image denoising setting. The minimum risk shrinkage operator of this kind effectively produces denoised wavelet coefficients with minimum attainable $L^2$ error.

Index Terms—Frame transform, Poisson distribution, Skellam distribution, wavelet transform.

I. INTRODUCTION

The ubiquity of digital imaging in consumer, scientific, and medical applications place ever greater demands on engineers to treat photons accumulated and counted by pixel sensors. Pixel sensor is a type of integrating detector—the distribution of measurement is well modeled as Poisson [1]. Fueled by the demand for shrinking pixel sensor size and increasing image resolution, the denoising of Poisson corrupted images has been a subject of intense study [2]–[25]. Specifically, the signal-to-noise ratio scales linearly with the Poisson intensity, meaning low-light regimes pose difficult challenges. Moreover, the normal approximation to Poisson breaks down in low-photon scenarios, rendering image denoising algorithms designed for additive white Gaussian noise (AWGN) useless.

A popular and effective approach to image denoising is multi-scale processing, where a) the noisy pixels $g$ are transformed into wavelet coefficients $\eta$; b) noise-free wavelet coefficients $x$ are estimated, typically by the way of shrinkage operator $\lambda(\eta)$; and c) inverse transform is applied to yield the estimate of the light intensity $f$. In the AWGN domain, an influential work by Donoho et al. [26] produced the Stein’s unbiased estimate of risk (SURE). This functional $\gamma[\lambda]$ predicts the $L^2$ error of a given wavelet estimator $\lambda(Y)$:

$$\gamma[\lambda] = \mathbb{E}[\|\lambda(Y) - x\|^2],$$

where upper case letters (e.g. $Y$) denote the use of a random variable. This can be used to search for parameters of $\lambda$ (e.g. threshold value) that minimize this risk.

Among Poisson image denoising methods, the two paradigms of multi-scale processing have emerged. First, the seminal work by Kolaczyk [2] and Timmermann et al. [3] interpret Haar frame coefficients as noisy partitioning of the “parent coefficient” into two “child coefficients” that are distributed as binomial. The latent success probability parameter—or the partitioning of the noise-free parent coefficient (which is referred to as multiplicative multi-scale innovation or MMI)—is estimated under the Bayesian setting. Second, the noise distribution of Haar wavelet coefficients is Skellam [4]. Latest developments in this area include Bayesian Skellam mean estimators [4] and unbiased estimate of risk [4], [5] aimed at the recovery of noise-free wavelet coefficients.

Alternatives to the Haar frame- and wavelet-based methods include the classical variance stabilization techniques. Dating back to Bartlett and Anscombe [6]–[10], [22], with [11], [12] providing more recent treatments, variance stabilization transforms yield an approach to Poisson mean estimation designed to attain homoscedasticity of noise. Here one seeks an invertible operator $\eta(\cdot)$, typically by way of a compressive nonlinearity such as the component-wise square root, that (approximately) maps the heteroscedastic realizations of an inhomogeneous Poisson process to the familiar additive white Gaussian setting:

$$g \sim \mathcal{P}(f) \quad \mapsto \quad \eta(g) \sim \mathcal{N}(\eta(f), 1).$$

Standard AWGN denoising techniques may then be used to estimate $\eta(f)$ directly (e.g. [27], [28]), with the inverse transform $\eta^{-1}(\cdot)$ applied post hoc.

Inhomogeneous Poisson data can also be treated directly. For instance, empirical Bayes approaches leverage the independence of Poisson variates via their empirical marginal distributions [13], [14], while multiparameter estimators borrow strength to improve upon maximum-likelihood estimation [15]–[17]; however, this ignores potential correlations amongst elements of $f$. Variational methods offer a generic approach for exploiting the correlations [18]–[21]. The relative merits of the various methods described above are well documented [10], [22]–[25] and will not be repeated here.

The goal of this article is to derive a shrinkage operator $\lambda_{\text{MR}}$ that minimizes the risk functional $\gamma[\lambda]$ in the multi-scale Poisson image denoising setting:

$$\forall \lambda \in L^2(\mathbb{R}), \quad \gamma[\lambda_{\text{MR}}] \leq \gamma[\lambda].$$

The minimum risk shrinkage operator (MRSO) of this kind effectively produces denoised frame/wavelet coefficients with minimum attainable $L^2$ error. The unbiased estimator of risk for Skellam mean estimation is already known (reiterated in Theorem II.1) [4], [5]. MRSO for Skellam is derived in Propositions III.1 below via the Euler-Lagrange equation.
Though it is known that MRSO for the natural/canonical parameter of exponential family “noise” is a subtraction by the score function [29], analogous MRSO is previously unknown for the Skellam setting. Following Raphan’s work in [29], the MRSO approach is proved to be nonparametric Bayes optimal (Proposition III.2).

II. BACKGROUND

A. Review: Calculus of Variations

Suppose for the moment that $\gamma[\lambda]$ is the “cost functional” associated with a function $\lambda(\cdot)$ (such as the $L^2$ risk that is the subject of this paper). Obviously, we are interested in minimizing this cost by the choice of $\lambda$. Previous works in this area have optimized parametric functions $\lambda_\pi$ by sweeping the value(s) of $\pi$ over the parameter set II [4], [5], [30]:

$$\hat{\pi} = \arg\min_{\pi \in \Pi} \gamma[\lambda_\pi]. \quad (3)$$

An alternative is to solve for a function $\lambda_{\text{MR}}$ that minimize $\gamma[\lambda]$ directly—the MRSO approach taken in this paper.

More specifically, let $\lambda : \mathbb{R} \rightarrow \mathbb{R}$ be a differentiable function, and consider a functional of the form $\gamma[\lambda] = \int u(y, \lambda, \lambda') dy$ where with $\lambda' = \frac{d\lambda}{dy}$. Then local minimum $\lambda_{\text{MR}}$ is attained if for an arbitrary function $\eta(y)$ and a sufficiently small $\epsilon$, we have

$$\gamma[\lambda_{\text{MR}}] \leq \gamma[\lambda_{\text{MR}} + \epsilon \eta(y)]. \quad (4)$$

$\lambda_{\text{MR}}$ is found by solving for $\lambda$ in the Euler-Lagrange equation:

$$\frac{\partial u}{\partial \lambda} - \frac{\partial}{\partial y} \frac{\partial u}{\partial \lambda'} = 0. \quad (5)$$

B. Review: Haar Wavelet Transform and Skellam Distribution

The wavelet transform is a type of multi-resolution analysis operator on $L^2(\mathbb{R})$. Owing to their compact support and vanishing moments, wavelets transform characterize local regularity of a signal efficiently [31], [32]. Denote $g = (g_0, \ldots, g_{N-1})$ the observed number of photons at pixel location $i$, where $g_i \sim \mathcal{P}(f_i)$ and $f = (f_0, \ldots, f_{N-1})$ is the latent light intensities. Let $t^0 = (t^0_0, \ldots, t^0_{N-1})$ where $t^0_i = g_i$. A $J$-level discrete Haar wavelet transform $g_i \mapsto (t^j_i, y^j_i, y^{j-1}_i, \ldots, y^1_i)$ where the scaling coefficients $t^j_i = (t^0_i, \ldots, t^j_{N/2^j-1})$ and the wavelet coefficients $y^j_i = (y^j_{0,1}, \ldots, y^j_{N/2^j-1})$ are comprised of sums and differences:

$$\begin{cases}
\text{scaling coefficients} & t^j_i = t^{j-1}_{2i} + t^{j-1}_{2i+1} \\
\text{wavelet coefficients} & y^j_i = t^{j-1}_{2i} - t^{j-1}_{2i+1}.
\end{cases} \quad (6)$$

Similarly, let $f_i \mapsto (s^j_i, x^j_i, x^{j-1}_i, \ldots, x^1_i)$ denote Haar wavelet transform of the noise-free image $f_i$. The focus of this paper is on the estimation of $x^j_i$ based on $t^j_i$ and $y^j_i$. The estimate of noise-free image $f_i$ is then recovered by taking the inverse transformation of $x^j$ and $s^j$.

A key observation here is that at every level, $t^j_i$ is a Poisson variable: $t^j_i \sim \mathcal{P}(s^j_i)$. The wavelet coefficient $y^j_i$ is hence a difference of two Poisson variables $t^{j-1}_{2i}$ and $t^{j-1}_{2i+1}$. Then the distribution of $y^j_i$ is said to be Skellam, defined by the relation [4]:

$$P[Y^j_i = y | X^j_i = x, S^j_i = s] = e^{-x} \left( \frac{s + x}{s - x} \right)^{\frac{y}{2}} I_y(\sqrt{s^2 - x^2}), \quad (7)$$

where $I_y(\cdot)$ is the modified Bessel function of the first kind. In the special case that $X^j_i = 0$, the distribution of $Y^j_i$ is known as Irwin distribution, taking the form:

$$P[Y^j_i = y | X^j_i = 0, S^j_i = s] = e^{-s} I_y(s). \quad (8)$$

The level index $j$ the subsequent discussion are omitted—instead we simply define the clean and noisy scaling coefficients $s = f_0 + f_1$ and $t = g_0 + g_1$, respectively, where Poisson counts $g_i \sim \mathcal{P}(f_i)$. Furthermore, let $x = f_0 - f_1$ and $y = g_0 - g_1$ denote clean and noisy wavelet coefficients, respectively. Thanks to the recursive relation of Equation (6), the results obtained for $t, y$ generalize to every level of Haar wavelet decomposition.

From the perspective of (7), the goal is to estimate the latent noise-free wavelet coefficient (Skellam mean) $x$ based on the “Skellam noise corrupted” variable $y$ and “Poisson noise corrupted” variable $t$; $f_i$ is the recovered signal via the inverse wavelet transform. Let $\lambda : \mathbb{Z}^2 \rightarrow \mathbb{R}$ and redefine the risk functional $\gamma[\lambda]$ in (1) as

$$\gamma[\lambda] = E\|\lambda(Y, T) - X\|^2. \quad (9)$$

In addition, denote by $\gamma[\lambda|x, s]$ the conditional risk

$$\gamma[\lambda|x, s] = E[\|\lambda(Y, T) - X\|^2 | X = x, S = s]. \quad (10)$$

The risk functionals of (9) and (10) do not depend on $\lambda'$, thus by (5) we are only concerned with $\frac{\partial u}{\partial \lambda} = 0$. The unbiased estimate of risk for Skellam mean estimation was first reported by Hirakawa et al. [4], [33]:

**Theorem II.1.** Define $t = g_0 + g_1$ and $y = g_0 - g_1$ where $g_i \sim \mathcal{P}(f_i)$, and let $x = f_0 - f_1$. Then

$$\hat{\gamma}[\lambda|x, s] := E[\|\lambda(Y, T)\|^2 + ||T||_1 + 2Y \lambda(Y, T)
\begin{align*}
&- (T + Y)\lambda(Y - 1, T - 1) \\
&+ (T - Y)\lambda(Y + 1, T - 1)|X = x, S = s\]
\end{align*} \quad (11)$$

is an unbiased estimate of the conditional risk functional $\gamma[\lambda|x, s]$, where $\lambda(y, t) = y + \theta(y, t)$.

The significance of Theorem II.1 is that the $L^2$ risk of a given wavelet shrinkage operator $\lambda(y, t)$ of $x$ is knowable without the explicit knowledge of the latent variables $x$ and $s$ (or their distributions). A well known application of Theorem II.1 is PURELET—$\lambda_B(y, s)$ takes a parametric form, whose parameter(s) $\pi$ is the minimizer of $\hat{\gamma}[\cdot|x, s]$ in (11) and (3) [5].

Other previous investigations have led to Bayes point estimates of Skellam mean. For instance, the shrinkage operator of the form [34]

$$\lambda_B(y, s) = y - s E\left[\frac{\partial}{\partial x} \ln p(Y = y | X, S = s) | Y = y, S = s\right] \quad (12)$$
has an error bound to the optimal minimum mean square error (MMSE) estimator \( \lambda_{\text{MMSE}}(y, s) = \mathbb{E}[X|Y = y, S = s] \) as
\[
\| \lambda_{\text{MMSE}}(y, s) - \lambda(y, s) \|_2 \leq \mathbb{E}[X^2|Y = y, S = s] \times \left[ \mathbb{E} \left[ \frac{\partial^2}{\partial y^2} \ln p(Y = y|X, S = s) \right]^2 | Y = y, S = s \right].
\]  
(13)

As is typical of Bayes estimation, (12) requires a proper elicitation of prior density. In addition, since clean scaling coefficients \( s \) are unavailable in practice, they must be replaced by noisy scaling coefficient \( t \) in implementation. An improvement to [34] is the Skellam empirical Bayes estimator [35]:
\[
\lambda_{\text{EB}}(y, s) = y - (s - 1) \times \frac{P[Y = y - 1, S = s] - P[Y = y + 1, S = s]}{2P[Y = y, S = s]},
\]  
(14)

Although (14) also approximates \( \lambda_{\text{MMSE}} \) without the prior density, we still lack the access to noise-free scaling coefficients \( s \) in practice.

### III. SKELLAM MINIMUM RISK SHRINKAGE OPERATOR

#### A. Main Results

In this section, we provide the main results of this paper. The proofs of the propositions are provided in Section VII.

By bivariate Skellam minimum risk shrinkage operator (BMRSO) we mean the function of the form \( \lambda_{\text{BMR}}(y, t) \) that minimizes the risk functional in (10). That is,
\[
\forall \lambda \in L^2(\mathbb{R}), \quad \gamma[\lambda_{\text{BMR}}] \leq \gamma[\lambda]
\]  
(15)

where \( \lambda(y, t) \) is any alternative function to \( \lambda_{\text{BMR}}(y, t) \). Thanks to Theorem II.1, one can also interpret \( \lambda_{\text{BMR}}(y, t) \) as the minimizer of the quantity in (11). By the calculus of variations, the local minimum estimator is attainable by solving the Euler-Lagrange equation.

#### Proposition III.1

When \( \phi(y, t) \) in Theorem II.1 is differentiable for all values of \( y \) and \( t \), then the minimizer \( \lambda_{\text{BMR}} \) of \( \gamma[\cdot] \) is:
\[
\lambda_{\text{BMR}}(y, t) = y - (s - 1) \times \frac{\phi(y + 1, t + 1) + \phi(y - 1, t + 1)}{2\phi(y, t)} + (t + 2) \frac{\phi(y + 1, t + 1) - \phi(y - 1, t + 1)}{2\phi(y, t)},
\]  
(16)

where \( \phi(y, t) = P[Y = y, T = t] \) denotes the joint marginal probability mass function of the Haar wavelet coefficient \( Y \) and the scaling coefficient \( T \).

Skellam BMRSO proposed in Propositions III.1 is also a nonparametric empirical Bayes MMSE estimator.

#### Proposition III.2

The nonparametric empirical Bayes MMSE estimator of Skellam mean is:
\[
\mathbb{E}[X|Y = y, T = t] = y \frac{\phi(y + 1, t + 1) + \phi(y - 1, t + 1)}{2\phi(y, t)} + (t + 2) \frac{\phi(y + 1, t + 1) - \phi(y - 1, t + 1)}{2\phi(y, t)},
\]  
(17)

where \( \phi(y, t) = P[Y = y, T = t] \) denotes the joint marginal probability mass function of the Haar wavelet coefficient \( Y \) and the scaling coefficient \( T \).

The proof (provided in Section VII) relies on the seminal work of [29]. One can understand (17) as an MMSE estimator whose prior density is implied by the marginal \( \phi(y, t) \).

Finally, the univariate Skellam minimum risk shrinkage operator \( \lambda_{\text{UMR}}(y) \) (UMRSO) we propose below is an alternative to Skellam BMRSO that reduces the complexity further. It makes use of two one dimensional functions—\( \tau(y) := \mathbb{E}[T|Y = y] \) and \( \psi(y) := P[Y = y] \).

#### Proposition III.3

The univariate nonparametric empirical Bayes MMSE estimator \( \lambda_{\text{UMR}}(y) := \mathbb{E}[X|Y = y] \) of Skellam mean is:
\[
\lambda_{\text{UMR}}(y) = \frac{y + 1}{2} \frac{\tau(y + 1)}{\psi(y)} + \frac{y - 1}{2} \frac{\tau(y - 1)}{\psi(y)}.
\]  
(18)

#### B. Significance and Practical Implementation

The Skellam BMRSO \( \lambda_{\text{BMR}} \) is a function of wavelet \( y \) and scaling \( t \) coefficients, as well as their joint marginal probability mass function \( \phi(y, t) \). The dependence on \( \phi(y, t) \) implies signal adaptivity, which is desirable for ensuring that Skellam BMRSO is effective for a diverse range of imaging applications. Owing to the fact that \( \lambda_{\text{BMR}} \) minimizes \( L^2 \) risk \( \gamma[\cdot] \), there is no bivariate estimation (e.g. Skellam mean estimation based on wavelet \( y \) and scaling \( t \) coefficients) that performs better (though it is still possible that non-bivariate forms of estimation can improve on BMRSO).

Skellam BMRSO and UMRSO are also Bayes optimal. This is significant since conventional Bayes estimation requires a proper elicitation of prior density of the latent variable \( x \), which remains an open area of research. The prior in BMRSO is implied by the marginal density \( \phi(y, t) \) instead, meaning the risk of model mismatch (assumed density of \( x \) does not match the scene) is minimal. Moreover, in contrast to the previously proposed estimators (e.g. methods in (12) and (14)) that approximate Bayes optimal estimators (albeit with known error bounds), BMRSO and UMRSO are exact MMSE estimators that do not require replacing the unobservable clean scaling coefficients \( s \) with observable noisy scaling coefficient \( t \) in implementation. Although the differences between (14) and (16) seem subtle, there is a clear difference in denoising performance (see Section IV).

Proposed denoising technique is not without weaknesses, however. The Skellam BMRSO \( \lambda_{\text{BMR}} \) is computable if a stable \( \phi(y, t) = P[Y = y, T = t] \) can be obtained (or approximated) from the empirical distribution of \( y \) and \( t \). Computing multivariate histogram is challenging due to the instability stemming from lack of sufficient data to robustly populate the histogram bins. To make the matters worse, \( \phi(y, t) \) is less stable for large \( y \) and \( t \) values because the distributions of \( y \) and \( t \) are heavy-tailed for natural images. As such, advantages to Skellam BMRSO over other Poisson
image denoising techniques are more pronounced when image resolution is higher (more pixels to stabilize empirical histogram) or when the photon count is low (smaller ranges of \( y \) and \( t \)).

For multivariate case, a conventional approach is to employ kernel density estimate (KDE) to smooth the empirical distribution of the observed values to stabilize the estimation. Surprisingly, denoising with joint empirical histogram \( \phi_E(y,t) \) of \( Y, T \) as approximation of \( \phi(y,t) \) outperforms KDE approximation \( \phi_K(y,t) \) of \( \phi(y,t) \). To understand the reasons behind this counterintuitive result, consider rewriting \( \lambda_{BMR}(y,t) \) in (16) as

\[
\lambda_{BMR}(y,t) = yA(y,t) + (t+2)B(y,t)
\]

where the attenuation \( A(y,t) \) and shrinkage \( B(y,t) \) factors are defined as

\[
A(y,t) = \frac{\phi(y+1,t+1) + \phi(y-1,t+1)}{2\phi(y,t)} \quad B(y,t) = \frac{\phi(y+1,t+1) - \phi(y-1,t+1)}{2\phi(y,t)}.
\]

The smoothing in KDE is problematic for denoising with (19) because the values of \( \hat{\phi}_K(y+1,t+1), \hat{\phi}_K(y-1,t+1) \) and \( \phi_K(y,t) \) are forced to be similar. It implies that \( \lambda_{BMR}(y,t) \) is approximately a unity and \( B(y,t) \) is close to 0, meaning a denoising function \( \lambda_{BMR}(y,t) \approx y \) passes noise to its output, rendering denoising ineffective (joint empirical histogram \( \phi_E(y,t) \) of \( y \) and \( t \) does not have this problem).

Another favorable property of joint histogram is its robustness to the outliers. If \( (y_0,t_0) \) pair is an outlier, then \( \hat{\phi}_E(y_0,t_0) \) will be nontrivial but \( \hat{\phi}_E(y_0+1,t_0+1) \) are likely to be small or zero. Denoising by (16) is aggressive in this scenario since \( A(y_0,t_0) \) and \( B(y_0,t_0) \) are small. Conversely, \( \hat{\phi}_E(y_0+1,t_0+1) \) will not be small when \( (y_0,t_0) \) is a representative signal, and hence the edge details will be preserved. In addition, \( \hat{\phi}_E(y_0,t_0) \) is nontrivial for all observed \( (y_0,t_0) \) pairs, ensuring that \( \lambda_{BMR}(y_0,t_0) \) is always computable (i.e., no divide by zero).

To illustrate by simulation the penalty paid by BMRSO to employ empirical histogram \( \hat{\phi}_E(y,t) \) in the place of a true marginal density \( \phi(y,t) \), consider Figure 1. The ground truth probability mass function \( \phi(y,t) \) is computable in simulation from the ideal image \( f \) or its wavelet coefficients \( (x,s) \); alternatively, one can obtain \( \phi(y,t) \) via the Monte-Carlo simulation by generating multiple instantiations of pseudorandom Poisson noise from the same source \( (x,s) \) repeatedly. Figure 1 reports the denoising performance on gray scale test images (‘bike,’ ‘car,’ and ‘office’) using Monte-Carlo simulation representing the best possible reconstruction of Proposition III.1. (The full experiment setup is described in Section IV.)

Compared to the Monte-Carlo Skellam BMRSO, the practical Skellam BMRSO using empirical histogram \( \hat{\phi}_E(y,t) \) (obtained from a single instantiation of Poisson noise) is a decent recovery. In general the edge details in the latter reconstruction are preserved to the same degree as the Monte-Carlo version, although the denoised homogeneous regions have a slightly noisier appearance. One notable difference can be seen in the specular highlights of the ‘car’ image—specularity cause strong edges, resulting in unusually large \( y \) values. The loss of contrast in Figure 1(d) (see wheel) is due in part to the coarse level scaling coefficients \( (y,t) \) which cover a larger dynamic range. The empirical histogram \( \hat{\phi}_E(y,t) \) is sparsely populated for large \( t \) and \( y \) ranges and therefore less stable. The \( (y,t) \) pairs corresponding to the specular behavior like outliers and thus they are aggressively attenuated.

The computational complexity of BMRSO and UMRSO is essentially the cost of forward/inverse wavelet transform and the histogram computation. The running times of (16) and (18) are negligible by comparison.

### C. Hybrid MRSO

To overcome the shortcoming of low contrast in Skellam BMRSO (\( \lambda_{BMR} \)), we propose a hybrid Skellam minimum risk shrinkage operator HMRSO (\( \lambda_{HMR} \)) that combines Skellam BMRSO and Skellam UMRSO (\( \lambda_{UMR} \)). Figure 2 shows the system diagram of HMRSO. As discussed, low contrast in \( \lambda_{BMR} \) is due to lack of data to stabilize histogram \( \phi(y,t) \) at coarse level. By comparison, the true advantage of the univariate Bayes denoising is that the empirical functions \( \psi(y) \) and \( \tau(y) \) are one-dimensional functions that are far more stable than the two dimensional function \( \phi(y,t) \). Empirical histogram can be used to estimate \( \psi(y) \) robustly in the usual way. As for \( \tau(y) \), one can take a “conditional sample average” of \( t \) for each \( y \) value that is observed. One disadvantage to the univariate Skellam mean estimation is that the dependence on \( t \) is significantly weakened compared to the bivariate version.

Since \( \lambda_{UMR} \) can estimate a stable one dimensional histogram at coarse level robustly with fewer pixels, the main idea of HMRSO \( \lambda_{HMR} \) is to multiplex \( \lambda_{UMR} \) at coarse levels to restore contrast while fine details are well preserved by \( \lambda_{BMR} \). Autoselection between \( \lambda_{BMR} \) and \( \lambda_{UMR} \) at each transform level relies on dynamic range (amount of data) is well represented by average scaling coefficient value.

\[
\lambda_{HMR} = \begin{cases} 
\lambda_{UMR} & \text{if } E[T_i^2] \approx \frac{2}{\tau} \sum t_i^2 > \text{threshold} \\
\lambda_{BMR} & \text{otherwise}
\end{cases}
\]

To train the threshold above, we evaluated MSE scores of BMRSO and UMRSO methods at each transform level on the McGill database [36] and apply Bayes classifier for an optimal threshold. See Figure 3. As reported in Figure 1(e), Skellam HMRSO restoration recovers the contrast while maintaining the same level of detail as in Skellam BMRSO. The overall performance of the Skellam HMRSO is competitive with the ideal Skellam BMRSO implementation.

### D. Sparse Image Model

It is by now well accepted that the majority of the wavelet coefficients \( X \) are extremely small in magnitude or zero. We introduce a technique to further improve the image denoising quality of MRSO by leveraging this notion of sparsity. By total
Fig. 1: Zoomed portion of simulation results with “bike” (top row, 2048×2560 pixels) “car” (middle row, 1280×1600 pixels) and “office” (bottom row 1280×1600 pixels). Showing denoising results with 0.25 average pixel intensity with “(PSNR / SSIM)” average scores: (a) original (∞/1), (b) input (6.0893/0.0608), (c) BMRSO with Monte-Carlo (19.9818/0.6032), (d) BMRSO (19.3984/0.5742), (e) HMRSO (19.7367/0.5815). BMRSO preserves image details to the same extent as the Monte-Carlo, but with a loss of contrast. HMRSO restores the contrast by borrowing from UMRSO.

Fig. 2: System diagram for hybrid MRSO denoising.
Fig. 3: MSE performance curve of Skellam UMRSO ($\lambda_{UMR}$) and Skellam BMRSO ($\lambda_{BMR}$) as a function of the scaling coefficient intensity, evaluated on McGill database [36] (700 test images) with an average photon count of 0.25. UMRSO performs better than BMRSO when the scaling coefficients are large.

Fig. 4: Impact of the sparsity mixture parameter $\alpha$. (a) Input, (b)~(f) HMRSO restoration with different $\alpha$ values: (b) $\alpha = 0$, (c) $\alpha = 0.2$, (d) $\alpha = 0.4$, (e) $\alpha = 0.8$, (f) $\alpha = 1$. First row: ‘fish’ patch; second row: ‘arch’ patch. The parameter $\alpha$ has little effect on the image detail recovery (e.g. edge in arch), but noise in smooth regions is greatly suppressed with small $\alpha$ value.

probability theorem, the marginal density $\phi(y,t)$ take the form

$$\phi(y,t) = P[Y = y, T = t]$$
$$= P[Y = y, T = t | X \neq 0] P[X \neq 0]$$
$$+ P[Y = y, T = t | X = 0] P[X = 0],$$

(22)

where $\phi_0(y,t) := P[Y = y, T = t | X = 0]$ corresponds to the probability mass function corresponding to the sparse component. Expanding $\phi_0(y)$ by the total probability:

$$\phi_0(y,t) = \int P[Y = y, T = t | X = 0, S = 2f]$$
$$\times P[F_0 = f | X = 0] df$$

(23)

Here, $P[Y = y, T = t | X = 0, S = 2f]$ is a computable quantity. For $P[F_0 = f | X = 0]$, we have the approximation

$$P[F_0 = f | X = 0] = P[F_0 = f | F_0 = F_1] \approx P[F_0 = f]$$

(24)

thanks to the notion of sparsity ($X$ small implies $F_0 \approx F_1$). The probability density function $P[F_0 = f]$ can be recovered from the empirical histogram $P[G_0 = g]$ using the so-called Poisson transform [37]. Hence $\phi_0(y,t)$ is approximable with high degrees of accuracy.

Sparsity also implies that $P[X = 0]$ is large, meaning $\phi(y,t) \approx \phi_0(y,t)$. By this fact, $\phi_0(y,t)$ makes an excellent regularization term that promote stability of $\phi(y,t)$, as follows:

$$\hat{\phi}(y,t) = \alpha \hat{\phi}_E(y,t) + (1 - \alpha) \phi_0(y,t),$$

(25)

where $\alpha$ is a proxy for $P[X = 0]$ that we treat as a parameter; and $\hat{\phi}_E(y,t)$ is the empirical histogram of $Y,T$ playing the role of $P[Y = y, T = t | X \neq 0]$. The regularization in (25) has a hierarchical Bayesian interpretation as well. By the total
in excited electrons that are particles in nature (as evidenced technically Poisson, stemming from the random perturbations noise is Gaussian [12], the distribution of the thermal noise is often the predominating source. Despite many factors contribute to this largely signal-independent the second setup, we simulate the background noise. Though

\[ \eta \]

\[ \psi \]

\[ \text{combines the empirical histogram } \hat{\psi}_E(y) \text{ with the computable density } \psi_0(y) := P[Y = y|X = 0] = \sum_{i=0}^{\infty} \phi_0(y, i) \text{ to promote sparse image recovery in UMRSO. The mixture parameter } \alpha \text{ could be trained on image database or chosen by quality assessment metrics [38], [39].} \]

IV. EXPERIMENTAL RESULTS

A. Setup

To verify the effectiveness of the MRSO denoising approaches, we conduct simulation and real sensor data experiments. The simulation study is performed on McGill Calibrated Color Image Database [36] comprised of 700 test images calibrated to be linear with respect to the incoming light intensities. Captured under ample light with little noticeable noise, these images are taken to be the “ground truth” or “ideal” image data \( f \). Pseudo-random Poisson noise is introduced to the images in two ways. In the first setup, we generate the Poisson image by simply scaling the ideal intensity image \( f \) by factor \( \eta \):

\[ g_i \sim \mathcal{P}(\eta f_i). \]  

(28)

The scale factor \( \eta \) is used to control the average intensity of an image (small \( \eta \) is indicative of the low light conditions). In the second setup, we simulate the background noise. Though many factors contribute to this largely signal-independent noise, thermal noise is often the predominating source. Despite some reports claiming that the distribution of the thermal noise is Gaussian [12], the distribution of the thermal noise is technically Poisson, stemming from the random perturbations in excited electrons that are particles in nature (as evidenced also by the noise variance that scales linearly with the voltage

across op-amps). Under this scenario, the observed Poisson image data is generated as follows:

\[ g_i \sim \mathcal{P}(\eta f_i + \theta). \]  

(29)

where the constant offset \( \theta \) is responsible for the background noise. In our experiments, \( \theta \) was set to twice the average pixel intensity value of \( \mathbb{E}f_i \), consistent with the “aquarium” raw image sensor data we describe below. Note that the average pixel intensity considered for the simulated data is far lower than the real sensor data images. The ranges of \( \mathbb{E}f_i < 2 \) may seem severe to some, but this level of noise is not unreasonable for scientific sensors such as the ones used in astronomy, fluorescent microscopy and hyper-spectral imaging. Peak signal-to-noise ratio (PSNR) and structural similarity index metric (SSIM) [40] are used to benchmark denoising performance.

The test images “aquarium” and “Gretag Macbeth Colorchecker” images shown in Figures 7(a)-(b) were taken by a FUJIFILM X-PRO 1 camera moments apart in a raw sensor mode with all manual settings in the same environment. One additional “blank” image was captured with a lens cap closed (not shown) to assess the level of the background noise with no light. Demosaicing was bypassed by subsampling green in the square \((3 \times 3)\) sampling lattice, yielding an effective resolution of \(1096 \times 1648\).

The raw sensor output value is assumed to be an affine transformed version of the Poisson count data \( g_i \) (as justified by Figure 7(c) and [1]). The affine transformation is parametrized by two numbers, which was learned by linearly regressing on the sample mean and variance of pixel values in each square of the Gretag Macbeth Colorchecker (see Figure 7(c)). Using the learned affine transformation, the raw sensor values of the “aquarium” and “blank” images were converted to Poisson count data by reversing the affine transform. The average Poisson count values of the “aquarium” and “blank” images were \(18.0411\) and \(13.1370\), respectively. Taking the latter to approximate the offset value \( \theta \) in (29), we conclude \( \mathbb{E}f_i \approx 4.9041 \) in Figure 7(a). Denoising was performed on the Poisson count data from the “aquarium” image.

We compared the proposed Skellam HMRSO to existing Poisson denoising methods—a combination of generalized Anscomb transform and the block-matching and 3D filtering (GAT+BM3D) in [27]; the Poisson unbiased risk estimate with linear expansion of threshold (PURE-LET) in [5]; the Skellam empirical Bayes estimator in [34] and [35]; MM estimation by Kolaczyk in [2] and by Timmermann in [3]; and the variational method in [19].

B. Synthetic Noisy Image Results

Tables I and II report PSNR/SSIM scores of restoration by HMRSO as well as the competing methods. With the noise model of (28), Skellam HMRSO provides optimal restoration at low photon levels. Although HMRSO is identical to BMRSO when the average photon level is low (as evidenced by the PSNR/SSIM scores), the improvements of the UMRSO-BMRSO hybridization are noticeable for higher average pixel intensity images. The performance of PURE-LET improves
also when photon counts increase, while GAT+BM3D needs a higher level of photons to be competitive. With the noise model of (29), the UMRSO plays a role in HMRSO at every pixel intensity level. Though the denoising performance of HMRSO is superior to PURE-LET and GAT+BM3D at low photon levels, method in [3] is surprisingly competitive also.

Figure 5 shows examples of denoising methods applied to Poisson corrupted images. Among the methods shown, GAT+BM3D retained the highest level of image contrast, but its output images have the noisiest appearance (see Figure 5(e)). Image details are ambiguous and are often indistinguishable from the artifacts that appear in the homogeneous region. Skellam HMRMRO and PURE-LET had similar appearance, but the image details in Skellam HMRMRO were superior (e.g. the resolution chart in Figure 5(c)) while PURE-LET retained a better contrast (e.g. specular highlight on the wheel in Figure 5(d)). With background noise incorporated into the observation. Figure 6(b) shows noisier images. Skellam HMRMRO still achieves smoothing in the homogeneous regions of the image while many of the edge details remain intact, though suffering from further contrast loss. By comparison, PURE-LET and GAT+BM3D retain a higher contrast image. However, PURE-LET suffers from pixilation/blocking artifacts that are common in Haar-based image denoising methods; and the edges in GAT+BM3D output are highly distorted.

C. Real Sensor Data Results

Figure 8 shows zoomed portions of the noisy and denoised “aquarium” image in Figure 7(a). In the picture of the arch, the edge details in GAT+BM3D appeared to be the best at first glance, though a closer examination reveals that the corners of objects are smoothed/rounded unnaturally to give it a waxy appearance. By contrast, Skellam HMRMRO, Timmermann and PURE-LET have less pronounced edges, but the edge details were preserved more accurately. Skellam HMRMRO preserves smaller image details than PURE-LET (see the widened widths of the vertical lines on the left arch’s columns). In the picture of the fish, noise in the homogeneous region of the image was best suppressed by Skellam HMRMRO, though Timmermann was close in appearance. Other methods—including PURE-LET and GAT+BM3D—introduced false textures to Poisson corrupted images. Among the methods shown, method in [3] is surprisingly competitive also. We thank Drs. Patrick Wolfe and Marianna Penske for the suggestion to incorporate Euler-Lagrange equation. We also thank authors of [2], [3], [5], [19], [27] for providing denoising code for comparison. This work was supported in part by the National Science Foundation under Grant No. 1422104 and the University of Dayton Graduate Student Summer Fellowship.

VI. ACKNOWLEDGEMENT

We thank Drs. Patrick Wolfe and Marianna Penske for the suggestion to incorporate Euler-Lagrange equation. We also thank authors of [2], [3], [5], [19], [27] for providing denoising code for comparison. This work was supported in part by the National Science Foundation under Grant No. 1422104 and the University of Dayton Graduate Student Summer Fellowship.

VII. APPENDIX: PROOFS

A. Proof of Proposition III.1

Proof. By Theorem II.1 and total expectation, we have

$$\gamma(\lambda) = E[\gamma(\lambda|x,s)] = E[\gamma(\lambda|x,s)].$$  (30)

Denote by $\phi(y,t|x,s)$ the marginal density $\phi(y,t|x,s) = P[Y = y, T = t|X = x, S = s]$. By (5) where $\frac{\beta_0}{y}$ is zero, we first solve

$$0 = \frac{\partial}{\partial \theta(y_0, t_0)} \gamma(\lambda|x,s) \bigg|_{(y_0, t_0) = (y, t)}$$  (31)

where $\lambda(y,t) = y + \theta(y,t)$ as before. Then,

$$0 = \sum_y \sum_t \phi(y,t|x,s) \frac{\partial}{\partial \theta(y_0, t_0)} \left[ \|\theta(y,t)\|^2_1 + \|t\|_1 
+ 2\theta(y,t) - (t + y)\theta(y - 1, t - 1)
+ (t - y)\theta(y + 1, t - 1) \right] \bigg|_{(y_0, t_0) = (y, t)}$$  (32)

$$= 2\phi(y,t|x,s)\theta(y,t) + 2\phi(y,t|x,s)y
- \phi(y + 1, t + 1|x,s) \left( (t + 1) + (y + 1) \right)
+ \phi(y - 1, t + 1|x,s) \left( (t + 1) - (y - 1) \right).$$

Evaluating this with the total probability theorem

$$\phi(y,t) = \int \int \phi(y,t|x,s)P[X = x, S = s]dxds,$$  (33)

it instead yields

$$0 = \frac{\partial}{\partial \theta(y_0, t_0)} \gamma(\lambda) \bigg|_{(y_0, t_0) = (y, t)}$$  (34)

Solving for $\theta(y_0, t_0)$ and setting $\lambda(y,t) = y + \theta(y,t)$ proves the Proposition.

B. Proof of Proposition III.2

Proof. A general form of Bayes MMSE estimator derived by Raphan and Simoncelli in [29] has the following:

$$E[X|Y = y, T = t] = \frac{L\{\phi(\cdot,y)\}(y,t)}{\phi(y,t)}$$  (35)

where the operator $L\{\cdot\}(\cdot)$ is the solution to the relation

$$x \cdot \phi(y,t|x,s) = L\{\phi(\cdot,|x,s)\}(y,t).$$  (36)

For Skellam, the solution to (36) can be found by rewriting:

$$x \cdot \phi(y,t|x,s) = \frac{s + x}{2} \phi(y,t|x,s) - \frac{s - x}{2} \phi(y,t|x,s).$$  (37)
Fig. 5: Zoomed portion of simulation results with “bike” (top row), “cafe” (middle row), and “car” (bottom row) images. Noise model of (28) was used to generate a pseudo-random Poisson corruption with average pixel intensity of 0.25. (a) original, (b) noisy, (c) HMRSO (proposed), (d) PURELET [5], (e) BM3D [27].

Fig. 6: Zoomed portion of simulation results with “bike” (top row), “cafe” (middle row), and “car” (bottom row) images. Noise model of (29) was used to generate a pseudo-random Poisson corruption with average pixel intensity of 0.25 and offset of 0.50. (a) original, (b) noisy, (c) HMRSO (proposed), (d) PURELET [5], (e) BM3D [27].
TABLE I: Reconstruction PSNR / SSIM scores for McGill Calibrated Colour Image Database [36], averaged over 700 images. Noisy input images simulated with model in (28).

<table>
<thead>
<tr>
<th>Average Pixel Intensity</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
<th>0.40</th>
<th>0.80</th>
<th>1.60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noisy</td>
<td>6.4195/0.0328</td>
<td>6.4795/0.0502</td>
<td>6.6355/0.0766</td>
<td>7.4469/0.1130</td>
<td>9.5369/0.1762</td>
<td>11.9117/0.2610</td>
</tr>
<tr>
<td>BMRSO (proposed)</td>
<td>18.1623/0.4449</td>
<td>18.9421/0.4897</td>
<td>19.6025/0.5328</td>
<td>20.0500/0.5729</td>
<td>20.2216/0.6041</td>
<td>20.1613/0.6256</td>
</tr>
<tr>
<td>BMRSO (proposed)</td>
<td>7.9701/0.0585</td>
<td>8.1745/0.0501</td>
<td>9.0848/0.0535</td>
<td>12.5870/0.1129</td>
<td>17.3825/0.2949</td>
<td>20.5117/0.4975</td>
</tr>
<tr>
<td>HMR20 (proposed)</td>
<td>18.1623/0.4449</td>
<td>18.9421/0.4897</td>
<td>19.6025/0.5328</td>
<td>20.0500/0.5729</td>
<td>20.2216/0.6041</td>
<td>20.1613/0.6256</td>
</tr>
<tr>
<td>PURE-LET</td>
<td>17.9663/0.4272</td>
<td>18.9148/0.4845</td>
<td>19.8789/0.5249</td>
<td>20.8435/0.6034</td>
<td>21.3717/0.6634</td>
<td>23.0253/0.7227</td>
</tr>
<tr>
<td>GAT+BM3D</td>
<td>14.9050/0.2939</td>
<td>15.7525/0.4049</td>
<td>19.1307/0.4943</td>
<td>20.5525/0.5808</td>
<td>22.0717/0.6617</td>
<td>23.5118/0.7320</td>
</tr>
<tr>
<td>Bayes Empirical</td>
<td>13.6430/0.2157</td>
<td>13.5481/0.2180</td>
<td>14.4727/0.2625</td>
<td>16.7871/0.3369</td>
<td>19.5400/0.4851</td>
<td>22.0136/0.6696</td>
</tr>
<tr>
<td>Kolaczyk</td>
<td>16.0747/0.3235</td>
<td>16.3503/0.3475</td>
<td>16.3749/0.3702</td>
<td>16.3802/0.3936</td>
<td>16.8690/0.4276</td>
<td>18.3382/0.4876</td>
</tr>
<tr>
<td>Timmermann</td>
<td>17.7553/0.4129</td>
<td>18.4084/0.4401</td>
<td>19.1434/0.4748</td>
<td>19.9253/0.5170</td>
<td>20.8142/0.5685</td>
<td>21.8436/0.6281</td>
</tr>
<tr>
<td>Total Variation</td>
<td>13.9103/0.2787</td>
<td>15.9912/0.3801</td>
<td>17.5511/0.4462</td>
<td>19.1362/0.5102</td>
<td>19.9382/0.5391</td>
<td>20.0334/0.5364</td>
</tr>
</tbody>
</table>

TABLE II: Reconstruction PSNR / SSIM scores for McGill Calibrated Colour Image Database [36], averaged over 700 images. Noisy input images simulated with model in (29).

<table>
<thead>
<tr>
<th>Average Pixel Intensity</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
<th>0.40</th>
<th>0.80</th>
<th>1.60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noisy</td>
<td>5.8840/0.2042</td>
<td>5.5991/0.0305</td>
<td>5.6603/0.0476</td>
<td>6.4370/0.0925</td>
<td>7.8614/0.0874</td>
<td>9.1100/0.1296</td>
</tr>
<tr>
<td>BMRSO (proposed)</td>
<td>16.4097/0.3270</td>
<td>17.2971/0.3633</td>
<td>17.9288/0.3924</td>
<td>18.3326/0.4150</td>
<td>18.5266/0.4353</td>
<td>18.5190/0.4495</td>
</tr>
<tr>
<td>BMRSO (proposed)</td>
<td>7.0976/0.0045</td>
<td>6.9661/0.0070</td>
<td>6.4599/0.0861</td>
<td>16.7794/0.2596</td>
<td>18.8800/0.3957</td>
<td>20.6645/0.5328</td>
</tr>
<tr>
<td>HMR20 (proposed)</td>
<td>16.6202/0.3362</td>
<td>17.5372/0.3728</td>
<td>18.2570/0.4130</td>
<td>18.7823/0.4387</td>
<td>19.7788/0.4862</td>
<td>20.3953/0.5176</td>
</tr>
<tr>
<td>PURE-LET</td>
<td>16.3527/0.3231</td>
<td>17.4317/0.3979</td>
<td>18.3758/0.4352</td>
<td>19.3124/0.4912</td>
<td>20.2910/0.5523</td>
<td>21.3126/0.6145</td>
</tr>
<tr>
<td>GAT+BM3D</td>
<td>13.8141/0.2235</td>
<td>15.2949/0.3991</td>
<td>17.0811/0.3882</td>
<td>18.9845/0.4768</td>
<td>20.2690/0.5460</td>
<td>21.3962/0.6085</td>
</tr>
<tr>
<td>Bayes Empirical</td>
<td>9.2369/0.0475</td>
<td>9.1137/0.0355</td>
<td>10.2711/0.0340</td>
<td>15.2752/0.2138</td>
<td>16.6750/0.3915</td>
<td>20.4480/0.5192</td>
</tr>
<tr>
<td>Kolaczyk</td>
<td>10.6059/0.0884</td>
<td>10.1732/0.0874</td>
<td>9.8393/0.0910</td>
<td>10.6137/0.1137</td>
<td>12.0690/0.1722</td>
<td>14.9710/0.2585</td>
</tr>
<tr>
<td>Timmermann</td>
<td>16.6422/0.3612</td>
<td>17.4906/0.3956</td>
<td>18.2062/0.4209</td>
<td>19.0949/0.4676</td>
<td>19.8938/0.5117</td>
<td>20.5138/0.5641</td>
</tr>
<tr>
<td>Total Variation</td>
<td>10.9290/0.1475</td>
<td>13.3080/0.2224</td>
<td>15.7624/0.3074</td>
<td>18.1013/0.4105</td>
<td>19.1947/0.4760</td>
<td>19.7873/0.5155</td>
</tr>
</tbody>
</table>

Fig. 7: Real sensor image data taken with FUJIFILM X-PRO 1 camera. Raw sensor data was downsampled by 3 × 3 to bypass demosaicing and keep the green channel only. (a) Test image ‘aquarium’, 18.04 pixel intensity on average (4.90 from light, 13.14 from sensor). (b) Gretag Macbeth Colorchecker image taken with the same camera and environment settings as (a). (c) Linear regression of the sample means and variances of pixels in Colorchecker squares.

By inverse wavelet transform \((f_0 = \frac{s+x}{2}, \text{etc.})\), we have the relation

\[
\phi(y, t|x, s) = P[G_0 = \frac{t+y}{2}, G_1 = \frac{t-y}{2}|F_0 = \frac{s+x}{2}, F_1 = \frac{s-x}{2}] = (g_0+1)P[G_0 = g_0 + 1, G_1 = g_1|F_0 = f_0, F_1 = f_1].
\]

Hence

\[
s + \frac{x}{2} = \phi(y, t|x, s) = \frac{s + x}{2} P[G_0 = \frac{t+y}{2}, G_1 = \frac{t-y}{2}|F_0 = \frac{s+x}{2}, F_1 = \frac{s-x}{2}] = \frac{t + y + 2}{2} P[G_0 = \frac{t+y}{2} + 1, G_1 = \frac{t-y}{2}|F_0 = \frac{s+x}{2}, F_1 = \frac{s-x}{2}] = \frac{t + y + 2}{2} \phi(y + 1, t + 1|x, s).
\]

where above, we used the property [14]

\[
f_0P[G_0 = g_0, G_1 = g_1|F_0 = f_0, F_1 = f_1]
\]

Similar analysis yields

\[
\phi(y, t|x, s) = \frac{t - y + 2}{2} \phi(y - 1, t + 1|x, s).
\]

Combining, (37) expands to

\[
x\phi(y, t|x, s) = y\phi(y + 1, t + 1|x, s) + \phi(y - 1, t + 1|x, s)
\]

\[
= L\{\phi(\cdot|x, s)\}(y, t).
\]
Fig. 8: Zoomed portions (“fish” on left; “arch” on right) of the denoised results with Gamma correction. (a) Noisy, (b) Total Variation [19], (c) PURE-LET [5], (d) GAT+BM3D [27], (e) Bayes Skellam [34], (f) Kolaczyk [2], (g) Timmermann [3], (h) HMRSO (proposed). A 5-level wavelet transform is performed for denoising. In HMRSO, mixture parameter $\alpha = [0.2, 0.3, 0.4, 0.5, 0.6]$ for each wavelet level from coarse to fine.
Hence,
\[
L\{\phi(\cdot, \cdot)\}(y, t) = y \frac{\phi(y + 1, t + 1) + \phi(y - 1, t + 1)}{2} + (t + 2) \frac{\phi(y + 1, t + 1) - \phi(y - 1, t + 1)}{2}.
\] (43)

Substituting \(L\{\phi(\cdot, \cdot)\}(y, t)\) into (35) proves the Proposition.

C. Proof of Proposition III.3

Proof. Owing to the fact that
\[
E[X|Y = y] = E[E[X|Y, T]|Y = y],
\] (44)
\[
\lambda_{UMR}(y)\] can be rewritten as
\[
\lambda_{UMR}(y) = E[\lambda_{BMR}(y, T)|Y = y] = \sum_{t=0}^{\infty} \lambda_{BMR}(y, t) P[T = t|Y = y].
\] (45)

Using the relation \(E[P[T = t|Y = y] = \frac{1}{\psi(y)}\), we rewrite
\[
\lambda_{UMR}(y) = y \frac{\sum_{t=0}^{\infty} [\phi(y + 1, t + 1) + \phi(y - 1, t + 1)]}{2\psi(y)}
\]
\[
+ \frac{\sum_{t=0}^{\infty} (t + 2) [\phi(y + 1, t + 1) - \phi(y - 1, t + 1)]}{2\psi(y)}.
\] (46)

By the fact
\[
\sum_{t=0}^{\infty} \phi(y \pm 1, t + 1) = \sum_{t=1}^{\infty} \phi(y \pm 1, t)
\]
\[
= \left( \sum_{t=0}^{\infty} \phi(y \pm 1, t) \right) - \phi(y \pm 1, 0)
\]
\[
= \begin{cases} 
\psi(y \pm 1) & \text{if } y \neq \mp 1 \\
\psi(y \pm 1) - P[T = 0] & \text{if } y = \mp 1
\end{cases}
\] (47)

and
\[
\sum_{t=0}^{\infty} (t + 1) \phi(y \pm 1, t + 1) = \sum_{t=1}^{\infty} t \phi(y \pm 1, t)
\]
\[
= \sum_{t=0}^{\infty} t P[T = t|Y = y \pm 1] \psi(y \pm 1)
\]
\[
= \tau(y \pm 1) \psi(y \pm 1)
\] (48)

(46) is equivalent to (18). Note that when \(y = \mp 1\), the extra \(P[T = 0]\) term in (47) conveniently cancels out when substituted into (46).


Wu Cheng received the B.S degree in Optoelectronics Information Engineering from Nanjing University of Science and Technology, Nanjing, China, in 2010, the M.S. and Ph.D. degrees in Electrical and Computer Engineering, in 2012 and 2015, respectively. He is currently an assistant professor at the Intelligent Signal Systems Laboratory at University of Dayton. His research focuses on statistical modeling and image processing.

Keigo Hirakawa (S’00–M’05–SM’11) received the B.S. degree in Electrical Engineering from Princeton University, Princeton, NJ, in 2000 with high honors, the M.S. and Ph.D. degrees in Electrical and Computer Engineering from Cornell University, Ithaca, NY, in 2003 and 2005, respectively, and the M.M. degree in Jazz Performance Studies from the New England Conservatory of Music, Boston, MA, in 2006 with high honors. He was a research associate at Harvard University, Cambridge, MA, from 2006 until 2009. He is currently an assistant professor at the University of Dayton, Dayton, OH. Prof. Hirakawa heads Intelligent Signal Systems Laboratory at University of Dayton, where the group focuses on statistical signal processing, color image processing, and computer vision.