BLIND FULL REFERENCE QUALITY ASSESSMENT OF
POISSON IMAGE DENOISING

Thesis
Submitted to
The School of Engineering of the
UNIVERSITY OF DAYTON

In Partial Fulfillment of the Requirements for
The Degree of
Master of Science in Electrical Engineering

By
Chen Zhang
Dayton, Ohio
May, 2014
BLIND FULL REFERENCE QUALITY ASSESSMENT OF
POISSON IMAGE DENOISING

Name: Zhang, Chen

APPROVED BY:

Keigo Hirakawa, Ph.D.
Advisory Committee Chairman
Assistant Professor, Electrical
and Computer Engineering

Raul Ordonez, Ph.D.
Committee Member
Professor, Electrical
and Computer Engineering

John G. Weber, Ph.D.
Associate Dean
School of Engineering

Tony E. Saliba, Ph.D.
Dean, School of Engineering
&Wilke Distinguished Professor
©Copyright by
Chen Zhang
All rights reserved
2014
ABSTRACT

BLIND FULL REFERENCE QUALITY ASSESSMENT OF POISSON IMAGE
        DENOISING

Name: Zhang, Chen
University of Dayton

Advisor: Dr. Keigo Hirakawa

The distribution of real camera sensor data is well approximated by Poisson, and the estimation of the light intensity signal from the Poisson count data plays a prominent role in digital imaging. It is highly desirable for imaging devices to carry the ability to assess the performance of Poisson image restoration. Drawing on a new category of image quality assessment called corrupted reference image quality assessment (CR-QA), we develop a computational technique for predicting the quality score of the popular structural similarity index (SSIM) without having the direct access to the ideal reference image. We verified via simulation that the CR-SSIM scores indeed agrees with the full reference scores; and the visually optimal denoising experiments performed on real camera sensor data give credibility to the impact CR-QA has on real imaging systems.
To my dear friends at UD

Thy friendship makes us fresh
ACKNOWLEDGEMENTS

This work is funded in part by the University of Dayton Graduate School Summer Fellowship Award, 2013.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>x</td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. BACKGROUND AND RELATED WORK</td>
<td>3</td>
</tr>
<tr>
<td>II.1 Image Quality Assessment</td>
<td>3</td>
</tr>
<tr>
<td>II.2 Structure Similarity(SSIM) Index</td>
<td>5</td>
</tr>
<tr>
<td>II.3 Training process</td>
<td>7</td>
</tr>
<tr>
<td>II.4 Haar Wavelet Transform and Skellam Distribution</td>
<td>9</td>
</tr>
<tr>
<td>III. BLIND STRUCTURAL SIMILARITY INDEX METRIC</td>
<td>11</td>
</tr>
<tr>
<td>III.1 Pixel Domain CR-SSIM</td>
<td>11</td>
</tr>
<tr>
<td>III.2 Wavelet Domain CR-SSIM</td>
<td>13</td>
</tr>
<tr>
<td>IV. EXPERIMENTS</td>
<td>16</td>
</tr>
<tr>
<td>IV.1 FR/CR-QA Performance Comparison</td>
<td>16</td>
</tr>
<tr>
<td>IV.2 Training-/CR-WSSIM Performance Comparison</td>
<td>20</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

II.1 Example of undersmoothing to oversmoothing image restorations and corresponding QA results. (a) Corrupted image. (b-e) Denoised with different parameters. (f) Noise free image. ........................................... 6

II.2 Optimizing denoising parameter $\tau$ using training images $f^{(n)}$. ............ 8

II.3 Example illustrating the dependence of denoising threshold parameter $\tau$ on the scene content. Shown here using wavelet-based denoising of [12] and wavelet FR-QA of [9]. ............................................................. 8

IV.1 Example of Poisson image denoising [3]. (a-b) Simulated ideal and noisy images. (d-e) CR-WSSIM optimal reconstruction is virtually identical to the FR-WSSIM optimal one. (c) By comparison, WMSE optimal image is oversmoothed. (f) Training optimal one is undersmoothed. (g) SSIM scores as a function of threshold value. (h) Finest diagonal subband WSSIM/MSE scores as a function of threshold value. Different color arrows indicate where to choose optimal parameters. FR-QA (green) achieves the maximum. (i) CR-WSSIM score as a function of FR-WSSIM. 45° line indicates a perfect match. (j-l) Some other subbands. ...................... 17

IV.3 Restoration results, Training-/CR-WSSIM vs. FR-WSSIM for testing image No.3 in Figure IV.2 (d-f) ............................... 23

IV.4 Restoration results, Training-/CR-WSSIM vs. FR-WSSIM for testing image No.2 in Figure IV.2 (d-f) ............................... 23

IV.5 Comparison of overall similarity (dB) between CR-WSSIM optimal restorations and Training-WSSIM optimal restorations. ............................... 24

IV.6 Undersmoothing is sensitive to low WSSIM score subbands: (a) Relatively smoother restoration. (b-c) Lowering higher quality subbands. (d-f) Lowering lower quality subbands. ............................... 24

IV.7 Affine parameters determination. (a) Sensor data with Colorchecker (color filled in). (b) After recovery, \( E[G_i] = E[(G_i - E[G_i])^2] \) ............................... 26

IV.8 Application of CR-QA to real image sensor data. (a) Taken with Nikon D90 calibrated to recover Poisson data from raw sensor data. (b-c) WS-SIM optimal denoising result maintains contrast and appears less noisy than the WMSE optimal reconstruction. (d) WSSIM optimal threshold value is smaller than the WMSE optimal values. (e-f) WSSIM optimal threshold results in more shrinkage than the WMSE optimal one. ............................... 26
LIST OF TABLES

IV.1 Averaged deviation of $\tau_\delta$, $\tau_\gamma$ and $\tau_T$ from the desired $\tau_\delta$. For any wavelet subband, $\tau_\delta = \text{arg max}_r \delta[\lambda_r, x]$, $\tau_\delta = \text{arg max}_r \hat{\delta}[\lambda_r|y,t]$, $\tau_\gamma = \text{arg min}_r \hat{\gamma}[\lambda_r|y,t]$, and $\tau_T = \text{arg max}_r \frac{1}{N} \sum_{n=1}^{N} [\delta[x^{(n)}, \lambda_T(y^{(n)}, t^{(n)})]]$. Numbers below are the averaged value of nine subbands deviations.

IV.2 Comparison of FR-, CR-, Training-WSSIM and CR-WMSE over 1000 images in [17]: We report the average and minimum WSSIM score of nine wavelet subbands of each image, averaged over 1000 images. $\delta[y,x]$ is the average qualities of the noisy image compared to the attainable WSSIM. The $\delta[\lambda_\gamma, x]$ is the maximum attainable WSSIM score for this database. Numbers below are (average/minimum).
CHAPTER I

INTRODUCTION

The ubiquity of integrating detectors in scientific and engineering applications suggests that a variety of real-world measurements are high-dimensional count data. Count data, however, is usually an indirect means of measuring the underlying vector-valued signal of interest (e.g., light intensity in a pixel sensor) that cannot be measured directly. As such, the estimation of this signal from the observed count data therefore plays a prominent role across diverse applications. Recent trends in imaging devices have spurred miniaturization of the physical dimensions of sensing devices in effort to increase resolution, practicalize portability, reduce power consumption, and cut fabrication costs. Miniaturization often leads to increased noise in integrating detector by reducing photon count, however. Previous work has already shown that the distribution of noise in image sensor is well approximated by Poisson [1, 2], and a number of Poisson image denoising techniques have already been developed [3, 4, 5, 6].

One unsolved problem is the objective quality assessment of the intensity image reconstructed from Poisson counts. Specifically, an objective visual quality assessment (QA) metric aims to predict the perceived quality in an unsupervised manner. QA can determine the visually optimal balance in denoising between undersmoothing noise and oversmoothing image features. QA also provides a way to compare multiple denoising algorithms and determine the best performing method for a given image.
The image quality assessment techniques that are available today fall into one of the four following categories—full reference assessment (FR-QA), reduced reference assessment (RR-QA), corrupted reference assessment (CR-QA), and no reference assessment (NR-QA). Despite their usefulness in other imaging applications, FR-/RR-/NR-QA modalities are not suited for image restoration tasks such as Poisson image denoising, where the notion of ideal reference (i.e. intensity image) exists in theory but we lack direct access to it (see Section II.1). In this thesis, we draw on CR-QA [7] to predict the quality score of the popular structural similarity index (SSIM) [8, 9] without having the direct access to the ideal reference image (SSIM is FR-QA) by exploiting properties of the noisy observation (i.e. corrupted reference). Experiments with synthetic noisy images and real camera sensor data prove that visually optimal Poisson image denoising is indeed superior to mean square error (MSE) optimal denoising. CR-QA optimal denoising also outperforms denoising optimized by FR-QA with the help of training images.
II.1 IMAGE QUALITY ASSESSMENT

Let $f$ (where $f_i \in \mathcal{H}_2(\mathbb{Z}^2)$, $\ell^2(\mathbb{Z}^2)$ denotes two dimension sample space) be the latent intensity image, or a sequence $f = (\ldots, f_i, \ldots)$ of intensity values $f_i$ at the pixel locations $i \in \mathbb{Z}^2$. Denote by $g \in \mathcal{H}_2(\mathbb{Z}^2)$ the Poisson count sequence $g = (\ldots, g_i, \ldots)$ with $g_i | f_i \sim \mathcal{P}(f_i)$, and $\phi(g)$ the estimate of $f$ based on $g$. Since FR-QA aims to quantify the perceived similarities between the ideal reference and the target images, it is an idealistic metric for assessing the extent to which $\phi$ recovers $f$ from $g$ [8]. Let $QA_{FR}\{f, \phi(g)\}$ be the FR-QA score of $\phi$ relative to $f$. Then, the appeal of (most) FR-QA is that the maximum is attained when the denoised image matches the latent intensity image:

$$QA_{FR}\{f, f\} \geq QA_{FR}\{f, \phi\} \quad \forall \phi \in \mathcal{H}_2(\mathbb{Z}^2).$$  \hspace{1cm} (II.1)

Thus visually optimal image denoising aims to achieve the following:

$$\phi(g)^{opt} = \arg\max_{\phi} QA_{FR}\{f, \phi\}. \hspace{1cm} (II.2)$$

We lack direct access to $f$ in real imaging systems, however, and thus FR-QA optimal image denoising is impractical. An alternative is to use training images representative
of natural images as a proxy for ideal reference image \( f \). The shortcomings of this approach are discussed in detail in Section 2.3. RR-QA is designed to predict the perceptual quality of corrupted image by comparing partial statistical information of the reference image which is passed via an ancillary channel [10]. This is useful in scenarios such as live video streaming, where the encoder and decoder has the means to transmit this sideband information. However this is impractical for image acquisition, as there are no encoders involved. NR-QA offers an alternative to FR-QA by doing away with \( f \) altogether [11]. Though NR-QA admits a way to assess the quality of \( \phi \), NR-QA score does not correspond to the “degree of faithfulness that \( \phi \) reproduces \( f \).” As such, not even a perfect reconstruction \((\phi = f)\) maximizes the NR-QA score:

\[
\exists \phi \in \ell^2(\mathbb{Z}^2) \ni Q_{NR}\{\phi\} \geq Q_{NR}\{f\}, \quad (II.3)
\]

where \( Q_{NR}\{\phi\} \) is the NR-QA score of the image \( \phi \).

By contrast, CR-QA is designed to meet the three desired attributes of QA for image restoration problems (not met by FR-/RR-/NR-QA) [7]:

(A) the metric describes perceptual similarity of the processed image to the ideal reference.

(B) the metric attains a maximum with the ideal reference.

(C) the metric computable without the ideal reference.

One can thus interpret CR-QA as a blind FR-QA [7]—with the help of the corrupted reference image \( g \), CR-QA predicts the FR-QA score of \( \phi \) without a direct access to
\[ f: \]
\[
QA_{CR}\{\phi(\cdot)|g}\} = QA_{FR}\{f, \phi(g)\}. \tag{II.4}
\]

CR-QA in [7] was developed for handling additive white Gaussian noise (AWGN). Though this was an adequate proof-of-concept, more work is needed before CR-QA can be used in real imaging systems. In this thesis, we extend the seminal work of [7] to Poisson corruption scenario, which characterizes the overall noise distribution better than AWGN [1, 2]. The denoising experiments performed on raw camera data in Chapter IV give credibility to the impact CR-QA has on real imaging systems.

II.2 STRUCTURE SIMILARITY (SSIM) INDEX

Structure similarity (SSIM) index is a subjective image quality assessment metric shown to be more consistent with human visual system than some traditional error metrics such as mean squared error (MSE) and mean absolute error (MAE) [8]. Figure II.1 shows a simple example of denoising images between undersmoothing and over-smoothing noisy images, where SSIM indicates a relatively good quality restoration in (c) with higher grade. By contrast, the MSE and MAE hardly register differences accurately among (a)-(e).

SSIM is designed to characterize the similarity of images \( f \) and \( g \) based on their
Figure II.1: Example of undersmoothing to oversmoothing image restorations and corresponding QA results. (a) Corrupted image. (b-e) Denoised with different parameters. (f) Noise free image.

The most common choice for the exponent parameters are $\alpha = \beta = \gamma = 1$; constants $C_1$, $C_2$, are small values that prevent division by zero. When $C_3 = C_2/2$, (II.5) simplifies to:

$$SSIM(f, g) = \frac{2\mathbb{E}[F_i] \cdot \mathbb{E}[G_i] + C_1}{(\mathbb{E}[F_i])^2 + (\mathbb{E}[G_i])^2 + C_1} \times \frac{2\mathbb{E}[F_i] \cdot \mathbb{E}[G_i] - 2\mathbb{E}[F_i] \cdot \mathbb{E}[G_i] + C_2}{\mathbb{E}[(F_i - \mathbb{E}[F_i])^2] + \mathbb{E}[(G_i - \mathbb{E}[G_i])^2] + C_2}. \tag{II.6}$$
between the target image $g$ and the ideal image $f$ in (II.5) can be interpreted as a proxy for FR-QA of $g$. In addition, SSIM has a few desirable characteristics for assessing image denoising performance, including:

1. $\text{SSIM}(f, g) = \text{SSIM}(g, f)$.
2. $|\text{SSIM}(f, g)| \leq 1$.
3. $\text{SSIM}(f, g) = 1$ if and only if $f = g$.

II.3 TRAINING PROCESS

Suppose $\phi_\tau$ is a parametric image denoising function, with parameter $\tau$. FR-QA is commonly used to find the visually optimal parameter $\tau_T$ with the help of training images [7]. This process is illustrated in Figure II.2. Given “ideal” training images $f^{(n)}$, $n \in [1, N]$, make $g^{(n)}|f^{(n)} \sim \mathcal{P}(f^{(n)})$ as a Poisson corrupted version of $f^{(n)}$. FR-QA optimal parameter $\tau_T$ is

$$\tau_T = \arg \max_\tau \frac{1}{N} \sum_{n=1}^{N} [Q_A_{FR}\{f^{(n)}, \phi_\tau(g^{(n)})\}],$$

(II.7)

Once the “training image optimal” denoising parameter $\tau_T$ is found, this parameter is fixed—i.e. $\phi_{\tau_T}$ is used in the camera once the camera is deployed. However, training image-based optimization is limited by whether the training images are representative of the actual data or images used in practice. Moreover, the visually optimal denoising parameter is highly dependent on the scene content. Figure II.3 shows the the wavelet domain SSIM (WSSIM) score [9] corresponding to the image denoising method of [12] over a various range of thresholding values $\tau$ (See (II.7) for further details). As evidenced by the differences in $Q_A_{FR}\{f^{(1)}, \phi_\tau(g^{(1)})\}$ and $Q_A_{FR}\{f^{(2)}, \phi_\tau(g^{(2)})\}$, the visually optimal $\tau$ value $\tau_{FR} = \arg \max_\tau Q_A_{FR}\{f^{(n)}, \phi_\tau(g^{(n)})\}$ is highly dependent on
Figure II.2: Optimizing denoising parameter $\tau$ using training images $f^{(n)}$.

Figure II.3: Example illustrating the dependence of denoising threshold parameter $\tau$ on the scene content. Shown here using wavelet-based denoising of [12] and wavelet FR-QA of [9].

Thus the training image optimal threshold value $\tau_T$ does not always yield a favorable denoising result. It is therefore more desirable to select a visually optimal $\tau$ for each scene. This is the subject of our investigation in the remainder of this thesis.
II.4 HAAR WAVELET TRANSFORM AND SKELLAM DISTRIBUTION

Let \( t^0 = g \). A \( J \) level discrete Haar wavelet\(^1\) transform coefficients \( g \mapsto (t^j, y^j, y^{j-1}, \ldots, y^1) \)
are comprised of sums and differences [13, 14]:

\[
\begin{align*}
\text{jth subband scaling coefficients} & \quad t_i^j = t_{2i}^{j-1} + t_{2i+1}^{j-1} \\
\text{jth subband wavelet coefficients} & \quad y_i^j = t_{2i}^{j-1} - t_{2i+1}^{j-1}.
\end{align*}
\]

(II.8)

Similarly, let \( f \mapsto (s^j, x^j, \ldots, x^1) \) denote Haar wavelet transform of the noise-free image \( f \):

\[
\begin{align*}
\text{jth subband scaling coefficients} & \quad s_i^j = s_{2i}^{j-1} + s_{2i+1}^{j-1} \\
\text{jth subband wavelet coefficients} & \quad x_i^j = s_{2i}^{j-1} - s_{2i+1}^{j-1}.
\end{align*}
\]

(II.9)

The focus of this thesis is on estimation of \( x^j \) based on \( t^j \) and \( y^j \). The estimate of noise-free image \( f \) is then recovered by taking inverse transformation of \( x^j \) and \( s^j \).

Recalling \( t_i^j | s_i^j \sim \mathcal{P}(s_i^j) \), \( y_i^j \) is a difference of two Poisson variables—the distribution of \( y_i^j | f \) is said to be Skellam [3]:

\[
P[Y_i^j = y | X_i^j = x, S_i^j = s] = e^{-s} \left( \frac{s + x}{s - x} \right)^{y/2} I_y(\sqrt{s^2 - x^2}),
\]

(II.10)

From this perspective, the goal is to estimate the “Skellam mean” (latent noise-free wavelet coefficient) \( x \) based on the “Skellam corrupted” variable \( y \) and “Poisson corrupted” variable \( t \); \( f \) is recovered via the inverse wavelet transform. Let \( \lambda_i(y^j, t^j) \) be any Skellam mean estimator of \( x_i^j \), and define the \( \ell^2 \) risk of \( \lambda(y^j, t^j) \) (where \( \lambda_i(y^j, t^j) \in 1\)Ideas presented in 1D wavelet here generalize to 2D straightforwardly.)
\( \ell^2(\mathbb{Z}^2) \) functional as:

\[
\gamma[\lambda, x^j] = \mathbb{E}[\|\lambda(Y^j, T^j) - X_i^j\|^2].
\] (II.11)

The unbiased estimate of risk for Skellam mean estimation \( \lambda(y^j, t^j) \) was first reported by Hirakawa et al. [3, 12]:

**Theorem 1.** Let \( \lambda_i(y^j, t^j) = y_i^j + \theta(y^j, t^j) \). Then \( \hat{\gamma}[\lambda|y^j, t^j] := \)

\[
\mathbb{E}[(T_i^j - Y_i^j)\theta(Y^j + e_i, T^j - e_i) + \theta(Y^j, T^j)^2 + T_i^j
\]

\[
- (T_i^j + Y_i^j)\theta(Y^j - e_i, T^j - e_i) + 2Y_i^j\theta(Y^j, T^j)]
\] (II.12)

is an unbiased estimate of \( \gamma[\lambda, x^j] \), where \( e_i \) is the \( i \)th standard basis.

Indeed, \( \hat{\gamma} \) in Theorem 1 is an example of a CR-WMSE for Poisson imagery based on FR-WMSE \( \gamma \) (here we use “W” to denote in wavelet domain). The significance of Theorem 1 is that the \( \ell^2 \) risk of a given wavelet shrinkage operator \( \lambda(y^j, t^j) \) is knowable without the access to the latent variables \( x^j \) and \( s^j \) (or their distributions). A well known application of Theorem 1 is PURELET—\( \lambda_\tau(y^j, t^j) \) is a parametric function, whose parameter \( \tau \) is the minimizer of the form [4]:

\[
(\tau_\gamma = \arg\min_{\tau} \gamma[\lambda_\tau, x^j]) = (\tau_\hat{\gamma} = \arg\min_{\tau} \hat{\gamma}[\lambda_\tau|y^j, t^j]).
\] (II.13)
CHAPTER III
BLIND STRUCTURAL SIMILARITY INDEX METRIC

III.1 PIXEL DOMAIN CR-SSIM

Thanks in part to the consistent performance against human visual system, SSIM is a popular FR-QA metric commonly used to benchmark performance of various image denoising algorithms. Let $\phi$ be any denoising operator, and simply replace $g$ in (II.6) by $\phi$, we define FR-SSIM functional as

$$\rho[\phi, f] := \frac{2E[F_i] \cdot E[\phi(G)]}{(E[F_i])^2 + (E[\phi(G)])^2} \times \frac{2E[F_i \cdot \phi(G)] - 2E[F_i] \cdot E[\phi(G)]}{E[(F_i - E[F_i])^2 + (\phi(G) - E[\phi(G)])^2]}.$$  \hspace{1cm} (III.1)

The dependence of $\rho[\phi]$ on the statistics of $f$, which are $E[F_i]$, $E[(F_i - E[F_i])^2]$, and $E[F_i \cdot \phi(G)]$, precludes direct application of FR-SSIM to infer the denoising quality of $\phi$ without the access to $f$.

However, these statics estimable indirectly with the help of the “corrupted reference” $g$. From moment matching:

$$E[F_i] = E[G_i].$$  \hspace{1cm} (III.2)

$$E[G_i|F_i] = E[G_i^2|F_i] - (E[G_i|F_i])^2$$

$$\Leftrightarrow E\left[ E[G_i|F_i]\right] = E\left[E[G_i^2|F_i]\right] - E\left[(E[G_i|F_i])^2\right]$$

$$\Leftrightarrow E[F_i^2] = E[G_i^2] - E[G_i].$$  \hspace{1cm} (III.3)
Thus the variance of $f$ can be computed as:

\[
\mathbb{E}[(F_i - \mathbb{E}[F_i])^2] = \mathbb{E}[F_i^2] - (\mathbb{E}[F_i])^2 \\
= \mathbb{E}[G_i^2] - \mathbb{E}[G_i] - (\mathbb{E}[G_i])^2 \\
= \mathbb{E}[(G_i - \mathbb{E}[G_i])^2] - \mathbb{E}[G_i]. \quad (\text{III.4})
\]

In addition, the following relation of the discrete exponential families applies to Poisson count variables [15, 16]:

\[
\mathbb{E}[F_i \cdot \phi(G)] = \mathbb{E}[G_i \cdot \phi(G - e_i)]. \quad (\text{III.5})
\]

Here $e_i$ is a standard vector $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ where $i$ denotes the location of the 1 in the vector. Substituting above to (III.1), we arrive at a CR-SSIM $\hat{\rho}$:

\[
\hat{\rho}[\phi|g] = \frac{2\mathbb{E}[G_i] \cdot \mathbb{E}[\phi(G)]}{(\mathbb{E}[G_i])^2 + (\mathbb{E}[\phi(G)])^2} \times \left( \frac{2\mathbb{E}[G_i \cdot \phi(G - e_i)] - 2\mathbb{E}[G_i] \cdot \mathbb{E}[\phi(G)]}{\mathbb{E}[(G_i - \mathbb{E}[G_i])^2] - \mathbb{E}[G_i] + (\phi(G) - \mathbb{E}[\phi(G)])^2} \right). \quad (\text{III.6})
\]

We emphasize that CR-SSIM in (III.6) is identical to the FR-SSIM in (III.1) —i.e. $\hat{\rho}[\phi, f] = \rho[\phi|g]$.

The main advantage to $\hat{\rho}[\cdot]$ is that it does not rely on the intensity image $f$ and is therefore computable from $g$ and $\phi(g)$. There are a few differences between $\rho[\cdot]$ and $\hat{\rho}[\cdot]$ in implementation, however. First, the expectation operator $\mathbb{E}[\cdot]$ must be approximated by ensemble average, which is computed as a Gaussian weighted average over a local window near location $i$ [8]. By the law of large numbers, SSIMs $\rho[\cdot]$ and $\hat{\rho}[\cdot]$ agree when the local window is larger. When the window size is too small, the empirical statistics may be unstable (in this case, FR-SSIM as well as CR-SSIM
scores fluctuate). Second, $\phi(G - e_i)$ in CR-SSIM refers to the result of denoising when $i$th Poisson count $G_i$ is replaced with $G_i - 1$ (while $G_j$ for $\forall j \neq i$ remains the same). Denoising by $\phi(G - e_i)$ must be repeated $N$ times (where $N$ is the number of pixels in an image). One may reduce the computational complexity by inferring the overall image quality from a random subset of $N$ pixels to compute the $\phi(G - e_i)$. Third, $\hat{\rho}[\phi|g]$ is a biased estimate of $\rho[\phi, f]$ when ensemble average replaces the image statistics (since $\hat{\rho}[\phi|g]$ is a ratio of sample averages). The bias vanishes as the window size is increased.

### III.2 WAVELET DOMAIN CR-SSIM

Many image denoising methods operate in a (linear) transform domain. Haar wavelet transform is well matched for the Poisson image denoising [3]. Though the wavelet version of FR-SSIM (FR-WSSIM) was originally developed for complex wavelet transform [9], we adopt this for the Haar wavelet domain. This is convenient for parametric Skellam mean estimation (e.g. [3, 4]), as parameters can be optimized one subband at a time (combining all subbands via the inverse wavelet transform makes the parameter search space prohibitively large). Let $\lambda(y^j, t^j)$ be any Skellam mean estimation operator. Then the FR-WSSIM is

$$
\delta[\lambda, x^j] = \frac{2\mathbb{E}[X^j_i\lambda(Y^j, T^j)]}{\mathbb{E}[(X^j_i)^2 + \lambda(Y^j, T^j)^2]}.
$$

(III.7)

The simplicity of wavelet domain FR-SSIM stems from the fact that mean wavelet value is zero ($\mathbb{E}[X^j_i] = \mathbb{E}[Y^j_i] = 0$). Again, the dependence of $\delta[\lambda, x^j]$ on the statistics of $x^j_i$ ($\mathbb{E}[X^j_i\lambda(Y^j, T^j)]$ and $\mathbb{E}[(X^j_i)^2]$) makes FR-WSSIM incompatible with real world scenarios.
To develop the CR-WSSIM, the moment matching yields [3]:

\[ \mathbb{E}[(X_j^i)^2] = \mathbb{E}[(Y_j^i)^2 - T_j^i]. \]  

(III.8)

Proof of (III.8) follows from (II.8) and (II.9). Specifically, denote \( T_{2i}^{j-1} = \frac{T_i^j + Y_i^j}{2} = Y_i^+, \) \( T_{2i+1}^{j-1} = \frac{T_i^j - Y_i^j}{2} = Y_i^- \), similarly \( S_{2i}^{j-1} = X_i^+ \) and \( S_{2i+1}^{j-1} = X_i^- \). Thus \( Y_i^+ | X_i^+ \sim \mathcal{P}(X_i^+) \) and \( Y_i^- | X_i^- \sim \mathcal{P}(X_i^-) [3] \). Therefore recalling (III.2):

\[ \mathbb{E}[(X_j^i)^2] = \mathbb{E}[(X_i^+ - X_i^-) \cdot (X_i^+ - X_i^-)] 
\quad = \mathbb{E}[(X_i^+)^2 + (X_i^-)^2 - \frac{(S_j^i)^2 - (X_j^i)^2}{2}] 
\quad \iff \mathbb{E}[(X_j^i)^2] = \mathbb{E}[2((Y_i^+)^2 - Y_i^+) + 2((Y_i^-)^2 - Y_i^-) - ((T_j^i)^2 - T_j^i)] 
\quad = \mathbb{E}[(Y_j^i)^2 - T_j^i]. \]  

(III.9)

Next, recalling (II.8), we have the relation

\[ \mathbb{E}[X_j^i \lambda(Y_j^j, T_j^j)] = \mathbb{E}[(S_{2i}^{j-1} - S_{2i+1}^{j-1}) \lambda(Y_j^j, T_j^j)], \]  

(III.10)

where by (III.5) and (II.8) we have

\[ \mathbb{E}[S_{2i}^{j-1} \lambda(Y_j^j, T_j^j)] = \mathbb{E}[T_{2i}^{j-1} \lambda(Y_j^j - e_i, T_j^j - e_i)] \]  

(III.11)

\[ = \mathbb{E}[(T_i^j + Y_i^j) \lambda(Y_j^j - e_i, T_j^j - e_i)]/2, \]

\[ \mathbb{E}[S_{2i+1}^{j-1} \lambda(Y_j^j, T_j^j)] = \mathbb{E}[(T_i^j - Y_i^j) \lambda(Y_j^j + e_i, T_j^j - e_i)]/2. \]
Combining, we arrive at the wavelet version of CR-SSIM:

\[
\hat{\delta}(\lambda|y^j,t^j) = \frac{\mathbb{E}[(Y_i^j + T_i^j)\lambda(Y^j - e_i, T^j - e_i)]}{\mathbb{E}[(Y_i^j)^2 - T_i^j + \lambda(Y^j, T^j)^2]} + \frac{\mathbb{E}[(Y_i^j - T_i^j)\lambda(Y^j + e_i, T^j - e_i)]}{\mathbb{E}[(Y_i^j)^2 - T_i^j + \lambda(Y^j, T^j)^2]}. \tag{III.12}
\]

As before, the ensemble averages within local window replaces the expectation operator. Unlike the pixel domain CR-SSIM \( \hat{\rho} \), the standard basis \( e_i \) in (III.11) is only a minor inconvenience. Since a typical wavelet-based denoising function \( \lambda(y^j,t^j) \) depends only a few coefficients in \( y^j = (y_1^j, \ldots, y_N^j) \) and \( t^j = (t_1^j, \ldots, t_N^j) \), denoising procedure \( \lambda(y^j \pm e_i, t^j - e_i) \) only needs to be repeated a few times.
CHAPTER IV

EXPERIMENTS

IV.1 FR/CR-QA PERFORMANCE COMPARISON

The following experiments use a bivariate Skellam thresholding [3]:

$$\lambda_i(Y^j_i, T^j_i) = \text{sgn}(Y^j_i) \max(Y^j_i - \tau T^j_i, 0).$$  \hspace{1cm} (IV.1)

By no means do we suggest that this is the best Poisson image denoising technique—with only one parameter in each subband, it is convenient for illustrating the behavior of CR-(W)SSIM. Pseudorandom Poisson count data was created from McGill Calibrated image dataset in [17] that was converted into a linear grayscale image. Figure IV.1 shows an example of selecting a threshold value $\tau \in [0, 1]$ in (IV.1) over a three level Haar wavelet decomposition (9 wavelet subbands in total). SSIM scores in Figure IV.1(g) are shown with a restriction that a single threshold value is applied to all subbands (to keep the search space size reasonable). (W)SSIM scores in Figure IV.1(g-h, j-l) depicts “per subband” quality assessment. The proposed CR-(W)SSIM yields a fairly accurate prediction of FR-(W)SSIM, especially when the WSSIM score is already high (see Figure IV.1(i)). Figure IV.1(h) also shows the WMSE scores, which has a strikingly different trajectory than WSSIM. Besides the fact that it is difficult to determine the optimal threshold due to its relative flatness, WMSE optimal threshold value clearly oversmoothes compared to the WSSIM optimal result. This
Figure IV.1: Example of Poisson image denoising [3]. (a-b) Simulated ideal and noisy images. (d-e) CR-WSSIM optimal reconstruction is virtually identical to the FR-WSSIM optimal one. (c) By comparison, WMSE optimal image is oversmoothed. (f) Training optimal one is undersmoothed. (g) SSIM scores as a function of threshold value. (h) Finest diagonal subband WSSIM/MSE scores as a function of threshold value. Different color arrows indicate where to choose optimal parameters. FR-QA (green) achieves the maximum. (i) CR-WSSIM score as a function of FR-WSSIM. 45° line indicates a perfect match. (j-l) Some other subbands.
is consistent with Figure IV.1(c-e), where “WSSIM optimal” image yields a sharper image (albeit slightly noisier in appearance) compared to the “WMSE optimal” one. However, same as we discussed in Chapter III.1, \( \hat{\delta}(\lambda|y^i,t^i) \) becomes a biased estimate of \( \delta(\lambda,x^i) \) since image statistic is replaced by limited samples average. The probability of such bias could be magnified when noise level is increasing. Table IV.2 summarizes the denoising results over 1000 images in database of [17]. The average intensity \( \mathbb{E}[G] \) is a proxy for the level of noise seen in the image (smaller \( \mathbb{E}[G] \) indicates more noise). We report the average and the minimum WSSIM score of nine wavelet subbands of each image, averaged over 1000 images. \( \delta[y,x] \) is the average qualities of the noisy coefficients compared to the attainable WSSIM. The \( \delta[\lambda_{\tau\delta},x] \) is the maximum attainable WSSIM score for this database. \( \delta[\lambda_{\tau\delta},x] \) is the WSSIM score (the performance) of corresponding CR-WSSIM optimization \( \lambda_{\tau\delta}(y,t) \). When \( \tau_\delta \approx \tau_\delta \), we achieve the desiderata: \( \delta[\lambda_{\tau\delta},x] \approx \delta[\lambda_{\tau\delta},x] \). WMSE optimal denoising by comparison yields far worse WSSIM scores. In Table IV.1, we report the deviation of \( \tau_\delta \) and \( \tau_T \) from the desired \( \tau_\delta \).
Table IV.1: Averaged deviation of $\tau_\delta$, $\tau_\gamma$ and $\tau_T$ from the desired $\tau_s$. For any wavelet subband, $\tau_\delta = \arg \max_r \delta[\lambda_r, x]$, $\tau_\delta = \arg \max_r \delta[\lambda_r|y, t]$, $\tau_\gamma = \arg \min_r \hat{\gamma}[\lambda_r|y, t]$, and $\tau_T = \arg \max_r \frac{1}{N} \sum_{n=1}^N [\delta[x^{(n)}, \lambda_r(y^{(n)}, t^{(n)})]]$. Numbers below are the averaged value of nine subbands deviations.

| $E[G]$ | $|\tau_\delta - \tau_s|$ | $|\tau_\delta - \tau_T|$ | $|\tau_\delta - \tau_\gamma|$ |
|--------|-----------------|-----------------|-----------------|
| 128.00 | 0.001           | 0.003           | 0.010           |
| 18.286 | 0.009           | 0.013           | 0.075           |
| 9.846  | 0.014           | 0.022           | 0.133           |
| 5.818  | 0.022           | 0.032           | 0.193           |
| 4.571  | 0.029           | 0.036           | 0.220           |
| 2.000  | 0.076           | 0.087           | 0.278           |

Table IV.2: Comparison of FR-, CR-, Training-WSSIM and CR-WMSE over 1000 images in [17]: We report the average and minimum WSSIM score of nine wavelet subbands of each image, averaged over 1000 images. $\delta[y, x]$ is the average qualities of the noisy image compared to the attainable WSSIM. The $\delta[\lambda_{r_3}, x]$ is the maximum attainable WSSIM score for this database. Numbers below are (average/minimum).

<table>
<thead>
<tr>
<th>$E[G]$</th>
<th>$\delta[y, x]$</th>
<th>$\delta[\lambda_{r_3}, x]$</th>
<th>$\delta[\lambda_{r_3}, x]$</th>
<th>$\delta[\lambda_{r_3}, x]$</th>
<th>$\delta[\lambda_{r_3}, x]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>128.00</td>
<td>0.719/0.465</td>
<td>0.728/0.500</td>
<td>0.727/0.499</td>
<td>0.726/0.498</td>
<td>0.708/0.449</td>
</tr>
<tr>
<td>18.286</td>
<td>0.495/0.196</td>
<td>0.516/0.236</td>
<td>0.513/0.232</td>
<td>0.512/0.229</td>
<td>0.480/0.143</td>
</tr>
<tr>
<td>9.846</td>
<td>0.420/0.136</td>
<td>0.442/0.172</td>
<td>0.440/0.168</td>
<td>0.437/0.167</td>
<td>0.398/0.074</td>
</tr>
<tr>
<td>5.818</td>
<td>0.357/0.096</td>
<td>0.381/0.129</td>
<td>0.378/0.124</td>
<td>0.377/0.122</td>
<td>0.330/0.039</td>
</tr>
<tr>
<td>4.571</td>
<td>0.329/0.081</td>
<td>0.354/0.112</td>
<td>0.350/0.107</td>
<td>0.350/0.105</td>
<td>0.299/0.029</td>
</tr>
<tr>
<td>2.000</td>
<td>0.239/0.043</td>
<td>0.267/0.075</td>
<td>0.261/0.063</td>
<td>0.261/0.062</td>
<td>0.207/0.029</td>
</tr>
</tbody>
</table>
IV.2 TRAINING-/CR-WSSIM PERFORMANCE COMPARISON

To compare the performance of our proposed method with training process described in Section II.3, we conducted a Four-fold cross-validation [18] over database in [17]. An example of denoising comparisons is shown in Figure IV.1. In Figure IV.1 (f) the restored image is undersmoothed. In Figure IV.1 (h,j-l), red line indicates Training-WSSIM for comparing with FR-/CR-WSSIM. For this image, training selected $\tau_T$s are lower than $\tau_\delta$s or $\hat{\tau}_\delta$s, resulting in undersmoothing. Figure IV.2 shows similar examples comparing QA performance of training- and CR-QA. It is clear from Figure IV.2 that the optimal image threshold depends on the scene. A fixed $\tau_T$ is only adequate when it is close to $\tau_\delta$. By comparison, $\tau_\delta$ is also image content dependent, and is a better approximation to $\tau_\delta$ in most cases. When noise is extremely high, the CR-QA approximation to FR-QA also deteriorates however, so the difference between $\tau_\delta$ and $\tau_\delta$ maybe widen.

The denoising results are illustrated in Figure IV.3 and IV.4. The visual quality of the training image optimal denoised image (undersmoothed) in Figure IV.3-4 (f) and WMSE optimal denoised image (oversmoothed) in Figure IV.3-4 (c) seem to be lower than the FR-WSSIM optimal image in (d) or than the CR-WSSIM optimal image in (e).

Table IV.1 and IV.2 also summarizes the performance of training image optimal denoising $\delta[\tau_T, x]$. Even in presence of high noise ($\mathbb{E}[G] = 2$), we observe the general trend $\delta[\tau_\delta, x] \geq \delta[\tau_\hat{\delta}, x] \geq \delta[\tau_T, x]$. Figure IV.5 plots $\log(\bar{\delta}[\tau_\delta, x])$ against $\log(\bar{\delta}[\tau_T, x])$. Here we use $\bar{\delta}[]$ to denote the overall coefficient similarity of nine subbands for each image [9], estimating the quality of target image. Obviously, it is clear that the reconstruction qualities $\bar{\delta}[\tau_\delta, x]$ and $\bar{\delta}[\tau_T, x]$ are comparable when WSSIM
score is already high; however, when WSSIM score is low, \( \delta[\lambda_{r_5}, x] \) clearly outperforms \( \delta[\lambda_{r_1}, x] \).

We argue that the overall visual quality of a denoised image, especially undersmooth phenomenon is more sensitive to the low WSSIM score subband. This is just the drawback that training-WSSIM encounters. Consider Figure IV.6. In Figure IV.6 (b-f), we deliberately lowered the quality of one wavelet subband. We can see from the images that the overall visual quality is highly sensitive to such processing when WSSIM score is already low. However, when WSSIM score is high, lowering the quality makes little difference.
Figure IV.2: QA Performance comparison of Training/CR- vs. FR-QA: (a-c) Lower noisy image ($\mathbb{E}[G] = 9.846$) finest diagonal/coarser diagonal/finest horizontal subbands. (d-f) Higher noisy image ($\mathbb{E}[G] = 2$) finest diagonal/coarser diagonal/finest horizontal subbands.
Figure IV.3: Restoration results, Training-/CR-WSSIM vs. FR-WSSIM for testing image No.3 in Figure IV.2 (d-f)

Figure IV.4: Restoration results, Training-/CR-WSSIM vs. FR-WSSIM for testing image No.2 in Figure IV.2 (d-f)
Figure IV.5: Comparison of overall similarity (dB) between CR-WSSIM optimal restorations and Training-WSSIM optimal restorations.

Figure IV.6: Undersmoothing is sensitive to low WSSIM score subbands: (a) Relatively smoother restoration. (b-c) Lowing higher quality subbands. (d-f) Lowing lower quality subbands.
IV.3 REAL CAMERA SENSOR DATA EXPERIMENT

Towards the goal of enabling FR-QA (SSIM in particular) in real image systems, we seek to understand the impact CR-QA has on the denoising of actual image sensor data. The image was captured using Nikon D90 in raw sensor mode with all manual settings; we downsampled Bayer data by $2 \times 2$ to avoid extra processing by demosaicking. Each scene was captured twice in succession—once with and once without the Gretag Macbeth Colorchecker [1]. The raw sensor data value $h$ is assumed to be an affine transformation of the Poisson count variable $h_i = \alpha g_i + \beta$ where the affine parameters $\alpha$ and $\beta$ were determined from the Colorchecker. Input to image denoising and CR-QA was the recovered Poisson count data $g = \frac{h-\beta}{\alpha}$. The process of finding $\alpha$ and $\beta$ is illustrated as Figure IV.7. Although the direct comparison of FR-QA and CR-QA is not possible, the general trends of CR-WSSIM seen in earlier simulated experiments remains in tact. Figure IV.8 (d-f) shows a unique peak in WSSIM score indicating a visually optimal threshold value. Comparing Figures IV.8 (d) and (e), WSSIM peak broadens and shifts to the left when denoising a coarser scale, presumably because the signal-to-noise ratio of Skellam variable improves in a coarse scale. Figures IV.8 (d-f) make it clear that WSSIM and WMSE based optimization yield different results—whether WMSE over-/under-smoothing relative to WSSIM depends on the subband. By visual inspection, “WSSIM optimal” denoising has less noisy appearance but yet maintains higher contrast. We captured and tested on a diverse set of real sensor data—they generally follow the trends shown in Figure IV.8.
Figure IV.7: Affine parameters determination. (a) Sensor data with Colorchecker (color filled in). (b) After recovery, $E[G_i] = E[(G_i - E[G_i])^2]$

Figure IV.8: Application of CR-QA to real image sensor data. (a) Taken with Nikon D90 calibrated to recover Poisson data from raw sensor data. (b-c) WSSIM optimal denoising result maintains contrast and appears less noisy than the WMSE optimal reconstruction. (d) WSSIM optimal threshold value is smaller than the WMSE optimal values. (e-f) WSSIM optimal threshold results in more shrinkage than the WMSE optimal one.
CHAPTER V

CONCLUSION

The goal of CR-QA is to enable FR-QA in the context of image restoration with the help of corrupted version of the ideal reference image. Although CR-QA version of SSIM for AWGN image denoising is already known [7], the CR-(W)SSIM for Poisson corruption we proposed is more appropriate for real imaging systems [1]. Though we focused exclusively on SSIM to meet the page limit, we emphasize that Poisson CR-QA is not limited to SSIM. Our experiments with real camera sensor data confirms that it is possible to determine how faithfully the denoised Poisson image reproduced the latent intensity image. This contrasts the current practices of evaluating denoising performance on real camera sensor data that largely rely on visual inspection, and NR-QA that has yet to gain widespread adoption. CR-QA clearly complements the existing workflow of a typical image denoising algorithm development involving FR-QA evaluation of denoised image from simulated sensor data. Hence CR-QA provides a systematic and meaningful way to bridge the evaluation of real and simulated denoising results.


2012.


