LEARNING OPTIMAL PARAMETERS FOR BINARY SENSING IMAGE RECONSTRUCTION ALGORITHMS

Renán A. Rojas[⋆]*, Wangyu Luo*[†]*, Victor Murray*[⋆‡]*, and Yue M. Lu*[†]

[⋆]Department of Electrical Engineering, Universidad de Ingeniería y Tecnología, Lima, Perú
[†]School of Engineering and Applied Sciences, Harvard University, MA, USA
[‡]Department of Electrical and Computer Engineering, University of New Mexico, NM, USA

ABSTRACT

A novel data-driven reconstruction algorithm for quantum image sensors is proposed. Binary observations are efficiently decoded by modeling the reconstruction structure as a two-layer neural network, where optimal coefficients are obtained via error backpropagation. Such a model encapsulates the structure of state-of-the-art algorithms, yet it presents a considerably faster alternative which adapts to input examples without a priori statistical information. Simulations on natural and synthetic datasets show accurate reconstructions with structural similarities consistent with the state of the art, while requiring approximately 5 times less computational cost.

Index Terms— Image reconstruction, quanta image sensors, error backpropagation, MLE, Anscombe transform

1. INTRODUCTION

The constant miniaturization of camera technology allows the fabrication of systems with less power consumption, reduced cost and higher spatial resolution. Yet, its demand for pixels reaching sizes below the diffraction limit of light implies a lower well capacity: each pixel can hold less photoelectrons. Consequently, the dynamic range and signal-to-noise ratio on typical solid-state sensors are severely reduced [1, 2]. A solution to this issue is proposed by a novel imaging paradigm referred to as Quanta Image Sensor (QIS) [3, 4]. Based on photographic film, binary representations are obtained by special photosensors (*jots*) capable of detecting a single photon. Information loss by extreme quantization is neglected by largely oversampling the light field optical resolution, discarding full-well capacity limitations and allowing further pixel shrinkage.

The current bottleneck in the QIS implementation is the high computational cost required to reconstruct the scene. Since it deals with quantized Poisson random processes, classic enhancement strategies are severely restricted. State-of-the-art algorithms include an iterative method based on the maximum likelihood estimation (MLE) [5]. By using the log-likelihood of the binary measurements as cost function, a greedy projection method [6] is adopted to estimate the underlying image. Similarly, a variant [7] exploits the separability of such a cost function via the Alternating Direction Method of Multipliers algorithm (ADMM) [8]. A set of subproblems is solved to estimate the underlying image with faster convergence rate, while a total variation (TV) prior [9, 10] is applied to improve the output quality. Finally, a novel reconstruction strategy [11] applies a variance-stabilizing transformation (VST) to convert distortions into additive noise with normal probability distribution, followed by classic denoising methods to enforce the output smoothness.

The present work proposes a novel QIS image reconstruction algorithm by modeling it as a simple feed-forward neural network and adapting its weights to independent data realizations. Inspired by state-of-the-art algorithms, a highly-efficient parametric representation is adopted, allowing a learning procedure via backpropagation to obtain optimal parameters without prior assumptions. Numerical evaluations on different imaging scenarios show accurate reconstructions comparable to the state of the art with much lower computational requirements. The rest of the paper is organized as follows. Section 2 provides the theoretical background to understand the algorithm. Section 3 shows a detailed description of the proposed reconstruction method. Section 4 presents experimental results on synthetic and natural images. The conclusions are stated in Section 5.

2. THEORETICAL BACKGROUND

2.1. Binary Sensing Scheme

Without loss of generality, let $\mathbf{c} = \{c_0, c_1, \ldots, c_{N-1}\}^T$ denote a set of 1D nonnegative ground-truth coefficients representing the image to be encoded. Assuming a QIS linear array of one spatial unit $x \in [0, 1]$, the captured light intensity field is modeled as $\lambda(x) = \frac{N}{\tau} \sum_{n=0}^{N-1} c_n \cdot \varphi_{\text{QIS}}(Nx - n)$, where τ is the exposure time and φ_{QIS} a nonnegative interpolation kernel. Let the array contain M pixels, such that the *mth* pixel covers the segment $\left[\frac{m}{M}, \frac{(m+1)}{M}\right]$, $m \in \{0, \ldots, M-1\}$. Then, by sampling $\lambda(x)$, the total light exposure value $s_m =$

This work was supported by the UTEC-Harvard Academic Collaboration Funds under the Grant Join Research Seed Fund 2015-03, and by the US NSF under grant CCF-1319140.

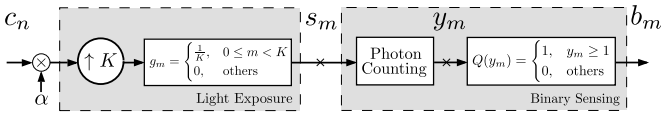
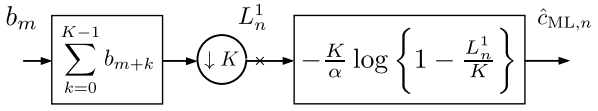
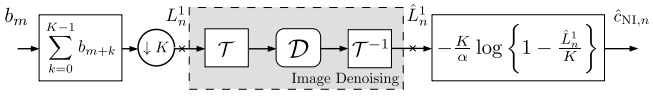


Fig. 1. QIS imaging scheme for $\varphi_{\text{QIS}}(x) = \beta(x)$ and $q = 1$.



(a) MLE algorithm.



(b) Non-iterative reconstruction algorithm.

Fig. 2. Conventional QIS image reconstruction algorithms for $\varphi_{\text{QIS}}(x) = \beta(x)$, and $q = 1$.

$\alpha \cdot \sum_{n=0}^{N-1} c_n \cdot g_{m-Kn}$ is obtained, where α is a gain factor, g_m is a linear filter defined as the standard L^2 inner product between $\varphi_{\text{QIS}}(x)$ and $\beta(Kx - m)$, $K \triangleq \frac{M}{N}$ is the spatial oversampling factor, and $\beta(x)$ is the *box function*:

$$\beta(x) = \begin{cases} 1 & , 0 \le x \le 1 \\ 0 & , \text{other cases} \end{cases}. \tag{1}$$

Particularly, for $\varphi_{\text{QIS}}(x) = \beta(x)$, g_m corresponds to:

$$g_m \triangleq \langle \varphi_{\text{QIS}}(x), \beta(Kx - m) \rangle = \begin{cases} \frac{1}{K} & , 0 \le m < K \\ 0 & , \text{other cases} \end{cases}. \tag{2}$$

To create the binary output, photons hitting each pixel surface are counted and denoted by y_m, which corresponds to realizations of a Poisson random variable Y_m: $\mathbb{P}(Y_m = y_m, s_m) = \frac{s_m^{y_m} e^{-s_m}}{y_m!}$, $\quad y_m \in \mathbb{Z}^+ \cup \{0\}$. The output is then defined as $b_m \triangleq Q(y_m)$, where $Q(y)$ is a binary quantifier with an integer threshold q:

$$Q(y) = \begin{cases} 1 & , y \ge q \\ 0 & , \text{other cases} \end{cases}. \tag{3}$$

Consequently, for random variable $B_m \triangleq Q(Y_m)$, the probability distribution $p_{b_m}(s) \triangleq \mathbb{P}(B_m = b_m, s_m)$, is defined by: $p_0(s) \triangleq \sum_{k=0}^{q-1} \frac{s^k}{k!} e^{-s}$, $p_1(s) \triangleq 1 - \sum_{k=0}^{q-1} \frac{s^k}{k!} e^{-s}$. Figure 1 describes the imaging model for $\varphi_{\text{QIS}}(x) = \beta(x)$ and $q = 1$. For the rest of the paper, this will be the scenario of interest. Likewise, K is assumed to be an integer greater than 1.

2.2. QIS Image Reconstruction Algorithms

2.2.1. Maximum Likelihood Estimation

Let s_m be written in matrix-vector notation as $s_m = \alpha e_m^T \mathbf{G}\mathbf{c}$ for a proper $\mathbf{G} \in \mathbb{R}^{M \times N}$ and the *mth* standard Euclidean basis vector e_m (1 in the *mth* position, 0 elsewhere). Given the log-likelihood function $\ell_{\mathbf{b}}(\mathbf{c}) \triangleq \sum_{m=0}^{M-1} \log\left(p_{b_m}(\alpha e_m^T \mathbf{G}\mathbf{c})\right)$, where $\mathbf{b} = \{b_0, \ldots, b_{M-1}\}^T$, and $\mathbf{c} \in [0, S]^N$ for some upper bound S, the image is decoded by iteratively solving [5]:

$$\hat{\mathbf{c}}_{\text{ML}}(\mathbf{b}) \triangleq \underset{\mathbf{c} \in [0,S]^N}{\text{argmax}} \quad \ell_{\mathbf{b}}(\mathbf{c}). \tag{4}$$

Specifically, if $\varphi_{\text{QIS}}(x) = \beta(x)$ and $q = 1$, s_m becomes piecewise constant: $s_m = \frac{c_n}{K}, m \in \{nK, \ldots, (n+1)K - 1\}$. As a result, coefficients $\{c_n\}$ can be estimated independently of each other by maximizing the likelihood function of observing a subset of K binary outputs, as follows [5, 11]:

$$\hat{c}_n(\mathbf{b}) = \begin{cases} -\frac{K}{\alpha}\log\left(1 - \frac{L_n^1}{K}\right) & , 0 \le L_n^1 \le K\left(1 - e(\frac{-\alpha S}{K})\right) \\ S & , \text{other cases} \end{cases}, \tag{5}$$

where $L_n^1 \triangleq \sum_{k=0}^{K-1} b_{Kn+k}$ is the sum of 1's in the subset.

2.2.2. Non-Iterative Reconstruction Algorithm

A denoising strategy [11] is incorporated to the MLE solution for $\varphi_{\text{QIS}}(x) = \beta(x)$ and $q = 1$. Let the *Anscombe transform* $\mathcal{T}$ and its algebraic inverse $\mathcal{T}^{-1}$ be defined by [12]:

$$Z_n = \mathcal{T}\{X_n\} \triangleq \left(K + \frac{1}{2}\right)^{\frac{1}{2}} \sin^{-1}\left[\left(\frac{X_n + \frac{3}{8}}{K + \frac{3}{4}}\right)^{\frac{1}{2}}\right], \tag{6}$$

$$X_n = \mathcal{T}^{-1}\{Z_n\} \triangleq \left(K + \frac{3}{4}\right) \sin^2\left[\frac{Z_n}{(K + \frac{1}{2})^{\frac{1}{2}}}\right] - \frac{3}{8}, \tag{7}$$

where X_n and Z_n are binomial and approximately-Gaussian random variables, respectively. For a large K, L_n^1 is an approximately-Gaussian random variable with nonstationary variance. Then, variance stabilization via the Anscombe transform is applied to L_n^1, converting distortions into i.i.d. Gaussian random noise and allowing the use of classic denoising methods. Figure 2 shows both reconstruction algorithms.

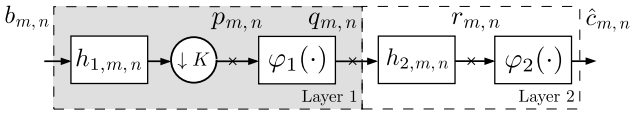


Fig. 3. Proposed two-layer reconstruction algorithm.

3. PROBLEM FORMULATION

3.1. Parametric Representation and Optimality Criterion

A two-layer structure comprised of a concatenation of linear-shift-invariant systems and pointwise nonlinearities is proposed. Figure 3 describes the suggested structure, which consists of the following components: (i) a downsampling process of factor K, which implies an effective downsampling factor of $\hat{K} \triangleq \sqrt{K}$ in each dimension [11], (ii) two linear, shift-invariant systems, $h_{1,m,n} \in \mathbb{R}^{M_{h_1} \times N_{h_1}}, h_{2,m,n} \in$

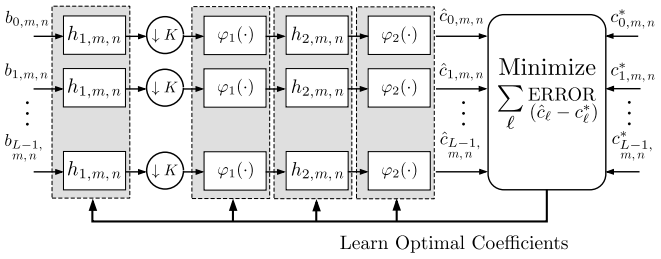


Fig. 4. Proposed learning approach based on the MMSE.

$\mathbb{R}^{M_{h_2} \times N_{h_2}}$, and (iii) two pointwise nonlinearities $\varphi_1(\cdot)$, $\varphi_2(\cdot)$, parametrized by: $\varphi_i(z) = \sum_k w_{i,k} \cdot \beta_3\big(\frac{z}{\Delta_i} - k\big)$, where $w_{i,k} \in \mathbb{R}^{N_{w_i}}$ are the expansion coefficients, the basis function β_3 corresponds to *cubic B-splines* [13], and $\Delta_i \in \mathbb{R}$ is a scaling factor. Therefore, the system is characterized by:

$$\hat{c}_{m,n} = \sum_k w_{2,k} \cdot \beta_3\left\{\frac{r_{m,n}}{\Delta_2} - k\right\}, \tag{8}$$

$$r_{m,n} = \sum_{s,t} h_{2,s,t} \cdot q_{m-s,n-t}, \tag{9}$$

$$q_{m,n} = \sum_k w_{1,k} \cdot \beta_3\left\{\frac{p_{m,n}}{\Delta_1} - k\right\}, \tag{10}$$

$$p_{m,n} = \sum_{s,t} h_{1,s,t} \cdot b_{\hat{K}m-s,\hat{K}n-t}. \tag{11}$$

The fine-tuning of each parameter based on backpropagation is similar to the online trainable structures presented in [14, 15, 16]: motivated by well-established QIS reconstruction schemes, a parametrized feed-forward architecture is adapted to the probability distribution of a representative training set, expressed as $\{\mathbf{b}_\ell, \mathbf{c}^*_\ell\}, \ell \in \{0, L-1\}$. Subsequently, for the scope of this work, we are interested in the minimum mean-square error (MMSE) as optimal criterion. Figure 4 describes the proposed optimization scenario.

Let the system components be expressed in vector form as $\mathbf{h_1} \in \mathbb{R}^{(M_{h_1} \times N_{h_1}) \times 1}$, $\mathbf{w_1} \in \mathbb{R}^{N_{w_1} \times 1}$, $\mathbf{h_2} \in \mathbb{R}^{(M_{h_2} \times N_{h_2}) \times 1}$, $\mathbf{w_2} \in \mathbb{R}^{N_{w_2} \times 1}$, where the 2D linear systems are sorted in column-wise order. Then, let $\mathbf{a} = \text{vec}(\{\mathbf{h_1}, \mathbf{w_1}, \mathbf{h_2}, \mathbf{w_2}\})$ be the column vector formed by their concatenation. Consequently, the optimal system parameters are defined by:

$$\hat{\mathbf{a}} = \underset{\mathbf{a} \in \mathcal{A}}{\text{argmin}} \quad \frac{1}{L} \sum_\ell \varepsilon(\mathbf{a}, \mathbf{b}_\ell, \mathbf{c}^*_\ell), \tag{12}$$

where $\mathcal{A}$ is the set of possible solutions and the cost function is defined as $\varepsilon(\mathbf{a}, \mathbf{b}, \mathbf{c}^*) \triangleq \frac{1}{2}||\mathbf{c}^* - \hat{\mathbf{c}}(\mathbf{a}, \mathbf{b})||^2_{\ell^2}$.

3.2. Optimization Problem and Algorithm Initialization

The greedy projection method is adopted to estimate the optimal parameters [6]. Let $\hat{\mathbf{c}}(\mathbf{a})$ be the vectorized version of system output $\hat{c}_{m,n}$ for parameters $\mathbf{a}$, sorted in column-wise order. Then, $\hat{\mathbf{a}}$ can be iteratively computed as:

$$\mathbf{a}^{(i)} = \underset{\mathcal{A}}{\text{proj}}\big\{\mathbf{a}^{(i-1)} - \mu \nabla \varepsilon\big(\mathbf{a}^{(i-1)}\big)\big\}, \tag{13}$$

where $\text{proj}_{\mathcal{A}}$ is an orthogonal projection operator onto set $\mathcal{A}$ and μ the step size. For each cost function in (12), the gradient $\nabla\varepsilon(\mathbf{a})$ and the Jacobian matrix $\frac{d}{d\mathbf{a}}\hat{\mathbf{c}}(\mathbf{a})$ are expressed as:

$$\nabla\varepsilon(\mathbf{a}) = \left[\frac{d}{d\mathbf{a}}\hat{\mathbf{c}}(\mathbf{a})\right]^T (\hat{\mathbf{c}}(\mathbf{a}) - \mathbf{c}^*), \tag{14}$$

$$\frac{d}{d\mathbf{a}}\hat{\mathbf{c}}(\mathbf{a}) \triangleq \left[\frac{d\hat{\mathbf{c}}(\mathbf{a})}{d\mathbf{h_1}}, \frac{d\hat{\mathbf{c}}(\mathbf{a})}{d\mathbf{w_1}}, \frac{d\hat{\mathbf{c}}(\mathbf{a})}{d\mathbf{h_2}}, \frac{d\hat{\mathbf{c}}(\mathbf{a})}{d\mathbf{w_2}}\right]. \tag{15}$$

Finally, by differentiating $\hat{c}_{m,n}$ w.r.t. each system component, the error backpropagation is determined by:

$$\frac{d\hat{c}_{m,n}}{dh_{1,s,t}} = \varphi_2'(r_{m,n}) \cdot \sum_{\hat{s},\hat{t}} h_{2,m-\hat{s},n-\hat{t}} \cdot \underbrace{\varphi_1'\{p_{\hat{s},\hat{t}}\} \cdot b_{\hat{K}\hat{s}-s,\hat{K}\hat{t}-t}}_{\triangleq z^{(s,t)}_{\hat{s},\hat{t}}}$$

$$= \varphi_2'(r_{m,n}) \cdot \big(h_2 * z^{(s,t)}\big)_{m,n}, \tag{16}$$

$$\frac{d\hat{c}_{m,n}}{dw_{1,k}} = \varphi_2'(r_{m,n}) \cdot \sum_{\hat{s},\hat{t}} h_{2,m-\hat{s},n-\hat{t}} \cdot \underbrace{\beta_3\left\{\frac{p_{\hat{s},\hat{t}}}{\Delta_1} - k\right\}}_{\triangleq \alpha^{(k)}_{\hat{s},\hat{t}}}$$

$$= \varphi_2'(r_{m,n}) \cdot \big(h_2 * \alpha^{(k)}\big)_{m,n}, \tag{17}$$

$$\frac{d\hat{c}_{m,n}}{dh_{2,s,t}} = \varphi_2'(r_{m,n}) \cdot q_{m-s,n-t}, \tag{18}$$

$$\frac{d\hat{c}_{m,n}}{dw_{2,k}} = \beta_3\left\{\frac{r_{m,n}}{\Delta_2} - k\right\}, \tag{19}$$

where $\varphi'(z)$ is the derivative of φ w.r.t. z.

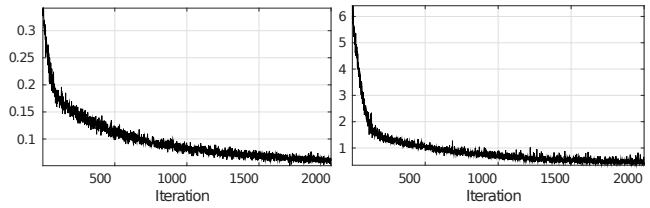
Initial weights $\mathbf{a}^{(0)}$ are set according to the non-iterative reconstruction algorithm with a lowpass filter as denoising method. For the 2D scenario, let L^1 be re-defined as: $L^1_{m,n} \triangleq \sum_{s=0}^{\hat{K}-1} \sum_{t=0}^{\hat{K}-1} b_{\hat{K}m+s,\hat{K}n+t}$. Then, by associating it to the sum of 1's in a region of $\hat{K} \times \hat{K}$ pixels followed by downsampling, $h_{1,m,n}$ is initialized as:

$$h^{(0)}_{1,m,n} = \begin{cases} 1, & -\hat{K} < m < 1 \wedge -\hat{K} < n < 1 \\ 0, & \text{other cases} \end{cases}. \tag{20}$$

$h_{2,m,n}$ is initialized as a Gaussian lowpass filter with standard deviation σ heuristically selected between $[0.25, 1]$. $\varphi_1(z)$ is initialized as the *Anscombe transform* $\mathcal{T}$. Finally, $\varphi_2(z)$ is initialized as the *inverse Anscombe transform* $\mathcal{T}^{-1}$, followed by the logarithmic function $-\log(1 - \frac{x}{K})$ and the factor $\frac{K}{\alpha}$: $\varphi_2^{(0)}(z) = -\frac{K}{\alpha} \log\big(1 - \frac{\mathcal{T}^{-1}(z)}{K}\big)$.

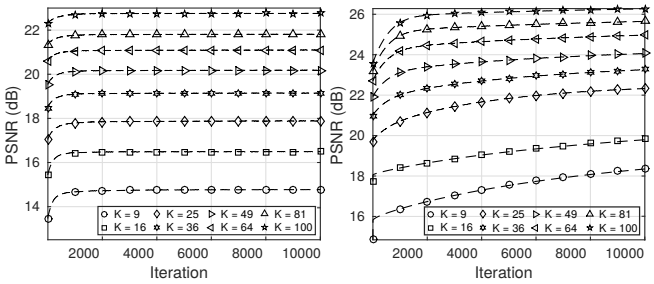
4. SIMULATION RESULTS

Simulations are reported on synthetic and natural images for $K = \{3^2, 4^2 \ldots, 10^2\}$. For the synthetic scenario, c^* is generated as a 32×32 random matrix with standard uniform



(a) Cost function. (b) Gradient norm.

Fig. 5. Gradient descent for the synthetic scenario ($K = 36$).



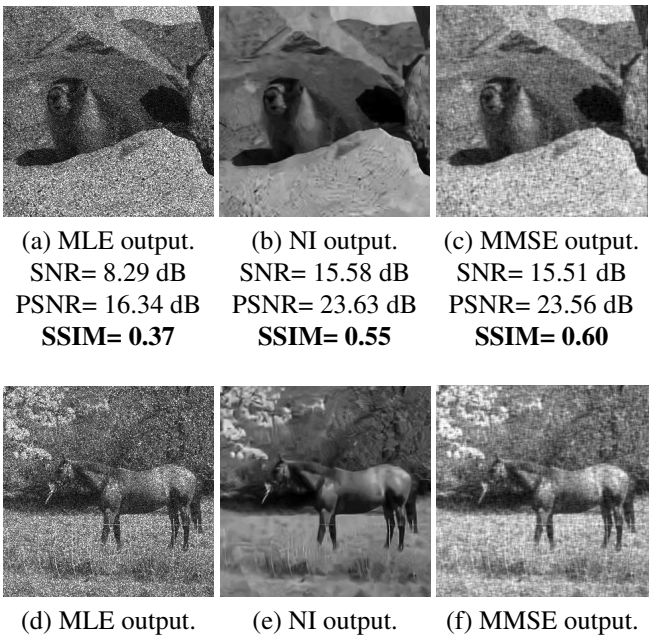
(a) Synthetic scenario. (b) Natural scenario.

Fig. 6. Learning process for synthetic and natural scenarios.

distribution. For the natural scenario, the Cifar-10 dataset [17] ($5 \cdot 10^4$ training samples and 10^4 test samples of 32×32 pixels) is considered. Reconstruction quality is measured by the signal-to-noise ratio (SNR), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) [18]. The experimental setup is as follows: step size $\mu = 5 \cdot 10^{-6}$, scaling factor $\alpha = K$, $\text{iter}_{\max} = 10^4$. Linear systems h_1 and h_2 are of dimensions $\hat{K} \times \hat{K}$ and 15×15, respectively. Expansion coefficients w_1, and w_2 have both 150 elements. Basis functions for φ_1 and φ_2 are spreaded uniformly over the dynamic range of $p_{m,n}$ and $r_{m,n}$, respectively.

For synthetic images, online learning is performed using 128 training samples per iteration and a test set of 512 samples. Figure 5 describes the cost function values and its gradient norm at each iteration for $K = 36$. Both functions show a decreasing behavior with minor oscillations, which reflects the reconstruction improvement along the learning procedure. Figure 6a describes the average PSNR of the reconstructed data set for different K values. Despite the natural improvement for large oversampling values, PSNR increases for all K, with its largest gain corresponding to 1.7 dB for $K = 9$.

For natural images, c^* corresponds to the grayscale version with normalized intensity of a test sample. Online learning is performed using 128 training samples per iteration. Figure 6b describes the average PSNR of the reconstructed data set for different K values. The improvement is more prominent than in the former case, with a maximum gain of 3.2 dB for $K = 9$. Also, performance on larger images is evaluated on the Berkeley Segmentation Dataset [19] (200 training samples and 100 test samples of 481×321 pixels) for $K = 16$, $\mu = 5 \cdot 10^{-8}$, and the described setup. Figure 7 shows results for the proposed algorithm (MMSE), the maximum likelihood estimation (MLE) and the non-iterative method (NI) with BM3D as denoising algorithm [20]. Despite considerable information loss for such a small K value, MMSE yields results comparable with NI, which applies a more complex denoising method, while better preserving the scene structure.



(a) MLE output. SNR= 8.29 dB PSNR= 16.34 dB **SSIM= 0.37**

(b) NI output. SNR= 15.58 dB PSNR= 23.63 dB **SSIM= 0.55**

(c) MMSE output. SNR= 15.51 dB PSNR= 23.56 dB **SSIM= 0.60**

(d) MLE output. SNR= 8.87 dB PSNR= 13.28 dB **SSIM= 0.37**

(e) NI output. SNR= 15.78 dB PSNR= 20.19 dB **SSIM= 0.41**

(f) MMSE output. SNR= 15.6 dB PSNR= 20.01 dB **SSIM= 0.51**

Fig. 7. Reconstruction accuracy for the proposed algorithm and alternative methods on natural images ($K = 16$).

Finally, processing time is compared using Matlab-only code (1.2 GHz Intel core i7, L2: 256K, RAM: 4G) for $K = 16$ and the described setup on the Berkeley dataset. On average, MLE requires 52 ms to process a single sample, whereas NI and MMSE require 2377 ms and 453 ms, respectively. This shows that the proposed algorithm achieves a comparable reconstruction quality but is 5.25 times faster than NI.

5. CONCLUSIONS

The proposed QIS image reconstruction algorithm obtains accurate estimations consistent with state-of-the-art methods, while showing a more efficient design. Its main contribution relies on its adaptive nature: modeled as a simple neural network, optimal components are learned directly from examples without any statistical assumptions, achieving reconstructions with coherent structural similarity 5 times faster than alternative methods, as experimentally demonstrated. Further work will focus on adding layers in the structure while preserving its computational efficiency, in order to explore more complex initial settings and alternative binary sensing scenarios.

6. REFERENCES

[1] Eric R Fossum, "What to do with sub-diffraction-limit (SDL) pixels?—A proposal for a gigapixel digital film sensor (DFS)," in *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, 2005, pp. 214–217.

[2] Eric R Fossum, "Modeling the performance of single-bit and multi-bit quanta image sensors," *IEEE Journal of the Electron Devices Society*, vol. 1, no. 9, pp. 166–174, 2013.

[3] Eric R Fossum, "The Quanta Image Sensor (QIS): Concepts and Challenges," in *Imaging and Applied Optics*. 2011, Optical Society of America.

[4] Luciano Sbaiz, Feng Yang, Edoardo Charbon, Sabine Susstrunk, and Martin Vetterli, "The gigavision camera," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*. IEEE, 2009, pp. 1093–1096.

[5] Feng Yang, Yue M Lu, Luciano Sbaiz, and Martin Vetterli, "Bits from photons: Oversampled image acquisition using binary poisson statistics," *IEEE Transactions on image processing*, vol. 21, no. 4, pp. 1421–1436, 2012.

[6] Martin Zinkevich, "Online Convex Programming and Generalized Infinitesimal Gradient Ascent," in *Machine Learning, Proceedings of the Twentieth International Conference (ICML 2003), August 21-24, 2003, Washington, DC, USA*, 2003, pp. 928–936.

[7] Stanley H Chan and Yue M Lu, "Efficient image reconstruction for gigapixel quantum image sensors," in *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on*. IEEE, 2014, pp. 312–316.

[8] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[9] Leonid I Rudin, Stanley Osher, and Emad Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[10] Manya V Afonso, José M Bioucas-Dias, and Mário AT Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2345–2356, 2010.

[11] Stanley H Chan, Omar A Elgendy, and Xiran Wang, "Images from Bits: Non-Iterative Image Reconstruction for Quanta Image Sensors," *Sensors*, vol. 16, no. 11, pp. 1961, 2016.

[12] Francis J Anscombe, "The transformation of Poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 3/4, pp. 246–254, 1948.

[13] Michael Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal processing magazine*, vol. 16, no. 6, pp. 22–38, 1999.

[14] Karol Gregor and Yann LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 399–406.

[15] Alexander M. Bronstein, Pablo Sprechmann, and Guillermo Sapiro, "Learning Efficient Structured Sparse Models," in *Proceedings of the 29th International Conference on Machine Learning*, 2012, p. 33.

[16] Ulugbek S Kamilov and Hassan Mansour, "Learning optimal nonlinearities for iterative thresholding algorithms," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 747–751, 2016.

[17] Alex Krizhevsky and Geoffrey Hinton, "Learning multiple layers of features from tiny images," Tech. Rep., University of Toronto, 2009.

[18] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[19] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. IEEE, 2001, vol. 2, pp. 416–423.

[20] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.