

Noise-Aware Merging of High Dynamic Range Image Stacks without Camera Calibration

Param Hanji, Fangcheng Zhong, and Rafał K. Mantiuk

Department of Computer Science and Technology, University of Cambridge
{pmh64,fz261,rkm38}@cam.ac.uk

Abstract. A near-optimal reconstruction of the radiance of a High Dynamic Range scene from an exposure stack can be obtained by modeling the camera noise distribution. The latent radiance is then estimated using Maximum Likelihood Estimation. But this requires a well-calibrated noise model of the camera, which is difficult to obtain in practice. We show that an unbiased estimation of comparable variance can be obtained with a simpler Poisson noise estimator, which does not require the knowledge of camera-specific noise parameters. We demonstrate this empirically for four different cameras, ranging from a smartphone camera to a full-frame mirrorless camera. Our experimental results are consistent for simulated as well as real images, and across different camera settings.

Keywords: high dynamic range reconstruction, exposure stacks, camera noise, computational photography

1 Introduction

The dynamic range of a scene may far exceed the range of light intensities that a standard digital sensor can capture. The conventional way of capturing all the information for such a High Dynamic Range (HDR) scene is with a stack of images taken with different exposure times. These are later combined in post-processing as part of the digital pipeline [19, 24, 10, 12, 2, 9]. The probabilistic photon registration and electronic processing in the camera will result in some variation in the values recorded in each pixel, which manifests as noise in images. Any method attempting to accurately estimate the scene radiance from multiple images strives to increase the dynamic range while simultaneously reducing such noise. In this paper, we provide a comprehensive analysis of how noise in images affects the performance of several scene radiance estimators [8, 10, 12]. This work is restricted to static and well-aligned images and we do not consider the problems of pixel alignment and deghosting [14, 25].

It has been shown that, under the assumption of a normal distribution, Maximum Likelihood Estimation (MLE) provides near-optimal estimates of the true radiance values [2]. However, it does not offer a closed-form solution and running non-linear solvers on large images is impractical. For this reason, MLE is typically approximated with an iterative Expectation Maximization (EM) algorithm [10]. We show that such a solver does not always converge to the correct MLE

solution and thus, may introduce an error in estimation. Another limitation of MLE is that it is highly sensitive to the correct calibration of noise parameters [2]. Motivated by these observations, we derive a much simpler, analytical estimator based on the Poisson nature of photon noise that is independent of camera-specific noise parameters and can, therefore, be used with any camera without requiring prior knowledge of its noise characteristic.

Starting with a multi-source noise model [12, 1], we describe a calibration procedure to determine camera-specific noise model parameters. For experimentation, we generated synthetic HDR stacks using physically accurate simulations with the noise parameters of real cameras. We rely on such simulations to compare the empirical biases and standard deviations of different estimators for the scene radiance.

The main contributions of this paper are:

- A simple yet practical camera noise model, fitted for several cameras with both large (full-frame) and small (smartphone) sensors.
- A recommendation to use an estimator based on the Poisson nature of photon noise, which performs as well as near-optimal MLE estimators for the usable dynamic range.
- An empirical validation showing that estimating the sensor noise characteristic is unnecessary when merging HDR images in a noise-aware manner.
- An extended analysis showing that the recommended estimator is robust to high camera noise and is a suitable choice for low-light HDR photography.

2 Related Work

Early HDR reconstruction methods [19, 21, 22, 3, 8] focused on inverting the Camera Response Function (CRF). This is because camera manufacturers did not historically provide access to unprocessed and uncompressed RAW images. Most estimators proposed were weighted averages of the linearized pixel values, where the weights were functions of the inverse and the derivative of the CRF. Debevec and Malik proposed a hat-shaped function that assigns higher weights to linearized pixels near the middle of the intensity range [8]. All these methods do not account for camera noise and therefore provide sub-optimal estimations of HDR pixel values. We refer the reader to chapter three of the book *HDR Video for a detailed discussion on Stack-Based Algorithms for HDR Capture and Reconstruction* [9].

The first HDR estimation method that used a noise model was proposed by Tsin et al. [24]. They proposed to combine images with weights equal to the ratios of the respective exposure times and standard deviations, measured directly from the images. In a later work, Granados et al. [10] showed that Tsin et al.’s method was sub-optimal under a compound-normal noise assumption as pixels near the saturation point were given smaller than ideal weights despite having the highest signal-to-noise ratio (SNR). Debevec and Malik’s hat-shaped weighting function [8] also suffers from this limitation of under-weighting pixels close to the saturation point. Kirk and Anderson [15] proposed an MLE based

weighting scheme using a simple noise model. This was later extended to a more complete model that incorporated noise from several sources by Hasinoff et al. [12]. Granados et al. [10] noted that the true MLE-based estimator does not have an analytical solution and used the EM algorithm for a more accurate estimation than other similar works.

Aguerreberre et al. [2] compared the previously mentioned HDR reconstruction methods and analyzed how far each of their variances were from the theoretical Cramér–Rao Lower Bound (CRLB). They concluded that the variance of MLE-based estimators, such as that of Granados et al. [10], were close to the CRLB but the estimation could be easily affected by errors in noise parameter calibration. In this paper, we show that a comparable performance can be achieved by a simpler estimator without the need for camera calibration.

3 Image Formation Pipeline

We begin this section by explaining the capturing process, highlighting the probabilistic nature of camera noise. A more comprehensive description can be found elsewhere [7, 1, 16], but we include this overview for completeness. After introducing the sensor noise model, we describe the calibration procedure to estimate camera-specific noise model parameters.

3.1 Sensor model

Photons from a scene are captured by the camera lens and pass through a Color Filter Array (CFA), before being focused on an imaging sensor such as a Charged Coupled Device (CCD) or Complementary Metal Oxide Semiconductor (CMOS) sensor as depicted in Fig. 1. When exposed for a fixed interval, the imaging sensor converts some incident photons to electrons. The number of electrons is proportional to the number of registered photons. These electrons accumulate to yield a voltage, which is processed by analog electronics. The next operation performed is amplification of the voltage based on the camera gain. The exposure time and gain are determined by the user-controllable shutter speed and ISO settings. Finally an Analog-to-Digital Converter (ADC) digitizes the signal into discrete pixel intensities. Modern digital cameras provide access to this uncompressed, minimally processed data directly from the electronic imaging sensor in the form of RAW images.

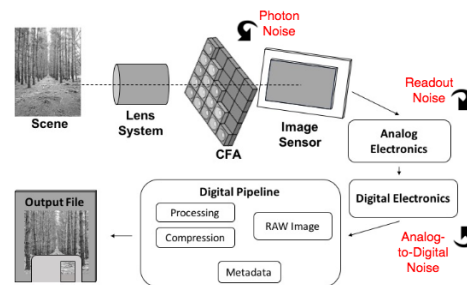


Fig. 1: The illustration of an image formation pipeline that converts photons from a scene into images. Variance in pixel values arises from the noise added at different stages (marked in red).

3.2 Noise model

RAW values are inaccurate measurements of the unknown scene radiance due to the potential saturation of the sensor and the addition of noise at various stages of the pipeline (Fig. 1). To correctly reconstruct a scene in a noise-optimal manner, we model the probabilistic nature of noise in RAW images.

The process of photon registration by the sensor inherently follows a Poisson distribution [13]. This leads to photon noise and is the first source of noise in our model. The other contributions to noise are signal-independent. Because the signal is amplified by some *gain* before reaching the ADC, the signal-independent noise is typically split into pre-amplifier and post-amplifier components [12]. Readout noise captures the voltage fluctuations while accumulating electrons and is amplified along with the signal. The last component, analog-to-digital noise, is added after amplification and is attributed to the quantization error. Digital sensors also exhibit fixed-pattern noise due to photo-response and dark-current non-uniformity [1]. These sources of noise, however, are easy to compensate for as they are fixed for every sensor and are often removed by camera firmware from RAW images. We do not model fixed-pattern noise as it was not present in the images captured by our cameras. Other random sources of noise, such as temperature-dependent dark-current shot noise [7], are accommodated in the signal-independent components. Moreover, previous works [12, 10] indicate that a simple, statistical noise model is sufficient for the problem of HDR radiance estimation.

Let the image be taken with an exposure time t and gain g , and let $Y(p)$ be a random variable representing the final recorded value of the unknown scene radiance $\phi(p)$ at pixel p . The different sources of noise depicted in Fig. 1 motivate the decomposition of this random variable into a sum of three independent random variables. The first random variable is sampled from a Poisson distribution with a parameter equal to the number of incoming photons; the other two are sampled from zero-mean normal distributions. The first normally distributed component accounts for readout noise and has a standard deviation equal to σ_{read} and the second component, parameterized by σ_{adc} , captures amplifier and quantization noise. Assuming that the pixel is not saturated,

$$Y(p) \sim \text{Pois}(\phi(p)t) g k_c + \mathcal{N}(0, \sigma_{\text{read}}) g k_c + \mathcal{N}(0, \sigma_{\text{adc}}) k_c. \quad (1)$$

Each color channel has a different quantum efficiency for photon-to-electron conversion due to differences in the sensitivity of the sensor across the light spectrum. This is accounted for by the color coefficient k_c where $c \in \{r, g, b\}$. Gain affects the Poisson random variable and the first normal random variable, while the k_c is a multiplier on all three terms. The expected value and variance of $Y(p)$ can be written as:

$$\begin{aligned} E[Y(p)] &= \phi(p) t g k_c \\ \text{var}(Y(p)) &= \phi(p) t g^2 k_c^2 + \sigma_{\text{read}}^2 g^2 k_c^2 + \sigma_{\text{adc}}^2 k_c^2 \\ &= \underbrace{E[Y(p)] g k_c}_{\text{photon noise}} + \underbrace{\sigma_{\text{read}}^2 g^2 k_c^2 + \sigma_{\text{adc}}^2 k_c^2}_{\text{static noise}}. \end{aligned} \quad (2)$$

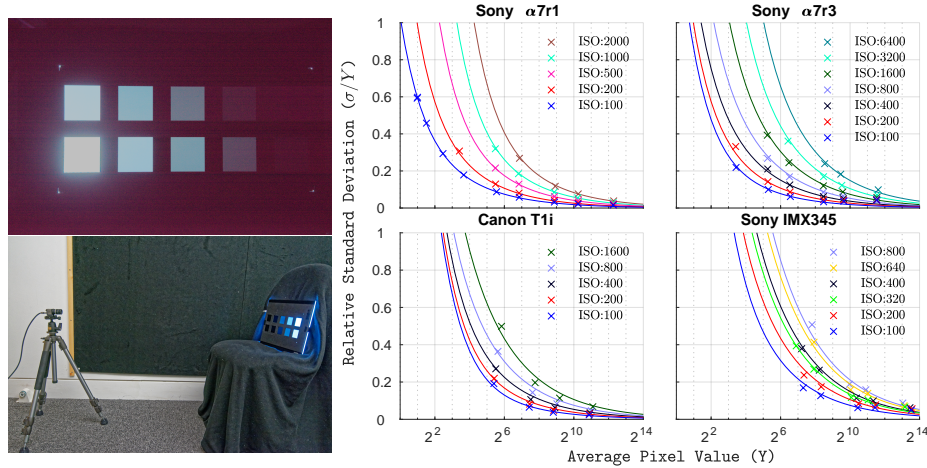


Fig. 2: The calibration target (top-left) and capture setup (bottom-left) used to measure the variance of sensor noise, and the corresponding fitted noise models for the green channel of different cameras (right). Relative standard deviation is plotted against the average pixel value recorded on a logarithmic scale. We control gain by changing ISO (different lines in each plot) since our noise equation (Eq. 2) models how the variance of a pixel changes with gain. The *crosses* represent measurements and the *lines* are the model predictions. The RAW pixel values of all the cameras sensors have been scaled to 14-bit values to enable the comparison of different sensors.

Notice that the variance can be conveniently represented as a function of the expected value. We refer to the signal-independent component of the variance as static noise. Static noise is the same for all pixels of an image. When the radiance of the scene is close to zero, static noise can result in the underestimation of the true radiance, effectively making some pixel values negative. Camera manufacturers typically add an offset, called black-level, to ensure that RAW pixel values are positive. In all our experiments we subtract black level to operate on the actual measurements.

3.3 Noise parameter estimation

Let us consider Eq. 2, the noise model, and how to estimate its camera specific parameters. Rather than measuring the noise added from various sources individually, we use a calibration target shown in the top-left image of Fig. 2. This is constructed by overlaying a uniform light source (a light box) with Neutral Density (ND) filters of different transmittance values so that each square region emits a different radiance. The variance of pixels within each captured square provides an empirical measure of noise for a specific value of radiance.

We extracted several data points (RAW pixel mean and standard deviation pairs) and plotted them using crosses in Fig. 2. The mean of a large number

Table 1: Noise parameters fitted for the tested cameras

Camera	Sensor Size (<i>mm</i>)	Pitch (μm)	Color Coefficients			σ_{read}	σ_{adc}
			k_r	k_g	k_b		
Sony $\alpha 7r1$	35.9×24	4.86	0.327	0.33	0.32	0.7	0.04
Sony $\alpha 7r3$	35.6×23.8	4.5	0.422	0.384	0.389	0.705	3.028
Canon T1i	22.3×14.9	4.69	1.363	1.183	1.153	0.928	5.005
Sony IMX345	8.27×5.51	1.4	0.303	0.313	0.321	1.063	2.373

of pixels contained within each square is used as a substitute for the expected value to fit Eq. 2. To minimize the error in the expected value of the RAW pixel intensity of each square, we computed the average of all pixels within each square from a set of five images captured using the same settings. We captured several such sets starting with a base ISO of 100 and an appropriate shutter speed such that none of the pixels were saturated. We assume that ISO 100 corresponds to the gain of 1. To capture subsequent sets of images, we doubled the ISO and halved the shutter speed every time to maintain the same mean intensity and to use the complete dynamic range of the camera sensor. The number of image sets varied from camera to camera and was typically between five and seven.

We then used a nonlinear solver [17] to estimate the 5 parameters of the noise model from Eq. 2. Very noisy samples with SNR less than one were excluded to ensure convergence. The fit for different cameras, shown in Fig. 2, demonstrates that the model can well explain the noise found in the tested cameras. Each plot shows the measured relative noise against recorded digital intensity, as well as fitted noise model for the green channel. Please refer to the supplementary material to view similar plots for the red and blue channels and also the individual contribution of each component of noise. For a better comparison, the digital values from each camera were rescaled such that the maximum pixel value of every image was set to $2^{14} - 1$. This is the largest bit-rate registered among all our cameras. Estimated parameters for our calibrated cameras are given in Table 1. Images with synthetic noise, generated using Eq. 1 can be found in the supplementary material.

A few interesting observations can be made about the measured cameras. If we define the dynamic range as the ratio between the largest registered value and the smallest value whose SNR is 1 (corresponding to $\sigma/Y = 1$ in the plots in Fig. 2), we notice that the dynamic range differs substantially between the sensors. But it should be noted that these sensors differ in their pixel pitches and resolutions (see Table 1). Hence, the effective amount of noise in images from the different sensors can vary even more when rescaled to the same resolution. The dynamic range of every sensor is also reduced with increased gain (ISO).

3.4 Digital pipeline

The RAW image from the sensor passes through the digital Image Signal Processing (ISP) pipeline starting with black-level subtraction, followed by demosaicing,

denoising, tone-mapping and compression [6]. Most of these stages perform non-linear operations and alter the original readings significantly. Any attempt to recover the scene radiance should thus omit these digital operations. Working with RAW values is preferable as they are linearly related to the scene radiance. The only step needed is black-level subtraction to ensure this linear relationship between radiance and pixel values.

4 HDR Radiance Estimation

In this section, we formulate the problem of estimating radiance values from a number of noisy sensor measurements. We use upper case letters to represent random variables and lower case letters to represent their observed values. Given a stack of RAW images $i = 1, 2, \dots, N$, let $y_i(p)$ represent the RAW value of pixel p and image i captured using a corresponding exposure time t_i and gain g_i . To compensate for the differences in exposure time and gain, and to bring all exposures in the stack to the same scale, we represent their relative radiance as:

$$X_i(p) = \frac{Y_i(p)}{t_i g_i k_c}, \quad (3)$$

such that each $x_i(p)$ is an observation of the true value $\phi(p)$. The expected value of $X_i(p)$ is thus $\phi(p)$ and its variance is obtained by scaling Eq. 2:

$$\sigma_i^2(p) = \frac{\text{var}(Y_i(p))}{t_i^2 g_i^2 k_c^2} = \frac{\phi_i(p)}{t_i} + \frac{\sigma_{\text{read}}^2}{t_i^2} + \frac{\sigma_{\text{adc}}^2}{t_i^2 g_i^2}. \quad (4)$$

All the estimators considered in this section assume that the images in the stack are perfectly aligned. This can be achieved by a global homography-based alignment [23] or a local alignment based on optical flow [18, 26, 4].

Uniform estimator The simplest estimator is the arithmetic mean of all the available samples and is referred to as the *Uniform* estimator. This is obviously a poor estimator as pixel values from different images in the stack are sampled from different distributions and have different SNRs.

Hat-shaped estimator The widely-used weighting scheme proposed by Debevec and Malik [8] assigns higher weights to image pixel values in the middle of the intensity range. The weights are functions of tone-mapped pixel values that pass through the whole pipeline. Let the smallest and largest intensities that can be recorded by a particular sensor be y_{\min} and y_{\max} respectively. Approximating the CRF by the gamma function, the weights are equal to

$$w_i(p) = \begin{cases} y_i(p)^{1/\gamma} - y_{\min}^{1/\gamma} + \epsilon, & y_i(p)^{1/\gamma} \leq \frac{y_{\min}^{1/\gamma} + y_{\max}^{1/\gamma}}{2} \\ y_{\max}^{1/\gamma} - y_i(p)^{1/\gamma} + \epsilon, & y_i(p)^{1/\gamma} > \frac{y_{\min}^{1/\gamma} + y_{\max}^{1/\gamma}}{2}, \end{cases} \quad (5)$$

where $\epsilon = 10^{-10}$ is a small constant added to ensure that the weights are strictly positive and prevent division by zero. γ is generally set to 2.2. Pixels with values close to the noise floor or the saturation point are assigned lower weights. The weights are applied to the linearized pixel values to obtain the *Hat-shaped* estimator:

$$\hat{\phi}_{\text{hat}}(p) = \frac{\sum_{i=1}^N w_i(p) x_i(p)}{\sum_{i=1}^N w_i(p)}. \quad (6)$$

4.1 Maximum Likelihood Estimation

Given the probabilistic image formation model described by Eq. 1, the best estimators are based on MLE, which has been shown to be near-optimal for this problem [2]. The problem is that the Probability Density Function (PDF) of each $X_i(p)$ is a convolution of the three independent density or mass functions, and does not have a close-form expression.

Variance-weighted estimator The noise-based estimator introduced in [15] and extended in [12] assumes that the Poisson component of each $X_i(p)$ can be approximated by a normal distribution. Eq. 1 then simplifies to the sum of three normally distributed random variables, which is also normally distributed. The log-likelihood function to be maximized simplifies to:

$$\ln \mathcal{L}_{\mathcal{N}}(\phi(p)) = \sum_{i=1}^N \ln \frac{1}{\sqrt{2\pi\sigma_i^2(p)}} + \sum_{i=1}^N \frac{-(x_i(p) - \phi(p))^2}{2\sigma_i^2(p)}. \quad (7)$$

where $\sigma_i^2(p)$ is the variance of $X_i(p)$ from Eq. 4. When $\sigma_i^2(p)$ and $\phi(p)$ are independent, the MLE has a simple form:

$$\hat{\phi}_{\text{var}}(p) = \frac{\sum_{i=1}^N \frac{x_i(p)}{\sigma_i^2(p)}}{\sum_{i=1}^N \frac{1}{\sigma_i^2(p)}}. \quad (8)$$

The problem is that $\sigma_i^2(p)$ and $\phi(p)$ are not independent as $\sigma_i^2(p)$ is a function of $\phi(p)$ due to photon noise (refer to Eq. 4). This requires another simplifying assumption, that $\sigma_i^2(p)$ can be estimated using a single observation $y_i(p)$. However, when $E[Y_i(p)]$ is approximated by $y_i(p)$, the variance computed using Eq. 4, can be zero or negative. This may happen because $y_i(p)$ can be negative and larger in magnitude than the static noise. In our implementation, the weights $\frac{1}{\sigma_i^2(p)}$ are replaced in such instances with a small value of $\epsilon = 10^{-10}$.

Iterative Expectation Maximization (EM) A simple method that produces a computationally efficient MLE solution for Eq. 7 is the EM algorithm. In this iterative approach, proposed by Granados et al. [10], $\phi(p)$ is initialized as the mean of $y_i(p)$. Then variances for every exposure are calculated according to Eq. 4 and a better estimate of $\phi(p)$ is found using the new variances according to Eq. 8. The alternating procedure is repeated until converge.

Full MLE The MLE for Eq. 7 can be estimated without making any additional assumptions using a non-linear optimization method. Although this is the most accurate estimator, running such a solver for each pixel is too computationally expensive to be used in practice. We include this estimator to show how close other estimator are to the true MLE solution.

Normal Photon Noise Estimator (NPNE) Eq. 7 has an analytical solution if we assume that the variance is only due to photon noise and there is no static noise. Such a simplification is justified because the overall noise is dominated by the photon noise component for the usable range of $\phi(p)$. Setting static noise to zero and maximizing Eq. 7 yields the estimator:

$$\hat{\phi}_{\text{npne}}(p) = \frac{\sqrt{\sum_{i=1}^N x_i^2(p) t_i \cdot \sum_{i=1}^N t_i + N^2 - N}}{\sum_{i=1}^N t_i}. \quad (9)$$

Poisson Photon Noise Estimator (PPNE) Setting static noise to zero also enables us to simplify the PDF of each $X_i(p)$ and derive an estimator that maximizes the likelihood without the normal approximation to the Poisson distribution. This means that the random variables are sampled from:

$$X_i(p) \sim \frac{\text{Pois}(\phi(p) t_i)}{t_i}. \quad (10)$$

And the new log-likelihood function to be maximized is:

$$\ln \mathcal{L}_{\text{Pois}}(\phi(p)) = \sum_{i=1}^N x_i t_i \ln \phi t_i - \sum_{i=1}^N \phi t_i - \sum_{i=1}^N \ln(x_i t_i)!. \quad (11)$$

The last sum does not depend on $\phi(p)$ and can safely be ignored. The resulting Poisson Estimator takes on the simple form:

$$\hat{\phi}_{\text{ppne}}(p) = \frac{\sum_{i=1}^N x_i(p) t_i}{\sum_{i=1}^N t_i} \quad (12)$$

Such an estimator is a classical choice in the imaging industry [2]. Here, we demonstrated how it can be derived from the assumption of Poisson noise. Additionally, in the supplementary material, we employ the *Lehmann-Scheffe theorem* to show that *PPNE* is the unique Minimum Variance Unbiased Estimator (MVUE) under the assumption of zero static noise. This is an important result, since estimators produced by MLE are generally not guaranteed to be unbiased or to have the minimum variance.

Estimators and noise parameters An advantage of *NPNE* and *PPNE* estimators that ignore static noise is that they do not require knowledge of the noise parameters to provide accurate estimates. Eqs. 9 and 12 are functions of $x_i(p)$, which depends only on the parameter k_c . And, since k_c is the same for all exposures, it is effectively a constant multiplier, which does not affect relative radiance values.

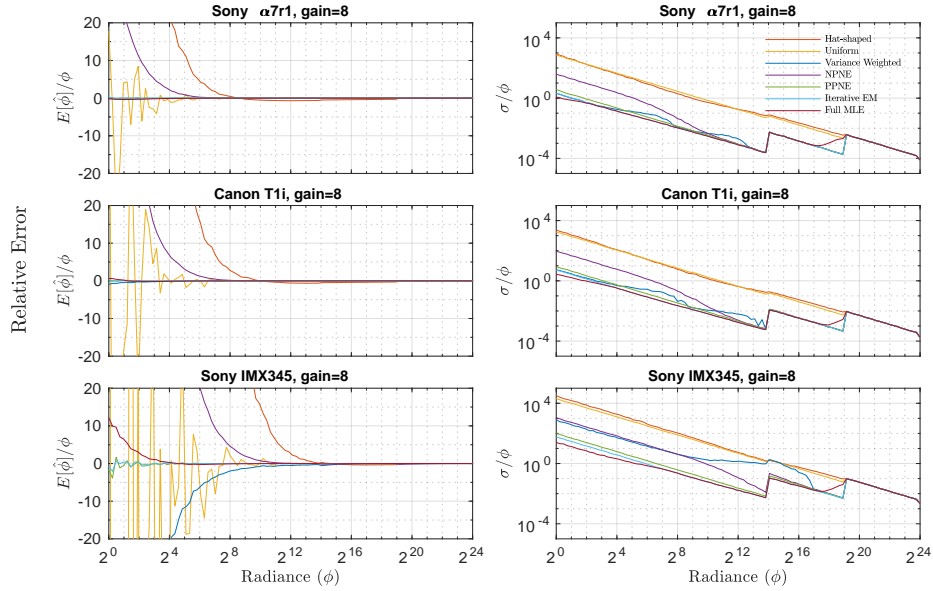


Fig. 3: Monte Carlo simulation results: relative errors of the different estimators (*color lines*) for logarithmically spaced values of the ground truth scene radiance, ϕ (*x-axis*). The errors arise due to non-zero biases (left column of plots) and non-zero standard deviations (right column of plots). The simulation was performed for three exposures, spaced five stops apart.

5 Comparison of Estimators — Simulation

We empirically compare the calibration-independent estimators, *NPNE* and *PPNE*, to the classical *Uniform* and *Hat-shaped* estimators [8] and state-of-the-art MLE-based estimators, *Variance-weighted* [12] and *Iterative EM* [10, 2]. We rely on Monte Carlo (MC) methods and simulate 10,000 HDR exposure stacks. Each stack consists of three exposures separated by five stops in exposure time and a constant gain of 8. Starting with 100 logarithmically-spaced ground truth values of radiance, spanning a dynamic range of 24 stops, we simulated the probabilistic camera capture process described in Section 3.2. Our noisy samples were generated using the parameters of calibrated cameras from Table 1.

Fig. 3 shows relative errors of the compared estimators for three calibrated cameras. We show relative quantities because they better correspond with perceived magnitudes (Weber’s law). Notice that when the radiance ϕ is large, all the estimators are unbiased, and the error is solely due to the relative standard deviation. As expected, both the bias and standard deviation are highest for the smartphone sensor (Sony IMX345), and smallest for the full-frame mirrorless camera (Sony α 7r3). The sawtooth patterns visible in the standard deviation plots are due to the three exposures: the highest five stops are captured only in

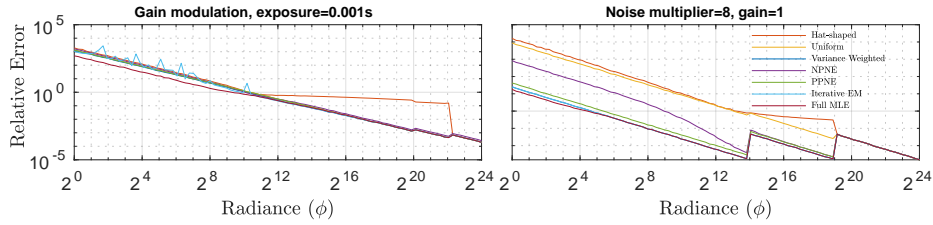


Fig. 4: Relative errors of estimators for two additional scenarios for the Sony $\alpha 7r3$: using a stack of images with different gain and same exposure time (left) and the effect of increasing the static noise by $8\times$ (right). The increased error of the *Hat-shaped* estimator is due to a small negative bias (see left column of Fig. 3).

images with the shortest exposure time and, therefore, the error drops at $\phi \approx 2^{19}$ and $\phi \approx 2^{14}$ when data from the second and third images become available.

The results in Fig. 3 confirm that the estimators that do not account for noise (*Uniform* and *Hat-shaped*) result in the highest bias and the highest amount of noise. The bias of the *Uniform* estimator appears unstable at low radiance values because the estimate is dominated by noise. The popular *Variance-weighted* estimator performs reasonably well for the high quality sensor (Sony $\alpha 7r3$) but results in a noisy estimate with a negative bias for the two other sensors. The negative bias is due to clamping of weights since variances can not be negative. As expected, the *Full MLE* achieves the best performance. It is, however, much more computationally expensive. Notice that, the *Iterative EM* estimator does not converge to the same solution as the *Full MLE* for lower radiance values.

The most interesting results are seen for the two estimators that account only for photon noise. The estimator that assumes photon noise is normally distributed (*NPNE*) introduces a positive bias and results in a higher standard deviation than the estimator that assumes Poisson photon noise distribution (*PPNE*). Overall, the error of the *PPNE* is comparable with that of the *Iterative EM* estimator, which is the best estimator used in practical applications. The relative standard deviation of noise is only marginally higher than that of the *Iterative EM* but only for very low pixel values. However, it has a major advantage over the *Iterative EM* as it does not require knowledge of the noise parameters. Therefore, it can be used with any camera without prior calibration, as long as the camera noise characteristic can be well explained by our model (Eq. 1).

5.1 Gain modulation

An alternate acquisition strategy for capturing HDR scenes is modulating gain while maintaining the same exposure time and consistent motion blur across images [11]. A detailed analysis of the increase in dynamic range and SNR with the number of images captured for different strategies is presented in the supple-

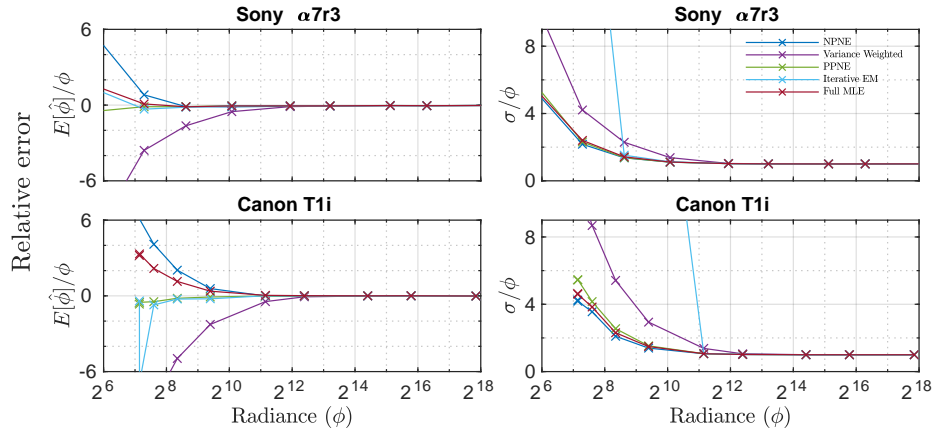


Fig. 5: Results for real images: observed relative biases and standard deviations for the green channel of two cameras. The sudden increase in standard deviation at $\phi \approx 2^{11}$ for the *Iterative EM* estimator in the top-right plot is due to the non-convergence of the EM algorithm for low inputs.

mentary material. Here, we compare the estimators on a stack of gain modulated captures. The plots on the left of Fig. 4 indicate that the performance of most estimators is very similar, except for the better performance of *Full MLE* for low radiance and failure of the *Hat-shaped* estimator for medium radiance values.

5.2 Robustness to noise

We validated the performance of the estimators for exposure stacks captured with increased camera noise. The noise in input image stacks was artificially increased by amplifying the contribution of static noise $2\times$, $4\times$ and $8\times$ the measured value for the Sony $\alpha 7r3$ sensor. See the right side of Fig. 4 for the simulation with $8\times$ static noise and the supplementary for other multipliers. These additional MC simulations, confirm that the relative performance of *PPNE* does not noticeably degrade with noise. Its relative error is very similar to the calibration-sensitive *Iterative EM* estimator even when the input is very noisy. Since the proportion of static noise is much greater in low-luminance conditions, these results indicate that *PPNE* is suitable for low-light HDR photography.

6 Comparison of Estimators — Real Images

To make sure that our simulation results are not the outcome of wrong assumptions about camera noise, we measured the errors of different estimators on real data using a stack of images of our calibration target (see Fig. 2-left). The stack is composed of images captured with different exposure times and gains as described in Section 3.3. The reference radiance ϕ was calculated as the average

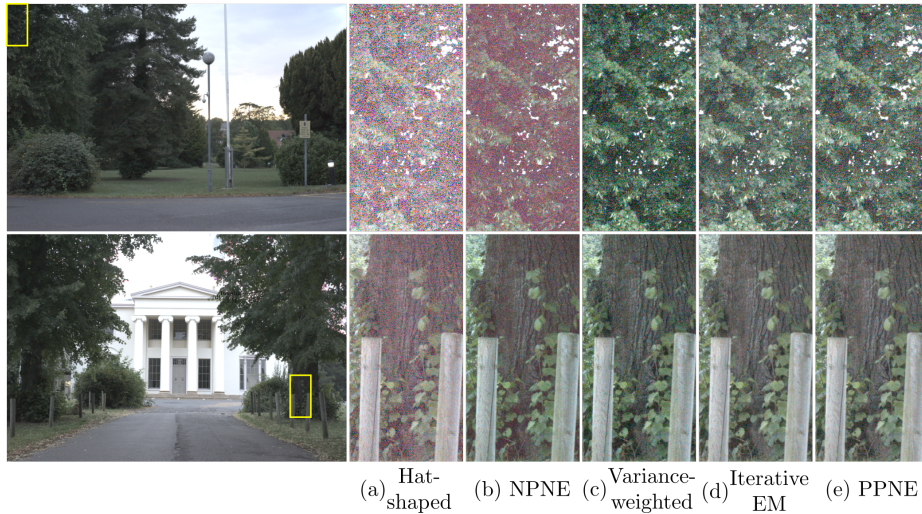


Fig. 6: HDR reconstructions of outdoor scenes, “Trees” and “House”, using different estimators given exposure stacks of three images captured by the Sony $\alpha 7r3$ at ISO 6400 and gamma-encoded for visualization ($\gamma = 2.2$). The positive bias of the *Hat-shaped* estimator and *NPNE* as well as the negative bias of the *Variance-weighted* estimator are visible in the dark regions. The images produced by the iterative EM estimator and *PPNE* are almost identical. Refer to the supplementary for the “Cottage” and “Street” scenes. In all the scenes, the shortest exposure times were deliberately set to a small value to produce noisy images and test the robustness of the estimators.

of all pixels in each square of the target. Fig. 5 shows the error, due to bias and standard deviation, of the HDR estimations for our calibrated cameras. Here we see a similar pattern as in Fig. 3, where the standard deviation of the *Variance-weighted* estimator is much higher than other MLE-based estimators and it has a large negative bias at low radiance. The performance of the analytical *PPNE* is very similar to that of the *Iterative EM* and the *Full MLE*.

6.1 Qualitative and quantitative comparison on complex images

Next, we show visual differences in HDR images due to the choice of the estimator in challenging conditions. We captured several scenes with three exposure times, spaced two stops apart. The images are processed with different estimators and show substantial differences, as depicted in Fig. 6. The *Hat-shaped* weights resulted in a noisy image with a positive bias (Fig. 6a) in dark regions. *NPNE* substantially reduced the amount of noise but still produced the bias (Fig. 6b). The *Variance-weighted* estimator produced a noisier image than *NPNE* (Fig. 6c) and also introduced a negative bias. This made some pixel values darker than they should be, resulting in an accidental increase in contrast. Merged images

Estimator	PU-PSNR				PU-SSIM				HDR-VDP-3			
	House	Trees	Cottage	Street	House	Trees	Cottage	Street	House	Trees	Cottage	Street
Hat-shaped	20.266	13.241	18.516	9.019	0.636	0.489	0.51	0.302	5.719	5.39	5.013	4.638
Var-weighted	16.412	18.191	22.476	13.267	0.624	0.497	0.578	0.512	6.049	5.946	5.846	5.581
NPNE	24.326	17.451	13.588	13.331	0.681	0.543	0.374	0.289	6.027	6.114	5.391	5.304
PPNE	24.428	18.876	22.336	13.361	0.72	0.538	0.607	0.432	6.255	6.253	5.904	5.802
EM	24.922	18.339	23.059	13.962	0.72	0.52	0.614	0.524	6.273	6.37	5.983	5.928

Table 2: The reconstruction error for images in Fig. 6 and Fig. 4 in the supplementary. The error is computed using HDR image quality metrics: PU-PSNR, PU-SSIM [5] and HDR-VDP-3 [20] (v3.0.6, Q-values). For all the metrics, a higher value denotes higher quality. In each column, the highest value has a gold background, the second-best has a silver background and the third has a bronze background. Overall, *PPNE* is the second-best estimator and it is narrowly outperformed by the calibration-sensitive *EM* estimator.

of *PPNE* (Fig. 6d) and the *EM* estimator (Fig. 6e) show the least amount of noise and smallest bias. For additional indoor and outdoor scenes, please refer to the supplementary material. Apart from the independence to noise-parameters, another advantage over the *Iterative EM* estimator is the reduced computation time of *PPNE* due to its analytical form. This does not make much of a difference for the simple logarithmic gradient, but is important for high-resolution images captured by a DSLR.

Finally, we report quality scores for three HDR image quality metrics in Table 2. The test images were obtained from three exposures, captured at high ISO setting and merged with each estimator while the reference images were obtained by merging five exposures with the *EM* estimator (the most accurate). The results confirm the findings of other experiments; *PPNE* produces results that are only marginally worse than those of *EM*, even though *EM* was used to generate the reference images.

7 Conclusions

Although the state-of-the-art HDR reconstruction methods advocate using MLE solvers that require accurate camera parameters, we demonstrate that they provide little advantage over the simple Poisson noise estimator, which does not require camera noise calibration. We show that the Poisson noise estimator is unbiased and its standard deviation is only marginally higher than that of the near-optimal MLE solution for very low pixel values. Such a difference is unlikely to be noticed in complex images. For a simplified noise model, the Poisson estimator is provably MVUE. Furthermore, we show how each estimator can be derived making different simplifying assumptions about the camera noise model, and we illustrate the relative errors of the estimators using gain modulation and under increased static noise. In all our experiments, the Poisson noise estimator was consistently among the best performing estimators.

Acknowledgement

We would like to thank Minjung Kim and Maryam Azimi for their advice on the paper. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement N° 725253–EyeCode).

References

1. Aguerrebere, C., Delon, J., Gousseau, Y., Musé, P.: Study of the digital camera acquisition process and statistical modeling of the sensor raw data. Tech. rep. (Aug 2013)
2. Aguerrebere, C., Delon, J., Gousseau, Y., Musé, P.: Best Algorithms for HDR Image Generation. A Study of Performance Bounds. *SIAM Journal on Imaging Sciences* **7**(1), 1–34 (jan 2014)
3. Akyüz, A.O., Reinhard, E.: Noise reduction in high dynamic range imaging. *Journal of Visual Communication and Image Representation* **18**(5), 366–376 (2007)
4. Anderson, R., Gallup, D., Barron, J.T., Kontkanen, J., Snavely, N., Hernández, C., Agarwal, S., Seitz, S.M.: Jump: virtual reality video. *ACM Transactions on Graphics (TOG)* **35**(6), 1–13 (2016)
5. Aydın, T.O., Mantiuk, R., Seidel, H.P.: Extending quality metrics to full dynamic range images. In: *Human Vision and Electronic Imaging XIII*. pp. 6806–10. Proceedings of SPIE, San Jose, USA (January 2008)
6. Buckler, M., Jayasuriya, S., Sampson, A.: Reconfiguring the Imaging Pipeline for Computer Vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 975–984 (may 2017)
7. Costantini, R., Susstrunk, S.: Virtual sensor design. In: *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications V*. vol. 5301, pp. 408–419. International Society for Optics and Photonics (2004)
8. Debevec, P.E., Malik, J.: Recovering high dynamic range radiance maps from photographs. In: *Proc. of SIGGRAPH ’97*. pp. 369–378. ACM Press, Los Angeles, CA, USA (1997)
9. Gallo, O., Sen, P.: Stack-Based Algorithms for HDR Capture and Reconstruction. In: *High Dynamic Range Video*, pp. 85–119. Elsevier (2016)
10. Granados, M., Ajdin, B., Wand, M., Theobalt, C., Seidel, H.P., Lensch, H.P.: Optimal hdr reconstruction with linear digital cameras. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. pp. 215–222. IEEE (2010)
11. Hajisharif, S., Kronander, J., Unger, J.: Adaptive dualiso hdr reconstruction. *EURASIP Journal on Image and Video Processing* **2015**(1), 41 (2015)
12. Hasinoff, S., Durand, F., Freeman, W.: Noise-optimal capture for high dynamic range photography. In: *CVPR*. pp. 553–560. IEEE (2010)
13. Healey, G., Kondepudy, R.: Radiometric CCD camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**(3), 267–276 (mar 1994)
14. Karadzovic-Hadziabdic, K., Telalovic, J.H., Mantiuk, R.: Expert evaluation of deghosting algorithms for multi-exposure high dynamic range imaging. In: *Second International Conference and SME Workshop on HDR imaging*. pp. 1–4. Citeseer (2014)

15. Kirk, K., Andersen, H.J.: Noise characterization of weighting schemes for combination of multiple exposures. In: BMVC. vol. 3, pp. 1129–1138. Citeseer (2006)
16. Konnik, M., Welsh, J.: High-level numerical simulations of noise in CCD and CMOS photosensors: review and tutorial (2014)
17. Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E.: Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on optimization* **9**(1), 112–147 (1998)
18. Liu, C., et al.: Beyond pixels: exploring new representations and applications for motion analysis. Ph.D. thesis, Massachusetts Institute of Technology (2009)
19. Mann, S., Picard, R.: Being "undigital" with digital cameras: extending dynamic range by combining differently exposed pictures. In: Proc. IS&T 48th Annual Conference. pp. 442–428 (1995)
20. Mantiuk, R., Kim, K.J., Rempel, A.G., Heidrich, W.: Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on graphics (TOG)* **30**(4), 1–14 (2011)
21. Mitsunaga, T., Nayar, S.K.: Radiometric self calibration. In: Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149). vol. 1, pp. 374–380. IEEE (1999)
22. Robertson, M.A., Borman, S., Stevenson, R.L.: Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *Journal of Electronic Imaging* **12**(2), 219–229 (2003)
23. Tomaszewska, A., Mantiuk, R.: Image registration for multi-exposure high dynamic range image acquisition (2007)
24. Tsin, Y., Ramesh, V., Kanade, T.: Statistical calibration of CCD imaging process. In: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. vol. 1, pp. 480–487. IEEE (2001)
25. Tursun, O.T., Akyüz, A.O., Erdem, A., Erdem, E.: The state of the art in hdr deghosting: a survey and evaluation. In: Computer Graphics Forum. vol. 34, pp. 683–707. Wiley Online Library (2015)
26. Zimmer, H., Bruhn, A., Weickert, J.: Optic flow in harmony. *International Journal of Computer Vision* **93**(3), 368–388 (2011)