Supplementary materials for:
Noise-Aware Merging of High Dynamic Range Image Stacks without Camera Calibration

Param Hanji, Fangcheng Zhong, and Rafał Mantiuk

University of Cambridge

Appendix 1: Proof of MVUE

Here we use the Lehmann-Scheffe theorem to show that with no static noise, the Poisson Estimator $\hat{\phi}_{\mathrm{ppne}}$ is the unique minimum variance unbiased estimator (MVUE). We state the theorem and a few related definitions first. In this section, the per pixel notation is dropped and unless explicitly stated otherwise, every summation and product is indexed by i going from 1 to N, which identifies each image in a stack of size N.

Definition 1 (Sufficient statistic). *Let $\vec{U}$ denote a random vector sampled from a distribution that has joint pdf or pmf $f(\vec{u}; \theta)$ parameterized by $\theta \in \Omega$. Let $S = g(\vec{U})$ be a statistic whose PDF or PMF is $f_S(s; \theta)$. S is a sufficient statistic for θ if $\frac{f(\vec{u}; \theta)}{f_S(s; \theta)} = h(\vec{u})$ where $h(\vec{u})$ does not depend on θ.*

Definition 2 (Complete family). *Let the random variable S have a PDF or PMF that is a member of the family $\{f_S(s; \theta) | \theta \in \Omega\}$. Let $w(S)$ be a function of S. If the expectation $E[w(S)]$ being zero for every $\theta \in \Omega$ requires that $w(S)$ be zero for every s in the support of $f_S(s; \theta)$, then the family $\{f_S(s; \theta) | \theta \in \Omega\}$ is called a complete family.*

Theorem 1 (Lehmann and Scheffe). *Let $\vec{U}$ denote a random vector sampled from a distribution that has joint PDF or PMF $f(\vec{u}; \theta)$ parameterized by $\theta \in \Omega$. Let $S = g(\vec{U})$ be a sufficient statistic for θ, and let the family $\{f_S(s; \theta) | \theta \in \Omega\}$ be complete. If there is a function of S that is an unbiased estimator of θ, then this function of S is the unique MVUE of θ.*

Proof. Let $x_i t_i$ be denoted as u_i, then $\hat{\phi}_{\mathrm{ppne}} = \frac{\sum u_i}{\sum t_i}$. Let $\vec{u}$ denote $(u_1, u_2, ...u_N)$, a random vector with joint PMF $f(\vec{u}; \phi)$. Let $s = \sum u_i$. We first show that s is a sufficient statistic for ϕ.

According to Equation 1, $u_i \sim \mathrm{Pois}(\phi\, t_i)$ if we assume no static noise. Since each u_i is independent,

$$f(\vec{u}; \phi) = \prod \frac{(\phi t_i)^{u_i} e^{-\phi t_i}}{u_i!} = e^{-\sum \phi t_i} \phi^{\sum u_i} \prod (t_i)^{u_i} \prod \frac{1}{u_i!}$$

$s \sim \mathrm{Pois}(\sum \phi t_i)$ is a compound Poisson distribution with PMF

$$f_S(s;\phi) = \frac{(\sum \phi t_i)^s e^{-\sum \phi t_i}}{s!} = \frac{\phi^s (\sum t_i)^s e^{-\sum \phi t_i}}{s!}$$

Then,

$$\frac{f(\vec{u};\phi)}{f_S(s;\phi)} = \frac{(\sum t_i)^s \prod (t_i)^{u_i}}{s! \prod u_i!},$$

which does not depend on ϕ. Therefore, $s = \sum u_i$ is a sufficient statistic for ϕ by definition. Next, we show that the family $\{f_S(s;\phi)|\phi > 0\}$ is complete.

Suppose a function $w(s)$ of s is such that $E[w(s)] = 0$ for every $\phi > 0$. We want to show that this requires $w(s) = 0$ for every $s >= 0$. First, we know that for all $\phi > 0$,

$$0 = E[w(s)] = \sum_{s=0}^{\infty} w(s) f_S(s;\phi) = \sum_{s=0}^{\infty} w(s) \frac{(\sum \phi t_i)^s e^{-\sum \phi t_i}}{s!}$$

$$= e^{-\sum \phi t_i} \sum_{s=0}^{\infty} w(s) \frac{\phi^s (\sum t_i)^s}{s!}$$

Since $e^{-\sum \phi t_i}$ is nonzero, $\sum_{s=0}^{\infty} w(s) \frac{(\sum t_i)^s}{s!} \phi^s$ is a power series of ϕ which converges to zero for all $\phi > 0$. Therefore its coefficients must be zero. Since $\frac{(\sum t_i)^s}{s!}$ is also non-zero, $w(s)$ must be zero for every $s >= 0$, which completes the proof of $\{f_S(s;\phi)|\phi > 0\}$ being a complete family.

At last, note that $\hat{\phi}_{\mathrm{ppne}}$ is an unbiased estimator of ϕ (this is also true for nonzero static noise):

$$E[\hat{\phi}_{\mathrm{ppne}}] = \frac{\sum E[u_i]}{\sum t_i} = \frac{\sum \phi t_i}{\sum t_i} = \phi$$

And since $\hat{\phi}_{\mathrm{ppne}}$ is a function of a sufficient statistic for ϕ, $\hat{\phi}_{\mathrm{ppne}}$ is the unique MVUE of ϕ by the *Lehmann-Scheffe* theorem.

Appendix 2: Noise Parameter Validation

The fits for all color channels of the tested cameras are depicted in Fig. 1. These are computed from the calibration box. We also qualitatively compare real versus simulated noise in Fig. 2. This comparison further confirms the accuracy of our noise parameters and shows that our model (Eq. 1) can effectively simulate the camera capture process. Since ground truth data is unavailable for real images, we relied on such simulations to compare different high dynamic range (HDR) estimators and different capture strategies.
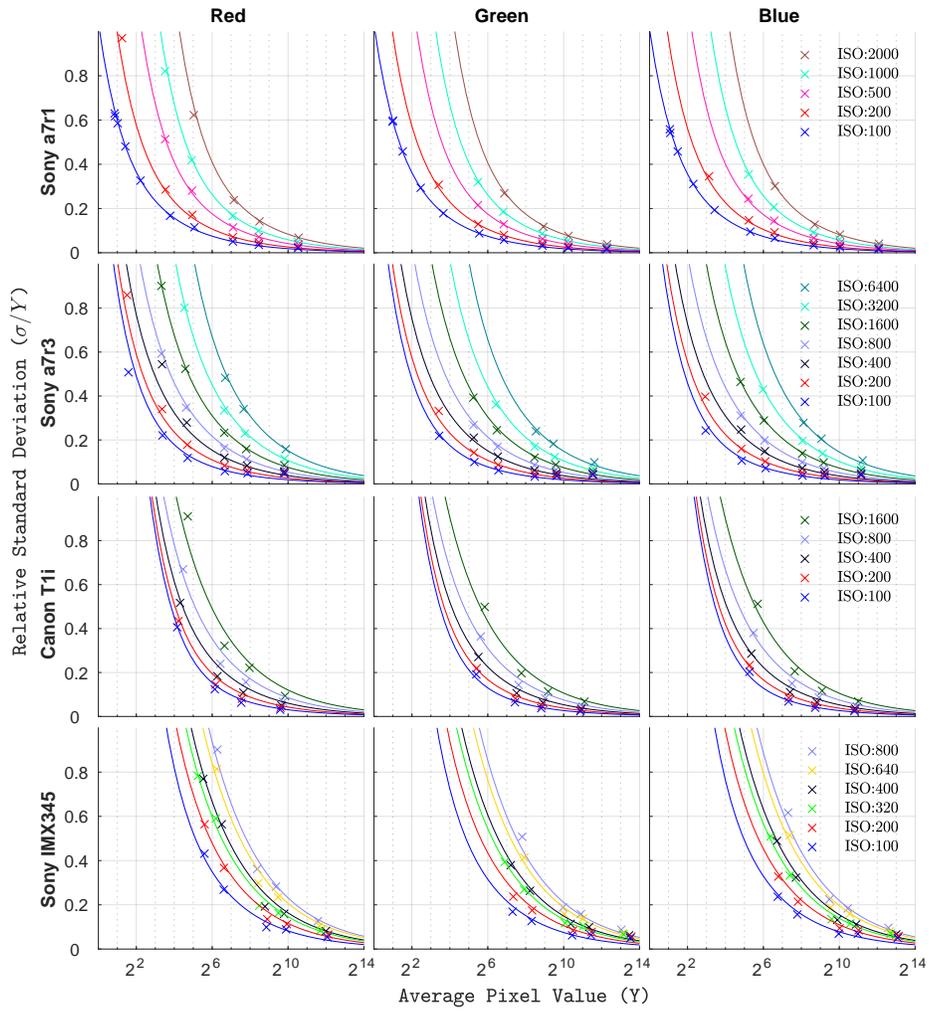
Fig. 1: Computed relative standard deviations plotted against average pixel value for all color channels of tested cameras

(a) Ground Truth

$t = \frac{1}{32}s,\ g = 32$

$t = \frac{1}{100}s,\ g = 32$

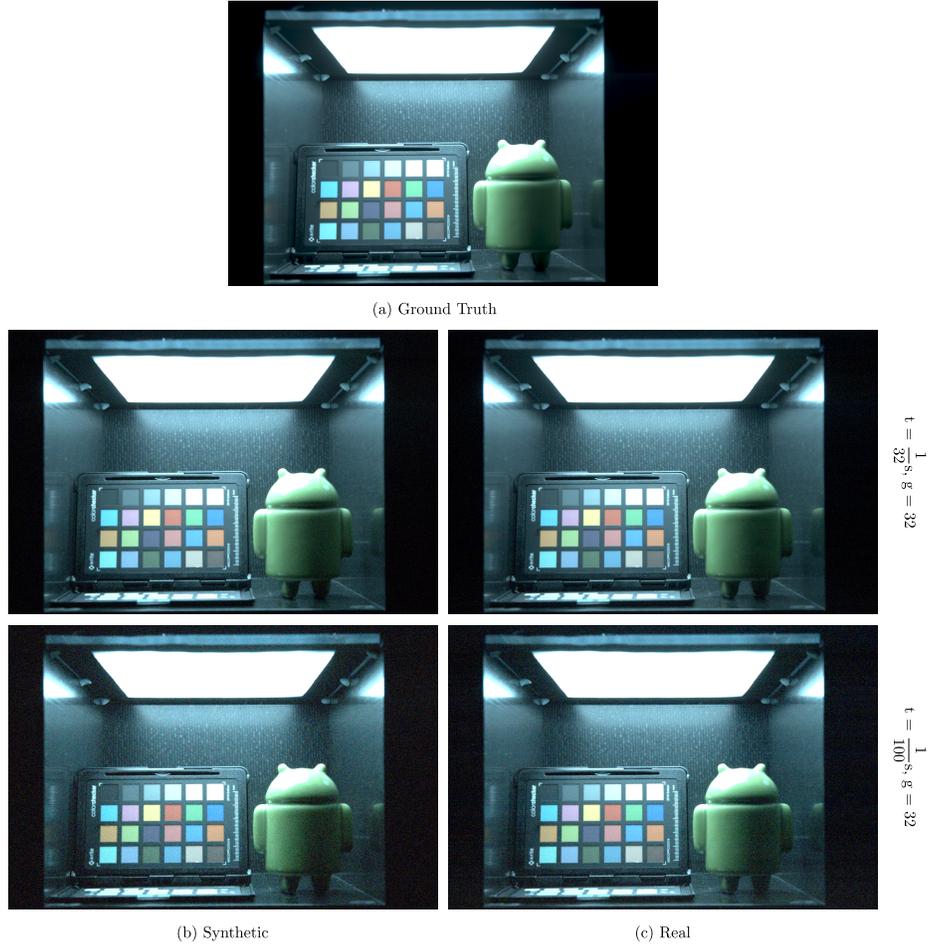(b) Synthetic   (c) Real

Fig. 2: The first image on the top (a) is the average of 100 images captured with exposure time and gain equal to 1. This is in practice a noise-free image that serves as the ground truth. The two images on the left (b) show a ground truth image with added synthetic noise (Eq. 1 in the main paper), and the images of the right (c) were captured with the camera parameters shown on the right. The images demonstrate that simulated noise is very close to real noise captured by the camera. All the images were captured using the Canon T1i and presented without white balance and using gamma correction to depict noise in RAW images.

(a) Hat-shaped   (b) NPNE   (c) Variance-weighted
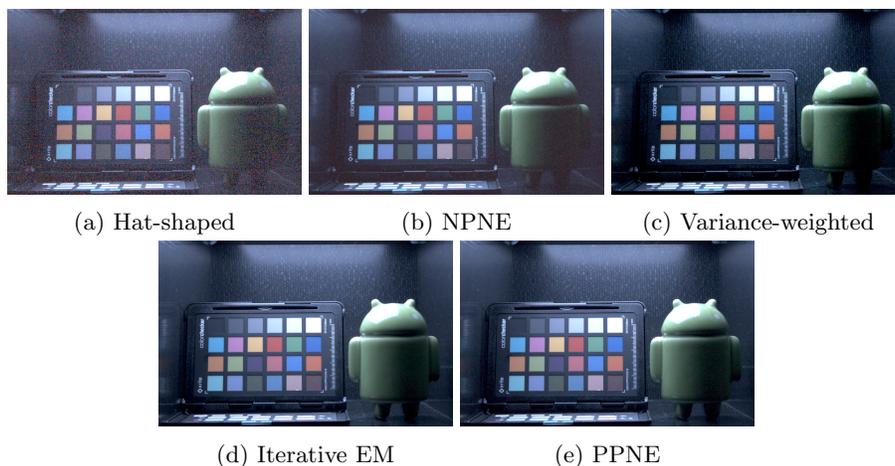
(d) Iterative EM   (e) PPNE

Fig. 3: HDR reconstructions using different estimators given an exposure stack of three images captured by the Canon T1i at ISO 3200. The images were gamma-encoded for visualization ($\gamma = 2.2$).
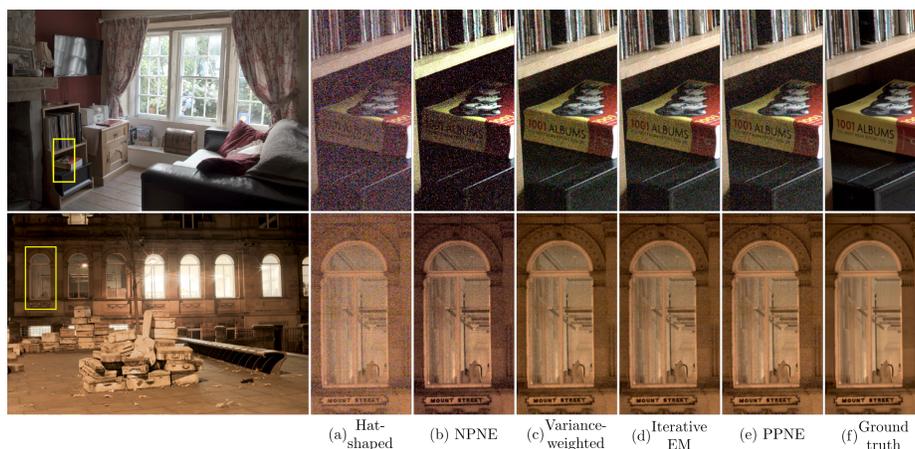

(a) Hat-shaped   (b) NPNE   (c) Variance-weighted   (d) Iterative EM   (e) PPNE   (f) Ground truth

Fig. 4: Comparison of the estimators on a few more challenging scenes with high noise. These scenes correspond to "Cottage" and "Street" in Table 2 in the main text.
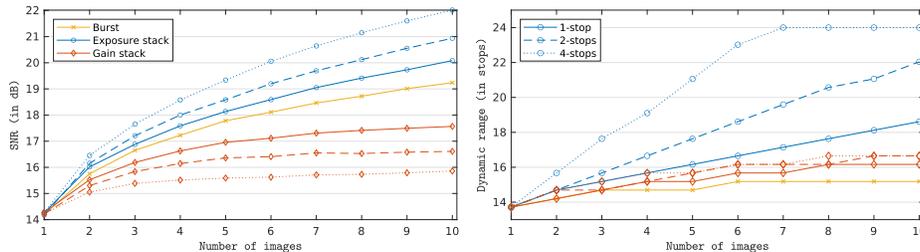
Fig. 5: Three HDR capture strategies: an exposure time stack, a gain stack and a burst of the same exposures. Different line styles represent the separation in either exposure time or gain between the images.

Appendix 3: Qualitative comparison

Here, we show a few more real scenes (Figs 3 and 4) for visual inspection. Quantitative comparison of the estimators for the "Cottage" and "Street" scenes depicted in Fig 4 are provided in Table 2. For these scenes, the ground truth is obtained by capturing five images separated by four stops with ISO 100. Subsequently, the noisy image stack for testing the estimators is captured using three images at a much higher ISO of 12800 for "Cottage" and 6400 for "Street". For the other two outdoor scenes "House" and "Trees", we could not employ such a strategy due to the unavoidable movement of leaves and clouds. Instead, we captured five images at ISO 800 and used all the images to obtain the ground truth. Then, only the three shortest exposures were used to compare the estimators.

Appendix 4: Different Capture Scenarios

In most of our examples, we used image stacks with varying exposure time. However, some cameras employ two other strategies instead: they take a stack of images with the same exposure but different gains [1], or a stack (burst) of images with the same settings [3, 2]. Both strategies share an advantage since all images have the same amount of motion blur, thus eliminating possible artifacts. The gain stack allows us to capture images with shorter exposure times and therefore less camera shake and image burst produces images of the same brightness that are easier to align. Here, we want to compare how each capture strategy compares in terms of average signal-to-noise ratio (SNR) and dynamic range.

We use the same Monte Carlo (MC) simulations as in Section 5 to quantify the increase in SNR and dynamic range with increasing number of images in a stack. We simulated the Sony α7r1 and used the *PPNE* estimator for all reconstructions. The SNR is computed as an average across the tested range of 24 stops in decibels. The dynamic range is reported as the number of stops between the highest registered radiance and the radiance at which the ratio $\sigma/Y = 1$. In Fig. 5 each scenario is shown with a different color line and the line-styles indicate the separations between exposures (1, 2 or 4 stops). The

results indicate that an exposure time stack is the best strategy to increase both the dynamic range and SNR. However, if shorter exposure times are required, the gain stack is the best way to increase dynamic range and burst is better for improving SNR.

Appendix 5: Extended plots

We conducted additional experiments in high-noise conditions and using gain modulated input stacks. These are analyzed in Sections. 5.2 and 5.1 of the main text. Extended versions of the plots are provided here. Fig. 6 splits the error for an input gain stack into bias and standard deviation. Finally, Fig. 7 depicts the effect of increasing static noise (by $2\times, 4\times$ and $8\times$) in the inputs.
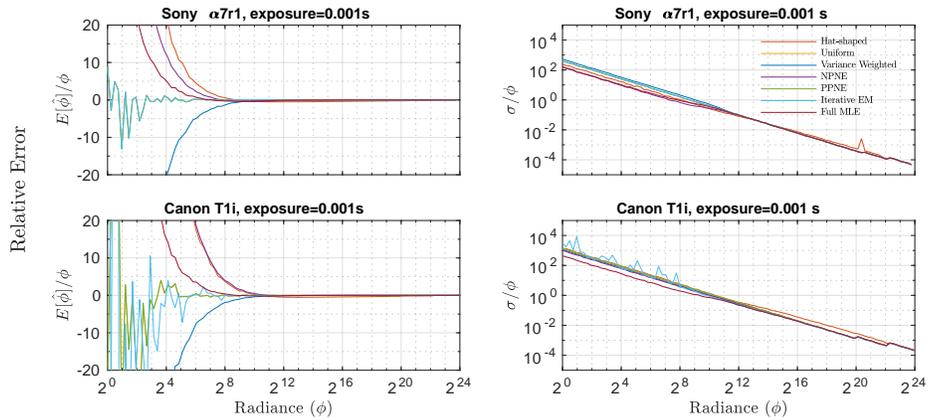

Fig. 6: Results for gain modulation: relative biases (left column) and standard deviations (right column) for gain stacks obtained using MC simulation. Gains of the inputs are logarithmically spaced between the minimum and maximum values recorded for each sensor. The minimum gain is fixed to 1, while the maximum is 32 for the Sony α7r1 and 16 for the Canon T1i.

References

1. Hajisharif, S., Kronander, J., Unger, J.: Adaptive dualiso hdr reconstruction. EURASIP Journal on Image and Video Processing **2015**(1), 41 (2015)
2. Hasinoff, S.W., Sharlet, D., Geiss, R., Adams, A., Barron, J.T., Kainz, F., Chen, J., Levoy, M.: Burst photography for high dynamic range and low-light imaging on mobile cameras. ACM Transactions on Graphics (TOG) **35**(6), 192 (2016)
3. Heide, F., Steinberger, M., Tsai, Y.T., Rouf, M., Pajak, D., Reddy, D., Gallo, O., Liu, J., Heidrich, W., Egiazarian, K., et al.: Flexisp: A flexible camera image processing framework. ACM Transactions on Graphics (TOG) **33**(6), 1–13 (2014)
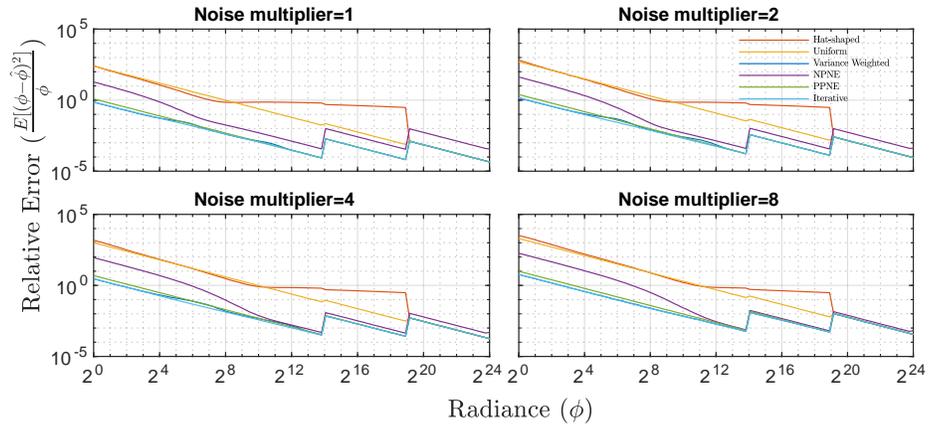
Fig. 7: Results with high-noise: observed relative errors for MC simulation of Sony α7r3 for the logarithmic input gradient. The combined effect of bias and standard deviation is plotted for all the analytical estimators. The deviation observed for the *hat-shaped* estimator is due to the negative bias visible in the earlier plots (Fig. 3 and Fig. 6).