Robust estimation of exposure ratios in multi-exposure image stacks

Param Hanji*, Rafał K. Mantiuk†

Department of Computer Science and Technology, University of Cambridge

Email: {*pmh64,†rkm38}@cam.ac.uk

Abstract—Merging multi-exposure image stacks into a high dynamic range (HDR) image requires knowledge of accurate exposure times. When exposure times are inaccurate, for example, when they are extracted from a camera's EXIF metadata, the reconstructed HDR images reveal banding artifacts at smooth gradients. To remedy this, we propose to estimate exposure ratios directly from the input images. We derive the exposure time estimation as an optimization problem, in which pixels are selected from pairs of exposures to minimize estimation error caused by camera noise. When pixel values are represented in the logarithmic domain, the problem can be solved efficiently using a linear solver. We demonstrate that the estimation can be easily made robust to pixel misalignment caused by camera or object motion by collecting pixels from multiple spatial tiles. The proposed automatic exposure estimation and alignment eliminates banding artifacts in popular datasets and is essential for applications that require physically accurate reconstructions, such as measuring the modulation transfer function of a display. The code for the method is available.

Index Terms—High dynamic range imaging; camera noise model, statistical estimation, multi-exposure fusion

I. INTRODUCTION

WHEN the dynamic range of a scene exceeds the operating range of a standard digital sensor, one can overcome this limitation by capturing a stack of images with different camera settings: modulating exposure times [13], sensor gain [16], [22] or by capturing and averaging a burst of images [20]. Then, the captured exposure stack can be merged into a single image to both expand the dynamic range and reduce noise.

Regardless of the approach, regions of the reconstructed high dynamic range (HDR) images that contain smooth intensity gradients often end up with banding artifacts, such as those shown in Figure 1. Banding artifacts are highly visible in both images and videos, and they have been the subject of extensive research, particularly in the field of video streaming. Banding tends to be more noticeable compared to other video compression artifacts [27], [40], [44]. The reason for banding in merged exposure stacks is a mismatch between the actual capture parameters and those reported by the camera, typically

Fig. 1. The first row depicts patches from images of overcast skies taken from large HDR datasets [10], [18]. Misalignment between the reported camera settings results in banding artifacts (pointed at by the red arrows) at the boundary between two exposures. This problem is not limited to the sky and affects other regions too, as shown by patches in the second row. We carefully picked exposure stacks consisting of RAW spatially-aligned images that do not require CRF inversion or motion compensation.

in the exchangeable image file format (EXIF) header. Such inaccuracies could be caused by

- limited accuracy of the (mechanical) aperture and shutter
- wrongly reported EXIF data due to rounding (e.g. 1/60 exposure time could be reported instead of 1/64)
- changes in scene illumination, for example, due to flickering lights or overcast skies with moving clouds

We stress that the artifacts shown in Figure 1 are not due to spatial misalignment because we carefully selected image stacks without camera or object motion. Further, since we merged linear RAW images, the artifacts are not due to incorrect camera response function (CRF) inversion.

The inaccuracies of exposure time found in the EXIF metadata are substantial, reaching 40%, as shown in Figure 2. To prepare this figure, we computed the relative error between the reported EXIF exposure times and the times estimated by the proposed method. The two plots show similar error distribution for two datasets of multi-exposure HDR image stacks [10], [18]. The errors of such magnitude can easily introduce banding artifacts, as shown in Figure 1.

In this work, we propose to estimate exposure ratios directly from a stack of images, which lets us eliminate the banding artifacts from the merged HDR images. We formulate the problem to be efficiently solved as a sparse linear system
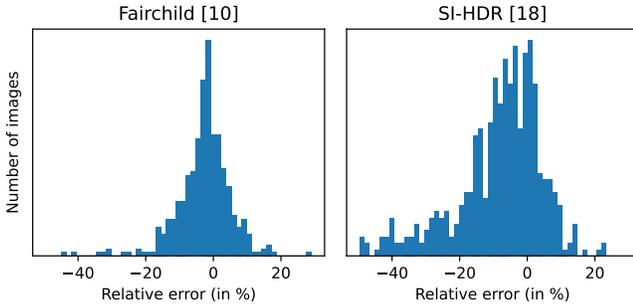
Fig. 2. Histograms of relative errors for two popular HDR datasets. The computed sample standard deviations of relative errors are 9% for the Fairchild survey [10] and 14% for the SI-HDR dataset [18].

while accounting for heteroskedastic camera noise with known or unknown noise parameters. Thus, our exposure estimation can be used to improve ground truth reconstructions of several existing stack-based HDR datasets. Further, algorithms that enhance merged reconstructions by iterative optimization [29], as well as deep neural networks that use HDR images for other related tasks [6], [46] stand to be improved by our more accurate exposures.

Correctly estimating exposures is also relevant for HDR merging with deghosting, which involves detecting and registering pixels that belong to moving objects [32], [41]. Erroneous capture parameters make this task more challenging, as pixel differences could be caused by either object movement or incorrect exposure time.

Although banding artifacts are visible only in smooth image regions in Figure 1, inaccurate exposures affect all pixels in the image. Incorrect pixel intensities pose a serious complication for applications that utilize cameras in place of expensive light measurement instruments. Multi-image exposure stacks often serve as substitutes for more accurate measurements from instruments such as spectrophotometers. In Section IV-D, we highlight potential discrepancies in the measured modulation transfer function (MTF) of an HDR display when the exposure stack of a slant-edge [36] is merged with incorrectly reported capture parameters.

From a practical standpoint, our method is suitable for large images (8k or more) from modern cameras. We achieve this by solving a smaller system of equations by collecting pixel-pairs with the lowest relative variances according to the noise model we consider. Contrary to existing methods that utilize pixel pairs from consecutive exposures in the stack, we introduce a greedy algorithm that provides optimal pixel pairs to ensure that all exposures are well-estimated. Our reduced linear system is a union of the highest weighted spanning trees induced on an *exposure multigraph*, resulting in a balanced linear system. Additionally, our proposed method is spatially balanced because we split the input image into tiles and collect pixels from all tiles. This improves the robustness of our estimator to ghosting caused by camera or object motion.

Here is a brief overview of the paper that summarizes our contributions. We highlight that camera metadata may be unreliable and motivate the need to estimate corrective per-exposure ratios to obtain artifact-free HDR reconstruction. We show that such exposure ratios can be estimated by solving a large linear system of equations (Section III-C), where each equation connects pixel intensities in the logarithmic domain. To deal with underexposed pixels in shorter exposures, we model heteroskedastic camera noise with inverse-variance weights (Section III-D). Then, in Section III-F, we show how to reduce the system of equations for faster equations without sacrificing the estimation quality. We finally validate our exposure estimation framework in Section IV, both on synthetic and real image captures.

II. RELATED WORK

The problem of inaccurate capture parameters was identified in very early works in HDR merging [7], [30]. However, most of these focused on the challenging task of inverting the CRF under the assumption of film or sensor *reciprocity* [38]. Mitsunaga and Nayar [30] used a polynomial model to jointly estimate the CRF and exposure times. Then, Grossberg and Nayar [15] demonstrated how to recover the brightness transfer function (BTF), a function that describes how the brightness transfers from one image to another. They recovered the BTF from image histograms and stipulated that it can be used to estimate exposure ratios if the CRF is known. More recently, Rodríguez at al. have shown that the previously mentioned methods rely on incorrect assumptions [35] about the independence of color channels and linearity of exposures, resulting in estimation errors and hue shifts. We avoid those problems by directly operating on demosaiced RAW images.

Based on [15], Cerman and Hlavac [3] assumed a linear CRF by relying on RAW pixel values. They computed the brightness-transferred image histograms and used them to weight a system of equations in the linear pixel domain.

In contrast to their work, we solve a weighted linear system in the logarithmic domain to estimate exposure ratios. The weights in our system of equations are derived from a popular statistical camera noise model and ensure a noise-optimal solution. Our approach systematically handles heteroskedastic camera measurements and provides accurate estimates even when many pixels are under-exposed or affected by noise.

A. Camera noise model

A key contribution of our work is the use of a parametric noise model to weigh some pixel correspondences more than others. Our weights are based on the popular Poisson-normal statistical camera noise model [2], [12], [28] that has signal-dependent and static (or signal-independent) components. Many earlier works have used approximations of similar noise models for noise-optimal HDR reconstructions [14], [19], [21]. Although deep generative networks [1], [4] model spatially-varying components of real camera noise better, the Poisson-normal statistical model and its normal approximation are better suited for our problem as they offer a convenient algebraic form under the assumption that noise is independent at each pixel.

Code for the method has been integrated into the software for noise-optimal HDR merging (*HDRutils*) and can be found at https://github.com/gfxdisp/HDRutils.

B. HDR datasets

Merged HDR images of many multi-exposure datasets [10], [17], [18], [24], [25], [32] can be improved with accurate exposure estimation. We observed the banding artifacts depicted in Figure 1 in various images from these datasets. Recent HDR deep learning reconstruction methods for tasks like HDR deghosting [5], [24], [33], [34], [47] and inverse tone-mapping [8], [37], [43] utilize these as well as other multi-exposure datasets for testing and evaluation. All these works are likely to produce better results when trained with accurate ground truth information due to our work on better exposure alignment.

III. METHODOLOGY

After introducing the camera model and some terminology, we describe how to estimate exposure ratios directly from the input image stack by solving a weighted linear system. We then derive noise-optimal weights for the system based on the widely-used Poisson-normal camera noise model. Finally, we discuss practical considerations for making the system agnostic to sensitive noise parameters and improve robustness to ghosting caused by camera or object motion.

A. Camera model

To digitally represent an HDR scene, we start by capturing N images with varying exposure times or gains (ISO). Although earlier works included CRF estimation in their image formation pipelines [7], [15], [30], we skip this step since modern cameras provide access to linear RAW pixel values. We model the captured RAW pixels as samples from independent random variables, which are linearly related to the scene radiance, and denote them as:

$$Y_i(p), \quad i = 1 \ldots N, \quad p = 1 \ldots M, \tag{1}$$

where i is the exposure index with N total exposures, and p is the pixel index with M total pixels. Throughout this work, we use upper case for random variables and lower case for observed values. Before merging (averaging) the exposure stack, we need to convert them to relative radiance units by compensating for exposure time t_i, gain g_i, and aperture f-number a_i:

$$X_i(p) = \frac{Y_i(p)}{t_i \, g_i \, \pi \left(\frac{f}{2 \, a_i} \right)^2} = \frac{Y_i(p)}{d_i}, \tag{2}$$

where f is the focal length. Typically, the scaling constant, $d_i = t_i \, g_i \, \pi \left(\frac{f}{2 \, a_i} \right)^2$, is computed directly from the camera EXIF header. The problem is that t_i, g_i, and a_i could be incorrect due to the limited accuracy of the mechanical shutter or the rounding of exposure values. Incorrect values of d_i will lead to inaccurate HDR reconstructions Thus, samples from $X_i(p)$, the exposure-compensated or absolute estimates in the captured images, may represent biased measurements of the true scene radiance. Merging images using the inaccurately reported parameters results in the banding artifacts depicted in Figure 1 and Figure 4. In all the images, the artifacts appear when a longer exposure image saturates.
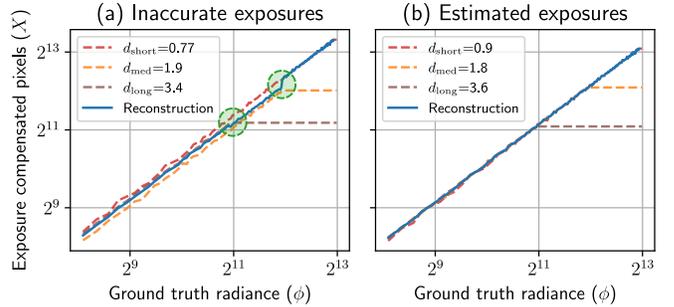
Fig. 3. A linear gradient spanning 13 stops is captured with exposure times $[0.5, 2, 4]$ (in seconds) by simulating the *Sony ILCE-7R* at ISO 3200. Each capture is quantized to 8 bits for easy visualization. The left plot depicts exposure-compensated pixels that represent samples from $X(p)$ (according to Eq. (2)) with inaccurate exposures (red, yellow, and brown dashed lines). The reconstructed gradient (blue line) is jagged around $\phi = 2^{11}$ and $\phi = 2^{12}$ (green circles) due to misaligned exposures caused by biased exposure values. We can reconstruct the smooth gradient by correctly aligning exposure ratios, as shown in the right plot.

B. Banding due to inaccurate exposure

To better understand the reason for banding, consider the one-dimensional linear gradient depicted in Figure 3 (left). The noisy measurements (dashed lines), obtained by simulating captures using calibrated noise parameters of the *Sony ILCE-7R*, become misaligned when scaled with inaccurate exposure times. Merging such images results in a jagged reconstruction (solid blue line), causing banding in an otherwise smooth output. Notice that although the reconstruction deviates from the ground truth for almost all pixels, artifacts will be visible only at transition points when an image in the stack saturates as highlighted by the green circles.

While Figure 3 demonstrates the problem with simple averaging, banding is further exaggerated when sophisticated algorithms based on physical noise models [14], [19], [21] are used. This is because pixels of longer exposures are more reliable (due to smaller noise variance), but they saturate at lower physical values. Thus, these estimators weigh longer exposure pixels more, and there is a sharp change upon saturation of any image in the stack.

If we are able to estimate relative exposure ratios w.r.t one of the images correctly, we can align all exposures to produce the reconstruction in Figure 3 (right). Note that the exposure-aligned reconstruction may still not coincide with the ground truth since we do not have the correct baseline. However, accurately estimating relative ratios is sufficient to align the capture parameters and eliminate banding.

C. Exposure estimation in real images

To prevent banding and obtain physically accurate pixel values, we align the exposures of all images in the stack by estimating all scaling constants d_i. However, computing them directly from input pixels is impossible since the correct absolute measurements representing observations of $X(p)$ from Eq. (2) are unknown. We thus eliminate these unknowns by estimating the ratio of exposures between any two images in the stack instead. For a scene with constant illumination and

no scene motion, this ratio should be the same for all pixels from a given pair of images. Its expected value is:

$$d_{ij} = \mathbb{E}_p \left[\frac{Y_i(p)}{Y_j(p)} \right] , \qquad (3)$$

where i and j index different images in the exposure stack, and $\mathbb{E}_p$ indicates that the expected value is computed over all pixels. To allow for fast computation using linear solvers, we operate on logarithmic values. Thus, let

$$e_{ij} = \log d_{ij} \quad \text{and } L_i(p) = \log Y_i(p) . \qquad (4)$$

Although we cannot write a closed-form expression for the density function of $L_i(p)$ because of the log transformation, it is possible to approximate the expected value of any transformed random variable using its Taylor expansion. Our results, detailed in Eq. (16) in the Appendix, are only applicable to normally distributed random variables. It is thus imperative to use only well-exposed pixels (we show how to do this in Section III-F), since the Poisson photon noise component of such pixels is well-approximated by a normal distribution. We can obtain an approximate expression for the expected value by applying the result for $\log Y(p)$:

$$\begin{aligned} e_{ij} = \log \mathbb{E}_p \left[\frac{Y_i(p)}{Y_j(p)} \right] &\approx \mathbb{E}_p \left[\log \frac{Y_i(p)}{Y_j(p)} \right] \\ &= \mathbb{E}_p \left[L_i(p) - L_j(p) \right] . \end{aligned} \qquad (5)$$

The equality approximately holds for the operating range of pixel values, as we will detail in Section III-D, Eq. (11). This allows us to compute the expected value over all pixels by setting up and solving a linear system. For the most reliable estimate, we should utilize information from all the available exposures (all possible values of i and j). Unlike previous work [3], we thus consider not only ratios between neighboring exposures but between all pairs. This results in a large yet sparse linear system:

$$\sqrt{\boldsymbol{W}} \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix} = \sqrt{\boldsymbol{W}} \begin{bmatrix} L_1(1) - L_2(1) \\ L_1(2) - L_2(2) \\ \vdots \\ L_1(1) - L_3(1) \\ \vdots \\ L_{N-1}(M) - L_N(M) \end{bmatrix} , \qquad (6)$$

or more compactly,

$$\sqrt{\boldsymbol{W}} \boldsymbol{O} \boldsymbol{e} = \sqrt{\boldsymbol{W}} \boldsymbol{m} , \qquad (7)$$

where $\boldsymbol{W}$ is a diagonal weight matrix denoting the relative importance of each row of the system. In practice, we found that the weighted system does not always provide a good solution for shorter exposures. Since a rough estimate of exposure values $\boldsymbol{e}_0$ is available in the image metadata, we introduce a Tikhonov penalty with weight λ and solve to get:

$$\begin{aligned} \hat{\boldsymbol{e}}_{\text{WLS}} &= \arg\min_{\boldsymbol{e}} \left\| \sqrt{\boldsymbol{W}} (\boldsymbol{O}\boldsymbol{e} - \boldsymbol{m}) \right\|_2^2 + \lambda \|\boldsymbol{e} - \boldsymbol{e}_0\|_2^2 \\ &= (\boldsymbol{O}^\mathsf{T} \boldsymbol{W} \boldsymbol{O} + \lambda \boldsymbol{I})^{-1} (\boldsymbol{O}^\mathsf{T} \boldsymbol{W} \boldsymbol{m} + \lambda \boldsymbol{e}_0) . \end{aligned} \qquad (8)$$

To demonstrate the effectiveness of exposure alignment, we captured an exposure stack of a simple HDR scene consisting of a bright light shining at an angle, which produces an
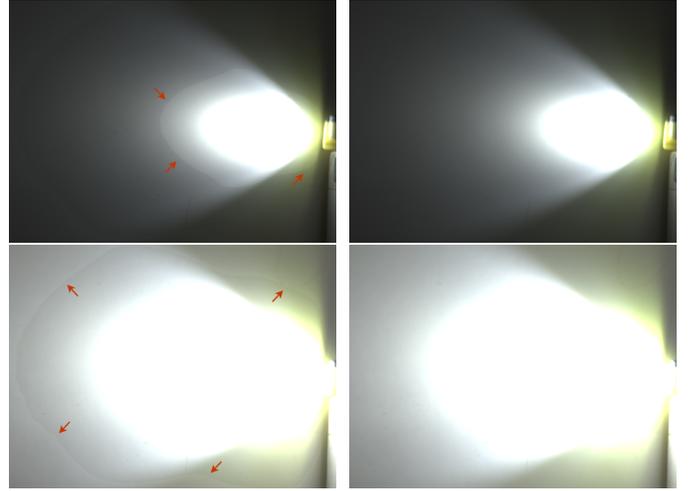
(a) Parameters from EXIF   (b) Estimated parameters

Fig. 4. The left column shows exposures (gamma-encoded for visualization, $\gamma = 2.2$) of the HDR image reconstructed with EXIF parameters, while the right column shows exposures (encoded with $\gamma = 2.2$) of the HDR image reconstructed by solving the linear system represented by Eq. (6) with $\boldsymbol{W} = \boldsymbol{I}$. Banding artifacts are visible in dark (top row) and bright (bottom row) regions of the reconstruction using EXIF metadata (a). Aligning exposures according to Eq. (8) fixes the problem (b).

inverse-square fall-off in intensity. Since most pixels are well-exposed, as shown in Figure 4, we could assume that they are equally reliable and set $\boldsymbol{W} = \boldsymbol{I}$. This works well for the carefully controlled scene and eliminates banding artifacts that would have appeared when merging with EXIF parameters. However, the constant noise assumption breaks down for real-world HDR stacks since the noise in camera pixels is heteroskedastic [11] and thus, different pixels provide different amounts of information.

D. Heteroskedastic pixels

To determine noise-optimal weights for images of real cameras, we need to derive an expression for the variance of each row of Eq. (6). Then, we populate $\boldsymbol{W}$ with inverse-variance weights (i.e., each weight is given by the reciprocal of variance) to obtain the weighted least-square estimate $\hat{\boldsymbol{e}}_{\text{WLS}}$. This is equivalent to maximum likelihood estimation (MLE) under the assumption that the system of equations models additive, normally distributed noise, where the variance is different for different rows. Since each entry of the output vector $\boldsymbol{m}$ is a difference of two random variables, the inverse-variance weight for each row of the system is given by:

$$w_{k,k} = \frac{1}{\mathbb{V}[L_i(p) - L_j(p)]} = \frac{1}{\mathbb{V}[L_i(p)] + \mathbb{V}[L_j(p)]} . \qquad (9)$$

Here, k indexes each row of Eq. (6) since $\boldsymbol{W}$ is diagonal. For instance, the second row corresponds to $k = 2, i = 1, j = 2$.

In order to get an expression for the denominator in Eq. (9), we refer to detailed studies of the noise characteristics of cameras [2], [28], which indicate that real camera noise follows a compound Poisson-normal distribution. For the working range of pixels in most images, this can be approximated by zero-mean additive noise that follows a normal distribution. Temporarily dropping the exposure index for brevity of notation,

the variance at each pixel thus consists of a signal-dependent component as well as a static component and is equal to,

$$\sigma_Y^2(p) = \mathbb{V}[Y(p)] = \alpha\,\mathbb{E}[Y(p)] + \beta\,, \tag{10}$$

where α and β are camera-specific noise parameters, which also depend on the sensor's gain.

Since $L(p)$ is a random variable obtained by applying the $\log$ transformation to $Y(p)$, its density function does not have an exact expression. This is because the domain of $Y(p)$ includes negative values for which the $\log$ function is undefined. As we will show in Section III-F, we operate on a small subset of available pixels, selected for the lowest relative variances, and whose intensities tend to be much greater than 0. We show, in Eq. (17) in the Appendix, how to approximate the variance when any random variable is transformed by an invertible function. Here, we apply the result for $Y(p)$ when it is transformed by the $\log$ function,

$$\begin{aligned} \mathbb{E}[L(p)] &\approx \log\mu_Y(p)\,, \\ \mathbb{V}[L(p)] &\approx \frac{\sigma_Y^2(p)}{\mu_Y^2(p)}\,, \end{aligned} \tag{11}$$

where $\sigma_Y^2(p)$ is the variance of the noise from Eq. (10) and $\mu_Y^2(p)$ is the expected pixel value, which we approximate by $\mathbb{E}[Y(p)] = \mu_Y(p) \approx y(p)$. We reiterate that we are able to use this result because we work with well-exposed pixels, that are approximately normal. After substituting the computed variance in Eq. (9), the diagonal weights become:

$$w_{k,k} = \left(\frac{\alpha y_i(p) + \beta}{y_i^2(p)} + \frac{\alpha y_j(p) + \beta}{y_j^2(p)}\right)^{-1}. \tag{12}$$

E. Camera noise calibration

While the inverse-variance weights in Eq. (12) help us compute noise-optimal estimates of the exposure ratios, a fundamental limitation is their dependence on calibrated noise parameters α and β. In the considered noise model [12], these are camera and gain specific and may not be available at the time of HDR merging. Moreover, the quality of HDR reconstructions is highly sensitive to accurate noise parameters [2], motivating the need for methods that do not rely on them.

If noise parameters are unavailable or inaccurate, it is still possible to solve the weighted linear system by assuming that the static noise parameter β is 0 as we work with well-exposed pixels (see Section III-F). Since the entries of $\boldsymbol{W}$ determine the relative importance of the different rows of Eq. (6), we can eliminate the common signal-dependent constant, α as well. The new weights which no longer depend on calibration-sensitive parameters are:

$$w'_{k,k} = \left(\frac{1}{y_i(p)} + \frac{1}{y_j(p)}\right)^{-1}. \tag{13}$$

In practice, we can get away with using camera-independent $\boldsymbol{W'}$ instead of camera-specific $\boldsymbol{W}$ from Eq. (12).

F. Reducing the linear system

The linear system given by Eq. (6) contains up to $M\binom{N}{2}$ equations corresponding to all pixels (M) and all pairs of exposures in the stack (N). Solving such a large system may be impossible for large images or deep exposure stacks due to computation and memory limitations. However, this system is strongly overdetermined as only $N-1$ exposure ratios need to be estimated. Therefore, we only need a small percentage of equations to solve Eq. (8). Another reason for reducing the system is to eliminate underexposed pixels pairs because they do not satisfy the assumptions made in the derivations of previous sections. For example, dark pixels may be negative or poorly approximated by a normal distribution.

A logical reduction strategy is to select pixels pairs with the highest weights since they are least affected by noise. However, such a strategy is heavily biased towards longer exposures because pixel pairs that include these images will have the highest weights. The shorter exposures will then be poorly represented and, thus, poorly estimated. Another issue is related to the spatial location of bright objects in the scene, such as the Sun or other light sources. If we select a small fraction of pixels (say 5% of the pixels per image), all of them are likely to be concentrated in one portion of the image, corresponding to these bright objects. If those objects happen to be in motion, the exposure estimation will fail.

We propose two orthogonal design choices to balance the system of equations and handle both these biases in estimated exposure ratios.

1) Spatial balance: Tiling: To ensure that the linear system contains samples from all parts of a scene, we split the input image stack into $t \times t$ pixel tiles (16×16 in our experiments). We can then select a fixed number of pixel pairs from each tile and pixels from a few bright objects will not dominate the system of equations.

The tiled processing provides a convenient way to vectorize the construction of the reduced linear system. Several tiles can be processed in parallel for faster execution with multi-core or multi-threaded systems.

Our noise-based solution helps provide a robust estimation even though some tiles may contain noisy pixels corresponding to dimly-lit parts of the scene. Variance-optimal balancing of exposures is crucial for such tiles, as we will show in the next section.

2) Exposures balance: Spanning trees: The next objective is guaranteeing that all $N-1$ exposure ratios are correctly estimated. Within each spatial tile, we need to include pixels from all exposure pairs, including noisy short exposures despite their relatively smaller weights.

Consider the *exposure graph*: an undirected weighted multi-graph (a graph that contains more than one edge between two vertices as shown in Figure 5a), whose vertices represent the N exposures and edges link pairs of co-located pixels from two exposures. Each edge corresponds to one row of Eq. (6), with its weight given by Eq. (12) or Eq. (13). The different colored vertices and edges in Figure 5a represent different pixel locations. Reducing the linear system from Eq. (6) is then equivalent to removing edges from this dense multigraph.

Fig. 5. Consider the exposure multigraph in (a) where different colored vertices and edges represent different pixels. The MST may connect pairwise exposures (b) or connect the longest exposure to all others (c). Algorithm 1 will provide different solutions for different inputs. For example, the MST for the blue horizontal tile near the Sun in (d) is the pairwise-connected spanning tree connected at pixels marked by the two triangles in plot (e). However, for the orange tile, the MST is given by the longest-exposure spanning tree at the pixel marked by the star in plot (e).

The optimal subset contains $(N-1)k$ edges with the largest weights that connect all exposures in a balanced manner. Such a subset can be found by computing the k highest-weighted spanning trees of the multigraph, where the weight of a spanning tree is the sum of the weights of its edges. The previous work directly linked pixels from neighboring exposures [3], resulting in pairwise connectivity as shown in Figure 5b. However, this solution is sub-optimal for some inputs (such as the orange tile in Figure 5d) because using edges linked to the longest exposures, as shown in Figure 5c), results in higher weights and, therefore better estimates.

An optimal solution would sequentially extract k MSTs using *Kruskal's* or *Prim's* algorithm. Better algorithms [9], [26] extract the k highest spanning trees more efficiently. However, explicitly creating a large multigraph and computing MSTs is both memory and computationally expensive. Below we show that the optimal solution can be found by a simpler greedy algorithm.

Algorithm 1: Greedy MST algorithm

input : $y_1, \ldots, y_N$ (Images sorted by exposure time)
output: MST (Maximimum spanning Tree)

1 Function greedyMST($y[\ ]$)
2 $N \leftarrow length(y)$
3   MST $\leftarrow [\ ]$
4   **for** $i \leftarrow 1$ *to* $N-1$ **do**
       /* Iterate over all exposures starting from the shortest */
5 $mask \leftarrow$ isValid(y_i) and isValid(y_{i+1})
6 $p_* \leftarrow$ maxWeight($y_i[mask], y_{i+1}[mask]$)
       // Location of the highest weighted edge connecting images y_i and y_{i+1}
7   **for** $j \leftarrow N$ *to* $i+1$ **do**
8   **if** isValid($y_j(p_*)$) **then**
9   MST.addEdge(i ●—● j)
10   break

11   **return** *MST*

Greedy MST solution: We start by extracting all valid pixels — those that are unsaturated and sufficiently above the noise floor. We iterate through all exposures from the shortest to the longest. For each exposure i, we identify p_*, the pixel location that contains the highest-weighted edge between exposures i and $i+1$. This is typically the brightest pixel that is not saturated in exposure $i+1$. Then, we find the longest exposure $j > i$ in which pixel p_* is not saturated. We add an edge between i and j. By selecting the longest exposure, we ensure that the weight of the edge is maximized.

This procedure, summarized in Algorithm 1, can be repeated k times to extract the k highest-weighted spanning trees without explicitly constructing a graph. If all longer exposures $y_N(p_*), y_{N-1}(p_*), \ldots y_{i+2}(p_*)$ are invalid due to saturation, the solution reduces to pairwise connectivity depicted in Figure 5b.

Note that while a pair of pixels forming an edge must have the same position, each edge in the MST can come from a different pixel position. Consider the 1-dimensional blue tile in the tonemapped image in Figure 5d and its horizontal scanline in Figure 5e. The MST is given by two edges: an edge between the medium and long exposures and an edge between the medium and short exposures (marked by triangles in the plot). Note that these are the brightest unsaturated pixels in the long and medium exposures. The MST for the blue tile is thus the pairwise connected spanning tree. The limited dynamic range of the orange tile means that a single pixel, marked by the star in Figure 5e, aligns all three exposures for the orange tile, resulting in the spanning tree in which the longest exposure is linked to all other exposures.

G. Handling outliers and pixel misalignment

The weighted least-squares solution Eq. (8) using rows of the reduced system will provide an accurate estimate only if there is no movement in the scene and all pixels are well aligned. If there is motion across exposures, for example, due to camera shake or object motion, those pairs of pixels will affect our estimates of exposure ratios, as shown in Figure 6. Notice the artifacts surrounding bright regions such as the windows and red lights in Figure 6a due to camera motion. Here, we estimated the exposures using the original stack and then registered the images with homography alignment [39] before merging. Thus, any visible artifacts are due to incorrect exposure alignment and not due to ghosting.

(a) Regular $\hat{e}_{\mathbf{WLS}}$ (b) $\hat{e}_{\mathbf{WLS}}$ with outlier removal

Fig. 6. For an exposure stack containing some spatially misaligned pixels, solving the weighted linear system Eq. (7) results in inaccurate estimates. In this scene, camera motion causes pixels of the building to be misaligned due to camera motion, producing artifacts near the windows in Figure 6a. However, pixels from other areas, such as the sky, can still be utilized for exposure estimation. Figure 6b demonstrates the advantage of using the outlier-removal procedure described in Section III-G to improve robustness.

The spatial tiling described in Section III-F1 ensures that motion in a few bright objects does not adversely affect the estimated exposures, i.e., our procedure is robust to localized object motion. In case of widespread pixel misalignment (e.g., due to camera motion), we can still provide reasonable exposure estimates if the scene contains uniform image patches. We identify *usable* tiles by collecting k MSTs within a tile and solving the smaller per-tile system of equations separately. If the predictions of a tile deviate too much from the EXIF values, we treat the tile as an outlier and do not include the corresponding MSTs in the final system.

The added overhead of solving per-tile linear systems is negligible because each system contains only k equations ($k = 50$ in our experiments). This outlier removal is much faster than an iterative algorithm such as the one described in [3].

IV. RESULTS AND APPLICATIONS

The primary application of our method is merging exposure stacks with varying exposure times or gains. We first validate our results on synthetically generated stacks, for which we know the ground truth, and then show qualitative comparisons on real captures. Finally, in Section IV-D, we show how our method can improve the estimate of the MTF of a display. All our results use $\lambda = 10$ for the Tikhonov penalty term described in Eq. (8). For merging exposures stacks after exposure estimation, we used a noise-aware HDR estimator [19]. However, our results hold for other simpler methods too.
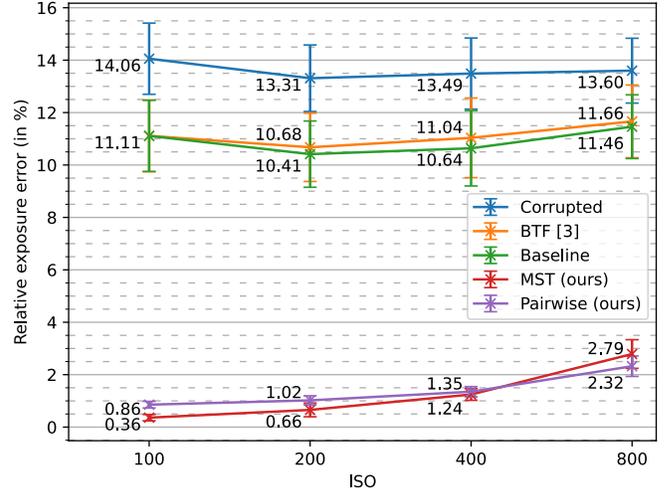
Fig. 7. Accuracies of exposure estimation methods with increasing ISO computed by simulating captures on an existing HDR dataset [10]. Error bars indicate 95% confidence intervals computed across all images. The blue line represents corrupted exposures according to Eq. (14), while the baseline, in green, was obtained by directly estimating exposures by solving a system based on Eq. (3).

TABLE I
NOISE PARAMETERS MEASURED FROM THE *Canon PowerShot S100* AT VARIOUS ISOS. THESE PARAMETERS HAVE BEEN CALCULATED AFTER NORMALIZING THE CAMERA SENSOR VALUES TO A RANGE OF 0 TO 1.

ISO	α_{R}	α_{G}	α_{B}	β_{R}	β_{G}	β_{B}
100	2.46e-5	1.67e-5	7.41e-5	3.58e-8	2.13e-8	1.28e-7
200	4.57e-5	3.02e-5	1.32e-4	9.89e-8	6.07e-8	2.66e-7
400	9.12e-5	5.95e-5	2.59e-4	2.21e-7	1.72e-7	5.61e-7
800	1.85e-4	1.19e-4	5.26e-4	4.94e-7	4.28e-7	1.14e-6

A. Validation on synthetic dataset

First, we rely on synthetic exposure stacks to compare the accuracy of different methods for exposure estimation. We simulated HDR exposure stacks using the noise parameters of the *Canon PowerShot S100* with exposure times [1/64, 1/8, 1, 8] (in seconds). All images were quantized to 14 bits to match the bit-depth of the camera, and the experiment was repeated for ISO settings between 100 and 800. We simulated noise according to Eq. (10), with noise parameters listed in Table I.

Source HDR images were taken from the Fairchild photographic survey [10] containing 105 scenes at a resolution of 4312×2868. All simulated captures are linearly related to the HDR image, i.e., we do not apply a CRF since it is likely to introduce artifacts.

Before merging the images, we corrupted the exposure times by introducing a small amount of normally distributed noise:

$$e'_i = e_i + \eta_i \quad \text{where } \eta_i \sim \mathcal{N}(0, 0.15\, e_i). \quad (14)$$

The relative standard deviation factor 0.15 was selected to match real camera EXIF errors plotted in Figure 2.

Figure 7 plots the relative root-mean-squared error (RMSE) (in percent) of the exposure ratios, for different ISO levels.

Code for noise simulation is available at https://github.com/gfxdisp/HDRutils/tree/main/HDRutils/noise_modeling.

Note that we can compute such relative errors only for synthetic datasets. We include results of a simple baseline by solving a linear system that realises Eq. (3) without relying on a noise model. Note that the baseline includes other components described in Section III, such as spatial balance via tiling and exposure balance via MSTs. The blue line shows the error of noisy exposures according to Eq. (14). The synthetic exposure stacks represent a challenging estimation problem with exposures three stops apart. As a result, the simple baseline (green) and the histogram-based BTF method [3] (orange) are unable to provide good results due to the adverse impact of pixels in shorter exposures. A key limitation of both methods is not explicitly modeling camera noise, which we address in our formulation.

Figure 7 depicts results for the two versions of our weighted least-squares solution (Eq. (7)), showing both pairwise connectivity (in red) and the greedy MST heuristic (in purple). Both our solutions use weights that model the noise characteristics of the camera. Thus, they result in better estimates with smaller errors. Overall, the greedy MST heuristic is the best-performing estimator of exposure ratios.

Execution times: Histogram matching used in [3] is computationally expensive for high-resolution images, resulting in an average execution time of 2.11 seconds. Further, we observed that the expensive iterative procedure for removing outliers is needed for good estimates increasing time to 8.54 seconds. Our tiled reduction (Section III-F1) results in a significantly faster average execution time of 0.265 seconds for pairwise connected exposures and 0.29 seconds for the greedy MST solution. The reported times were computed on an *Intel i7-8700 CPU* for images of resolution 4312×2868.

B. Performance on real captures

Next, we show the advantage of our proposed exposure estimation over naive merging using EXIF data and compare it with the histogram-based BTF approach [3]. The RAW images have a bit-depth of 14 and are linearly related to the scene radiance. We did not apply CRF or tone-mapping for any image shown in this section. Thus, all artifacts visible are due to incorrect exposures. For all results in this section, we use calibration-free weights given by Eq. (13).

Banding artifacts are visible for scenes containing a smooth gradient at the transition point when one of the input exposures saturates. This frequently occurs very close to bright light sources such as the Sun during the day or street lights at night, as indicated by red arrows in Figure 8 (a). Pixels close to light sources tend to be unsaturated only in shorter exposures where other parts of the image are strongly affected by noise. Aligning exposures under these conditions is challenging, resulting in the poor performance of the baseline as well as the histogram-based BTF weights [3]. The baseline method obtained by directly solving Eq. (3) completely fails to recover exposure ratios. When using the BTF weights, the banding artifacts also persist at the same locations or are introduced in other locations as shown in Figure 8 (c). By accounting for heteroskedastic noise with inverse-variance weights in Figure 8 (d), we can correctly align all the exposures to produce banding-free results.
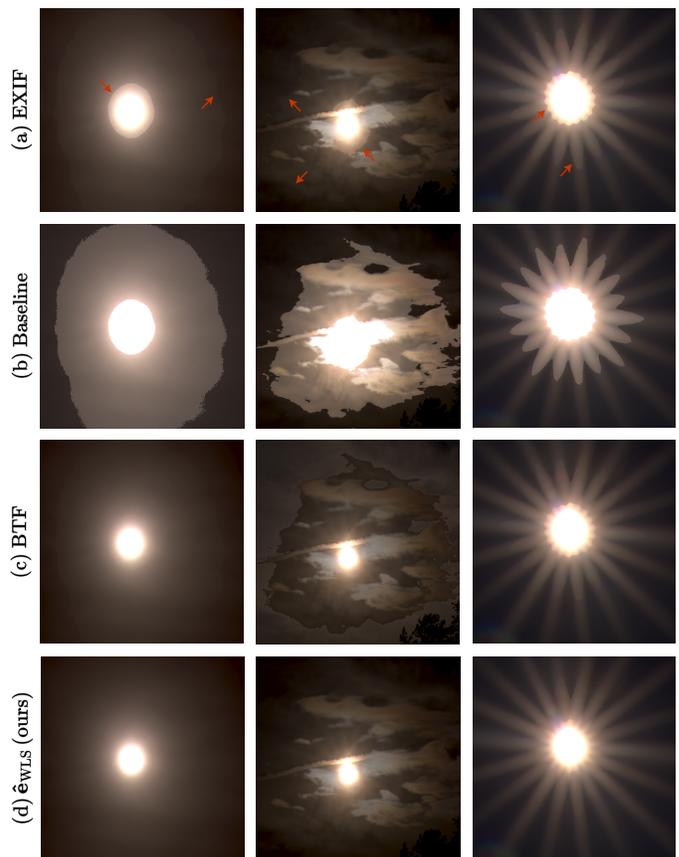
Fig. 8. Zoomed in patches of light sources, appropriately exposed to highlight banding close to the source due to inaccurately reported EXIF exposures (first row). The baseline method (second row) that does not model camera noise is unable to estimate shorter exposures, resulting in artifact-ridden reconstructions. Similarly, only some exposures can be aligned with histogram weights based on BTF (third row). By modeling camera noise, we simultaneously align all exposures (last row).

Even in the absence of point light sources, banding can appear at natural smooth gradients or due to defocus blur, as shown in Figure 9. Large regions of the images are well-exposed in such scenes, and most methods work reasonably well. However, the baseline and BTF weights can still sometimes fail to recover the exposure ratio for shorter exposures (for example, see the red patch in the third column and green and blue patches in the second column of the first scene). Our noise-model motivated approach consistently aligns all exposures to produce banding-free reconstructions across the scenes.

C. Deghosting for scenes with motion

When multi-image stacks with misaligned exposures are fused, the resulting HDR reconstructions often exhibit accentuated ghosting artifacts. Even recent deep learning-based methods [24], [42], [45] are unable to account for inaccurate exposure ratios. Figure 10b illustrates the artifacts that arise when exposures are corrupted according to Eq. (14). It is important to note that these artifacts are in addition to the ghosting artifacts visible in Figure 10a. Our proposed algorithm, which includes tile-based outlier detection to account for scene motion, effectively aligns exposures.
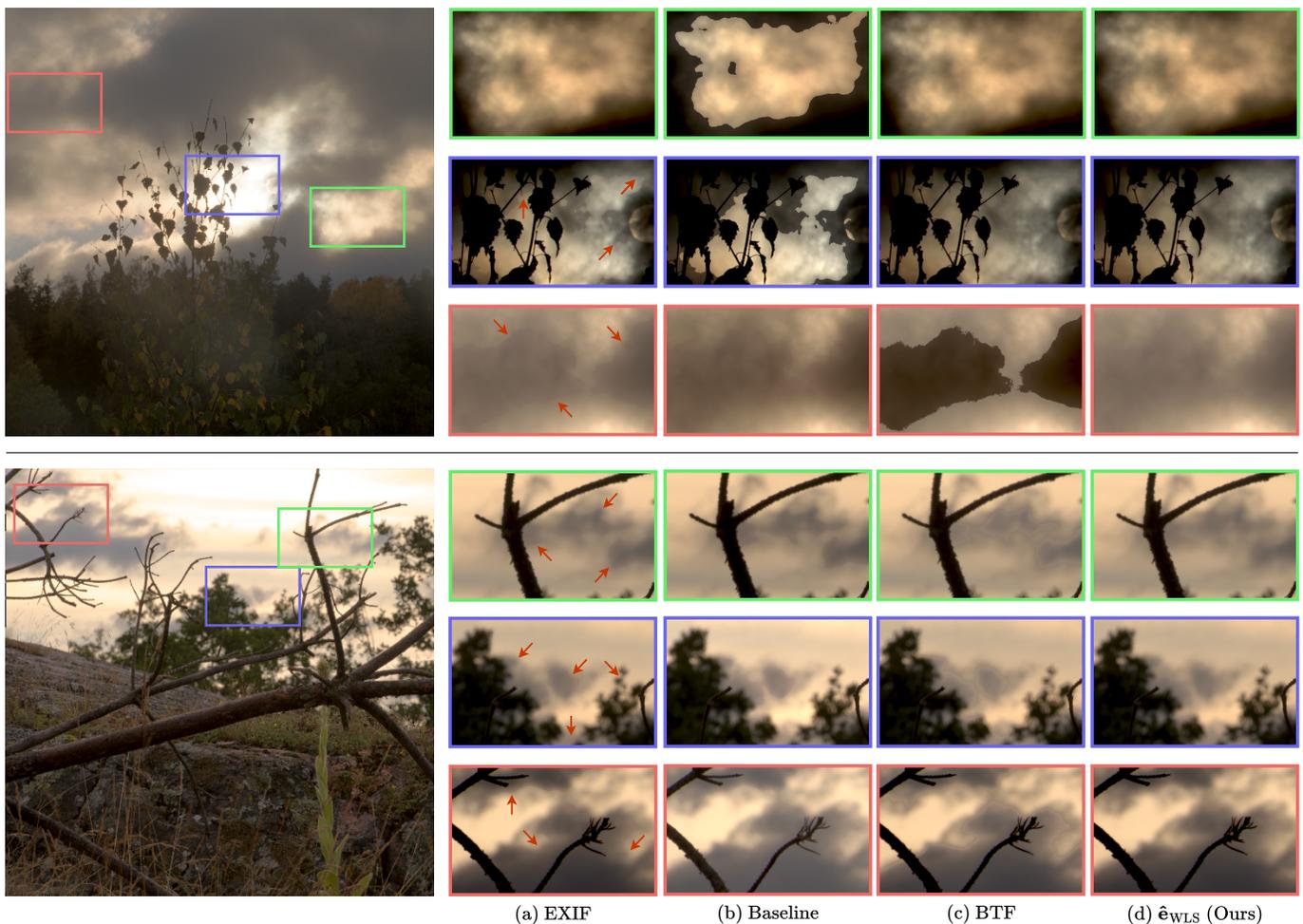
Fig. 9. Fixing banding in smooth image gradients when images contain sufficient well-exposed pixels to populate the linear system. Our inverse-variance solution (last column) is more robust and produces better reconstructions.
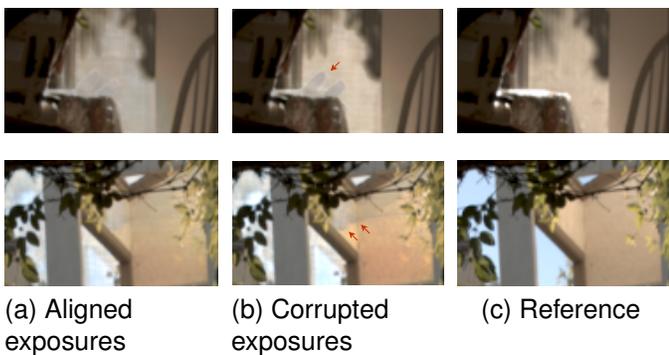
Fig. 10. Comparison of HDR reconstructions of an attention-based deep network [42] with aligned (first column) and corrupted (middle column) exposures. These insets are from the HDR deghosting dataset [24], which the deghosting network was trained on.

D. Measuring display MTF

Digital cameras are often used as inexpensive light measuring instruments, for example, for measuring the spatial characteristic of an electronic display, such as spatial uniformity of a modulation transfer function (MTF) [48]. Inaccuracies in pixel values due to exposure misalignment may lead to errors
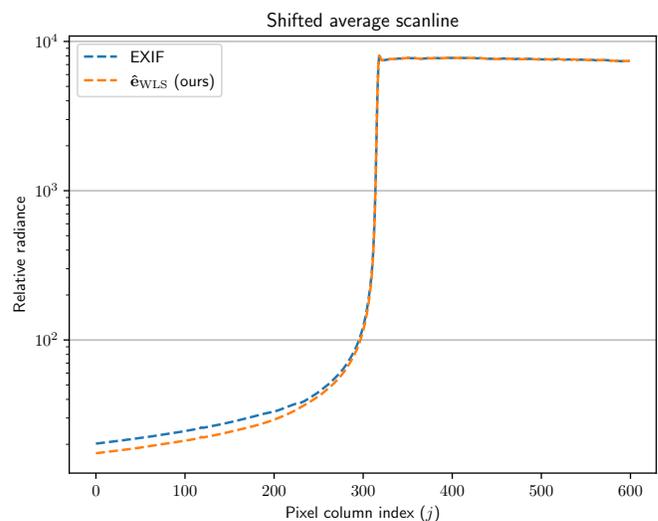
Fig. 11. Edge profile of an HDR display recovered from a slant edge captured with an exposure stack. Errors in EXIF metadata lead to deviation from the actual edge spread function, affecting MTF estimation. On inspecting the captured images, we observed the banding artifacts close to the captured slant-edge when the image was merged with EXIF exposures.

in the estimated display characteristics. Here, we demonstrate the need for exposure alignment for measuring MTF of HDR displays.

The MTF measurement involves taking an image of a slanted edge [36], where one side has its pixels set to 0 and the other side to the maximum pixel value. Modern displays can reach very high contrast levels, so such a slanted-edge image can be faithfully captured only using an exposure stack. The inaccuracies of exposure times can easily introduce bias to our measurements.

Figure 11 shows the edge profile computed with (blue) and without (orange) exposure estimation. The difference in reconstructions is due to inaccurate EXIF exposure values, visible as banding in the merged image. This translates to the bias at low-radiance pixels in Figure 11 and shows that errors in exposure times can easily lead to biases in the final MTF estimates.

V. Conclusions

We propose to estimate relative exposures directly from images of an HDR exposure stack instead of relying on inaccurate camera EXIF metadata. The problem can be formulated as a weighted linear system of equations with Tikhonov regularization. Our formulation considers the noise characteristics of a camera to derive weights. We show that good performance on a large number of scenes is possible without the need for camera- and gain-specific noise parameters.

We also describe an efficient approach based on the union of the highest weighted spanning trees of the exposure multigraph to reduce the size of the system for large images. When applied to multi-image HDR stacks, our method eliminates banding artifacts at smooth image gradients close to light sources or due to defocus blur. The exposure-aligned images are closer to physical quantities, making our work essential for using a camera as a light-measurement instrument.

Acknowledgments

Appendix
Moments of a transformed random variable

Here, we use the Taylor expansion to derive expressions for the first and second moments (expected value and variance respectively) of a random variable under an arbitrary invertible function. In Section III-D we apply these results to obtain expressions for the $\log$ of a normally distributed random variable.

To get an approximate variance for a nonlinear function of a random variable, we describe the Delta method [23], [31] that uses a Taylor expansion. Let Z be an asymptotically normal random variable with a known mean μ_Z and variance σ_Z^2, and let f be an invertible and differentiable function. To get

expressions for the expected value and variance of $f(Z)$ we use its first order Taylor expansion about the mean μ_Z,

$$f(Z) = f(\mu_Z) + f'(\mu_Z)(Z - \mu_Z) \\ + \text{higher-order terms}. \tag{15}$$

Ignoring the diminishing contributions of higher-order terms, the expected value is,

$$\begin{aligned} \mathbb{E}[f(Z)] &\approx \mathbb{E}[f(\mu_Z)] + \mathbb{E}[f'(\mu_Z)(Z - \mu_Z)] \\ &= \mathbb{E}[f(\mu_Z)] + f'(\mu_Z)(\mathbb{E}[Z] - \mathbb{E}[\mu_Z]) \\ &= f(\mu_Z) + f'(\mu_Z)(\mu_Z - \mu_Z) \\ &= f(\mu_Z), \end{aligned} \tag{16}$$

and the variance is

$$\begin{aligned} \mathbb{V}[f(Z)] &\approx \mathbb{V}[f(\mu_Z)] + \mathbb{V}[f'(\mu_Z)(Z - \mu_Z)] \\ &= \mathbb{V}[f(\mu_Z)] + f'(\mu_Z)^2 \, \mathbb{V}[Z - \mu_Z] \\ &= \mathbb{V}[f(\mu_Z)] + f'(\mu_Z)^2 \, (\mathbb{V}[Z] + \mathbb{V}[\mu_Z]) \\ &= 0 + f'(\mu_Z)^2 \, (\sigma_Z^2 + 0) \\ &= f'(\mu_Z)^2 \, \sigma_Z^2. \end{aligned} \tag{17}$$

References

[1] A. Abdelhamed, M. A. Brubaker, and M. S. Brown, "Noise Flow: Noise Modeling with Conditional Normalizing Flows," in *International Conference on Computer Vision (ICCV)*, 2019. 2

[2] C. Aguerrebere, J. Delon, Y. Gousseau, and P. Musé, "Study of the digital camera acquisition process and statistical modeling of the sensor raw data," Laboratoire Traitement et Communication de l'Information, Instituto de Ingeniería Eléctrica, Tech. Rep., Sep. 2012. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00733538 2, 4, 5

[3] L. Cerman and V. Hlavac, "Exposure time estimation for high dynamic range imaging with hand held camera," in *Proc. of Computer Vision Winter Workshop, Czech Republic*. Citeseer, 2006. 2, 4, 6, 7, 8

[4] K.-C. Chang, R. Wang, H.-J. Lin, Y.-L. Liu, C.-P. Chen, Y.-L. Chang, and H.-T. Chen, "Learning camera-aware noise models," in *Proceedings of European Conference on Computer Vision (ECCV)*, 2020. 2

[5] J. Chen, Z. Yang, T. N. Chan, H. Li, J. Hou, and L.-P. Chau, "Attention-guided progressive neural texture fusion for high dynamic range image restoration," *IEEE Transactions on Image Processing*, vol. 31, pp. 2661–2672, 2022. 3

[6] Y. Chen, G. Jiang, M. Yu, Y. Yang, and Y.-S. Ho, "Learning stereo high dynamic range imaging from a pair of cameras with different exposure parameters," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1044–1058, 2020. 2

[7] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '97. USA: ACM Press/Addison-Wesley Publishing Co., 1997, p. 369–378. [Online]. Available: https://doi.org/10.1145/258734.258884 2, 3

[8] G. Eilertsen, J. Kronander, G. Denes, R. Mantiuk, and J. Unger, "Hdr image reconstruction from a single exposure using deep cnns," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 6, 2017. 3

[9] D. Eppstein, "Finding the k smallest spanning trees," in *Scandinavian Workshop on Algorithm Theory*. Springer, 1990, pp. 38–47. 6

[10] M. D. Fairchild, "The hdr photographic survey," in *Color and imaging conference*, vol. 2007. Society for Imaging Science and Technology, 2007, pp. 233–238. 1, 2, 3, 7

[11] A. Foi, "Clipped noisy images: Heteroskedastic modeling and practical denoising," *Signal Processing*, vol. 89, no. 12, pp. 2609–2629, 2009, special Section: Visual Information Analysis for Security. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168409001996 4

[12] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical poissonian-gaussian noise modeling and fitting for single-image raw-data," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1737–1754, 2008. 2, 5

[13] O. Gallo and P. Sen, "Stack-Based Algorithms for HDR Capture and Reconstruction," in *High Dynamic Range Video*. Elsevier, 2016, pp. 85–119. 1

[14] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H.-P. Seidel, and H. P. Lensch, "Optimal hdr reconstruction with linear digital cameras," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 215–222. 2, 3

[15] M. D. Grossberg and S. K. Nayar, "What can be known about the radiometric response from images?" in *Computer Vision — ECCV 2002*, A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 189–205. 2, 3

[16] S. Hajisharif, J. Kronander, and J. Unger, "Adaptive dualiso hdr reconstruction," *EURASIP Journal on Image and Video Processing*, vol. 2015, no. 1, p. 41, 2015. 1

[17] P. Hanji, M. Z. Alam, N. Giuliani, H. Chen, and R. K. Mantiuk, "Hdr4cv: High dynamic range dataset with adversarial illumination for testing computer vision methods," *Journal of Imaging Science and Technology*, 2021. [Online]. Available: http://www.cl.cam.ac.uk/research/rainbow/projects/hdr4cv-dataset/ 3

[18] P. Hanji, R. K. Mantiuk, G. Eilertsen, S. Hajisharif, and J. Unger, "Comparison of single image hdr reconstruction methods — the caveats of quality assessment," in *Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings (SIGGRAPH '22 Conference Proceedings)*, 2022. [Online]. Available: https://www.cl.cam.ac.uk/research/rainbow/projects/sihdr_benchmark/ 1, 2, 3

[19] P. Hanji, F. Zhong, and R. K. Mantiuk, "Noise-aware merging of high dynamic range image stacks without camera calibration," in *Advances in Image Manipulation (ECCV workshop)*. Springer, 2020, pp. 376–391. [Online]. Available: https://doi.org/10.1007/978-3-030-67070-2_23 2, 3, 7

[20] S. W. Hasinoff, D. Sharlet, R. Geiss, A. Adams, J. T. Barron, F. Kainz, J. Chen, and M. Levoy, "Burst photography for high dynamic range and low-light imaging on mobile cameras," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 6, p. 192, 2016. 1

[21] S. Hasinoff, F. Durand, and W. Freeman, "Noise-optimal capture for high dynamic range photography," in *CVPR*. IEEE, 2010, pp. 553–560. 2, 3

[22] F. Heide, M. Steinberger, Y.-T. Tsai, M. Rouf, D. Pajak, D. Reddy, O. Gallo, J. Liu, W. Heidrich, K. Egiazarian *et al.*, "Flexisp: A flexible camera image processing framework," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 6, pp. 1–13, 2014. 1

[23] J. M. V. Hoef, "Who invented the delta method?" *The American Statistician*, vol. 66, no. 2, pp. 124–127, 2012. [Online]. Available: https://doi.org/10.1080/00031305.2012.687494 10

[24] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2017)*, vol. 36, no. 4, 2017. 3, 8, 9

[25] K. Karaduzovic-Hadziabdic, J. H. Telalovic, and R. Mantiuk, "Subjective and Objective Evaluation of Multi-exposure High Dynamic Range Image Deghosting Methods," in *Eurographics 2016 - Short Papers*, 2016. 3

[26] N. Katoh, T. Ibaraki, and H. Mine, "An algorithm for finding k minimum spanning trees," *SIAM Journal on Computing*, vol. 10, no. 2, pp. 247–255, 1981. 6

[27] M. Kim, M. Azimi, and R. K. Mantiuk, "Perceptually motivated model for predicting banding artefacts in high-dynamic range images," *Color and Imaging Conference*, vol. 2020, no. 28, pp. 42–48, nov 2020. [Online]. Available: https://www.ingentaconnect.com/content/10.2352/issn.2169-2629.2020.28.8 1

[28] M. Konnik and J. Welsh, "High-level numerical simulations of noise in ccd and cmos photosensors: review and tutorial," *arXiv preprint arXiv:1412.4031*, 2014. 2, 4

[29] K. Ma, Z. Duanmu, H. Yeganeh, and Z. Wang, "Multi-exposure image fusion by optimizing a structural similarity index," *IEEE Transactions on Computational Imaging*, vol. 4, no. 1, pp. 60–72, 2018. 2

[30] T. Mitsunaga and S. Nayar, "Radiometric self calibration," in *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, vol. 1, 1999, pp. 374–380 Vol. 1. 2, 3

[31] G. W. Oehlert, "A note on the delta method," *The American Statistician*, vol. 46, no. 1, pp. 27–29, 1992. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/00031305.1992.10475842 10

[32] E. Perez-Pellitero, S. Catley-Chandar, A. Leonardis, and R. Timofte, "Ntire 2021 challenge on high dynamic range imaging: Dataset, methods and results," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2021, pp. 691–700. 2, 3

[33] K. R. Prabhakar, S. Agrawal, and R. V. Babu, "Self-gated memory recurrent network for efficient scalable hdr deghosting," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 1228–1239, 2021. 3

[34] Z. Pu, P. Guo, M. S. Asif, and Z. Ma, "Robust high dynamic range (hdr) imaging with complex motion and parallax," in *Proceedings of the Asian Conference on Computer Vision (ACCV)*, November 2020. 3

[35] R. G. Rodríguez, J. Vazquez-Corral, and M. Bertalmío, "Issues with common assumptions about the camera pipeline and their impact in hdr imaging from multiple exposures," *https://doi.org/10.1137/19M1250248*, vol. 12, pp. 1627–1642, 10 2019. [Online]. Available: https://epubs.siam.org/doi/10.1137/19M1250248 2

[36] E. Samei, M. J. Flynn, and D. A. Reimann, "A method for measuring the presampled mtf of digital radiographic systems using an edge test device," *Medical physics*, vol. 25, no. 1, pp. 102–113, 1998. 2, 10

[37] M. S. Santos, R. Tsang, and N. Khademi Kalantari, "Single image hdr reconstruction using a cnn with masked features and perceptual loss," *ACM Transactions on Graphics*, vol. 39, no. 4, 7 2020. 3

[38] T. Tani, *Photographic sensitivity: theory and mechanisms*. Oxford University Press on Demand, 1995, no. 8. 2

[39] A. Tomaszewska and R. Mantiuk, "Image registration for multi-exposure high dynamic range image acquisition," in *Proceedings of the International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, 01 2007. 6

[40] Z. Tu, J. Lin, Y. Wang, B. Adsumilli, and A. C. Bovik, "Bband index: A no-reference banding artifact predictor," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 2712–2716. 1

[41] O. T. Tursun, A. O. Akyüz, A. Erdem, and E. Erdem, "The state of the art in hdr deghosting: a survey and evaluation," in *Computer Graphics Forum*, vol. 34, no. 2. Wiley Online Library, 2015, pp. 683–707. 2

[42] T. V.V, "High dynamic range image synthesis via attention non-local network," 2020. [Online]. Available: github.com/tuvovan/ANL-HDRI 8, 9

[43] L. Wang and K.-J. Yoon, "Deep learning for hdr imaging: State-of-the-art and future trends," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 3

[44] Y. Wang, S.-U. Kum, C. Chen, and A. Kokaram, "A perceptual visibility metric for banding artifacts," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 2067–2071. 1

[45] Q. Yan, L. Zhang, Y. Liu, Y. Zhu, J. Sun, Q. Shi, and Y. Zhang, "Deep hdr imaging via a non-local network," *IEEE Transactions on Image Processing*, vol. 29, pp. 4308–4322, 2020. 8

[46] X. Yang, K. Xu, Y. Song, Q. Zhang, X. Wei, and R. W. Lau, "Image correction via deep reciprocating hdr transformation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1798–1807. 2

[47] Q. Ye, J. Xiao, K.-m. Lam, and T. Okatani, *Progressive and Selective Fusion Network for High Dynamic Range Imaging*. New York, NY, USA: Association for Computing Machinery, 2021, p. 5290–5297. [Online]. Available: https://doi.org/10.1145/3474085.3475651 3

[48] X. Zhang, T. Kashti, D. Kella, T. Frank, D. Shaked, R. Ulichney, M. Fischer, and J. P. Allebach, "Measuring the modulation transfer function of image capture devices: what do the numbers really mean?" in *Image Quality and System Performance IX*, vol. 8293. SPIE, 2012, pp. 64–74. 9