

Visualizing Errors in Rendered High Dynamic Range Images

P. Andersson, J. Nilsson, P. Shirley, and T. Akenine-Möller

NVIDIA

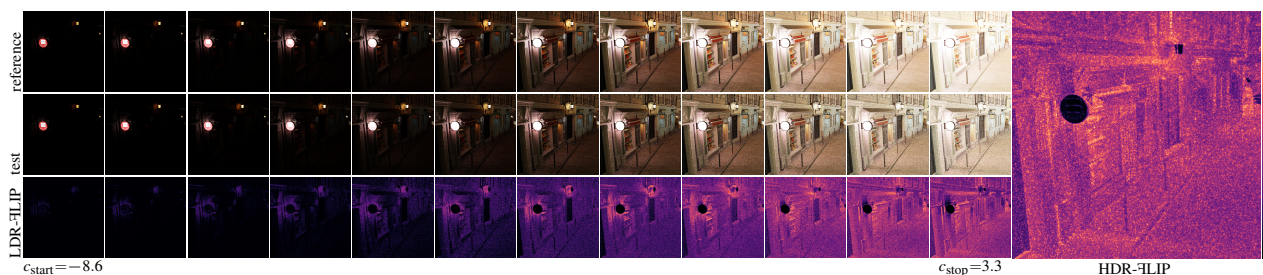


Figure 1: Illustration of the difficulties of revealing errors in rendered HDR images. A reference HDR image of the BISTRONIGHT scene is compared against a test HDR image, where the latter was path traced with 16 samples per pixel (SPP) and the former with 2M SPP. Columns 1–12 are LDR images, which were generated from the HDR images using an exposure compensation value between c_{start} and c_{stop} and then tone mapped. From the LDR-FLIP [ANA*20] error maps at the bottom, it is clear that some errors are visible in bright regions using short exposures (left), while other errors are visible in dark areas with longer exposures (right). Note also that yet different errors are largest for medium exposures. This leads to the insight that the errors cannot be visualized using a single exposure. Right: the per-pixel error map generated by our new method, HDR-FLIP, which attempts to signal errors at any relevant exposure and of any common tone mapper.

Abstract

A new error metric targeting rendered high dynamic range images is presented. Our method computes a composite visualization over a number of low dynamic range error maps of exposure compensated and tone mapped image pairs with automatically computed, or manually provided, parameters. We argue that our new error maps predict errors substantially better than metrics previously used in rendering. Source code is released with the hope that our work can be a useful tool for future research.

CCS Concepts

• **Computing methodologies** → **Rendering**; • **General and reference** → **Metrics**;

1. Introduction

Almost all rendering algorithms compute high dynamic range (HDR) images [RHD*10]. Until there is a ubiquitous HDR display ecosystem, however, the vast majority of HDR images are *exposure compensated*, sometimes dynamically so in applications such as video games, and *tone mapped* into a low dynamic range (LDR) image before they are displayed. This procedure—mainly the exposure compensation, as most common tone mappers perform similar compression of the HDR image’s dynamic range—has significant impact on what is shown to the observer. Naturally, this also affects the observer’s perception of errors in images. Despite this, common methods of HDR image quality assessment in rendering research include visual comparisons of LDR images using a single exposure compensation, the use of statistical measures (e.g., variants of the mean squared error and the symmetric mean absolute percentage error), directly on the HDR input, and the application of LDR metrics on perception-adjusted HDR data [AMS08]. As-

suming that the HDR image will be exposure compensated using one of several possible exposures within a reasonable (but possibly unknown) range and tone mapped with a common (but possibly unknown) tone mapper before it is displayed, neither of these approaches are reliable in communicating the errors in the image.

In this paper, we present an algorithm for visualizing the errors in HDR test images, when compared to HDR references. Our algorithm, named HDR-FLIP, is developed based on the assumptions stated in the previous paragraph. It uses LDR-FLIP [ANA*20] to evaluate the perceived differences between multiple LDR versions, generated using different exposure compensation values, of the reference and test images. A single visualization is then created by taking the maximum per pixel error over exposures, yielding an error map that indicates the largest error the HDR image may show on an LDR display under any reasonable exposure. An example is shown in Figure 1. In Section 3, we argue that, under the given assumptions, our approach exceeds those mentioned previously.

2. HDR-FLIP

Simply making a single LDR image out of an HDR image does not reveal all rendering errors, as illustrated in both Figures 1 and 2. Our approach is based on computing a set of exposure compensated and tone mapped LDR image pairs from the HDR image pair, then applying an LDR error metric to those, and, finally, extracting a single error image from the LDR error metric images. This is similar to what Munkberg et al. [MCHAM06] did for HDR texture compression, but with the following key differences:

1. We use the perception-motivated LDR-FLIP [ANA*20] error metric instead of using MSE, which is not reliable [WB09].
2. We present a method that automatically computes the start and stop exposures based on the reference image.
3. Instead of averaging the error of all exposure compensated images, we keep only the maximum error per pixel as it represents the largest error that may be presented to an observer.
4. We take exposure compensation *and* tone mapping into account and not only the exposure compensation.
5. Our metric reveals *where* errors are located via error maps.

We refer to our supplemental material for additional support of the choices made in this section.

2.1. Computing Start and Stop Exposures

This section describes our procedure for automatically calculating the start exposure, c_{start} , and the stop exposure, c_{stop} . These will be derived as the exposures that map, with a chosen tone mapper, the reference image’s maximum and median luminance, denoted Y_{max} and Y_{med} , respectively, to pre-defined LDR luminance targets. By using the reference, we guarantee that the same c_{start} and c_{stop} values are used for all possible test images, at the cost of symmetry.

Exposure compensation with a factor, c , is computed as $T(I) = 2^c I$, where I is an HDR value, which we assume is nonnegative. Many tone mapper functions [RSSF02, Hab10, Nar16] can be described as rational polynomials, or approximated as such, i.e.,

$$y(x) = (k_0 x^2 + k_1 x + k_2) / (k_3 x^2 + k_4 x + k_5), \quad (1)$$

where x is an exposure compensated HDR value (luminance or an R/G/B value), and y is the tone mapped, LDR version of x . In this paper, all renderings are tone mapped with an ACES approximation [Nar16] applied per channel, while our source code allows for other tone mappers. The ACES approximation is also the default tone mapper used by HDR-FLIP. To derive c_{start} and c_{stop} , we set $y = t$ in Equation 1, where t is the target LDR luminance, and get

$$(tk_3 - k_0)x^2 + (tk_4 - k_1)x + tk_5 - k_2 = 0. \quad (2)$$

This is a second degree equation. Since tone mapping functions (Equation 1) are monotonic and positive in $[0, \infty)$, one of the roots will be negative and can be neglected (because $x \geq 0$ since $I \geq 0$). For the positive root, x_p , and for a certain I , we can solve for c as

$$x_p = 2^c I \iff c = \log_2(x_p/I). \quad (3)$$

To compute c_{start} and c_{stop} , we now use Equations 2 and 3. For c_{start} , we use $t = 0.85$ and $I = Y_{\text{max}}$. The rationale for c_{start} is that the brightest pixel should be tone mapped, using exposure, to a reasonably bright value. For c_{stop} , we use the same procedure and

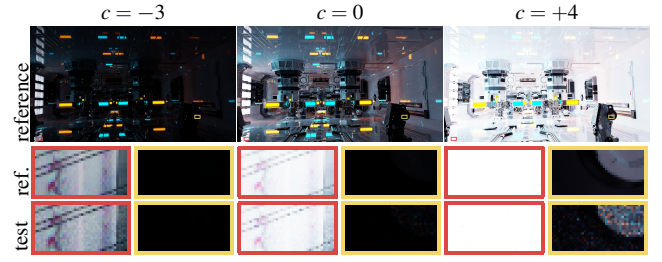


Figure 2: With ZERO DAY path traced using 2M (reference) and 256 samples per pixel (test), we illustrate that a single exposure, c , is not sufficient to convey all errors in an HDR image, since different exposures reveal different errors, as is emphasized in the insets.

target LDR luminance, but with $I = Y_{\text{med}}$, as it generally results in an overall bright image. Our choices give aesthetical results for all our test images, i.e., the image with c_{start} is dark, but not overly so, and the image with c_{stop} is bright, but not excessively so. The choices are subjective and some use cases may require other exposure ranges. Therefore, our code also offers the option to supply c_{start} and c_{stop} manually.

2.2. Computing and Visualizing the Error

Given c_{start} and c_{stop} , we create $N = \max(2, \lceil c_{\text{stop}} - c_{\text{start}} \rceil)$ LDR versions of the HDR test and reference images. The LDR versions are exposure compensated, using exposures uniformly distributed over the range $[c_{\text{start}}, c_{\text{stop}}]$, tone mapped, and clamped to $[0, 1]$. We denote the set of exposures C . To capture the largest errors the HDR image may display over C , we define our per-pixel error map as

$$E(x, y) = \max_{c \in C} \Delta E(x, y, c), \quad (4)$$

where $\Delta E(x, y, c)$ is the LDR-FLIP error [ANA*20] between the LDR reference and test versions using exposure compensation c . The resulting HDR-FLIP error map is in $[0, 1]$ and so the pooling techniques proposed by Andersson et al. [ANA*20] can be used.

After the numeric errors are computed, they can be directly visualized. Our main method is “false coloring” (also called “pseudocolor” or “color map”), which is the de facto approach in many fields. A good set of effective false color maps have been developed and provided in widely available environments. Among the maps recommended by visualization research, two characteristics are especially suited to rendering research: zero error should map to black, and high errors should look “hot.” The *magma* color map has these characteristics [LH18] and is used throughout the paper.

While the error’s magnitude and location usually are the most important aspects of error communication, knowing which exposures are responsible for the largest errors may also be of interest in our setting. As we compute the maximum error over the generated set of LDR image pairs, we output an additional false color map indicating, per pixel, which exposure yielded the maximum error. We use the *viridis* map, which is part of the same set as *magma* [LH18], but which does not map the lowest value to black. This serves intuition, as the shortest exposure is not necessarily 0, as was the case for the error. An example of this is shown in Figure 3.

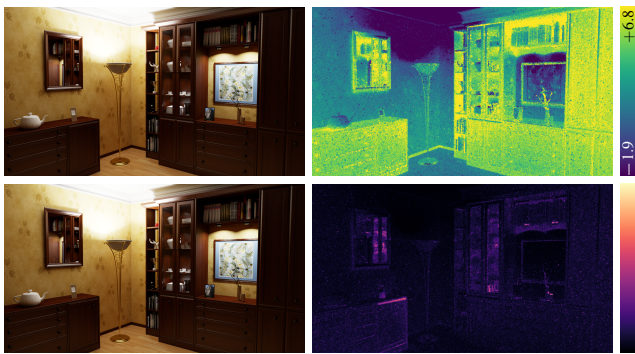


Figure 3: Many errors between the reference (top left) and test image (bottom left) of the BOOKSHELF scene are perceived as largest under long exposures, but some instead for medium and short ones. This can be seen in the exposure map (top right), where dark indicates short exposures ($c_{\text{start}} = -1.9$) and bright indicates long exposures ($c_{\text{stop}} = 6.8$). In the bottom right is the HDR-FLIP map.

3. Results

We compare HDR-FLIP to symmetric mean absolute percentage error (SMAPE), relative mean absolute error (RMAE), and relative MSE (relMSE), since those are the metrics most often used in recent rendering papers. Each of these metrics is detailed in the supplemental material. We avoid pure MSE, since it is not reliable [WB09], and mPSNR [MCHAM06], since it is based on MSE. HDR-VDP-2 [MKRH11], which is an image metric specifically developed for HDR images, is evaluated, but suffers from poor localization and detection probability levels. Due to its limitations, we choose to not present and discuss the HDR-VDP-2 results in the paper, but some results are included in the supplemental material. Metrics that combine perceptually uniform curves and LDR metrics [AMS08] are excluded, as applying LDR metrics to HDR data is expected to only capture parts of the errors due to the curve being a type of compression of the signal, similar to tone mapping operators, and different exposures are not considered. Note that, unless otherwise stated, all intermediate LDR-FLIP images were generated under the assumption that observation is done with 67 pixels per degree. All reference images were path traced using 2M SPP.

In Figure 4, we show that RMAE and relMSE clearly do not generate satisfying error maps, compared to the errors seen in the test image. SMAPE performs better, but fails to pick up errors that are visible in some regions under high exposure. In our supplemental material, we show an example where SMAPE again incorrectly indicates small errors in dark regions of a scene, but also examples where SMAPE, partly, corresponds well with the perceived errors.

We expect a common use case of our metric to be the comparison of different algorithms. In Figure 5, we use HDR-FLIP to compare two direct lighting rendering algorithms, denoted A and B. By directly examining the error images, we find that A seems to yield larger and more concentrated errors than B, while, on the other hand, suffering from less noise than the other algorithm.

Figure 6 shows HDR-FLIP maps computed between a 2M SPP HDR reference and a low SPP, denoised HDR test image. We

compare the error maps generated using different underlying tone mappers. In addition to HDR-FLIP's default tone mapper, i.e., the ACES approximation [Nar16], we selected the Uncharted 2 tone mapper [Hab10] and the Reinhard tone mapper [RSSF02] for the comparison, as these are some of the most commonly used tone mapping operators in rendering currently. The figure shows that, while the error maps are not identical and do show differences in error magnitudes, they look qualitatively similar in that they often indicate the same relative errors in the same parts of the image. This indicates that, even if the tone mapping operation that follows the user's rendering is unknown, the HDR-FLIP map, using the default tone mapper, will still aid the user in evaluating the error of the test image almost as well as if the tone mapper had been known.

4. Conclusions

Our methodology addresses a practical challenge: how to examine and communicate the accuracy of rendering algorithms. We invite researchers to use our work and hope it can be further improved. In the future, we wish to adapt our research to HDR displays.

Acknowledgments We are grateful to the people sharing their scenes: ZERODAY ©beep, BISTRONIGHT ©Amazon Lumberyard, SUBWAY ©silvertm. BOOKSHELF ©Tiziano Portenier. This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

References

- [AMS08] AYDIN T. O., MANTIUK R., SEIDEL H.-P.: Extending Quality Metrics to Full Dynamic Range Images. In *Human Vision and Electronic Imaging XIII* (San Jose, USA, January 2008), Proceedings of SPIE, pp. 6806–10. 1, 3
- [ANA*20] ANDERSSON P., NILSSON J., AKENINE-MÖLLER T., OSKARSSON M., ÅSTRÖM K., FAIRCHILD M. D.: FLIP: A Difference Evaluator for Alternating Images. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 3, 2 (2020), 15:1–15:23. 1, 2
- [Hab10] HABLE J.: Uncharted2 HDR Lighting. In *Game Developers Conference* (2010). 2, 3, 4
- [LH18] LIU Y., HEER J.: *Somewhere Over the Rainbow: An Empirical Assessment of Quantitative Colormaps*. Association for Computing Machinery, 2018, p. 1–12. URL: <https://doi.org/10.1145/3173574.3174172>. 2
- [MCHAM06] MUNKBERG J., CLARBERG P., HASSELGREN J., AKENINE-MÖLLER T.: High Dynamic Range Texture Compression for Graphics Hardware. *ACM Transactions on Graphics* 25, 3 (2006), 698–706. 2, 3
- [MKRH11] 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* 30, 4 (2011), 40:1–40:14. 3
- [Nar16] NARKOWICZ K.: ACES Filmic Tone Mapping Curve, 2016. URL: <https://knarkowicz.wordpress.com/2016/01/06/aces-filmic-tone-mapping-curve/>. 2, 3, 4
- [RHD*10] REINHARD E., HEIDRICH W., DEBEVEC P., PATTANAİK S., WARD G., MYSZKOWSKI K.: *High Dynamic Range Imaging*, 2nd ed. Elsevier, 2010. 1
- [RSSF02] REINHARD E., STARK M., SHIRLEY P., FERWERDA J.: Photographic Tone Reproduction for Digital Images. *ACM Transactions on Graphics* 21, 3 (2002), 267–276. 2, 3, 4
- [WB09] WANG Z., BOVIK A. C.: Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. *IEEE Signal Processing Magazine* 26, 1 (January 2009), 98–117. 2, 3

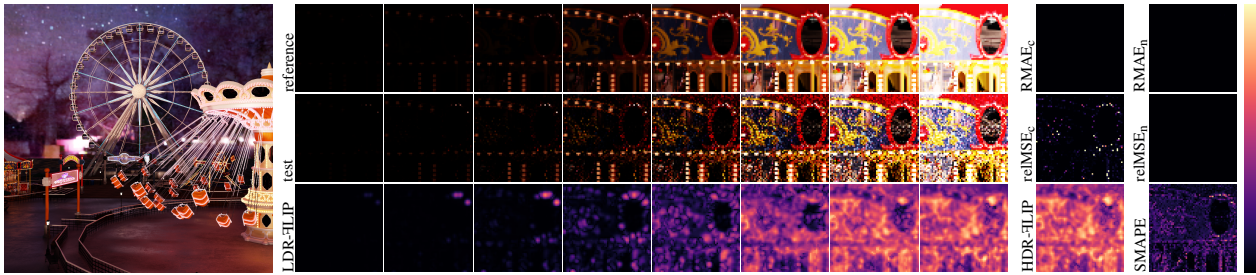


Figure 4: Left: an exposure compensated and tone mapped image of the AMUSEMENTPARK scene. The next eight columns show exposure compensated and tone mapped zoomins for the reference (top) and test image (middle), and the corresponding LDR-FLIP maps (bottom). Even though $N = 18$ ($c_{\text{start}} = -8.8$ and $c_{\text{stop}} = 8.7$), we only show every other exposure after the first three to save space. To the right, we show several standard metrics, including our HDR-FLIP, where it is clear that the standard metrics fail to detect all the errors seen in the eight columns. The subscript c means the metric has been clamped to $[0, 1]$, while n means it has been normalized to $[0, 1]$, for visualization.

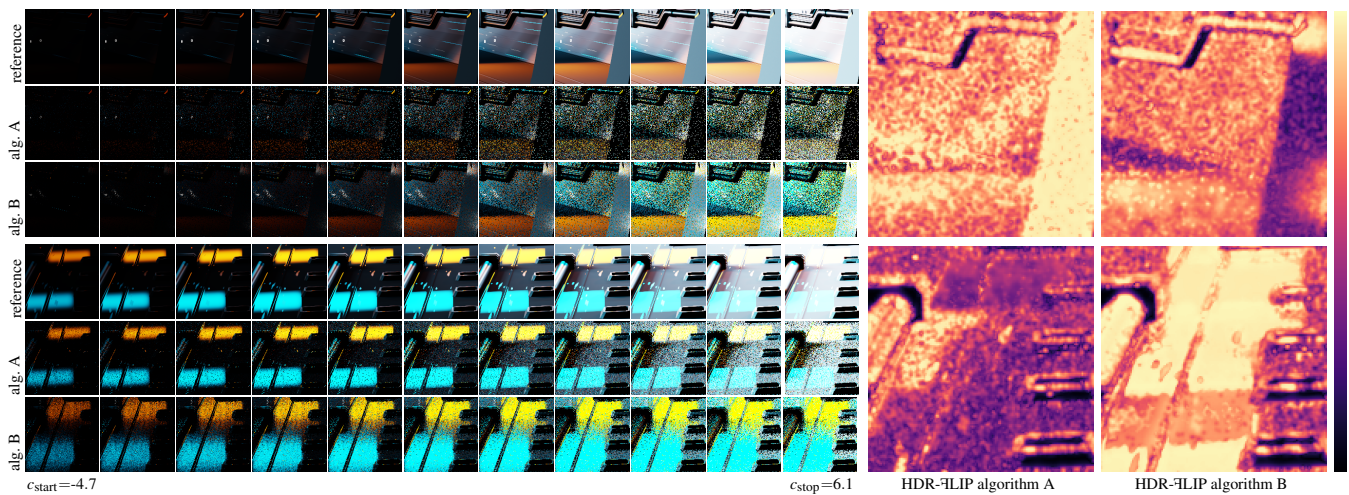


Figure 5: Two crops of the ZERODAY scene showing how HDR-FLIP can be used to evaluate and compare algorithms. The first 11 columns show exposure compensated and tone mapped reference images, images produced with algorithm A, and images generated with algorithm B. On the right are corresponding HDR-FLIP images for the two algorithms. We see how the algorithms struggle in different regions, with A sometimes generating black pixels where the reference is rather bright (see the right part of the first crop), whereas B tends to erroneously spread bright values over nearby pixels (see the reflection of the emitters in the second crop) and introduce false highlights (see the right part of the first crop). The full LDR images and corresponding LDR-FLIP maps that were used to composite the HDR-FLIP maps can be found in the supplemental material.



Figure 6: Example showing how the HDR-FLIP image depends on the underlying tone mapper. From left to right: exposure compensated and tone mapped reference and test image, HDR-FLIP computed using the ACES [Nar16] tone mapper, HDR-FLIP given with the Uncharted 2 (also known as Hable) [Hab10] tone mapper, and HDR-FLIP after using the Reinhard [RSSF02] tone mapper as underlying tone mapping operator. While showing differences in error magnitudes, the three maps look qualitatively similar, indicating that analyzing the error map generated with the default underlying tone mapper (ACES), can be beneficial to assess the perceived quality of HDR images that will be subject to an unknown tone mapper before it is displayed to the observer.