# Perceptually Adaptive Real-Time Tone Mapping

Taimoor Tariq
Meta, University of Lugano
USA, Switzerland
tariqt@usi.ch

Nathan Matsuda
Meta
USA
nathan.matsuda@meta.com

Eric Penner
Meta
USA
epenner@meta.com

Jerry Jia
Meta
USA
jerry.jia@meta.com

Douglas Lanman
Meta
USA
douglas.lanman@meta.com

Ajit Ninan
Meta
USA
ajitninan@meta.com

Alexandre Chapiro
Meta
USA
alex@chapiro.net


**Figure 1: Three tone mapping control methods are applied to an HDR-VR scene (Section 6). While the *fixed* and *heuristic* parameter estimations fail to faithfully reproduce appearance, *our* method successfully finds an optimal tone-curve. Note that HDR scenes cannot be accurately represented in this document format as they were seen by our users. The rightmost image shows the unmapped HDR reference, with luminances above 100 nits clipped and marked with dashes.**

## ABSTRACT

Tone mapping operators aim to remap content to a display's dynamic range. Virtual reality is a popular new display modality that has significant differences from other media, making the use of traditional tone mapping techniques difficult. Moreover, real-time adaptive estimation of tone curves that faithfully maintain appearance remains a significant challenge. In this work, we propose a real-time perceptual contrast-matching framework, that allows us to optimally remap scenes for target displays. Our framework is optimized for efficiency and runs on a mobile Quest 2 headset in under 1*ms* per frame. A subjective study on an HDR-VR prototype demonstrates our method's effectiveness across a wide range of display luminances, producing imagery that is preferred to alternatives tone mapped at peak luminances an order of magnitude higher. This result highlights the importance of good tone mapping for visual quality in VR.


*SIGGRAPH Asia '23 Conference Papers, Dec 12–15, 2023, Sydney, Australia*

ACM ISBN 978-1-4503-9470-3/22/12.
https://doi.org/000-000

## CCS CONCEPTS

• **Computing methodologies** → **Perception**.

## KEYWORDS

perception, virtual reality, high-dynamic-range

**ACM Reference Format:**
Taimoor Tariq, Nathan Matsuda, Eric Penner, Jerry Jia, Douglas Lanman, Ajit Ninan, and Alexandre Chapiro. 2023. Perceptually Adaptive Real-Time Tone Mapping. In *SIGGRAPH Asia 2023 Conference Papers (SIGGRAPH Asia '23 Technical Papers), Dec 12–15, 2023, Sydney, Australia.* ACM, New York, NY, USA, 10 pages. https://doi.org/000-000

## 1 INTRODUCTION

Tone mapping is an essential part of high-quality presentation in modern displays. Traditionally, a tone mapping operator (TMO) is the computational re-mapping of an image or video with high-dynamic-range (HDR) to a smaller (standard) dynamic range (SDR).

More broadly, TMOs can be thought of as functions that bring content to utilize the dynamic range of a given display optimally.

Remapping content to a smaller dynamic range has a big effect on visual appearance, and maintaining a perceptually equivalent look after tone mapping is challenging. Virtual reality (VR) displays introduce further difficulties as they require high refresh-rates at a wide field-of-view, adding latency risks that can cause discomfort and temporal artifacts. Furthermore, standalone VR-HMDs have significant restrictions on display brightness and processing power, due to form-factor requirements. In this work, we revisit the familiar topic of tone mapping, as we believe solving these challenges will bring us closer to the goal of realistic immersive display.

We draw inspiration from recent advances in perceptual image difference metrics [Mantiuk et al. 2021], and the understanding of human contrast perception across different luminance scales [Ashraf et al. 2022], to build a framework that estimates real-time multi-scale supra-threshold contrast losses due to remapping, at minimal computational cost. To demonstrate this system's effectiveness, we implemented an optimized version of the popular Photographic TMO curve [Reinhard et al. 2002]. Our method was developed and tested on an HDR-VR prototype, and runs on a commercially-available standalone Quest-2 HMD in under $1ms$ per frame. We proceed to show through a subjective study that our system can adaptively control tone-mapping to bring the perceived appearance significantly closer to the HDR reference. Our primary contributions are:

- A real-time perceptual framework for preserving supra threshold contrast across the frustum during tone mapping.
- An efficient real-time VR-oriented TMO implementation using our framework.
- A perceptual study showing our method's effectiveness over different display luminances and content.

## 2 RELATED WORK

### 2.1 Tone mapping

Tone mapping is an established topic of exploration, with many techniques published in the literature. A prominent example is Reinhard's Photographic tone mapper [2002], inspired by print photography exposure techniques. This popular method has been extended for real-time video processing [Krawczyk et al. 2005] using a physiologically-motivated filtering operator for temporal stability. The Photographic tone mapper is robust under a variety of conditions due to its straightforward formulation, which leads to its continued popularity for real-time implementations in applications such as gaming [2016]. This led us to select this method as the base for the system presented in this paper.

Display Adaptive tone mapping and its extensions [Eilertsen et al. 2015; Mantiuk et al. 2008] propose an optimization to generate a display-aware tone curve using conditional contrast probability histograms, which do no capture local contrast information, and are challenging to implement in real-time on mobile platforms. In contrast, our method can be efficiently parallelized using MipMaps. Aydin et al. [2014] employ spatio-temporal filtering prior to tone-mapping to ensure temporal consistency, also adding prohibitive costs to real-time implementation. In their work, tone mapping curves developed by Drago et al. [2003] and Tumblin et al [1999] are used as bases for a manually calibrated mapping. To show our framework's robustness, we demonstrate it on these TMOs in Section 7.2. For an overview of HDR and TMOs, we recommend the book by Reinhard et al. [2010].

### 2.2 HDR in VR

HDR for VR has received far less attention compared to traditional HDR displays. Recent work by Matsuda et al. [2022] demonstrates a high-luminance VR prototype, and a user-study on luminance preference; but tone mapping is not discussed. In contrast, rather than focusing on general luminance preference, we employ a high-luminance VR-HMD prototype to investigate the effectiveness of tone mapping in immersive HDR environments.

A natural choice for inherently binocular systems like VR are dichoptic tone mappers, which take advantage of binocular vision to present differently mapped images to each eye to generate an improved precept. While exciting new methods have been proposed [Zhang et al. 2018], they have been shown to have unstable performance in many conditions [Wang and Cooper 2021].

A TMO for VR panoramas was proposed by Zadeh et al. [2017]. Their method tone-maps by subdividing an immersive panorama into tiles. For a given frustum, parameters are set based on metadata of visible regions. Similarly, Goude et al. [2020] propose an approach blending TMO parameters between the entire panorama and the current frustum. Both these techniques employ a manually calibrated Photographic TMO [2002]. Additionally, these approaches rely on prior knowledge of the entire 360° scene, which may not be available in real-time VR scenarios. An overview of panorama TMOs was published by Melo and colleagues [2018]. In contrast, our work focuses on real-time tone mapping with perceptually optimized content-adaptive parameter estimation, which is more widely applicable to VR scenarios.

### 2.3 Perceptual rendering techniques

Perception-in-the-loop algorithms have become popular in real-time rendering. In particular, techniques like foveated rendering are geared for VR, leveraging reduced contrast sensitivity to high frequencies in the periphery of vision for efficiency. Tursun et al. [2019] derive a predictor of foveated rendering parameters based on minimizing the perceived contrast error of a foveated image in relation to a reference. Walton et al. [2021] use a pyramid decomposition and parallel processing to compute foveated metamers.

The perceptual difference metric FovVideoVDP was recently proposed by Mantiuk and colleagues [2021]. This metric utilizes a general-approach pipeline to predict visible differences using perceptual models. Notably, FovVideoVDP scales contrast linearly via threshold multiples, but TMOs modify content luminance at significant supra-threshold levels, and our aim is to compare the perception of supra-threshold contrast at very different luminance scales. Recent work by Ashraf et al. [2022] shows that the Kulikowski contrast-constancy formula [1976] is more accurate when matching contrast within our range of interest for VR display (20-2000 $cd/m^2$), leading us to adopt it in our pipeline, similarly to day/night appearance compensation by Wanat et al. [2014].

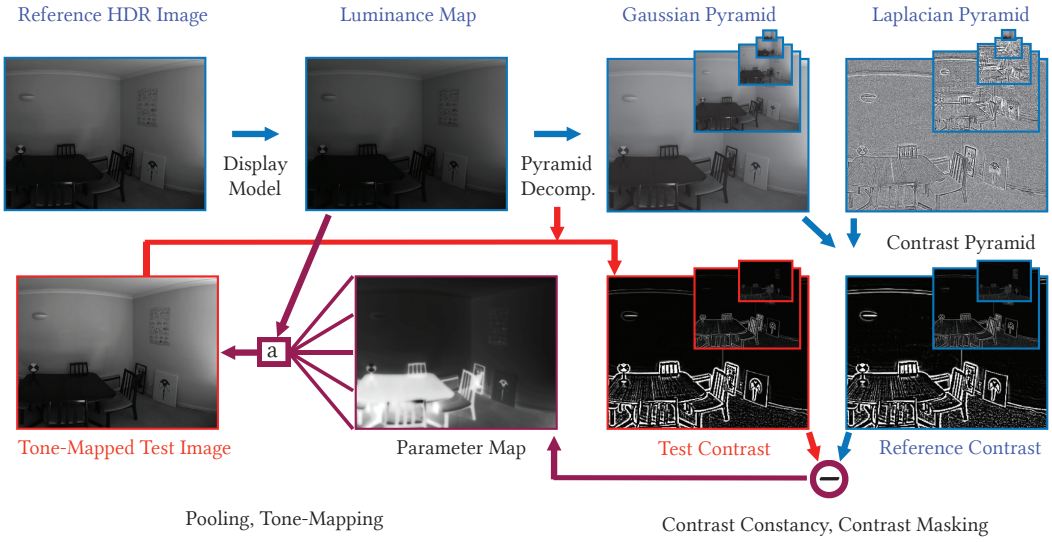


**Figure 2: A visual representation of our perceptual framework, detailed in Section 3.**

## 3 PERCEPTUAL FRAMEWORK

An HDR image can be mapped onto an SDR display in many ways, which can dramatically alter the appearance of the displayed content. Notably, our goal in this work is to create a general real-time framework in which a tone mapper can be optimally employed to reproduce the target appearance on a given display. This problem is especially relevant for novel display modes, such as VR, whose characteristics differ significantly from the traditional displays for which most tone mapping methods were developed.

To achieve this, we create a model of the human visual system, which is used to find tone-curves minimizing the perceived difference between the HDR image, and the tone-mapped image on the target display. Our perceptual framework is depicted in Figure 2, and we begin our explanation following the blue line in the schematic.

*Display model.* The reference image is processed via a standard display photometry and geometry model. The pixel values are linearized, and luminance and chrominance channels are separated via the YUV color space. The luminance channel is then converted to physical values in *nits*. All the tone mapping operations described below are applied only to this channel, while the chrominance channels are compensated using color-to-luminance ratios [Schlick 1995] to account for the hunt effect [Mantiuk et al. 2008].

*Pyramid decomposition.* The luminance image is converted into a Gaussian pyramid, which is used to generate a Laplacian pyramid. Pyramid decomposition is used as a computationally efficient estimate of local image frequencies, avoiding a costly Fourier transform. Peak frequency of relevant pyramid bands are calculated following Mantiuk et al. [2021].

*Contrast encoding.* The human visual system is attuned to contrast rather than absolute luminance values [Shapley et al. 1993]. After pyramid decomposition, Gaussian and Laplacian pyramids are used to generate a contrast-pyramid following the Weber formula ($\Delta L/L$), in which the feature $\Delta L$ is given by the pixel value at image location $x$, examined at the $i$-th level of the Laplacian pyramid, and the adapting surround luminance $L$ is taken from the Gaussian pyramid at a higher level. Mantiuk et al. [2021] and Tursun et al. [2019] adopt a similar approach to quantify multi-scale contrast. Following Tursun et al. [2019], we represent local adaptation as the value 2 levels higher in the pyramid, as this is nearer to the 0.5° adapting area suggested by the results of Vangorp et al. [2015] for our display.

$$C(x,i) = \frac{L(x,i)}{G(x,i+2)+\epsilon} \tag{1}$$

where $\epsilon$ is a small quantity used to avoid a possible null division.

*Tone mapping.* As we aim to minimize the perceptual difference between a reference and tone mapped image, we need to simulate the tone mapping operation. We can then find ideal tone mapping parameters via an iterative technique. Details on the efficient real-time computation of this step are described in Section 4 and Section 5. For the moment, assume that given a TMO, we calculate the mapped image, and following identical pyramid decomposition and contrast encoding steps as above (depicted in Figure 2 via the red line) we obtain a contrast pyramid representation for our test image.

*Perceptual Scaling.* The relative weight of each level of the contrast pyramid is given by the threshold value found via the contrast sensitivity function (CSF) [Barten 2003], computed for the previously calculated adapting luminance, the level's peak spatial frequency $F(i)$, and area $A(i)$:

$$T(x,i) = 1/\mathrm{CSF}(G(x,i+2), F(i), A(i)) \tag{2}$$

Notably, models like Mantiuk et al. [2021] and Tursun et al. [2019] rely on contrast scaling by the visibility threshold to account for the change in sensitivity between pyramid levels. However, as our application involves minimizing a supra-threshold difference between

Threshold Scaling Contrast Constancy Reference (Clipped)

**Figure 3: Implementations of our method using contrast constancy and threshold scaling are compared. Threshold scaling does not maintain fidelity, making this dark tunnel scene too bright.**

a reference and test image at very different luminances, threshold-based scaling is no longer accurate (see Section 2.3). Instead, we employ Kulikowski's contrast constancy formula [1976]. The resulting relationship with subscripts $r$ and $t$ corresponding to the HDR reference and SDR test is as follows:

$$\tilde{C}_r(x,i) = C_r(x,i) - T_r(x,i) \approx C_t(x,i) - T_t(x,i) = \tilde{C}_t(x,i) \quad (3)$$

Figure 3 shows an ablation comparing these two approaches.

*Contrast Masking.* Finally, following the purple line in Figure 2, we compute the effective difference in perceived contrasts between reference and test at a given pixel location. A contrast masking model (Mantiuk et al. [2021], $p = 2.4$, $k = 0.2854$, $q_c = 3.237$) is used to avoid overestimation of overlapping patterns:

$$D_{r,t}(x,i) = \frac{|\tilde{C}_t(x,i) - \tilde{C}_r(x,i)|^p}{1 + (k\tilde{C}_m(x,i))^{q_c}} \quad (4)$$

where $\tilde{C}_m(x,i)$ is the mutual masking signal given by:

$$\tilde{C}_m(x,i) = min\{|\tilde{C}_t(x,i)|, |\tilde{C}_r(x,i)|\} \quad (5)$$

The result of this operation is a difference energy value $D_{r,t}(x,i)$ for each pixel at each pyramid level. This energy value will increase or decrease depending on how accurately the tone mapping operator was able to reproduce the perception of the reference image's local contrast $\tilde{C}_r(x,i)$ at pixel location $x$ at the pyramid level $i$. Our goal becomes finding the tone mapper that minimizes this difference.

To maintain real-time performance, we avoid a costly temporal frequency decomposition and operate on each frame separately, opting instead for a temporal stabilization scheme described in Section 5.3. Similarly, models like FovVideoVDP [2021] are often used without their temporal component to quantify spatial distortions for images.

## 4 TONE MAPPING

The perceptual framework presented in Section 3 can be applied to any monotonic tone mapping curve. In practice, minimizing the perceived difference to the reference as per Eq. (4) can be costly. A lightweight real-time implementation necessary for VR applications can be achieved by searching along the parameter space of a TMO. Based on our exploration of prior art (see Section 2.1), we chose the Photographic global TMO as our main demonstrator in this work. A short primer follows; we point the reader to the original article [2002] for detail. More TMOs are explored in Section 7.2.

This operator begins by computing the log-mean of the reference luminance $L_r(x)$ for a given pixel location $x$, taking care to add a small quantity $\delta$ to avoid numerical error:

$$\bar{L}_r = \exp(\sum_{x=1}^{N} \frac{\log(\delta + L_r(x))}{N}) \quad (6)$$

A user controllable parameter '$a$' is used to re-scale values:

$$L(x) = \frac{a}{\bar{L}_r} L_r(x) \quad (7)$$

Finally, the tone mapper is applied as follows:

$$L_t(x) = \frac{L(x)}{1 + L(x)} \quad (8)$$

By default, a value of $a = 0.18$ is used, corresponding to the log-mean of a unity-scaled range. Alternatively, the authors suggest manual selection fitting the artistic intent of the user. Our goal is precisely the automatic selection of '$a$' in such a way that the reference image is reproduced as faithfully as possible on the target display.

## 5 REAL-TIME IMPLEMENTATION

Our demonstrator is based on the TMO described in Section 4, and optimized using the framework in Section 3. In this section, we focus on a novel formulation that emphasizes computational efficiency, vital for the system's adoption in standalone VR.

### 5.1 Parameter Optimization

As seen in Figure 2, we need to estimate the test-contrast, which must be done efficiently for a real-time implementation. We address this challenge by implementing a parallel computation on a fragment shader for the Photographic TMO, which does not require memory re-allocation and contrast pyramid regeneration for each iteration. Examining Eq. (7), note that $\bar{L}_r$ is fixed by the reference, and:

$$L_t(x) = L_{\max}(\frac{aL_r(x)}{\bar{L}_r + aL_r(x)}) \quad (9)$$

where $L_{\max}$ is the peak display luminance to which the image is being mapped. By calculating a differential wrt. to $L_r(x)$ and dividing by Eq. (9), we get:

$$\frac{\delta L_t(x)}{L_t(x)} = \frac{\delta L_r(x)}{L_r(x)} \frac{\bar{L}_r}{\bar{L}_r + aL_r(x)} \quad (10)$$

Finally, we express test contrast as a function of reference contrast:

$$C_t(x,i) = C_r(x,i) \frac{\bar{L}_r}{\bar{L}_r + aL_r(x)} \quad (11)$$

It is worth noting that this computation can also be done numerically for any monotonic tone-mapping curve. Now, at each pixel location $x$ and pyramid level $i$, we use this estimate and Eq. (4) to calculate the perceived error for a given value $a(x,i)$. Finally, per-pixel $a(x,i)$ values are pooled globally to obtain a single parameter. After experimenting with pooling techniques (e.g. weighted

Minkowski sums) a simple average was found to have the best performance:

$$\bar{a} = \sum_{x=1}^{N} \sum_{i=1}^{M} (a(x,i)/NM) \tag{12}$$

## 5.2 Implementation Details

We implemented our framework using the scriptable render pipeline in Unity 3D. The implementation is centered around two main fragment shaders: the parameter shader, and the tone-mapping shader.

Assuming a linear 16-bit RGBA float texture, we place the log luminance in the alpha channel and compute MipMaps. The highest mip-level contains mean RGB, as well as the log-mean adapting luminance. The parameter shader then computes the optimal parameter map as per Section 5.1, whose MipMaps are also computed to obtain $\bar{a}$ as per Eq. (12) (using $M = 3$ in our implementation). This value is used by the tone-mapping shader along with the adapting luminance to compute the final tone-mapped output. For stereo rendering, the pipeline is applied to a single cyclopean render between the two views, ensuring that consistent optimized parameters are applied to both eye buffers, mitigating any possibility of binocular rivalry.

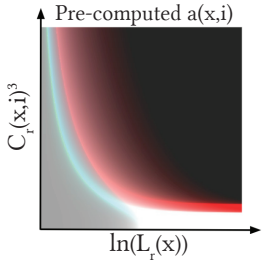


**Figure 4: Rather than re-computing $a(x,i)$ for every pixel, we pre-compute into a small lookup texture, and look up desired values using $L_r(x)$ and $C_r(x,i)$ in the parameter shader. RGB colors are used to store the $M = 3$ bands used.**

In the prototype implementation, our pipeline runs at 5*ms* per frame (2560x1620 resolution) on an Nvidia GeForce 3080 RTX GPU using a binary search to find optimal $a(x,i)$ value at each pixel. To optimize the technique for a stand-alone HMD such as the Quest 2, a key insight is that the per-pixel optimization has only 3 free parameters: mean luminance, local luminance, and local contrast. We leverage this by pre-computing $a(x,i)$ for all possible inputs into either a static 3D texture or a dynamic 2D texture recomputed each frame (as mean luminance varies per frame). To parameterize the lookup texture, we use log-luminances, and contrast to the power of 3. This parameterization allows reducing the lookup texture dimensions to 64 or lower in size without affecting the result. Finally, since the optimization is performed at multiple mip-levels with different frequencies $f(i)$, it is performed $M = 3$ times, and the results are stored in the RGB channels of the lookup texture, as depicted in Figure 4.

This optimization means only a few operations and texture fetches are needed per-pixel. Our implementation on the commercially available Quest 2 VR-HMD runs under 1*ms* per-frame. If access to a linear texture as input (either an eye buffer or an intermediate render pass) is not available; we use a separate cyclopean camera to probe scene luminance, and apply tone-mapping inline while rendering the final eye buffers. The additional render pass adds another 1-2*ms*, if required.

## 5.3 Temporal Consistency

As our method operates on each frame independently, objects entering or exiting the frustum may cause a significant change in tone mapping parameters $a$ and $\bar{L}_r$, causing noticeable temporal inconsistencies with brusque changes of the tone curve. To avoid this, we temporally stabilize these parameters via a leaky integrator model proposed for video extensions of the Photographic TMO by Kiser et al. [2012]:

$$k_{f+1} = \alpha k_f + (1-\alpha) k_{f-1} \tag{13}$$

This method maintains a buffer $k$ for each variable, updated per frame $f$ with a weighted contribution $\alpha$. We ran a small-scale study to test acceptable ranges for this parameter in our framework. 5 users viewed the *sunset* scene in VR, rendered at 100 nits using our TMO. After training, $\alpha$ was varied in a staircase procedure (implemented using psychtoolbox [Kleiner et al. 2007]). Subjects were prompted to answer whether temporal inconsistencies were visible. After 30 responses, the PSE was drawn (see Figure 5 for individual results), for a mean of 0.1026. As this is significantly higher (less stable) than the standard value $\alpha = 0.01$ used in prior art [Goudé et al. 2020; Kiser et al. 2012], we default to the latter value.

It is important to note that the goal of our temporal consistency model was to mitigate brightness flicker, and we did not make targeted efforts to preserve brightness coherency (as defined by Boitard et al. [2014]). A natural trade-off exists between temporal stability (flicker mitigation) and accurate perceptual reproduction of temporal changes: lower $\alpha$ values tend towards temporal stability over accurate reproduction of temporal changes. An example is moving away from a stable frustum, and then returning to it: lower values of $\alpha$ may produce a different instantaneous appearance to the initial one, converging back only after a period of time. An extended implementation may instead aim to preserve brightness ratios during temporal changes (brigntness coherency) after tone-mapping at potential the risk of failing to remove flicker. Exploring the range of the $\alpha$ parameter and overall efficacy of the leaky integrator technique over variables like head-movement speed, scene statistics, and display characteristics is left as future work.

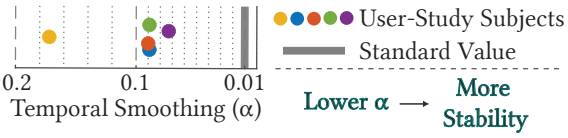


**Figure 5: Our subjective study of the temporal stability parameter.**

# 6 SUBJECTIVE VALIDATION

To validate our technique's effectiveness and understand the impact of tone mapping on quality for VR, we ran a subjective study.

## 6.1 Hardware

Our study was conducted on a hardware setup following Matsuda et al. [2022]. This design has a resolution of 20 ppd over a 62 degree field of view, with a checkerboard contrast ratio around 78:1 and sequential contrast ratio over 400,000:1. Crucially, the platform's LED backlights are controlled via Thorlabs DC2200 power supplies and are calibrated so that we can accurately emulate different backlight luminances via precise current control. This allows us to maintain a constant bit depth and avoids variable pulse width modulation that could introduce perceptual artifacts. This setup can reproduce a peak luminance of over 20,000 nits with a black level of 0.05 nits, though for this study we limit the reference luminance to 5,000 nits to be closer to plausible values for future devices in a headset form factor. During our study, the peak brightness of the device was modulated by adjusting the backlight luminance. This produces a proportional change to the black level, leaving overall contrast ratio unaffected.

## 6.2 Stimuli

Our stimuli consisted of five 360° HDR images (16,384 x 8,192 resolution in latitude-longitude mapping) selected from an online repository of CC0-licensed spherical images [PolyHaven 2023]. During rendering, our method does not have access to any information outside the viewer's current frustum, simulating a real-time rendering scenario. As the source images are uncalibrated, they were manually graded to a maximum value of 5,000 nits. Stimuli used for our study were selected to ensure ample contrast is present to avoid situations where no tone mapping is necessary. Additionally, we attempted to select scenes that are representative of various qualitative scenarios, such as day and night, dark or bright, indoor and outdoor scenes.

Our goal is to demonstrate the effectiveness of our framework in adaptively controlling tone-mapping parameters, so that the visual appearance of the HDR scene is maintained. To avoid qualitative comparisons on distinct families of TMOs, we focus our study on different methods for selecting the '$a$' parameter for the Photographic curve, for which popular heuristics had previously been proposed in the literature. This restriction allows for an objective comparison. The first alternative method is a *fixed* mapping, and follows the recommendation by Reinhard et al. [2002] to set $a = 0.18$; which serves as a useful baseline for comparison. The second alternative is a *heuristic* technique proposed by Krawczyk et al. [2005]. The authors target video tone mapping, and manually tuned a formula to obtain an optimal value for '$a$' based on the mean luminance of the reference. For all three methods, temporal consistency is maintained as described in Section 5.3. Finally, we included the non-mapped HDR *reference* scenes in the set. This helps us ensure that the quality range reported in the study is scaled appropriately, and would allow for the calculation of a DMOS score if desired [Streijl et al. 2016].

In addition to different techniques, we also want to understand the impact of tone mapping for different luminance ranges and test our framework on a wide range of mapping scenarios. With the reference always scaled to a maximum of 5,000 nits, we selected four more values: 50 nits was picked to depict a low-brightness display, possibly exemplifying a power-saving scenario. 100 nits represents a typical luminance found in commercially available headsets [Mehrfard et al. 2019]. Going further, we picked 500 and 1,000 nits as examples of hypothetical high-luminance cases. An example can be seen in Figure 7(a).

In sum, here are all the conditions present in our study:

- **Scenes**: Neon, Car, Street, Tunnel, Hall
- **Luminance**: 50, 100, 500, 1000 (nits)
- **Mapping**: Ours, Fixed, Heuristic, and Reference (the latter presented at 5,000 nits with no mapping)

This amounted to a total of $5 \times 4 \times 3 = 60$ trials, with an additional 5 reference cases for a total of 65 randomized trials per participant.

## 6.3 Procedure

Following piloting, the study was conducted with 24 participants recruited among colleagues of the authors, who were naïve to the purpose of the study (10F 14M, aged 24-51 with a mean of 35). The study was approved by an external ethics committee, and subjects were screened for normal or corrected-to-normal vision and signed informed consent. One participant experienced a hardware malfunction and was consequently excluded from further analysis.

Participants were instructed to rate how similar scenes appeared in relation to the reference on a scale of 0-10, where 10 is exactly and 0 is nothing alike. In addition, instructions encouraged an unhurried exploration of the scenes and evaluating the scene as a whole (as opposed to choosing one element, like a window or a chair, to focus on). Considering both dark and bright regions and paying attention to the context of the scene (e.g. day or night time) was recommended. Responses were collected using controllers built into the prototype VR display handles, and consisted of easy to use buttons and a scroll wheel for the rating task.

The experiment began with a training session comprised of 10 scenes which showed a comprehensive combination of all the experiment variables and reference cases, for which results were discarded. The entire study took on average 36 minutes per participant, which was judged a good compromise on duration to avoid fatigue. A qualitative interview was conducted post study.

For each individual trial, participants were first shown a clearly marked reference version of each scene. On selecting to proceed, the test version of the scene was displayed. Participants also had the option of returning to the reference scene as many times as necessary during a trial. Whenever scenes were changed, either between trials or between reference and test versions, a 200 ms gray screen was shown to avoid direct comparison and assist with adaptation. This short duration was selected based on bright-to-bright adaptation timing data by Hayhoe et al. [1987]. As our experiment did not contain the much slower bright-to-dark adaptation case, this interval was appropriate and no maladaptation was observed by the authors during extensive testing. A subtle audio chime was played 40 seconds into each comparison to help participants keep track of trial duration, but no time limit was enforced.

## 6.4 Results

The results of our study are shown in Figure 6. On the x-axis, the maximum luminance for each trial is shown. The y-axis depicts the mean opinion score of users for each condition. The horizontal dashed line at the top shows the value for the references, which

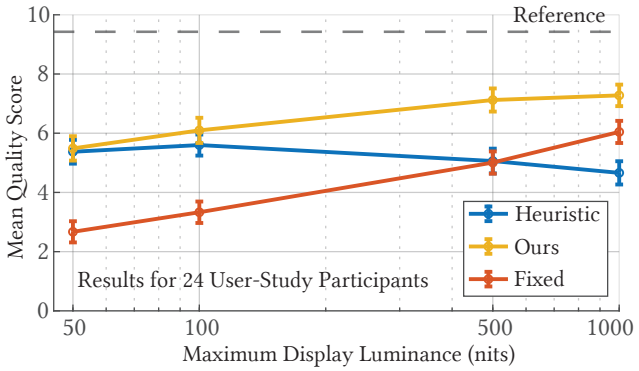


**Figure 6: The results of our subjective study (Section 6.4). Vertical lines depict mean standard error for all participants.**

were themselves included as test scenes in a subset of trials, and unsurprisingly obtained a nearly perfect mean opinion score of 9.7.

The orange line represents the *fixed* baseline tone mapping parameter. As this method is unable to make even simple distinctions for different scenes, it was rated at a low 2.6 for the 50 nit condition, but the score gradually rises to 5.9 for 1,000 nits, as luminance approaches the 5,000 nit reference. The blue line represents the *heuristic* algorithm for the selection of tone mapping parameters, and shows interesting and unintuitive behavior. At 50 nits, it obtains an excellent score of 5.2, clearly improving upon the baseline, and further rising to 5.6 at 100 nits. However, as luminance grows further the quality experiences a reversal, dropping to 4.5 at 1,000 nits, first being matched and then overtaken by the *fixed* condition. The authors observed that this method often over-estimated the key value for a faithful representation of scenes at the top end of the luminance scale, resulting in excessively lifted black levels. Sample results are shown in Figure 1 and Figure 7. Note how the heuristic algorithm makes night-time scenes appear overly bright, revealing excessive detail of regions meant to be in shadows.

Finally, the yellow line shows the results for our method, which significantly outperforms both alternatives (ANOVA analysis: $F = 176$, $p \ll 0.01$). Our adaptive method successfully mapped to the entire range of luminances effectively. It beats the *fixed* baseline at the low end of the range, selecting appropriate parameter to optimally represent contrast in the frustum for a mean improvement of 2.3 points on our quality scale accross the range. It also avoids the mischaracterization of the scenes by the *heuristic*, keeping shadow regions dark even when mapping was done to 1,000 nits maxima, and providing a mean improvement of 1.4, centered especially around brighter mappings.

This study leads us to conclude that when displaying HDR scenes on an SDR headset, choices in tone mapping can be extremely important. For example, *our* tone mapper at a low 100 nits maximum (Q=5.9) was seen as preferable to either *fixed* (Q=5.8) or *baseline* (Q=4.5) conditions at 1,000 nits, despite a tenfold increase in display brightness. As brighter VR displays are released, it will be critical to ensure that content is appropriately and automatically mapped to make optimal use of content properties and display capabilities.

# 7 EXTENSIONS & APPLICATIONS

So far, we have presented the application of our perceptual framework for global tone-mapping in VR. In this section, we demonstrate other important applications of our framework.

## 7.1 Artistic Guidance

In our framework, each pixel has the same perceptual weight. A prioritization scheme could be devised where pixels are weighted based on artistic intent. During the pooling phase of Eq. (12), values of the tone-mapping parameter can be weighted by an artist-generated mask, which could be used to increase the weight of regions considered important or reduce unimportant ones (see Figure 9, left).

## 7.2 Alternative TMOs

Our framework can be employed for any monotonic TMO. To demonstrate, Figure 8, shows the application of our method to popular tone-mapping curves by Drago et al. [2003] and Tumblin et al. [1999]. The fixed parameter comparison was implemented based on the work of Aydin et al. [2014], who use manually tuned values for these two operators in their framework. It is worth noting that the choice of TMO can be artistic in nature, and many modern methods rely on traditional TMO curves to achieve the desired results: Chuang et al. [2002] employ Tumblin's TMO after base-detail decomposition by bilateral filtering. VR-oriented methods by Najaf et al. [2017] and Goude et al. [2020] use the Photographic TMO [2002] with a fixed $a = 0.18$, equivalent to the *fixed* baseline described in Section 6.2. Our perceptual framework can be used to augment these methods by providing real-time content-aware parameter tuning, regardless of the base TMO employed.

## 7.3 Local Tone Mapping

Global tone mappers employ a single curve to map the entire image, while local tone mappers may employ different curves on regions of the frustum with the goal of increasing local contrast. The latter can be beneficial to retain visibility in scenes where significant contrast is present, but also incurs some risks: increasing local contrast may reduce overall contrast. In addition, strong edges may be affected by the local region size, resulting in perceptible haloing.

Our real-time tone mapping implementation described in Section 5 can be easily modified to produce a local version of the Photographic TMO. Instead of globally pooling the optimized values of $a$, each value is pooled across levels of the pyramid, but not across different pixel locations (see Figure 9, right). Note that this has no computational overhead compared to the global version. We have provided visual examples of local tone-mapping applications in Figure 10, but a controlled user study to quantify improvements over alternatives is left as future work.

*7.3.1 Local Tone Mapping Region Size.* The size of the region used for local tone mapping has a significant effect on visual appearance. Small region sizes fail to maintain global contrast, while very large region sizes converge toward global tone mapping, as shown in Figure 10(left). Our method can be applied to adaptively estimate the region size for faithful representation.

*7.3.2 Global & Local TMO Blending.* Another popular technique for good tone reproduction is combining global and local tone-mapping by blending using a control parameter $k$:

$$I_{out} = k \cdot f_{\text{TMO}}^{\text{GLOBAL}}(I_{in}) + (1-k) \cdot f_{\text{TMO}}^{\text{LOCAL}}(I_{in}) \tag{14}$$

We can apply our framework to adaptively control the blending parameter for optimal visual reproduction of the HDR image after tone mapping (shown in Figure 10, right).

## 8 CONCLUSIONS AND FUTURE WORK

We propose a real-time perceptual framework that can be used to adaptively minimize the perceived difference between an HDR image and its tone-mapped counterpart. We apply this framework to design a real-time, temporally consistent tone mapping parameter optimizer for VR. Finally, we demonstrate extensions of our implementations for local tone mapping and alternative TMOs.

We run a validation study demonstrating that not only does our framework maintain the appearance of the HDR reference, but the right application of a tone mapper can dramatically increase the perceived quality of high-contrast scenes when presented on SDR displays. Future work could build on this foundation by finding optimal tone-mapping methods for VR. In addition, our framework maintains global contrast across the field-of-view, but a gaze-tracked foveated version giving more importance to foveal regions is a possible extension: only frequencies visible to the user would be considered in the computation [Tursun et al. 2019] by adapting the formulation in Section 3 for a foveated CSF (e.g. stelaCSF [2022]).

## ACKNOWLEDGMENTS

We thank the artists who made their HDRI images available. *Car*: Andreas Mischok, *Street*: Dimitrios Savva & Jarod Guest, *Neon*, *Tunnel*, and *Hall*: Greg Zaal. We thank Fartun Sheygo, Cameron Wood, and Helen Ayele for their assistance in running the subjective study. A special thank you to Ken Koh for his invaluable help in setting up the user-study functionality.

## REFERENCES

2016. Preparing for Real HDR. https://developer.nvidia.com/preparing-real-hdr.

Maliha Ashraf, Rafal Mantiuk, Jasna Martinovic, and Sophie Wuerger. 2022. Suprathreshold contrast matching between different luminance levels. In *Color Imaging Conference (CIC)*.

Tunç Ozan Aydin, Nikolce Stefanoski, Simone Croci, Markus Gross, and Aljoscha Smolic. 2014. Temporally coherent local tone mapping of HDR video. *ACM Transactions on Graphics (TOG)* 33, 6 (2014), 1–13.

Peter G. J. Barten. 2003. Formula for the contrast sensitivity of the human eye. In *Proc. SPIE 5294, Image Quality and System Performance*, Yoichi Miyake and D. Rene Rasmussen (Eds.). 231–238. https://doi.org/10.1117/12.537476

Ronan Boitard, Rémi Cozot, Dominique Thoreau, and Kadi Bouatouch. 2014. Zonal brightness coherency for video tone mapping. *Signal Process. Image Commun.* 29 (2014), 229–246.

Yung-Yu Chuang, Alexei A. Efros, John Ruskin, and Frédo Durand. 2002. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics* (2002).

Frédéric Drago, Karol Myszkowski, Thomas Annen, and Norishige Chiba. 2003. Adaptive Logarithmic Mapping For Displaying High Contrast Scenes. *Computer Graphics Forum* 22 (2003).

Gabriel Eilertsen, Rafał K Mantiuk, and Jonas Unger. 2015. Real-time noise-aware tone mapping. *ACM Transactions on Graphics (TOG)* 34, 6 (2015), 1–15.

Ific Goudé, Rémi Cozot, and Olivier Le Meur. 2020. A perceptually coherent TMO for visualization of 360 HDR images on HMD. In *Transactions on Computational Science XXXVII: Special Issue on Computer Graphics*. Springer, 109–128.

Mary M Hayhoe, Norma I Benimoff, and DC Hood. 1987. The time-course of multiplicative and subtractive adaptation process. *Vision Research* 27, 11 (1987), 1981–1996.

Chris Kiser, Erik Reinhard, Mike Tocci, and Nora Tocci. 2012. Real time automated tone mapping system for HDR video. In *IEEE International Conference on Image Processing*, Vol. 134. 2749–2752.

Mario Kleiner, David Brainard, and Denis Pelli. 2007. What's new in Psychtoolbox-3? (2007).

Grzegorz Krawczyk, Karol Myszkowski, and Hans-Peter Seidel. 2005. Perceptual effects in real-time tone mapping. In *Proceedings of the 21st spring conference on Computer graphics*. 195–202.

JJ Kulikowski. 1976. Effective contrast constancy and linearity of contrast sensation. *Vision research* 16, 12 (1976), 1419–1431.

Rafał Mantiuk, Scott Daly, and Louis Kerofsky. 2008. Display adaptive tone mapping. In *ACM SIGGRAPH 2008*. 1–10.

Rafał K. Mantiuk, Maliha Ashraf, and Alexandre Chapiro. 2022. StelaCSF: A Unified Model of Contrast Sensitivity as the Function of Spatio-Temporal Frequency, Eccentricity, Luminance and Area. *ACM Transactions on Graphics (TOG)* 41, 4 (2022).

Rafał K. Mantiuk, Gyorgy Denes, Alexandre Chapiro, Anton Kaplanyan, Gizem Rufo, Romain Bachy, Trisha Lian, and Anjul Patney. 2021. FovVideoVDP: A Visible Difference Predictor for Wide Field-of-View Video. *ACM Trans. Graph.* 40, 4 (2021).

Nathan Matsuda, Alex Chapiro, Yang Zhao, Clinton Smith, Romain Bachy, and Douglas Lanman. 2022. Realistic Luminance in VR. In *SIGGRAPH Asia 2022 Conference Papers*. 1–8.

Arian Mehrfard, Javad Fotouhi, Giacomo Taylor, Tess Forster, Nassir Navab, and Bernhard Fuerst. 2019. A comparative analysis of virtual reality head-mounted display systems. *arXiv preprint:1912.02913* (2019).

Miguel Melo, Kadi Bouatouch, Maximino Bessa, Hugo Coelho, Remi Cozot, and Alan Chalmers. 2018. Tone Mapping HDR Panoramas for Viewing in Head Mounted Displays. In *VISIGRAPP (1: GRAPP)*. 232–239.

Hossein Najaf-Zadeh, Madhukar Budagavi, and Esmaeil Faramarzi. 2017. VR+ HDR: A system for view-dependent rendering of HDR video in virtual reality. In *2017 IEEE International Conference on Image Processing (ICIP)*. 1032–1036.

PolyHaven. 2023. HDRIs. https://www.polyhaven.com.

Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. 2010. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann.

Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. 2002. Photographic tone reproduction for digital images. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*. 267–276.

Christophe Schlick. 1995. Quantization techniques for visualization of high dynamic range pictures. In *Photorealistic rendering techniques*. Springer, 7–20.

Robert Shapley, Ehud Kaplan, and Keith Purpura. 1993. Contrast sensitivity and light adaptation in photoreceptors or in the retinal network. *Contrast sensitivity* 5 (1993), 103–116.

Robert C Streijl, Stefan Winkler, and David S Hands. 2016. Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives. *Multimedia Systems* 22, 2 (2016), 213–227.

Jack Tumblin, Jessica K. Hodgins, and Brian K. Guenter. 1999. Two methods for display of high contrast images. *ACM Transactions on Graphics (TOG)* 18 (1999), 56–94.

Okan Tarhan Tursun, Elena Arabadzhiyska-Koleva, Marek Wernikowski, Radosław Mantiuk, Hans-Peter Seidel, Karol Myszkowski, and Piotr Didyk. 2019. Luminance-contrast-aware foveated rendering. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–14.

Peter Vangorp, Karol Myszkowski, Erich W Graf, and Rafał K Mantiuk. 2015. A model of local adaptation. *ACM Transactions on Graphics (TOG)* (2015).

David R Walton, Rafael Kuffner Dos Anjos, Sebastian Friston, David Swapp, Kaan Akşit, Anthony Steed, and Tobias Ritschel. 2021. Beyond blur: Real-time ventral metamers for foveated rendering. *ACM Transactions on Graphics* 40, 4 (2021), 1–14.

Robert Wanat and Rafał K Mantiuk. 2014. Simulating and compensating changes in appearance between day and night vision. *ACM Transactions on Graphics (TOG)* 33, 4 (2014), 1–12.

Minqi Wang and Emily A Cooper. 2021. A Re-examination of Dichoptic Tone Mapping. *ACM Transactions on Graphics (TOG)* 40, 2 (2021), 1–15.

Zhuming Zhang, Xinghong Hu, Xueting Liu, and Tien-Tsin Wong. 2018. Binocular tone mapping with improved overall contrast and local details. In *Computer Graphics Forum*, Vol. 37. Wiley Online Library, 433–442.

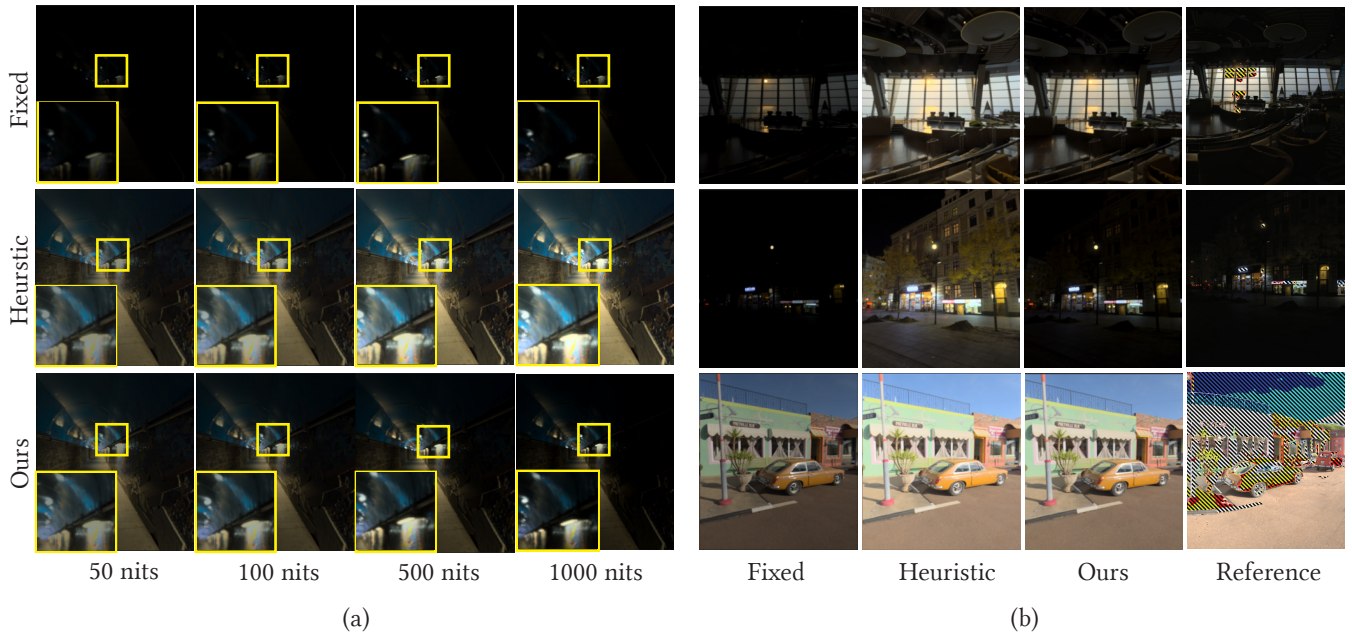


**Figure 7: (a): A sample of all the luminance and mapping combinations in our study. The *fixed* method makes most of the scene dark, especially at low luminances. The *heuristic* parameter estimation tends to make the background overly bright for this dark *tunnel* scene. *Our* method adapts the tone-mapping to different display luminances effectively, maintaining the visual appearance of the HDR scene. Note that we cannot accurately depict the scenes in this document format. In this figure, images are linearly scaled for presentation.**

**(b): Methods described in Section 6.2 are compared when mapping to 100 nits. The *fixed* method fails to faithfully represent the ambient, while the *heuristic* tends to over-expose dark scenes. *Our* method faithfully represents arbitrary scenes through adaptive perceptual control, as demonstrated in our subjective study. Note that HDR scenes cannot be faithfully represented in this document format as they were seen by our users. The rightmost column shows the unmapped HDR references, with highlights above 100 nits clipped and marked with dashes.**

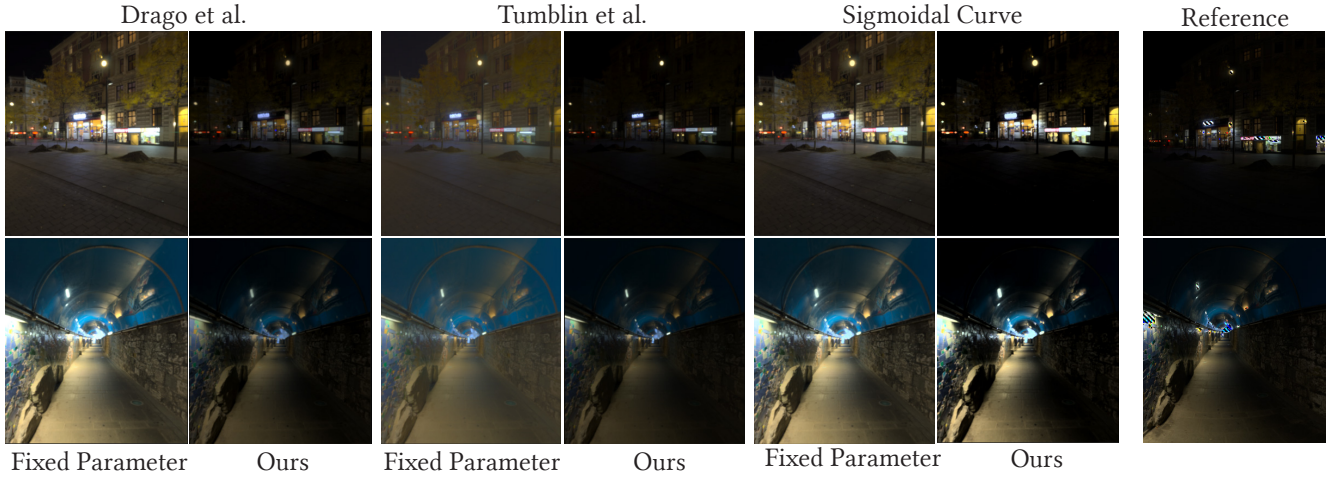


**Figure 8: Our framework can be applied to any monotonic TMO. Here, we showcase results on the Drago [2003] and Tumblin [1999] mappers (see Section 7.2). In addition, a simple sigmoidal curve described by Reinhard [2010] (page-291) is used with $a = 0.18$ and $b = 1$ as the baseline, and $b$ optimized in our version. Our framework achieves consistent results across tone-mapping techniques, obtaining a faithful representation of the HDR reference.**

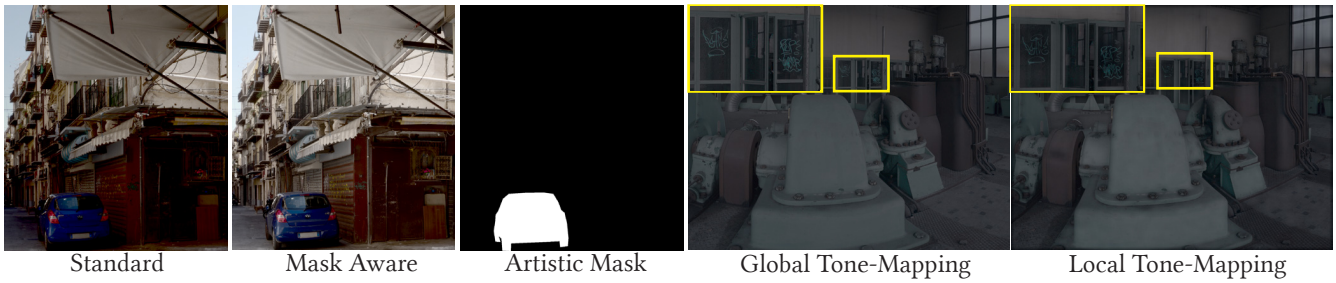


**Figure 9: (Left) Artistic intent provided via an optional weight mask guides parameter selection, increasing the weight given to the car. (Right) The global method applies the same curve for the entire frustum, while the local method aims to preserve local contrast**


**Figure 10: (Left) The region size of a local tone mapper is optimized using our framework, as described in Section 7.3.1. (Right) Linear blending of local and global tone mapping, with the blending coefficient optimized using our method as described in Section 7.3.2. For an objective comparison, all images in this figure were tone-mapped with a fixed '$a$' parameter of the Photographic TMO, and differences are due only to changes in region size (Left) and blending (Right).**