TONE-MAPPING REQUIREMENTS IN REAL-TIME VIDEOS FOR STUDYING THE DYNAMISM OF VIEWS-OUT IN VIRTUAL REALITY

Yunni Cho[1], Stephen Wasilewski[1], Caroline Karmann[2], Marilyne Andersen[1]
[1] École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland,
[2]Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
yunni.cho@epfl.ch

Abstract

Current view representation methods for Virtual Reality (VR), monitors, and projection screens predominantly rely on static photographs and tone-mapping operators without temporal variations, limiting the exploration of dynamic features in a view. Capturing views with real-time video overcomes this limitation but has a limited dynamic range and relies on an unverified built-in tone-mapping process. This study investigates the adequacy and reliability of using real-time video captures of dynamic daylit views in VR. To evaluate the suitability of the camera's built-in tone-mapping procedures, a comparison to normative tone-mapping operators is conducted with a focus on brightness and contrast, based on pixel-by-pixel luminance values from calibrated high dynamic range images. This study finds that real-time video reliably maintains relative contrast, enabling more comprehensive investigations in future studies involving dynamic movements and temporal changes in VR. Furthermore, the proposed method provides the foundations to establish a workflow generalizable to new experimental approaches for the validation of tone-mapping in VR environments.

Keywords: Daylighting Scenes, Views Out, Tone-Mapping, Virtual Reality, Real-Time Videos

1 Introduction

Access to a satisfactory view of the outside is crucial for the health and well-being of building occupants (Beute and De Kort, 2013). Previous research has demonstrated that certain visual elements, such as natural landscapes (Kaplan, 1995), distant features (Matusiak and Klöckner, 2016), and diverse visual elements (Li and Samuelson, 2020), can enhance the perceived quality of a view-out and improve occupant satisfaction. Although dynamic movement and temporal changes in the content of views-out have been suggested to further improve the perceived view quality by fostering occupants' awareness and connection to their surroundings (Ko et al., 2021), motion and temporal changes are largely missing from most studies to date on this topic. Existing view rating metrics and representation methods typically rely on static, time-independent views, failing to address the significance of dynamic features in a view (Rodriguez et al., 2021).

To address this gap in view-out research, our study assesses the suitability of a novel workflow that employs real-time videos and Virtual Reality (VR) technology to capture dynamic daylit views i.e., daylit scenes with changes in view content over time (whether from agentive movements or changes in the lighting conditions). VR technology has significantly advanced the study of light perception and of views out, offering immersive depth perception and three-dimensional stereoscopic vision (Abd-Alhamid et al., 2019). Compared to experiments conducted in physical spaces, VR provides greater control over environmental factors, reproducibility, a broader range of visual content options, and greater consistency in experimental conditions. However, VR experiments face limitations from its hardware characteristics, such as the available luminance range, the screen resolution, and the field of view. The realism in VR is further compromised due to constraints on movement and limited interactions within a virtual environment that extend beyond mere observation. These limitations tend to hinder the ability of VR technology to accurately replicate physical environments (Bellazzi et al., 2022), particularly when scenes involve natural light and its dynamics.

To overcome the limitations related to luminance range, researchers often prepare images for VR display by capturing High Dynamic Range (HDR) photographs and then compressing them to a reduced dynamic range (Low Dynamic Range or LDR) using a variety of tone-mapping operators (TMOs). Tone mapping is the process of translating HDR values to the range of the output device or format to best reproduce the perception of the real view. For conducting experiments where the output image is used as a proxy for the real scene, the suitability of a TMO depends on its ability to mimic the stimulus of the source data. These tone curves can be either functional (such as linear, logarithmic, or sigmoid) or histogram-based (distributing output according to the input distribution) and may include localized adjustments to maintain contrast (Eilertsen et al., 2017). While TMOs are capable of recreating real scenes on output devices more effectively than a camera's native tone curve (Petit et al., 2013), there are certain challenges associated with capturing dynamic movement. Due to limited sensor capabilities and exposure time needed for producing HDR images, most cameras struggle to capture and process changing daylight conditions in real-time videos (Eilertsen et al., 2017).

In a prior pilot study by the same authors (Cho et al., 2023), we compared perceptual impressions of window views from a real office and the same views projected in VR using the proposed scene collection method with 34 subjects. The study's outcomes suggested that combining real-time videos with a framing contextualization provided by a scale model yielded perceptually accurate view-out experience in VR. Although these results are promising with regards to the potential of this method to produce a satisfactory subjective appraisal, a further evaluation of the method was warranted in terms of its influence on actual pixel values. We thus conducted a direct comparison study between HDR photographs tone-mapped using reference TMOs and the same scenes tone-mapped in real-time video output using the camera's automatic settings (i.e., the method used in the pilot study).

Our present study proposes a way to incorporate VR in daylighting and view-out research without losing the ability to represent temporal dynamics. The aim is to understand the trade-off between perceptual accuracy of the images presented to the viewer and technical feasibility of producing videos. To investigate this trade-off, we assessed the performance of each TMO by analyzing its capacity to maintain contrast and brightness in VR environments.

A critical aspect of researching perceptual impressions using images displayed in VR is to ensure that the relationship between the features and details of a scene closely matches how the same scene would appear in reality. Visual difference prediction models (VDP) address this issue by analyzing the differences between HDR source data and output display values (Daly, 1993). A dynamic range invariant VDP can be employed to assess perceptible differences while accounting for the highly adaptable nature of the human visual system (Aydin et al., 2008). By conducting a comparative analysis using this model in the context of the output display device, we aim to quantify perceived differences in brightness and contrast between the real and VR scenes, thereby offering valuable insights into perceptual accuracy for VR projection.

2 Methods

To establish a baseline for the capabilities offered by current tone-mapping approaches, we first needed to examine the performance of TMOs in terms of their ability to accurately replicate the relationships between scene features and details when displayed in VR, in comparison to how the same scene would be perceived in reality.

2.1. Scene collection

A set of ten view-out scenes was produced for this purpose. Our method for collecting those view-out scenes relied on a Canon R5 camera equipped with a Canon RF 5.2m F2.8L Dual Fisheye lens. This lens, featuring a 60mm interpupillary distance, approximates human vision and natural parallax, which enables capturing accurately aligned 180-degree stereoscopic videos with resolutions up to 8K. During scene collection, outdoor illuminance values were recorded with an LMT lux meter positioned next to the camera lens and HDR photographs were generated from multiple exposure series of photographs taken before and after collecting the video footage using the exact same camera setup (see Figure 1). The illuminance sensor readings were used to ensure consistent conditions throughout the capture of both the HDR images and the video.

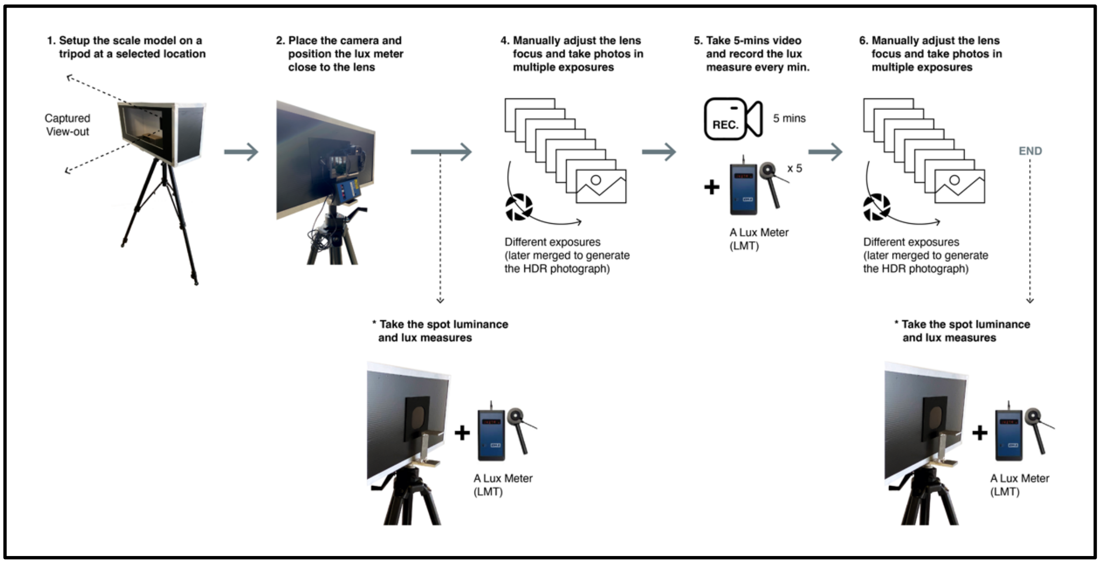

Figure 1 – Scene collection method

To identify potential limitations of our method, we consciously chose a variety of different weather conditions and environmental contexts for these ten scenes, so as to embed enough variability when comparing the impact of TMOs on their pixel values. As can be seen in Figure 2, each of the ten scenes includes sky, ground, and landscape layers but differ in view content and covered a wide range of luminance, going from 2,400-36,000 cd/m2 for their respective brightest pixel. Note that while the scenes encompass different variations of overcast to clear skies, we made sure that direct sun never enters the field of view or the model. We considered this selection of ten scenes thus representative enough of conditions relevant for conducting VR experiments on views-out.

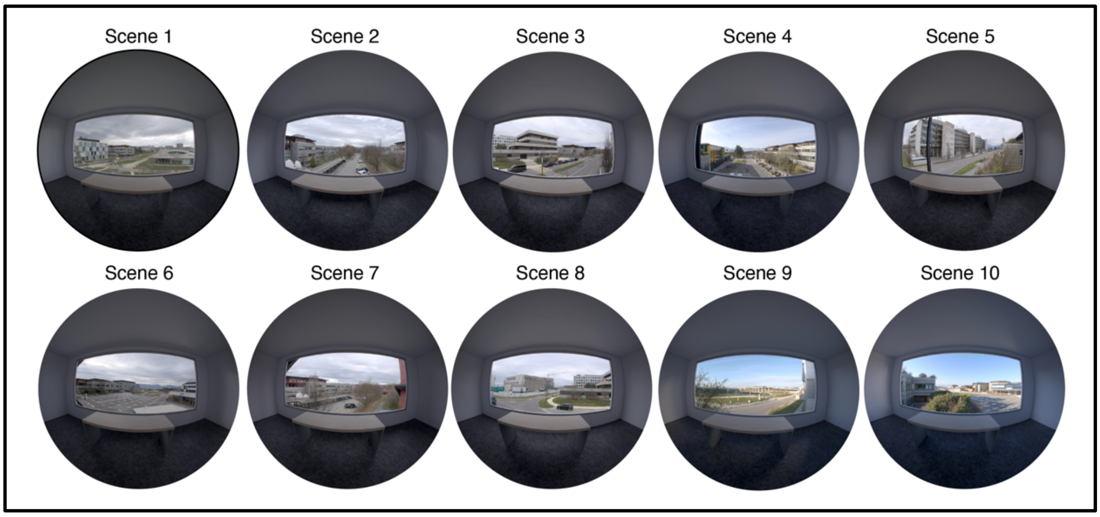

Figure 2 – Selected scenes

2.2. Tone mapping comparisons

Nine TMOs were then selected to conduct performance comparisons with the built-in video camera output (hereafter referred to as Video), based on their coverage of a wide range of algorithmic approaches, their usage in previous VR research (Petit et al., 2012; Melo et al., 2018; Chamilothori, 2019), and their availability through open-source implementations (*Radiance* and *pfstools*). Seven of these TMOs are sourced from publications (Ward, 1997; Pattanaik, 2000; Durand 2002; Reinhard, 2002; Drago, 2003; Reinhard, 2005; and Mantiuk, 2008) and are labeled by primary author and two-digit year for brevity. The last two TMOs, Reinhard02L and Reinhard05L, include a local adaptation algorithm of their original implementation.

We used the HDR images of the ten scenes as the “ground truth” reference to evaluate the performance of the nine TMOs and Video. We compared the output images based on their ability to maintain perceptible contrast using a VDP called the contrast invariant visual difference metric or CIVDM[1] (Aydin et al., 2008) that was updated to include a validated visual difference predictor (Mantiuk et al., 2023). This metric combines visual system models and psychophysical data to report the probability of tone-mapped image detecting or failing to detect contrast on a per-pixel basis, given that it would be visible in the HDR image displayed on an ideal HDR output device. The comparison is made between the raw HDR values and the tone-mapped values as they will appear on the output device. To estimate the output device values, we measured the screen luminance of the VR headset (Pico Neo3 Pro Eye) with a tripod mounted luminance meter while viewing eleven blank images with different lightness values ranging from black to white. The source values of the grey images and recorded luminance values were linearly interpolated to form an output response curve. Multiplying this curve against each TMO output yields the luminance values of input images as observed on the VR display.

We performed the comparison on the captured fisheye projection (single eye, see Figure 3A), where pixels represent nearly equal solid angles in the output display. This approach avoids bias towards the floor and ceiling, which are stretched in the equirectangular projection. The results are presented as the distribution of pixels according to their detection probabilities, ranging from negative one for complete loss of detection to one for complete amplification of undetectable differences in the source HDR, see Figure 3B for an example image. To gain a deeper understanding of the performance results of tone-mapping operators, we created four additional masks on the fisheye image, dividing the area into wall, view (window opening), desk, and floor sections (indicated in Figure 3C). This segmentation allows for a more nuanced understanding of how the TMOs perform in different parts of the scene. Using CIVDM, we first evaluated the entire images to assess the overall performance and then focused on view mask to assess specific performance of different TMOs for view-out applications. We also compared the average luminance across different mask regions (view to wall, wall to floor, view to desk) to measure the contrast preservation between higher-level scene elements.

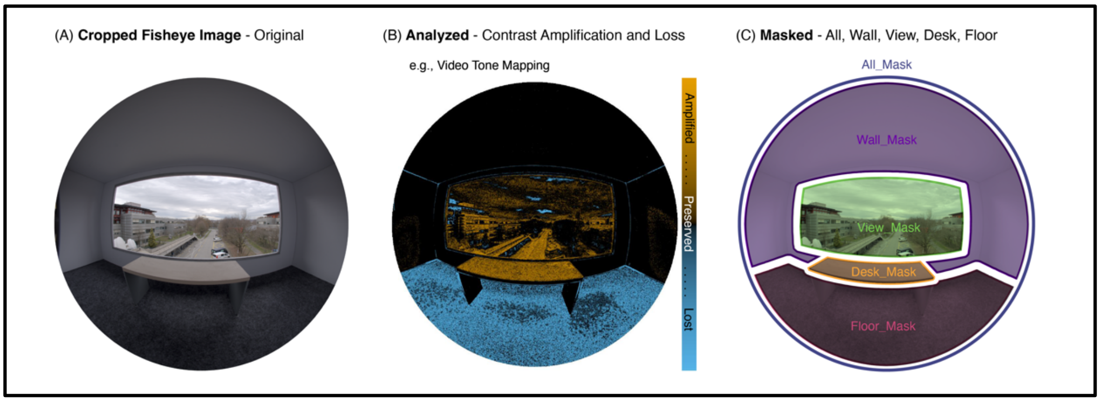

Figure 3 – Components of the comparative analysis

Several important considerations were taken into account when preparing the images for comparison. First, since all TMO implementations operate on rectangular image spaces and some TMOs construct their response curves based on image histograms, it was essential to exclude the masked corners of the image outside the fisheye. To achieve this, all images were transformed to square coordinates using a low distortion area preserving transformation (Shirley and Chiu, 1997) before applying the TMO and then transformed back, as illustrated in Figure 4, top. The reference images were subjected to the same transformation but remained otherwise unaltered to prevent the introduction of interpolation errors. Second, the actual

[1] CIVDM was calculated using Matlab code HDR-VDP-3 version 3.0.6 (hdrvdp.sourceforge.net).

captured video had different source data from the HDR sequence due to the presence of moving clouds, trees, cars, and people, resulting in a significant number of artificial differences when comparing the video still to the reference HDR. Instead of using the actual video output, we recovered the video tone curve by comparing the video still to the HDR, using only the values between the 40th and 60th percentile to eliminate physical mismatches. A smoothed running average was then fitted to these values (see Figure 4, bottom). The comparisons were run on the fisheye images, and the regions outside the fisheye were excluded from the results. This approach facilitated a more accurate assessment of the images while addressing the inherent challenges of comparing dynamic scenes.

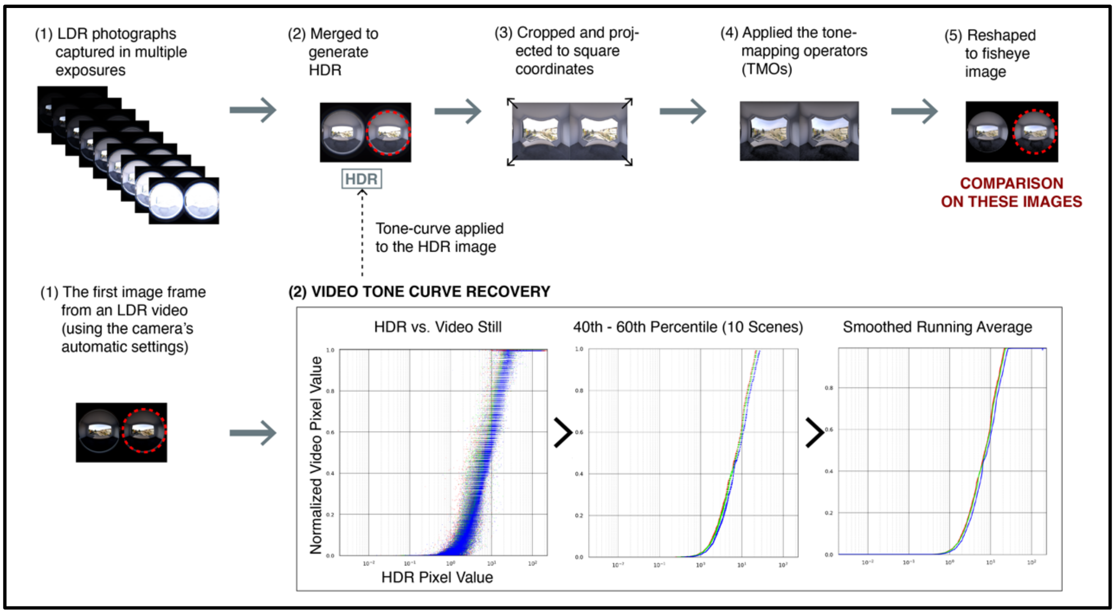

Figure 4 – Input image transformation for comparative analysis

3 Results

In this study, we evaluated the performance of various TMOs and camera's automatic tone-mapping for capturing dynamic views-out, with the aim of assessing the potential and limitations of camera-based tone-mapping as an efficient method. This section first analyzes the images in their entirety and then offers a detailed examination of the performance results within specific regions of the image.

Figure 5 illustrates the average distribution of contrast loss, preservation and amplification across all ten scenes, as determined by the CIVDM for each pixel (Figure 3C). Thresholding these distributions at +0.1 (a 10% chance of seeing an imperceptible difference) and -0.1 (a 10% chance of missing a perceptible difference) as indicated by the grey band visible in Figure 5A, we can evaluate each TMO based on how much of the scene information is lost, preserved and amplified (Figure 5B). We can see that Ward97, Reinhard02(L), and Drago03 tend to emphasize details the CIVDM model suggests would not be perceptible in the real scene, but also that they exhibit minimal contrast loss. Meanwhile, Video, Reinhard05, and Pattanaik00 display little contrast amplification but substantial contrast loss. The remaining TMOs, Durand02, Reinhard05L, and Mantiuk08, have less than an average of 10% loss and 10% amplification across all scenes. These results generally align with the specific intent and approach of each TMO. Ward97 maximizes the available range by allocating contrast according to the image distribution. Reinhard02 and Drago03 choose to prioritize the best subjective quality, which tends to favor more detail. The contrast loss observed for Pattanaik00 and Reinhard05 can be explained by the fact that they visualize the contrast loss based on adaptation and veiling processes in the human eye, thereby producing images of human perception rather than for inducing a realistic perceptual response. Lastly, it is expected that Video would have the highest contrast loss, given its operation on a more limited dynamic range of input data.

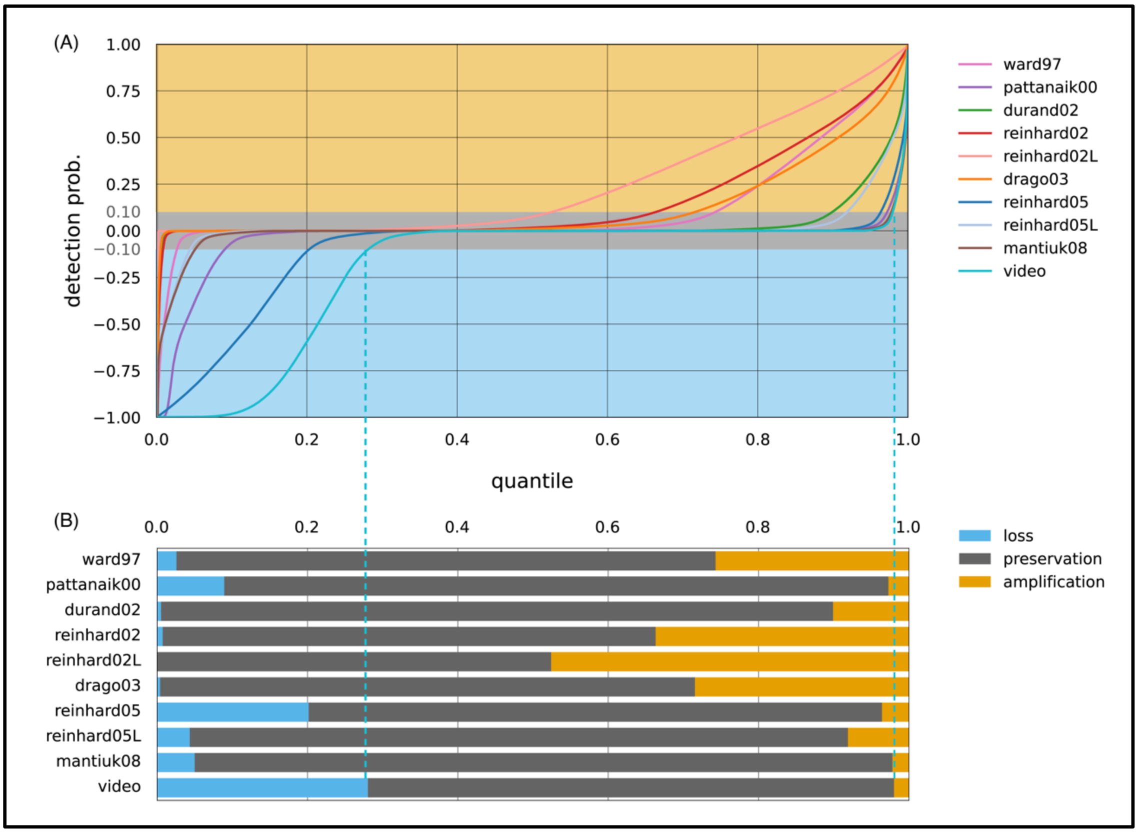

Figure 5 – Tone-mapping comparisons of the entire image area averaged across 10 scenes (acceptance threshold of +0.1 for contrast amplification and -0.1 for contrast loss)

As shown in Figure 6, this contrast loss from Video is primarily concentrated on the floor. Other TMOs exhibit similar tendencies to either overpredict or underpredict contrast within specific ranges or features of the image (Figure 6). This highlights an important caveat in interpreting whole image values such as those in Figure 5: the CIVDM VDP is calibrated for an observer's central foveal acuity. Given that the primary purpose of these images is to evaluate the view out the window, it is unlikely that observers will focus their gaze on variations in the carpet pattern. To address this issue, Figure 7 shows the same set of plots, but only for the area of the image encompassing the window (see Figure 3B). Aside from Pattanaik00, which has more contrast loss in the view, the other TMOs maintain or reduce contrast loss, with Video and Reinhard05 showing significantly better performance than for the entire view. In the view area, Ward97, Reinhard02 and Drago03 all have less contrast amplification compared to the whole scene. Except for Reinhard02L, all other TMOs and Video have similar or improved performance in the view area. The view area is typically the most dynamic and has the highest contrast within the region; thus, it makes sense that the Video performs as well as the other TMOs in this area. The camera metering appears to favor this central viewing area, as indicated by the balance of contrast loss and amplification. Pattanaik00, on the other hand, is adapted to the entire scene and does not favor one region over the other, resulting in the loss of contrast in this region with more dynamic content. Similar to Video, Ward97 also performs better in the view area than the surrounding image. Due to the large areas of even lighting in the interior portions of the scene, the Ward97 tone-curve allocates a larger proportion of the output range to these lower values, making small details, such as the grain of the carpet and texture on the walls, hyper-visible.

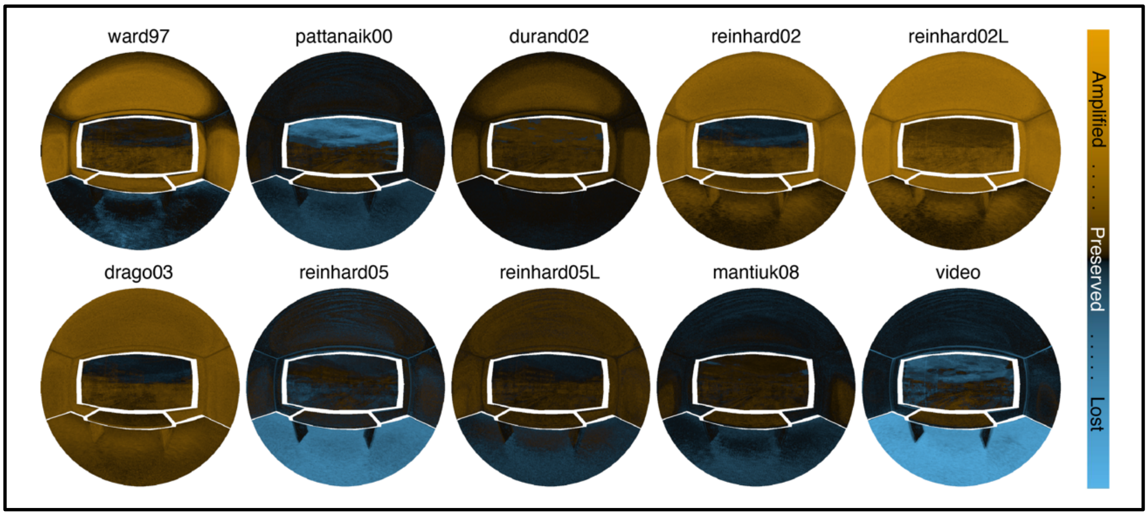

Figure 6 – Average Contrast Loss and Amplification across all scenes for each TMO, segmented by view region.

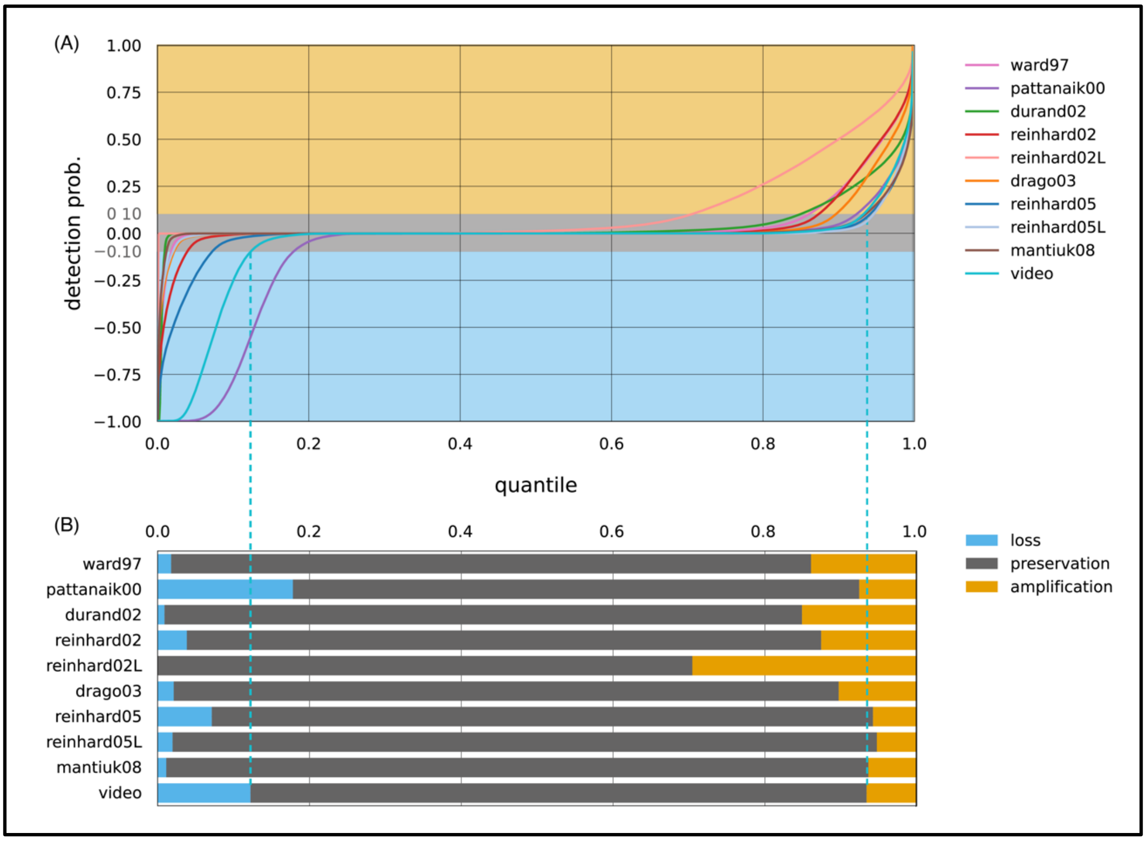

Figure 7 – Tone-mapping comparisons of the view area averaged across 10 scenes (acceptance threshold of +0.1 for contrast amplification and -0.1 for contrast loss)

Figure 8 displays the contrast preservation at the 90% level for each view and TMO, across the entire image (all) and the view area. In the examined scenes, specific to evaluating views from window-adjacent spaces, Mantiuk08 demonstrates the highest performance. The automatic camera tone-mapping of the video achieves a performance comparable to the full range of TMOs, preserving contrast for more than 80% of the image in over half of the scenes. This underlines the potential of camera-based tone-mapping as an efficient method for maintaining crucial visual information in view-out scenes.

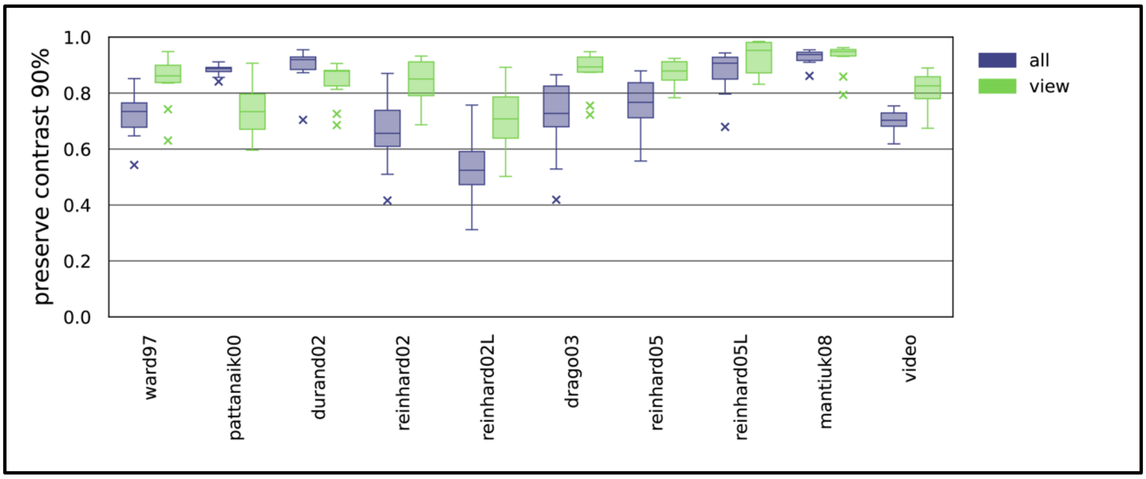

Figure 8 – Comparative analysis between the entire image area and view area. Boxes show the distribution of contrast preservation across all scenes.

4 Discussion and Limitations

When evaluating the results above, it is essential to acknowledge that the camera's built-in tone-mapping results in a large amount of contrast loss in darker regions (wall, floor, and desk). These areas also appear perceptibly darker compared to the view when using other TMOs. Consequently, this leads to amplified contrast ratios between these higher level scene features. Figure 9 illustrates the relative difference in contrast ratio between the reference HDR and TMO outputs for the ratios view to wall, wall to floor, and view to desk. Interestingly, it is the local TMO Reinhard02L that comes closest to maintaining these higher-level contrast ratios. Nonetheless, no TMO avoids scenes with more than twice or less than half the contrast of the reference HDR image. This reflects the inherent limitations of the low dynamic range output of the VR headset, which has a total available range of 0.2 to 86 cd/m2 (430:1). For reference, the ratio between the maximum luminance and average floor luminance in each image ranges from 320:1 to 7,700:1 with a median contrast of 1,200:1. While this result is specific to the camera we used, its S-shaped tone curve is typical for other LDR cameras (Petit et al., 2013). Likewise, the output range of the VR headset used in this study is also representative of other commercially available headsets. These results, therefore, are indicative of what is currently achievable with the current state of digital camera and VR headset technologies.

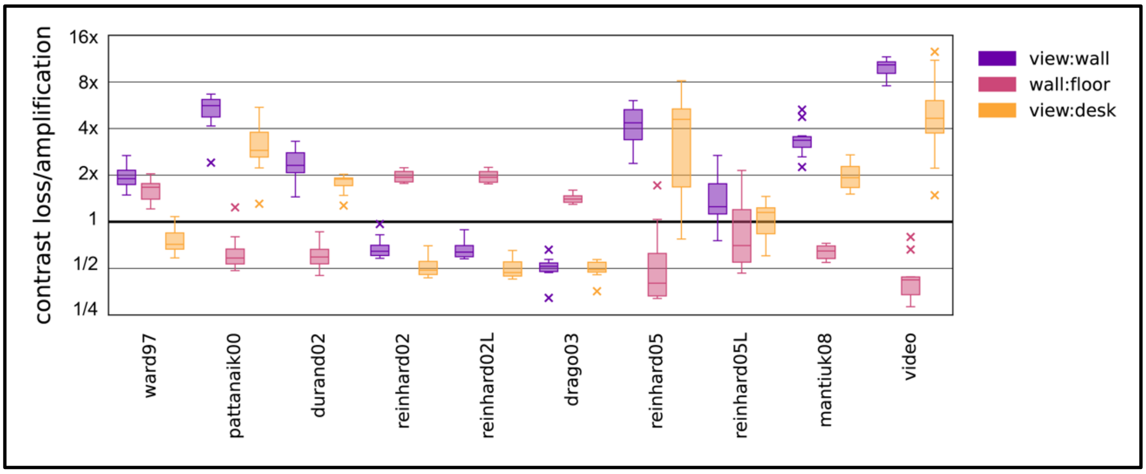

Figure 9 – Comparative analysis between the view, wall, floor, and desk regions. Boxes show the distribution of contrast preservation across all scenes.

Hence, the choice of a tone-mapping approach should be carefully considered depending on the specific focus and objectives of the experiment. For experiments emphasizing visualization of the view content and the efficient capture of dynamic views, the camera's automatic tone-mapping offers a suitable solution. However, if maintaining contrast between regions for an accurate representation of an interior environment is a priority, further research is required to determine whether any tone-mapping can adequately reproduce the higher scale contrast ratios necessary for such an assessment.

5 Conclusion

This study presents the first attempt to assess the suitability of using real-time videos in VR for studying the perception of daylit views-out, utilizing original data based on illuminance and luminance measurements. Our research explores the applicability of a simplified visualization tool for consistently representing daylit window views from office environments in a controlled manner. Through comparative studies, we have evaluated both the strengths and limitations of a workflow that relies on the camera's automatic tone-mapping procedures rather than tone-mapping constructed HDR images to represent dynamic views. At the same time, our research underscores the usefulness of the CIVDM as a tool for comparing TMOs, with an emphasis on the human visual system, utilizing HDR images as a benchmark. Moreover, our methodology further underlines the importance of considering contrast amplification alongside contrast loss as another measure of perceptual accuracy when comparing HDR to tone-mapped images.

The findings from our analysis suggest that our camera's automatic tone-mapping paired with our headset can effectively serve as a reliable method for capturing dynamic views as real-time videos in view-out studies. The embedded tone-mapping preserves contrast levels within the window region, thereby making it comparable to alternative TMOs. Moreover, the use of the camera's Low Dynamic Range (LDR) outputs, as opposed to HDR video, offers significant advantages in terms of video quality, time, and effort. The LDR outputs deliver higher frame rates necessary for capturing fluid motion, reducing the need for extensive post-processing, thereby streamlining the workflow in view-out research. The reduced time and effort associated with utilizing the camera's LDR outputs and automatic tone-mapping make it a more practical and cost-effective solution for both researchers and practitioners.

The combination of preserving contrast and the practical benefits of using LDR outputs establishes the camera's automatic tone-mapping as a valuable tool in view-out research. Our work suggests that directly captured LDR outputs perform similarly to tone-mapped HDR outputs for capturing window view contents. As the dynamic range of cameras and VR headsets expands with further technological developments, the direct capture workflow will continue to improve. Meanwhile, our findings open new experimental possibilities for reliably representing dynamic movements and temporal changes in view-out studies. Simultaneously, our analysis can be repeated with other scenes, cameras, and display outputs, including VR headsets and monitor screens, to further test suitability of low dynamic range hardware for studies on visual perception. This provides an innovative approach to assess perceptual accuracy and crucial parameters of realism using computer-based visual perception models like CIVDM prior to conducting human subject experiments. Future research can build upon these findings by investigating additional factors that influence perceptual accuracy and realism of view-out scenes across both physical and virtual environments.

References

ABD-ALHAMID, F., KENT, M., BENNETT, C., CALAUTIT, J., WU, Y. 2019. Developing an Innovative Method for Visual Perception Evaluation in a Physical-Based Virtual Environment. Building and Environment, 162, 106278.

AYDIN, T.O., MANTIUK, R., MYSZKOWSKI, K., SEIDEL, H.P. 2008. Dynamic range independent image quality assessment. ACM Trans. Graph, 27, 1–10.

BEAUTE, F. and DE KORT, Y.A.W. 2013. Let the Sun Shine! Measuring Explicit and Implicit Preference for Environments Differing in Naturalness, Weather Type and Brightness. Journal of Environmental Psychology, 36, 162–178.

BELLAZZI, A., BELLIA, L., CHINAZZO, G., CORBISIERO, F., D'AGOSTINO, P., DEVITOFRANCESCO, A., FRAGLIASSO, F., GHELLERE, M., MEGALE, V., SALAMONE, F. 2022. Virtual Reality for Assessing Visual Quality and Lighting Perception: A Systematic Review. Building and Environment, 209, 108674.

CHAMILOTHORI, K. 2019. Perceptual Effects of Daylight Patterns in Architecture. Lausanne: EPFL PhD thesis.

CHO, Y., KARMANN, C., ANDERSEN, M. 2023. A VR-Based Workflow to Assess Perception of Daylit Views-Out with a Focus on Dynamism and Immersion. Proceedings of the CISBAT International Scientific Conference.

DALY, S. 1993. The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity, Digital Images and Human Vision (179–206). Cambridge: MIT Press.

DRAGO, F., MYSZKOWSKI, K., ANNEN, T., CHIBA, N. 2003. Adaptive Logarithmic Mapping For Displaying High Contrast Scenes. Computer Graphics Forum, 22, 419–426.

DURAND, F. and DORSEY, J. 2002. Fast Bilateral Filtering for the Display of High-Dynamic-Range Images. ACM Transactions on Graphics. 257–266.

EILERTSEN, G., MANTIUK, R.K., UNGER, J. A. 2017. Comparative Review of Tone-Mapping Algorithms for High Dynamic Range Video. Computer Graphics Forum, 36, 565–592.

KAPLAN, S. 1995. The Restorative Benefits of Nature: Toward an Integrative Framework. Journal of Environmental Psychology, 15, 169–182.

KO, W.H., KENT, M.G., SCHIAVON, S. LEVITT, B., BETTI, G. 2021. A Window View Quality Assessment Framework. LEUKOS, 18, 268–293.

LI, W. and SAMUELSON, H. 2020. A New Method for Visualizing and Evaluating Views in Architectural Design. Developments in the Built Environment, 1, 100005.

MANTIUK, R., DALY, S., KEROFSKY, L. 2008. Display adaptive tone mapping. ACM Trans. Graph, 27, 1–10.

MANTIUK, R.K., HAMMOU, D., HANJI, P. 2023. HDR-VDP-3: A multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content.

MATUSIAK, B.S. and KLÖCKNER, C.A. 2016. How We Evaluate the View Out Through the Window. Architecture Science Review, 59, 203–211.

MELO, M., BOUATOUCH, K., BESSA, M., COELHO, H., COZOT, R., CHALMERS, A. 2018. Tone Mapping HDR Panoramas for Viewing in Head Mounted Displays. Science and Technology, 232–239.

PATTANAIK, S.N., TUMBLIN, J., YEE, H., GREENBERG, D.P. 2000. Time-Dependent Visual Adaptation for Fast Realistic Image Display. Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques.

PETIT, J., BRÉMOND, R., TOM, A. 2012. Evaluation of Tone Mapping Operators in Night-Time Visual Worlds. 2013. Virtual Reality, 17, 253–262.

PETIT, J. and MANTIUK, R.K. 2013. Assessment of Tone-Mapping: Are Cameras' S-Shaped Tone-Curves Good Enough? Journal of Visual Communication and Image Representation, 21, 1020–1030.

REINHARD, E., DEVLIN, K. 2005. Dynamic Range Reduction Inspired by Photoreceptor Physiology. IEEE Trans. Visual. Comput. Graphics, 11, 13–24.

REINHARD, E., STARK, M., SHIRLEY, P., FERWERDA, J. 2002. Photographic tone reproduction for digital images. ACM Transactions on Graphics, 267–276.

RODGRIGUEZ, F., GARCIA-HANSEN, V., ALLAN, A., ISOARDI, G. 2021. Subjective Responses Toward Daylight Changes in Window Views: Assessing Dynamic Environmental Attributes in an Immersive Experiment. Building and Environment, 195, 107720.

SHIRLEY, P. and CHIU, K. 1997. A Low Distortion Map Between Disk and Square. Journal of Graphics Tools, 2, 45–52.

WARD, G., RUSHMEIER, H., PIATKO, C. 1997. A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes. IEEE Transactions on Visualization and Computer Graphics, 3, 291–306.