Preprint: Article in *Journal of Digital Video*, vol. 5, no. 1, pp. 74-83, Dec. 2020

Diagnosing Visual Quality Impairments in High-Dynamic-Range/Wide-Color-Gamut Videos

Article to the Editor prepared for SCTE•ISBE
by

Zhou Wang, PhD, Professor, University of Waterloo
Chief Scientist, SSIMWAVE Inc.
200 University Ave W, Dept. of ECE, University of Waterloo
Waterloo, Ontario, N2L 3G1, Canada
zhou.wang@uwaterloo.ca
519-888-4567 ex. 35301

Hojatollah Yeganeh, PhD, Senior Member Technical Staff, SSIMWAVE Inc.
375 Hagey Boulevard, Suite 310
Waterloo, Ontario, N2L 6R5, Canada
hojat.yeganeh@ssimwave.com
519-489-2688

Kai Zeng, PhD, Senior Member Technical Staff, SSIMWAVE Inc.
375 Hagey Boulevard, Suite 310
Waterloo, Ontario, N2L 6R5, Canada
kai.zeng@ssimwave.com
519-489-2688

Jiheng Wang, PhD, Senior Member Technical Staff, SSIMWAVE Inc.
375 Hagey Boulevard, Suite 310
Waterloo, Ontario, N2L 6R5, Canada
jiheng.wang@ssimwave.com
519-489-2688

1. Introduction

Content producers have been making great progress in the past few years adopting advanced technologies to create video content of ultra-high definition (UHD), high dynamic range (HDR), and wide color gamut (WCG). Meanwhile, there has been an accelerated growth of UHD/HDR/WCG content being delivered to consumers’ home TVs and mobile devices. While consumers are enjoying the improved resolution/contrast/brightness/bit depth and enhanced richness of colors, maintaining the quality in video production and distribution becomes an ever greater challenge that content producers and cable, satellite, IPTV and OTT video distribution service providers have to face. The challenge arises not only because of the much larger bandwidth that is required to guarantee seamless UHD/HDR/WCG video transmission, but also because of the difficulty in preserving the visual quality during video production, post-production and delivery, and in maintaining consistent presentation over a wide variety of display devices at the studios and at the user ends.

It is important to note that HDR/WCG video content is much more prone to quality issues than Standard Dynamic Range (SDR) content of smaller color gamut, not only because of the higher user expectations, but also because any loss of texture/highlight/shadow details, any shifts of exposures, colors and skin tones, and any artifacts of blocking and banding, could be much more manifest on HDR/WCG displays and visual settings. Moreover, the perceptual visibility of such artifacts could vary depending on the display devices and settings. Taking banding, an annoying artifact frequently occurring in HDR/WCG content, as an example, its visibility could vary drastically across scenes and across viewing devices. To detect banding, a good understanding about the human visual system, the content attributes, the display characteristics, and the interplay between them is crucial. To remove or reduce banding, the video scenes and their associated metadata need to be processed, encoded and decoded in a consistent manner throughout the video production and distribution workflows.

Here we discuss the common quality issues in HDR/WCG video production and distribution. We will then discuss how such quality issues may be diagnosed through an end-to-end quality control framework, for which the most critical technology is an objective quality assessment method that can help identify, localize and assess the visual quality impairments in HDR/WCG videos.

2. Quality Issues in HDR/WCG Videos

HDR/WCG video distribution should aim for preserving the creative intent of content producers. Ideally, the video quality at any point along the video delivery chain should be compared against the pristine video graded by the artists, colorists and directors of photography at the grading suites. The journey of the source HDR/WCG videos from the grading suites to end users’ viewing devices consists of multiple sophisticated stages, and quality issues may arise in any of these stages. A brief summary of the key stages is discussed below.

Mezzanine preparation. Immediately after color grading, motion pictures often go through the process of generating mezzanine files. Figure 1 demonstrates a typical workflow in studios in mezzanine file generation. Once grading content is done, an Output Display Transform (ODT) defined for a target delivery platform is applied to encode linear RGB using perceptual quantization (PQ) Opto-Electrical Transfer Function (OETF) [1]. The output of this stage is PQ encoded RGB or equivalently R’G’B’ in 16bits TIFF format. Assuming P3 color gamut [2] is used for grading and a theatrical ODT is used, there is a need for converting P3 color gamut to Rec. 2020 [3], which is the container of HDR videos in the distribution pipeline. Eventually, an intra frame coding is used to quantize and compress HDR frames. If the codec does not support R'G'B' color space like Apple ProRes, there is also a color space conversion involved in the process. Almost all such transformations are lossy and could potentially create visual quality impairment.

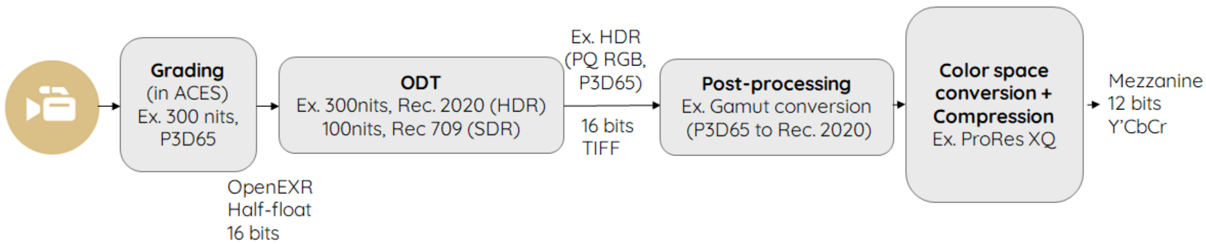

Figure 1 – An example of a practical workflow to generate a mezzanine file.

Transcoding and transmission. HDR Mezzanine files are often of high bitrate that cannot fit into the pipeline of subsequent video delivery chains. Therefore, transcoders such as HEVC, together with bit-depth mapping (often from 12 bits to 10 bits) and chroma subsampling, are typically used to further compress mezzanine files, leading to even stronger visual artifacts.

Display on user devices. With the advent of HDR filming techniques, display manufacturers sought to support HDR EOTFs and roll out displays that are capable of producing luminance above 100nits. Displays play an important role in the rendition of HDR videos. Consumer TVs and other viewing devices have huge variations in terms of hardware and software design, which is in addition to the largely varying physical limitations on the display technologies being used. As a result, the same HDR video stream may have significantly different visual appearances on different displays under different display modes, often creating dramatic deviation from what the content producers see at studios on calibrated professional displays.

HDR formats. A variety of HDR formats/standards have been used in practice for delivering HDR content, among which the common ones include HDR10 [1], HLG [4], Dolby Vision [5] and HDR 10+ [6]. These standards differ in the EOTF transfer function (e.g., PQ vs. Hybrid Log-Gamma), allowed bit depths (10 bits vs 12 bits), metadata formats (static vs. dynamic), and color representations. All of these variations are entangled with the variations in color grading platform, mezzanine file generation, transcoding technology, and display design, creating significant variations in the appearance of HDR content.

Eventually, human eyes are the ultimate consumers and the final judges of the visual quality of HDR/WCG videos. The most common visual quality impairments include:

- *Texture detail loss* is often caused by compression and bit depth reduction. In addition, HDR displays that cannot accommodate content peak brightness, often apply tone mapping techniques to reduce the dynamic range of the HDR content, leading to texture detail loss.

- *Blocking* is often produced by quantization in heavy compression, where artificial blocky structures may be produced and their visibility varies depending on the underlying texture content near the blocky regions.

- *Banding* is a common artifact that often appears in large regions of low textures and slow gradients, where large smooth regions are divided into flat bands with long and visually strong fake contours separating them. One complication in banding is that its visual appearance may vary not only depending on the bit-depth quantization and compression, but also on the noise level, the denoising effect of compression, and the HDR formats and displays.

- *Highlight and shadow detail loss* refers to the reduction of visibility of fine details in the brightest and darkest regions. They may be caused by any mismatch between the maximum/minimum content brightness, the OETF and EOTF functions, the bit-depth mapping algorithm, the maximum/minimum brightness of the display, and the physical limitation of the display technology. Highlight and shadow detail loss is critically important because the purpose of HDR technology is to improve the capability to expand the scene dynamic range, such that fine details in bright and dark regions that are invisible in SDR are clearly discerned.

- *Color distortion and skin tone deviation*. Color gamut mapping, color space conversion, chroma subsampling/upsampling, mismatching OETF and EOTF transformations, and compression may all contribute to color distortions, for which skin tone deviation is often the most noticeable.

- *Tone mapping artifacts*. In video distribution practice, we often encounter a mismatch between the higher dynamic range of video content (e.g., 10bits or 12bits HDR) and the lower dynamic range capability of displays (e.g., 8bits SDR), tone mapping has to be applied for bit-depth conversion, which is a source of severe quality degradations. Visually, such degradations may appear to be a combination of the loss of structural details, the loss of naturalness, and the loss of temporal consistency and smoothness.

3. Diagnosing Visual Quality Impairments

HDR/WCG video is prone to quality degradations throughout the video delivery chain. The most effective approach for quality control is to monitor it using an end-to-end framework [7], where quality testing probes are deployed at all transition points in the delivery chain, so that any quality deterioration can be captured and localized in a timely manner. The most important technology to enable such an end-to-end quality control system is an objective video quality assessment (VQA) metric that faithfully reflects the perceptual video quality. Such objective metrics need to be validated through comparisons with subjective testing on dedicated HDR/WCG video databases [8]. Unfortunately, conventional SDR VQA metrics, such as PSNR, SSIM [9], VQM [10] and VMAF [11], do not produce accurate assessment of HDR/WCG videos. This has inspired significant effort in the past years to overcome the challenge, represented by advancements over the SSIM method like SSIMPLUS [12], [13]. Meanwhile, the VMAF project has also been continuously making progress along this direction [14]. Here we show how some of the most significant visual impairments in HDR/WCG videos may be diagnosed.

Banding significantly degrades the visual Quality of Experience (QoE) of end users. Banding effect is particularly annoying as it frequently exhibits even in high definition, high bitrate HDR/WCG content, which otherwise appears to have nearly perfect quality. What often frustrates many practitioners is that simply increasing bit-depth and bitrate of a video does not necessarily lead to removal or reduction of banding. Figure 2 (left) gives an example of a video frame with severe banding artifacts in the background region. Two substantially different types of approaches have shown notable success at detecting banding. The first approach is based learning deep neural network (DNN) predictors from large-scale datasets [15], where convolutional neural networks (CNNs) are trained in an end-to-end manner to classify local image patches into having or not having banding, and the local predictions are then aggregated to produce a global banding estimation of an image or video frame. The approach has been tested to be highly accurate when trained on large-scale databases, though the banding diagnosis does not go deep into pixel level. In addition, such a purely learning based approach does not offer much meaningful insight about why and how banding is happening, insight that may help correct or reduce banding. The second type of approach is based on detecting abrupt local activities in smooth image regions, and then analyzing the visibility of such activities from the perspectives of human visual system characteristics, display modes and properties, OETF and EOTF transfer features, color features, and viewing environment. Pixels that correspond to abrupt activities deemed visible as banding are then marked, which collectively constitute a banding map of the image or video frame. An example of such a banding map is shown in Fig. 2 (right), where the banding artifacts are highlighted by the white pixels. It appears that this knowledge-driven approach (as opposed to the learning-based DNN approach) not only detects banding, but also precisely localizes the banding impairment at pixel-precision.

Figure 2 – Video frame (left) with banding artifacts, and banding map (right) created by an objective banding detection algorithm.

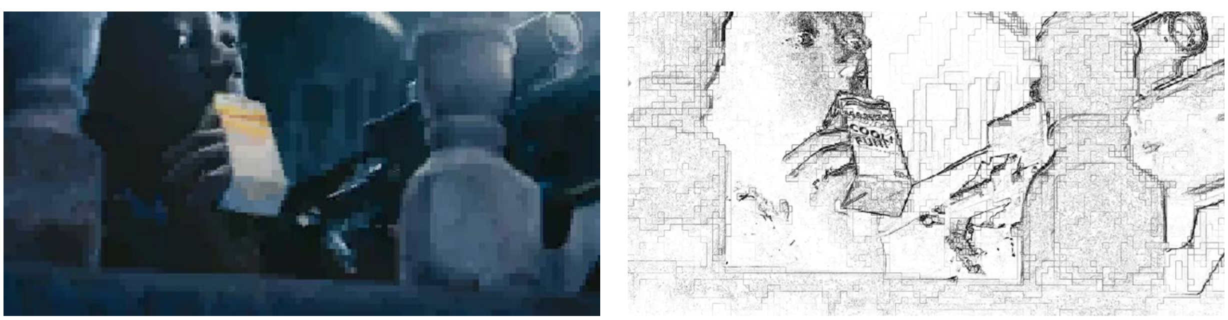

Figure 3 – Video frame (left) with strong blocking artifacts, and quality map (right) created by an objective video quality metric.

Blocking is mainly caused by block-based video compression, which is a common technology adopted in a majority of video coding standards used in practice. Unlike banding, blocking may occur in both smooth and texture regions of an image, as exemplified in Figure 3 (left). Blocking may be diagnosed by either blind (or no-reference) or non-blind (or full-reference) video quality metrics – the former detect blocking artifacts without referencing to the distortion-free original video, while the latter uses the original video as a reference. Successful blind blocking effect detection methods often quantify the magnitude of block edges based on human visual system features such as the texture masking effect [16]. In general, blind blocking detection methods are more convenient to deploy in practical systems, but less accurate, as image content may contain blocky features too. Non-blind or full-reference video quality metrics, especially those that measure the local structural similarities [9] [12] between the reference and test images, may be able to detect blocking effects in their quality maps. An example of such a quality map is given in Fig. 3 (right), where darker pixels indicate stronger visual artifacts. It can be observed that the blocking artifacts created by block-based hybrid video compression schemes are well detected and precisely localized on the quality map.

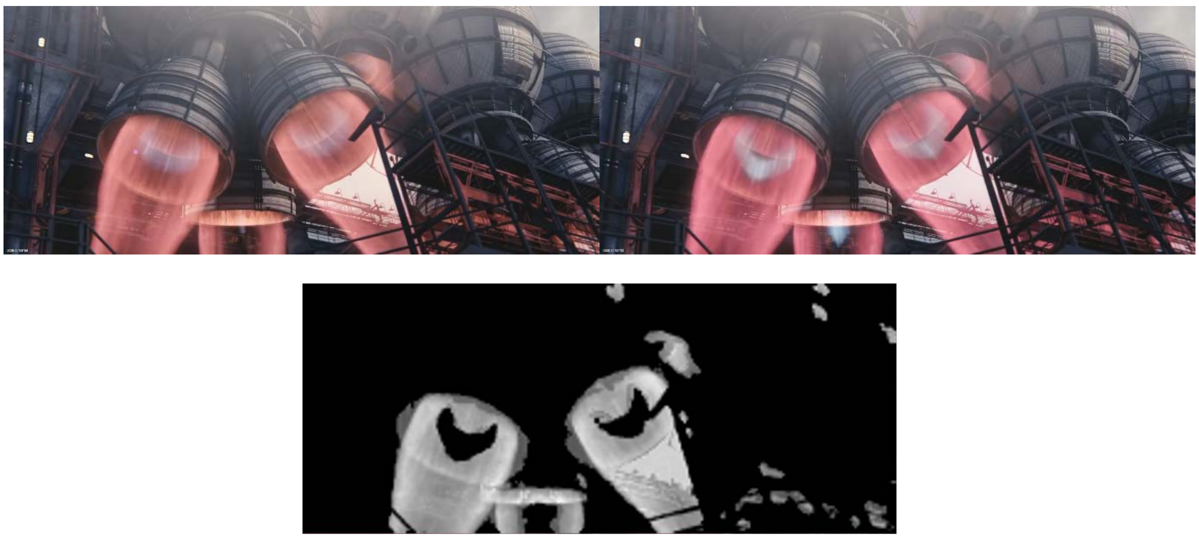

Figure 4 – Frames from the source (left) and degraded test (right) videos, and the color deviation map (bottom) created by an objective metric.

Due to the sophisticated video production and distribution workflows where color changes are often non-invertible (e.g., in color gamut mapping, color space conversion, chroma subsampling/upsampling, OETF and EOTF transformations, bit-depth conversion, and compression) and the studio professional displays and consumer viewing devices are largely mismatched, reproducing the exact colors seen at grading suites on common consumers' viewing devices is deemed impossible. However, there is a large playground on how to reduce the color deviations, for which the first step is to assess the color disparity in a perceptual meaningful way. One common approach towards this goal is to develop a perceptually uniform color space, within which ideally the same level of numerical change leads to the same level of visual difference and such uniformity should hold across the whole color space. It should be noted that none of the color representations used for creation/exchange of mezzanine files or transmission of compressed video streams, e.g., PQ-encoded R'B'G' or YCbCr, were derived for perceptual color uniformity. There has been a hundred-year effort of designing color spaces of perceptual uniformity [17] [18]. Some recent study suggests that great progress has been achieved over the years but there is still significant space for improvement [19]. Figure 4 compares two sample video frames, one from the source and the other from the test (degraded) videos, together with the color deviation map created by an objective video quality metric, where brighter regions signal higher levels of color deviations. Such color deviation maps play an important role in identifying and assessing how color/skin tone information is preserved during video distribution.

HDR to SDR tone mapping often produces complicated quality issues that have to be captured/described in multiple dimensions in terms of the loss of structural fidelity, the loss of naturalness, and the loss of temporal consistency and smoothness, with interactions between them. Specifically, to avoid fine structural details being washed out via bit-depth reduction, tone mapping operators strive to preserve the structural details, but the process often produces unnatural looking textures and colors. It is important for a video quality metric to be able to capture unnatural appearance of tone-mapped images. A joint consideration of both of the structural fidelity and statistical naturalness factors leads to the tone-mapped image quality index (TMQI) [20], which has been widely used in both industry and academic research for assessing the performance of tone mapping operators, and has also been successfully used to drive the optimal design and optimal fusion of tone mapping operators [21]. Figure 5 depicts two examples of visual artifacts generated through HDR to SDR tone-mapping. The left column shows images created from different tone-mapping parameters, and the right column is the corresponding visual distortion map, where darker regions indicate more significant information loss. When tone mapping operators are applied to videos, cross-frame consistency becomes a critical factor. Inconsistent inter-frame tone mapping often creates severe flickering across frames and unnatural luminance change across scenes. Thus a joint consideration of spatial ante temporal artifacts is desirable in objective video quality metrics [22].

Figure 5 – Example of tone-mapping artifacts and the structural fidelity map created by an objective video quality metric.

As HDR/WCG video becomes increasingly more popular, new artifacts will emerge that will pose new challenges to the HDR/WCG VQA metric. Continuous efforts need to be spent on studying the impact of the new artifacts and developing updated objective VQA algorithms to capture those impacts.

4. Conclusions

As the HDR/WCG video promises unprecedented viewer experience to consumers, quality control becomes crucial to ensure such promises deliver. Nevertheless, quality control of HDR/WCG content production and distribution is challenging, much more so as compared to delivering video content of SDR and smaller color gamut. Here we discuss the reasons that cause visual quality impairments in the production and distribution workflows of HDR/WCG videos, and lay out the common visual quality artifacts. We have also discussed the end-to-end quality control framework and the recent effort at developing dedicated objective video quality metrics that are effective at diagnosing the visual quality impairments in HDR/WCG videos. These methods and discussions pave the ways for the development of reliable quality control and quality optimization systems for HDR/WCG video production and distribution.

5. Bibliography and References

[1] *BT.2100: Image parameter values for high dynamic range television for use in production and international programme exchange*, International Telecommunication Union. 2018.

[2] *The C.I.E. colorimetric standards and their use*, Smith, Thomas, Guild, John. Transactions of the Optical Society. 1931-1932.

[3] *BT.2020: Parameter values for ultra-high definition television systems for production and international programme exchange*, International Telecommunication Union. 2014.

[4] *High Dynamic Range*, European Broadcasting Union. 2015.

[5] *An Introduction to Dolby Vision*. Dolby Vision, Dolby Laboratories. 2016.

[6] *HDR10+ System Whitepaper*, HDR10+ Technologies, LLC. September 2019.

[7] *Begin with the end in mind: a unified end-to-end quality-of-experience monitoring, optimization and management framework*, Z. Wang and A. Rehman, SMPTE Motion Imaging Journal, 2019.

[8] *Perceptual colour difference uniformity in high dynamic range and wide colour gamut*, T. Costa, V. Gaudet, E. R. Vrscay, and Z. Wang, IEEE International Conference on Image Processing, Sept. 2020.

[9] *Image quality assessment: from error visibility to structural similarity*, Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, IEEE Transactions on Image Processing, Apr. 2004.

[10] *A new standardized method for objectively measuring video quality*, M. H. Pinson, *IEEE Transactions on Broadcasting*, Sept. 2004.

[11] Toward a practical perceptual video quality metric, Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy and M. Manohara, Netflix TechBlog, Jun 2016.

[12] *Display device-adapted video quality-of-experience assessment*, A. Rehman, K. Zeng and Z. Wang, IS&T/SPIE Electronic Imaging: Human Vision & Electronic Imaging, Feb. 2015.

[13] *SSIMPLUS: The most accurate video quality measure*, https://www.ssimwave.com/from-the-experts/ssimplus-the-most-accurate-video-quality-measure/

[14] *VMAF: the journey continues*. Z. Li. C. Bampis, J. Novak, A. Aaron, K. Swanson, A. Moorthy and J. De Coc, Netflix TechBlog, 2019.

[15] *Capturing banding in images: database construction and objective assessment*, A. Kapoor, J. Sapra and Z. Wang, IEEE International Conference on Acoustic, Speech and Signal Processing, Jun 2021.

[16] *Blind measurement of blocking artifacts in images*, Z. Wang, A. C. Bovik and B. L. Evans, IEEE International Conference on Image Processing, Sept. 2000.

[17] *How the CIE 1931 color-matching functions were derived from wright-guild data*, H. S. Fairman, M. H. Brill, and H Hemmendinger, Color Research & Application, Feb. 1997.

[18] *Hitting the mark – A new color difference metric for HDR and WCG imagery*, E Pieri and J Pytlarz, SMPTE Motion Imaging Journal, 2018.

[19] *Perceptual quality assessment of UHD-HDR-WCG videos*, S. Athar, T. Costa, K. Zeng and Z. Wang, IEEE International Conference on Image Processing, Sept. 2019.

[20] *Objective quality assessment of tone mapped images*, H. Yeganeh and Z. Wang, IEEE Transactions on Image Processing, Feb. 2013.

[21] *High dynamic range image compression by optimizing tone mapped image quality index*, K. Ma, H. Yeganeh, K. Zeng and Z. Wang, IEEE Transactions on Image Processing, Oct. 2015.

[22] *Objective quality assessment of tone-mapped videos*, H. Yeganeh, S. Wang, K. Zeng, M. Eisapour and Z. Wang, IEEE International Conference on Image Processing, Sept. 2016.