On the Calculation and Usage of HDR Static Content Metadata

By Michael D. Smith and Michael Zink

Abstract

HDR10 is a common platform-agnostic video format that is widely used for ultrahigh-definition (UHD-1)/4K content distribution. HDR10 includes 10-bit video data formatted for the perceptual quantization (PQ) electro-optical transfer function (EOTF) using color encoding primaries corresponding to International Telecommunication Union - Radiocommunication (ITU-R) BT.2020 (and BT.2100) and includes optional "static metadata." HDR10 static metadata typically comprises two different types of metadata, mastering display metadata and content metadata. This paper discusses the meaningful calculation and usage of static content metadata for the display and processing of HDR10 content. Outlier values in the high-dynamic-range (HDR) pixel values can exist in content being mastered today and a simple statistical approach is presented to reduce the outlier's influence on the static content metadata values. Additionally, this paper describes how a display or process can utilize good and reliable maximum content light level (MaxCLL) values to improve the reproduction of creative intent.

Keywords

High dynamic range (HDR), maximum content light level (MaxCLL), maximum frame average light level (MaxFALL), metadata, outlier rejection, outliers, static metadata

The industry was introduced to the content-related high dynamic range (HDR) static metadata items of MaxCLL and MaxFALL via the CTA 861.3 video interface standard, which was initiated in response to a request from the Blu-ray Disc Association (BDA). Modern consumer display interfaces like high-definition multimedia interface (HDMI) and DisplayPort use the CTA 861 video interface standard. The MaxCLL calculation pseudocode appearing in CTA 861.3 Annex A will find the very brightest pixel in the entire content sequence. While this can be an important metric to assist a display showing an optimal HDR image, it can also be irrelevant if the very brightest pixel is a statistical outlier.

Introduction

HDR10 static metadata typically comprises two different types of metadata—mastering display metadata and content metadata. The first corresponds to SMPTE ST 2086 Mastering Display Color Volume metadata,[1] which describes the color primaries and minimum and maximum luminance of the mastering display. Maximum content light level (MaxCLL) and maximum frame average light level (MaxFALL) are the static content metadata values distributed with HDR10 content.

The industry was introduced to the content-related high-dynamic-range (HDR) static metadata items of MaxCLL and MaxFALL via the CTA 861.3[2] video interface standard, which was initiated in response to a request from the Blu-ray Disc Association (BDA). Modern consumer display interfaces like high-definition multimedia interface (HDMI) and DisplayPort use the CTA 861 video interface standard. The MaxCLL calculation pseudocode appearing in CTA 861.3 Annex A will find the very brightest pixel in the entire content sequence. While this can be an important metric to assist a display showing an optimal HDR image, it can also be irrelevant if the very brightest pixel is a statistical outlier.

HDR content mastering practices are evolving rapidly but current practices in use today often generate outlier values unintentionally. The presence of outliers can increase the MaxCLL value for such content when using the algorithm in CTA 861.3 Annex A to calculate

Digital Object Identifier 10.5594/JMI.2021.3090176
Date of publication: 6 August 2021

Authorized licensed use limited to: IEEE Xplore. Downloaded on November 09,2024 at 03:27:01 UTC from IEEE Xplore. Restrictions apply.

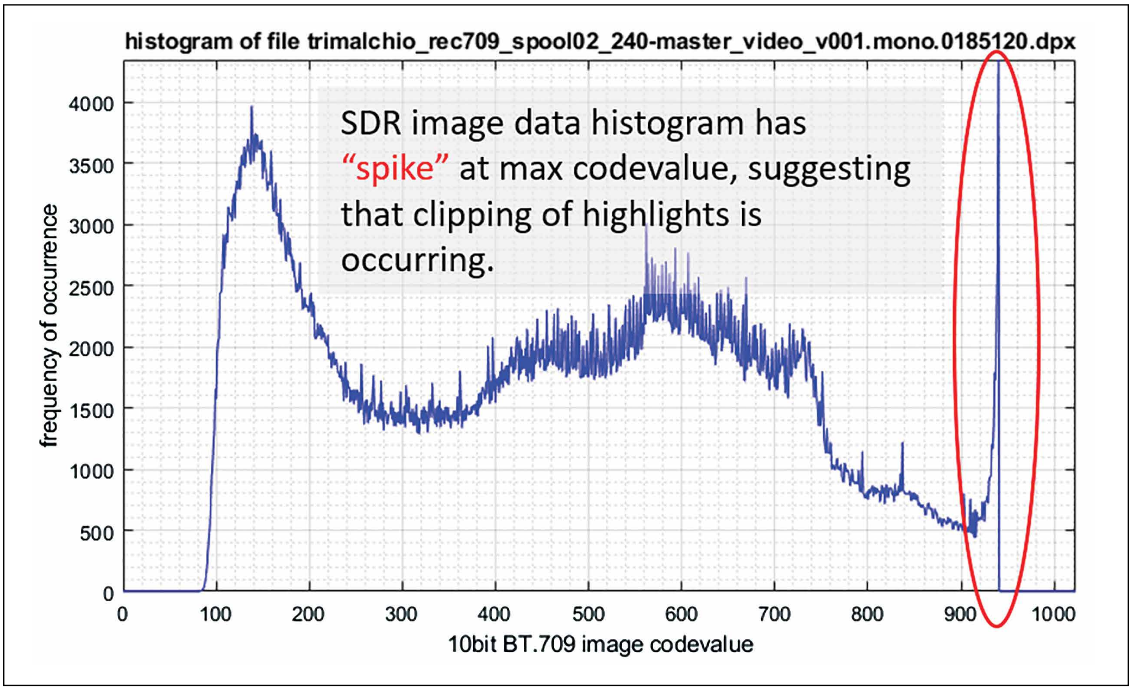


FIGURE 1. SDR histogram example.

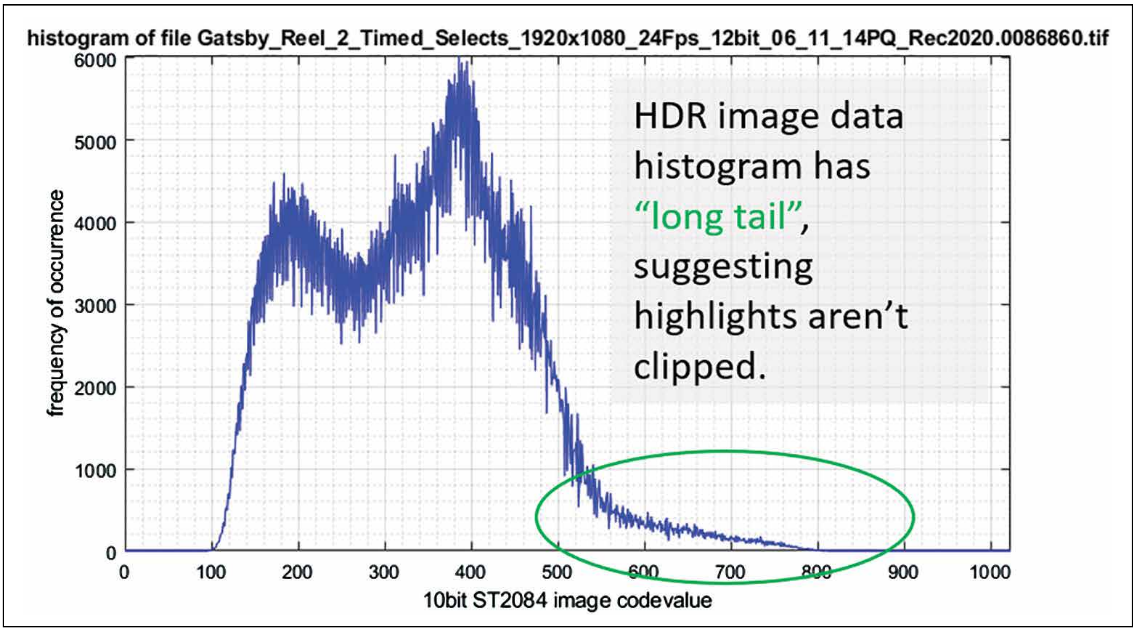


FIGURE 2. HDR histogram example.

MaxCLL. The MaxCLL value would be lower if HDR mastering practices did not generate unintentional outlier values. This paper describes an alternative process that uses a simple statistical analysis to generate a MaxCLL value that is more representative of the statistically significant brightest pixels contained in the content sequence, making it more meaningful for display processing. This simple technique has been used to prepare the static content metadata values for hundreds of ultra-high-definition (UHD-1)/4K HDR titles distributed by Warner Bros. over the past several years.

HDR Content Analysis

A simple histogram analysis shows the distribution of pixel values within an image and can be used to explain some fundamental differences between standard dynamic range (SDR) and HDR imagery. Histograms of the same frame in SDR and HDR formats are shown in **Figs. 1** and **2** and their corresponding images are shown in **Figs. 3** and **4**, respectively. The SDR histogram in **Fig. 1** shows a "spike" at the maximum code value, suggesting that clipping of the highlights is occurring, which is quite common in SDR imagery.

Authorized licensed use limited to: IEEE Xplore. Downloaded on November 09,2024 at 03:27:01 UTC from IEEE Xplore. Restrictions apply.

FIGURE 3. SDR frame example.

FIGURE 4. HDR frame example (tone mapped with output luminance reduced by 2.5 stops for display as SDR in this article).

The "long tail" in the HDR histogram in **Fig. 2** suggests that the highlights from the same scene are not clipped in the HDR version.

Outlier Examples

Outliers in Single Frames

Examining the histogram statistics of a single frame can reveal outliers that occur beyond the natural "long tail" representing unclipped HDR highlights, as shown in **Fig. 5**. Information from the histogram analysis of this example frame that contains the brightest pixel in the whole film is used to generate corresponding percentile statistics for this frame, which illustrate what percentage of the pixels in the frame are below a specified light level. A heat map corresponding to pixels in the 90%–95%, 95%–99%, 99%–99.9%, 99.9%–99.99%, 99.99%–99.999%, and 100% percentile ranges is overlaid on the image in **Fig. 6.**

This analysis illustrates that the brightest pixel of the entire film corresponds to a 7707 nits single pixel in the reflection of a key light in an actor's eye within a relatively dim scene. In this case, it is known that the mastering monitor used was limited to a peak luminance of 4000 nits and used dual-modulation full-array (local-dimming) liquid crystal display (LCD) display technology with a significant number of dimming zones but was limited to HD resolution. Such a screen cannot even display a single 7707 nits UHD pixel accurately; so in this case, it is reasonable to expect that the single brightest pixel in the feature was very likely not intended to be shown at 7707 nits even though it existed in the so-called *"creatively approved" master.*

Temporal Outliers

Outliers can also exist in the temporal dimension when the statistics of one short scene are significantly different than that of other scenes within the same feature or episode. **Figure 7** shows an example that plots the

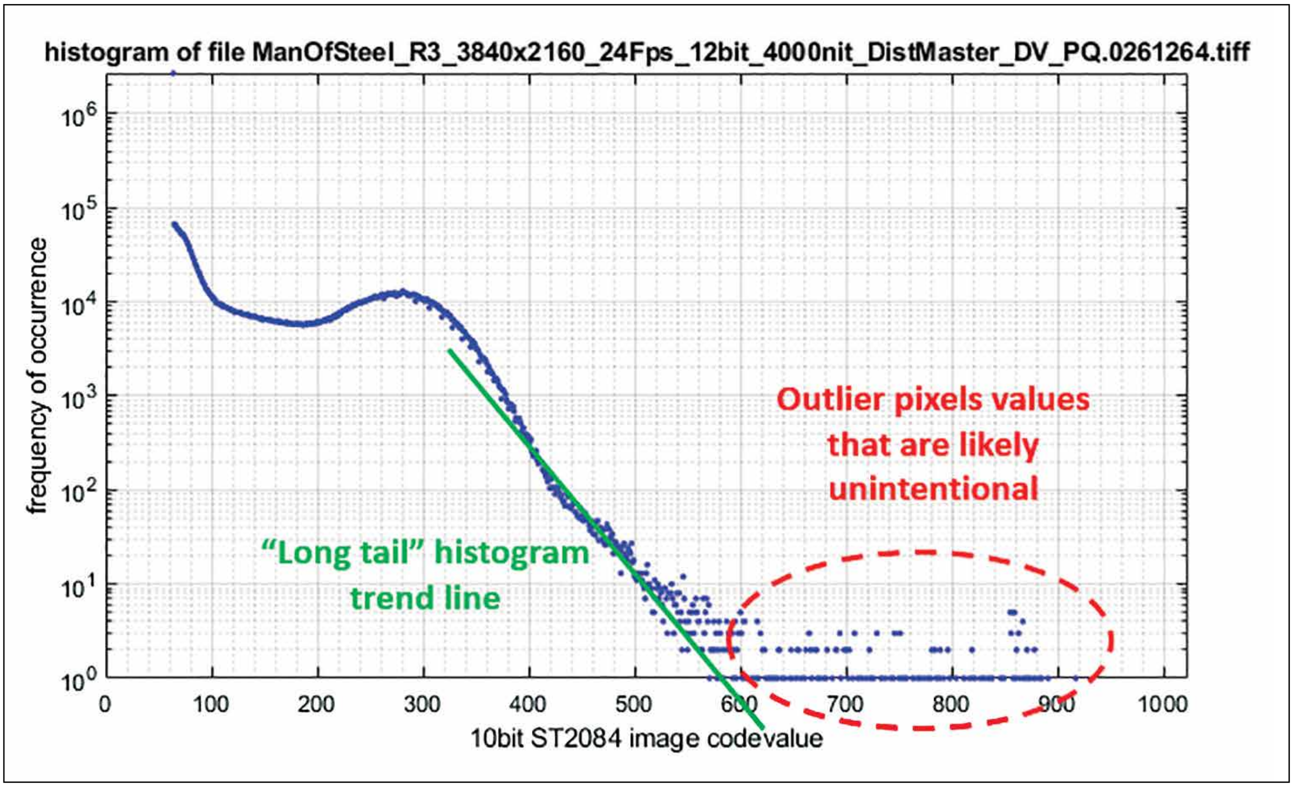


FIGURE 5. Example histogram analysis of HDR frame with outliers.

Authorized licensed use limited to: IEEE Xplore. Downloaded on November 09,2024 at 03:27:01 UTC from IEEE Xplore. Restrictions apply.

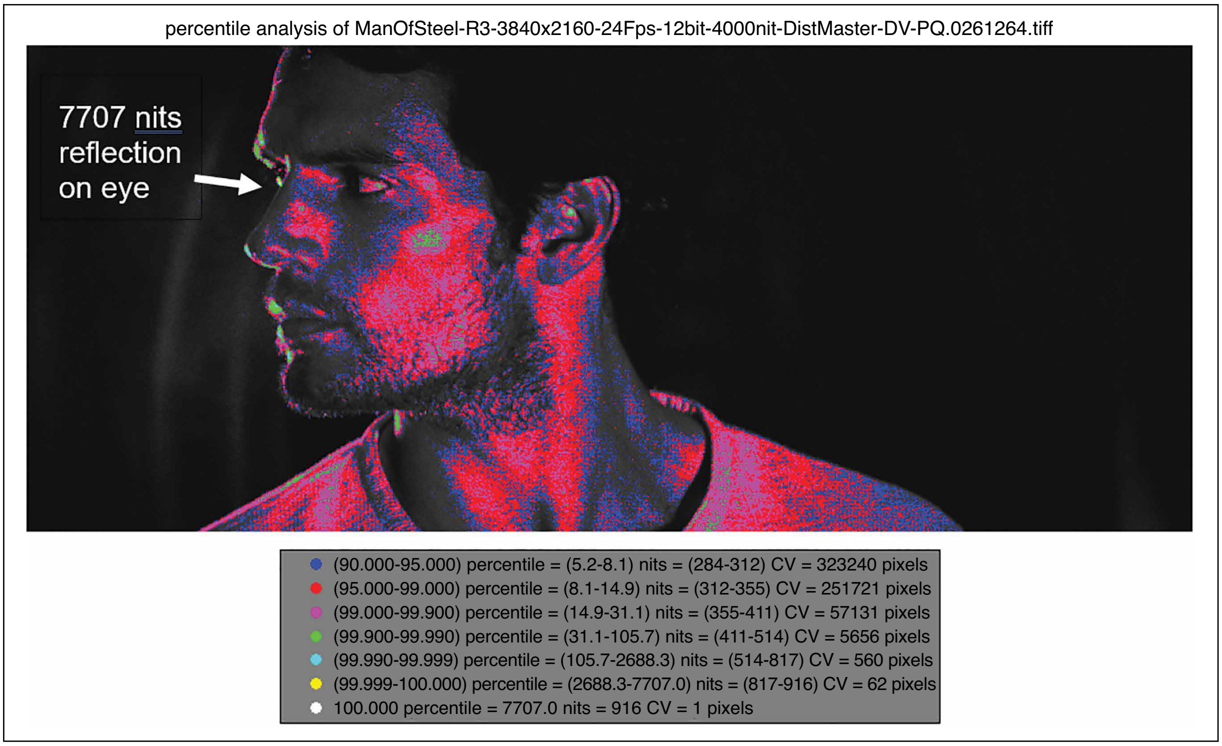


FIGURE 6. Example frame containing the single brightest pixel of the film overlaid with a percentile heat map.

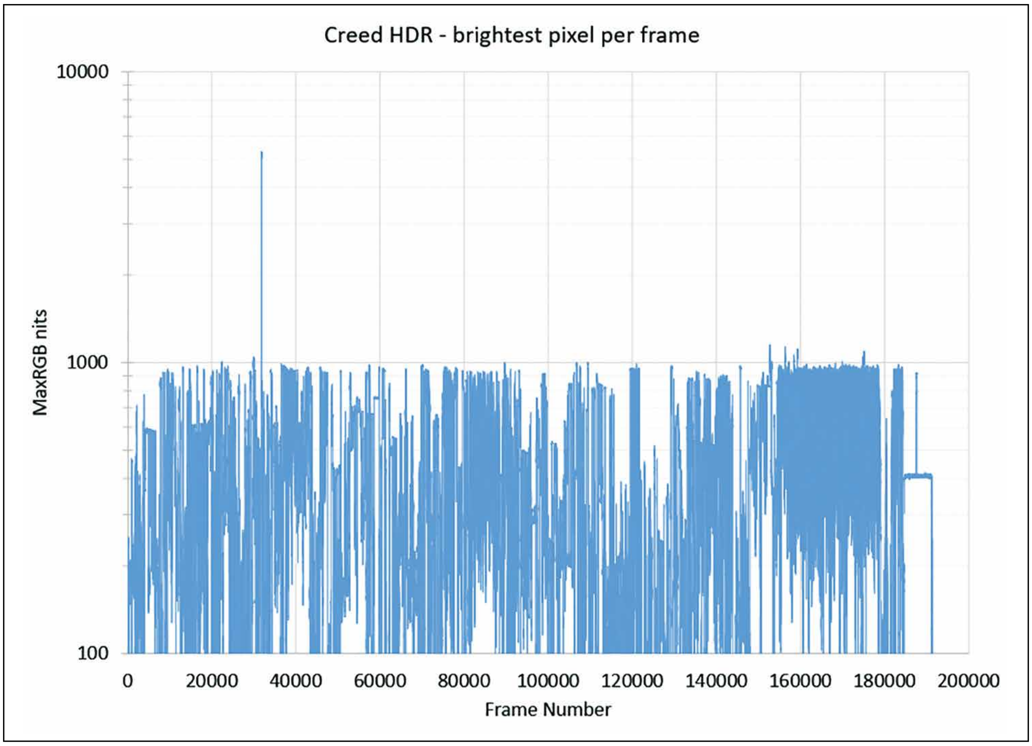


FIGURE 7. Example of the brightest pixel from each frame of an entire film.

brightest pixel in each frame across all the frames of the film. In this case, the brightest pixel in each frame appears to vary from frame to frame in a somewhat consistent manner between 100 and 1000 nits, but a single outlier scene stands out. The outlier scene corresponds to a short establishing shot in front of an apartment building at night, the streetlight in the scene contains pixels with a light level over 5000 nits, which was significantly different than other scenes in the film that had maximum values ranging between 100 and 1000 nits.

Authorized licensed use limited to: IEEE Xplore. Downloaded on November 09,2024 at 03:27:01 UTC from IEEE Xplore. Restrictions apply.

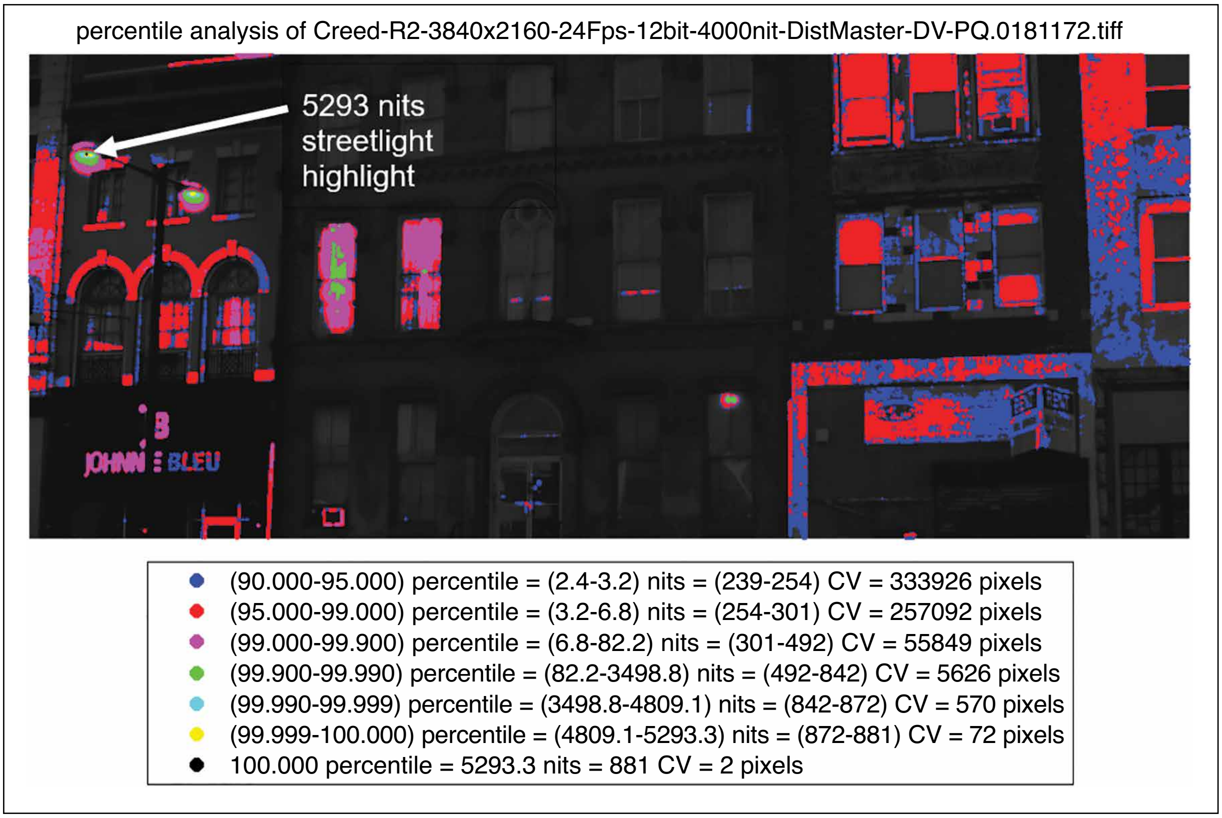


FIGURE 8. Another example frame containing the single brightest pixel of the film overlaid with a percentile heat map.

Figure 8 shows a percentile heat map overlaid on the frame containing the brightest pixel in the film.

Figure 9 shows the same data plotted with only the frames in the shots adjacent to the frame that contains the single brightest pixel in the film. In this case again, it is known that the mastering monitor was limited to 4000 nits and thus a 5293 nits pixel value could not have been shown accurately during the creative approval process

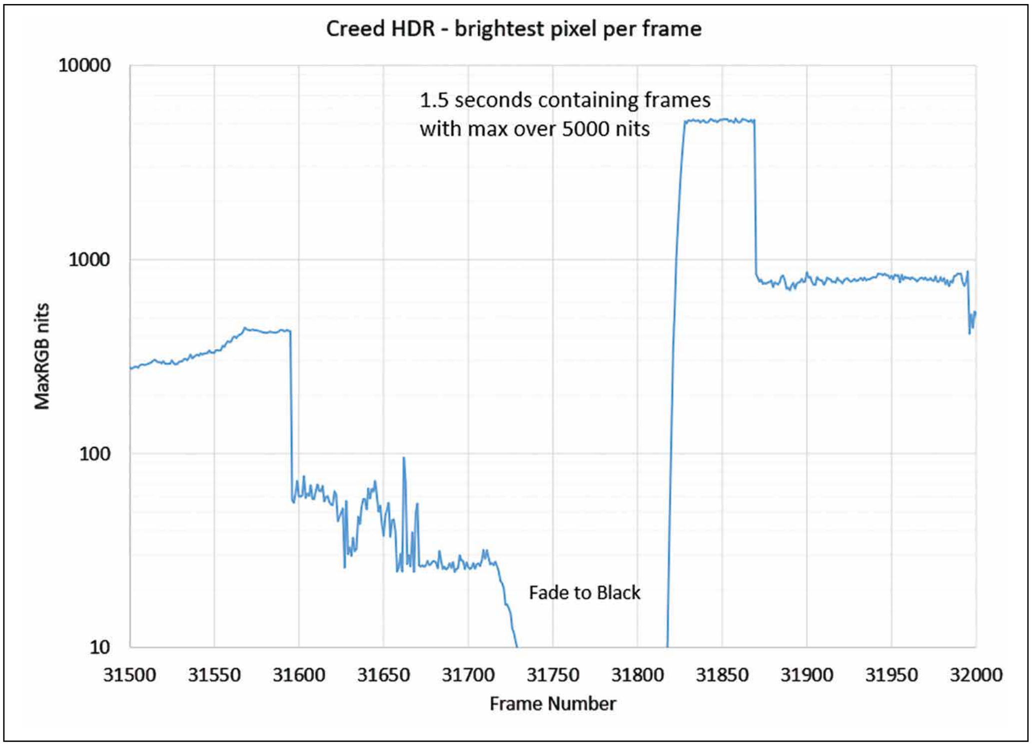


FIGURE 9. Section of film with temporal outlier.

Authorized licensed use limited to: IEEE Xplore. Downloaded on November 09,2024 at 03:27:01 UTC from IEEE Xplore. Restrictions apply.

for the master. It serves as a great reminder of the importance to ensure that sufficient time is spent during the mastering process to also review the content statistics and for the color grading team to leverage waveform analysis monitors and other tools showing a percentile heat map overlay similar to what is shown in **Figs. 5** and **8**.

Outliers From Resize Filter Overshoot

Outliers can be introduced in HDR content from resize filter overshoots. A basic tenet of color grading is that it should be a resolution-independent operation, which means that the color grading is typically performed at the available source resolution of the content without resizing the images received from production. After color grading is completed, cropping and resizing operations are performed to convert the finished content into the release aspect ratio and distribution container format.

Filter overshoots are common artifacts of resize filters that can generate outliers due to the highly nonlinear response of the perceptual quantization (PQ) electro-optical transfer function (EOTF). The step responses of several commonly used resize filters are shown in **Fig. 10**, illustrating the overshoot values of +2% to +10% for most of the filters used in this example, sampled from a 1:2 upsampling resize using The Foundry NUKE compositing software. Different filters and different resizing ratios have different overshoot behavior. Common filters that do not generate overshoots (like linear or cubic) do not generally have as desirable frequency response characteristics as common filters that do generate overshoots (like Lanzcos).

Table 1 provides numerical examples of the impact of overshoots when filtering in the PQ domain versus filtering in the display linear domain for different peak input values of 500, 1000, 2000, and 4000 nits. At 1000 nits, a 10% filter overshoot due to a resize in the PQ domain translates to a 1992 nits pixel value, while the same percentage overshoot due to resize in display linear translates to 1100 nits. Based on this, it is easy to understand that a simple way to avoid introducing significant outliers due to resize operations is to perform the resize operations in the display linear domain rather than the PQ nonlinear domain or to simply not use resize filters that generate overshoots. A well-known benefit of performing the resizing operation in display linear is that the average luminance of an area of an image will not be modified by the resize operation. Resizing in other domains like scene-linear or scene-non-linear (log) have other properties but will not be discussed further in this paper.

Outliers From Lossy Compression Distortion

Outliers can also be introduced by lossy mezzanine compression of distribution service masters (e.g., ProRes or JPEG 2000), even when using high data rates that result in very high visual quality levels. The lossy compression distortion can introduce small changes in PQ nonlinear

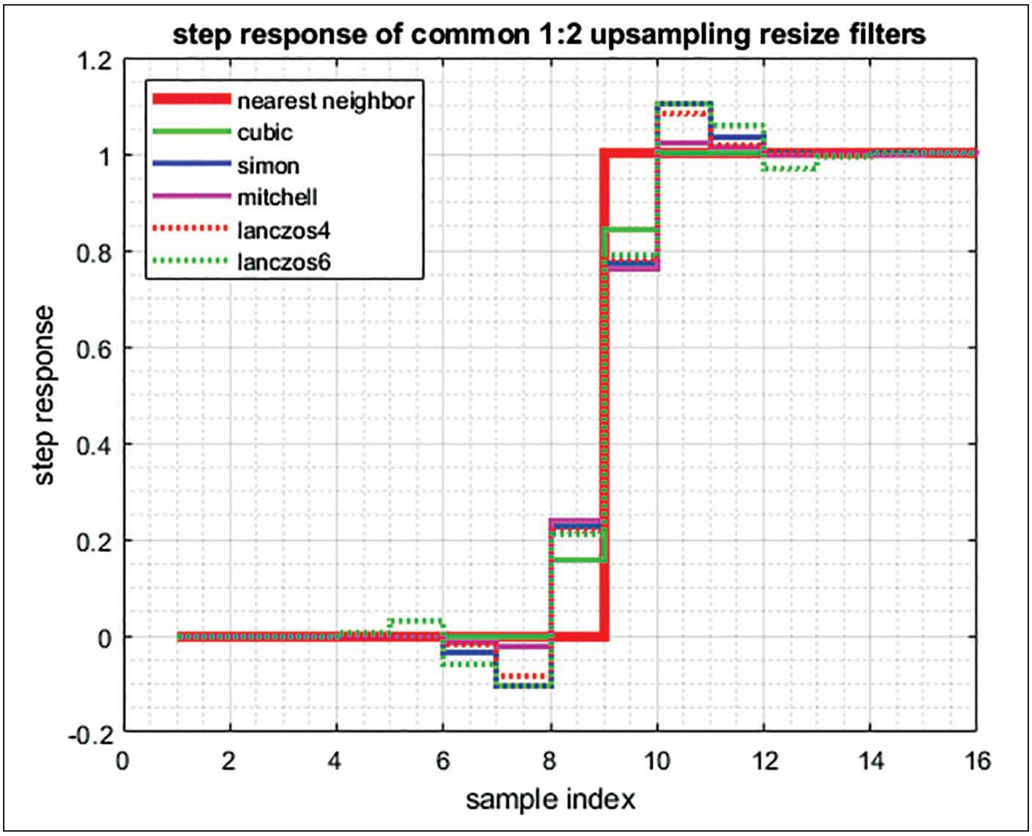


FIGURE 10. Step response of resize filters.

Authorized licensed use limited to IEEE Xplore. Downloaded on November 09,2024 at 03:27:01 UTC from IEEE Xplore. Restrictions apply.

TABLE 1. Impact of overshoot due to resize filters.

	Overshoot percentage due to resize filters			
	0%	2%	5%	10%
Filtering in display linear domain (nits)	500	510	525	550
	1000	1020	1050	1100
	2000	2040	2100	2200
	4000	4080	4200	4400
Filtering in PQ domain (nonlinear value)	0.6766	0.6901	0.7104	0.7442
	0.7518	0.7669	0.7894	0.8270
	0.8274	0.8440	0.8688	0.9102
	0.9026	0.9206	0.9477	0.9928
Filtering in PQ domain (nits)	500	567	683	933
	1000	1148	1411	1992
	2000	2328	2926	4292
	4000	4731	6095	9339

values. As described in the preceding section about filter overshoot, small changes in bright PQ-encoded pixel values can lead to large changes in corresponding light levels, which can lead to unintended outliers in the decoded data. This issue can be avoided if content metadata values are computed on uncompressed imagery or with files that have been subjected to mathematically lossless compression rather than lossy compression.

Alternative Calculation of MaxCLL and MaxFALL Values

When MaxCLL and MaxFALL static content metadata were invented by the BDA as part of the UHD Blu-ray format development process in 2014, the only HDR content that existed at the time were short test clips demonstrating the new format's potential capabilities. These test clips did not have outliers. The industry only became aware of the potential existence of outliers in HDR content after real production and distribution started on a broader scale. Mastering practices vary by facility and by title, and the creative approval process often cannot be revisited easily due to scheduling limitations of the talent in addition to the schedule pressure that often exists related to meeting preannounced title release dates. These challenges make fixing outliers in the master difficult if they are discovered after the creative approval process completes.

We developed a simple method to determine more meaningful MaxCLL and MaxFALL values that are not influenced by outliers. The process is summarized as follows.

MaxCLL:

- Calculate histogram for each frame and use the histogram to compute the 99.99% percentile.
- Calculate the 99.5% percentile of the 99.99% frame percentiles for the whole sequence.

MaxFALL:

- Calculate the frame average light level of each frame in the sequence.
- Calculate the 99.75% percentile of the frame average light levels for the whole sequence.

A more detailed pseudocode description is provided in **Tables 2** and **3**. Using this alternative approach leads to MaxCLL and MaxFALL values that are not significantly affected by spatial or temporal outliers. While we have studied percentile statistics for hundreds of titles and feel comfortable suggesting the percentiles described above, content owners are also encouraged to do their own analysis and use the suggested method with suitable percentile values with which they feel comfortable. Additionally, other more advanced techniques not purely based on fixed percentile values could also be used to reject outliers.

Example MaxCLL and MaxFALL Values From Warner Bros. Releases

Figures 11 and **12** show the calculated MaxCLL and MaxFALL values, respectively, for 231 Warner Bros. titles using both the original method described in CTA 861.3 that includes outliers if they exist and the alternative method suggested in this paper that rejects outliers. Note that in **Figs. 11** and **12**, the same title is not plotted with the same title number in both figures; the plotted data was sorted by MaxCLL and MaxFALL without outliers to make it easier to see the differences between outlier and nonoutlier calculations.

Note that some titles do not have significant outliers, which causes the two plotted points for a title to be close together.

Authorized licensed use limited to: IEEE Xplore. Downloaded on November 09,2024 at 03:27:01 UTC from IEEE Xplore. Restrictions apply.

TABLE 2. Alternative MaxCLL calculation with outlier rejection.

Alternative MaxCLL calculation pseudocode:

```
CalculateMaxCLLwithOutlierRejection()
{
    for each (frame in the sequence)
    {
        for each (pixel in the active image area of the frame)
        {
            convert the pixel's non-linear (R', G', B') values to linear values (R,G,B) calibrated to cd/m^2
            compute maxRGB = max(R, G, B)
            store maxRGB in temporary buffer of the frame's maxRGB values
        }
        compute the 99.99% percentile of the frame's maxRGB values
        store the 99.99% percentile value in a temporary buffer of the sequence's frames 99.99% percentile values.
    }
    compute the 99.5% percentile of the set of the sequence's frames  99.99% percentile values.
    set MaxCLL = 99.5% percentile of the sequence's frames 99.99% percentile values.
    return MaxCLL
}
```

TABLE 3. Alternative MaxFALL calculation with outlier rejection.

Alternative MaxFALL calculation pseudocode:

```
CalculateMaxFALLwithOutlierRejection()
{
    for each (frame in the sequence)
    {
        set runningSum = 0
        for each (pixel in the active image area of the frame)
        {
            convert the pixel's non-linear (R', G', B') values to linear values (R,G,B) calibrated to cd/m^2
            set maxRGB = max(R,G,B)
            set runningSum = runningSum + maxRGB
        }
        set frameAverageLightLevel = runningSum / numberOfPixelsInActiveImageArea
        store frameAverageLightLevel in temporary buffer of the sequence's frames frameAverageLightLevel values
    }
    compute the 99.75% percentile of the sequence's frames frameAverageLightLevel values.
    set MaxFALL = 99.75% percentile of the sequence's frames frameAverageLightLevel values.
    return MaxFALL
}
```

Using MaxCLL to Improve Reproduction of Creative Intent

Ideally, viewers would perceive the same appearance when watching the content that the content creator experienced on the mastering display when the content was finished and approved for distribution. The simplest way to achieve this goal is to replicate the viewing conditions that occurred in the mastering environment by using a display that matches or exceeds the capabilities of the mastering display. This was the motivation for including ST 2086 Mastering Display Color Volume metadata in HDR10.

If the viewer display has capabilities lesser than the mastering display, achieving the goal of appearance matching is more challenging. Gamut mapping is a technique that has a long history in print reproduction and has become more important to the video industry as wide gamut displays have been paired with wide gamut encodings for distribution. Gamut mapping can be optimized more efficiently if the gamut boundaries of the source material are well known. An early discussion of this topic appears in SMPTE EG 432-1 Chapter 8 "Gamut Mapping and the Value of Mastering Projector Metadata."[3] Gamut mapping can be used to compensate for the viewer's display if it has a smaller gamut

Authorized licensed use limited to: IEEE Xplore. Downloaded on November 09,2024 at 03:27:01 UTC from IEEE Xplore. Restrictions apply.

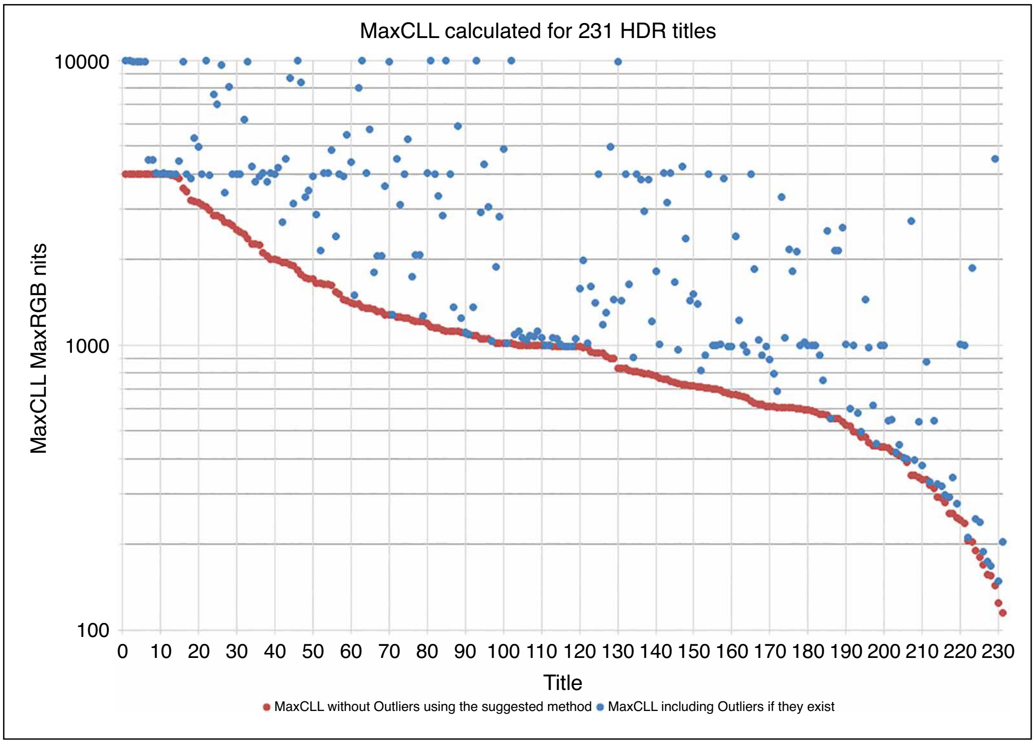


FIGURE 11. Calculated MaxCLL values for 231 titles.

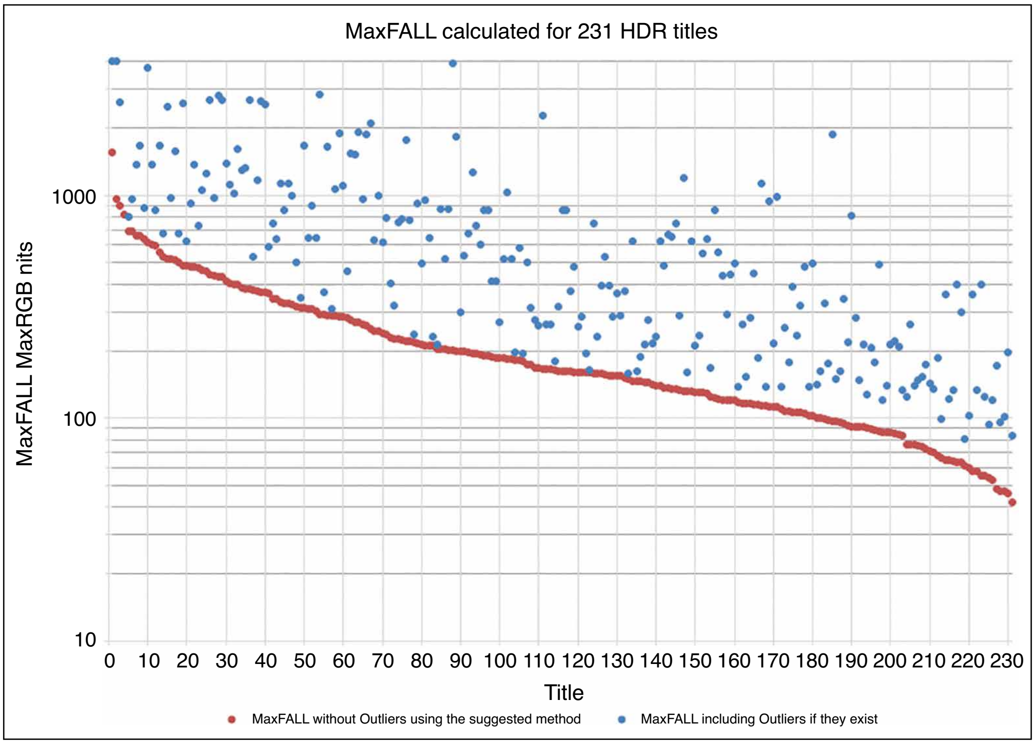


FIGURE 12. Calculated MaxFALL values for 231 titles.

than the mastering display. Tone mapping[4] can be used to compensate for the viewing display that has less luminance range than the mastering display.

Tone mapping typically remaps the luminance range of the mastering display to that of the viewing display. One example of nonlinear tone mapping is the EETF described in International Telecommunication Union – Radiocommunication (ITU-R) Report BT.2390-7.[5] If the MaxCLL value that is delivered with the HDR10 content is reliable and was not influenced by outliers in the content, the MaxCLL value should be used as the upper bound of the input luminance range (Lw in BT.2390-7 EETF)

Authorized licensed use limited to: IEEE Xplore. Downloaded on November 09,2024 at 03:27:01 UTC from IEEE Xplore. Restrictions apply.

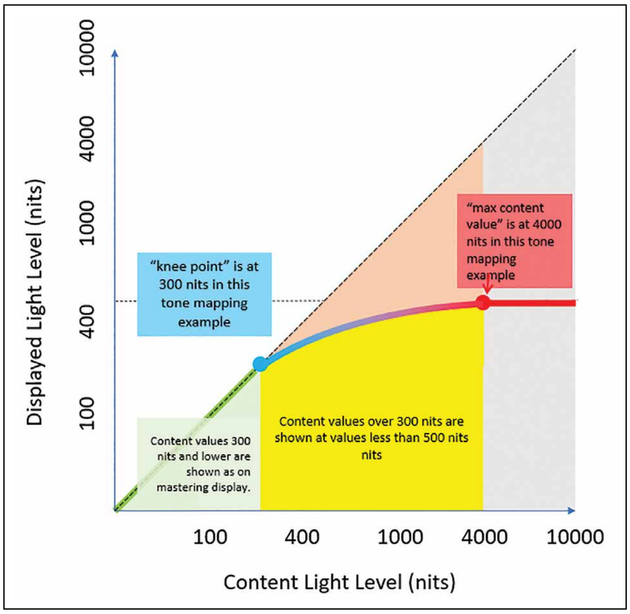


FIGURE 13. Five hundred nits tone mapping with input range defined by Mastering Display Maximum Luminance.

in the remapping process instead of the maximum luminance of the mastering display. Using the MaxCLL in this way helps improve the appearance because typically the MaxCLL value is lower than the maximum luminance of the mastering display, which leads to less extreme tone mapping. If the MaxCLL is less than the viewing display's maximum luminance, then no tone mapping is necessary.

The example shown in **Figs. 13** and **14** illustrates the concept of how using MaxCLL helps improve the reproduction of creative intent on a 500 nits display. If the content that was mastered on a 4000 nits display only has pixel values that go to 800 nits, the tone

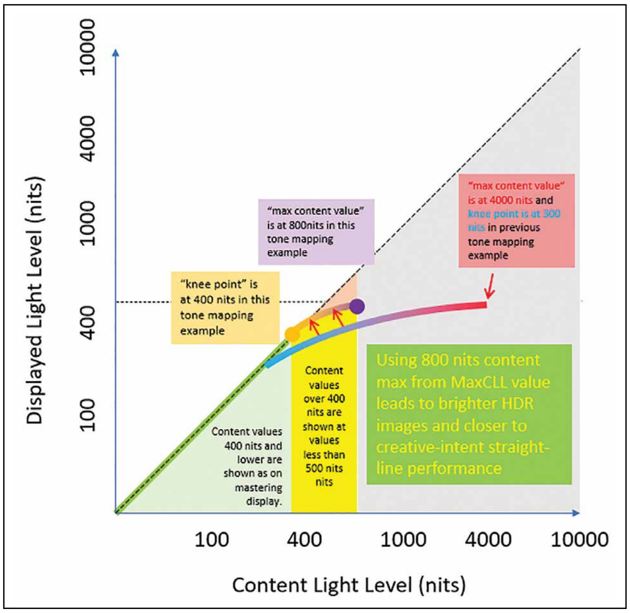


FIGURE 14. Five hundred nits tone mapping with input range defined by MaxCLL.

mapping curve shown in **Fig. 13** that remaps 4000 nits luminance range to a 500 nits display is reserving some output range for an input range 800–4000 that will not occur because the content does not go higher than 800 nits. If, instead, the MaxCLL of 800 nits is used to design the tone mapping curve, as shown in **Fig. 14**, the "knee point" of the curve will be shifted up by about 100 nits because the tone mapping is less extreme. This makes the HDR content appear brighter and the displayed luminance will more closely match the luminance that was displayed during creative approval on the mastering display, while also utilizing the full luminance range of the viewer's 500 nits display.

Dynamic metadata can be used in a similar manner to improve the reproduction of creative intent, but it typically operates on a scene-by-scene basis instead of on a title-by-title basis. Some advanced consumer displays can also perform a realtime analysis of the input video statistics to optimize the reproduction without receiving predetermined scene-based dynamic metadata.

Conclusion

HDR10's MaxCLL and MaxFALL static content metadata carriage mechanisms have been widely deployed in various distribution standards, devices, and interfaces. It is up to content distributors to ensure that the MaxCLL and MaxFALL values are relevant to their content in the same way they ensure the image looks correct before it is distributed. Since MaxCLL and MaxFall are optional, they do not have to be used; in fact, a special value (0) is reserved for "unknown." If the values are used, it is obvious that the values should not represent outliers in the content and instead represent the intended upper bound of the creatively approved content. This paper described a simple technique that can be used to help content distributors ensure that static content metadata values distributed with HDR10 content are robust to unintended outliers that can occur for multiple reasons. This paper has also described how the presence of good statistical metadata values can help improve the reproduction of creative intent.

References

1. SMPTE, ST 2086:2018, "Mastering Display Color Volume Metadata Supporting High Luminance and Wide Color Gamut Images."
2. Consumer Technology Association (CTA)-861.3-A, "HDR Static Metadata Extensions." Accessed: Mar. 1, 2021. [Online.] Available: https://shop.cta.tech/products/hdr-static-metadata-extensions
3. SMPTE, EG 432-1:2010, "Digital Source Processing—Color Processing for D-Cinema."
4. A. Burke, M. D. Smith, and M. Zink, "Color Volume and Hue-Preservation in HDR Tone Mapping," *SMPTE Mot. Imag. J.*, 129(4):45–50, May 2020.
5. International Telecommunication Union – Radiocommunication (ITU-R), Report BT.2390-7, "High Dynamic Range Television for Production and International Programme Exchange," Jul. 2019.

Authorized licensed use limited to: IEEE Xplore. Downloaded on November 09,2024 at 03:27:01 UTC from IEEE Xplore. Restrictions apply.

About the Authors

Michael D. Smith has been a digital imaging and intellectual property consultant since 2003. He is currently a co-editor of the JPEG 2000 High-Throughput image compression standard. In 2018, he received a screen credit for his work on *Mary Poppins Returns*, which was related to the color science involved in integrating a traditional 2D animation workflow with a modern ACES production pipeline. From 2013 to 2015, he was the co-chair of Blu-ray Disc Association's UHD-TF Video Subgroup, which defined the video-related requirements for the Ultra HD Blu-ray disc format. His work on more than 35 intellectual property matters related to infringement and validity of patents has resulted in payments of approximately $1.7 billion. He was the editor of the book "*3D Cinema and Television Technology: The First 100 Years*" published by SMPTE in 2011. He received BS and MS degrees in electrical engineering from the University of California Los Angeles (UCLA), Los Angeles, CA, in 2001 and 2004, respectively.

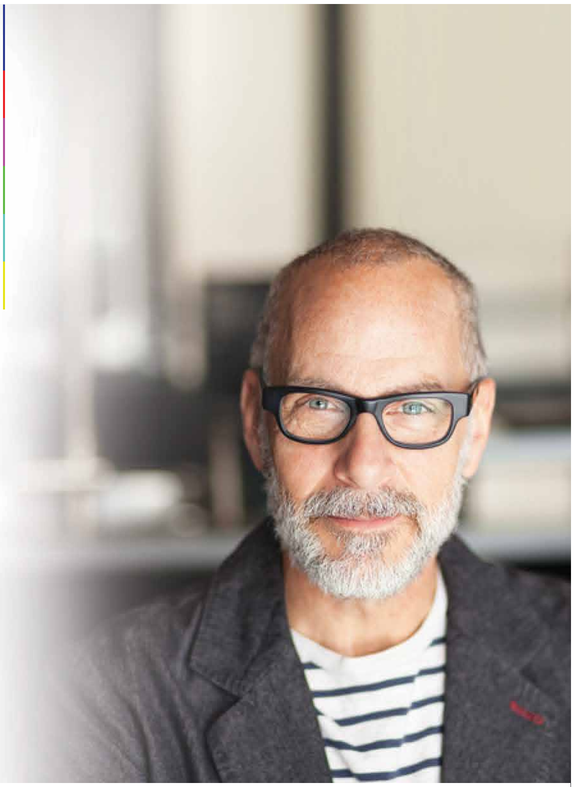
Michael Zink is the vice president of technology at Warner Bros. (WB), Burbank, CA, where he is responsible for exploring emerging technologies to enhance WB's capabilities for production, post-production, and distribution. This includes assessing new technologies and assisting with the setup and integration of digital workflows. He also participates in a number of standards associations such as the Consumer Technology Association (CTA), Digital Cinema Initiatives (DCI), and SMPTE among others and also serves as the chairman for the Ultra High Definition Alliance (UHDA). Prior to joining WB, in 2014, he worked at Technicolor in London, U.K., and Los Angeles, CA, for more than ten years, most recently as the vice president of technology strategy, where he was responsible for launching the production efforts around various new optical disc formats. Additionally, he was responsible for the promotion and adoption of Technicolor technology solutions within industry groups. Earlier in his career, he worked for several media production facilities in Germany.

A contribution received for the SMPTE May 2021. Copyright © 2021 by SMPTE.

Gain Exposure. Make Your Mark.

Submit your technical paper to the SMPTE Motion Imaging Journal

The Motion Imaging Journal publishes new and interesting perspectives on the technology driving the evolution of media and entertainment. With our in-depth, monthly exploration of the latest engineering and scientific developments, we help SMPTE members stay current in their specialty and aware of related technologies. Even better, we provide the opportunity to share their expertise and engage with the brightest minds in the industry. The Journal accepts both thought-leadership and technical papers for peer review and possible publication. We encourage leading academics, researchers and engineers worldwide to collaborate and share their expertise, while they help shape the Journal and the industry itself.



SMPTE

See complete submission guidelines at
smpte.org/publications/submit_paper

Authorized licensed use limited to: IEEE Xplore. Downloaded on November 09,2024 at 03:27:01 UTC from IEEE Xplore. Restrictions apply.