Distribution of HDR in an SDR World

Keith Slavin
isovideo LLC
keith@isovideo.com

Written for presentation at the
SMPTE 2016 Annual Technical Conference & Exhibition

Abstract.

HDR media is now possible due to improved camera dynamic range and logarithmic transfer functions. An inverse EOTF at the TV gives a linear relation between camera photons and display luminance. Local contrast is thereby preserved. Unfortunately linearity requires more display dynamic range, which increases cost/power (impractical for mobile devices). HDR displays ultimately limit HDR systems, and careful camera setup or tone-mapping is required to avoid display saturation. Sports events need HDR cameras and real-time tone-mapping for uncontrolled scenes. We propose a system for automatically tone-mapping unsaturated HDR down to vivid Rec709, allowing HDR production of "future-proof" end-products. Instead of a linear relationship between camera and display, the system takes full advantage of the eye's logarithmic response to preserve local contrast while limiting dynamic range. HDR production and SDR distribution makes practical and economic sense, and HDR production can go main-stream, while distributed SDR, without supplementary metadata, can now be inverse tone-mapped to vivid HDR.

Keywords. HDR, HDR Distribution, HDR to SDR Tone Mapping, SDR to HDR Conversion, Noise reduction, Frame-rate conversion, Remastering, Up-scaling, GPU, IP-based, UHD, 4K, HD, AVC/HEVC Compression, Online distribution, Workflow, Low latency, Scalability, Flexibility, Automation, Metadata.

1. Introduction

Cameras, post production processing, reference displays, and consumer displays have been advancing steadily [1]. It is now possible to shoot an uncontrolled scene and capture both highlights and shadow, and represent the result in ST-2084 (aka Dolby PQ[2] EOTF) which takes advantage of the eye's logarithmic perceptual response at brightness levels up to 10,000 nits, with 10 or more bits per encoded component for reduced contouring. When combined with a wider color gamut encoding (e.g. BT2020), higher frame rates, effective noise reduction, high quality processing and compression, the results can be stunning. However, if even one element is below par, then the human vision system can be very good at sensing “something isn't right”.

With a variety of modified logarithmic output representations, HDR cameras can process and represent a much wider dynamic range within each image. HDR setup is actually much simpler than for SDR, in that fewer stop adjustments are needed during filming. All aspects of the scene can be captured, without having to make tricky over/under exposure decisions during shooting, and quite apart from the future-proof benefits of filming in HDR.

2. HDR in an SDR World

The penetration of HDR content and HDR TVs has been increasing steadily, but as of 2016, we are still very much in an SDR (Rec709) world:

- The vast majority of newly generated content is still SDR,
- The vast majority of display devices are SDR only,
- Nearly all of broadcast infrastructure targets SDR.

If we could all somehow switch over to an HDR ecosystem overnight, life could be simple again. Unfortunately, we can't, and it won't. Differences in what pixel values represent is in many ways much harder to keep track of than frame rates, widths and heights, where erroneous metadata is much more obvious. Metadata may not be preserved well in a world where Rec709 is assumed, and often not even signaled. How do we smoothly transition to an HDR world from today's SDR world? This paper touches three main aspects:

1. Bridging HDR to the SDR world from a post production perspective,
2. Factors and attributes that affect visual immersiveness of HDR, and
3. Options for HDR/SDR distribution, and issues that arise.

The paper concludes with a brief discussion of a new server cluster that has the capability to meet some of the challenges, along with a few examples showing its HDR/SDR conversion capability.

3. Bridging HDR to SDR – a Post-Production Perspective

3.1 Unifying HDR/SDR Creative Intent

Currently all production targets SDR. There simply isn't the audience numbers to justify HDR-only production and distribution. Today, an “HDR production” implies the creation of an HDR and SDR version, each with their own creative intent. With the tools available up to now, making sure both “intents” match is not easy – however we may expect that the mid-tones from say 0.2 to 40 nits (which can work well in both 10-bit SDR and HDR) should correspond closely.

Up to now, automated conversion from HDR intent to SDR has been compromised due to the limited tools available for high quality HDR to SDR conversion. Using HDR sources to fabricate a high quality SDR version with its own creative intent can be very costly and time consuming.

Without doubt, the increased capability of HDR to deal with both darker and brighter scenes allows for an equal or improved expression of creative intent at all light levels. The differences in intent expressed in the SDR version is mainly to deal with SDR limitations, rather than imbue what was *really* wanted. It therefore makes both artistic and economic sense to express creative intent in HDR, and then use an automated tone-mapping system to generate an SDR version with the creative intent essentially preserved. This can be achieved even though absolute luminance values are less than the HDR original, as the human vision luminance response is logarithmic. See Figures 3a/b and c/d which compare 4000 nit peak video images: tone-mapped to 100 nits versus tone mapped to 1000 and clipped to 100 nits.

Figure 3a: *Bistro* (original: BT2020 ST-2084, graded to 4000 nits) tone mapped to rec709

Figure 3b: Original tone mapped to HDR10 at 1000 nits, then clipped to rec709

Figure 3c: *Beerfest_lightshow* (original: BT2020 ST-2084, graded to 4000 nits) tone mapped to rec709

Figure 3d: Original tone mapped to HDR10 at 1000 nits, then clipped to rec709

3.2 Challenges and Benefits of Vivid HDR to SDR Tone Mapping

In an ideal HDR system, light intensity into the camera has a corresponding linear relationship to intensity in the display. Therefore, with sufficient pixel resolution and an appropriate shutter speed, local contrast in features and edges are *automatically* preserved, so the resulting HDR image should look as sharp as the original scene.

If HDR is converted to SDR directly using a simple transfer function, then middle and dark tones need to be preserved in order to preserve creative intent. However, brighter levels have to be squashed down in order to fit into SDR. Inevitably this leads to significantly reduced local contrast. An extreme squashing scenario is hard clipping, which removes local contrast altogether, results in gross loss of detail.

The peak luminance of a production may be less than the luminance that can be represented within an ST-2084 container (up to 10,000 nits), or HLG [3][4] (up to 1200 nits). For HDR that has gone through post-production, such as film or TV shows, a single maximum luminance value is mandated as a goal. For live sports, the maximum is usually determined by the cameras and the (often uncontrolled) scene, so live normalization to a target peak luminance (prior to distribution) has been most problematic up to now.

Preserving local contrast is key to preserving creative intent when tone mapping HDR to SDR, and it is an inherently complex nonlinear operation that has effects over all areas of an image. However, this can make SDR converted from HDR capture much more immersive than direct SDR capture. Optimal tone mapping to the full target output range can now be performed automatically as long as the peak input and output luminances are known. The output peak is usually set by the media requirements for tone-mapping downwards, and set by the TV for any inverse tone mapping at the TV itself. If the input peak luminance is unknown, then the container maximum peak nits can be used, which can still give very good results.

Most tone mapping software is very slow, and designed to work on single frames, so temporal continuity is not addressed. The result could be excessive flicker.

Ideally then, high quality HDR to SDR tone mapping includes:

- Preserving local contrast,
- Preserving HDR creative intent in the low/mid-tones,
- Avoiding the introduction of objectionable artifacts,
- Minimizing gross area saturation,
- Running fully automatically and efficiently to save time and money,
- Leveraging existing SDR distribution infrastructure, without requiring additional metadata,
- Allowing inversion back to HDR that looks very close to the original.

Unsaturated HDR material, when properly converted to SDR, typically looks much better than from an all-SDR system! This enhanced SDR appearance is referred to here as **SDR+**, and its improved immersiveness is mainly due to HDR capture, 10-bit or more encoding, and the ability of tone mapping to satisfy the seemingly contradictory goals of preserving local contrast while minimizing saturation.

4. HDR Video Quality and Impairments

The following system specifications [5] (note: these are *not* signal quality attributes) correspond closely to HDR10, and are capable of achieving a truly immersive viewing experience:

- Bit depth accuracy: 10 bits,
- Wide color gamut color primaries: ITU-R BT. 2020,
- Color sub-sampling: 4:2:0,
- Modified logarithmic electro-optic transfer function: SMPTE ST-2084,
- Frame rate: up to 120p (depending on the degree of motion),
- Spatial resolution: 3840x2160 pixels (for evaluation at 1.6 picture heights).

With high quality sources, the parameters of HDR10 can meet a quality level that allows the vision system to drop its guard and completely relax – an effect often characterized as "*immersiveness*". However, the above define only what a system is capable of. They do not guarantee a given production will *feel* immersive. The following factors and/or artifacts degrade accuracy, and can easily spoil the desired effect:

- Noise: shot noise is inherent in high spatial resolution cameras, and film grain from legacy SDR,
- Blur: our eyes can normally choose to focus on anything in our field of view,
- Motion judder: the camera capture frame rate was not high enough for the scene,
- Added flicker: usually caused by poor processing methods,
- Contouring: caused by the source material not having enough bit depth (e.g. 8-bit SDR),
- Compression artifacts, including blocking and loss of fine detail.

Motion judder can be reduced either by increasing capture frame rates, or by increasing shutter angles at the camera. However, increased shutter angles add blur for objects moving with respect to the camera frame. If the eye is tracking these objects, they will appear more blurred – again losing immersiveness.

Noise is inevitable during capture. The degree of noise is inversely related to immersiveness, as well as compressed bitrate. Typical noise reduction techniques apply spectral or spatial filtering to each frame separately. These techniques are limited in their ability to distinguish noise from fine detail, resulting in a blurred appearance due to washed-out low-level details. High quality motion-compensated noise reduction can do a much better job at reducing noise without impacting local contrast, and should be the first choice if at all possible.

Without visible quality loss, effective MC noise reduction can also improve PSNR in lossy compression by between 5 to 8 db at Blu-ray quality levels [6].

When transitioning to HDR, it is important to at least maintain or even improve SDR production quality, while at the same time delighting consumers who invest in HDR TVs. This may requires improved quality processing in the workflow chain over what has been the broadcast norm. For UHD systems, viewing distances of around one and half picture height need be assumed for home theater setups - much closer than the three-picture height "common usage" scenario assumed up to now.

Without proper handling, contouring is also likely to occur when converting 8-bit SDR sources to HDR, also spoiling immersiveness. In this case, adding dither is not enough to eliminate contouring artifacts.

5. SDR/HDR Distribution Issues

5.1 Legacy content

A large, high quality HDR display can reveal issues with legacy SDR material which may have been previously masked by smaller, dimmer SDR displays. All newer broadcast systems will use progressive scan, not interlaced. This means that conversion of legacy interlaced or 3:2 pulldown content to high quality progressive frames will now be required *prior* to any distribution intended for HDR viewing. Furthermore, artifacts from poor quality deinterlacing or low quality frame-rate conversion become much more apparent on an HDR display.

5.1.1 Remastering SDR in Preparation for HDR Distribution

For legacy SDR, it makes sense to go back to the original sources and remaster at a higher quality level using state-of-the-art, computationally intensive algorithms, which can show dramatic improvements over older, more computationally frugal methods. GPU-based system have shown themselves to be effective in this arena.

Current US ATSC SDR broadcasting systems are 8 bits and use 1080i at 59.94 fields/sec or 720p at 59.94 frames/second, and either deinterlace or scale to 1080p 59.94. ATSC's MPEG2 compression can't really be used to send a high enough quality SDR suitable for conversion directly to HDR at the TV.

It is now possible to implement an HDR broadcast system with vivid, backwards compatible 10-bit SDR+ as an intermediary, and also to convert high quality broadcasts of progressive 10-bit SDR to 10-bit HDR at the TV or set-top box. All this can be done while using an open standard SDR broadcast stream with no additional metadata or additional layers needed to assist in the conversion.

When the larger SDR TV market moves to ATSC3.0, it makes sense to broadcast contrast-enhanced, sharper and cleaner SDR+ as a replacement for conventional SDR.

For SDR+ broadcasts, HDR TV purchasers can either use the TV's existing SDR viewing mode, or they could purchase an external high quality SDR to HDR set-top box, giving them improved specular highlights. Even if the final conversion process within a set-top box or TV is proprietary, an open stream standard and workflow means there is no fundamental legal or technical barrier to manufacturers either achieving similar results themselves, or licensing competing alternatives, while allowing the post-production chain to be completely open.

5.1.2 SDR to HDR Conversion for HDR Distribution

For HDR distribution, much the same applies as for SDR distribution, but with the added step of high quality SDR to HDR up-conversion after any necessary remastering. This requires luminance expansion at higher intensities in order to get best results with specular highlights, and best results are usually obtained by mapping the brightest SDR level to the brightest HDR level. However, without proper design, simple expansion can artificially boost bright areas, resulting in unnatural looking images.

Top-range expansion can also cause problems with HDR over-brightness where peak white is used in text, causing some eye strain. This is particularly true for computer monitors, where peak white for text, icons, etc is “baked in” to a lot of computer windows management software.

To support a higher level of absolute nits for specular highlights, TV manufacturers have implemented systems to limit the extent of peak brightness within a scene.

A newly developed automatic inverse tone mapping solution from SDR to HDR has been shown to work well. It is able to map from SDR to any peak luminance supported in an HDR TV in real-time using CUDA GPU acceleration, and it works without a second layer or input metadata. It is also able to convert 8-bit SDR to 10-bit HDR without visible contouring artifacts. This technology can be deployed in a set-top box or HDR TVs. When combined with HDR to SDR+ tone mapping, the two solutions offer an attractive open content delivery alternative, especially for HDR broadcasting.

5.2 Metadata and its Role in Distribution

HEVC encoders now support metadata that specifies the transfer function, color matrix and white point, in addition to maximum possible brightness levels expected in the mastering display luminance. The open HDR10 standard uses ST-2084 for the EOTF and BT2020 for the color space, in conjunction with ST-2086 [7]. It also specifies additional MaxCLL (maximum content light level) and MaxFALL (maximum frame-average light level) parameters [8].

In practice, MaxCLL and MaxFALL cannot be calculated correctly *until all stream content has been processed* [8]. Only then can the final values be calculated and passed in at the *start* of the encoding process, in order to be useful to the TV right from the first frame. However, this is a two-pass algorithm, which *cannot be applied to live signals,* so calculated MaxCLL and MaxFALL values are simply not available from live content.

5.3 HDR Delivery Options

5.3.1 OTT Services

Since OTT services have the flexibility in implementing delivery mechanisms, they have the luxury of choosing the best-of-breed solutions in order to offer the highest quality viewing experience for both SDR and HDR devices. The problem of generating high quality HDR is essentially the same as for other broadcast systems, but they don't have broadcast channel limitations – they simply redirect IP requests to the appropriate SDR or HDR files or streams, as they do now when serving multiple image resolutions.

5.3.2 Broadcast Services

For backward-compatible SDR/SDR+ broadcast, SDR to HDR inverse tone mapping at the TV/set-top box is possible using three approaches:

- Send an entire extra encoded layer with metadata to assist with inverse tone mapping, or
- Use an invertible process for HDR reconstruction with metadata, or
- Use an invertible process for HDR reconstruction without metadata.

For a good review of the first two approaches, see [9]. The third approach, with vivid SDR+, is now possible. It offers the most open workflow, can produce good results from legacy SDR, and is implementable in a TV or set-top box using modern chip technology at a reasonable cost.

The HLG standard is also invertible without metadata, and is simple to implement, but it uses dynamic range squashing rather than tone mapping, which causes brighter highlights to look depressed for SDR viewing. For example, 100 nits gets remapped to roughly 35% dimmer than maximum on an SDR TV. Furthermore, local contrast is severely impacted in both dark and bright areas. See Figures 5a/b, which compare 4,000 nit video images tone-mapped directly to 100 nit Rec709 (SDR+) versus the same 4,000 nits tone mapped to 1200 nits, and then HLG encoded and decoded as Rec709 (SDR compatible decoding). The latter result looks comparatively dull on an SDR TV. This simple observation seems at odds with the relative numbers of SDR versus HDR viewers.

Figure 5a: *Bistro* (original: BT2020 ST-2084, graded to 4000 nits) tone mapped to rec709

Figure 5b: *Bistro* (original: BT2020 ST-2084, graded to 4000 nits) tone mapped to HLG10 and shown as rec709

5.4 SDR/HDR Multi-Region Delivery

5.4.1 Proposed HDR10 Dynamic Metadata – a Double-Edged Sword?

Vivid HDR/SDR tone mapping (up or down) is now possible if the peak input and output luminance is known. File-based systems (including Blu-ray) already have the peak input luminance available at the start of the file, either explicitly via ST-2086 data, or from implicit default values. In broadcast systems, this information may not be available on-the-fly, so the simplest solution is to have broadcast TV stations agree on a predetermined target peak luminance that is within the range supported by the ST-2084 container. This assumption of a maximum target luminance value has already been a part of SDR and will also be used for future HLG10 TV broadcasting.

If a TV's peak luminance happens to agree with the TV broadcast standard, then no further tone mapping (up or down) at the TV is required, which simplifies TV processing. Otherwise, further tone mapping might be needed to match its peak luminance capabilities. Note that large steps downward in tone mapping luminance are more complex to do well, so from that perspective, broadcasting SDR or low HDR peak input luminance values are attractive. Each broadcast channel could use either the standard peak value, or a value for that channel that is predetermined and could be automatically obtained via the Internet. This scheme would be future proof, allowing higher peak luminance values at a later date.

The intention of the dynamic metadata proposal ST-2094 is to make per-frame metadata of master display information instantly available for color/tone mapping, and preserving creative intent when tone mapping HDR to SDR [10]. In practice, this extra flexibility offered by ST-2094 in regards to per-frame dynamic metadata in broadcast standards may only add needless complexity, and create market confusion. Dynamic metadata in an open HDR10 workflow is subject to mishandling or loss in the processing paths during any transition. Wrong metadata can have a very significant negative impact on picture quality. For example, it could cause unexpected and excessive flicker if it is unreliable, or used incorrectly. It may even inadvertently alter the creative intent. To be on the safe side, dynamic metadata should be optional. Tone mapping up or down has been shown to work well without it!

If broadcasting unnormalized HDR, then each program or channel may have different peak luminance values. This could be based on ST-2094 per-frame metadata so that an HDR TV or set-top box can attempt to do a correct conversion from the first frame after tuning in. While this is an ambitious goal, unnormalized HDR moves the major burden of high quality tone mapping onto a set-top box for SDR TVs (up to a 100:1 down-ratio). A difficult trade-off between quality and price may have to be made by set-top box manufacturers.

5.4.2 Metadata Across Regions and Between Formats

Metadata is created in the mastering process in order to help with tone mapping from HDR to SDR, but it seems that its role has not stopped there. Content may need to be distributed across international boundaries, so frame-rate conversion (FRC) may also be needed. High quality FRC has always been a difficult technical problem. With increased TV sizes and brightness, the quality of conversion between broadcast standards is more critical than ever. FRC creates retimed frames which requires that dynamic metadata has also to be accurately recalculated and retimed at the output.

If the metadata is proprietary, then conversion needs to be nested within a proprietary decode and re-encode. This is particularly problematic for dual-layer encoding systems.

HLG or Rec709 do not carry dynamic metadata, so if either are converted to HDR10, then the dynamic metadata is simply not available. Dynamic metadata will tend to get lost in a mixed production environment where other types of sources are involved, as the process of normalization will inevitably drop dynamic metadata in the final result.

5.5 HDR/SDR Bitrates

AVC, HEVC and VP9 all work well for HDR compression [6], and 10-bit HDR encoding typically adds 10-20% to the bitrate vs SDR [11], so either system is viable for mass delivery.

Compressed bit rates can be further reduced using high quality motion-based processing to (a) reduce noise [12], and (b) reduce deinterlacing artifacts such as flicker, which manifests itself as large frame-to-frame changes, which increases bit rates.

The most important decision by broadcasters is whether to encode HD or UHD. UHD increases bitrate vs HD by about 300% [5]. UHD can offer some improvement, but it is only worthwhile *if the quality of every other signal attribute is at a comparable level.* Otherwise, it makes more sense to use HD and take advantage of TV scalers.

6. Transitioning from SDR/SDR+ to HDR

Transitioning broadcast systems to HDR cannot be done in one step. However, it can be done in such a way as to provide an improved SDR+ viewing experience during the transition. This improvement is largely due to the increased dynamic range of HDR cameras, 10-bit encoding, and vivid HDR to SDR tone mapping. The transition from HDR to SDR+ could move out from camera capture in a "*wave-front*", through post-production to delivery systems. There it can wait for the switch from ATSC to ATSC 3.0, and possibly an additional switch from SDR/SDR+ to HDR broadcasting.

This upgrade process is entangled with the transition from SDI to IP-based ST-2022-6, so it is important to make sure that new systems handle both 10-bit SDR and HDR10, particularly as a mixed system requires additional signaling to distinguish them, which is not explicitly supported in legacy SDI. IP-based standards are far more flexible, so audio and other signaling can be carried along with the video.

If handled correctly, the SDR viewer will see a genuine improvement in picture quality during an SDR+ to HDR transition. On the other hand, if simple top-level squashing is used, they will see a degradation.

Main factors and trade-offs to consider when selecting an HDR/SDR solution are:

- Ease of migration from SDR to HDR,
- Flexibility to handle legacy SDR formats,
- SDR/SDR+ quality during migration,
- Ease of interoperability between open encoded formats,
- Cost, including additional royalties,
- Adaptability for future open formats.

7. HDR/SDR Conversion Implementation

Automatic HDR to SDR+ tone mapping, as well as SDR to HDR up-conversion have been encapsulated and implemented in a GPU accelerated system. It can be configured for file-based or on-the-fly applications. The system may be deployed either in the cloud, or stand-alone with I/O combinations of SDI (up to 3g), or 10gbit IP (regular TCP/IP protocols, or SMPTE ST 2022-6), allowing flexible interfacing and conversion to any open format.

8. Conclusions

HDR/SDR+ immersive viewing experience in an SDR world is almost here, thanks to continued advancement in new cameras, HDR recording, post-production, encoding and display technology. Newly available automatic SDR to HDR up conversion breathes new life into valuable legacy content for HDR displays. Together with vivid HDR to SDR+ tone mapping, these solutions present OTT service providers an attractive alternative, using completely open standards to deliver both HDR and SDR+ from any source type, while also efficiently preserving HDR creative intent for SDR+.

High quality 10-bit SDR+ could be used with ATSC3.0 for broadcasting. HDR TVs or set-top boxes can be designed to support SDR/SDR+ to HDR inverse tone mapping without additional metadata. This allows both legacy SDR content, as well as multi-region processing and delivery to work seamlessly.

Acknowledgements

Special thanks to Jan Froelich and Stefan Grandinetti of Stuttgart Media University for making high quality HDR test sequences available at: https://hdr-2014.hdm-stuttgart.de/.

Addendum

All of the Stuttgart HDR test sequences (including the image excerpts used in this paper) will be shown at the SMPTE Exhibition hall.

References

[1] Mark Schubin, NAB 2016 Wrap-up, 08/06/2016.
[2] Scott Miller, "A Perceptual EOTF for Extended Dynamic Range Imagery", 5/2014.
[3] T. Borer; A. Cotton. "A Display Independent HDR Television System", 11/1/2015.
[4] "ARIB STD-B67", 7/3/2015.
[5] ITU-R BT.2100-0, 07/2016.
[6] How HEVC/H.265 supports High Dynamic Range (HDR) and Wide Color Gamut (WCG), 11/16/2015
[7] SMPTE, "Study Group Report High-Dynamic-Range (HDR) Image Ecosystem", 10/2015.
[8] Ultra HD Forum: Phase A Guidelines, 7/15/2016.
[9] Jim DeFilippis: "A New Day Dawning - HDR Delivery", 5/27/2016.
[10] "10E Dynamic Metadata for Color Transforms of HDR and WCG Images", https://kws.smpte.org/kws/public/projects/project/details?project_id=294", 6/14/2016 (draft)
[11] Next Generation DTV: ATSC 3.0, 2015.
[12] http://isovideo.com/isovideo_denoise_results_April_2016.php, 4/2016.