Live HDR video broadcast production

Igor G. Olaizola, Angel Martín, Josu Gorostegui

June 20, 2016

Abstract

Among the multiple use cases of HDR video imaging, live events can get a big benefit from HDR video imaging, especially when they are recorded outdoors under uncontrolled light conditions. HDR technology can provide a better adaptation of cameras to the rapidly changing light conditions such as scenes that combine bright sunny areas with dark shadows (football), balls that fly with a very bright sky in the background (golf, football), extremely rapidly changing light conditions (subjective cameras in Formula 1, concerts with flashing lights), etc. All these use cases introduce two main technological challenges: on the one hand the real-time factor which does not allow any manual intervention nor computationally demanding image data processing step. On the other hand, the end-to-end production pipeline has to preserve all the dynamic range information.

1 Introduction

Among the multiple use cases of HDR video imaging, live events can get a big benefit from HDR video imaging, especially when they are recorded outdoors under uncontrolled light conditions. HDR technology can provide a better adaptation of cameras to the rapidly changing light conditions such as scenes that combine bright sunny areas with dark shadows (football), balls that fly with a very bright sky in the background (golf, football), extremely rapidly changing light conditions (subjective cameras in Formula 1, concerts with flashing lights), etc. All these use cases introduce two main technological challenges: on the one hand the real-time factor which does not allow any manual intervention nor computationally demanding image data processing step. On the other hand, the end-to-end production pipeline has to preserve all the dynamic range information.

1.1 SDTV, HDTV and UHDTV

Digital TV and Video technologies started based on analog video formats. Therefore, Standard Definition TV (SDTV) takes its visual characteristics from NTSC (480/60i) and PAL/SECAM (575/50i) as it is defined by the ITU Recommendation BT.601[3] as well as by SMPTE 259M[13]. High Definition TV (HDTV) introduces a big change in spacial resolution (720p, 1080i, 1080p) but the color gamut is similar to SDTV and other aspects as the frame-rate are note increased. HDTV is defined by the ITU-R BT.709[4] and shares the same color primaries as sRGB. While the resolution in HDTV can be considered as acceptable for usual distances between the TV screen and the viewer, both, the color gamut and dynamic range of the Human Visual System (HVS) are far beyond the capabilities of HDTV. Therefore, Ultra High Definition TV improves all the dimensions involved in video: Resolution, frame-rate, color gamut and dynamic range. UHDTV has been defined by ITU in ITU-R BT.20202[2]. This recommendation can by summarized as:

- Aspect Ratio: 16:9 (square pixels)
- Resolution: 4K (3 840 2 160) or 8K (7 680 4 320)
- Framerate (Hz): 120, 120/1.001, 100, 60, 60/1.001, 50, 30, 30/1.001, 25, 24, 24/1.001
- Only progressive scan mode
- Coding format: 10 or 12 bits per component
- Wide Color Gamut:
- Opto-electronic transfer
- High Dynamic Range (not defined yet)

Regarding the HDR capabilities of BT.2020, even if there are not specific details about how to process and represent HDR data, there is a real constraint introduced by the 12 bits per component defined as the maximum bit-depth.

1.2 Real-time HDR video processing

Real-time processing of HDR data is demands extremely high computational power. A preliminary analysis of the throughput required of raw video is shown in Equation 1:

$$T_h = R_{xy} \cdot F_r \cdot 3b_d \tag{1}$$

For HD quality, we can consider R_{xy} as 1080i (BT.709 SMTP 292M 1920x1080 interlace). In this case at 50fps and 10 bits per channel we obtain 1.45Gb/s. For 720p, the same equation gives 1.29Gb/s. In the case of software based HDR video processing, GPUs are the solution to provide the required real-time performance. For dedicated video processing hardware, the current main limitation comes from the bit-depth that such solutions are able to accept. Typically, hardware solutions do not go beyond 10bit, a depth that might not suffice for high quality HDR imaging.



Figure 1: BT.2020 Color Space (Creative Commons BY-SA 3.0)

1.3 Dynamic Range preservation in live broadcast production workflows

Professional broadcast production workflows are mostly hardware based. This specific hardware design provides a very good performance together with a high reliability. However, they lack of flexibility to introduce changes in formats or signal characteristics and therefore legacy aspects become a key issue in this kind of infrastructures. Moreover, the big economical investments that these infrastructures imply allow only incremental changes that must ensure the complete interoperability with the rest of the elements in the pipeline. In this sense, standards take a key role to ensure this interoperability. Current production systems (HDTV) do not support more than 10bit flows that will be reduced to 8bit when the signal is going to be distributed.

1.4 Main aspects of quality preservation

There are several aspects that become critical in terms of content quality when introducing HDR flows into current broadcasting infrastructures. All these aspects are well known in HD or SD but become even more critical for HDR.

1.4.1 Banding

Banding effects are directly related to quantization errors. As the dynamic range increases, the quantization effects become more noticeable and banding artifacts arise. To avoid this, smaller quantization steps must be taken, but this implies a higher bitdepth (from 8 to 10 or 12 in order to become backwards com-



Figure 2: Basic Live TV Broadcasting Workflow

patibility with existing infrastructures, and ideally 14 or 16 bits) and non-linear transform functions based on the Human Visual System (HVS) that minimize the observable banding effects. Nowadays, this is partially solved by the EOTF and OETF mapping curves that will be introduced in a further section.

1.4.2 Chromatic deviations

Chromatic deviations are side effects of tone mapping operators. Ideally tone mappers only affect the image luminance and chromatic features are preserved as they operate in color spaces where luminance and chrominance values are decorrelated[17]. However, there is always a degree of correlation that can be altered if only the luminance channel is transformed. Boitard et al. [14] evaluated this chromatic deviations on the most popular video tone mapper or transfer functions. According to their study the most robust color space is YD_uD_v derived fro CIE L'u'v' (Eq. 2).

$$Y = L, \ D_u = \frac{u'}{0.62} - 0.5, \ D_v = \frac{v'}{0.62} - 0.5$$
 (2)

2 Digital TV basics

Live Broadcasting technologies imply to main steps, contribution and distribution.

2.1 Capture

Content can be created either by capturing real scenes with HDR cameras or through synthetic generation by using computer graphics (CG). While CG systems theoretically do not have actual limitations for the dynamic range, cameras are still facing strong technical problems for UHD.

There have been several techniques to extend the dynamic range of sensors. Bracketing (process of taking several frames at different exposures in order to get a single fused frame with a higher dynamic range) has been extensively used for still images and is used as well in video capturing. However, bracketing introduces ghosting effects over movement and limits the framerate. The use of two or more cameras at different exposure values combined with bracketing techniques is another solution to obtain a higher dynamic range. Experimental cases have been able to record 18 f-stops by using this approach. However in this case parallax problems arise. A possible solution to remove the parallax is the use of mirror based stereo rigs adjusted at 0 parallax. However, all these setups become complex and not feasible for regular TV broadcasts.

The use of multi-sensor cameras has been the solution adopted by some manufacturers (SpheronVR) that even if useful for other use cases such us engineering or cultural heritage, are too complex and not accurate (blurry) for live broadcasts.

Camera sensors (mostly based on CMOS technology) have rapidly evolved towards very high resolution values reaching 4K. However, this fact implies a smaller sensor size with some negative side effects for dynamic range such as noisier response and smaller dynamic range. Once 4K has become a common feature in most professional camcorders, dynamic range is becoming the next big trend where manufacturers such as Black Magic, Sony, GrassValley, ARRI, etc. already can capture up to 15 f-stops.

This quick evolution of sensors and cameras has introduced some issues that are still being solved by manufacturers. Global shutting (relevant for sports) is still not a feature that all HDR cameras include. Moreover, there is a lack of standard interfaces to get the captured content at full quality (raw or high quality encoding), where manufacturer are mainly applying different OETF curves.

2.2 Manipulation

Manipulation includes applications such as color grading, post production, image based lighting and scene analysis. While off-line processes are done in general by using commercial editing tools such as Avid Media Composer, Da Vinci Resolve from BlackMagic, Speed Grade from Adobe, Sony Vegas, Baselight from FilmLight, Autodesk Lustre, etc. There are still very few tools that allow the real-time processing of HDR video: SGOs Mistika and Vicomtech's Alchemist are two specific solutions for real-time HDR processing. Most of these solutions use GPU based acceleration to optimize the performance, for example, Blackmagic Da Vinci Resolve can operate with 12 bit RAW content, and apply different techniques for tone mapping.

For software based video manipulation, $OpenCV^1$ is probably the most popular image processing tool. The last version at this moment of OpenCV (3.0) includes GPU functions and specific instruction for HDR manipulation, specially tone-mapping operators. Halide² shows a novel approach (it was created in 2012) as a functional language that targets different compilers including CUDA³ and OpenCL⁴.

OpenGL⁵ and OpenSceneGraph⁶ at a higher abstraction level offer advanced GPU based graphics that can be used for real-time HDR video rendering.

2.3 Contribution

For real-time processes one of the main limitation comes from the physical interfaces. SDI supports up to 12 bit video, and there are commercial solutions able to deal with this bit depth. However, most of the real mixing and processing workflows support operations up to 10 bit and as they are based on specific hardware designs, it is not straight forward to upgrade to 12 bit. This can be a real barrier for the deployment of one of the most relevant use cases for HDR video: live sport transmissions. GPU based solutions are therefore becoming the main trend to provide this kind of high-end solution with real-time performance. However, there are still no commercial solutions that preserve the whole HDR data without tone-mapping operations along the manipulation processes.

2.4 Distribution

TV signal distribution is currently performed in 4 main ways. Satellite, cable or terrestrial broadcast and broadband infrastructures. Classical distribution has been done using broadcast networks which are dedicated infrastructures that require big investments in deployment. Wireless broadcast systems (satellite and terrestrial) do not have scaling limitations as they can provide the content to an unlimited amount or receivers located within the coverage area. On the other hand, cable TV and IP networks can be used for different purposes and are inherently bidirectional and might face problems for massive distribution. The widespread of the Internet availability and the improvement in bandwidth and QoS of IP networks has enabled new business models based on services that are offered directly by Internet Service Providers (in a similar model to cable operators) or even by content providers that do not own any specific infrastructure and operate Over The Top (OTT). Youtube and Netflix are two of the main OTT providers where Youtube offers prosuming content while Netflix acts as a OTT based broadcaster. Internet based delivery networks offer two

¹http://opencv.org/

²http://halide-lang.org/

³http://www.nvidia.com/object/cuda_home_new.html

⁴https://www.khronos.org/opencl/

⁵https://www.opengl.org/

⁶http://www.openscenegraph.org/



Figure 3: Single server centralized delivery system vs. CDN architecture

big advantages compared with classic broadcasting infrastructures. First, the continuous evolution and deployment of Internet infrastructures (both wired and wireless) are driven by a huge business model that includes all connected media, services and devices, ensuring their improvement in a long term. Second, there are not technical or regulatory constraints in content transmission. These two advantages make OTT based solutions the easiest environment for UHDTV and HDR content distribution. The scalability problem of Internet distribution infrastructures is nowadays solved by CDNs (Content Delivery Network) that basically create intermediate proxies as it can be seen in Figure 3.

2.4.1 Broadcasting Standards

Broadcasting systems vary between countries, specially for terrestrial transmission. DVB-T[5] (Digital Video Broadcasting) is the most widespread standard, ATSC (Advanced Television Systems Committee) has been adopted in North America and the Japanese System ISDB-T (Integrated Services Digital Broadcasting)[11] is the main standard for South America. DTMB (Digital Terrestrial Multimedia Broadcast)[6] is a system developed and adopted in China and few other countries such as Cuba. Most systems share similar modulation principles (QPSK, OFDM, VSB for ATSC) and include MPEG-2[8] and MPEG-4[12] encoding standards. DVB-T is based on COFDM modulation and for HDTV the most adopted solution has been H.264 for video encoding and AAC(Advanced Audio Coding)[15, 9] for audio. Typical DVB-T parameter configurations establish a bandwidth of around 20Mb/s in each multiplex and the maximum bandwidth allowed by the standard is 31.668Mb/s (64QAM FEC:7/8, Guard Interval: 1/32). The new version of DVB-T, DVT-T2 extends the bandwidth up to 50Mb/s.

2.4.2 Encoding

For the highest quality configurations of UHDTV bandwidth requirements can exceed the capabilities of existing distribution infrastructures. 8K@120fps and



Figure 4: DTT Systems Distribution Worldwide (Public Domain License)

12 bits per component means 57.6 times FullHD $(1080@50p)^7$ and 115.2 times 1080@50i. resolution means 16 times more pixels than in Full HD (1080p) and 32 times more than 1080i.

Most proposed solutions for HDR video encoding in broadcast distribution are based on existing LDR video encoders. In order to adapt HDR signals to existing encoders the captured dynamic range has to be quantized to the range where encoding methods can deal with the input signals. As it will be presented in Section 3.1, there are two main approaches to fit HDR data into LDR workflows, the single layer approach where the original signal is mapped through transfer function curves to the desired number of bits and the dual layer approach where the signal is split into two LDR flows.

For distribution, encoding is one of the most critical parts as encoders have to compress the signal in one order of magnitude from contribution formats, they have to process the signals in real-time and they have to ensure a high level of quality. In the previously proposed workflow and for the sake of backwards compatibility with currently existing broadcast infrastructures, the use of commonly adopted coding standards is proposed. While contribution formats can be used at 12 bit, it is not expected to go beyond 10 bit in distribution. H.264/MPEG-4 AVC[10] is the most extended codec right now and it's successor H.264/HEVC[7] is planned to be as one of the main formats for UHD. There are several organizations such as the Blu-ray Disc Association (BDA), the High-Definition Multimedia Interface (HDMI) Forum, and the Ultra-High Definition (UHD) Alliance that have adopted HEVC Main 10 Profile as single layer solution codec[16] where ITU-R BT.2020 wide color gamut is included. This HEVC profile has been defined as HDR10 by the Consumer Technology Association and sets the parameters for the current capabilities of commercial LCD displays (10 bit).

 $^{{}^{7}}R = \frac{8K_{res} \cdot 120_{fps} \cdot 12_{bpc}}{HD_{res} \cdot 50_{fps} \cdot 8_{bpc}} = \frac{3840 \cdot 2160 \cdot 120 \cdot 12}{1920 \cdot 1080 \cdot 50 \cdot 8} = 57.6$

2.5 Reproduction

The rendering and reproduction of HDR content requires displays with two main capabilities. First, the end device has to decode the transmitted signal mapping the information to the dynamic range capabilities of the display that might vary. Second, the display has to be able to reproduce a higher dynamic range than standard commercial devices. To increase the dynamic range, the key aspect of displays is the brightness. While average commercial displays do not go beyond $500Cd/m^2$, professional HDR displays can provide $4,000Cd/m^2$ (Dolby Pulsar) or $6000Cd/m^2$ (SIM2⁸)). Last consumer electronics devices announced in 2015 (e.g: Samsung and LG displays)) claim $1000Cd/m^2$ of peak luminance at 10 bit.

In general, the aforementioned professional devices offer around 14 f-stops which is pretty aligned with the capabilities of current professional video cameras suitable for HDR (e.g.:Sony F65, ARRI Alexa, Grass Valley LDX 86, Canon EOS C300, Black-Magic URSA, etc.)

3 Integration of HDR flows in standard broadcasting pipelines

Broadcast TV workflows encode the different streams (audio, video, metadata) in packets that are signalized according to the corresponding standard (DVB, ATSC, etc. introduces some specific metadata while most content related metadata are defined by MPEG-2 and MPEG-4).

3.1 Single layer vs. dual layer

In order to introduce HDR content to these distribution workflows, signals have to fit in up to 12 bit for contribution and up to 10 bit for distribution purposes. Two address these interoperability issues two main strategies can be followed.

3.1.1 Single layer

One of them is based on mapping HDR signals into 8, 10 or 12 bit by using the so called OETF (Opto-Electronic Transfer Function) curves that compress the information while trying to keep both the artistic intent and the qualitative perception of the Human Visual System (HVS). At the displaying stage, the compressed signal is mapped backed and expanded to HDR by using the EOTF(Electro-Optical Transfer Function). The whole system can be modeled as the OOTF (Opto-Optical Transfer Function) as it can be seen in Figure 5.

There are different OETF/EOTF proposals created by Dolby, BBC/NHK, Philips, Technicolor, Sony, ARRI, University of Warwick, etc. Figure 6 shows

⁸http://hdr.sim2.it/

¹⁰https://www.flickr.com/photos/alanfranco/5002543978/in/

photolist-8C4m1u-8SRBX8-poK7kR-8C4m3f-8C4kZd-8RD2ua-98acdd-5M2rxE-5M9p4e-8TqtyQ-8C1fjX)



Figure 5: OOTF Workflow (HDR Picture under Creative Commons License. Author: Alan Franco¹⁰)



Figure 6: Main OETF and EOTF proposals

the most popular ones, Dolby's PQ published as SMPTE ST 2084[18], and the Hybrid Log Gamma (HLG) presented by BBC and NHK, recently standardized as ARIB STD-B67[1].

4 GPU-based real-time manipulation and monitoring

HDR processing algorithms can be typically implemented pixel or regionwise. This fact allows the use of GPUs to boost processes that tend to be computationally intensive. This way, GPUs can provide high performance and much more flexibility than dedicated hardware. Following this approach, Vicomtech-IK4 has developed a prototype (Tebas Platform, Figure 7) able to perform real-time analysis of 16bit HDR video flows (Histograms, wave forms, vectorscopes,

etc.) as well as to manipulate this content for color transformation, mixing, effects, etc.



Figure 7: Screenshot of the GPU based real-time HDR flow analysis and manipulation platform [Vicomtech-IK4].

GPU based solutions use commercial graphic cards that currently are improving faster than CPUs and moreover, offer high flexibility as any new functionality or format interface can be introduced just by adding a new software module without hardware design constraints. Therefore many video processing functionalities (coding, filtering) are migrating more and more to GPU architectures.

5 Conclusions

High Dynamic Range is one of the most relevant characteristics of Ultra HD that introduces a dramatic difference from HD quality and thus it can be a key factor to leverage the UHD market. Furthermore, the technological state of the art is ready to provide end-to-end solutions at 14 stops. However, there are interoperability and backwards compatibility aspects that have to be taken into account to ensure a smooth migration from HD infrastructures to HDR workflows. Currently the industry has accepted the use of 12 bits for production and contribution while distribution will be set to 10 bit. Due to this relatively low bitdepth, OETF/EOTF transfer function become critical to reduce quantization effects and maximize the quality according to the Human Visual System.

As said, backwards compatibility and cost-effectiveness are two main aspects for the successful deployment of HDR video technology, however, there must be a trade-off with the risk of obsolescence of the adopted solutions. Selected approaches must be future proofing in order to convince the entire value chain



Figure 8: Operation set-up of the GPU based real-time HDR flow analysis and manipulation platform [Vicomtech-IK4].

(technology manufacturers, content producers, broadcasters, network operators, etc.) to join the HDR video business.

References

- [1] ARIB STD-B67 Essential parameter values for the extended image dynamic range television (EIDRTV) system for programme production.
- [2] BT.2020 : Parameter values for ultra-high definition television systems for production and international programme exchange.
- [3] BT.601 : Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios.
- [4] BT.709 : Parameter values for the HDTV standards for production and international programme exchange.
- [5] ETSI EN 300 744 European Standard (Telecommunications series) Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television.
- [6] GB 20600-2006 Framing structure, channel coding and modulation for digital television terrestrial broadcasting system.
- [7] H-265 SERIES H: AUDIOVISUAL AND MULTIMEDIA SYSTEMS Infrastructure of audiovisual services Coding of moving video High efficiency video codingitu-t.

- [8] ISO/IEC 13818-2:2013 Information technology Generic coding of moving pictures and associated audio information Part 2: Video.
- [9] ISO/IEC 13818-3:1998 Information technology Generic coding of moving pictures and associated audio information – Part 3: Audio.
- ISO/IEC 14496-10:2010 Information technology Coding of audio-visual objects – Part 10: Advanced Video Coding.
- [11] Transmission system for digital terrestrial television broadcasting ARIB standard.
- [12] ISO/IEC 14496-2, Information technology Coding of audio-visual objects
 Part 2: Visual, 2001.
- [13] For Television Serial Digital Fiber Transmission System for SMPTE 259M, SMPTE 344M, SMPTE 292 and SMPTE 424M Signals. SMPTE ST 297:2006 (Nov 2006), 1–24.
- [14] BOITARD, R., MANTIUK, R. K., AND POULI, T. Evaluation of color encodings for high dynamic range pixels. In SPIE/IS&T Electronic Imaging (2015), International Society for Optics and Photonics, pp. 93941K– 93941K.
- [15] BOSI, M., BRANDENBURG, K., QUACKENBUSH, S., FIELDER, L., AK-AGIRI, K., FUCHS, H., DIETZ, M., HERRE, J., DAVIDSON, G., AND OIKAWA, Y. ISO/IEC MPEG-2 Advanced Audio Coding. In *AES-101st* (1996). Preprint 4382.
- [16] DONG, Y., NASIOPOULOS, P., AND POURAZAD, M. T. Hdr video compression using high efficiency video coding (hevc). *reproduction 6* (2012), 7.
- [17] PEDZISZ, M. Beyond bt. 709. In Annual Technical Conference & Exhibition, SMPTE 2013 (2013), SMPTE, pp. 1–13.
- [18] SMPTE, S. 2084: 2014. High dynamic range electro-optical transfer function of mastering reference displays (2014).