HIGH DYNAMIC RANGE VIDEO CHAINS

Erik Reinhard, Edouard François, Ronan Boitard,
Christel Chamaret, Catherine Serré, Tania Pouli

Technicolor Research & Innovation, France

ABSTRACT

Over the last decade, much research experience has been gained in the realm of high dynamic range imaging (HDRI). On account of the significantly improved visual experience that HDRI offers, industry has recently taken a strong interest in this set of technologies. This is evident in the active consideration of HDR by standardization bodies, for instance DVD and EBU in the context of UHDTV, and the emergence of demos at tradeshows.

To enable a more engaging/immersive viewing experience in the home or in the cinema, a range of technologies, workflows and processes should be adapted, and HDRI would need to evolve to become HDRV: high dynamic range video. There exists a good number of solutions that each solves part of the problem. However, as a significant number of problems remain open, there is, up to now, no integrated pipeline that includes everything from capture to post-production, archival/storage, compression, transmission, and display.

In this paper, we review the requirements for building high dynamic range video pipelines, and discuss how and where HDR components can and should be inserted. This leads to a range of different scenarios, each with their own expected quality levels, associated cost and required effort to implement in practice. In particular, we consider an ideal end-to-end HDR video chain, as well as several variants that involve either legacy content or conventional decoding and rendering devices.

INTRODUCTION

The video industry has been permanently trying to offer improved viewing experience to attract consumers to their products. Much effort is currently expended to create infrastructure to support 4K resolution. The UltraHD community, however, has realized that the transition from HD to 4K could not be sufficient for the end-user, as under normal viewing conditions these resolutions are beyond the resolving power of the human visual system. For this reason, UltraHD is aligning itself with additional technologies, of which wide color gamut and high dynamic range video are perhaps most prominent.

This paper discusses one aspect of this, namely high dynamic range imaging. We note that this set of technologies is complementary to high pixel resolutions, and may therefore be adopted independently of the transition to 4K resolution. At the same time, we see wide color gamut and high dynamic range imaging as two sides of the same coin. High dynamic range imaging/video concerns itself predominantly with extending the range of representable luminance values, whereas wide color gamut technologies aim to increase the range of representable chromaticities. These are therefore orthogonal dimensions.

However, due to interactions between saturation, hue and luminance in the human visual system, it would be beneficial to treat wide color gamut and high dynamic range imaging as one integrated set of techniques.

For those who have seen material displayed on a high dynamic range display, it is immediately clear that the visual experience is significantly richer, more life-like, and produces a better sense of immersion. Currently, however, such experiences are only proposed at trade-shows and private demos in a select set of universities and companies. After at least a decade and a half of research on this technology, the fundamentals of high dynamic range imaging are relatively well known. What is less clear, however, is how to leverage this knowledge, and how to design high dynamic range video pipelines for the purpose of bringing the aforementioned level of immersion to consumers in their home or in their local cinema. We see several stumbling blocks along the way, which are briefly enumerated here:

1. Video data needs to have a sufficiently high fidelity to maintain its immersive qualities. From the capture and all the way to its delivery in the home/cinema, the video stream will have to be represented in high dynamic range. Whatever this means, and however this is done, it will have to involve more than 8 bits per pixel per channel. This requires a redesign of several parts of the pipeline, although part-way solutions already exist. How best to achieve this remains to some extent an open question. In this paper we review current work- and data-flows and assess existing transmission/communication bottlenecks. We then discuss possible solutions that crucially avoid the need for upgrading current infrastructure.
2. In post-production, color grading will happen on a display with fixed capabilities. These capabilities will be high-end, but in the near future, these may not surpass the abilities of consumer displays anymore. For instance, Dolby's professional grading monitor [1] supports a DCI P3 gamut with a peak luminance of 700 nits. On the other hand, some displays currently being readied for the market already exceed this peak luminance. As a result, color grading will increasingly happen in a studio and on a display that does not match the consumer's home environment. Both room illumination and display capabilities will increasingly differ. Moreover, the variability in home viewing environments will increase. As a result, conveying the director's intent will become increasingly difficult. In this paper we discuss possible solutions to this problem.
3. In the realm of 3D cinema it is reasonably common to create 3D movies by means of post-processing a single stream of footage. A similar approach is followed in '4K ready' televisions which upscale HD content to their native resolutions. This practice could also be applied to high dynamic range video, creating opportunities for simulating high dynamic range video content either during post-production or by devices located in the home. We outline the requirements of such algorithms and suggest reasonable use cases.

In the following, we discuss these issues in more detail, focusing on cinema productions. The concepts by and large extend to (live) broadcasting as well, and this will be indicated where appropriate.

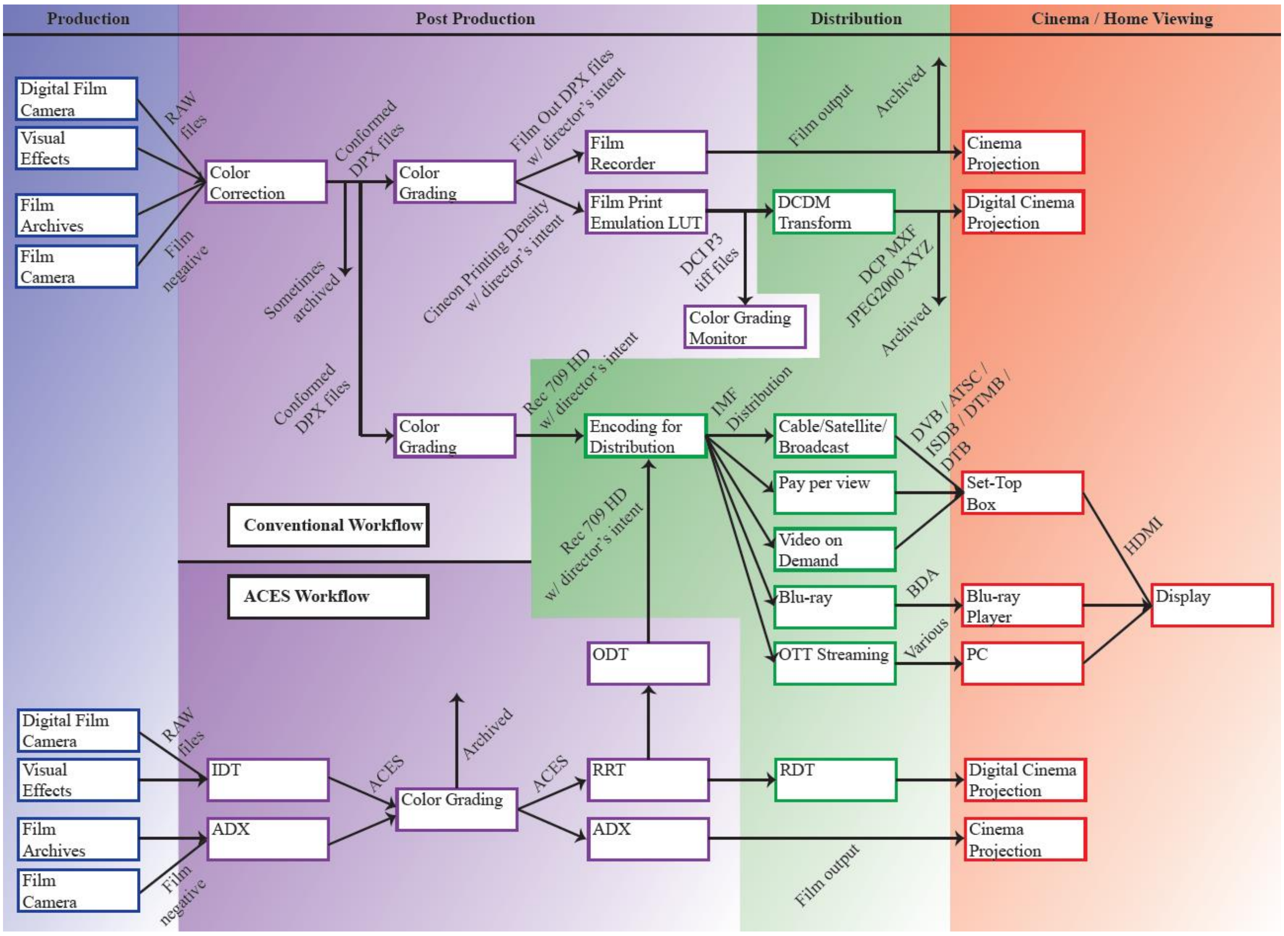

Figure 1. The flow of video data from capture to post-production, distribution, and finally display both in the cinema and at home. Conventional post-production as shown at the top, while the ACES workflow is shown at the bottom.

CURRENT DATA FLOWS

With this paper we are interested in facilitating the adoption of high dynamic range video technologies. There are generally four broad stages of processing involved in getting film and video from its source to the final consumer. These are production, post-production, distribution, and finally in-home processing and display. The general (and simplified) flow of video data through these four stages is depicted in Figure 1. Some parts of the path data takes from capture to home delivery are already compatible with high dynamic range video, whereas others provide functional and/or conceptual bottlenecks. Here, we discuss the various stages as well as their suitability for high dynamic range video delivery.

Production

Video and film can be captured using a camera, or can be computer-synthesized. VFX production adds computer-generated content to previously captured footage. Capture devices are rapidly becoming more powerful, with the likes of Arri [2], Red [3] and Sony [4] marketing film cameras that produce around 14 to 16 claimed f-stops of dynamic range. In practice several f-stops are lost due to noise and the extra headroom is often used to allow set designers to light their sets for dramatic effect, rather than to provide sufficient light to allow the camera to capture the scene. In the context of live capture and nature photography, where only natural light is available, further headroom may be desirable.

Post-Production

In post-production, footage from different origins are normally first corrected. In a second color grading step, the director's intent is applied. This step is separate for different target formats. In particular, color grading for cinema and home viewing requires separate work as the viewing environment in the cinema is dark, whereas in the home it is usually dim (albeit generally much less controlled). We note that there exist many different workflows and most colorists adopt their own personalized approach to color grading. As an alternative, ACES [5] brings a more unified workflow, which also fully supports wide color gamut and high dynamic range processing. As many color grading systems now support the ACES system, color grading can supposedly be applied to high dynamic range footage.

In practice, color grading for high dynamic range imaging is not straightforward, in part due to the lack of high dynamic range color grading monitors. Here, the options are the Sim2 HDR display [6], which has a very high peak luminance (2500 cd/m^2 sustained, and up to 4000 cd/m^2 for short periods of time), the Dolby professional reference monitor [1] (DCI P3 color gamut with 600 cd/m^2 peak luminance), and Dolby's Pulsar, which should achieve 4000 cd/m^2. All these displays are based on the dual-modulation technology [7].

A larger problem that is beginning to emerge in the post-production studio is that home displays are becoming increasingly more varied. This means that the combination of target display and target viewing environment is increasingly unlike the color grading set-up. It will be impractical to ask colorists to re-grade the same material for a wide variety of different target environments. This means that some automatic processing, possibly with the aid of metadata, will become increasingly necessary to enable the effective communication of the director's intent, while at the same time leveraging the capabilities of the consumer's display devices. This could take the form of tone mapping ([8, 9]), inverse tone mapping ([10, 11]), and/or gamut processing (e.g. [12]).

Distribution

Once the colorist is finished, the movie is adapted for cinema, broadcast/television, airplane entertainment systems, different language/subtitle regions, and censored for certain markets. These hundreds of different versions then have to be transported from the studio to the end consumer.

In the case of digital cinema, the Digital Cinema Distribution Master (DCDM) has replaced film printing, and allows the studios to deliver a Digital Cinema Package (DCP), a well-defined set of files on a hard-drive to cinema operators. At the heart of this format, individual frames are encoded as 16-bit TIFFs.

Distribution to end-users happens in two steps, with wholesalers and retailers sitting in-between the studio and the consumer. These include broadcasters, streaming, pay-per-view and video-on-demand services, as well as disk-based delivery methods. To deliver a movie to such retailers, the Interoperable Master Format can be used, which is being standardized by SMPTE. This format helps to streamline the management of the many different versions that are produced for each movie. The format uses either constrained or unconstrained MXF containers with in it JPEG2000 [13] encoded frames. The JPEG2000 image format is integer-based, but accepts input data in a range of bit-depths.

Broadcasting, streaming, Blu-ray Disks and other methods for home delivery have limits on bandwidth. Broadcasters, for instance, are commonly targeting bit-rates around 5-10 Mbps for HD video. UHD distribution is foreseen at bitrates typically around 10-15 Mbps [14]. This means that high compression is required. This is governed by standards such as DVB, BDA, ATSC and others, as shown in Figure 1. Most of these standards utilize the ITU-T H.264 / MPEG-4 Part 10 'Advanced Video Coding' (AVC) specification [15] as the video compression scheme (for HD delivery), with ITU-T H.265 / MPEG-H Part 2 'High Efficiency Video Codec' (HEVC) [16] following as its more capable replacement (envisioned for the UHD deployment).

To enable high dynamic range video to be broadcasted, streamed or delivered to the home via other methods, the current codecs should either be replaced or augmented. Due to the time it takes to agree on new standards, the most practical route is to find ways to shoehorn high dynamic range video data into existing framework. This can be achieved by pre-processing the signal before its compression. The pre-processing mostly consists in quantizing (after a non-linear transformation) the HDR signal with limited bit-depth (typically 10 or 12 bits per component) before encoding it with limited bit-depth encoders [17]. This of course implies a de-quantization of the decoded signal before rendering. This approach is the most straightforward solution, but does not offer backward compatibility with SRD devices (decoders and displays). In addition the interactions between the HDR signal quantization and the compression steps have to be carefully studied and evaluated. They could lead to limited compression efficiency and typical compression artefacts such as banding. Alternatively, it would be possible to split the signal in an intelligent manner into a Low Dynamic Range signal adapted to legacy encoders and decoders, with light weight side information allowing the conversion of the LDR signal into an HDR signal [18]. This approach can offer the backward compatibility with SRD devices (decoders and displays) and can guarantee a higher coding efficiency by better taking into account the HDR signal properties. In either case, at the consumer side, the signal will be reconstructed by inverting the compression process.

As both the codec and the additional processing will be lossy, it is paramount that the combined compression/decompression algorithm degrades gracefully for transmission rates that can be achieved with current infrastructure. It is expected that the target bit-rates for High Dynamic Range video would be of same order as for current Standard Dynamic Range video. Typically bitrates around 5-10 Mbps for HD signal using H.264/AVC, or around 10-15 Mbps for UHD signal using H.265/HEVC, should be targeted; otherwise an upgrade to the current distribution infrastructures would have to take place (this will in any case be most probably required for the transition from HD to UHD).

Cinema and Home Viewing

The DCP is delivered on hard-drive to cinemas, who then load the data onto their own computers. As all the mastering is performed in an environment that matches the cinema itself, good control over the final visual quality is maintained.

It would be desirable to extend the peak luminance displayed in cinemas beyond the current 48 cd/m^2 so that a wide audience can sample high dynamic range movies before eventually upgrading their equipment in the home. As the cinema environment is dark, it would not be necessary to increase peak luminance levels as dramatically as would be necessary in uncontrolled home viewing environments. A ten-fold increase would in all likelihood be more than sufficient. However, achieving such an increase is not a trivial matter. At the moment we can only speculate as to what technology would be suitable for high dynamic range cinema.

One could, for instance, think about replacing the projection screen with an active display such as used in advertising (for instance in stadiums and at large concerts). Combined with projection technology, this could create an additive increase in luminance. However, such a solution is problematic as speaker system tends to be located behind the screen. An active display would therefore affect sound quality.

Alternatively, it may be possible to update the cinema projectors themselves. The trend toward laser-based projection may yield both wider gamut as well as higher luminance levels. It may also be possible to consider multiple displays, insofar space in the projection room allows.

Display technology for home viewing is advancing rapidly. Many large OLED displays are being demonstrated, and these have both good black levels and peak luminance. LED-based displays of up to 1000 cd/m^2 peak luminance are coming to the market as well. The emerging Quantum-dot technology is also arriving in consumer displays, with potentially improved performance compared to LED or OLED technologies. This raises the question as to how content should be created for such devices. The proliferation of these displays means that the viewing environment in the post-production studio is increasingly less representative of the home viewing environment. Traditionally, there is no control over the illumination in the home, but now colorists and directors have less control as to how their material is displayed. Conveying the director's intent is therefore becoming more difficult as more high-end displays come onto the market. As a result, there is a need to apply some form of display-specific regrading somewhere along the chain of processing. In the following section, we discuss several scenarios to achieve this.

STUDIO-SIDE OR DISPLAY-SIDE REGRADING?

As it is not practically feasible to color grade a piece of footage for every display device, a solution is required to bring content to a diverse range of display devices without significantly increasing the workload of the colorist. In addition, it may not be possible to always ensure that the equipment used for color grading is more capable than what is available in the home. This pertains to both dynamic range and color gamut (although color gamut mismatches could be minimized by using a standard wide color gamut, such as BT.2020). A reasonable approach is to target basically two gradings, one for SDR displays and another one for a reference standardized HDR display (the standardization of the HDR video signal format is a critical issue that is being addressed in standardization bodies such as ITU-R or SMPTE). The adaptation of the video signal to the display capabilities is anyway required. Note that it therefore becomes important to provide the display with metadata related to the mastering display properties (if no reference HDR mastering format is specified) and to the HDR content (e.g. peak luminance, color volume). Metadata related to the conversion from one format to another (e.g. SRD-to-HDR or HDR-to-SDR, Rec.709-to-Rec.2020 or Rec.2020-to-Rec.709), with respect to the artistic intent, would also be needed.

Figure 2. An example of inverse tone reproduction. The input is shown to the left, and simulated output is shown on the right.

In the following we assume that color grading has been applied to source material using a display that has a lower peak luminance than a hypothetical consumer display. In this case, it would be desirable to apply an automatic regrading step to take advantage of the display's capabilities. At the same time, the regrading should respect the director's intent. Such regrading is also referred to as inverse tone mapping. A (simulated) example of our in-house solution is shown in Figure 2. There are at least two places where this procedure could be applied, namely in the post-production studio and in a consumer device.

Inverse tone mapping, if applied in the post-production studio, has the advantage that the colorist can exert control over the process, manually adjusting the algorithm's parameters where necessary, or even applying a trim pass afterwards. This approach is practical if the number of different target environments/displays remains limited. Of course, the director and colorist have full control over the final output. The inverse tone mapping algorithm will have to be fast and robust, but it does not have to operate at real-time frame-rates. This means that the algorithm can be relatively advanced.

If applied by a consumer device, the restriction on processing speed is absolute: inverse tone mapping must achieved at guaranteed frame rates. In this scenario the algorithm also needs to be exceptionally robust, as there is no opportunity for user intervention to suppress possible artefacts. The usage of previously mentioned metadata would be needed to help the display for a proper signal adaptation. The processing power available depends on the type of consumer device in which the algorithm is embedded. Consumer displays would be the most natural type of device to host inverse tone mapping. However, their on-board processing capabilities are exceptionally limited in terms of both speed of processing and availability of memory. To our knowledge no consumer display manufacturer uses internal data paths of more than 10 bits. This number may have to increase in the future.

Set-top boxes tend to have somewhat more processing power, but then high dynamic range data would have to be communicated between set-top box and consumer display. Standards for this are currently lacking. Finally, in a streaming scenario, a PC could host the inverse tone mapping algorithm. Communication with the final display device would still

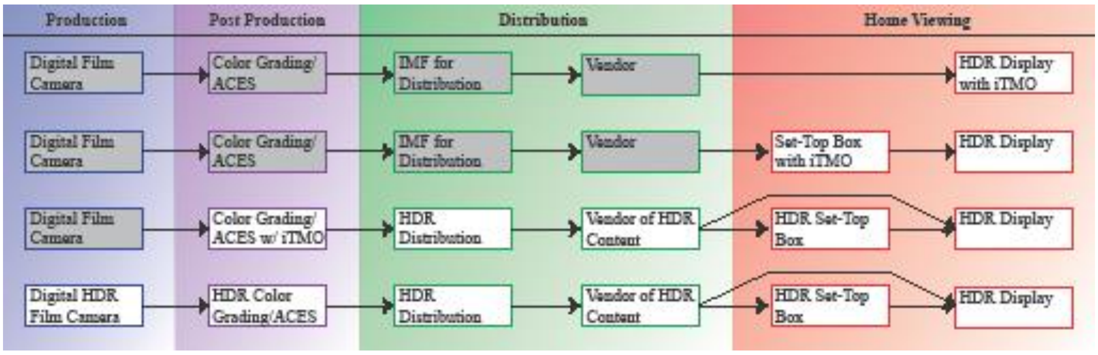

Figure 3. Four possible upgrade paths to create an HDR viewing experience in the home.

require standardization, however. Consumer-side processing does by-pass the need to transmit high dynamic range footage from source to consumer, which means that this would be a good way to serve highly capable consumer displays in the interim while the remainder of the video pipeline is being updated.

AN UPGRADE PATH TO FULL HDR VIDEO PIPELINES

This brings us to the question of what is stopping us today from taking high dynamic range footage from its source to the customer. The answer depends to some extent on which scenario is adopted. In order of the amount of effort required to implement, we think that the following sequence of four scenarios provide a reasonable path toward re-engineering video processing to accommodate high dynamic range footage (see also Figure 3):

1. Augment the current processing pipeline with consumer-side inverse tone mapping, embedded in consumer displays. This means that none of the capture, post-production or distribution blocks are affected and that display manufacturers can leverage the capabilities of their high dynamic range displays. The success of this approach depends strongly on the quality of the inverse tone mapping algorithm applied, as well as the bit-depth of the circuitry driving the panel itself.
2. Set-top boxes and PC's could also be used to create high dynamic range footage from conventional content, alleviating processing and memory limitations, but then communication with consumer displays will have to be standardized, for instance in a revision of the HDMI standard.
3. Standard content to be inverse tone mapped in the post-production studio. This requires a high dynamic range reference display, as well as a re-engineering and subsequent standardization of the distribution and transmission channels. While much effort is currently expended on designing high dynamic range compression solutions, which may be adopted in a future revision of the MPEG and ITU-T series of standards, less effort appears to be going into the design of a high dynamic range capable distribution solution.
4. The final step is to introduce a full high dynamic range workflow that incorporates the capture of full HDR data. While current film cameras are becoming increasingly capable, especially in terms of spatial resolution, the range of light that is effectively captured could (and should) be improved further. Camera manufacturers claim a dynamic range of up to 16.5 f-stops for their cameras, but discounting noise leads to an actual dynamic range that is several stops less than the claimed figures. We

do note that in terms of data paths, the ACES standard has in effect already paved the way for full high dynamic range and wide color gamut processing and archiving of film footage.

The first two solutions are relatively straightforward and would have a minor or even possibly no impact on the current distribution infrastructure, but would not at all guarantee the respect of the artistic intent of the color grading process, leading to significant variability in the end-user experience. The fourth solution is scientifically speaking the most relevant, but has impact of the full workflow, and would still require several years to be practically deployed. The solution 3 is an intermediate step that anyway may have impacts on the distribution infrastructures.

CONCLUSIONS

In this paper we have discussed the process of video acquisition, post-production, distribution and consumption, showing which stages currently form bottlenecks in the adoption of HDR technologies. We have identified a fundamental problem in that consumer displays are becoming increasingly varied, making color grading for home viewing increasingly less precise. New adaptive solutions are required, enabling to distribute and render properly a content graded on a reference master display on various rendering devices. These solutions involve the generation and specification of adapted metadata (characterizing the HDR content) in the production/post-production, the development of adapted HDR video format signals and of coding solutions (e.g. HEVC/H.265 extensions) for the distribution, and the development of content adaptation algorithms exploiting the metadata to adapt the content to the rendering capabilities. Tone management algorithms, including inverse tone reproduction, are able to mitigate these problems to some extent. Such algorithms could be inserted into various parts of the pipeline, leading to a sequence of four video chains that together form a reasonably smooth upgrade path with the final aim of arriving at a fully HDR compliant system. We do foresee that inverse tone reproduction will continue to play an important role in color grading for reasons of convenience and cost, even if HDR film cameras have become available.

The deployment of HDR video content will require to coordinate and synchronize the standardization efforts for the different components of the workflow. One of the challenges is to specify solutions that should have an unavoidable but as limited as possible impact on the production, post-production, distribution, and rendering sides, while being future-proof enough to support the technological evolutions of the rendering devices.

REFERENCES

[1] www.dolby.com/us/en/professional/hardware/video-monitors/prm-4220-professional-reference-monitor.html

[2] www.arri.com/alexa

[3] www.red.com/products/red-one

[4] www.pro.sony.eu/pro/lang/en/eu/product/broadcast-products-camcorders-digital-motion-picture-camera/f65/overview/

[5] Science and Technology Council, Academy of Motion Picture Arts and Sciences, “The Academy Color Encoding System”, 2013.

[6] http://www.sim2.com/HDR/hdrdisplay/hdr47e_s_4k

[7] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. A. Whitehead, M. Trentacoste, and A. Vorozcovs, “High dynamic range display systems”. ACM Transactions on Graphics, 23(3), 760, 2004.

[8] Eilertsen, R. Wanat, R. Mantiuk and J. Unger. “Evaluation of Tone Mapping Operators for HDR-Video”. Computer Graphics Forum, 32, 7, 2013.

[9] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward and K. Myszkowski, “High Dynamic Range Imaging, Second Edition: Acquisition, Display, and Image-Based Lighting”, Morgan-Kaufman, 2010.

[10] F. Banterle, A. Artusi, T. Aydin, P. Didyk, E. Eisemann, D. Gutierrez, R. Mantiuk and K. Myszkowski, “Multidimensional Image Retargeting”, ACM Siggraph ASIA 2011 Courses - December 2011.

[11] A. O. Akyüz, R. Fleming, B. Riecke, E. Reinhard, H. Bülthoff, “Do HDR displays support LDR content? A psychophysical evaluation”, ACM Transactions on Graphics, 26, 3, 2007.

[12] T. Pouli, A. Artusi, F. Banterle, E. Reinhard, A. O. Akyüz and H. P. Seidel, “Color Correction for Tone Reproduction”, Proceedings of the 21st IS&T Color Imaging Conference, 2013.

[13] C. Christopoulos, “The JPEG2000 still image coding system: an overview”, IEEE Trans on Consumer Electronics, Vol.46, No.4, pp 1103-1127, Nov.2000.

[14] E. François, S. Lasserre, F. Le Léannec and Y. Chen, “Use cases and requirements for HDR and WCG video coding”, ISO/IEC JTC1/SC29/WG11 M31956, Oct 2014, San Jose, USA.

[15] T. Wiegand, G. Sullivan, G. Bjontegaard and A. Luthra, “Overview of the H.264/AVC video coding standard”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, N. 7, pp 560-576, July 2003.

[16] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, Y.-K. Wang and T. Wiegand, “High efficiency video coding (HEVC) text specification draft 10 (for FDIS & Consent)”, ISO/IEC JTC1/SC29/WG11 JCTVC-L1003, Jan 2013.

[17] S. Miller, S. Daly, and M. Nezamabadi, “Perceptual signal coding for more efficient usage of bit codes”, 2013 May/June issue of SMPTE Motion Imaging Journal.

[18] D. Touze, Y. Olivier, D. Thoreau and C. Serre, “High dynamic range video distribution using existing video codecs”, Proc. PCS 2013, pp 349-352, San Jose, USA, Dec 2013.

ACKNOWLEDGEMENTS

The authors want to thank the French ‘Images and Réseaux’ cluster and the NEVEx project.

The authors want to thank P. Andrivon, S. Lasserre, F. Le Léannec, Y.Olivier, D. Thoreau, D.Touzé from Technicolor.