Fifty Shades of HDR

Alan Chalmers   Brian Karr   Rossella Suma   Kurt Debattista
WMG, University of Warwick, UK
alan.chalmers@warwick.ac.uk

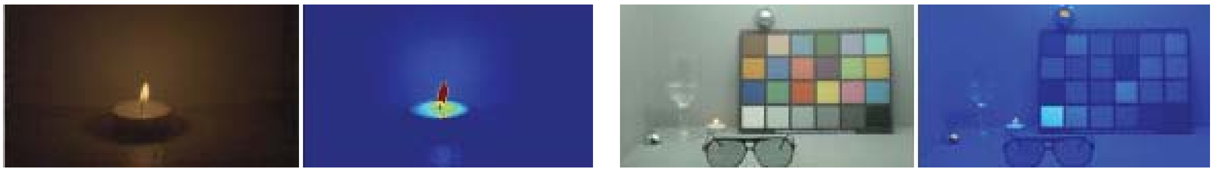


Fig. 1.   A single candle lit in dark a room has a dynamic range 16.8 million to1, while items evenly lit by sunlight only have a dynamic range of 256 to 1 (Image courtesy of Jassim Happa).

***Abstract*— From relatively unknown, just 5 years ago, High Dynamic Range (HDR) video is now having a major impact on most aspects of imaging. Although one of the five components of the specification for UHDTV, ITU-R Recommendation BT.2020 in 2012, it is only when it became apparent that HDR could help accelerate the slow penetration of 4K into the TV and home-cinema market, that HDR suddenly started to gain significant attention. But what exactly is HDR? Dynamic range is defined as the difference between the largest and smallest useable signal. In photography this has meant the luminance range of the scene being photographed. However, as HDR grows as a "marketing tool" this definition is becoming less "black & white".**

This paper considers the different ways in which the term HDR is now being exploited; the challenges of achieving a complete efficient HDR pipeline from capture to display for a variety of applications; and, what could be done to help ensure HDR algorithms are future proof as HDR technology rapidly improves.

Keywords—HDR; HDR video; UHDTV, compression

I. Introduction

UHDTV, the specification of Ultra High Definition (UHD) for the broadcasting sector, is seen as the next major step in the future of television. Defined in ITU-R Recommendation BT.2020 [1], UHDTV attempts to overcome the limitations of current TVs by introducing higher spatial resolution (4K and 8K) and framerates, wider colour gamuts, and increased dynamic range. The widespread consumer uptake of UHDTV is dependent on the perceived added value when compared with current HD formats. Many user studies, for example those by the 4Ever project [2], have clearly shown that of all enhancements introduced by UHDTV, dynamic range is the feature that users most notice.

The dynamic range of a scene is the ratio of the maximum light intensity to the minimum light intensity [3]. As Fig. 1 shows, a candle in dark room has a much higher dynamic range than a scene which is evenly lit with sun light. Agreed in September 2013 by EU COST Action IC1005 on HDR [4] the following definitions of Dynamic Range were subsequently adopted by the MPEG adhoc committee on HDR in their Call for Evidence document [5]. The term f-stop refers to the following contrast ratios:

X f-stops = difference of $2^X = 2^X : 1$

So

16 f-stops = difference of 2^{16} = 65,536 : 1
This is normally noted as 100,000:1; approximately what the eye can see in a scene with no adaptation.

20 f-stops = difference of 2^{20} = 1,048,576 : 1
This is normally noted as 1,000,000 : 1; approximately what the eye can see in a scene with minimal (no noticeable) adaptation.

From this the following were defined:

Standard Dynamic Range (SDR) is ≤ 10 f-stops (aka Low Dynamic Rage (LDR))

Enhanced Dynamic Range (EDR) is > 10 f-stops and ≤ 16 f-stops

High Dynamic Range (HDR) is > 16 f-stops

This is, unfortunately, by no means the only definition of HDR.

The UHD Alliance [6] was formed by a consortium of TV manufacturers, broadcasters and content producers, to present their view of UHD. They have proposed Ultra HD Premium; the criteria by which future platforms can be judged by them to be suitable for delivering a "premium 4K experience". UHD Alliance has proposed two definitions of HDR:

1. 1,000nits peak brightness and < 0.05nits black level (contrast ratio 20,000:1, 14.3 f-stops),

or

2. > 540nits brightness and < 0.0005nits black level (assuming it is possible to measure 0.0005nits, this give a contrast ratio 1,080,000:1, ≈20 f-stops)

This is to satisfy a range of TV manufacturers. Option 1 is for those with LED based TVs which have higher brightness but inferior black levels, while Option 2 is for OLED based TVs which have deep blacks but much lower peak brightness levels [7]. In the real world, the peak brightness of the sun at midday is approximately 1.6×10^9nits.

II. HDR- THE POTENTIAL

HDR imaging offers the potential of capturing the full range of lighting in a scene and delivering it in a digital format along a pipeline to a display. Known as "scene referred", this full range of lighting can be kept, while maintaining physical accuracy, if 32 bit IEEE floating point values are used to represent each colour channel. This means that 96 bits per pixel (bpp) are needed compared with a standard image of just 24bpp [8]. A single HDR frame of uncompressed 4K UHD resolution (3840×2160 pixels) requires approximately 94.92Mbytes of storage, and a minute of data at 30fps needs 166Gbytes. This is currently prohibitive on existing ICT infrastructure. Efficient data formats and compression techniques are thus essential if HDR video is to be widely adopted.

While linear encoding is desired to ensure all the data in a scene is preserved, this is particularly inefficient when providing HDR video for human consumption. The Human Visual System (HVS) is able to discern luminance threshold differences more clearly in darker areas than brighter areas. Described by Poynton as the "code 100" problem [9], this means, for 8 bit linear encoding, pixel values below code 100, have increasingly perceivable luminance differences, resulting in visible artefacts such as contouring or banding. On the other hand, for pixel values above 100, codes eventually fall beyond the 1% luminance difference threshold of the HVS resulting in unnecessary codes. Increasing the number of bits, for example from 8 to 16 bits increases the overall contrast ratio, but does not solve the "code 100" problem [9]. Minimising the visible artefacts and wasted codes may be achieved by exploiting the near logarithmic response of the HVS. This uses codes spaced at the ratio of 1.01 (or smaller depending on desired accuracy); the 1% perceivable luminance step.

Several pixel formats have been proposed to represent HDR pixel values, including shared exponent RGBE encoding used in the Radiance image format [10] and logLuv encoding used in a custom HDR-version of the TIFF images [11]. The current widely used file format for individual HDR frames is OpenEXR, developed within the movie industry. OpenEXR has been released as free software [12], however, it is not suitable for HDR video as each frame at HDR resolution typically requires 8MBytes. The ACES system uses OpenEXR to handle HDR images [13].

III. HDR – THE REALITY

Although HDR has the potential to capture, transmit and display the full range of light in a scene, the currently available HDR video technology is not yet able to live up to this promise.

A. Capture

Well known techniques of HDR capture use multiple exposures with different exposure times to create a static HDR image [14]. Systems that use such a multiple exposure approach to capture HDR video of dynamic scenes, such as [15], require sophisticated, real-time de-ghosting algorithms to minimise any artefacts that arise from objects moving while the multiple exposures for a single frame are captured.

To minimise artefacts and still capture a large dynamic range, dedicated HDR video systems use, for example, multiple

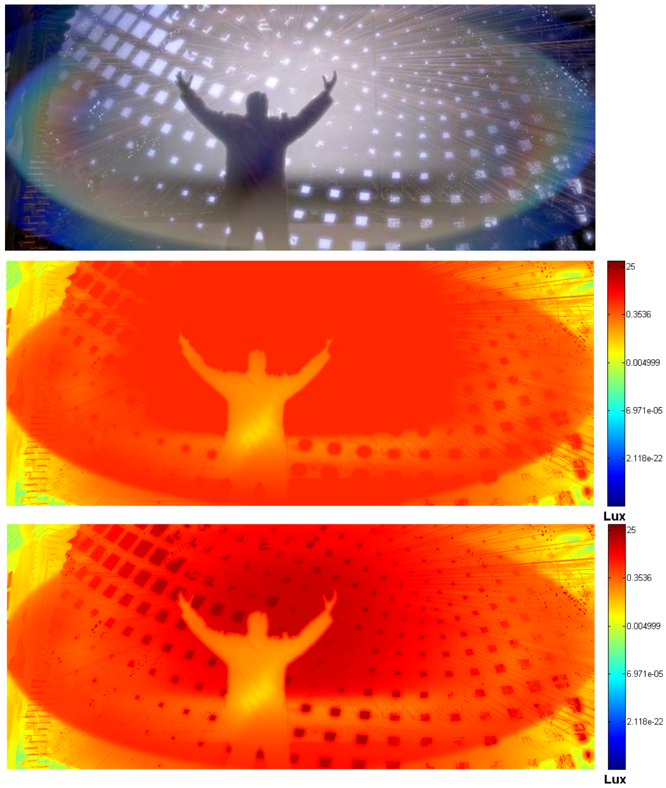


Fig. 2. Frame from Tears of Steel (top) Tone mapped image of full HDR data (middle) false colour of 10 bit content (bottom) false colour of full HDR

sensors with the same integration time through a single lens [16, 17]. The sensors do need different sensitivities to cover the whole dynamic range of the scene which can be created by analogue-gain and/or neutral density-filters. Such an approach has the advantage that artefacts from moving objects are avoided and correct motion blur is achieved due to identical integration-time. However, matching the images from the sensors is difficult to achieve due to restrictions on sensor alignment precision. Furthermore, building cameras with multiple sensors having a larger format is complex and the result could be large and heavy.

While research prototype HDR video cameras, such as [16], have been able to achieve 20 f-stops at 30fps, commercial HDR video cameras can currently only capture a much more modest dynamic range. A recent survey of the dynamic range of 18 commercial cameras for possible use in imaging rocket launches [18], concluded that commercial systems were only able to capture < 18 f-stops, and thus not yet suitable to adequately image a rocket launch (> 20 f-stops). Table I shows the dynamic range (measured and vendor claimed) of some commercial cameras [18].

TABLE I. DYNAMIC RANGE OF DIFFERENT CAMERAS [18]

Camera	**Stops @ 0.5 RMS Noise**	**Vendor Spec.**
RED DRAGON – HDR x6, Log Film, Total	11.8 + ~3	16.5+
ARRI Alexa – Log	13.9	14.0
Black Magic Cinema 4K - Film	9.0	12.0

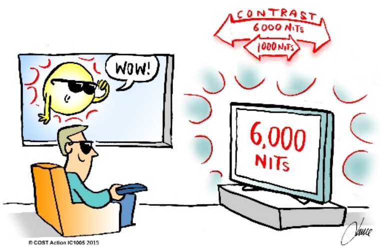
B. *Compress*

HDR video compression may be classified as *one-stream* or *two-stream* [19]. The one-stream approach utilises a single layer transfer function to map the HDR content to a fixed number of bits, typically 10 or 12. Examples of one-stream methods include [20, 21, 22, 23, 24]. While it is true that many scenes do contain a dynamic range of lighting that is sufficiently low to be adequately contained within a lower number of bits, there are many others for which 10 or even 12 bits are not enough [19]. Fig. 2 shows a frame from the HDR film Tears of Steel. As the false colour image (to represent the dynamic range in the image) in the middle shows, significant detail can be lost if, in this case, only 10 bits are used to hold the HDR content.

In two-stream compression methods there is one input HDR video stream to the encoder which produces two bit streams as output. These streams can consist of (1) a standard compliant bit stream, for example HEVC Main 10, H.264 etc. and (2) one other stream corresponding to additional data to reconstruct the HDR video. At the decoder, these two streams are recombined to produce the HDR video stream. Examples of two-stream compression for HDR video include [25, 26, 27, 28].

C. *Display*

The UHD Alliance's "displayed referred" definition of HDR is encouraging a number TV manufacturers to market HDR displays. However, the 1,000nit peak luminance of these displays are significantly lower than the HDR displays currently available from the Italian company, SIM2. Derived from the pioneering HDR displays developed by Brightside [29], SIM2's latest HDR47 display has a peak luminance of 6,000nits. This is achieved via a backlight of 2202 LEDs each of which has 12 bits of brightness resolution. A key feature of the SIM2 display is the attention to high power management and heat cooling. This allows the 6,000nit display to have a 1.5kW power supply limit with a standard 110V plug.



IV. HDR – THE APPLICATIONS

HDR video is having a major impact on all forms of imaging, from broadcast and entertainment to industrial applications, such as security or imaging laser welding, Fig. 3. One major challenge, however, is the different needs of these applications. While applications intended for human viewing can exploit the HVS for efficient compression of the large data requirements of HDR, this can have undesirable consequences when considering computer vision applications.

A. *Computer vision applications*

Object tracking in a video stream is a key part of many applications, including automotive, robot vision and surveillance. Such object tracking requires a number of tasks: object representation (how to construct robust object descriptors using different types of visual features), object recognition (build effective mathematical models for object identification using statistical learning techniques) and object tracking (dynamic state estimation problem) [30]. A key part of object detection and subsequent tracking is Feature Point (FP) detection. Low level FPs in an image include points, edges or blobs [31]. The extraction of these low level FPs is strongly dependent on the illumination within a scene when it is captured. In extreme lighting, under- or over-exposed pixels from LDR video will cause FP detectors to fail. HDR video, of course, avoids this problem. However, popular FP detection algorithms, such as FAST [32], SURF [33] and SIFT [34] have been developed for LDR data.

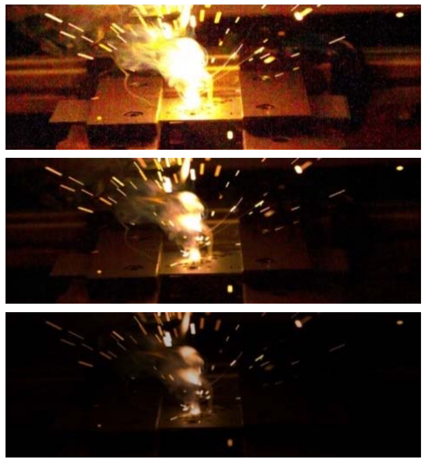


Fig. 3. Three exposures from HDR imaging of robot laser welding

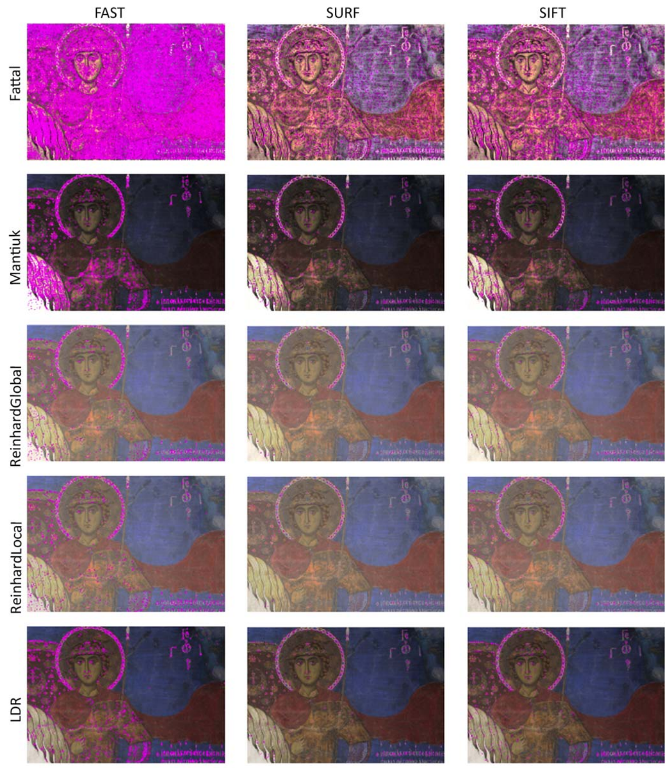


Fig. 4. FP detection in an LDR and tone mapped HDR images [39]

Although tone mapping the HDR video has been shown to enable these algorithms to be used directly [35], care needs to be taken in the choice of tone mapping operator (TMO). Most

TMOs have been developed for humans to view HDR content [8], therefore when processed with computer vision algorithms the results might be different from what is expected. As Fig. 4 shows, such perceptually based TMOs might give worse results than the LDR image (eg Reinhard [36]) because parts of the scene which are perceptually important are highlighted, or the TMO is simply poor at reproducing detail in bright or dark areas of the scene [37]. Furthermore, when a TMO is finely-tuned towards enhancing small contrast or details (eg Fattal [38]), even sensor noise can be detected as a feature [39].

B. *Perceptual compression*

For broadcast applications, there are two current standards for HDR video compression: SMPTE ST2084 [40] and ARIB STD-B67 [41]. Central to both of these one-stream methods are perceptually based transfer functions. In SMPTE ST2084 this is the PQ curve [23], while ARIB STD-B67 uses Hybrid Log Gamma (HLG) [24]. Both of these exploit knowledge of the HVS to compress the input HDR video stream into 10 bits. However, the transfer functions in both cases are relatively computationally complex (for example HLG contains a branch instruction) which may preclude their efficient implementation on time critical ICT infrastructure. More efficient transfer functions include the straightforward Power Transfer Function (PTF):

$$L = AV^{\gamma}$$

where: A is a constant, L and V are contained by the set $\mathbf{R} \in [0; 1]$, and $\gamma \in \mathbf{R}^{+}$. This has been recently shown, for a $\gamma=4$ to achieve better quality than either PQ or HLG and is capable of being decoded at over 380 fps. The method outperforms an analytic implementation of PQ by a factor of over 29.5 and a look-up implementation by a factor of nearly 1.5. Encoding performance outperforms PQ by a factor of 16.6 and is only slightly slower than a LUT [42].

V. PSEUDO HDR VS TRUE HDR

A recent investigation of a number of consumer HDR displays by HDTVtest [43] has highlighted a key limitation of these devices. In particular these consumer HDR displays do not provide an overall brighter image, but rather maintain the average picture level (APL) of the frames, and achieve an HDR effect by providing (a) more details in the darker regions of the scene and (b) having (a few) brighter highlights. The key implications of this are:

1. The HDR metadata means that the TV has to be driven at its maximum backlight capacity in order to provide the required highlights. However, this means that the overall brightness of the screen can't simply be increased as you can with an SDR display.
2. HDR content on these consumer HDR displays should not be viewed in ambient lighting levels exceeding 5nits [44]. The situation is even worse for HDR displays based on OLED technology. For such displays, darker than 5nits viewing conditions and even the absence of white walls [44] is recommended.

One of the contributing factors to this problem is the lack of flexibility in the ST.2084 EOTF to adapt to different ambient lighting conditions because every input signal value is mapped directly to the same output luminance level on all HDR displays.

True HDR was defined by EU COST Action IC1005 as an approach that ensures the full range of HDR data is preserved throughout the entire pipeline from capture to display [4]. This implies tone mapping does not occur anywhere in the pipeline prior to the display. True HDR provides great flexibility in how the HDR content can be handled at the display. In addition to enabling the direct display of the HDR content to an HDR display, on an LDR display, True HDR provides the ability to dynamically alter the tone mapping to suit the current scene, creative intent and ambient light conditions [19]. Furthermore, the desired exposure value of every pixel – if required – can be chosen. So, while watching a football match, a favourite player can be selected and then this object tracked and its pixels modulated allowing the player to always be clearly seen no matter what the lighting conditions of the scene are. Alternatively, an automatic "ball-mode" can be chosen, enabling the ball up against the sky to be visible (at one exposure) while the sky is in displayed in another exposure. As Fig. 5 shows, because the pixels around the goal can be modulated, it is possible to set them at a different exposure, and thus ensure the ball is visible as it enters the goal.

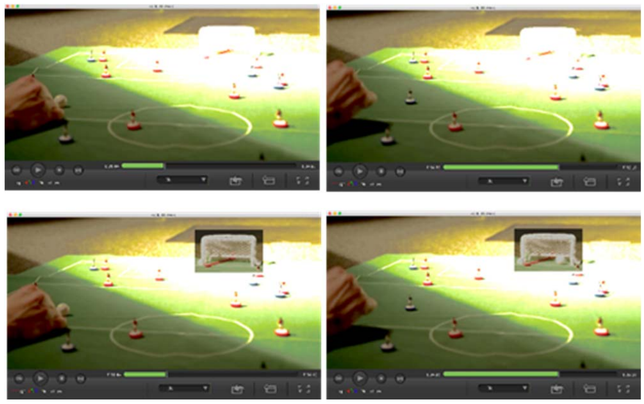


Fig. 5. true-HDR (a) Ball is visible as it is struck, but (b) Goal cannot be seen due to over-exposed area (c) Pixels around goal modulated to allow visibility (d) Goal!!! [19]

The ability to "personalise" pixels can also provide content creators additional creativity by enabling them to deliberately hide detail, such as clues, in areas of the scene which can only be discovered by the user interacting with the content by exploring the exposures of those pixels, Fig. 6.

Fig. 6. **Personalised pixels:** The hidden character can be seen in the image on the right by exploring the pixels of the HDR image.

VI. Conclusions

HDR video has exploded in the last 3 years from relative obscurity to become widely popular and "the main reason why you would want to buy a new television". In the rush to provide some form of standard to facilitate the widespread uptake of new consumer HDR displays, the industry is adopting ST.2084 [40] and ARIB STD B-67 [41], despite clear evidence that there are more efficient ways of encoding HDR video, and the fact that sooner or later someone is going to want to watch HDR video in a room which is brighter than 5nits. Furthermore, these two methods are for 10 bit infrastructures, which precludes their use with the large numbers of 8 bit devices, including legacy displays, and most importantly, mobile devices. The growth of online video request from mobile devices has risen from a modest 6% of all requests in 2002 to over 50% by the end of 2015 [45]. Indeed more than 51% of traffic on mobile devices is video viewing [46]. Viewing HDR video on mobile devices does, of course, introduce additional complications, including dynamically varying ambient lighting conditions, reflections on the screen [47] etc.

If HDR video is to provide a significantly enhanced viewing experience, not only for a few premier devices in a dark room, then two-stream HDR video compression should be reconsidered. Such approaches are able to work equally well on 8 bit as well higher bit infrastructures and can provide backwards compatibility in a straightforward manner, [25, 28].

Furthermore, to harness the potential of HDR, not just for current displays, but for the future, it is important to maintain the full range of HDR video throughout the entire pipeline from capture to display. This will enable, not only this content to be played directly onto HDR displays of the future, but also to achieve the best viewing experience now, by taking into account *at the display*, any creative intent, the current display capabilities, and importantly, the actual viewing conditions.

It is clear that HDR can provide a step-change in viewing experience if done properly. It would be a pity if it simply became "yet another abused marketing term".

References

[1] ITU-R BT.2246-2 "The present state of ultra-high definition television", http://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-BT.2246-2-2012-PDF-E.pdf, 2012

[2] http://www.4ever-project.com/home/

[3] Boitard R., Pourazad M.T., Nasiopoulos P. Slevinsky J. "Demystifying High-Dynamic-Range Technology: A new evolution in digital media." IE

[4] http://www.cost.eu/COST_Actions/ict/IC1005

[5] Luthra A., François E., Husak W. (Editors), "Requirements and Use Cases for HDR and WCG Content Distribution", ISO/IEC JTC 1/SC 29/WG 11 (MPEG) Doc. N15084

[6] www.uhdalliance.org

[7] http://www.trustedreviews.com/opinions/ultra-hd-premium

[8] Banterle F., Artusi A., DebattistaK., Chalmers A. *Advanced High Dynamic Range Imaging*, AKPeters, ISBN: 978-156881-719-4, March 2011.

[9] Poynton C.A. Digital Video and HDTV: Algorithms and Interfaces, Morgan Kaufmann Publishers, 2003.

[10] G. Ward "Real Pixels", Graphics Gems II, pp 80-83, Academic Press, 1991.

[11] Ward Larson G. "Logluv encoding for full-gamut, high-dynamic range images". Journal of Graphics Tools, 3(1):15–31, 1998.

[12] http://www.openexr.com/

[13] http://www.oscars.org/science-technology/sci-tech-projects/aces

[14] Debevec, P. E. and Malik, J. "Recovering high dynamic range radiance maps from photographs" SIGGRAPH '97, ACM Press1997.

[15] McNamee J., Hatchett J., Debattista K., Chalmers A.G. Live HDR Video Streaming on Commodity Hardware, SPIE 2015, August 2015.

[16] Chalmers, A., Bonnet, G., Banterle, F., Dubla, P., Debattista, K., Artusi, A., and Moir, C. "High Dynamic-range video solution", ACM SIGGRAPH ASIA 2009 Art Gallery & Emerging Technologies, ACM, 2009.

[17] Tocci, M. D., Kiser, C., Tocci, N., and Sen, P. "A versatile HDR video production system", ACM Transactions on Graphics, 30(4), ACM, 2011.

[18] Karr, B., Chalmers, A.G., and Debattista, K. "High dynamic range digital imaging of spacecraft", In High Dynamic Range - From Acquisition to Display and Applications, chapter 20, Dufaux, F., Le Callet, P., Mantiuk, R., and Mrak, M. (eds), Elsevier, 2016.

[19] Chalmers A., McNamee J., Hatchett J., Mukherjee R., Olaizola I., Debattista K. "12 bits is simply not enough for HDR video!", BEC 2015, NAB, 2015.

[20] Mantiuk, R., Krawczyk, G., Myszkowski, K., and Seidel, H.-P. "Perception-motivated high dynamic range video encoding", ACM Trans. Graph. 23(3), 733-741, 2004.

[21] Motra, A. and Thoma, H. "An adaptive logluv transform for high dynamic range video compression", 17th IEEE International Conference on Image Processing (ICIP), IEEE, 2010.

[22] Zhang, Y., Reinhard, E., and Bull, D. "Perception-based high dynamic range video compression with optimal bit-depth transformation", 18th IEEE International Conference on Image Processing (ICIP), IEEE, 2011.

[23] Miller, S., Nezamabadi, M., and Daly, S. "Perceptual signal coding for more efficient usage of bit codes", SMPTE Conference, 2012.

[24] Borer, T. "Non-linear opto-electrical transfer functions for high dynamic range television", BBC White Paper, 2014.

[25] Mantiuk, R., Efremov, A., Myszkowski, K., and Seidel, H.-P. "Backward compatible high dynamic range MPEG video compression", ACM Transactions on Graphics, 25(3), 713-723, 2006.

[26] Lee, C. and Kim, C.-S. "Rate-distortion optimized compression of high dynamic range videos", 16th European Signal Processing Conference, 2008.

[27] "HDR video data compression devices and methods" EP Patent 2144444, 2009.

[28] Debattista K., Bashford-Rogers T., Selmanovic E., Mukharjee R., Chalmers A.G. "Optimal Exposure Compression for High Dynamic Range Content." The Visual Computer, 31 (6-8) 1089-1099, 2015.

[29] Seetzen H., Heidrich W., Stuerzlinger W., Ward G., Whitehead L., Trentacoste M., Ghosh A., Vorozcovs A. "High dynamic range display systems." ACM Transaction on Graphics 23:3, 2004.

[30] Kobayashi, T.; Otsu, N. "Color image feature extraction using color index local auto-correlations", IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP), 2009.

[31] Schmid C., Mohr R., Bauckhage C. "Evaluation of interest point detectors", International Journal of computer vision 37 (2), 151–172, 2000.

[32] Rosten E., Drummond T. "Machine learning for high-speed corner detection". In Computer Vision–ECCV 2006, Springer Berlin Heidelberg, pp. 430-443, 2006.

[33] Bay H., Tuytelaars T., Van Gool L. "Surf: Speeded up robust features". Computer Vision and Image Understanding, 110(3), pp. 346–359, 2008.

[34] Lowe D.G. "Object recognition from local scale-invariant features". In Computer vision. The proceedings of the seventh IEEE international conference on, 2, pp. 1150–1157, 1999.

[35] Přibyl B., Chalmers A.G., Zemčík P., Hooberman L., Čadík M. "Evaluation of feature point detection in high dynamic range imagery". Journal of Visual Communication and Image Representation, 2016.

[36] Reinhard E., Stark M., Shirley P., Ferwerda J. "Photographic tone reproduction for digital images". In ACM Transactions on Graphics (TOG), 21(3), ACM, pp. 267-276, 2002.

[37] Ledda P., Chalmers A.G., Troscianko T., Seetzen H. "Evaluation of tone mapping operators using a high dynamic range display". In ACM SIGGRAPH'05, 24(3), pp. 640-648, 2005.

[38] Fattal R., Lischinski D., Werman M. "Gradient domain high dynamic range compression. In ACM Transactions on Graphics (TOG), 21(3), ACM, pp. 249-256, 2002.

[39] Suma R., Stavropoulou G., Stathopoulou, E., van Gool L., Georgopoulos A., Chalmers A.G. "Evaluation of the effectiveness of HDR tone mapping operators for photogrammetics applications" Arqueologica 2.0, 2016.

[40] ST 2084:2014 : High Dynamic Range Electro-Optical Transfer Function of Mastering Refference Displays. Society of Motion Picture and Television Engineers (SMPTE), August 2014.

[41] Association of Radio Industries and Businesses (ARIB). STD-B67: Essential parameter values for the extended image dynamic range television (EIDRTV) system for programme production, July 2015

[42] Hatchett J., Debattista K., Mukherjee R., Bashford Rogers T., Chalmers A.G. "Highly efficient HDR video compression", ISO/IEC JTC1/SC29/WG11 MPEG2016/JCTVC-W0072, Feb 2016.

[43] http://www.hdtvtest.co.uk/news/4k-vs- 201604104279.htm

[44] Schulte T., Barsotti J. HDR Demystified: Emerging UHDTV systems", Technical paper, SpecraCal Inc, 2016.

[45] Ooyala: Global video index q3 2014. Report, International Telecommunication Union, Silicon Valley, USA, 2014.

[46] CISCO: Cisco visual networking index: Global mobile data traffic forecast update, 2012-2017. http://www.cisco.com/c/en/us/solutions/collateral/serv\ice-provider/visual-networking-index-vni/white_paper_c11-520862.html, 2013.

[47] Melo M., Bessa M., Barbosa L., Debattista K., Chalmers A.G. Screen Reflections Impact on HDR Video Tone-mapping for Mobile Devices: An Evaluation Study. EURASIP Journal on Image and Video Processing. Accepted for publication November 2015.