PRODUCTION OF HIGH DYNAMIC RANGE VIDEO

Marc Price[1], David Bull[2], Terry Flaxton[3], Stephen Hinde[2], Richard Salmon[1], Alia Sheikh[1], Graham Thomas[1], Aaron Zhang[2]

[1]BBC R&D, Wood Lane, London, W12 7SB, UK
[2]Bristol Vision Institute, University of Bristol, Bristol, BS8 1TH, UK
[3]University of the West of England, Bristol, BS16 1QY, UK

ABSTRACT

In this paper we discuss the production of high dynamic range video and assess how high dynamic range video influences audience experience. The paper describes our method of assessing audience experience, which involves measuring the degree of audience 'immersion'. This is followed by a discussion of the method and workflow we used for capturing a short high dynamic range movie for use in our ongoing assessments. Some promising preliminary results are noted.

INTRODUCTION

Over the past decade, the development of video technology has mainly been focused on increases in spatial resolution, with the standardisation of HDTV formats and the present work on standardising UHDTV formats. We have seen some progress in temporal resolution too, with up to 50Hz progressive in HDTV, and 120Hz in UHDTV.

Developments are now emerging along a third dimension; that of dynamic range (i.e. the word length of each pixel colour). Manufacturers are producing high-end cameras that are claimed to be able to capture digital video at up to 14 bits per pixel colour (1), 16 bits per pixel colour (2), and 18 bits per pixel colour (3). Furthermore, some high-end displays are capable of displaying high dynamic range (HDR, i.e. >12 bits per pixel colour) videos (4), technology which could potentially find its way into future consumer equipment.

Increasing the dynamic range presents content-makers with significant challenges, but arguably for significant gains – the human optical system has an instantaneous luminance sensitivity of ~4 orders of magnitude (Kunkel and Reinhard (5)). At the Bristol Immersive Technology Lab (BITL), a collaboration between the Bristol Vision Institute and BBC Research and Development, we have been running experiments to find out how much the audience experience may be improved by increasing the dynamic range. The experiments use a method we have developed for measuring audience 'immersion', which is discussed in the first section of this paper.

In order to run the experiments, we needed a short broadcast quality movie, shot and produced in HDR. To this end we have recently produced a digital HDR movie, which gave us the opportunity to explore the challenges of producing HDR video content and how they can be addressed. In the second and third sections of this paper, we give a short review of the general process of capturing high dynamic range images, followed by a discussion of how we applied this process to capture and produce our HDR movie.

MEASURING AUDIENCE EXPERIENCE OF HDR

Measurement of Audience Immersion

In this section, we discuss how an audience's level of immersion can be measured when watching HDR video. We assume that the audience's main motivation for watching video content is to maximize the experience of being transported into the space of the video (Bordwell and Thompson (6)). There has been a great deal of work on a construct relating to the user experience of transportation into video, called *presence*, defined as 'being there' in the mediated space (IJsselsteijn et al (7)), and also relates to enjoyment.

Initially presence research was targeted at teleworking applications (Draper et al (8)), and then later the study of virtual reality applications (Barfield & Weghorst (9)). Cinema and TV are considered to be the visual media that have the maximum ability to capture an audience into a mediated space (Anderson & Burns (10)). It is proposed therefore that the investigation of HDR video can benefit from this prior work on using presence as an instrument for measuring immersion.

The International Society for Presence Research's guidelines (2000) (ISPR (11)) split presence measurement into two categories: either subjective measurement methods or objective measurement methods. Many of the subjective presence measurement methods consist of questionnaires given after the experience of watching a film. However there are many methodological problems caused by the indirect nature of these offline, post-experience methods, which necessarily rely on long-term memory and give results that are often unstable across subjects, groups and time (Freeman et al (12)).

For these reasons a number of online subjective measurement methods have been developed. Techniques include: a spoken commentary, a hand-held potentiometer, and a pencil-and-paper line bisection method (Troscianko et al (13)). The advantage of the real-time subjective measurement techniques is that they give a direct subjective report of presence from a subject who is still in the experience.

The ISPR (2000) document (11) also refers to so-called direct objective measures of presence; these attempt to measure presence by recording physiological and/or behavioural responses during the experience e.g. ocular responses, EEG, skin conductance, heart-rate, blood pressure, muscle tension, respiration (7). However there appears to be no clear evidence that physiological measures correlate well with subjective reports of presence (Ellis (14)). Also, by definition presence is a perceptual construct rather than a physiological quantity, hence subjective measures would be the best direct measurement.

Another important point brought out by the ISPR (2000) (11) guidelines is that there is no standard way of measuring presence, and a comparison of presence measures across methodologies is consequently hazardous. Instruments and experimental conditions vary, as do stimuli settings and participant groups (for details see: Lassiter et al (15)). Hence, for evaluating audience immersion in video, a measurement method is required that: is consistent with related studies; gives a presence construct that is both reliable and valid; and is based on a real-time measurement method (7), (13).

Application of Presence Measure to HDR Video Content Evaluation

To evaluate the degree of improvement to audience experience afforded by HDR video, we propose a comparative study of HDR and standard HD video using the presence line bisection method as applied in (13). The proposed test will test two conditions, namely HDR and a HD baseline using the same video content at the two formats. We shall use ecologically valid conditions consistent with TV broadcast, in other words:

1. We shall use professionally produced, 1080p50 HDR video content, of roughly the length of a short TV programme, (e.g. 30 minutes) – the test tracks the presence measure at intervals through-out the duration of the programme.
2. The content will be played-out in a typical state-of-the art high-end home cinema environment matching, where possible, the living room environment specification in ITU-R BT500 (16).
3. Testing will be conducted on individual subjects and on small family-size audiences, (e.g. 2-5), consistent with a home-cinema scenario.

The hypothesis of this experiment is that there will be a significant improvement in the presence measure in the HDR experience as compared to the HD experience, if HDR is truly a more immersive format. As we make no assumption about the relationship between presence and subjective picture quality, we shall also run subjective picture quality assessments of clips from the HDR movie, in accordance with (16).

Although these experiments are still ongoing, preliminary picture quality assessment tests with expert viewers have shown that HDR can give a discernible improvement in picture quality over the standard HD baseline. Full results will be available later in the year.

HDR VIDEO CAPTURE

As the experiments described in the preceding section require a broadcast quality programme of HDR content, of at least 30 minutes duration, we now turn to address the problem of capturing and producing such a programme. We begin in this section with a discussion of the theory of capturing HDR video. In the first sub-section, we consider HDR imaging, and in the following sub-section, we extend that to capturing HDR video.

Concept of HDR Imaging

Most methods of capturing HDR images are based on the theory that the HDR image of a subject can be constructed from a set of two or more, low dynamic range (LDR) images of the subject.

This works by ensuring each LDR image in the set has a different exposure setting (i.e. the amount of light captured), so that the dynamic ranges of the set of LDR exposures together encompass the full dynamic range that is to be constructed. This assumes that the LDR images are of exactly the same subject, taken from the same viewpoint, and at the same instant in time – they ideally have to match in every respect other than relative photometric exposure, although in practice this can be difficult to achieve.

The algorithm commonly used for constructing the HDR image from the set of LDR images is discussed in detail in Reinhard et al (17). It consists of two stages: the first stage recovers the response function g of the image capture system (lens, sensor/film, etc.); and the second stage applies that response function to compute the HDR pixel values R_{ij} from the LDR pixel values L_{ijk} of the set of N LDR images, having exposure e_k thus:

$$R_{ij} = \sum_{k=1}^{N} g(L_{ijk}) / e_k$$

Hence alongside each LDR image, we need EXIF or similar metadata detailing exposure settings. In the above equation, the exposure e_k is expressed in terms of equivalent exposure time. In practice, a weighting function is also applied to filter-out pixels that are approaching under-exposed and over-exposed values.

Exposure can be varied by altering the camera aperture, the shutter speed, the camera gain (ISO value for film), or the ND filter if one is fitted. However, varying the aperture will also vary the depth of field, introducing a dis-similarity between the images, causing multiple-image artefacts in the resulting HDR image. Similar issues arise if we vary the shutter speed or the gain: the shutter speed affects the amount of motion blur in an image, and the gain setting affects the amount of noise. Altering the exposure via the ND filter has no side-effects.

Traditionally, digital HDR images are constructed from sets of digital photographs taken sequentially of a still-life subject at different shutter speeds. However for HDR video, where we need to capture/reconstruct at least 24 HDR images per second of a subject that is likely to be moving, the process is somewhat more involved.

Approaches for Capturing HDR Video

There are a number of approaches to capturing HDR video, which are summarised as:

1. **Read-out multiple exposures from the sensor.** This solution effectively provides multiple exposures per frame, each having a different shutter speed and hence a different photometric exposure. It works by reading-out the pixel data from the sensor multiple times during each frame period without clearing the pixel data from the sensor until the end of the frame period. For example, assume the frame rate is set to 25 frames per second with a 100% shutter, and 3 exposures spaced apart by 2 stops each are being captured per frame. To achieve this, in addition to the usual sensor-read after 40ms of the start of each frame, the camera would make two further intermediate sensor-reads after 2.5ms and 10ms of the start of each frame.

   In the next section, we describe in further detail how we used a commercially-available digital cinema camera (3) to employ this technique for capturing footage for our own experimental HDR movie.

2. **Use two cameras mounted on a 3D mirror rig with zero intraocular distance.** As with a standard stereo 3D set-up, the cameras must be genlocked, have matched lenses, and have tracked aperture, focus, and zoom controls. Additionally, one of the cameras should be fitted with a ND filter to provide the exposure difference between the two cameras. For this set-up, the rigidity and accuracy of the rig is paramount, as the quality of the resulting HDR image is highly sensitive to misalignment between the LDR images.

3. **Use multiple sensors with an optical beam-splitter between the lens and sensors.** As above, this is a direct solution to the problem of obtaining multiple exposures at different exposure settings at the same instant in time, with the same

shutter speed and aperture. An optical beam-splitting prism divides the light from the lens to two or more sensors. The amount of light fed to each sensor is controlled so that there is an even spread of exposure settings across the dynamic range. In Tocci et al (18), this was achieved by using an optical beam-splitter system that divided the light to each of the three sensors in the appropriate proportions (92%, 7.5%, and 0.44%).

4. **Increase the dynamic range of the sensor.** Recent developments in camera sensor technology are facilitating larger pixel word lengths, and we are now seeing CMOS sensors with word lengths of up to 16 bits per pixel colour. Even after accounting for noise, there is a real possibility of directly capturing video with dynamic ranges extending beyond 4 orders of magnitude, without the need for capturing and post-processing multiple images per frame.

PRODUCTION

In this section we discuss how we employed the above theory to produce our experimental HDR movie. Our full production workflow is summarised in figure 1, and it is discussed in further detail in the following sub-sections.

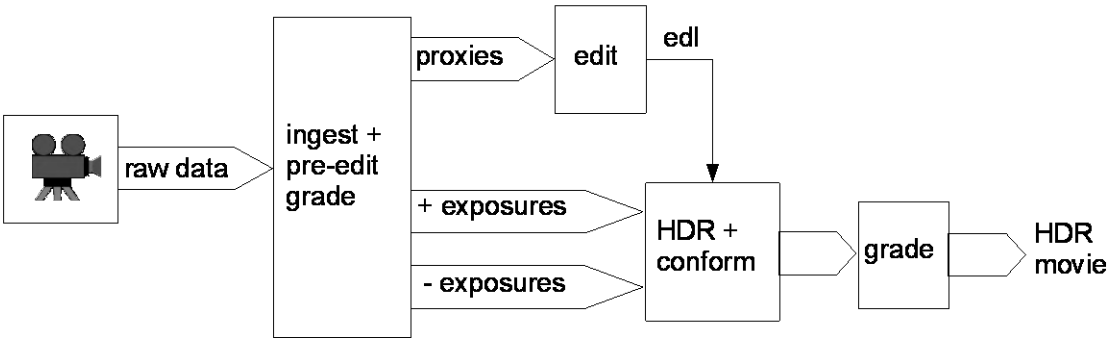

Figure 1 – Our HDR Movie Production Workflow

If producing a HDR movie under less experimental circumstances with the camera we used, the proprietary software tool (19) (referred to herein as ‘RCX’) that accompanies the camera is likely to be at the centre of the workflow (Price and Corp (20)). We chose to use an alternative workflow to this, primarily because we wanted more control over the HDR formation process than the RCX tool provided.

Capturing Footage

The footage for our HDR movie needed to have lots of inherently HDR content without being unduly challenging. Hence we shot a range of subjects, including a night-time carnival and a choreographed pyrotechnics show, having details in both light and dark areas.

We used a camera (3) that captures two exposures per frame by reading twice from the same sensor per frame: the main exposure is read-out of the sensor at the end of its shutter period; the intermediate exposure is read-out at a user-controlled instant before the main exposure. The user sets the timing of the intermediate exposure in terms of number of stops in exposure below the main exposure, of between 1 and 6 stops.

The manufacturer's recommended usage of the camera in its HDR capture mode, is to set the aperture as though filming normally, allowing the main exposure to be used directly on standard dynamic range displays. However as our footage was specifically for a HDR application, we set the aperture and the HDR setting so that the main exposure was over exposed by 2 or 3 stops (depending on light levels in the subject), and the intermediate exposure was 4 or 6 stops below the main exposure. This placed the main and intermediate exposures approximately 2 or 3 stops either side of what would have been the normal exposure setting, giving us an equal expansion in dynamic range above and below 'normal'.

Some training/re-education of the production crew was required to achieve this, as setting exposure for our HDR method breaks fundamental rules of conventional photography.

All footage was shot at a resolution of 3840x2160 at 50 frames per second with a 180^{o} shutter.

Ingest

After colour corrections were made, suitable proxies were generated using RCX. The HDR composition tool within RCX was used for pre-visualising the footage within the tool while making colour corrections, and for tuning the proxies to give a fair LDR representation of the footage.

Edit, Conform and HDR Composition

As our footage was captured as two LDR images per frame, an extra stage was required in the process to compose the HDR video frames from the LDR images. The decision of where in the workflow this conversion happened was influenced by a number of factors, including the amount of time available and the capabilities of other tools (e.g. edit) used in the workflow. We therefore chose to convert to HDR after the edit stages.

Demosaicing algorithms, used to convert the raw sensor data to standard image formats, can fail to produce the correct colours in the resulting image for localised regions that have some saturated pixels ('highlights'). With the conventional HDR imaging algorithm described above, the artefacts from this are pulled-through to the resulting HDR image, especially when using only two LDR images per frame. Hence we would have ideally preferred to have implemented our own HDR composition software based on the algorithm presented in (18), which constructs the HDR image from the raw LDR images prior to demosaicing, and applies the demosaicing algorithm to the resulting HDR image instead. However, we resorted to using 'off-the-shelf' tools, as implementing HDR software was outside the remit of this stage of the project, and is a focus of our on-going R&D work.

A number of HDR construction tools were widely available, including the open source 'pfstools' (21) as well as a number of proprietary products, including the HDR functions provided as a part of the RCX tool. We used a variety of these tools to maximise the resulting subjective quality of our clips.

For the edit, we used a tool capable of exporting to EDL (22). We developed our own conform tool which applied the same edits to both the main and intermediate exposures of the footage. In preparation for this, the main and intermediate versions of each clip were converted and stored as separate final resolution (1920x1080) TIF image stacks during ingest. The resulting main and intermediate versions of the final edit were then used to construct the HDR movie.

Grading

It may be best to grade the final edit after converting the movie to HDR format, to maintain precise control of the demosaicing and HDR conversion process – indeed, it may be best if the HDR formation process were incorporated into the grading tool. However, we have found that commercial grading tools presently have limited import capabilities for high dynamic range content.

The new Academy of Motion Picture Arts and Sciences 'ACES' (23) architecture and format for digital movie production is based on the OpenEXR format, and has options for managing high dynamic range assets. As ACES is undergoing standardisation with SMPTE, it is likely that commercial grading tools will be capable of importing files stored in this format in the near future.

For our HDR movie, pre-edit grading was applied prior to HDR formation, taking care to apply the same colour corrections and gamma alterations to both main and intermediate exposures of the footage. Final grading has been deferred for the time being.

SUMMARY

This paper has described our method for assessing audience experience with the use of an 'immersion' measurement technique. It also discussed the methodology for capturing HDR video. This was followed by a detailed discussion of the workflow we used in producing our own short HDR video, which we are using for our ongoing audience experience tests. Our preliminary picture quality assessments indicate that extending dynamic range through the entire chain from camera to display can yield significant benefits.

REFERENCES

1. ARRI, 2013. 'ALEXA Digital Cameras', http://www.arri.com/camera.
2. Sony, 2013. 'PMW-F55 CineAlta 4K', http://www.sony.co.uk/pro/product/xdcamcamcorders/pmw-f55/
3. RED Digital Cinema, 2011. 'Epic Mysterium X', http://www.red.com/products/epic-mx.
4. Sim2, 2013. 'HDR47 High Dynamic Range Display Series', http://www.sim2.com/HDR/
5. T. Kunkel and E. Reinhard, 2010. A reassessment of the simultaneous dynamic range of the human visual system. Proceedings of the ACM Symposium on Applied Perception in Graphics and Visualization, Los Angeles, 2010, pp. 17-24.
6. D. Bordwell, K. Thompson, 2008. Film Art: An Introduction. McGraw-hill, New York.
7. W. A. Ijsselsteijn, H. de Ridder, J. Freeman, S. E. Avons, 2000. Presence: concept, determinants and measurement. In B. E. Rogowitz and T. N. Pappas (Eds.), Human

Vision and Electronic Imaging V, vol 3959, pp.520-529. Bellingham: Spie-Int Soc. Optical Engineering.

8. J. V. K. Draper, D. B. Kaber, J. M. Usher, 1998. Telepresence. Human Factors, vol 40(3), pp. 354-375.
9. W. Barfield and S. Weghorst, 1993. The Sense of Presence Within Virtual Environments: a conceptual framework. In: G. Salvendy and M. Smith (Ed.), Human-computer interaction: software and hardware interfaces (pp. 699-704). Elsevier.
10. D. R. Anderson and J. Burns, 1991. Paying Attention to Television. In: J. Bryant and D. Zillman (Eds.), Responding to the screen: perception and reaction processes (pp. 2-26). Hillsdale, New Jersey.
11. ISPR, 2000. International Society for Presence Research, 'Measures Statement and Compendium', http://sct.temple.edu/blogs/ispr/about-presence-2/tools-to-measure-presence/ispr-measures-compendium/
12. J. Freeman, S. E. Avons, R. Meddis, D. E. Pearson, W. A. Ijsselsteijn, 2000. Using Behavioural Realism to Estimate Presence: a study of the utility of postural responses to motion-stimuli. Presence: Teleoperators and Virtual Environments, vol 9(2), pp. 149-165.
13. T. Troscianko, T. S. Meese, S. Hinde, 2012. Perception While Watching Movies: effects of physical screen size and scene type. i-Perception vol 3, pp. 414-425.
14. S. R. Ellis, 1996. Presence of Mind: A reaction to Thomas Sheridan's 'Further Musings on the Psychophysics of Presence'. Presence, vol 5(2), pp. 247-259.
15. J. Lassiter, J. Freeman, E. Keogh, J. Davidoff, 2001. A Cross-media Presence Questionnaire: the ITC sense of presence inventory. Presence: Teleoperators and Virtual Environments, vol 10(3), pp. 282-297.
16. ITU Recommendation ITU-R BT.500-13, 2012. Methodology for the subjective assessment of the quality of television pictures. International Telecommunications Union.
17. E. Reinhard, G. Ward, S. Pattanaik, P. Debevec, W. Heidrich, K. Myszkowski, 2010. High Dynamic Range Imaging: acquisition, display, and image-based lighting. Elsevier.
18. M. Tocci, C. Kiser, N. Tocci, and P. Sen, 2011. A Versatile HDR Video Production System. ACM Trans Graph., vol 30(4), article 41.
19. RED Digital Cinema, 2013. 'REDCINE-X PRO', http://www.red.com/products/redcine-x
20. M. Price and A. Corp, 2012. Private discussions with BBC Natural History Unit.
21. 2011. 'pfscalibration', http://pfstools.sourceforge.net/pfscalibration.html.
22. Apple Inc., 2009. 'Final Cut Pro 7', http://www.apple.com/uk/support/finalcutpro7.
23. The Academy of Motion Picture Arts and Sciences, 2013. http://www.oscars.org/science-technology/council/projects/aces.html.

ACKNOWLEDGEMENTS

The authors would like to thank Peter Milner of the University of Bristol Drama Department for his assistance with the making of the HDR movie.