First published in the Proceedings of the 21st European Signal Processing Conference (EUSIPCO-2013) in 2013, published by EURASIP.

CREATING HDR VIDEO CONTENT FOR VISUAL QUALITY ASSESSMENT USING STOP-MOTION

P. Zolliker[1,2], *Z. Barańczuk*[1], *D. Küpper*[1], *I. Sprow*[1]

[1] Laboratory for Media Technology
[2] Electronics/Metrology/Reliability Laboratory
Swiss Federal Laboratories for Materials Science and Technology (Empa), Dübendorf, Switzerland

T. Stamm

Manderim GmbH
Zürich, Switzerland

ABSTRACT

A complete workflow from capture via processing to the display of video sequences, entirely in high dynamic range (HDR), is described. This workflow is set up to allow for image quality assessment of HDR content with renderings for smaller dynamic ranges. Special emphasis is on the controlled reproduction of lightness and color. Our method allows capturing scenes with a high dynamic range, storing the HDR content in physically meaningful units and displaying it on a commercially available HDR display.

***Index Terms*—** high dynamic range, video, capture, display, bracketing, stop-motion, photography

1. INTRODUCTION

High dynamic range images store real-world lighting information to display a more realistic image. Novel capture and display technologies enable an imaging pipeline that encompasses 5 orders of magnitude between minimum and maximum luminance, a contrast of 100 000:1. This covers the range the human eye is capable to perceive simultaneously [1]. Current HDR procedures however still consist of reproducing HDR input data accurately to the available low dynamic range (LDR) display devices by mapping the content accordingly. The definition of 'accurate' is defined by the best fitting tone mapping operator (TMO) for a certain imaging task.

While tone mapping operators (TMO) have been a major research topic in the past decade [2], of necessity to fit HDR content to traditional low dynamic range (LDR) display limitations, the emergence of HDR displays opens up new ways to assess perceived image quality during TMO development. Such algorithms have been studied widely but only few studies included evaluations of their visual performance in regard to the original scene for obvious difficulties of setting up a controlled test environment [3].

Using an HDR display [4] allows to simultaneously compare mapped images with the closest-to-reality representation of the scene, the HDR image. Much effort is put into the development of a high quality HDR workflow, such as the development of HDR cameras, displays and software components, for example within the COST action IC1005 on HDRi[1]. However, for a setup of a HDR test environment for image quality assessment, key parts such as a high quality HDR camera or a video viewer for uncompressed data are not commercially available.

In this paper, we describe a full HDR workflow that is developed for psychophysical research of TMOs. The challenge is to process HDR data with agreeing luminance values between the captured scene and the resulting displayed HDR image. For this, devices need to be characterized by means of luminance measurements. The image quality of TMOs shall be visually evaluated on both, still images and video material. For video material we need a method to bypass the current limitations of HDR video camera availability and the display of uncompressed HDR video data. A method is presented to create HDR video from still images using multi exposure imaging technology [5] [6] and stop motion. The display is driven by a newly developed HDR tool named *Viewy* which is able to prepare videos from still images and can display several short-sequences of uncompressed HDR video in parallel. This allows us to set up a visual testing environment where unmapped and mapped HDR content is displayed on a Sim2 HDR47E display [7] simultaneously for subjective image quality assessment. The main contributions of this paper are: controlled HDR workflow, creation of HDR video sequences without HDR video camera, verification of workflow by means of luminance measurements.

2. WORKFLOW

The workflow is summarized in Fig.1. Images of a scene are captured as low dynamic range (LDR) exposure series. Then the images are fused together into a device-independent HDR image, based on a previously determined camera calibration.

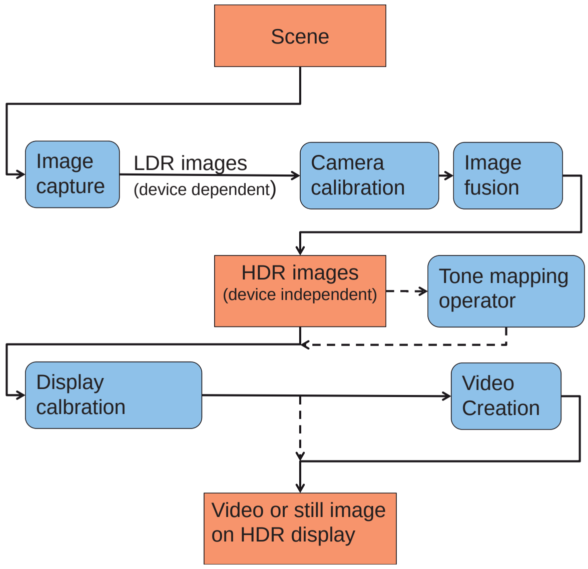

Fig. 1. Workflow from scene capture to video display

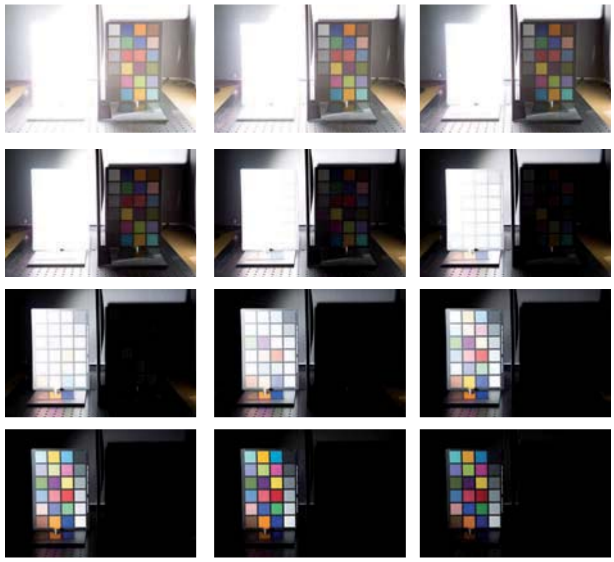

Fig. 2. Exposure series of two test charts in extreme lighting condition, 12 images @1 EV spacing captures a dynamic range of 23 EV.

Fig. 3. Luminance measurements of scene inside capture booth (left) and display (right) for device characterization. Stray light is also measured (with and without black tubus).

Depending on the application, different tone mapping operators can be applied, or the HDR data is left unchanged if images have to be displayed as scene data. The HDR images have to be adapted to the target display for which tone mapping has to be optimized. They are either put together into a video or shown as still images. In the following, the different steps are described in more detail.

2.1. Device Characterization

To calibrate the camera, an exposure series of a scene with two identical Gretag Macbeth test charts in extreme lighting conditions is captured (see Fig 2) similar to the set up used by Fairchild [8]. The 48 gray and color patches cover a lightness range of 4 to 5 orders of magnitude. The patches in the scene are measured with a Minolta CS 2000 spectrophotometer at the 63 cm measuring distance. To minimize the effect of stray light, a black tubus with 35 cm length, a 6 cm diameter and a 4 cm opening was placed around the lens (see Fig 3 for measurement set-up).

An offset, a color transform matrix and a compensation for stray light are determined in order to match the measured X, Y, Z intensities with those in the HDR image. The section about image fusion gives further details.

The same scene is then shown on the SIM2 display and the 48 patches are then measured on the display with the same spectrophotometer (see Fig 3 right) and the parameters of a calibration transform are calculated. The transform is analogous to matrix profiles used in an ICC workflow, i.e. it consists of a scale, a color transform matrix and a gamma correction exponent for the lightness adaption.

2.2. Image Capture

HDR still images are photographed in a controlled capturing environment with defined lighting, object- and camera position. The set-up consists of an X-Rite SpectraLight light booth with two additional dedolight DLH4 spots, a rotation stage and camera mount (see Fig 2). Objects are placed on the 2-axes rotation stage which is moved on a defined, incremental path. Sample images for a typical scene are shown in Fig 5). At each of the 180 increments, the motion is stopped just long enough to capture an exposure bracketing series.

Each exposure bracketing covers the dynamic range of the scene. With the employed DSLR Canon 1D Mark4 which has a sensor that covers 12EV for a single exposure [9], a dynamic range of 20EV is approximately achieved which covers the measured scene contrast of about 14EV, a contrast of 1:20.000. All settings on the camera are fixed, only the exposure time is varied during bracketing.

Surveying HDR images on the HDR display that are created from either Raw, sRaw and JPEG images, reveal that

Raw images lead to the best image quality. Therefore single images are captured in the Raw format to preserve highest accuracy and are downsampled using bicubic interpolation outside the camera before HDR image fusion to the target resolution shown on the display.

Due to the nature of this set-up, only relatively small objects can be placed in the light box using the rotation stage. Additional video material was collected from slowly changing natural scenes (e.g. sun sets) using bracketing and time lapse methods.

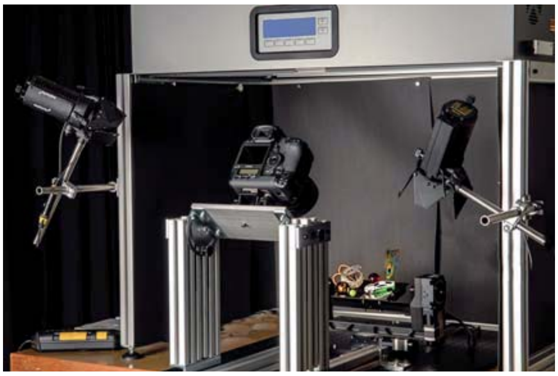

Fig. 4. Capture environment with lightbox and camera mount.

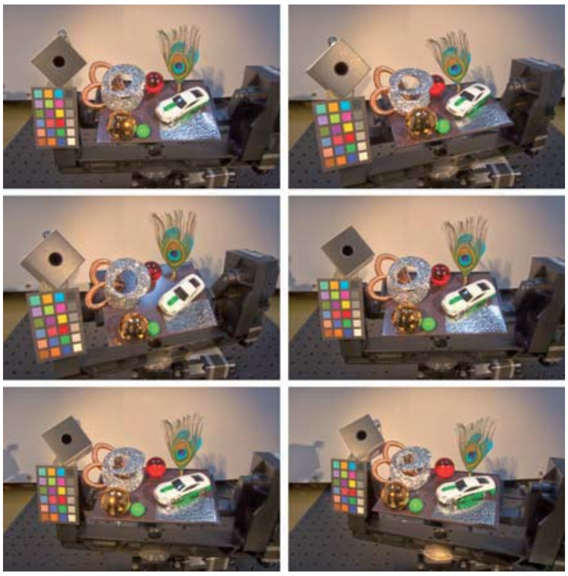

Fig. 5. Incrementally rotated object. One movie sequence consists of 180 such HDR still images, created with the stop-motion technique.

2.3. Image Fusion

In a first step, the images of the exposure bracketing series are converted from the native Raw format to linear camera RGB, without applying any white balancing. This is done using dcraw[2], an open source decoder for Raw images. Evaluating the resulting camera sensor characteristics showed that for our purposes, the responses are sufficiently linear to allow for the assumption that measured intensities scale with exposure time, except if the sensor is saturated. This means, that given a sensor value v and an exposure time t, we can safely assume that the irradiance at the sensor is proportional to v/t.

A pixel in the final fused image is a weighted average of the pixels at the same pixel location in the image, but acquired using different exposure times. Assuming sensor values in $[0, 1]$, we use a triangular weighting function centered at 0.5, i.e. we favor the exposure time that leads to a sensor value midway between under- and overexposure. Similar to [6], we have

$$w(x) = \begin{cases} x & : x < 0.5 \\ 1 - x & : x \geq 0.5 \end{cases}$$

To guard against crosstalk in cases where one color channel is fully saturated while another is not, for a pixel p we compute a single weight as the minimum of the three channel weights, i.e. $w(p) = \min\{w(p_R), w(p_G), w(p_B)\}$. Putting it all together, for a given sensor location, we have pixel values p_i, exposed for time t_i, determining the fused value p^* according to

$$p^* = \frac{\sum_i p_i \cdot w(p_i)/t_i}{\sum_i w(p_i)},$$

where i indexes the individual images of the exposure series.

The resulting values have not been corrected for different sensor gains, and the data also suggests that these values are affected by stray light. To correct for this, we use the gray patches in the color checker, and for each individual channel we make the ansatz that the sensor data can be explained by the relation $v = a \cdot Y + \lambda$, where v is the channel value, a is the sensor gain for that channel, and Y is the lightness of the patch, measured in $\mathrm{cd/m^2}$. This leads to a simple least squares problem. However, since the data spans several orders of magnitude, we do not minimize absolute but rather relative errors. In a final step, we determine a matrix that transforms the fused RGB values into XYZ values. This is again a least squares problem, based on the 48 XYZ measurements in our reference scene. The resulting fit is shown in Fig 6.

2.4. HDR Video creation

Short video sequences are now built from the calibrated HDR stop-motion still image series. When processing images for display, the 32 Bit values of the HDR files are converted to RGB float values (3 × 32 Bit). Arbitrary tone mapping algorithms can now be applied as well as any calibration desired. The DVI-Plus mode of the SIM2 display expects input color values in a vendor-specific 24 Bit LogLuv format. The

[2]http://www.cybercom.net/~dcoffin/dcraw/

conversion from RGB float values to the LogLuv format can be done using a fragment shader in the GPU. With current GPU drivers, the transfer of float values to the GPU is very time consuming and therefore limits the frame rate when displaying videos. By using preprocessing, all necessary tasks like image scaling, tone mapping, calibration and finally the LogLuv conversion can be done on the CPU. Later on, the preprocessed data is transferred directly to the display without any further processing. Due to much faster transfer rates of integer data (3×8 Bit), the speed of our CPU-based implementation even surpasses the GPU-solution.

When rendering videos, all frames are converted to the LogLuv binary representation and are stored on disk as an image sequence. As for intended application, only short video sequences are used, the amount of memory available in modern hardware configurations is sufficient to hold all image data. Therefore, all frames can be loaded from disk once and then be accessed reliably fast. When displaying multiple videos in parallel, they need to be synchronized. This has been solved with a busy-loop which selects the appropriate frame-number for each video, loads the images and displays them simultaneously.

3. RESULTS AND DISCUSSION

Dynamic range. Using our characterization data, we have shown that our workflow can handle HDR images with a dynamic range exceeding 1:20 000. There are two steps in our workflow where colors could possibly be distorted. First, when fitting the fused camera images to the measurements of the actual scene. Second, when displaying the reconstructed measurements on the monitor. Fig. 6 shows the measured intensities in the scene and on the display (ordinate) as a function of the reference intensities (abscissa).The log-log plot shows that the lightness is well controlled over an order of magnitude of 4 to 5. A few of the darkest patches show a lightness which is higher than modeled. A closer inspection showed that the global stray light correction used for the camera characterization is not accurate enough. The left row of the test chart in the dark (see Fig. 3, left) was affected by stray light of the bright scene on the left side much more than the other patches, leading to lightness differences in the order of the error seen in the plot.

Color reproduction. Color gamut on a display is restricted by two factors, 1) the primaries of the display and 2) the maximum intensities of the three channels. Our Sim2 display has primaries that are close to the sRGB primaries, thus it cannot display colors which lie outside the primaries-based triangle in the x-y diagram (see Fig. 7 left). However due to high intensities attainable with the HDR display, the color capabilities for light colors are much larger than in conventional displays. The lightness of reference white (typically 300 $\mathrm{cd/\,m^2}$) is usually set well below the display's maximal capabilities (typically 4000 $\mathrm{cd/\,m^2}$). (see Fig. 7 right).

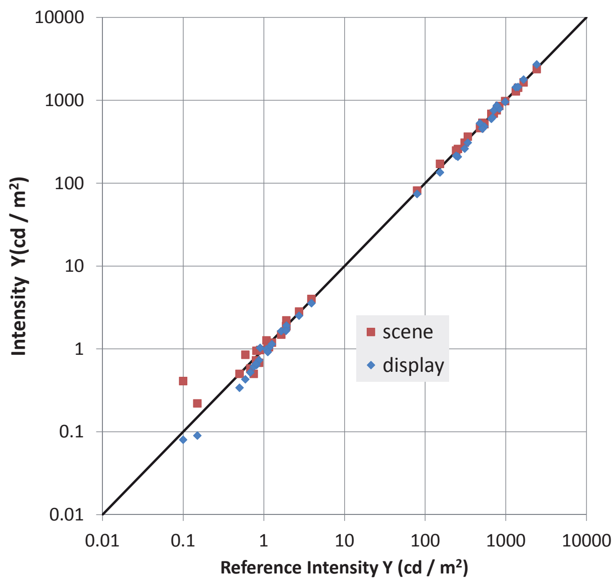

Fig. 6. Measured lightness range for display and scene compared to intensities in HDR image.

Video speed. HDR-video rendering in two FullHD-sized windows as well as showing three videos on one screen showed very good synchronization and no stuttering with 30 FPS and more. Note that this speed is realized without using any video compression algorithm. Playing videos in a smooth, synchronized manner is an important prerequisite for psychophysical testing of moving scenes.

4. CONCLUSION AND FUTURE WORK

The described workflow is a basis for setting up psychophysical tests comparing representations of moving scene content in HDR with tone mapped versions on the same display. It is also a well controlled reproduction of lightness in a dynamic range exceeding conventional display by one to two orders of magnitude. Results of a first such psychophysical study using this workflow were presented by Sprow et.al. [10]. The infrastructure is also suitable for other types of quality assessment such as comparing videos using different compression algorithms (JPEG XR or video compression) with uncompressed videos. Nevertheless there are several issues, which cause color and lightness reproduction to be far from perfect and need to be considered carefully if psychophysical tests are evaluated:

1) The technology of the available display using a limited number of LEDs for back light illumination reduces the available local dynamic range of the monitor. However the human

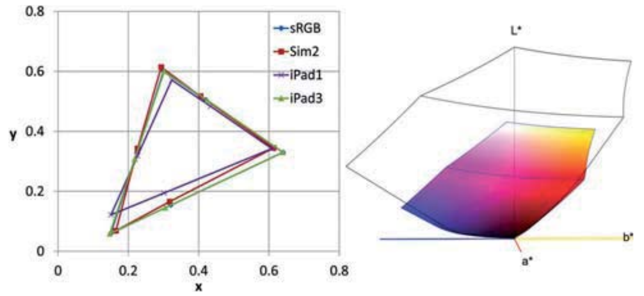

Fig. 7. Left: Gamut in x-y coordinates: Sim2 (red) compared to typical displays (iPad1 (violett), iPad3 (green); sRGB gamut is almost identical to iPad3 gamut . Right: Gamut in CIELAB coordinates: Sim2 (grid) with iPad3 (colored).

visual system also has a reduced perception of local intensity ratios [11] .

2) Differences due to stray light effects for the spectrophotometric measurement, the capture with a camera and the viewing by humans are compensated only partially. Here a model for local compensation of stray light for the measurement devices as well as the visual system could substantially improve the matching of colors in a scene and its reproduction.

3) Note that for an assessment of perceived color differences in an HDR environment the currently used color difference models such as CIELab [12], CIECAM [13] and iCAM [14] are not sufficient to describe color differences over the dynamic range in question. They can not accurately predict perceived color difference in the required dynamic range. What would be needed is a model with local white adaptation.

ACKNOWLEDGEMENTS

The project was partially funded by the Swiss Commission for Technology and Innovation (CTI) and supported by X-Rite Europe GmbH and the EU COST Action IC1005 HDRi.

5. REFERENCES

[1] E. Reinhard, G. Ward, S. Pattanaik, P. Debevec, W. Heidrich, and K. Myszkowski. *High Dynamic Range Imaging: Acquisition, Display, and Image-based Lighting*. The Morgan Kaufmann series in Computer Graphics. Elsevier (Morgan Kaufmann), Burlington, MA, 2nd edition, 2010.

[2] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen. Evaluation of tone mapping operators using a high dynamic range display. *ACM TRANSACTIONS ON GRAPHICS*, 24:640–648, 2005.

[3] A. Yoshida, V. Blanz, K. Myszkowski, and H.-P. Seidel. Perceptual evaluation of tone mapping operators with real-world scenes. In *Human Vision and Electronic Imaging X, SPIE*, 2005.

[4] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs. High dynamic range display systems. In *ACM SIGGRAPH 2004 Papers*, SIGGRAPH '04, pages 760–768, New York, NY, USA, 2004. ACM.

[5] S. Mann and W. Picard. On being 'undigital' with digital cameras: extending dynamic range by combining differently exposed pictures. In *Proceedings of Society for Imaging Science and Technology's 48th Annual Conference*, pages 442–448, 1995.

[6] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '97, pages 369–378, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.

[7] Sim2. HDR47E technical specifications. `http://www.sim2.com/HDR/hdrdisplay/hdr47e_s_4k`, 2013.

[8] M. D. Fairchild. The HDR photographic survey. In *IS&T/SID 15th Color Imaging Conference*, pages 233–238, 2007.

[9] DXO Image Science. Tests and reviews for the camera Canon EOS 1D Mark IV. `http://www.dxomark.com/index.php/Cameras/Camera-Sensor-Database/Canon/EOS-1D-Mark-IV`, 2013.

[10] I. Sprow, D. Küpper, Z. Barańczuk, and P. Zolliker. Image quality assessment using a high dynamic range display. In *Proceedings of the 12th Congress of the International Colour Association, Newcastle upon Tyne, England*, 2013.

[11] J.E. Dowling. *The Retina: An Approachable Part of the Brain*. Belknap Press of Harvard University Press, 2012.

[12] CIE. *Recommendations on uniform color spaces—color-difference equations, psychometric color terms*, 1978.

[13] N. Moroney, M. D. Fairchild, R. W. G. Hunt, C. Li, M. R. Luo, and T. Newman. The CIECAM02 color appearance model. In *IS&T/SID 10th Color Imaging Conference*, pages 23–27, 2002.

[14] M. D. Fairchild and G. M. Johnson. The iCAM framework for image appearance, image differences, and image quality. *Journal of Electronic Imaging*, 13:126–138, 2004.