VR+HDR: A SYSTEM FOR VIEW-DEPENDENT RENDERING OF HDR VIDEO IN VIRTUAL REALITY

Hossein Najaf-Zadeh, Madhukar Budagavi and Esmaeil Faramarzi

{h.najafzadeh, m.budagavi, e.faramarzi}@samsung.com
Samsung Research America, 1301 E. Lookout Drive, Richardson, TX 75082

ABSTRACT

This paper introduces a view-dependent method for rendering high-dynamic-range (HDR) video in virtual reality (VR) on VR displays such as head mounted displays (HMD), mobile phones, TVs, and computer monitors. The user's view direction is taken into account to design a tone-mapping operator which appropriately displays the HDR content on the display device. The proposed method can be utilized if HDR capturing and playback (i.e. HDR camera and 10-bit video codec) are available. However, it can also be used on the 8-bit pipeline (i.e. 8-bit camera and 8-bit video codec). A VR+HDR prototype was implemented using Samsung Gear360 camera and SamsungVR Android app on Samsung Galaxy Note 5 smartphone. The subjective comparison between the proposed method and rendering of 360-degree VR video through global tone mapping was performed using this prototype and showed that for all test sequences, the proposed technique significantly improves the video quality in VR rendering.

Index Terms— VR+HDR, virtual reality (VR), view-dependent tone mapping, high dynamic range (HDR), 360-degree video

1. INTRODUCTION

A real-world scene can cover very bright spots such as the Sun (above 1000s nits) to very dark regions. As such, the dynamic range is significantly higher for a 360-degree video when compared to a regular video, especially for outdoor scenes and user generated content. For instance, cameras pointing to the Sky/Sun have too much peak brightness while cameras pointing away from the sun have less peak brightness. Normalizing the brightness in the images from multiple cameras and stitching them into a single 360-degree Standard Dynamic Range (SDR) image leads to lost details in dark and/or bright regions in the image as shown in Figure 1. So, HDR functionality is needed in VR for improving video quality.

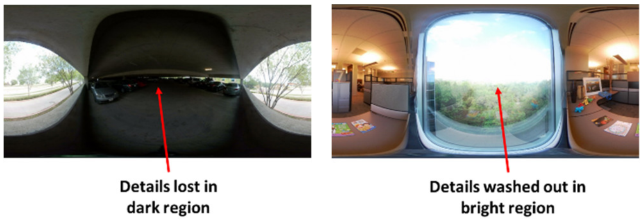

Details lost in dark region
Details washed out in bright region

Fig. 1. Lost details in SDR 360-degree images

VR+HDR content (i.e. 360-degree HDR video) is easily generated by using existing projections and mappings in conjunction with HEVC Main 10 codec or other codecs such as AVC along with coding options described in JCT-VC-Y0046 [1][2]. However, the full dynamic range of the scene (which can vary from 0 nit to 4000 nits, e.g. in curated content) cannot be shown on current video displays such as mobile phones and HMDs. So, a tone-mapping operator (i.e. a color transform) is used to transform the VR+HDR content to be shown on such displays.

To map the dynamic range of the input HDR content to a given display, a fixed tone-mapping operator depending on the dynamic range of both the input content and the display can be designed and utilized. An example of such tone-mapping operators includes the Knee Function SEI and the Colour remapping information SEI message of HEVC [3]. The tone-mapping function can be made scene-adaptive too to improve performance. Recently, the Society of Moving Pictures and Television Engineers (SMPTE) has published

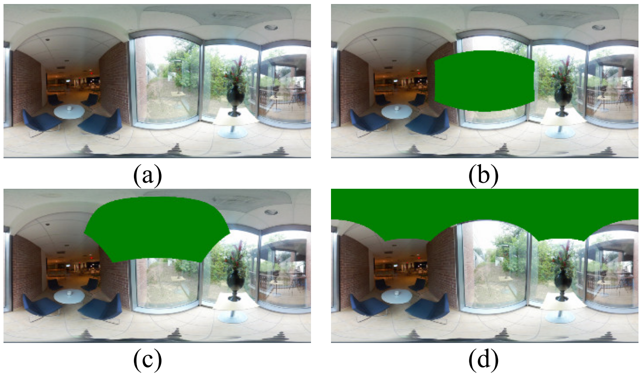

(a) (b) (c) (d)

Fig. 2. The area (green in color) projected into the viewport for different viewpoints. (a) Original image; (b) yaw = 0^{0}, pitch = 0^{0}; (c) yaw = 0^{0}, pitch = 45^{0}; (d) yaw = 0^{0}, pitch = 90^{o}.

978-1-5090-2175-8/17/$31.00 ©2017 IEEE

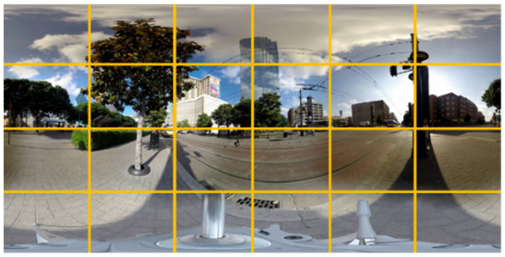


Fig. 3. Example of image tiling.

ST2094, a new standard on dynamic metadata for color transforms of HDR and WCG images [4]. In ST2094, metadata is dynamically generated for each scene and to be used for dynamic range conversion at the receiver side. These tone mappings (i.e. color transforms) can be used in the VR context too.

However, in VR, a user's field of view is limited to around 100 degrees, allowing the user to see a part of the 360-degree video at any time (i.e. viewport) as depicted in Figure 2. This feature implies that an optimal tone-mapping operator for VR video rendering can take into account the dynamic range (or other statistics) of the part of the 360-degree video seen by the user at any time. This approach, compared to a global or scene-adaptive tone mapping, will adapt the tone-mapping operator to the user's field of view which results in better video quality.

Currently, MPEG is developing the omnidirectional media application format (OMAF) for storing and rendering omnidirectional content (i.e. 360-degree video). The official requirements for OMAF include post-processing metadata for visual quality improvement (e.g. HDR tone mapping). It also supports view-dependent processing (encoding, delivery and rendering) according to user's view direction (i.e. viewpoint) [5]. The VR+HDR system described in this paper has been proposed to OMAF.

The approach proposed in this paper is based on view-dependent tone mapping of the HDR content being displayed in the viewport. Section 2 describes a VR+HDR system in which HDR capturing and playback (i.e. HDR camera and 10-bit video codec) are available. However, in Section 3, a VR+HDR system is proposed for devices with an 8-bit pipeline (i.e. 8-bit camera and 8-bit video codec). The results and conclusion are represented in Section 4.

2. VR+HDR SYSTEM FOR 10-BIT PIPELINE

As stated earlier, a user will see a part of the 360-degree video in VR at a time. To design an optimal tone-mapping operator, the receiver needs to know the peak luminance (or other statistics) of the input HDR content in the viewport. However, computing such statistics at the receiver side results in some challenges. First, the receiver which could be a battery-operated head-mounted device (HMD) or a VR player normally does not have enough processing power or battery capacity. Second, flicker can happen due to an abrupt change in peak luminance when the user's viewport changes. One way to overcome these challenges is to compute the luminance statistics at the transmitter and send them out to the receiver for the current viewport of the current frame. However, this requires a very fast communication link such that the receiver sends the user's current viewpoint to the transmitter, and the transmitter computes the luminance statistics in real time and sends them to the receiver.

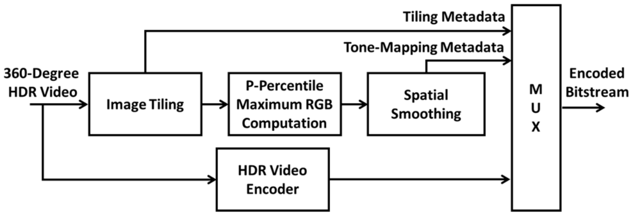


Fig. 4. Block diagram of the transmitter in the proposed VR+HDR system with 10-bit pipeline.

This paper proposes an alternative approach which will not require a fast communication link and real-time computation of peak luminance at the receiver. It divides the entire 360-degree HDR image/video frame into multiple tiles. The tiling map and the peak luminance for each tile are transmitted to the receiver as metadata along with the encoded HDR content. Figure 3 illustrates an example of image tiling.

2.1. Transmitter

Figure 4 displays the block diagram of the transmitter in the proposed VR+HDR system with 10-bit pipeline. A video encoder with 10-bit functionality such as HEVC Main 10 encodes the HDR content. In parallel, some metadata on image tiling and tone-mapping parameters are extracted from the input content and multiplexed with the encoded HDR video sequence. The metadata will include the following parts:

Tile Size: a pair of integers that specify the width and height of a tile (in number of pixels).

Smoothed Percentile Maximum RGB Values: a matrix containing smoothed *P*-percentile maximum RGB value in each tile. The value of this metadata in each tile will be larger than the maximum RGB values of *P*% of the pixels in the tile. Note that RGB values are linearized color components. This metadata is calculated in two steps. First, *P*-percentile maximum RGB value in each tile is calculated. Then, some spatiotemporal smoothing is performed on the percentile maximum RGB values to avoid flicker artifact due to fast illumination change between the viewports in space and/or time.

$$M_{i,n} = \underset{i\in \mathrm{I}}{\mathrm{Filt}}(L_{i,n}) \qquad (1)$$

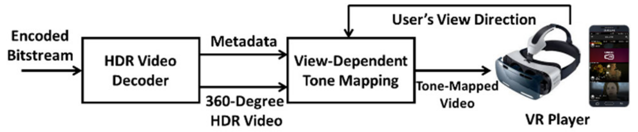


Fig. 5. Block diagram of receiver in the proposed VR+HDR system with 10-bit pipeline.

where $\mathrm{Filt}(\cdot)$ denotes the spatial smoothing operation, $L_{i,n}$ and $M_{i,n}$ are respectively the P-percentile maximum RGB values for the i-th tile in the n-th frame before and after smoothing, and I is the total number of tiles in each video frame.

The use of percentile maximum RGB value (instead of absolute maximum RGB value) will reduce the impact of a few very bright pixels in the tile. A value of P (i.e. percentage) near 100 will better preserve details in the bright parts of rendered video/image, whereas a smaller value of P will better preserve details in the dark parts of rendered video/image at the cost of some saturation in the bright pixels. In the proposed method, P is set to 99%.

2.2. Receiver

Figure 5 demonstrates the block diagram of the receiver in the proposed VR+HDR system with 10-bit pipeline. The decoder outputs the VR+HDR video and the associated metadata. Based on the user's view direction obtained from sensors in an HMD or a VR player device, the metadata (i.e. $M_{i,n}$ in (1)) related to the tiles overlapping the viewport are determined. The percentile maximum RGB in the user's viewport is then defined as a weighted average of the percentile maximum RGB values of the overlapping tiles as follows:

$$V_n = \sum_{i=1}^{I} \frac{c_i}{N_v} M_{i,n} \qquad (2)$$

where V_n represents the percentile maximum RGB in the viewport when frame n is being decoded, c_i is the number of pixels overlapping the viewport in tile i, $M_{i,n}$ is the smoothed percentile maximum RGB in tile i, and N_v is the number of pixels in the viewport.

2.2.1. View-dependent tone mapping

Once the percentile maximum RGB in the user's viewport is determined, a tone-mapping operator depending on the user's view direction can be designed. Various tone-mapping operators can be used to fit the dynamic range of the HDR content into the dynamic range of VR player's display [4][6]. However, for the proof of concept we use a simple linear tone-mapping operator for dynamic range conversion as follows.

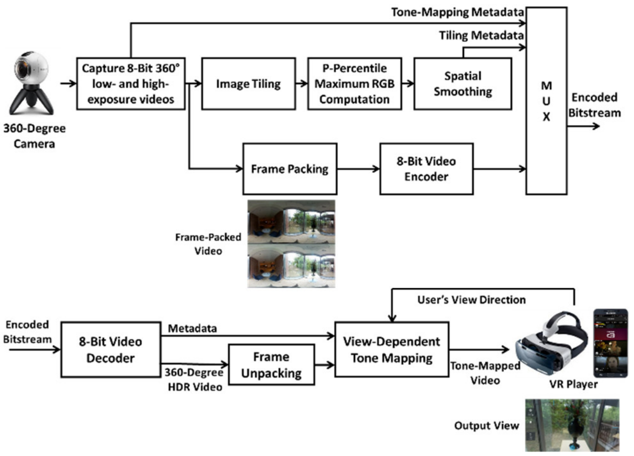


Fig. 6. Block diagram of the transmitter and the receiver in the proposed VR+HDR system with 8-bit pipeline.

$$\hat{\mathrm{C}}_n = \begin{cases} \min(\mathrm{C}_n , D) & \text{if } V_n \le D \\ \min(\frac{D}{V_n} \mathrm{C}_n , D) & \text{if } V_n > D \end{cases} \qquad (3)$$

where C_n is a linear primary color component (i.e. red, green or blue) of a pixel in frame n, $\hat{\mathrm{C}}_n$ is the tone-mapped color component, and D denotes the peak luminance of VR player's display. If a color component exceeds D, it will be clipped to the peak luminance of the player display.

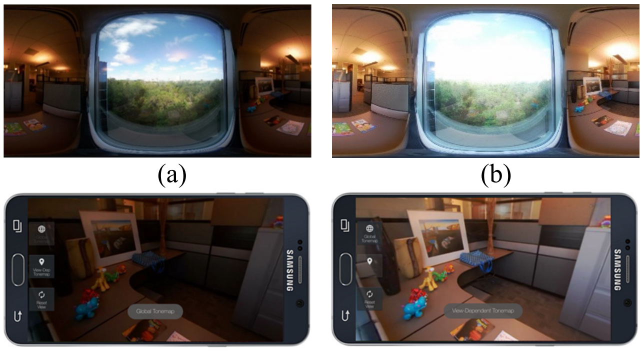
Note that for global tone mapping (with no dependency on the user's view direction), V_n in (3) will be replaced with the percentile maximum RGB across the entire input image/video frame.

3. VR+HDR SYSTEM FOR 8-BIT PIPELINE

The view-dependent tone mapping described in the previous section requires 10-bit functionality (e.g. HDR content, 10-bit video codec). However, if such 10-bit functionality is not supported, this section proposes frame-packing HDR rendering in VR. Figure 6 demonstrates a block diagram of the proposed method.

In the proposed method, one low-exposure and one high-exposure 360-degree images/videos are captured such that the former has a right exposure in bright area while the latter has a right exposure in dark area. These two videos are frame packed in a top-down approach. Then, similar to Section 2, the image is split into non-overlapping tiles and smoothed P-percentile maximum RGB values $M_{i,n}$ are computed for all tiles according to (1) and added to the encoded video bitstream as metadata. The number of tiles is determined based on a compromise between the size of metadata and the smoothness of transition among different viewpoints.

At the receiver, the bitstream is decoded, the metadata and the frame-packed video frames are extracted, and the frames are unpacked. Then, the percentile maximum RGB for the



(a) (b)

(c) (d)

Fig. 7. Test result of SamsungVR Android app; (a), (b) 360-degree low- and high-exposure frames. (c) One viewport using global tone mapping; (d) Same viewport with view-dependent tone mapping.

current viewport is computed from the metadata according to (2). Next, a blend factor to blend the n-th low- and high-exposure frames is calculated according to (4).

$$\beta_n = \frac{\max_i(M_{i,n}) - V_n}{\max_i(M_{i,n}) - \min_i(M_{i,n})} \tag{4}$$

where $\max_i(\cdot)$ and $\min_i(\cdot)$ are the maximum and minimum operators applied over all tiles in the n-th frame. The blend factor is roughly inversely proportional to the brightness in the viewport in the user's current field of view. The higher is the blend factor, a higher weight is assigned to the low-exposure frame and vice versa. The output frame is given by:

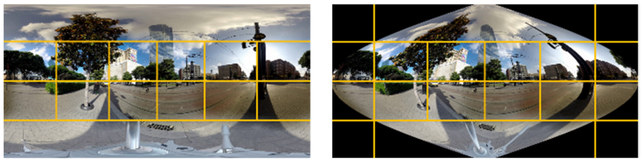
$$\mathrm{C}_{n,out} = \beta_n \, \mathrm{C}_{n,LE} + (1 - \beta_n)\, \mathrm{C}_{n,HE} \tag{5}$$

where $\mathrm{C}_{n,LE}$ and $\mathrm{C}_{n,HE}$ denote the linear primary color component in the low- and high-exposure frames respectively.

4. RESULTS AND CONCLUDING REMARKS

We have evaluated the proposed methods on eight 360-degree test images and two 360-degree video test sequences. The evaluation was based on an informalsubjective quality assessment of the rendered images/videos. More than 20 subjects participated in the subjective assessment and compared the quality of the rendered content using global tone mapping versus view-dependent tone mapping. No quality score such as Mean Opinion Score (MOS) was collected as the purpose of the assessment was simply to identify which rendered content was preferable to the subject.

The subjects unanimously preferred the quality of all the rendered content using view-dependent tone mapping. The proposed method succeeded to retain details in both bright and dark regions of the test sequences. Figures 7(a) and 7(b) show two 360-degree low- and high-exposure frames captured using Samsung Gear360 camera. Figures 7(c) and 7(d) show a viewport of the scene displayed in SamsungVR Android app [7] viewed on Samsung Galaxy Note 5 smartphone using global and view-dependent tone mappings, respectively. Due to the lack of space, only one example image is provided in the paper. A list of test material and more example images can be found in [12]. Note that the metadata for the proposed method can be generated either on-the-fly in the encoder or off-line. As such, the proposed method can be used for both real-time and off-line video processing.



(a) (b)

Fig. 8. Irregular tiling with large tiles at poles; (a) Rectangular projection; (b) Sinusoidal projection; the four black corner regions are not transmitted.

A number of ways that might improve the proposed method are discussed below. Although for the sake of simplicity, the view-dependent tone-mapping method introduced in this paper is based on linear tone mapping, advanced techniques can be used as alternatives to linear tone mapping [3][4][8]-[11].

In the proposed method, the input image is uniformly divided. However, image tiling can be made more efficiently by using irregular tiling to reduce the number of tiles which in turn reduces the size of metadata. For instance, larger tiles can be used at the location of poles, as shown in Figure 8(a). This idea can be combined with using a different projection such as sinusoidal, as depicted in Figure 8(b). However, the cost associated with this approach would be to find the optimal tiling of the input image and send the tiling map.

Another alternative to the regular image tiling would be to calculate percentile maximum RGB values for different view directions and send metadata corresponding to these values to the receiver. At the receiving end, depending on the viewer's view direction the appropriate percentile maximum RGB value can be directly used for tone mapping. The precision of this method depends on the granularity of the view directions for the calculation of percentile maximum RGB values at the transmitter.

Other modifications to the proposed method can be using multiple images (more than two) taken at different exposure levels to be frame-packed for the 8-bit pipeline, joint optimization of tone-mapping design and irregular image tiling, exploiting user's motion trajectory to calculate tone-mapping parameters at any time, etc. Due to the lack of enough space, a detailed description of these modifications and related algorithms cannot be presented in this paper.

5. REFERENCES

[1] V. Sze, M. Budagavi, and G. J. Sullivan (Editors), *High Efficiency Video Coding (HEVC): Algorithms and Architectures*, Springer, 2014.

[2] Samuelsson, J., et al. Suggested new draft text of Conversion and Coding Practices for HDR/WCG Y′CbCr 4:2:0 Video with PQ Transfer Characteristics. Document: JCTVC-Y0046. ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC), Chengdu, October 2016.

[3] Boyce, Jill, et al. Draft high efficiency video coding (HEVC) version 2. Document: JCTVC-R1013. ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC), Sapporo, July 2014.

[4] SMPTE OV 2094-0:2017 - SMPTE Overview Document - Dynamic Metadata for Color Volume Transformation — Overview for the SMPTE ST 2094 Document Suite, 2017.

[5] Requirements for OMAF. Document: N15907. ISO/IEC JTC1/SC29/WG11, San José, February 2016.

[6] Banterle, Francesco, et al., *Advanced high dynamic range imaging: theory and practice*, CRC Press, 2011.

[7] Samsung VR Android app, https://play.google.com/store/apps/details?id=com.samsung.android.video360

[8] R. Mantiuk, S. Daly, and L. Kerofsky, "Display adaptive tone mapping," *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, Aug. 2008.

[9] G. Eilertsen, et al., "Evaluation of Tone Mapping Operators for HDR-Video," *Computer Graphics Forum*, vol. 32. no. 7, pp. 275-284, Oct. 2013.

[10] T. O. Aydin, et al., "Temporally coherent local tone mapping of HDR video," *ACM Transactions on Graphics (TOG),* vol. 33, no. 6, Nov. 2014.

[11] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *ACM Transactions on Graphics (TOG),* vol. 21. no. 3, July 2002.

[12] E. Faramarzi, M. Budagavi, and H. Najaf-Zadeh, VR+HDR demo. Document: m40535. ISO/IEC JTC1/SC29/WG11, Hobart, April 2017.