

Introducing a Public Stereoscopic 3D High Dynamic Range (SHDR) Video Database

Amin Banitalebi-Dehkordi

Received: 6 December 2016 / Revised: 16 January 2017 / Accepted: 18 January 2017 / Published online: 24 January 2017
© 3D Research Center, Kwangwoon University and Springer-Verlag Berlin Heidelberg 2017

Abstract High dynamic range (HDR) displays and cameras are paving their ways through the consumer market at a rapid growth rate. Thanks to TV and camera manufacturers, HDR systems are now becoming available commercially to end users. This is taking place only a few years after the blooming of 3D video technologies. MPEG/ITU are also actively working towards the standardization of these technologies. However, preliminary research efforts in these video technologies are hampered by the lack of sufficient experimental data. In this paper, we introduce a Stereoscopic 3D HDR database of videos that is made publicly available to the research community. We explain the procedure taken to capture, calibrate, and post-process the videos. In addition, we provide insights on potential use-cases, challenges, and research opportunities, implied by the combination of higher dynamic range of the HDR aspect, and depth impression of the 3D aspect.

Keywords High dynamic range video · Stereoscopic video · 3D TV · SHDR database

1 Introduction

Over the past few years industry and research communities have been showing an increasing amount of interest in immersive media technologies. These technologies aim at providing a higher, closer to natural, quality of experience (QoE) to end users. Examples of these technologies include 3D video, high dynamic range (HDR) video, 360° displays, virtual reality (VR), and 4–8 K video. MPEG and ITU are also holding active discussions in regards to the standardization of such types of media content [1]. These efforts received a boost after the introduction of the first draft of high efficiency video coding (HEVC) [2, 3].

Considering that some of these technologies are fairly new, free and public availability of these databases is still a challenge. This paper focuses on two specific state-of-the-art video technologies namely 3D and HDR video.

3D video technologies entered the consumer market a few years back, by introducing stereoscopic video. Stereo video contains two left and right views, which are calibrated and aligned in such a way that provide an impression of depth (through disparity) to viewers. Initial commercialization efforts in the realm of 3D technology included the standardization of stereo and multi-view video. Later many showed interest in glass-free auto-stereo TVs and several models became commercially available. Since then, many 3D movies were produced and displayed at

A. Banitalebi-Dehkordi (✉)
University of British Columbia (UBC), Vancouver, BC,
Canada
e-mail: dehkordi@ece.ubc.ca

theaters. However, there are still only limited number of public databases of original uncompressed 3D video available. Note that 3D movies produced for cinema, are almost all the time converted from 2D. In other words, the movie is initially captured in 2D, but then later 2D shots are manually converted to 3D in post-production studios. That being said, compared to other kinds of immersive video technologies mentioned earlier, there are more 3D databases (mainly stereo video) available [4–6]. Consequently, research in areas of 3D video compression [1, 7–12], quality assessment [13–19], and 3D visual attention modeling [20–22] is very active.

High dynamic range video technologies have been receiving significant recognition over the last couple of years. HDR video provides a wide range of dynamic range, through a close to human visual system (HVS) perceived gamut. As opposed to standard dynamic range (SDR) video that spans over only a contrast range of 100:1–1000:1, HDR video can potentially cover up to 10^5 :1 contrast ratio [23]. Each pixel in HDR video frame stores more than 8 bits of information, often up to 16 bits. This results in ordinary SDR displays not being able to show HDR content. To show HDR content on regular displays a technique known as tone mapping is employed to reduce the dynamic range to 8 bits, in a way that most of the color and brightness information is preserved [24]. In addition, capturing HDR video requires special camera technologies that support capturing of many exposure levels simultaneously [23, 24]. This make it harder to capture and produce freely available HDR databases. As a result there are only very few HDR databases publicly available. As well, research on HDR video compression [25–29], HDR quality assessment [30–32], and HDR visual attention modeling [33] are at preliminary stages.

One of advanced video technologies that is recently being discussed within the research and industry communities is the combination of 3D and HDR video, known as 3D-HDR video. Stereoscopic HDR (SHDR) is a two-view video with each view being captured in HDR. It is worth noting that early efforts in combining 3D and HDR content was shaped around creating synthetic HDR images using multiple views of a same scene [34, 35]. Rufenacht [36] proposes to use side-by-side inexpensive low dynamic range cameras to generate HDR stereo data

after further processing of the two. Selmanovic et al. [37] takes this one step ahead by proposing to use one LDR and one HDR camera to generate stereo HDR data, that was shown in [37] to be very closely similar to the ground truth captured stereo HDR content. However, it was only recently that HDR video cameras became commercially available (and affordable). As a result, none of the mentioned studies used multiple HDR cameras to capture stereo or multi view HDR content. As 3D-HDR video receives more attention, it would be necessary to have standard public databases for performing research activities. To the best of the author's knowledge, there is no public large-scale database of SHDR video available to this date.

This paper introduces a public database of SHDR videos. Stereoscopic videos are captured using two side by side HDR cameras. After that, a process of calibration and rectification is performed to align the videos to ensure a pleasant 3D quality is achieved. Our SHDR video database is available at: <http://ece.ubc.ca/~dehkordi/databases.html>.

The proposed database contains 11 videos from a wide range of brightness, depth, color, temporal, and spatial complexity. The videos are captured and calibrated accurately both at the hardware (physically locking and aligning cameras) and post-processing (disparity correction and frame alignment) levels.

The main contribution of this paper is the introduction of a new 3D-HDR database to the community, to facilitate the investigation of challenges involved with processing and understanding this kind of content. In addition, the creation of the database is explained in reproducible steps so the same approach can be carried out by colleagues to generate new 3D-HDR content from different scenes with different capturing or lighting settings, or with other cameras. It is worth mentioning that this paper does not attempt to solve challenges involved with processing or understanding of 3D-HDR data, but rather it only re-iterates the primary challenges and potential areas of research for interested readers.

Section II introduces our database and explains the procedure to capture and post-process the videos. Section III elaborates on the use cases and challenges involved in 3D-HDR video processing, that exist solely in 3D-HDR video. Section IV concludes the paper.

2 Preparation of Our 3D-HDR Video Database

This section provides a description on the capturing setup, database scene content, as well as post-processing stages involved with the content creation.

2.1 Capturing Configuration

The two RED Scarlet-X cameras used for capturing are mounted in a side-by-side fashion on top of a tripod. As Fig. 1 illustrates, we use an adjustable metal bar to mount the cameras that can support up to 4 side-by-side cameras. The bar, joints, and screws are built to a sub-millimeter precision. Since we only capture two view (stereo) HDR video, we use only two stands on the bar (multi-view HDR video can be captured similarly). The cameras use identical firmware and settings, and are of a same model and build date. A single remote control is used to control both cameras. The two cameras were synchronized using a video Genlock input signal (RS170A Tri-Level Sync Input) [38]. There was about 8 cm horizontal disparity between the centers of the lenses. Each camera records videos at 30 fps and in floating points HDR (18 F-stops). For indoor sequences an artificial light source was used to provide increased dynamic range and brightness over the target objects.

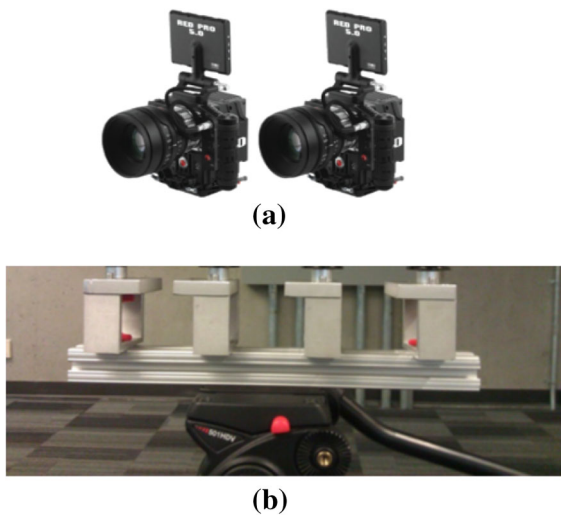


Fig. 1 Capturing with two side-by-side HDR cameras. Cameras are mounted to an adjustable aluminum bar on a tripod

2.2 Capturing Content

Using our capturing setup described in the previous sub-section, over 24 sequences were captured initially. All videos are shot in full HD resolution at 1920×1080 (each view) and 30 fps. From the initially captured sequences, a total of 11 videos were selected for our database, to cover different ranges of scene content, lighting conditions, and pleasant 3D quality. Figure 2 demonstrates a snapshot of the videos and Table 1 provides specifications on the sequences. Note that the snapshots in Fig. 2 are only from one view of each video. Also, in order to be able to show the snapshots here, we converted floating point frames to 8 bits through tone-mapping [24].

Generally the selected videos provide an impression of depth, as well as an impression of a higher dynamic range, that comes with not just a brighter video, but also more information at different exposures. Each video has about 10 s length which makes it possible to use this database for various types of subjective experiments.

As it is observed from Table 1, videos cover a different ranges of spatial and temporal complexity. Spatial complexity, measured as Spatial Information (SI), is a measure of spatial complexity of a scene [39]. SI is calculated by applying a Sobel filter to each frame (luminance plane) to detect edges first. Then standard deviation on the edge map provides a measure of spatial complexity of the associated frame. Maximum standard deviation value over all frames results in the SI value of the sequence. Temporal Information (TI) on the other hand, measures temporal complexity, and is based on motion among consecutive frames [39]. To measure the TI, first the difference between the pixel values (of the luminance plane) at the same coordinates in consecutive frames is calculated. Then, the standard deviation over pixels in each frame is computed and the maximum value over all the frames is set as the measure of TI. Note that SI and TI were calculated over the entire dynamic range of the SHDR videos, after they were converted to 12 bits (more details later in this Section). Therefore, the SI and TI metrics measure the complexity in spatial and temporal domain in the brightness intensity units as digital numbers and thus can be considered unit-less. Figure 3 shows the SI and TI distribution over the proposed video database.

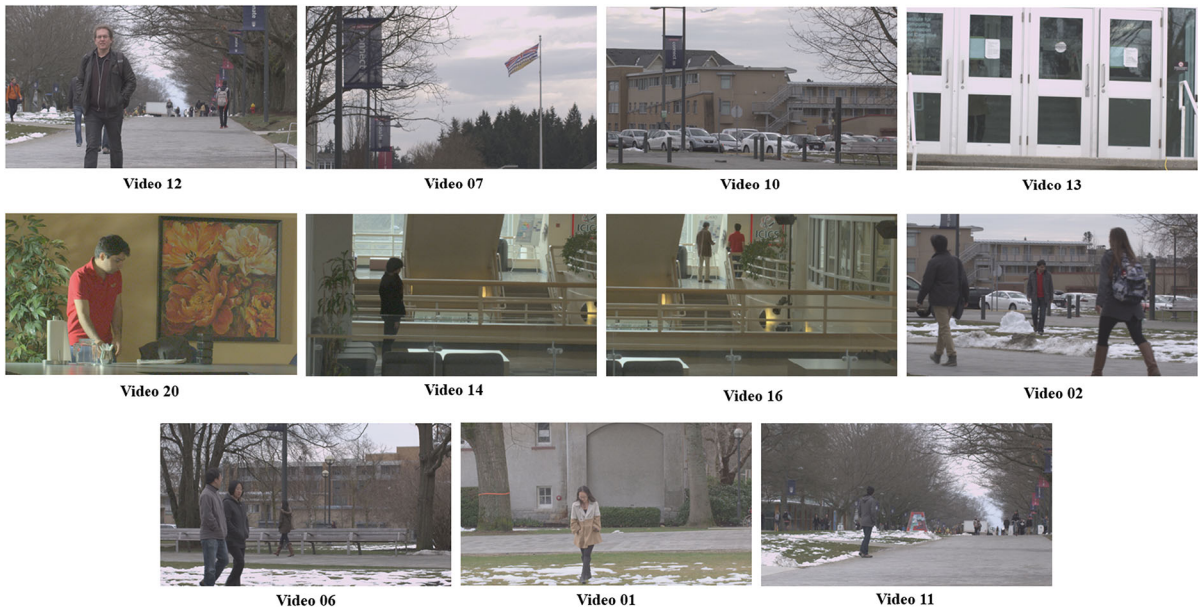


Fig. 2 Snapshots of tone-mapped video frames. Only right view frames are shown here

Table 1 Specifications of our 3D-HDR video database

Sequence	Resolution	Frame rate (fps)	Number of frames	Spatial complexity (spatial information)	Temporal complexity (temporal information)	Depth bracket
Video_01	1920 × 1080	30	300	Low (26.8121)	Low (5.8977)	Medium
Video_02	1920 × 1080	30	510	Medium (32.5871)	High (26.5389)	Wide
Video_06	1920 × 1080	30	300	High (49.3520)	High (20.8599)	Wide
Video_07	1920 × 1080	30	300	High (50.7216)	Low (5.3762)	Wide
Video_10	1920 × 1080	30	460	Medium (34.6945)	Low (2.4159)	Wide
Video_11	1920 × 1080	30	440	High (45.0262)	Low (7.5869)	Wide
Video_12	1920 × 1080	30	300	Medium (35.9981)	Medium (12.4612)	Wide
Video_13	1920 × 1080	30	400	Medium (39.1949)	Medium (11.2539)	Medium
Video_14	1920 × 1080	30	400	Low (25.5698)	Low (5.5410)	Medium
Video_16	1920 × 1080	30	300	Low (23.3037)	Medium (9.7458)	Medium
Video_20	1920 × 1080	30	170	Low (23.3551)	Medium (10.6995)	Narrow

Table 1 also provides information about “depth bracket” of each scene. The depth bracket for each scene is defined as the amount of 3D space used in a shot (or a sequence). In other words, it is a rough estimate of difference between the distance of the closest and the farthest visually important objects from the camera in each scene [40].

When capturing, we tried to avoid any 3D window violation. Window violation (in 3D) occurs when objects only partially appear in a scene. As a result,

poor 3D quality (due to conflict between depth cues) is expected around the edges of the 3D display.

The captured sequences are first stored in “.hdr”, floating point format, and in linear RGB space. However, to make these videos useful for 3D and HDR video compression studies, they were converted to 12 bit “4:2:0” raw format according to ITU-R Rec. BT.709 color conversion primaries [41] (the HEVC and H.264/AVC video encoders currently only support the compression of video in “.yuv” raw file

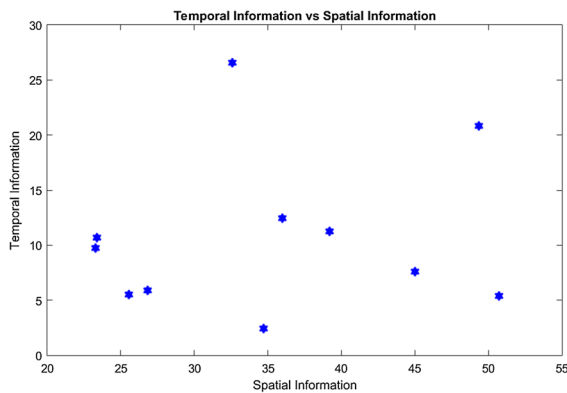


Fig. 3 Spatial and temporal information associated to the SHDR video database

format). To this end, the RGB values were first normalized to [0, 1]. Then, they were converted to Y-Cb-Cr color space in accordance to ITU-R Rec. BT.709 [41]. At last, using the separable filter values provided by [42], chroma-subsampling was applied to the Y-Cb-Cr values to linearly quantize the resulted signals to the desired bit depth (i.e., 12 bits) [43]. Both the raw “.hdr” stored format, and the “.yuv” format will be made publicly available along with this paper.

2.3 Post-processing

Several stages of post-processing were applied to the captured sequences to ensure they have a high 3D and HDR quality and are comfortable to watch.

2.3.1 Temporal Synchronization

Although a same remote control was used, and camera firmware versions were exactly the same, but due to very small hardware differences, it is always possible that the two left and right cameras are out of sync, even for a couple of frames. This makes more sense considering that video frame rate for each view is 30 fps, so even if there is 1/30th of a second difference, that can cause temporal mismatch of one frame.

Two ensure temporal synchronization, we manually marked objects of interest, and compared them in the two views. Wherever needed, a few frames were taken out to make sure the views are perfectly synchronized.

2.3.2 3D Alignment

It was mentioned previously that our camera configuration hardware was built up to a sub-millimeter precision. However, since cameras are manually mounted (and taken apart after), there is a small possibility that there is very minor vertical misalignment in the configuration. Any vertical mismatch can make the 3D content uncomfortable to watch.

To remove any vertical mismatch, we extract SIFT [44] features of the two views, and match them against each other. Top 10% of the matches are selected as reliable matches, and average height difference between those is an amount that is cropped from one of the views. This provides vertical alignment and stability of frames [45].

2.3.3 Disparity Correction

Using two side-by-side cameras results in negative parallax as objects converge in infinity (objects pop out of the display). This can cause visual discomfort to the viewers as subjective experiments have shown viewers show preference in watching objects appearing behind the display (positive parallax), as opposed to popping out of the display.

To correct for the negative parallax, we crop from the left and right views, in a way that object of interest appears on the screen, and rest of objects appear behind the screen. More details regarding this correction can be found in [45].

3 Challenges, Use Cases, and Potential Research Opportunities in Processing 3D-HDR Video

After post-processing, our SHDR video database is ready for use. That being said, there are multiple challenges with 3D-HDR data and requires special considerations. This section iterates some of these challenges.

3.1 Tone Mapping and Dynamic Range Compression

Tone mapping of HDR image content is a challenge, and still a very hot area of interest to research community. In the case of HDR video, it would be even more challenging as temporal coherency of tone-

mapped video is crucial. Figure 4 shows two difference scenes when they are over-exposed, under-exposed, and tone-mapped. It is observed from this figure, that “clouds” or “tree colors” can be hidden in the first two cases for the top row of this figure. In the bottom row of the figure, the “airplane” or building details can get lost in the two extremes of exposure. This figure shows the importance of an accurate tone-mapping technique.

In the case of 3D-HDR video, tone-mapping will be even more challenging, as not only temporal coherency is essential, but also inter-view coherency has to be assured. While human eyes can tolerate a minimal degree of difference between the tones of the two views due to binocular suppression, drastic tone differences between the views can severely degrade the overall video quality.

3.2 3D Specific Challenges

Depth map extraction can potentially be a different task for stereoscopic HDR video compared to that of stereo SDR video. Most of the existing state-of-the-art depth synthesis methods are designed or tested against only SDR video. Therefore, depth estimation from stereo HDR video has room for further investigations. There will be questions such as: Do we need to tone map the SHDR video first, and then generate depth maps? Or do we generate depth maps from floating point values stored in HDR files? How should one

consider dynamic range locally or globally when estimating a depth map? Figure 5 shows sample depth maps from the presented stereo HDR videos. For demonstration purposes, depth maps are generated by tone-mapping frames using the MPEG Depth Estimation Reference Software (DERS) [46, 47].

Another challenge with 3D-HDR video is handling crosstalk (of 3D aspect) and ghosting (of HDR aspect). In 3D video, sometimes objects appear as double shadowed, that is called crosstalk, and is mainly resulted from the leakage of one eye view to the other eye due to imperfections of the 3D display or glasses. In HDR video, on the other hand, fast moving objects may appear blurry, because each frame is resulted from combining over a dozen F-stops within which a moving object may have different positions. In 3D-HDR domain, reducing the crosstalk and ghosting together will likely be a much more difficult task as they may both target a same object.

3.3 Other Challenges

One other major challenge in the processing of 3D-HDR video is encoding such type of video content. MPEG/ITU are developing video compression standards for 3D or HDR video, but standards for compression of 3D-HDR video are yet to come.

Quality assessment of 3D-HDR video also opens new doors for potential introduction of new quality metrics that are dynamic range independent [30], and

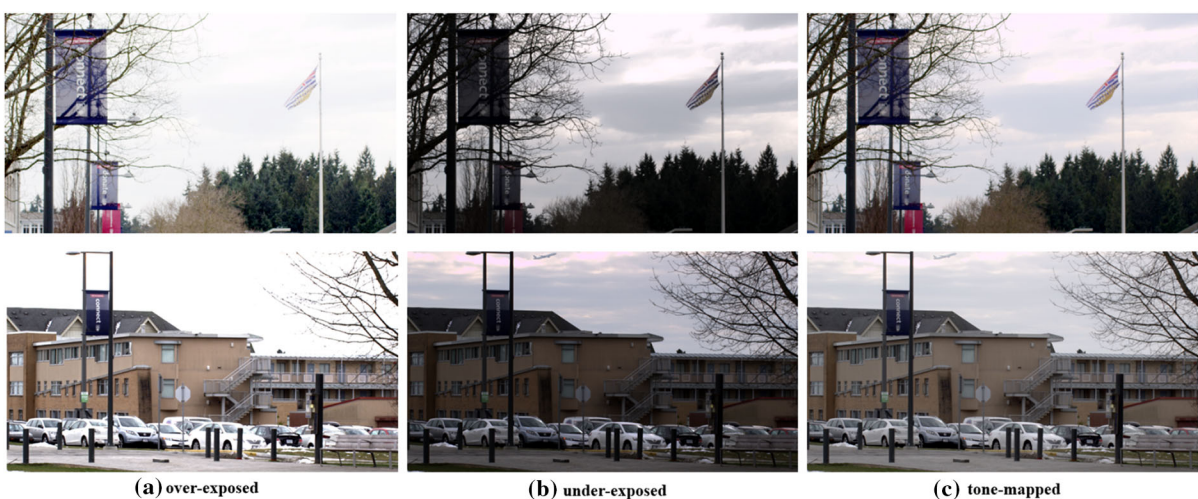


Fig. 4 Capturing with different exposure levels: **a** over-exposed, **b** under-exposed, and **c** tone-mapped. Note that tree colors are not visible in **b-top**, and airplane is not visible in **a-bottom**. Only right view frame is shown here



Fig. 5 Synthesized depth map samples

also consider binocular properties of human vision [13–16].

Last but not the least, visual attention modeling in 3D-HDR video needs to be studied, as it will be different from that of either 3D [20, 21] or HDR [30, 33] video.

4 Conclusion

With the research and industry communities showing an increasing amount of interest in advanced immersive media technologies, it is essential to have publicly available database to facilitate studying such media systems. This paper introduces a Stereoscopic 3D HDR video database made of 11 stereo HDR sequences from scenes with different characteristics. The capturing configuration and database characteristics are described first. Then post-processing stages were explained in details. It is important to provide reproducible steps so that interested readers can carry out the same approach to create similar databases from different scenes or under different capturing settings. At the end, some challenges and potential future research opportunities were discussed. This database is made publicly available to the research community.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. MPEG document repository. (2017). <http://phenix.int-evry.fr/jct/index.php>
2. Sullivan, G. J., et al. (2012). Overview of the high efficiency video coding (HEVC) standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12), 1649–1668. doi:10.1109/TCSVT.2012.2221191.
3. Ohm, J. R., et al. (2012). Comparison of the coding efficiency of video coding standards—including high efficiency video coding (HEVC). *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12), 1669–1684. doi:10.1109/TCSVT.2012.2221192.
4. Multimedia Signal Processing Group at EPFL. (2017). <http://mmspg.epfl.ch/downloads>
5. Winkler, S. (2017). *Image and video quality resources*. <http://stefan.winklerbros.net/resources.html>
6. IRCCyN lab at Institut de Recherche en Communications et Cybernétique de Nantes. (2017). <http://ivc.univ-nantes.fr/en/>
7. Smolic, A., et al. (2007). Coding algorithms for 3DTV—A survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(11), 1606–1621.
8. Muller, K., et al. (2013). 3D high-efficiency video coding for multi-view video and depth data. *IEEE Transactions on Image Processing*, 22(9), 3366–3378. doi:10.1109/TIP.2013.2264820.
9. Sullivan, G. J., et al. (2013). Standardized extensions of high efficiency video coding (HEVC). *IEEE Journal of Selected Topics in Signal Processing*, 7(6), 1001–1016. doi:10.1109/JSTSP.2013.2283657.
10. Hannuksela, M., et al. (2013). Multiview-video-plus-depth coding based on the advanced video coding standard. *IEEE Transactions on Image Processing*, 22(9), 3449–3458. doi:10.1109/TIP.2013.2269274.
11. Jiang, L., He, J., Zhang, N., et al. (2010). An overview of 3D video representation and coding. *3D Research*, 1, 43. doi:10.1007/3DRes.01(2010)6.

12. Rusanovskyy, D., Hannuksela, M. M., & Su, W. (2013). Depth-based coding of MVD data for 3D video extension of H.264/AVC. *3D Research*, 4, 6. doi:[10.1007/3DRes.02\(2013\)6](https://doi.org/10.1007/3DRes.02(2013)6).
13. Banitalebi-Dehkordi, A., Pourazad, M. T., & Nasiopoulos, P. (2012). A human visual system based 3D video quality metric. In *2nd international conference on 3D imaging, IC3D*, December 2012, Belgium.
14. Banitalebi-Dehkordi, A., Pourazad, M. T., & Nasiopoulos, P. (2015). An efficient human visual system based quality metric for 3D video. *Springer Journal of Multimedia Tools and Applications*, 75(8), 4187–4215. doi:[10.1007/s11042-015-2466-z](https://doi.org/10.1007/s11042-015-2466-z).
15. Banitalebi-Dehkordi, A., Pourazad, M. T., & Nasiopoulos, P. (2013). 3D video quality metric for 3D video compression. 11th IEEE IVMSWP workshop: 3D Image/Video Technologies and Applications, June 2013, Seoul, Korea.
16. Banitalebi-Dehkordi, A., Pourazad, M. T., & Nasiopoulos, P. (2013). A study on the relationship between the depth map quality and the overall 3D video quality of experience. In *International 3DTV conference: vision beyond depth*, October 2013, Scotland, UK.
17. Hewage, C., et al. (2009). Quality evaluation of color plus depth map-based stereoscopic video. *IEEE Journal of Selected Topics in Signal Processing*, 3(2), 304–318. doi:[10.1109/JSTSP.2009.2014805](https://doi.org/10.1109/JSTSP.2009.2014805).
18. Shao, F., et al. (2013). Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics. *IEEE Transactions on Image Processing*, 22(5), 1940–1953. doi:[10.1109/TIP.2013.2240003](https://doi.org/10.1109/TIP.2013.2240003).
19. Zhang, W., et al. (2016). Using saliency-weighted disparity statistics for objective visual comfort assessment of stereoscopic images. *3D Research*, 7, 17. doi:[10.1007/s13319-016-0079-6](https://doi.org/10.1007/s13319-016-0079-6).
20. Banitalebi-Dehkordi, A., Nasiopoulos, E., Pourazad, M. T., & Nasiopoulos, P. (2017). Benchmark three-dimensional eye-tracking dataset for visual saliency prediction on stereoscopic three-dimensional video. *SPIE Journal of Electronic Imaging*, 25(1), 013008. doi:[10.1117/1.JEI.25.1.013008](https://doi.org/10.1117/1.JEI.25.1.013008). <http://ece.ubc.ca/~dehkordi/databases.html>
21. Banitalebi-Dehkordi, A., Pourazad, M. T., & Nasiopoulos, P. (2016). A learning-based visual saliency prediction model for stereoscopic 3D video (LBVS-3D). *Multimedia Tools and Applications*. doi:[10.1007/s11042-016-4155-y](https://doi.org/10.1007/s11042-016-4155-y).
22. Chagnon-Forget, M., Rouhafzay, G., Cretu, A. M., et al. (2016). Enhanced visual-attention model for perceptually improved 3D object modeling in virtual environments. *3D Research*, 7, 30. doi:[10.1007/s13319-016-0106-7](https://doi.org/10.1007/s13319-016-0106-7).
23. Ferwerda, J. A. (2001). Elements of early vision for computer graphics. *Computer Graphics and Applications*, 21(5), 22–33.
24. Salih, Y., et al. (2012). Tone mapping of HDR images: A review. In *4th international conference on intelligent and advanced systems (ICIAS)*, 2012.
25. Azimi, M., Banitalebi-Dehkordi, A., Dong, Y., Pourazad, M. T., & Nasiopoulos, P. (2014). Evaluating the performance of existing full-reference quality metrics on high dynamic range (HDR) video content. In *ICMSP 2014: XII international conference on multimedia signal processing*, November 2014, Venice, Italy.
26. Banitalebi-Dehkordi, A., Azimi, M., Pourazad, M. T., & Nasiopoulos, P. (2014). Compression of high dynamic range video using the HEVC and H.264/AVC standards. In *2014 10th international conference on heterogeneous networking for quality, reliability, security and robustness (QShine)*, Rhodes Island, Greece, August 2014 (invited paper).
27. Yu, Sh., et al. (2016). Adaptive PQ: Adaptive perceptual quantizer for HEVC main 10 profile-based HDR video coding. In *2016 visual communications and image processing (VCIP)* (pp. 1–4). doi:[10.1109/VCIP.2016.7805499](https://doi.org/10.1109/VCIP.2016.7805499)
28. Jung, Ch., et al. (2016). HEVC encoder optimization for HDR video coding based on perceptual block merging. *Visual Communications and Image Processing (VCIP)*. doi:[10.1109/VCIP.2016.7805536](https://doi.org/10.1109/VCIP.2016.7805536).
29. Bouzidi, I., et al. (2016). On the selection of residual formula for HDR video coding. In *2016 6th European workshop on visual information processing (EUVIP)* (pp. 1–5). doi:[10.1109/EUVIP.2016.7764590](https://doi.org/10.1109/EUVIP.2016.7764590)
30. Banitalebi-Dehkordi, A., Azimi, M., Pourazad, M. T., & Nasiopoulos, P. (2016). Visual saliency aided high dynamic range (HDR) video quality metrics. In *International conference on communications (ICC)*, 2016.
31. Korshunov, P., et al. (2015). Subjective quality assessment database of HDR images compressed with JPEG XT. In *2015 seventh international workshop on quality of multimedia experience (QoMEX)* (pp. 1–6). doi:[10.1109/QoMEX.2015.7148119](https://doi.org/10.1109/QoMEX.2015.7148119)
32. Mantel, C., et al. (2014). Comparing subjective and objective quality assessment of HDR images compressed with JPEG-Xt. In *2014 IEEE 16th international workshop on multimedia signal processing (MMSP)* (pp. 1–6). doi:[10.1109/MMSP.2014.6958833](https://doi.org/10.1109/MMSP.2014.6958833)
33. Banitalebi-Dehkordi, A., Dong, Y., Pourazad, M. T., & Nasiopoulos, P. (2015). A learning based visual saliency fusion model for high dynamic range video (LBVS-HDR). In *23rd European signal processing conference (EUSIPCO)*, 2015.
34. Vavilin, A., & Jo, K.-H. (2011). Fast HDR image generation from multi-exposed multiple-view LDR images. In *3rd European workshop on visual information processing (EUVIP)*, July 2011.
35. Sun, N., Mansour, H., & Ward, R. (2010). HDR image construction from multi-exposed stereo LDR images. In *17th international conference on image processing (ICIP)*, September 2010.
36. Rufenacht, D. (2011). *Stereoscopic high dynamic range video*. Master Thesis, EPFL, August 2011.
37. Selmanovic, E., et al. (2014). Enabling stereoscopic high dynamic range video. *Signal Processing: Image Communication*, 29(2), 216–228 (Special Issue on Advances in High Dynamic Range Video Research).
38. RED Scarlet-X Operation Guide. (2017). <https://red.com>
39. Recommendation ITU P.910. (1999). Subjective video quality assessment methods for multimedia applications, ITU.
40. Xu, D., Coria, L. E., & Nasiopoulos, P. (2012). Guidelines for an improved quality of experience in 3D TV and 3D mobile displays. *Journal of the Society for Information Display*, 20(7), 397–407. doi:[10.1002/jsid.99](https://doi.org/10.1002/jsid.99).

41. Recommendation ITU-R BT.709-5. (2002). Parameter values for the HDTV standards for production and international programme exchange.
42. Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG. (2005). New test sequences in the VIPER 10-bit HD data. JVTQ090, 2005.
43. Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG. (2007). Donation of tone mapped image sequences. JVT-Y072, October, 2007.
44. Lowe, D. G. (1999) Object recognition from local scale invariant features. In *Proceedings of the international conference on computer vision* (Vol. 2, pp. 1150–1157).
45. Banitalebi-Dehkordi, A., Pourazad, M. T., & Nasiopoulos, P. (2015). The effect of frame rate on 3D video quality and bitrate. *Springer Journal of 3D Research*, 6(1), 5–34. doi:[10.1007/s13319-014-0034-3](https://doi.org/10.1007/s13319-014-0034-3).
46. Wgner, K., & Stankiewicz, K. (2014). DERS software manual. ISO/IEC JTC1/SC29/WG11 MPEG2014/M34302, July 2014, Sapporo, Japan.
47. Tanimoto, M., Fujii, T., & Suzuki, K. (2009). Video depth estimation reference software (DERS) with image segmentation and block matching. ISO/IEC JTC1/SC29/WG11 MPEG/M16092, 2009.