

A NOVEL APPROACH FOR MULTI-VIEW 3D HDR CONTENT GENERATION BASED ON DIBR & DEPTH ADAPTIVE TONE MAPPING

Mansi Sharma¹, M. S. Venkatesh², Rohan Lal³

Indian Institute of Technology Madras^{1,3}, Indian Institute of Technology Delhi²

ABSTRACT

In this work, we propose a novel DIBR scheme for stereo HDR imaging and 3D display. The proposed scheme endows a new depth-adaptive cross-trilateral filter (DA-CTF) for recovering High Dynamic Range (HDR) images from multiple Low Dynamic Range (LDR) images captured at different exposure levels. Explicitly leveraging additional depth information in the tone mapping operation correctly identify global contrast change and detail visibility change by preserving the edges and reducing halo artifacts in the synthesized 3D views by depth-image-based rendering (DIBR) procedure. The experiments show that the proposed DIBR scheme and DA-CTF outperform state-of-the-art operators in the enhanced depiction of tone mapped synthesized HDR stereo images on LDR displays.

Index Terms — Multi-view multi-exposure sequence, tone mapping, stereo HDR video, halo artifacts, DIBR, 3DTV

1. INTRODUCTION

A multiple-camera setup allows us to acquire the scene from different camera viewpoints and at different exposure levels. Acquired multi-view images or video may typically have a limited dynamic range, which yields under or over-exposed regions. It is not practical, for example, to capture the details in sunny areas and in dark shadow areas altogether in a single shot from a multi-camera setup [1, 2, 3, 4]. The regular way of composition multi-camera LDR images to create a high quality HDR image will not always give desirable results due to viewpoint variation [3, 5, 9]. The problems of 3D HDR multi-view image synthesis and tone-mapping are well-suited for the new capturing and 3D display environment, and have not been widely addressed as an integrated system [5, 6, 7, 8]. The objective of this work is to develop a non-standard approach that offers both HDR effect and stereo personalization in an MPEG standardized Multi-view video-plus-depth (MVD) representation format for 3D display technologies. The major contributions of this paper are:

- A novel flexible end-to-end production pipeline incubating DIBR for the restoration of multi-view HDR content generation and view synthesis from the set of multi-view LDR camera-captured images.
- A new tone reproduction operator based on a depth-adaptive cross-trilateral filter, explicitly leveraging scene geometry for preserving detail visibility and mitigating undesirable halo artifacts.

2. RELATED WORKS

The conventional HDR restoration techniques in 2D settings are particularly designed for recovering radiance maps from differently exposed photographs acquired from a single camera output

and common viewpoint. A wider state-of-the-art overview of 2D HDR imaging algorithms is given in [2, 4]. However, 3D HDR is only just beginning to emerge [5]. Recent trends in computational approaches for 3D video creation using camera arrays or multi-camera dome are gaining commercial acceptance [3, 6]. A very few studies hypothesizes HDR depth illusion and real binocular depth cues [5, 8]. There is still considerable scope of expansion for multi-camera systems. Seshadrinathan and Nestares [3] adopt a popular exposure fusion approach for merging input synchronized videos and altering appearance of the image. Their approach accounts for scene disparity in HDR image generation, however, lack of radiometric calibration, higher bit-depth representations and tone mapping specifically restricts to rendering using short baseline or closely placed cameras. Moreover, it demands sophisticated alignment algorithm that minimizes visual artifacts while accommodating disparity errors.

Our objective is to create a generalized 3D HDR imaging model which does not enforce any constraints on the camera shooting geometries. Thus, we explicitly leverage rich MVD signal representation [6, 13, 15], where each camera viewpoint of the multi-camera systems provides some additional information about the scene. In addition, the per viewpoint depth signals could be useful for providing a dense correspondence of pixels between views and virtual viewpoint rendering. Previously, such aspects of MVD and DIBR have not been analysed for the construction of 3D HDR video generation. Besides, the proposed system is endowed with a depth-adaptive 3D tone mapping operator. A comparative overview of popular 2D tone-mapping algorithms is found in [7]. However, it is critical to note that existing tone mapping methods [7, 10, 11, 12, 16, 17] function only in the intensity domain. To the best of our knowledge, no state-of-the-art tone reproduction operators have been introduced that exploits intensity and per-pixel depth information to better recover minute details of the scene and attenuate undesirable halo artifacts.

3. PROPOSED MULTI-VIEW HDR CONTENT CREATION SCHEME

The novel DIBR scheme for rendering 3D HDR images from multi-view LDR images taken from different camera viewpoints and with different exposures is depicted in Fig. 1. The proposed multi-view HDR imaging model is a three step process comprising of the following: 1) synthesize novel views at the virtual camera viewpoint using depth-image-based rendering procedure, 2) recover camera response function and the radiance map from multi-view multi-exposure LDR images, 3) depth-adaptive tone reproduction of the HDR radiance maps for LDR displays.

Suppose there are N views I_1, \dots, I_N and their corresponding associated depth maps D_1, \dots, D_N respectively. Since scene is acquired from different camera viewpoints under different exposure/illumination conditions, the objective is to recover camera response function from multi-view multi-exposure LDR images.

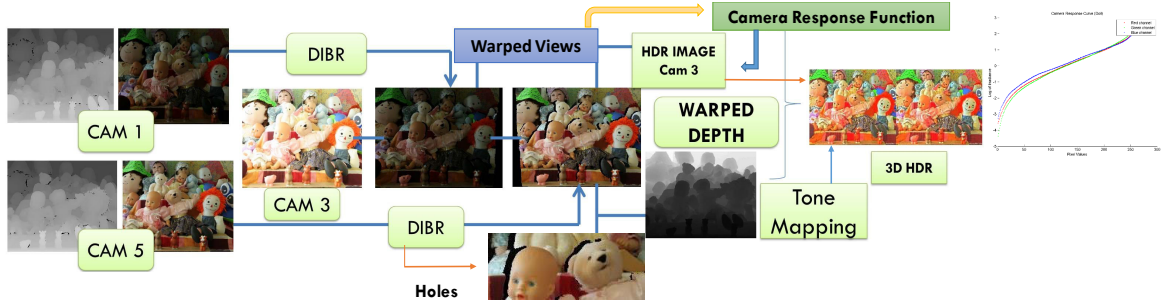


Figure 1. Generalised workflow of generating HDR tone mapped image from a novel viewpoint using multi-view LDR images. Here a novel 3D HDR view is rendered from camera viewpoint 3 using reference cameras 1 and 5. **Note: scheme works for any number of input images and cameras.**

However, the existing methods for recovering camera response function rely on the sequence of multi-exposure images acquired from the same camera viewpoint [4]. In case of multi-camera capturing, such a sequence is not available. Thus, we apply DIBR algorithm [1] to warp differently exposed multiple views onto a common camera viewpoint using depth maps. The holes appear in warped images due to disocclusion and resampling problems associated with 3D warping [15], as shown in Fig. 1. There is no hole filling operation performed after warping. Only actual intensity samples are considered from the warped images for calculation of the camera response function. Further, Debevec and Malik [9] algorithm is modified for multi-camera scenario to obtain camera response function from warped views with radiometric differences. The procedure is made robust to compensate for certain radiometric differences or noise while measuring the similarity of image locations and handling inconsistencies in depth maps.

4. CONSISTENT CAMERA RESPONSE RECOVERY

We modified Debevec and Malik method [9] to work in multi-camera settings. The camera response is recovered from virtual synthesized image sequence. A robust zero mean normalised cross correlation measure is adopted for accurate matching among multi-exposed LDR image sequence. The film reciprocity equation is computed using warped images, given domain of corresponding pixels $Z = \{Z_{ij}\}$ over different known exposure times δ_j

$$g(Z_{ij}) = \ln(E_i) + \ln(\delta_j) \quad (1)$$

where, i is a spatial index over pixels, j indexes over different known exposure times δ_j , $j = 1, \dots, L$, \ln is natural logarithm. The function g is defined as $g = \ln f^{-1}$, where f is the camera response function which is assumed to be monotonic, hence its inverse f^{-1} is well defined. The objective is to recover g and irradiances E_i at each pixel i that best satisfy the constraint (1). The problem is formulated as one of finding $Z_{max} - Z_{min} + 1$ values of $g(Z)$ and C_N values of $\ln(E_i)$ that minimizes the quadratic objective function.

$$O = \sum_{i=1}^{C_N} \sum_{j=1}^L (w(z)(g(Z_{ij}) - \ln(E_i) - \ln(\delta_j)))^2 + \lambda \sum_{z=Z_{min}+1}^{Z_{max}-1} (w(z)g''(z))^2 \quad (2)$$

where, Z_{max} and Z_{min} be the greatest and least pixel values in domain Z , C_N be the number of pixel locations, and $g'' = g(z-1) - 2g(z) + g(z+1)$. The first term in (2) ensures the condition (1) is satisfied while the second term ensures the smoothness of function g . The readers refer to [9] for additional details about

the objective function O . The function is quadratic in the E_i 's and $g(z)$'s. Thus, the minimization of O is a simple linear least squares problem. The solution of an overdetermined system of linear equations is robustly obtained by applying standard singular value decomposition (SVD) method [9]. The solution further improves by imposing some additional constraints:

- The constraint $g(Z_{mid}) = 0$ and $Z_{mid} = \frac{Z_{min} + Z_{max}}{2}$ are imposed to take care of scale factor. The objective function (2) and the system of equations (1) would remain unchanged, if each log irradiance value $\ln(E_i)$ be replaced by $\ln(E_i) + \alpha$ and the function g be replaced by $g + \alpha$, α is the scale factor. This constraint is imposed to establish a scale factor. It implies a pixel with value midway be assumed to have the unit exposure.
- Apply a weighting function $w(z)$ as described in [9] to emphasize the smooth fitting of the curve, *i.e.*

$$w(z) = \begin{cases} z - Z_{min} & \text{if } z \leq Z_{mid} \\ Z_{max} - z & \text{if } z > Z_{mid} \end{cases} \quad (3)$$

- To recover a robust solution of an overdetermined system of equations, the chosen samples must encompass a range of pixel intensities from multiple camera images. Thus, sampling in low intensity regions is recommended. The saturated pixels are left out to obtain an exact solution. However, in the multi-view case, there is another constraint. The holes appear in warped images due to disocclusion and resampling issues (Fig. 1). These holes should not be considered as image samples. We constructed a mask,

$$Mask_{(u,v)}^i = \begin{cases} 0 & \text{if } D_{(u,v)}^i \text{ is unknown} \\ 1 & \text{otherwise,} \end{cases} \quad (4)$$

to remove the pixels whose depth values are not defined in any of the warped images. The warping errors due to inconsistencies in depth maps are reduced by taking twice the number of samples. The objective function (2) is minimized using SVD for each color channel separately to recover the $g(z)$. A more robust estimate of the response curve is recovered by averaging over all recovered response functions.

The irradiance values are recovered after obtaining $g(z)$ by averaging all the images using a weighting function [9].

$$\ln(E_{ij}) = \frac{\sum_{j=1}^L (w(z_{ij})(g(z_{ij}) - \ln(\delta t_j))}{\sum_{j=1}^L (w(z_{ij}))} \quad (5)$$

To obtain the irradiance values in present multi-view case of an image j , all L images of the set are warped onto j and then the averaging is performed. There might be some missing values (*i.e.*, holes) present in the warped images. In such cases, the irradiance

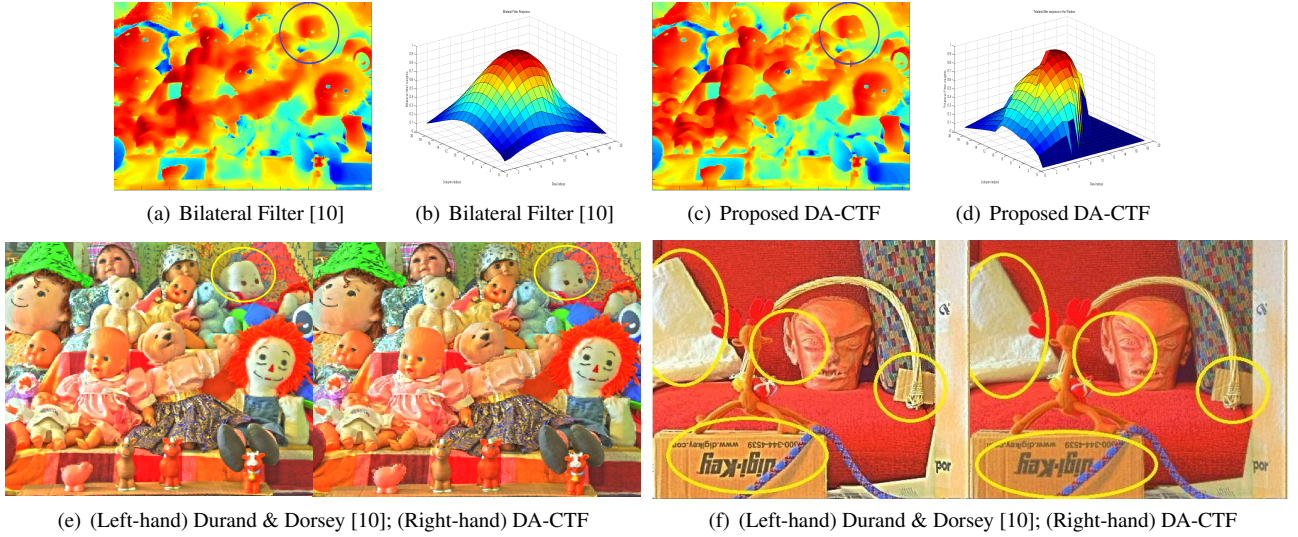


Figure 2. Extracted base layers of “Dolls” data using (a) bilateral filter [10] and (c) proposed DA-CTF filter. (b,d) Response of filter’s around pixel (36, 369) on “Dolls” data. (e,f) Synthesized tone mapped 3D HDR views at camera viewpoint 3. The halo artifacts are clearly reduced by proposed DA-CTF tone mapping approach (see inside of the highlighted regions). **Note: some minor artifacts appear in rendered 3D views because no hole filling is performed in DIBR. Our main objective here is to present the effect of proposed depth-adaptive tone mapping operation. Visit project page for video results [14]. Here only left views of rendered stereo pairs are shown.**

values are selected corresponding to the j^{th} image only. The response curves determined for each RGB channel of synthesized “Dolls” sequence is shown in Fig. 1. The exact benefit of this extension is to make our 3D HDR content production scheme more flexible for image-based rendering and compatible to adopt any existing 2D HDR radiance recovery methods.

5. NOVEL DEPTH ADAPTIVE CROSS-TRILATERAL FILTER FOR TONE MAPPING

The HDR map of each image of synthesized sequence is calculated after the recovery of camera response function. An another major novelty of our system lies in the tone mapping process for the realistic display of 3D HDR images, as depicted in Fig. 1. Many tone mapping methods have been developed over the years [7, 10, 11, 12, 16, 17]. A non-linear bilateral filter is presented in [10] to decompose an image into a base layer and a detail layer. Prasun and Tumblin [11] uses a trilateral filter and Farbman et al. [12] apply a weighted least square optimization framework to recover the base layer. In these methods, the base layer is subtracted from the image to obtain the detail layer. Further, the base layer is compressed and added to the detail layer to recover a high contrast image. However, it is critical to note that existing tone mapping methods function in the intensity domain [7]. Our framework considers a novel exploration of intensity and per-pixel depth information in tone mapping. The proposed depth adaptive cross-trilateral filter (**DA-CTF**) image is obtained as

$$I_{filt}^C(u, v) = \frac{\sum_{(\Delta u, \Delta v) \in W} C_{\Delta u, \Delta v}^{TF} * I_{\Delta u, \Delta v}^{log}}{\sum_{(\Delta u, \Delta v) \in W} C_{\Delta u, \Delta v}^{TF}} \quad (6)$$

where, I_{filt}^C represents the base image and proposed **DA-CTF** is defined as

$$C_{\Delta u, \Delta v}^{TF} = S_{\delta u, \delta v} * L(I_{u, v}^{log}, I_{\Delta u, \Delta v}^{log}) * K(D_{u, v}, D_{\Delta u, \Delta v})$$

The I^{log} represents logarithm of RGB image, D represents the depth map, (u, v) represents the pixel location, $(\delta u, \delta v)$ indicates the pixel displacement within a window W around (u, v) , $\Delta u = u + \delta u$, $\Delta v = v + \delta v$, and functions S , L and K are defined as:

$$S(\delta u, \delta v) = \mathfrak{K}\left(\frac{-(\delta u^2 + \delta v^2)}{\sigma_s^2}\right), \quad (7)$$

$$L(I_{u, v}^{log}, I_{\Delta u, \Delta v}^{log}) = \mathfrak{K}\left(\frac{-(I_{\Delta u, \Delta v}^{log} - I_{u, v}^{log})^2}{\sigma_r^2}\right), \quad (8)$$

$$K(D_{u, v}, D_{\Delta u, \Delta v}) = \mathfrak{K}\left(\frac{-(D_{\Delta u, \Delta v} - D_{u, v})^2}{\sigma_c^2}\right) \quad (9)$$

where, the kernel is selected as $\mathfrak{K}(\mathfrak{S}) = \exp\{-\frac{\mathfrak{S}}{2}\}$ for an arbitrary term \mathfrak{S} . The σ_r , σ_s and σ_c denote range support, spatial support, and geometry support respectively. Further, we calculated detail layer $Detail_{layer} = I^{log} - I_{filt}^{log}$. The final tone-mapped LDR image I^{LDR} is obtained as $I^{LDR} = \exp(\log \mathfrak{J})$. Here $\mathfrak{J} = I_{filt}^C \times C_f + Detail_{layer}$ and C_f denotes the compression factor. Our proposed DA-CTF tone mapping approach recovers contrast in the real world with minute details and attenuate undesirable halo artifacts.

6. EXPERIMENTAL RESULTS

The performance of DIBR-based multi-view HDR content generation and depth-adaptive tone mapping is demonstrated on Middlebury multi-view multi-exposure data sets. The camera views captured from three different exposure levels (0.25ms, 1ms, 4ms) are used in rendering 3D HDR views. The synthesized HDR views from novel camera viewpoints 3 and 4 are shown in Fig. 2 and Fig. 3. The content is generated considering texture images and depth maps of reference cameras 1 and 5 respectively. The parameters selected in DA-CTF rendering are $\sigma_r = 0.9$, $\sigma_s = 6$, $\sigma_c = 4$, and $K_{compress} = 0.3$ for “Dolls”, and $\sigma_r = 0.45$, $\sigma_s = 3$, $\sigma_c = 3.6$, and $K_{compress} = 0.4$ for “Reindeer” data.

In Fig. 2, we compare DA-CTF operator with bilateral tone reproduction approach [10]. The σ_r and σ_s is chosen common for both bilateral and proposed DA-CTF for fair comparison. Utilizing additional information from the depth signal in our DA-CTF improves the result. The DA-CTF prevents blurring over the edges. It is clearly seen in extracted base layer in Fig. 2(c). The halo artifacts get significantly reduced in Fig. 2(e) and Fig. 2(f). To better visualise, consider a window around the pixel (36, 369) on “Dolls” view, where the effects can be profoundly observed. The response of DA-CTF is dominant near depth discontinuities as compared to bilateral filter. This reduces halo artifacts in rendered HDR 3D views. Further, in Fig. 3, the visual results clearly

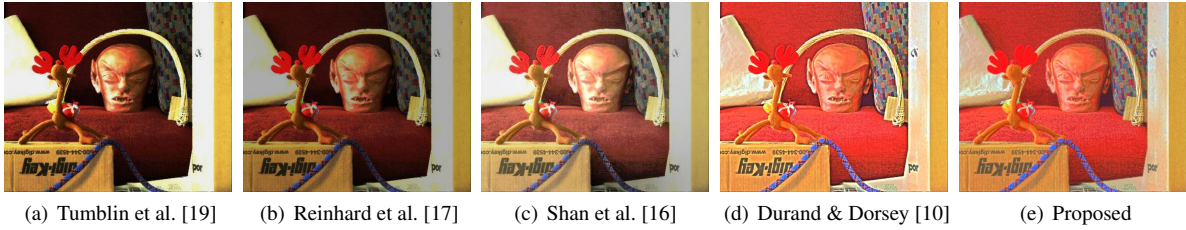


Figure 3. Comparison with different tone reproduction operators. The views presented are synthesized at virtual viewpoint 4 from input cameras 1 & 5.

Table 1. No-reference quality assessment using different objective metrics of tone-mapped HDR pictures.

Model	Tumblin et al. [19]		Reinhard et al. [17]		Shan et al. [16]		Durand & Dorsey [10]		Proposed	
	Dolls	Reindeer	Dolls	Reindeer	Dolls	Reindeer	Dolls	Reindeer	Dolls	Reindeer
HIGRADE-1 [18]	0.17760	0.18725	0.21733	0.16955	0.20945	0.17291	0.23768	0.18779	0.26798	0.23537
HIGRADE-2 [18]	0.31379	0.23171	0.27277	0.25953	0.32176	0.26154	0.31571	0.28478	0.35626	0.28879
IL-NIQE [20]	23.3005	22.6345	20.9424	22.9477	22.0128	20.4164	22.2277	20.8017	24.4233	24.3910

show our proposed DA-CTF tone mapping approach outperforms widely used state-of-the-art tone reproduction operators: Durand & Dorsey [10], Reinhard et al. [17], Shan et al. [16], Tumblin et al. [19]. We performed no-reference image quality assessment of tone-mapped HDR images using an objective metric, called HDR Image GRADient based Evaluator (HIGRADE), derived by Kundu et al. [18]. In addition, we computed Integrated Local NIQE metric used for “completely blind” image quality evaluation [20]. The scores summarized in Table I established that the proposed scheme outperforms state-of-the-art operators.

7. CONCLUSIONS

We proposed a novel DIBR scheme for multi-view 3D HDR content generation. This is a first-of-its-kind scheme that makes use of the MVD format for 3D HDR content creation. Besides, an end-to-end rendering pipeline is supported by a new HDR tone mapping operator based on a depth-adaptive cross-trilateral filter. The proposed tone mapping explicitly leverages additional depth information to recover HDR images from multiple LDR images. The best effect of depth adaptive tone mapping operator is observed in preserving detail visibility (edges) changes and mitigating the challenging halo artifacts in the synthesized LDR 3D views. This requirement is critical for realistic display of 3D HDR images. We displayed generated HDR content on a Samsung 3D LED TV.

More significantly, proposed scheme could play an important role in designing a display-invariant 3D TV HDR content production pipeline [13]. Because it offers HDR depth-contrast and stereo personalization in 3D display technologies. Supporting multi-user and personalized HDR stereo 3D applications increased realism and “3D-ness” - depth-from-HDR effect reproducible in LDR conditions. This is feasible in multi-camera setting, since scene parameterization in MPEG standardized MVD and DIBR format is future proofed for existing and emerging 3D multi-view video formats like depth-enhanced stereo and compatible with 2D TV, stereoscopic, multi-view, or head-tracking 3D and VR displays [13]. In the future, we will improve the proposed scheme for light field HDR tone mapping by interpreting plenoptic images as multi-view sequences.

8. REFERENCES

- [1] C. Fehn, “A 3D-TV approach using depth-image-based rendering (DIBR),” in *Proc. VIIP*, pp. 482-487, 2003.
- [2] Dufaux et al., *High Dynamic Range Video From Acquisition to Display and Applications.*, Elsevier Ltd., 2016.
- [3] K. Seshadrinathan, O. Nestares, “High dynamic range imaging using camera arrays,” *ICIP*, 2017, pp. 725-729.
- [4] Reinhard et al., *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005.
- [5] Orozco et al., *Chapter 3 - HDR Multiview Image Sequence Generation: Toward 3D HDR Video*, Academic Press, 2017.
- [6] Wojciech Matusik and Hanspeter Pfister, “3D TV: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes,” *SIGGRAPH'04*.
- [7] Eilertsen et al., “A comparative review of tone-mapping algorithms for high dynamic range video,” *Comput. Graph. Forum*, 36(2), 2017, 565-592.
- [8] Vangorp et al., “Depth from HDR: depth induction or increased realism?,” in *Proc. ACM SAP'14*, pp. 71-78.
- [9] P. Debevec and J. Malik, “Recovering high dynamic range radiance maps from photographs,” in *Proc. SIGGRAPH'97*.
- [10] Fredo Durand and Julie Dorsey, “Fast bilateral filtering for the display of high-dynamic-range images,” *ACM Trans. Graph.*, 21(3), pp. 257-266, 2002.
- [11] P. Choudhury and J. Tumblin, “The trilateral filter for high contrast images and meshes,” *ACM SIGGRAPH*, 2005.
- [12] Farbman et al., “Edge-preserving decompositions for multi-scale tone and detail manipulation,” *ACM ToG*, 27(3), 2008.
- [13] Bartczak et al., “Display-Independent 3D-TV Production and Delivery Using the Layered Depth Video Format,” in *IEEE ToB*, 57(2), pp. 477-490, 2011.
- [14] <https://sites.google.com/site/mansisharmaiitd/publications/hdr3d>
- [15] Sharma et al., A flexible architecture for multi-view 3DTV based on uncalibrated cameras, *JVCIR*, 25(4), 2014.
- [16] Q. Shan, J. Jia, M. S. Brown, “Globally Optimized Linear Windowed Tone Mapping,” *TVCG*, 16(4), 663-675, 2010.
- [17] Reinhard et al. “Photographic tone reproduction for digital images,” *ToG*, 21(3), 267-276, 2002.
- [18] Kundu et al., “No-Reference Quality Assessment of Tone-Mapped HDR Pictures,” *TIP*, 26(6), 2957-2971, 2017.
- [19] Tumblin et al., “Two methods for display of high contrast images,” *ToG*, 56-94, 1999.
- [20] Zhang et al., “A feature-enriched completely blind local image quality analyzer,” *IEEE TIP*, 24(8), 2579-2591, 2015.