HDR-Plenoxels: Self-Calibrating High Dynamic Range Radiance Fields

Kim Jun-Seong[1*] Kim Yu-Ji[2*] Moon Ye-Bin[1] Tae-Hyun Oh[1,2,†]

[1]Dept. of Elect. Eng. [2]Grad. School of AI
Pohang University of Science and Technology (POSTECH)
{junseong.kim, ugkim, ybmoon, taehyun}@postech.ac.kr
https://github.com/postech-ami/HDR-Plenoxels

Abstract. We propose high dynamic range radiance (HDR) fields, HDR-Plenoxels, that learns a plenoptic function of 3D HDR radiance fields, geometry information, and varying camera settings inherent in 2D low dynamic range (LDR) images. Our voxel-based volume rendering pipeline reconstructs HDR radiance fields with only multi-view LDR images taken from varying camera settings in an end-to-end manner and has a fast convergence speed. To deal with various cameras in real-world scenario, we introduce a tone mapping module that models the digital in-camera imaging pipeline (ISP) and disentangles radiometric settings. Our tone mapping module allows us to render by controlling the radiometric settings of each novel view. Finally, we build a multi-view dataset with varying camera conditions, which fits our problem setting. Our experiments show that HDR-Plenoxels can express detail and high-quality HDR novel views from only LDR images with various cameras.

Keywords: high dynamic range (HDR), novel view synthesis, plenoptic function, voxel-based volume rendering, neural rendering

1 Introduction

The human eyes can respond to a wide range of brightness in the real-world scene, from very bright to very dark, *i.e.*, high dynamic range (HDR). Human can see an object with its color and texture even in dark and dim conditions. However, standard digital cameras capture a limited range of scenes due to the low dynamic range (LDR) limits of the sensors. HDR imaging and display techniques have been developed to overcome the sensors' limits and to share the beauty of the world as humans see.

Existing studies on HDR recovery [5, 20, 32, 36] have been mainly focused on a static view HDR from a monocular perspective or HDR video recovery. HDR images are typically reconstructed by merging multi-exposure LDR images in a fixed camera pose. To recover an HDR image from LDR images taken from various viewpoints, the prior work [3, 37] suggests accumulating images after alignment.

* Authors contributed equally to this work.
† Joint affiliated with Yonsei University, Korea.

Fig. 1. Qualitative results of static and varying camera settings. The static and varying mean camera conditions include exposure, white balance, and CRF. The static camera condition is a controlled environmental setting, *i.e.*, all views of the scene have the same components of exposure, white balance, and CRF. The varying camera condition is alternated environmental settings, *i.e.*, all views of the scene have different components. Each row represents camera conditions, and each column represents the class of synthetic datasets.

However, the HDR reconstruction results of the prior work are still limited to a given view. To overcome the limitations, we propose a method of restoring HDR radiance fields with only multi-view LDR images. The LDR images taken from varying cameras are used, where various radiometric conditions exist, including different exposure, white balance, and camera response functions (CRFs).

The novel view synthesis requires additional information to reconstruct and synthesize unseen views, given a sparse set of images. Previous arts inject prior knowledge by voxel-based [14,47,48], mesh-based [2,4,40,45], multi-plane [43], and volume rendering [16,23] to cope with the problem. Recently, Plenoxels [47] have shown outstanding efficiency of the voxel-based method by assigning spherical harmonics to each voxel corners. While maintaining comparable qualitative results, the Plenoxels achieve two orders of magnitude faster rendering speed than the Neural Radiance Fields (NeRF) [23], which utilizes the implicit neural functions to conduct volume rendering.

In this work, we extend Plenoxels by proposing HDR-Plenoxels that can restore HDR radiance fields with LDR images under diverse camera conditions in an end-to-end manner. Although many saturated regions are appeared due to a wide dynamic range of a scene during training, our HDR-Plenoxels are robust to saturated regions and represents accurate geometry and color at rendering. This is achieved by proposing a tone mapping module that approximates the in-camera pipeline (*i.e.*, camera ISP) from HDR radiance to LDR intensity, allowing flexible modeling of various radiometric and non-linear camera conditions. The tone mapping parameters are spontaneously learned during training. In addition, once 3D HDR radiance fields and the CRFs are fitted, our tone-mapping module can be freely controlled to synthesize different radiometric conditions of rendering in any view. Our tone mapping module can easily be attached to most volume rendering model variants as well.

Our HDR-Plenoxels mainly consist of two parts: 1) HDR radiance fields modeled by Plenoxels followed by 2) the tone mapping module. The differentiable tone mapping module renders HDR radiance values composited from Plenoxels into LDR intensity, which allows to back-propagate gradients to the voxel grid so that spherical harmonics (SH) coefficients and opacity are learned to span the HDR radiance with the scene geometry. The tone mapping module explicitly models CRFs, which enables to self-calibrate CRFs of each view during training. In addition, thereby, we can easily edit the rendering property by just controlling the radiometric curves of each novel view by virtue of the disentangled parameterization of the module. Our experiments show that our method achieves preferable performance on novel view synthesis with varying radiometric conditions of input. Our main contributions are summarized as follows:

- We propose an end-to-end HDR radiance field learning method, HDR-Plenoxels, that allow learning 3D HDR radiance fields from only multi-view and varying radiometric conditioned LDR images as input.
- We model the tone mapping module based on a physical camera imaging pipeline that maps HDR to LDR with explicit radiometric functions.
- We build a multi-view dataset containing varying camera conditions. The dataset includes synthetic and real scenes with various camera settings such as white balance, exposure, and CRF.

2 Related Work

The scope of our work contains HDR imaging, voxel-based volume rendering, and its calibration. We overview the prior work in each perspective in this section.

HDR Imaging. A standard HDR recovery [5] directly accumulates multi-exposure LDR images taken from a fixed camera pose, which are prone to ghosting artifacts in dynamic scenes. To cope with the limitation, several studies [3, 36] suggest a method to recover an HDR image from LDR images taken from moving cameras by using image alignment methods, such as image warping or optical flows. However, they still suffer from large camera motion and occlusion due to the imperfect warping model in the alignment step. In contrast, our work exploits multi-view geometric information, which enables to obtain radiometrically calibrated HDR radiance of an entire 3D scene and to be robust even with large camera motion and occlusion.

Typical digital cameras can only deal with LDR due to the limited dynamic range and the inherent nonlinear components, which represent the real-world scene irradiance inferiorly in pixel values and cause discrepancy to the real scene during the image processing [4]. To obtain an accurate HDR, we have to understand an inherent nonlinear relationship of the camera, *i.e.*, the radiometric properties of camera ISPs. Traditional radiometric calibration models the components of physical pipelines of cameras, including white balance and CRF, and optimizes to reconstruct HDR from only given LDR images [5] or HDR-LDR image pairs [12]. The latest learning-based approaches [6, 7, 18] suggest an implicit model-based method but require ground truth HDR images paired with LDR images for training. Liu *et al.* [15] replace an implicit function with an explicit physical camera model, enhancing the HDR image reconstruction quality. Our method shares the same advantages by adopting the explicit tone-mapping module. Note that our HDR-Plenoxels learn HDR radiances up to scale, given only LDR images but without ground truth HDR images and camera parameters.

Volume Rendering. Volume rendering is the method of understanding the 3D information inherited in two-dimensional images to render images at unseen views, called novel views. Existing methods [23, 38] show high performance in complex geometric shapes but require high memory for high expressiveness.

The recent volume rendering methods utilize multi-layer perceptron (MLP) based implicit neural function to predict the signed distance fields [8, 26, 46] and occupancy [21, 28, 34], and demonstrate the high expressiveness with high compression power. In particular, Neural Radiance Fields (NeRF) [24] shows fine-detailed rendering performance unprecedently. However, the NeRF-related studies [19, 27, 29, 42] have a limitation of high training and rendering time complexity due to the forward process in every sampling point. Several studies [9, 25, 31, 39] try to modify the neural network to reduce the computation at each sampling point to reduce the time cost.

Octree structure-based methods [14, 47, 48] are efficient methods that reduce rendering time by virtue of their structure. Plenoxels [47] optimize the octree structure with spherical harmonics instantaneously, requiring only tens of minutes of training time to achieve detailed rendering results comparable to NeRF. Our method uses Plenoxels as a volume rendering backbone for efficient rendering, and further expands the expression power of Plenoxels to 3D HDR radiance fields with negligible computational cost.

Calibrated Volume Rendering. Several methods [10, 19, 22, 33] are proposed for better performance with relaxed assumptions in volume rendering. NeRF in the Wild (NeRF-W) [19] uses web images in the wild setting for reconstruction, which deals with varying camera conditions and occluded objects by introducing a handling mechanism by appearance and transient embeddings.

ADOP [33] is a point-based HDR neural rendering pipeline that consists of a differentiable and physically-based tone mapping. Due to the differentiable property, all the varying camera conditions can be optimized. However, ADOP requires dense COLMAP structure-from-motion package [35] results as input, which have expensive time costs. Also, the method itself has expensive time and memory costs during training due to the point-based method. Distinctive from ADOP, our method is cost efficient by utilizing an octree-based structure and camera poses without dense COLMAP.

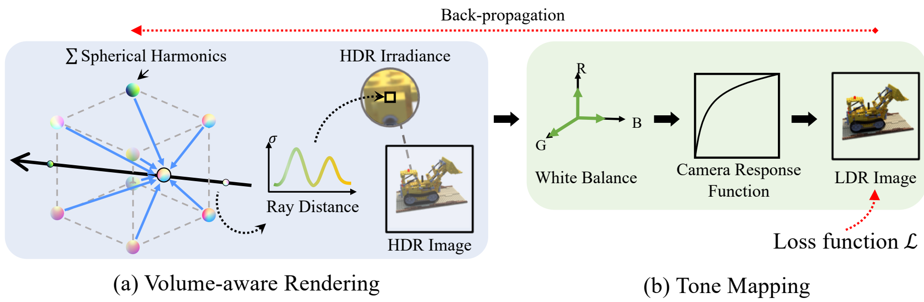

Fig. 2. Overall pipeline of HDR-Plenoxels. 1) Plenoxels synthesizes an HDR image from HDR radiance by ray-marching, then 2) the differentiable tone-mapping function maps from HDR to LDR in an end-to-end manner. The self-calibration is done by minimizing the residual between the synthesized LDR image and the captured one with regularizations.

Recently, HDR-NeRF [10] tackles a similar HDR radiance field problem with ours by NeRF, which is concurrent work with us. The work requires known exposure information, and does not take into account white balance parameters, in contrast to ours.

3 HDR Radiance Fields from Multi-view LDR Images

Our work aims to reconstruct HDR radiance fields of a visual scene from multi-view LDR input images. In this section, we first present the overall pipeline of our method, which is composed of two parts, 1) volume-aware HDR image rendering (Sec. 3.1), and 2) synthesis of LDR images through the tone mapping module (Sec. 3.2). We then explain the details of optimization (Sec. 3.3).

3.1 Volume-aware Rendering to HDR Images

To reconstruct the HDR irradiance fields from the multi-view LDR images, we parameterize HDR radiance fields by voxel grid with spherical harmonics (SH) called Plenoxels [47]. A bounded three-dimensional space of interest is represented as a sparse voxel grid, each of which has opacity and SH coefficients. The volume rendering method adopts a coarse-to-fine training scheme similar to NSVF [14]. The learning process starts with a broad and uniformly divided sparse voxel grid, and the voxels are upsampled to make a denser grid as learning progresses. The voxels are pruned according to the occupancy threshold to reduce the computational cost, as training iterations go. Upsampling and pruning are applied simultaneously during training and repeated several times.

As SH can compactly represent any functions on a sphere well with a few SH bases [1, 48], it has been vastly used in graphics for HDR environmental lighting [30] and glossy representation [17], which motivates our HDR radiance representation. Furthermore, recent work has adopted SH in volume contents and has demonstrated its effectiveness in implicitly expressing the non-Lambertian effects [13, 44]. Our HDR irradiance field modeling by SH based volume-aware rendering exploits these advantages.

The SH in each voxel grid is used for view representation. All the colors over every direction of a sphere are spanned by a weighted sum of pre-defined spherical basis functions and coefficients corresponding to each function. Therefore, the corresponding color can be defined for each specific angle. Each vertex of the voxel grid stores 28-dimensional vectors: 27 for SH coefficients (9 coefficients per color channels) and 1 for voxel opacity σ. Empirically, the values of the SH coefficients change significantly during training which makes the training unstable. To mitigate this, we initialize the color to grey by adding offset color 0.5 and the voxel opacity value to 0.1.

The HDR volume rendering part determines the color of a rendered HDR image pixel $\hat{C}(\mathbf{r})$ by ray-marching the color and opacity of points sampled along a ray in a bounded three-dimensional voxel grid volume. At any 3D point $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ and normalized viewing angle $(\mathbf{v_x}, \mathbf{v_y}, \mathbf{v_z})$ inside the voxel grid of Plenoxels, the color and opacity of the point are trilinearly-interpolated from eight nearby voxel vertices. For a camera center $\mathbf{o}$ and given an image pixel grid, we can define the ray $\mathbf{r} = \mathbf{o} + t\mathbf{d}$ starting from $\mathbf{o}$ to each pixel in the camera along the direction $\mathbf{d}$. After that, N sampling points are sampled over the ray at regular intervals $\delta_i = t_{i+1} - t_i$. The color and opacity of each sampled point are denoted as $\mathbf{c_i}$ and σ_i, respectively, and T_i the accumulated transmittance value up to the i-th point. The ray-marching proceeds as follows:

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^{N} T_i \left(1 - \exp\left(-\sigma_i \delta_i\right)\right) \mathbf{c}_i, \text{ where } T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right). \tag{1}$$

The ray sampling randomly selects rays among the set of rays toward all pixels of an image for efficient training.

3.2 Tone Mapping

The tone mapping stage converts an HDR image into a LDR image. We denote a pixel value of an HDR image and a LDR image as as I_h and I_l, respectively. The output of volume rendering is HDR radiance fields, and we obtain I_h as output by ray marching HDR radiance fields. We represent our explicit tone mapping module as a function $\mathcal{T}$ with radiometric parameters θ, *i.e.*, $I_l = \mathcal{T}(I_h, \theta)$.

The tone mapping function $\mathcal{T}$ consists of two stages. Each stage is parameterized by the physical property of its components and represented as separate functions: white balance function w and camera response function (CRF) g. Note that we regard the white balance scale parameters are merged with the exposure value and learn at once. Two sub functions are applied sequentially as $I_l = \mathcal{T}(I_h) = g \circ (w(I_h))$ which follows the image acquisition process of common digital cameras.

Specifically, first, for a specific ray $\mathbf{r}$, the pixel color of an HDR image $C_h(\mathbf{r})$ is calculated through ray-marching. The white balance function $w(\cdot)$ is applied to $C_h(\mathbf{r})$ with $\theta_w = [w_r, w_g, w_b]^\top \in \mathbb{R}^3$, and the function output is a pixel of white balance calibrated image I_w. That is, given each channel components of

$C_h(\mathbf{r}) = [c_h^r, c_h^g, c_h^b] \in \mathbb{R}^3$,

$$I_w = w(C_h(\mathbf{r}), \theta_w) = C_h(\mathbf{r}) \odot \theta_w = \begin{bmatrix} c_h^r \\ c_h^g \\ c_h^b \end{bmatrix} \odot \begin{bmatrix} w_r \\ w_g \\ w_b \end{bmatrix} = \begin{bmatrix} w_r c_h^r \\ w_g c_h^g \\ w_b c_h^b \end{bmatrix}, \tag{2}$$

where the operator $\odot$ stands for an element-wise product. To make white balance physically proper, we regularize θ_w to be a positive value.

The CRF g is applied to I_w. We paramterize non-linear CRFs with an approximated discrete piece-wise linear function. The function $g(\cdot)$ is divided into 256 intervals which are allocated for uniformly sampled points in [0,1], and parameterized by 256 control points. The pixel value of white balance corrected image I_w is mapped to I_l by interpolating corresponding CRF values of nearby control points in domain. To make the CRF $g(\cdot)$ differentiable, we adapt 1D grid-sampling used in [11]. According to Debevec *et al*. [4], the CRF is enforced to follow the following boundary condition: $I_l = g\left(I_w; \theta_g\right)$, $g(0) = 0$, $g(1) = 1$. A range beyond the dynamic range [0,1] is thresholded when applying the CRF g. To propagate a loss on the saturation region of the rendered images during training HDR radiance fields, we apply the leaky-thresholding method:

$$g_{leaky}(x) = \begin{cases} \alpha x, & x < 0 \\ \mathrm{g}(x), & 0 \leq x \leq 1 \\ -\frac{\alpha}{\sqrt{x}} + \alpha + 1, & 1 < x, \end{cases} \tag{3}$$

where α is the thresholding coefficient.

3.3 Optimization

Leaky Saturation Mask. When taking a scene with a wide dynamic range, the LDR image may contain over- and under-saturation. In the saturated region, no cue exists to guess correct geometric and photometric information due to missing texture. This acts as outliers when taking into account it in the loss computation during optimization. To suppress the impact of saturation regions and prevent our recovery from being biased, we use saturation masking in the loss computation. We define our leaky saturation mask as follows:

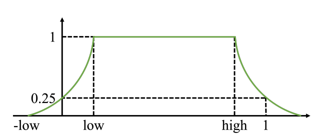

Fig. 3. Leaky saturation mask.

$$\mathtt{mask}(x) = \begin{cases} \left(\frac{x+\mathsf{low}}{2\mathsf{low}}\right)^2 & x < \mathsf{low}, \\ 1, & \mathsf{low} \leq x \leq \mathsf{high}, \\ \left(\frac{2-x}{2(1-\mathsf{high})}\right)^2 & x > \mathsf{high}, \end{cases} \tag{4}$$

We empirically set $\mathsf{low} = 0.15$ and $\mathsf{high} = 0.9$ for the experiment.

White Balance Initialization. There exists inherent ambiguity in the camera imaging pipeline, which occurs due to the inherent entangled relationship across model components in the pipeline, *e.g.*, for an image of a view, if we increase its exposure time twice while reducing the white balance of the view by half, the resulting image appears same with the original setting of exposure time; *i.e.*, there are multiple solutions that can produce the same LDR images.

We avoid such cumbersome ambiguity between exposure and white balance in our method. We use only the white balance module w to express both exposure as a scale and white balance ratio following the study [12] to workaround the scale ambiguity between exposure and white balance. However, with this representation, the overall scale of the white balance is trained extremely small or large. Therefore, we calculate the averaged color of all inputs $(r_a, g_a, b_a) \in \mathbb{R}^3$, select a reference image which has the closest value of (r_a, g_a, b_a), and fix the white balance of the reference image $(r_{ref}, g_{ref}, b_{ref}) \in \mathbb{R}^3$. This acts as regularization. This helps white balance be learned on the proper scale, which also means we have a suitable exposure value.

However, still there exists a similar ambiguity between *SH coefficients* and *white balance*. We observe that when the exposure differences are significantly dynamic among neighborhood views, observed LDR intensity differences by exposure times are misunderstood as the cause of the high-frequency reflectance* of the scene and different white balances. This tends to produce wrong geometry as the rays have reached different parts of the scene. We also found that the more abrupt the intensity changes among neighborhood views are, the more dominant the coefficients corresponding to high-frequency SH components become.

As a simple workaround, we use white balance information for each camera as a prior to resolve the ambiguity in the SH side. We introduce white balance initialization for each camera that guides initial solutions to physically plausible solutions; thereby, the optimization process becomes more stable and faster to converge to desirable solutions, and robust to such harsh input conditions. We estimate a reference color ratio by comparing per-image averaged pixel values to averaging each rgb value from the entire image set $\mathcal{S}$. Thenb the initial white balances $wb_{c,i}$ for each camera, where $c \in \{r, g, b\}$ of each image I_i are initialized as $wb_{c,i} = \frac{\text{mean}_{k \in I_i}(c_k)}{\text{mean}_{j \in \mathcal{S}}(c_j)}$.

Spherical Harmonics Regularization. In the harsh input condition case, where LDR images obtained from neighborhood views have significantly dynamic exposure differences, the above initialization stabilizes early optimization steps. If the optimization speed of the white balance does not match that of the SH coefficients, the ambiguity may arise again in later optimization steps. In order to regularize this, we introduce SH coefficient masking that allows scheduling to learn from diffuse reflectance property (view direction invariant radiance) first to view direction sensitive ones, *i.e.*, low frequency order SH to high frequency ones. We apply SH masking to the coefficient of SH of degrees 2 and 3. We decrease the rate of SH masking by 1/5 per epoch during the early five epochs for gradual learning. After the early five epochs, we update SH of all degrees with full rate,

*The high-frequency reflectance refers to the case that a subtle view direction change results in drastic reflectance ratio changes, such as glossy materials.

i.e., no SH masking. This scheduling notably stabilizes the optimization for the harsh condition input case.

Loss Functions. We optimize our pipeline *w.r.t.* voxel opacity, SH coefficients, white balance and CRF, given multi-view LDR images as input, with the following objective function:

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + \lambda_{\text{TV}}\mathcal{L}_{\text{TV}} + \lambda_{\text{smooth}}\mathcal{L}_{\text{smooth}}, \tag{5}$$

where each term is defined as follows. The LDR reconstruction loss for i-th image is defined as:

$$\mathcal{L}_{\text{recon}} = \frac{1}{|\mathcal{R}|}\sum\nolimits_{\mathbf{r}\in\mathcal{R}} M_i(\mathbf{r})\|I_i\left(\Pi_i(\mathbf{r})\right) - \mathcal{T}\left(\hat{C}(\mathbf{r})\right)\|_2^2, \tag{6}$$

where $\Pi_i(\cdot)$ denotes the camera projection operator from a ray to the 2D pixel coordinate of the i-th LDR image, and $M_i(\mathbf{r}) = \texttt{mask}(I_i\left(\Pi_i(\mathbf{r})\right))$ denotes the saturation mask computed from input LDR images. We randomly sample rays $\mathbf{r}_{sampled}$ among the possible set of rays $\mathcal{R}$ from $\mathcal{N}$ images. The loss is calculated through a color difference between the rendered results and the ground truth LDR values along each ray $\mathbf{r}_{sampled}$ considering the saturation masking $M_i(\mathbf{r})$. The total loss is applied by normalizing the number of ray sampled. The other two terms are for regularization. The total variation loss is defined as:

$$\mathcal{L}_{TV} = \frac{1}{|\mathcal{V}|}\sum\nolimits_{\mathbf{v}\in\mathcal{V},d\in[D]}\sqrt{\Delta_x^2(\mathbf{v},d) + \Delta_y^2(\mathbf{v},d) + \Delta_z^2(\mathbf{v},d) + \epsilon}, \tag{7}$$

where the differences Δ. are calculated between successive voxels along each respective (x, y, z)-axis, *e.g.*, the d-th voxel value at (x, y, z) and the d-th voxel value at $(x+1, y, z)$ for x-axis. The total variation loss is applied for opacity σ and SH coefficients separately. This encourages spatial and color consistency in the voxel space. In implementation, we use different weighting for SH coefficients $\lambda_{\text{TV,SH}}$ and opacity $\lambda_{\text{TV},\sigma}$.

The smoothness loss is for obtaining a physically appropriate CRF [4] such that CRFs increase smoothly, which is defined as:

$$\mathcal{L}_{smooth} = \sum\nolimits_{i=1}^{N}\sum\nolimits_{e\in[0,1]} g_i''(e)^2, \tag{8}$$

where $g''(e)$ denotes the second order derivative of CRFs *w.r.t.* the domain of CRFs. We set $\lambda_{\text{TV},\sigma} = 5\cdot 10^{-4}$, $\lambda_{\text{TV,SH}} = 1\cdot 10^{-2}$ and $\lambda_{\text{smooth}} = 1\cdot 10^{-3}$.

4 Experiments

We compare the qualitative and quantitative results of our HDR-Plenoxels from three perspectives: LDR image rendering accuracy, HDR irradiance image, and 3D structure reconstruction quality. We also conduct an ablation study on our tone-mapping components to demonstrate that our tone mapping components efficiently understand camera settings.

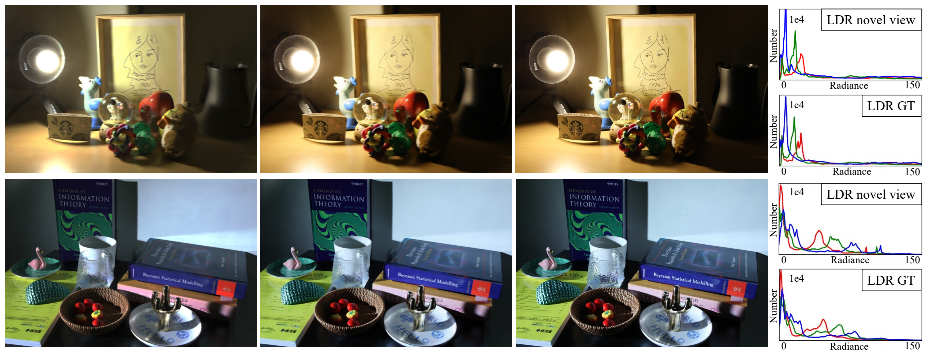

Fig. 4. Qualitative novel view results in real scenes with various camera conditions. Each row represents different real scenes, and the column represents HDR novel view rendering, LDR novel view rendering, LDR GT (*i.e.*, inputs of training), and RGB histograms of LDR novel view and LDR GT in order. Histograms of RGB are clipped from 0 to 150 at RGB radiance and from 0 to 10^4 at the number of RGB radiance for visual better visibility.

4.1 Experimental Settings

Due to the lack of open datasets proper to our experiment setting, *i.e.*, images under varying camera conditions, we collect synthetic and real data that fits our experimental settings: images with various exposure, white balance, and vignetting taken from the same camera poses. The details of the synthetic and real datasets are explained in the supplementary material.

Baselines. We compare our proposed method against the following three models: original Plenoxels [47], NeRF-A [19], and Approximate Differentiable One-Pixel Point Rendering (ADOP) [33]. We conduct experiments for two environmental settings: the varying camera, which has various environments that are exposure, white balance, and CRF, and the static camera, which has controlled environments. Plenoxels with inputs of the static camera are assumed to be an upper bound of the task performance. We perform experiments with two environmental settings at baselines.

Evaluation. HDR-Plenoxels can learn the 3D HDR radiance fields and freely control the appearance of a novel view, *i.e.*, unseen viewpoint, according to camera conditions. We compute the similarity between the LDR image synthesized in the novel view and the ground truth LDR image to measure the performance for the novel view synthesis tasks. For quantitative evaluation, we use three metrics, PSNR, SSIM [41], and LPIPS [49]. Our method and NeRF-A need to learn tone mapping function and appearance embedding parameters separately at each image. We cannot predict the tone mapping parameters at novel views; thus, we use the left half of the test image (*i.e.*, unseen view) at training and evaluate the performance of the right half with learned parameters.

Table 1. Quatitative results of novel view synthesis on synthetic data. Values are average of the results for the test data of each scene. For evaluation, ADOP exploits all the input images as it needs dense reconstruction. For the other models, the left half of the image was included in the learning data and learned, and tested on the unseen right half of the image. Our model shows overall high performance compared to other models. $\mathcal{S}$ denote the static and $\mathcal{V}$ is the varying datasets. The blue and red color stand for the **best** and the **second best**, respectively.

Type	Method	Book			Classroom			Monk			Room			Kitchen		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
$\mathcal{S}$	Baseline	22.53	0.796	0.293	28.71	0.902	0.261	27.15	0.848	0.281	30.70	0.912	0.183	33.43	0.957	0.138
$\mathcal{V}$	Baseline	11.92	0.454	0.597	12.83	0.542	0.660	15.81	0.535	0.542	13.28	0.599	0.643	18.24	0.718	0.496
	ADOP	22.15	0.824	**0.291**	21.04	0.800	0.345	21.92	0.764	**0.392**	19.25	0.834	0.329	20.13	0.827	0.280
	NeRF-A	**28.44**	**0.873**	0.310	**29.30**	**0.895**	**0.295**	**27.33**	**0.793**	0.398	**30.32**	**0.891**	**0.234**	**31.30**	**0.928**	**0.233**
	Ours	**27.49**	**0.837**	**0.292**	**29.87**	**0.908**	**0.284**	**28.27**	**0.852**	**0.297**	**28.70**	**0.900**	**0.291**	**31.53**	**0.936**	**0.156**

4.2 High Dynamic Range Radiance Fields

We evaluate the effectiveness of our HDR-Plenoxels by comparison against the counterpart models, which handle images with varying appearances.

Comparison. Our HDR-Plenoxels learn HDR radiance fields from LDR input images with various camera conditions at real scenes. The results in Fig. 4 represent the novel view synthesis of HDR images and tone mapping from HDR to LDR images in real datasets. Compared to previous work focusing on single perspective HDR [5], our method can reconstruct HDR radiance fields from varying camera LDR images with several multi-view. We can render novel LDR views from reconstructed HDR radiance fields with an explicitly controllable tone mapping module.

Quantitative results are summarized in Table 1. NeRF-A [19] shows comparable novel view synthesis results. However, it cannot explicitly decompose each camera condition, such as exposure, white balance, and CRF, because those information is implicitly entangled in the embedding; thus, it cannot predict *radiance* in contrast to ours. NeRF-A also needs considerable training time compared to our neural networks free method as shown in Fig. 5. To reach PSNR 29, our method takes 30 minutes, but nerf-a takes 6 hours and 30 minutes, which is 13 times larger. We use RTX 3090 for training.

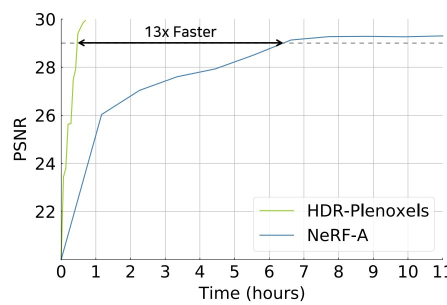

Fig. 5. Computation time.

The qualitative results of rendered novel LDR views on our synthetic dataset are shown in Fig. 6 and Fig. 7. ADOP tends to incorrectly estimate the camera components such as white balance, vignetting, and CRF, leading to inaccurate color-mapped rendering results. The rendering result of NeRF-A shows comparable quality to ours overall, but the details tend to be deficient. On the contrary, our HDR-Plenoxels can render relatively accurate LDR images in both aspects of color and geometry.

Fig. 6. Comparison in synthetic scenes with varying camera conditions. The left half of an image is used for training the tone mapping module $\mathcal{T}$, and the right half is for the test. By applying the trained $\mathcal{T}$ at left half, we can synthesize the novel view images. MSE maps of each result are on the right of the corresponding render results.

Fig. 7. Qualitative results from the experiments between other baselines. Our result represents fine-grained and proper color rendering results. All experiments are trained with various radiometric conditions. The left half of an image is used in training, and the right half is only used to measure quantitative results.

ADOP [33] shows deficient novel view rendering results compared to our method. In the Fig. 7, they cannot reconstruct fine-grained geometry structure compared to ours, and color information is also ambiguous. The error map results of ours also show better than ADOP results as shown in Fig. 6. Therefore the visual quality of the novel view results of ADOP is not satisfying, and we outperform quantitative results with PSNR metrics at Table 1.

To verify that our model can restore HDR radiance field robust to under- or over- saturated points, we qualitatively compared HDR details on saturation points as shown in Fig. 8. There are severe dark or bright points, so some regions are saturated and cannot distinguish color or geometry. We train with these saturated LDR input images at our HDR-Plenoxels. They can render high-quality HDR novel views, which means we can handle wide dynamic ranges and render non-saturated novel HDR views.

Fig. 8. High dynamic range radiance rendering at saturation points. Our novel view rendering results show robustness to under-saturation (1^{st} row) and over-saturation (2^{nd} row) points because we train the HDR radiance fields.

We also compare our HDR-Plenoxels and original Plenoxels [47] with images of different camera settings in Fig. 1. The novel view synthesis results of the Plenoxels with static camera condition are considered as our upper bound performance and represent fine-detailed rendering results with clear color estimation. With the varying camera setting, the Plenoxels fail to reconstruct proper 3D geometry, especially in the right-half of images unseen view during training. These results imply that previous volume rendering methods such as Plenoxels are prone to degrade at varying camera conditions due to their static photometric assumption, which needs to be compensated with additional regularization. With our tone mapping module, the quality of novel view synthesis improves considerably by showing comparable details to original Plenoxels trained with static LDR images. This demonstrates that our tone mapping module precisely disentangles varying cameras and properly reconstructs informative HDR radiance fields.

Ablation Studies. To verify the effectiveness of our tone mapping module components, we conduct an ablation study by removing each component as in Table 2. We report the averaged results of five synthetic datasets. Comparing (A) with (B), the considerable performance improvement by the white balance module means that disentanglement of exposure and white balance helps to learn accurate geometry and color. The performance degradation from (B) to (C) and from (D) to (E) implies that decomposing the vignetting effect of images is inconducive. According

Table 2. Effect of tone mapping components on novel view synthesis. WB stands for white balance, **VIG** for vignetting, and **CRF** for camera response function. (A) is the baseline.

	WB	VIG	CRF	PSNR↑	SSIM↑	LPIPS↓
(A)				14.42	0.569	0.587
(B)	✓			23.03	0.811	0.355
(C)	✓	✓		21.12	0.799	0.352
(D)	✓		✓	**29.34**	**0.876**	**0.294**
(E)	✓	✓	✓	26.73	0.878	0.264

to the characteristics of our tone mapping module, which trains quickly according to the learning speed of SH, the vignetting function significantly increases the complexity of the model and hinders training. The result in (D) demonstrates that all components of our tone mapping module are essential.

To confirm whether the CRF expression is suitable for defining camera non-linearity we proceed an ablation study by replacing CRF function. Our method adopt a piece-wise linear CRF model that considers the mapping relationship of the pixel value. NeRF-W and HDR-NeRF [10] suggest that an implicit function can represent CRF. We compare the performance between our explicit piece-wise linear function and implicit function such as multi-layer perceptrons (MLPs). Following the concurrent work, HDR-NeRF, we replace our CRF with three MLPs to predict each color channel. As shown in Fig. 9, rendering results are degraded with MLP-based CRF. The averaged PSNR of all synthetic datasets with ours is 28.18, and the MLP-based method is 26.11. We verify that our physically-based explicit tone mapping module outperforms than MLP-based method since our CRF satisfies real-world physical conditions and can disentangle non-linear components correctly.

Fig. 9. Implicit CRF. GT (left), MLP (middle), and ours (right).

5 Conclusion

We present an HDR-Plenoxels, which learns to synthesize 3D HDR radiance field from multi-view LDR images of the varying camera by self-calibrating radiometric characteristics. Distinctive from conventional HDR reconstruction methods, ours can get a novel view, depth, and 3D HDR radiance fields simultaneously in an end-to-end manner. We investigate that the white balance and CRF functions are critical factors among the in-camera components and find an effective representation of the CRF function. With these observations, we present a simple and straightforward physical-based tone-mapping module, which can be easily attached to various volume rendering models extending one's usability. Using the tone mapping module, we can take fine-grained HDR rendering results as well as LDR images of varying cameras, *e.g.*, exposure, white balance, and CRF. The HDR radiance fields represent real-world scenes more realistically with a wide dynamic range similar to the way human sees. Our work could improve the experiences of many HDR-based applications, such as movie post-production. Since we focus on covering HDR radiance fields of static scenes, it has room to improve in dealing with dynamic objects as future work.

Acknowledgment. This work was partly supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2021R1C1C1006799), Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2022-0-00290, Visual Intelligence for Space-Time Understanding and Generation based on Multi-layered Visual Common Sense; and No.2019-0-01906, Artificial Intelligence Graduate School Program(POSTECH)).

References

1. Basri, R., Jacobs, D.: Lambertian reflectance and linear subspaces. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) **25**(2) (2003) 5
2. Buehler, C., Bosse, M., McMillan, L., Gortler, S., Cohen, M.: Unstructured lumigraph rendering. ACM Transactions on Graphics (SIGGRAPH) (2001) 2
3. Chen, G., Chen, C., Guo, S., Liang, Z., Wong, K.Y.K., Zhang, L.: HDR Video Reconstruction: A Coarse-to-fine Network and A Real-world Benchmark Dataset. In: IEEE International Conference on Computer Vision (ICCV) (2021) 1, 3
4. Debevec, P.E., Taylor, C.J., Malik, J.: Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. ACM Transactions on Graphics (SIGGRAPH) (1996) 2, 4, 7, 9
5. Debevec, P.E., Taylor, C.J., Malik, J.: Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach. In: ACM Transactions on Graphics (SIGGRAPH) (1996) 1, 3, 4, 11
6. Eilertsen, G., Kronander, J., Denes, G., Mantiuk, R.K., Unger, J.: Hdr image reconstruction from a single exposure using deep cnns. ACM Transactions on Graphics (SIGGRAPH) **36**(6) (2017) 4
7. Endo, Y., Kanamori, Y., Mitani, J.: Deep reverse tone mapping. ACM Transactions on Graphics (SIGGRAPH) **36**(6) (2017) 4
8. Gropp, A., Yariv, L., Haim, N., Atzmon, M., Lipman, Y.: Implicit Geometric Regularization for Learning Shapes. In: International Conference on Machine Learning (ICML) (2020) 4
9. Hedman, P., Srinivasan, P.P., Mildenhall, B., Barron, J.T., Debevec, P.: Baking Neural Radiance Fields for Real-Time View Synthesis. In: IEEE International Conference on Computer Vision (ICCV) (2021) 4
10. Huang, X., Zhang, Q., Feng, Y., Li, H., Wang, X., Wang, Q.: HDR-NeRF: High Dynamic Range Neural Radiance Fields. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2022) 4, 5, 14
11. Jaderberg, M., Simonyan, K., Zisserman, A., Kavukcuoglu, K.: Spatial Transformer Networks. In: Advances in Neural Information Processing Systems (NeurIPS) (2016) 7
12. Kim, S.J., Lin, H.T., Lu, Z., Süsstrunk, S., Lin, S., Brown, M.S.: A New In-Camera Imaging Model for Color Computer Vision and Its Application. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) **34**(12) (2012) 4, 8
13. Kuang, Z., Olszewski, K., Chai, M., Huang, Z., Achlioptas, P., Tulyakov, S.: NeROIC: Neural object capture and rendering from online image collections. Computing Research Repository (CoRR) (2022) 5
14. Liu, L., Gu, J., Lin, K.Z., Chua, T.S., Theobalt, C.: Neural Sparse Voxel Fields. In: Advances in Neural Information Processing Systems (NeurIPS) (2021) 2, 4, 5
15. Liu, Y.L., Lai, W.S., Chen, Y.S., Kao, Y.L., Yang, M.H., Chuang, Y.Y., Huang, J.B.: Single-Image HDR Reconstruction by Learning to Reverse the Camera Pipeline. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020) 4
16. Lombardi, S., Simon, T., Saragih, J., Schwartz, G., Lehrmann, A., Sheikh, Y.: Neural Volumes: Learning Dynamic Renderable Volumes from Images. ACM Transactions on Graphics (SIGGRAPH) **38**(4) (2019) 2
17. Mahajan, D., Ramamoorthi, R., Curless, B.: A theory of frequency domain invariants: Spherical harmonic identities for brdf/lighting transfer and image consistency. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (2008) 5

18. Marnerides, D., Bashford-Rogers, T., Hatchett, J., Debattista, K.: ExpandNet: A Deep Convolutional Neural Network for High Dynamic Range Expansion from Low Dynamic Range Content. European Association for Computer Graphics (Eurographics) (2019) 4
19. Martin-Brualla, R., Radwan, N., Sajjadi, M.S.M., Barron, J.T., Dosovitskiy, A., Duckworth, D.: NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2021) 4, 10, 11
20. Mertens, T., Kautz, J., Van Reeth, F.: Exposure fusion. In: 15th Pacific Conference on Computer Graphics and Applications (PG'07) (2007) 1
21. Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A.: Occupancy Networks: Learning 3D Reconstruction in Function Space. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019) 4
22. Mildenhall, B., Hedman, P., Martin-Brualla, R., Srinivasan, P., Barron, J.T.: NeRF in the Dark: High Dynamic Range View Synthesis from Noisy Raw Images. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2022) 4
23. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In: European Conference on Computer Vision (ECCV) (2020) 2, 4
24. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: European Conference on Computer Vision (ECCV) (2020) 4
25. Neff, T., Stadlbauer, P., Parger, M., Kurz, A., Mueller, J.H., Chaitanya, C.R.A., Kaplanyan, A., Steinberger, M.: DONeRF: Towards Real-Time Rendering of Compact Neural Radiance Fields using Depth Oracle Networks. Computer Graphics Forum **40**(4) (2021) 4
26. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019) 4
27. Park, K., Sinha, U., Barron, J.T., Bouaziz, S., Goldman, D.B., Seitz, S.M., Martin-Brualla, R.: Nerfies: Deformable Neural Radiance Fields. In: IEEE International Conference on Computer Vision (ICCV) (2021) 4
28. Peng, S., Niemeyer, M., Mescheder, L., Pollefeys, M., Geiger, A.: Convolutional Occupancy Networks. In: European Conference on Computer Vision (ECCV) (2020) 4
29. Pumarola, A., Corona, E., Pons-Moll, G., Moreno-Noguer, F.: D-NeRF: Neural Radiance Fields for Dynamic Scenes. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020) 4
30. Ramamoorthi, R., Hanrahan, P.: An efficient representation for irradiance environment maps. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (2001) 5
31. Reiser, C., Peng, S., Liao, Y., Geiger, A.: KiloNeRF: Speeding up Neural Radiance Fields with Thousands of Tiny MLPs. In: IEEE International Conference on Computer Vision (ICCV) (2021) 4
32. Robertson, M.A., Borman, S., Stevenson, R.L.: Dynamic range improvement through multiple exposures. In: In Proc. of the Int. Conf. on Image Processing (ICIP'99 (1999) 1
33. Rückert, D., Franke, L., Stamminger, M.: ADOP: Approximate Differentiable One-Pixel Point Rendering. arXiv:2110.06635 (2021) 4, 10, 12

34. Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., Li, H.: PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization. IEEE International Conference on Computer Vision (ICCV) (2019) 4
35. Schonberger, J.L., Frahm, J.M.: Structure-from-Motion Revisited. In: CVPR (Jun 2016) 4
36. Sen, P., Kalantari, N.K., Yaesoubi, M., Darabi, S., Goldman, D.B., Shechtman, E.: Robust patch-based hdr reconstruction of dynamic scenes. ACM Transactions on Graphics (SIGGRAPH) **31**(6) (2012) 1, 3
37. Sen, P., Kalantari, N.K., Yaesoubi, M., Darabi, S., Goldman, D.B., Shechtman, E.: Robust patch-based hdr reconstruction of dynamic scenes. ACM Transactions on Graphics (SIGGRAPH) **31**(6) (2012) 1
38. Sitzmann, V., Zollhöfer, M., Wetzstein, G.: Scene Representation Networks: Continuous 3D-Structure-Aware Neural Scene Representations. In: Advances in Neural Information Processing Systems (NeurIPS) (2020) 4
39. Song, W., Shi, C., Xiao, Z., Duan, Z., Xu, Y., Zhang, M., Tang, J.: AutoInt: Automatic Feature Interaction Learning via Self-Attentive Neural Networks. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management (2019) 4
40. Waechter, M., Moehrle, N., Goesele, M.: Let There Be Color! Large-Scale Texturing of 3D Reconstructions. In: European Conference on Computer Vision (ECCV) (2014) 2
41. Wang, Z., Simoncelli, E., Bovik, A.: Multiscale structural similarity for image quality assessment. In: Asilomar Conference on Signals, Systems & Computers (2003) 10
42. Wang, Z., Wu, S., Xie, W., Chen, M., Prisacariu, V.A.: NeRF–: Neural Radiance Fields Without Known Camera Parameters. In: arXiv:2102.07064 [cs] (2021) 4
43. Wizadwongsa, S., Phongthawee, P., Yenphraphai, J., Suwajanakorn, S.: NeX: Real-time View Synthesis with Neural Basis Expansion. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2021) 2
44. Wizadwongsa, S., Phongthawee, P., Yenphraphai, J., Suwajanakorn, S.: Nex: Real-time view synthesis with neural basis expansion. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2021) 5
45. Wood, D.N., Azuma, D.I., Aldinger, K., Curless, B., Duchamp, T., Salesin, D.H., Stuetzle, W.: Surface light fields for 3D photography. ACM Transactions on Graphics (SIGGRAPH) (2000) 2
46. Yariv, L., Kasten, Y., Moran, D., Galun, M., Atzmon, M., Basri, R., Lipman, Y.: Multiview Neural Surface Reconstruction by Disentangling Geometry and Appearance. In: Advances in Neural Information Processing Systems (NeurIPS) (2020) 4
47. Yu, A., Fridovich-Keil, S., Tancik, M., Chen, Q., Recht, B., Kanazawa, A.: Plenoxels: Radiance Fields without Neural Networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2022) 2, 4, 5, 10, 13
48. Yu, A., Li, R., Tancik, M., Li, H., Ng, R., Kanazawa, A.: PlenOctrees for Real-time Rendering of Neural Radiance Fields. In: IEEE International Conference on Computer Vision (ICCV) (2021) 2, 4, 5
49. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In: CVPR (2018) 10