Pano-NeRF: Synthesizing High Dynamic Range Novel Views with Geometry from Sparse Low Dynamic Range Panoramic Images

Zhan Lu[1], Qian Zheng[2 3]*, Boxin Shi[4 5], Xudong Jiang[1]

[1] School of Electrical and Electronic Engineering, Nanyang Technological University
[2] College of Computer Science and Technology, Zhejiang University
[3] The State Key Lab of Brain-Machine Intelligence, Zhejiang University
[4] National Key Laboratory for Multimedia Information Processing, School of Computer Science, Peking University
[5] National Engineering Research Center of Visual Technology, School of Computer Science, Peking University
zhan007@e.ntu.edu.sg, qianzheng@zju.edu.cn, shiboxin@pku.edu.cn, exdjiang@ntu.edu.sg

Abstract

Panoramic imaging research on geometry recovery and High Dynamic Range (HDR) reconstruction becomes a trend with the development of Extended Reality (XR). Neural Radiance Fields (NeRF) provide a promising scene representation for both tasks without requiring extensive prior data. However, in the case of inputting sparse Low Dynamic Range (LDR) panoramic images, NeRF often degrades with under-constrained geometry and is unable to reconstruct HDR radiance from LDR inputs. We observe that the radiance from each pixel in panoramic images can be modeled as both a signal to convey scene lighting information and a light source to illuminate other pixels. Hence, we propose the irradiance fields from sparse LDR panoramic images, which increases the observation counts for faithful geometry recovery and leverages the irradiance-radiance attenuation for HDR reconstruction. Extensive experiments demonstrate that the irradiance fields outperform state-of-the-art methods on both geometry recovery and HDR reconstruction and validate their effectiveness. Furthermore, we show a promising byproduct of spatially-varying lighting estimation. The code is available at https://github.com/Lu-Zhan/Pano-NeRF.
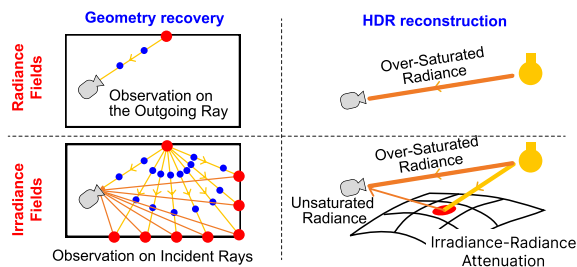
Figure 1: Comparison between the existing radiance fields (top) and the proposed irradiance fields (bottom) on geometry recovery (left) and HDR reconstruction (right) from sparse LDR panoramic images. Radiance fields suffer from poor geometry recovery due to a few observation counts (number of blue dots) and cannot reconstruct HDR radiance with LDR inputs. In contrast, the proposed irradiance fields recover faithful geometry by increasing the observation counts (number of blue dots) from incident rays (yellow lines) and infer over-saturated radiance from the unsaturated area by considering irradiance-radiance attenuation.

Introduction

Panoramic imaging stands as a trend with the rise of extended reality (XR) for achieving immersive experiences, such as virtual walks in 360° scenes and inserting virtual objects with 360° lighting information. These XR applications motivate panoramic imaging techniques, particularly in the tasks of geometry recovery (da Silveira et al. 2022) and High Dynamic Range (HDR) reconstruction (Yu et al. 2021b) under a panoramic scene. Previous research yields promising outcomes via supervised learning on extensive pre-collected datasets or under specific predetermined conditions, *e.g.*, a stereo camera system for geometry recovery (Wang et al. 2020b) or controllable multi-exposure for HDR reconstruction (Shen et al. 2011). However, the efficacy of these approaches is profoundly influenced by the quality of the underlying training data and the imposed conditions.

Recently, Neural Radiance Fields (NeRF) (Mildenhall et al. 2021) emerged with a promising scene representa-

tion to recover geometry and radiance information through self-supervised training from multi-view images, which avoids extensive pre-collected data requirement. Later research on sparse-view NeRF further improves the practicability of NeRF technique by reducing the number of multi-view images, where the recovered geometry is constrained by several priors, *e.g.*, depth (Niemeyer et al. 2022), visibility (Somraj and Soundararajan 2023), and semantic feature (Jain, Tancik, and Abbeel 2021). Besides, several NeRF-based models exhibit the capability to reconstruct HDR radiance by requiring casual multi-exposed LDR images (Huang et al. 2022; Jun-Seong et al. 2022) or a pre-trained HDR reconstruction model (Gera et al. 2022).

However, NeRF-based models encounter two primary limitations: 1) existing sparse-view NeRF might fail to recover accurate geometry for a panoramic scene by using priors from objects rather than a panoramic scene. And the scale variety of nearby/far objects in the panoramic scene further increases the trouble (Barron et al. 2022); 2) they are unable to reconstruct HDR radiance from LDR inputs captured under a fixed exposure, due to the lack of a mechanism

---

to address the ill-posed problem of HDR reconstruction.

Conversely, panoramic images exhibit a remarkable feature, where the radiance emitted by each pixel serves both as **a signal**, conveying scene lighting information to cameras through intrinsic factors (*e.g.*, position, surface normal, and albedo) and as **a source light**, illuminating other pixels within the scene. Based on this observation, this paper introduces irradiance fields to model the inter-reflection in panoramic scenes, via surface rendering (Kajiya 1986). The irradiance fields consider the irradiance received from all incident light directions upon a given surface point and integrate them with the intrinsic factors of the surface point to yield the observed outgoing radiance. Hence, the irradiance fields bring a distinct capability to recover faithful geometry through the augmentation of the observation counts for volumetric particles with sparse inputs and reconstruct HDR radiance from multi-view LDR inputs by considering irradiance-radiance attenuation, shown in Figure 1. Furthermore, considering the feature of the radiance within panoramic images, the proposed irradiance fields could be within the existing radiance fields. Therefore, we integrate our irradiance fields into radiance fields and perform joint optimization. Our contributions are summarized as

- We propose the irradiance fields from sparse LDR panoramic images and explain how irradiance fields contribute to geometry recovery and HDR reconstruction.
- We demonstrate the irradiance fields could be integrated into and jointly optimized with radiance fields, based on which we propose Pano-NeRF for geometry recovery and HDR novel views synthesis.
- We show that Pano-NeRF achieves state-of-the-art performance regarding geometry recovery and HDR reconstruction. We further provide a byproduct of spatially varying lighting estimation.

Related Work

Panoramic imaging. We introduce previous works on geometry recovery and HDR reconstruction dealing with panoramic images. For geometry recovery, (Wang et al. 2020b) requires a stereo panoramic camera setup. Several works estimate depth from the video recorded under free viewpoints (Im et al. 2016; Zhang et al. 2020). Recent works focus on the single-image panoramic geometry recovery via deep learning techniques and share the idea of the fusion of perspective depth estimation for a panoramic image (Wang et al. 2020a; Yun, Lee, and Rhee 2022; Zhuang et al. 2022; Ai et al. 2023). For HDR reconstruction, previous methods require multi-exposure LDR images for static poses (Debevec, Taylor, and Malik 1996) or dynamic camera poses (Chen et al. 2021b). Deep learning boosts the research on single-image HDR reconstruction on perspective images (Li and Fang 2019; Liu et al. 2020; Santos, Ren, and Kalantari 2020) or panoramic images (Yu et al. 2021b)). Our method only requires sparse LDR panoramic images for geometry recovery and HDR reconstruction, to avoid extensive prior data or imposed conditions.

Neural radiance fields. For geometry recovery by NeRF-based models, we only study sparse-view conditions. Several works solve this problem by aggregating prior knowledge through pre-trained conditional radiance fields, *e.g.*, deep features (Yu et al. 2021a; Chibane et al. 2021), and 3D cost volume (Chen et al. 2021a). Other works constrain appearance consistency on the seen/unseen views to avoid pre-training (Niemeyer et al. 2022; Wynn and Turmukhambetov 2023). However, this work still recovers a blurry geometry since appearance regularizers cannot directly constrain the scene geometry. Recent efforts on applying geometry priors would improve the recovered geometry, such as pre-estimated depth (DS-NeRF (Deng et al. 2022) and DD-NeRF (Roessle et al. 2022)), depth smoothness (RegNeRF (Niemeyer et al. 2022) and FreeNeRF (Yang, Pavone, and Wang 2023)), and geometry visibility (Ref-NeRF (Verbin et al. 2022) and ViP-NeRF (Somraj and Soundararajan 2023)). 360FusionNeRF (Kulkarni, Yin, and Scherer 2023) requires a single RGBD panoramic image and trains NeRF by warping the RGBD inputs. IndoorPanoDepth (Chang, Zhang, and Xiong 2023) builds a Signed Distance Field (SDF) and recovers geometry at the input views from sparse panoramic images by introducing a well geometry initialization. For HDR reconstruction, Raw-NeRF (Mildenhall et al. 2022) proposes a model working with raw images for HDR radiance fields. HDR-NeRF (Huang et al. 2022) builds HDR radiance fields with multi-exposure LDR inputs and paired exposure time. HDR-Plenoxels (Jun-Seong et al. 2022) reduces the exposure time requirement and conducts self-calibrated for multi-exposure inputs. PanoHDR-NeRF (Gera et al. 2022) reconstructs HDR radiance from pre-estimated HDR panoramic images by an existing panoramic HDR reconstruction method. The proposed irradiance fields help the geometry recovery and HDR reconstruction model by modeling inter-reflection within a panoramic scene, which is free of any pre-trained model, prior training data, or extra information.

Implicit reflectance representation. Previous methods propose reflectance fields to estimate intrinsic factors under several conditions, *e.g.*, known lighting (Bi et al. 2020; Srinivasan et al. 2021), or known geometry((Yao et al. 2022)). Some methods distill the trained radiance fields to bake intrinsic factors and lighting conditions (NeRD (Boss et al. 2021) and NeRFactor (Zhang et al. 2021)), or indirect illumination (Zhang et al. 2022). Besides, TexIR (Li et al. 2023) models the irradiance information with several HDR panoramic images (*i.e.*, 14), pre-estimated mesh-based geometry, and semantic priors. Our irradiance fields only require LDR inputs without any prior knowledge and exhibit the simplicity of sharing the scene representation with the same radiance fields.

Irradiance Fields

Modeling Irradiance Fields

Preliminary of radiance fields. The radiance fields in NeRF (Mildenhall et al. 2021) assume that the scene is composed of a cloud of light-emitting particles and the volumetric particles have emission and transmittance to pass through the radiance from other volumetric particles (Tagliasacchi and Mildenhall 2022), as shown in Figure 2 (a). The outgo-
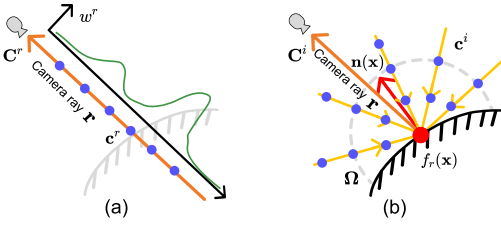
Figure 2: An illustration of different outgoing radiance modeling between the radiance fields and the proposed irradiance fields. (a) Radiance fields integrate the radiance $\mathbf{c}^r$ of each volumetric particle (blue dots) with weight w^r along the camera ray $\mathbf{r}$ (original line). (b) Irradiance fields integrate the radiance $\mathbf{c}^i$ from the incident light directions (yellow lines) with weight $f_r(\mathbf{x}, \boldsymbol{\omega}_o, \boldsymbol{\omega}_i) \cdot \boldsymbol{\omega}_i \mathbf{n}^\top$.

ing radiance $\mathbf{C}^r$ of position $\mathbf{x}$ is computed through volume rendering, *i.e.*, integrate radiance $\mathbf{c}^r$ of different volumetric particles (weighted by w^r) distributed along the camera ray $\mathbf{r}$ between viewpoint $\mathbf{o}$ and $\mathbf{x}$. Due to the unavailability of $\mathbf{x}$, $\mathbf{r}$ is formulated as $\mathbf{r}(t) = \mathbf{o} + t\boldsymbol{\omega}_o, t \in [t_n, t_f]$ with view direction $\boldsymbol{\omega}_o$, and near/far bounds t_n/t_f. The calculation of outgoing radiance $\mathbf{C}^r$ is formulated as[1],

$$\mathbf{C}^r(\mathbf{x}) = \mathbf{C}^r(\mathbf{r}) = \int_{t_n}^{t_f} w^r(\mathbf{o}, \boldsymbol{\omega}_o, t) \cdot \mathbf{c}^r(\mathbf{o}, \boldsymbol{\omega}_o, t)dt,$$

$$\text{s.t.} w^r(\mathbf{o}, \boldsymbol{\omega}_o, t) = \exp(-\int_{t_n}^{t} \sigma(\mathbf{o}, \boldsymbol{\omega}_o, t)dt) \cdot \sigma(\mathbf{o}, \boldsymbol{\omega}_o, t),$$

$$(1)$$

where $\mathbf{c}^r$ represents the radiance color of each volumetric particle that determines emission, and w^r is the weight calculated based on the density σ of volumetric particles. The radiance fields in NeRF (Mildenhall et al. 2021) are implemented based on a multilayer perception (MLP) network,

$$\text{MLP}(\mathbf{r}) \rightarrow (\mathbf{c}^r, \sigma), \qquad (2)$$

Irradiance fields formulation. Our irradiance fields assume that the scene is composed of several surface pieces with reflection effect only and the radiance is the interaction result of incoming light, reflectance, and surface normal, as shown in Figure 2 (b). Different from the radiance fields, our irradiance fields compute the outgoing radiance of position $\mathbf{x}$ through surface rendering (Kajiya 1986), *i.e.*, integrate irradiance $\boldsymbol{\omega}_i \mathbf{n}^\top \cdot \mathbf{c}^i$ around different incoming light directions $\boldsymbol{\omega}_i$ distributed on sphere $\boldsymbol{\Omega}$ centered at $\mathbf{x}$,

$$\mathbf{C}^i(\mathbf{x}) = \int_{\boldsymbol{\omega}_i \in \boldsymbol{\Omega}} f_r(\mathbf{x}, \boldsymbol{\omega}_o, \boldsymbol{\omega}_i) \cdot \boldsymbol{\omega}_i \mathbf{n}^\top \cdot \mathbf{c}^i(\mathbf{x}, \boldsymbol{\omega}_i)d\boldsymbol{\omega}_i, \quad (3)$$

where $\mathbf{c}^i$ indicates the radiance from each incoming light direction and f_r is the bidirectional reflection distribution function (BRDF), $\mathbf{n}$ is the surface normal at point $\mathbf{x}$. Note $\boldsymbol{\omega}_i \mathbf{n}^\top$ is replaced as $\max(\boldsymbol{\omega}_i \mathbf{n}^\top, 0)$ in the implementation.
Geometry recovery with sparse inputs. As illustrated in Figure 1 (left-top), sparse inputs bring a small number of

---

[1]We provide a more comprehensive formulation showing the relationship between different variables according to the implementation in (Mildenhall et al. 2021).

observation counts, which might cause unconstrained geometry (Niemeyer et al. 2022). As can be observed from Figure 1 (left-bottom), the irradiance fields generate several rays along the incident light direction $\boldsymbol{\omega}_i$ for each position $\mathbf{x}$. The $\mathbf{c}^i$ is equal to the radiance $\mathbf{C}^r$ of each incident ray that could be calculated in Equation (1). Thus, the proposed irradiance fields are expected to increase the observation counts for volumetric particles (Figure 1 (left)) by considering the inter-reflection between irradiance and outgoing radiance. Although $\mathbf{c}^i$ is not directly supervised, its integral is constrained through the training error between the observed radiance $\mathbf{C}_{\text{gt}}$ and $\mathbf{C}^i$, hence facilitates the optimization of σ and $\mathbf{c}^r$. Therefore, the proposed irradiance fields could achieve faithful geometry recovery even with sparse inputs.
HDR reconstruction with LDR inputs. The HDR radiance could be well restored from LDR one if it is unsaturated (*e.g.*, applying inverse tone mapping (Rempel et al. 2007)). The primary difficulty of HDR reconstruction is from oversaturated regions, where the clipping operation drops vital information. Fortunately, over-saturated regions do not always cover the whole panoramic image (*e.g.*, most are light sources in the indoor scene), and the proposed irradiance fields could leverage the radiance in unsaturated regions to reconstruct HDR radiance. As illustrated in Figure 1 (right), the basic idea is to back-propagate the training error (from points in unsaturated regions) between the observed radiance $\mathbf{C}_{\text{gt}}$ (from input images) and $\mathbf{C}^i$ to optimize HDR $\mathbf{c}^i$ in over-saturated regions, based on Equation (3). This can be achieved due to the attenuation effect brought by the BRDF (*i.e.*, $\phi < 1$) and the cosine of the incident angle (*i.e.*, $\boldsymbol{\omega}_i \mathbf{n}^\top < 1$). Additionally, $\mathbf{c}^i$ in over-saturated regions could be further optimized based on the observed radiance $\mathbf{C}^r$ according to Equation (8).

Optimizing Irradiance Fields

This section introduces how to integrate our irradiance fields into the radiance fields and perform joint optimization. Specifically, we focus on the calculation of outgoing radiance $\mathbf{C}^i$ based in Equation (3).
Obtaining intrinsic factors. We follow a numerical solution (Zhang et al. 2021) sharing a similar idea as volume rendering (Kajiya and Von Herzen 1984) in Equation (1) to calculate the position of surface point $\mathbf{x}$ by

$$\mathbf{x} = \mathbf{o} + (\int_{t_n}^{t_f} w^r(\mathbf{o}, \boldsymbol{\omega}_o, t) \cdot tdt)\boldsymbol{\omega}_o, \qquad (4)$$

where $w^r(\mathbf{o}, \boldsymbol{\omega}_o, t)$ is same as that in Equation (1).

As suggested by many scene understanding works (*e.g.*, (Xing et al. 2018; Zhou, Yu, and Jacobs 2019; Wang et al. 2021)), we assume the BRDF in our irradiance fields to be diffuse (*i.e.*, Lambertain model), *i.e.*, f_r only depends on the position of 3D point $\mathbf{x}$ while being free from the view direction,

$$f_r(\mathbf{x}) = f_r(\mathbf{r}) = \int_{t_n}^{t_f} w^r(\mathbf{o}, \boldsymbol{\omega}_o, t) \cdot \phi(\mathbf{o}, \boldsymbol{\omega}_o, t)dt, \quad (5)$$

where $\phi_r(\mathbf{o}, \boldsymbol{\omega}_o, t)$ is the albedo for a volumetric particle. Though the diffuse assumption limits the expression of non-Lambertain surfaces, it affects less since 1) non-Lambertian
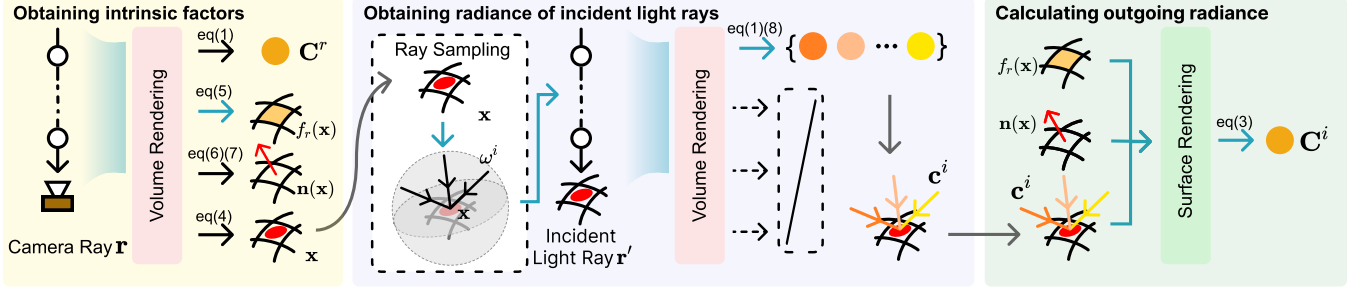
Figure 3: Illustration of calculating the outgoing radiance $\mathbf{C}^i$ from the conventional radiance fields. Firstly, we obtain the intrinsic factors on a surface point by integrating the BRDF output Φ, derived density $d\sigma$, and distance t along the camera ray $\mathbf{r}$ via volume rendering, as same as the computation of radiance $\mathbf{C}^r$. Secondly, We sample the incident light rays $\mathbf{r}'$ on the surface point $\mathbf{x}$ from directions distributed uniformly on the sphere centered at $\mathbf{x}$, and then compute the radiance $\mathbf{c}^i$ of sampled incident light rays. At last, we integrate obtained BRDF $f_r(\mathbf{x})$, surface normal $\mathbf{n}(\mathbf{x})$, and incident light rays (with $\mathbf{c}^i$ and direction ω^i) to calculate outgoing radiance $\mathbf{C}^i$.

surfaces occupy less area than the Lambertian ones in practice and 2) we leverage the outputs from radiance fields (non-Lambertain model) as the reconstructed results.

According to (Verbin et al. 2022), the density $\sigma(\mathbf{o}, \boldsymbol{\omega}_o, t)$ for each volumetric particle can be leveraged to calculate its surface normal $\mathbf{n}(\mathbf{o}, \boldsymbol{\omega}_o, t)$,

$$\mathbf{n}(\mathbf{o}, \boldsymbol{\omega}_o, t) = -\frac{d\sigma(\mathbf{o}, \boldsymbol{\omega}_o, t)}{\|d\sigma(\mathbf{o}, \boldsymbol{\omega}_o, t)\|}, \tag{6}$$

and the surface normal for position $\mathbf{x}$ can be numerically calculated by

$$\mathbf{n}(\mathbf{x}) = \int_{t_n}^{t_f} w^r(\mathbf{o}, \boldsymbol{\omega}_o, t) \cdot \mathbf{n}(\mathbf{o}, \boldsymbol{\omega}_o, t) dt, \tag{7}$$

Note that each surface normal is normalized to be a unit vector before further calculation.

Obtain radiance of incident light rays. As shown in Figure 3, we first sample the incident light directions $\boldsymbol{\omega}_i$ on the surface point $\mathbf{x}$ to determine the incident light rays $\mathbf{r}'(t) = \mathbf{x} + t\boldsymbol{\omega}_i$, then obtain the radiance $\mathbf{c}^i(\mathbf{x}, \boldsymbol{\omega}_i)$,

$$\mathbf{c}^i(\mathbf{x}, \boldsymbol{\omega}_i) = \mathbf{C}^r(\mathbf{r}'), \tag{8}$$

Calculating outgoing radiance. The outgoing radiance $\mathbf{C}^i(\mathbf{x})$ is then calculated with the estimated BRDF f_r, surface normal $\mathbf{n}(\mathbf{x})$ and a grouped radiance $\{\mathbf{c}^i\}$ of incident light rays via surface rendering in Equation (3).

In summary, the proposed irradiance fields can be jointly optimized with radiance fields since all variables except ϕ can be directly calculated from radiance fields. We modify the MLP in a NeRF-based method to output an additional variable ϕ, and Equation (2) is revised as

$$\text{MLP}(\mathbf{r}) \rightarrow (\mathbf{c}^r, \sigma, \phi), \tag{9}$$

Pano-NeRF

The overview of irradiance fields is displayed in Figure 3.
Network structure. We build our irradiance fields upon the model of Mip-NeRF (Barron et al. 2021) and modify it in the following aspects. First, we increase the output range of $\mathbf{c}^r$

to consider HDR radiance, by changing the output activation function of ReLU to Softplus. Second, we add output $\phi \in \mathbf{R}^3$ to the MLP for albedo estimation.

Geometry prior. The recovered geometry in the radiance fields often reveals a thick surface ('fog' as demonstrated in (Verbin et al. 2022), even with dense inputs) and rough surface (Niemeyer et al. 2022). Such geometry might limit the reconstruction of irradiance fields as the irradiance fields require a rough geometry for ray sampling and surface rendering computation. Therefore we adopt the geometry prior $\mathcal{R}_v$ indicated in (Verbin et al. 2022) to produce thin and smooth surface, that is,

$$\mathcal{R}_v =$$
$$\sum_{\mathbf{o}, \boldsymbol{\omega}_o} \left(\int_{t_n}^{t_f} w^r(\mathbf{o}, \boldsymbol{\omega}_o, t) \cdot \max(-\boldsymbol{\omega}_o \mathbf{n}^\top(\mathbf{o}, \boldsymbol{\omega}_o, t), 0)^2 dt \right), \tag{10}$$

The minimization of $\mathcal{R}_v$ ensures a thin and smooth surface by preventing the model from expressing too many visible volumetric particles around the surface.

Albedo priors. As we only leverage the observed pixel color for optimization, the freedom of the new variable ϕ might cause the radiance color distortion as the neural network is more easily to express color variety by ϕ rather than the irradiance. To mitigate the impact from ϕ, we take two priors to constrain the albedo estimation of our model. First, we limit the range of the estimated ϕ with $[0.03, 0.8]$ as suggested in (Ward and Shakespeare 1998). Second, we encourage the chromaticity of the estimated f_r to be similar to that of corresponding observed radiance $\mathbf{C}_{\text{gt}}$,

$$\mathcal{R}_c = \sum_{\mathbf{x} \in \Phi} \left\| \frac{f_r(\mathbf{x})}{\|f_r(\mathbf{x})\|} - \frac{\mathbf{C}_{\text{gt}}(\mathbf{x})}{\|\mathbf{C}_{\text{gt}}(\mathbf{x})\|} \right\|_2^2, \tag{11}$$

where set Φ contains positions of all surface points in a scene. Such constraints of the albedo are able to maintain the chromaticity of the reconstructed radiance color, which is enough to produce pleasurable results.

Loss function. We adopt the same loss function from Mip-NeRF to constrain the estimated radiance via volume render-

ing. To convert HDR radiance output to LDR value, we apply empirical tone-mapping (Arrighetti 2017; Chen, Wang, and Liu 2022), clipping (into $[0, 1]$), and gamma correction (factor of 2.2) on the output radiance to obtain mapped LDR value. MSE loss $\mathcal{L}_r$ in the radiance fields is calculated in coarse and fine phases,

$$\mathcal{L}_r = \sum_{\mathbf{o}, \boldsymbol{\omega}_o} (\alpha_1 \left\| g(\mathbf{C}_c^r(\mathbf{o}, \boldsymbol{\omega}_o)) - \mathbf{C}_{\text{gt}}(\mathbf{o}, \boldsymbol{\omega}_o) \right\|_2^2 + \tag{12}$$
$$\left\| g(\mathbf{C}_f^r(\mathbf{o}, \boldsymbol{\omega}_o)) - \mathbf{C}_{\text{gt}}(\mathbf{o}, \boldsymbol{\omega}_o) \right\|_2^2),$$

where $\mathbf{C}_c^r$ and $\mathbf{C}_f^r$ are the estimated radiance in coarse and fine phases, respectively. $g(\cdot)$ represents the process of HDR to LDR conversion. α_1 is set to 0.1 as in Mip-NeRF.

The MSE loss $\mathcal{L}_i$ for the irradiance fields is calculated only in the fine phase as,

$$\mathcal{L}_i = \sum_{\mathbf{o}, \boldsymbol{\omega}_o} \left\| g(\mathbf{C}^i(\mathbf{o}, \boldsymbol{\omega}_o)) - \mathbf{C}_{\text{gt}}(\mathbf{o}, \boldsymbol{\omega}_o) \right\|_2^2, \tag{13}$$

The overall loss function is calculated with the geometry prior $\mathcal{R}_v$, and the albedo prior $\mathcal{R}_c$,

$$\mathcal{L} = \mathcal{L}_r + \alpha_2 \mathcal{L}_i + \alpha_3 \mathcal{R}_v + \alpha_4 \mathcal{R}_c, \tag{14}$$

We empirically set $\alpha_2 = 1$, $\alpha_3 = 0.1$, and $\alpha_4 = 1$ to balance loss scale and stabilize the model training.

Experiments

Setup

Synthetic data. We render evaluation data from 5 3D scene models, including 'classroom', 'barbershop', 'living room', 'bedroom', and 'gallery'. Besides, we collect the data from the 3D scanned HDR dataset Replica (Straub et al. 2019) and select 8 enclosed scenes, including 'apartment', 'hotel', 'bathroom', 'office-0', 'office-1', 'office-4', 'room-0', and 'room-1'. For each scene, we sample 100 camera poses and render the LDR panoramic image and the paired ground truths (*i.e.*, HDR panoramic image, depth maps, and surface normal map) at each sampled camera pose.

Real captured data. We capture 30 HDR panoramic images in 3 real scenes, including 'real-meeting room', 'real-bedroom', and 'real-classroom', representing the large, medium, and small indoor spaces. Besides, we scan each scene with the LiDAR camera in iPhone 13 Pro to obtain the depth and surface normal for reference. We implement existing tool openMVG[2] to calibrate camera poses and merge HDR panoramic images with 9 bracketed exposures.

Evaluation Settings To evaluate irradiance fields with sparse inputs, for each time training, we randomly select 3 LDR panoramic images to train models as the worst case in (Niemeyer et al. 2022). After optimization, the rest panoramic images are used for evaluation. We repeat this procedure 10 times for each scene and compute the mean performance to alleviate the randomness of the results. The resolution of the panoramic image is set to 128×256.

Training details The sample numbers of volumetric particles along a camera ray and an incoming lighting ray are 64

Table 1: Comparison of geometry recovery and novel view synthesis. **Bold** and underline numbers indicate the best and the second best results, respectively.

Methods	Synthetic					Real		
	Depth	Normal	Image			Image		
	RMSE ↓	MAE ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Omnifusion	1.13	-	-	-	-	-	-	-
IndoorPanoDepth	1.29	45.35	18.57	0.61	0.43	15.91	0.59	0.47
Mip-NeRF	1.06	65.44	19.91	0.66	0.41	16.77	0.60	0.45
Mip-NeRF w/ $\mathcal{R}_v$	1.05	<u>36.81</u>	20.54	0.69	0.40	17.23	0.62	0.42
RegNeRF	<u>0.77</u>	63.98	<u>22.22</u>	<u>0.74</u>	0.34	<u>19.88</u>	0.68	0.36
FreeNeRF	1.88	66.76	21.23	0.68	<u>0.32</u>	19.23	0.67	<u>0.35</u>
Ours	**0.72**	**29.03**	**23.10**	**0.78**	**0.31**	**21.08**	**0.76**	**0.33**

and 10, respectively. The sample number of directions for incoming lighting rays towards each surface point $\mathbf{x}$ is 80. We train Pano-NeRF with the following settings: 44k optimization iterations, using Adam (Kingma and Ba 2015) optimizer with hyper-parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-6}$, a log-linearly annealed learning rate from 2×10^{-4} to 2×10^{-5} with a warm-up phase of 2500 iterations. Our model and ablations take 512 as the batch size to keep a close number of sampled volumetric particles per iteration with Mip-NeRF (Barron et al. 2021), whose batch size is 4096. We optimize the only radiance fields at the first 8.8k iterations and then perform joint optimization on two fields.

Overall performance

Geometry Recovery from sparse inputs. To validate the effectiveness of geometry recovery, we compare our method with two panoramic depth estimation methods, including Omnifusion (Li et al. 2022) and IndoorPanoDepth (Chang, Zhang, and Xiong 2023). Omnifusion is a deep learning-based method trained with large-scale data, while Indoor-PanoDepth is an SDF-based method using the same sparse panoramic inputs as Pano-NeRF. We also add the two state-of-the-art sparse-view NeRF-based methods, *i.e.*, RegN-eRF (Niemeyer et al. 2022), FreeNeRF (Yang, Pavone, and Wang 2023). RegNeRF is trained only with its proposed geometry regularization since appearance regularization is not available in their released codes. We set the frequency attenuation as used for facing forward scenes as suggested in FreeNeRF. Besides, we take Mip-NeRF (Barron et al. 2021) and its ablated version Mip-NeRF w/ $\mathcal{R}_v$ as the baseline methods. We report the same geometry metrics used in Table 3, and LDR image quality metrics including PSNR, SSIM (Wang et al. 2004), and LPIPS (Zhang et al. 2018). Note that geometry metrics are only computed on synthetic datasets as the real one has no aligned ground truth.

As demonstrated in Table 1, Pano-NeRF outperforms others on both geometry recovery. Figure 4 demonstrate the qualitative comparison of geometry recovery, where we have the same conclusion that Pano-NeRF could achieve more faithful depth and accurate surface normal. Besides, Pano-NeRF provides the best performance on novel view synthesis, as shown in Table 1 and Figure 4 (estimating a clear panoramic image with fewer artifacts.). This reason could be that a well-recovered geometry is crucial to novel view synthesis in the case of sparse inputs. Furthermore, by comparing the Mip-NeRF and Mip-NeRF w/ $\mathcal{R}_v$, we observe that directly applying geometry prior $\mathcal{R}_v$ could only benefit
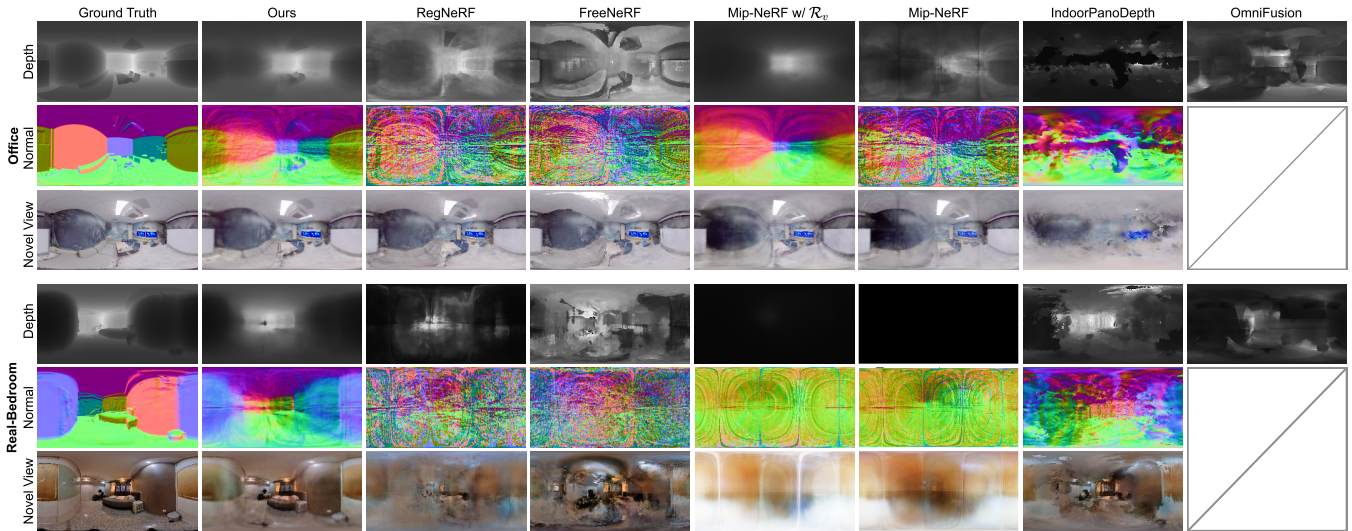
Figure 4: Comparison of geometry recovery in term of depth maps (1st&4th rows), surface normal maps (2nd&5th rows), and panoramic images (3rd&6th rows). The ground truth of depth and surface normal for the real scene is only for visual reference.
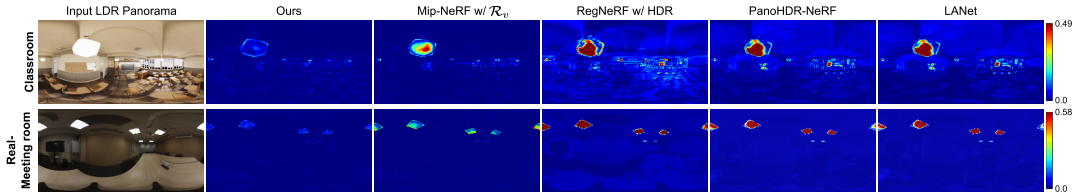

Figure 5: Comparison of HDR reconstruction on input views. The first column is the input LDR panoramic image, and 2nd to 5th columns show the error maps between the reconstructed HDR panoramic images and the ground truth.

to a smoother surface but no improvement on depth estimation or cause the severe degradation on novel views (shown in Figure 4). In contrast, the irradiance fields not only benefit from the smooth surface for better geometry but also prevent the collapse of geometry recovery. These results validate the proposed irradiance fields do help to improve the geometry recovery from sparse inputs.

HDR Reconstruction from LDR inputs. To validate the effectiveness of HDR reconstruction, we compare our method with the state-of-the-art panoramic HDR reconstruction method LANet (Yu et al. 2021b) and three NeRF-based methods, including PanoHDR-NeRF (Gera et al. 2022), RegNeRF (Niemeyer et al. 2022), and our baseline Mip-NeRF (Barron et al. 2021) with $\mathcal{R}_v$. Since RegNeRF is not implemented for the HDR reconstruction, we take pre-estimated HDR panoramic images from LANet for its training, namely 'RegNeRF w/ HDR'. In this experiment, only the reconstructed HDR panoramic images at input viewpoints are evaluated on each method to keep the same view requirement of LANet. We adopt commonly used HDR metrics for HDR image evaluation, including PU-PSNR (Azimi et al. 2021), PU-SSIM (Azimi et al. 2021), HDR-VDP3 (Mantiuk, Hammou, and Hanji 2023), and RMSE.

Table 2 demonstrates the quantitative comparison. LANet reports limited performance of HDR reconstruction, as they

Table 2: Comparison of HDR reconstruction on input views. **Bold** and underline numbers indicate the best and the second best results, respectively.

Methods	PU-PSNR ↑	PU-SSIM ↑	HDR-VDP3 ↑	RMSE ↓
LANet	31.9274	0.9317	7.7622	0.9195
PanoHDR-NeRF	33.0560	0.9415	7.9560	0.7343
RegNeRF w/ HDR	31.3735	0.8638	7.3694	0.9064
Mip-NeRF w/ $\mathcal{R}_v$	<u>41.6452</u>	<u>0.9875</u>	<u>9.6121</u>	<u>0.3110</u>
Ours	**43.4662**	**0.9899**	**9.6745**	**0.2686**

might suffer from poor generalization from their training data to testing scenarios. NeRF-based methods PanoHDR-NeRF and RegNeRF w/ HDR keep a close performance with LANet since they could only learn HDR information from the noisy HDR reconstructed results. Besides, RegNeRF w/ HDR drops significantly in terms of PU-SSIM, which indicates that the inconsistent HDR reconstruction of multi-views might cause degradation in the novel view synthesis. Mip-NeRF w/ $\mathcal{R}_v$ demonstrates better results as it takes advantage of the tone mapping procedure. In contrast, Pano-NeRF achieves the best performance on HDR reconstruction even compared with the baseline. Observed from in Figure 5, we find that Pano-NeRF accurately predicts the light source where often over-saturated as the HDR reconstruction results are close to the ground truth. The reason could be that
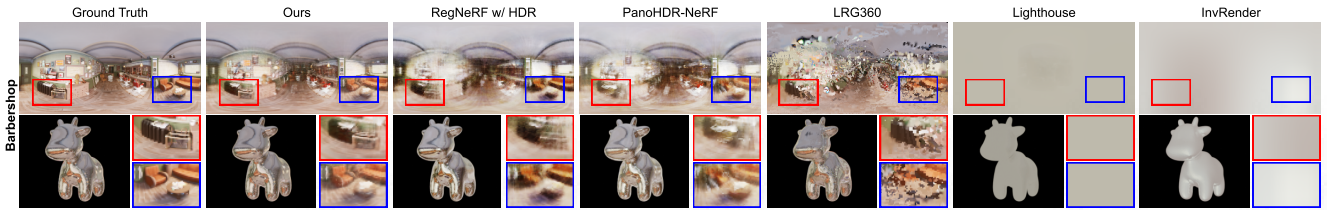
Figure 6: Comparison of HDR reconstruction on novel views and inserted mirror-like objects. Zoom in for better details.

Table 3: Validation of the proposed irradiance fields. **Bold** and underline numbers indicate the best and the second best results, respectively.

Methods	HDR-VDP3 ↑	Depth RMSE ↓	Normal MAE ↓
Mip-NeRF	7.1165	1.0576	65.4398
Mip-NeRF + $\mathcal{R}_v$	7.2328	1.0524	<u>36.8086</u>
Mip-NeRF + Irr	<u>7.5852</u>	<u>0.8565</u>	56.6353
Mip-NeRF + Irr + $\mathcal{R}_v$	**7.6693**	**0.8146**	**30.4295**

Figure 7: Comparison of the recovered geometry observed from an overhead angle. The blue point represents weight ω^r of the volumetric particle at that position (The brighter indicates a larger ω^r). The red line is the ground truth.

our irradiance fields constrain these over-saturated pixels by the information learned from the unsaturated area.

Validation of Irradiance Fields

To validate the effectiveness of the proposed irradiance fields, we take the Mip-NeRF (Barron et al. 2021) as our baseline method representing radiance fields. Then, we simply integrate the proposed irradiance fields into the baseline, namely 'Mip-NeRF + Irr', as the counterpart. Besides, we study the impact of the geometry prior $\mathcal{R}_v$ as it could be applied to both methods, namely 'Mip-NeRF + $\mathcal{R}_v$' and 'Mip-NeRF + Irr + $\mathcal{R}_v$', respectively. We report linear Root Mean Square Error (RMSE) for the estimated depth, Mean Angle Error (MAE) (in degree) for the estimated surface normal ('normal' for short), and HDR-VDP3 (Mantiuk, Hammou, and Hanji 2023) for HDR novel view synthesis. Note we conduct this experiment only on 13 synthetic scenes with the ground truth of geometry.

Observed from Table 3, irradiance fields benefit the performance on both tasks. Specifically, irradiance fields improve significantly in the depth estimation and HDR novel view quality by simply adding to the radiance fields and performing joint optimization. Besides, we find that $\mathcal{R}_v$ could further boost the performance of irradiance fields, in contrast, it only contributes to the better surface normal when applying to the radiance fields. We demonstrate an example of reconstructed geometry in Figure 7. We found that with the help of irradiance fields, the reconstructed geometry would be more accurate to the ground truth than that in radiance fields (Mip-NeRF) from sparse inputs. Besides, $\mathcal{R}_v$ could lead to a smoother surface as shown in Figure 7 but might significantly degrade the depth estimation, which is not suitable for novel view synthesis with an unconstrained depth. In contrast, the irradiance fields benefit from the smooth surface from $\mathcal{R}_v$ and predict a more faithful depth.
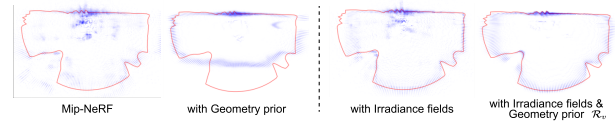
Byproduct of Spatially-varying Lighting Estimation

We demonstrate that Pano-NeRF could be further used for spatially varying lighting estimation, as a byproduct of HDR novel view synthesis. We compare Pano-NeRF with aforementioned NeRF-based methods and several spatially-varying lighting estimation methods, including InvIndoor (Li et al. 2020), Lighthouse (Srinivasan et al. 2020), and LRG360 (Li, Li, and Matsushita 2021)[3]. As shown in Figure 6, Lighthouse and InvRender produce low-frequency lighting maps, and LRG360 brings the artifacts into the lighting maps. For NeRF-based methods, RegNeRF w/ HDR and PanoHDR-NeRF keep similar results on HDR novel view synthesis, while they suffer from inconsistent HDR reconstruction for different views to produce blurry lighting maps. Conversely, Pano-NeRF outperforms others as it provides spatially-varying HDR lighting maps with the richest details close to the ground truth, which benefits more realistic mirror-like objects inserted with accurate specular.

Conclusion

In summary, this paper introduces irradiance fields as a novel solution for geometry recovery and HDR reconstruction in the context of sparse LDR panoramic images. The irradiance fields consider the inter-reflection in panoramic scenes, to recover faithful geometry from sparse inputs by increasing observation counts of volumetric particles and to reconstruct accurate HDR radiance from LDR inputs by yielding irradiance-radiance attenuation. We integrate the irradiance fields into a common scene representation of radiance fields and perform joint optimization. Extensive experiments demonstrate superior performance both on the tasks of geometry recovery and HDR reconstruction and validate the effectiveness of irradiance fields. We further provide a byproduct of spatially varying lighting estimation.

---

[3]Implementation is at https://github.com/Lu-Zhan/Pano-NeRF.

Acknowledgments

References

Ai, H.; Cao, Z.; Cao, Y.-P.; Shan, Y.; and Wang, L. 2023. HRDFuse: Monocular 360deg depth estimation by collaboratively learning holistic-with-regional depth distributions. In *CVPR*, 13273–13282.

Arrighetti, W. 2017. The academy color encoding system (ACES): A professional color-management framework for production, post-production and archival of still and motion pictures. *Journal of Imaging*, 3(4): 40.

Azimi, M.; et al. 2021. PU21: A novel perceptually uniform encoding for adapting existing quality metrics for HDR. In *2021 Picture Coding Symposium (PCS)*, 1–5. IEEE.

Barron, J. T.; Mildenhall, B.; Tancik, M.; Hedman, P.; Martin-Brualla, R.; and Srinivasan, P. P. 2021. Mip-NeRF: A multiscale representation for anti-aliasing neural radiance fields. In *ICCV*, 5855–5864.

Barron, J. T.; Mildenhall, B.; Verbin, D.; Srinivasan, P. P.; and Hedman, P. 2022. Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. In *CVPR*, 5470–5479.

Bi, S.; Xu, Z.; Srinivasan, P.; Mildenhall, B.; Sunkavalli, K.; Hašan, M.; Hold-Geoffroy, Y.; Kriegman, D.; and Ramamoorthi, R. 2020. Neural reflectance fields for appearance acquisition. *arXiv preprint arXiv:2008.03824*.

Boss, M.; Braun, R.; Jampani, V.; Barron, J. T.; Liu, C.; and Lensch, H. 2021. Nerd: Neural reflectance decomposition from image collections. In *ICCV*, 12684–12694.

Chang, W.; Zhang, Y.; and Xiong, Z. 2023. Depth estimation from indoor panoramas with neural scene representation. In *CVPR*, 899–908.

Chen, A.; Xu, Z.; Zhao, F.; Zhang, X.; Xiang, F.; Yu, J.; and Su, H. 2021a. MVSNeRF: Fast generalizable radiance field reconstruction from multi-view stereo. In *ICCV*, 14124–14133.

Chen, G.; Chen, C.; Guo, S.; Liang, Z.; Wong, K.-Y. K.; and Zhang, L. 2021b. HDR video reconstruction: A coarse-to-fine network and a real-world benchmark dataset. In *ICCV*, 2502–2511.

Chen, Z.; Wang, G.; and Liu, Z. 2022. Text2Light: Zero-shot text-driven HDR panorama generation. *ACM TOG*, 41(6).

Chibane, J.; Bansal, A.; Lazova, V.; and Pons-Moll, G. 2021. Stereo radiance fields (SRF): Learning view synthesis for sparse views of novel scenes. In *CVPR*, 7911–7920.

da Silveira, T. L.; Pinto, P. G.; Murrugarra-Llerena, J.; and Jung, C. R. 2022. 3D scene geometry estimation from 360 imagery: A survey. *ACM Computing Surveys*, 55(4): 1–39.

Debevec, P. E.; Taylor, C. J.; and Malik, J. 1996. Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach. In *SIGGRAPH*, 11–20.

Deng, K.; Liu, A.; Zhu, J.-Y.; and Ramanan, D. 2022. Depth-supervised NeRF: Fewer views and faster training for free. In *CVPR*, 12882–12891.

Gera, P.; Dastjerdi, M. R. K.; Renaud, C.; Narayanan, P.; and Lalonde, J.-F. 2022. Casual indoor HDR radiance capture from omnidirectional images. In *BMVC*.

Huang, X.; Zhang, Q.; Feng, Y.; Li, H.; Wang, X.; and Wang, Q. 2022. HDR-NeRF: High dynamic range neural radiance fields. In *CVPR*, 18398–18408.

Im, S.; Ha, H.; Rameau, F.; Jeon, H.-G.; Choe, G.; and Kweon, I. S. 2016. All-around depth from small motion with a spherical panoramic camera. In *ECCV*, 156–172. Springer.

Jain, A.; Tancik, M.; and Abbeel, P. 2021. Putting NeRF on a diet: Semantically consistent few-shot view synthesis. In *ICCV*, 5885–5894.

Jun-Seong, K.; Yu-Ji, K.; Ye-Bin, M.; and Oh, T.-H. 2022. HDR-Plenoxels: Self-calibrating high dynamic range radiance fields. In *ECCV*, 384–401. Springer.

Kajiya, J. T. 1986. The rendering equation. In *SIGGRAPH*, 143–150.

Kajiya, J. T.; and Von Herzen, B. P. 1984. Ray tracing volume densities. *ACM SIGGRAPH computer graphics*, 18(3): 165–174.

Kingma, D. P.; and Ba, J. 2015. Adam: A method for stochastic optimization. In *ICLR*.

Kulkarni, S.; Yin, P.; and Scherer, S. 2023. 360FusionNeRF: Panoramic neural radiance fields with joint guidance. In *IROS*, 7202–7209.

Li, J.; and Fang, P. 2019. HDRNET: Single-image-based HDR reconstruction using channel attention cnn. In *ICMSSP*, 119–124.

Li, J.; Li, H.; and Matsushita, Y. 2021. Lighting, reflectance and geometry estimation from 360 panoramic stereo. In *CVPR*, 10586–10595.

Li, Y.; Guo, Y.; Yan, Z.; Huang, X.; Duan, Y.; and Ren, L. 2022. Omnifusion: 360 monocular depth estimation via geometry-aware fusion. In *CVPR*, 2801–2810.

Li, Z.; Shafiei, M.; Ramamoorthi, R.; Sunkavalli, K.; and Chandraker, M. 2020. Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and svbrdf from a single image. In *CVPR*, 2475–2484.

Li, Z.; Wang, L.; Cheng, M.; Pan, C.; and Yang, J. 2023. Multi-view inverse rendering for large-scale real-world indoor scenes. In *CVPR*, 12499–12509.

Liu, Y.-L.; Lai, W.-S.; Chen, Y.-S.; Kao, Y.-L.; Yang, M.-H.; Chuang, Y.-Y.; and Huang, J.-B. 2020. Single-image HDR reconstruction by learning to reverse the camera pipeline. In *CVPR*, 1651–1660.

Mantiuk, R. K.; Hammou, D.; and Hanji, P. 2023. HDR-VDP-3: A multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content. *arXiv preprint arXiv:2304.13625*.

Mildenhall, B.; Hedman, P.; Martin-Brualla, R.; Srinivasan, P. P.; and Barron, J. T. 2022. NeRF in the dark: High dynamic range view synthesis from noisy raw images. In *CVPR*, 16190–16199.

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2021. NeRF: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99–106.

Niemeyer, M.; Barron, J. T.; Mildenhall, B.; Sajjadi, M. S.; Geiger, A.; and Radwan, N. 2022. RegNeRF: Regularizing neural radiance fields for view synthesis from sparse inputs. In *CVPR*, 5480–5490.

Rempel, A. G.; Trentacoste, M.; Seetzen, H.; Young, H. D.; Heidrich, W.; Whitehead, L.; and Ward, G. 2007. Ldr2Hdr: on-the-fly reverse tone mapping of legacy video and photographs. *ACM TOG*, 26(3).

Roessle, B.; Barron, J. T.; Mildenhall, B.; Srinivasan, P. P.; and Nießner, M. 2022. Dense depth priors for neural radiance fields from sparse input views. In *CVPR*, 12892–12901.

Santos, M. S.; Ren, T. I.; and Kalantari, N. K. 2020. Single image HDR reconstruction using a CNN with masked features and perceptual loss. *ACM TOG*, 39(4): 80–1.

Shen, R.; Cheng, I.; Shi, J.; and Basu, A. 2011. Generalized random walks for fusion of multi-exposure images. *IEEE TIP*, 20(12): 3634–3646.

Somraj, N.; and Soundararajan, R. 2023. ViP-NeRF: Visibility prior for sparse input neural radiance fields. In *ACM SIGGRAPH*.

Srinivasan, P. P.; Deng, B.; Zhang, X.; Tancik, M.; Mildenhall, B.; and Barron, J. T. 2021. NeRV: Neural reflectance and visibility fields for relighting and view synthesis. In *CVPR*, 7495–7504.

Srinivasan, P. P.; Mildenhall, B.; Tancik, M.; Barron, J. T.; Tucker, R.; and Snavely, N. 2020. Lighthouse: Predicting lighting volumes for spatially-coherent illumination. In *CVPR*, 8080–8089.

Straub, J.; Whelan, T.; Ma, L.; Chen, Y.; Wijmans, E.; Green, S.; Engel, J. J.; Mur-Artal, R.; Ren, C.; Verma, S.; Clarkson, A.; Yan, M.; Budge, B.; Yan, Y.; Pan, X.; Yon, J.; Zou, Y.; Leon, K.; Carter, N.; Briales, J.; Gillingham, T.; Mueggler, E.; Pesqueira, L.; Savva, M.; Batra, D.; Strasdat, H. M.; Nardi, R. D.; Goesele, M.; Lovegrove, S.; and Newcombe, R. 2019. The Replica dataset: A digital replica of indoor spaces. *arXiv preprint arXiv:1906.05797*.

Tagliasacchi, A.; and Mildenhall, B. 2022. Volume rendering digest (for NeRF). *arXiv preprint arXiv:2209.02417*.

Verbin, D.; Hedman, P.; Mildenhall, B.; Zickler, T.; Barron, J. T.; and Srinivasan, P. P. 2022. Ref-NeRF: Structured view-dependent appearance for neural radiance fields. In *CVPR*, 5491–5500.

Wang, F.-E.; Yeh, Y.-H.; Sun, M.; Chiu, W.-C.; and Tsai, Y.-H. 2020a. Bifuse: Monocular 360 depth estimation via bi-projection fusion. In *CVPR*, 462–471.

Wang, N.-H.; Solarte, B.; Tsai, Y.-H.; Chiu, W.-C.; and Sun, M. 2020b. 360SD-Net: 360 stereo depth estimation with learnable cost volume. In *2020 IEEE ICRA*, 582–588.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4): 600–612.

Wang, Z.; Philion, J.; Fidler, S.; and Kautz, J. 2021. Learning indoor inverse rendering with 3D spatially-varying lighting. In *ICCV*, 12538–12547.

Ward, G.; and Shakespeare, R. 1998. Rendering with radiance: The art and science of lighting visualization.

Wynn, J.; and Turmukhambetov, D. 2023. DiffusioNeRF: Regularizing neural radiance fields with denoising diffusion models. In *CVPR*.

Xing, G.; Liu, Y.; Ling, H.; Granier, X.; and Zhang, Y. 2018. Automatic spatially varying illumination recovery of indoor scenes based on a single RGB-D image. *IEEE TVCG*, 26(4): 1672–1685.

Yang, J.; Pavone, M.; and Wang, Y. 2023. FreeNeRF: Improving few-shot neural rendering with free frequency regularization. In *CVPR*, 8254–8263.

Yao, Y.; Zhang, J.; Liu, J.; Qu, Y.; Fang, T.; McKinnon, D.; Tsin, Y.; and Quan, L. 2022. NeILF: Neural incident light field for physically-based material estimation. In *ECCV*, 700–716. Springer.

Yu, A.; Ye, V.; Tancik, M.; and Kanazawa, A. 2021a. pixelNeRF: Neural radiance fields from one or few images. In *CVPR*, 4578–4587.

Yu, H.; Liu, W.; Long, C.; Dong, B.; Zou, Q.; and Xiao, C. 2021b. Luminance attentive networks for HDR image and panorama reconstruction. In *Computer Graphics Forum*, volume 40, 181–192. Wiley Online Library.

Yun, I.; Lee, H.-J.; and Rhee, C. E. 2022. Improving 360 monocular depth estimation via non-local dense prediction transformer and joint supervised and self-supervised learning. In *AAAI*, volume 36, 3224–3233.

Zhang, M.; Ye, X.; Fan, X.; and Zhong, W. 2020. Unsupervised depth estimation from monocular videos with hybrid geometric-refined loss and contextual attention. *Neurocomputing*, 379: 250–261.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*.

Zhang, X.; Srinivasan, P. P.; Deng, B.; Debevec, P.; Freeman, W. T.; and Barron, J. T. 2021. NeRFactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM TOG*, 40(6): 1–18.

Zhang, Y.; Sun, J.; He, X.; Fu, H.; Jia, R.; and Zhou, X. 2022. Modeling indirect illumination for inverse rendering. In *CVPR*, 18643–18652.

Zhou, H.; Yu, X.; and Jacobs, D. W. 2019. Glosh: Global-local spherical harmonics for intrinsic image decomposition. In *ICCV*, 7820–7829.

Zhuang, C.; Lu, Z.; Wang, Y.; Xiao, J.; and Wang, Y. 2022. ACDNet: Adaptively combined dilated convolution for monocular panorama depth estimation. In *AAAI*, volume 36, 3653–3661.