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

Zhan Lu¹, Qian Zheng²,³*, Boxin Shi⁴,⁵, Xudong Jiang¹

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.

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 representation 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 cap-
tured 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 radiance 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 (Roesle et al. 2022)), depth smoothness (RefNeRF (Niemeyer et al. 2022) and FreeNeRF (Yang, Pavone, and Wang 2023)), and geometry visibility (RefNeRF (Verbin et al. 2022) and ViP-NeRF (Sonrav 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. HDRPlenoxels (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 use reflectance fields to estimate intrinsic factors under 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 outgoing radiance \( C' \) of position \( x \) is computed through volume rendering, i.e., integrate radiance \( c' \) of different volumetric particles (weighted by \( w' \)) distributed along the camera ray \( r \) between viewpoint \( o \) and \( x \). Due to the unavailability of \( x, r \) is formulated as \( r(t) = o + tw_o, t \in [t_n, t_f] \) with view direction \( \omega_o \), and near/far bounds \( t_n, t_f \). The calculation of outgoing radiance \( C' \) is formulated as:

\[
C'(x) = \int_{t_n}^{t_f} w'(o, \omega_o, t) \cdot c'(o, \omega_o, t) dt,
\]

s.t. \( w'(o, \omega_o, t) = \exp(- \int_{t_n}^{t} \sigma(o, \omega_o, t) dt) \cdot \sigma(o, \omega_o, t) \),

where \( c' \) represents the radiance color of each volumetric particle that determines emission, and \( w' \) is the weight calculated based on the density \( \sigma \) of volumetric particles. The radiance fields in NeRF (Mildenhall et al. 2021) are implemented based on a multilayer perceptron (MLP) network,

\[
MLP(r) \rightarrow (c', \sigma),
\]

**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 \( x \) through surface rendering (Kajiya 1986), i.e., integrate radiance \( \omega_i n^\top \cdot c' \) around different incoming light directions \( n \) distributed on sphere \( \Omega \) centered at \( x \),

\[
C'(x) = \int_{\omega_i \in \Omega} f_r(x, \omega_o, \omega_i) \cdot \omega_i n^\top \cdot c'(x, \omega_i) d\omega_i,
\]

where \( c' \) indicates the radiance from each incoming light direction and \( f_r \) is the bidirectional reflection distribution function (BRDF). \( n \) is the surface normal at \( x \). Note \( \omega_i n^\top \) is replaced as \( \max(\omega_i n^\top, 0) \) in the implementation.  

**Geometry recovery with sparse inputs.** As illustrated in Figure 1 (left-top), sparse inputs bring a small number of 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 \( \omega_i \) for each position \( x \). The \( c' \) is equal to the radiance \( C' \) of each incident ray that could be calculated in Equation (1). Through the proposed irradiance fields, the radiance fields are expected to be increased to the observation counts for volumetric particles (Figure 1 (left)) by considering the inter-reflection between irradiance and outgoing radiance. Although \( c' \) is not directly supervised, its integral is constrained through the training error between the observed radiance \( C_{gt} \) and \( C' \), hence facilitates the optimization of \( \sigma \) and \( c' \). 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 over-saturated 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 \( C_{gt} \) (from input images) and \( C' \) to optimize HDR \( c' \) 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., \( \omega_i n^\top < 1 \)). Additionally, \( c' \) in over-saturated regions could be further optimized based on the observed radiance \( C' \) 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 \( C' \) 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 \( x \) by

\[
x = o + \left( \int_{t_n}^{t_f} w'(o, \omega_o, t) \cdot dt \right) \omega_o.
\]

where \( w'(o, \omega_o, t) \) 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 irradiance fields to be diffuse (Lambertian model), i.e., \( f_r \) only depends on the position of 3D point \( x \) while being free from the view direction,

\[
f_r(x) = f_r(r) = \int_{t_n}^{t_f} w'(o, \omega_o, t) \cdot \phi(o, \omega_o, t) dt,
\]

where \( \phi(o, \omega_o, t) \) is the albedo for a volumetric particle. Though the diffuse assumption limits the expression of non-Lambertian surfaces, it affects less since 1) non-Lambertian
The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24)

Obtaining intrinsic factors

Camera Ray \( r \)

Volume Rendering

\[
C^r \rightarrow \text{eq}(1)
\]

\[
C^r \rightarrow \text{eq}(5)
\]

\[
f_r(x) \rightarrow \text{eq}(6)\]

\[
f_r(x) \rightarrow \text{eq}(7)
\]

Obtaining radiance of incident light rays

Ray Sampling

Incident Light Ray \( r' \)

Volume Rendering

\[
x \rightarrow \text{eq}(8)
\]

\[
x \rightarrow \text{eq}(9)
\]

Calculating outgoing radiance

Surface Rendering

\[
C^i \rightarrow \text{eq}(3)
\]

\[
C^i \rightarrow \text{eq}(4)
\]

\[
C^i \rightarrow \text{eq}(10)
\]

Figure 3: Illustration of calculating the outgoing radiance \( C^i \) from the conventional radiance fields. Firstly, we obtain the intrinsic factors on a surface point by integrating the BRDF output \( \Phi \), derived density \( d\sigma \), and distance \( t \) along the camera ray \( r \) via volume rendering, as same as the computation of radiance \( C^r \). Secondly, We sample the incident light rays \( r' \) on the surface point \( x \) from directions distributed uniformly on the sphere centered at \( x \), and then compute the radiance \( c' \) of sampled incident light rays. At last, we integrate obtained BRDF \( f_r(x) \), surface normal \( n(x) \), and incident light rays (with \( c' \) and direction \( \omega' \)) to calculate outgoing radiance \( C^i \).

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

According to (Verbin et al. 2022), the density \( \sigma(o, \omega_o, t) \) for each volumetric particle can be leveraged to calculate its surface normal \( n(o, \omega_o, t) \),

\[
n(o, \omega_o, t) = -\frac{d\sigma(o, \omega_o, t)}{\|d\sigma(o, \omega_o, t)\|},
\]

and the surface normal for position \( x \) can be numerically calculated by

\[
n(x) = \int_{t_n}^{t_f} w^r(o, \omega_o, t) \cdot n(o, \omega_o, t) dt,
\]

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 \( \omega_i \) on the surface point \( x \) to determine the incident light rays \( r_i(t) = x + t\omega_i \), then obtain the radiance \( c'(x, \omega_i) \),

\[
c'(x, \omega_i) = C^r(r_i),
\]

Calculating outgoing radiance. The outgoing radiance \( C^i(x) \) is then calculated with the estimated BRDF \( f_r \), surface normal \( n(x) \) and a grouped radiance \( \{c^i\} \) 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 \( \phi \) can be directly calculated from radiance fields. We modify the MLP in a NeRF-based method to output an additional variable \( \phi \), and Equation (2) is revised as

\[
\text{MLP}(r) \rightarrow (C^r, \sigma, \phi),
\]

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

Geometry prior. The recovered geometry in the radiance fields often reveals a thick surface (‘foggy’ even with dense inputs (Verbin et al. 2022)) 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. Hence we adopt the geometry prior \( R_v \) indicated in (Verbin et al. 2022) to produce thin and smooth surface, that is

\[
R_v = \sum_{o, \omega_o, t} \left( \int_{t_n}^{t_f} w^r(o, \omega_o, t) \cdot \max(-\omega_o n^r(o, \omega_o, t), 0)^2 dt \right),
\]

The minimization of \( R_v \) 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 \( \phi \) might cause the radiance color distortion as the neural network is more easily to express color variety by \( \phi \) rather than the irradiance. To mitigate the impact from \( \phi \), we take two priors to constrain the albedo estimation of our model. First, we limit the range of the estimated \( \phi \) 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 \( C_{gt} \),

\[
R_c = \sum_{x \in \Phi} \left\| \frac{f_r(x)}{\|f_r(x)\|} - \frac{C_{gt}(x)}{\|C_{gt}(x)\|} \right\|_2^2,
\]

where set \( \Phi \) 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 rendering. 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 \( L_r \) for the radiance fields is calculated in coarse and fine phases as,
\[
L_r = \sum_{\mathbf{r}, \omega, o} (\alpha_1 || g(C^r_c(\mathbf{r}, \omega, o)) - C^r_g(\mathbf{r}, \omega, o)||_2^2 + || g(C^f_c(\mathbf{r}, \omega, o)) - C^f_g(\mathbf{r}, \omega, o)||_2^2),
\]
where \( C^r_c \) and \( C^f_c \) are the estimated radiance in coarse and fine phases, respectively. \( g(\cdot) \) represents the process of HDR to LDR conversion. \( \alpha_1 \) is set to 0.1 as in Mip-NeRF. The MSE loss \( L_i \) for the irradiance fields is calculated only in the fine phase as,
\[
L_i = \sum_{\mathbf{r}, \omega, o} || g(C^i_c(\mathbf{r}, \omega, o)) - C^i_g(\mathbf{r}, \omega, o)||_2^2,
\]
Therefore, the overall loss function is,
\[
L = L_r + \alpha_2 L_i + \alpha_3 R_v + \alpha_4 R_c,
\]
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 an open-source tool openMVG to calibrate camera poses and merge HDR panoramic images with 9 bracketed exposures.

**Evaluation settings** For each scene, we randomly select 3 LDR panoramic images to train models, and the rest are used for evaluation. We repeat this procedure 10 times and compute the mean performance to alleviate the randomness of the results. The resolution of images is set to 128 \( \times \) 256.

**Training details** The sample numbers of volumetric particles along a camera ray and an incoming lighting ray are 64 and 10, respectively. The sample number of directions for incoming lighting rays towards each surface point \( x \) is 80. We train Pano-NeRF with the following settings: 44k optimization iterations, using Adam (Kingma and Ba 2015) optimizer with hyper-parameters \( \beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8} \), a log-linearly annealed learning rate from \( 2 \times 10^{-4} \) to \( 2 \times 10^{-5} \) 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 MipNeRF (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.** 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 IndoorPanoDepth is an SDF-based method using the same sparse panoramic inputs. We additionally add the two state-of-the-art sparse-view NeRF-based methods, i.e., RegNeRF (Niemeyer et al. 2022), FreeNeRF (Yang, Pavone, and Wang 2023). Besides, we take Mip-NeRF (Barron et al. 2021) and its ablated version Mip-NeRF+\( R_v \) as the baseline methods. We report the geometry metrics, and LDR image quality metrics (including PSNR, SSIM (Wang et al. 2004), and LPIPS (Zhang et al. 2018)). Geometry metrics are not calculated for real scenes due to no aligned ground truth.

As shown in Table 1, Pano-NeRF outperforms others on the results of geometry recovery. Figure 4 demonstrates 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, PanoNeRF provides the best performance on novel view synthesis, as shown in Table 1 and Figure 4 (estimating a clear panoramic image with fewer artifacts.). The reason could be that a well-recovered geometry is crucial to novel view synthesis in the case of sparse inputs. Furthermore, by comparing the results of Omnifusion and IndoorPanoDepth, we find that our method can achieve more accurate geometry recovery, especially in the case of large indoor spaces. Figure 3 demonstrates the qualitative comparison of novel view synthesis, where we have the same conclusion that Pano-NeRF could achieve more realistic and accurate novel views compared to other methods.
Comparing the Mip-NeRF and Mip-NeRF+$\mathcal{R}_v$, we observe that directly applying geometry prior $\mathcal{R}_v$ could only benefit 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.

**HDR Reconstruction from LDR inputs.** 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+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 might suffer from poor generalization from their training data to testing scenarios. NeRF-based methods PanoHDR-NeRF and RegNeRF+HDR keep a close performance with LANet since they could only learn HDR information from the noisy HDR prediction. Besides, RegNeRF+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+$\mathcal{R}_v$ demonstrates better results as it takes advantage of the tone mapping operation. 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 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-
Table 2: Comparison of HDR reconstruction on input views.

<table>
<thead>
<tr>
<th>Methods</th>
<th>PU-PSNR↑</th>
<th>PU-SSIM↑</th>
<th>HDR-VDP3↑</th>
<th>RMSE↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>LANet</td>
<td>31.9274</td>
<td>0.9317</td>
<td>7.7622</td>
<td>0.9195</td>
</tr>
<tr>
<td>PanoHDR-NeRF</td>
<td>33.0560</td>
<td>0.9415</td>
<td>7.9560</td>
<td>0.7343</td>
</tr>
<tr>
<td>RegNeRF+HDR</td>
<td>31.3735</td>
<td>0.8638</td>
<td>7.3694</td>
<td>0.9064</td>
</tr>
<tr>
<td>Mip-NeRF+R_v</td>
<td>41.6452</td>
<td>0.9875</td>
<td>9.6121</td>
<td>0.3110</td>
</tr>
<tr>
<td>Ours</td>
<td>43.4662</td>
<td>0.9899</td>
<td>9.6745</td>
<td>0.2686</td>
</tr>
</tbody>
</table>

Table 3: Validation of the proposed irradiance fields.

<table>
<thead>
<tr>
<th>Methods</th>
<th>HDR-VDP3↑</th>
<th>D-RMSE↓</th>
<th>N-MAE↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mip-NeRF</td>
<td>7.1165</td>
<td>1.0576</td>
<td>65.4398</td>
</tr>
<tr>
<td>Mip-NeRF+R_v</td>
<td>7.2328</td>
<td>1.0524</td>
<td>36.8086</td>
</tr>
<tr>
<td>Mip-NeRF+Irr</td>
<td>7.5852</td>
<td>0.8565</td>
<td>56.6353</td>
</tr>
<tr>
<td>Mip-NeRF+Irr+R_v</td>
<td>7.6993</td>
<td>0.8146</td>
<td>30.4295</td>
</tr>
</tbody>
</table>

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 with NeRF-based methods and additional spatially-varying lighting estimation methods, including InvIndoor (Li et al. 2020), Lighthouse (Srinivasan et al. 2020), and LRG360 (Li, Li, and Matsushita 2021). Figure 6 shows that Lighthouse and InvRender produce low-frequency lighting maps, and LRG360 brings the artifacts into the lighting maps. RegNeRF+HDR and PanoHDR-NeRF keep similar results and 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 with 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 the scene representation of radiance fields and perform joint optimization. Extensive experiments validate the effectiveness of irradiance fields and demonstrate superior performance on geometry recovery and HDR reconstruction. We further investigate the potential of the proposed irradiance fields for estimating spatially varying lighting.
References


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


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. In CVPR.


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


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.


