Evaluate Geometry of Radiance Fields with Low-Frequency Color Prior

Qihang Fang[1,2,*], Yafei Song[3,*], Keqiang Li[1,2], Li Shen[3], Huaiyu Wu[1], Gang Xiong[1,†], Liefeng Bo[3]
[1]State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences
[3]Alibaba Group
fangqihang2020@ia.ac.cn, huaizhang.syf@alibaba-inc.com, likeq98@gmail.com, lshen.lsh@gmail.com, huaiyu.wu@ia.ac.cn, gang.xiong@ia.ac.cn, liefeng.bo@alibaba-inc.com

Abstract

A radiance field is an effective representation of 3D scenes, which has been widely adopted in novel-view synthesis and 3D reconstruction. It is still an open and challenging problem to evaluate the geometry, i.e., the density field, as the ground-truth is almost impossible to obtain. One alternative indirect solution is to transform the density field into a point-cloud and compute its Chamfer Distance with the scanned ground-truth. However, many widely-used datasets have no point-cloud ground-truth since the scanning process along with the equipment is expensive and complicated. To this end, we propose a novel metric, named Inverse Mean Residual Color (IMRC), which can evaluate the geometry only with the observation images. Our key insight is that the better the geometry, the lower-frequency the computed color field. From this insight, given a reconstructed density field and observation images, we design a closed-form method to approximate the color field with low-frequency spherical harmonics, and compute the inverse mean residual color. Then the higher the IMRC, the better the geometry. Qualitative and quantitative experimental results verify the effectiveness of our proposed IMRC metric. We also benchmark several state-of-the-art methods using IMRC to promote future related research. Our code is available at https://github.com/qihangGH/IMRC.

Introduction

A radiance field has been popular for representing 3D objects or scenes, especially the Neural Radiance Field (NeRF) (Mildenhall et al. 2020). However, as the ground-truth of a radiance field is hard to obtain, we could not directly evaluate its reconstructed geometry. Alternatively, novel-view synthesis methods usually measure the similarity between the rendered image and its observed counterpart using image quality metrics such as peak signal-to-noise ratio (PSNR). As shown in the top two rows of Fig. 1, this metric may be enough to evaluate the synthetic images but not appropriate to evaluate the geometry.

Besides novel-view synthesis, a lot of works exploit radiance fields for 3D reconstruction (Wang et al. 2021a; Long et al. 2022; Oechsle, Peng, and Geiger 2021; Sun et al. 2022; Zhang et al. 2021, 2022; Wu et al. 2022b; Wang et al. 2022; Wu et al. 2022a, 2023). They need a proper metric to quantitatively evaluate the geometry result. To this end, surface reconstruction methods usually transform the density field into a point-cloud using the marching cubes algorithm and compute Chamfer Distance (CD) with the scanned ground-truth. However, the scanning process along with the equipment is expensive and complicated. Thus, few datasets have this ground-truth, and many widely-used ones do not.

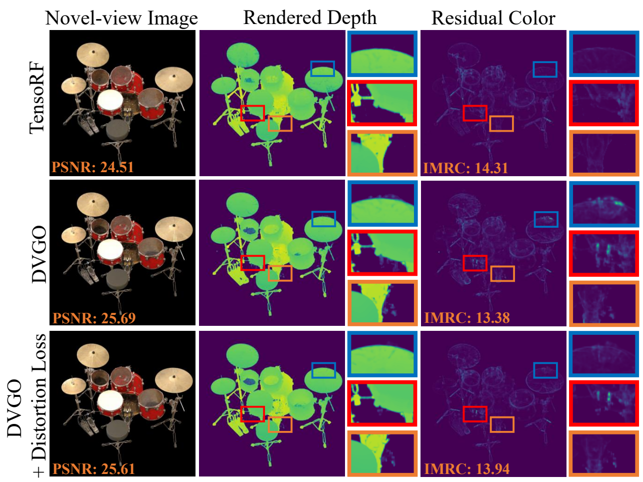


Figure 1: One example of novel-view images, rendered depth from reconstructed density fields, and residual color of TensoRF (Chen et al. 2022a), DVGO (Sun, Sun, and Chen 2022a), and DVGO (Sun, Sun, and Chen 2022a) + Distortion Loss (Barron et al. 2022). From top two rows, although DVGO achieves a better PSNR↑, its geometry is qualitatively worse than TensoRF. From the bottom row, distortion loss could be qualitatively verified to improve the geometry. Our IMRC↑ quantitatively evaluates these results correctly.

To alleviate above difficulties and introduce a new metric that does not require the geometric ground-truth, our key observation is that the color of any point on the ground-truth surface of an object tends to be low-frequency. This phenomenon is named as low-frequency color prior. Specifically, we adopt a closed-form method to compute the color field given the observation images and density field of a scene. We name the result as the computed color field to

[*]These authors contributed equally.
[†]Corresponding author.
This paper has been accepted by AAAI2024.

distinguish it from the reconstructed one given by the radiance field reconstruction method. Via exploring the computed color field, we find that the low-frequency color prior may be invalid for the points with inaccurate density values. Thus, the density field could be evaluated by the mean frequency of a computed color field.

However, it is difficult to directly evaluate the color frequency even for a single point, not to mention the whole field. To this end, we approximate the color with low-frequency spherical harmonics and compute the residual color. Then, the smaller mean residual color (MRC) implies lower color frequency, which indicates better geometry. Moreover, as the MRC is usually very small, which is not convenient for presentation and comparison, we transform it into decibel (dB) as PSNR does and name it as inverse mean residual color (IMRC). Then the higher the IMRC, the better the geometry. As demonstrated in Fig. 1, our IMRC metric could quantitatively evaluate the geometry correctly.

Our main contributions are concluded as follows: 1) We present the low-frequency color prior via analysing the ideal radiance field. We further find that this prior may be invalid for the computed color field if the density field is inaccurate. 2) To quantitatively evaluate the prior, we propose to approximate the color with low-frequency spherical harmonics and design the inverse mean residual color as a new metric. Qualitative and quantitative experimental results verify its effectiveness. 3) We further benchmark several state-of-the-art radiance field reconstruction methods using inverse mean residual color to promote future related research.

Related Work

We firstly review two types of radiance field reconstruction works, and then discuss the geometry metrics.

Novel View Synthesis. Mildenhall *et al.* (Mildenhall et al. 2020) proposed NeRF to synthesize novel-view images from posed images and dramatically improved the performance. They adopted a multilayer perceptron (MLP) to present the radiance field, which contains two components, *i.e.*, the density field and color field. Each point in the density field has a scalar controlling how much the color is accumulated. Each point in the color field encodes the view-dependent color. The image can be rendered using the volume rendering algorithm (Max 1995). From this path-breaking work, there have been many works aiming at improving NeRF from several aspects, such as various challenging scenarios (Zhang et al. 2020; Martin-Brualla et al. 2020; Barron et al. 2021; Jain, Tancik, and Abbeel 2021; Mildenhall et al. 2022; Tancik et al. 2022; Weng et al. 2022; Zhao et al. 2022; Shao et al. 2022; Niemeyer et al. 2022), inference speed (Yu et al. 2021a; Hedman et al. 2021; Reiser et al. 2021; Garbin et al. 2021; Chen et al. 2022b), training efficiency (Deng, Barron, and Srinivasan 2020; Hu et al. 2022; Yu et al. 2022; Sun, Sun, and Chen 2022a,b; Müller et al. 2022; Chen et al. 2022a; Wang et al. 2023b,a), generalization ability (Yu et al. 2021b; Wang et al. 2021b; Chen et al. 2021), and so on.

Since these works focused on novel view synthesis tasks, to evaluate the results, they usually adopted image quality metrics, such as PSNR, SSIM, and LPIPS. These metrics are suitable for the task. However, it remains questionable how to evaluate the radiance field itself and whether these metrics are adequate. Evaluating the radiance field directly is challenging due to the difficulty in obtaining its ground-truth. Existing works have not well explored this problem.

Surface Reconstruction. Some other works proposed to restore the surfaces via reconstructing the latent radiance field by inverse volume rendering (Wang et al. 2021a; Oechsle, Peng, and Geiger 2021; Sun et al. 2022; Wu et al. 2022b; Zhang et al. 2022; Wu et al. 2022a, 2023). These methods usually obtain the surfaces via transforming the latent density field into an occupancy field or signed distance field (SDF). Then the Chamfer Distance between the points on the surfaces and the ground-truth can be computed to evaluate the results. However, it is expensive and complicated to set up the hardware environment and scan the ground-truth point-cloud. Moreover, during the transformation process, some information would be discarded as the transformed result only contains the iso-surface. Therefore, the CD metric may not be suitable to evaluate a density field. In contrast, IMRC is not suitable to evaluate the surfaces, but it could evaluate the density field.

Geometry Metrics. Due to the unavailable ground-truth of a density field, we cannot use direct evaluation metrics, such as mean squared error. Thus, only indirect metrics could be considered. In previous Structure-from-Motion (SfM) systems, *e.g.*, (Hartley and Zisserman 2004; Snavely, Seitz, and Szeliski 2006; Agarwal et al. 2009), the mean re-projection error is well-known and widely adopted as a metric and optimization objective to evaluate the reconstructed structures and motions. The mean re-projection error is defined as the mean distance between each observed image feature point and the re-projection point of its reconstructed 3D point. In other words, it could evaluate the reconstructed geometry without the corresponding ground-truth. Inspired by this, we argue that the geometry of a radiance field also could be evaluated only with observation images.

Low-Frequency Color Prior

According to NeRF (Mildenhall et al. 2020), a radiance field $\mathcal{F}$ is defined on a 3D space $\mathbf{V}$, which has two components, *i.e.*, the density field $\mathcal{F}^\sigma$ and the color field $\mathcal{F}^c$. For a 3D point $\mathbf{v}$ in the space $\mathbf{V}$, the density $\mathcal{F}^\sigma_\mathbf{v}$ is a scalar that controls how much color of this point is accumulated. And the color $\mathcal{F}^c_\mathbf{v}$ encodes all view-dependent color information. For a specific 3D direction $\mathbf{d}$, the color vector is denoted as $\mathcal{F}^c_{\mathbf{v},\mathbf{d}}$. To render the color observed from a ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ with near and far bounds t^n, t^f, we can follow the function

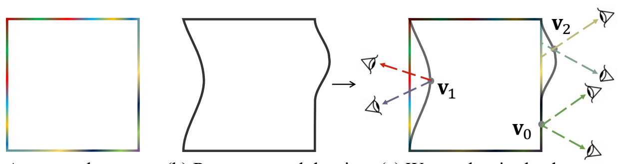
$$C(\mathbf{r}) = \int_{t^n}^{t^f} T(t)\mathcal{F}^\sigma_{\mathbf{r}(t)}\mathcal{F}^c_{\mathbf{r}(t),\mathbf{d}}\, dt, \tag{1}$$

where $\mathbf{o}$ is the camera original point, $\mathbf{d}$ is the direction from $\mathbf{o}$ to the corresponding pixel center, and the transmittance is

$$T(t) = \exp\left(\int_{t^n}^{t} -\mathcal{F}^\sigma_{\mathbf{r}(s)}\, ds\right). \tag{2}$$

(a) Side view of an example scene. (b) Hemisphere light field at different locations with the surface.

Figure 2: The color frequency tends to be lower and lower when the point approaches a surface.



(a) An example scene. (b) Reconstructed density. (c) Wrong density leads to high-frequency color.

Figure 3: As the points $\mathbf{v}_1$ and $\mathbf{v}_2$ demonstrate, inaccurate surfaces lie inside or outside the ground-truth surface will lead to higher-frequency colors.

To numerically estimate this integral, NeRF resorts to quadrature (Max 1995). Then the rendering equation is

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^{N} \hat{T}(t_i) \left(1 - \exp(-\mathcal{F}^{\sigma}_{\mathbf{r}(t_i)} \delta_{\mathbf{r}(t_i)})\right) \mathcal{F}^{c}_{\mathbf{r}(t_i),\mathbf{d}}, \quad (3)$$

where N is the number of sample points along the ray,

$$\hat{T}(t_i) = \exp\left(-\sum_{j=1}^{i-1} \mathcal{F}^{\sigma}_{\mathbf{r}(t_j)} \delta_{\mathbf{r}(t_j)}\right), \quad (4)$$

and $\delta_{\mathbf{r}(t_i)}$ is the distance between the point $\mathbf{r}(t_i)$ and its previous neighbour $\mathbf{r}(t_{i-1})$. Specifically, $\delta_{\mathbf{r}(t_1)} = 0$.

To illustrate the low-frequency color prior, we construct an example scene in Fig. 2, which shows that the color frequency of a point tends to be lower when it approaches a surface. Since the ground-truth density field should be the surface of the cube, and we do not care about the density values inside the surface as they do not affect the shape, a straight-forward hypothesis is that, for an ideal radiance field, the transmittance-density-weighted mean of color frequency should be small. Moreover, a wrong shape will break the low-frequency color prior, as illustrated by Fig. 3. Intuitively, in Fig. 3 (c), the observation color of a point from a ray should be identical with the color at the intersection of this ray and ground-truth surface. The color frequency of the points on wrong surfaces that lie inside ($\mathbf{v}_1$) or outside ($\mathbf{v}_2$) the ground-truth tends to be higher than that of the ground-truth surface ($\mathbf{v}_0$).

From these observations, if we have the color frequency field computed from a density field and observation images, the transmittance-density-weighted mean color frequency should be small for a good reconstructed density field. However, it is not easy to figure out the color frequency of even a single point. We therefore resort to the conjugate problem. According to the frequency domain transformation theory, if a signal is low-frequency, it can be well approximated by a group of low-frequency basis, and the residual would be quite small. Otherwise, the residual would be large. Therefore, the residual could be adopted to replace the frequency. Suppose that the residual color of a point $\mathbf{v}$ at a direction $\mathbf{d}$ is $\mathcal{F}^{r}_{\mathbf{v},\mathbf{d}}$, and the transmittance of $\mathbf{v}$ along $\mathbf{d}$ is $T_{\mathbf{v},\mathbf{d}}$, the transmittance-density-weighted mean residual color

$$r_{\mathbf{V}} = \frac{\int_{\mathbf{v}} \int_{\mathbf{d}} T_{\mathbf{v},\mathbf{d}} \mathcal{F}^{\sigma}_{\mathbf{v}} \mathcal{F}^{r}_{\mathbf{v},\mathbf{d}} \, d\mathbf{v} d\mathbf{d}}{\int_{\mathbf{v}} \int_{\mathbf{d}} T_{\mathbf{v},\mathbf{d}} \mathcal{F}^{\sigma}_{\mathbf{v}} \, d\mathbf{v} d\mathbf{d}} \quad (5)$$

should be small for a good reconstructed density field. We use the product of transmittance and density, i.e., $T_{\mathbf{v},\mathbf{d}} \mathcal{F}^{\sigma}_{\mathbf{v}}$, as the residual color's weight, which is the same as the color weight used in the rendering equation (1). In this way, the high color frequency of a point inside the reconstructed surface is not considered, as it does not affect the shape.

Inverse Mean Residual Color

In this section, we present the algorithm to compute the inverse mean residual color. The difficulty is that we could not use the color field reconstructed by the radiance field reconstruction method. If so, the geometry is not evaluated individually. To overcome this, we propose to utilize the observation images. For a 3D point, its projection points on all images are its observations. We can approximate its observations with a group of low-frequency basis, and the residual color could be obtained. We illustrate the whole pipeline in Fig. 4 and first introduce the observation acquisition process, then transform them into the low-frequency domain, and compute the IMRC at last.

Observation Acquisition

As illustrated in Fig. 4 (a), for any 3D point $\mathbf{v}$ in the scene, given the reconstructed density field $\mathcal{F}^{\sigma}$ and K cameras, we could obtain at most K observations corresponding to the images. For a specific camera, we denote its projection matrix and original point as $\mathbf{P}_k$ and $\mathbf{o}_k$ respectively. Then the point $\mathbf{v}$ could be projected on the image plane as follows,

$$v = \mathbf{P}_k \mathbf{v}, \quad (6)$$

where v is the coordinate on an image. Its observed color $\mathbf{c}_k$ as illustrated in Fig. 4 (b) could be calculated using bilinear interpolation. Besides the color, we should also record the observation direction since the view-dependent color is a function of it. The observation direction is calculated as

$$\mathbf{d}_k = \frac{\mathbf{o}_k - \mathbf{v}}{\|\mathbf{o}_k - \mathbf{v}\|}, \quad (7)$$

where $\|\cdot\|$ is the L_2 normalization operator.

It is important to note that, due to the occlusion, observations from different cameras should not be regarded equally. As illustrated in Fig. 4 (c), it is easy to understand that, if the accumulated density between a 3D point $\mathbf{v}$ and camera original point $\mathbf{o}_k$ is small, the corresponding observation confidence should be high, and vice versa. Coincidentally, this measurement is actually the transmittance between the 3D point and the camera. Thus, we directly take the transmittance to measure the confidence of each observation, which

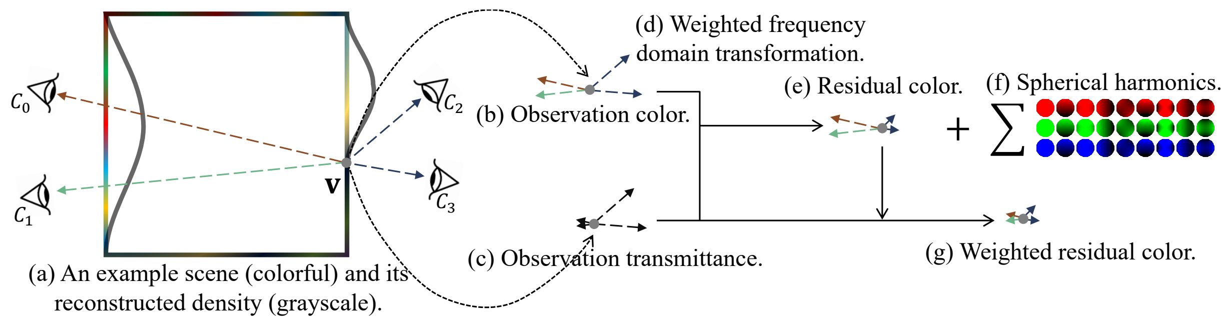


Figure 4: Residual color computation. (a) demonstrates an example scene and its reconstructed density field. For a point $\mathbf{v}$, we can obtain its observation colors (b) given captured images. To tackle the occlusion, the transmittance between the point $\mathbf{v}$ and each camera C is taken as the confidence of each observation, *i.e.*, (c). Based on (b) and (c), the color is weighted-transformed (d) into frequency domain (f) with the residual color (e). Due to the confidence (c), the residual color will be large if its corresponding confidence is low. To this end, the final residual color (g) should be weighted by observation confidence (c).

could be calculated as

$$T_k = \exp\left(\int_0^{\|\mathbf{o}_k - \mathbf{v}\|} -\mathcal{F}^{\sigma}_{\mathbf{r}(s|\mathbf{v},\mathbf{d}_k)}\, ds\right), \tag{8}$$

where $\mathbf{r}(s|\mathbf{v},\mathbf{d}_k) = \mathbf{v} + s\mathbf{d}_k$. As this integral is not easy to compute, we numerically estimate it following NeRF (Mildenhall et al. 2020) as follows,

$$\hat{T}_k = \exp\left(-\sum_{i=1}^{N} \mathcal{F}^{\sigma}_{\mathbf{r}(t_i|\mathbf{v},\mathbf{d}_k)} \delta_{\mathbf{r}(t_i|\mathbf{v},\mathbf{d}_k)}\right), \tag{9}$$

where N is the number of sample points, and $\mathbf{r}(t_i|\mathbf{v},\mathbf{d}_k) = \mathbf{v} + t_i\mathbf{d}_k$ is the i-th point.

Overall, for any 3D point, we will have K observations where each has three cells, *i.e.*, color, direction, and confidence, which is denoted as $\langle \mathbf{c}_k, \mathbf{d}_k, \hat{T}_k \rangle$. Specially, if the point is outside the viewing cone of the k-th camera, the confidence $\hat{T}_k$ should be set as 0.

Frequency Domain Transformation

In this subsection, we approximate the observations with a group of low-frequency basis. A natural choice is to use spherical harmonics (SH) (Green 2003), as they have been widely adopted to represent low-frequency colors (Basri and Jacobs 2003; Ramamoorthi and Hanrahan 2001) and also applied to radiance field reconstruction (Yu et al. 2021a, 2022). In the following, we briefly review the frequency coefficient estimation process following the well-known Monte Carlo method and present the difference under our situation. Note that, partial contents of this subsection have been presented in our previous work (Fang et al. 2023).

From the frequency domain transformation theory, given a signal F defined on the unit sphere $\mathbf{S}$, we can obtain its coefficient h_ℓ^m of the basis function Y_ℓ^m as

$$\begin{aligned} h_\ell^m &= \int_{\mathbf{d}\in\mathbf{S}} \frac{F(\mathbf{d})Y_\ell^m(\mathbf{d})}{p(\mathbf{d})} p(\mathbf{d})\, d\mathbf{d} \\ &= \mathbb{E}\left(\frac{F(\mathbf{d})Y_\ell^m(\mathbf{d})}{p(\mathbf{d})}\right) = 4\pi\, \mathbb{E}\left(F(\mathbf{d})Y_\ell^m(\mathbf{d})\right) \end{aligned}, \tag{10}$$

where $\ell \in \mathbb{N}, m \in \mathbb{N} \cap [-\ell, \ell]$ is the degree and order of the SH basis, $p(\mathbf{d})$ is the sampling probability of direction $\mathbf{d}$, $\mathbb{E}(\cdot)$ is the expectation. We set $p(\mathbf{d}) \equiv \frac{1}{4\pi}$ for evenly sampled directions. Given K observations obtained in the previous subsection, the coefficient can be estimated as

$$\hat{h}_\ell^m = 4\pi \frac{1}{K} \sum_{k=1}^{K} \mathbf{c}_k Y_\ell^m(\mathbf{d}_k). \tag{11}$$

However, this equation treats all observations equally, ignoring their varying confidence levels. To account for confidence T_k, the equation can be revised as

$$\hat{h}_\ell^m = 4\pi \frac{1}{\sum_k T_k} \sum_{k=1}^{K} T_k \mathbf{c}_k Y_\ell^m(\mathbf{d}_k). \tag{12}$$

Equation (12) above still has one problem. It implicitly assumes that the direction $\mathbf{d}$ distributes uniformly on the unit sphere $\mathbf{S}$. Therefore, the estimation process of each $\hat{h}_\ell^m$ is independent. However, this is not true under our settings as we cannot control the observation directions at all. In practice, they are usually not uniformly distributed.

To overcome this, we should estimate the coefficients in turn and eliminate the influence of previous estimated coefficients. Then Eq. (12) could be further updated as

$$\hat{h}_\ell^m = 4\pi \frac{1}{\sum_k T_k} \sum_{k=1}^{K} T_k \cdot {}^m_\ell\tilde{\mathbf{c}}_k \cdot Y_\ell^m(\mathbf{d}_k), \tag{13}$$

where

$${}^m_\ell\tilde{\mathbf{c}}_k = \mathbf{c}_k - \sum_{0\le i\le \ell}^{-i\le j<\tilde{m}} \hat{h}_i^j \cdot Y_i^j(\mathbf{d}_k), \tag{14}$$

where $\tilde{m} = m$ if $i = \ell$, else $\tilde{m} = i+1$. In practice, the implementation involves incrementally increasing the degree and order while estimating the coefficients of the corresponding SH basis. Before each estimation, we remove the color encoded by previous basis functions and only use the residual color ${}^m_\ell\tilde{\mathbf{c}}_k$.

Note that, when the directions are uniformly distributed, Eq. (13) is equivalent to Eq. (11) since the basis functions are orthogonal to each other. In this case, we have

Scene	JaxNeRF	Plenoxels	DVGO	TensoRF
Scan24	29.19/1.415/20.11/1	27.72/1.592/14.96/3	27.96/2.028/17.45/2	28.76/2.144/11.73/4
Scan37	26.77/1.502/15.43/1	26.22/1.813/12.99/3	26.47/1.466/14.05/2	26.00/1.778/ 8.86/4
Scan40	29.43/1.495/20.74/1	28.72/2.030/15.75/3	28.38/2.017/17.43/2	28.85/2.381/13.39/4
Scan55	31.06/0.637/19.55/1	30.50/0.892/15.55/3	31.66/1.247/19.37/2	29.57/1.765/ 9.78/4
Scan63	34.43/1.689/20.68/1	34.09/1.931/17.63/3	35.42/1.574/19.94/2	34.69/2.216/10.92/4
Scan65	29.90/1.212/17.59/1.3	31.16/1.548/14.55/3	30.63/1.482/17.32/1.7	30.63/1.947/10.78/4
Scan69	29.25/1.306/19.07/1	30.15/2.346/15.40/3	29.67/1.489/17.81/2	29.43/2.254/ 7.91/4
Scan83	35.28/1.478/15.98/1	37.00/2.245/16.63/2.7	36.14/1.610/17.95/2.3	35.96/2.712/ 7.95/4
Scan97	28.02/1.600/17.10/1	29.56/2.809/15.29/3	29.11/1.629/15.85/2	29.25/2.278/ 9.06/4
Scan105	31.86/1.136/19.99/1	33.36/1.907/16.67/3	33.07/1.352/18.85/2	32.92/2.219/ 9.19/4
Scan106	33.97/0.903/19.70/1	33.82/2.317/16.37/3	34.70/1.595/19.29/2	33.61/2.524/ 9.01/4
Scan110	32.96/1.842/16.00/1.7	32.49/3.349/14.86/3	34.00/1.770/16.13/1.3	33.67/3.396/ 7.99/4
Scan114	29.44/1.091/18.89/1	30.52/1.657/16.53/3	30.47/1.481/18.18/2	29.79/2.024/ 9.54/4
Scan118	36.68/0.998/21.77/1.7	36.10/2.719/17.82/3	37.58/1.297/22.57/1.3	35.76/2.436/11.34/4
Scan122	35.06/0.761/18.00/2	36.42/2.349/17.15/3	37.05/1.244/20.62/1	36.18/2.258/ 9.60/4
Average	31.55/1.271/18.71/1.18	31.86/2.100/15.88/2.98	32.15/1.552/18.19/1.84	31.67/2.289/ 9.80/4

Table 1: The PSNR↑/CD↓/IMRC↑/UserRank↓ results of 4 methods on the DTU dataset.

$T_k = 1, \sum_{k=1}^{K} Y_i^j(\mathbf{d}_k) \cdot Y_\ell^m(\mathbf{d}_k) = 0$, and Eq. (13) could be transformed to Eq. (11) because

$$\begin{aligned}&\hat{h}_\ell^m = 4\pi \frac{1}{K} \sum_{k=1}^{K} \left(\mathbf{c}_k - \sum_{0\le i\le\ell}^{-\ell\le j<\tilde{m}} \hat{h}_i^j Y_i^j(\mathbf{d}_k) \right) Y_\ell^m(\mathbf{d}_k) = \\ &\frac{4\pi}{K} \sum_{k=1}^{K} \mathbf{c}_k Y_\ell^m(\mathbf{d}_k) - \frac{4\pi}{K} \sum_{k=1}^{K} \sum_{0\le i\le\ell}^{-\ell\le j<\tilde{m}} \hat{h}_i^j Y_i^j(\mathbf{d}_k) Y_\ell^m(\mathbf{d}_k)\end{aligned}, \quad (15)$$

and

$$\begin{aligned}&\sum_{k=1}^{K} \sum_{0\le i\le\ell}^{-\ell\le j<\tilde{m}} \hat{h}_i^j Y_i^j(\mathbf{d}_k) Y_\ell^m(\mathbf{d}_k) = \\ &\sum_{0\le i\le\ell}^{-\ell\le j<\tilde{m}} \hat{h}_i^j \sum_{k=1}^{K} Y_i^j(\mathbf{d}_k) Y_\ell^m(\mathbf{d}_k) = \sum_{0\le i\le\ell}^{-\ell\le j<\tilde{m}} \hat{h}_i^j \cdot 0 = 0\end{aligned}. \quad (16)$$

The detailed proof and comparison results could be found in our previous work (Fang et al. 2023).

Final Metric Computation

With the calculated residual color at all locations, the MRC could be obtained following Eq. (5). As it is a continuous integral, to numerically estimate it, we resort to quadrature (Max 1995) once again. Specifically, we discretize the 3D space $\mathbf{V}$ into a volume, and denote all voxel vertexes as $\mathcal{V}$. Then the mean residual color could be estimated as

$$\text{MRC} = \frac{\sum_{\mathbf{v}\in\mathcal{V}} \sum_{k=1}^{K} T_{\mathbf{v},k} \left(1 - \exp\left(-\mathcal{F}_{\mathbf{v}}^{\sigma} \cdot \delta_{\mathbf{v}}\right)\right) \left({}_{L}^{L}\tilde{\mathbf{c}}_{\mathbf{v},k}\right)^2}{\sum_{\mathbf{v}\in\mathcal{V}} \sum_{k=1}^{K} T_{\mathbf{v},k} \left(1 - \exp\left(-\mathcal{F}_{\mathbf{v}}^{\sigma} \cdot \delta_{\mathbf{v}}\right)\right)}, \quad (17)$$

where ${}_{L}^{L}\tilde{\mathbf{c}}_{\mathbf{v},k}$ is the final residual color at vertex $\mathbf{v}$ as defined in Eq. (14), $T_{\mathbf{v},k}$ is the transmittance along the ray that connects $\mathbf{v}$ and the k-th camera center $\mathbf{o}_k$ as defined in Eq. (4), and $\delta_{\mathbf{v}}$ is the half voxel size.

Finally, we found that the MRC would be very small, which makes the value not convenient to be presented and compared. To this end, we further transform it into decibel as the PSNR metric does. As the theoretically maximal residual color is 1, the final metric could be obtained as

$$\text{IMRC} = -10 \cdot \log_{10}(\text{MRC}). \quad (18)$$

As the transformation includes an inverse operation, we denote it as IMRC shorted for Inverse Mean Residual Color. Then, the larger the IMRC, the better the geometry.

Experiments

In this section, we first verify the effectiveness of IMRC metric, then explore the influence of different settings. Finally, we benchmark several state-of-the-art methods.

Validation of IMRC

To exploit the effectiveness of the proposed IMRC metric, we conduct a series of experiments on the DTU dataset (Jensen et al. 2014) as it provides the point-cloud ground-truth. Though the point-cloud is essentially different from the radiance field, it still could be taken as a reference. As previous surface reconstruction works did, *e.g.*, (Yariv et al. 2020; Wang et al. 2021a), we use the selected 15 scenes. Each scene consists of 49 or 64 calibrated images and scanned point-cloud ground-truth. We select 5 or 6 images from the total 49 or 64 images respectively as testing ones, and use the remaining as training ones. The background of each image has been removed in advance.

For radiance field reconstruction, we select 4 well-known methods, including JaxNeRF (Mildenhall et al. 2020; Deng, Barron, and Srinivasan 2020), Plenoxels (Yu et al. 2022), DVGO (Sun, Sun, and Chen 2022a), and TensoRF (Chen et al. 2022a). We adopt the code released by the authors to optimize each scene. Additionally, we add a transmittance loss to the JaxNeRF to supervise the transmittance of a ray according to the foreground mask. After the optimization process, we only export the density field obtained by each method. For a fair comparison, we discrete the density field via sampling a 512^3 density volume, then the IMRC metric could be calculated. The experimental results are presented in Table 1, where the PSNR on testing images and the CD calculated with the point-cloud ground-truth are also reported. Specifically, we search the optimal threshold needed by the marching cubes algorithm for each method on each scene and report the lowest CD. Because the ground-truth

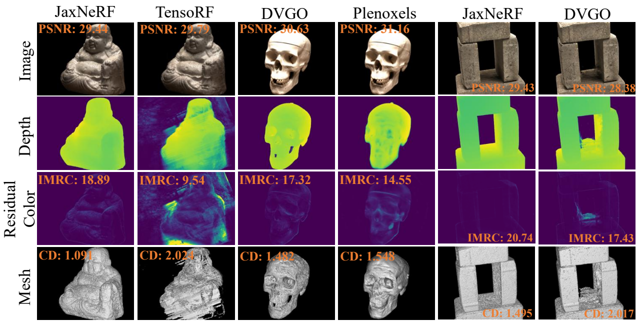


Figure 5: Examples of the CD↓/IMRC↑/UserRank↓ consistent results on the DTU dataset. For each example, left is better. From left to right, Scan 114, 65, and 40.

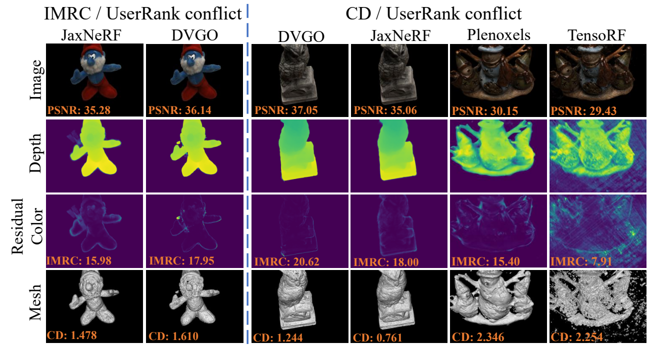


Figure 6: The IMRC↑/UserRank↓ and CD↓/UserRank↓ conflict results. For each example, left has better UserRank. From left to right, Scan 83, 122, and 69.

of the density field is not available, we perform a user study inspired by previous work (Li et al. 2015). Specifically, 3 experts evaluate and rank the scene geometry produced by the 4 methods from 1 (best) to 4 (worst) based on the depth images, residual colors, and meshes. They are unknown about the metric values and the method that produces the results.

In most cases, IMRC, CD, and UserRank agree with each other. The average results (last row) in Table 1 show that they rank 4 methods consistently. We showcase three CD/IMRC/UserRank consistent cases in Fig. 5. They also indicate that PSNR cannot well evaluate a density field. Besides, there are 2 IMRC/UserRank conflict pairs and 11 CD/UserRank conflict pairs out of all $90 = 15 \cdot \binom{4}{2}$ pairs, respectively. In the left panel of Fig. 6, one IMRC/UserRank conflict case is shown. The density field of the JaxNeRF is not sharp, and its surface is surrounded by many low density floaters. In contrast, the DVGO has a distinct floater as highlighted by the residual color. Although the IMRC of DVGO is better, such a distinct floater may lead to worse UserRank. The CD/UserRank conflicts mainly stem from two reasons. First, a NeRF model does not guarantee that its surface points all have the same density value. Thus, by extracting one iso-surface at a typical density level with the marching cubes algorithm, some surface information and near surface floaters that have different density values are inevitably discarded. As a result, the CD for such a surface may not well reflect real geometry. As shown by the middle two columns of Fig. 6. The JaxNeRF produces many low density floaters around the surface, while these floaters are missing after applying the marching cubes algorithm. On the other hand, the iso-surface of DVGO is inferior as some of its surface points that have different density values are discarded, which leads to poor CD. Another reason is that an object mask is applied in CD calculation, so points that lie outside the mask are neglected. As shown in the last two columns of Fig. 6, the TensoRF's mesh is qualitatively worse than Plenoxels', but lots of floaters in it are out of the mask. As a result, the CD of TensoRF is erroneously better than that of Plenoxels. The IMRC metric successfully evaluate the last two scenes and is consistent with UserRank. We visualize all the rest conflict cases in the Appendix. They can be analyzed similarly.

Overall, the qualitative and quantitative results verify the efficacy of IMRC to evaluate density fields. On one hand, PSNR could only evaluate the radiance field as a whole. On the other hand, CD suffers from information loss that only an iso-surface is considered. It also needs the point-cloud ground-truth that is usually not available. The IMRC metric, in contrast, is more suitable to evaluate the whole field, rather than only a surface, as it can deal with density values at all locations. It is also more consistent with UserRank.

Different Settings of IMRC

In this subsection, we exploit the influences of the parameters defined in IMRC. Only 2 parameters need to be configured manually, *i.e.*, the SH degree and the volume resolution.

The first and foremost parameter is the degree of SH used to compute the residual color for each point. In Table 2, we present the IMRC results on the DTU dataset with different SH degrees. We can see that a higher SH degree leads to a better IMRC. This is reasonable since the observations could be approximated better with more SH basis functions. However, we could not set a very large SH degree, because, if so, the residual color will be very small at all positions in the field, which makes the metric indistinct to evaluate the geometry. We also observe that the IMRC increases quickly when the SH degree is less than 2. And the increased margin becomes smaller starting from degree 3. Therefore, we set the SH degree as 2 in all other experiments. This setting is also usually used in previous works such as (Yu et al. 2022; Chen et al. 2022a). Notably, the relative ranks of all methods

SH Degree	0	1	2	3
SH Basis #	1	4	9	16
JaxNeRF	15.93	17.18	18.71	19.57
Plenoxels	13.81	14.78	15.88	16.52
DVGO	15.66	16.80	18.19	18.97
TensoRF	9.00	9.40	9.80	10.06
Average	13.60	14.54	15.64	16.28

Table 2: IMRC results of the selected methods on the DTU dataset with different SH degrees, which indicates that the higher SH degree leads to better IMRC results.

Resolution	64^3	128^3	256^3	512^3	768^3
JaxNeRF	17.92	18.42	18.61	18.71	18.54
Plenoxels	15.96	16.04	16.04	15.88	16.00
DVGO	17.05	17.72	18.14	18.19	18.27
TensoRF	9.45	9.62	9.75	9.80	9.88
Average	15.09	15.45	15.64	15.64	15.67

Table 3: IMRC results of the selected methods on the DTU dataset with different density volume resolutions.

remain unchanged with different SH degrees. This indicates that the IMRC is somehow stable with different SH degrees.

Notably, when the number of observations is small, we should decrease the SH degree. This will happen for the methods aiming at reconstructing the radiance field from sparse-views, such as (Niemeyer et al. 2022). According to the frequency domain transformation theory, the SH degree should be smaller than the number of observations.

The second parameter is the density volume's resolution. As presented in Table 3, with the increasing resolution, the IMRC metric is slightly better. The increase may come from the more accurate density field with higher space resolution. We also observe that the relative ranks of all methods remain unchanged at different resolutions. This indicates that the IMRC is also somehow stable with different resolutions. As the average IMRC does not change from resolution 256^3 to 512^3, we adopt 512^3 in other experiments.

Benchmarking State-of-the-arts

Besides the DTU dataset, we also benchmark the 4 state-of-the-art methods on NeRF Synthetic (Mildenhall et al. 2020) and LLFF (Mildenhall et al. 2019) datasets. For the NeRF Synthetic dataset, we use the released code and train on the black background as done in the DTU dataset. For the LLFF dataset, we directly use the models released by the authors. As there is no point-cloud ground-truth on NeRF Synthetic and LLFF datasets, we could not compute the CD metric and only present the PSNR and IMRC results in Table 4 and Table 5, respectively.

Benefiting from the IMRC metric, we could quantitatively but not only qualitatively evaluate how a regularization loss improves the density field. Specifically, we train DVGO (Sun, Sun, and Chen 2022a) with distortion loss (Barron et al. 2022), which could improve the density field qualitatively by observations. The results are reported in Table 6. With distortion loss, the better IMRC for all the three datasets verified the effectiveness of both the distortion loss and the metric itself.

Scene	JaxNeRF	Plenoxels	DVGO	TensoRF
Chair	30.28/19.21	30.66/17.60	33.79/21.59	30.46/19.10
Drums	24.16/13.41	24.26/12.65	25.69/13.38	24.51/14.32
Ficus	27.99/14.58	28.31/14.57	33.27/18.29	28.56/11.59
Hotdog	35.49/ 4.59	35.11/20.02	36.90/21.16	35.43/22.41
Lego	31.41/ 3.73	30.87/17.54	34.83/19.27	31.83/19.21
Materials	29.67/16.09	28.39/15.15	29.94/17.16	28.91/12.63
Mic	31.22/ 2.15	32.50/16.47	28.41/ 8.09	32.88/14.41
Ship	28.76/18.77	28.52/18.37	29.82/19.08	28.68/19.24
Average	29.87/11.57	29.83/16.55	31.58/17.25	30.16/16.61

Table 4: The PSNR↑/IMRC↑ results on NeRF synthetic dataset.

Scene	JaxNeRF	Plenoxels	DVGO	TensoRF
Fern	24.83/21.58	25.47/19.07	25.07/20.04	25.31/19.68
Flower	28.07/23.03	27.83/19.46	27.61/21.42	28.22/21.15
Fortress	31.76/26.97	31.09/22.32	30.38/24.82	31.14/24.57
Horns	28.10/23.58	27.60/19.09	27.55/21.20	27.64/21.80
Leaves	21.23/17.53	21.43/16.32	21.04/16.43	21.34/16.58
Orchids	20.27/18.06	20.26/16.03	20.38/16.59	20.02/16.37
Room	33.04/26.08	30.22/21.64	31.45/24.64	31.80/25.19
Trex	27.42/24.22	26.49/19.23	27.17/20.90	26.61/21.23
Average	26.84/22.63	26.30/19.15	26.33/20.76	26.51/20.82

Table 5: The PSNR↑/IMRC↑ results on LLFF dataset.

Distortion Loss	DTU	NeRF Synthetic	LLFF
×	32.15/18.19	**31.58**/17.25	26.23/18.87
✓	**32.20/18.47**	31.50/**17.94**	**26.33/20.76**

Table 6: PSNR↑/IMRC↑ results on three datasets produced by DVGO with and without distortion loss.

Discussion and Conclusion

We first discuss the limitations of IMRC and then conclude this paper in the following.

Limitations. There are also some degrade situations where the proposed IMRC metric may fail. For example, the low-frequency color prior no longer holds for surfaces that are purely specular or exhibit high levels of specular reflection, such as mirrors, leading to the failure of IMRC. Another degrade situation is that, if the whole density field is empty except for a few points that have been observed only by several cameras, the IMRC metric will be high. However, this density field is really bad. This type of degrade situations also exist for other geometry metrics, such as the reprojection error. Therefore, when using the proposed IMRC metric, it should be combined with other metrics, such as PSNR and SSIM. Only in this way, we can evaluate the results with less bias.

Conclusion. In this paper, we aim at evaluating the geometry information of a radiance field without ground-truth. This problem is important as the radiance field has been widely used not only on novel view synthesis but also 3D reconstruction tasks. For 3D reconstruction tasks, the geometry information is essential. However, due to the unavailable ground-truth and unsuitable metrics, there is no proper metric to quantitatively evaluate the geometry. To alleviate this dilemma, we propose the Inverse Mean Residual Color (IMRC) metric based on our insights on the properties of the radiance field. Qualitative and quantitative experimental results verify the effectiveness of the IMRC metric. We also benchmark 4 state-of-the-art methods on 3 datasets and hope this work could promote future related research.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFB3301504, in part by the National Natural Science Foundation of China under Grant U19B2029, Grant U1909204, and Grant 92267103, in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2021B1515140034, and in part by the CAS STS Dongguan Joint Project under Grant 20211600200022.

References

Agarwal, S.; Snavely, N.; Simon, I.; Seitz, S. M.; and Szeliski, R. 2009. Building Rome in a Day. In *ICCV*, 72–79.

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.

Basri, R.; and Jacobs, D. W. 2003. Lambertian Reflectance and Linear Subspaces. *IEEE TPAMI*, 25(2): 218–233.

Chen, A.; Xu, Z.; Geiger, A.; Yu, J.; and Su, H. 2022a. TensoRF: Tensorial Radiance Fields. In *ECCV*.

Chen, A.; Xu, Z.; Zhao, F.; Zhang, X.; Xiang, F.; Yu, J.; and Su, H. 2021. MVSNeRF: Fast Generalizable Radiance Field Reconstruction From Multi-view Stereo. In *ICCV*, 14124–14133.

Chen, Z.; Funkhouser, T.; Hedman, P.; and Tagliasacchi, A. 2022b. MobileNeRF: Exploiting the Polygon Rasterization Pipeline for Efficient Neural Field Rendering on Mobile Architectures. *arXiv preprint arXiv:2208.00277*.

Deng, B.; Barron, J. T.; and Srinivasan, P. P. 2020. JaxNeRF: an efficient JAX implementation of NeRF.

Fang, Q.; Song, Y.; Li, K.; and Bo, L. 2023. Reducing Shape-Radiance Ambiguity in Radiance Fields with a Closed-Form Color Estimation Method. In *NIPS*.

Garbin, S. J.; Kowalski, M.; Johnson, M.; Shotton, J.; and Valentin, J. 2021. FastNeRF: High-fidelity Neural Rendering at 200fps. In *ICCV*, 14346–14355.

Green, R. 2003. Spherical harmonic lighting: The gritty details. In *Archives of the game developers conference*, volume 56, 4.

Hartley, R. I.; and Zisserman, A. 2004. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition.

Hedman, P.; Srinivasan, P. P.; Mildenhall, B.; Barron, J. T.; and Debevec, P. 2021. Baking Neural Radiance Fields for Real-Time View Synthesis. In *ICCV*.

Hu, T.; Liu, S.; Chen, Y.; Shen, T.; and Jia, J. 2022. EfficientNeRF - Efficient Neural Radiance Fields. In *CVPR*, 12892–12901.

Jain, A.; Tancik, M.; and Abbeel, P. 2021. Putting Nerf on A Diet: Semantically Consistent Few-shot View Synthesis. In *ICCV*, 5885–5894.

Jensen, R.; Dahl, A.; Vogiatzis, G.; Tola, E.; and Aanæs, H. 2014. Large scale multi-view stereopsis evaluation. In *CVPR*, 406–413. IEEE.

Li, J.; Xia, C.; Song, Y.; Fang, S.; and Chen, X. 2015. A Data-Driven Metric for Comprehensive Evaluation of Saliency Models. In *ICCV*.

Long, X.; Lin, C.; Wang, P.; Komura, T.; and Wang, W. 2022. SparseNeuS: Fast Generalizable Neural Surface Reconstruction from Sparse views. In *ECCV*.

Martin-Brualla, R.; Radwan, N.; Sajjadi, M. S. M.; Barron, J. T.; Dosovitskiy, A.; and Duckworth, D. 2020. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. In *CVPR*.

Max, N. 1995. Optical Models for Direct Volume Rendering. *IEEE TVCG*, 1(2): 99–108.

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.; Ortiz-Cayon, R.; Kalantari, N. K.; Ramamoorthi, R.; Ng, R.; and Kar, A. 2019. Local Light Field Fusion: Practical View Synthesis with Prescriptive Sampling Guidelines. *ACM Trans. Graph.*, 38(4).

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *ECCV*.

Müller, T.; Evans, A.; Schied, C.; and Keller, A. 2022. Instant Neural Graphics Primitives With A Multiresolution Hash Encoding. *arXiv preprint arXiv:2201.05989*.

Niemeyer, M.; Barron, J. T.; Mildenhall, B.; Sajjadi, M. S. M.; Geiger, A.; and Radwan, N. 2022. RegNeRF: Regularizing Neural Radiance Fields for View Synthesis from Sparse Inputs. In *CVPR*.

Oechsle, M.; Peng, S.; and Geiger, A. 2021. UNISURF: Unifying Neural Implicit Surfaces and Radiance Fields for Multi-View Reconstruction. In *ICCV*, 5569–5579.

Ramamoorthi, R.; and Hanrahan, P. 2001. On the Relationship Between Radiance and Irradiance: Determining the Illumination From Images of a Convex Lambertian Object. *JOSA A*, 18(10): 2448–2459.

Reiser, C.; Peng, S.; Liao, Y.; and Geiger, A. 2021. KiloNeRF: Speeding up Neural Radiance Fields with Thousands of Tiny MLPs. In *ICCV*.

Shao, R.; Zhang, H.; Zhang, H.; Chen, M.; Cao, Y.; Yu, T.; and Liu, Y. 2022. DoubleField: Bridging the Neural Surface and Radiance Fields for High-fidelity Human Reconstruction and Rendering. In *CVPR*.

Snavely, N.; Seitz, S. M.; and Szeliski, R. 2006. Photo Tourism: Exploring Photo Collections in 3D. *ACM TOG*, 25(3): 835–846.

Sun, C.; Sun, M.; and Chen, H.-T. 2022a. Direct Voxel Grid Optimization: Super-fast Convergence for Radiance Fields Reconstruction. In *CVPR*.

Sun, C.; Sun, M.; and Chen, H.-T. 2022b. Improved Direct Voxel Grid Optimization for Radiance Fields Reconstruction. *arXiv preprint arXiv:2206.05085*.

Sun, J.; Chen, X.; Wang, Q.; Li, Z.; Averbuch-Elor, H.; Zhou, X.; and Snavely, N. 2022. Neural 3D Reconstruction in the Wild. In *SIGGRAPH*.

Tancik, M.; Casser, V.; Yan, X.; Pradhan, S.; Mildenhall, B.; Srinivasan, P. P.; Barron, J. T.; and Kretzschmar, H. 2022. Block-NeRF: Scalable Large Scene Neural View Synthesis. In *CVPR*.

Wang, F.; Chen, Z.; Wang, G.; Song, Y.; and Liu, H. 2023a. Masked Space-Time Hash Encoding for Efficient Dynamic Scene Reconstruction. In *NIPS*.

Wang, F.; Tan, S.; Li, X.; Tian, Z.; Song, Y.; and Liu, H. 2023b. Mixed Neural Voxels for Fast Multi-view Video Synthesis. In *ICCV*, 19706–19716.

Wang, P.; Liu, L.; Liu, Y.; Theobalt, C.; Komura, T.; and Wang, W. 2021a. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. In *NeurIPS*.

Wang, Q.; Wang, Z.; Genova, K.; Srinivasan, P. P.; Zhou, H.; Barron, J. T.; Martin-Brualla, R.; Snavely, N.; and Funkhouser, T. 2021b. IBRNet: Learning Multi-view Image-based Rendering. In *CVPR*, 4690–4699.

Wang, Y.; Han, Q.; Habermann, M.; Daniilidis, K.; Theobalt, C.; and Liu, L. 2022. NeuS2: Fast Learning of Neural Implicit Surfaces for Multi-view Reconstruction. *arXiv*.

Weng, C.-Y.; Curless, B.; Srinivasan, P. P.; Barron, J. T.; and Kemelmacher-Shlizerman, I. 2022. HumanNeRF: Free-viewpoint Rendering of Moving People from Monocular Video. In *CVPR*.

Wu, Q.; Liu, X.; Chen, Y.; Li, K.; Zheng, C.; Cai, J.; and Zheng, J. 2022a. Object-Compositional Neural Implicit Surfaces. In *ECCV*.

Wu, Q.; Wang, K.; Li, K.; Zheng, J.; and Cai, J. 2023. ObjectSDF++: Improved Object-Compositional Neural Implicit Surfaces. In *ICCV*.

Wu, T.; Wang, J.; Pan, X.; Xu, X.; Theobalt, C.; Liu, Z.; and Lin, D. 2022b. Voxurf: Voxel-based Efficient and Accurate Neural Surface Reconstruction. arXiv:2208.12697.

Yariv, L.; Kasten, Y.; Moran, D.; Galun, M.; Atzmon, M.; Ronen, B.; and Lipman, Y. 2020. Multiview Neural Surface Reconstruction by Disentangling Geometry and Appearance. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *NeurIPS*, volume 33, 2492–2502. Curran Associates, Inc.

Yu, A.; Fridovich-Keil, S.; Tancik, M.; Chen, Q.; Recht, B.; and Kanazawa, A. 2022. Plenoxels: Radiance Fields Without Neural Networks. In *CVPR*.

Yu, A.; Li, R.; Tancik, M.; Li, H.; Ng, R.; and Kanazawa, A. 2021a. PlenOctrees for Real-time Rendering of Neural Radiance Fields. In *ICCV*.

Yu, A.; Ye, V.; Tancik, M.; and Kanazawa, A. 2021b. PixelNeRF: Neural Radiance Fields From One or Few Images. In *CVPR*, 4578–4587.

Zhang, J. Y.; Yang, G.; Tulsiani, S.; and Ramanan, D. 2021. NeRS: Neural Reflectance Surfaces for Sparse-view 3D Reconstruction in the Wild. In *NeurIPS*.

Zhang, K.; Luan, F.; Li, Z.; and Snavely, N. 2022. IRON: Inverse Rendering by Optimizing Neural SDFs and Materials from Photometric Images. In *CVPR*.

Zhang, K.; Riegler, G.; Snavely, N.; and Koltun, V. 2020. NeRF++: Analyzing and Improving Neural Radiance Fields. *arXiv preprint arXiv:2010.07492*.

Zhao, F.; Yang, W.; Zhang, J.; Lin, P.; Zhang, Y.; Yu, J.; and Xu, L. 2022. HumanNeRF: Efficiently Generated Human Radiance Field from Sparse Inputs. In *CVPR*.