SUNDIAL: 3D Satellite Understanding through Direct, Ambient, and Complex Lighting Decomposition

Nikhil Behari[1,2], Akshat Dave[2], Kushagra Tiwary[2], William Yang[2], Ramesh Raskar[2]
[1]Harvard University, [2]Massachusetts Institute of Technology
nikhilbehari@g.harvard.edu, {nbehari,ad74,ktiwary,wyyang,raskar}@mit.edu

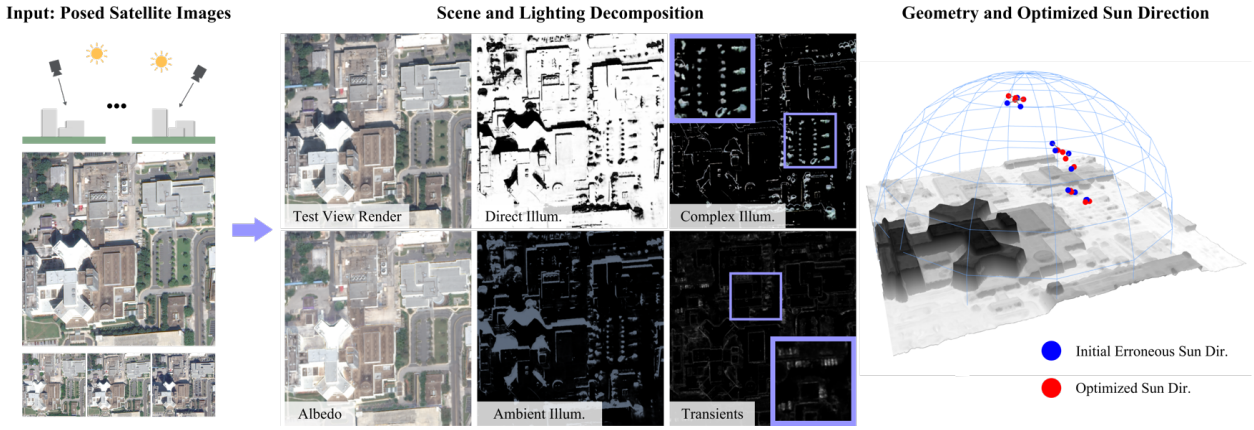


Figure 1. SUNDIAL scene and lighting disentanglement. SUNDIAL resolves ambiguous direct, ambient, and complex illumination effects and estimates accurate surface albedo and transient features in satellite imagery, resulting in high-quality renders under novel viewpoints and sun positions. By adapting priors from remote sensing literature, SUNDIAL adheres to a more physically-accurate scene rendering, and enables joint estimation of scene geometry and sun direction.

Abstract

Estimating 3D geometry of terrain from satellite imagery is essential for environmental science, urban planning, agriculture, and disaster response. However, traditional 3D modeling techniques face unique challenges in the remote sensing context, including limited multi-view baselines over extensive regions, varying direct, ambient, and complex illumination conditions, and time-varying scene changes across captures. In this work, we introduce SUNDIAL, a comprehensive approach to 3D reconstruction of satellite imagery using neural radiance fields. We jointly learn satellite scene geometry, illumination components, and sun direction in this single-model approach, and propose a secondary shadow ray casting technique to 1) improve scene geometry using oblique sun angles to render shadows, 2) enable physically-based disentanglement of scene albedo and illumination, and 3) determine the components of illumination from direct, ambient (sky), and complex sources. To achieve this, we incorporate lighting cues and geometric priors from remote sensing literature in a neural rendering approach, modeling physical properties of satellite scenes such as shadows, scattered sky illumination, and complex illumination of vegetation and water. We evaluate the performance of SUNDIAL against existing NeRF-based techniques for satellite scene modeling and demonstrate improved scene and lighting disentanglement, novel view and lighting rendering, and geometry and sun direction estimation on challenging scenes with small baselines, sparse inputs, and variable illumination.

1. Introduction

1.3 million images are taken daily from satellites orbiting the Earth [31]. This imagery has wide-scale applications across fields of environmental science, urban development, agriculture, and disaster response. Extracting 3D information from this remote sensing imagery, in particular, remains an open problem with significant impact in each field. For example, 3D scene understanding can help provide greater spatial awareness and context while monitoring global phenomena such as deforestation, climate change, and urban expansion.

3D scene reconstruction from satellite images, however, presents multiple challenges. First, satellite captures typically have limited multi-view baselines over large areas, with multi-view cues only visible in low-altitude, high-resolution, off-nadir satellite captures [25]. Second, images captured over time may contain transient, or changing, features such as shifting cars or construction. Third, illumination in satellite imagery includes components of direct, atmospheric, reflected, and scattered irradiance [2]; each component can vary greatly across captures, causing inconsistencies across images. Fourth, accurately estimating camera pose and relative sun direction can be challenging, especially due to small baselines and large scene distances. Small errors in these poses can propagate through 3D modeling resulting in large geometry estimation errors [26]. Despite these challenges, properly disentangling 3D scene geometry and illumination components is essential to minimize ambiguity and loss of spatial, lighting, and temporal information. [23, 39, 45].

To overcome the unique constraints of 3D satellite imagery, shadow-based height estimation has been proposed to overcome insufficient multi-view samples, harnessing oblique sun angles and shadow rendering to uncover hidden scene geometry [20, 22]. Recent techniques have applied neural radiance fields (NeRFs) [30] to model and separate these shadows in satellite imagery. S-NeRF [13] first proposed estimating a sun-dependent shading scalar across images; SatNeRF [27] extended this work, modeling transient features in addition to shading. EO-NeRF [27] then proposed a geometric NeRF-based shadow rendering approach, casting secondary sun-directed rays to determine regions with primary occluders (i.e. in shadow). However, these existing methods largely overlook the unique aspects of illumination specific to satellite-captured scenes. In particular, multi-bounce irradiance and reflections can create complex, soft shadows that cannot be modeled *only* with secondary ray casting. As a result, these approaches struggle to accurately decompose illumination and scene geometry (Figure 5), causing ambiguities in separation. Overcoming these ambiguities is critical; accurate shadow separation in satellite imagery can help identify occluded objects [39], enable more accurate land cover classification [23], and uncover shadow-related surface temperature cooling effects [50].

In this work, we present SUNDIAL, a NeRF-based framework that reduces ambiguity in satellite scene decomposition by estimating three key illumination components: direct illumination, ambient illumination, and complex illumination. We propose a geometric shadow ray casting approach, using the transmittance of these rays to render physically-accurate, disentangled shadows. However, in addition to estimating direct and indirect illumination component, we use these secondary rays to predict an additional complex illumination component that captures the challenging, multi-bounce light transport of satellite scenes that may be computationally infeasible to explicitly model. By estimating this additional scene lighting component, we are also able to reduce ambiguity in shadow rendering, thereby improving estimated scene geometry and lighting decomposition in satellite-captured scenes.

We propose several techniques to achieve this scene decomposition. First, we use secondary shadow ray transmission to estimate the components of illumination at a scene point, and apply loss functions inspired by prior research in remote sensing to accurately predict ambient and complex illumination. Second, we enforce more confident shadow predictions, thereby refining both scene geometry and shadow rendering jointly, and apply geometric regularization to improve scene modeling. Third, given the critical importance of direct illumination and shadow rendering in our approach, we jointly estimate sun direction with scene geometry during training, enabling refined sun direction prediction and mitigating propagated error from incorrect initial poses.

In summary, our contributions are the following:

- We disentangle and estimate direct, ambient, and complex illumination in satellite imagery, enabling more accurate, physically-based 3D scene modeling
- We jointly estimate sun directions and scene geometry, and propose remote sensing-based illumination and geometric priors to improve scene decomposition
- We enable more accurate separation of satellite scene albedo, shadow, and complex shading through our decomposition technique, improving novel view and lighting synthesis for satellite imagery

2. Related Work

2.1. Neural Radiance Fields

Neural Radiance Fields (NeRFs) have gained significant traction in the field of multi-view 3D reconstruction. By modelling the scene as coordinate-based neural networks, Mildenhall et al. [30] demonstrated that NeRFs can synthesize novel views of complex scenes with unprecedented detail and photo-realism. Subsequent works have scaled NeRFs from smartphone captures to in-the-wild internet images [29], large unbounded scenes [3], drone videos [4], city-scale regions [40, 47], and satellite imagery [27, 48]. NeRFs also show significant promise in 3D computer vision, with applications ranging from simultaneous localization and mapping [59], robotics control [21], manipulation [37] and navigation [1], semantic segmentation [44] to scene understanding [55].

Disentangling appearance in NeRFs Conventional NeRFs model the total radiance emitted by the scene that comprises of the lighting, geometry and material properties of the scene. Several works have explored disentangling the emit-

ted radiance into its constituents by leveraging reflectance priors [9, 10, 51, 53], specular reflections [24, 42, 43], shading [6, 52] and shadow cues [41, 49]. In this work, we focus on shadow-based disentanglement given the prominence of hard shadows in satellite imagery. We show how transmittance along the shadow ray can allow us to decompose not just the direct and shadow components, but also the components corresponding to ambient environment illumination and complex lighting effects.

2.2. Remote Sensing

While remote sensing research has made significant strides, there remain key challenges in building generalized, illumination-aware 3D reconstructions from satellite imagery. Particularly in urban areas, large shadows from buildings can cause significant ambiguity and classification errors [12, 46]. Both thresholding and geometric modeling techniques have been applied to this problem of shadow removal [23, 56]. Multi-view stereo (MVS) is another standard approach to 3D scene modeling, utilizing stereo pairs across images [5, 16]. However, MVS can produce errors due to variability in sun direction, atmospheric conditions, and seasonal features across images [11, 32, 39], requiring manual filtering of stereo pairs [14, 16].

3D Satellite Modeling with NeRF Neural radiance fields (NeRFs) have recently been applied to overcome these challenges. S-NeRF [13] introduced a technique for separating shading from albedo; Sat-NeRF [27] extended this work, highlighting the importance of bundle adjustment for improving reconstruction. EO-NeRF [28] further implemented a geometric shadow rendering approach, casting secondary rays in input sun directions to refine geometry. However, these methods still result in scene disentangling ambiguities: S-NeRF and SatNeRF estimate scene shading irrespective of geometry, resulting in inaccuracies in separation; EO-NeRF uses a geometry-based shadow rendering but still captures scene shading in non-shadowed regions, indicating that separating direct and indirect lighting alone is insufficient for remote sensing reconstruction. By modeling complex illumination effects, sun position estimation, and geometry priors, we achieve more accurate, physically-based scene and lighting disentanglement compared to prior NeRF-based techniques for satellite imagery (Fig. 5).

3. Formulation

We first describe our approach for *modelling* the complex geometries and illumination of satellite image-captured scenes. The key challenges in 3D satellite image reconstruction are handling (1) small baselines, (2) occlusions from top-down captures, and (3) complex mixed illumination sources. We handle challenges (1) and (2) by leveraging shadows, which provide secondary 3D cues that can be used to constrain the 3D reconstruction (Section 3.2). We handle (3) by modeling, and separating, different illumination types using a neural network and a physics-based forward model (Section 3.3).

3.1. NeRF Preliminaries

NeRF [30] is a volumetric representation, with inputs spatial position $\mathbf{x}$ and viewing direction $\mathbf{d}$, and outputs volume density σ and color $\mathbf{c}$: $F : (\mathbf{x}, \mathbf{d}) \rightarrow (\sigma, \mathbf{c})$. Volumetric rendering is achieved by uniformly sampling along rays $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$, enabling color $C(\mathbf{r})$ and depth $d(\mathbf{r})$ estimation of a ray as:

$$C(\mathbf{r}) = \sum_i w_i \mathbf{c}_i \quad d(\mathbf{r}) = \sum_i w_i t_i \tag{1}$$

$$T_i = \exp(-\sum_{j=1}^{i} \sigma_j \delta_j) \quad w_i = T_i(1 - \exp(-\sigma_i \delta_i)) \tag{2}$$

where T_i is the accumulated transmittance by step size δ_i, the distance between sampled points. We extend this base NeRF formulation to account for unique challenges of reconstruction from satellite imagery, including leveraging shadows to improve 3D reconstruction for occluded regions (Section 3.2) and rendering different illumination types (Section 3.3).

3.2. Small Baselines and Hidden Regions

When limited, top-down satellite views are sampled, 3D reconstruction of vertical surface geometries such as building facades becomes challenging [41, 54]. We address this problem by leveraging shadows; while vertical surfaces may be hidden from primary satellite capture angles, oblique lighting angles from the sun can help exploit and model this hidden geometry. Specifically, we separate shadows using secondary shadow rays cast from from scene surface points to the sun direction (shown Figure 2). These secondary shadow rays $\mathbf{s}$, drawn from the surface intersection point of each primary ray $\mathbf{r}$, are used to: 1) render sun-dependent shadows, 2) provide an additional cue to constrain 3D reconstruction, and 3) help estimate the weight of a *complex illumination* component, to account for regions with challenging multi-bounce light that cannot be directly modeled with a single-bounce secondary ray casting (Section 3.3 below).

3.3. Disentangling Illumination

There are four primary radiative components in satellite imagery: 1) direct solar irradiance, 2) scattered atmospheric irradiance, 3) irradiance from reflection between objects, and 4) coupled irradiance from multiple scattering events [2]. To better disentangle these lighting components, we use secondary shadow rays to estimate the weights of three illumination components for a given ray $\mathbf{r}$:

1. $W_{\text{di}}(\mathbf{r})$: Direct illumination component weight

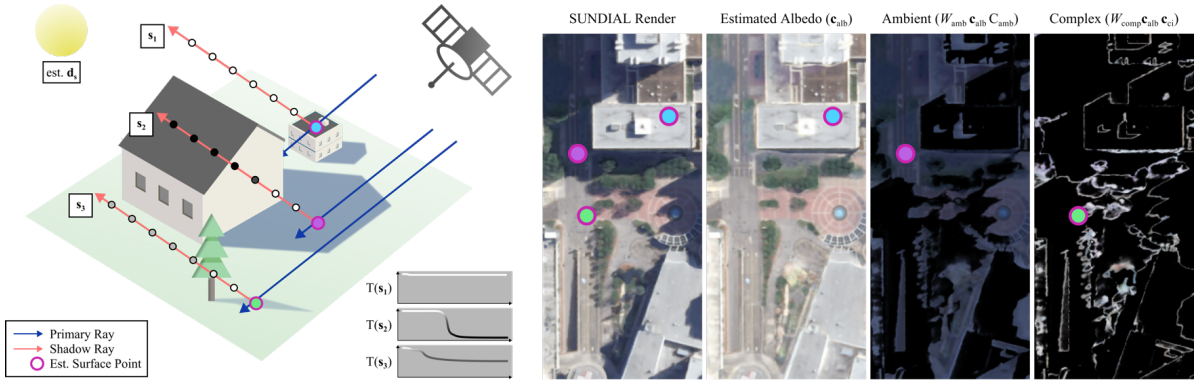


Figure 2. Accumulated transmittance from secondary shadow rays as illumination cues. Final transmittance values of secondary shadow rays are used to determine the proportion of direct, ambient, and complex illumination values. Direct light is the primary component of rays with direct sun visibility (e.g. ray $\mathbf{s}_1$); ambient light is the primary component of rays distinctly in shadow (e.g. ray $\mathbf{s}_2$); complex lighting is the primary component of rays with ambiguous or challenging geometry occluders (e.g. ray $\mathbf{s}_3$).

2. $W_{\text{amb}}(\mathbf{r})$: Ambient illumination component weight
3. $W_{\text{comp}}(\mathbf{r})$: Complex illumination component weight

In general, W_{di} is highest for regions with direct sun visibility, W_{amb} is highest in shadowed areas with strong atmospheric irradiance, and W_{comp} is highest in complex-geometry regions such as trees and water, where interreflections and multiple scattering events create intricate shading [7, 18, 36]. Secondary shadow rays are used to determine weightings for each illumination component, which we apply to three predicted rendering colors.

Direct Illumination Direct illumination is the primary irradiance component for regions with direct sun visibility. In these regions, the primary component of our final color rendering is predicted as position-dependent $\mathbf{c}_{\text{alb}}(\mathbf{x})$, the albedo color. We compute this direct lighting component in the final rendering as:

$$W_{\text{di}}(\mathbf{r}) \cdot \mathbf{c}_{\text{alb}} \tag{3}$$

which is a weighted quantity representing the direct illumination component of the final ray color. For pixels in complete direct illumination (i.e. not in shadow), $W_{\text{di}}(\mathbf{r}) = 1$, and the predicted albedo color is therefore the primary color component of the final rendering.

Ambient Illumination For regions completely in shadow, direct solar radiation is blocked, so scattered atmospheric irradiance contributes most to total irradiance [2, 58]. To model this atmospheric irradiance, we predict an ambient color $C_{\text{amb}}(\mathbf{d}_s)$. This color is a function of the estimated sun position $\mathbf{d}_s$ for simplicity; in reality, atmospheric scattering may depend on a variety of additional scatting features, such as air, smoke, and cloud particles [34]. This value $C_{\text{amb}}(\mathbf{d}_s)$ is a single RGB color value (typically a dark blue hue) multiplied by the base albedo to render regions in shadow where atmospheric irradiance is strongest. We therefore apply this ambient light as a weighted component (strongest for regions in shadow):

$$W_{\text{amb}}(\mathbf{r}) \cdot \mathbf{c}_{\text{alb}} \otimes C_{\text{amb}} \tag{4}$$

where $\otimes$ denotes elementwise multiplication.

Complex Illumination Captured irradiance in satellite imagery may also be explained by more complex multi-bounce illumination features such as interreflections and scattered light. Thus, we additionally predict a "complex illumination" feature $\mathbf{c}_{\text{ci}}(\mathbf{x}, \mathbf{d}_s)$ as a function of position and sun direction. This complex illumination feature $\mathbf{c}_{\text{ci}}$ is applied in the final rendering equation with weighted quantity $W_{\text{comp}}(\mathbf{r})$, which is highest when occluder geometry is ambiguous (i.e. in trees, bushes, and water). We apply this spatially and sun position-dependent complex illumination component in the final rendering as:

$$W_{\text{comp}}(\mathbf{r}) \cdot \mathbf{c}_{\text{alb}} \cdot \mathbf{c}_{\text{ci}} \tag{5}$$

3.4. Final Rendering Equation

We combine the direct, indirect, and complex illumination components to form the final rendering equation:

$$\mathbf{C}(\mathbf{r}) = W_{\text{di}}\mathbf{c}_{\text{alb}} + W_{\text{amb}}\mathbf{c}_{\text{alb}} \otimes C_{\text{amb}} + W_{\text{comp}}\mathbf{c}_{\text{alb}}\mathbf{c}_{\text{ci}} \tag{6}$$

In this final rendering equation, the first term is strongest for regions with well-defined geometry and that are directly visible to the sun. The second term is strongest for regions with well-defined geometry that are entirely in shadow (occluded from direct sun visibility). The third term is strongest for regions requiring multi-bounce light shading, for example in regions with trees or water; in this case, the complex illumination color can provide nuanced shading, potentially capturing interreflection and multiple scattering components.

3.5. Refining Poses

Accurate camera poses are critical for 3D satellite scene reconstruction, especially given the large scene distances

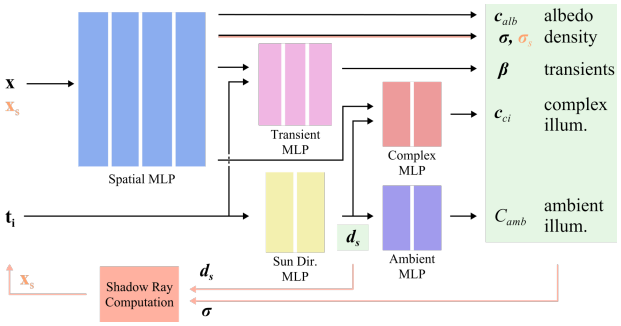


Figure 3. Overview of our proposed architecture. Outputs of SUNDIAL are highlighted in green. Primary ray densities σ and estimated sun directions are used to query secondary shadow rays with density σ_s; final transmittance of these secondary ray are used to compute direct, ambient, and complex illumination weights.

and small multi-view baselines. Prior work has demonstrated the importance of bundle adjustments in this setting [17, 27, 33]. We rely on direct sun illumination and shadows as a secondary geometry cue; because camera pose and relative sun direction are closely related, we propose a joint geometry and sun direction estimation technique in SUNDIAL. By ensuring accurate relative sun directions, we can avoid propagating error from imprecise estimated sun direction inputs.

4. Our Approach

We now describe how we *estimate* the components of our proposed rendering equation using a NeRF-based approach. We visualize our proposed architecture in Figure 3. The proposed MLP takes as input spatial position $\mathbf{x}$ and a per-image embedding $\mathbf{t}_i$. It outputs albedo $\mathbf{c}_{\text{alb}}$, complex $\mathbf{c}_{\text{ci}}$, and ambient illumination C_{amb}, along with density σ, sun direction $\mathbf{d}_{\text{s}}$, and a transient scalar β. Critically, we then use density σ and estimated sun direction $\mathbf{d}_{\text{s}}$ to query secondary shadow rays $\mathbf{s}$, from which we estimate density σ_s. Density along the shadow rays are used to estimate the three illumination component weights $W_{\text{di}}, W_{\text{amb}}, W_{\text{comp}}$, which account for shadows and complex illumination.

4.1. Primary Ray Components

Primary rays are rays mapped directly from scene points to camera pixels. The color resulting from the primary ray can be expressed as the decomposition of three variables: albedo, ambient color, and complex color.

Albedo We model scene albedo $\mathbf{c}_{\text{alb}}$ using a coordinate-based MLP. We exclude view dependence in the spatial MLP inputs, assuming Lambertian surface properties. Although non-Lambertian modeling is critical for topography-influenced lighting variation in certain terrains [35], assuming Lambertian surfaces is often suitable for urban aerial imagery [8, 19], the primary target of our evaluation.

Ambient Color Ambient color C_{amb}, the primary color component for regions in shadow, is predicted using a two-layer MLP with input the estimated sun directions $\mathbf{d}_{\text{s}}$. This single RGB color prediction is therefore uniform for a given input image.

Complex Color We further model the complex color $\mathbf{c}_{\text{ci}}$, the primary color component for regions with challenging-geometry occluders, using a 2-layer MLP with input spatial features and estimated sun directions $\mathbf{d}_{\text{s}}$. This modeling technique aims to capture intricate shading, multi-bounce reflections, and ambient occlusions that are difficult to capture with direct and indirect light modeling alone. Instead, shading is applied using coordinate and sun position-based complex color, weighted with estimated quantity W_{comp}.

4.2. Sun Direction, Shadows, and Illumination

For each primary ray $\mathbf{r}$ we cast secondary shadow rays $\mathbf{s}$. These secondary shadow rays are used to determine the proportion of direct, ambient, and complex illumination values. We visualize this shadow ray casting procedure in Figure 2.

Surface Estimation for Shadow Rays The secondary shadow rays have origin $\mathbf{o}_{\text{s}}$ estimated using primary ray depth predictions. Given estimated depth $d(\mathbf{r})$ of ray $\mathbf{r}$, we compute a surface point $\mathbf{o}_{\text{s}}$:

$$\mathbf{o}_{\text{s}} = \mathbf{o} + (d(\mathbf{r}) - \Delta(t))\mathbf{d} \tag{7}$$

Padded by $\Delta(r) = \Delta_0 e^{-kr}$, a function of the training iteration r and Δ_0, some initial small padding constant. In early stages of training, the estimated implicit surface may not be well-formed, resulting in partial initial occlusion in secondary shadow rays. This padding on the estimated depth enables more confident shadow rendering (and downstream sun position estimation) in early stages of training, and decays as the estimated surface becomes more refined.

Sun Direction Estimation We estimate sun directions $\mathbf{d}_{\text{s}}$ jointly with scene geometry during training. We first consider original (input) sun directions $\mathbf{d}_{\text{s0}}$ with initial "up" directions $\mathbf{u}$. This formulation allows for initial estimates to be used for faster convergence when sun direction is known. For training settings without known sun directions, these values may be randomly initialized for each image.

Given an input image embedding $\mathbf{t}_i$, the SUNDIAL MLP predicts $\mathbf{u}_i = [u_{i1}, u_{i2}, u_{i3}]$, a new "up" direction for the rotated sun direction reference frame in image i. We compute $\mathbf{V}_i$, the skew-symmetric matrix from $\mathbf{v} = \mathbf{u} \times \mathbf{u}_i$, and compute a rotation matrix $\mathbf{R}_i$ from:

$$\mathbf{V}_i = \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix} \quad \mathbf{R}_i = \mathbf{I} + \mathbf{V}_i + \frac{\mathbf{V}_i^2}{1+c} \tag{8}$$

where $\mathbf{I}$ is the 3x3 identity matrix, c is the dot product $\mathbf{u} \cdot \mathbf{u}_i$, and $\mathbf{R}_i$ is the final rotation matrix for the sun directions in image i. We then obtain the estimated sun direction, used as the direction $\mathbf{d}_{\text{s}}$ of the secondary shadow rays:

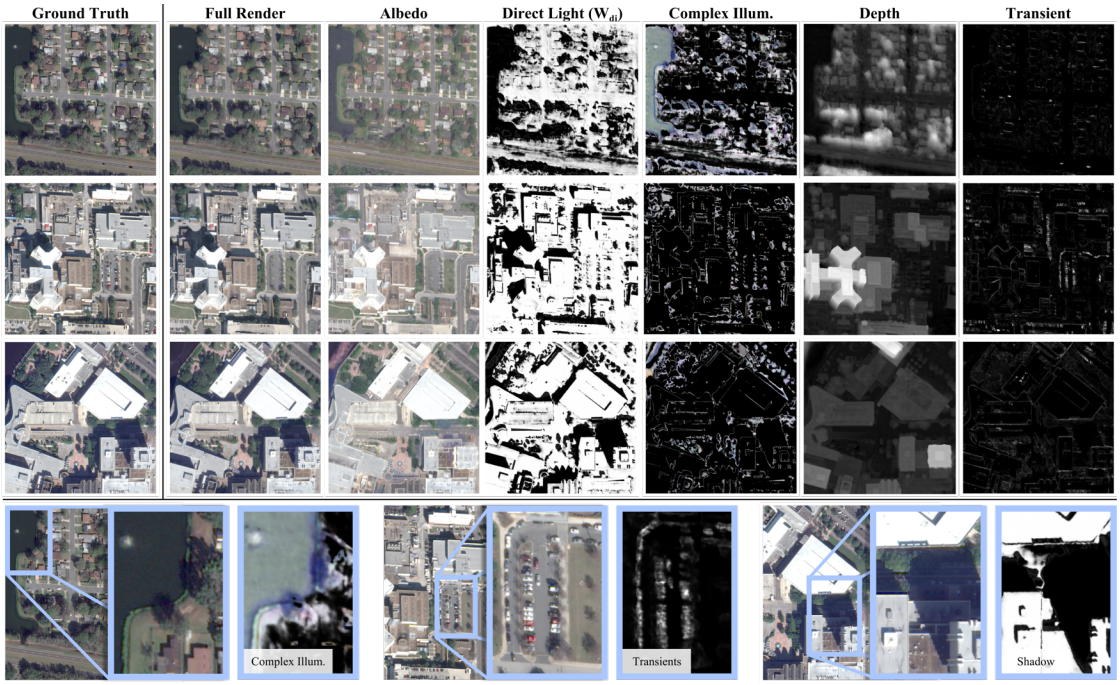


Figure 4. SUNDIAL scene disentangling results. SUNDIAL can disentangle complex illumination (soft shadows, shading) in challenging-geometry regions such as trees, bushes, and water. This complex illumination separation disambiguates albedo, transient, and shadow predictions, refining SUNDIAL-estimated scene decomposition and improving novel viewpoint and sun direction rendering.

$$\mathbf{d}_{\text{s}} = \mathbf{R}_i \cdot \mathbf{d}_{\text{s0}} \tag{9}$$

Secondary Shadow Rays With the network-estimated surface points $\mathbf{o}_s$ and predicted sun directions $\mathbf{d}_{\text{s}}$, we then cast secondary rays $\mathbf{s}$ to predict shadowed regions. We determine the probability of the pixel being in shadow by the transmittance at the end of the secondary shadow ray:

In shadowed regions, this secondary ray transmittance will be low, as the ray $\mathbf{s}$ will be occluded from direct sun visibility. Thus, we obtain the probability that the surface of ray $\mathbf{r}$ is in direct sun lighting as: $P_{\text{di}}(\mathbf{r}) = T(\mathbf{s})$.

Estimating Illumination Weights Using the probability of direct lighting $P_{\text{di}}(\mathbf{r})$, we then compute the weights for the three illumination components. First, we compute W_{comp}:

$$W_{\text{comp}} = 1 - \left(\frac{1}{e^{(\kappa-\xi)-\kappa P_{\text{di}}(\mathbf{r})}+1} + \frac{1}{e^{\kappa P_{\text{di}}(\mathbf{r})-\xi}+1}\right) \tag{10}$$

Where κ and ξ are hyperparameters that can be tuned according to the desired sensitivity to complex illumination. This quantity is highest when predicted direct lighting probability $P_{\text{di}}(\mathbf{r})$ is near 0.5, where occluder geometry may be ambiguously defined). Then, we obtain W_{di} and W_{amb} as:

$$W_{\text{di}} = (1 - W_{\text{comp}}) \cdot P_{\text{di}}(\mathbf{r}) \tag{11}$$

$$W_{\text{amb}} = (1 - W_{\text{comp}}) \cdot (1 - P_{\text{di}}(\mathbf{r})) \tag{12}$$

Intuitively, complex illumination weight decreases as $P_{\text{di}}(\mathbf{r})$ probabilities become more confident (i.e. closer to 0 or 1), and increases when predictions are ambiguous (when $P_{\text{di}}(\mathbf{r}) = 0.5$). This weighting method has two key advantages: 1)this approach enables more confident shadow predictions, reducing the need for SUNDIAL to handle semi-shaded areas with partial volume occlusions that may negatively impact estimated geometry; 2) this approach represents complex illumination in both directly lit and shadowed regions, increasing complex illumination weight as the precision of geometry-based occlusion prediction decreases.

4.3. Loss Functions

We incorporate photometric loss, normal orientation loss [43], and distortion loss [3] in our architecture. We additionally introduce three loss functions that help improve rendering specifically for satellite imagery scenes by constraining ambient and complex color predictions.

Ambient Color Loss Prior work in atmospheric modeling indicates that, due to the Rayleigh effect of atmospheric scattering, the ambient lighting component of shadowed regions has a higher saturation of blue and violet colors [34, 57]. We therefore apply two physics-based loss functions on the predicted ambient color C_{amb}:

$$\mathcal{L}_{\text{white}}(C_{\text{amb}}) = \lambda_{\text{w}} \cdot \max(0, \frac{1}{3}(r+g+b) - \tau)^2 \tag{13}$$

$$\mathcal{L}_{\text{blue}}(C_{\text{amb}}) = \lambda_{\text{b}} \cdot \max(0, \psi \cdot \frac{r+g}{2} - b)^2 \tag{14}$$

Where τ is a "brightness" threshold, ψ is a "blue dominance" factor, and λ_{w} and λ_{b} are scaling terms. These terms ensure that predicted ambient light color C_{amb} is sufficiently dark and blue in color, adhering to establish principles of atmospheric modeling. We find this prior is especially critical in cases where input images have shadows in similar positions across captures, making it difficult to separate albedo and shadows. The proposed ambient color loss ensures that rendered shadows are sufficiently dark blue (from ambient light), enabling accurate disentangling of albedo in shadowed regions.

Complex Color Loss To prevent $P_{\text{di}}(\mathbf{r})$ from converging to 0.5 for all regions, and to enable more confident direct lighting probabilities, we similarly apply a "complex illumination" loss for $W_{\text{comp}}(\mathbf{r})$ values near 0.5:

$$\mathcal{L}_{\text{ci}} = \lambda_{\text{ci}} \cdot W_{\text{comp}}(\mathbf{r}) \tag{15}$$

This loss term helps disambiguate direct lighting probabilities by penalizing values near 0.5; for challenging-geometry regions that cannot be properly resolved, the estimated complex illumination component enables more nuanced shading and soft shadows.

Normal Estimation We estimate normal vector fields for our scenes following [9, 38, 43], using the gradient of the volume density with respect to position:

$$\hat{\mathbf{n}}(\mathbf{x}) = -\frac{\nabla\sigma(\mathbf{x})}{\|\nabla\sigma(\mathbf{x})\|} \tag{16}$$

And further predict a 3-vector output $\hat{\mathbf{n}}'_i$ using the output spatial MLP features, applying a regularization term as follows to predict more accurate normal vector fields:

$$\mathcal{L}_{\text{normal}} = \sum_i w_i \|\hat{\mathbf{n}}_i - \hat{\mathbf{n}}'_i\|^2 \tag{17}$$

In addition to smoother surface normal predictions, we find this regularization term enables smoother geometry and depth estimation, acting as an effective geometry prior particularly for urban scenes characterized by smooth, cuboidal buildings.

Transient Loss Satellite imagery captures are both multi-view and multi-time, capturing transient features such as cars, construction events, and seasonal vegetation changes. To account for transient features, we adopt the technique proposed in NeRF-W [29], predicting $\beta(\mathbf{t}_{\text{i}}, \mathbf{x})$ as a function of per-input transient embedding. We incorporate the NeRF-W loss to ensure β uncertainty captures transients without converging to infinity:

$$\mathcal{L}_\beta = \sum_{\mathbf{r}} \frac{\|\mathbf{c}(\mathbf{r}) - \mathbf{c}_{gt}(\mathbf{r})\|_2^2}{2(\beta(\mathbf{r}) + \beta_{min})^2} + \frac{\log(\beta(\mathbf{r}) + \beta_{min}) + \eta}{2} \tag{18}$$

Comparison of Satellite Reconstruction Techniques								
Approach	JAX 004		JAX 068		JAX 214		JAX 260	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
S-NeRF	21.80	0.641	21.13	0.691	18.70	0.577	17.62	0.546
SatNeRF	22.25	0.669	20.56	0.701	18.55	0.600	17.42	0.553
EO-NeRF*	20.52	0.619	18.37	0.693	19.92	0.684	17.96	0.569
SUNDIAL	21.75	0.646	21.35	0.735	20.41	0.704	18.31	0.581

Table 1. Evaluation metrics for SUNDIAL compared to previous approaches. Novel view rendering on real, unseen, *future* satellite images. SUNDIAL provides improvements across test scenes for novel view and lighting synthesis.

Satellite Depth Estimation MAE				
Approach	JAX 004	JAX 068	JAX 214	JAX 260
S-NeRF	1.890	1.800	4.510	3.070
SatNeRF	1.508	1.432	2.898	2.413
EO-NeRF*	2.275	1.928	2.557	1.953
SUNDIAL	2.241	1.533	2.289	1.893

Table 2. Digital surface model estimation for real satellite scenes. SUNDIAL, trained without initial sun directions, demonstrates comparable depth estimation to previous approaches (that use sun direction), and consistently outperforms EO-NeRF.

5. Experiments

5.1. Dataset

We evaluate SUNDIAL on real satellite imagery captured by WorldView-3, a high resolution satellite operating at altitude 617km. The dataset includes images from four locations in Jacksonville, Florida, spanning 2014-2016. Each location has 10-25 multi-date captures at 0.3m resolution. Image metadata was used to verify sun position and capture time. In some cases, multiple image captures of a scene were taken in a single day. To avoid overfitting on one of these instances, we train SUNDIAL on the earliest 75% of sequentially captured images from each scene, and evaluate on the remaining 25% of images (captured at subsequent times). Additional details are provided in the supplement.

5.2. Implementation Details

We use bundle-adjusted rational polynomial camera (RPC) models to cast scene sampling rays, following [15, 27]. We train SUNDIAL with a batch size of 2048 rays for 200k iterations, using an initial learning rate of 5×10^{-4} using a step scheduler with decay factor 0.9. Additional training details are provided in the supplement.

5.3. Disentangling Appearance

We show results for the disentangled outputs of SUNDIAL in Figure 4. We observe that SUNDIAL can effectively resolve scene albedo, physically-accurate shadows, transient features, scene depth, as well as complex illumination features, which primarily enable shading for challenging geometry regions in trees and water.

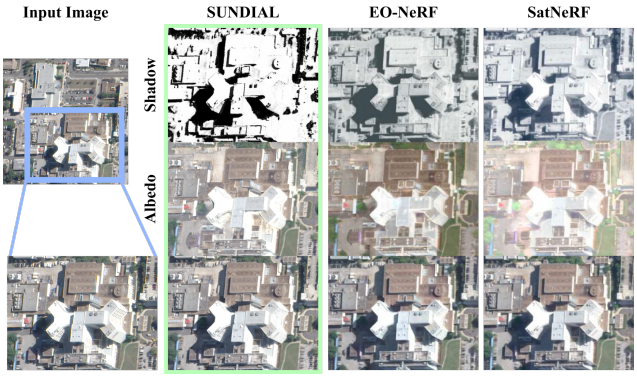
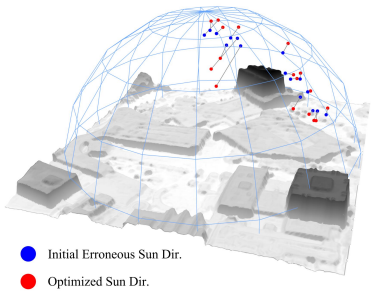


Figure 5. Scene disentangling comparison. SUNDIAL can resolve physically-accurate, disambiguated shadows by estimating an additional complex illumination component. EO-NeRF, despite casting geometric shadow rays, cannot properly disentangle soft shadows and shading, instead estimating an ambiguous shading component similar to the non-geometric SatNeRF predictions.

5.4. Comparisons with Baseline

We provide quantitative comparisons to previous approaches in Tables 1 and 2. EO-NeRF does not have a publicly available code release, so we reimplement the proposed model (EO-NeRF*). Our evaluation demonstrates that SUNDIAL accurately disentangles scene geometry, illumination, and sun direction, enabling accurate novel view and lighting synthesis for future, unseen, test images.

We further demonstrate this decomposition accuracy in Figure 5. SUNDIAL is able to resolve accurate, confident shadow and albedo predictions compared to EO-NeRF and SatNeRF. In particular, SatNeRF shading scalars encompass both shadows and complex illumination components; while EO-NeRF implements a geometric approach to shadow rendering, this simple direct and indirect lighting separation cannot account for these nuanced, soft shadows, resulting in ambiguous shadow predictions.



(a) JAX 214 estimated geometry and sun dir.

Data ID	Sun Dir. MAE ↓ (°)
JAX 004	8.404 ±3.026
JAX 068	1.928 ±1.332
JAX 214	1.608 ±1.177
JAX 260	5.468 ±4.043

(b) Sun dir. MAE

Figure 6. SUNDIAL-estimated scene geometry and sun directions. SUNDIAL can jointly estimate accurate scene geometry and sun positions (a); SUNDIAL-estimated sun directions are accurate compared to initial dataset-provided directions (b).

Sun-Based Image Location and Time Estimation				
Dataset	Location (lon,lat)		Error (miles)	Time Error (mins)
	True	Predicted	Great-Circle	L1 Error
JAX 004	(-81.706, 30.358)	(-82.200, 29.611)	34.86	12.89 ± 14.34
JAX 068	(-81.664, 30.349)	(-81.700, 30.159)	3.14	6.55 ± 5.88
JAX 214	(-81.664, 30.316)	(-81.576, 30.129)	6.32	4.97 ± 4.57
JAX 260	(-81.663, 30.312)	(-81.433, 29.820)	16.67	14.66 ± 18.35

Table 3. Image time and location prediction using estimated sun position. SUNDIAL accurately predicts the time-of-capture and location of images (within a few minutes and miles of ground truth values) using model-estimated sun directions.

5.5. Applications

Sun Estimation SUNDIAL accurately estimates sun directions jointly with scene geometry (Figure 6) during training, with reasonable accuracy compared to dataset-provided sun directions. This joint estimation helps disentangle scene geometry and illumination even when sun direction estimations are not available.

Image Time and Location Estimation Given known scene location information, we can use SUNDIAL-estimated sun directions to predict image time-of-capture information. Similarly, given time-of-capture information and SUNDIAL-estimated sun directions, we can estimate image location information. We demonstrate these results in Table 3. SUNDIAL can estimate image time to within 5 minutes, and location within 4 miles, of ground truth values.

6. Conclusion

We present SUNDIAL, a technique for reconstructing 3D satellite-captured scenes by decomposing direct, ambient, and complex illumination effects. Our method can accurately disentangle scene albedo, shadows, ambient illumination, complex illumination, and transient features in satellite imagery. We achieve this by jointly estimating sun direction with scene geometry, and casting geometry-based shadow rays that 1) refine hidden scene geometry using oblique sun angles, 2) render separated, geometric shadows, and 3) determine illumination component weights. By disentangling scene geometry and illumination components, we enable accurate novel view and lighting synthesis, unlocking new applications in wide-scale environmental modeling. For example, SUNDIAL can help build detailed illumination and transient-aware city models using drone imagery, or render novel scenes with weather features such as fogs, clouds, and weather events that may otherwise inhibit satellite captures. Future work may consider additional geometric cues to overcome small baselines in low-resolution satellite imagery, such as from near-nadir Sentinel-2 images, or in captured scenes with low sun direction variance, which may provide weaker secondary shadow cues. Our approach demonstrates the importance of incorporating both physics and remote sensing-based insights to extract hidden information from global satellite captures.

References

[1] Michal Adamkiewicz, Timothy Chen, Adam Caccavale, Rachel Gardner, Preston Culbertson, Jeannette Bohg, and Mac Schwager. Vision-only robot navigation in a neural radiance world. *IEEE Robotics and Automation Letters*, 7(2): 4606–4613, 2022. 2

[2] Karine RM Adeline, M Chen, X Briottet, SK Pang, and N Paparoditis. Shadow detection in very high spatial resolution aerial images: A comparative study. *ISPRS Journal of Photogrammetry and Remote Sensing*, 80:21–38, 2013. 2, 3, 4

[3] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5479, 2022. 2, 6

[4] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Zip-nerf: Anti-aliased grid-based neural radiance fields. *arXiv preprint arXiv:2304.06706*, 2023. 2

[5] EF Berra and MV Peppa. Advances and challenges of uav sfm mvs photogrammetry and remote sensing: Short review. In *2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS)*, pages 533–538. IEEE, 2020. 3

[6] Sai Bi, Zexiang Xu, Pratul Srinivasan, Ben Mildenhall, Kalyan Sunkavalli, Milos Hasan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. Neural reflectance fields for appearance acquisition. *arXiv preprint arXiv:2008.03824*, 2020. 3

[7] Michael P Bishop and Jeffrey D Colby. Topographic normalization of multispectral satellite imagery. *Journal of Glaciology*, 55:131–146, 2011. 4

[8] Stefania Bonafoni and Aliihsan Sekertekin. Albedo retrieval from sentinel-2 by new narrow-to-broadband conversion coefficients. *IEEE Geoscience and Remote Sensing Letters*, 17 (9):1618–1622, 2020. 5

[9] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T Barron, Ce Liu, and Hendrik Lensch. Nerd: Neural reflectance decomposition from image collections. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12684–12694, 2021. 3, 7

[10] Mark Boss, Varun Jampani, Raphael Braun, Ce Liu, Jonathan Barron, and Hendrik Lensch. Neural-pil: Neural pre-integrated lighting for reflectance decomposition. *Advances in Neural Information Processing Systems*, 34: 10691–10704, 2021. 3

[11] Gong Cheng, Xingxing Xie, Junwei Han, Lei Guo, and Gui-Song Xia. Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:3735–3756, 2020. 3

[12] Paul M Dare. Shadow analysis in high-resolution satellite imagery of urban areas. *Photogrammetric Engineering & Remote Sensing*, 71(2):169–177, 2005. 3

[13] Dawa Derksen and Dario Izzo. Shadow neural radiance fields for multi-view satellite photogrammetry. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1152–1161, 2021. 2, 3

[14] Gabriele Facciolo, Carlo De Franchis, and Enric Meinhardt-Llopis. Automatic 3d reconstruction from multi-date satellite images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 57–66, 2017. 3

[15] Jian Gao, Jin Liu, and Shunping Ji. Rational polynomial camera model warping for deep learning based satellite multi-view stereo matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6148–6157, 2021. 7

[16] Alvaro Gómez, Gregory Randall, Gabriele Facciolo, and Rafael Grompone von Gioi. An experimental comparison of multi-view stereo approaches on satellite images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 844–853, 2022. 3

[17] Jacek Grodecki and Gene Dial. Block adjustment of high-resolution satellite images described by rational polynomials. *Photogrammetric Engineering & Remote Sensing*, 69 (1):59–68, 2003. 5

[18] Jianhong Guo, Lu Liang, and Peng Gong. Removing shadows from google earth images. *International Journal of Remote Sensing*, 31(6):1379–1389, 2010. 4

[19] Jaroslav Hofierka and Katarína Onačillová. Estimating visible band albedo from aerial orthophotographs in urban areas. *Remote Sensing*, 14(1):164, 2021. 5

[20] R Bruce Irvin and David M McKeown. Methods for exploiting the relationship between buildings and their shadows in aerial imagery. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6):1564–1575, 1989. 2

[21] Yunzhu Li, Shuang Li, Vincent Sitzmann, Pulkit Agrawal, and Antonio Torralba. 3d neural scene representations for visuomotor control. In *Conference on Robot Learning*, pages 112–123. PMLR, 2022. 2

[22] Gregoris Liasis and Stavros Stavrou. Satellite images analysis for shadow detection and building height estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 119: 437–450, 2016. 2

[23] Wen Liu and Fumio Yamazaki. Object-based shadow extraction and correction of high-resolution optical satellite images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(4):1296–1302, 2012. 2, 3

[24] Yuan Liu, Peng Wang, Cheng Lin, Xiaoxiao Long, Jiepeng Wang, Lingjie Liu, Taku Komura, and Wenping Wang. Nero: Neural geometry and brdf reconstruction of reflective objects from multiview images. *arXiv preprint arXiv:2305.17398*, 2023. 3

[25] Nathan Longbotham, Chad Bleiler, Chuck Chaapel, Chris Padwick, William Emery, and Fabio Pacifici. Spatial classification of worldview-2 multi-angle sequence. In *2011 Joint Urban Remote Sensing Event*, pages 105–108. IEEE, 2011. 2

[26] Roger Marí, Carlo de Franchis, Enric Meinhardt-Llopis, and Gabriele Facciolo. To bundle adjust or not: A comparison of relative geolocation correction strategies for satellite

multi-view stereo. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019. 2

[27] Roger Marí, Gabriele Facciolo, and Thibaud Ehret. Sat-nerf: Learning multi-view satellite photogrammetry with transient objects and shadow modeling using rpc cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1311–1321, 2022. 2, 3, 5, 7

[28] Roger Marí, Gabriele Facciolo, and Thibaud Ehret. Multi-date earth observation nerf: The detail is in the shadows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2034–2044, 2023. 3

[29] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7210–7219, 2021. 2, 7

[30] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 2, 3

[31] Doug Mohney. Terabytes from space: Satellite imaging is filling data centers, 2020. 1

[32] Yasser Mostafa. A review on various shadow detection and compensation techniques in remote sensing images. *Canadian journal of remote sensing*, 43(6):545–562, 2017. 3

[33] Ozge C Ozcanli, Yi Dong, Joseph L Mundy, Helen Webb, Riad Hammoud, and Tom Victor. Automatic geo-location correction of satellite imagery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 307–314, 2014. 5

[34] Airton Marco Polidorio, Franklin César Flores, Nilton Nobuhiro Imai, Antonio MG Tommaselli, and Clelia Franco. Automatic shadow segmentation in aerial color images. In *16th brazilian symposium on computer graphics and image processing (SIBGRAPI 2003)*, pages 270–277. IEEE, 2003. 4, 6

[35] Purnama Budi Santosa. Evaluation of satellite image correction methods caused by differential terrain illumination. In *Forum Geografi*, pages 1–13, 2016. 5

[36] JD Shepherd and JR Dymond. Correcting satellite imagery for the variance of reflectance and illumination with topography. *International Journal of Remote Sensing*, 24(17):3503–3514, 2003. 4

[37] Anthony Simeonov, Yilun Du, Andrea Tagliasacchi, Joshua B Tenenbaum, Alberto Rodriguez, Pulkit Agrawal, and Vincent Sitzmann. Neural descriptor fields: Se (3)-equivariant object representations for manipulation. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 6394–6400. IEEE, 2022. 2

[38] Pratul P Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T Barron. Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7495–7504, 2021. 7

[39] Nan Su, Ye Zhang, Shu Tian, Yiming Yan, and Xinyuan Miao. Shadow detection and removal for occluded object information recovery in urban high-resolution panchromatic satellite images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(6):2568–2582, 2016. 2, 3

[40] Matthew Tancik, Vincent Casser, Xinchen Yan, Sabeek Pradhan, Ben Mildenhall, Pratul P Srinivasan, Jonathan T Barron, and Henrik Kretzschmar. Block-nerf: Scalable large scene neural view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8248–8258, 2022. 2

[41] Kushagra Tiwary, Tzofi Klinghoffer, and Ramesh Raskar. Towards learning neural representations from shadows. In *European Conference on Computer Vision*, pages 300–316. Springer, 2022. 3

[42] Kushagra Tiwary, Akshat Dave, Nikhil Behari, Tzofi Klinghoffer, Ashok Veeraraghavan, and Ramesh Raskar. Orca: Glossy objects as radiance-field cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20773–20782, 2023. 3

[43] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5481–5490. IEEE, 2022. 3, 6, 7

[44] Suhani Vora, Noha Radwan, Klaus Greff, Henning Meyer, Kyle Genova, Mehdi SM Sajjadi, Etienne Pot, Andrea Tagliasacchi, and Daniel Duckworth. Nesf: Neural semantic fields for generalizable semantic segmentation of 3d scenes. *arXiv preprint arXiv:2111.13260*, 2021. 2

[45] Tianxing Wang, Jiancheng Shi, Husi Letu, Ya Ma, Xingcai Li, and Yaomin Zheng. Detection and removal of clouds and associated shadows in satellite imagery based on simulated radiance fields. *Journal of Geophysical Research: Atmospheres*, 124(13):7207–7225, 2019. 2

[46] Qihao Weng. Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sensing of Environment*, 117:34–49, 2012. 3

[47] Yuanbo Xiangli, Linning Xu, Xingang Pan, Nanxuan Zhao, Anyi Rao, Christian Theobalt, Bo Dai, and Dahua Lin. Bungeenerf: Progressive neural radiance field for extreme multi-scale scene rendering. In *European conference on computer vision*, pages 106–122. Springer, 2022. 2

[48] Ziyang Xie, Junge Zhang, Wenye Li, Feihu Zhang, and Li Zhang. S-nerf: Neural radiance fields for street views, 2023. 2

[49] Wenqi Yang, Guanying Chen, Chaofeng Chen, Zhenfang Chen, and Kwan-Yee K Wong. S3-nerf: Neural reflectance field from shading and shadow under a single viewpoint. *Advances in Neural Information Processing Systems*, 35:1568–1582, 2022. 3

[50] Ke Yu, Yunhao Chen, Dandan Wang, Zixuan Chen, Adu Gong, and Jing Li. Study of the seasonal effect of building shadows on urban land surface temperatures based on remote sensing data. *remote sensing*, 11(5):497, 2019. 2

[51] Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5453–5462, 2021. 3

[52] Kai Zhang, Fujun Luan, Zhengqi Li, and Noah Snavely. Iron: Inverse rendering by optimizing neural sdfs and materials from photometric images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5565–5574, 2022. 3

[53] Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (ToG)*, 40(6):1–18, 2021. 3

[54] Li Zhao, Haiyan Wang, Yi Zhu, and Mei Song. A review of 3d reconstruction from high-resolution urban satellite images. *International Journal of Remote Sensing*, 44(2):713–748, 2023. 3

[55] Shuaifeng Zhi, Tristan Laidlow, Stefan Leutenegger, and Andrew J Davison. In-place scene labelling and understanding with implicit scene representation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15838–15847, 2021. 2

[56] Tingting Zhou, Haoyang Fu, Chenglin Sun, and Shenghan Wang. Shadow detection and compensation from remote sensing images under complex urban conditions. *Remote Sensing*, 13(4), 2021. 3

[57] Tingting Zhou, Haoyang Fu, Chenglin Sun, and Shenghan Wang. Shadow detection and compensation from remote sensing images under complex urban conditions. *Remote Sensing*, 13(4):699, 2021. 6

[58] Xiaolin Zhu and Eileen H Helmer. An automatic method for screening clouds and cloud shadows in optical satellite image time series in cloudy regions. *Remote sensing of environment*, 214:135–153, 2018. 4

[59] Zihan Zhu, Songyou Peng, Viktor Larsson, Weiwei Xu, Hujun Bao, Zhaopeng Cui, Martin R Oswald, and Marc Pollefeys. Nice-slam: Neural implicit scalable encoding for slam. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12786–12796, 2022. 2