Multi-Date Earth Observation NeRF: The Detail Is in the Shadows

Roger Marí, Gabriele Facciolo, Thibaud Ehret

Multi-Date Earth Observation NeRF: The Detail Is in the Shadows

Roger Marí Gabriele Facciolo Thibaud Ehret
Université Paris-Saclay, CNRS, ENS Paris-Saclay, Centre Borelli, 91190, Gif-sur-Yvette, France
https://rogermm14.github.io/eonerf

Abstract

We introduce Earth Observation NeRF (EO-NeRF), a new method for digital surface modeling and novel view synthesis from collections of multi-date remote sensing images. In contrast to previous variants of NeRF proposed in the literature for satellite images, EO-NeRF outperforms the altitude accuracy of advanced pipelines for 3D reconstruction from multiple satellite images, including classic and learned stereovision methods. This is largely due to a rendering of building shadows that is strictly consistent with the scene geometry and independent from other transient phenomena. In addition to that, a number of strategies are also proposed with the aim to exploit raw satellite images. We add model parameters to circumvent usual pre-processing steps, such as the relative radiometric normalization of the input images and the bundle adjustment for refining the camera models. We evaluate our method on different areas of interest using sets of 10-20 pre-processed and raw pansharpened WorldView-3 images.

1. Introduction

Today, hundreds of satellite Earth observation missions are in operation and the number continues to grow [54]. Many Earth observation satellites acquire optical images periodically over the same areas, contributing to large multi-date collections of satellite images. These collections offer the opportunity to observe the evolution of a site over time and to identify permanent and transient structures. However, when it comes to digital surface modeling from remote sensing, multi-date imagery is usually disregarded in favor of a few stereo or tri-stereo image products. The latter are more expensive, as images are acquired almost simultaneously to facilitate the use of classic photogrammetry.

The state of the art in digital surface modeling from satellite images relies on the availability of stereo image products. The reason of this is that the Multi-View Stereo (MVS) logic is predominantly adopted to address the task as a multi-pair problem instead of a real multi-view problem [18]. This is the case of the NASA Ames stereo pipeline [1], MicMac [40] by the the French National Geographic Institute (IGN), CARS [34] by the French Space Agency (CNES), S2P [11] by the Centre Borelli research center, or the CATENA [23] multi-stereo chain by the German Aerospace Center (DLR). In MVS, each stereo product is used to obtain an independent digital surface model (DSM). When several pairs are available, the resulting DSMs are merged into a single model of higher accuracy.

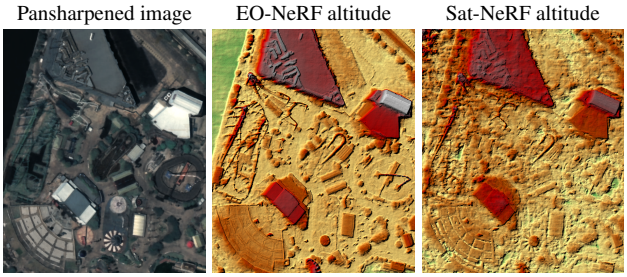

Figure 1. EO-NeRF can be used for novel view synthesis and 3D reconstruction from an input set of multi-date satellite images. Elevation models obtained with EO-NeRF reveal scene geometry with an unprecedented level of detail that surpasses previous NeRF variants [32] and is comparable to that of intrusive aerial acquisitions. This includes narrow and irregular structures such as the arches and roller coasters observed in the pansharpened image.

Parallel to this scenario, MVS approaches for 3D reconstruction have lost hegemony for common close-range imagery. Since the release of NeRF [35] in 2020, neural rendering methods have become extremely popular for conventional image collections, as they represent a highly accurate, unsupervised, and truly multi-view solution to the 3D modeling problem. Satellite imagery is a niche where neural rendering still has a long way to go, but some pioneering work has already started to bridge the gap between the two worlds. In particular, S-NeRF [13] and Sat-NeRF [32] explored NeRF for multi-date satellite images, but their performance remained inferior to state-of-the-art MVS using a set of manually selected pairs.

We introduce EO-NeRF, the *Earth Observation NeRF*, for digital surface modeling and novel view synthesis using sets of multi-date remote sensing images. The altitude accuracy of EO-NeRF outperforms state-of-the-art stereo-

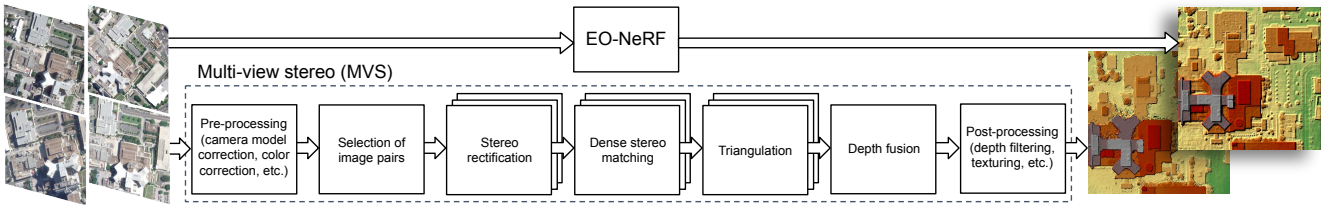

Figure 2. Stereovision methods used in modern production chains for 3D reconstruction from multiple satellite images require control of a multitude of stages, each of which can induce fatal errors. EO-NeRF replaces the entire chain with a single, more practical stage.

vision methods and previous NeRF approaches. Figure 1 shows the level of detail that EO-NeRF achieves in narrow and irregular structures that are normally lost using concurrent methods. This is made possible thanks to the following contributions:

- A NeRF approach for multi-date satellite images that does not *predict* shadows, but renders them according to the geometry and the position of the sun. The geometrically consistent shadows provide hints that permit to refine the geometry, which in turn refines the shadows. This approach also ensures that the model generalizes for novel view synthesis using solar directions completely different from those of the input images.
- A variety of strategies to use unprocessed satellite images. These include the use of UTM coordinates to handle georeferenced data in a more appropriate way and the addition of network parameters to address inaccuracies in the camera models and color biases in the images during the optimization process.

We test EO-NeRF on 7 areas of interest covering 256×256 m each, using $\sim$10-20 crops from multi-date WorldView-3 images with a resolution of 30 cm/pixel. Differently from previous NeRF variants, our evaluation is not limited to pre-processed RGB images, as EO-NeRF can directly handle raw pansharpened products. The input data and EO-NeRF results are released to contribute to the creation of a NeRF benchmark for multi-date satellite images.

2. Related work

Stereovision methods and NeRF can be used for multi-view 3D reconstruction under the assumption that the scene geometry and radiance are invariant or nearly invariant. On the one hand, the strengths of NeRF are its simplicity (no need to select or merge pairs and no geometry supervision), its impressive level of detail and the ability to solve the colorization problem simultaneously with 3D estimation. On the other hand, the weaknesses of NeRF are the need for larger sets of input images and a long optimization time (usually hours), whereas stereo methods can quickly produce coarse geometry estimates with as few as two images.

In this section we review the fundamentals of NeRF, stereovision satellite 3D reconstruction, and NeRF variants for multi-date images that are most relevant to our work.

2.1. NeRF in a nutshell

A NeRF, or Neural Radiance Field [35], is a continuous function $\mathcal{F}$ that represents the geometry and appearance of a 3D scene. $\mathcal{F}$ is encoded using a fully-connected neural network or multi layer perceptron (MLP) that is specific to each scene and does not generalize to others. In its simplest form, a NeRF MLP takes as input some 3D coordinates $\mathbf{x}$ and, optionally, a viewing direction $\mathbf{d}$, which are used to predict the observed RGB color $\mathbf{c}$ and a non-negative scalar volume density σ at the input point.

$$\mathcal{F} : (\mathbf{x}, \mathbf{d}) \mapsto (\sigma, \mathbf{c}). \tag{1}$$

Using a collection of input views and their camera poses, the MLP encoding (1) is optimized by casting individual rays that project onto the known pixels. Each ray $\mathbf{r}$ is defined by a point of origin $\mathbf{o}$ and a direction vector $\mathbf{d}$. The color $\mathbf{c}(\mathbf{r})$ of a ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ is computed as

$$\mathbf{c}(\mathbf{r}) = \sum_{i=1}^{N} T_i \alpha_i \mathbf{c}_i. \tag{2}$$

The rendered color $\mathbf{c}(\mathbf{r})$ results from integrating the colors $\mathbf{c}_i$ predicted at different points of the ray $\mathbf{r}$. Each ray $\mathbf{r}$ is therefore discretized into N 3D points $\mathbf{x}_i$ between the near and far bounds of the scene, t_n and t_f. Each point $\mathbf{x}_i$ in $\mathbf{r}$ is obtained as $\mathbf{x}_i = \mathbf{o} + t_i\mathbf{d}$, where $t_i \in [t_n, t_f]$. The contribution of each point $\mathbf{x}_i$ in $\mathbf{r}$ to the rendered color (2) depends on the opacity α_i and the transmittance T_i:

$$\alpha_i = 1 - \exp(-\sigma_i \delta_i) \quad \text{and} \quad T_i = \prod_{j=1}^{i-1} (1 - \alpha_j). \tag{3}$$

Both factors take values in the interval $[0, 1]$ and depend only on the volume density σ that defines the geometry. The opacity increases with σ and is the probability that $\mathbf{x}_i$ belongs to a non-transparent surface. The transmittance is the probability that light reaches $\mathbf{x}_i$ without hitting previous opaque points in the ray $\mathbf{r}$ and is used to handle occlusions.

Using (3), the depth $d(\mathbf{r})$ observed by casting a ray $\mathbf{r}$ can be determined in a similar manner to the color (2) as

$$d(\mathbf{r}) = \sum_{i=1}^{N} T_i \alpha_i t_i. \tag{4}$$

The MLP is optimized by minimizing the mean squared error (MSE) between the rendered color $\mathbf{c}(\mathbf{r})$ and the real color $\mathbf{c}_{\text{GT}}(\mathbf{r})$ of the image pixel intersected by each ray $\mathbf{r}$:

$$\sum_{\mathbf{r} \in \mathcal{R}} \|\mathbf{c}(\mathbf{r}) - \mathbf{c}_{\text{GT}}(\mathbf{r})\|_2^2, \tag{5}$$

where $\mathcal{R}$ is a batch of arbitrarily selected rays.

2.2. Classic and learned stereovision methods

Stereovision 3D reconstruction pipelines usually share some fundamental steps illustrated in Figure 2. The most critical task is dense matching or disparity estimation.

Among the classic methods for disparity estimation, Semi-Global Matching (SGM) [21] is the preferred choice for satellite imagery due to its efficiency to exploit spatial regularity along different cardinal directions. A good number of DSM production pipelines use SGM with the census transform [51] as matching cost for its robustness to illumination changes [10, 34, 43]. Other pipelines employ SGM variants: MGM [15, 16] increases the directions taken into account for regularization without loss of efficiency, tSGM [27, 39] adopts a coarse-to-fine strategy to limit the disparity search range, Semi-Global Block Matching [25] uses windows instead of individual pixels, etc.

Learned methods for disparity estimation use deep neural networks. These networks comprise different modules dedicated to feature extraction, cost volume construction, cost volume regularization and disparity regression (based on minimum cost). Architectures such as PSM [5], HSM [45] or GA-Net [52] have been tested on remote sensing images with encouraging results [19, 31, 44]. Unlike classic algorithms, deep neural networks are prone to fail or lose accuracy in unseen scenarios (e.g., different shapes, disparity range, viewing angles or rectification criteria). In addition to this generalization issue, learned methods (including multi-view ones, such as MVSNet [47]) are subject to long training times (in the order of days or weeks) and require large disparity benchmarks for supervision [44].

2.3. NeRF for multi-date image collections

NeRF [35] opened the door to a large number of variants that address the limitations of the original method. This includes speeding up the optimization process [6, 17, 36, 42], reducing the number of input views [14, 22, 37] or generalizing to unseen scenes [7, 9, 49]. Certain variants have also addressed the use of in-the-wild photo collections, such as multi-date images, presenting changes in appearance and geometry across the input views [13, 32, 33].

The trend in NeRF variants for unconstrained photo collections is to add auxiliary networks or heads to the single MLP architecture originally used in NeRF. The color rendering operation (2) and/or the loss function (5) are modified to account for image-dependent reflectance models that can accommodate the complex inconsistencies between the input views while recovering the underlying shared geometry [24, 53]. Image-dependent features are normally extracted using embedding modules learned during the NeRF optimization [33] or independently pre-trained convolutional neural networks [8].

NeRF in the Wild (or NeRF-W) [33] was created for collections of hundreds or thousands of multi-date street view images. It added auxiliary heads to predict transient colors and densities in parallel to the original multi-view consistent outputs (1). *Shadow NeRF* (or S-NeRF) [13] worked with tens of satellite images modeled locally as pinhole cameras. An auxiliary head was used to render building shadows in the input images as a function of the direction of solar rays. *Satellite NeRF* (or Sat-NeRF) [32] improved the results of S-NeRF by directly employing the actual RPC (Rational Polynomial Coefficients) camera models associated with the satellite images and by adding an auxiliary head to handle transient objects similarly to NeRF-W.

3. Method

EO-NeRF is optimized using the ray casting strategy proposed in NeRF [35] (Section 2.1). The original NeRF function (1) is transformed into:

$$\mathcal{F} : (\mathbf{x}, \mathbf{d}_{\text{sun}}, \mathbf{t}_j) \mapsto (\sigma, \mathbf{c}_{\text{a}}, \mathbf{a}, \beta, \tau, \mathbf{A}_j, \mathbf{b}_j). \tag{6}$$

The inputs of (6) are the same as in Sat-NeRF [32]: the 3D point coordinates $\mathbf{x}$, the solar direction vector $\mathbf{d}_{\text{sun}}$ and an image-dependent embedding vector $\mathbf{t}_j$ (where j is the image index). The outputs of (6) are the following:

- σ: Volume density scalar at location $\mathbf{x}$.
- $\mathbf{c}_{\text{a}}$: Albedo RGB, related only to the geometry, i.e. the spatial coordinates $\mathbf{x}$.
- $\mathbf{a}$: Ambient RGB color that defines a global hue according to the solar direction $\mathbf{d}_{\text{sun}}$.
- β: Uncertainty scalar related to the probability that the color of $\mathbf{x}$ corresponds to a transient object.
- τ: Transient scalar, learned as a function of $\mathbf{x}$ and $\mathbf{t}_j$.
- $\mathbf{A}_j$, $\mathbf{b}_j$: RGB vectors encoding a color affine transformation between the albedo and the current image j.

EO-NeRF uses the outputs of (6) to render the color $\mathbf{c}(\mathbf{r})$ of each ray $\mathbf{r}$ as follows:

$$\mathbf{c}(\mathbf{r}) = \mathbf{A}_j \left(\ell(\mathbf{r}) \cdot \sum_{i=1}^{N} T_i \alpha_i \mathbf{c}_{\text{a}} \right) + \mathbf{b}_j \tag{7}$$

where

$$\ell(\mathbf{r}) = s(\mathbf{r}) + (1 - s(\mathbf{r}))\mathbf{a} \quad \text{with } s(\mathbf{r}) \text{ defined in (9).} \tag{8}$$

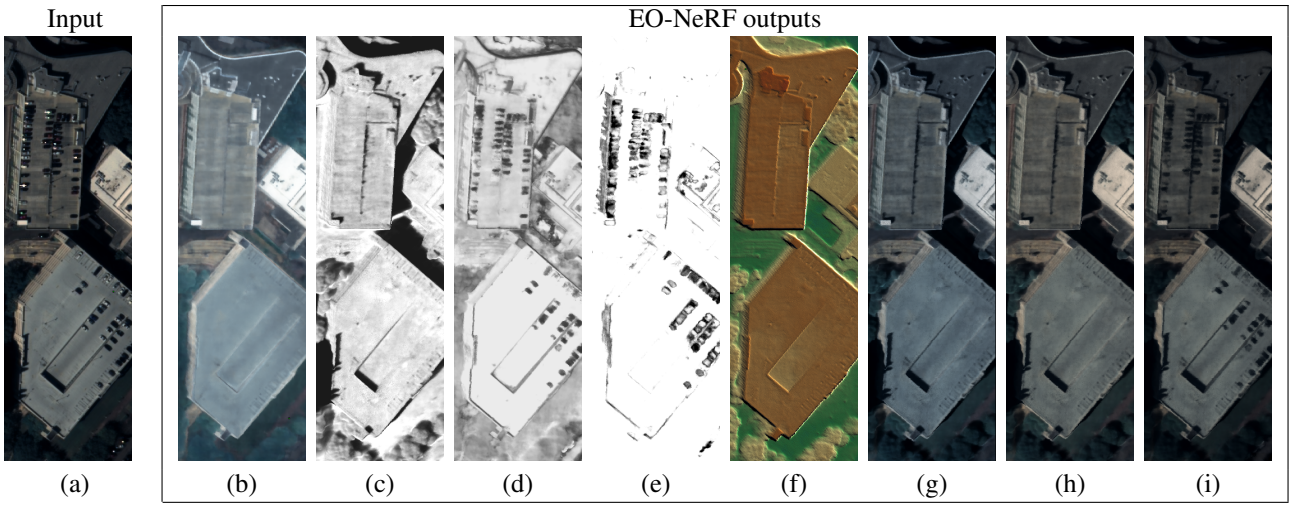

Figure 3. Left to right: (a) Input image, (b) albedo, (c) geometric shadows, (d) transient scalar, (e) uncertainty scalar, (f) altitude, (g) albedo with geometric shadows irradiance, (h) albedo with geometric shadows irradiance after the affine transformation defined by $[\mathbf{A}_j, \mathbf{b}_j]$, (i) albedo with geometric shadows and transient objects irradiance after the affine transformation, i.e., the rendered color $\mathbf{c}(\mathbf{r})$ from (7).

Figure 3 illustrates the different outputs of EO-NeRF. According to (7), the output color $\mathbf{c}(\mathbf{r})$ results from the product of the albedo and an irradiance model $\ell(\mathbf{r})$, subject to an image-specific affine transformation defined by $\mathbf{A}_j$ and $\mathbf{b}_j$. Affine correction models are a common practice in the literature for relative radiometric normalization of multi-date satellite imagery [20, 50]. The key element in the irradiance model (8) is $s(\mathbf{r})$, a scalar in the interval $[0, 1]$. Low values of $s(\mathbf{r})$ correspond to points in the shade or involving transient objects, where the ambient color $\mathbf{a}$ is allowed to be stronger. High values of $s(\mathbf{r})$ correspond to points of the scene that can be fully explained by the albedo color. S-NeRF and Sat-NeRF used irradiance models similar to (8), but the way $s(\mathbf{r})$ is defined in EO-NeRF is critically different as explained in Section 3.1.

3.1. Geometrically consistent shadow model

Both S-NeRF and Sat-NeRF predict shadows as a color property depending on the solar direction [13, 32]. This leads the model to overfit the input views, taking advantage of that freedom to render other complex changes such as vegetation color or transient objects. As a result, it does not generalize well to solar directions different from those in the input images (Figure 4). S-NeRF and Sat-NeRF attempted to minimize this misbehavior by adding a solar correction term to the loss that seeks to bring shadow renderings closer to binary masks strictly related to object shadows. At the same time, this choice forces the network to explain differently the other transient phenomena, which may have side effects, so it is necessary to set an appropriate weight for the auxiliary solar correction term.

In EO-NeRF, we propose a model that does not predict shadows but instead renders them based on the geometry at each optimization step. We denote the rendering of geometrically consistent shadows as s_{geo}. The value of s_{geo} is obtained by casting rays from each surface point to the source of light, similarly to [2, 41, 46]. For each camera ray $\mathbf{r}$, we find the point on the surface using (4), i.e., $\mathbf{x}_S$ in Figure 5. This point is taken as the origin to cast a solar ray, $\mathbf{r}_{\text{sun}}$, towards the position of the sun. The transmittance T (3) of the last point in $\mathbf{r}_{\text{sun}}$ is equal to the amount of geometric shadow, $s_{\text{geo}} = T(\mathbf{r}_{\text{sun}}(t_N))$, which reveals whether $\mathbf{x}_S$ is visible in the direction of the sun or not.

The value of s_{geo} renders shadows based on the geometry of the scene and ignores the other transient phenomena in each image. To represent the latter, we use the transient scalar τ predicted by the transient head of EO-NeRF (MLP 3 in Figure 6). τ is an arbitrary scalar between 0 and 1 that is not bound by geometry. The product of s_{geo} with τ defines the $s(\mathbf{r})$ of the irradiance model (8):

$$s(\mathbf{r}) = s_{\text{geo}}(\mathbf{r})\tau(\mathbf{r}) = T(\mathbf{r}_{\text{sun}}(t_N)) \sum_{i=1}^{N} T_i \alpha_i \tau(\mathbf{x}_i), \quad (9)$$

where the dependency of τ on $\mathbf{t}_j$ is omitted for simplicity and $\mathbf{x}_i \in \mathbf{r}$. The use of a product in (9) guarantees that $s(\mathbf{r})$ always contains geometrically consistent shadows in addition to the rest of transient phenomena. $\tau(\mathbf{r}) = 1$ can be set at test time for transient object removal, as in Figure 3(h).

Compared to previous work, the strength of (9) is that it decouples shadows from other transient phenomena. Note that casting solar rays is also necessary to compute the solar correction term used in S-NeRF and Sat-NeRF [13, 32]. We use solar rays in a more direct way, which does not require modifying the loss function or tuning any hyperparameters to balance the optimization process.

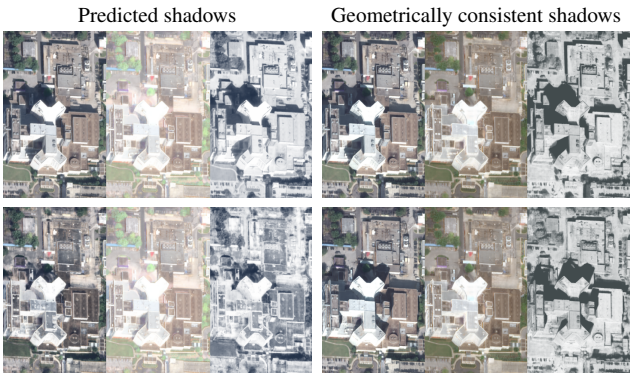

Figure 4. Left to right (in each set): output RGB, albedo and irradiance $\ell(\mathbf{r})$ (8). Predicted shadows, as in Sat-NeRF or S-NeRF, can correctly interpolate within the range of solar directions observed in the input images. However, they produce unrealistic results outside this range. Top row: Solar direction taken from the input data. Bottom row: Solar direction perpendicular to the sun path, which could never be observed in a real satellite image of the area. The geometrically consistent shadow model used in EO-NeRF can realistically render arbitrary solar directions as that of the bottom row and simultaneously improve the geometry estimation.

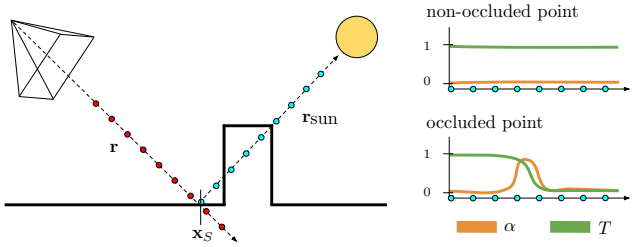

Figure 5. EO-NeRF shadow model. The transmittance T of the last point in a ray $\mathbf{r}_{\text{sun}}$ in the direction of the sun indicates the amount of shadow at the surface point $\mathbf{x}_S$. The pinhole model is adopted to represent the camera geometry for simplicity.

3.2. Network architecture and loss function

The EO-NeRF network architecture is shown in Figure 6. The main MLP 1 takes as input the 3D point coordinates $\mathbf{x}$ and predicts the volume density σ, as in the original NeRF [35]. The features extracted by MLP 1 are then plugged into the auxiliary heads MLP 2 and MLP 3. MLP 2 uses them to predict the albedo, while MLP 3 merges them with $\mathbf{t}_j$ to predict the transient-related magnitudes, i.e., β and τ. MLP 4 is disconnected from the rest as it predicts the ambient color $\mathbf{a}$, which is constant for all points $\mathbf{x}$ and only depends on the solar direction $\mathbf{d}_{\text{sun}}$. All image-dependent vectors $\mathbf{t}_j$, $\mathbf{A}_j$, $\mathbf{b}_j$ are learned using embedding modules that take as input the image index j. We use positional encoding [35] for inputs $\mathbf{x}$ and $\mathbf{d}_{\text{sun}}$, as well as ReLU nonlinearity between MLP layers.

EO-NeRF uses a loss function similar to NeRF-W [33] and Sat-NeRF [32], where the uncertainty scalar β reduces the contribution of camera rays emitted from pixels with a high probability of representing transient phenomena:

$$\sum_{\mathbf{r}\in\mathcal{R}} \frac{\|\mathbf{c}(\mathbf{r}) - \mathbf{c}_{\text{GT}}(\mathbf{r})\|_2^2}{2\beta'(\mathbf{r})^2} + \left(\frac{\log \beta'(\mathbf{r}) + \eta}{2}\right), \qquad (10)$$

where $\beta'(\mathbf{r}) = \beta(\mathbf{r}) + \beta_{\text{min}}$. As in [32], $\beta_{\text{min}} = 0.05$ and $\eta = 3$ are used in (10) to avoid negative values in the logarithm. The logarithm prevents β from converging to infinity to minimize (10) and encourages the network to correctly use β to identify which points are worthy of being regarded as transient or not depending on each image.

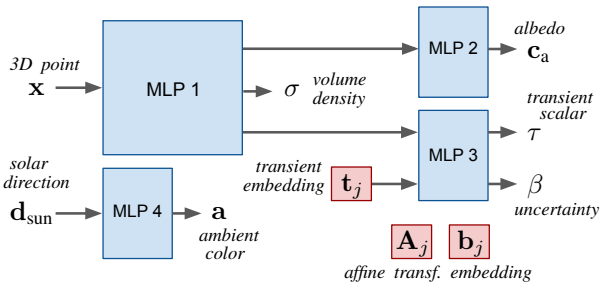

Figure 6. EO-NeRF network architecture, with the parameters to be optimized delimited by boxes. Multi layer perceptron components are shown in blue and embedding vectors are shown in red.

3.3. UTM-based point coordinates

S-NeRF [13] approximated the RPC camera models of the input images as pinhole camera models. Sat-NeRF [32] demonstrated the benefits in accuracy of directly using the RPC models to sample points in the 3D space. For that purpose, Sat-NeRF uses Earth-centered Earth-fixed (ECEF) coordinates to represent points in the 3D space. We follow the RPC-based sampling but using UTM coordinates and altitude to represent 3D points, which locally preserves the properties of a Cartesian system and offers the advantage that the altitude of the scene is aligned with the z-axis. As shown in Figure 7, this ensures a better use of the space occupied by the scene with respect to the 3D volume that contains it.

3.4. Internal bundle adjustment

Sat-NeRF stressed the importance of bundle adjustment to ensure consistency between the input RPC camera models associated with the satellite images. However, it proposed to perform the bundle adjustment correction *before* the MLP optimization, using an independent software [29]. Recent work with NeRF shows that it is possible to refine the camera models simultaneously *during* the MLP optimization [28,48]. Especially if a good initialization is available, as in the case of satellite RPC camera models [30].

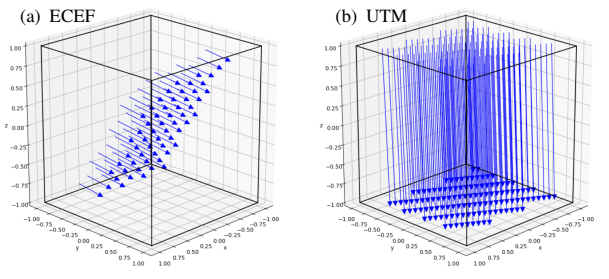

Figure 7. ECEF vs. UTM-based representation of geographic 3D point coordinates. The scene has an altitude range of $[h_{\min}, h_{\max}]$. Using the same RPC camera, the rays in blue originate at different pixels, localized at $h_{\max}$, and end at the same pixel, localized at $h_{\min}$. All coordinates are normalized in the interval $[-1, 1]$.

We propose a simple way to integrate the refinement of RPC camera models into the EO-NeRF optimization, based on the assumption that RPC cameras can be locally approximated as affine projection models [11]. Under such assumption it is common to refine each RPC by composition of the projection function with an offset correction (i.e., a translation on the image plane) [30, 38]. In our UTM-based coordinate system the ray origin related to each pixel lies on the upper plane of the 3D volume (see Figure 7(b)), which is coincident with the maximum altitude of the scene [32]. Thus, the offset correction can be approximated as a displacement in the x and y axis. If $\mathbf{r}$ is a ray intersecting the j-th camera, with origin $\mathbf{o}$ and direction $\mathbf{d}$, the bundle adjusted version of $\mathbf{r}$ can be approximated as

$$\mathbf{r}(t) = (\mathbf{o} + \mathbf{q}_j) + t\mathbf{d} \quad \text{such that} \quad \mathbf{q}_j = (q_1, q_2, 0)_j. \tag{11}$$

Expression (11) replaces the usual $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ and is constrained to points inside the boundaries of the cube containing the scene to prevent large displacements. The offset coefficients q_1 and q_2 of each camera are learned using a two-dimensional embedding module in the same fashion as $\mathbf{t}_j$, $\mathbf{A}_j$ and $\mathbf{b}_j$. In Section 4, the results show that using (11) significantly narrows the gap compared to correcting the camera models separately before optimizing the MLP.

4. Experiments

We evaluate EO-NeRF on 7 areas of interest (AOI) covering 256 × 256 m each, using ∼10-20 crops from multi-date WorldView-3 images with a resolution of 30 cm/pixel. The images are taken from the public data of the *2019 IEEE GRSS Data Fusion Contest* (DFC2019) [3, 26] and *2016 IARPA Multi-View Stereo 3D Mapping Challenge* (IARPA2016) [4]. The specific number of images and other details of each AOI are shown in Table 1. In both cases, we used the panchromatic and multispectral products to create raw pansharpened images without any color correction. The DFC2019 data additionally provides true color RGB images over the AOIs with reduced dynamic range in 8-bit unsigned integer format. We assess the EO-NeRF elevation models using lidar DSMs with a resolution of 30-50 cm/pixel. The PSNR of the image renderings and the altitude mean absolute error (MAE) with respect to the lidar data are used as evaluation metrics. Geometrically corrected RPC camera models obtained with the bundle adjustment package [29] were used in all experiments unless *raw RPCs* is mentioned.

DFC2019 data — Location: Jacksonville (United States)

Area index	004	068	214	260
Input images	9	17	21	15
Alt. bounds [m]	[-24, 1]	[-27, 30]	[-29, 73]	[-30, 13]
Latitude	30.357	30.348	30.316	30.311
Longitude	-81.706	-81.663	-81.663	-81.663

IARPA2016 data — Location: Buenos Aires area (Argentina)

Area index	001	002	003
Input images	25	21	21
Alt. bounds [m]	[12, 59]	[12, 80]	[12, 80]
Latitude	-34.490	-34.447	-34.417
Longitude	-58.584	-58.575	-58.575

Table 1. Number of input images used for each area, altitude bounds of the scene and approximate latitude and longitude.

4.1. Implementation details

We use a batch size of 1024 rays, Adam optimizer and an initial learning rate of $5e^{-4}$. The first optimization steps are performed using the standard MSE loss (5). The complexity of the optimization problem is gradually increased: first by plugging β and adopting the full loss (10) and then by adding the internal bundle adjustment offsets (11). Both camera rays $\mathbf{r}$ and solar rays $\mathbf{r}_{\text{sun}}$ are discretized into $N = 128$ evenly distributed samples to encourage detailed shadows. For novel view synthesis from unseen viewpoints, we use affine projection camera models to locally approximate unknown RPC functions [30]. Convergence takes about 250-300k steps (∼10-20 h, similarly to Sat-NeRF, depending on the number of images using a 12 GB GPU).

4.2. DFC2019 areas - Results and discussion

Figure 8 shows the DSMs of the DFC2019 areas obtained with EO-NeRF and concurrent methods. We propose three categories of experiments, corresponding to the rows of Table 2 discussed below. EO-NeRF achieves the lowest MAE on average, as indicated in bold in Table 2. Due to the difficulty of computing PSNR for unseen views (e.g. change of ambient color or transients), it is computed using the input views and thus might not improve over Sat-NeRF. The latter has more freedom to overfit the input views since shadows are not forced to be consistent with geometry.

Category 1. We test the state-of-the-art Sat-NeRF model using $N = 64$ points per ray (row 1), as in [32]. Using UTM-based point coordinates (Section 3.3) we obtain similar or better altitude MAE with respect to the results originally reported in [32], while using half of units in the MLP

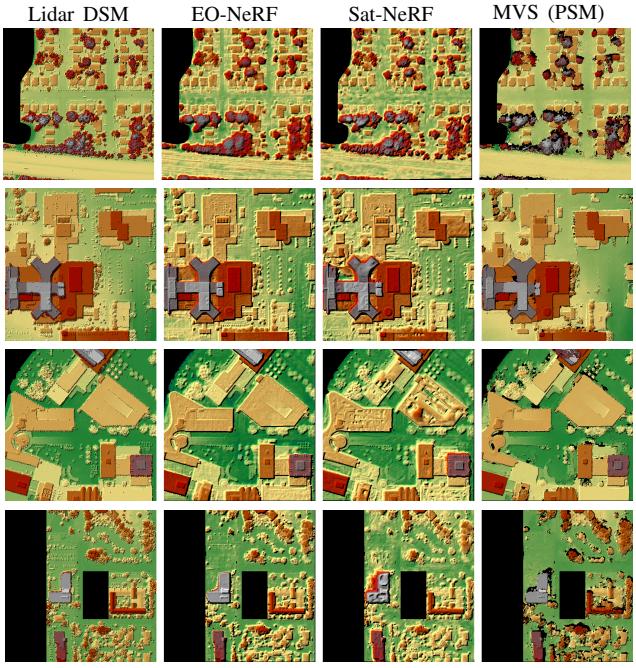

Figure 8. DFC2019 areas. Top to bottom: 004, 068, 214, 260. DSM resolution: 50 cm/pixel. For each method, the DSM that achieved the best altitude MAE is chosen between true color RGB and raw pansharpened inputs.

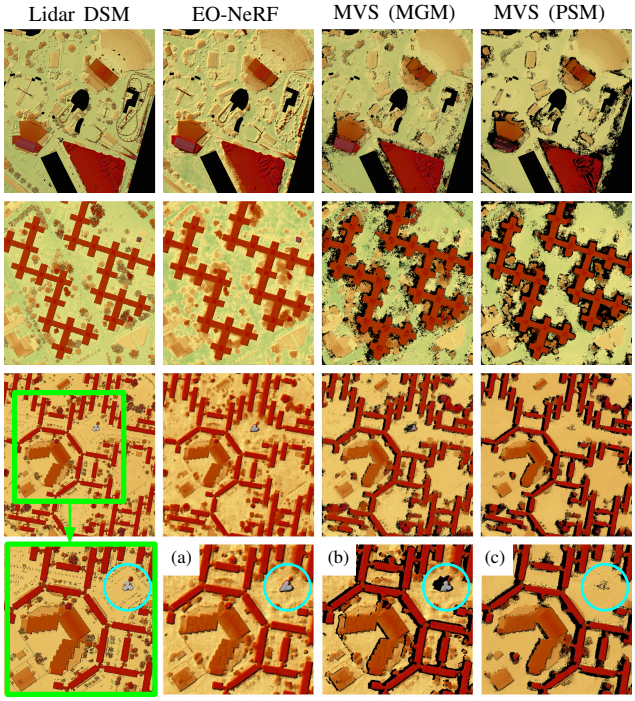

Figure 9. IARPA2016 areas. Top to bottom: 003, 002, 001. DSM resolution: 30 cm/pixel. Close-up views at the bottom: (a) full EO-NeRF prediction, (b) EO-NeRF prediction at points observed by more than 3/4 of all input cameras, (c) PSM prediction. Similarly to the left-right consistency check in stereo methods such as PSM, (b) rejects frequently occluded points near object borders. The altitude MAE is computed using the surface points in (b).

layers ($h = 256$ instead of $h = 512$). Increasing the number of points per ray to $N = 128$ (row 2) does not necessarily improve the performance. We note that Sat-NeRF does not adapt to non-radiometrically normalized inputs, as the MAE increases significantly using raw pansharpened data. Visual inspection of the DSMs in Figure 8 shows that Sat-NeRF is more prone to surface irregularities than EO-NeRF. In fact, Sat-NeRF explored an auxiliary depth supervision term to prevent geometry irregularities [12]. The geometrically consistent shadow model of EO-NeRF is a natural solution to prevent undesired holes or blobs in the geometry that would result in unrealistic shadows.

Category 2. We test EO-NeRF (row 3) using the same externally bundle adjusted RPC cameras as in the Sat-NeRF experiments. The altitude MAE improves dramatically by an average of more than half a meter. The improvement is especially noticeable for raw pansharpened images. Using the unrefined camera models from the image metadata significantly reduces performance in both PSNR and MAE (row 4), unless the EO-NeRF internal bundle adjustment is active (row 5). Externally corrected RPCs provide the most consistent geometry by a small margin, which we attribute to the fact that the proposed offset correction is highly dependent on the affine approximation of the RPC functions.

Category 3. We run the open-source stereo pipeline S2P [11] to reconstruct 10 manually selected image pairs. The pairs are selected following the heuristic criterion that won the 2016 IARPA multi-view stereo challenge using S2P [16]. All pairs are reconstructed using bundle adjusted RPCs and they are merged into a single surface model by taking the median altitude along the z-axis. For disparity estimation, we use the classic MGM algorithm (row 6) [15] and the Pyramid Stereo Matching (PSM) network (row 7) [5]. Pre-trained PSM weights were taken from a public benchmark for aerial imagery [44]. MGM is clearly outperformed by EO-NeRF in terms of altitude MAE, by an average difference of several dozen centimeters. PSM provides better MAE in certain areas using pansharpened inputs, but visual inspection reveals missing structures in the DSMs that only EO-NeRF manages to capture. For instance, certain tall buildings as highlighted in Figure 9 (bottom row, cyan circle) are lost because they fall outside the maximum disparity range used to train the PSM network.

It should be noted that the MVS surface models in Figures 8 and 9 present incomplete regions (in black) beyond the water zones (also in black), which we ignore to compute the altitude MAE. As explained in Figure 9, for a fair comparison, we remove surface points non-observed by the majority of cameras before computing the altitude error.

Area index	True color RGB (uint8) PSNR ↑ / Alt. MAE [m] ↓ 004	068	214	260	Raw pansharpened (float32) PSNR ↑ / Alt. MAE [m] ↓ 004	068	214	260	Mean
1. Sat-NeRF ($N=64$, $h=256$, UTM)	29.94 / 1.34	27.72 / 0.94	26.86 / 1.90	26.91 / 1.70	38.67 / 3.02	33.74 / 1.24	31.84 / 2.53	35.97 / 2.43	31.46 / 1.89
2. Sat-NeRF ($N=128$, $h=256$, UTM)	28.89 / 1.48	27.18 / 0.98	26.08 / 2.29	27.75 / 1.76	38.59 / 2.64	32.80 / 1.39	30.89 / 2.38	34.66 / 2.37	30.86 / 1.91
3. Ours ($N=128$, $h=256$, UTM)	28.56 / 1.25	27.25 / 0.91	26.59 / 1.52	26.09 / 1.43	38.13 / 1.33	33.33 / 0.89	31.48 / 1.41	34.77 / 1.34	30.78 / **1.26**
4. Ours, raw RPCs	27.78 / 1.39	26.11 / 1.42	25.62 / 1.74	25.07 / 1.95	37.55 / 1.72	31.54 / 1.08	30.21 / 1.58	32.08 / 1.53	29.49 / 1.55
5. Ours, raw RPCs+BA	27.93 / 1.30	27.27 / 0.90	26.36 / 1.38	25.75 / 1.48	37.56 / 1.68	32.88 / 0.95	31.43 / 1.44	33.89 / 1.36	30.38 / 1.31
6. Classic MVS (MGM)	— / 1.97	— / 1.71	— / 2.49	— / 2.16	— / 1.37	— / 1.17	— / 1.81	— / 1.64	— / 1.79
7. Learned MVS (PSM)	— / 2.28	— / 0.88	— / 1.77	— / 1.67	— / 1.16	— / 0.83	— / 1.48	— / 1.20	— / 1.41

Table 2. Numerical results, DFC2019 areas. Sat-NeRF fails to handle the raw pansharpened images, while MVS methods struggle with the less textured true color RGB images. EO-NeRF achieves the best overall MAE with no major differences between the two input types.

Area index	Raw pansharpened (float32) PSNR ↑ / Alt. MAE [m] ↓ 001	002	003	Mean
1. Sat-NeRF ($N=128$, $h=256$, UTM)	30.55 / 2.12	31.78 / 2.27	32.79 / 2.75	31.71 / 2.38
2. Ours ($N=128$, $h=256$, UTM)	31.70 / 1.23	31.71 / 1.43	33.07 / 1.19	32.16 / **1.28**
3. Ours, raw RPCs	29.68 / 2.07	29.78 / 2.39	30.97 / 1.99	30.14 / 2.15
4. Ours, raw RPCs+BA	31.76 / 1.37	31.72 / 1.55	33.28 / 1.30	32.25 / 1.41
5. Classic MVS (MGM)	— / 1.40	— / 2.48	— / 1.34	— / 1.74
6. Learned MVS (PSM)	— / 1.15	— / 1.44	— / 1.27	— / 1.29

Table 3. Numerical results, IARPA2016 areas.

4.3. IARPA2016 areas - Results and discussion

Figure 9 shows the DSMs of some IARPA2016 challenge areas obtained with EO-NeRF. We repeat the same categories of experiments of Section 4.2. Only raw pansharpened images are available in this case. The numerical results in the different rows of Table 3 are discussed below.

Category 1. We test Sat-NeRF using UTM-based point coordinates and $N = 128$ points per ray (row 1). Consistent with the DFC2019 results, the model does not fit non-radiometrically normalized inputs and the altitude MAE is above 2 m in all areas. A detailed view of the altitude provided by Sat-NeRF in area 003 is shown in Figure 1.

Category 2. We test EO-NeRF using externally bundle adjusted and internally bundle adjusted RPC camera models (row 2 and 4 respectively). When the input RPCs are consistent, because of the external or the internal bundle adjustment, EO-NeRF outperforms Sat-NeRF by about one meter of altitude in average. This advantage is lost if the unrefined camera models are used instead (row 3).

Category 3. As in Section 4.2 (Category 3), we run the stereo pipeline S2P using 10 manually selected image pairs and MGM and PSM for dense matching. Again, EO-NeRF achieves better overall altitude MAE with respect to MGM and PSM. The DSMs in Figure 9 show the superior level of detail in the shapes reconstructed by EO-NeRF. Note that the geometry of narrow and irregular structures, such as the arches or roller coasters in the area 003 (top row) is only visible in the lidar and EO-NeRF DSMs.

5. Conclusion

We presented EO-NeRF, a variant of NeRF for Earth observation adapted to multi-date satellite images. Our method achieves state-of-the-art novel view synthesis and digital surface modeling from this kind of data. Using input views with a resolution of 30 cm/pixel, EO-NeRF elevation models reveal scene geometry with a level of detail comparable to intrusive aerial acquisitions.

EO-NeRF improves on previous concurrent NeRF variants mainly thanks to a geometrically consistent rendering of shadows that does not modify the loss function and the addition of network parameters to handle raw satellite images. The fine-scale details recovered by EO-NeRF are lost in modern multi-view stereo pipelines for satellite 3D reconstruction, which are not adapted for multi-date inputs. Classic matching algorithms are usually subject to strong regularization, whereas deep neural networks may have generalization problems. While there is much room for improvement, EO-NeRF demonstrates that neural rendering has the potential to take Earth observation to another level.

Acknowledgements. Work partly financed by Office of Naval research grant N00014-17-1-2552, MENRT, and Kayrros. It was also performed using HPC resources from GENCI–IDRIS (grants 2023-AD011014015 and AD011011801R3) and from the "Mésocentre" computing center of CentraleSupélec and ENS Paris-Saclay supported by CNRS and Région Île-de-France (http://mesocentre.centralesupelec.fr). Centre Borelli is also with Université Paris Cité, SSA and INSERM.

References

[1] Ross A Beyer, Oleg Alexandrov, and Scott McMichael. The Ames Stereo Pipeline: NASA's open source software for deriving and processing terrain data. *Earth and Space Science*, 5(9):537–548, 2018. 1

[2] Sai Bi, Zexiang Xu, Pratul Srinivasan, Ben Mildenhall, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. Neural reflectance fields for appearance acquisition. *arXiv preprint arXiv:2008.03824*, 2020. 4

[3] Marc Bosch, Kevin Foster, Gordon Christie, Sean Wang, Gregory D Hager, and Myron Brown. Semantic stereo for incidental satellite images. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1524–1532, 2019. 6

[4] Marc Bosch, Zachary Kurtz, Shea Hagstrom, and Myron Brown. A multiple view stereo benchmark for satellite imagery. In *2016 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, pages 1–9, 2016. 6

[5] Jia-Ren Chang and Yong-Sheng Chen. Pyramid stereo matching network. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5410–5418, 2018. 3, 7

[6] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. TensoRF: Tensorial radiance fields. In *Computer Vision – ECCV 2022*, pages 333–350, 2022. 3

[7] Anpei Chen, Zexiang Xu, Fuqiang Zhao, Xiaoshuai Zhang, Fanbo Xiang, Jingyi Yu, and Hao Su. MVSNeRF: Fast generalizable radiance field reconstruction from multi-view stereo. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14104–14113, 2021. 3

[8] Xingyu Chen, Qi Zhang, Xiaoyu Li, Yue Chen, Ying Feng, Xuan Wang, and Jue Wang. Hallucinated neural radiance fields in the wild. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12933–12942, 2022. 3

[9] Julian Chibane, Aayush Bansal, Verica Lazova, and Gerard Pons-Moll. Stereo radiance fields (SRF): Learning view synthesis for sparse views of novel scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7907–7916, 2021. 3

[10] Pablo d'Angelo and Peter Reinartz. Digital elevation models from stereo, video and multi-view imagery captured by small satellites. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43-B2-2021:77–82, 2021. 3

[11] Carlo de Franchis, Enric Meinhardt-Llopis, Julien Michel, Jean-Michel Morel, and Gabriele Facciolo. An automatic and modular stereo pipeline for pushbroom images. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2-3:49–56, 2014. 1, 6, 7

[12] Kangle Deng, Andrew Liu, Jun-Yan Zhu, and Deva Ramanan. Depth-supervised NeRF: Fewer views and faster training for free. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12872–12881, 2022. 7

[13] Dawa Derksen and Dario Izzo. Shadow neural radiance fields for multi-view satellite photogrammetry. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1152–1161, 2021. 1, 3, 4, 5

[14] Thibaud Ehret, Roger Marí, and Gabriele Facciolo. Nerf, meet differential geometry! *arXiv preprint arXiv:2206.14938*, 2022. 3

[15] Gabriele Facciolo, Carlo de Franchis, and Enric Meinhardt. MGM: A significantly more global matching for stereovision. In *Proceedings of the British Machine Vision Conference (BMVC)*, number 90, pages 1–12, 2015. 3, 7

[16] Gabriele Facciolo, Carlo de Franchis, and Enric Meinhardt-Llopis. Automatic 3D reconstruction from multi-date satellite images. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1542–1551, 2017. 3, 7

[17] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5491–5500, 2022. 3

[18] Yasutaka Furukawa and Carlos Hernández. Multi-view stereo: A tutorial. *Foundations and Trends in Computer Graphics and Vision*, 9(1-2):1–148, 2015. 1

[19] Alvaro Gómez, Gregory Randall, Gabriele Facciolo, and Rafael von Gioi Grompone. An experimental comparison of multi-view stereo approaches on satellite images. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 707–716, 2022. 3

[20] Charles Hessel, R Grompone von Gioi, Jean-Michel Morel, Gabriele Facciolo, Pablo Arias, and Carlo de Franchis. Relative radiometric normalization using several automatically chosen reference images for multi-sensor, multi-temporal series. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 5(2):845–852, 2020. 4

[21] Heiko Hirschmuller. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):328–341, 2007. 3

[22] Ajay Jain, Matthew Tancik, and Pieter Abbeel. Putting NeRF on a diet: Semantically consistent few-shot view synthesis. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5865–5874, 2021. 3

[23] Thomas Krauß, Pablo d'Angelo, Mathias Schneider, and Veronika Gstaiger. The fully automatic optical processing system CATENA at DLR. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40-1/W1:177–181, 2013. 1

[24] Zhengfei Kuang, Kyle Olszewski, Menglei Chai, Zeng Huang, Panos Achlioptas, and Sergey Tulyakov. NeROIC: Neural rendering of objects from online image collections. *ACM Transactions on Graphics (TOG)*, 41(4):1–12, 2022. 3

[25] Lorenzo Lastilla, Roberta Ravanelli, Francesca Fratarcangeli, Martina Di Rita, Andrea Nascetti, and Mattia Crespi. FOSS4G DATE for DSM generation: Sensitivity analysis of the semi-global block matching parameters. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42-2/W13, 2019. 3

[26] Bertrand Le Saux, Naoto Yokoya, Ronny Hansch, Myron Brown, and Greg Hager. 2019 data fusion contest [technical committees]. *IEEE Geoscience and Remote Sensing Magazine*, 7(1):103–105, 2019. 6

[27] Matthew J Leotta, Chengjiang Long, Bastien Jacquet, Matthieu Zins, Dan Lipsa, Jie Shan, Bo Xu, Zhixin Li, Xu Zhang, Shih-Fu Chang, et al. Urban semantic 3D reconstruction from multiview satellite imagery. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1451–1460, 2019. 3

[28] Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. BARF: Bundle-adjusting neural radiance fields. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5721–5731, 2021. 5

[29] Roger Marí, Carlo de Franchis, Enric Meinhardt-Llopis, Jérémy Anger, and Gabriele Facciolo. A generic bundle adjustment methodology for indirect RPC model refinement of satellite imagery. *Image Processing On Line*, 11:344–373, 2021. 5, 6

[30] 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 multiview stereo. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 2188–2196, 2019. 5, 6

[31] Roger Marí, Thibaud Ehret, and Gabriele Facciolo. Disparity estimation networks for aerial and high-resolution satellite images: A review. *Image Processing On Line*, 12:501–526, 2022. 3

[32] Roger Marí, Gabriele Facciolo, and Thibaud Ehret. Sat-NeRF: Learning multi-view satellite photogrammetry with transient objects and shadow modeling using rpc cameras. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1310–1320, 2022. 1, 3, 4, 5, 6

[33] 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 *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7206–7215, 2021. 3, 5

[34] Julien Michel, Emmanuelle Sarrazin, David Youssefi, Myriam Cournet, Fabrice Buffe, Jean-Marc Delvit, Aurélie Emilien, Julien Bosman, Olivier Melet, and Céline L'Helguen. A new satellite imagery stereo pipeline designed for scalability, robustness and performance. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 5-2-2020:171–178, 2020. 1, 3

[35] 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. In *Computer Vision – ECCV 2020*, pages 405–421, 2020. 1, 2, 3, 5

[36] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4), 2022. 3

[37] Michael Niemeyer, Jonathan T Barron, Ben Mildenhall, Mehdi SM Sajjadi, Andreas Geiger, and Noha Radwan. RegNeRF: Regularizing neural radiance fields for view synthesis from sparse inputs. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5470–5480, 2022. 3

[38] Ozge C Ozcanli, Yi Dong, Joseph L Mundy, Helen Webb, Riad Hammoud, and Tom Victor. Automatic geo-location correction of satellite imagery. In *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 307–314, 2014. 6

[39] Mathias Rothermel, Konrad Wenzel, Dieter Fritsch, and Norbert Haala. SURE: Photogrammetric surface reconstruction from imagery. In *Proceedings LC3D Workshop, Berlin*, volume 8, 2012. 3

[40] Ewelina Rupnik, Mehdi Daakir, and Marc Pierrot-Deseilligny. MicMac–a free, open-source solution for photogrammetry. *Open Geospatial Data, Software and Standards*, 2(14), 2017. 1

[41] 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. 4

[42] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5449–5459, 2022. 3

[43] Jurgen Wohlfeil, Heiko Hirschmuller, Björn Piltz, Anko Börner, and Michael Suppa. Fully automated generation of accurate digital surface models with sub-meter resolution from satellite imagery. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 39-B3:75–80, 2012. 3

[44] Teng Wu, Bruno Vallet, Marc Pierrot-Deseilligny, and Ewelina Rupnik. A new stereo dense matching benchmark dataset for deep learning. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43-B2-2021:405–412, 2021. 3, 7

[45] Gengshan Yang, Joshua Manela, Michael Happold, and Deva Ramanan. Hierarchical deep stereo matching on high-resolution images. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5510–5519, 2019. 3

[46] Wenqi Yang, Guanying Chen, Chaofeng Chen, Zhenfang Chen, and Kwan-Yee K Wong. S^3-NeRF: Neural reflectance field from shading and shadow under a single viewpoint. *arXiv preprint arXiv:2210.08936*, 2022. 4

[47] Yao Yao, Zixin Luo, Shiwei Li, Tian Fang, and Long Quan. MVSNet: Depth inference for unstructured multi-view stereo. In *Computer Vision – ECCV 2018*, pages 785–801, 2018. 3

[48] Lin Yen-Chen, Pete Florence, Jonathan T Barron, Alberto Rodriguez, Phillip Isola, and Tsung-Yi Lin. iNeRF: Inverting neural radiance fields for pose estimation. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1323–1330, 2021. 5

[49] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelNeRF: Neural radiance fields from one or few images. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4576–4585, 2021. 3

[50] Ding Yuan and Christopher D Elvidge. Comparison of relative radiometric normalization techniques. *ISPRS Journal of Photogrammetry and Remote Sensing*, 51(3):117–126, 1996. 4

[51] Ramin Zabih and John Woodfill. Non-parametric local transforms for computing visual correspondence. In *Computer Vision — ECCV '94*, pages 151–158, 1994. 3

[52] Feihu Zhang, Victor Prisacariu, Ruigang Yang, and Philip H.S. Torr. GA-Net: Guided aggregation net for end-to-end stereo matching. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 185–194, 2019. 3

[53] Jason Zhang, Gengshan Yang, Shubham Tulsiani, and Deva Ramanan. NeRS: Neural reflectance surfaces for sparse-view 3D reconstruction in the wild. *Advances in Neural Information Processing Systems*, 34:29835–29847, 2021. 3

[54] Qiang Zhao, Le Yu, Zhenrong Du, Dailiang Peng, Pengyu Hao, Yongguang Zhang, and Peng Gong. An overview of the applications of Earth observation satellite data: impacts and future trends. *Remote Sensing*, 14(8):1863, 2022. 1