POLITECNICO DI TORINO
Repository ISTITUZIONALE

Gaussian Splatting for Efficient Satellite Image Photogrammetry

Original
Gaussian Splatting for Efficient Satellite Image Photogrammetry / Aira, L.S., Facciolo, G., Ehret, T.. - (2025), pp. 5959-5969. (2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Nashville (USA) 10-17 June 2025) [10.1109/cvpr52734.2025.00559].

Availability:
This version is available at: 11583/3005598 since: 2025-12-02T16:19:47Z

Publisher:
IEEE

Published
DOI:10.1109/cvpr52734.2025.00559

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2025 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

26 November 2026

EOGS: Gaussian Splatting for Earth Observation

Luca Savant Aira[1] Gabriele Facciolo[2] Thibaud Ehret[3]

https://mezzelfo.github.io/EOGS/

[1] Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129 Torino TO, Italia

[2] Universite Paris-Saclay, CNRS, ENS Paris-Saclay, Centre Borelli, 91190, Gif-sur-Yvette, France

[3] AMIAD, Pôle Recherche, France

Abstract

Recently, Gaussian splatting has emerged as a strong alternative to NeRF, demonstrating impressive 3D modeling capabilities while requiring only a fraction of the training and rendering time. In this paper, we show how the standard Gaussian splatting framework can be adapted for remote sensing, retaining its high efficiency. This enables us to achieve state-of-the-art performance in just a few minutes, compared to the day-long optimization required by the best-performing NeRF-based Earth observation methods. The proposed framework incorporates remote-sensing improvements from EO-NeRF, such as radiometric correction and shadow modeling, while introducing novel components, including sparsity, view consistency, and opacity regularizations.

1. Introduction

Since the mid-20th century, the number of active satellites in orbit has increased exponentially. Today, hundreds of satellites are dedicated to Earth observation, and this number is expected to continue growing [9]. These satellites frequently acquire optical images of the same areas at regular intervals. As a result, the availability of these datasets will continue to grow in the near future, requiring efficient algorithms to handle this expansion.

One common task for these image datasets is photogrammetry [13, 19, 27], which aims to recover the 3D geometry (*e.g.* Digital Surface Model - DSM), and the appearance (*e.g.* an albedo map) of the Earth's surface for this area using only the available 2D satellite images. Specifically, we focus on performing digital surface modeling from remote sensing images.

Historically, binocular stereovision and tri-stereo methods have been used for this purpose. However, these methods rely on image acquisitions being nearly simultaneous and with specific relative positions, which is often impractical, with limited acquisition opportunities, and/or costly.

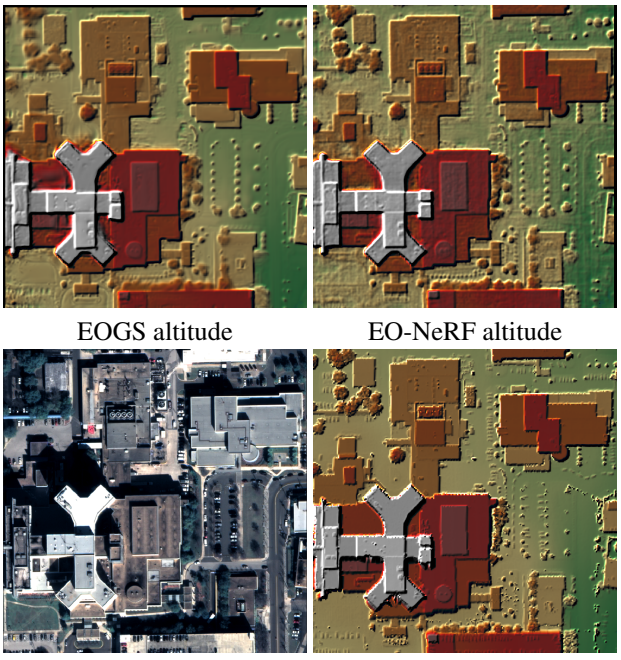


Figure 1. Using a limited number of satellite images of a given scene, the proposed EOGS method estimates the appearance and geometry of the scene. It achieves the same level of detail as EO-NeRF [32], such as the group of fans or the thin structures on top of the tall building on the left. However, EOGS requires only a few minutes of optimization, compared to the day-long training time required by EO-NeRF [32].

Therefore, our goal is to develop a method capable of handling multi-date images captured from arbitrary satellite positions.

More recently, multi-view stereo methods tailored for novel-view synthesis (NVS) have been used to solve this task as they naturally handle diverse camera positions. Indeed, as the NVS task also requires understanding the geometry of the scene, these methods can be used to recover a digital elevation model. These methods are usually based

on the concept of radiance field, *i.e*. a representation of the scene modeling the radiance emitted by each point, of which NeRF [36] is the main representative. Among those, EO-NeRF [32] has established itself as a state-of-the-art approach for digital surface modeling thanks to an improved shadow model. However, because it is NeRF-based, it also inherits NeRF's computational slowness. In recent years, 3D Gaussian Splatting (3DGS) [23] has been proposed as an alternative to NeRF. This method offers much faster training and rendering while reaching comparable reconstruction accuracy.

In this work, we introduce EOGS, the Earth-observation Gaussian splatting, the first method for digital elevation modeling based on 3DGS. EOGS achieves accuracy comparable to previous state-of-the-art approaches while being approximately 300× faster. Keys to the success of EOGS are the following contributions, all of which are compatible with the original 3DGS framework's efficiency:

- Approximating locally the pushbroom satellite sensors as affine cameras.
- Introducing a shadow-mapping-based pipeline for rendering the shadows in a physically accurate manner.
- Adding three new regularization terms that promote sparsity in the Gaussians opacities, view consistency, and completely opaque objects. This reduces the training time and improves the quality of the results.

2. Related Work

2.1. Stereovision for Earth Observation

Stereovision is at the heart of many tools for 3D estimation from series of satellite images. Examples of such pipelines are Ames stereo pipeline [4], MicMac [43], CARS [35], S2P [11], or CATENA [25]. Traditionally, these multi-view stereo methods are applied to well-chosen (either manually or automatically) image pairs. Since they process each pair independently (for example in the dense stereo matching step), a crucial step is the fusion of all the generated pairwise 3D models into a single one.

The recent trend has been to replace classic dense matching methods, such as semi-global matching (SGM) [20] or more global matching (MGM) [14], with deep learning based methods such as PSM [8], HSM [53] or GA-Net [54]. A review of these methods and a comparative study for satellite images is performed in [30].

2.2. NeRF for Earth Observation

Recently, Mildenhall *et al*. [36] have shown that it is possible to learn a volumetric model of a scene, called neural radiance fields (NeRF), using differentiable inverse rendering. Given a sparse set of views of the scene, NeRF learns in a self-supervised manner by maximizing the photoconsistency across the predicted renderings corresponding to the available viewpoints. After convergence, the volumetric model can then be used to render realistic novel views of the scene. In practice, this volumetric model is represented by an MLP that predicts, for each position $\mathbf{p}$ of the space, the local density of the scene $\sigma(\mathbf{p})$ as well as its appearance (*i.e*. color) $c(\mathbf{p})$. The rendering is performed using an approximation of the volumetric integral I from optical physics estimated using ray casting,

$$I(\mathbf{o},\mathbf{d}) = \int_{t_n}^{t_f} e^{-\int_{t_n}^{t}\sigma(\mathbf{r}(s))ds} c\left(\mathbf{r}(t),\mathbf{d}\right)\sigma\left(\mathbf{r}(t)\right)dt \quad (1)$$

with $\mathbf{o}$ the camera center from which the ray r originates and $\mathbf{d}$ unit direction.

NeRF-based methods were then extended to the remote sensing case, and in particular to perform multi-view and multi-date satellite photogrammetry, namely S-NeRF [12], Sat-NeRF [31], and EO-NeRF [32]. S-NeRF [12] exploits the solar direction, information typically available in the metadata of each observation or that can be easily retrieved knowing the location of the scene as well as the acquisition hour and date, to predict the direct sun light reaching each point in the scene. This is done by adding the solar direction as an input to the MLP and predicting the amount of sun light reaching a point as a new output. In this way, the shadows cast by buildings can be learned by the MLP and generated accordingly during the novel-view rendering step. Sat-NeRF [31] extends S-NeRF by modeling the transient parts of the scenes (*e.g*. cars, construction sites, or foliage) as done in NeRF-in-the-wild [33] and improves the camera representation of S-NeRF, from pinhole to RPC [1, 46]. EO-NeRF [32] improves the handling of shadows of S-NeRF and Sat-NeRF by defining physically plausible shadows directly from the geometry. These shadows are then rendered by additional raycasting from the surface in the direction of the sun. More recent work focuses on modeling difficult seasonal effects [17], extending the proposed volumetric models to surface models [40], using the raw pre-pansharpened data provided directly by the satellite operators [38], and accelerating the training step by taking advantage of faster NeRF versions [5].

2.3. 3D Gaussian Splatting

Following the growing interest in NVS, 3DGS [23] was proposed as an alternative to NeRF-based methods. While both NeRF and 3DGS use alpha compositing as their image formation model, 3DGS represents the scene using a set of discrete Gaussian-shaped primitives placed in the 3D space, as opposed to a continuous black-box MLP representation of NeRF. In the following, we use the notation of Bulò *et al*. [7].

A *Gaussian primitive* is a tuple $\gamma = (\boldsymbol{\mu}, \Sigma, \alpha, \mathbf{f})$ representing a single Gaussian-shaped volume element in the scene, where $\boldsymbol{\mu} \in \mathbb{R}^3$ is the primitive center, $\Sigma \in \mathcal{M}^{3\times3}(\mathbb{R})$

its 3D shape and orientation, $\alpha \in [0,1]$ its opacity, and $\mathbf{f} \in \mathbb{R}^d$ its feature vector (*e.g.* color when $d=3$).

3DGS uses K independent Gaussian primitives to represent the scene, so we will index them using $k = 1, \dots, K$ as subscript. Each Gaussian primitive k has an associated 3D Gaussian kernel $\mathcal{G}_k$ defined as

$$\mathcal{G}_k(\mathbf{x}) = \exp\left\{-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}_k\right)^T \Sigma_k^{-1}\left(\mathbf{x}-\boldsymbol{\mu}_k\right)\right\}. \quad (2)$$

To render a view (characterized by its associated camera), 3DGS "splats" each 3D Gaussian kernel onto the camera image plane. This process is called the *splatting* operation and, mathematically, it associates a Gaussian primitive γ_k and a camera model/projection $\mathcal{A}: \mathbb{R}^3 \to \mathbb{R}^2$ to a 2D Gaussian kernel $\mathcal{G}_k^{\mathcal{A}} : \mathbb{R}^2 \to \mathbb{R}$. The original 3DGS method deals only with pinhole camera models. During the splatting operation, each Gaussian is projected according to the first-order approximation of the perspective projection computed at $\boldsymbol{\mu}_k$, $J^{\mathcal{A}}(\boldsymbol{\mu}_k)$. In this way, the mean vector and covariance matrix of the 2D Gaussian kernel are:

$$\boldsymbol{\mu}_k^{\mathcal{A}} = \mathcal{A}(\boldsymbol{\mu}_k) \qquad \Sigma_k^{\mathcal{A}} = J^{\mathcal{A}}(\boldsymbol{\mu}_k)\,\Sigma\left(J^{\mathcal{A}}(\boldsymbol{\mu}_k)\right)'. \quad (3)$$

Once all the primitives are splatted, they are sorted front-to-back with respect to the camera reference. Then, they are aggregated using the traditional alpha compositing, accounting also for the Gaussian kernel decay such that

$$I^{\mathcal{A}}(\mathbf{u}) = \sum_{k=1}^{K} \tilde{\mathbf{f}}_k \omega_k^{\mathcal{A}}(\mathbf{u}), \quad (4)$$

where $I^{\mathcal{A}}$ is the rendered image associated with camera $\mathcal{A}$, $\mathbf{u} \in \mathbb{R}^2$ is a point in the 2D image plane, $\tilde{\mathbf{f}}_k$ are features that depend on $\mathbf{f}_k$ and the view direction (for modeling view-direction dependent color effect, such as shiny objects), and the alpha-compositing coefficient is

$$\omega_k^{\mathcal{A}}(\mathbf{u}) = \alpha_k \mathcal{G}_k^{\mathcal{A}}(\mathbf{u}) \prod_{j=1}^{k-1}\left(1-\alpha_j \mathcal{G}_j^{\mathcal{A}}(\mathbf{u})\right). \quad (5)$$

Recent literature expanded the original 3DGS [23] in many directions. As 3DGS has focused only on pinhole camera models and relied on a first-order local approximation of the model, recent works propose different camera models (*e.g.* fish-eye [29]) or more complex splatting strategies [21]. Another line of work focused on improving 3DGS in challenging input regimes, such as sparse-view input [52], and input without camera parameters [16]. Moreover, many recent methods focus on controllable texture and lighting [18, 22, 28, 44]. Further references can be found in recent surveys [2, 51].

3. Method

The proposed Earth-observation Gaussian splatting method, referred to as EOGS, specializes and adapts 3DGS for the satellite photogrammetry task. Given N non-orthorectified satellite images and their corresponding RPC camera model coefficients, a set of Gaussian-shaped 3D primitives is optimized to recover both the 3D geometry and appearance of the scene.

The general learning problem is to find the set of K Gaussian primitives that best approximates the N satellite images, with the rendering process of Eq. (4). This can be formulated as:

$$\underset{(\gamma_k)_{[\![1,K]\!]}}{\arg\min} \sum_{i=1}^{N} \ell(\hat{I}_i, I_i), \quad (6)$$

where I_i is the i-th input satellite observation, $\hat{I}_i$ is the corresponding synthesized view (in the original 3DGS $\hat{I}_i = I^{\mathcal{A}_i}$), and ℓ is the same photometric distance function used in 3DGS.

In the following sections we highlight the differences between EOGS and previous 3DGS and NeRF-based approaches.

3.1. Projections and Coordinate Systems

We define the coordinate system in which the Gaussian primitive centers and shapes are expressed as *world-space coordinates*. This coordinate system is a uniformly rescaled and recentered version of the *Universal Transverse Mercator (UTM) coordinate system* [49], such that the center of the scenes coincides with the origin, the scene is contained in a unit cube, and it is east-north-up aligned similarly to EO-NeRF [32]. At the other end of the transformation pipeline lies the *2D NDC-space*, where the Gaussian primitives are splatted.

The correct mapping between these two spaces is a composition of transformations: world-space to UTM to longitude-latitude-altitude. Using the RPC coefficients (that model the satellite position and 3D attitude) associated with each observation, the latter coordinate system is mapped onto the image row-column coordinates. Finally, the coordinates are normalized to range in $[-1,1]$ to get to NDC-space. As shown in Fig. 2, we instead compute a per-scene affine approximation of the whole transformation, introducing a negligible mean error of ≈ 0.012 pixels while being more computationally efficient than previous works and compatible with a Gaussian Splatting formulation. Specifically, for an affine camera model $\mathcal{A}: \mathbf{x} \in \mathbb{R}^3 \to A\mathbf{x}+\mathbf{a} \in \mathbb{R}^2$, the Eq. (3) simplifies to $\boldsymbol{\mu}_k^{\mathcal{A}} = A\boldsymbol{\mu}_k + \mathbf{a} \in \mathbb{R}^2$ and $\Sigma_k^{\mathcal{A}} = A\Sigma A' \in \mathbb{R}^{2\times 2}$. This simplification eliminates the need for the local first-order approximation used in the original 3DGS method, as we moved the approximation to the camera models.

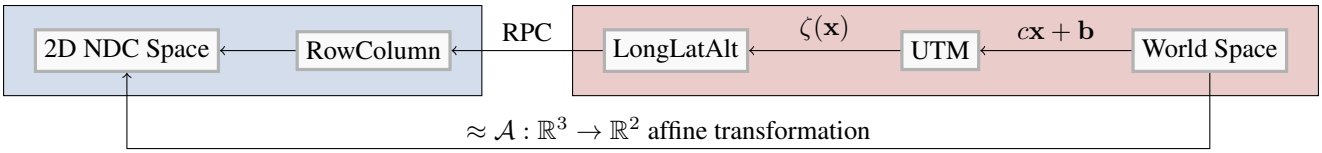


Figure 2. Summary of the transformation from world-space to NDC-space and its affine approximation. The affine approximation is computationally efficient, compatible with the Gaussian splatting formulation, and well-suited for satellite images. The coordinate systems in the right red box represent 3D world coordinates (camera-independent), while the left blue box shows 2D coordinates (camera-dependent).

3.2. Shadow Mapping

As in EO-NeRF, we want to explicitly model the shadow phenomena in the images, as the solar direction is available for each image in the scene. Unlike previous literature, our method uses a custom variant of shadow mapping to cast geometrically consistent shadows. Introduced in [50], Shadow Mapping is a well-known technique in the field of 3D graphics for adding shadows to a computer graphic rendering. It is particularly suited for EOGS, as it requires just the ability to render the scene from different points of view, as opposed to the shadow casting technique used in EO-NeRF that requires ray marching (which is not defined in Gaussian splatting).

Before introducing our variant of Shadow Mapping, we define the *elevation render*, *localization* function, and the *homologous point* function.

Given a camera model/projection $\mathcal{A}$, the elevation render is defined as the 3DGS rendering Eq. (4) using the real elevations instead of colors

$$E^{\mathcal{A}}(\mathbf{u}) = \sum_{k=1}^{K} [\mathcal{E}\mu_k]\, \omega_k^{\mathcal{A}}(\mathbf{u}), \tag{7}$$

where $\mathcal{E} : \mathbb{R}^3 \to \mathbb{R}$ is an affine operation mapping 3D points expressed in the "native" world coordinates to the corresponding real altitude, expressed in meters. We remark that this is not the depth nor the inverse depth, typically found in the MVS literature.

Given a camera model/projection $\mathcal{A}$, the localization function that maps a pixel of the camera and a given absolute altitude to its associated point in the native 3D world is defined as

$$\begin{aligned} &\mathrm{loc}^{\mathcal{A}} : \left(\mathbb{R}^2 \times \mathbb{R}\right) \to \mathbb{R}^3, \quad (\mathbf{u}, h) \mapsto \mathbf{x} \\ &\text{s.t.} \quad \mathcal{A}(\mathbf{x}) = \mathbf{u} \quad \text{and} \quad \mathcal{E}(\mathbf{x}) = h. \end{aligned} \tag{8}$$

Given two cameras $\mathcal{A}$ and $\mathcal{B}$, the homologous point function maps a pixel of the first camera to the corresponding pixel of the second camera, taking into consideration the 3D geometry:

$$\begin{aligned} &\mathrm{hom}^{\mathcal{A},\mathcal{B}} : \mathbb{R}^2 \to \mathbb{R}^2, \quad \mathbf{u} \mapsto \tilde{\mathbf{u}} \\ &\text{s.t.} \quad \tilde{\mathbf{u}} = \mathcal{B}\left(\mathrm{loc}^{\mathcal{A}}\left(\mathbf{u}, E^{\mathcal{A}}(\mathbf{u})\right)\right). \end{aligned} \tag{9}$$

In our shadow mapping approach (depicted in Fig. 3), we assume that the sun is the only light source present in the scene. Moreover, since it is far from the scene, it can be approximated as a directional light. Following the classic shadow mapping approach, we construct a camera $\mathcal{S}$, called *sun camera*, placed at and aligned with the light source (detailed in the supplementary). As the camera model corresponding to a directional light is the affine camera, we can handle uniformly the sun cameras and the satellite cameras.

Then, we consider a second camera, $\mathcal{A}$, from which we want to synthesize a novel view and apply shadows according to the sun direction. Given a point $\mathbf{u}$ in the $\mathcal{A}$ NDC-space, and its corresponding altitude $E^{\mathcal{A}}(\mathbf{u})$, we first localize it, obtaining a 3D point in world-space. We then project this point according to $\mathcal{S}$, obtaining the homologous point of $\mathbf{u}$ in $\mathcal{S}$. We then resample the elevation rendering of $\mathcal{S}$ at this projected point and compare it with $E^{\mathcal{A}}(\mathbf{u})$. Mathematically, this corresponds to

$$\Delta h^{\mathcal{A},\mathcal{S}}(\mathbf{u}) = E^{\mathcal{S}}\left(\mathrm{hom}^{\mathcal{A},\mathcal{S}}(\mathbf{u})\right) - E^{\mathcal{A}}(\mathbf{u}). \tag{10}$$

If these two elevations, $E^{\mathcal{A}}(\mathbf{u})$ and $E^{\mathcal{S}}(\tilde{\mathbf{u}})$ are the same, it means that both the camera and the sun camera are imaging the same 3D point, hence this point is in light. If the two elevations are not the same, then the sun camera is not able to "see" the 3D point, hence it is in shadows. To represent this shading, the color of points in the shadows is multiplied by a darkening coefficient computed from $\Delta h^{\mathcal{A},\mathcal{S}}(\mathbf{u})$ as

$$s^{\mathcal{A},\mathcal{S}}(\mathbf{u}) = \min\left\{\exp\left\{-\rho \Delta h^{\mathcal{A},\mathcal{S}}(\mathbf{u})\right\}, 1\right\}. \tag{11}$$

We argue that this formulation is physically plausible as this would be the correct equation for a homogeneous medium of density ρ, as shown in [34].

Following [32], we also model a per-camera ambient light $\psi^{\mathcal{A}}$ so that in-shadow objects do not appear completely black. The shading to be applied to a given pixel $\mathbf{u}$ is given by the following *lighting coefficient*

$$l^{\mathcal{A},\mathcal{S}}(\mathbf{u}) = s^{\mathcal{A},\mathcal{S}}(\mathbf{u}) + (1 - s^{\mathcal{A},\mathcal{S}}(\mathbf{u}))\psi^{\mathcal{A}}. \tag{12}$$

Finally, EOGS image formation equation is:

$$I^{\mathcal{A},\mathcal{S}}(\mathbf{u}) = l^{\mathcal{A},\mathcal{S}}(\mathbf{u}) \sum_{k=1}^{K} \phi^{\mathcal{A}}(\mathbf{f}_k)\omega_k^{\mathcal{A}}(\mathbf{u}), \tag{13}$$

where $\phi^{\mathcal{A}}(\cdot)$ is a camera-specific affine color correction applied to the intrinsic primitive colors $\mathbf{f}_k$. We remark that, differently from 3DGS, we drop the view-direction dependencies of the primitive colors and introduce a camera-dependent color correction.

It is useful to define the *albedo rendering*, where we do not use the shadows or the camera-specific color correction:

$$I^{\mathcal{A}}(\mathbf{u}) = \sum_{k=1}^{K} \mathbf{f}_k \omega_k^{\mathcal{A}}(\mathbf{u}). \tag{14}$$

While the image formation model defined in Eq. (13) is equivalent to EO-NeRF, the shadow definition is quite different. In EO-NeRF case, shadows are defined as the sun visibility for all points on the surface. Because of possible occlusions, two points of the scene can correspond to the same point seen from the sun direction. Therefore, it is not possible to define the sun visibility as an "image" that could be estimated using a Gaussian splatting-like process. Trying to compute an irregularly sampled "image" corresponding to these points would break the locality assumption used in Gaussian splatting during the rasterization step and thus reduce the computational efficiency. On the contrary, the proposed shadow mapping verifies all the assumptions made by Gaussian splatting.

3.3. Regularizers

It is well-known that deep neural networks are implicitly regularized [39, 41, 45, 48], meaning that despite being used in the overparametrized regime, they show generalization capabilities.

On the other hand, we found out that primitives in 3DGS-based methods are almost independently optimized one from the other. This is probably due to the fact that the primitives in 3DGS are initialized as small spheres, spread out in the entire scene. This results in 3DGS being less regularized than NeRF-based methods and "lacking" constraints during the optimization phase.

Hence we are free to add additional regularization constraints to the general optimization problem Eq. (6) that induce smoother and more regular solutions. In particular, we introduce constraints that promote our solution to be sparse (*i.e.* we encourage solutions that require fewer Gaussian primitives), view consistent, and mostly composed of completely opaque objects.

As common ML pipelines are specialized for unconstrained optimization problems, we argue to use a Lagrangian relaxation approach and re-formulate each constraint as a new loss term, each with its own experimentally-found Lagrangian multiplier.

Promoting Sparsity. Training time is directly proportional to the number of Gaussian primitives considered during the optimization process. As we want to recover the geometry of the scene as fast as possible, we want as few Gaussian primitives as possible, hence a sparse solution.

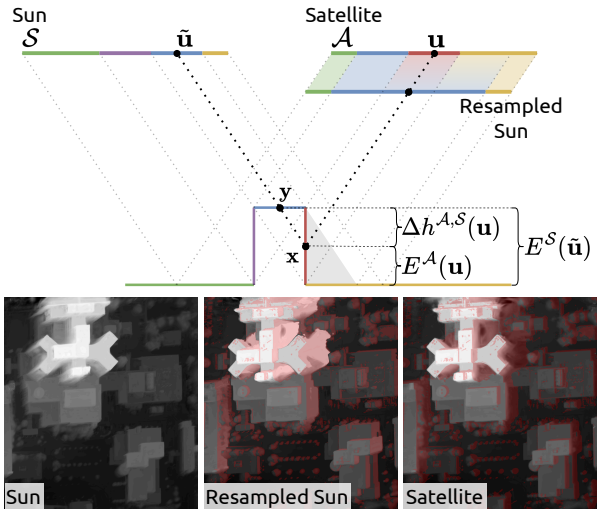


Figure 3. Shadow mapping illustration. The point $\mathbf{u}$ in the satellite image (affine camera $\mathcal{A}$) corresponds to a the the 3D point $\mathbf{x} = \text{loc}^{\mathcal{A}}(\mathbf{u})$ on the vertical wall. Projecting $\mathbf{x}$ to the sun camera (affine camera $\mathcal{S}$), $\tilde{\mathbf{u}} = \mathcal{S}\mathbf{x}$ is obtained. Then $\mathbf{y} = \text{loc}^{\mathcal{S}}(\tilde{\mathbf{u}})$ is obtained localizing $\tilde{\mathbf{u}}$. The point $\mathbf{x}$ and its pixel $\mathbf{u}$ are in shadow because the elevation of $\mathbf{y}$ is greater than the elevation of $\mathbf{x}$. Indeed, all and only the points where the satellite elevation and the resampled sun elevation do not match should be shaded. On the bottom of the illustration are shown examples of the sun elevation, the resampled sun elevation, and the satellite elevation renderings, with shadows highlighted in red.

Inspired by the well-known LASSO regularization in linear regression [47] that promotes a sparse solution, and by recent works such as 3DGSMCMC [24], we consider a L^1 regularization of the opacities

$$\mathcal{L}_o = \frac{1}{K} \sum_{k=1}^{K} \alpha_k. \tag{15}$$

This regularization promotes sparsity in the primitive opacities distribution, hence only "useful" primitives will be visible at the end of the optimization. We pair this regularization with a simple thresholding pruning procedure that discards any primitive with $\alpha < \alpha_{\min}$. In this way, unused primitives are actually discarded, yielding faster splatting and overall faster training (specifically, we recorded speedups of up to $2\times$ on the considered datasets).

We remark that many works [7, 24] have proposed replacements to the original 3DGS densification/pruning procedure. Here, instead, we aim only at lowering the number of primitives, so we do not need a densification strategy as long as we instantiate enough of them at the beginning of the optimization. Moreover, we set $\alpha_{\min} = 0.0025$ as

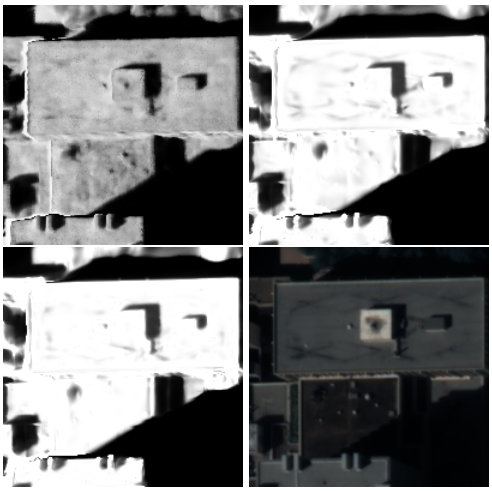


Figure 4. From top-left to bottom-right, shadow maps of EO-NeRF, EOGS without the $\mathcal{L}_s$ penalizer, EOGS with the $\mathcal{L}_s$ penalizer, and the corresponding satellite image. Textures corresponding to the image content can be observed in the shadow map of EO-NeRF and EOGS without the $\mathcal{L}_s$ penalizer, but not in EOGS.

primitives with lower opacities are already discarded in the original 3DGS implementation of the front-to-back splatting procedure.

Promoting View Consistency. Differently from the classical NVS context, in remote sensing the available views are low-count and sparse, resulting in Eq. (6) being even less constrained. Paired with the fact that 3DGS does not benefit from the implicit regularization of NeRF, we argue that an additional constraint promoting view consistency is needed.

We propose a "local view consistency" loss based on the intuition that if the same 3D point is visible from two cameras and the cameras are close to each other, then the color and elevation resampled at the corresponding pixels should be the same. Otherwise, the object is occluded or outside the camera boundaries.

Mathematically, given a camera $\mathcal{A}$ we randomly perturb it and obtain a camera $\mathcal{B}$. Assuming that there is no view-direction dependent color effect, this constraint reads:

$$\Delta h^{\mathcal{A},\mathcal{B}}(\mathbf{u}) < \Delta h_{\min} \Rightarrow \begin{cases} I^{\mathcal{A}}(\mathbf{u}) = I^{\mathcal{B}}\left(\hom^{\mathcal{A},\mathcal{B}}(\mathbf{u})\right) \\ E^{\mathcal{A}}(\mathbf{u}) = E^{\mathcal{B}}\left(\hom^{\mathcal{A},\mathcal{B}}(\mathbf{u})\right), \end{cases} \tag{16}$$

where we reused the same notation of the shadow mapping explanation.

This constraint results in two loss terms, the color (albedo) consistency and the altitude consistency:

$$\mathcal{L}_{cc} = \sum_{\mathbf{u}} M^{\mathcal{A},\mathcal{B}}(\mathbf{u}) \left| I^{\mathcal{A}}(\mathbf{u}) - I^{\mathcal{B}}\left(\hom^{\mathcal{A},\mathcal{B}}(\mathbf{u})\right)\right| \tag{17}$$

$$\mathcal{L}_{ac} = \sum_{\mathbf{u}} M^{\mathcal{A},\mathcal{B}}(\mathbf{u}) \left| E^{\mathcal{A}}(\mathbf{u}) - E^{\mathcal{B}}\left(\hom^{\mathcal{A},\mathcal{B}}(\mathbf{u})\right)\right|, \tag{18}$$

where $M^{\mathcal{A},\mathcal{B}}(\cdot)$ is a binary mask that selects all pixels $\mathbf{u}$ such that $\Delta h^{\mathcal{A},\mathcal{B}}(\mathbf{u}) < \Delta h_{\min}$ and $\hom^{\mathcal{A},\mathcal{B}}(\mathbf{u})$ is inside the image boundaries. We remark that we always choose $\mathcal{A}$ from the input posed images and we set $\Delta h_{\min} = 30$cm. Moreover, we obtain $\mathcal{B}$ by independently sampling $q_1, q_2 \in \mathbb{R}$ from a ± 1-truncated standard distribution and defining

$$\mathcal{B}(\mathbf{x}) = \mathcal{A}(\mathbf{x}) + 0.05\, \mathcal{E}(\mathbf{x}) \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}. \tag{19}$$

In the generic NVS literature, many works [10, 37] proposed different methods for increasing the view consistency. RegNeRF [37] is the first work that deals with sparse camera poses by introducing a loss term that maximizes the likelihood of rendered RGB patches from virtual cameras with a pre-trained deep normalizing-flow model, while also adding a total variation regularization on the rendered depth. Furthermore, [10] introduces a reprojection mechanism such that only the geometry needs to be learned from the NeRF, as the colors are resampled from the input images. Note that EOGS differs from [37] as we ask for consistency (RGB and depth) between two cameras (one real and one virtual), hence we do not need any pre-trained model for the RGB renders nor prior on the elevation renders. EOGS also differs from [10] as we learn the colors and do not resample input images that may contain transients or color shifts.

Promoting Opaqueness. Looking at the output from Sat-NeRF and EO-NeRF (see Fig. 4), we can see that much of the texture of the scene is embedded in the geometric shadows. This geometry misuse is caused by semi-transparent objects casting semi-transparent shadows. In order to lessen this effect, we propose to add an entropy-based penalty $\mathcal{L}_s$ for incorrect use of the shadows. This penalty is defined as

$$\mathcal{L}_s = \sum_{\mathbf{u}} H\left(s^{\mathcal{A},\mathcal{S}}(\mathbf{u})\right), \tag{20}$$

where $H(x) = -\left(x \log_2(x) + (1-x)\log_2(1-x)\right)$. Indeed, the shadow map $s^{\mathcal{A},\mathcal{S}}$ should contain only 0 or 1 values. This is the case for $\rho \to +\infty$ in Eq. (11), as a building should not cast a semi-transparent shadow. Hence we add this entropy-based penalizer to discourage the use of semi-transparent shadows, which in turn encourage objects to be either completely transparent or fully opaque. Note that choosing a large ρ during training is not an option since it would make the training unstable as Eq. (11) would be close to a non-differentiable step function.

We acknowledge that the problem of promoting hard surfaces in NeRF has been studied in previous works, such as [3, 42].

3.4. Implementation Details

The implementation of EOGS is based on the original 3DGS code base. Other than the aforementioned novel

		JAX					IARPA				
		004	068	214	260	Mean ↓	001	002	003	Mean ↓	Time ↓
	Number of views	8	16	20	14	-	24	20	21	-	-
No mask	EO-NeRF [32]	1.37	1.05	1.61	1.37	1.35	1.43	1.79	1.31	1.51	15 hours
	SAT-NGP [5]	1.63	1.27	2.18	1.79	1.72	1.54	2.11	1.69	1.78	25 minutes
	Sat-Mesh [40]	1.55	1.15	2.02	1.36	1.52	N.A.	N.A.	N.A.	N.A.	8 minutes
	S2P [15]	1.45	1.19	1.82	1.66	1.53	1.48	2.48	1.38	1.78	20 minutes
	EOGS (ours)	1.45	1.10	1.73	1.55	1.46	1.58	2.00	1.27	1.62	3 minutes
Foliage mask	EO-NeRF [32]	1.02	1.03	1.55	1.24	1.21	1.32	1.63	1.18	1.38	15 hours
	SAT-NGP [5]	1.03	1.26	2.17	1.43	1.47	1.34	1.85	1.62	1.60	25 minutes
	EOGS (ours)	0.89	1.01	1.63	1.24	1.19	1.38	1.70	1.03	1.37	3 minutes

Table 1. Mean absolute error on the elevation [meters] and the corresponding training time for various baseline methods, when considering the whole AOI (no mask) or when ignoring foliage areas (foliage mask). Results for Sat-Mesh are reported from the paper since the authors did not share their code.

contributions, the main differences lie in disabling the per-Gaussian view-direction color dependency and initializing all the Gaussians with white color and as low as possible opacity (1%). Moreover, we reduce the number of iterations to 5000 and enable the shadow mapping and all three regularizations at iteration 1000. Furthermore, the Gaussians centers are initialized uniformly in the 3D scene such that the initial density is 0.13 Gaussians per m^3.

We use the same optimizer and scheduler of 3DGS for the primitives and use a second Adam scheduler with 10^{-2} learning rate for learning the camera-dependent parameters: the affine color-correction $\boldsymbol{\phi}^{\mathcal{A}}$ and the ambient color $\boldsymbol{\psi}^{\mathcal{A}}$.

The Lagrangian coefficients of the regularization constraints have been found experimentally on a single scene, rounded to the nearest power of ten, and applied to all scenes. This highlights the robustness of EOGS to the specific values of these coefficients. The final loss is:

$$\min \sum_{i=1}^{N} \ell(\hat{I}_i, I_i) + 0.1\mathcal{L}_o + 0.1\mathcal{L}_{cc} + 0.01\mathcal{L}_{ac} + 0.01\mathcal{L}_s, \tag{21}$$

where $\hat{I}_i$ is now, differently from 3DGS in Eq. (6), a shorthand notation for $I^{\mathcal{A}_i,\mathcal{S}_i}$ from Eq. (13), which also depends on the sun camera $\mathcal{S}_i$.

4. Experiments

We evaluate EOGS in the same experimental setting as the most recent related work in the literature, EO-NeRF.

We are using datasets provided in the 2019 IEEE GRSS Data Fusion Contest (DFC2019) [6, 26] and 2016 IARPA Multi-View Stereo 3D Mapping Challenge (IARPA2016). These datasets, comprising a total of 7 areas of interest (AOI), contain cropped non-orthorectified multidate WorldView-3 observations, along with metadata such as the 3D satellite attitude (encoded in the RPC coefficient) and the local sun direction. We use the bundled-adjusted version of the RPC coefficient used in EO-NeRF. Each image covers approximately 256×256 meters squared of terrain with a resolution of $30 \sim 50$ cm per pixel, while each AOI is imagined by $10 \sim 20$ crops.

4.1. Main Experiment Results

Tab. 1 show the main experimental results of EOGS. To assess the accuracy of EOGS we report the mean absolute error (MAE) between a lidar scan included in the dataset and the elevation render aligned to this nadir view. We argue that the volume of these data will grow in the near future, so we are also interested in the time required to recover the geometry from the input images. Hence, we also report the training time.

If the entire AOIs are considered, as reported in Tab. 1 (top), EOGS performs slightly worse than the state of the art EO-NeRF but it is approximately $300\times$ faster. For reference, we also report all available results of other methods from the literature (EO-NeRF [32], SAT-NGP [5], Sat-Mesh [40], and S2P [11, 15]). We see that EOGS is Pareto-optimal with respect to elevation MAE and training time. If instead we use available ground truth semantic maps to ignore prediction in the foliage areas, EOGS performance is equivalent to EO-NeRF, showing higher accuracy for structural objects, as reported in Tab. 1 (bottom). We present visual results in Fig. 5, more results, as well as details on the number of Gaussians per scene and memory usage, are provided in the supplement.

4.2. Ablation and Parameter Studies

Impact of the Different Losses. Tab. 2 reports an ablation study of the loss terms in EOGS. Each column corresponds to a different ablation experiment, while each row corresponds to a different component of EOGS being ablated. The first row indicates whether the Shadow Mapping technique is enabled or not. The following three rows indi-

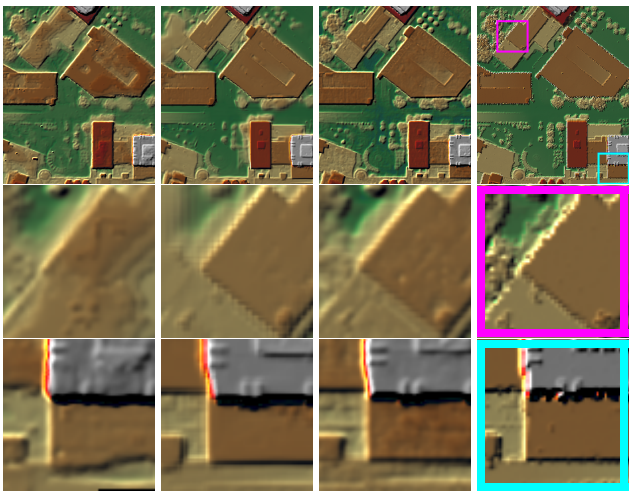


Figure 5. From left to right: visual results on JAX_214 comparing SAT-NGP [5], EOGS, EO-NeRF [32] and the ground truth.

Shadowmap	✗	✓	✓	✓	✓	✓	✓	✓	✓
Sparsity	✗	✗	✓	✗	✗	✗	✓	✓	✓
Consistency	✗	✗	✗	✓	✗	✓	✗	✓	✓
Opaqueness	✗	✗	✗	✗	✓	✓	✓	✗	✓
MAE [m] ↓	5.03	1.86	1.83	1.69	1.79	1.57	1.76	1.64	1.54
Train [min] ↓	4.18	-	-	-	-	4.27	-	-	2.85

Table 2. Ablation study of each proposed component of EOGS.

cate, respectively, the presence of the sparsity, consistency, and opaqueness regularizers. We remark that the first column is equivalent to 3DGS with affine cameras, learnable per-camera affine color correction, and different primitives initialization. For each column, we report the grand mean elevation MAE of JAX and IARPA scenes.

To quantitatively measure the impact of each single component, we linearly regress the MAE from the presence of the components, obtaining a coefficient for each component that expresses an elevation MAE gain with respect to the base case (reported in the first column of Tab. 2). The introduction of shadow mapping is the most impactful component, gaining 3.16 meters of accuracy. Then, the consistency regularizer and the opaqueness regularizer further improve the accuracy of EOGS by 0.20 and 0.09 meters, respectively. Lastly, the sparsity regularizer, while being necessary for achieving efficient training as shown in the last row of Tab. 2, also reports an improvement of 0.04 meters. Hence, all components independently contribute to the quality of the geometry reconstruction.

Regularization Parameters. Fig. 6 shows the results of a grid search on the coefficients of $\mathcal{L}_{cc}$ and $\mathcal{L}_{ac}$ in Eq. (21). It shows that both altitude regularization and color regularization are necessary to achieve the best performance. We remark that, in order to reduce overfitting to a particular scene, we choose the same "round coefficients" for all scenes.

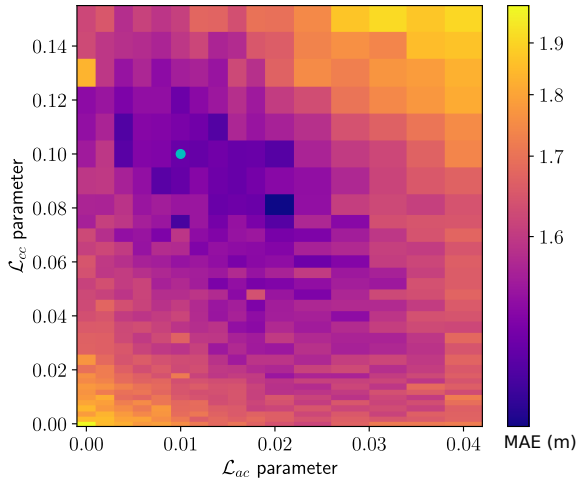


Figure 6. View consistency regularization parameter ablation study. Selected parameter set is shown with the cyan dot. Estimation performed on the JAX_260 sequence.

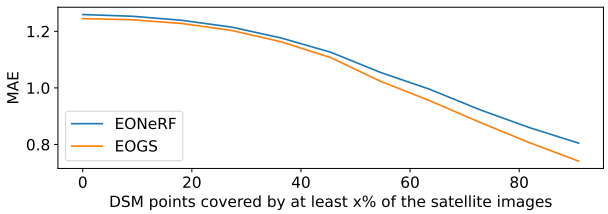


Figure 7. Impact of the visibility on the performance (using foliage mask).

Impact of visibility. Fig. 7 shows the impact of the visibility (*i.e.* the number of cameras that can see a given point of the scene) on the performance. While EOGS and EO-NeRF are comparable on average, this test shows that EOGS performs better for regions that are visible in most images but struggles in the regions observed in only a few images.

5. Conclusion

This study presents EOGS, the first Gaussian-Spatting-based framework for earth observation. By introducing remote sensing requirements in 3DGS, such as shadow modeling and camera-specific corrections, EOGS achieves comparable accuracy to the existing state-of-the-art method, EO-NeRF, while being over 300 times faster. Hence, EOGS will be a practical solution for near-future large-scale datasets.

Our analysis shows that EOGS excels in high-coverage regions, where it produces fine details at a fraction of the computational cost. Addressing regions with low image coverage or with foliage, by adding additional regularization or employing better initialization schemes, could further improve EOGS accuracy.

Acknowledgments

This publication is part of the project PNRR-NGEU which has received funding from the MUR - DM 352/2022. This work was partially supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART"). It was performed using HPC resources from GENCI-IDRIS (grant 2024-AD011012453R3). Centre Borelli is also with Université Paris Cité, SSA and INSERM.

References

[1] Emmanuel P Baltsavias and Dirk Stallmann. Metric information extraction from spot images and the role of polynomial mapping functions. In *XVII ISPRS Congress, Commission IV*. Swiss Federal Institute of Technology, Institute of Geodesy and Photogrammetry, 1992. 2

[2] Yanqi Bao, Tianyu Ding, Jing Huo, Yaoli Liu, Yuxin Li, Wenbin Li, Yang Gao, and Jiebo Luo. 3d gaussian splatting: Survey, technologies, challenges, and opportunities. *arXiv preprint arXiv:2407.17418*, 2024. 3

[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. 6

[4] 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. 2

[5] Camille Billouard, Dawa Derksen, Emmanuelle Sarrazin, and Bruno Vallet. Sat-ngp : Unleashing neural graphics primitives for fast relightable transient-free 3d reconstruction from satellite imagery. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 8749–8753, 2024. 2, 7, 8

[6] 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. 7

[7] Samuel Rota Bulò, Lorenzo Porzi, and Peter Kontschieder. Revising densification in gaussian splatting. *arXiv preprint arXiv:2404.06109*, 2024. 2, 5

[8] 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. 2

[9] Jeff Clauson, Simon Cantrell, Jim Vrabel, John Oeding, Bhaskar Ranjitkar, Tom Rusten, Shankar Ramaseri, and Kimberly Casey. Earth observing satellites online compendium. U.S. Geological Survey, 2024. 1

[10] François Darmon, Bénédicte Bascle, Jean-Clément Devaux, Pascal Monasse, and Mathieu Aubry. Improving neural implicit surfaces geometry with patch warping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6260–6269, 2022. 6

[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. 2, 7

[12] 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

[13] eoPortal. WorldView Legion, 2024. https://www.eoportal.org/satellite-missions/worldview-legion. 1

[14] 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. 2

[15] 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. 7

[16] Yang Fu, Sifei Liu, Amey Kulkarni, Jan Kautz, Alexei A. Efros, and Xiaolong Wang. Colmap-free 3d gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20796–20805, 2024. 3

[17] Michael Gableman and Avinash Kak. Incorporating season and solar specificity into renderings made by a nerf architecture using satellite images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 2

[18] Jian Gao, Chun Gu, Youtian Lin, Hao Zhu, Xun Cao, Li Zhang, and Yao Yao. Relightable 3d gaussian: Real-time point cloud relighting with brdf decomposition and ray tracing. *arXiv:2311.16043*, 2023. 3

[19] Richard I Hartley and Joseph L Mundy. Relationship between photogrammmetry and computer vision. *Integrating photogrammetric techniques with scene analysis and machine vision*, 1944:92–105, 1993. 1

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

[21] Letian Huang, Jiayang Bai, Jie Guo, Yuanqi Li, and Yanwen Guo. On the error analysis of 3d gaussian splatting and an optimal projection strategy. *arXiv preprint arXiv:2402.00752*, 2024. 3

[22] Yingwenqi Jiang, Jiadong Tu, Yuan Liu, Xifeng Gao, Xiaoxiao Long, Wenping Wang, and Yuexin Ma. Gaussianshader: 3d gaussian splatting with shading functions for reflective surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5322–5332, 2024. 3

[23] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42 (4), 2023. 2, 3

[24] Shakiba Kheradmand, Daniel Rebain, Gopal Sharma, Weiwei Sun, Jeff Tseng, Hossam Isack, Abhishek Kar, Andrea Tagliasacchi, and Kwang Moo Yi. 3d gaussian splatting as markov chain monte carlo. *arXiv preprint arXiv:2404.09591*, 2024. 5

[25] 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. 2

[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. 7

[27] Laurent Lebègue, Eric Cazala-Hourcade, Florie Languille, Stéphanie Artigues, and Olivier Melet. Co3d, a worldwide one one-meter accuracy dem for 2025. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B1-2020:299–304, 2020. 1

[28] Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, and Kui Jia. Gs-ir: 3d gaussian splatting for inverse rendering. *arXiv preprint arXiv:2311.16473*, 2023. 3

[29] Zimu Liao, Siyan Chen, Rong Fu, Yi Wang, Zhongling Su, Hao Luo, Linning Xu, Bo Dai, Hengjie Li, Zhilin Pei, et al. Fisheye-gs: Lightweight and extensible gaussian splatting module for fisheye cameras. *arXiv preprint arXiv:2409.04751*, 2024. 3

[30] 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. 2

[31] 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

[32] 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 2035–2045, 2023. 1, 2, 3, 4, 7, 8

[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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7210–7219, 2021. 2

[34] N. Max. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 1(2): 99–108, 1995. 4

[35] 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. 2

[36] 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

[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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5480–5490, 2022. 6

[38] Emilie Pic, Thibaud Ehret, Gabriele Facciolo, and Roger Marí. Pseudo pansharpening nerf for satellite image collections. In *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 2650–2655. IEEE, 2024. 2

[39] Tomaso Poggio, Qianli Liao, Brando Miranda, Andrzej Banburski, Xavier Boix, and Jack Hidary. Theory iiib: Generalization in deep networks. *arXiv preprint arXiv:1806.11379*, 2018. 5

[40] Yingjie Qu and Fei Deng. Sat-mesh: Learning neural implicit surfaces for multi-view satellite reconstruction. *Remote Sensing*, 15(17):4297, 2023. 2, 7

[41] Sameera Ramasinghe, Lachlan E MacDonald, and Simon Lucey. On the frequency-bias of coordinate-mlps. *Advances in Neural Information Processing Systems*, 35:796–809, 2022. 5

[42] Daniel Rebain, Mark Matthews, Kwang Moo Yi, Dmitry Lagun, and Andrea Tagliasacchi. Lolnerf: Learn from one look. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1558–1567, 2022. 6

[43] 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. 2

[44] Yahao Shi, Yanmin Wu, Chenming Wu, Xing Liu, Chen Zhao, Haocheng Feng, Jingtuo Liu, Liangjun Zhang, Jian Zhang, Bin Zhou, et al. Gir: 3d gaussian inverse rendering for relightable scene factorization. *arXiv preprint arXiv:2312.05133*, 2023. 3

[45] Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *NeurIPS*, 2020. 5

[46] C Vincent Tao and Yong Hu. A comprehensive study of the rational function model for photogrammetric processing. *Photogrammetric engineering and remote sensing*, 67(12): 1347–1358, 2001. 2

[47] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288, 1996. 5

[48] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018. 5

[49] U.S. Geological Survey. The universal transverse mercator (utm) grid. Technical report, U.S. Geological Survey, Reston, VA, 2001. Report. 3

[50] Lance Williams. Casting curved shadows on curved surfaces. In *Proceedings of the 5th annual conference on Computer graphics and interactive techniques*, pages 270–274, 1978. 4

[51] Tong Wu, Yu-Jie Yuan, Ling-Xiao Zhang, Jie Yang, Yan-Pei Cao, Ling-Qi Yan, and Lin Gao. Recent advances in 3d gaussian splatting. *Computational Visual Media*, 10(4):613–642, 2024. 3

[52] Haolin Xiong, Sairisheek Muttukuru, Rishi Upadhyay, Pradyumna Chari, and Achuta Kadambi. Sparsegs: Real-time 360 {\deg} sparse view synthesis using gaussian splatting. *arXiv preprint arXiv:2312.00206*, 2023. 3

[53] 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. 2

[54] 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. 2