

EOGS++: Earth Observation Gaussian Splatting with Internal Camera Refinement and Direct Panchromatic Rendering

Pierrick Bournez¹, Luca Savant Aira², Thibaud Ehret³, Gabriele Facciolo^{1,4}

¹ Université Paris-Saclay, CNRS, ENS Paris-Saclay, Centre Borelli, Gif-sur-Yvette, France

² Politecnico di Torino, Corso Duca degli Abruzzi, Torino TO, Italia

³ AMIAD, Pôle Recherche, France

⁴ Institut Universitaire de France

Keywords: Gaussian Splatting, Satellite Photogrammetry, Remote Sensing, Digital Surface Modeling.

Abstract

Recently, 3D Gaussian Splatting has been introduced as a compelling alternative to NeRF for Earth observation, offering competitive reconstruction quality with significantly reduced training times. In this work, we extend the Earth Observation Gaussian Splatting (EOGS) framework to propose EOGS++, a novel method tailored for satellite imagery that directly operates on raw high-resolution panchromatic data without requiring external preprocessing. Furthermore, leveraging optical flow techniques we embed bundle adjustment directly within the training process, avoiding reliance on external optimization tools while improving camera pose estimation. We also introduce several improvements to the original implementation, including early stopping and TSDF post-processing, all contributing to sharper reconstructions and better geometric accuracy. Experiments on the IARPA 2016 and DFC2019 datasets demonstrate that EOGS++ achieves state-of-the-art performance in terms of reconstruction quality outperforming the original EOGS method and other NeRF-based methods while maintaining the computational advantages of Gaussian Splatting. Our model demonstrates an improvement from 1.33 to 1.19 mean MAE errors on buildings compared to the original EOGS models. The code is publicly available at <https://gardiens.github.io/EOGS2>.

1. Introduction

Since the mid-20th century, the number of satellites orbiting the Earth has grown steadily, leading to an unprecedented availability of remote sensing data. Today, high-resolution panchromatic and multispectral imagery is regularly acquired over the same areas under varying conditions, and the volume of such data is expected to continue increasing in the coming years. As a result, developing scalable methods to exploit these datasets efficiently has become a key challenge for the remote sensing community. One of the main applications of such data is 3D reconstruction through photogrammetry, which aims to recover both the geometry (e.g. the digital surface models - DSM) and the appearance of the Earth's surface from 2D images. Traditional stereovision pipelines rely on temporally consistent acquisitions and precise calibrations, which are difficult to guarantee in practice. More recent approaches, such as implicit neural representations, have shown the ability to handle diverse imaging conditions while producing accurate reconstructions. In particular, Neural Radiance Fields (NeRF) (Mildenhall et al., 2021), and its Earth-observation extensions (Zhang and Rupnik, 2023, Marí et al., 2023) have demonstrated high-quality results, but remain computationally demanding.

To address these limitations, 3D Gaussian Splatting (Kerbl et al., 2023), often abbreviated as 3DGS, has recently emerged as an alternative representation offering both competitive accuracy and significant computational efficiency. The EOGS (Savant Aira et al., 2025) framework successfully adapted 3DGS to satellite imagery by introducing shadow modeling, affine camera approximations, and per-camera color corrections into the pipeline. However, similar to most NeRF-based methods, it still relied on external preprocessing steps, such as pansharpening and camera poses bundle adjustment. Moreover, the original implementation was restricted to RGB data, although satellite

imagery typically includes high-resolution panchromatic data alongside lower-resolution multispectral data, limiting its applicability to real-world scenarios.

The main contribution of this work is the introduction of EOGS++, a new framework for Earth observation that extends the applicability of EOGS while eliminating its dependencies. Figure 1 illustrates qualitative results produced by our pipeline. More specifically, our contributions can be summarized as follows:

- We introduce an internal bundle adjustment based on optical flow to correct camera localization errors.
- We show that raw panchromatic images are sufficient to achieve a good 3D reconstruction, thus eliminating the need to pansharpen the images beforehand.
- We improve the EOGS framework by introducing an early stopping mechanism (Chung et al., 2024) and a truncated signed distance function-based postprocessing operation (Curless and Levoy, 1996).

2. Related Works

2.1 3D reconstruction from satellite images

Traditionally, stereoscopic 3D reconstruction from satellite imagery has relied on conventional processing pipelines, such as S2P (De Franchis et al., 2014a) or Catena (Krauß et al., 2013). More recently, there has been a growing trend towards integrating deep learning models (Chang and Chen, 2018) to directly regress the digital surface models. While promising, these methods face limitations in terms of computational efficiency and require image acquisitions to be temporally close.

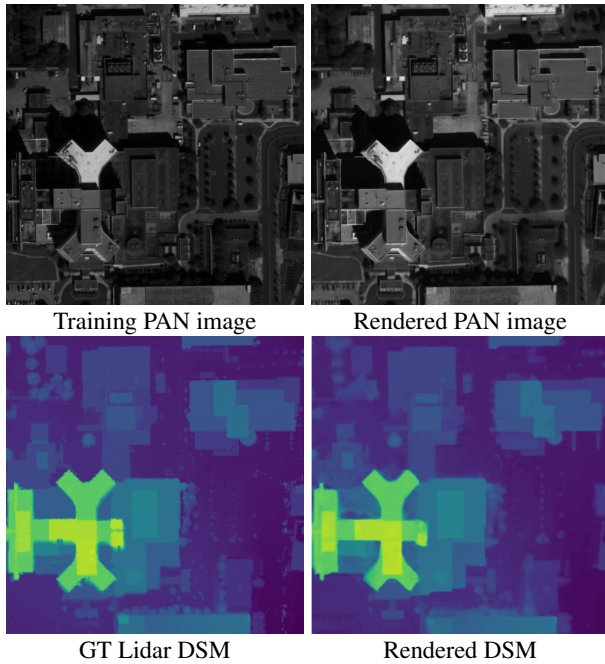


Figure 1. Qualitative result of the EOGS++ pipeline.

2.2 Implicit representations for 3D reconstruction

A different line of research focuses on implicit neural representations, such as NeRF (Mildenhall et al., 2021) and Gaussian Splatting (Kerbl et al., 2023), which reconstruct 3D scenes from calibrated image collections. Unlike traditional pipelines, these approaches optimize a continuous scene representation and have demonstrated improved performance across diverse imaging conditions (Yan et al., 2024).

Adapting implicit methods to satellite imagery, however, presents additional challenges due to sparse viewpoints, limited variation in viewing angles across acquisitions, and varying illumination. Several works have addressed these limitations by extending NeRF-like models to pushbroom cameras (Derksen and Izzo, 2021, Marí et al., 2022), explicitly modeling shadows (Marí et al., 2023, Behari et al., 2024), or filtering transient elements to improve reconstruction robustness (Huang et al., 2025).

While NeRF-based techniques have been actively explored, literature on Gaussian Splatting in the satellite domain remains limited (Savant Aira et al., 2025, Huang et al., 2025). This approach achieves performance comparable to NeRF-like methods while requiring shorter training times.

2.3 Pansharpening

Pansharpening aims to fuse high-resolution panchromatic images with low-resolution multispectral images to produce high-resolution multispectral outputs.

Classical approaches can be grouped into three main categories: Component Substitution (CS), Multi-Resolution Analysis (MRA), and variational methods. CS methods rely on spectral transformations such as IHS (Carper et al., 1990) or PCA (Ghadjati et al., 2019), while MRA techniques use spatial filters (e.g., Laplacian pyramids (Li et al., 1995)) to extract high-frequency details from the panchromatic image and inject them into the multispectral image. Variational approaches (Palsson et al., 2013,

Meng et al., 2020) formulate the fusion as an energy minimization problem to ensure spectral and spatial consistency.

More recently, deep learning-based methods such as PNN (Citola et al., 2022) have gained attention. However, these models often face generalization issues across different sensors or geographic regions.

In the context of inverse rendering, most pipelines (Marí et al., 2022) perform pansharpening as a preprocessing step prior to 3D reconstruction. Alternative strategies, however, directly leverage multispectral data, either by predicting spectral channels independently (Sprintson et al., 2025a) or by explicitly modeling inter-band dependencies (Pic et al., 2024).

2.4 Bundle adjustment

Accurate 3D reconstruction requires reliable camera poses. While satellite providers supply approximate orientations, these are often affected by geolocation errors (Grodecki and Dial, 2003), which can degrade reconstruction quality (Marí et al., 2023). Bundle Adjustment (BA) (Triggs et al., 1999) is commonly used to refine these estimates by minimizing reprojection errors (Marí et al., 2021).

Classical BA pipelines rely on interest-point matching (e.g., SIFT (Lowe, 1999)), but are computationally expensive and sensitive to outliers (Szeliski, 2022). More recent approaches embed pose refinement within learning-based frameworks, jointly optimizing poses and geometry, for instance in NeRF (Lin et al., 2021, Marí et al., 2023). In the context of Gaussian Splatting, BA has been explored mainly in SLAM (Yan et al., 2024) and for correcting motion blur (Deng et al., 2025), with recent work proposing BA directly on Gaussian parameters (Zhang et al., 2024).

3. Gaussian splatting for earth observation

Our approach is based on Earth Observation Gaussian Splatting (EOGS) (Savant Aira et al., 2025), which is an extension of 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) to satellite images. In this section, we present in detail the main ideas of these two methods. The core idea behind 3DGS is that, given a set of N images and their corresponding camera models, a set of K Gaussian 3D primitives is optimized to recover the appearance of the scene. In practice, this collection of primitives $\{\gamma_k\}_{k=1}^K$ is parametrized by their respective position $\boldsymbol{\mu}_k \in \mathbb{R}^3$, covariance $\Sigma_k \in \mathcal{M}^{3 \times 3}(\mathbb{R})$, color features $\mathbf{c}_k \in \mathbb{R}^3$ and opacity $\alpha_k \in [0, 1]$. Each primitive has an associated Gaussian density function defined as

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

The link with the observations and the set of primitives is given by the rendering pipeline. All primitives are first projected onto the image plane thanks to the camera model. The projected Gaussians \mathcal{G}_k^C are then sorted front-to-back (based on their center position) and aggregated using an alpha-compositing scheme:

$$I^C(\mathbf{u}) = \sum_{k=1}^K \mathbf{c}_k \omega_k^C(\mathbf{u}), \quad (2)$$

where \mathbf{u} is the coordinate of a pixel on the image plane I^C associated to the camera C and ω_k^C is the accumulated opacity of

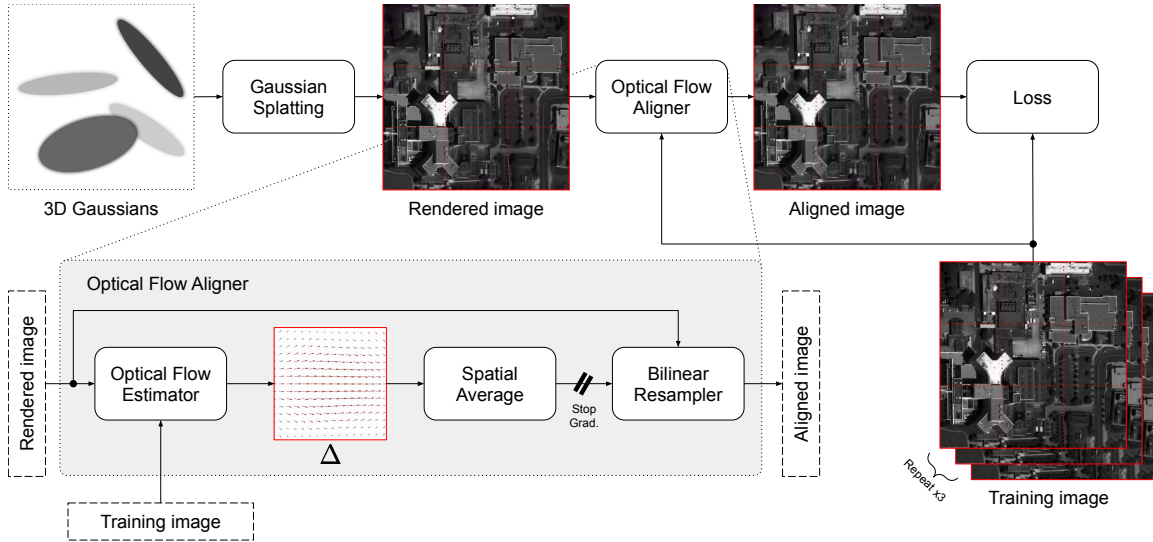


Figure 2. Schematic overview of the proposed training pipeline. From the 3D Gaussian primitives, an image is rendered. The rendered image is then aligned with the training observation using an optical flow algorithm, after which the model is trained accordingly.

the k -th Gaussian at pixel \mathbf{u} , defined as

$$\omega_k^C(\mathbf{u}) = \alpha_k \mathcal{G}_k^C(\mathbf{u}) \prod_{j=1}^{k-1} (1 - \alpha_j \mathcal{G}_j^C(\mathbf{u})). \quad (3)$$

The learning problem is therefore to find the set of K Gaussian primitives that best approximates the N images, with the rendering process of Eq. (2). This can be formulated as:

$$\arg \min_{\{\gamma_k\}_{k=1}^K} \sum_{n=1}^N \ell_1(I^{C_n} - I^n) + \lambda \ell_{D-SSIM}(I^{C_n}, I^n), \quad (4)$$

where λ is the weight term, I^n is the n -th input observation, corresponding to camera C_n , I^{C_n} is the corresponding synthesized view, ℓ_1 is the pixelwise L_1 reconstruction loss, and ℓ_{D-SSIM} is the differentiable SSIM-based loss introduced in the original 3D Gaussian Splatting framework (Kerbl et al., 2023).

EOGS extends the original 3DGS framework to the satellite imaging setting. We summarize here the key differences with the original 3DGS framework that will be useful for the rest of paper. First of all, the camera model used is different. Indeed, 3DGS is designed to work with pinhole camera model, which is the standard model for most consumer cameras. However this camera model is not appropriate in the case of satellite images where the camera model is usually represented with a Rational Polynomial Coefficient (RPC) camera model (De Franchis et al., 2014b). Because EOGS argues that the projection step is computationally inefficient when applied to RPCs, it instead employs an approximate affine camera model \mathcal{A} . Hence, each camera is described as an affine function $\mathcal{A} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$, mapping 3D points \mathbf{x} to 2D image coordinates $\mathbf{u} = \mathcal{A}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{a}$ with $\mathbf{A} \in \mathbb{R}^{2 \times 3}$ and $\mathbf{a} \in \mathbb{R}^2$. In that case, The projection operation assigns to each Gaussian a 2D Gaussian density function, with mean $\boldsymbol{\mu}_k^A = \mathcal{A}(\boldsymbol{\mu}_k) \in \mathbb{R}^2$ and covariance $\Sigma_k^A = \mathbf{A}\Sigma_k\mathbf{A}^T \in \mathcal{M}^{2 \times 2}(\mathbb{R})$.

Similarly to mesh-based variants of 3DGS, e.g. (Guédon et al., 2025), EOGS aims to recover the 3D geometry of the scene. To this end, it defines an elevation rendering that uses the true

elevation rather than RGB values:

$$E^A(\mathbf{u}) = \sum_{k=1}^K \mathcal{E}(\mu_k) \omega_k^A(\mathbf{u}), \quad (5)$$

where $\mathcal{E} : \mathbb{R}^3 \rightarrow \mathbb{R}$ is an affine mapping that returns the real altitude, expressed in meters. In the original implementation, this elevation rendering is used to evaluate the quality of the reconstructed 3D geometry.

EOGS also improves the alpha-compositing scheme that was described in (2). It introduces the concept of shadow mapping inside the rendering pipeline. Modeled as a binary mask l^A , it represents the visible regions where lighting coming from the sun is occluded by the scene. It also introduces a per-camera color-correction, as satellite image colors can change due to difference in atmospheric conditions at the acquisition dates. To model this effect, the authors use a learnable $d \times 3$ per-camera affine transformation, ϕ^A , that maps Gaussian attributes to color channels appropriate for the given acquisition. This leads to the following alpha-compositing scheme

$$I^A(\mathbf{u}) = l^A(\mathbf{u}) \sum_{k=1}^K \phi^A(\mathbf{f}_k) \omega_k^A(\mathbf{u}). \quad (6)$$

Finally, the authors also add a regularization term to encourage view consistency. Given a small perturbation of a camera model, view consistency is encouraged by penalizing differences between the perturbed view and the original view (after alignment and masking of occluded regions). Two terms are proposed, one for the elevation (ℓ_{ec}) and one for the color (ℓ_{cc}). Additional terms to promote sparsity and opaqueness (referred as R in the following) are also included (Savant Aira et al., 2025). The complete EOGS learning problem can then be formulated as

$$\arg \min_{\{\gamma_k\}_{k=1}^K, \{\phi^{A_n}\}_{n=1}^N} \sum_{n=1}^N \ell_1(I^{A_n} - I^n) + \lambda_1 \ell_{D-SSIM}(I^{A_n}, I^n) + \lambda_2 \ell_{ec}(I^{A_n}) + \lambda_3 \ell_{cc}(I^{A_n}) + \lambda_4 R(\{\gamma_k\}_{k=1}^K, \mathbf{A}_n, l^{A_n}), \quad (7)$$

with λ_s the different weighting terms for each corresponding loss. In the following, we define the *photometric loss* ℓ as the combination of the ℓ_1 loss and the ℓ_{D-SIMM} loss.

4. Method

The proposed EOGS++ extends and improves EOGS during preprocessing, training and postprocessing phases. The required preprocessing steps in EOGS, such as bundle adjustment and pansharpening, are integrated in EOGS++, as explained respectively in Sections 4.1 and 4.2. This enables EOGS++ to have virtually zero required preprocessing steps. The training phase of EOGS is also improved with the introduction of opacity reset and early stopping, as explained in Section 4.3. Finally, a postprocessing step is proposed to improve the final DSM prediction, as explained in Section 4.4. The training pipeline is illustrated in Figure 2.

4.1 Bundle adjustment

Bundle adjustment (BA) in satellite imagery is commonly used to refine camera models estimates by minimizing reprojection errors (Grodecki and Dial, 2003), and it is an essential preprocessing step in EOGS to ensure good reconstruction results, as demonstrated in Table 1. EOGS++ integrates bundle adjustment directly within the training process, eliminating the need for external BA software. This is achieved by leveraging optical flow to align rendered and reference images.

First, the image I^A is rendered, using Eq. (2). Then an off-the-shelf optical flow estimation model computes a displacement field $\Delta : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, between I^A and the corresponding training image I , such that the pixel $\mathbf{u} + \Delta(\mathbf{u})$ in I^A corresponds to the pixel \mathbf{u} in I .

Since camera miscalibration primarily manifests as pixel offsets in the context of satellite imagery (Marí et al., 2021), a constant averaged version of Δ is considered, namely $\bar{\Delta} \in \mathbb{R}^2$. $\bar{\Delta}$ is used to bilinearly resample the rendered image: $\hat{I}^A \approx I^A(\mathbf{u} + \bar{\Delta})$. This is equivalent to consider the shifted version of the rendered image that better matches the training image. To keep the optical-flow estimator isolated from the image optimizer, we block its gradients by applying a stop-gradient to $\bar{\Delta}$ so that no back-propagation flows through it. Finally, the role of I^{A_n} is replaced by \hat{I}^{A_n} in Eq. (7).

4.2 Handling panchromatic images

While EOGS handles only RGB images, satellite imagery typically includes high-resolution panchromatic (PAN) images alongside lower-resolution multispectral (MSI) images. Hence a preprocessing step, named pansharpening, is required to transform the MSI images into high-resolution RGB images before inputting them into the model. This step, however, can introduce artifacts and inconsistencies, especially when the PAN and MSI images are not perfectly aligned or captured under different conditions. EOGS++ discards the low-resolution MSI images and reconstructs the scene only from the original high-resolution PAN data. The single-channel panchromatic training images are repeated into a three-channel RGB representation, and the rendered outputs I^A are compared directly to the replicated training images. This strategy is referred to as 3-PAN. We acknowledge that this strategy disregards the spectral information from the MSI images, which could be valuable for certain applications. However, as we will show experimentally in Section 5.4, the inclusion of MSI data does not yield a significant gain for surface model estimation.

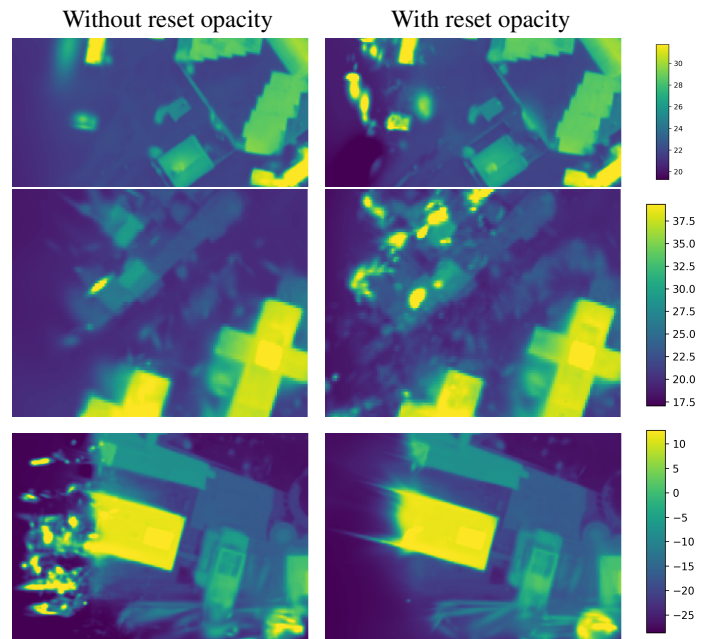


Figure 3. Zoomed-in DSM comparison for respectively the IARPA 001, IARPA 002, and JAX 214 with and without opacity reset. The reset operation helps eliminate Gaussian floaters.

4.3 Opacity reset and early stopping

We observed that while EOGS converges rapidly, the final reconstructions sometimes lack sharpness and fine details as shown in Figure 3. To address this, we reintegrated a technique from the original 3DGS paper, called *opacity reset*, which involves cyclically resetting the opacity of all Gaussians every 3'000 iterations, i.e., $\alpha_k \leftarrow \min(\alpha_k, 0.05)$. In order to fully benefit from this mechanism, we also increase the maximum number of iterations from 5'000 to 40'000. However, we noticed that by enabling the reset opacity mechanism during the optimization process, the photometric loss ℓ of Eq. (7) exhibits a strong non-monotonic behavior: it first decreases, but after a number of iterations starts to increase. We speculate that this is caused by the regularizers in Eq. (7) which take hold of the training dynamics. To mitigate this issue, we implemented an early stopping mechanism based on photometric loss ℓ itself. Training is therefore halted as soon as the photometric loss ℓ starts to increase, as a minimum in the photometric loss indicates a high-fidelity reconstruction. Thus, this approach enhances sharpness while minimizing the loss of fine and useful details. In practice, this criterion leads to convergence after approximately 10'000 to 20'000 iterations

4.4 Imposing explicit depth consistency across views

In the original EOGS implementation, a virtual nadir-view camera was used to compute the final elevation maps for MAE evaluation. However, since this camera is not part of the training process, it can introduce artifacts, particularly when training views deviate significantly from the nadir configuration.

To address this, we propose an optional post-processing step based on the truncated signed distance function (TSDF) fusion technique, akin to many current methods (Huang et al., 2024). Our method aggregates depth maps from all training views to create a more accurate and consistent digital surface model. The process involves rendering depth maps from each training view using the final optimized Gaussian primitives, then fusing these

depth maps into TSDF, with a fixed truncation margin and with weights proportional to the angle between the training view direction and an estimated surface normal, as described in (Curless and Levoy, 1996). Finally, we extract a fixed-resolution voxelization of the TSDF using marching cubes (Lorenson and Cline, 1998), resulting in a mesh. We enhance this mesh by removing isolated voxels (stemming from artifacts known as floaters in the Gaussian Splatting literature) and bottom-up filling holes. The final DSM is then obtained by sampling the height of this refined mesh at each ground position.

5. Experiments

5.1 Datasets

We evaluate EOGS++ under experimental conditions similar to those used in the original EOGS article. The datasets employed are extracted from the 2016 IARPA Multi-View Stereo 3D Mapping Challenge (Le Saux et al., 2019), denoted as IARPA2016, and the 2019 IEEE GRSS DATA Fusion Contest (Bosch et al., 2019), denoted as DFC2019. Overall, these datasets consist of 7 areas of interest and include multi-date non-orthorectified PAN and MSI WorldView-3 satellite observations. Each area covers a terrain of $256 \times 256 \text{ m}^2$, with a spatial resolution ranging from 30 to 50 cm per pixel for the PAN images and four times coarser for the MSI images. Each site is observed from approximately 10 to 30 different points of view.

5.2 Implementation details

The implementation builds upon the publicly available EOGS code. EOGS++ uses RAFT_sma11 (Teed and Deng, 2020) as the optical flow estimator for camera calibration, that represents a good tradeoff between speed and accuracy.

Similarly to EOGS, we report the Mean Absolute Error (MAE) between the LiDAR scan provided in the dataset and the predicted elevation map obtained either as a nadir depth rendering or through a TSDF post-processing step, as described in Section 4.4. Because foliage conditions vary across seasons and over time, and the reconstruction primarily targets buildings, the MAE is computed using a foliage mask, following the practice introduced by EOGS.

5.3 Internal bundle adjustment

Table 1 shows multiple ways of handling errors in the camera pointing, by comparing raw RPCs (*none*), internally learned RPC corrections (*learn wv*), the proposed optical-flow-based method (*optical flow*) and externally bundle adjusted RPCs (*external B.A.*). During the early development phase of EOGS++, we implemented a straightforward bundle adjustment algorithm, referred to as *learn wv*, which involves applying backpropagation to the affine projection matrix. In this strategy, both the Gaussian positions and the projection matrices of each camera are optimized during training. This enables the model to implicitly recalibrate the camera models. Similarly to the proposed optical flow approach, we restrict the optimization to the constant terms \mathbf{a} of the affine transformation, in order to learn just a per-camera constant shift in the image plane. The original implementation did not support gradient descent with respect to projection matrices. We extended the framework to allow gradient-based updates for affine camera models.

B.A. algorithm	JAX ↓	IARPA ↓	Mean ↓
None	1.33	2.54	1.93
Learn wv	1.21	2.43	1.82
Optical flow	1.23	1.50	1.36
External B.A. (Reference)	1.19	1.46	1.33

Table 1. Raw RPC handling comparison. Reported metric is the mean absolute error of the predicted elevation map [meters], when ignoring foliage areas. To ensure a fair comparison, all methods use the same pre-computed pansharpened images and are trained for 5'000 iterations

While the *learn wv* method is able to correctly align and render images (see Figure 5), its MAE performance does not show significant improvement compared to the raw RPC initialization. We hypothesize that the joint optimization of Gaussian parameters and affine matrices leads to convergence toward bad local minima. Conversely, the optical flow algorithm effectively registers the images, achieving comparable results (within 3 cm) when compared to bundle-adjusted cameras (see the DSM in Figure 4). Eventually the best performance is obtained with externally bundle-adjusted cameras, which we attribute to the improved stability of explicit triangulation in the bundle adjustment procedure compared with RGB-based optimization.

5.4 The need for multispectral images

Table 2 summarizes the results obtained with different strategies for handling panchromatic and MSI data, comparing the *Brovvey* pansharpening method, the *linear combination* method, the *single channel* and the 3-PAN method defined earlier. The *Brovvey* method corresponds to the classical *Brovvey* pansharpening technique, which is used in EOGS as a preprocessing step. It fuses the high-resolution PAN image with the low-resolution MSI image to produce a high-resolution MSI image, which is then used as the sole input for training EOGS++.

The second approach, referred to as *linear combination*, involves first rendering a high-resolution MSI image $I_{\text{MSI}}^{A_n}$, from which the corresponding PAN image I^{A_n} is derived through an "MSI-to-PAN" linear combination defined as $I^{A_n} = \alpha \cdot I_{\text{MSI}}^{A_n}$, where $\alpha \in \mathbb{R}^3$ are fixed parameters pre-estimated using standard linear regression (Pic et al., 2024). In this configuration, training is explicitly performed using both the PAN and low-resolution MSI inputs, by computing the loss between the synthesized and training PAN images, as well as between a rendered low-resolution MSI and its corresponding training data.

The third approach, *single channel*, omits the low-resolution MSI inputs as in the 3-PAN configuration. However, it reconstructs the scene directly from PAN images only, such that the color features \mathbf{f}_k lie in \mathbb{R}^d with $d = 1$.

The final approach, 3-PAN, uses only the panchromatic data and still achieves reconstruction accuracy comparable to the *Brovvey* configuration. In contrast, the *linear combination* method, which jointly trains on both MSI and panchromatic images, leads to a noticeable performance degradation of approximately 0.5 m. When comparing the *single channel* and 3-PAN approaches, both operating on panchromatic data only, the *single channel* strategy produces slightly worse results (by approximately 0.14 m on average). This observation suggests that the modeling of panchromatic rendering has a non-negligible impact on reconstruction quality. Although the applied pansharpening technique is the *Brovvey* method, these findings indicate that Gaussian Splatting cannot effectively exploit both

modalities simultaneously during training. Instead, panchromatic and MSI inputs should be processed separately. This observation is consistent with previous work (Sprintson et al., 2025b), where the two modalities were also treated independently.

5.5 Comparing EOGS++ with the state of the art

The results are reported in Table 3 and visual results are reported in Figure 6. Additional results obtained without the foliage

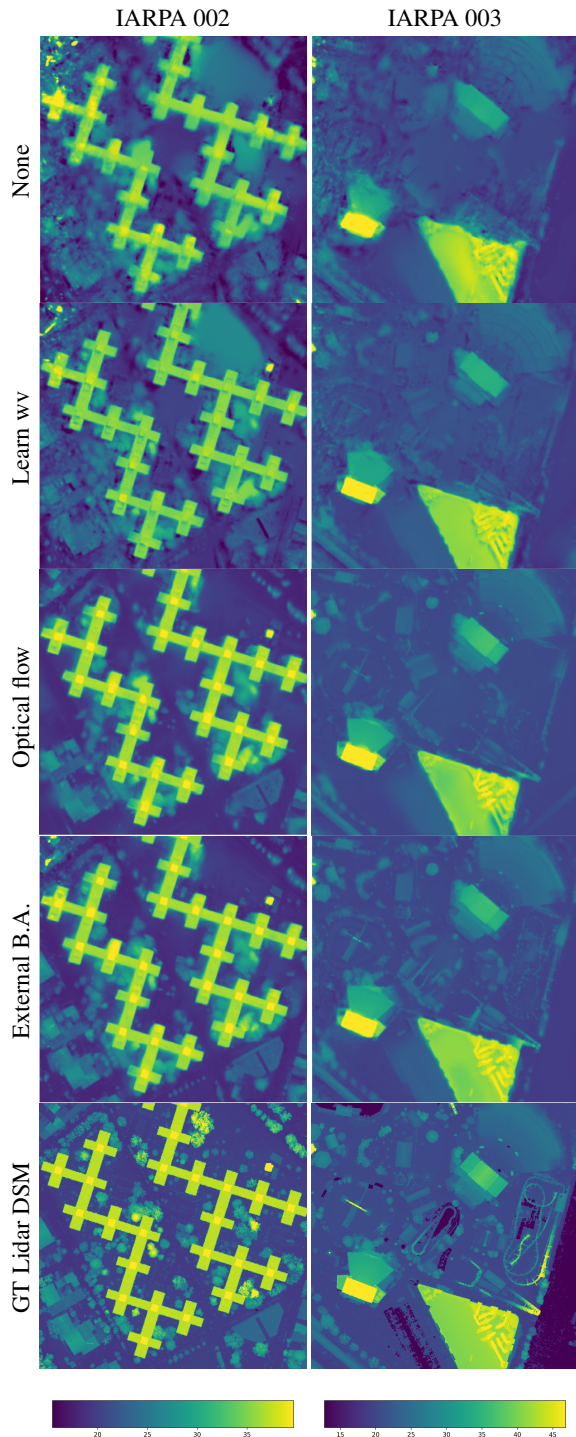


Figure 4. Qualitative comparison of different methods for handling errors in the camera pointing (cf. Table 1) on IARPA 002 and IARPA 003 scenes.

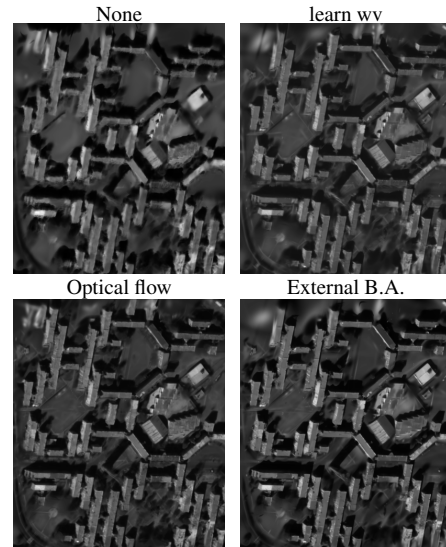


Figure 5. Rendered images on IARPA 001 using different B.A. strategies. Notice how the images become sharper.

	HR MSI	LR MSI	HR PAN	JAX ↓	IARPA ↓	Mean ↓
Brovoy	✓	✗	✗	1.19	1.46	1.33
Linear combination	✗	✓	✓	1.95	1.58	1.77
Single channel	✗	✗	✓	1.36	1.58	1.47
3-PAN	✗	✗	✓	1.19	1.47	1.33

Table 2. Mean absolute error (MAE) on elevation for different pansharpening strategies. HR denotes high-resolution, and LR denotes low-resolution data. Lower values indicate better performance. For a fair comparison, all methods are trained for exactly 5'000 iterations.

mask, as well as further comparative figures, are provided in the supplementary material. EOGS++, when using the same

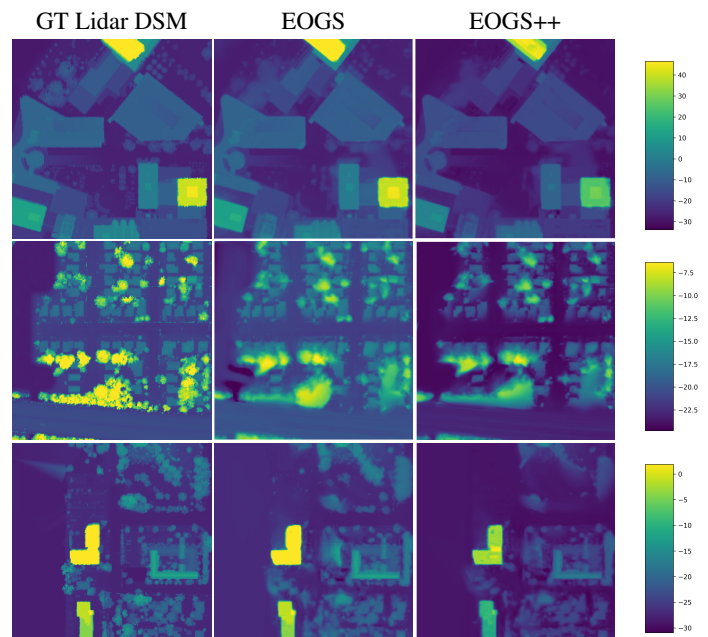


Figure 6. Qualitative comparison of EOGS and EOGS++ predicted DSMs on the JAX 214, JAX 004 and JAX 260 scenes.

	B.A.RPC Brovey Image	Opacity Reset TSDF	JAX				IARPA				Pooled mean ↓	Time [min] ↓	
			004	068	214	260	mean ↓	001	002	003			mean ↓
EOGS	✓	✗	0.77	1.07	1.65	1.26	1.19	1.41	1.86	1.13	1.46	1.33	9
EOGS (40k iters)	✓	✗	0.70	1.03	1.65	1.27	1.16	1.39	1.91	1.09	1.46	1.31	72
SAT-NGP	✓	✗	1.03	1.26	2.17	1.43	1.47	1.34	1.85	1.62	1.60	1.53	25
EO-NeRF	✓	✗	1.02	1.03	1.55	1.24	1.21	1.32	1.63	1.18	1.38	1.29	900
EOGS++	✓	✓	0.73	0.95	1.38	1.02	1.02	1.31	1.51	1.23	1.35	1.19	25
EOGS++	✗	✓	0.72	0.90	1.64	1.16	1.10	1.42	1.61	1.42	1.49	1.29	32
Ablation	✗	✗	0.88	1.25	1.63	1.46	1.30	1.57	1.93	1.33	1.61	1.46	31

Table 3. Quantitative comparison of EOGS++ and baseline methods on elevation reconstruction accuracy and runtime. Baselines include EOGS (Savant Aira et al., 2025), SAT-NGP (Billouard et al., 2024), and EO-NeRF (Marí et al., 2023). The first column denote the use of preprocessed data. The two last row reports the performance when training directly with the raw RPCs provided by the satellite vendor and raw panchromatic data.

data modalities as prior work, consistently achieves the best performance on buildings compared to existing models by 0.1 meters.

The early stopping mechanism effectively removes misplaced Gaussians and results in a more coherent scene reconstruction.

Instead, when relying solely on the non-bundle-adjusted panchromatic data, the model achieves performance comparable to the original EOGS, despite EOGS relying on camera priors and pre-pansharpened MSI data.

However, when evaluating results without the foliage mask, performance remains inferior to Eo-NeRF. As illustrated in Figure 6, certain fine vegetation structures, particularly in the JAX 260 scene, are not fully preserved. As components such as early stopping and opacity reset act as regularization mechanisms, they tend to suppress high-frequency elements with limited geometric stability. Given that vegetation is inherently transient and varies with seasons and acquisition dates, these details are therefore less critical for our primary objective, which is the reconstruction of permanent structures such as buildings.

The last row of Table 3 reports the performance of EOGS++ when trained directly on raw panchromatic data and raw RPCs, without any of the proposed enhancements. As PAN imagery provides weaker geometric cues than pansharpened inputs, the resulting MAE is higher than when using pansharpened data (shown in table 1). Nevertheless, integrating the full set of improvements yields a substantial reduction in error (about 0.2 m on average), indicating that each component contributes meaningfully to training stability and reconstruction accuracy.

6. Conclusion

In this work, we presented EOGS++, an enhanced Gaussian Splatting framework specifically designed for Earth observation. Building upon the original EOGS, our approach eliminates preprocessing requirements, such as external bundle adjustment and pansharpening, by directly integrating pose refinement through optical flow and by operating natively on raw panchromatic imagery. Additional mechanisms, including opacity reset, early stopping, and TSDF-based postprocessing, further improve reconstruction sharpness and geometric accuracy.

Experiments conducted on the IARPA2016 and DFC2019 datasets demonstrate that EOGS++ achieves state-of-the-art reconstruction accuracy. The framework performs robustly across varying acquisition conditions, achieving consistent improvements over existing methods, though challenges remain in capturing fine vegetation structures.

Acknowledgments

This work was granted access to the HPC resources of IDRIS under the allocation AD011012453R4 made by GENCI.

References

- Behari, N., Dave, A., Tiwary, K., Yang, W., Raskar, R., 2024. SUNDIAL: 3D Satellite Understanding through Direct, Ambient, and Complex Lighting Decomposition. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, Seattle, WA, USA, 522–532.
- Billouard, C., Derksen, D., Sarrazin, E., Vallet, B., 2024. Satngp : Unleashing neural graphics primitives for fast relightable transient-free 3d reconstruction from satellite imagery.
- Bosch, M., Foster, K., Christie, G., Wang, S., Hager, G. D., Brown, M., 2019. Semantic stereo for incidental satellite images. *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, 1524–1532.
- Carper, W., Lillesand, T., Kiefer, R. et al., 1990. The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data. *Photogrammetric Engineering and remote sensing*, 56(4), 459–467.
- Chang, J.-R., Chen, Y.-S., 2018. Pyramid stereo matching network. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5410–5418.
- Chung, J., Oh, J., Lee, K. M., 2024. Depth-regularized optimization for 3d gaussian splatting in few-shot images. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 811–820.
- Ciotola, M., Vitale, S., Mazza, A., Poggi, G., Scarpa, G., 2022. Pan sharpening by Convolutional Neural Networks in the Full Resolution Framework. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-17.
- Curless, B., Levoy, M., 1996. A volumetric method for building complex models from range images. *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, 303–312.
- De Franchis, C., Meinhardt-Llopis, E., Michel, J., Morel, J.-M., Facciolo, G., 2014a. An automatic and modular stereo pipeline for pushbroom images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.
- De Franchis, C., Meinhardt-Llopis, E., Michel, J., Morel, J.-M., Facciolo, G., 2014b. On stereo-rectification of pushbroom images. *2014 IEEE international conference on image processing (ICIP)*, IEEE, 5447–5451.

- Deng, Y., Wang, Y., Xiao, R., Tang, C., Zhou, J., Fan, J., Xiong, D., Lv, J., Tang, H., 2025. Ebad-gaussian: Event-driven bundle adjusted deblur gaussian splatting. *arXiv preprint arXiv:2504.10012*.
- Derksen, D., Izzo, D., 2021. Shadow neural radiance fields for multi-view satellite photogrammetry. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1152–1161.
- Ghadjati, M., Moussaoui, A., Boukharouba, A., 2019. A novel iterative PCA-based pansharpening method. *Remote sensing letters*, 10(3), 264–273.
- Grodecki, J., Dial, G., 2003. Block adjustment of high-resolution satellite images described by rational polynomials. *Photogrammetric Engineering & Remote Sensing*, 69(1), 59–68.
- Guédon, A., Gomez, D., Maruani, N., Gong, B., Drettakis, G., Ovsjanikov, M., 2025. MILo: Mesh-In-the-Loop Gaussian Splatting for Detailed and Efficient Surface Reconstruction. *arXiv preprint arXiv:2506.24096*.
- Huang, B., Yu, Z., Chen, A., Geiger, A., Gao, S., 2024. 2d gaussian splatting for geometrically accurate radiance fields. *ACM SIGGRAPH 2024 conference papers*, 1–11.
- Huang, X., Liu, X., Wan, Y., Zheng, Z., Zhang, B., Xiong, M., Pei, Y., Zhang, Y., 2025. SkySplat: Generalizable 3D Gaussian Splatting from Multi-Temporal Sparse Satellite Images. *arXiv:2508.09479 [cs]*.
- Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Trans. Graph.*, 42(4), 139–1.
- Krauß, T., d'Angelo, P., Schneider, M., Gstaiger, V., 2013. The fully automatic optical processing system CATENA at DLR. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40, 177–183.
- Le Saux, B., Yokoya, N., Hansch, R., Brown, M., Hager, G., 2019. 2019 data fusion contest [technical committees]. *IEEE Geoscience and Remote Sensing Magazine*, 7(1), 103–105.
- Li, H., Manjunath, B., Mitra, S. K., 1995. Multisensor image fusion using the wavelet transform. *Graphical models and image processing*, 57(3), 235–245.
- Lin, C.-H., Ma, W.-C., Torralba, A., Lucey, S., 2021. Barf: Bundle-adjusting neural radiance fields. *Proceedings of the IEEE/CVF international conference on computer vision*, 5741–5751.
- Lorensen, W. E., Cline, H. E., 1998. Marching cubes: A high resolution 3d surface construction algorithm. *Seminal graphics: pioneering efforts that shaped the field*, 347–353.
- Lowe, D. G., 1999. Object recognition from local scale-invariant features. *Proceedings of the seventh IEEE international conference on computer vision*, 2, Ieee, 1150–1157.
- Marí, R., de Franchis, C., Meinhardt-Llopis, E., Anger, J., Facciolo, G., 2021. A generic bundle adjustment methodology for indirect RPC model refinement of satellite imagery. *Image Processing On Line*, 11, 344–373.
- Marí, R., Facciolo, G., Ehret, T., 2022. Sat-nerf: Learning multi-view satellite photogrammetry with transient objects and shadow modeling using rpc cameras. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1311–1321.
- Marí, R., Facciolo, G., Ehret, T., 2023. Multi-date earth observation nerf: The detail is in the shadows. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2035–2045.
- Meng, X., Xiong, Y., Shao, F., Shen, H., Sun, W., Yang, G., Yuan, Q., Fu, R., Zhang, H., 2020. A large-scale benchmark data set for evaluating pansharpening performance: Overview and implementation. *IEEE Geoscience and Remote Sensing Magazine*, 9(1), 18–52.
- Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., Ng, R., 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1), 99–106.
- Palsson, F., Sveinsson, J. R., Ulfarsson, M. O., 2013. A new pansharpening algorithm based on total variation. *IEEE Geoscience and Remote Sensing Letters*, 11(1), 318–322.
- Pic, E., Ehret, T., Facciolo, G., Marí, R., 2024. Pseudo pansharpening nerf for satellite image collections. *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 2650–2655.
- Savant Aira, L., Facciolo, G., Ehret, T., 2025. Gaussian Splatting for Efficient Satellite Image Photogrammetry. *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5959–5969.
- Sprintson, M., Chellappa, R., Peng, C., 2025a. Fusionrf: High-fidelity satellite neural radiance fields from multispectral and panchromatic acquisitions. *IEEE Journal of Selected Topics in Signal Processing*.
- Sprintson, M., Chellappa, R., Peng, C., 2025b. FusionRF: High-Fidelity Satellite Neural Radiance Fields from Multispectral and Panchromatic Acquisitions. *arXiv:2409.15132 [cs]*.
- Szeliski, R., 2022. *Computer vision: algorithms and applications*. Springer Nature.
- Teed, Z., Deng, J., 2020. Raft: Recurrent all-pairs field transforms for optical flow. *European conference on computer vision*, Springer, 402–419.
- Triggs, B., McLauchlan, P. F., Hartley, R. I., Fitzgibbon, A. W., 1999. Bundle adjustment—a modern synthesis. *International workshop on vision algorithms*, Springer, 298–372.
- Yan, C., Qu, D., Xu, D., Zhao, B., Wang, Z., Wang, D., Li, X., 2024. Gs-slam: Dense visual slam with 3d gaussian splatting. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19595–19604.
- Zhang, J., Zheng, Y., Li, Z., Dai, Q., Yuan, X., 2024. Gbr: Generative bundle refinement for high-fidelity gaussian splatting and meshing. *arXiv preprint arXiv:2412.05908*.
- Zhang, L., Rupnik, E., 2023. Sparsesat-nerf: Dense depth supervised neural radiance fields for sparse satellite images. *arXiv preprint arXiv:2309.00277*.