drones

MDPI

Article

Reconstruction-Aware Observation Selection for Bandwidth-Constrained Multi-UAV Gaussian Splatting Reconstruction

Zhen Ding [1,*], Tao Ma [2] and Guanxin Jia [1]

[1] College of Computer and Artificial Intelligence, Northeast Forestry University, No. 26 Hexing Road, Xiangfang District, Harbin 150040, China; jiaguanxin@nefu.edu.cn
[2] College of Geomatics, Liaoning Technical University, No. 47 Zhonghua Road, Xihe District, Fuxin 123000, China; 472510052@stu.lntu.edu.cn
* Correspondence: dingzhen@nefu.edu.cn

Highlights

What are the main findings?

- RAOS-GS selects reconstruction-relevant RGB-D observations for bandwidth-constrained multi-UAV Gaussian Splatting reconstruction.
- RAOS-GS substantially reduces transmitted observations while maintaining competitive reconstruction quality on simulated and real-world UAV scenes.

What are the implications of the main findings?

- Reconstruction-aware observation selection can reduce air–ground communication demands without requiring high-fidelity onboard Gaussian mapping.
- The proposed framework provides a practical solution for collaborative 3D reconstruction under limited communication and onboard computing resources.

Abstract

Multi-UAV observation improves the efficiency and coverage of three-dimensional scene acquisition, but continuous onboard Gaussian Splatting reconstruction is computationally expensive, whereas transmitting dense RGB-D streams wastes communication bandwidth. We propose RAOS-GS (Reconstruction-Aware Observation Selection for Gaussian Splatting), an air–ground collaborative framework for bandwidth-constrained multi-UAV reconstruction. Each UAV maintains a lightweight Gaussian proxy solely to estimate the reconstruction value of candidate observations. The proposed Rendering-Aware Keyframe Selection (RAKS) evaluates uncovered scene content, appearance discrepancy within reliably rendered regions, and relative viewpoint change to retain complementary, non-redundant RGB-D keyframes. The ground station then constructs per-UAV Gaussian submaps, aligns their coordinate frames through cross-UAV RGB-D registration, and jointly refines camera poses and Gaussian parameters to produce a global representation. Experiments on three Replica scenes, two OpenFly-UE virtual outdoor scenes, and three real-world multi-UAV scenes demonstrate a favorable trade-off between observation volume and reconstruction quality. RAOS-GS retains only 5.1%–7.9% of the input observations on Replica and the real-world sequences and 18.7%–24.0% on OpenFly-UE. It consistently outperforms 10% uniform temporal sampling and remains competitive with or surpasses 30% sampling despite using fewer observations. Ablation studies confirm that rendering-aware selection identifies reconstruction-relevant observations, viewpoint constraints suppress redundancy, and ground-side pose–Gaussian refinement improves final reconstruction quality. These results show that reconstruction-aware observation selection can reduce air–ground data transmission without requiring high-fidelity onboard Gaussian mapping.

Check for updates

Academic Editor: Emmanouel T. Michailidis

Received: 3 September 2026
Revised: 31 October 2026
Accepted: 3 November 2026
Published: 5 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: multi-UAV reconstruction; 3D Gaussian Splatting reconstruction-aware observation selection; keyframe selection; communication-efficient reconstruction; air–ground collaboration

1. Introduction

Large-scale 3D scene reconstruction supports applications such as disaster inspection, infrastructure monitoring, environmental mapping, and scene digitization. Multi-UAV systems can accelerate data acquisition and improve coverage through parallel trajectories and complementary viewpoints. Meanwhile, 3D Gaussian Splatting (3DGS) provides an explicit, differentiable scene representation that supports high-quality reconstruction and efficient novel-view rendering [1]. Gaussian SLAM systems such as SplaTAM and Photo-SLAM further demonstrate that Gaussian representations can support online tracking and incremental mapping from RGB-D observations [2,3].

Applying this capability to resource-constrained UAV teams remains difficult because both onboard computation and air–ground bandwidth are limited. Existing multi-agent Gaussian SLAM systems typically require each agent to construct and optimize a local Gaussian submap before communicating keyframes, compressed submaps, or Gaussian summaries for cross-agent alignment and fusion [4–8]. This design reduces raw-data transmission but retains substantial onboard computation, memory, and energy demands. At the other extreme, continuously transmitting multiple RGB-D streams to a ground server causes communication and backend processing costs to grow with mission duration and team size.

We therefore consider a different allocation of responsibilities: high-fidelity Gaussian reconstruction is performed at the ground server, while each UAV handles tracking, lightweight scene-state maintenance, and communication-oriented observation selection. The central problem is to determine which RGB-D observations should be transmitted before their contribution to the final reconstruction is known. Fixed-rate sampling can discard informative views or retain many near-duplicate frames, while conventional motion- and tracking-based keyframe criteria are designed primarily for local localization and optimization rather than downstream reconstruction quality [9]. Effective selection must instead retain views that reveal insufficiently modeled regions or appearance information that the current scene representation cannot explain while explicitly suppressing viewpoint redundancy.

Selection also makes ground-side reconstruction more challenging. The retained observations are sparse in time and space, remain expressed in separate UAV-local coordinate frames, and may be affected by trajectory drift, RGB-D noise, and cross-view matching errors. Communication-aware observation selection must therefore be coupled with cross-UAV registration and global refinement so that a limited set of uploaded observations can still produce a consistent scene representation.

To address these challenges, we propose RAOS-GS (Reconstruction-Aware Observation Selection for Gaussian Splatting), an air–ground collaborative framework for communication-constrained multi-UAV 3D reconstruction. Each UAV maintains a lightweight Gaussian proxy map solely to evaluate candidate observations; the proxy is neither transmitted nor used in final map fusion. Based on this proxy, Rendering-Aware Keyframe Selection (RAKS) combines accumulated-opacity coverage, appearance discrepancy within reliably rendered regions, and relative viewpoint change. It therefore prioritizes observations that complement the current scene representation while rejecting spatially redundant views.

The ground server receives the selected RGB-D keyframes and their local poses, constructs a high-quality Gaussian submap for each UAV, and estimates cross-UAV transformations from overlapping RGB-D observations. The aligned submaps then initialize joint pose–Gaussian refinement, which reduces residual trajectory and registration inconsistencies. This design separates low-cost online observation assessment from computationally intensive reconstruction while maintaining a direct connection between transmission decisions and final reconstruction requirements.

The main contributions of this study are as follows:

1. We propose RAOS-GS, an air–ground collaborative framework that concentrates high-fidelity Gaussian reconstruction on the ground server and uses the onboard Gaussian representation only as a lightweight observation-assessment proxy.
2. We propose Rendering-Aware Keyframe Selection (RAKS), which jointly evaluates proxy-map coverage, appearance discrepancy, and relative viewpoint change to retain reconstruction-relevant RGB-D observations and suppress near-duplicate transmissions.
3. We evaluate RAOS-GS on Replica, OpenFly-UE virtual outdoor scenes, and real-world multi-UAV sequences through reconstruction-quality comparisons, observation-reduction analysis, uniform-sampling baselines, and component ablations.

2. Related Work

2.1. 3D Gaussian Splatting and Gaussian SLAM

3D Gaussian Splatting (3DGS) represents a scene using anisotropic Gaussian primitives parameterized by position, covariance, opacity, and appearance and supports high-quality rendering through differentiable visibility-aware splatting [1]. Its explicit representation and efficient renderer have motivated its integration with visual SLAM for joint camera tracking and dense scene reconstruction.

Gaussian SLAM systems primarily differ in how they couple tracking, Gaussian-map updates, and geometric optimization. GS-SLAM combines RGB-D tracking with adaptive Gaussian expansion and coarse-to-fine pose optimization, while SplaTAM uses a unified Gaussian representation to jointly estimate camera poses and a renderable map from unposed RGB-D sequences [2,10]. Photo-SLAM integrates feature-based tracking with photorealistic Gaussian mapping across monocular, stereo, and RGB-D settings, whereas MonoGS couples direct tracking and incremental Gaussian reconstruction [3,11]. Gaussian-SLAM and RTG-SLAM further improve photometric consistency and the scalability of real-time RGB-D mapping [12,13].

These methods establish 3DGS as an effective representation for localization, dense reconstruction, and novel-view synthesis. Nevertheless, they are primarily designed for a single continuous sensor stream and assume that the sensing platform can maintain and optimize a reconstruction-quality Gaussian map online. They do not address how reconstruction computation should be divided between resource-constrained UAVs and a ground server, or how a lightweight, non-final Gaussian representation can be used to evaluate observations before transmission.

2.2. Collaborative Multi-Agent SLAM and Gaussian Splatting Reconstruction

Collaborative SLAM combines observations from multiple agents through inter-robot place recognition, coordinate alignment, pose-graph optimization, and map fusion. CCM-SLAM and COVINS adopt centralized architectures in which agents upload keyframes, features, or local states for server-side loop closure and global optimization [14,15]. Kimera-Multi instead combines inter-robot loop closures with distributed pose-graph optimization to construct a globally consistent metric-semantic mesh [16]. These systems provide estab-

lished mechanisms for cross-agent association and global consistency, but their primary objective is geometric mapping rather than photorealistic Gaussian reconstruction.

Recent collaborative systems incorporate neural or Gaussian scene representations to obtain dense, renderable global maps. CP-SLAM employs neural point fields within a collaborative pipeline that includes odometry, loop closure, submap fusion, and global refinement [17]. Gaussian-based systems generally construct local Gaussian submaps at each agent and subsequently align or fuse them. MAGiC-SLAM performs loop closure and merging of RGB-D Gaussian submaps, while MAC-Ego3D associates, aligns, and jointly optimizes local maps through multi-agent Gaussian consensus [4,5]. CoMA-SLAM uses 2D Gaussian surfels, distributed keyframe optimization, and submap-centered updates to improve consistency and communication efficiency [6]. Coko-SLAM compresses Gaussian submaps and combines keyframe optimization with centralized loop closure, whereas MAGS-SLAM communicates compact submap summaries and uses a Sim(3) pose graph and occupancy-aware fusion for monocular multi-agent reconstruction [7,8]. Surveys consequently identify global consistency, communication management, map fusion, and scalable optimization as central challenges in collaborative 3DGS-SLAM [18].

Despite differences in representation and backend design, these approaches generally treat a locally constructed and optimized submap as the unit communicated or incorporated into the final global map. Submap compression and compact summaries reduce transmitted data, but the agents must still perform substantial local Gaussian reconstruction. The setting considered here is complementary: UAVs maintain only a lightweight proxy for observation assessment, while reconstruction-quality Gaussian mapping and cross-UAV refinement are performed at the ground server.

2.3. Communication-Constrained Observation Selection

Keyframe selection in visual SLAM is commonly used to bound computation and map growth. Conventional insertion criteria rely on camera translation or rotation, feature-track quality, local covisibility, and bundle-adjustment requirements and are designed primarily to preserve tracking and pose-estimation accuracy [9]. Collaborative systems additionally reduce the keyframes, features, or local states exchanged among agents. For example, COVINS manages redundant visual–inertial states at a central server, while optimization-based map summarization can select measurements that preserve valuable localization constraints under storage or processing budgets [15,19].

Keyframe insertion for tracking, backend map summarization, and observation selection before transmission serve distinct objectives. The first maintains local estimator stability, and the second preserves representative poses or geometric constraints after data have been collected. Neither is primarily designed to estimate, before upload, whether an RGB-D observation reveals insufficiently represented regions or appearance information that would improve subsequent Gaussian reconstruction. Motion-only criteria may therefore retain nearby views with limited additional scene content, whereas fixed-rate sampling may discard observations that are important to reconstruction.

Communication-constrained multi-UAV 3DGS requires an observation-selection criterion that connects data transmission to downstream reconstruction value without imposing final-quality mapping on each UAV. RAOS-GS addresses this gap by using an onboard Gaussian proxy only to assess rendering coverage, appearance discrepancy, and viewpoint redundancy. Selected RGB-D observations, rather than locally reconstructed final submaps, are transmitted for ground-side Gaussian reconstruction, cross-UAV alignment, and joint refinement.

3. Materials and Methods

3.1. Problem Formulation and System Overview

Consider S UAVs observing the same static scene, where $s \in \{1, \ldots, S\}$ denotes the UAV index. At time t, the s-th UAV captures an RGB-D frame and estimates its camera pose in the UAV-local world frame:

$$F_t^{(s)} = \left(I_t^{(s)}, D_t^{\text{obs},(s)}\right), \qquad T_{wc,t}^{(s)} \in SE(3), \tag{1}$$

where $I_t^{(s)}$ and $D_t^{\text{obs},(s)}$ are the RGB image and observed depth map, respectively. $T_{wc,t}^{(s)}$ transforms points from the camera frame to the local world frame, and $T_{cw,t}^{(s)} = (T_{wc,t}^{(s)})^{-1}$. The camera-intrinsic matrix $K^{(s)}$ is assumed to be calibrated.

For each frame, a communication-keyframe indicator $z_t^{(s)} \in \{0, 1\}$ determines whether the observation is uploaded. If $z_t^{(s)} = 1$, the UAV generates the ordered packet

$$\mathcal{P}_t^{(s)} = \left(I_t^{(s)}, D_t^{\text{obs},(s)}, K^{(s)}, T_{wc,t}^{(s)}, t, s\right). \tag{2}$$

As illustrated in Figure 1, each UAV performs tracking and maintains a lightweight Gaussian proxy for selecting keyframe; this proxy is neither uploaded nor fused into the final map. The ground server reconstructs one Gaussian submap from each uploaded packet set, aligns the UAV-local coordinate frames, and jointly refines the camera poses and Gaussian parameters.

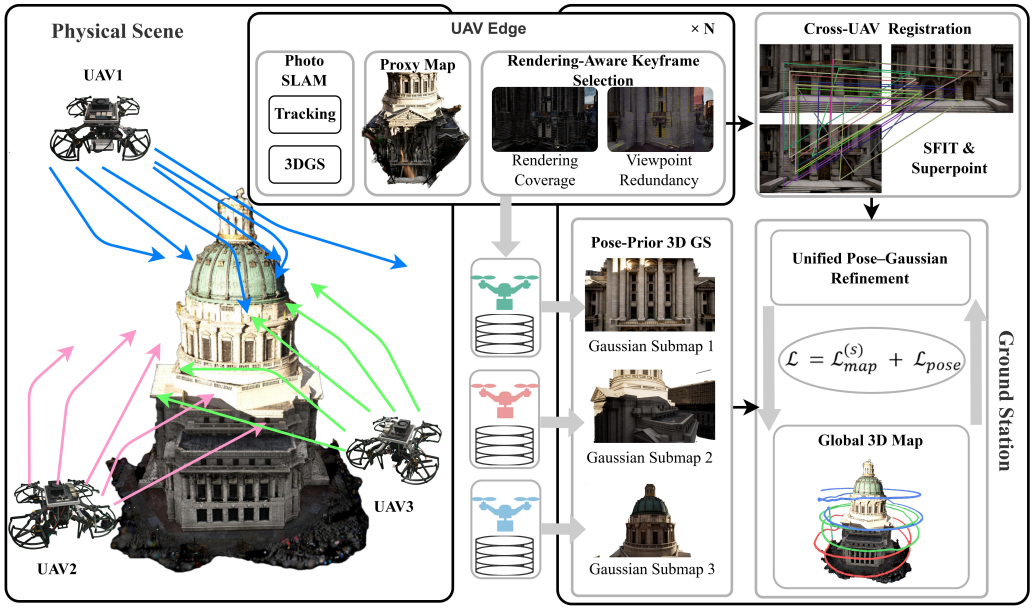


Figure 1. The RAOS-GS air–ground collaborative reconstruction framework.

3.2. Communication-Efficient Onboard Keyframe Selection

Rendering-Aware Keyframe Selection (RAKS) operates on the Gaussian proxy maintained by the RGB-D Photo-SLAM front end [3]. For each candidate frame, it measures proxy-map coverage and appearance discrepancy and then applies a relative-viewpoint gate to suppress nearby redundant observations. This selection affects only communication; all frames remain available to the onboard tracking front end.

3.2.1. Rendering Coverage and Observation Discrepancy Evaluation

At time t, the s-th UAV renders a predicted color image from its latest proxy map $\mathcal{G}_{e,t}^{(s)}$ at the candidate pose:

$$\hat{I}_t^{(s)} = \mathcal{R}(\mathcal{G}_{e,t}^{(s)}, T_{cw,t}^{(s)}; \mathbf{c}, \mathbf{b}). \tag{3}$$

where $\mathcal{R}$ is the Gaussian renderer, and $\mathbf{c}$ and $\mathbf{b}$ denote the Gaussian colors and rendering background, respectively.

Because color residuals are unreliable in regions with insufficient Gaussian support, RAKS first evaluates the accumulated opacity along each pixel ray:

$$A_t^{(s)}(p) = 1 - \prod_{i \in \mathcal{N}(p)} (1 - \alpha_i(p)). \tag{4}$$

Here, $\mathcal{N}(p)$ is the depth-ordered set of Gaussians contributing to pixel p, and $\alpha_i(p)$ is their effective opacity. Thus, $A_t^{(s)}(p) \in [0,1]$ measures rendering support, with lower values indicating insufficiently modeled regions. The opacity map is rendered by assigning a constant attribute of 1 to the Gaussians and 0 to the background [2].

The covered and reliably rendered regions are defined by two opacity thresholds:

$$\begin{aligned} \Omega_{\mathrm{cov},t}^{(s)} &= \left\{ p \in \Omega \mid A_t^{(s)}(p) > \tau_{\mathrm{cov}} \right\}, \quad C_t^{(s)} = \frac{\left|\Omega_{\mathrm{cov},t}^{(s)}\right|}{|\Omega|}, \\ \Omega_{\mathrm{app},t}^{(s)} &= \left\{ p \in \Omega \mid A_t^{(s)}(p) > \tau_{\mathrm{app}} \right\}, \quad \tau_{\mathrm{cov}} < \tau_{\mathrm{app}}. \end{aligned} \tag{5}$$

Here, Ω is the image domain, $C_t^{(s)}$ is the proxy-map coverage ratio, and the higher threshold τ_{app} restricts appearance evaluation to pixels with reliable rendering support.

After normalizing the observed and rendered images to $[0,1]$, their pixelwise RGB residual is

$$e_{\mathrm{rgb},t}^{(s)}(p) = \left\| I_t^{(s)}(p) - \hat{I}_t^{(s)}(p) \right\|_1. \tag{6}$$

Sobel-gradient magnitudes additionally measure local structural disagreement:

$$e_{\nabla,t}^{(s)}(p) = \left| \left\| \nabla_S I_t^{g,(s)}(p) \right\|_2 - \left\| \nabla_S \hat{I}_t^{g,(s)}(p) \right\|_2 \right|. \tag{7}$$

where I^g is the grayscale image and ∇_S is the Sobel operator. The appearance-discrepancy score over the reliable region is

$$E_{\mathrm{app},t}^{(s)} = \lambda_q \, \mathrm{Quantile}_q\left(\left\{ e_{\mathrm{rgb},t}^{(s)}(p) \mid p \in \Omega_{\mathrm{app},t}^{(s)} \right\} \right) + \frac{\lambda_g}{\left|\Omega_{\mathrm{app},t}^{(s)}\right|} \sum_{p \in \Omega_{\mathrm{app},t}^{(s)}} e_{\nabla,t}^{(s)}(p). \tag{8}$$

where $\lambda_q, \lambda_g \geq 0$, $\lambda_q + \lambda_g = 1$, and q is the selected upper quantile. The quantile term preserves sensitivity to localized high-residual regions that could be diluted by a global mean. The rendering-value gate is then

$$h_{R,t}^{(s)} = \mathbf{1}\left[C_t^{(s)} < \tau_C \ \vee \ E_{\mathrm{app},t}^{(s)} > \tau_E \right]. \tag{9}$$

Thus, a candidate is considered reconstruction-relevant if it either exposes insufficiently covered content or contains reliably rendered regions with a large appearance discrepancy.

3.2.2. Viewpoint Redundancy Suppression and Upload Decision

Rendering discrepancies may be triggered repeatedly at nearby poses. Let $\ell_{t-1}^{(s)}$ denote the most recently uploaded frame. RAKS suppresses such redundancy using the relative translation and rotation

$$\begin{aligned} \Delta T_t^{(s)} &= T_{cw,t}^{(s)} \left(T_{cw,\ell_{t-1}^{(s)}}^{(s)} \right)^{-1}, \\ d_t^{(s)} &= \left\| \mathrm{trans}\left(\Delta T_t^{(s)} \right) \right\|_2, \qquad \theta_t^{(s)} = \frac{180}{\pi} \left\| \log\left(\mathrm{rot}\left(\Delta T_t^{(s)} \right) \right) \right\|_2, \\ h_{M,t}^{(s)} &= \mathbf{1}\left[d_t^{(s)} \geq \tau_d \ \vee \ \theta_t^{(s)} \geq \tau_\theta \right]. \end{aligned} \tag{10}$$

Here, τ_d and τ_θ are the translation and rotation thresholds. For the first frame, $h_{M,t}^{(s)}$ is set to 1.

The final upload decision combines the rendering and viewpoint gates:

$$z_t^{(s)} = \begin{cases} h_{M,t}^{(s)}, & \Omega_{\mathrm{cov},t}^{(s)} = \varnothing \vee \Omega_{\mathrm{app},t}^{(s)} = \varnothing, \\ h_{R,t}^{(s)} \wedge h_{M,t}^{(s)}, & \Omega_{\mathrm{cov},t}^{(s)} \neq \varnothing \wedge \Omega_{\mathrm{app},t}^{(s)} \neq \varnothing. \end{cases} \tag{11}$$

When both opacity-derived regions are valid, upload requires reconstruction value and sufficient viewpoint change. During proxy initialization or invalid rendering, RAKS falls back to the viewpoint gate. If $z_t^{(s)} = 1$, $\mathcal{P}_t^{(s)}$ is queued and $\ell_t^{(s)}$ is updated; otherwise, the frame remains available only for onboard tracking and proxy maintenance.

3.3. Global Gaussian Map Construction and Joint Optimization

The ground server reconstructs the scene directly from $\{\mathcal{P}^{(s)}\}_{s=1}^{S}$. It first builds an independent Gaussian submap in each UAV-local frame, estimates rigid cross-UAV transformations from overlapping RGB-D observations, and then jointly refines the aligned poses and Gaussian parameters in a common frame.

3.3.1. Pose-Prior-Guided Local Gaussian Submap Reconstruction

For each UAV, the server reconstructs a Gaussian submap from the uploaded frames indexed by $\mathcal{K}^{(s)}$ in its local world frame w_s. Before cross-UAV registration, the submaps remain independent:

$$\mathcal{G}_s = \left\{ g_j^{(s)} \right\}_{j=1}^{N_s}, \qquad g_j^{(s)} = \left(\mu_j^{(s)}, \Sigma_j^{(s)}, \alpha_j^{(s)}, c_j^{(s)} \right). \tag{12}$$

where $\mu_j^{(s)}$, $\Sigma_j^{(s)}$, $\alpha_j^{(s)}$, and $c_j^{(s)}$ are the center, covariance, opacity, and appearance of the j-th Gaussian.

For each $t \in \mathcal{K}^{(s)}$, the uploaded local pose initializes the camera state. The differentiable renderer produces color, depth, and accumulated opacity:

$$\left(\widehat{\mathbf{I}}_t^{(s)}, \widehat{\mathbf{D}}_t^{(s)}, \widehat{\mathbf{A}}_t^{(s)} \right) = \mathcal{R}\left(\mathcal{G}_s, \mathbf{K}^{(s)}, \mathbf{T}_{cw,t}^{(s)} \right), \qquad \mathbf{T}_{cw,t}^{(s)} = \left(\mathbf{T}_{wc,t}^{(s)} \right)^{-1}. \tag{13}$$

Within a local keyframe window $\mathcal{W}_s$, the server jointly refines the poses, exposure parameters, and Gaussian parameters using

$$\begin{aligned} \mathcal{L}_{\mathrm{map}}^{(s)} = \sum_{t \in \mathcal{W}_s} \Big[& \lambda_{\mathrm{rgb}} \left\| \mathbf{M}_{I,t}^{(s)} \odot \left(\widetilde{\mathbf{I}}_t^{(s)} - \mathbf{I}_t^{(s)} \right) \right\|_1 \\ & + \lambda_{\mathrm{dep}} \left\| \mathbf{M}_{D,t}^{(s)} \odot \left(\widehat{\mathbf{D}}_t^{(s)} - \mathbf{D}_t^{\mathrm{obs},(s)} \right) \right\|_1 \Big] + \lambda_{\mathrm{iso}} \mathcal{L}_{\mathrm{iso}}. \end{aligned} \tag{14}$$

Here, $\widetilde{\mathbf{I}}_t^{(s)}$ is the exposure-compensated rendering; $\mathbf{M}_{I,t}^{(s)}$ and $\mathbf{M}_{D,t}^{(s)}$ mask valid color and depth pixels; and $\mathcal{L}_{\text{iso}}$ penalizes degenerate Gaussian scales. The weights λ_{rgb}, λ_{dep}, and λ_{iso} balance the three terms. This optimization follows the differentiable mapping strategy of Gaussian Splatting SLAM [11].

Local optimization yields

$$\{\mathcal{G}_1^*, \mathcal{G}_2^*, \ldots, \mathcal{G}_S^*\}, \tag{15}$$

which are subsequently mapped to the global frame using $\mathbf{G}_s \equiv {}^{g}\mathbf{T}_{w_s}$.

3.3.2. Cross-UAV RGB-D Registration and Global Map Initialization

RAOS-GS estimates cross-UAV transformations with a coarse-to-fine RGB-D registration pipeline. SIFT correspondences first screen cross-stream keyframe pairs for potential overlap. For the retained pairs, SuperPoint extracts keypoints and descriptors [20], and LightGlue produces the final image correspondences [21].

Matches with valid depth are back-projected using the corresponding camera intrinsics and transformed into their UAV-local world frames with the uploaded poses. The resulting cross-frame 3D–3D correspondences are used in an SVD-based least-squares rigid registration, and transformations with excessive three-dimensional residuals are rejected.

The first UAV-local frame is chosen as the global reference. Each remaining $\mathbf{G}_s$ is obtained from a validated direct registration or propagated through a validated transformation chain. Applying these transformations to the local camera poses and Gaussian submaps produces an initial global map. Because this rigid initialization does not remove trajectory drift, depth noise, or residual registration error, it is followed by unified refinement.

3.3.3. Unified Pose–Gaussian Refinement

The aligned submaps initialize a unified Gaussian model. Short, temporally ordered keyframe blocks from the different UAV streams are introduced alternately so that optimization preserves local trajectory continuity while revisiting different scene regions.

For keyframe t from stream s, the initial global pose is

$$\overline{\mathbf{T}}_{wc,t}^{(s),g} = \mathbf{G}_s \mathbf{T}_{wc,t}^{(s)}. \tag{16}$$

When the current global map provides sufficient rendering coverage, rendered depth establishes 2D–3D constraints for geometric pose correction; photometric consistency then refines the pose.

A six-dimensional perturbation in $\mathfrak{se}(3)$ is defined as

$$\boldsymbol{\xi}_t^{(s)} = [\boldsymbol{v}_t^{(s)\mathsf{T}}, \boldsymbol{\omega}_t^{(s)\mathsf{T}}]^{\mathsf{T}} \in \mathbb{R}^6,$$

where $\boldsymbol{v}_t^{(s)}$ and $\boldsymbol{\omega}_t^{(s)}$ are the translational and rotational components. The current world-to-camera estimate $\overline{\mathbf{T}}_{cw,t}^{(s)}$ is updated with the left-multiplicative model

$$\mathbf{T}_{cw,t}^{(s)}\left(\boldsymbol{\xi}_t^{(s)}\right) = \exp\left(\widehat{\boldsymbol{\xi}}_t^{(s)}\right)\overline{\mathbf{T}}_{cw,t}^{(s)}, \qquad \widehat{\boldsymbol{\xi}}_t^{(s)} = \begin{bmatrix} [\boldsymbol{\omega}_t^{(s)}]_\times & \boldsymbol{v}_t^{(s)} \\ \mathbf{0}^{\mathsf{T}} & 0 \end{bmatrix}. \tag{17}$$

where $[\cdot]_\times$ is the skew-symmetric operator and $\exp(\cdot)$ maps $\mathfrak{se}(3)$ to $SE(3)$. Substitution into Equation (13) gives the rendering used to optimize

$$\mathcal{L}_{\text{pose}} = (1-\lambda_s)\left\|\widehat{\mathbf{I}}_t^{(s)} - \mathbf{I}_t^{(s)}\right\|_1 + \lambda_s\left[1 - \text{SSIM}\left(\widehat{\mathbf{I}}_t^{(s)}, \mathbf{I}_t^{(s)}\right)\right] + \lambda_{\tilde{\xi}}\left\|\boldsymbol{\xi}_t^{(s)}\right\|_2^2. \tag{18}$$

Here, λ_s balances the photometric and structural terms, while $\lambda_{\tilde{\xi}}$ regularizes the pose increment. A correction is accepted only if it improves full-image rendering quality and remains within the trusted translation and rotation ranges. Accepted poses then participate in the joint optimization of Equation (14), producing the final global map $\mathcal{G}^*$.

4. Experimental and Results

4.1. Experimental Setup

4.1.1. Datasets and Scenarios

RAOS-GS was evaluated in three types of scenes Figure 2with different observation conditions: indoor simulated scenes from Replica, Unreal Engine-based virtual outdoor environments from the OpenFly benchmark, and indoor RGB-D sequences collected by real UAVs. Replica served as the primary controlled benchmark for method comparisons and the main ablation experiments. OpenFly-UE was used to further examine the behavior of the method under outdoor visual conditions and different temporal sampling densities. The real-world UAV sequences are used to evaluate its performance in the presence of real sensor noise and flight-induced perturbations.

- **Replica simulated scenarios.** We selected the *apartment_0* , *apartment_1*, and *apartment_2* scenes from Replica. Three UAV observation trajectories with complementary spatial coverage were constructed for each scene. Partial view overlap was retained between trajectories to support cross-UAV observation association and subsequent global Gaussian reconstruction. The stable and controllable imaging conditions of Replica make it suitable as the primary environment for quantitative comparisons.
- **UE outdoor virtual scenarios.** One building and its surrounding area were selected from each of two Unreal Engine environments provided by the OpenFly benchmark, producing two outdoor experimental scenes denoted as OpenFly-UE-1 and OpenFly-UE-2. Each scene contains three UAV trajectories with partial overlap. These experiments do not reconstruct the complete city-scale OpenFly environments. Instead, they use local outdoor building regions to evaluate RAOS-GS under visual conditions that differ from those of the indoor Replica scenes.
- **Real-world UAV scenes.** Three real UAVs were used to collect RGB-D sequences in three indoor environments: Table, Lecture Hall, and Corridor. Each UAV was equipped with an Intel RealSense D435i RGB-D camera(Intel Corporation, Santa Clara, CA, USA). The Table scene primarily contains nearby objects and substantial local viewpoint variations. Lecture Hall covers a larger spatial area and contains longer observation trajectories. Corridor has an elongated structure with repeated textures and long-range motion. These scenes therefore provide different spatial structures and observation difficulties. The recorded RGB-D sequences were used for offline evaluation of RAOS-GS and the compared methods.

For the supplementary fleet-size evaluation, we also compare two- and three-UAV configurations on one representative scene from each dataset group; the results are reported in Table A10.

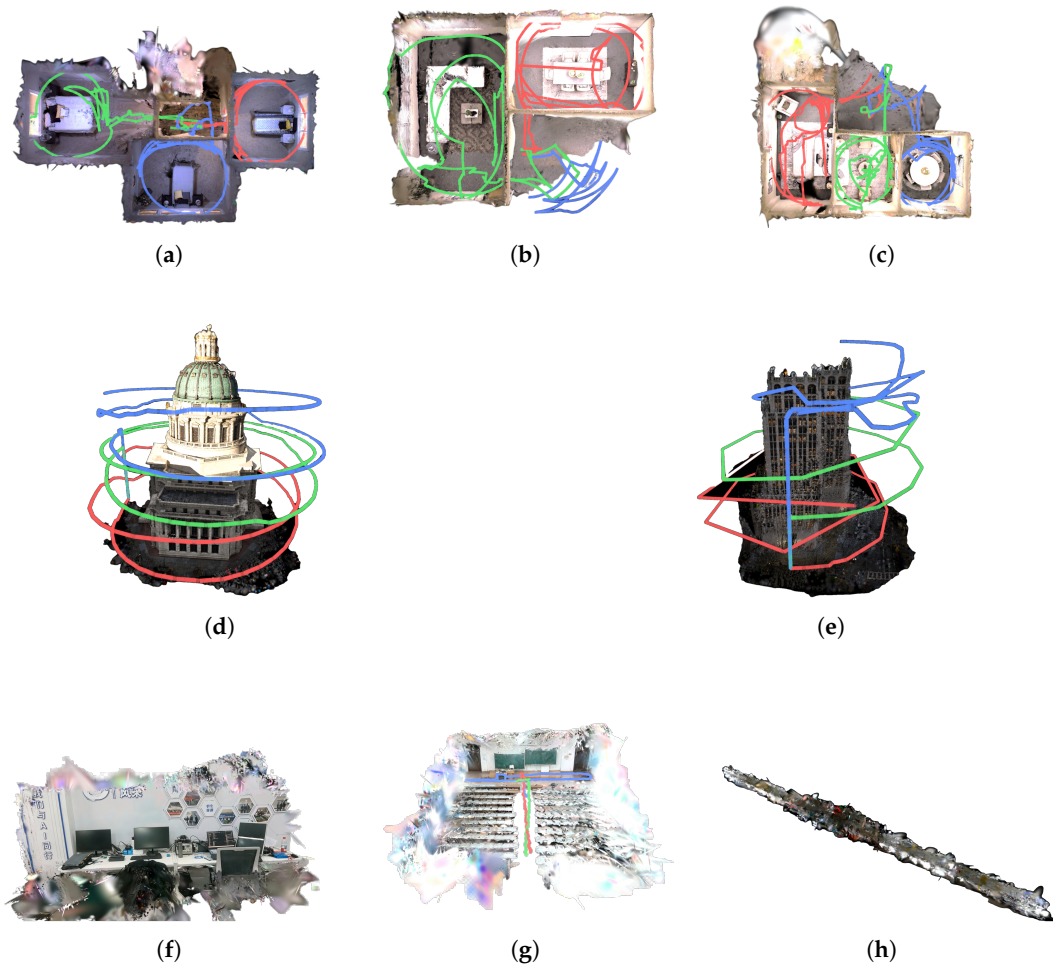


Figure 2. Representative scenes used in the experiments: (**a**) Replica Apt-0; (**b**) Replica Apt-1; (**c**) Replica Apt-2; (**d**) OpenFly-UE1; (**e**) OpenFly-UE2; (**f**) real-world table scene; (**g**) real-world lecture hall; and (**h**) real-world corridor.

4.1.2. Compared Methods and Evaluation Protocol

To evaluate the reconstruction performance of RAOS-GS under communication-constrained multi-UAV settings, we compare it with MonoGS [11], MAGiC-SLAM [4], MAC-Ego3D [5], and Coko-SLAM [7]. These methods represent different technical paradigms, including single-agent Gaussian reconstruction, multi-agent Gaussian map fusion, and communication-aware multi-agent Gaussian SLAM. MonoGS is included solely as a full-input, single-agent reference for reconstruction quality: each UAV trajectory is reconstructed independently, without cross-UAV alignment or map fusion. It is not treated as a fused multi-UAV communication baseline.

Comparative experiments were conducted on Replica and the real UAV sequences using the same source RGB-D data, camera intrinsics, and image resolution. Because MonoGS does not include cross-agent coordinate alignment or global-map fusion, it was run independently on the three UAV sequences and serves as a full-input single-agent reference rather than a fused multi-UAV map. The multi-agent methods were run according to their original processing pipelines.

For OpenFly-UE, only the complete RAOS-GS pipeline is evaluated. These experiments are used as a complementary assessment of observation selection and Gaussian reconstruction under outdoor virtual environments.

4.1.3. Evaluation Metrics

RAOS-GS was evaluated primarily in terms of reconstruction quality and the number of observations requiring transmission. Reconstruction quality was measured using peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and learned perceptual image patch similarity (LPIPS). Higher PSNR and SSIM values indicate better pixel fidelity and structural similarity, whereas a lower LPIPS value indicates greater perceptual similarity to the reference image. For the Replica scenes, camera-pose accuracy is additionally evaluated using ATE RMSE, with the results reported in Table A9.

To quantify the reduction in RGB-D observations resulting from reconstruction-aware observation selection, we record the total number of frames in the input sequence, N_{total}, and the number of keyframes ultimately admitted to the communication queue, N_{upload}. The observation-retention ratio is defined as

$$\rho_{\text{ret}} = \frac{N_{\text{upload}}}{N_{\text{total}}}. \tag{19}$$

A lower ρ_{ret} indicates that fewer RGB-D observations need to be transmitted to the ground server. It should be noted that this metric characterizes the number of observations selected for transmission rather than the actual throughput, latency, or packet-loss performance of the wireless communication link.

4.1.4. Implementation Details

The reconstruction-quality comparisons and ablation experiments were conducted offline on a workstation equipped with an NVIDIA RTX 5090 GPU(NVIDIA Corporation, Santa Clara, CA, USA). Additional onboard runtime and edge-to-PC transmission experiments were conducted on an NVIDIA Jetson Orin NX 16 GB(NVIDIA Corporation, Santa Clara, CA, USA) using the recorded real-UAV sequences.

The principal RAKS parameters used for Replica, OpenFly-UE, and the real-UAV experiments are summarized in Table A1, including coverage, opacity, appearance, rendering-scale, translation, and rotation settings.

4.2. Comparative Reconstruction Performance

Tables 1 and 2 report the quantitative reconstruction results on Replica and the real UAV scenes, respectively. Across the reported experiments, RAOS-GS outperforms the three evaluated multi-agent Gaussian reconstruction methods, namely, MAGiC-SLAM, MAC-Ego3D, and Coko-SLAM, in every listed scene and metric. The consistent trends across both simulated and real data suggest that reconstruction-aware observation selection and the ground-side reconstruction framework can operate under different scene scales and observation conditions, rather than providing improvements only in a particular environment.

Compared with Coko-SLAM, RAOS-GS achieves higher PSNR and SSIM and lower LPIPS in all three Replica scenes. On Apt-0, for example, PSNR increases by 2.23 dB, while LPIPS decreases from 0.183 to 0.023. On the more challenging Apt-2 scene, PSNR increases by 3.07 dB. Similar improvements are observed in the real UAV experiments. In Corridor, the PSNR improvement is 2.80 dB, while RAOS-GS also achieves better SSIM and LPIPS. These results suggest that prioritizing raw observations with greater potential value for subsequent high-quality reconstruction can support ground-side Gaussian scene construction more effectively than strategies centered on keyframe management and communication of local Gaussian submaps.

The comparison with MonoGS using complete observation sequences presents a more nuanced pattern. On Apt-0 and Apt-2 of Replica, RAOS-GS outperforms MonoGS on all three metrics. On Apt-1, however, MonoGS achieves higher PSNR and SSIM, while RAOS-

GS obtains a lower LPIPS. A similar pattern appears in the real scenes. MonoGS achieves higher PSNR on Table and Lecture Hall and slightly better LPIPS, whereas RAOS-GS achieves higher SSIM in all three real scenes and obtains the best PSNR, SSIM, and LPIPS simultaneously in Corridor. Because MonoGS uses complete input sequences while the ground server of RAOS-GS reconstructs the scene using only selected observations, these results indicate that reducing the number of input observations does not necessarily cause a substantial loss of reconstruction quality. They also illustrate the different sensitivities of PSNR, SSIM, and LPIPS to pixel fidelity, structural consistency, and perceptual quality.

Table 1. Quantitative reconstruction comparison on the Replica scenes.

Method	Apt-0			Apt-1			Apt-2		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
MonoGS	34.39 (3.8725)	0.94 (0.0592)	0.140 (0.0842)	33.38 (3.4652)	0.95 (0.0650)	0.14 (0.0990)	30.15 (3.7530)	0.92 (0.0573)	0.095 (0.0820)
MAGiC-SLAM	24.23 (2.319)	0.812 (0.0226)	0.351 (0.1451)	22.43 (3.032)	0.803 (0.0625)	0.319 (0.1942)	25.35 (2.415)	0.793 (0.0903)	0.391 (0.1852)
MAC-Ego3D	26.57 (1.182)	0.814 (0.185)	0.293 (0.124)	28.01 (2.158)	0.825 (0.092)	0.319 (0.175)	29.24 (3.814)	0.849 (0.158)	0.283 (0.108)
Coko-Slam	36.43 (6.138)	0.95 (0.0714)	0.183 (0.1048)	30.52 (5.915)	0.91 (0.0823)	0.085 (0.1124)	31.26 (5.125)	0.93 (0.0791)	0.075 (0.1258)
Ours	38.66 (4.1742)	0.98 (0.0155)	0.023 (0.0271)	32.01 (4.2284)	0.93 (0.0638)	0.070 (0.0666)	34.33 (3.8014)	0.96 (0.0318)	0.044 (0.0399)

Darker cell shading indicates better performance. The arrows indicate the preferred direction for each metric: ↑ means a higher value is better, and ↓ means a lower value is better.

Table 2. Quantitative reconstruction comparison on the real-world UAV scenes.

Method	Table			Lecture Hall			Corridor		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
MonoGS	32.54 (2.72)	0.92 (0.028)	0.163 (0.0239)	33.57 (3.0264)	0.92 (0.0392)	0.177 (0.0679)	28.89 (3.3539)	0.83 (0.0361)	0.25 (0.1643)
MAGiC-SLAM	21.19 (4.14)	0.79 (0.136)	0.23 (0.165)	23.14 (3.92)	0.85 (0.159)	0.25 (0.153)	22.53 (2.15)	0.82 (0.094)	0.36 (0.164)
MAC-Ego3D	24.81 (3.94)	0.86 (0.127)	0.25 (0.093)	26.93 (3.29)	0.83 (0.092)	0.26 (0.103)	23.84 (2.27)	0.75 (0.113)	0.31 (0.126)
Coko-Slam	28.35 (3.35)	0.90 (0.036)	0.192 (0.036)	30.48 (5.25)	0.90 (0.053)	0.209 (0.079)	27.38 (4.23)	0.88 (0.041)	0.234 (0.27)
Ours	30.40 (2.89)	0.93 (0.021)	0.164 (0.0250)	32.06 (2.3940)	0.93 (0.0185)	0.186 (0.0357)	30.18 (2.4539)	0.91 (0.0130)	0.213 (0.128)

Darker cell shading indicates better performance. The arrows indicate the preferred direction for each metric: ↑ means a higher value is better, and ↓ means a lower value is better.

Overall, RAOS-GS achieves better reconstruction metrics than the evaluated multi-agent methods in the reported scenes, whereas its results relative to full-input MonoGS vary across scenes and metrics. The principal finding is that RAOS-GS maintains competitive reconstruction quality using a substantially reduced set of RGB-D observations. The following observation-reduction experiments quantify the input reduction associated with this reconstruction quality.

As shown in Figure 3, RAOS-GS preserves more complete geometry and appearance details across indoor and real-world scenes.

4.3. Observation Reduction Analysis

Table 3 summarizes the numbers of RGB-D observations before and after reconstruction-aware observation selection in each scene. Overall, the proposed method substantially reduces the number of observations entering the air–ground communication and subsequent ground-side reconstruction pipelines. Only approximately 5.1∼7.5% of the original RGB-D frames are retained in the three Replica scenes. The retention ratios in the real UAV Table, Lecture Hall, and Corridor scenes are similarly limited to approximately 6.3∼7.9%. Thus, more than 90% of the continuous observations are classified by the onboard selector as redundant data with limited estimated contribution to the current reconstruction and do not enter the transmission set or the subsequent ground-side reconstruction process.

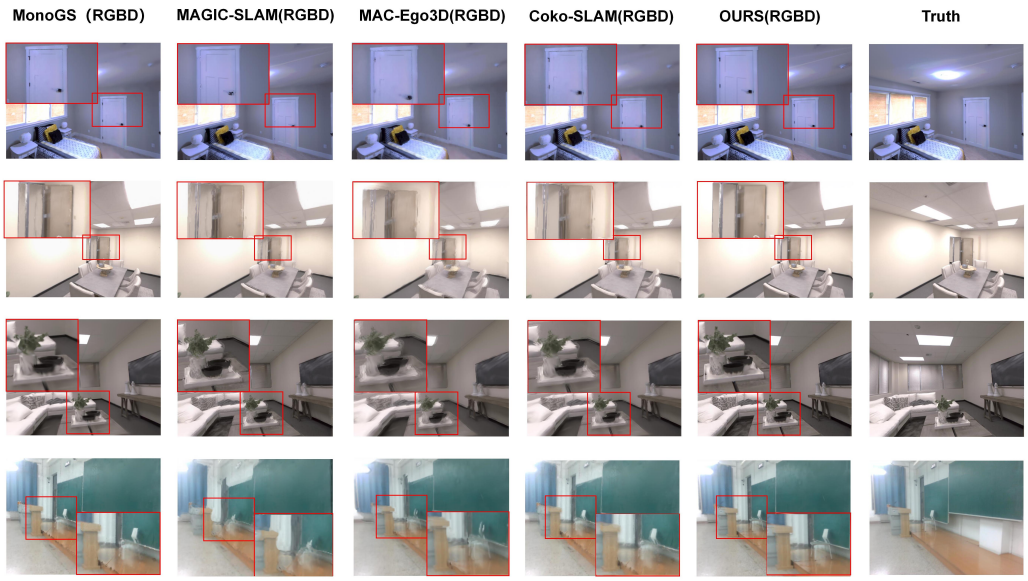


Figure 3. Qualitative comparison of reconstruction results on Replica and the real-world UAV scenes.

Table 3. RGB-D observation retention after reconstruction-aware selection.

Dataset	Scene	Total Frames	Selected Frames	Retention (%)
Replica	Apt-0	20,969	1067	5.1
	Apt-1	18,830	1410	7.5
	Apt-2	20,465	1540	7.5
OpenFly-UE	OpenFly-UE-1	5474	1021	18.7
	OpenFly-UE-2	4725	1133	24.0
Real UAV	Table	7937	628	7.9
	Lecture Hall	18,825	1195	6.3
	Corridor	26,286	1990	7.6

OpenFly-UE exhibits different observation-retention characteristics, with retention ratios of 18.7% and 24.0%, which are substantially higher than those on Replica and the real UAV scenes. The OpenFly-UE data were generated using a higher virtual UAV flight speed. Consequently, the complete trajectories contain fewer RGB-D frames, and neighboring frames exhibit larger spatial and angular changes. This lower temporal redundancy makes candidate observations more likely to contain information that has not been sufficiently covered or explained by the proxy map, causing the reconstruction-aware selector to retain a larger proportion of the observations. This behavior indicates that the RAKS retention ratio depends on the redundancy of the input sequence rather than on a fixed sampling rate.

To distinguish the effect of reducing the number of observations from that of selecting observations with greater reconstruction value, RAKS was compared with uniform temporal sampling. The same ground-side Gaussian reconstruction pipeline was applied after both selection strategies, thereby excluding differences in the backend algorithm. Because the fixed uniform-sampling ratios do not exactly match the adaptive retention ratios produced by RAKS, Table 4 reports, for each scene, the closest tested uniform setting whose retention ratio is no lower than that of the proposed method. Uniform sampling therefore uses the same number or more input observations in these comparisons. Results for all tested sampling rates are provided in Appendix D.

On Replica and OpenFly-UE, RAKS outperforms uniform temporal sampling on all three metrics while retaining fewer observations. Across the three Replica scenes,

RAOS-GS retains 5.1∼7.5% of the observations, compared with 10% for uniform sampling, and improves PSNR by 0.66∼2.21 dB. The differences are more pronounced on OpenFly-UE. With lower retention than the 30% uniform-sampling setting, RAKS improves PSNR by 1.58 dB on UE-1 and 3.27 dB on UE-2.

Table 4. Comparison with uniform temporal sampling. Best results are shown in bold.

		Uniform Temporal Sampling				Ours			
Dataset	**Scene**	**Ret. (%)**	**PSNR ↑**	**SSIM ↑**	**LPIPS ↓**	**Ret. (%)**	**PSNR ↑**	**SSIM ↑**	**LPIPS ↓**
Replica	Apt-0	10.0	36.45	0.94	0.045	5.1	**38.66**	**0.98**	**0.023**
	Apt-1	10.0	31.35	0.90	0.097	7.5	**32.01**	**0.93**	**0.070**
	Apt-2	10.0	32.89	0.93	0.063	7.5	**34.33**	**0.96**	**0.044**
OpenFly-UE	UE-1	30	23.87	0.81	0.245	18.7	**25.45**	**0.85**	**0.135**
	UE-2	30	22.97	0.79	0.268	24.0	**26.24**	**0.84**	**0.115**
Real UAV	Table	10	28.45	0.86	0.193	7.9	**30.40**	**0.93**	**0.164**
	Lecture Hall	10	29.35	0.85	0.236	6.3	**32.06**	**0.93**	**0.186**
	Corridor	10	28.89	0.79	0.359	7.6	**30.18**	**0.91**	**0.213**

The comparisons on the real UAV sequences exhibit a different pattern. Relative to uniform sampling with 10% retention, RAOS-GS improves all three metrics while retaining fewer observations. In Lecture Hall, PSNR increases by 2.71 dB, while LPIPS in Corridor decreases from 0.359 to 0.213. At a 30% retention ratio, uniform sampling achieves PSNR values comparable to those of RAOS-GS, with differences below 0.6 dB in all three scenes. RAOS-GS nevertheless achieves lower LPIPS in Lecture Hall and Corridor. Under the evaluated real observation conditions, the benefit of RAKS is therefore reflected in its ability to achieve reconstruction quality comparable to that of a higher-rate uniform strategy using a substantially smaller observation budget.

Across the three datasets, the final reconstruction quality depends not only on the number of retained observations but also on whether the selected viewpoints complement the current scene representation. Uniform sampling selects observations solely according to their temporal positions in the sequence and cannot distinguish highly redundant viewpoints from observations containing new scene information. In contrast, RAKS evaluates the potential reconstruction value of a candidate frame using the coverage and rendering-explanation capability of the Gaussian proxy map and further applies a viewpoint-redundancy constraint. These differences enable RAKS to achieve comparable or better reconstruction quality with lower retention. Its purpose is to improve the reconstruction relevance of observations transmitted under a limited communication budget rather than merely lower the temporal sampling rate.

4.4. Evaluation on OpenFly-UE Virtual Outdoor Environments

Table 4 reports the reconstruction results of RAOS-GS on the two OpenFly-UE virtual outdoor scenes. Figure 4 shows the corresponding qualitative reconstructions. The two scenes exhibit similar overall rendering quality. OpenFly-UE-2 achieves slightly better PSNR and LPIPS, whereas OpenFly-UE-1 achieves slightly higher SSIM. The three metrics do not vary substantially between the two scenes, suggesting that RAOS-GS maintains relatively stable Gaussian reconstruction performance under outdoor visual conditions that differ considerably from those of the indoor Replica scenes.

The absolute PSNR values of the two outdoor scenes are approximately 9 dB lower than those of the indoor Replica scenes. The larger spatial extent of the outdoor scenes, extensive repeated textures on building facades, and strong illumination contrasts increase the difficulty of fitting the Gaussian appearance parameters. Under the same scene conditions, uniform temporal sampling with 30% retention achieves PSNR values of only 23.87 and

22.97 dB and LPIPS values of 0.245 and 0.268, respectively, all of which are worse than those of RAOS-GS. These results indicate that the lower absolute reconstruction quality is primarily associated with the imaging characteristics of the evaluated scenes rather than being explained solely by the observation-selection strategy.

Although the retention ratios on OpenFly-UE are substantially higher than those on Replica and the real UAV scenes, RAKS still achieves better values on all three metrics using fewer observations than uniform sampling on the same scenes. When the temporal redundancy of an input sequence decreases, RAKS does not maintain a fixed sampling ratio. Instead, it increases the retention ratio to cover regions exhibiting faster viewpoint changes. Its selected observations nevertheless outperform uniformly sampled observations at a lower retention ratio. This behavior is consistent with the previously observed adaptation of the retention ratio to input redundancy and suggests that RAOS-GS remains applicable under virtual outdoor visual conditions that differ substantially from the indoor Replica environments.

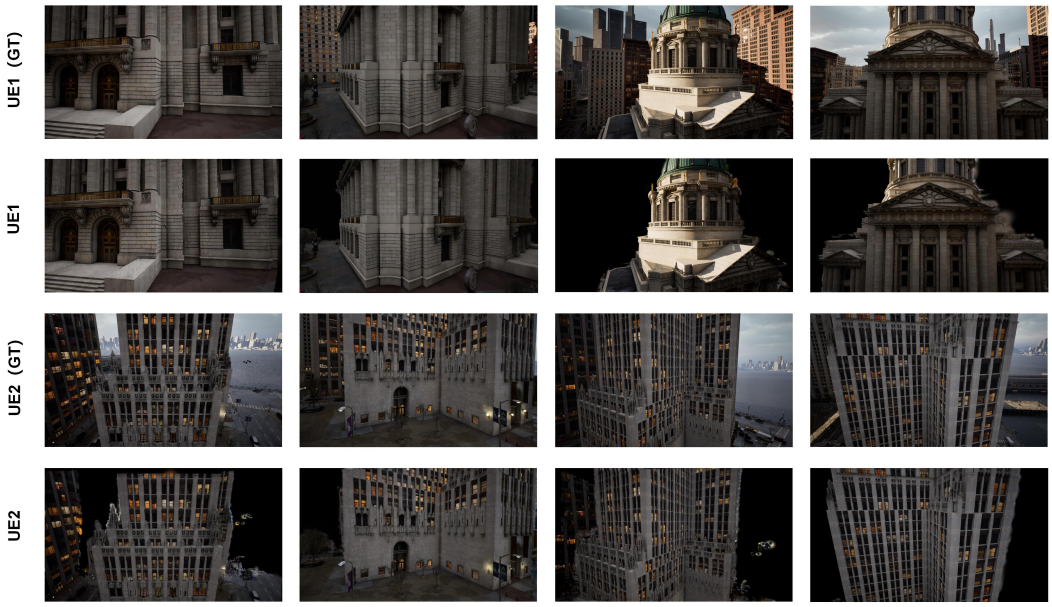


Figure 4. Reconstruction results on the OpenFly-UE virtual outdoor scenes.

4.5. Ablation Study

To analyze the roles of the principal components in the proposed framework, ablation experiments were conducted on the three Replica scenes. Three variants were considered. The first removes the viewpoint-redundancy constraint to examine how the relative-viewpoint-change constraint suppresses redundant observations. The second removes rendering-aware observation selection to assess the contribution of Gaussian-proxy rendering feedback to identifying valuable observations. The third removes ground-side pose–Gaussian refinement to evaluate the effect of final joint refinement on reconstruction quality. The first two variants affect observation selection and therefore report both retention and reconstruction quality. Ground-side refinement does not alter the onboard selection result, so its retention ratio remains identical to that of the complete method.

As shown in Table 5, the three components perform different but complementary functions. Removing the viewpoint-redundancy constraint increases the retention ratio to approximately 45% in all three Replica scenes while causing only limited changes in PSNR, SSIM, and LPIPS. On Apt-0, for example, retention increases from 5.1% to 45.4%, whereas PSNR remains nearly unchanged. These results indicate that the viewpoint-redundancy constraint does not primarily improve rendering quality directly. Instead, it suppresses

spatially redundant observations without substantially degrading reconstruction performance, thereby reducing the amount of data entering the communication and ground-side reconstruction pipelines.

Table 5. Ablation study of the proposed observation-selection and ground-side refinement components on Replica.

Scene	Variant	Retention(%)	PSNR ↑	SSIM ↑	LPIPS ↓
Apt-0	**Full model**	5.1	38.66	0.98	0.023
	w/o viewpoint redundancy constraint	45.4	38.62	0.98	0.022
	w/o rendering-aware selection	9.4	37.18	0.96	0.031
	w/o pose–Gaussian refinement	5.1	28.78	0.87	0.240
Apt-1	**Full model**	7.5	32.01	0.93	0.070
	w/o viewpoint redundancy constraint	45.3	32.46	0.93	0.068
	w/o rendering-aware selection	11.4	30.92	0.92	0.082
	w/o pose–Gaussian refinement	7.5	29.45	0.88	0.230
Apt-2	**Full model**	7.5	34.33	0.96	0.044
	w/o viewpoint redundancy constraint	45.0	34.97	0.95	0.047
	w/o rendering-aware selection	10.7	33.25	0.94	0.051
	w/o pose–Gaussian refinement	7.5	30.36	0.87	0.251

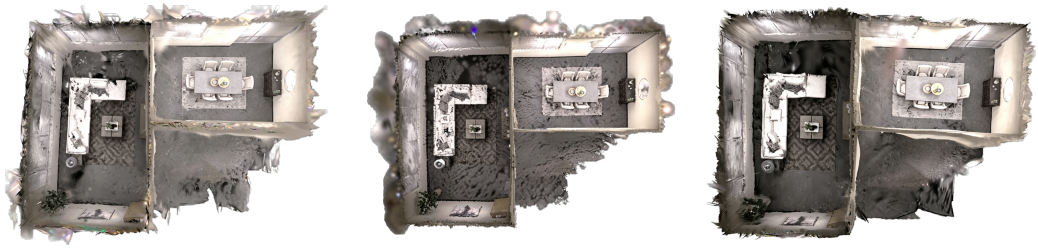
In contrast, removing rendering-aware observation selection degrades reconstruction quality to different degrees in all three Replica scenes. Although the viewpoint-change constraint in this variant can still filter some redundant observations, the system no longer has access to the coverage state and rendering-explanation errors provided by the Gaussian proxy. It therefore cannot effectively determine whether a candidate frame genuinely complements the current scene representation. Viewpoint change alone is consequently insufficient to ensure that the retained observations make substantial contributions to subsequent Gaussian reconstruction.

This behavior is even more evident in the extended ablation experiments. As shown in Appendix D, removing rendering-aware selection further decreases the retention ratio in some real scenes while simultaneously degrading the final reconstruction quality. A lower retention ratio alone therefore does not indicate a more effective communication strategy. The objective is not merely to minimize the number of transmitted frames but to preferentially retain observations that contribute to subsequent reconstruction under a reduced communication requirement. The rendering-aware cues primarily determine which observations are worth retaining, rather than simply increasing or decreasing the number of keyframes.

Ground-side pose–Gaussian refinement has a more substantial effect on the final reconstruction quality. When the admitted observations remain identical to those of the complete method, removing this component degrades PSNR, SSIM, and LPIPS in all three scenes. On Apt-0, PSNR decreases by 9.88 dB. This degradation indicates that cross-UAV observations still require joint refinement of their poses and Gaussian representations after initial coordinate-frame unification. Ground-side refinement is therefore an important component for converting a limited set of valuable observations into a high-quality global Gaussian representation.

Overall, the three principal components have a clear division of responsibilities. Rendering-aware observation selection identifies candidate observations that contribute to subsequent reconstruction. The viewpoint-redundancy constraint further removes spatially redundant views from these candidates. Ground-side pose–Gaussian refinement improves the quality of the final scene representation after observation selection.

Figure 5 qualitatively compares the reconstruction results obtained after removing each major component of RAOS-GS.



(**a**) (**b**) (**c**)

Figure 5. Qualitative reconstruction results for the ablation variants: (**a**) without pose–Gaussian refinement; (**b**) without rendering-aware selection; and (**c**) without the viewpoint-redundancy constraint.

4.6. Discussion

RAOS-GS improves the reconstruction quality–communication trade-off by selecting observations according to their reconstruction relevance rather than their temporal positions alone. Across the three datasets, RAKS maintains or improves reconstruction quality while often retaining fewer observations because it evaluates whether a candidate view contributes information beyond the current proxy representation. Uniform sampling distributes observations according to time and therefore cannot distinguish newly observed content from repeated coverage. By jointly considering proxy-map coverage, appearance discrepancy within reliably rendered regions, and viewpoint change, RAKS prioritizes observations that complement the existing representation. Accordingly, its objective is not to minimize the retention ratio itself but to preserve the information required for downstream reconstruction under a limited communication budget.

Rendering-aware assessment and the viewpoint-redundancy constraint play complementary roles in balancing observation informativeness and spatial redundancy. Removing the viewpoint constraint substantially increases the number of retained observations while producing only limited quality changes in most scenes, indicating that it primarily filters spatially redundant views. In contrast, removing rendering-aware selection degrades reconstruction quality even when the retention ratio becomes lower in some scenes. Fewer observations therefore do not necessarily imply more effective selection. The proxy-map coverage and appearance cues determine which observations are informative, whereas the viewpoint constraint identifies which nearby observations can be safely discarded. Their combination enables RAKS to reduce redundant transmissions without pursuing data reduction at the expense of reconstruction quality.

Separating lightweight onboard observation assessment from high-fidelity ground-side reconstruction is essential to the efficiency and accuracy of RAOS-GS. The Gaussian proxy is maintained only as a lightweight representation of the current local scene state for assessing candidate observations; it is not intended to serve as the final reconstruction. The selected multi-UAV observations are processed at the ground server to construct the final Gaussian representation. Moreover, the consistent degradation observed after removing pose–Gaussian refinement shows that observation selection and initial cross-UAV alignment alone cannot fully eliminate residual pose and map inconsistencies. Observation selection determines which data enter the communication and reconstruction pipelines, while ground-side refinement converts the resulting sparse and imperfectly aligned observations into a coherent global representation. This separation reduces onboard processing and transmission demands while preserving the optimization capacity required for accurate reconstruction.

RAKS adapts its retention behavior to the redundancy and motion characteristics of each sequence rather than enforcing a fixed transmission ratio. Replica and the real indoor sequences contain considerable consecutive-frame redundancy, permitting aggressive

observation reduction. OpenFly-UE contains larger viewpoint changes because of higher virtual flight speeds and lower temporal sampling densities and, therefore, requires a larger fraction of the observations. These decisions depend on the current proxy state and the redundancy of incoming observations. Such scene-dependent scheduling is better aligned with reconstruction needs that vary across environments, motion patterns, and mission stages. We did not compare RAKS with external observation-selection methods using the same reconstruction backend, as SLAM-oriented keyframe selection [19] differs in focus from our objective of reducing transmitted observations while preserving ground-side reconstruction quality. The absence of such controlled comparisons remains a limitation of the current evaluation.

Several limitations remain. Cross-UAV coordinate alignment requires sufficient trajectory overlap and may become unreliable when overlap is limited, scene texture is weak or repetitive, or RGB-D measurements are noisy. Because appearance discrepancy is used as an observation-value cue, dynamic objects, illumination changes, and other non-static variations may be mistaken for previously unseen scene content. The reconstruction-quality experiments were conducted offline with up to three UAVs. Supplementary experiments additionally measured RGB-D transmission volume and edge-to-PC delivery latency, as well as onboard runtime and CUDA memory usage on a Jetson Orin NX 16 GB. Scalability to larger UAV teams, radio-level packet loss, and communication or onboard energy consumption remain to be evaluated. Future work will therefore investigate scalable multi-UAV coordination, online link-aware scheduling, real-network deployment, and more robust scene-change discrimination in dynamic environments.

5. Conclusions

This paper presented RAOS-GS, a reconstruction-aware framework for communication-efficient multi-UAV Gaussian Splatting reconstruction. Each UAV maintains a lightweight Gaussian proxy for Rendering-Aware Keyframe Selection (RAKS), which evaluates scene coverage, reliable-region appearance discrepancy, and viewpoint change to retain informative RGB-D observations. The selected observations are processed at the ground server through cross-UAV registration and pose–Gaussian refinement to obtain a unified global representation.

Experiments on Replica, OpenFly-UE, and real-world UAV sequences demonstrate that RAOS-GS substantially reduces the observations required for transmission and reconstruction while maintaining competitive quality. In the Replica and real-world scenes, only approximately 5.1∼7.9% of the original observations were retained. Comparisons with uniform temporal sampling and the ablation studies confirm that the gains arise from reconstruction-aware selection and the complementary operation of redundancy suppression and ground-side refinement.

Future work will focus on online link-aware scheduling, scalable coordination for larger UAV fleets, real-network and onboard-efficiency evaluation, and robust observation assessment in dynamic scenes.

Author Contributions: Conceptualization, Z.D. and G.J.; methodology, T.M.; software, G.J.; validation, G.J., T.M., and Z.D.; formal analysis, G.J., Z.D., and T.M.; investigation, G.J. and T.M.; resources, G.J. and T.M.; data curation, G.J.; writing—original draft preparation, G.J.; writing—review and editing, Z.D.; visualization, G.J., T.M., and Z.D.; supervision, G.J.; project administration, Z.D.; funding acquisition, Z.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Heilongjiang Provincial Natural Science Foundation of China (No. LH2024F004)

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The synthetic datasets, processing scripts, and processed camera trajectories supporting this study are openly available at https://github.com/jiaguanxin73-debug/Multi-UAV-Gaussian-Splatting, (accessed on 29 August 2026). The real-world datasets are not publicly archived due to large storage requirements. Processed results and metadata for the real-world experiments may be obtained from the corresponding author upon reasonable request.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT5.6 (OpenAI) solely for English translation and language polishing. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. RAKS Parameter Settings

Table A1 summarizes the principal RAKS parameters used for Replica, OpenFly-UE, and the real-UAV experiments. The coverage and appearance settings are shared across the three dataset groups, while the rendering scale and pose-novelty thresholds differ between their experimental configurations.

Table A1. Principal RAOS parameters used in the experiments.

Parameter	Replica	OpenFly-UE	Real-UAV	Role
Minimum rendered coverage	0.75	0.75	0.75	Coverage threshold
Strong-opacity threshold	0.30	0.30	0.30	Valid-region for appearance evaluation
Appearance-score threshold	0.025	0.025	0.025	Appearance-based selection threshold
Appearance score	RGB residual P90, weights 0.60 RGB/0.40 gradient	Same	Same	Appearance-discrepancy measure
Rendering score scale	1.0	0.5	1.0	Resolution scale for candidate-view evaluation
Translation novelty threshold	0.10 m	0.04 m	0.10 m	Pose-novelty threshold
Rotation novelty threshold	8.0°	3.0°	5.0°	Pose-novelty threshold

Appendix B. Communication and Onboard Performance Measurements

Table A2 reports the transmission results for all nine trajectories in the three real-UAV scenes. In addition to the number and proportion of transmitted RGB-D observations, it presents the total transmitted data volume and mean edge-to-PC delivery latency for each trajectory.

Table A2. Actual transmission results on the three real-UAV scenes.

Real-UAV Scene	Trajectory	Input Frames	Transmitted Frames	Retention/Transmission Ratio	Total Transmitted Data (MB)	Mean Edge-to-PC Delivery Latency (ms)
Table	UAV1	2786	202	7.3%	28.39	21.14
	UAV2	2652	217	8.2%	32.29	25.62
	UAV3	2499	209	8.4%	31.79	24.45
Lecture Hall	UAV1	5085	339	6.7%	53.12	23.28
	UAV2	5161	314	6.1%	49.48	24.87
	UAV3	8579	542	6.3%	80.67	21.78
Corridor	UAV1	10,060	556	5.5%	69.69	22.04
	UAV2	8861	666	7.5%	93.19	27.93
	UAV3	7365	768	10.4%	118.25	25.65

Table A3 reports the onboard processing measurements obtained on the Jetson Orin NX 16 GB using the recorded real-UAV sequences. Per-frame time and processing rate describe the complete onboard pipeline, while tracking time and peak CUDA allocated and reserved memory provide additional detail on its computational cost.

Table A3. Onboard runtime and CUDA-memory measurements on the three real-UAV scenes.

Real-UAV Scene	Trajectory	Avg. Frame Time (ms)	P95 Frame Time (ms)	Processing Rate (FPS)	Avg. Tracking Time (ms)	Peak CUDA Allocated (MB)	Peak CUDA Reserved (MB)
Table	UAV1	247.31	440.43	4.04	45.03	1023.94	1086
	UAV2	273.43	461.23	3.65	46.25	979.32	1136
	UAV3	272.38	459.32	3.67	44.83	935.84	1256
Lecture Hall	UAV1	263.89	460.50	3.79	50.42	1588.58	2158
	UAV2	283.45	477.18	3.53	48.90	1684.20	2274
	UAV3	292.10	481.91	3.42	51.65	2033.75	2468
Corridor	UAV1	301.75	503.21	3.31	38.99	3909.05	4572
	UAV2	315.32	513.46	3.17	40.31	2308.52	4150
	UAV3	309.45	504.68	3.23	39.97	2035.36	4320

Appendix C. Detailed Observation Reduction Statistics

Table A4 reports the detailed per-UAV observation statistics before and after reconstruction-aware selection. The retention ratio is defined as the number of selected RGB-D observations divided by the total number of input observations for each UAV trajectory.

Table A4. Per-UAV RGB-D observation statistics before and after reconstruction-aware selection.

Dataset	Scene/UAV	Total	Selected	Retention (%)	Reduction (%)
Real UAV	Table/UAV1	2786	202	7.3	92.7
	Table/UAV2	2652	217	8.2	91.8
	Table/UAV3	2499	209	8.4	91.6
	Lecture Hall/UAV1	5085	339	6.7	93.3
	Lecture Hall/UAV2	5161	314	6.1	93.9
	Lecture Hall/UAV3	8579	542	6.3	93.7
	Corridor/UAV1	10,060	556	5.5	94.5
	Corridor/UAV2	8861	666	7.5	92.5
	Corridor/UAV3	7365	768	10.4	89.6
OpenFly-UE	UE1/UAV1	1777	383	21.6	78.4
	UE1/UAV2	1848	321	17.4	82.6
	UE1/UAV3	1849	317	17.1	82.9
	UE2/UAV1	1448	381	26.3	73.7
	UE2/UAV2	1580	370	23.4	76.6
	UE2/UAV3	1697	382	22.5	77.5
Replica	Apt-0/UAV1	7099	328	4.6	95.4
	Apt-0/UAV2	7447	344	4.6	95.4
	Apt-0/UAV3	6423	395	6.1	93.9
	Apt-1/UAV1	5888	522	8.9	91.1
	Apt-1/UAV2	6920	641	9.3	90.7
	Apt-1/UAV3	6022	247	4.1	95.9
	Apt-2/UAV1	7006	583	8.3	91.7
	Apt-2/UAV2	6679	539	8.1	91.9
	Apt-2/UAV3	6780	418	6.2	93.8

Appendix D. Extended Ablation Results

To complement the main ablation study on Replica, we further evaluatethe same variants on the two OpenFly-UE scenes and the three real-world UAV sequences. The complete results are reported in Table A5. The overall trends are consistent with those observed on Replica: removing the viewpoint redundancy constraint substantially increases the number of retained observations while producing only limited changes in rendering quality; removing the rendering-aware selection generally degrades reconstruction performance; and disabling ground-side pose–Gaussian refinement leads to the most pronounced deterioration in the final Gaussian representation.

Tables A6–A8 report the complete results for all tested uniform temporal sampling rates.

Table A9 reports the ATE RMSE of the final camera trajectories estimated by RAOS-GS on the three Replica scenes.

Table A5. Extended ablation results on OpenFly-UE and real-world UAV sequences.

Scene	Variant	Retention (%)	PSNR ↑	SSIM ↑	LPIPS ↓
OpenFly-UE-1	**Full model**	18.7	25.45	0.85	0.135
	w/o viewpoint redundancy constraint	70.6	26.36	0.84	0.138
	w/o rendering-aware selection	35.4	24.38	0.81	0.143
	w/o pose–Gaussian refinement	18.7	21.38	0.80	0.218
OpenFly-UE-2	**Full model**	24.0	26.24	0.84	0.115
	w/o viewpoint redundancy constraint	79.2	26.79	0.85	0.126
	w/o rendering-aware selection	43.4	25.13	0.83	0.114
	w/o pose–Gaussian refinement	24.0	21.13	0.78	0.238
Table	**Full model**	7.9	30.40	0.93	0.164
	w/o viewpoint redundancy constraint	41.4	30.15	0.93	0.165
	w/o rendering-aware selection	5.8	28.38	0.88	0.220
	w/o pose–Gaussian refinement	7.9	26.35	0.82	0.280
Lecture Hall	**Full model**	6.3	32.06	0.93	0.186
	w/o viewpoint redundancy constraint	52.2	32.55	0.93	0.179
	w/o rendering-aware selection	3.1	29.97	0.91	0.230
	w/o pose–Gaussian refinement	6.3	25.46	0.81	0.290
Corridor	**Full model**	7.6	30.18	0.91	0.213
	w/o viewpoint redundancy constraint	54.6	30.29	0.92	0.245
	w/o rendering-aware selection	7.8	27.39	0.89	0.290
	w/o pose–Gaussian refinement	7.6	22.35	0.77	0.350

Table A6. Detailed comparison with uniform temporal sampling on Replica.

Scene	Strategy	Retention (%)	PSNR ↑	SSIM ↑	LPIPS ↓
Apt-0	Uniform temporal sampling	30.0	37.97	0.97	0.034
	Uniform temporal sampling	10.0	36.45	0.94	0.045
	Ours	5.1	38.66	0.98	0.023
Apt-1	Uniform temporal sampling	30.0	30.96	0.91	0.071
	Uniform temporal sampling	10.0	31.35	0.90	0.097
	Ours	7.5	32.01	0.93	0.070
Apt-2	Uniform temporal sampling	30.0	33.99	0.96	0.053
	Uniform temporal sampling	10.0	32.89	0.93	0.063
	Ours	7.5	34.33	0.96	0.044

Table A7. Detailed comparison with uniform temporal sampling on OpenFly-UE.

Scene	Strategy	Ret. (%)	PSNR ↑	SSIM ↑	LPIPS ↓
UE-1	Uniform temporal sampling	30	23.87	0.81	0.245
	Uniform temporal sampling	10	20.52	0.75	0.382
	Ours	18.7	25.45	0.85	0.135
UE-2	Uniform temporal sampling	30	22.97	0.79	0.268
	Uniform temporal sampling	10	19.34	0.62	0.405
	Ours	24.0	26.24	0.84	0.115

Table A8. Detailed comparison with uniform temporal sampling on the real-world UAV sequences.

Scene	Strategy	Retention (%)	PSNR ↑	SSIM ↑	LPIPS ↓
Table	Uniform temporal sampling	30	30.97	0.93	0.155
	Uniform temporal sampling	10	28.45	0.86	0.193
	Ours	7.9	30.40	0.93	0.164
Lecture Hall	Uniform temporal sampling	30	32.18	0.92	0.192
	Uniform temporal sampling	10	29.35	0.85	0.236
	Ours	6.3	32.06	0.93	0.186
Corridor	Uniform temporal sampling	30	30.45	0.92	0.247
	Uniform temporal sampling	10	28.89	0.79	0.359
	Ours	7.6	30.18	0.91	0.213

Table A9. Absolute trajectory error of the proposed method on Replica.

Dataset	Apt-0	Apt-1	Apt-2
ATE RMSE (m)	0.08474	0.04743	0.08327

Appendix E. Qualitative Reconstruction Results

This section presents the three-dimensional Gaussian reconstruction results produced by RAOS-GS on all eight evaluation scenes, including three Replica scenes, two OpenFly-

UE scenes, and three real-world UAV scenes. The qualitative Replica results are shown in Figure A1.

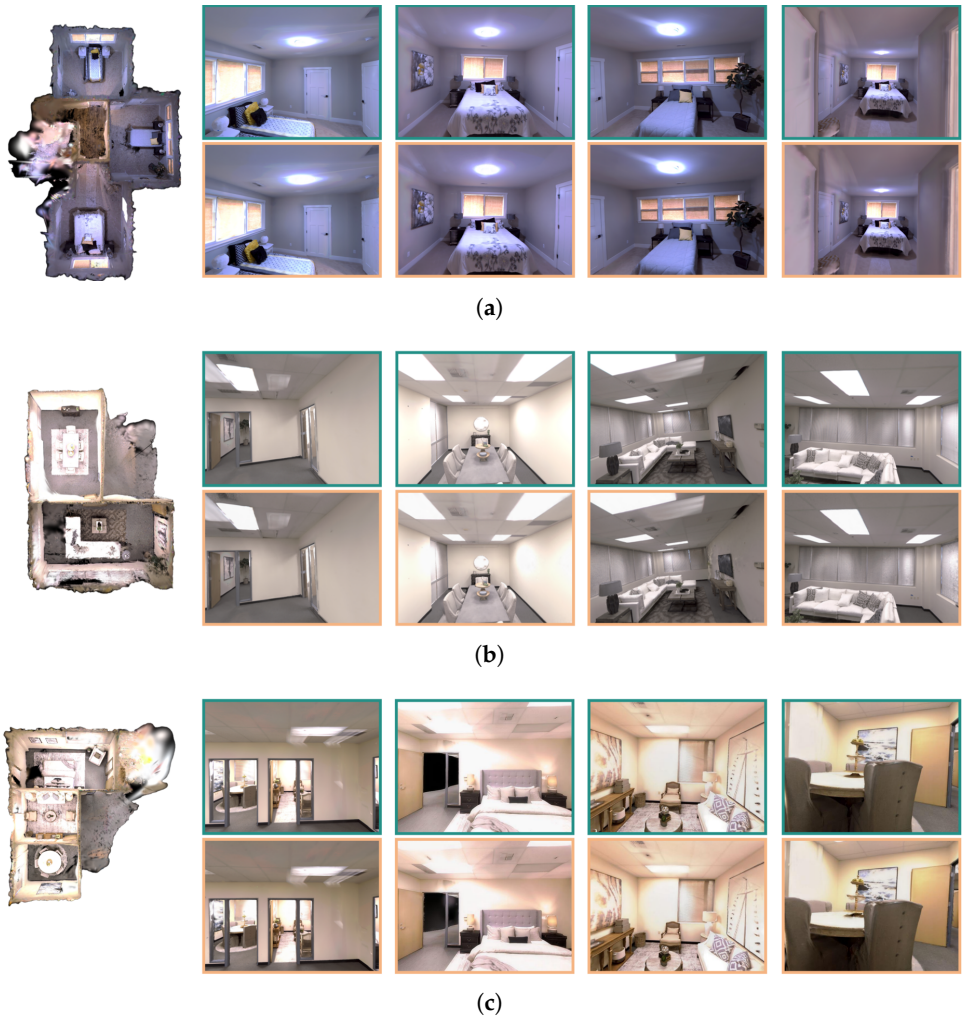


Figure A1. Qualitative reconstruction results of RAOS-GS on the Replica scenes: (**a**) Apt-0; (**b**) Apt-1; and (**c**) Apt-2.

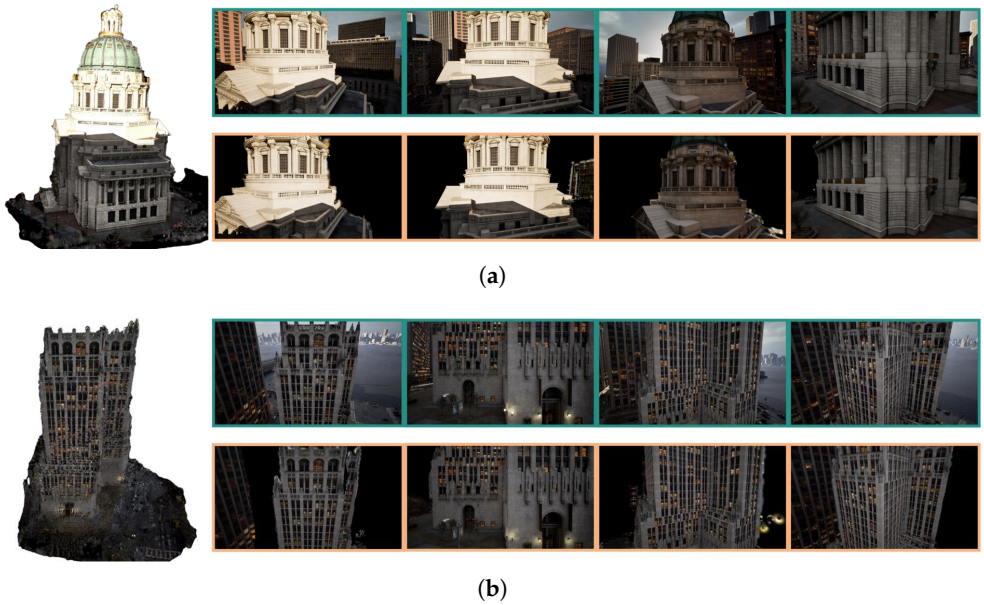


Figure A2. Qualitative reconstruction results of RAOS-GS on the OpenFly-UE scenes: (**a**) OpenFly-UE-1; and (**b**) OpenFly-UE-2.

https://doi.org/10.3390/drones10110828



Figure A3. Qualitative reconstruction results of RAOS-GS on the real-world UAV scenes: (**a**) Table; (**b**) Lecture Hall; and (**c**) Corridor.

Appendix F. Reconstruction Results with Two and Three UAVs

To examine the effect of UAV team size, we compare reconstructions using UAV1 and UAV2 with those using all three UAV trajectories in one representative scene from each dataset group: Replica Apt-0, OpenFly-UE UE1, and the real-UAV Lecture Hall scene. Both configurations are evaluated using PSNR, SSIM, and LPIPS, as reported in Table A10. The results are comparable across the tested two- and three-UAV configurations; this comparison does not evaluate teams larger than three UAVs or systematically vary trajectory overlap.

Table A10. Reconstruction results with two and three UAVs on representative scenes.

Dataset	Trajectories	UAV Number	PSNR	SSIM	LPIPS
Apt-0	UAV1 and UAV2	2	38.40	0.97	0.021
	All	3	38.66	0.98	0.023
UE1	UAV1 and UAV2	2	25.30	0.84	0.128
	All	3	25.45	0.85	0.135
Lecture Hall	UAV1 and UAV2	2	31.98	0.93	0.192
	All	3	32.06	0.93	0.186

References

1. Kerbl, B.; Kopanas, G.; Leimkühler, T.; Drettakis, G. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* **2023**, *42*, 139.
2. Keetha, N.; Karhade, J.; Jatavallabhula, K.M.; Yang, G.; Scherer, S.; Ramanan, D.; Luiten, J. Splatam: Splat, track & map 3d gaussians for dense rgb-d slam. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2024; pp. 21357–21366.
3. Huang, H.; Li, L.; Cheng, H.; Yeung, S.K. Photo-slam: Real-time simultaneous localization and photorealistic mapping for monocular, stereo, and rgb-d cameras. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2024; pp. 21584–21593.
4. Yugay, V.; Gevers, T.; Oswald, M.R. Magic-slam: Multi-agent gaussian globally consistent slam. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2025; pp. 6741–6750.
5. Xu, X.; Xue, F.; Zhao, S.; Pan, Y.; Scherer, S.; Huang, X. Mac-ego3d: Multi-agent gaussian consensus for real-time collaborative ego-motion and photorealistic 3d reconstruction. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2025; pp. 854–863.
6. Chen, L.; Su, Y.; Wang, J.; Han, P.; Xia, Z.; Bu, S.; Li, K.; Hu, B.; Meng, S.; Wang, G. CoMA-SLAM: Collaborative Multi-Agent Gaussian SLAM with Geometric Consistency. In *Proceedings of the AAAI Conference on Artificial Intelligence*; AAAI Press: Palo Alto, CA, USA, 2026; Volume 40, pp. 2922–2929.
7. Li, M.M.; Lajoie, P.Y.; Liu, J.; Beltrame, G. Compact Keyframe-Optimized Multi-Agent Gaussian Splatting SLAM. *arXiv* **2026**, arXiv:2604.00804.
8. Cao, Z.; Shao, Q.; Zhai, S.; Zhang, J.; Nguyen, A.; Wang, H.; Huang, B. MAGS-SLAM: Monocular Multi-Agent Gaussian Splatting SLAM for Geometrically and Photometrically Consistent Reconstruction. *arXiv* **2026**, arXiv:2605.10760.
9. Mur-Artal, R.; Tardós, J.D. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Trans. Robot.* **2017**, *33*, 1255–1262.
10. Yan, C.; Qu, D.; Xu, D.; Zhao, B.; Wang, Z.; Wang, D.; Li, X. Gs-slam: Dense visual slam with 3d gaussian splatting. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2024; pp. 19595–19604.
11. Matsuki, H.; Murai, R.; Kelly, P.H.; Davison, A.J. Gaussian splatting slam. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2024; pp. 18039–18048.
12. Yugay, V.; Li, Y.; Gevers, T.; Oswald, M.R. Gaussian-slam: Photo-realistic dense slam with gaussian splatting. *arXiv* **2023**, arXiv:2312.10070.
13. Peng, Z.; Shao, T.; Liu, Y.; Zhou, J.; Yang, Y.; Wang, J.; Zhou, K. Rtg-slam: Real-time 3d reconstruction at scale using gaussian splatting. In *Proceedings of the ACM SIGGRAPH 2024 Conference Papers*; ACM: New York, NY, USA, 2024; pp. 1–11.
14. Schmuck, P.; Chli, M. CCM-SLAM: Robust and efficient centralized collaborative monocular simultaneous localization and mapping for robotic teams. *J. Field Robot.* **2019**, *36*, 763–781.
15. Schmuck, P.; Ziegler, T.; Karrer, M.; Perraudin, J.; Chli, M. Covins: Visual-inertial slam for centralized collaboration. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*; IEEE: Piscataway, NJ, USA, 2021; pp. 171–176.
16. Tian, Y.; Chang, Y.; Arias, F.H.; Nieto-Granda, C.; How, J.P.; Carlone, L. Kimera-multi: Robust, distributed, dense metric-semantic slam for multi-robot systems. *IEEE Trans. Robot.* **2022**, *38*, 2022–2038.
17. Hu, J.; Mao, M.; Bao, H.; Zhang, G.; Cui, Z. Cp-slam: Collaborative neural point-based slam system. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 39429–39442.
18. Xuan, P.N.; Canh, T.N.; Nguyen, H.H.; Chong, N.Y.; HoangVan, X. A survey on collaborative SLAM with 3D Gaussian splatting. *arXiv* **2025**, arXiv:2510.23988.
19. Thorne, D.; Chan, N.; Ma, Y.; Robison, C.S.; Osteen, P.R.; Lopez, B.T. Submodular optimization for keyframe selection & usage in SLAM. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*; IEEE: Piscataway, NJ, USA, 2025; pp. 5033–5039.
20. DeTone, D.; Malisiewicz, T.; Rabinovich, A. Superpoint: Self-supervised interest point detection and description. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*; IEEE: Piscataway, NJ, USA, 2018; pp. 337–33712.
21. Lindenberger, P.; Sarlin, P.E.; Pollefeys, M. Lightglue: Local feature matching at light speed. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: Piscataway, NJ, USA, 2023; pp. 17581–17592.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.