Active Gaussian Splatting for Autonomous Reconstruction Using D-Optimal Next Best View Planning

Kanav Prashar, Carlos Torre, Rodney Staggers Jr., Bharath Vedantha Desikan, Jnaneshwar Das
DREAMS Lab, Arizona State University, Tempe, AZ, USA

***Abstract*—Autonomous 3D reconstruction of geological features is a critical capability for planetary surface exploration, where communication latency prohibits manual view selection. Conventional approaches survey the environment exhaustively before selecting informative viewpoints, incurring unnecessary observation overhead. We propose an active Gaussian Splatting framework in which a drone autonomously selects the Next Best View (NBV) to maximise reconstruction quality of a target rock. Our approach assumes that an initial survey yields a high-fidelity point cloud of the target, from which a mesh model is derived and used to initialise 100,000 3D Gaussians on the surface. Gaussian positions are frozen at mesh-surface locations, and only appearance parameters (colour, scale, rotation, and opacity) are learned online from incrementally acquired images. Four evenly-spaced seed images initialise the model, after which, at each step, a diagonal Fisher Information Hessian is accumulated over the current training set and the candidate viewpoint that maximises the D-optimal information gain is selected as the next observation. We further provide a theoretical justification showing that convergence between steps is a *mathematical necessity* for the Taylor linearisation underlying the Fisher criterion, not merely an engineering heuristic.**

I. Introduction

When exploring extraterrestrial terrain, transmitting compressed mesh models back to Earth is insufficient for scientific analysis and mission planning. Agents operating on planetary surfaces must build rich, queryable 3D representations of their surroundings autonomously and efficiently. Three-dimensional Gaussian Splatting (3DGS) [1] has opened new possibilities in this direction: a well-trained 3DGS model supports real-time novel view synthesis, improving situational awareness for remote operators and enabling downstream tasks such as hazard detection, sample selection, and volumetric mapping. Unlike volumetric neural representations such as NeRF [2], 3DGS uses an explicit set of primitives whose parameters are directly amenable to gradient-based uncertainty analysis.

Planetary exploration, however, imposes severe operational constraints. Communication bandwidth between the surface and Earth is limited to a few kilobits per second, and agents (whether rovers or UAVs) carry finite battery reserves. Both constraints require that every captured image contributes maximal information towards the construction of the 3D-GS model. Autonomous planetary science [3] has long identified this as a core challenge, and it makes the *Next Best View* (NBV) problem central: the agent must decide, at each step, either which image to select from a pre-collected pool or from which pose to capture the next image, so that the uncertainty of the 3DGS model decreases as rapidly as possible.

To address this, we formulate NBV selection as a D-optimal Fisher information problem over the parameters of a mesh-initialised 3DGS model. Our experimental setup mirrors a realistic planetary scenario: a UAV orbits a rock along a fixed circular trajectory, yielding a dataset of 100 images and their associated camera poses. Starting from four evenly-spaced seed images, the system greedily selects the 50 most informative viewpoints from the remaining 96 candidates based on the Fisher information each would add to the model. This is equivalent to constraining the drone to a fixed observation budget; in practice the stopping criterion could be a target uncertainty level or a maximum information gain threshold, but here we demonstrate the 50-step regime to characterise the full active selection trajectory.

Prior active-reconstruction work for NeRF [4] and for Gaussian Splatting [5]–[8] relies on Fisher information, uncertainty quantification, or coverage metrics to score candidate views. None of these works, however, (i) exploit a coarse geometric mesh prior to initialize and constrain Gaussian positions, (ii) provide a formal analysis of when the Fisher approximation is theoretically valid, or (iii) target the constrained UAV-orbit scenario typical of planetary science.

This paper makes the following contributions:

1) **Mesh-prior 3DGS for active reconstruction.** We show that freezing Gaussian positions on an orbital mesh prior dramatically reduces the learnable parameter space, enabling reliable Fisher Hessian estimation from as few as four seed images.
2) **Theoretical validity analysis of Fisher NBV.** We show that the Taylor linearisation underlying the D-optimal information-gain criterion is self-consistent only when the model has converged between steps, and we introduce *surface coverage* as a diagnostic for prior quality.

The remainder of the paper is organised as follows. Section II surveys related work. Section III derives the D-optimal NBV criterion and analyses its validity conditions: Section IV describes the simulation setup. Section V presents results and discussion. Section VI concludes with future directions.

II. Related Work

A. 3D Gaussian Splatting

Kerbl *et al.* [1] introduced 3DGS, representing scenes as collections of anisotropic 3D Gaussians with learnable position, covariance, colour (via spherical harmonics), and opacity. Differentiable tile-based rasterisation enables real-time rendering and fast per-parameter gradient computation via backpropagation, making 3DGS attractive for online robotic reconstruction. Each Gaussian $\mathbf{g}_j$ contributes to a pixel through alpha-compositing of sorted depth-ordered splats: $C = \sum_n \mathbf{c}_n \alpha_n \prod_{j<n}(1-\alpha_j)$, where the contribution α_n is computed from the projected 2D covariance. Despite its rendering quality and explicit parameter structure, standard 3DGS provides no native measure of uncertainty over its parameters, a prerequisite for any principled active-view policy. Our work exploits the explicit parameter structure of 3DGS to compute a diagonal Fisher Hessian from which view informativeness is derived.

B. POp-GS: P-Optimality for 3D Gaussian Splatting

Wilson *et al.* [6] (CVPR 2025) place the 3DGS active-reconstruction problem within the classical theory of P-optimal experimental design. Starting from a maximum-likelihood formulation in which pixel errors are modelled as IID Gaussian, they derive the parameter covariance $\mathbf{\Sigma} = \sigma_e^2(\mathbf{J}^\top\mathbf{J})^{-1}$ via a first-order Taylor expansion around the converged model $\boldsymbol{\theta}_*$, and show that uncertainty can be characterised by the P-optimality functional $U_p(\mathbf{\Sigma}) = \left(\frac{1}{l}\operatorname{tr}(\mathbf{\Sigma}^p)\right)^{1/p}$. D-optimality ($p \to 0$, volume of the covariance ellipsoid) and T-optimality ($p = 1$, average variance) achieve the best performance on the Mip-NeRF360 and Blender benchmarks, outperforming both FisherRF and uniform sampling. A-optimality ($p = -1$) performs worst, consistent with its divergence for unobserved Gaussians ($H_{jj} \approx 0$). A critical property of POp-GS that we inherit is *pose-only scoring*: the Jacobian $\mathbf{J}_i$ is obtained by rendering from the current model at candidate pose $\mathbf{p}_i$ without consulting the ground-truth image, enabling deployment in scenarios where the drone has not yet physically visited the candidate viewpoint.

Our work builds on POp-GS but differs in two key respects. First, POp-GS operates on a fully trained unconstrained 3DGS model; our setting begins with only four seed images, where Gaussian positions are under-constrained and the Fisher information criterion cannot be meaningfully applied from the outset. Freezing Gaussian positions on a mesh prior sidesteps this instability and enables uncertainty-driven view selection from the very first step. Second, POp-GS targets general scene reconstruction with dense training images; our setting starts from four seed views, making the initialisation condition geometrically critical and the choice of prior non-trivial.

C. ActiveSplat: Active Gaussian Mapping

Li *et al.* [9] (RA-L 2025) propose ActiveSplat, an autonomous high-fidelity reconstruction system that maintains a live Gaussian map updated online from posed RGB-D input and couples it with a sparse Voronoi graph (a topology classically used for NBV planning [10]) for workspace abstraction and path planning. Viewpoint selection is driven by a view-dependent completeness score: Gaussians are splatted onto candidate panoramas and the fraction of covered area serves as the information metric. A hierarchical planning strategy traverses the Voronoi graph to extend the frontier while avoiding redundant trajectories, balancing coverage efficiency with reconstruction accuracy. ActiveSplat demonstrates high completion ratios ($> 93\%$) in room-scale environments within a limited frame budget.

The key difference from our approach is the information metric. ActiveSplat uses geometric coverage (rendered opacity/completeness) to drive exploration in an *unknown* environment. We target a *known-geometry* object (a rock whose mesh prior is given) and maximise the D-optimal Fisher information gain over appearance parameters, providing a principled uncertainty reduction guarantee rooted in statistical estimation theory. Furthermore, ActiveSplat allows Gaussian positions to float freely during online updates, which would violate the Taylor validity condition we establish in Section III-C; our frozen-position design is specifically motivated by this analysis.

III. Methodology

A. Scene Representation and Mesh Prior

The agent first performs an autonomous survey of the scene. It follows a circular orbit while capturing RGB-D images and accumulating a dense point cloud via onboard SLAM. Upon detecting a loop closure the agent returns to its home position, at which point the survey point cloud is considered complete. The raw point cloud contains both the ground plane and the target rock. We apply RANSAC to segment and remove the dominant planar surface, then DBSCAN to cluster the remaining points and isolate the rock cluster. The isolated cluster is passed to Poisson surface reconstruction [11], which fits a smooth, watertight mesh $\mathcal{M}$ to the noisy point set. The mesh is a smoothed approximation of the underlying surface geometry: small-scale sensor noise is absorbed while large-scale shape is preserved.

Given $\mathcal{M}$, we sample N points uniformly on the surface. The sampling density is governed by a target inter-point spacing δ, so $N \approx A_{\mathcal{M}}/\delta^2$, where $A_{\mathcal{M}}$ is the total mesh area. This gives direct control over the number of Gaussians: a finer spacing yields more Gaussians and a higher-fidelity representation, at the cost of increased memory and computation. In our experiments we use $N = 100{,}000$.

Each sampled point initialises one 3D Gaussian $\mathbf{g}_j$:

$$\mathbf{g}_j = (\boldsymbol{\mu}_j, \mathbf{c}_j, \mathbf{s}_j, \mathbf{q}_j, o_j), \tag{1}$$

where $\boldsymbol{\mu}_j \in \mathbb{R}^3$ is the position (frozen throughout training), $\mathbf{c}_j \in \mathbb{R}^3$ the RGB colour, $\mathbf{s}_j \in \mathbb{R}^3$ the log-scale, $\mathbf{q}_j \in \mathbb{R}^4$ the rotation quaternion, and $o_j \in \mathbb{R}$ the opacity logit. The learnable parameter vector is therefore $\boldsymbol{\theta} = (\mathbf{c}_j, \mathbf{s}_j, \mathbf{q}_j, o_j)_{j=1}^N \in \mathbb{R}^{11N}$.

Initial colours are assigned by projecting each Gaussian onto the four seed images: for each Gaussian visible in a seed view (depth-consistent and within the camera frustum), the corresponding pixel colour is sampled and averaged across all visible seed cameras. Gaussians not visible in any seed image are initialised to a neutral grey.

B. Fisher Information and View Scoring

Maximum likelihood formulation. Following POp-GS [6], let $h(\boldsymbol{\theta})$ denote the 3D-GS rendering function mapping parameters to a vectorised pixel image $\mathbf{c} \in \mathbb{R}^P$, $P = HW$. The per-image residual is $\mathbf{e} = \mathbf{c} - h(\boldsymbol{\theta})$. Assuming pixel errors are zero-mean IID Gaussian, $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma_e^2 \mathbf{I})$, the likelihood is $p(\mathbf{c}|\boldsymbol{\theta}) = \exp\left(-\frac{1}{2\sigma_e^2}\mathbf{e}^\top\mathbf{e}\right)$, and maximising it is equivalent to minimising

$$-\sigma_e^2 \log p(\mathbf{c}|\boldsymbol{\theta}) = \tfrac{1}{2}\,\mathbf{e}^\top\mathbf{e}. \tag{2}$$

Taylor expansion and the information matrix. Given a current estimate $\boldsymbol{\theta}_*$, expanding h to first order yields $\mathbf{e} \approx \mathbf{e}_* - \mathbf{J}\,\Delta\boldsymbol{\theta}$, where $\mathbf{e}_* = \mathbf{c} - h(\boldsymbol{\theta}_*)$, $\Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}_*$, and

$$\mathbf{J} \;\triangleq\; \left.\frac{\partial h}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}_*} \;\in\; \mathbb{R}^{P\times 11N}. \tag{3}$$

Substituting into (2) and applying first-order optimality gives the Gauss–Newton update $\Delta\boldsymbol{\theta} = (\mathbf{J}^\top\mathbf{J})^{-1}\mathbf{J}^\top\mathbf{e}_*$. Assuming $\boldsymbol{\theta}_*$ is unbiased ($\mathbb{E}[\mathbf{e}_*] = \mathbf{0}$), the parameter covariance is

$$\mathbb{E}[\Delta\boldsymbol{\theta}\,\Delta\boldsymbol{\theta}^\top] = \sigma_e^2\,(\mathbf{J}^\top\mathbf{J})^{-1} = \sigma_e^2\,\mathbf{H}^{-1}, \tag{4}$$

where $\mathbf{H} \triangleq \mathbf{J}^\top\mathbf{J} \in \mathbb{R}^{l\times l}$, $l = 11N$, is the *information matrix*. A large H_{jj} means parameter θ_j is well-constrained by the current training data; $H_{jj} \approx 0$ means it is essentially unobserved.

Uncertainty decrease due to an added image. Given the prior Hessian $\mathbf{H}_-$ accumulated from the current training set $\mathcal{D}$, adding a candidate image i captured at pose $\mathbf{p}_i$ yields a Jacobian $\mathbf{J}_i = \partial h/\partial\boldsymbol{\theta}|_{\boldsymbol{\theta}_*,\mathbf{p}_i}$ and updates the Hessian as [6]:

$$\mathbf{H}_i = \mathbf{H}_- + \mathbf{J}_i^\top\mathbf{J}_i. \tag{5}$$

The best view is the one that maximally reduces the uncertainty of the posterior covariance $\boldsymbol{\Sigma}_i = \mathbf{H}_i^{-1}$:

$$\arg\max_i\ \mathcal{I}(\mathbf{H}_i) = \arg\min_i\ U(\boldsymbol{\Sigma}_i) = \arg\min_i\ U(\mathbf{H}_i^{-1}). \tag{6}$$

D-optimal information gain. We adopt the D-optimal criterion from the theory of optimal experimental design [12], which POp-GS [6] empirically identifies as the best-performing design among classical optimality criteria for Gaussian Splatting NBV. D-optimality minimises the volume of the posterior covariance ellipsoid (equivalently, maximising the log-determinant of $\mathbf{H}_i$) and connects to the Laplace approximation of the parameter posterior [13]. The competing A-optimal criterion $\frac{1}{l}\sum_j[1/(H_{-,jj}+\lambda) - 1/(H_{i,jj}+\lambda)]$ diverges whenever $H_{-,jj} \approx 0$ (which holds for any Gaussian not yet observed), making it numerically unstable early in the active phase.

Since $\mathbf{H} \in \mathbb{R}^{11N\times 11N}$ is intractable to store, we retain only its diagonal. The update (5) then reduces to $H_{i,jj} = H_{-,jj} + J_{ij}^2$, and the D-optimal information gain factorises as:

$$\mathrm{IG}_D(i) = \frac{1}{l}\sum_{j=1}^{l}\log\left(1 + \frac{J_{ij}^2}{H_{-,jj} + \lambda_\theta}\right), \tag{7}$$

where $H_{-,jj} = \sum_{k\in\mathcal{D}}(\partial L_k/\partial\theta_j)^2$ is accumulated by backward passes over the training set and λ_θ is a small regularisation constant. The logarithm saturates naturally for unobserved Gaussians, avoiding the divergence of A-optimality while preserving the correct ranking among candidates.

The NBV is selected greedily:

$$c^* = \underset{i\,\in\,\mathcal{C}\setminus\mathcal{D}}{\operatorname{argmax}}\ \mathrm{IG}_D(i), \tag{8}$$

requiring one forward-backward pass per candidate, with $O(|\mathcal{C}|)$ cost with $O(11N)$ memory.

Pose-only scoring. A critical property of (7) is that $\mathbf{J}_i$ is computed by rendering from the *current model* at pose $\mathbf{p}_i$; the ground-truth image at that pose is never used during scoring. As noted by POp-GS [6]: *"the contents of the i-th image are not necessary for our formulation, only the pose at which we plan to take the image from."* In our implementation, $\mathbf{J}_i$ is obtained by a single forward-backward pass $h(\boldsymbol{\theta}_*; \mathbf{p}_i)$ with a dummy photometric loss; no ground-truth pixel is consulted. This makes the NBV policy deployable in a true active setting: the drone requires only a set of candidate *poses*; it need not have physically captured those images before selecting which one to visit next.

C. Taylor Validity Without a Conventional Prior

What POp-GS requires. The entire information-gain framework rests on the Taylor expansion (3): $h(\boldsymbol{\theta}) \approx h(\boldsymbol{\theta}_*) + \mathbf{J}\,\Delta\boldsymbol{\theta}$. This is accurate only when $\boldsymbol{\theta}_*$ is a *true local minimum* of the training loss, i.e., the model has genuinely converged. At a non-converged $\boldsymbol{\theta}_*$, the Jacobian $\mathbf{J}_i$ reflects transient gradient directions rather than stable rendering sensitivity, and IG_D rankings become arbitrary. POp-GS implicitly assumes this condition is met by starting from a fully-trained 3D-GS model.

The problem with unconstrained 3D-GS and few images. In standard 3D-GS, $\boldsymbol{\theta}$ includes Gaussian positions $\boldsymbol{\mu}_j$. With only four seed images, positions drift aggressively toward the seen cameras to minimise photometric loss, and the optimiser finds a degenerate minimum that geometrically over-fits the seed views. The Jacobian evaluated at this collapsed $\boldsymbol{\theta}_*$ accurately describes the gradient for seen views but gives no reliable signal for unseen poses. The covariance $\mathbf{H}^{-1}$ does not represent the true parameter uncertainty; it represents the curvature of a distorted loss landscape. The "good prior" condition is violated, and IG_D may be unreliable.

Why our formulation restores Taylor validity. By initialising Gaussians on the Poisson-reconstructed mesh $\mathcal{M}$ and *freezing* $\boldsymbol{\mu}_j$, we remove positions from the optimisation entirely. The learnable space $\boldsymbol{\theta} = (\mathbf{c}_j, \mathbf{s}_j, \mathbf{q}_j, o_j)_{j=1}^{N}$ is the *appearance subspace only*. Because positions are geometrically correct from the outset, the loss landscape over $\boldsymbol{\theta}$ is smooth and well-conditioned: there is no mechanism for the model to collapse onto the seed views. Convergence to a genuine local minimum is therefore achievable from as few as four seed images, and the Taylor linearisation holds at $\boldsymbol{\theta}_*$.

Concretely, a Gaussian j that lies within the frustum of at least one seed view will have $H_{jj} > 0$ after seed training; its appearance has converged locally and its Jacobian J_{ij} faithfully encodes how visible it is from any candidate pose $\mathbf{p}_i$. We formalise how much of the surface has reached this state via *surface coverage*:

$$C_0 = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}[H_{jj} > \epsilon]\,, \tag{9}$$

the fraction of Gaussians with sufficient gradient signal. The four seed images at $0°, 90°, 180°, 270°$ are chosen to maximise C_0: wide azimuthal separation ensures the majority of the rock surface is observed from step zero, giving IG_D a large, reliable base from which to score candidates. Gaussians with $H_{jj} \approx 0$ lie on the unobserved patches, and those are precisely the targets the NBV policy will steer toward next, bootstrapping C_0 upward with every added view.

Corner case: single-seed initialisation. Although we use four seed images in practice, it is instructive to analyse the single-seed case ($|\mathcal{D}_0| = 1$) formally. Let $\mathcal{V}_0 \subset \{1, \ldots, N\}$ denote the set of M Gaussians that fall within the frustum of the single seed, $M \ll N$. After training on this one image, the diagonal Hessian satisfies

$$H_{jj} > 0 \;\; \forall j \in \mathcal{V}_0, \qquad H_{jj} \approx 0 \;\; \forall j \notin \mathcal{V}_0, \tag{10}$$

so the information matrix is non-degenerate only on the $11M$-dimensional subspace corresponding to observed Gaussians. Partitioning $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\mathcal{V}_0}, \boldsymbol{\theta}_{\bar{\mathcal{V}}_0})$ and the Jacobian accordingly, $\mathbf{J} = \big[\mathbf{J}_{\mathcal{V}_0} \;\big|\; \mathbf{0}\big]$, the full information matrix takes the block form

$$\mathbf{H} = \mathbf{J}^\top \mathbf{J} = \begin{array}{c} \left[\begin{array}{c|c} \mathbf{J}_{\mathcal{V}_0}^\top \mathbf{J}_{\mathcal{V}_0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \end{array} \begin{array}{l} {\scriptstyle 11M} \\ {\scriptstyle 11(N-M)} \end{array}, \tag{11}$$

where the top-left block $\mathbf{J}_{\mathcal{V}_0}^\top \mathbf{J}_{\mathcal{V}_0} \in \mathbb{R}^{11M \times 11M}$ is positive semi-definite (full rank once M provides sufficient view diversity), and the bottom-right block is exactly zero; unobserved parameters carry no gradient signal whatsoever.

Now consider evaluating IG_D for a candidate pose $\mathbf{p}_i$ with visible set $\mathcal{V}_i$. The sum in (7) splits into two regimes:

$$\begin{aligned} \mathrm{IG}_D(i) = &\underbrace{\frac{1}{l} \sum_{j \in \mathcal{V}_i \cap \mathcal{V}_0} \log\Big(1 + \tfrac{J_{ij}^2}{H_{-,jj} + \lambda}\Big)}_{\text{observed (Taylor valid)}} \\ &+ \underbrace{\frac{1}{l} \sum_{j \in \mathcal{V}_i \setminus \mathcal{V}_0} \log\Big(1 + \tfrac{J_{ij}^2}{\lambda}\Big)}_{\text{unobserved (saturated)}}. \end{aligned} \tag{12}$$

The first term is reliable: the Taylor expansion holds over $\mathcal{V}_0$ and the score meaningfully discriminates among candidates that overlap with the trained region. For candidates *near* the seed ($\mathcal{V}_i \approx \mathcal{V}_0$), the first term dominates and D-optimality correctly ranks viewpoints that would further refine the already-observed surface.

The second term is degenerate: unobserved Gaussians ($j \notin \mathcal{V}_0$) have not been trained, so their appearance parameters remain at grey initialisation and their opacity logits carry no signal from any real image. Because these parameters have never been touched by a gradient step, $J_{ij} \approx 0$ for any candidate i whose visible set lies entirely outside $\mathcal{V}_0$. Consequently the second term collapses to ≈ 0, and so does the entire score:

$$\mathcal{V}_i \cap \mathcal{V}_0 = \emptyset \implies \mathrm{IG}_D(i) \approx 0. \tag{13}$$

The NBV policy therefore assigns *zero information gain* to all candidates that view only unobserved surface patches, and positive gain only to candidates that overlap with the already-seen region $\mathcal{V}_0$. The direct consequence is that the drone *stays pinned near the first seed image*: it will refine the already-observed surface rather than explore the rest of the rock, leaving the far side permanently unvisited within any finite budget.

This is the principal motivation for our four-seed initialisation. By placing seeds at $0°, 90°, 180°, 270°$, we ensure $\mathcal{V}_0 \approx \{1, \ldots, N\}$, eliminating the zero-IG subspace entirely and making (7) reliable from the early steps.

IV. Experimental Setup

A. Simulation Environment

Experiments are conducted in Gazebo Classic with PX4 and ROS 2. A photo-realistic rock model (GLB mesh) is placed in a flat arena. An X500 quadrotor UAV equipped with a simulated RGB-D camera executes a circular orbit around the rock at a fixed radius and altitude. Images are captured at a fixed rate of one frame every 0.5 s, yielding a dataset of 100 keyframes spanning the full orbit. At the moment of each capture the exact 6-DoF camera pose is recorded from the PX4 ground-truth state estimator; no additional pose noise is added.

B. Mesh Reconstruction from RGB-D Data

Prior to the active phase, the same drone performs a dense mapping pass using RTAB-Map [14] with the onboard RGB-D camera. RTAB-Map performs simultaneous loop-closure detection and pose-graph optimisation, producing a globally consistent, dense point cloud of the scene. The rock cluster is extracted via RANSAC (ground removal) and DBSCAN (cluster isolation), and the resulting point cloud is passed to Poisson surface reconstruction to yield the mesh $\mathcal{M}$ used for Gaussian initialisation (Section IIIA).

C. NBV Evaluation Protocol

Because all 100 images and their poses are available offline, both methods are evaluated in simulation replay mode: at each active step, the scoring pass renders from the *current model* at every candidate pose using only pose information (no ground-truth pixel is consulted), selects the best candidate, and immediately retrieves the corresponding pre-captured image for training. This mirrors a real deployment where the drone would physically fly to the selected pose and capture a new image; the offline protocol simply removes flight-time overhead to allow rapid comparative evaluation. PSNR and SSIM [15] are computed by rendering from the trained model and comparing against the held-out ground-truth images.

V. Results and Discussion

A. Spatial Coverage and Information Gain

Figure 1 shows all accumulated camera positions at step 50 overlaid on the 3DGS reconstruction of the rock. The 50 NBV-selected cameras encircle the full azimuth of the rock rather than clustering near the four seed positions, confirming that the D-optimal criterion drives the drone away from already-observed regions. This spatial spread is not a design choice but an emergent consequence of the Fisher criterion: once a surface patch is well observed, its Gaussian subspace carries low uncertainty, and the IG_D score for nearby views collapses. The policy is therefore naturally repelled from revisiting the same angles.

Figure 2 provides the complementary quantitative evidence. IG_D of the selected view decays monotonically from 0.091 at step 1 to ≈ 0.002 at step 50. Crucially, the decay is *smooth* throughout – there is no early plateau followed by a sharp drop, which would indicate the criterion had stagnated and was revisiting redundant viewpoints. Instead, every selected view contributes genuine information, consistent with the spatial spread observed in Figure 1. Taken together, the two figures confirm that the D-optimal criterion achieves global exploration of the rock surface without explicit coverage constraints.

B. Reconstruction Quality

Table I summarises the final reconstruction quality of both methods, evaluated by rendering all 100 available viewpoints and comparing against ground-truth images. Both methods

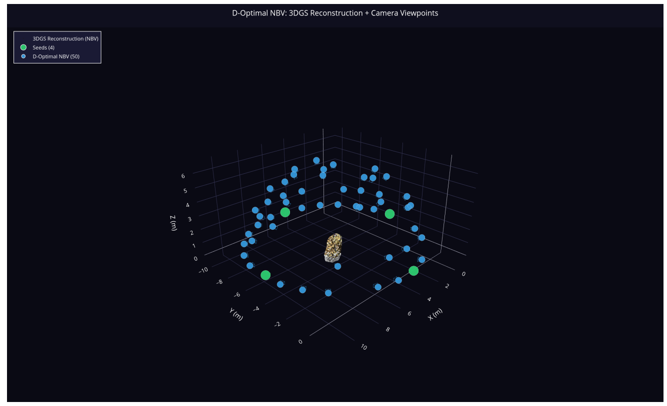

Fig. 1. 3DGS reconstruction and NBV-selected camera positions at step 50. Green dots mark the four seed cameras; blue dots mark the 50 D-optimal NBV cameras. Cameras encircle the full azimuth of the rock, confirming global exploration with no local-minima trapping.

TABLE I
Reconstruction quality evaluated on all 100 viewpoints.

Method	#Images	PSNR (dB)	SSIM
D-Optimal NBV (ours)	54	25.97	0.9654
Coverage-Greedy	50	26.06	0.9648

operate under the same 50-image selection budget on top of the four seed views. D-optimal NBV achieves 25.97 dB PSNR and 0.9654 SSIM; coverage-greedy achieves 26.06 dB PSNR and 0.9648 SSIM. The two are within 0.09 dB of each other.

The key distinction is not in final PSNR but in what each criterion measures at each selection step. As shown in Figure 3, the coverage-greedy criterion reaches 98.17 % surface coverage with the four seed images alone; the remaining 46 images add only 1.28 % additional geometric coverage. From step 5 onward, the marginal geometric contribution of each additional image falls below 1 %, leaving the coverage criterion with limited signal to guide further selection. That comparable PSNR is still achieved suggests those additional views contributed useful appearance variation, though not necessarily because the criterion identified them as such.

The D-optimal criterion operates differently. As shown in Figure 2, IG_D remains non-zero across all 50 steps, decaying smoothly from 0.091 to 0.002. Every image NBV selects is chosen because it measurably reduces uncertainty in the learned appearance parameters. The 50-image budget is a practical constraint, not a consequence of the criterion running out of signal.

Figure 4 shows a qualitative comparison on a representative frame. The original 3DGS model (d) achieves the highest per-frame PSNR, though this should be interpreted with care: `kf_0041` is one of the 100 images the model was trained on, so the score reflects a seen view rather than a novel one. The visual difference between (b), (c), and (d) is relatively modest – the rock shape, surface texture, and shadow are

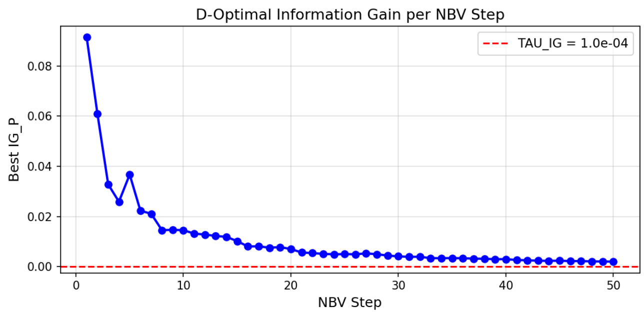

Fig. 2. D-optimal information gain IG_D of the view selected at each active step. The smooth monotone decay from 0.091 to $\approx$0.002 confirms that the criterion identifies genuinely novel viewpoints at every step and approaches an asymptotic minimum as global surface coverage saturates.

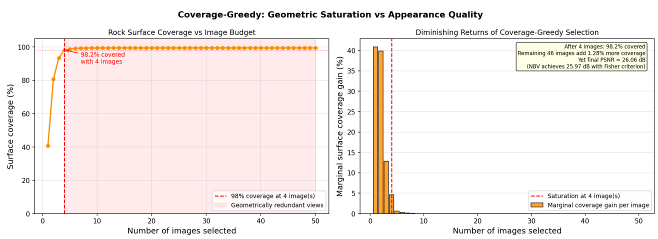

Fig. 3. Marginal surface coverage gain per image under coverage-greedy selection. After the four seed views, each additional image contributes less than 1 % new geometric coverage, leaving the criterion without meaningful signal for the remaining 46 selections.

broadly recovered by both active methods using roughly half the images. This is consistent with the observation that fixing Gaussian positions on a mesh prior, while restrictive, may be sufficient to learn a reasonable appearance model from a limited set of views, and that geometric initialisation can partially offset a reduced image budget.

VI. Conclusion

We presented an active Gaussian Splatting framework for autonomous planetary rock reconstruction. Fixing Gaussian positions on a coarse mesh prior enables reliable Fisher Hessian estimation from sparse seed images. The D-optimal criterion selects views that maximally reduce posterior uncertainty, which we formally show requires sequential retraining to convergence. Over 50 simulation steps, D-optimal NBV and coverage-greedy achieve comparable reconstruction quality (25.97 dB vs. 26.06 dB PSNR). However, our principled D-optimal information gain remains actively non-zero across all 50 steps, whereas the geometric coverage criterion saturates after four seed views.

Future work. We identify three primary extensions:

1) **Incremental Hessian updates** to reduce computational overhead.
2) **Seed count ablation** to characterise initial coverage sensitivity.
3) **Real-hardware deployment** to validate under physical sensor noise.

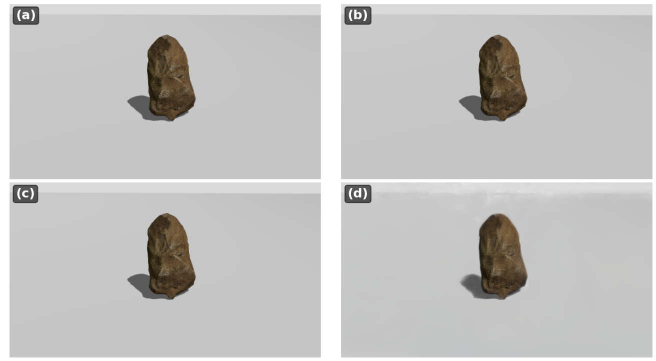

Fig. 4. Qualitative comparison on frame `kf_0041`. (a) Ground truth. (b) D-Optimal NBV (54 images, PSNR = 28.50 dB, SSIM = 0.9530). (c) Coverage-Greedy (50 images, PSNR = 26.90 dB, SSIM = 0.9480). (d) Original 3DGS trained on all 100 images for 15k iterations (PSNR = 34.63 dB, SSIM = 0.9792). Note that (d) is evaluated on a training frame and reflects memorisation rather than generalisation.

References

[1] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3D Gaussian splatting for real-time radiance field rendering," in *ACM Transactions on Graphics (SIGGRAPH)*, vol. 42, no. 4, 2023, pp. 139:1–139:14.

[2] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," in *European Conference on Computer Vision (ECCV)*. Springer, 2020, pp. 405–421.

[3] I. A. D. Nesnas, M. Hentges, J. Muller *et al.*, "Autonomy for space robots: Past, present, and future," in *Current Robotics Reports*, vol. 2, 2021, pp. 217–229.

[4] X. Pan, Z. Lai, S. Song, and G. Huang, "ActiveNeRF: Learning where to see for active sensing," in *European Conference on Computer Vision (ECCV)*. Springer, 2022, pp. 59–75.

[5] W. Jiang, B. Lei, and K. Daniilidis, "FisherRF: Active view selection and uncertainty quantification for radiance fields using fisher information," in *European Conference on Computer Vision (ECCV)*, 2024.

[6] J. Wilson, M. Almeida, S. Mahajan, M. Labrie, M. Ghaffari, O. Ghasemalizadeh, M. Sun, C.-H. Kuo, and A. Sen, "POp-GS: Next best view in 3D Gaussian splatting with P-optimality," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

[7] K. Jun-Seong, T.-H. Oh, E. Pérez-Pellitero, and Y. Jang, "Sa-resgs: Self-augmented residual 3d gaussian splatting for next best view selection," 2026. [Online]. Available: https://arxiv.org/abs/2601.03024

[8] B. Lee, J. Park, K. T. Giang, and S. Song, "Online 3d gaussian splatting modeling with novel view selection," 2025. [Online]. Available: https://arxiv.org/abs/2508.14014

[9] Y. Li, Z. Kuang, T. Li, Q. Hao, Z. Yan, G. Zhou, and S. Zhang, "ActiveSplat: High-fidelity scene reconstruction through active Gaussian splatting," *IEEE Robotics and Automation Letters*, 2025.

[10] C. Connolly, "The determination of next best views," in *IEEE International Conference on Robotics and Automation (ICRA)*, 1985, pp. 432–435.

[11] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," in *Eurographics Symposium on Geometry Processing*, 2006, pp. 61–70.

[12] F. Pukelsheim, *Optimal Design of Experiments*. Philadelphia, PA: SIAM, 2006.

[13] D. J. C. MacKay, "A practical Bayesian framework for backpropagation networks," *Neural Computation*, vol. 4, no. 3, pp. 448–472, 1992.

[14] M. Labbé and F. Michaud, "RTAB-Map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation," *Journal of Field Robotics*, vol. 36, no. 2, pp. 416–446, 2019.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.