Hybrid Sparse-Model Next-Best-View Planning for Active 3D Reconstruction for Digital Twins

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

***Abstract*—We present a next-best-view (NBV) pipeline for active 3D reconstruction with an RGB UAV in simulation that plans directly from an incrementally updated Structure-from-Motion (SfM) sparse point cloud, avoiding dense or learned representations during flight. The system initializes with a structured seed acquisition and then operates in a closed-loop adaptive phase where candidate viewpoints are generated, filtered, and ranked using sparse-model-aware scoring. We evaluate two base scorers: co-visibility and baseline-aware repair, and introduce hybrid switching strategies that combine their complementary behaviors. On a simulated lunar rock inspection task, the oracle hybrid achieves the strongest ground-truth-referenced fidelity across all metrics, while the heuristic hybrid reveals that the switching criterion requires further refinement. The resulting dense point cloud constitutes a lightweight digital twin of the target object, enabling downstream inspection without re-flying the mission. These results support the use of online sparse SfM as an efficient planning state for image-budgeted digital twin construction within autonomous robotic systems. Code is available at: https://github.com/Mulleky/Hybrid-Sparse-Model-Next-Best-View-Planning-for-Active-3D-Reconstruction**

I. Introduction

Active 3D reconstruction requires performing actions that improve geometric understanding under constrained missions. Many recent NBV methods rely on volumetric maps, meshes, or learned scene representations such as Neural Radiance Fields (NeRF) [1], [2] and 3D Gaussian Splatting (3DGS) [3]. These approaches can be highly effective, but often require heavier online scene representations or scene-specific optimization. In contrast, we pursue a lightweight photogrammetry-native pipeline that plans directly from an incrementally updated Structure-from-Motion (SfM) point cloud [4].

NBV methods for 3D reconstruction include classical uncertainty-reduction strategies, geometric rule-based planners, volumetric and surface-based methods, and more recent learning-based policies [5]. Classical model-improvement methods select the next view by estimating which observation most improves an incomplete reconstruction. Surface and occupancy-based planners instead reason over coverage, frontier expansion [6], visibility, or expected information gain.

Our approach is organized as a single closed-loop mission as seen in Fig. 1. It first acquires a structured seed set of 24 views around a target lunar rock sample to bootstrap sparse reconstruction. It then enters an adaptive phase in which candidate viewpoints are generated, filtered by feasibility and motion constraints, and ranked by their expected utility for improving the current sparse model. Dense reconstruction is deferred to a post-flight stage. We study two base sparse-model scorers that expose a coverage-versus-geometry tradeoff, and introduce hybrid switching strategies that combine co-visibility with baseline-aware repair cues. The dense point cloud produced at mission end constitutes a digital twin of the target object, synchronized to the physical world via the captured image set and supporting inspection tasks without requiring repeated physical access.

II. Problem Statement

We consider active reconstruction of an object of interest with a monocular aerial platform in simulation. The platform is given an approximate target center and must acquire a limited number of images that maximize reconstruction utility. At each adaptive step, the planner must choose the next viewpoint using only the partial model available online. No dense mesh, voxel occupancy map [7], or learned scene prior is assumed during flight.

The planner must improve the reconstruction where it is weak while avoiding redundant observations and respecting practical motion constraints. In particular, it must balance observing already reconstructed structure from informative baselines, repairing under-supported regions of the sparse model, maintaining directional and spatial diversity in selected views, and limiting unnecessary travel.

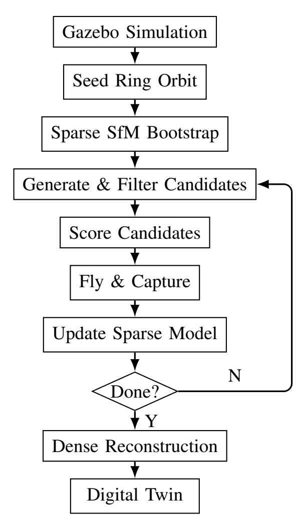

Fig. 1. System pipeline. The adaptive loop (Phase 2) iterates until the image budget or KNN convergence threshold is reached. Then, COLMAP dense reconstruction is performed

III. Methodology

The system is implemented as a unified PX4–ROS2–Gazebo mission with a PX4 X500 quadcopter equipped with a gimbal RGB camera and with online COLMAP updates [4]. Gazebo serves as a digital twin of the deployment environment, enabling safe validation of the full autonomy stack before physical deployment. During Phase 1, the vehicle executes a stratified ring orbit at radius 4 m around the target, spanning four elevation rings (15°, 22°, 29°, 36°) with 8, 6, 6, and 4 azimuths per ring, capturing 24 seed images to bootstrap COLMAP.

During Phase 2, the planner generates candidate viewpoints using a two-band stratified pool at radius 3.5 m around the target. A dense lateral band spans four lower elevation rings (10°, 15°, 25°, 35°) with viewpoints distributed using inverse-rank weighting, concentrating more candidates at the lowest rings where side geometry is most likely to remain under-observed. A sparser upper band covers three higher elevation rings (35°, 45°, 55°) to preserve overhead support. Candidates are filtered by altitude limits, maximum travel distance, and minimum angular spacing. Phase 2 terminates when the 95th-percentile k-nearest-neighbor (KNN) distance in the sparse cloud drops below a convergence threshold, or a hard image budget is reached, whichever occurs first. Fig. 2 shows the stratified seed sampling and NBV candidate generation and filtering.

At each iteration, the full candidate pool is regenerated and re-filtered against the current UAV position and the set of previously visited viewpoints, ensuring that revisit exclusion and travel constraints reflect the evolving mission state.

A. Scorer Formulation

All scorers share the same structure. For a candidate viewpoint c, the score is:

$$S(c) = W_{\text{main}} M_{\text{main}}(c) + W_{\text{nov}}\, \text{nov}(c) - W_{\text{move}}\, \text{move}(c) - W_{\text{ang}}\, \text{ang}(c) \tag{1}$$

with weights W_{main}=1.0, W_{nov}=1.2, W_{move}=0.1, W_{ang}=0.4, shared across all scorers. The three shared terms are defined as follows.

The *novelty* term rewards candidates far from previously visited viewpoints:

$$\text{nov}(c) = \min\left(1, \frac{\min_j \|c - v_j\|_2}{d_{\text{scale}}}\right), \quad d_{\text{scale}} = 2\,\text{m} \tag{2}$$

The *movement cost* penalizes large displacements from the current pose:

$$\text{move}(c) = 0.7\,\frac{\|\Delta \boldsymbol{x}\|}{d_{\text{travel}}} + 0.3\,\frac{|\Delta\psi|}{\pi}, \quad d_{\text{travel}} = 10\,\text{m} \tag{3}$$

The *angular penalty* ang(c) linearly penalizes any candidate whose viewing direction falls within $\theta_{\min}$=20° of an already-visited direction:

$$\text{ang}(c) = \max\left(0,\ 1 - \frac{\min_j \angle(c,\, v_j)}{\theta_{\min}}\right) \tag{4}$$

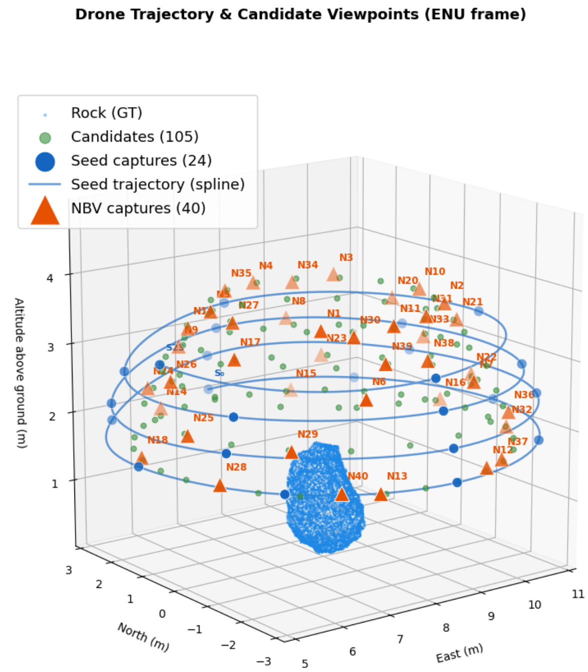

Fig. 2. Stratified seed ring and adaptive candidate viewpoint pool in the ENU frame. Blue markers show the 24 Phase 1 seed captures used to bootstrap the sparse SfM model and orange markers show the selected NBV captures around the ground-truth rock geometry

where $\angle(c, v_j)$ denotes the angle between the viewing direction of candidate c and that of previously visited viewpoint v_j, each defined as the unit vector from the viewpoint position toward the target center.

All scorers differ only in how M_{main} is defined. Let $\mathcal{P}$ be all tracked SfM points within a 2 m half-extent bounding box around the target, and $\mathcal{P}(c) \subseteq \mathcal{P}$ the subset visible from candidate c via frustum projection.

Co-visibility (Covis): The main term is the fraction of tracked points visible from the candidate:

$$M_{\text{main}}^{\text{Covis}}(c) = \frac{|\mathcal{P}(c)|}{|\mathcal{P}|} \tag{5}$$

This rewards viewpoints that observe the broadest slice of existing structure, preserving geometric support across the model.

Baseline-aware repair (B-aware): Each point p is assigned a weakness score $w_p \in [0,1]$ combining inverse track support and local density sparsity:

$$w_p = \alpha_\tau\, \tilde{\tau}_p + \alpha_\rho\, \tilde{\rho}_p \tag{6}$$

with α_τ=0.6, α_ρ=0.4. Here $\tilde{\tau}_p$=clip$\left(\frac{T^* - \tau_p}{T^*}, 0, 1\right)$ with target track length T^*=6, and $\tilde{\rho}_p$ is the percentile-normalized mean 5-nearest-neighbor distance. Each visible point's contribution is further weighted by a geometry gain $g(p,c) \in [0,1]$

measuring the angular novelty of the candidate's ray to p relative to existing observation rays:

$$g(p, c) = \text{clip}\left(\frac{\theta_{\min}(p, c) - \theta_\ell}{\theta^* - \theta_\ell}, 0, 1\right) \tag{7}$$

where $\theta_{\min}(p, c)$ is the minimum angle between the candidate's ray to p and any existing observation ray to p, θ_ℓ=5° is the redundancy floor, and θ^*=25° is the saturation angle. The main term is:

$$M_{\text{main}}^{\text{B-aware}}(c) = \frac{\sum_{p \in \mathcal{P}(c)} w_p \, g(p, c)}{\sum_{p \in \mathcal{P}} w_p} \tag{8}$$

where the denominator sums over all tracked points $\mathcal{P}$, which is non-empty during Phase 2 by construction since COLMAP must succeed before scoring begins. This rewards candidates that observe fragile regions from angles that are genuinely informative for triangulation, rather than directions already well-covered by existing cameras.

B. Hybrid Scorers

Both hybrid strategies are *hard switches*: at each iteration they select one scorer's complete candidate ranking unchanged rather than blending scores. The underlying scorers are always Covis and B-aware.

Hybrid-Heuristics (Hyb-H) uses current sparse model statistics to decide at each iteration. Let f_b=$N_{\text{used}}/N_{\text{budget}}$ be the budget fraction consumed and d_{95} the 95th-percentile KNN distance in the sparse cloud:

$$\text{scorer} = \begin{cases} \text{B-aware} & f_b < 0.4 \text{ or } d_{95} > 0.14\,\text{m} \\ \text{Covis} & \text{otherwise} \end{cases} \tag{9}$$

The gate thresholds ($f_b < 0.4$, $d_{95} > 0.14\,\text{m}$) were designed to balance mission time between both scorers, allocating the early sparse phase to geometry-focused repair and the later dense phase to co-visibility.

Hybrid-Oracle (Hyb-O) uses ground-truth mesh supervision to select the scorer at each iteration. It simulates which scorer's top-ranked candidate would cover more uncovered GT surface samples (sampled at 30,000 points, coverage threshold 0.02 m), and selects that scorer's full ranking. Because it requires access to the GT mesh during mission execution, Hyb-O is not deployable; it establishes the performance ceiling for any data-driven switching policy.

The sparse model is updated after each adaptive capture, and dense reconstruction is performed offline once the stopping criterion is met.

IV. Results

Prior to metric evaluation, the dense cloud is cleaned offline: background points are removed via a bounding box crop around the target, the ground plane is removed using RANSAC [8], and statistical outlier removal (SOR) is applied. This yields the cleaned cloud used for all ground-truth comparisons.

TABLE I
DIGITAL TWIN FIDELITY METRICS FOR EACH NBV SCORING STRATEGY. THRESHOLDS OF 48 MM AND 80 MM CORRESPOND TO APPROXIMATELY 3% AND 5% OF THE ROCK HEIGHT (≈1.6 M).

Metric	Covis	B-aware	Hyb-H	Hyb-O
Dense cloud vs. GT				
Hausdorff ↓ (mm)	480.9	522.6	535.1	**465.1**
Mean C2C ↓ (mm)	66.0	79.4	79.7	**59.8**
Median C2C ↓ (mm)	39.0	48.9	47.3	**38.7**
P95 C2C ↓ (mm)	245.6	293.0	298.4	**223.3**
Thresholded GT metrics @ 48 mm				
Completeness ↑ (%)	56	49	50	**59**
F-score ↑ (%)	72	66	67	**73**
Thresholded GT metrics @ 80 mm				
Completeness ↑ (%)	75	67	66	**80**
F-score ↑ (%)	86	80	79.5	**89**

The Hausdorff distance reflects a single maximum deviation and can be inflated by outliers even when overall reconstruction is accurate; we therefore treat the C2C distance family and thresholded completeness as the primary twin fidelity indicators.

As seen in Table I, Hyb-O achieves the best performance across all distance-based and completeness metrics, confirming that oracle-guided switching between co-visibility and baseline-aware repair produces the most faithful digital twin. Covis is the strongest base scorer, with a mean C2C of 66.0 mm and completeness of 56% at 48 mm, demonstrating that broad re-observation of reconstructed structure remains an effective strategy even without geometry-focused repair. B-aware improves triangulation quality for under-supported regions but yields worse coverage than Covis, indicating that geometry-focused targeting alone does not compensate for reduced breadth of observation.

Hyb-H underperforms all other strategies. Two compounding sources of miscalibration likely explain this outcome. First, the gate thresholds in Eq. (9) were hand-tuned without systematic optimization against mission outcomes, and may hold the B-aware phase too long early in the mission, crowding out the coverage-preserving Covis phase. Second, the scorer weight vector $(W_{\text{main}}, W_{\text{nov}}, W_{\text{move}}, W_{\text{ang}})$ is shared and fixed across both base scorers, meaning neither scorer is independently optimized when called by the switch. Together, these two factors suggest that the gate criterion and per-scorer weight configuration are both design challenges that must be addressed before a heuristic hybrid policy can be reliably deployed.

The primary differentiator between strategies is surface coverage rather than point-level precision: all strategies achieve comparable geometric faithfulness for reconstructed points, and the main gap between them is how much of the target surface is reached at all.

The qualitative overlay in Fig. 4 shows Hyb-O recovering

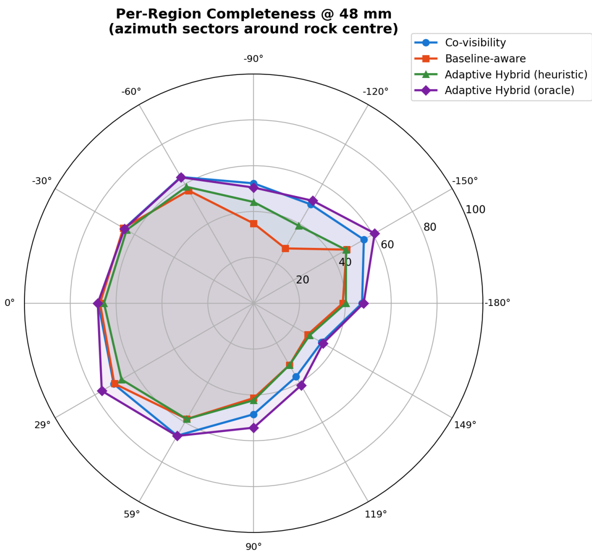

Fig. 3. Per-region completeness at 48 mm across azimuth sectors for all NBV strategies. The polar plot highlights the broader sector-wise completeness of Hybrid-Oracle and Co-visibility, while Baseline-aware and Hybrid-Heuristic strategies leave larger coverage gaps in several regions

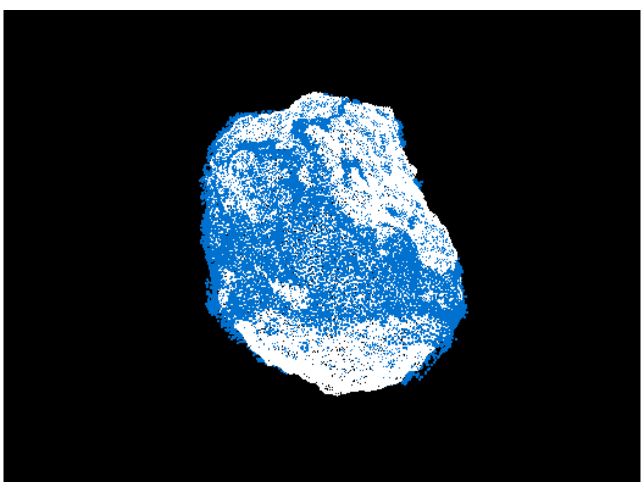

Fig. 4. Qualitative overlay of the ground-truth rock surface and the Hybrid-Oracle dense reconstruction. White points represent the ground truth, while blue points represent the reconstructed dense cloud, illustrating strong overall alignment and surface recovery while also showing remaining unreconstructed regions.

broad portions of the target surface, although notable regions remain unreconstructed. The per-region breakdown in Fig. 3 shows that Hyb-O and Covis achieve broadly similar sector-wise completeness, both outperforming B-aware across most azimuths. B-aware shows the sharpest drop in completeness in the lower-hemisphere sectors, consistent with its tendency to sacrifice broad coverage in favor of geometry-focused repair. Hyb-H underperforms across most sectors, reflecting the miscalibration discussed above.

V. Conclusion and Future Work

Hyb-O achieves the strongest performance across all metrics, demonstrating that adaptive scorer switching outperforms either base scorer alone. In contrast, Hyb-H underperforms both base scorers, revealing two critical gaps: (1) gate thresholds require mission-specific tuning, and (2) the shared weight vector $(W_{\text{main}}, W_{\text{nov}}, W_{\text{move}}, W_{\text{ang}})$ is unoptimized for individual scorers. A joint approach using a Gaussian process surrogate model over the combined space of thresholds and weights, optimized against reconstruction metrics could identify generalizable configurations without per-scene search. An alternative is to train a learned switching policy from the oracle's per-iteration scorer selections, replacing hand-tuned gates with data-driven criteria. The resulting dense point cloud serves as a digital twin enabling inspection and change detection without repeated physical access, directly integrating autonomous robotic systems with digital twin workflows.

Acknowledgments

The authors acknowledge NASA's Astromaterials 3D initiative for providing access to the lunar sample datasets used in this work. Apollo Lunar Sample Credit: NASA/These 3D reconstructed image data were produced at the Lunar Sample Laboratory Facility for Astromaterials 3D in NASA's Acquisition and Curation Office and were funded by the NASA Planetary Data Archiving, Restoration, and Tools (PDART) Program, Proposal No.: 15-PDART15_2-0041.

References

[1] 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," *Commun. ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[2] S. Lee, L. Chen, J. Wang, A. Liniger, S. Kumar, and F. Yu, *Uncertainty Guided Policy for Active Robotic 3D Reconstruction using Neural Radiance Fields*, arXiv:2209.08409 [cs], Sep. 2022.

[3] B. Kerbl, G. Kopanas, T. Leimkuehler, and G. Drettakis, "3D Gaussian Splatting for Real-Time Radiance Field Rendering," en, *ACM Transactions on Graphics*, vol. 42, no. 4, pp. 1–14, Aug. 2023.

[4] J. L. Schonberger and J.-M. Frahm, "Structure-From-Motion Revisited," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016.

[5] M. Maboudi, M. Homaei, S. Song, S. Malihi, M. Saadatseresht, and M. Gerke, "A Review on Viewpoints and Path Planning for UAV-Based 3-D Reconstruction," en, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 5026–5048, 2023.

[6] B. Yamauchi, "A frontier-based approach for autonomous exploration," in *Proceedings of the 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation*, ser. CIRA '97, USA: IEEE Computer Society, 1997, p. 146.

[7] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "OctoMap: An efficient probabilistic 3D mapping framework based on octrees," en, *Autonomous Robots*, vol. 34, no. 3, pp. 189–206, Apr. 2013.

[8] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.