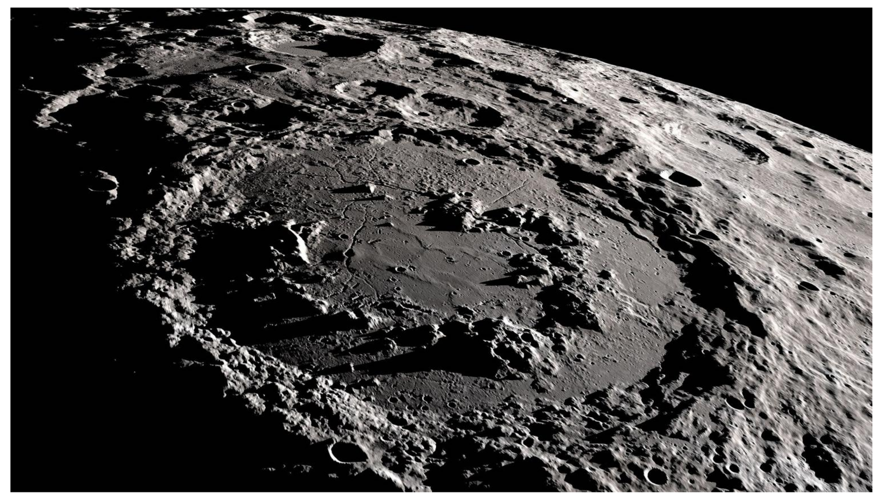
Hybrid Sparse-Model Next-Best-View Planning for Active 3D Reconstruction for Digital Twins

Carlos Torre, Kanav Prashar, Bharath Vedantha Desikan, Rodney Staggers Jr., Jnaneshwar Das

PROBLEM & MOTIVATION

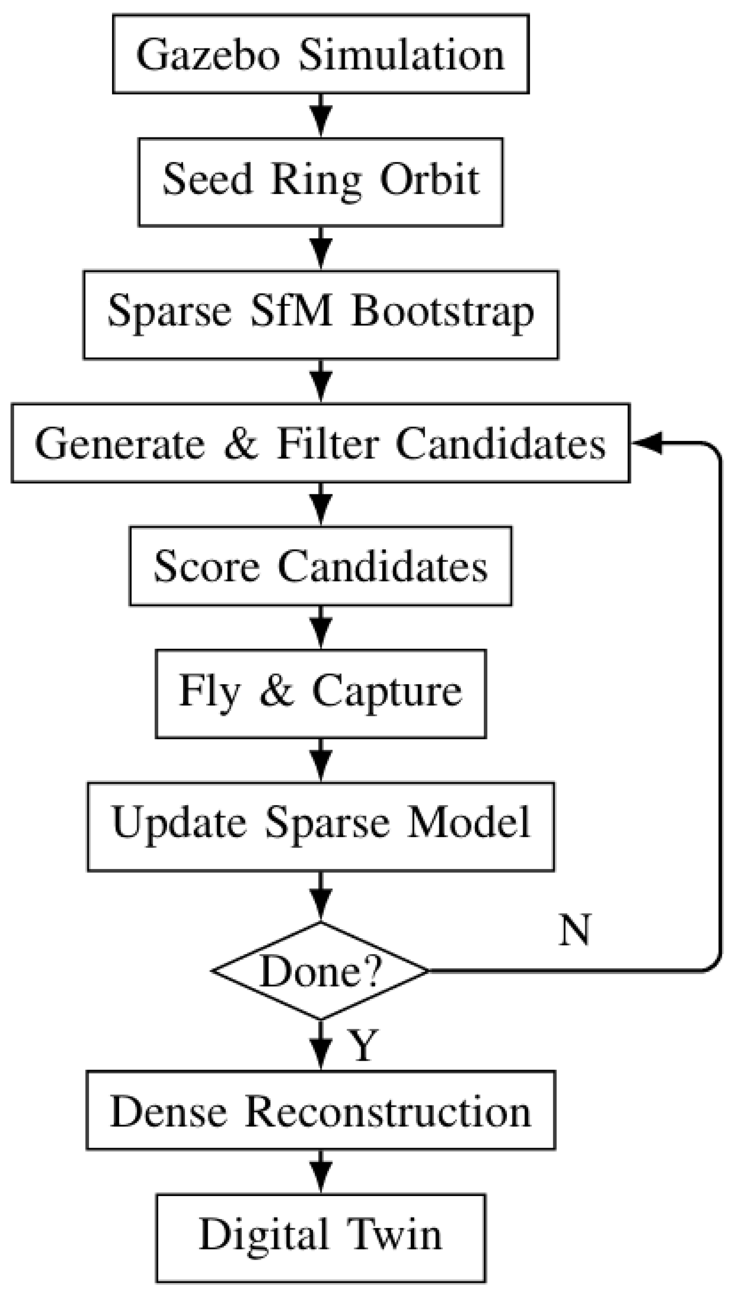
- Current lunar south pole exploration efforts by NASA require accurate terrain mapping strategies
- Next Best View (NBV) selects optimal camera poses to iteratively improve a 3D model.
- NeRF [1] and 3DGS [2] enable high-quality novel-view synthesis but often require heavier representations.
- We propose a lightweight photogrammetry-native pipeline that plans directly from an incrementally updated Structure-from-Motion [3] (SfM) sparse point cloud with no dense representation required online.



SYSTEM PIPELINE

The mission consists of two main phases:

- **Phase 1 – Seed collection:** The vehicle executes a stratified ring orbit around the target capturing 24 seed images to bootstrap COLMAP.
- **Phase 2 – Active View Selection:** Candidate views are filtered by flight constraints, scored, and visited iteratively until the image budget or sparse-cloud convergence threshold is reached



SCORER FORMULATION

Shared candidate (C) scorer

$S(c) = W_{main} * M_{main}(c) + W_{nov} * nov(c) - W_{move} * move(c) - W_{ang} * ang(c)$

$W_{main} = 1.0,\ W_{nov} = 1.2,\ W_{move} = 0.1, W_{ang} = 0.4$

$nov(c)$ – Rewards candidates far from previously visited viewpoints

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

move(c) – Penalized large displacements from current pose

$$move(c) = 0.7\frac{\|\Delta x\|}{d_{travel}} + 0.3\frac{|\Delta\psi|}{\pi}, d_{travel} = 10,$$

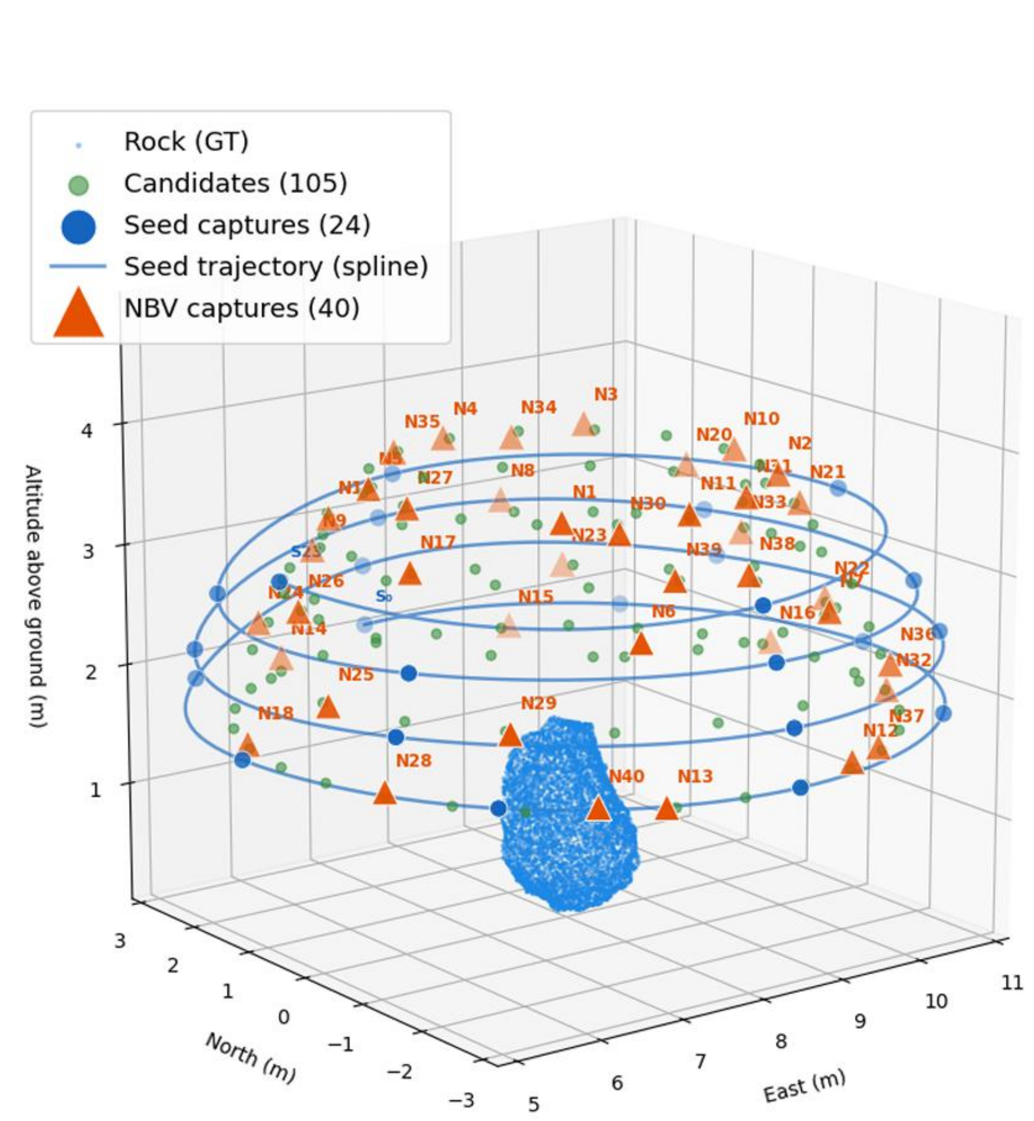
ang(c) –Penalizes views within $\theta_{min} = 20°$

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

Where:

- $c = candidate\ viewpoint$
- $\angle(c - v_j)$ =angle between viewing direction of candidate c and that of previously visited viewpoint v_j
- $\Delta\psi$ = yaw change
- Δx = 3D Euclidean distance change

Drone Trajectory & Candidate Viewpoints (ENU frame)



Hybrid scorers: Two hybrid hard switch scorers are used to determine which M_{main} is used in the scorer formulation:

- Hybrid-Heuristics (Hybrid-H) uses image budget fraction consumed f_b and d_{95} the 95th-percentile KNN distance in the sparse point cloud to decide at each iteration :

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

- Hybrid-Oracle (Hyb-O) uses ground-truth mesh supervision to choose the best scorer at each iteration by selecting the scorer whose top candidate covers the most previously uncovered surface.

M_{main} differs in two different ways

$$M^{covis}_{main}(c) = \frac{|P(c)|}{|P|}$$

- Rewards viewpoints that observe the broadest slide of existing structure, preserving geometric support across the model.
- Let P be all tracked SfM points withing a 2m half-extent bounding box around the target, and $P(c) \subseteq P$ the subset visible from candidate c via frustum projection

$$M^{B-aware}_{main}(c) = \frac{\sum_{p \in P(c)} w_p g(p,c)}{\sum_{p\in P(c)} w_p} \qquad g(p,c) = clip\left(\frac{\theta_{min}(p,c) - \theta_\ell}{\theta^* - \theta_\ell}, 0{,}1\right)$$

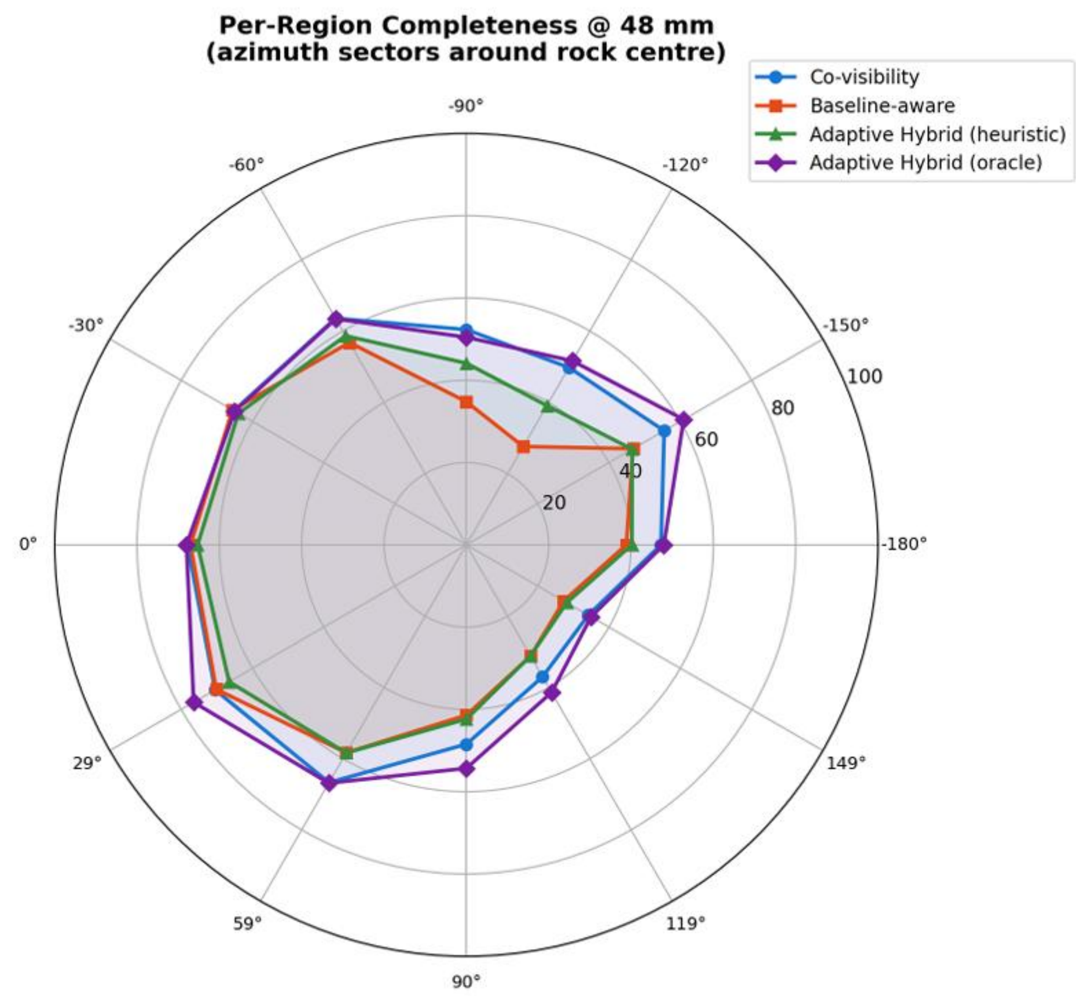
- w_p = Weakness score per viewpoint. Point utility combining track quality and local sparsity
- $g(p,c)$ = Geometry gain

Where:

- $w_p = \alpha_\tau \tilde{\tau}_p + \alpha_\rho \tilde{\rho}_p,\ \alpha_\tau = 0.6, \alpha_\rho = 0.4$
- $\tilde{\rho}_p$ is the percentile-normalized mean 5-nearest-neighbor distance
- $\tilde{\tau}_p = clip\left(\frac{T^* - \tau_p}{T^*}, 0{,}1\right)$, track length $T^* = 6$
- $\theta_{min}(p, c)$ is the minimum angle between the candidate's ray to p and any existing observation ray to p
- $\theta_\ell = 5°$ is the redundancy floor and $\theta^* = 25°$ is the saturation angle

RESULTS:

- Hyb-O achieves the best overall performance across all evaluated metrics.
- Adaptive scorer switching improves results, outperforming either base scorer alone.
- Surface coverage is the main differentiator, not point-level precision.
- The largest performance gap comes from how much of the target surface is reconstructed, rather than how accurately individual points are reconstructed.



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**

PLATFORM & STACK

Vehicle: PX4 X500 quadcopter with RGB gimbal camera

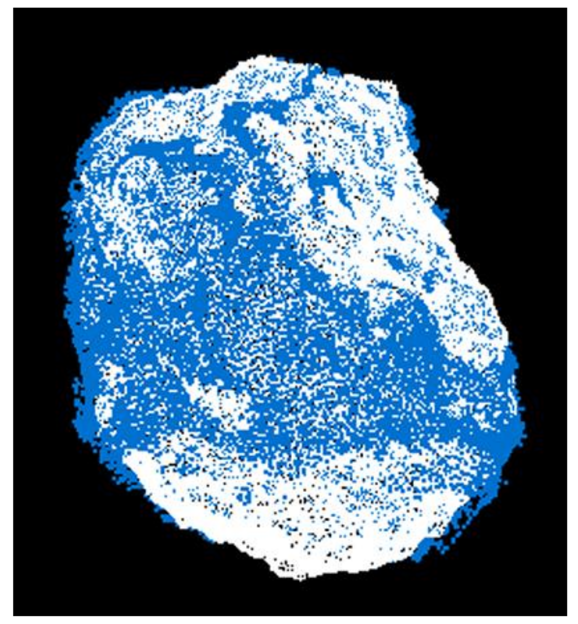
Simulation: Gazebo

Autonomy:

- PX4: Flight controller
- ROS2: Robot Operating System
- MicroXRCE: Middleware

COLMAP: Structure-from-Motion

QUALITATIVE COMPARISON



BLUE: 3D reconstruction
WHITE: Ground Truth

CODE

SCAN ME

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] 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. [3] 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. [4] [4] P. Moore, "south-pole-schrodinger-cropped-annotated.jpg," NASA, image uploaded Apr. 21, 2020. Image taken Apr. 21, 2020, 20:02:32.

Contact information:
ctorre1@asu.edu
carlostorresada@gmail.com
LinkedIn:
https://www.linkedin.com/in/carlos-torre-761a6a154/