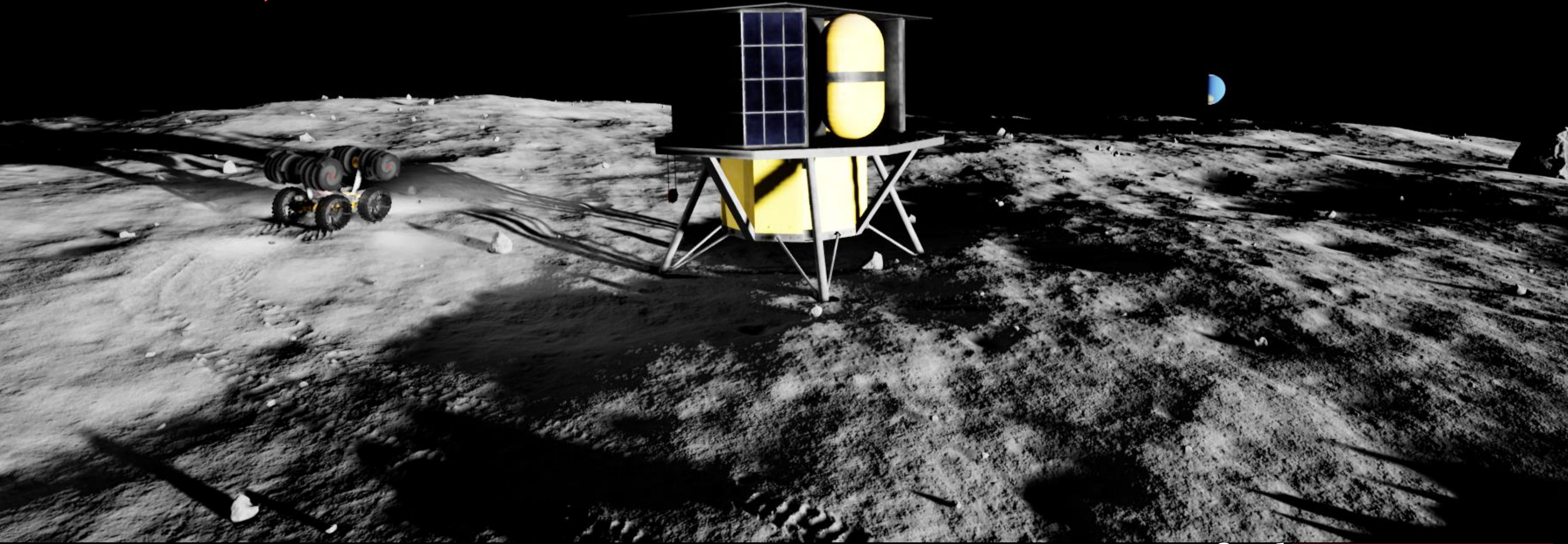
Full Stack Navigation, Mapping, and Planning for the Lunar Autonomy Challenge

Adam Dai, Asta Wu, Keidai Iiyama, Guillem Casadesus Vila, Kaila Coimbra, Thomas Deng, and Grace Gao



September 12, 2025 - ION GNSS+ 2025, Baltimore, MD

Stanford University

Autonomy for Future Lunar Missions

Autonomy refers to a rover's ability to perceive, decide, and act without human input

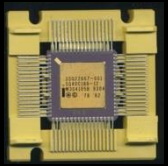
- **Enable scalable rover operations** without real-time human oversight
- **Extend access to challenging terrain**, including permanently shadowed regions
- **Lower operations burden on Earth-based teams** to reduce cost and increase resilience


Credit: NASA

Autonomy is essential for sustained lunar surface activity

Challenges of Lunar Surface Autonomy

Limited Sensing and Compute



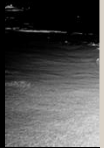
- Vision
- Low-p
  proce
- Power constraints

GNSS Denied

- No reliable global localization

→ Optical Navigation

Lighting and Shadows



Lack of Features

less

ks for

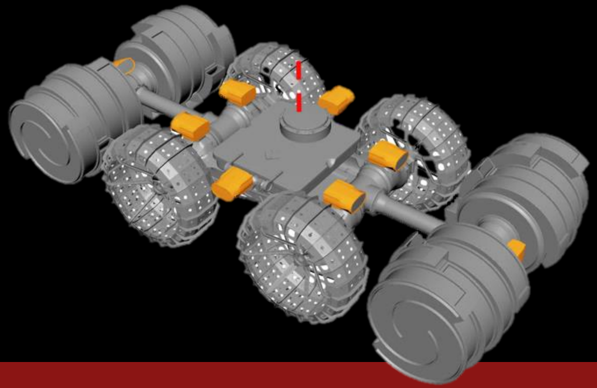
Traditional visual SLAM methods (e.g., ORB-SLAM, Kimera-VIO) struggle under these conditions

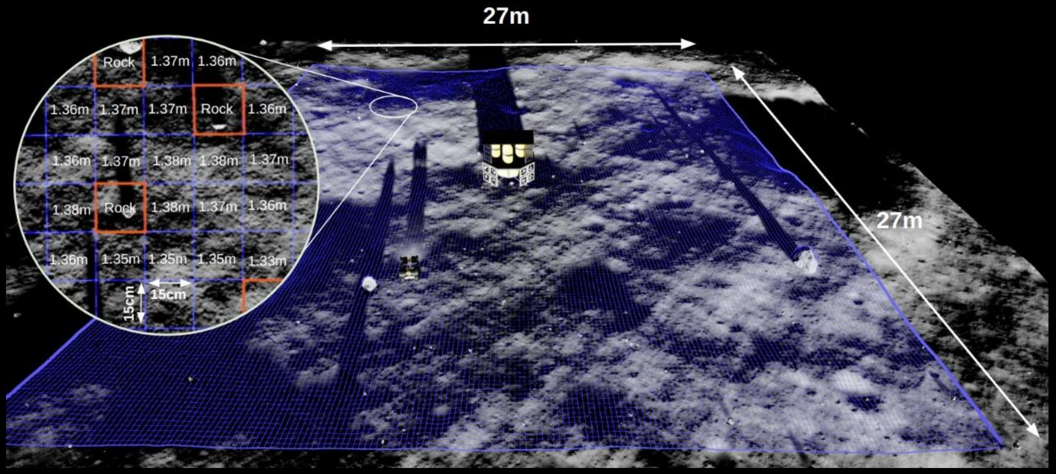
The Lunar Autonomy Challenge

Credit: Lunar Autonomy Challenge

Challenge Overview

- **Objective:** map 27 m × 27 m region of terrain in a 180 × 180 grid (15 cm cell resolution)
- **Geometric map:** elevation per cell. Score based on % of cells mapped within 5 cm error
- **Rock map:** binary rock presence per cell. Evaluated with F1 score (precision and recall)
- **Rover:**
  - Sensors: 8 cameras, IMU
  - Linear and angular velocity control





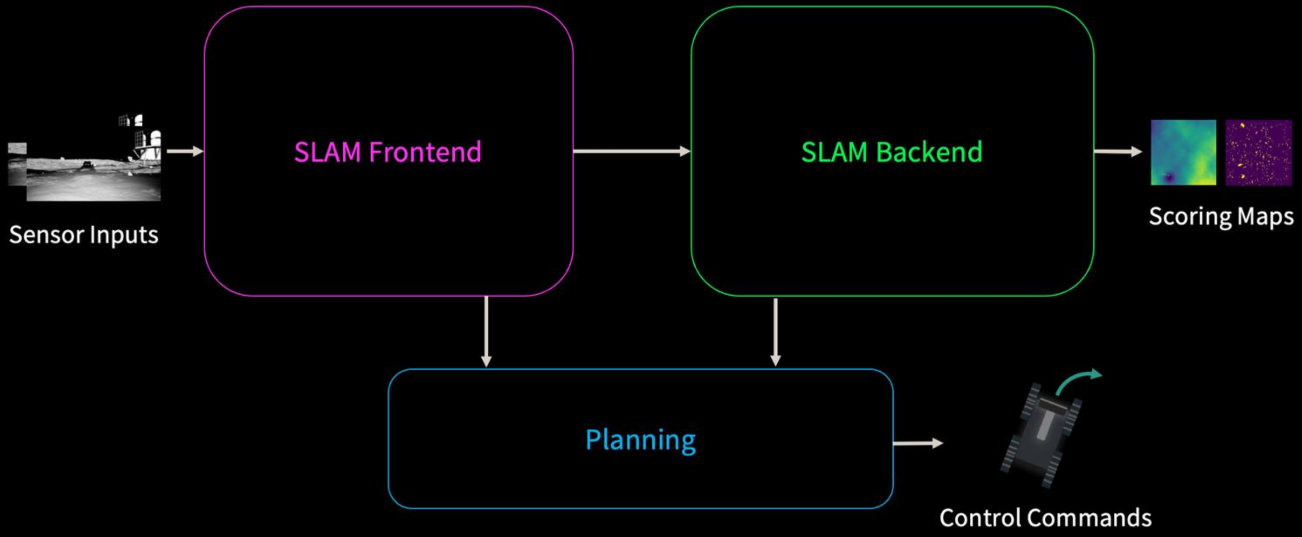
Credit: Lunar Autonomy Challenge

Outline

- Approach
  - Perception
  - SLAM
  - Planning
- Results
- Conclusion

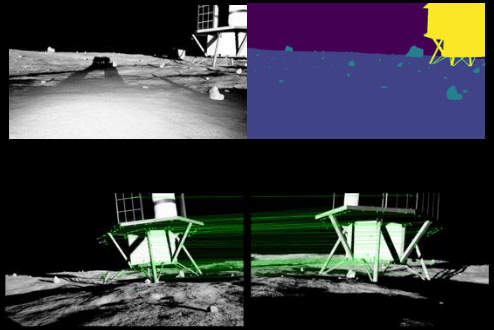
Outline

- **Approach**
  - Perception
  - SLAM
  - Planning
- Results
- Conclusion

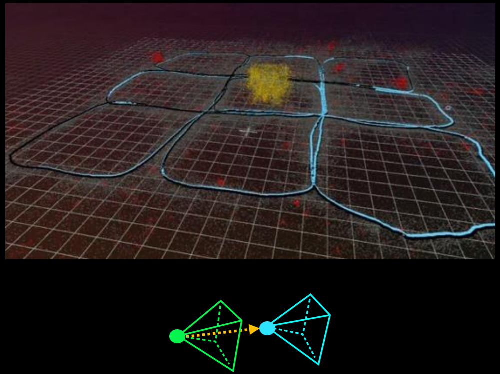


Key Contributions

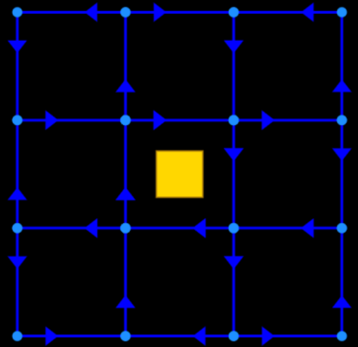
Learning-based Perception under Harsh Conditions



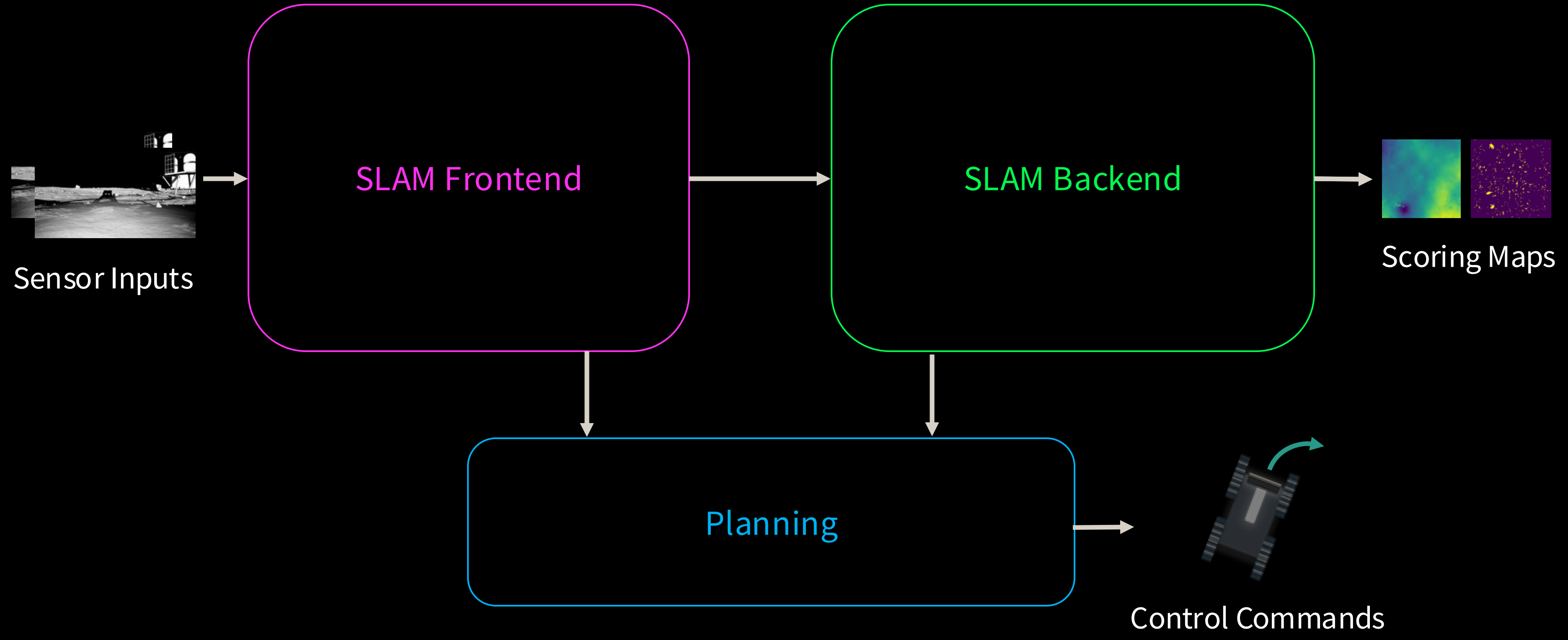
Lightweight Vision-Only Simultaneous Localization and Mapping **(SLAM)**



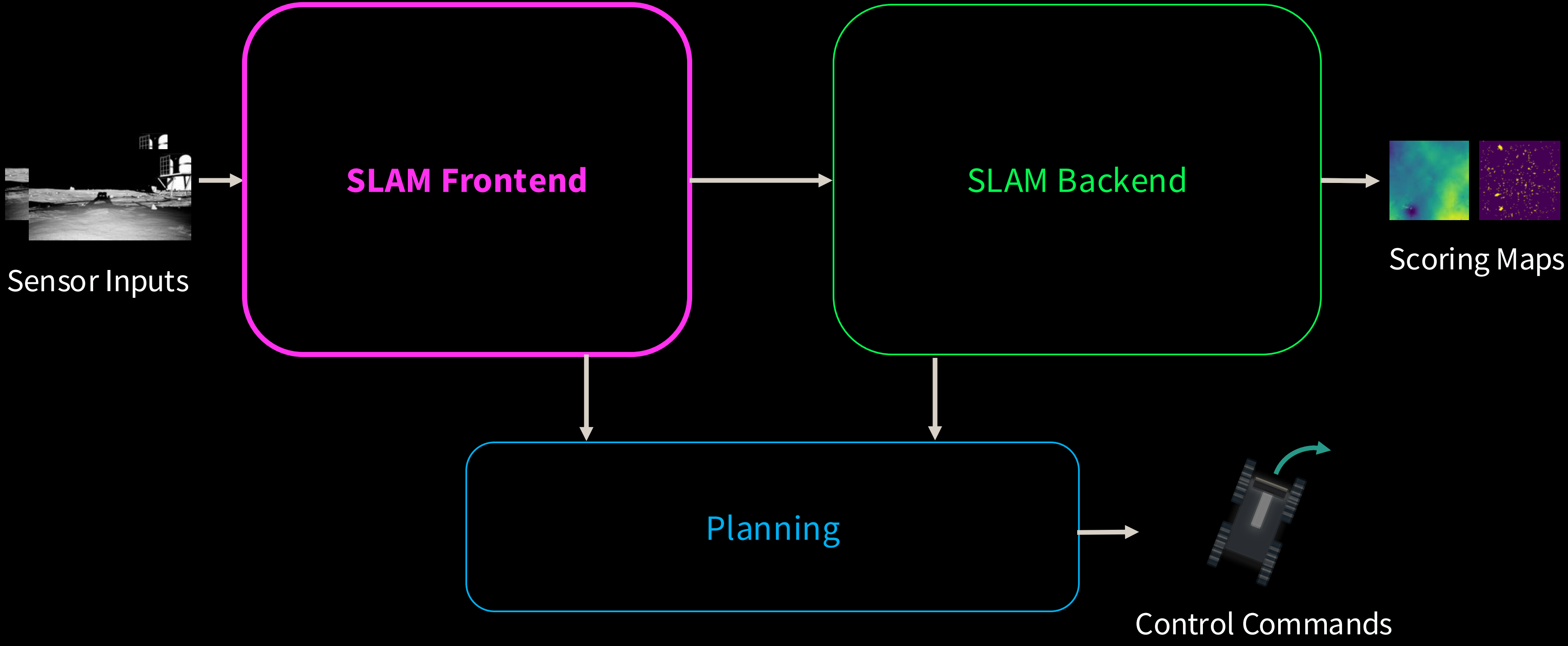
Trajectory Design for Loop Closure and Coverage



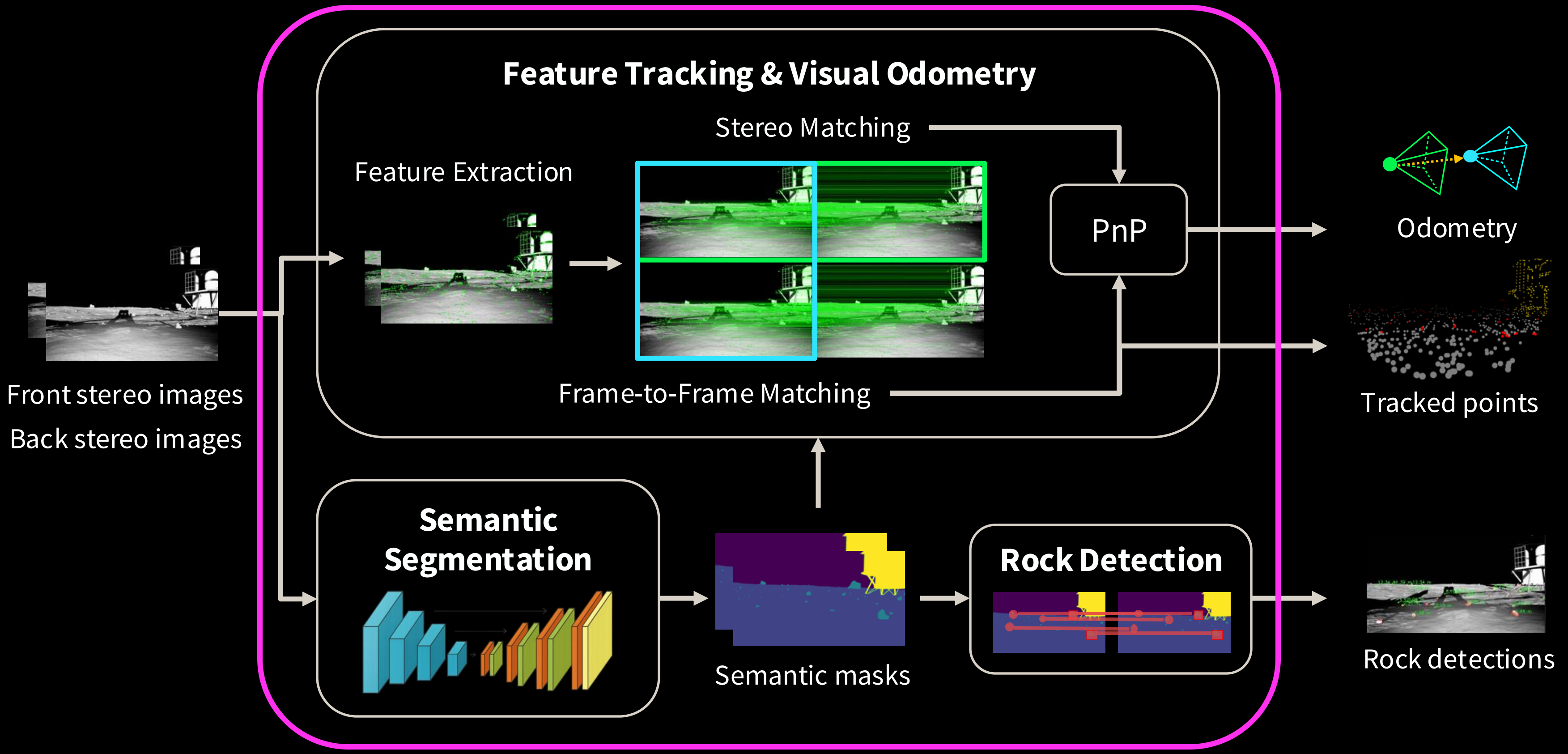
Autonomy Stack



Autonomy Stack

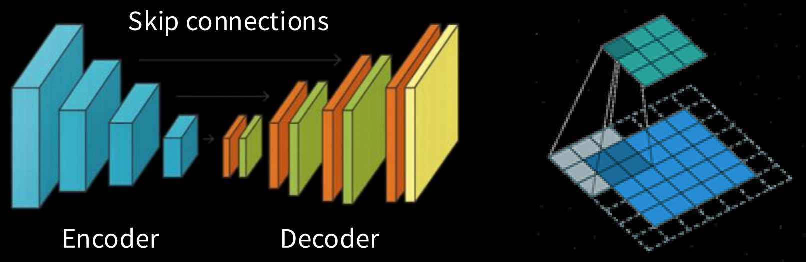


SLAM Frontend

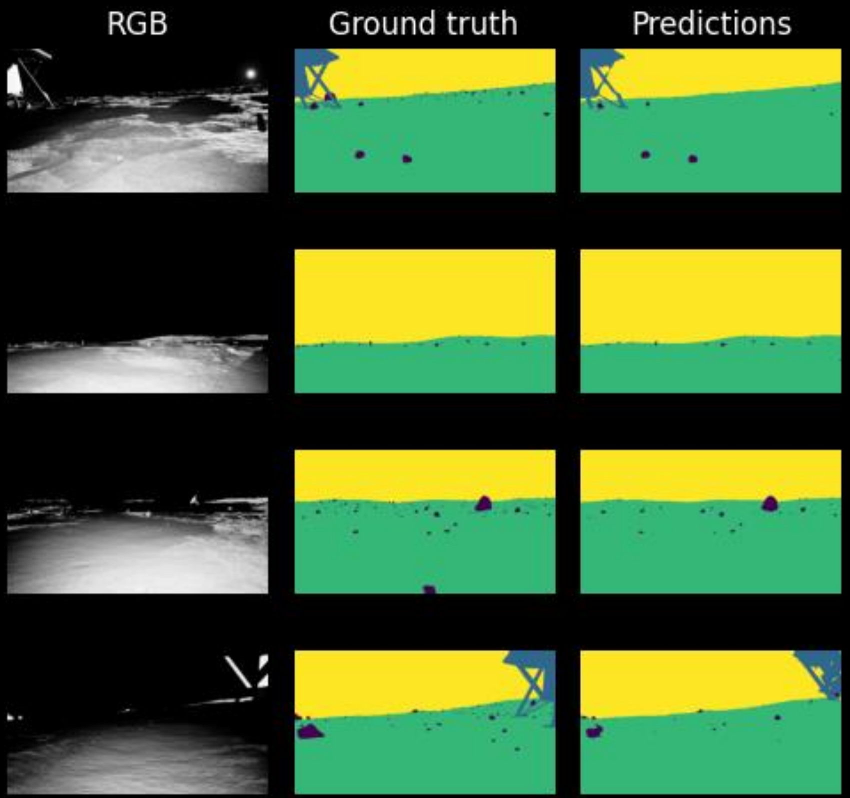


Semantic Segmentation

- Model: **Unet++** [1]
  - Convolutional Neural Network (CNN)



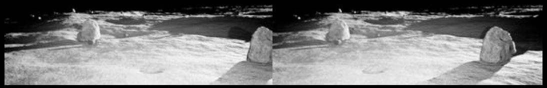
- Finetuned on ground-truth semantic masks
- Outperformed newer transformer-based methods in speed and accuracy



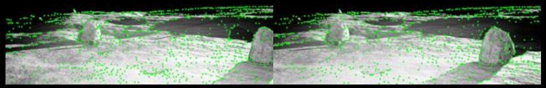
[1] Zhou et al., *ML-CDS* 2018

Feature Extraction and Matching

Left Right



Detected keypoints



Matches

Feature Extraction: **SuperPoint** [2]

- *CNN-based keypoint detector and descriptor*
- *Detects repeatable and distinctive 2D features under varied lighting and texture*

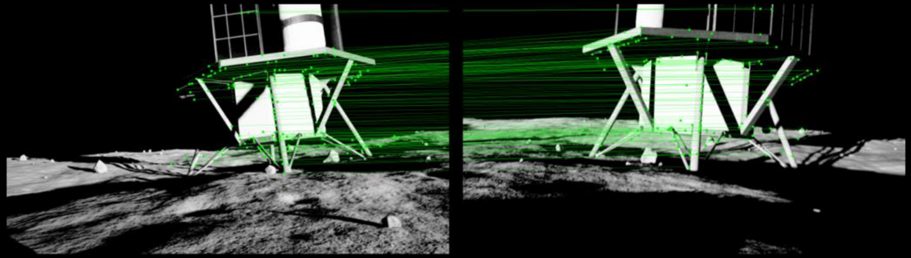
Feature Matching: **LightGlue** [3]

- *Transformer-based feature matcher*
- *Robust to large viewpoint and appearance changes*

Stereo matching gives depth

In addition to left-right stereo matching, we can also match features across:

1. consecutive frames (feature tracking, motion estimation)
2. non-consecutive frames (loop closure)



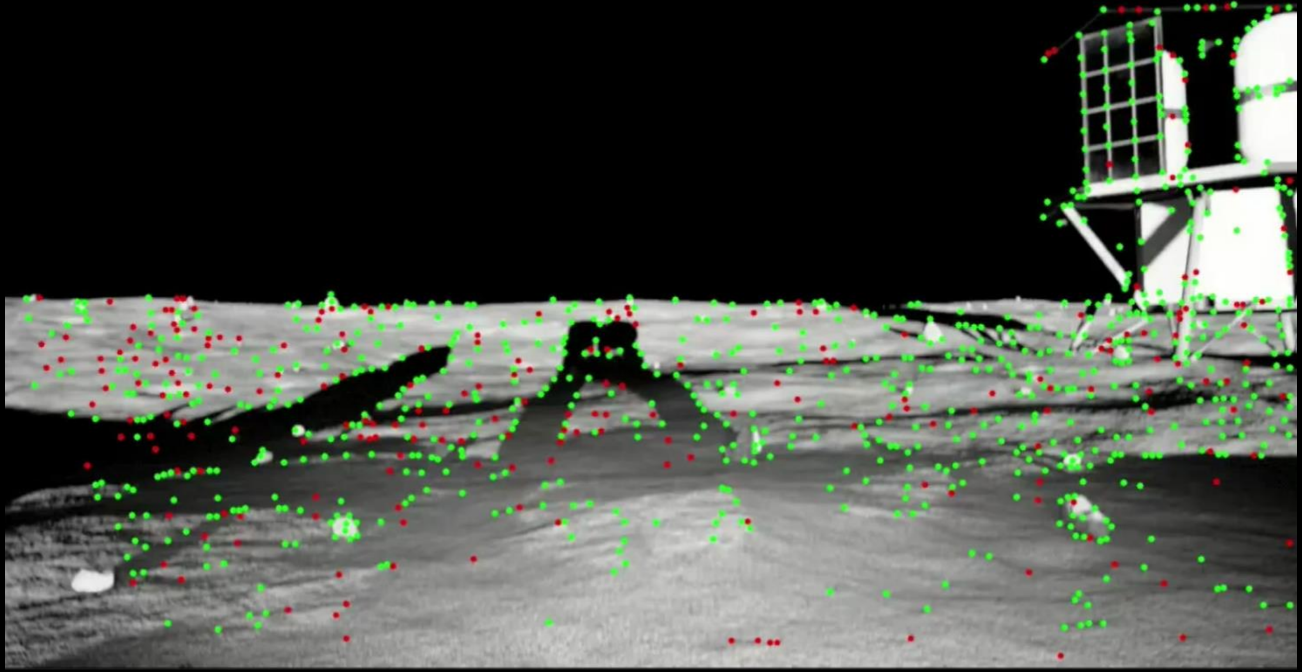
[2] DeTone et al., *CVPR* 2018

[3] Lindenberger et al., *ICCV* 2023

Feature Tracking

Features matched across frames via LightGlue matching

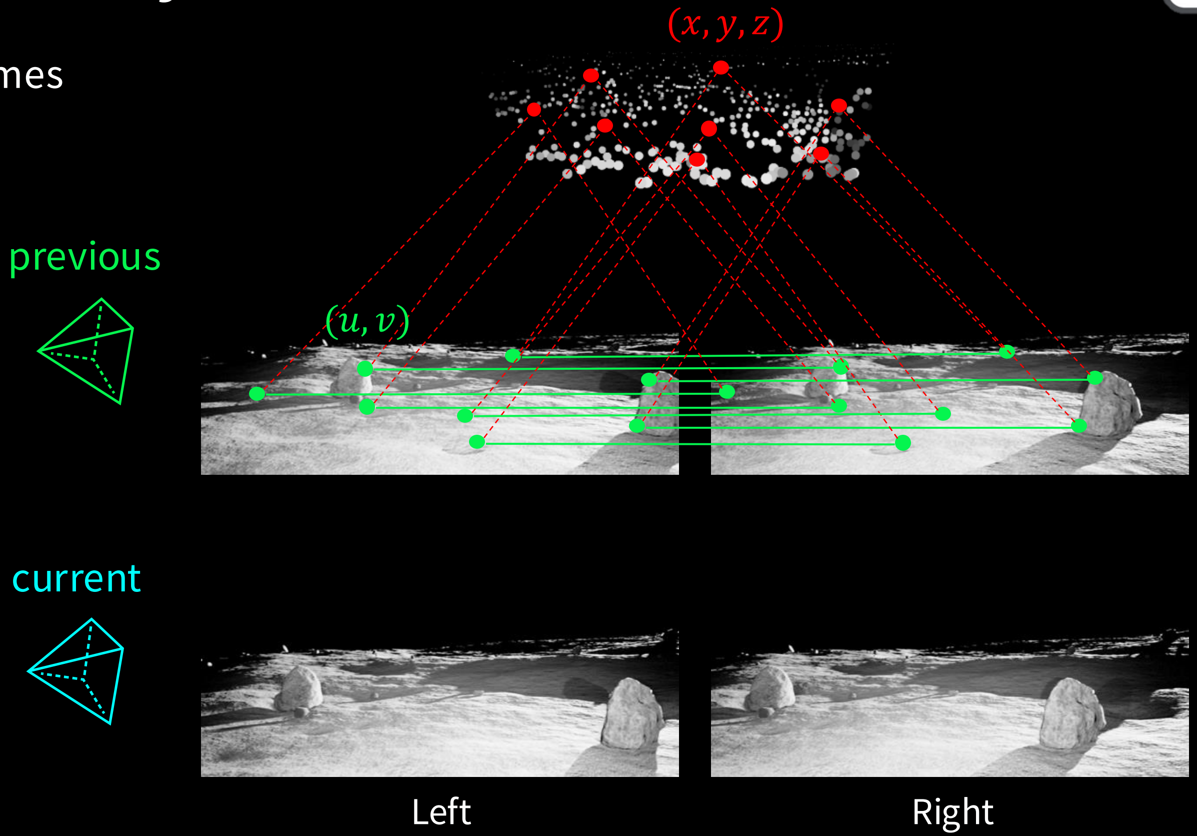
(green = tracked, red = newly initialized)



Stereo Visual Odometry

Estimate motion between frames

1. **Stereo matching**

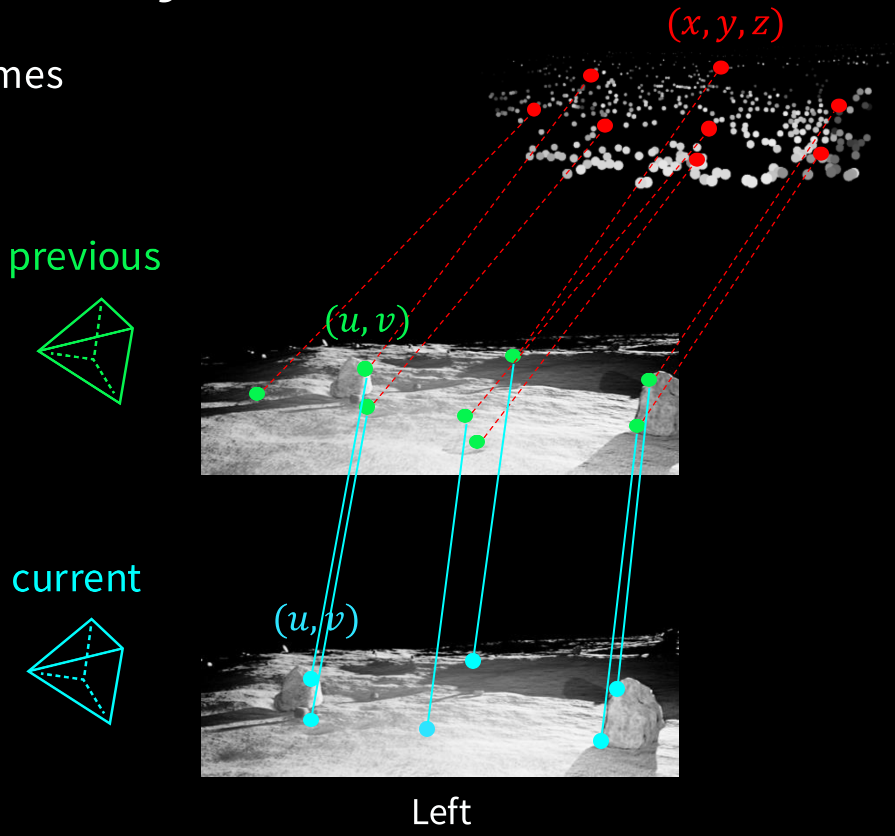


Geiger et al., *IV* 2011

Stereo Visual Odometry

Estimate motion between frames

1. Stereo matching
2. **Frame-to-frame matching**



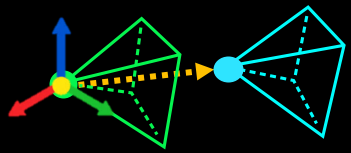
Geiger et al., *IV* 2011

Stereo Visual Odometry

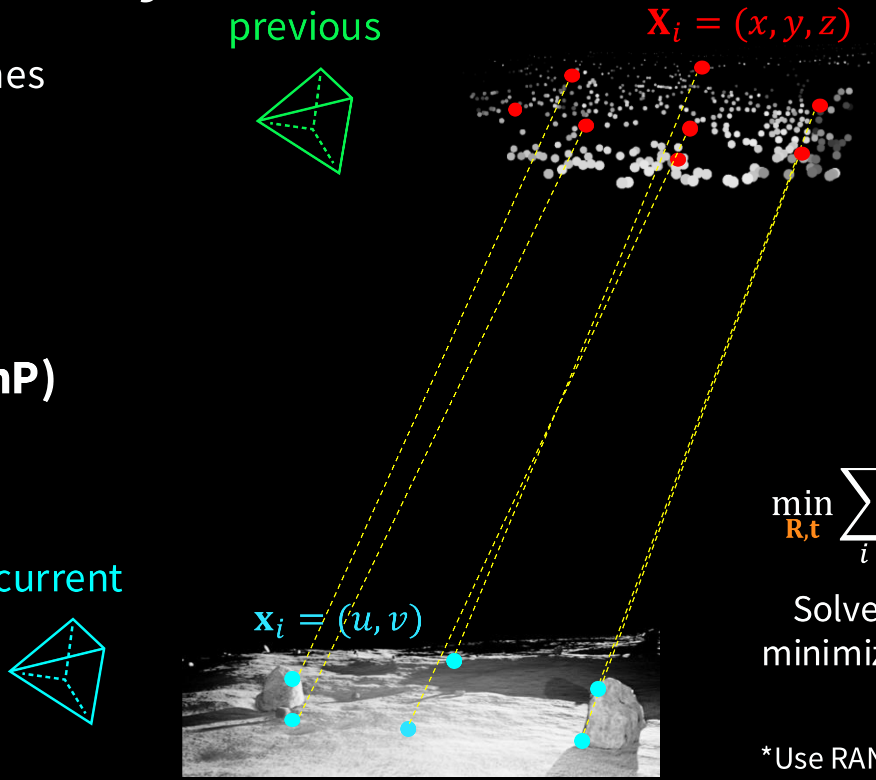
Estimate motion between frames

1. Stereo matching
2. Frame-to-frame matching
3. **Perspective-n-Point (PnP)**

➢ Current camera pose in the frame of previous



$(\mathbf{R}, \mathbf{t})$



Left

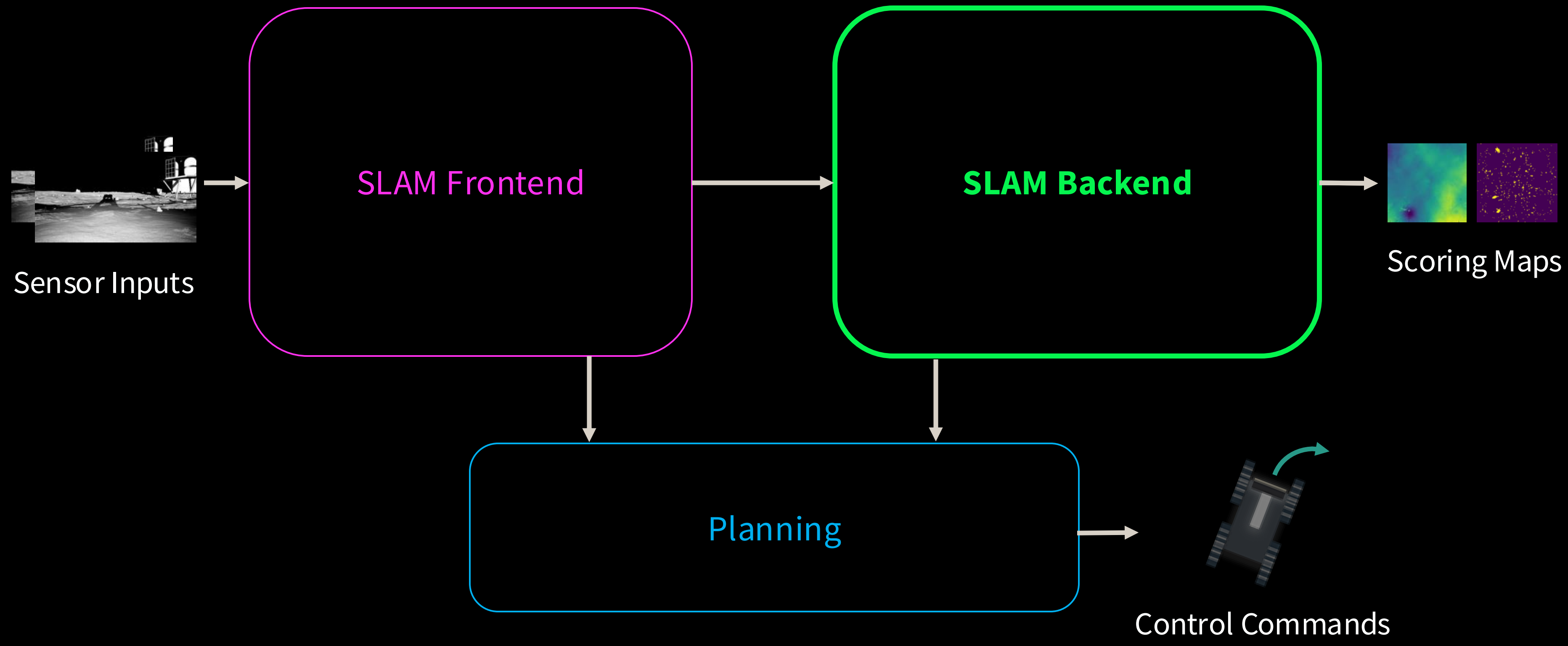
$$\min_{\mathbf{R},\mathbf{t}} \sum_i \|\pi(\mathbf{R}\mathbf{X}_i + \mathbf{t}) - \mathbf{x}_i\|^2$$

Solve for pose $(\mathbf{R}, \mathbf{t})$ that minimizes reprojection error

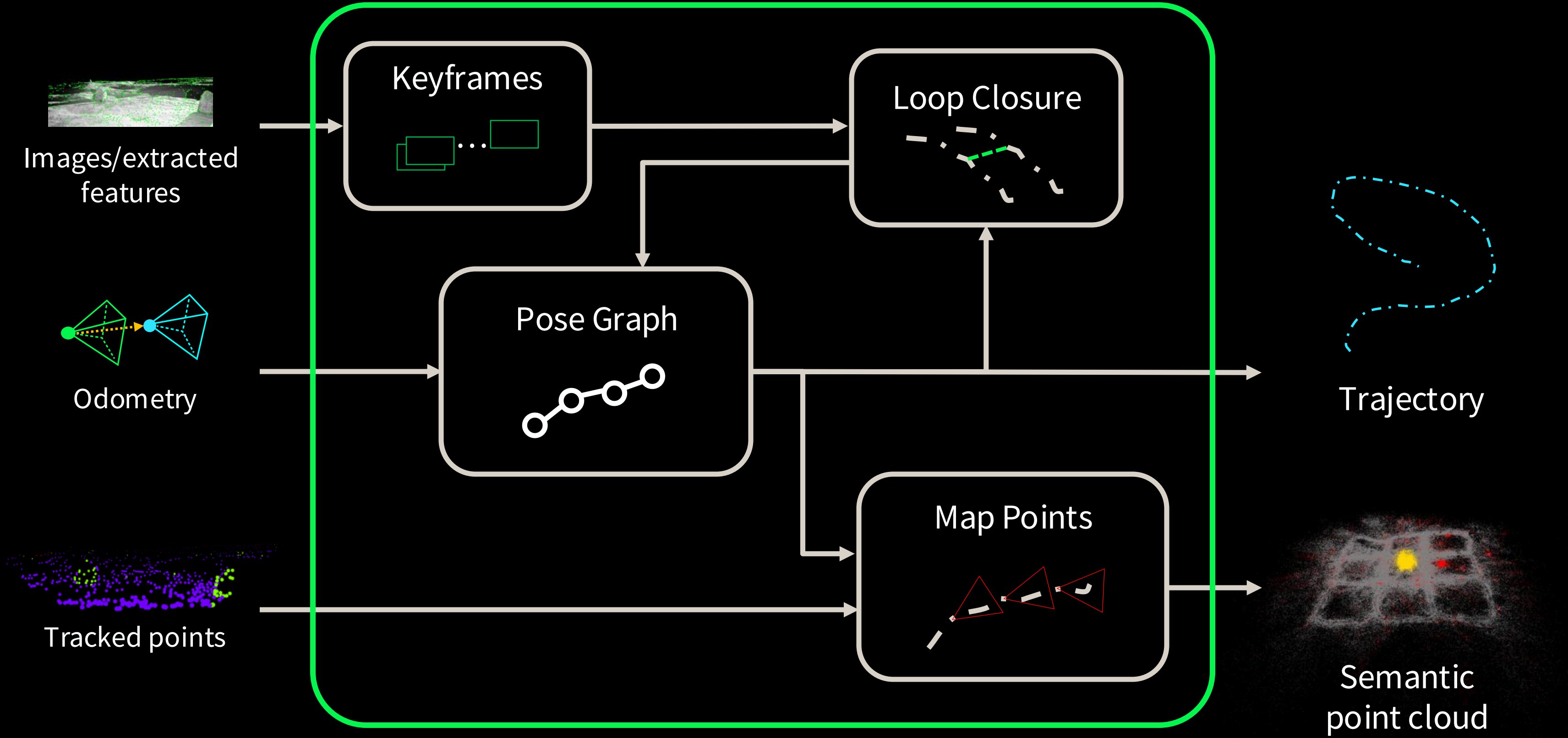
*Use RANSAC for outlier rejection

Geiger et al., *IV* 2011

Autonomy Stack

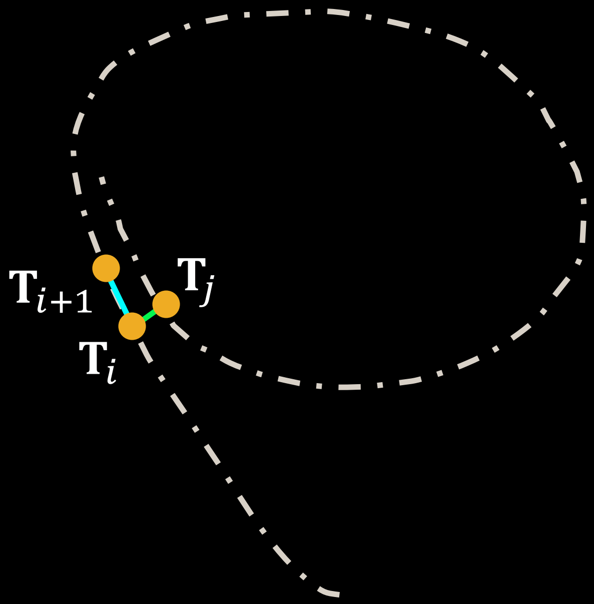


SLAM Backend



Pose Graph

Graph of 3D poses (position and orientation) connected by spatial constraints



Nodes: 3D pose - $\mathbf{T}_i \in SE(3)$

Edges (factors):

- Odometry: $(\mathbf{T}_i, \mathbf{T}_{i+1})$ consecutive poses
- Loop closure: $(\mathbf{T}_i, \mathbf{T}_j)$ non-consecutive poses

Optimization:

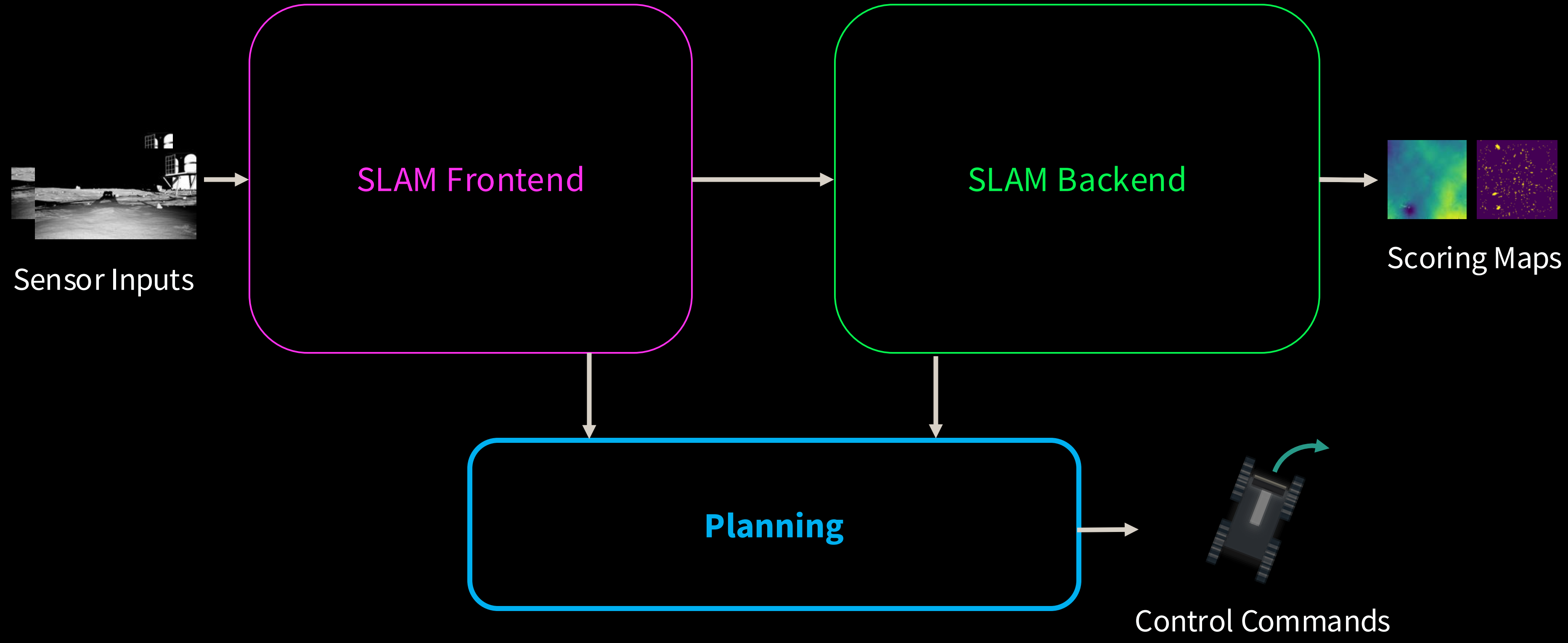
$$\underbrace{\min_{\{\mathbf{T}_i\}}}_{\text{all poses}} \underbrace{\sum_{(i,j)\in\mathcal{E}}}_{\text{edge set}} \left\| \overbrace{\mathbf{T}_i^j}^{\text{measured}} - \overbrace{(\mathbf{T}_i)^{-1}\mathbf{T}_j}^{\text{expected}} \right\|^2$$

Minimize residuals of measured vs. expected relative poses

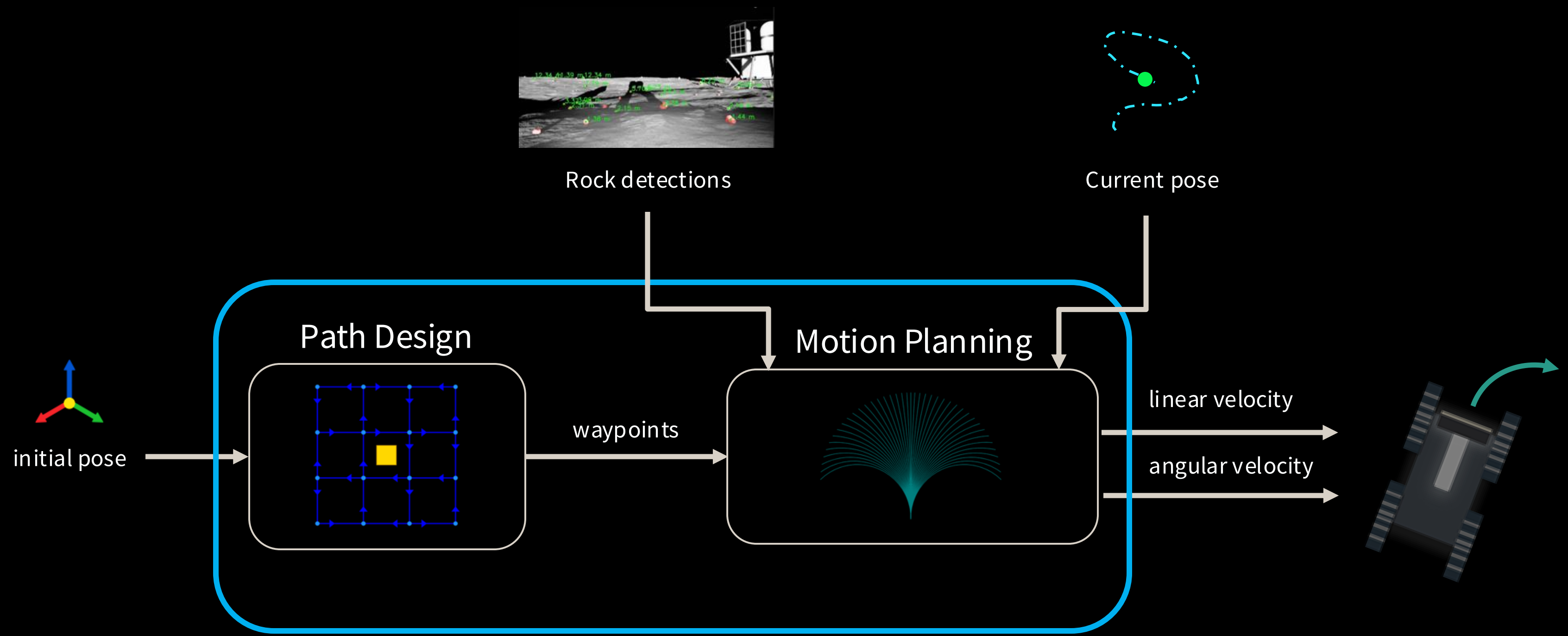
Implemented with GTSAM [4]

[4] Dellaert et al., *Georgia Institute of Technology* 2012

Autonomy Stack

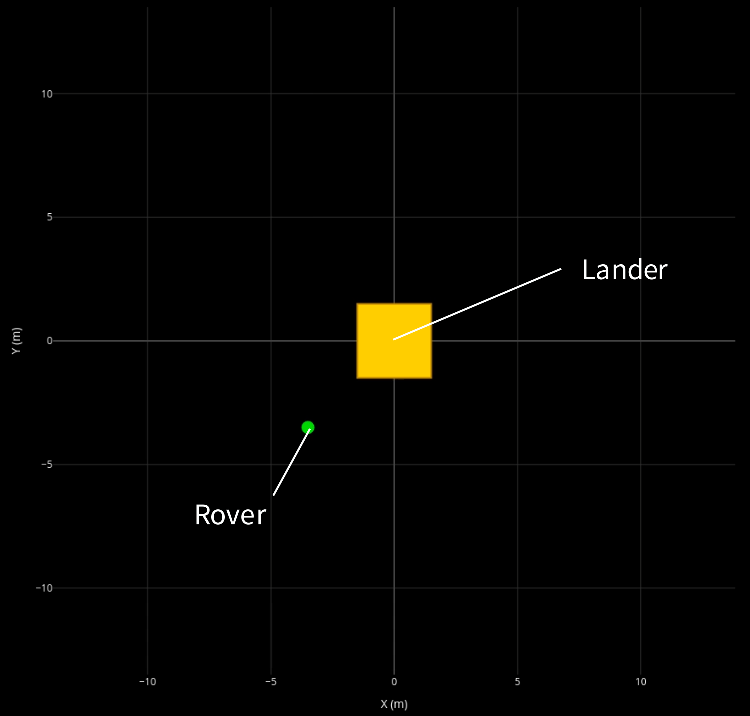


Planning Modules



High-level Path Design

- Grid pattern to maximize coverage while incentivizing loop closure

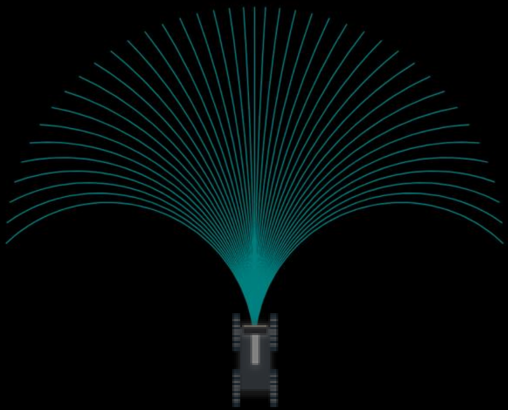


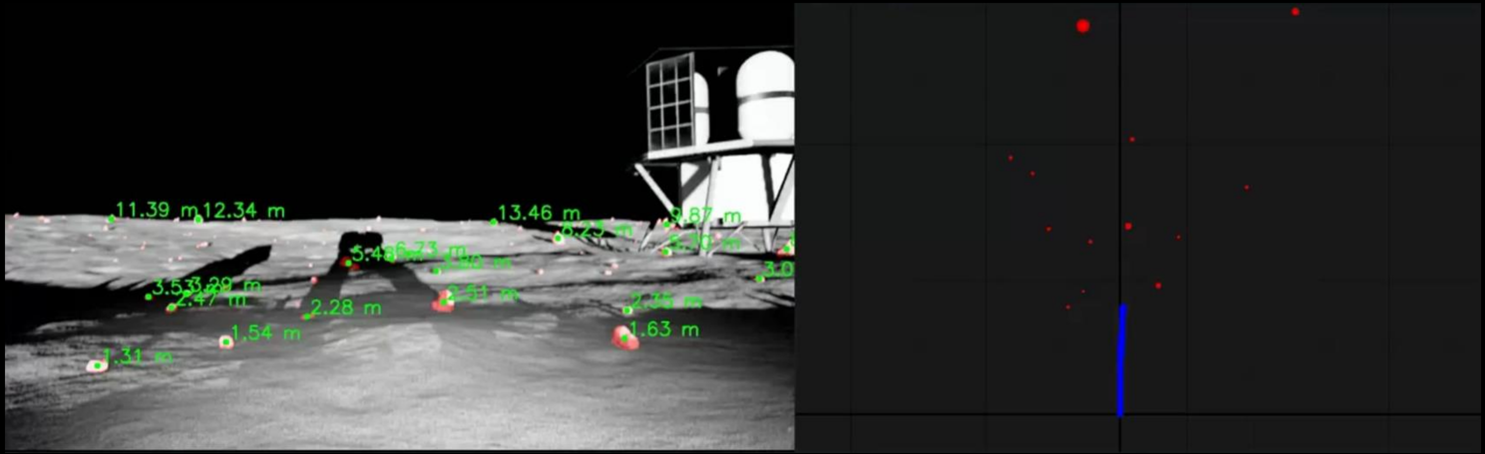
Stanford University

Motion Planning

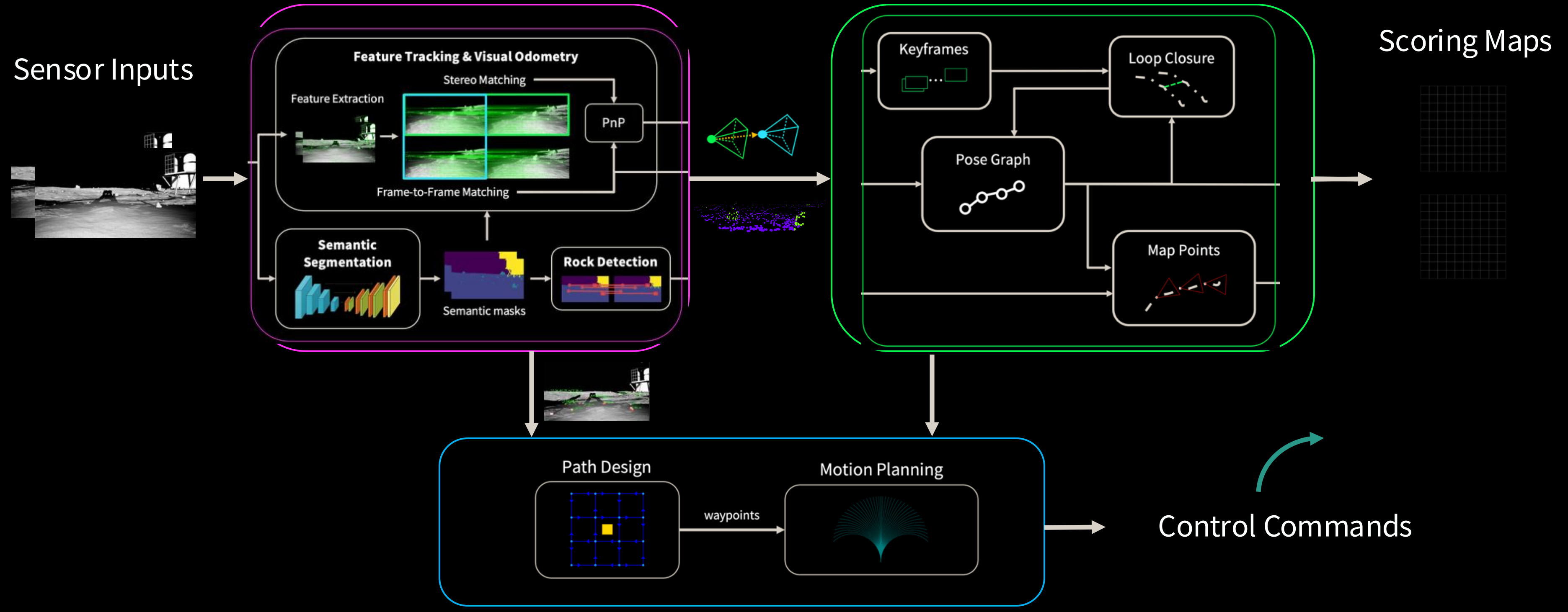
Plan safe path to current waypoint while avoiding rocks

- **Sample** **candidate arcs** parameterized by angular velocity
- **Score and sort arcs** based on distance to waypoint
- **Select** **safe arc** which does not intersect with **rocks**
  - ➢ *Triggers backup maneuver if stuck or no safe arcs found*



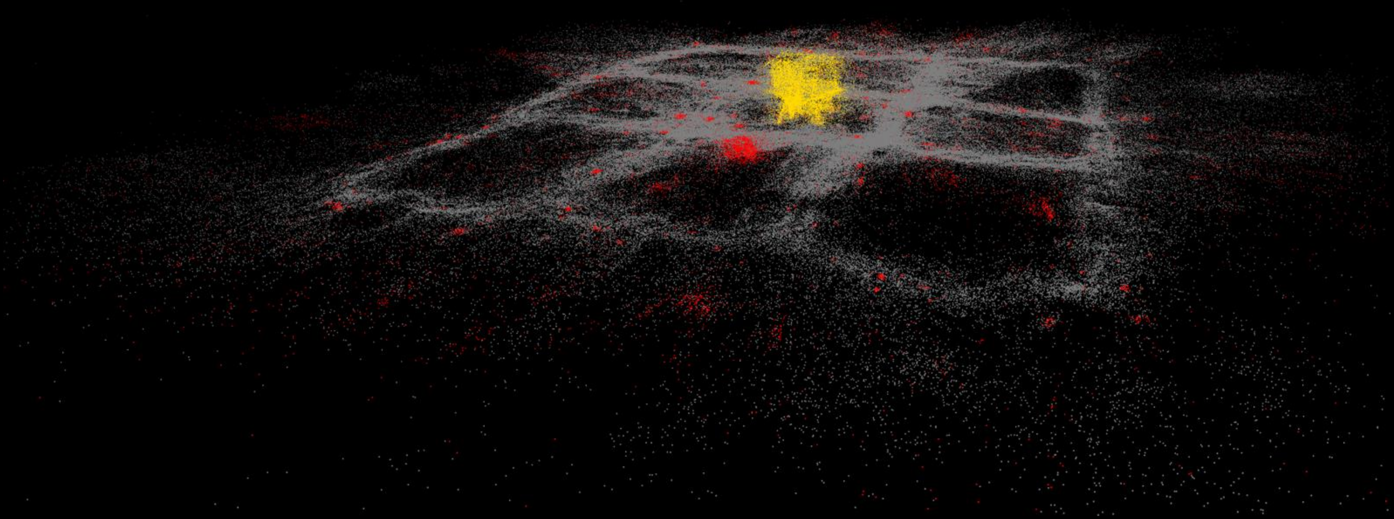


Autonomy Stack

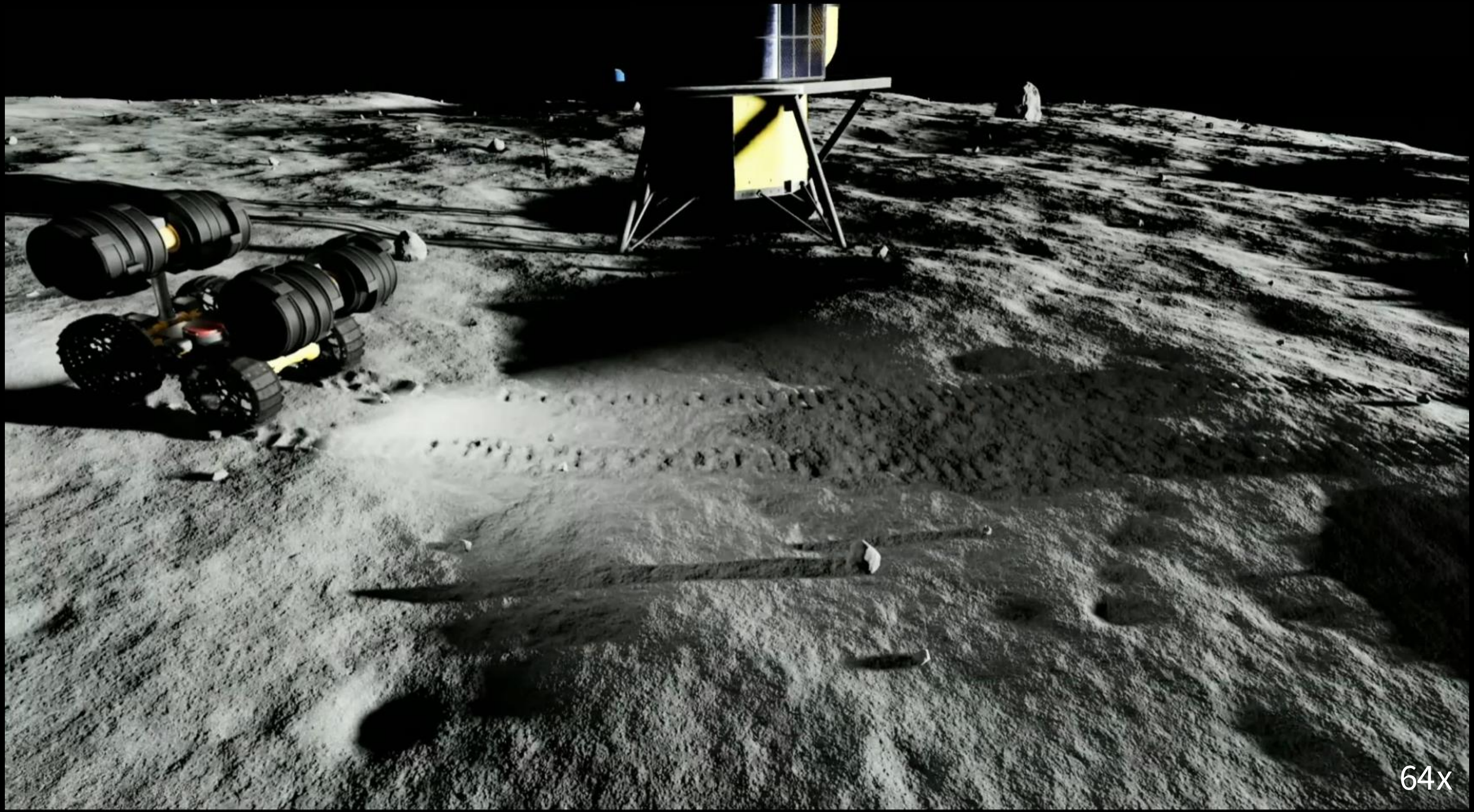


Outline

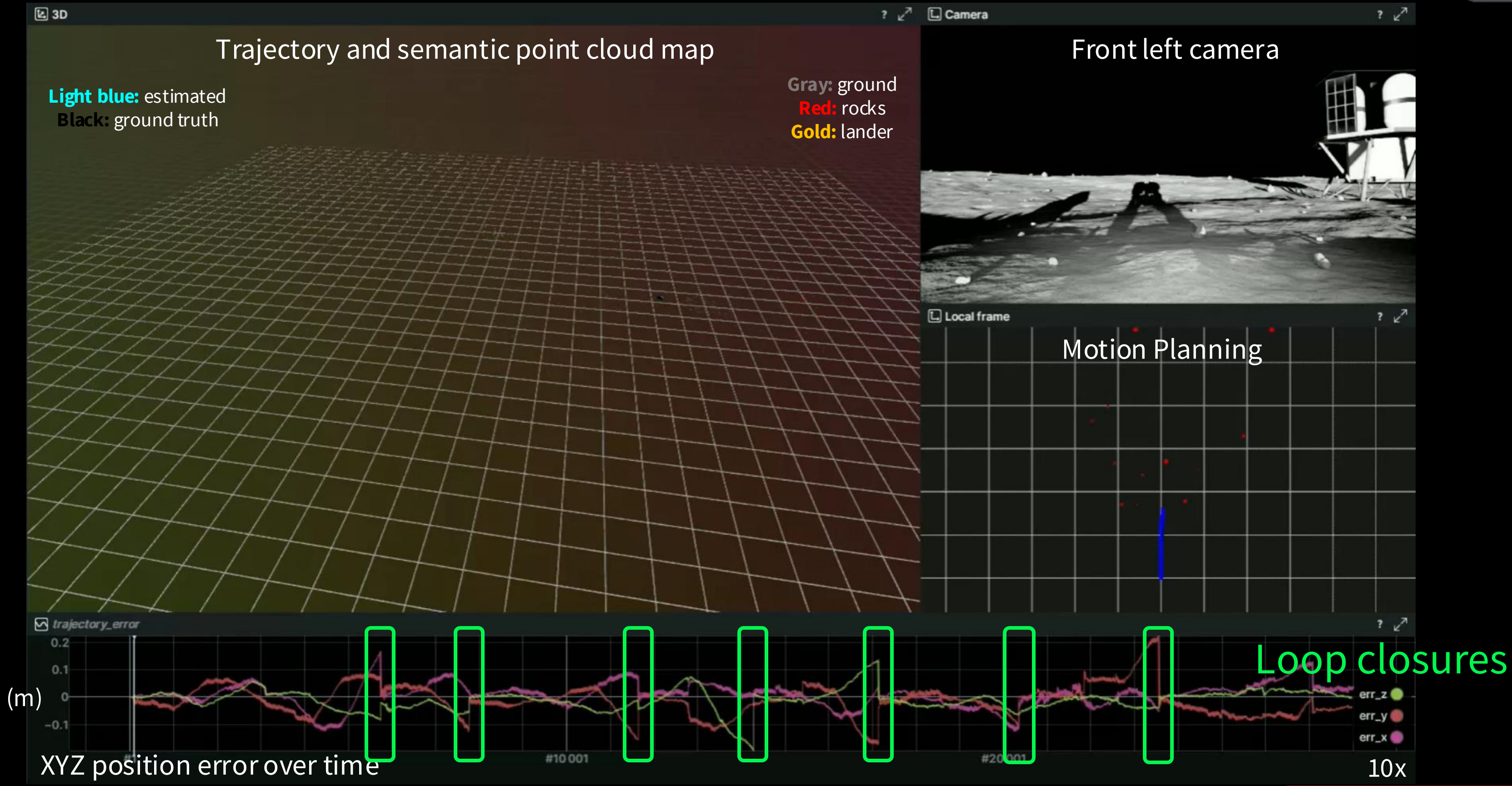
- Approach
  - Perception
  - SLAM
  - Planning
- **Results**
- Conclusion



Simulator Recording



Rerun Visualization



Stanford University

Competition Results

Rank	Participant team	Total score (↑)	Geometric Map score (↑)	Rocks Map score (↑)	Mapping Productivity score (↑)	Localization score (↑)
1	NAV Lab	571.800	111.400	60.400	250.000	150.000
2	MAPLE	544.400	91.700	52.700	250.000	150.000
3	Moonlight	543.700	96.900	46.800	250.000	150.000
4	LunatiX	440.100	24.400	15.600	250.000	150.000
5	AIWVU	438.500	22.000	16.500	250.000	150.000
6	Lunar Explorers	433.400	17.300	16.100	250.000	150.000
7	Lunar Pathfinders	420.400	4.800	15.700	250.000	150.000
8	Rose-Hulman Institute of Technology LAC	400.000	0.000	0.000	250.000	150.000

Local Testing

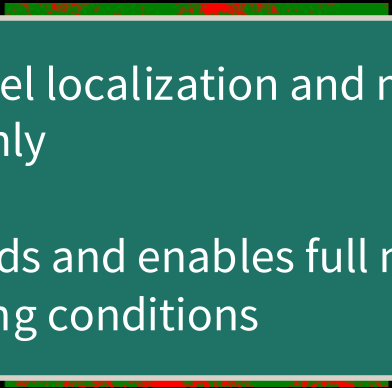
Performance over different rock distributions (presets), initial rover locations

Preset	Trajectory RMSE (m)	Geometric Map Score (max 300)	Rocks Map Score (max 300)	Total Score (max 1000)
1				
1				
1				
1				
1				
2	0.0379	272.3	155.2	827.5
3	0.0605	200.8	146.1	746.9
4	0.0612	190.2	154.8	745.0
5	0.0510	224.7	150.6	775.3

Our SLAM consistently achieves cm-level localization and mapping with vision only

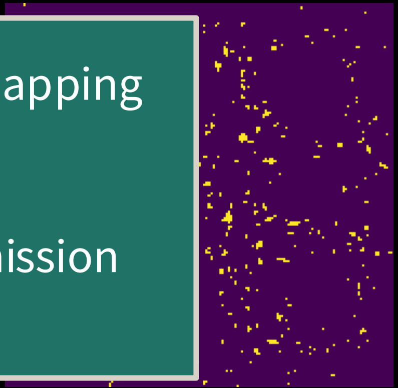
Motion planning reliably avoids hazards and enables full mission completion under varying conditions

Geometric map



90% of cells mapped within 5 cm error

Rock map

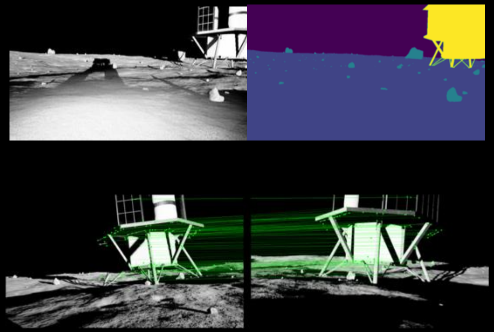


Contributions

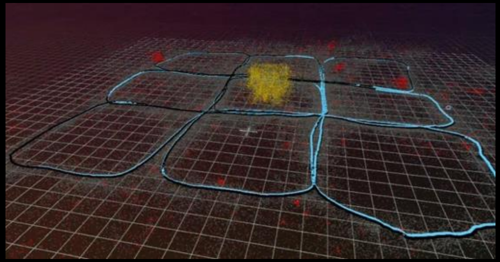
- Develop a full autonomy stack for a lunar surface mission with:
  - Learning-based perception under harsh conditions
  - Lightweight vision-only pose graph SLAM
  - Trajectory design for loop closure and coverage
- Extensive testing and validation in high fidelity simulation
- Placed 1st in the competition

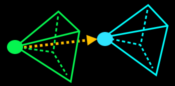
Contributions

Learning-based Perception under Harsh Conditions

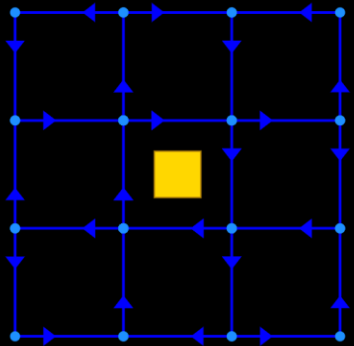


Lightweight Vision-Only Simultaneous Localization and Mapping **(SLAM)**





Trajectory Design for Loop Closure and Coverage



Acknowledgements

NAV LAB

NASA

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

CATERPILLAR®

EMBODIED AI
FOUNDATION

Related Presentations from the NAV Lab

Session F1: K. Iiyama et al., “Constellation Design and Staged Deployment for the Lunar Navigation Satellite System.”

Session F1: G. Casadesus Vila et al., “Lunar Surface Station to Support Lunar Positioning, Navigation, and Timing Services.”

Session B3: K. Coimbra et al., “Single-Satellite Doppler-Based Localization for Lunar Rovers in Motion.”

Full Stack Navigation, Mapping, and Planning for the Lunar Autonomy Challenge

Thank you!

Questions?

website

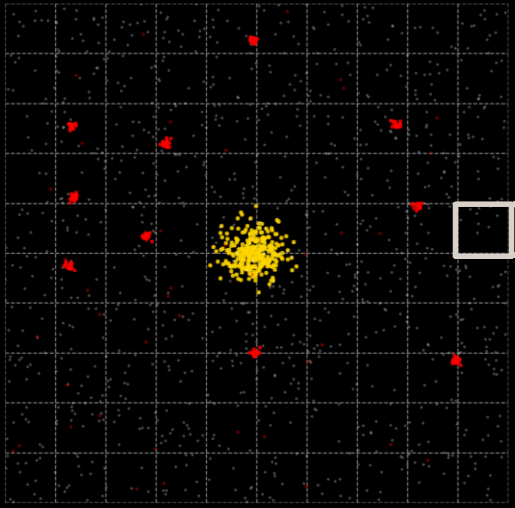
code

Adam Dai • addai@stanford.edu

Map Generation

Given the semantic point cloud from SLAM, how to produce the final geometric and rock maps?

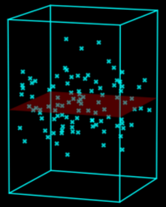
Bin points into 2D grid cells



(not to scale)

For each cell:

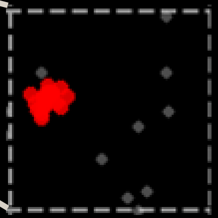
Geometric Map



$$z_{\text{cell}} = \text{median}(z_i)$$

z_i: height of points in the cell

Rock Map

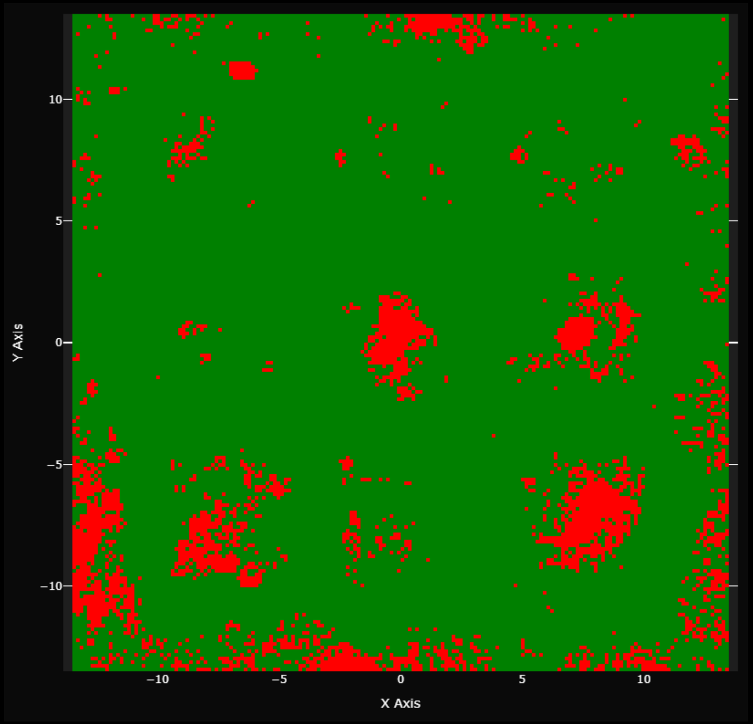


$$P(\text{rock}) = \frac{\#\ \text{rock points}}{\#\ \text{total points}}$$

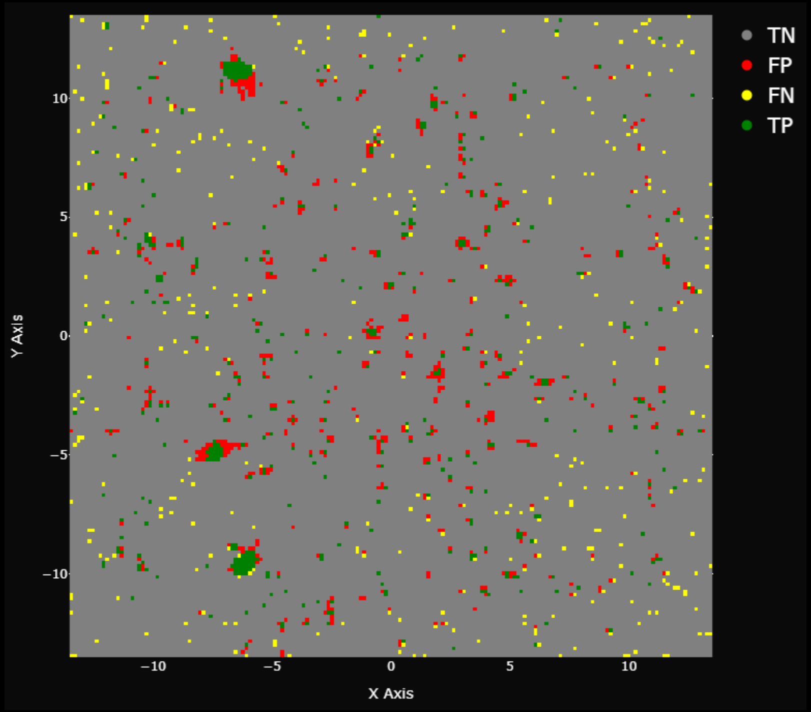
$$o_{\text{cell}} = \begin{cases} 1 & P(\text{rock}) \geq 0.5 \\ 0 & P(\text{rock}) < 0.5 \end{cases}$$

Robustness to outliers and noisy detections

Final Scoring Maps



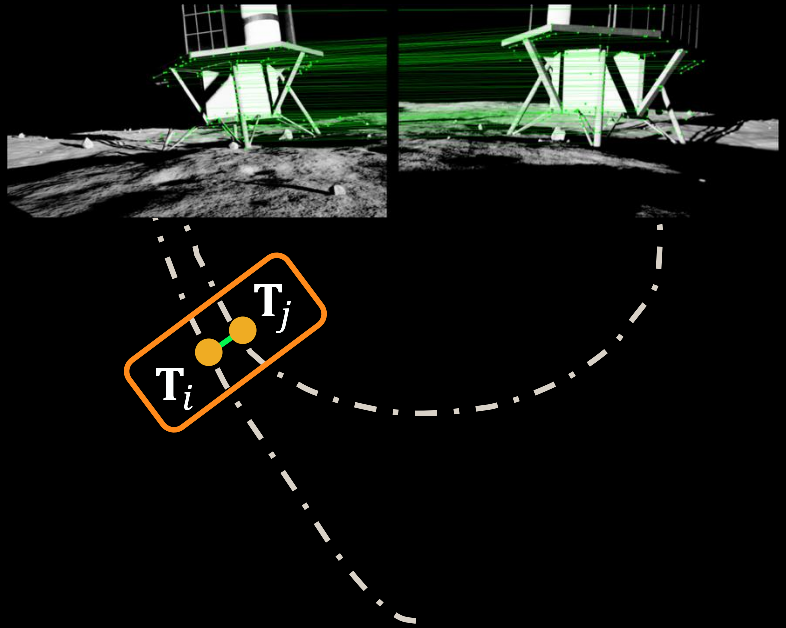
Geometric
(score: 269.6)



Rock
(score: 153.6)

Loop Closure

Detect similar viewpoints and add relative pose constraints to correct drift



Proximity check:

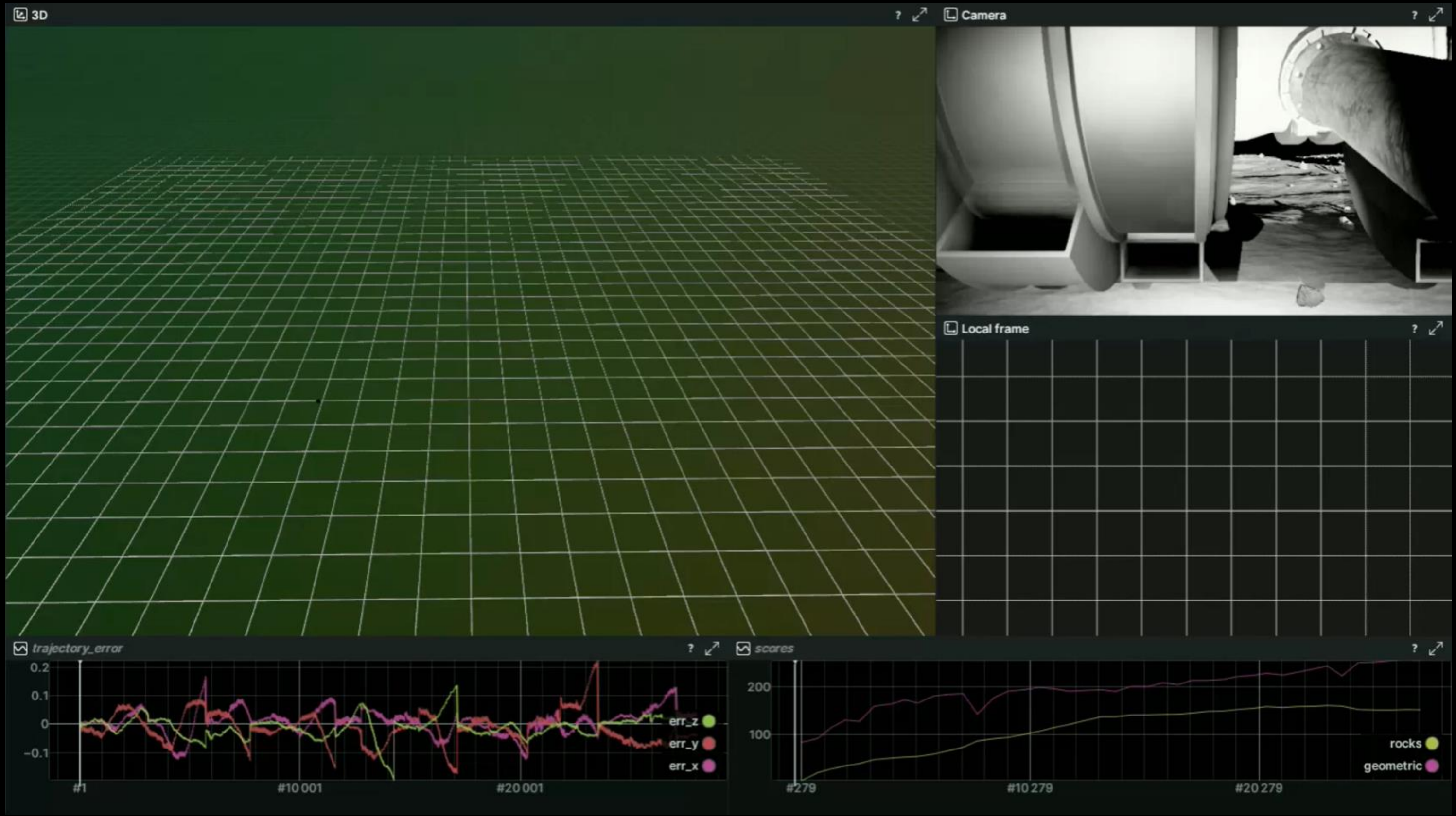
- Distance: L_2 norm
- Angle: rotation matrix error

Relative pose estimation:

- Stereo triangulation and PnP (same as VO)

Loop closure factor added to graph and graph is re-optimized

Rerun Visualization (static)



Backup Maneuver

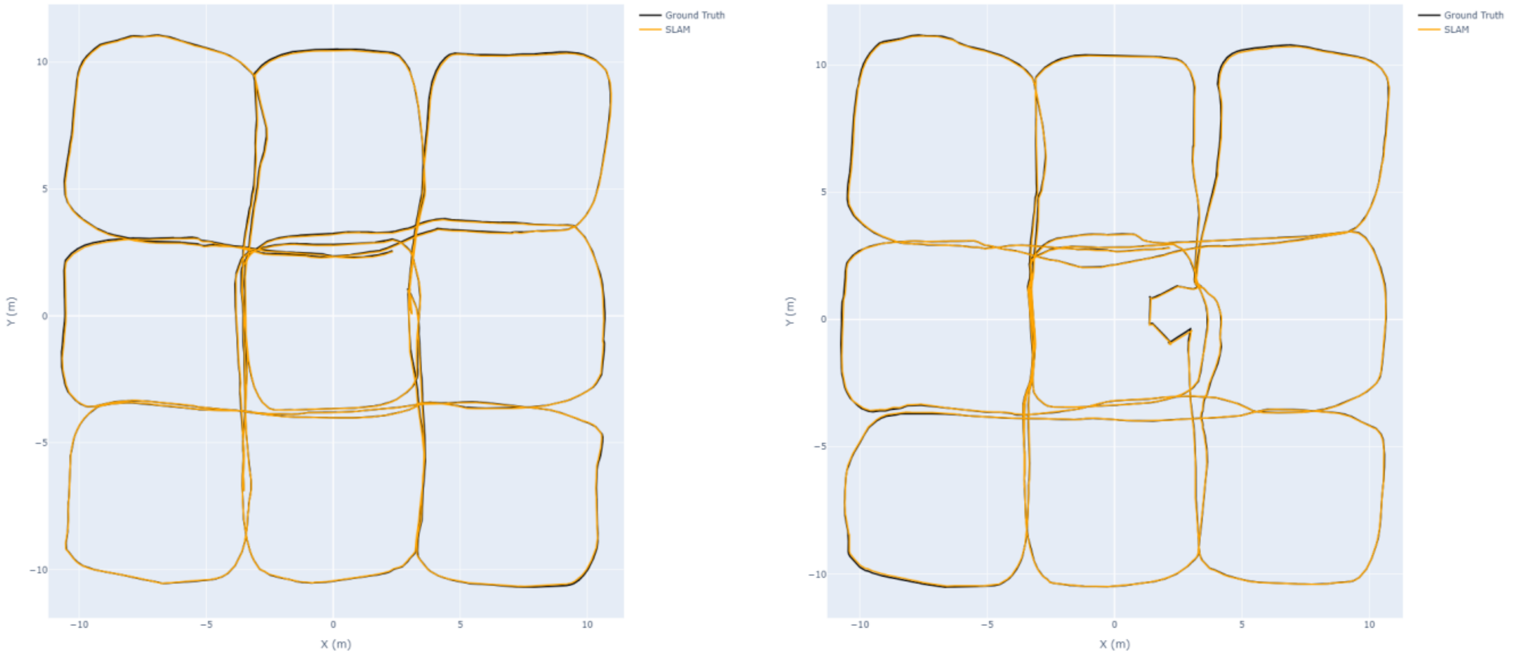
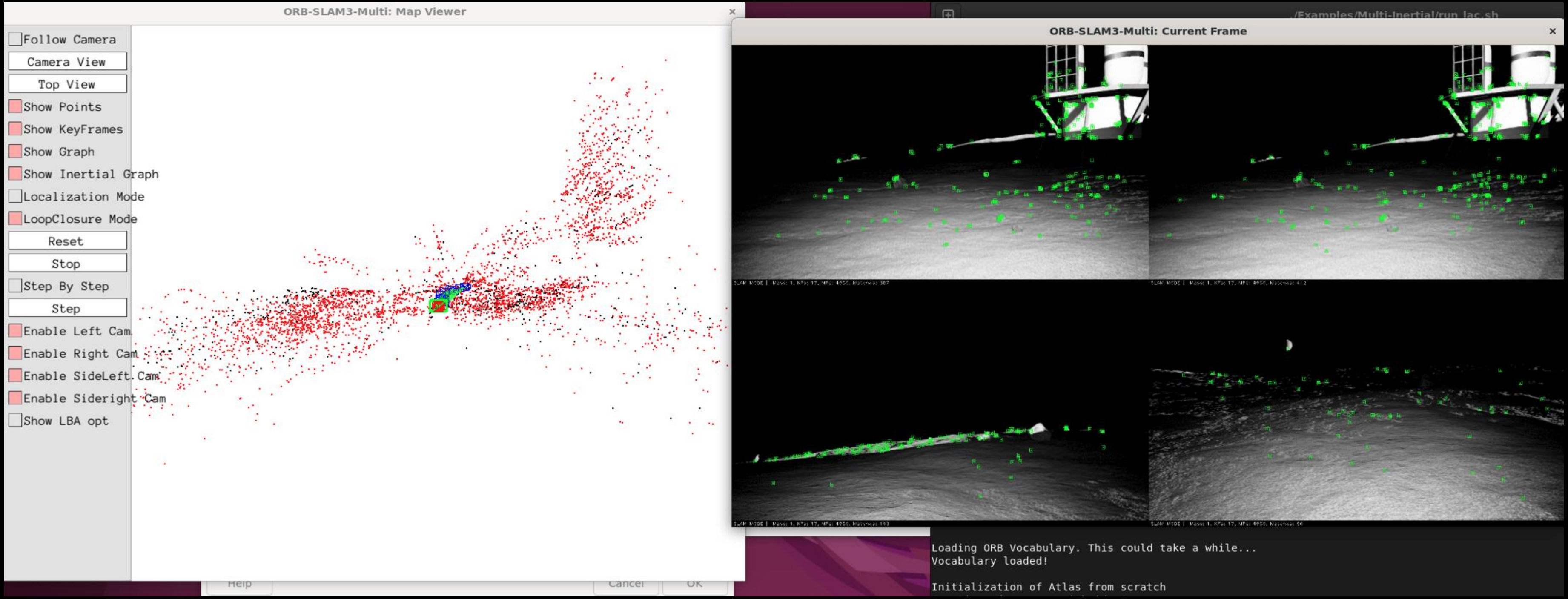


Figure 10: 2D trajectories from two different runs. In the right plot, multiple backup maneuvers were triggered as the planner disengaged from local obstacles. Despite this, our SLAM maintains low localization error through the entire trajectory.

ORB-SLAM3


ORB-SLAM3-Multi: Map Viewer
Follow Camera
Camera View
Top View
Show Points
Show KeyFrames
Show Graph
Show Inertial Graph
Localization Mode
LoopClosure Mode
Reset
Stop
Step By Step
Step
Enable Left Cam
Enable Right Cam
Enable SideLeft Cam
Enable Sideright Cam
Show LBA opt
ORB-SLAM3-Multi: Current Frame
Loading ORB Vocabulary. This could take a while...
Vocabulary loaded!
Initialization of Atlas from scratch

Competition Spiral

SuperPoint and LightGlue