A Dual-Simulator Framework for Physics Based Locomotion and Vision Based Navigation of Planetary Rovers

Umut Kurt[1], Ege Ecevit[1], Denge Uzel[1], Büşra Burhan[2], Mustafa Mert Ankaralı[1], Uluç Saranlı[1]

***Abstract*— Autonomous mobility for micro rovers is becoming critical for lunar exploration, where robots must operate under extreme conditions such as low gravity, abrasive regolith, and rough terrain. Because reproducing these conditions on Earth is difficult and resource-intensive, simulation is widely used for repeatable and cost-effective experimentation. However, existing simulators typically prioritize either physically accurate contact dynamics or photorealistic sensing, making it difficult to evaluate locomotion and perception within a unified lunar environment. This creates a gap for integrated frameworks that simultaneously support contact-rich rover dynamics and realistic perception-based autonomy evaluation. To address this limitation, we propose a Dual-Simulator framework that combines MuJoCo and Isaac Sim within a spatially consistent lunar environment. MuJoCo is used for legged rover locomotion, gait control, and contact-dynamics analysis, while Isaac Sim provides photorealistic rendering and sensor simulation along the same rover trajectory transferred from MuJoCo. In Isaac Sim, synchronized RGB-D data are generated with spatial randomization to create diverse synthetic lunar datasets.**

I. Related Work

A. Simulation Environments

Rather than relying on a single criterion, robotic simulators are commonly categorized based on task-specific priorities, such as physical accuracy and visual realism, into three main groups: physics-oriented, photorealism-oriented, and hybrid.

1) Physics-Oriented: Physics-oriented simulators such as Gazebo support general rigid-body dynamics [1], [2], while specialized platforms extend these capabilities to extraterrestrial navigation and complex terramechanics for lunar construction [3], [4]. However, these frameworks face severe computational challenges during complex collision and deformable terrain calculations [4]. To overcome these limitations, MuJoCo utilizes a constrained based architecture that guarantees real-time computational efficiency and numerical stability, making it highly suitable for contact-rich robotic control [5].

2) Photorealism-Oriented: Unreal Engine and Unity are widely used for perception and synthetic data generation [6], [7], [8], [9], [10]. LunarSim and DUST extend these rendering capabilities to lunar environments [11], [12]. However, the built-in physics infrastructures of such engines are not designed for high-precision robotic contact modeling. Therefore, although these platforms are powerful for visual tasks, they are not sufficient on their own for robotic scenarios that require precise rigid-body physics and contact analysis.

[1]Middle East Technical University, Turkey
[2]Hacettepe University, Turkey

3) Hybrid: Hybrid simulators aim to balance visual fidelity and physical realism. Platforms such as OmniLRS and SensorSim incorporate advanced rendering and terrain interaction capabilities for lunar robotics research [13], [14], [4]. Isaac Sim [15] further combines rigid-body dynamics, RTX-based sensor simulation, providing a strong foundation for perception-oriented robotic validation.

Existing simulators still struggle to jointly provide accurate contact dynamics and photorealistic perception. This limits the evaluation of vision-based robotic control, particularly for unconventional rovers with custom leg morphologies that depend on complex ground contact interactions.

B. Lunar Surface Perception and Navigation

Autonomous lunar exploration requires reliable perception and localization because lunar missions operate without GPS infrastructure and must support long-range navigation in visually challenging environments [3], [16]. Since purely relative localization methods accumulate drift over time, recent research increasingly focuses on surface-aware localization and mapping frameworks.

1) Datasets and Benchmark Environments: Recent datasets and benchmark environments support lunar perception and SLAM research through photorealistic rendering, orbital terrain modeling, and IMU-supported annotations. OmniLRS, and RSIM generate visually detailed lunar scenes using ray-tracing or path-tracing pipelines [13], [17], while MoonAnything improves geometric and photometric consistency through orbital digital elevation models and physically based rendering [18]. LunaPolaris and POLAR-Sim additionally provide geometric annotations and inertial measurements for localization research [17], [19]. However, these frameworks mainly target perception and navigation evaluation rather than contact-rich robotic interaction.

2) Localization and Navigation Architectures: SLAM systems play a key role in autonomous lunar navigation under conditions such as extreme illumination changes, deep shadows, and sparse surface features [20]. Several lunar navigation frameworks have been proposed for both relative and global localization. Lo-SLAM performs target-oriented localization using image matching and segmentation [21], while LunarNav estimates global position through crater-map matching [3]. ShadowNav further extends crater-based localization to permanently shadowed regions using active illumination and particle-filter-based estimation [16].

Overall, existing studies separately address either contact-rich simulation or perception-oriented lunar navigation. However, the joint evaluation of realistic contact dynamics and photorealistic perception remains limited in the literature. To address this gap, the proposed framework integrates MuJoCo for precise contact-rich physics with Isaac Sim for photorealistic rendering and sensor simulation, enabling realistic vision-based robotic evaluation in lunar environments.

II. Methods

A. System Overview

The proposed framework is illustrated in the diagram in Figure 1. The framework leverages MuJoCo for accurate locomotion dynamics and gait control, while using Isaac Sim to generate visual and depth data under realistic lunar lighting conditions.

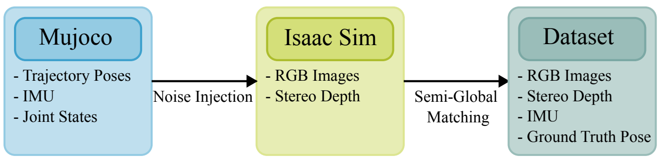


Fig. 1. Dual-Simulation data generation pipeline.

B. Physics Simulation in MuJoCo

MuJoCo is selected as the physics engine for locomotion simulation due to its efficient and accurate rigid body contact dynamics which is crucial for modeling the ground contacts of spoke-leg model of our rover. MuJoCo's constrained based contact solver provides stable simulation of the repeated impact and release events at each spoke tip, while its generalized coordinate formulation enables precise computation of joint torques and reaction forces even under low gravity conditions. These properties make MuJoCo well suited for generating physically realistic motion trajectories.

We model the legged micro rover in MuJoCo and develop a phase based gait controller for stable locomotion over the simulated lunar terrain. Leg morphology is inspired from RHex robot [22], each leg consists of three spokes, a rigid multi-spoke structure driven by a single motor, where spokes alternately contact the ground as legs rotate. The lunar terrain is obtained from the OmniLRS LargeScale simulation environment, and imported into the MuJoCo as surface terrain by a high-resolution mesh conversion.

The controller implements a periodic walking gait across four legs, where each leg's motion is governed by an independent normalized phase variable $\phi_i \in [0, 1)$. A shaping function maps the raw phase to a spoke-angle trajectory, which parameterized by the duty factor (the fraction of the cycle spent in stance), foot width (the angular displacement per stance phase) and smoothing coefficients for flight and stance segments of the trajectory. This approach provides flexibility in continuous tuning gait characteristics without redefining the underlying trajectory profile.

Leg phase offsets are specified via a configurable gait array, enabling trot or other multi-leg coordination patterns. In our experiments, robot's specific gait pattern is set to be diagonal trot where off-diagonal legs having a phase offset of 0.5. During locomotion, each leg's phase advances at a rate proportional to the commanded speed. Turning is realized differentially by introducing a duty factor offset between left and right legs, which causes asymmetric stance duration while ensuring the phase offsets between the legs to be kept constant. Control loop consists of: (i) phase error feedback term that corrects motor tracking errors which ensures the physical leg angles remains synchronized, (ii) a feedforward velocity term obtained by the slope of the shaping function, and (iii) and a velocity feedback term.

Throughout the simulation, we obtained the following data at each timestep (i) **body pose trajectory**: position and orientation of the rover chassis in the world frame; (ii) **joint states**: per-leg angles and angular velocities; and (iii) **IMU data**: simulated body frame linear acceleration and angular velocities. These signals are passed to IsaacSim for photorealistic rendering and serve as ground truth data for evaluating SLAM and navigation algorithms.

To simulate realistic sensor degradation we have augmented the simulated IMU outputs with a standard noise model following the continuous time formulation of [23]. The accelerometer and gyroscope measurements are corrupted by additive white noise and a slowly drifting bias

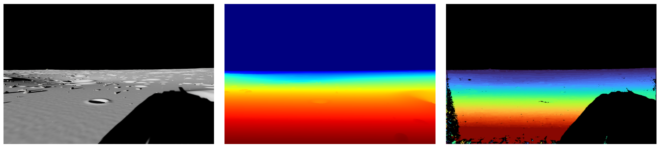
$$\begin{aligned} \tilde{\mathbf{a}}(t) &= \mathbf{a}(t) + \mathbf{b}_a(t) + \mathbf{n}_a(t), \\ \tilde{\boldsymbol{\omega}}(t) &= \boldsymbol{\omega}(t) + \mathbf{b}_g(t) + \mathbf{n}_g(t) \end{aligned} \tag{1}$$

where $\mathbf{a}(t)$ and $\boldsymbol{\omega}(t)$ are the true linear acceleration and angular velocity, $\mathbf{n}_a(t)$, $\mathbf{n}_g(t)$ are zero mean Gaussian white noise with spectral densities σ_a^2, σ_g^2, and the bias evolves as a random walk driven by coefficients σ_{b_a}, σ_{b_g}. These parameters are set to values representative of an IMU.

C. Photorealistic Rendering in Isaac Sim

While MuJoCo provides accurate contact modeling and a physical foundation for locomotion, it is limited in terms of photorealism. In our framework, we use Isaac Sim to generate realistic vision and depth data required to evaluate the autonomous navigation capabilities of rovers under lunar analog conditions.

An important contribution of our framework is that, instead of using scene depth as seen from that camera position, we take left and right stereo pairs and apply a Semi-Global Matching (SGM) [24]. This process introduces artifacts such as occlusion holes, matching noise in low-texture regions, and temporal inconsistencies similar to a real stereo camera hardware. The resulting disparity map is illustrated in Figure 2, along with the corresponding RGB frame and default disparity map provided by Isaac Sim. This design choice creates a challenge for the robustness of the SLAM pipeline by exposing it to realistic failures inherent in physical sensors. Isaac Sim provides a realistic lunar environment for modeling extreme illumination conditions on the Moon. The absence of an atmosphere reduces light scattering, resulting in directional illumination and sharp shadow boundaries. Within our framework, we randomize the incident light direction, focusing on low elevation ranges between $0°$ to $15°$ to replicate conditions near the lunar poles. These



(a) RGB Observation (b) Isaac Sim Disparity Map (c) SGM Disparity Map

Fig. 2. An example RGB observation taken from Isaac Sim, corresponding stereo depth output, and SGM stereo matching output.

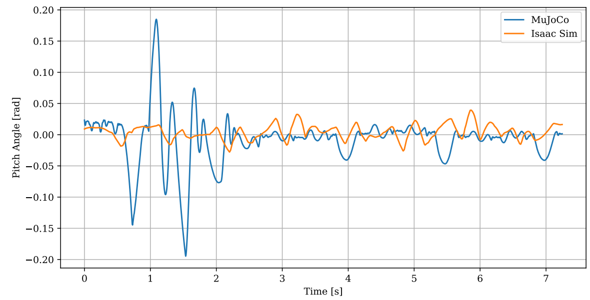


Fig. 3. Pitch-angle response of the rover body on flat terrain under identical high-level velocity commands.

conditions are known to degrade the performance of vision-based navigation algorithms due to elongated shadows and high dynamic range. We systematically vary the simulation parameters to generate a diverse and physically consistent dataset.

The output of this stage is synchronized data collected by our rover's sensor suite over simulated trajectories. This data is valuable for evaluating perception algorithms under challenging visual conditions encountered on lunar exploration missions. Our dataset includes photorealistic stereo RGB-D data, noisy IMU measurements, and 6-DOF ground truth pose information.

III. Results

The results include a comparison of rover locomotion dynamics in MuJoCo and Isaac Sim under identical conditions, as well as an evaluation of autonomous lunar navigation performance using ORB-SLAM3 under different simulated terrain conditions.

A. *Dual-Simulator Locomotion Validation*

To motivate the use of MuJoCo for locomotion in Dual-Simulator design, we compared the simulated pitch response of the rover body in MuJoCo and Isaac Sim under the same high-level velocity commands on perfectly flat terrain. As shown in Figure 3, the MuJoCo response contains a sharp transient at the beginning of the motion. This behavior is caused by the initial body jerk when the robot starts moving from rest, which is also observed on the real robot, and is captured by the low-level gait controller implemented directly in MuJoCo. This transient is followed by a repetitive oscillatory pattern caused by the periodic contact sequence of the spoke-like legs. On the other hand, the Isaac Sim response is smoother and less periodic, suggesting that the velocity-command interface and internal motion control in Isaac Sim do not capture the same contact dynamics.

B. *Autonomous Navigation and SLAM Evaluation*

To further demonstrate the ability of our framework for benchmarking autonomous lunar navigation, we conducted a representative experiment using state-of-the-art SLAM algorithm ORB-SLAM3 [25]. Table I summarizes the ORB-SLAM3 evaluation under monocular and stereo configurations for three simulated terrain conditions. In the baseline terrain, both configurations achieve similar sub-decimeter accuracy, with ATE RMSE values of $0.0519\ m$ for monocular and $0.0507\ m$ for stereo. The feature-sparse terrain produces the largest error in both cases, indicating that reduced

TABLE I

ORB-SLAM3 evaluation under monocular and stereo configurations.

Scenario	Mono. ATE [m]	Mono. Scale	Stereo ATE [m]	Stereo Scale
Baseline Terrain	0.0519	23.068	0.0507	0.967
Feature-Sparse Terrain	0.1365	48.532	0.1445	0.833
Low Albedo Terrain	0.0511	4.745	0.0263	0.965

structure and fewer distinctive visual landmarks make pose estimation more challenging. In the low-albedo terrain, stereo achieves the lowest error of $0.0263\ m$, showing that metric stereo constraints can remain effective even under reduced appearance contrast.

The scale estimates also highlight the expected difference between monocular and stereo SLAM. The monocular configuration requires Sim(3) alignment and produces large arbitrary scale factors due to scale ambiguity, while the stereo configuration estimates scales close to one in the baseline and low-albedo terrains. Overall, these results show that the proposed simulation framework can generate controlled visual conditions with different levels of SLAM difficulty, making it suitable for evaluating visual navigation algorithms in lunar-like environments.

IV. Conclusion

This work presents a dual-simulator framework with high-fidelity locomotion of MuJoCo with photorealistic perception of Isaac Sim for lunar exploration missions.

Experimental results demonstrated that MuJoCo captures locomotion dynamics that are not clearly represented in Isaac Sim under identical motion commands, supporting the use of MuJoCo for locomotion-focused studies. In addition, ORB-SLAM3 experiments under multiple simulated lunar terrain conditions showed that the framework can generate controlled visual environments with varying levels of navigation difficulty, enabling systematic evaluation of visual SLAM performance.

Overall, the proposed framework provides a flexible platform for studying both rover locomotion and autonomous navigation in lunar-like environments. Future work will focus on integrating additional sensor modalities, more realistic terrain interaction models, reinforcement learning-based navigation and locomotion policies, and sim-to-real gap evaluation within the Dual-Simulator framework.

ACKNOWLEDGMENT

The authors acknowledge the support of the Challenging Environment Exploration Robot for Intelligence (CHERI) project conducted at the Middle East Technical University (METU) Robotics and Artificial Intelligence Technologies Application and Research Center (ROMER) within an international collaboration involving Zhejiang University, STAR.VISION, and the China National Space Administration (CNSA). This study was supported in part by the ADEP project under grant no. ADEP-313-2025-11656.

REFERENCES

[1] N. Koenig and A. Howard, "Design and use paradigms for gazebo, an open-source multi-robot simulator," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2149–2154, 2004.

[2] C. E. Agüero, N. Koenig, I. Chen, H. Boyer, S. Peters, J. Hsu, B. Gerkey, S. Paepcke, J. L. Rivero, J. Manzo, E. Krotkov, and G. Pratt, "Inside the virtual robotics challenge: Simulating real-time robotic disaster response," *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 2, pp. 494–506, 2015.

[3] S. Daftry, Z. Chen, Y. Cheng, S. Tepsuporn, S. Khattak, L. Matthies, B. Coltin, U. Naal, L. M. Ma, and M. Deans, "Lunarnav: Crater-based localization for long-range autonomous lunar rover navigation," in *2023 IEEE Aerospace Conference*, pp. 1–15, 2023.

[4] M. Linde, D. Lindmark, S. Ålstig, and M. Servin, "A simulation framework for autonomous lunar construction work," 2025.

[5] Google DeepMind, "Mujoco documentation: Computation," 2026.

[6] A. Juliani, V.-P. Berges, E. Teng, A. Cohen, J. Harper, C. Elion, C. Goy, Y. Gao, H. Henry, M. Mattar, and D. Lange, "Unity: A general platform for intelligent agents," 2020.

[7] S. Shah, D. Dey, C. Lovett, and A. Kapoor, "Airsim: High-fidelity visual and physical simulation for autonomous vehicles," 2017.

[8] J. Tremblay, T. To, and S. Birchfield, "Falling things: A synthetic dataset for 3d object detection and pose estimation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2119–21193, 2018.

[9] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proceedings of the 1st Annual Conference on Robot Learning* (S. Levine, V. Vanhoucke, and K. Goldberg, eds.), vol. 78 of *Proceedings of Machine Learning Research*, pp. 1–16, PMLR, 13–15 Nov 2017.

[10] X. Puig, E. Undersander, A. Szot, M. D. Cote, T.-Y. Yang, R. Partsey, R. Desai, A. Clegg, M. Hlavac, S. Y. Min, V. Vondruš, T. Gervet, V.-P. Berges, J. M. Turner, O. Maksymets, Z. Kira, M. Kalakrishnan, J. Malik, D. S. Chaplot, U. Jain, D. Batra, A. Rai, and R. Mottaghi, "Habitat 3.0: A co-habitat for humans, avatars, and robots," in *The Twelfth International Conference on Learning Representations*, 2024.

[11] D. Pieczyński, B. Ptak, M. Kraft, and P. Drapikowski, "Lunarsim: Lunar rover simulator focused on high visual fidelity and ros 2 integration for advanced computer vision algorithm development," *Applied Sciences*, vol. 13, no. 22, p. 12401, 2023.

[12] L. Bingham, J. Kincaid, B. Weno, N. Davis, E. Paddock, and C. Foreman, "Digital lunar exploration sites unreal simulation tool (dust)," in *2023 IEEE Aerospace Conference*, pp. 1–12, 2023.

[13] A. Richard, J. Kamohara, K. Uno, S. Santra, D. van der Meer, M. Olivares-Mendez, and K. Yoshida, "Omnilrs: A photorealistic simulator for lunar robotics," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 16901–16907, 2024.

[14] N. M. Batagoda, B.-H. Chen, H. Zhang, R. Serban, and D. Negrut, "A physics-based sensor simulation environment for lunar ground operations," in *2025 IEEE Aerospace Conference*, pp. 1–20, 2025.

[15] NVIDIA, "Isaac sim documentation," 2023. Version 4.5.0.

[16] D. Atha, R. M. Swan, A. Cauligi, A. Bettens, E. Goh, D. Kogan, L. Matthies, and M. Ono, "Shadownav: Autonomous global localization for lunar navigation in darkness," *IEEE Transactions on Field Robotics*, vol. 1, pp. 213–230, 2024.

[17] B.-H. Chen, P. Negrut, T. Liang, N. Batagoda, H. Zhang, and D. Negrut, "Polar-sim: Augmenting nasa's polar dataset for data-driven lunar perception and rover simulation," *IEEE Aerospace and Electronic Systems Magazine*, vol. 41, no. 1, pp. 4–18, 2026.

[18] C. Grethen, Y. Shi, S. Gasparini, and G. Morin, "Moonanything: A vision benchmark with large-scale lunar supervised data," in *Proceedings of the ACM Multimedia Systems Conference 2026*, MMSys '26, (New York, NY, USA), pp. 457–463, Association for Computing Machinery, 2026.

[19] D. van der Meer, U. Wong, and M. A. Olivares Mendez, "Lunapolaris: A stereo camera, point cloud and imu dataset for future lunar exploration in polar regions," in *Proceedings of the related conference/workshop if applicable*, (Sendai, Japan), 2026. Dataset and publication record.

[20] Z. Liang, H. Long, Z. Zhu, Z. Cao, J. Yi, Y. Ma, E. Liu, and R. Zhao, "High-precision disparity estimation for lunar scene using optimized census transform and superpixel refinement," *Remote Sensing*, vol. 16, no. 21, 2024.

[21] Y. Tian, X. Wan, S. Zhang, J. Zuo, Y. Shao, B. Liu, and M. Yang, "Lo-slam: Lunar target-oriented slam using object identification, relative navigation, and multilevel mapping," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1–16, 2025.

[22] U. Saranli, M. Buehler, and D. E. Koditschek, "Rhex: A simple and highly mobile hexapod robot," *The International Journal of Robotics Research*, vol. 20, no. 7, pp. 616–631, 2001.

[23] P. T. Furgale, J. Rehder, and R. Y. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1280–1286, 2013.

[24] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, p. 328–341, Feb. 2008.

[25] C. Campos, R. Elvira, J. J. Gómez Rodríguez, J. M. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual-inertial and multi-map slam," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.