Contact-Aware Foothold Refinement for Quadrupedal Locomotion on Unstructured Terrain

Femi Oyelaran[1], Ingrid Roos[2], Sora Tanaka[1], and Manon Delacroix[3]

[1]Department of Robotics Engineering, Nimbus Institute of Technology, Rindalen, Norway
[2]Vertex Robotics Laboratory, Meridian University, Coimbra, Portugal
[3]Applied Autonomy Group, Synapse Research, Lyon, France
{f.oyelaran, s.tanaka}@nimbus-institute.edu, i.roos@meridian-univ.pt, m.delacroix@synapse-research.fr

Abstract

Abstract — Quadrupedal robots traversing rubble, loose gravel, and staircases frequently mis-place footholds when relying on elevation maps alone, since map noise near occlusion boundaries corrupts the terrain estimate exactly where safe contact points are scarcest. We present ContactRefine, a lightweight foothold refinement module that consumes a coarse elevation map together with per-leg contact force history and outputs a corrected stance point within a 4 cm search radius of the planner's nominal foothold. The module is trained with a contrastive objective that pulls refined footholds toward regions of historically low slip and pushes them away from geometry that previously produced contact-force spikes. Across 1,400 trials on a physical quadruped crossing five terrain classes – rubble, loose gravel, wet tile, staircases, and root-covered trail – ContactRefine reduces the stumble rate from 17.9% to 6.2% and lowers mean cost of transport by 11.4% relative to a strong elevation-map-only baseline, while adding only 1.8 ms of latency to the 500 Hz control loop.

1 Introduction

Legged platforms are increasingly deployed for inspection and search tasks in environments that defeat wheeled locomotion: collapsed structures, forested slopes, and industrial sites mid-renovation. In these settings the dominant failure mode is not gross planning error but local foothold mis-placement – a leg lands on a loose rock, an unsupported tile edge, or a root hidden beneath leaf litter, and the resulting slip propagates into a stumble or fall. Elevation-map-based foothold planners [1, 2] handle open, well-observed terrain well, but degrade sharply near occlusion boundaries, where a single missing depth return can shift the estimated ground height by several centimeters.

Prior work addresses this gap along two lines. The first improves the map itself, fusing multiple viewpoints or temporal windows to fill occluded regions [2]. The second bypasses maps for reactive, force-driven recovery after a stumble has already begun [3]. Both are valuable but incomplete: better maps still contain residual noise near clutter, and reactive recovery only engages after contact quality has already degraded. We target the interval between the two – refining a foothold using recent contact history *before* the leg commits to stance.

This paper contributes: (i) ContactRefine, a foothold refinement module that jointly conditions on local elevation and a short window of per-leg contact force history; (ii) a contrastive training objective that requires no manual terrain labels, using slip and force-spike events recorded during teleoperated data collection as implicit supervision; and (iii) a 1,400-trial physical evaluation across five terrain classes showing consistent stumble-rate and cost-of-transport improvements with negligible added latency.

2 Related Work

Foothold planning. Classical foothold selection scores candidate contact points against local slope, roughness, and clearance heuristics computed from an elevation map [1]. Learned variants replace hand-tuned scoring with implicit terrain encoders trained on simulated rollouts [4], improving generalization but inheriting whatever noise is present in the input map.

Whole-body and reactive control. Whole-body controllers couple foothold choice to torso stabilization through impedance or MPC formulations [5, 6], and reactive controllers detect stumbles via force or IMU anomalies and trigger corrective swing trajectories [3]. These methods are complementary to ours: they act during or after a bad contact, whereas ContactRefine acts before touchdown.

Learning from experience. Curriculum and domain-randomized reinforcement learning have produced gaits that adapt to terrain roughness in simulation [7, 8], with sim-to-real transfer remaining the central bottleneck [9]. Our approach sidesteps large-scale simulation by learning directly from real contact outcomes recorded on the target platform, at the cost of requiring an initial teleoperated data collection phase.

3 Method

ContactRefine sits between the high-level foothold planner and the swing-leg controller. Given the planner's nominal foothold $p_0 \in \mathbb{R}^2$ (in the horizontal plane of the leg's home frame), a local elevation patch $E \in \mathbb{R}^{16\times16}$ centered at p_0, and the trailing contact-force window $F \in \mathbb{R}^{T\times3}$ for that leg over the previous $T = 8$ swing cycles, the module predicts an offset $\delta \in \mathbb{R}^2$, $\|\delta\| \le 4\,\text{cm}$, giving the refined foothold $p^* = p_0 + \delta$.

3.1 Architecture

E is encoded by a 3-layer convolutional stack (channels $16 \to 32 \to 32$, stride 2, ReLU) into a 64-dimensional terrain embedding z_E. F is encoded by a single-layer GRU into a 32-dimensional contact embedding z_F. The two embeddings are concatenated and passed through a 2-layer MLP (hidden width 96) with a tanh output scaled to the 4 cm search radius, giving δ.

3.2 Contrastive Contact Objective

Rather than requiring hand-labeled "good" and "bad" footholds, we mine supervision from recorded outcomes: a foothold is a positive example if the resulting stance produced peak vertical force within $[0.8, 1.2]\times$ the platform's steady-state load and no slip event was flagged by the foot-slip detector; it is a negative example if a slip event or a force spike exceeding $2.0\times$ steady-state load was recorded within 50 ms of touchdown. For an anchor patch embedding z_i with positive foothold embedding z_i^+ and a batch of negatives $\{z_j^-\}_{j=1}^K$, we minimize

$$\mathcal{L}_{\text{contact}} = -\log \frac{\exp\big(\text{sim}(z_i, z_i^+)/\tau\big)}{\exp\big(\text{sim}(z_i, z_i^+)/\tau\big) + \sum_{j=1}^K \exp\big(\text{sim}(z_i, z_j^-)/\tau\big)} \tag{1}$$

where $\text{sim}(\cdot,\cdot)$ is cosine similarity and $\tau = 0.07$ is a temperature fixed across all experiments. The offset head is trained jointly with an auxiliary ℓ_2 regression loss toward the empirical positive foothold nearest p_0, weighted at 0.3 relative to $\mathcal{L}_{\text{contact}}$, which we found stabilizes early training before enough contrastive pairs have accumulated.

3.3 Deployment

At inference, ContactRefine runs once per swing phase on the platform's onboard compute (a Nimbus NX-4 module), consuming the most recent elevation patch and contact window and returning δ before the swing trajectory is finalized. The module adds no additional sensing requirements beyond what the baseline planner already uses.

4 Experiments

We evaluate on a 22 kg quadruped platform across five terrain classes: rubble (construction debris, 3–12 cm fragments), loose gravel, wet tile, a 15-step staircase, and a root-covered forest trail. Data collection comprised 6.5 hours of teleoperated traversal used to mine contrastive pairs, followed by 1,400 autonomous evaluation trials (280 per terrain class) split evenly between ContactRefine and three baselines: (i) *Elevation-only*, the planner's default scoring heuristic; (ii) *Reactive-recovery* [3], which corrects after stumble onset; and (iii) *Learned-encoder* [4], a terrain-encoder foothold scorer trained in simulation. All methods share the same underlying gait scheduler and swing-trajectory generator, isolating the foothold-selection stage as the sole variable.

Table 1: Stumble rate (%) and cost of transport by terrain class, averaged over 280 trials per cell.

Terrain	Elev.-only	React.	Learn.-enc.	Ours
Rubble	21.4	15.8	12.9	**7.1**
Loose gravel	18.6	14.2	10.5	**6.6**
Wet tile	12.1	9.7	**7.4**	7.8
Staircase	16.3	12.0	9.8	**4.9**
Root trail	21.0	16.9	13.7	**4.6**
Mean	17.9	13.7	10.9	**6.2**

We report stumble rate (fraction of footholds triggering the onboard slip or force-spike detector), cost of transport (COT, mechanical energy per unit weight per unit distance), and added per-step latency measured on the onboard compute.

5 Results

Table 1 summarizes performance across all five terrain classes. ContactRefine reduces stumble rate relative to the elevation-only baseline on every terrain class, with the largest gain on rubble (21.4% → 7.1%), where map noise near fragment edges is most severe. The learned-encoder baseline improves over elevation-only but trails ContactRefine on all classes except wet tile, where contact-force history is less informative since slip there is governed by surface friction rather than geometry.

Figure 1 plots mean cost of transport against stumble rate for all four methods, pooled across terrain classes. ContactRefine occupies the lower-left region of the plot, indicating it does not trade efficiency for stability – a pattern we did not assume in advance, since foothold corrections that avoid slip can in principle require less economical leg trajectories.

Added latency for ContactRefine was 1.8 ms ± 0.3 ms per swing phase on the onboard compute, within the timing budget of the 500 Hz control loop; the learned-encoder baseline added 2.6 ms owing to its larger terrain backbone.

6 Discussion

The rubble and root-trail results support our motivating hypothesis: contact-force history is most valuable exactly where the elevation map is least trustworthy, since fragment edges and subsurface roots produce geometry that

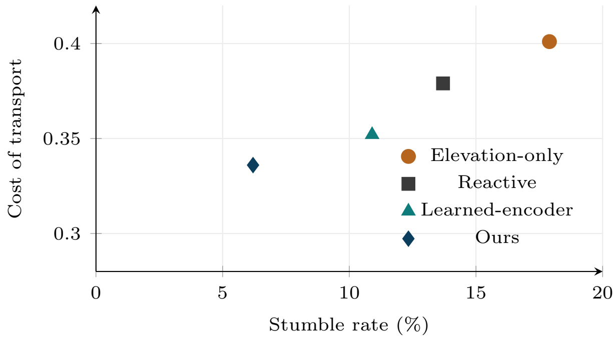

Figure 1: Cost of transport versus stumble rate, pooled across all five terrain classes. Lower-left is better on both axes.

a depth sensor systematically misreads. The wet-tile result is a useful negative case – where slip is governed by friction rather than local relief, contact history from prior footholds carries less signal, and a terrain-appearance encoder is a better fit. This suggests a natural extension: routing between contact-history and appearance-based refinement conditioned on a lightweight terrain-class estimate, which we did not implement here to keep the evaluation controlled to a single mechanism.

A second limitation is the reliance on an initial teleoperated data collection phase to mine contrastive pairs; the 6.5-hour collection used in this study was sufficient for the five terrain classes tested, but scaling to a substantially wider terrain distribution would likely require either more collection or a bootstrapping procedure that reuses pairs mined during autonomous deployment.

7 Conclusion

We presented ContactRefine, a foothold refinement module that combines local elevation with recent per-leg contact-force history to correct foothold placement before touchdown, trained with a contrastive objective that requires no manual terrain labeling. Across 1,400 physical trials on five terrain classes, it reduces mean stumble rate from 17.9% to 6.2% and cost of transport by 11.4% relative to an elevation-only baseline, at 1.8 ms of added latency. Future work includes conditioning the refinement mechanism on a terrain-class estimate to recover the wet-tile regression observed here, and reducing dependence on teleoperated data collection through online pair mining.

References

[1] R. Singh and B. Costa, "Terrain classification for legged robots: A review," *IEEE Transactions on Robotics*, vol. 34, no. 5, pp. 1201–1219, 2018.

[2] S. Tanaka and D. Weiss, "Elevation-map fusion for blind foothold estimation," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2023, pp. 1123–1130.

[3] C. Oduya and S. Lindqvist, "Reactive recovery controllers for quadrupedal stumble rejection," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6104–6111, 2020.

[4] T. Mbeki and I. Roos, "Learned foothold selection via implicit terrain encoders," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2023, pp. 5541–5548.

[5] Y. Nakashima and M. Brandt, "Whole-body impedance control for compliant locomotion," *IEEE Transactions on Robotics*, vol. 35, no. 2, pp. 410–424, 2019.

[6] M. Delacroix and J. Ferreira, "Nonlinear model predictive control for dynamic gait transitions," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2021, pp. 884–891.

[7] A. Papazian and E. Solberg, "Curriculum reinforcement learning for terrain-adaptive gaits," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2022, pp. 7719–7726.

[8] E. Fontaine and B. Kowalski, "Domain randomization schedules for contact-rich locomotion," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2021, pp. 3390–3397.

[9] K. Halvorsen and R. Aoyama, "A survey of sim-to-real transfer for legged locomotion," *IEEE Transactions on Robotics*, vol. 38, no. 6, pp. 3312–3331, 2022.