Perceptive Behavior Foundation Model: Adapting Human Motion Priors to Robot-Centric Terrain

Zifan Wang[1,2] **Yizhao Li**[1] **Teli Ma**[1,2] **Qiang Zhang**[4]
Yudong Fan[1] Hao Xu[1] Shuo Yang[1,*] Junwei Liang[2,*]
[1]Mondo Robotics
[2]The Hong Kong University of Science and Technology (Guangzhou)
[4]Artificial General Intelligence Institute, University of Science and Technology of China
[*]Corresponding authors
Project page: https://acodedog.github.io/perceptive-bfm/

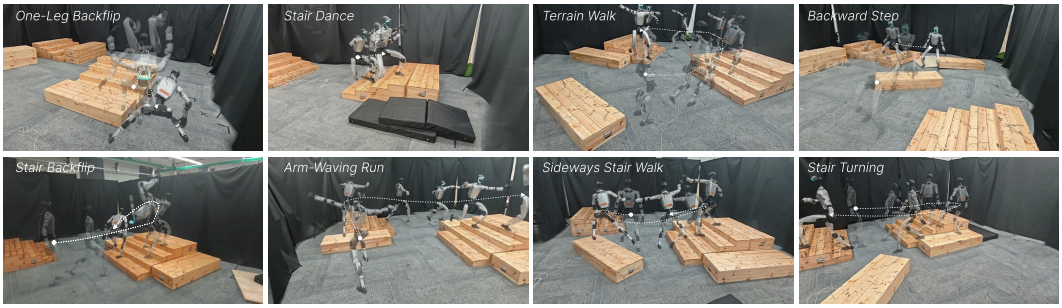

Figure 1: **Single-policy terrain grounding.** A single Perceptive BFM tracks diverse flat-ground human-motion commands while adapting them to randomly placed robot-side terrains. Robot-centric perception adjusts footholds, swing clearance, posture, and contact timing online.

Abstract: Humanoid behavior foundation models aim to acquire reusable whole-body control policies from broad human motion priors, enabling a single controller to produce diverse and expressive behaviors. However, existing motion-centric foundation policies largely assume that the reference motion is already physically compatible with the robot's surroundings. This assumption breaks when the demonstrator, operator, and robot inhabit different environments: a human motion may specify the intended behavior, but not the footholds, clearance, body height, or contact timing required by the robot's local terrain. We introduce *Perceptive Behavior Foundation Model* (Perceptive BFM), a terrain-aware humanoid control framework that grounds human motion priors in robot-centric perception. The model preserves raw kinematic motion references as the behavioral interface, while using local terrain observations to adapt contacts, posture, and timing. To provide scalable terrain supervision, we develop *terrain-conformal reference synthesis* (TCRS), which converts locomotion-oriented human motion clips into terrain-consistent references through contact-aware foothold construction, foot-geometry-aware swing optimization, support-aware root reconstruction, collision repair, and multi-point inverse kinematics. We then train a blind adapted-reference teacher and transfer its terrain-conformal behavior to a deployed raw-reference student through target-frame action alignment. The student is an identity-gated Transformer tracker whose terrain features enter through residual pathways initialized to preserve the motion-tracking prior and trained to produce local corrections only when needed. Across quantitative TCRS analysis, matched-compute training ablations, and qualitative real-robot rollouts, a single policy tracks locomotion, stylistic motions, acrobatic maneuvers, and motion-capture teleoperation across stairs, slopes, sparse supports, recessed obstacles, grass, and irregular indoor/outdoor terrain. The results indicate that robot-centric perception can turn human motion priors into terrain-compatible whole-body behavior without changing the raw motion command interface.

Keywords: perceptive humanoid control, behavior foundation models, motion tracking

1 Introduction

Humanoid control is rapidly shifting from isolated, task-specific skills toward generalist behavior foundation models and large motion-tracking policies. Recent whole-body trackers reproduce diverse motions with a single learned controller [1, 2, 3], while newer foundation-control systems further scale behavior priors, command encoders, policy capacity, and downstream interfaces [4, 5, 6, 7]. This trend is important because it turns human motion into a reusable command interface: instead of designing a separate controller for each maneuver, the robot can learn a broad prior over human-like whole-body intent.

This progress, however, exposes a hidden assumption: most reference-centric formulations ask how accurately the robot can reproduce the supplied motion, not how that motion should be physically grounded in the robot's own environment. These are different problems. A flat-ground walking reference does not specify stair footholds; a teleoperator in a control room does not encode sparse supports or recessed terrain at a remote site; a clean-floor demonstration does not tell the robot how much swing clearance is needed to avoid obstacles. The reference conveys intent and style, but it may not be a terrain-valid trajectory in the robot's local world. This is the operator–environment mismatch: the human supplies the desired behavior, while the robot must resolve terrain-specific contacts, body height, balance, and timing from its own perception.

Existing work leaves a gap between two successful directions. General motion-tracking and behavior-foundation models excel at diverse, expressive whole-body behavior, but their command interfaces obscure the need for environment-conditioned contact adaptation. Perceptive locomotion and parkour policies use height maps, depth, and terrain observations to traverse obstacles or sparse footholds [8, 9, 10], but they are organized around traversal skills, motion matching, or system-selected maneuvers rather than preserving an arbitrary human motion command. The missing capability is the perceptual grounding of behavior priors: using robot-centric terrain perception to reinterpret what a human motion command physically requires.

We introduce *Perceptive Behavior Foundation Model* (Perceptive BFM), a single-policy framework for terrain-aware humanoid motion tracking. In this work, the foundation interface is the kinematic motion reference: a unified command representation that lets one policy reuse broad whole-body motion priors while grounding their terrain-dependent realization in robot-centric perception. The underlying control problem is *perceptive motion tracking*: given a raw kinematic reference, robot proprioception, and local terrain observation, the policy must generate whole-body actions that are both behaviorally faithful and environmentally feasible. The raw reference remains the deployment command. Terrain perception provides only the local realization: footholds, clearance, posture, and contact timing.

The method is built around a staged *Perceptive Motion Tracking* (PMT) training algorithm (Figure 2). First, an offline *terrain-conformal reference synthesis* (TCRS) module converts raw motion clips and sampled height fields into terrain-consistent supervision. Rather than presenting this module as a pair of low-level optimizers, we formulate it as structured reference synthesis: contact-aware foothold construction, foot-geometry-aware swing optimization in a mid-foot frame, support-aware root reconstruction, collision repair, and multi-point Jacobian IK. Second, a blind Transformer teacher learns to track the synthesized terrain-conformal references. Third, a vision-conditioned student receives the original raw reference and a local terrain scan, and imitates the teacher through target-frame action alignment, which expresses the teacher's effective joint target in the student's raw-command frame. Finally, PPO fine-tunes the student with identity-gated terrain residuals, updating the transferred tracking prior conservatively while learning local perception-conditioned corrections.

The architecture follows the same separation of roles. Command and proprioceptive histories form a motion-tracking latent, while terrain observations enter through zero-initialized intent and action-

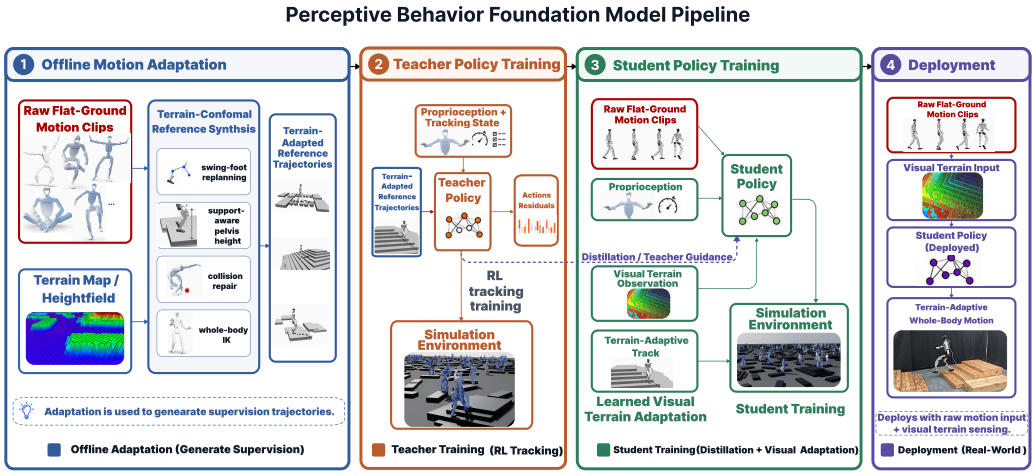

Figure 2: **Perceptive BFM overview.** TCRS synthesizes terrain-conformal references *offline only*; it is never queried at deployment. A blind teacher learns adapted-reference tracking on this supervision; the deployed identity-gated Transformer student receives the raw reference and a robot-centric terrain scan, and learns local residual corrections through target-frame action alignment. The deployment command remains the raw kinematic reference.

residual pathways, so at initialization the student behaves as a raw-reference tracker and terrain features only contribute through residual branches learned during distillation and fine-tuning. Our experiments reflect the current evidence hierarchy: quantitative measurements isolate the quality of TCRS supervision and the effect of architecture/training ablations, while real-robot indoor, outdoor, and motion-capture mismatch rollouts provide qualitative deployment coverage. Figure 1 shows the same policy accepting diverse flat-ground commands and adapting each to a randomly placed terrain layout.

Contributions. This paper introduces *Perceptive BFM*, a motion-reference-conditioned humanoid behavior foundation model that grounds human motion priors in robot-centric terrain. We make three contributions:

- We introduce *TCRS*, a scalable offline synthesis pipeline that converts raw human motion and sampled height fields into terrain-consistent supervision through contact-aware foothold construction, foot-geometry-aware swing optimization, support-aware root reconstruction, collision repair, and multi-point leg IK.
- We propose *PMT*, a raw-reference teacher–student algorithm for deploying terrain-aware behavior without changing the command interface. A blind teacher tracks TCRS references, while a vision student receives the original raw reference; target-frame action alignment transfers the teacher's terrain-conformal behavior into the student's raw-reference action frame.
- We design and evaluate an identity-gated Transformer policy for single-policy terrain grounding of broad human motion priors. The policy initializes as a raw-reference tracker and learns terrain-conditioned residual corrections from robot-centric perception, enabling the same command interface to support locomotion, expressive motions, acrobatics, and mocap-based operator–environment mismatch across diverse robot terrains.

2 Related Work

Generalist humanoid behavior and expressive tracking. DeepMimic and AMP established reinforcement-learning approaches for tracking motion clips and imitation priors [11, 12]. Recent humanoid systems scale these ideas to richer corpora and robot embodiments: H2O and OmniH2O learn teleoperation and universal tracking policies [13, 1]; expressive whole-body controllers reproduce diverse human motions on hardware [2]; and PHC, PULSE, HOVER, HumanPlus, reference-guided motion tracking, OmniXtreme, and SONIC broaden the representation, command encoder, model

scale, or residual refinement stack [14, 15, 3, 16, 17, 18, 7]. Robot foundation and behavior-foundation models push the same trend toward reusable whole-body policies and promptable control [4, 5, 6]. Perceptive BFM builds on this tracking tradition but targets a complementary failure mode: broad motion priors can encode human intent without specifying terrain-valid contacts in the robot's world.

Perceptive locomotion and terrain-aware whole-body control. Terrain-aware legged policies use height maps, depth, or images to traverse obstacles and sparse footholds, avoid collisions, and maintain safe, comfortable locomotion in dynamic environments [19, 20, 21, 22, 23, 24, 8, 25, 26]. Recent whole-body parkour systems integrate exteroception into motion tracking or distill depth policies from expert controllers, enabling contact-rich skills such as climbing, vaulting, rolling, or obstacle traversal [9, 10]. These systems demonstrate the value of perception for terrain interaction. Our setting is stricter in a different sense: the user-provided motion remains the behavior to preserve, so perception should ground the command rather than replace it with a generic terrain skill or an autonomously selected parkour maneuver.

Generated, repaired, and terrain-conditioned references. A closely related interface lets the system choose, generate, or repair a feasible reference. Generator–tracker systems synthesize terrain-conditioned motions online [27]; navigation-oriented reference-guided RL modulates trajectories to be consistent with terrain geometry [28]; motion-matching parkour retrieves and chains skill clips before distilling a perceptive controller [10]; and physics-consistent tracking pipelines roll out privileged policies to filter infeasible references before training a deployable tracker [29]. These methods are strong when feasibility, reference repair, or autonomous skill selection is the primary objective. We instead keep the raw clip fixed at deployment and use terrain-conformal references only as offline supervision for a raw-reference student.

Retargeting, reference synthesis, and residual distillation. Human-to-robot retargeting and trajectory optimization convert demonstrations into robot-executable references [30, 31, 32, 33, 34]. Teacher–student training, curriculum learning, and residual policies are common tools for transferring privileged or easier-to-train behavior into deployable policies [35, 36, 37, 38, 39, 40, 41, 42]. Our use of these tools is interface-specific: TCRS synthesizes terrain-conformal supervision, the teacher tracks that adapted reference, the student receives the original raw reference, and identity-gated residuals make terrain correction an explicitly learned departure from the raw-command tracker.

3 Method

Perceptive BFM is trained with a staged *PMT* algorithm. The key interface contract is that the raw motion reference remains the deployment command: terrain-conformal references are used to supervise learning, but they are not supplied to the final policy at test time. PMT has four stages. Stage 1 synthesizes terrain-conformal references offline with TCRS. Stage 2 trains a blind teacher on those synthesized references. Stage 3 distills a vision student that receives the raw reference and robot-centric terrain observation. Stage 4 fine-tunes the student with PPO while updating the transferred motion-tracking prior conservatively.

3.1 Problem Formulation

Let $\mathbf{o}_t^{\mathrm{prop}}$ denote robot proprioception, $\mathbf{o}_t^{\mathrm{vis}}$ denote the local terrain observation, and $\mathbf{m}_{t:t+H}^{\mathrm{raw}}$ denote a future window of the raw kinematic reference. In our implementation, the deployment command is represented by target joint positions and velocities, together with local motion-anchor displacements used by the tracker. Perceptive motion tracking asks for a policy

$$\mathbf{a}_t = \pi_\theta\left(\mathbf{o}_t^{\mathrm{prop}}, \mathbf{o}_t^{\mathrm{vis}}, \mathbf{m}_{t:t+H}^{\mathrm{raw}}\right), \tag{1}$$

whose action preserves the behavior encoded in the reference while adapting contact, posture, and local feasibility to the robot's terrain. The policy outputs a residual joint-position target

$$\mathbf{q}_t^{\mathrm{pd}} = \mathbf{q}_t^{\mathrm{cmd}} + \mathbf{a}_t, \tag{2}$$

where $\mathbf{q}_t^{\mathrm{cmd}}$ is the command-frame joint reference. The teacher uses $\mathbf{q}_t^{\mathrm{cmd}} = \mathbf{q}_t^{\mathrm{tcrs}}$, the terrain-conformal reference synthesized offline. The deployed student uses $\mathbf{q}_t^{\mathrm{cmd}} = \mathbf{q}_t^{\mathrm{raw}}$. This convention makes the transfer problem explicit: the teacher learns to track an adapted command, while the student must express the same terrain correction around the unmodified user command.

The deployable actor receives projected gravity, base angular velocity, joint positions, joint velocities, previous actions, a 10-step proprioceptive history, a 21-step reference-command window, and a 21-step motion-anchor displacement window. Terrain perception is a torso-centered ray-cast height scanner over a 1.6 m $\times$ 1.0 m region at 0.1 m resolution, represented as a normalized 17×11 height map with a validity mask. Privileged ground-truth body-pose, global-anchor, and base-linear-velocity terms are used only by the teacher, critic, or auxiliary losses, and are removed from the deployable student actor, which instead reads detached estimator outputs (Appendix A). The full PMT training pipeline is given in Appendix A (Alg. 1): TCRS produces paired $(\mathbf{m}^{\mathrm{raw}}, \mathbf{m}^{\mathrm{tcrs}}, \tau)$ data, a blind teacher trains with PPO on $\mathbf{m}^{\mathrm{tcrs}}$, an identity-gated vision student is distilled with target-frame labels from the teacher, and the student is fine-tuned with PPO under raw-reference commands.

3.2 Terrain-Conformal Reference Synthesis

A raw human motion is behaviorally informative but not necessarily terrain-conformal when placed in the robot's environment. We therefore introduce a synthesis operator

$$\mathbf{m}_{1:T}^{\mathrm{tcrs}} = \mathcal{S}_{\mathrm{TCRS}}\left(\mathbf{m}_{1:T}^{\mathrm{raw}}, \tau\right), \tag{3}$$

which converts a raw motion clip and terrain height field $h_\tau(x, y)$ into kinematic supervision for teacher training. TCRS is not intended to solve full contact-rich dynamics; instead, it constructs contact-consistent, smooth, and style-preserving references that make terrain adaptation learnable for the downstream policy.

Contact-aware terrain reference. TCRS first estimates stance and swing intervals from foot height, velocity, and hysteresis thresholds. Stance feet are latched to terrain support surfaces, while swing endpoints inherit the raw liftoff and landing timing. This step produces terrain-aware foot targets without changing the global behavior phase of the input clip.

Foot-geometry-aware swing optimization. For each foot f and swing phase s, TCRS optimizes a virtual mid-foot trajectory rather than the ankle origin. Let $\mathbf{r}^{\mathrm{toe}}$ and $\mathbf{r}^{\mathrm{heel}}$ be toe and heel offsets in the ankle frame, and define

$$\mathbf{r}^{\mathrm{mid}} = \frac{1}{2}(\mathbf{r}^{\mathrm{toe}} + \mathbf{r}^{\mathrm{heel}}), \qquad \mathbf{p}_{f,t}^{\mathrm{mid}} = \mathbf{p}_{f,t}^{\mathrm{ankle}} + \mathbf{R}_{f,t}\mathbf{r}^{\mathrm{mid}}. \tag{4}$$

Planning in the mid-foot frame balances toe and heel clearance near terrain discontinuities. For control knots $\mathbf{Y} = \{\mathbf{y}_k\}_{k=1}^{K}$, the swing objective is

$$\begin{aligned} J_s(\mathbf{Y}; \tau) = {} & \lambda_{\mathrm{ref}} \sum_k \|\mathbf{y}_k - \mathbf{y}_k^{\mathrm{raw}}\|^2 + \lambda_{\mathrm{sm}} \sum_k \|\Delta^2 \mathbf{y}_k\|^2 \\ & + \lambda_{\mathrm{clr}} \sum_k [h_\tau(x_k, y_k) + \delta - z_k]_+^2 + \lambda_{\mathrm{edge}} \Phi_{\mathrm{edge}}(\mathbf{Y}; \tau) \\ & + \lambda_{\mathrm{end}} \|\mathbf{y}_1 - \bar{\mathbf{y}}_1\|^2 + \lambda_{\mathrm{end}} \|\mathbf{y}_K - \bar{\mathbf{y}}_K\|^2. \end{aligned} \tag{5}$$

where $[x]_+ = \max(x, 0)$. The terms preserve the raw swing, encourage smoothness, enforce terrain clearance margin δ, discourage penetration near vertical faces, and keep liftoff/landing endpoints fixed; Φ_{edge} penalizes toe/heel samples whose neighboring terrain queries indicate a vertical height discontinuity within the foot support footprint. We instantiate this optimizer with batched sampling-based trajectory optimization. With perturbations $\epsilon^{(j)}$ and temperature η, the knot update is

$$\mathbf{Y} \leftarrow \mathbf{Y} + \sum_j \frac{\exp(-J_s(\mathbf{Y} + \epsilon^{(j)})/\eta)}{\sum_\ell \exp(-J_s(\mathbf{Y} + \epsilon^{(\ell)})/\eta)} \epsilon^{(j)}. \tag{6}$$

The optimized mid-foot path is then transformed back to ankle, toe, and heel targets for IK.

Support-aware root reconstruction. After foot replanning, TCRS reconstructs root height from the support contacts. For support weights $w_{f,t}$ and adapted foot positions $(x^{\mathrm{tcrs}}_{f,t}, y^{\mathrm{tcrs}}_{f,t})$, a target root height is

$$z^{\star}_{\mathrm{root},t} = \frac{\sum_f w_{f,t}\left[h_\tau(x^{\mathrm{tcrs}}_{f,t}, y^{\mathrm{tcrs}}_{f,t}) + z^{\mathrm{raw}}_{\mathrm{root},t} - z^{\mathrm{raw}}_{f,t}\right]}{\sum_f w_{f,t} + \epsilon}. \tag{7}$$

The value is clamped by leg reachability and smoothed across support transitions. During flight or weak-contact phases, the filter falls back to the raw vertical profile with limited per-frame displacement.

Collision repair and multi-point leg IK. Finally, TCRS repairs lower-leg and foot collisions, attenuates unsupported toe/heel constraints near step edges, and solves a damped support-aware multi-point Jacobian IK problem over the twelve leg joints. Root translation is fixed to the support-aware reconstruction stage above, and root orientation as well as non-leg joints are preserved from the raw reference. The IK residual stacks ankle, toe, and heel point Jacobians for both feet, with support-aware point weights, posture and continuity regularization, penetration penalties, and damped-least-squares regularization. Multiseed fallback and continuity guards reject high-error or discontinuous branches. The full IK objective is given in Appendix A (Eq. (12)). The output is the paired dataset $(\mathbf{q}^{\mathrm{raw}}, \mathbf{q}^{\mathrm{tcrs}}, \tau)$ for teacher training and student distillation.

3.3 Policy Architecture and Training

A detailed module-level diagram of the deployable policy is given in Appendix A.7 (Figure 7); we summarize the key components here.

Blind teacher. The blind teacher uses a Transformer actor–critic with tokenized proprioceptive history and reference-command windows. It receives TCRS references as commands and is trained with PPO to track adapted anchor, orientation, foot-position, and foot-velocity targets, with energy and lateral foot/shin contact penalties. Appendix A gives the observation and reward contract.

Identity-gated vision student. The student inherits the same command/history Transformer and adds a height-map encoder. Denote the pooled command-history latent by $\mathbf{u}_t$ and the visual latent by $\mathbf{z}^{\mathrm{vis}}_t = E_{\mathrm{vis}}(\mathbf{o}^{\mathrm{vis}}_t)$. Terrain affects the actor through two zero-initialized residual pathways. The intent latent is modulated as

$$\mathbf{u}'_t = \mathbf{u}_t + \tanh(\boldsymbol{\alpha}_u) \odot f_u(\mathbf{z}^{\mathrm{vis}}_t), \tag{8}$$

and the action mean is

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}^{\mathrm{base}}_t + \tanh(\boldsymbol{\alpha}_a) \odot f_a([\mathbf{o}^{\mathrm{prop}}_t, \mathbf{z}^{\mathrm{vis}}_t]). \tag{9}$$

Both gate vectors and final residual layers are initialized at zero, so the terrain pathway is inactive at initialization. The student therefore begins as a raw-reference tracker and learns terrain-conditioned corrections only when they improve the tracking objective.

Target-frame distillation and PPO fine-tuning. Because teacher and student act around different command frames, the student cannot imitate the teacher residual directly. We instead distill the teacher's effective PD target, expressed relative to the raw reference:

$$\mathbf{a}^{\star}_t = \left(\mathbf{q}^{\mathrm{tcrs}}_t + \boldsymbol{\mu}^{\mathrm{tea}}_t\right) - \mathbf{q}^{\mathrm{raw}}_t. \tag{10}$$

The student minimizes $\|\boldsymbol{\mu}^{\mathrm{stu}}_t - \mathbf{a}^{\star}_t\|_2^2$. During DAgger-style rollouts, the teacher-control probability is annealed from 1 to 0; when the teacher controls the simulator, the applied action is the aligned target-frame action rather than the teacher's native adapted-reference residual. PPO fine-tuning then uses raw-reference commands, height maps, and auxiliary estimator losses, with a lower learning-rate scale on the transferred backbone than on the terrain encoder, critic, and residual branches.

Table 1: **TCRS reference quality on STEPPING STONES WITH STAIRS.** TCRS is evaluated before any policy rollout, aggregated over 30 motion clips on the same terrain family. Lower is better for all metrics; bold marks the best per column. TCRS targets collision-free, physically feasible swing trajectories, so it leads on the collision-sensitive metrics (penetration depth, clearance violation); collision-agnostic baselines can score better on float rate and smoothness, which do not penalize collision.

Reference	Pen. Depth (cm) ↓	Float Rate (%) ↓	Clear. Viol. (%) ↓	Foot Smooth (m/s^2) ↓	Upper Dev. (cm) ↓
Z-offset (FK projection)	5.48	**12.4**	33.8	15.1	6.51
Cubic Interp + IK	2.69	31.7	14.3	**6.9**	**3.98**
TCRS (ours)	**2.38**	32.3	**7.4**	8.6	4.00

4 Experiments

We evaluate Perceptive BFM at three levels. First, we measure the quality of TCRS outputs before any policy rollout, isolating whether the reference synthesizer produces terrain-conformal supervision (Sec. 4.1). Second, we compare training-time ablations under matched reward, observation, task, and compute budgets (Sec. 4.2). Third, we document real-robot deployment of a single policy across diverse behaviors and terrain layouts (Fig. 1), including mocap-based operator–environment mismatch (Sec. 4.3). The first two levels provide quantitative evidence; the real-robot results provide qualitative deployment coverage rather than repeated-trial statistics.

Setup. TCRS supervision is generated from locomotion-oriented human motion clips on procedural terrain with height offsets in $[-0.10, 0.25]$ m, support widths in $[0.10, 1.00]$ m, stair risers 5–15 cm, treads in $[0.20, 0.35]$ m, and slopes up to $30°$. All variants are trained with PPO on 48 A800 GPUs for 10k iterations under the same task, reward, and observation contract; only the listed network or component changes (Sec. 4.2). Our main run resumes from a checkpoint at $\approx$ 5k with identical optimizer settings, while ablations run as a single segment. Real-robot trials follow Appendix A.8.

4.1 Quality of Terrain-Conformal Reference Synthesis

Table 1 evaluates TCRS on STEPPING STONES WITH STAIRS, aggregated over 30 motion clips. The baselines isolate simpler geometry edits: *Z-offset* forward-kinematically lifts ankles to the local terrain height, while *Cubic Interp+IK* smooths a height-projected ankle path with cubic interpolation and solves single-point IK. TCRS instead replans a mid-foot swing trajectory and reconstructs the body through support-aware root shaping plus multi-point ankle/toe/heel IK, primarily targeting physically feasible, collision-free swing along the path. It therefore leads on the collision-sensitive metrics, cutting penetration depth by 56.6% relative to Z-offset ($5.48 \rightarrow 2.38$ cm) and clearance violation by 48.3% relative to Cubic Interp+IK ($14.3 \rightarrow 7.4\%$). On metrics that do not penalize collision, simpler edits can score better by design: collision-agnostic baselines pin contacts to terrain labels and add no extra swing, so TCRS shows a slightly higher float rate and foot acceleration precisely because it lifts around stair faces to clear vertical geometry (Figure 3).

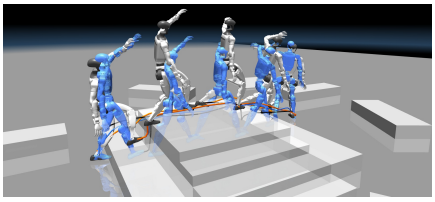

Figure 3: **TCRS trajectory synthesis.** The blue ghost is the raw reference placed on terrain; the opaque robot is the TCRS output. Foot traces compare the sampling-based (model predictive path integral, MPPI) foot-end optimization used in TCRS (yellow), Cubic Interp (blue), and direct terrain-height z-lifting (black).

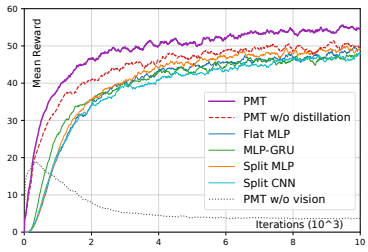

Figure 4: **Training reward diagnostics.**

4.2 Training-Time Ablations

We use training reward as a diagnostic for the PMT architecture, not as a substitute for deployed-rollout performance. All variants share the same task, reward, observation contract, and 48-A800-GPU budget, and we report mean reward over the last 1k iterations. The full PMT model reaches 54.6. Removing terrain perception is the most damaging change by far: the blind variant collapses to 3.6, an order of magnitude below every other configuration, confirming that perception—not capacity—drives terrain grounding. Replacing the Transformer backbone costs 5–8 points (Split MLP 49.0, Flat MLP 48.5, MLP-GRU 47.3, Split CNN 47.0), all far above the blind variant, so the gains are architectural. Removing target-frame distillation drops the reward by about 4.5 points (to 50.1), so aligning the teacher target into the raw-reference frame measurably improves optimization. These trends support the deployable design.

4.3 Real-Robot Deployment Across Behaviors and Terrains

Figure 1 summarizes a *single* raw-reference policy across eight maneuver–terrain pairs: a one-leg backflip and a step-to-stair backflip on raised blocks, a stair dance, an arm-waving run, a free-arm walk over uneven terrain, a sideways stair walk, a backward step over obstacles, and a turning gait on stairs. Each panel pairs a distinct flat-ground command with a different, randomly placed terrain; the policy preserves the commanded upper-body behavior while perception resolves the footholds, clearance, posture, and timing the reference never specifies. Acrobatics, expressive motion, and omnidirectional locomotion are all realized without per-command tuning or a per-terrain controller. Figure 5 isolates the operator–environment mismatch case: the human performs a mocap command on flat ground while the robot executes it over randomly placed terrain.

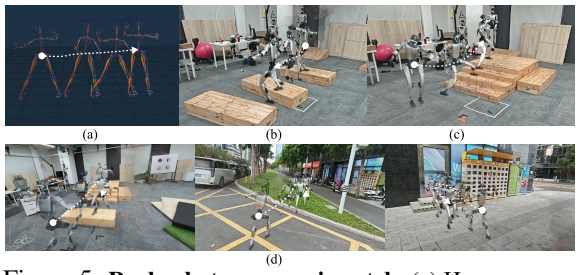

Figure 5: **Real-robot mocap mismatch.** (a) Human mocap motion captured on flat ground; (b,c) the robot tracks the (a) command over robot-side terrain; (d) a separate walk-and-dance motion deployed in the wild.

5 Conclusion

We presented Perceptive BFM, a motion-reference-conditioned humanoid control framework that keeps the raw kinematic reference as the deployment command while grounding its terrain-dependent realization in robot-centric perception. TCRS supplies terrain-conformal supervision offline, PMT transfers that behavior to a raw-reference vision student, and identity-gated residuals let perception correct footholds, clearance, posture, and timing without replacing the command. Our evidence spans quantitative TCRS quality, matched-compute ablations, and qualitative real-robot deployment.

6 Limitations

Assumptions. TCRS is a *kinematic* synthesizer: it builds contact-consistent, style-preserving references without solving contact-rich dynamics, and assumes a static, rigid, observable height field, so it does not model deformable, granular, or slippery media, and assumes the upper-body command stays feasible after lower-body correction.

Failure modes. Because adaptation is foot-centric while the upper-body command is preserved as-is, the arms and torso can strike nearby obstacles (Figure 6). **Future work**: collision-aware upper-body adaptation and quantitative rollouts.

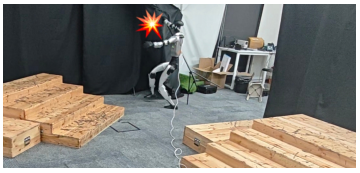
Figure 6: **Representative failure.** The upper-body command is collision-unaware, so arms or torso can strike obstacles.

References

[1] T. He, Z. Luo, X. He, W. Xiao, C. Zhang, W. Zhang, K. M. Kitani, C. Liu, and G. Shi. OmniH2O: Universal and dexterous human-to-humanoid whole-body teleoperation and learning. In *Proceedings of the 8th Conference on Robot Learning*, volume 270 of *Proceedings of Machine Learning Research*, pages 1516–1540. PMLR, 2025. URL https://proceedings.mlr.press/v270/he25b.html.

[2] X. Cheng, Y. Ji, J. Chen, R. Yang, G. Yang, and X. Wang. Expressive whole-body control for humanoid robots. In *Proceedings of Robotics: Science and Systems*, Delft, Netherlands, July 2024. doi:10.15607/RSS.2024.XX.107.

[3] T. He, W. Xiao, T. Lin, Z. Luo, Z. Xu, Z. Jiang, J. Kautz, C. Liu, G. Shi, X. Wang, L. Fan, and Y. Zhu. HOVER: Versatile neural whole-body controller for humanoid robots. *arXiv preprint arXiv:2410.21229*, 2024. doi:10.48550/arXiv.2410.21229.

[4] J. Bjorck, F. Castañeda, N. Cherniadev, X. Da, R. Ding, L. Fan, Y. Fang, D. Fox, F. Hu, S. Huang, J. Jang, Z. Jiang, J. Kautz, K. Kundalia, L. Lao, Z. Li, Z. Lin, K. Lin, G. Liu, E. Llontop, L. Magne, A. Mandlekar, A. Narayan, S. Nasiriany, S. Reed, Y. L. Tan, G. Wang, Z. Wang, J. Wang, Q. Wang, J. Xiang, Y. Xie, Y. Xu, Z. Xu, S. Ye, Z. Yu, A. Zhang, H. Zhang, Y. Zhao, R. Zheng, and Y. Zhu. GR00T N1: An open foundation model for generalist humanoid robots. *arXiv preprint arXiv:2503.14734*, 2025. doi:10.48550/arXiv.2503.14734.

[5] W. Zeng, S. Lu, K. Yin, X. Niu, M. Dai, J. Wang, and J. Pang. Behavior foundation model for humanoid robots. *arXiv preprint arXiv:2509.13780*, 2025. doi:10.48550/arXiv.2509.13780.

[6] Y. Li, Z. Luo, T. Zhang, C. Dai, A. Kanervisto, A. Tirinzoni, H. Weng, K. Kitani, M. Guzek, A. Touati, A. Lazaric, M. Pirotta, and G. Shi. BFM-Zero: A promptable behavioral foundation model for humanoid control using unsupervised reinforcement learning. *arXiv preprint arXiv:2511.04131*, 2025. doi:10.48550/arXiv.2511.04131.

[7] Z. Luo, Y. Yuan, T. Wang, C. Li, S. Chen, F. Castaneda, Z.-A. Cao, J. Li, D. Minor, Q. Ben, X. Da, L. Fan, and Y. Zhu. SONIC: Supersizing motion tracking for natural humanoid whole-body control. *arXiv preprint arXiv:2511.07820*, 2025. doi:10.48550/arXiv.2511.07820.

[8] S. Zhu, Z. Zhuang, M. Zhao, K.-Y. Lee, and H. Zhao. Hiking in the wild: A scalable perceptive parkour framework for humanoids. *arXiv preprint arXiv:2601.07718*, 2026. doi:10.48550/arXiv.2601.07718.

[9] Z. Zhuang, S. Zhu, M. Zhao, and H. Zhao. Deep whole-body parkour. *arXiv preprint arXiv:2601.07701*, 2026. doi:10.48550/arXiv.2601.07701.

[10] Z. Wu, X. Huang, L. Yang, Y. Zhang, K. Sreenath, X. Chen, P. Abbeel, R. Duan, A. Kanazawa, C. Sferrazza, G. Shi, and C. K. Liu. Perceptive humanoid parkour: Chaining dynamic human skills via motion matching. *arXiv preprint arXiv:2602.15827*, 2026. doi:10.48550/arXiv.2602.15827.

[11] X. B. Peng, P. Abbeel, S. Levine, and M. van de Panne. Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions on Graphics*, 37 (4):143, 2018. doi:10.1145/3197517.3201311.

[12] X. B. Peng, Z. Ma, P. Abbeel, S. Levine, and A. Kanazawa. AMP: Adversarial motion priors for stylized physics-based character control. *ACM Transactions on Graphics*, 40(4):1–20, 2021. doi:10.1145/3450626.3459670.

[13] T. He, Z. Luo, W. Xiao, C. Zhang, K. Kitani, C. Liu, and G. Shi. Learning human-to-humanoid real-time whole-body teleoperation. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8944–8951, 2024. doi:10.1109/IROS58592.2024.10801984.

[14] Z. Luo, J. Cao, A. Winkler, K. Kitani, and W. Xu. Perpetual humanoid control for real-time simulated avatars. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10895–10904, 2023. doi:10.1109/ICCV51070.2023.01000.

[15] Z. Luo, J. Cao, J. Merel, A. Winkler, J. Huang, K. Kitani, and W. Xu. Universal humanoid motion representations for physics-based control. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=0r0d8PxO02.

[16] Z. Fu, Q. Zhao, Q. Wu, G. Wetzstein, and C. Finn. HumanPlus: Humanoid shadowing and imitation from humans. In *Proceedings of the 8th Conference on Robot Learning*, volume 270 of *Proceedings of Machine Learning Research*, pages 2828–2844. PMLR, 2025. URL https://proceedings.mlr.press/v270/fu25a.html.

[17] Y. Ma, H. Yu, J. Xie, C. Lv, Q. Luo, C. Zhang, Y. Yin, B. Xing, X. Ren, and D. Zheng. Robust and generalized humanoid motion tracking. *arXiv preprint arXiv:2601.23080*, 2026. doi:10.48550/arXiv.2601.23080.

[18] Y. Wang, S. Zhu, P. Zhi, Y. Li, J. Li, Y.-L. Li, Y. Xiao, X. Wang, B. Jia, and S. Huang. OmniXtreme: Breaking the generality barrier in high-dynamic humanoid control. *arXiv preprint arXiv:2602.23843*, 2026. doi:10.48550/arXiv.2602.23843.

[19] A. Agarwal, A. Kumar, J. Malik, and D. Pathak. Legged locomotion in challenging terrains using egocentric vision. In *Proceedings of the 6th Conference on Robot Learning*, volume 205 of *Proceedings of Machine Learning Research*, pages 403–415. PMLR, 2023. URL https://proceedings.mlr.press/v205/agarwal23a.html.

[20] Z. Zhuang, Z. Fu, J. Wang, C. G. Atkeson, S. Schwertfeger, C. Finn, and H. Zhao. Robot parkour learning. In *Proceedings of the 7th Conference on Robot Learning*, volume 229 of *Proceedings of Machine Learning Research*, pages 73–92. PMLR, 2023. URL https://proceedings.mlr.press/v229/zhuang23a.html.

[21] X. Cheng, K. Shi, A. Agarwal, and D. Pathak. Extreme parkour with legged robots. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11443–11450, 2024. doi:10.1109/ICRA57147.2024.10610200.

[22] Z. Zhuang, S. Yao, and H. Zhao. Humanoid parkour learning. *arXiv preprint arXiv:2406.10759*, 2024. doi:10.48550/arXiv.2406.10759.

[23] H. Wang, Z. Wang, J. Ren, Q. Ben, T. Huang, W. Zhang, and J. Pang. BeamDojo: Learning agile humanoid locomotion on sparse footholds. *arXiv preprint arXiv:2502.10363*, 2025. doi:10.48550/arXiv.2502.10363.

[24] W. Sun, Y. Su, L. Huang, A. Zhang, D. Wei, M. San, D. Tian, E. Cao, F. Yan, E. Xie, and Z. Xie. Now you see that: Learning end-to-end humanoid locomotion from raw pixels. *arXiv preprint arXiv:2602.06382*, 2026. doi:10.48550/arXiv.2602.06382.

[25] Z. Wang, T. Ma, Y. Jia, X. Yang, J. Zhou, W. Ouyang, Q. Zhang, and J. Liang. Omni-perception: Omnidirectional collision avoidance for legged locomotion in dynamic environments. *arXiv preprint arXiv:2505.19214*, 2025. doi:10.48550/arXiv.2505.19214.

[26] Z. Wang, X. Yang, J. Zhao, J. Zhou, T. Ma, Z. Gao, A. Ajoudani, and J. Liang. End-to-end humanoid robot safe and comfortable locomotion policy. *arXiv preprint arXiv:2508.07611*, 2025. doi:10.48550/arXiv.2508.07611.

[27] Z. Zhang, K. Wen, M. Xu, J. He, C. Li, T. Miki, C. Schwarke, C. Zhang, X. B. Peng, and M. Hutter. Learning whole-body humanoid locomotion via motion generation and motion tracking. *arXiv preprint arXiv:2604.17335*, 2026. doi:10.48550/arXiv.2604.17335.

[28] W. D. Compton, Z. Olkin, and A. D. Ames. Terrain consistent reference-guided RL for humanoid navigation autonomy. *arXiv preprint arXiv:2605.15517*, 2026. doi:10.48550/arXiv.2605.15517.

[29] Y. Li, P. Zhi, Y. Wang, T. Liu, S. Yan, W. Liu, X. Wang, B. Jia, and S. Huang. OmniTrack: General motion tracking via physics-consistent reference. *arXiv preprint arXiv:2602.23832*, 2026. doi:10.48550/arXiv.2602.23832.

[30] S. Choi, M. K. X. J. Pan, and J. Kim. Nonparametric motion retargeting for humanoid robots on shared latent space. In *Robotics: Science and Systems*, 2020. doi:10.15607/RSS.2020.XVI.071.

[31] R. Villegas, J. Yang, D. Ceylan, and H. Lee. Neural kinematic networks for unsupervised motion retargetting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8639–8648, 2018. doi:10.1109/CVPR.2018.00901.

[32] L. Yang, X. Huang, Z. Wu, A. Kanazawa, P. Abbeel, C. Sferrazza, C. K. Liu, R. Duan, and G. Shi. OmniRetarget: Interaction-preserving data generation for humanoid whole-body loco-manipulation and scene interaction. *arXiv preprint arXiv:2509.26633*, 2025. doi:10.48550/arXiv.2509.26633.

[33] E. Dantec, M. Naveau, P. Fernbach, N. A. Villa, G. Saurel, O. Stasse, M. Taix, and N. Mansard. Whole-body model predictive control for biped locomotion on a torque-controlled humanoid robot. *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, pages 638–644, 2022. doi:10.1109/Humanoids53995.2022.10000129.

[34] A. Pajon, S. Caron, G. De Magistris, S. Miossec, and A. Kheddar. Walking on gravel with soft soles using linear inverted pendulum tracking and reaction force distribution. In *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*, pages 432–437, 2017. doi:10.1109/HUMANOIDS.2017.8246909.

[35] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter. Learning quadrupedal locomotion over challenging terrain. *Science Robotics*, 5(47):eabc5986, 2020. doi:10.1126/scirobotics.abc5986.

[36] A. Kumar, Z. Fu, D. Pathak, and J. Malik. RMA: Rapid motor adaptation for legged robots. In *Robotics: Science and Systems*, 2021. doi:10.15607/RSS.2021.XVII.011.

[37] T. Miki, J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter. Learning robust perceptive locomotion for quadrupedal robots in the wild. *Science Robotics*, 7(62):eabk2822, 2022. doi:10.1126/scirobotics.abk2822.

[38] T. He, J. Gao, W. Xiao, Y. Zhang, Z. Wang, J. Wang, Z. Luo, G. He, N. Sobanbab, C. Pan, Z. Yi, G. Qu, K. Kitani, J. Hodgins, L. J. Fan, Y. Zhu, C. Liu, and G. Shi. ASAP: Aligning simulation and real-world physics for learning agile humanoid whole-body skills. *arXiv preprint arXiv:2502.01143*, 2025. doi:10.48550/arXiv.2502.01143.

[39] T. Silver, K. Allen, J. Tenenbaum, and L. P. Kaelbling. Residual policy learning. *arXiv preprint arXiv:1812.06298*, 2018. doi:10.48550/arXiv.1812.06298.

[40] T. Johannink, S. Bahl, A. Nair, J. Luo, A. Kumar, M. Loskyll, J. A. Ojea, E. Solowjow, and S. Levine. Residual reinforcement learning for robot control. *IEEE International Conference on Robotics and Automation (ICRA)*, pages 6023–6029, 2019. doi:10.1109/ICRA.2019.8794127.

[41] S. Zhao, Y. Ze, Y. Wang, C. K. Liu, P. Abbeel, G. Shi, and R. Duan. ResMimic: From general motion tracking to humanoid whole-body loco-manipulation via residual learning. *arXiv preprint arXiv:2510.05070*, 2025. doi:10.48550/arXiv.2510.05070.

[42] Z. Wang, Y. Jia, L. Shi, H. Wang, H. Zhao, X. Li, J. Zhou, J. Ma, and G. Zhou. Arm-constrained curriculum learning for loco-manipulation of a wheel-legged robot. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 10770–10776. IEEE, 2024. doi:10.1109/IROS58592.2024.10802062.

A Additional Implementation Details

A.1 PMT Training Algorithm

Algorithm 1 Perceptive Motion Tracking (PMT) training algorithm.

Input: raw motion clips $\mathcal{M}$, terrain generator $\mathcal{G}$, robot model, height scanner
for clip $m \in \mathcal{M}$ and terrain $\tau \sim \mathcal{G}$ **do**
    synthesize $\mathbf{m}^{\mathrm{tcrs}} = \mathcal{S}_{\mathrm{TCRS}}(m, \tau)$ using Alg. 2
    export paired raw/adapted references $(\mathbf{m}^{\mathrm{raw}}, \mathbf{m}^{\mathrm{tcrs}}, \tau)$
end for
train blind teacher π_T with PPO to track $\mathbf{m}^{\mathrm{tcrs}}$
compute target-frame labels from π_T using Eq. (10)
distill identity-gated vision student π_S using raw-reference observations
fine-tune π_S with PPO under raw-reference commands and height-map observations
Output: deployable Perceptive BFM policy π_S

A.2 Training Stages

Table 2: **PMT stage map.** Each stage in the PMT training algorithm corresponds to a distinct role in the pipeline.

Stage	Role
TCRS data generation	Synthesizes paired raw/TCRS references on sampled terrain.
Blind teacher PPO	Tracks terrain-conformal references with privileged training groups.
Raw-reference distillation	Distills adapted-reference teacher targets into the raw-reference student frame.
Vision PPO fine-tuning	Fine-tunes the deployable vision student with raw commands and height scans.

A.3 Observation Contract

Table 3: **Observation groups used by PMT.** Privileged groups supervise critics or auxiliary heads but are not direct deployable actor inputs.

Group	Main contents	Shape or history	Deployment role
Proprio	projected gravity, base angular velocity, joint pos/vel, previous action	93D	actor input
Proprio history	unflattened proprioceptive history	10 steps	actor temporal input
Command window	future reference velocity, gravity, joint command tokens	21×38	actor command input
Anchor delta window	local anchor displacement window	21×3	actor command input
Vision	height scan plus validity mask	17×11 cells, mask appended	vision actor input
Critic	privileged reference/body/base information	current frame	training only
Auxiliary targets	base velocity, anchor, and foot-trajectory targets	current frame/window	auxiliary losses

A.4 Reward and Optimization Details

The PPO reward used by the adapted-reference teacher and the fine-tuning stage combines four exponential tracking terms with two penalties (Eq. (11)). The tracking residuals are anchor position $\Delta_a = \mathbf{p}_t^A - \bar{\mathbf{p}}_t^A$, body orientation $\Delta_R = d_R(\mathbf{R}_t, \bar{\mathbf{R}}_t)$, and ankle position/velocity Δ_f, Δ_v on F. E_t is action/torque energy and C_t is lateral foot/shin contact. Teacher rewards are evaluated against TCRS references, while student fine-tuning uses the raw-reference command frame with terrain-conditioned residual actions; weights and standard deviations are listed in Table 4.

$$r_t = \sum_{k \in \{a,R,f,v\}} w_k \, e^{-\|\Delta_k\|^2/\sigma_k^2} - c_E E_t - c_C C_t, \tag{11}$$

A.5 Network and Data-Flow Summary

Architecture. The actor consumes the raw reference command tokens and a 10-step proprioceptive history through the Transformer backbone, producing the motion-tracking latent. The 17×11 height

Table 4: **PPO reward terms for the teacher and fine-tuning tasks.** Exponential tracking terms use the listed standard deviation.

Term	Weight	Notes
Global anchor position tracking	1.0	std 0.2 m
Relative body orientation tracking	0.5	std 0.35
Foot position tracking	1.0	ankle bodies, std 0.1 m
Foot linear-velocity tracking	0.5	ankle bodies, std 1.0 m/s
Energy	-2×10^{-5}	action/torque energy penalty
Foot/shin lateral contact	-0.03	threshold 5 N on ankles and knees

Table 5: **Training hyperparameters.** The fine-tuning stage starts from the distilled vision checkpoint and updates the inherited tracker more conservatively than the terrain branch.

Stage	Learning rate	Entropy	Additional losses or schedule
Blind teacher PPO	5×10^{-4}	0.005	5 epochs, 4 minibatches, KL target 0.01, velocity/anchor Huber losses
Distillation	1×10^{-4}	–	MSE action loss, teacher-control mix annealed $1.0 \rightarrow 0.0$
Vision PPO fine-tuning	1×10^{-4}	0.001	backbone LR scale 0.3, foot-trajectory Huber loss with delta 0.05

map together with its validity mask is processed by the terrain encoder into a terrain latent. Two zero-initialized residual branches inject this latent into the actor: an intent gate that modulates the pooled command-history latent (Eq. 8) and an action-residual branch that adds to the action mean (Eq. 9). The actor output is a residual joint-position target around the raw reference, applied through the convention in Eq. (2). The actor also reads two learned estimator heads driven by internal latents only (no privileged observation input): a base-velocity estimator and a motion-anchor-position estimator whose (detached) 3D outputs are concatenated into the actor trunk, so the deployable actor never consumes ground-truth base velocity or anchor position. These estimators are supervised by the corresponding ground-truth targets (base velocity and robot-frame anchor position), which serve as auxiliary losses only; a foot-trajectory head is likewise supervised against the teacher target and is not an actor input. The future motion-anchor displacement window, in contrast, remains a direct command-token input (concatenated with the reference-command tokens), not an estimated quantity. Figure 2 in the main paper illustrates the corresponding overview pipeline.

Data flow. Each raw motion clip is paired with a sampled terrain to form (m^{raw}, τ), which TCRS converts into a terrain-conformal reference m^{tcrs} (Algorithm 2). The blind teacher is trained with PPO on m^{tcrs}. During distillation, the teacher's action is converted into the raw-reference frame using the target-frame relabeling in Eq. (10), and the raw-reference student is fit by MSE. The student is then fine-tuned with PPO under raw-reference commands and onboard height-scan observations. At deployment, the policy receives only the raw reference, proprioception, and the onboard terrain observation; TCRS supervision is never queried online.

A.6 Terrain-Conformal Reference Synthesis Details

Multi-point leg IK objective. Each frame, TCRS solves a damped support-aware multi-point Jacobian IK problem over the twelve leg joints; root translation is fixed to the support-aware reconstruction stage and root orientation and non-leg joints are preserved from the raw reference. For foot point set $\mathcal{P} = \{\mathrm{ankle}, \mathrm{toe}, \mathrm{heel}\}$, the local IK update solves

$$\begin{aligned}\Delta\mathbf{q}^{\mathrm{leg},\star} = \arg\min_{\Delta\mathbf{q}^{\mathrm{leg}}} & \sum_{f\in\{L,R\}}\sum_{p\in\mathcal{P}} \left\|\mathbf{W}_{f,p}\left(\mathbf{J}_{f,p}\Delta\mathbf{q}^{\mathrm{leg}} - \mathbf{e}_{f,p}\right)\right\|^2 \\ & + \lambda_{\mathrm{post}}\|\mathbf{q}^{\mathrm{leg}} - \mathbf{q}^{\mathrm{raw,leg}}\|^2 + \lambda_{\mathrm{cont}}\|\mathbf{q}^{\mathrm{leg}} - \mathbf{q}^{\mathrm{prev,leg}}\|^2 \\ & + \lambda_{\mathrm{pen}}\Psi_{\mathrm{pen}}(\mathbf{q}^{\mathrm{leg}}) + \lambda_{\mathrm{dls}}\|\Delta\mathbf{q}^{\mathrm{leg}}\|^2,\end{aligned} \tag{12}$$

where $\Delta\mathbf{q}^{\text{leg}}$ contains the twelve leg-joint increments, $\mathbf{e}_{f,p}$ are ankle/toe/heel residuals, $\mathbf{W}_{f,p}$ are support-aware point weights, and Ψ_{pen} penalizes foot or lower-leg penetration. Multiseed fallback and continuity guards reject high-error or discontinuous branches.

Algorithm 2 Terrain-Conformal Reference Synthesis (TCRS).

Input: raw clip m^{raw}, terrain height field h_τ, contact offsets, swing-optimization parameters, IK weights
detect contact masks using foot height, speed, and hysteresis thresholds
build stance references by anchoring support feet to terrain surfaces
for each foot f and swing phase $s = [t_0, t_1]$ **do**
    convert ankle targets to a mid-foot frame using Eq. (4)
    initialize trajectory knots from the raw mid-foot path
    for sampling iteration $i = 1, \ldots, N_{\text{iter}}$ **do**
        sample temporally smoothed knot perturbations
        evaluate tracking, smoothness, clearance, vertical-face, and endpoint costs using Eq. (5)
        update knots with the softmin-weighted sampling update in Eq. (6)
    end for
    convert optimized mid-foot trajectory back to ankle, toe, and heel targets
end for
reconstruct root height from support contacts using Eq. (7)
repair lower-leg and foot collisions; attenuate unsupported toe/heel point weights near edges
for frame $t = 1, \ldots, T$ **do**
    solve multi-point Jacobian IK using Eq. (12)
    retry with secondary seeds if foot error, penetration, or joint jumps exceed thresholds
end for
Output: terrain-conformal reference m^{tcrs}, contact masks, realized feet/root, and diagnostics

Reference-quality metrics. We evaluate TCRS before policy rollout, following the principle that a reference generator should be tested independently from the tracker it trains. Let

$$d_\tau(\mathbf{p}) = p_z - h_\tau(p_x, p_y) \tag{13}$$

be signed terrain clearance. Terrain penetration is

$$E_{\text{pen}} = \frac{1}{T|\mathcal{B}|} \sum_{t=1}^{T} \sum_{b \in \mathcal{B}} [-d_\tau(\mathbf{p}_{b,t})]_+, \tag{14}$$

where $\mathcal{B}$ contains foot and lower-leg points. Stance contact error and floating rate are

$$E_{\text{contact}} = \frac{1}{|\mathcal{C}|} \sum_{(f,t) \in \mathcal{C}} |d_\tau(\mathbf{p}_{f,t})|, \qquad R_{\text{float}} = \frac{1}{|\mathcal{C}|} \sum_{(f,t) \in \mathcal{C}} \mathbb{I}[d_\tau(\mathbf{p}_{f,t}) > \epsilon_{\text{float}}]. \tag{15}$$

Swing-clearance violation is

$$R_{\text{clear}} = \frac{1}{|\mathcal{S}|} \sum_{(f,t) \in \mathcal{S}} \mathbb{I}[d_\tau(\mathbf{p}_{f,t}) < c_{\min}], \tag{16}$$

and style deviation is measured by upper-body deviation from the raw clip,

$$E_{\text{upper}} = \frac{1}{T|\mathcal{J}_{ub}|} \sum_{t, j \in \mathcal{J}_{ub}} \|\mathbf{x}_{j,t}^{\text{tcrs}} - \mathbf{x}_{j,t}^{\text{raw}}\|_2. \tag{17}$$

Foot smoothness is measured as mean finite-difference foot acceleration,

$$E_{\text{smooth}} = \frac{1}{(T-2)|\mathcal{F}|} \sum_{t=2}^{T-1} \sum_{f \in \mathcal{F}} \left\| \frac{\mathbf{p}_{f,t+1} - 2\mathbf{p}_{f,t} + \mathbf{p}_{f,t-1}}{\Delta t^2} \right\|_2. \tag{18}$$

A good synthesizer should reduce penetration and clearance violations, avoid excessive nominal-contact floating, keep E_{upper} small, and maintain smooth foot trajectories. The corresponding values are reported in the main paper as Table 1.

Implementation specifics. In our implementation, TCRS preserves the root orientation and upper-body joint references from the raw clip. Root translation is shaped by the support-aware reconstruction stage (Eq. (7)) and is then held fixed during IK, so the IK solver only updates the twelve leg joints through a multi-point Jacobian system over ankle, toe, and heel targets for both feet. The multi-point structure is what enables foot-geometry-aware contact near step edges. This decomposition makes terrain adaptation primarily a lower-body contact adjustment while limiting distortion of the commanded whole-body style; the upper-body trajectory is therefore preserved across both synthesis and deployment.

A.7 Network Architecture

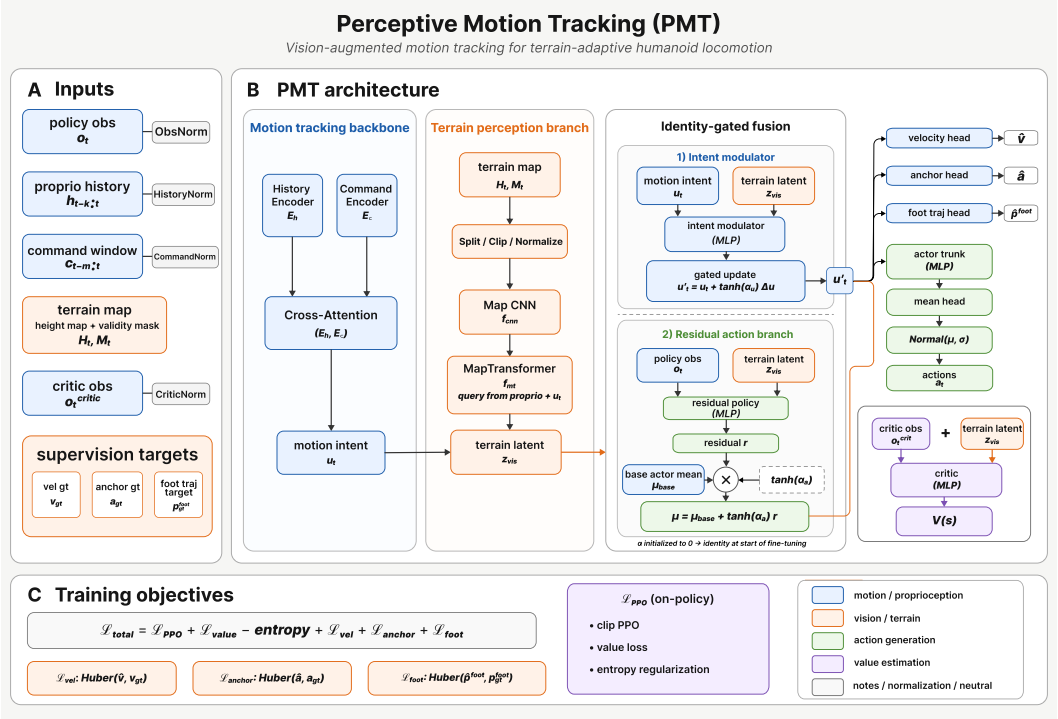

Figure 7: **Detailed PMT network architecture.** (A) Inputs: policy observation $\mathbf{o}_t$, 10-step proprioceptive history $\mathbf{h}_{t-k:t}$, command window $\mathbf{c}_{t-m:t}$, terrain map $(\mathbf{H}_t, \mathbf{M}_t)$, critic observation, and supervision targets. (B) PMT actor: a Transformer motion-tracking backbone with cross-attention encoders E_h, E_c produces a motion intent $\mathbf{u}_t$; the terrain perception branch (Map CNN f_{cnn} followed by a query-conditioned MapTransformer f_{mt}) produces a terrain latent $\mathbf{z}^{\text{vis}}$; the identity-gated fusion module updates the intent through $\mathbf{u}'_t = \mathbf{u}_t + \tanh(\alpha_u)\Delta\mathbf{u}$ and adds a residual to the action mean $\boldsymbol{\mu} = \boldsymbol{\mu}^{\text{base}} + \tanh(\alpha_a)\, r$, with α initialized at zero so the policy starts as a pure raw-reference tracker. Auxiliary heads predict base velocity, motion anchor, and foot trajectory; the critic operates on privileged inputs. (C) Training objective: PPO + value + entropy losses combined with Huber auxiliary losses on velocity, anchor, and foot trajectory. The corresponding text definitions are in Section 3.

A.8 Real-Robot Deployment Protocol

Platform and perception. The deployable policy runs on a 29-DoF Unitree G1 humanoid. Robot-centric terrain perception is implemented as a torso-mounted depth-to-height-map pipeline that produces a 17×11 height grid over a 1.6 m $\times$ 1.0 m footprint at 0.1 m resolution. The same height-map representation is used in simulation training and during real-robot deployment so that the observation contract is preserved across the sim-to-real transfer.

Indoor protocol. Indoor trials place rectangular obstacles, raised steps, virtual-lawn surfaces, and mixed obstacle layouts in a motion-capture–equipped lab. Each terrain configuration is exercised with multiple commanded behaviors drawn from the broad behavior set, including walking, running, side-walking, and gesture-rich motions. These rollouts are used as qualitative deployment coverage in the main paper.

Outdoor protocol. Outdoor trials cover stairs, grass, isolated steps, recessed flower beds, and sidewalk transitions. Hardware safety is supervised by a runtime watchdog that triggers a soft fall-over recovery if the estimator detects a torque saturation or a base-orientation excursion outside the training-time envelope.

Mocap mismatch protocol. The mocap mismatch setup deliberately separates the human and robot environments: the human performs the commanded motion on flat ground while the robot executes the corresponding kinematic command over randomly placed steps and cube obstacles. This protocol directly tests whether robot-centric terrain perception can ground a human command that contains no matching terrain information.

A.9 Training and Compute Details

All policy variants in Figure 4 and the Section 4.2 ablation share the same task definition, reward terms (Eq. (11)), observation contract (Table 3), action space, episode horizon, environment count, and PPO hyperparameters; only the actor network or the listed component differs. All variants are trained on 48 NVIDIA A800 GPUs for 10k iterations. Our deployable PMT model is resumed from an intermediate checkpoint at iteration $\approx$ 5k under identical optimizer settings, producing the two-segment continuation in the plotted curve; baselines run as a single 10k-iteration segment from scratch. Reported *mean reward* values average the per-iteration mean reward over the last 1k iterations, after a 21-step rolling-mean smoothing identical to the curves in Figure 4. *PMT w/o vision* removes the height-map encoder and the two identity-gated residual pathways while preserving every other component; *PMT w/o distillation* removes the target-frame distillation stage so the Transformer student is trained directly with PPO from initialization; the four MLP/CNN baselines replace only the actor backbone.

Scope of deployed evaluation. The current main paper uses quantitative measurements for TCRS reference quality (Table 1) and training-time ablations (Figure 4, Section 4.2), and qualitative real-robot rollouts for deployment evidence (Figures 1, 5). Repeated deployed-rollout statistics, including completion rate, foot penetration, leg collision, and target-frame imitation error, require a separate standardized rollout campaign and are not used as main-paper evidence in this submission.