Proceedings of CONF-FMCE 2025 Symposium: Semantic Communication for Media Compression and Transmission
DOI: 10.54254/2755-2721/2025.GL25269

Control Strategies for Embodied Humanoid Robots: Reinforcement Learning vs. Multi-Contact Planning

Haobo Wang

School of Mechanical Science & Engineering of HUST, Huazhong University of Science and Technology, Wuhan, China
u202410600@hust.edu.cn

Abstract. As the application demand for humanoid robots in complex and unstructured environments increases, how to balance adaptability and stability in control strategies has become increasingly critical. This paper compares and analyzes two typical humanoid robot control methods: a reinforcement learning-based controller on the Digit V3 platform and a multi-contact planning and control (MCPC) framework on the COMAN+ platform. This work addresses a gap in the literature that lacks a comparative perspective on cross-method and cross-platform practical verification. It first introduces the design concept and training mechanism of the reinforcement learning controller, which removes reliance on gait clocks and achieves the natural switching between standing and walking under external disturbances. Then it analyzes the MCPC framework, which combines posture sampling, nonlinear programming (NLP) trajectory optimization, and torque-level balance control to support the robot to stably perform complex multi-contact tasks. Experimental results show that the reinforcement learning controller exhibits excellent robustness in disturbance response and command switching, while the MCPC method shows higher accuracy and repeatability in structured tasks. The results of this study show that reinforcement learning is suitable for dealing with scenarios with strong dynamic adaptability, while planning control emphasizes interpretability and physical feasibility. The comparative analysis presented in this article provides a reference for understanding the trade-offs in humanoid robot control strategies and also offers guidance for truly realizing embodied intelligence in humanoid robots in the future.

Keywords: Humanoid robots, Embodied intelligence, Reinforcement learning, Multi-contact planning and control

1. Introduction

Compared to wheeled or crawling robots, humanoid robots are more suitable for working in human living environments due to their anthropomorphic design [1]. However, in daily life, tasks such as walking, maintaining balance, and avoiding obstacles place extremely high demands on humanoid robots' precise motion control, real-time environmental adaptation, and balance stability [1]. Therefore, in recent years, the concept of embodied intelligence has received increasing attention. This concept emphasizes that the realization of intelligence depends not only on algorithms, but also

DOI: 10.54254/2755-2721/2025.GL25269

on the interaction between the robot's physical structure and the environment in which it is located [2].

Many studies have been devoted to improving the control capabilities of humanoid robots. For example, the zero-moment point (ZMP) control method proposed by Kajita et al. can generate smooth center-of-mass trajectories, support arbitrary foot placement, and effectively compensate for the ZMP error of the multi-body model through preview control. It was successfully applied to achieve the dynamic walking of the HRP-2P bipedal robot on a spiral staircase in a simulated environment, providing a theoretical basis for stable bipedal gait, and is still widely used today [3]. Later, some scholars introduced reinforcement learning to enable robots to automatically learn actions in changing environments. For example, Xie et al. demonstrated the great potential of reinforcement learning in controlling the Cassie bipedal robot in a simulated environment. The learning controller was able to learn a robust policy in 2.5 hours, which enabled it to handle tasks such as 0.22-meter sinusoidal terrain and 140-N disturbances, while supporting dynamic speed adjustment, demonstrating significant superiority over a manually tuned reference controller in 3D walking, terrain adaptation, disturbance recovery, and speed regulation. However, there is still a gap in transferring these skills to real robots [4]. In addition, Sentis and Park proposed a multi-contact flexible control strategy that demonstrated high-precision center of mass tracking, compliant contact behavior, and internal force control, as well as robustness to unmodeled disturbances and model uncertainties [5]. However, there is still a lack of research that integrates these methods into a complete system and systematically verifies them on a real platform. This shows that the understanding of how the various modules in the embodied intelligence system work together is not deep enough.

This paper examines two representative humanoid robot studies: a reinforcement learning-based control method on the Digit platform and a multi-contact control framework verified on the COMAN+ platform. By comparing their different designs and experimental performances in task switching, body support control, and environmental adaptation, this paper aims to analyze the key issues and technical challenges currently faced in realizing embodied intelligence in real humanoid robots.

2. Theoretical basis and principles

Humanoid robots are essentially dynamically unstable systems. Robots often rely on rapid adjustment of body posture during short, intermittent contact with the ground to maintain balance, so they face great challenges in control. In addition, robots face multiple challenges in real environments from terrain, load changes, external interference, and human-machine interaction [6]. Therefore, developing a strategy that can achieve whole-body coordinated control in a changing environment is the key to achieving stable operation of humanoid robots.

2.1. Reinforcement learning control method

Reinforcement learning is a self-learning method based on trial and error. The robot gradually explores which actions can obtain higher rewards through interaction with the environment, thereby forming a stable and effective strategy. Unlike traditional control methods that rely on detailed modeling and clear rules, reinforcement learning guides the robot to autonomously learn to complete tasks such as maintaining balance, walking forward, or avoiding obstacles by setting a reward function [7].

DOI: 10.54254/2755-2721/2025.GL25269

In humanoid robot control, commonly used reinforcement learning algorithms include Proximal Policy Optimization (PPO), Deep Deterministic Policy Gradient (DDPG), and Soft Actor-Critic (SAC). Among them, PPO is more stable and has high sample efficiency, which is suitable for use on real robot platforms, while DDPG and SAC are more suitable for controlling robots with many joints and continuous actions. They are more reliable in complex environments and in the presence of noisy or uncertain data.

2.2. Planning-based multi-contact control method

Different from the strategy generation method of reinforcement learning, the planning-based control method emphasizes the interpretability of the control process and the determinism of task execution. In typical multi-contact control tasks, this method generally consists of three main modules: (1) a posture sampling planner is used to generate multi-contact sequences; (2) a trajectory optimizer ensures the dynamic feasibility and smoothness of the path; (3) a controller uses a torque feedback strategy to achieve balance control and execution accuracy [5]. Due to its strong predictability and clear action structure, this method shows strong stability and execution efficiency in tasks that require the robot to adjust the contact force and support surface by coordinating body movements. It is particularly suitable for actions such as climbing, wall support, or four-point support.

3. Case analysis

3.1. Learning and evaluation of Standing and Walking (SaW) control strategy on digit V3 platform

Digit V3 is a humanoid robot with high dynamic motion capability launched by Agility Robotics. Researchers developed a new controller structure on this platform to further improve the response flexibility of the reinforcement learning control strategy. Although reinforcement learning has freed robots from the dependence on traditional rules or trajectories and achieved "state-driven" strategy generation, many existing methods still rely on gait clocks or phase-based state representations, which limit the controller's ability to adapt to disturbances and instructions in real time. In contrast, the single-contact reinforcement learning controller (based on a single foot contact assumption) proposed in this study makes autonomous decisions based entirely on the current state of the robot. It can achieve natural switching between standing and walking without relying on external clock signals or behavioral cycles, and can still maintain stable operation when subjected to external interference, showing stronger versatility and robustness [8].

To this end, the researchers have done three main tasks: (1) Designed a low-cost and reusable SaW performance evaluation system to compare the performance of controllers in three aspects: disturbance rejection ability (fall rate), command tracking accuracy (rotation accuracy in place, speed accuracy) and energy efficiency; (2) Proposed a reward function based on the single contact assumption to avoid dependence on reference trajectory or clock signal. By adjusting the disturbance distribution in training (force: 20-200 N, duration: 200-500 ms) and increasing the command duration (increased from the original 40-100 timesteps to 100-300 timesteps), an improved version (Single Contact++) was designed. (3) The PPO reinforcement learning algorithm combined with the LSTM network structure was used to train the policy model [8].

As shown in Table 1, the researchers compared the standing and walking (SaW) performance of four controllers on the Digit platform, including the original controller (Agility controller), the reinforcement learning controller based on the gait clock, the single-contact RL controller, and its

DOI: 10.54254/2755-2721/2025.GL25269

improved version (Single-Contact++). The evaluation dimensions include: fall rate in different directions, rotation accuracy in place (displacement and angle drift), target speed tracking accuracy, and energy consumption per unit distance.

Table 1. Standing and Walking (SaW) performance evaluation table of four controllers [8]

Controller Metric	Manufacturer Controller (Agility controller)	Clock-Based RL	Single-Contact RL	Single-Contact++ RL
Fall rate in x-direction [%]	22	60	6	0
Fall rate in y-direction [%]	56	56	47	0
XY drift at 1s [m]	0.0	0.1	0.0	0.0
XY drift at 5s [m]	0.8	0.3	0.2	0.0
XY drift at 30s [m]	1.7	0.7	0.5	0.2
Angular deviation at 1s [°]	-11	0	0	0
Angular deviation at 5s [°]	-9	0	0	0
Angular deviation at 30s [°]	-17	0	0	0
Velocity tracking accuracy [m/s] (Target Command of 1.00 m/s)	0.74	1.04	1.13	Not recorded
Energy efficiency [J/m]	147	182	180	Not recorded

Judging from the results, the single-contact controller and its improved version outperform Digit's original default controller and gait clock reinforcement learning controller in core indicators such as disturbance rejection ability and rotation command tracking accuracy, reflecting that the controller designed by the researchers is more robust under complex disturbances than the previous controllers.

Although this method has obvious advantages in anti-interference and command response, it still has certain limitations. For example, the energy consumption of the single-contact reinforcement learning control strategy is significantly higher than that of the manufacturer's controller; the speed control that performs well in simulation still has Sim-to-Real deviation in reality. The researchers propose that these problems can be solved by introducing energy penalty terms and more realistic physical modeling, providing direction for the design of controllers that are more energy-efficient and more adaptable to changing environments in the future.

3.2. Design and verification of a Multi-Contact Planning and Control (MCPC) framework on the COMAN+ platform

COMAN+ is a humanoid robot platform with torque control capabilities that can perform complex whole-body movements. To meet the requirements of tasks such as ladder climbing, wall support, and narrow space movement, researchers designed a multi-contact control framework that integrates posture planning, trajectory optimization, and execution control [9].

The researchers adopted a two-stage strategy: at the planning level, a posture sampling method based on RRT-like (RRT: Rapidly-exploring Random Tree) was used to generate a preliminary action sequence, and trajectories were optimized with the help of a nonlinear programming (NLP) algorithm to ensure that the path is feasible and the action is coherent; at the control level, a posture switching manager and a reactive balance controller were used to generate torque commands in real time, so that the robot can maintain closed-loop balance during actual execution [9].

DOI: 10.54254/2755-2721/2025.GL25269

As shown in Table 2, COMAN+ successfully completed the complete action sequence planning and control execution in four types of multi-contact tasks: ladder climbing, parallel walls climbing, quadrupedal walking, and standing up. The performance evaluation covers five key indicators: total path planning time, posture segment trajectory generation time, number of trajectory optimization iterations, number of search graph nodes, and number of support transition steps.

Table 2. Average performance data of posture planner [9]

Task	Planning time (s)	Transition Generation time (s)	Number of iterations	Number of search graph nodes	number of support transition steps
Ladder climbing	43.60	37.64	1926.14	205.34	39.04
Parallel walls climbing	175.35	171.01	1557.37	151.28	44.12
Quadrupedal walking	62.69	55.99	2540.10	228.32	53.19
Standing up	7.56	6.02	459.24	62.19	17.00

From the data, the higher the complexity of the task, the greater the planning time and computational cost. For example, the total planning time for the "parallel walls climbing" task is 175.35 seconds, which is much higher than the 7.56 seconds of the "standing recovery" task, and corresponds to more optimization iterations (1557.37 times) and search nodes (151.28). In the "ladder climbing" and "quadrupedal walking" tasks, the number of supporting switching steps reached 39 and 53, respectively, indicating that the system has a strong continuous support transformation planning capability. Overall, the system has demonstrated strong multi-contact action organization capabilities in different task scenarios, providing a verification basis for complex body coordination and action generation in the embodied intelligence system.

A major advantage of the MCPC framework is that it does not rely on predefined contact points or manually configured motion templates and is applicable to a variety of task environments. However, there are still certain limitations. The system only supports static balance and does not have the ability to quickly adjust dynamically, and when trajectory optimization fails, there is a lack of an effective fallback mechanism or route reconstruction function. Future improvements can introduce dynamic balance control, strengthen trajectory quality assessment mechanisms, and expand robustness verification under a wider range of task types, providing ideas for realizing "multimodal body coordination" in embodied intelligence.

4. General discussion

From the above two cases, it can see that these two studies represent two typical paths of current embodied intelligence humanoid robot control methods: data-driven reinforcement learning control and structured planning-based control. Both reflect the key elements of embodied intelligence control systems from different perspectives: the former emphasizes adaptability and learning ability, while the latter emphasizes coordination and stability under structural constraints. See Table 3 for details:

DOI: 10.54254/2755-2721/2025.GL25269

Table 3. Comparative analysis of reinforcement learning and planning-based control strategies

Comparison Dimensions	Reinforcement Learning Control	Planning-Based Control
Control Core	Reinforcement learning strategy, autonomously learn actions from experience.	Planner + Optimizer, design the trajectory, and then execute it.
Data Source	Interact with the environment and train through trial and error	Has clear models, mission objectives, and contact sequences
flexibility	High, can adapt to unknown disturbances	Low, need to re-plan or fail when encountering changes
interpretability	Poor, unclear about the principles of RL strategy	Strong, visible path, clear control objectives
Simulation and reality consistency	Poor, serious Sim-to-Real problem	Good, depends on model accuracy, but low transfer cost
Learning and generalization	Generalizable, one set of controls can handle multiple action requirements	Limited by task settings, lack of versatility
Applicable scenarios	A dynamically changing and constantly disruptive environment	Complex structure, static multi-contact tasks

5. Conclusion

This paper discusses two specific implementation paths of embodied intelligence in humanoid robots through the analysis and comparison of the reinforcement learning control method on the Digit platform and the multi-contact planning control framework on the COMAN+ platform. The two cases represent the two ideas of "learning" and "planning" in the current embodied control strategy, each of which is suitable for different task requirements, which also reflects the trade-off relationship between adaptability, versatility, and interpretability of the control system.

Through comparison, it can be seen that if it wants to realize a humanoid robot with real embodied intelligence, it needs to bridge the gap between perception, learning, and control and achieve overall optimization driven by tasks. Future research can consider combining the flexibility of reinforcement learning with the stability of planning methods to build a hybrid control architecture that is both adaptive and predictable, providing new solutions for autonomous robots in complex environments.

References

[1] Tong, Y., Liu, H., & Zhang, Z. (2024). Advancements in humanoid robots: A comprehensive review and future prospects. IEEE/CAA Journal of Automatica Sinica, 11(2), 301–328.
[2] Pfeifer, R., & Bongard, J. (2006). How the body shapes the way we think: A new view of intelligence. MIT Press.
[3] Kajita, S., Kanehiro, F., Kaneko, K., et al. (2003). Biped walking pattern generation by using preview control of zero-moment point. In Proceedings of the 2003 IEEE International Conference on Robotics and Automation (Vol. 2, pp. 1620–1626). IEEE.
[4] Xie, Z., Berseth, G., Clary, P., Hurst, J., & van de Panne, M. (2018). Feedback control for Cassie with deep reinforcement learning. In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 1241–1246). IEEE.
[5] Sentis, L., Park, J., & Khatib, O. (2010). Compliant control of multicontact and center-of-mass behaviors in humanoid robots. IEEE Transactions on Robotics, 26(3), 483–501.
[6] Sheng, Q., Zhou, Z., Li, J., et al. (n.d.). A comprehensive review of humanoid robots. SmartBot, e12008.

DOI: 10.54254/2755-2721/2025.GL25269

[7] Kober, J., Bagnell, J. A., & Peters, J. (2013). Reinforcement learning in robotics: A survey. The International Journal of Robotics Research, 32(11), 1238–1274.
[8] van Marum, B., Shrestha, A., Duan, H., et al. (2024). Revisiting reward design and evaluation for robust humanoid standing and walking. In 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 11256–11263). IEEE.
[9] Ferrari, P., Rossini, L., Ruscelli, F., et al. (2023). Multi-contact planning and control for humanoid robots: Design and validation of a complete framework. Robotics and Autonomous Systems, 166, 104448.