

INTISARI

OPTIMALISASI KENDALI CERDAS ROBOT NON-HOLONOMIC MENGUNAKAN DEEP REINFORCEMENT LEARNING DENGAN PENDEKATAN WAYPOINT-A* DI LINGKUNGAN LABIRIN

Oleh

R Cahya Hidayat

24/547510/PPA/06893

Penelitian ini mengusulkan pendekatan navigasi robot *non-holonomic* pada lingkungan labirin dengan mengombinasikan *Deep Reinforcement Learning (DRL)* berbasis *Twin Delayed Deep Deterministic Policy Gradient (TD3)* dan panduan *waypoint* berbasis perencanaan jalur *A**. Permasalahan utama pada navigasi labirin adalah eksplorasi yang tidak terarah akibat aksi yang bersifat kontinu serta sinyal pembelajaran yang kurang informatif ketika target hanya berupa *goal* akhir, sehingga episode cenderung berakhir *timeout* meskipun tanpa tabrakan. Tujuan penelitian ini adalah meningkatkan efektivitas dan stabilitas pembelajaran agen agar mampu mencapai tujuan secara lebih konsisten pada labirin dengan tingkat kompleksitas yang meningkat, yang ditunjukkan melalui peningkatan *success rate*, penurunan *collision rate* serta penurunan kegagalan akibat kehilangan progres navigasi.

Metode yang diusulkan menggunakan *state* berdimensi 40 yang terdiri atas 36 sektor hasil *downsample* pemindaian *LiDAR* dan empat fitur internal robot, yaitu jarak *Euclidean* ke tujuan (d), perbedaan arah/*heading angle* terhadap tujuan (θ), kecepatan linear saat ini (v), dan kecepatan sudut saat ini (ω). Aksi direpresentasikan sebagai dua nilai kontinu berupa kecepatan linear dan kecepatan sudut. Data *LiDAR* dan odometri digunakan untuk membangun *Occupancy Grid Map (OGM)*, kemudian peta dikonversi menjadi *grid* biner dan dilakukan dilatasi untuk memberikan margin aman. *Grid* tersebut menjadi masukan algoritma *A** untuk menghitung rute menuju *goal*, lalu rute digunakan untuk membentuk

rangkaian *waypoint* sebagai target lokal bertahap yang *feasible* di koridor labirin. Evaluasi dilakukan pada tiga variasi lingkungan (sederhana, dua cabang, dan tiga cabang) dengan pembandingan TD3 standar (tanpa *waypoint*) dan TD3 dengan *waypoint* berbasis POI.

Hasil pengujian menunjukkan bahwa konfigurasi TD3-A* memberikan performa terbaik dan konsisten pada seluruh skenario uji, dengan *success rate* berturut-turut sebesar 0.96 (lingkungan sederhana), 0.98 (dua cabang), dan 0.78 (tiga cabang), serta *collision rate* 0 pada seluruh lingkungan. Sebaliknya, TD3 standar menghasilkan *success rate* 0 pada semua lingkungan (kegagalan didominasi oleh *timeout*), sementara konfigurasi TD3-POI menunjukkan penurunan kinerja pada lingkungan yang lebih kompleks hingga tidak menghasilkan episode sukses pada tiga cabang dan mengalami *collision rate* yang meningkat. Temuan ini menunjukkan bahwa *waypoint* berbasis A* efektif mengarahkan eksplorasi melalui target lokal yang *feasible*, sehingga mengurangi eksplorasi acak dan menekan kegagalan akibat kehilangan progres saat menghadapi koridor sempit dan percabangan.

Kata Kunci: navigasi robot, *non-holonomic*, *deep reinforcement learning*, TD3, A*, *occupancy grid map*, *waypoint*, labirin.

ABSTRACT

OPTIMIZATION OF INTELLIGENT CONTROL FOR A NON-HOLONOMIC ROBOT USING DEEP REINFORCEMENT LEARNING WITH A WAYPOINT-A* APPROACH IN A MAZE ENVIRONMENT

by

R Cahya Hidayat

24/547510/PPA/06893

This study proposes a navigation approach for non-holonomic robots in maze environments by combining Deep Reinforcement Learning (DRL) based on Twin Delayed Deep Deterministic Policy Gradient (TD3) with waypoint guidance derived from A path planning. The main challenge in maze navigation is unguided exploration due to continuous action spaces and insufficiently informative learning signals when the target is only the final goal, causing episodes to end in timeouts even without collisions. The objective of this study is to improve the effectiveness and stability of the agent's learning so that it can reach the goal more consistently in mazes of increasing complexity, as reflected by higher success rates, lower collision rates, and fewer failures caused by loss of navigational progress.*

The proposed method uses a 40-dimensional state consisting of 36 sectors obtained from downsampled LiDAR scans and four internal robot features: the Euclidean distance to the goal (d), the relative heading angle to the goal (θ), the current linear velocity (v), and the current angular velocity (ω). The action is represented as two continuous values: linear velocity and angular velocity. LiDAR and odometry data are used to build an Occupancy Grid Map (OGM); the map is then converted into a binary grid and dilated to provide a safety margin. This grid serves as the input to the A algorithm to compute a route to the goal, and the resulting route is used to form a sequence of waypoints as feasible stepwise local targets within the maze corridors. The evaluation is conducted on three*

environment variants (simple, two-branch, and three-branch) with TD3 standard (without waypoints) and TD3 with POI-based waypoints as baselines.

The results show that the TD3-A configuration achieves the best and most consistent performance across all test scenarios, with success rates of 0.96 (simple environment), 0.98 (two-branch), and 0.78 (three-branch), and a collision rate of 0 in all environments. In contrast, standard TD3 yields a success rate of 0 in all environments (failures are dominated by timeouts), while the TD3-POI configuration exhibits degraded performance as maze complexity increases, resulting in no successful episodes in the three-branch environment and an increased collision rate. These findings indicate that A*-based waypoints effectively guide exploration through feasible local targets, thereby reducing random exploration and mitigating failures due to loss of progress when encountering narrow corridors and branching structures.*

Keywords : robot navigation, non-holonomic, deep reinforcement learning, TD3, A*, occupancy grid map, waypoint, maze.