Journal of the KNST

2025; Vol.8, No.3; pp.354-360

ISSN: 2635-4926

https://doi.org/10.31818/JKNST.2025.9.8.3.354

Check for updates

Received: 2025/07/14
Revised: 2025/07/26
Accepted: 2025/08/31
Published: 2025/09/30

*Corresponding Author:

Hyunseung Kim

Marine R&D Center, LIG Nex1
333 Pangyo-ro, Bundang-gu, Seongnam-si,
Gyeonggi-do, 13488, Republic of Korea

Tel: +82-31-5179-7272

Fax: +82-31-5179-7086

E-mail: hyunseung.kim2@lignex1.com

경로계획과 PPO 기반 심도제어를 통합한 수중자율운동체의 해저지형 회피 알고리즘 개발

Development of Integrated Path Planning and PPO-based Depth Control for Seafloor Avoidance of Underwater Autonomous Vehicle

김현승[1*], 현철[2], 이성균[2], 고진용[2], 김창환[2]

[1]LIG넥스원 해양연구소 선임연구원
[2]LIG넥스원 해양연구소 수석연구원

Hyunseung Kim[1*], Chul Hyun[2], Sungkyun Lee[2], Jinyong Go[2], Changhwan Kim[2]

[1]Research engineer, Marine R&D Center, LIG Nex1
[2]Chief research engineer, Marine R&D Center, LIG Nex1

Abstract

본 논문에서는 수중 자율 운동체의 안전한 주행을 위해 경로계획과 심도제어를 통합한 해저지형 회피 알고리즘을 개발하였다. 성능 검증을 위해 1차원의 수심 지형 환경에서 기존 A* 알고리즘 또는 PPO 알고리즘 단독 방식과의 비교를 수행하였고, 충돌률과 추종 오차 측면에서의 우수함을 확인하였다. 본 논문은 복잡한 해저 환경에서도 심도 안정성과 에너지 효율을 극대화 하기 위한 경로 수립에 활용할 수 있을 것으로 기대된다.

This paper proposed a hybrid algorithm that integrates A* path planning and PPO(proximal Policy optimization)-based depth control for seafloor avoidance of underwater autonomous vehicles. To evaluate a performance in 1D depth scenario, the proposed hybrid algorithm outperformed standalone A* algorithm and PPO approaches in terms of collision rate and tracking accuracy. This study is expected to be used to establish a route for stable and energy-efficient navigation in complex underwater environments.

Keywords

수중자율운동체(Underwater Autonomous Vehicle),
경로계획(Path Planning),
심도제어(Depth Control),
PPO 알고리즘(Proximal Policy Optimization Algorithm)

1. 서론

수중운동체는 다양한 해양 임무에서 중요한 역할을 수행한다. 이때, 심도제어는 장애물 회피, 센서 안정성 확보 등 안전하고 효율적인 주행을 통한 임무 성공에 직접적인 영향을 미치는 핵심 요소로 꼽힌다. 기존에는 PID 제어기법이나 단일 목적 기반의 강화학습인 PPO(proximal policy optimization) 알고리즘을 이용해 일정 심도를 유지하거나 특정 수심대를 추종하는 방식이 주를 이루었다. 하지만 실제 해저지형은 복잡한 지형과 다양한 장애물을 포함한 비선형적인 특징이 있고, 조류나 잡음 등의 외란이 존재하는 실제 환경에서는 한계가 있기 때문에, 단순한 심도제어만으로는 안정적인 경로 확보가 어려운 상황이다.

따라서 본 논문에서는 경로계획 알고리즘과 PPO 기반 심도제어 기법을 통합하여 수중운동체가 해저지형을 회피하면서 목표점까지 빠르고 안정적으로 도달할 수 있는 새로운 제어 프레임워크를 제안한다. 먼저 경로계획은 전역적으로 안전한 경유점을 제공하여 수중운동체의 주행 시나리오 수립에 활용되고, 각 경유점까

지 PPO 기반 심도제어를 수행함으로써 기존 방식의 한계를 극복하고 복잡한 지형에서도 높은 제어 성공률과 낮은 충돌률을 달성할 수 있도록 하였다. 이에 따라 경로계획과 심도제어를 결합한 하이브리드 제어 시스템 설계를 제안하고, 시뮬레이션 기반 성능 평가를 통해 기존 방식과 비교 분석하였다.

이를 통해 본 연구는 다목적 보상함수를 적용한 PPO 알고리즘을 활용하여 수중운동체의 심도제어 성능을 향상시키고, A* 기반의 경로계획 알고리즘과의 하이브리드 구조 설계를 통해 다양한 지형 복잡도 조건에 대한 시뮬레이션을 실시하여 성능을 검증하였다. 이로부터 강화학습과 전통적 경로계획 기법의 융합을 통해 수중자율운동체가 명령 심도를 추종하며 주행하는지 사전에 분석하고, 주행시간과 배터리 효율을 최대화하는 경로 수립에 활용할 수 있을 것으로 기대된다.

2. 관련 연구

기존의 수중운동체 제어를 위한 연구는 주로 PID 기반의 전통적 제어기 또는 심층 강화학습 기반의 단일 목적 보상함수를 가진 심도제어에 집중되었다[1]. 그 중에서 PPO 기반 심도제어 기법은 단순한 구현과 안정적인 학습 성능 측면에서 널리 활용되고 있으며, 최근에는 다목적 보상함수를 적용한 응용 연구가 수행 중이다. 특히 수중자율운동체의 심도 추종 제어 문제에 PPO를 적용하여 기존의 PID 제어 대비 우수한 추종 성능을 확보하고, PPO 기반 정책이 고차원의 연속 제어 문제에서 기존 알고리즘보다 우수한 수렴 특성을 보이는 연구가 이루어졌다[2-4]. 하지만 단일 목적 보상함수 기반의 PPO 알고리즘을 이용한 심도제어는 대부분의 경우 복합적인 해저지형에서 한계를 보일 수 있다. 이에 따라 다목적 보상함수를 통합한 강화학습 제어 정책이 필요하다[5].

경로계획 측면에서는 A*, RRT(rapidly-exploring random tree)와 같이 고전적인 탐색 알고리즘이 널리 사용되고 있으며, 특히 해저지형 수심도맵 기반의 경로 탐색 기법이 수중운동체에 효과적으로 적용된다[6,7]. 하지만 기존 연구는 경로 생성과 심도제어를 별도로 처리하여 동적 환경이나 연속적인 제어 성능 측면에서 불규칙한 요소가 많아질수록 주행 효율이 떨어진다는 단점이 있다. 한편으로 강화학습 기반 경로계획 연구 또한 수행되었지만, 대부분 고차원 공간에서의 복잡도 문제로 인해 경로계획 자체를 모듈화했기 때문에 실제 환경과의 유사성이 떨어진다는 단점이 있다.

이러한 단점을 보완하기 위해 본 논문은 다목적 보상함수를 기반으로 한 PPO 심도제어기와 경로계획 알고리즘을 결합하였다. 경로계획을 통해 전역적인 안전 경로를 제공하고, PPO 기반 심도제어기가 로컬에서 정밀하게 동작하도록 하이브리드 구조를 제안함으로써 차별성을 두었다.

3. 하이브리드 구조 제안

제안하는 구조는 해저지형 수심도맵 기반 경로계획 모듈, PPO 기반 심도제어 모듈, 시뮬레이션 환경 및 평가 모듈 등 총 3개의 모듈로 구성된다. 경로계획 모듈은 해저지형 수심도맵과 수중운동체의 시작·목표 지점을 입력받아서 위험 지역을 회피하는 안전한 경유점을 생성하여 심도제어 모듈에 전달한다. PPO 기반 심도제어 모듈은 현재 위치와 다음 경유점 간의 상대 위치 정보를 이용하여 적절한 심도제어 명령을 출력한다. 여기서 보상은 경유점 접근 성공과 충돌 회피, 제어 에너지 소비 등을 고려한 다목적 보상 함수로 구성하였다. 학습은 시뮬레이션 평가 모듈에서 이루어지며, 제어 성능은 도달률, 충돌 횟수, 에너지 소비량을 기준으로 정량적으로 평가된다. 이러한 하이브리드 구조에 대한 흐름도는 Fig. 1에 도시하였다.

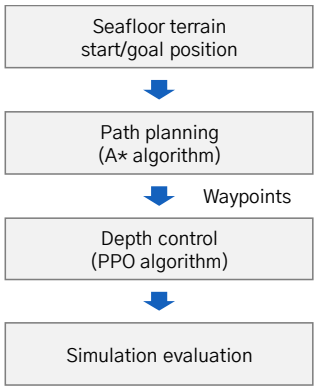

Fig. 1. Hybrid depth control framework

3.1 경로계획 모듈

본 연구에서는 해저지형 수심도맵을 기반으로 전

역 경로 생성을 위해 A* 알고리즘을 경로계획 모듈에 적용하였다. A* 알고리즘은 시작 지점에서 목표 지점까지의 최단 경로를 탐색하는 데 효과적인 휴리스틱 기반의 탐색 기법으로, 수심 조건과 제약적인 지형을 반영하여 구성한다. 수심도가 특정 임계값 이하인 지역은 장애물 또는 고위험 지역으로 간주하여 탐색 시 제외하고, 각 셀 간 이동 비용은 수심 안전성과 이동 거리를 복합적으로 고려하여 산정한다. 노드 n의 전체 비용 함수 $f(n)$을 식 (1)에 정의하였다. $h(n)$은 휴리스틱 함수로 유클리디안 거리를 사용하며 식 (2)와 같다. 또한 수중환경 특성을 반영하기 위해 실제 이동 비용인 $g(n)$은 식 (3)과 같이 계산된다. 이때, 수심 제약을 고려한 가중치 기반 모델인 식 (4)와 같이 적용하여 안전성을 확보하였다. 수심이 얕을수록 위험하므로 더 높은 가중치를 부여하였고, 비용이 급증해 경로에서 회피하는 원리이다.

$$f(n) = g(n) + h(n) \tag{1}$$

$$h(n) = \sqrt{(x_n - x_g)^2 + (y_n - y_g)^2} \tag{2}$$

$$g(n) = \sum_{i=1}^{n} d_i \cdot w(h_i) \tag{3}$$

$$d_i = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} \tag{4a}$$

$$w(h_i) = \begin{cases} \infty & h_i < h_{\min} \\ \alpha \dfrac{1}{h_i} & \text{otherwise} \end{cases} \tag{4b}$$

여기서, $g(n)$: 시작 지점에서 현재 노드 n까지 실제 이동 비용,
$h(n)$: 현재 노드 n에서 목표 지점까지 예상 휴리스틱 비용,
$f(n)$: 시작 지점에서 목표 지점까지 총 비용,
x_n, y_n: 현재 노드 n의 좌표,
x_g, y_g: 목표 지점 좌표,
x_0, y_0: 시작 지점 좌표,
α: 수심에 따른 가중치 계수.

탐색이 완료되면 시작 지점부터 목표 지점까지의 전체 경로는 일정 간격으로 분할되어 경유 지점으로 변환되며, 이는 심도제어 모듈의 입력값으로 활용된다. 이러한 방식은 수중운동체가 위험 지역을 회피하면서 목표 지점까지 효율적으로 접근하도록 한다.

3.2 심도제어 모듈

심도제어 모듈은 강화학습 알고리즘인 PPO를 기반으로 구성된다. 해당 모듈은 경로계획 모듈로부터 전달된 경유 지점을 따라 수중운동체가 안정적이고 효율적으로 심도를 조절하여 이동할 수 있도록 학습된 정책을 수행하는 역할을 한다. PPO는 안정적인 수렴 성능과 쉬운 구현으로 복잡한 연속 제어 문제에 효과적으로 적용되며, 본 연구에서는 이를 심도 조절 문제에 최적화하였다. 정책 업데이트 시 급격한 변화를 방지하면서 학습 효율을 높이기 위해 클리핑 기법을 도입한 것이 특징이다.

이를 이용하여 식 (5)와 같이 손실함수를 정의할 수 있다.

$$L^{CLIP}(\theta) = E_t[\min(r_t(\theta)A_t, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)A_t)] \tag{5}$$

여기서 θ는 현재 정책의 파라미터이고, $r_t(\theta)$은 식 (6)과 같이 현재와 이전 정책의 확률 비를 의미한다. ϵ은 급격한 변화에 따른 학습의 불안정성을 방지하는 계수로 본 논문에서는 0.1을 적용하였다.

$$r_t(\theta) = \frac{\pi_\theta(\delta_e^{PPO} \mid s_t)}{\pi_{old}(\delta_e^{PPO} \mid s_t)} \tag{6}$$

시점 t에서의 상태변수 s_t는 식 (7)과 같다. 이 값이 1보다 크면 해당 행동을 더 선호하는 것이다. A_t는 어드밴티지 함수로 식 (8)과 같이 현재 상태에서 해당 행동이 얼마나 유리한지 나타내는 값이다. 이 값이 0보다 크면 해당 행동이 좋은 행동으로, 0보다 작으면 나쁜 행동으로 판단하게 된다.

$$s_t = [d, \dot{d}, d_{cmd}] \tag{7}$$

$$A_t = \sum_{l=0}^{\infty} (\gamma\lambda)^l \{r_{t+l} + \gamma V(s_{t+l+1}) - V(s_{t+l})\} \tag{8}$$

현재 시점을 기준으로 l시간 후의 상태변수의 기댓값 $V(s_{t+l})$은 식 (9)와 같이 정의할 수 있다.

$$V(s_{t+l}) = E\left[\sum_{k=0}^{\infty} \gamma^k r_{t+l+k} \mid s_{t+l}\right] \tag{9}$$

심도 추종 정확도, 제어 입력 변화율의 최소화, 배터리 소모 최소화를 위한 다목적 보상 함수는 식 (10)

과 같이 정의된다. 이때, 제어 입력 변화율은 식 (11)과 같다. 각각의 가중치 α, β, γ의 경우, 심도 오차를 보상하기 위해 α는 1로 설정하고, 배터리 소모와 제어 진동 정도에 따라 β는 0.05, γ는 0.1로 설정하였다.

$$r_t = -\alpha|d(t) - d_{cmd}| - \beta\left(\delta_e^{PPO}\right)^2 - \gamma|\Delta_e^{PPO}| \quad (10)$$

$$\Delta_e^{PPO} = \delta_e^{PPO}(t) - \delta_e^{PPO}(t-1) \quad (11)$$

식 (5)의 손실함수가 0에 가까워지도록 학습하여 산출되는 제어 입력값이 타각 명령에 반영되도록 피드백한다. 이때, 타각 명령이 수중운동체의 종축 주행에 미치는 영향을 분석하기 위해 선형 운동방정식을 식 (12) – 식 (17)과 같이 정의하였다.

$$O_{long}\dot{X}_{long} = P_{long}X_{long} + Q_{long}u_{long} \quad (12)$$

$$O_{long} = \begin{bmatrix} m - Z_{\dot{w}} & -Z_{\dot{q}} & 0 & 0 \\ -M_{\dot{w}} & I_{yy} - M_{\dot{q}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (13)$$

$$P_{long} = \begin{bmatrix} Z_w & Z_q + mV_{tot} & -(W-B)\sin\theta_0 & 0 \\ M_w & M_q & -Bb_x\sin\theta_0 + Bb_z\cos\theta_0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & -V_{tot} & 0 \end{bmatrix} \quad (14)$$

$$Q_{long} = \begin{bmatrix} Z_{\delta e} \\ M_{\delta e} \\ 0 \\ 0 \end{bmatrix} \quad (15)$$

$$X_{long} = \begin{bmatrix} w \\ q \\ \theta \\ z \end{bmatrix} \quad (16)$$

$$u_{long} = \delta_e \quad (17)$$

여기서, $Z_{\dot{w}}$, $Z_{\dot{q}}$, $M_{\dot{w}}$, $M_{\dot{q}}$: 수중운동체의 형상 및 운동 특성을 반영한 선형 유체력 계수,

m: 질량(단위: kg),

W: 중량($= mg$, 단위: N),

B: 부력(단위: N),

V_{tot}: 속력(단위: m/s),

I_{yy}: y축에 대한 관성모멘트(단위: kg·m^2)

b_x, b_z: 무게중심과 부력중심 간의 x, z축 거리(단위: m)

θ_0: 정상상태 받음각(단위: rad)

4. 해저지형 회피 및 심도 추종 성능 분석

3장에서 기술한 하이브리드 구조의 A* 알고리즘 기반 경로계획 및 PPO 기반 심도제어 제어기의 성능을 다양한 해저 환경 시나리오에서 검증하고, 기존 방식들과의 정량적 비교를 수행하였다. 이때, 수중운동체의 횡축 제어는 반영하지 않으므로 한 방향에 해당하는 해저지형을 기준으로 삼았고, 시나리오 입력 변수는 Table 1에 기술하였다.

Table 1. Input parameters for scenario

Parameters	Value	Unit	Description
$[x_0, z_0]$	[0 10]	m	Start position
$[x_g, z_g]$	[200 10]	m	Goal position
V_{tot}	2	m/s	Speed
Δ_t	0.1	sec	Simulation time step

실험 환경은 3가지 유형의 지형 조건을 기준으로 구성하였다. 첫 번째는 일정한 수심이 유지되는 평탄 지형, 두 번째는 완만한 경사와 봉우리가 혼재된 중간 복잡도 지형, 세 번째는 얕은 수역과 깊은 골, 급경사가 불규칙하게 분포된 지형이다. 평가 지표는 충돌률, 에너지 소비량, 명령 추종 오차를 정량적 지표로 삼았다. 이때, 충돌률은 주행 시작부터 끝까지 지형과의 접촉이나 최소 운용 심도를 벗어나는 횟수를 카운트하였다. 에너지 소비량은 제어 입력에 대한 심도 변화량을 누적 합산하였다. 또한 명령 추종 오차는 실제 추종해야 하는 심도 값과의 차이를 평균내었다.

Fig. 2는 평탄 지형에 대한 시나리오 설정을 도시한 것이다. 이때, PPO, A*, 하이브리드 알고리즘을 적용한 각각의 심도 추종 결과는 Fig. 3와 같다. PPO와 A* 알고리즘의 경우 명령 심도와의 마진을 두어 주행하도록 하였고, 하이브리드 구조를 적용했을 때 최적화된 심도로 주행하는 것을 확인하였다.

Fig. 4는 중간 복잡도 지형에 대한 시나리오를 도시한 것으로 주행 방향으로 완만한 경사의 봉우리가 혼재되어 있는 상태이고, 심도 추종 결과는 Fig. 5와 같다. 여기서, 전진 방향에 따른 심도 명령의 차이가 임계치를 넘어가는 경우 이전 심도 명령을 추종하도록 PPO 알고리즘을 적용하였기 때문에 산봉우리 부근에서 직진 주행하는 것을 확인할 수 있다. 하이브리드

구조를 적용하였을 때, 충돌 위험을 감소하면서도 에너지 효율을 위해 A* 알고리즘을 적용하였을 때보다 명령 심도와의 차이를 50 cm 적게 두고 주행하는 것 또한 확인할 수 있다.

Fig. 6는 불규칙한 지형의 시나리오를 도시한 것으로 심도 추종 결과는 Fig. 7과 같다. 중간 복잡도 지형 시나리오를 분석한 결과와 유사함을 확인하였다.

Table 2는 평탄, 중간 복잡도, 고(高)복잡도 지형 시나리오에 적용한 PPO, A*, 하이브리드 알고리즘에 대한 충돌률, 에너지 소비량, 명령 추종 오차를 기술한 것이다. 세 가지 시나리오 모두 최소 운용 심도 이내로 주행하는 경우는 없으므로 충돌률은 0 %이다. 에너지 소비량은 PPO 알고리즘을 적용하였을 때 중간 복잡도, 고복잡도 지형 시나리오에 대해 산봉우리 형태의 심도를 추종하지 않고 직진 주행하는 구간이 존재했기 때문에 나머지 알고리즘을 적용한 경우보다

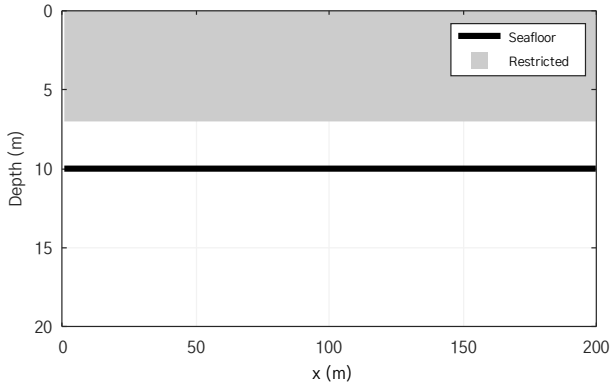

Fig. 2. Scenario for depth tracking paths in flat terrain

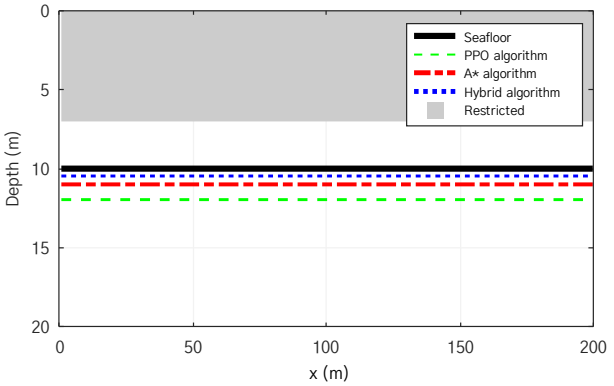

Fig. 3. Depth following performance under flat terrain

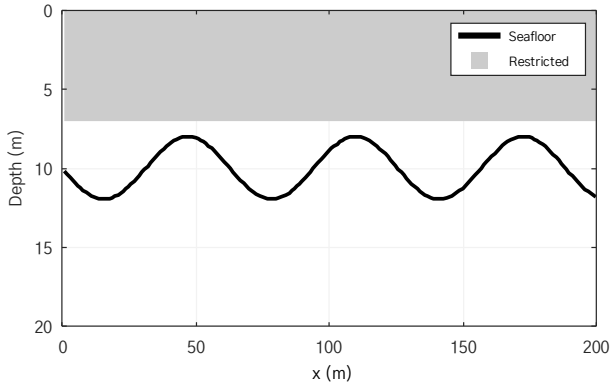

Fig. 4. Scenario for depth tracking paths in moderate terrain

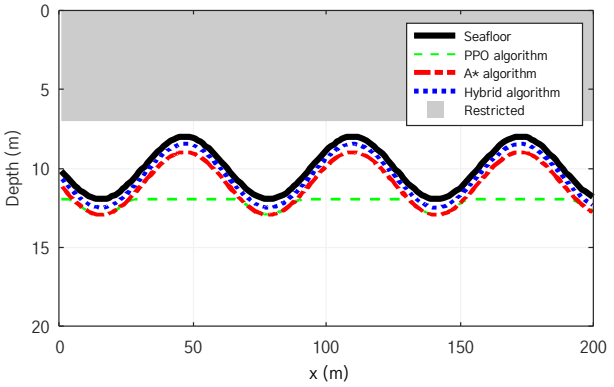

Fig. 5. Depth following performance under moderate terrain

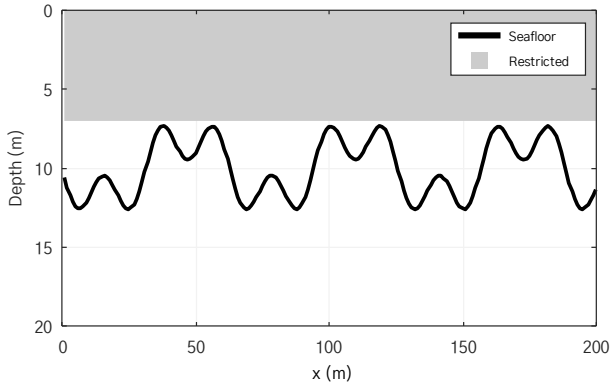

Fig. 6. Scenario for depth tracking paths in complex terrain

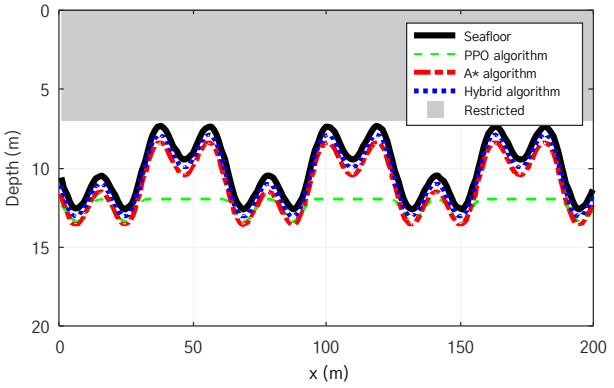

Fig. 7. Depth following performance under complex terrain

작음을 확인할 수 있다. 하지만 명령 심도 추종 오차의 경우 명령 심도와의 유격이 줄어드는 PPO, A*, 하이브리드 순으로 작아짐을 확인하였다. 이러한 에너지 소비량과 명령 추종 오차를 Figs. 8-9에 나타냈다.

Table 2. Comparison of quantitative evaluation result for PPO, A*, and hybrid algorithm

Scenario	Algorithm	Collision rate (%)	Energy consumption	Mean tracking error (m)
Flat	PPO	0	200	2.1
	A*	0	200	1.2
	Hybrid	0	200	0.5
Moderate	PPO	0	206.5	2.2
	A*	0	225.6	1.1
	Hybrid	0	225.6	0.4
Complex	PPO	0	218.4	2.3
	A*	0	259.5	1.2
	Hybrid	0	259.5	0.6

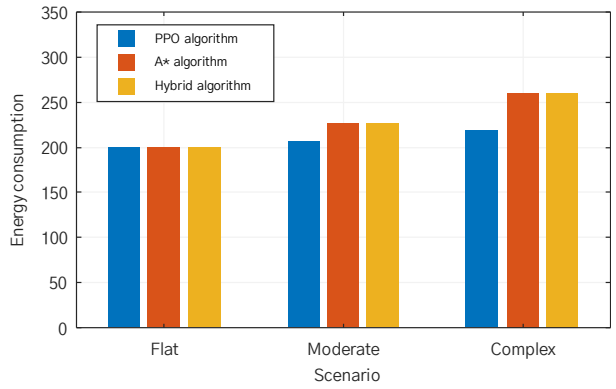

Fig. 8. Energy consumption for flat, moderate, and complex terrain Scenarios

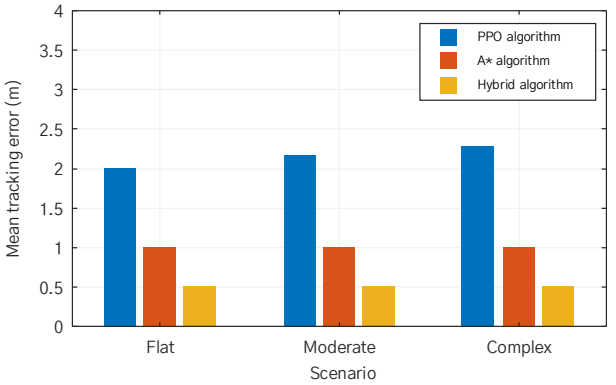

Fig. 9. Mean tracking error for flat, moderate, and complex terrain Scenarios

PPO 단독 제어기를 적용하는 경우 환경 적응성은 있으나 경로점 추종 성능이 낮고, A* 알고리즘 단독 방식은 전역 경로는 확보되지만 충돌률이 커지는 위험이 있다. 반면 본 논문에서 제안한 하이브리드 구조를 적용했을 때 두 기법의 장점이 유기적으로 결합됨으로써 주행 성공률, 안정성, 에너지 효율성 측면에서 모두 우수함을 확인하였다.

5. 결론

수중자율운동체가 주행할 때, 해저산과 같은 특정 장애물을 회피하여 심도를 추종하는 것이 중요하다. 이 때, 시간당 배터리 소모량을 최소로 하여 효율적으로 주행하는 것이 핵심이다.

이에 따라 본 논문에서는 복잡한 해저지형 환경에서 수중운동체의 안정적인 경로 추종을 달성하기 위해 A* 알고리즘 기반의 전역적인 경로계획과 PPO 알고리즘 기반의 강화학습 심도제어기를 통합한 하이브리드 제어 구조를 제안하였다. A* 알고리즘은 전역적인 최단경로 생성을 통해 전체적인 지형 회피 경로를 제공하고, PPO 제어기는 실시간 심도 조절을 통해 제한 수심을 회피하면서 목표 심도를 추종하도록 학습하는 방식을 적용하였다. 특히 운동체의 종축에 대한 안정성, 충돌 회피, 에너지 효율성을 통합적으로 고려한 다목적 보상함수를 강화학습에 적용함으로써 기존 제어기의 한계를 보완하였다.

평탄하거나 복잡한 지형에 대한 주행 시나리오를 분석한 결과, 제안한 하이브리드 방식은 경로계획이나 심도제어 단일 적용에 비해 높은 주행 성공률, 낮은 충돌률, 에너지 소비 측면에서의 균형적인 성능을 보였다. 특히 고복잡 해저지형에서도 안정적인 심도 추종이 가능함을 통해 실해역 응용 가능성을 확인하였다.

향후 연구에서는 현실적인 해양 환경을 고려하여 조류나 외란, 센서 잡음 등의 요인을 모델에 반영하여 강화학습 정책의 안정성을 검증할 필요가 있다. 또한 경로계획과 심도제어, 침로 제어를 분리된 모듈이 아닌 통합된 강화학습 구조로 학습시키는 연구가 필요하다.

본 논문의 설계 및 분석을 통해 수중자율운동체의 해저지형 회피 및 추종 성능을 확인함으로써, 제어기 성능을 주행 전에 사전 분석하고, 주행 성능과 안정적

인 경로계획 수립에 본 연구의 결과를 활용할 수 있을 것으로 기대된다.

참고문헌

[1] H. S. Kim, C. Hyun, S. K. Lee, J. Y. Go and C. H. Kim, “Research on Terrain Tracking Algorithm for Seafloor Topography Avoidance of Underwater Autonomous Vehicle,” Journal of the Korea Society for Naval Science and Technology, Vol. 7, No. 3, pp. 304-308, Sep. 2024.
[2] J. Du, D. Zhou and W. Wang, “Reference Model-based Deterministic Policy for Pitch and Depth Control of Autonomous Underwater Vehicle,” Journal of Marine Science and Engineering, Vol. 11, No. 3, pp. 1-23, Mar. 2023.
[3] A. Zhang, W. Wang, W. Bi and Z. Huang, “A Path Planning Method based on Deep Reinforcement Learning for AUV in Complex Marine Environment,” Ocean Engineering, Vol. 313, No. 1, pp. 1-9, Dec. 2024.
[4] H. Wu, S. Song, K. You and C. Wu, “Depth Control of Model-free AUV via Reinforcement Learning,” IEEE Transactions on Systems, Man, and Cybernetics, Vol. 49, No. 12, pp. 2499-2510, Dec. 2019.
[5] H. S. Kim, C. Hyun, S. K. Lee and J. Y. Go, “A Study on Underwater Autonomous Vehicle Terrain-following Depth Control Algorithm Using Reinforcement Learning Method,” Journal of the Korean Institute of Defense Technology, Vol. 7, No. 1, pp. 7-12, Jul. 2025.
[6] Li. Y, He. X, Lu. Z, Jing. P and Su. Y, “Comprehensive Ocean Information Enabled AUV Motion Planning based on Reinforcement Learning,” Remote Sensing, Vol. 15, No. 12, pp. 1-22, Dec. 2023.
[7] B. Hadi, A. Khosravi and P. Sarhadi, “Deep Reinforcement Learning for Adaptive Path Planning and Control of an Autonomous Underwater Vehicle,” Applied Ocean Research, Vol. 129, No. 9, pp. 1-8, Dec. 2022.