

강화학습 기반 변침점 최적화를 고려한 자율 수중 차량 제어

노지민, 박수현* 김중헌

고려대학교, *숙명여자대학교

emilyjroh@korea.ac.kr, *soohyun.park@sookmyung.ac.kr, joongheon@korea.ac.kr

Waypoint Optimization based Autonomous Underwater Vehicle Control using Reinforcement Learning

Emily Jimin Roh, Soohyun Park*, Joongheon Kim

Korea Univ., *Sookmyung Womans' Univ.

요약

현재 본 논문은 해양 환경에서의 자율 수중 차량(autonomous underwater vehicle, AUV)의 목표 지점 도달을 위한 강화학습 기반 변침점 최적화 방법을 제안한다. AUV의 제어에 있어 실제 해양 환경에서의 지형물로 인한 장애물과 어뢰의 각도 전환 시 발생하는 변침점의 최소화를 고려하여 3차원 환경에서의 어뢰 기동의 자율성과 효율성을 증대한다. 본 연구에서는 최대 90도로 회전할 수 있는 AUV의 특성을 고려한 행동(action)을 정의하였다. 이후 Markov decision process 기반 강화학습 알고리즘인 Proximal Policy Optimization (PPO)를 사용하여 AUV의 변침점 최적화를 통한 목표지점까지의 경로를 산출한다. 최종적으로 본 알고리즘의 보상 함수 수렴성 및 변침점 생성 결과를 통해 본 알고리즘의 우수성과 실제 해양 환경에서의 적용 가능성을 제시한다.

I. 서론

자율 수중 차량(autonomous underwater vehicle, AUV)은 해양 탐사, 군사 및 상업적 목적을 위한 다양한 임무에서 중요한 역할을 담당하고 있다 [1]. 이러한 AUV는 인간의 직접적인 조작 없이도 스스로 임무를 수행할 수 있는 능력을 갖추고 있어, 해양 환경에서의 탐사 및 작업의 효율성을 크게 향상시킨다. 그러나, 실제 해양 환경은 복잡하고 예측하기 어려운 지형적 장애물과 다양한 환경 조건들로 인해 AUV의 자율적인 경로 계획 및 제어에 있어 많은 도전 과제를 제기한다. 또한, 이러한 경로 계획에서 AUV의 각도 전환 시 발생하는 변침점을 최소화하는 것은 매우 중요하다 [2]. 변침점이 많아지면 에너지 소모가 증가하고, 임무 수행 시간이 길어질 뿐만 아니라, AUV의 기동 효율성 또한 저하되기 때문이다. 따라서 본 연구에서는 이러한 문제를 해결하기 위해 강화학습 기반의 변침점 최적화 방법을 통한 AUV 제어 방법을 제안한다. 강화학습은 에이전트가 환경과 상호작용하며 보상을 최대화하는 행동 전략을 학습하는 방법으로, AUV의 환경과 같이 복잡한 의사 결정 문제를 해결하는 데 효과적으로 사용될 수 있다[3-4]. 특히, 본 연구에서는 최대 90도로 회전할 수 있는 AUV의 특성을 고려하여 행동(action)을 정의하고, Markov decision process (MDP) 기반의 강화학습 알고리즘인 proximal policy optimization (PPO)을 사용하여 AUV의 변침점 최적화 방법을 제안한다. 제안된 알고리즘은 에피소드 진행에 따른 누적 보상의 수렴성과 변침점 생성 결과를 통해 우수성과 실제 해양 환경에서의 활용 가능성을 제시한다.

II. Proximal Policy Optimization

강화학습은 에이전트(agent)와 환경(environment)으로 구성되며, 에이전트가 현재 상태(state)에 행동을 취함에 따라 환경으로부터 보상(reward)을 얻어 다음 상태로 넘어가게 된다. 이러한 과정을 반복하는 것이 순차적 의사결정 문제이며, 이러한 문제를 해결하기 위한 최적의 정책(policy)을 학습하는 것이 강화학습의 목표이다[5]. 본 연구에서는 에이전

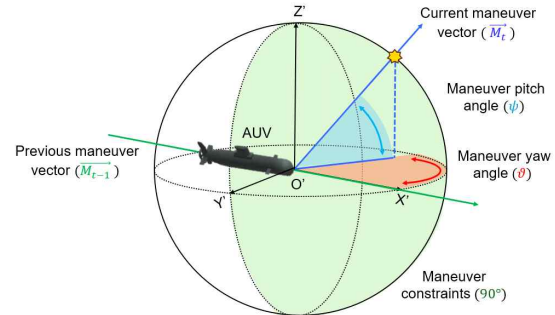


그림 1 자율 수중 차량 특성에 따른 행동

트인 AUV의 특정 상태에서 얻을 수 있는 가치(value)를 학습하며 최종적으로 강화학습 알고리즘 중 PPO를 사용하여 변침점 최적화를 고려한 AUV 제어 방법을 제안한다. PPO는 AUV가 정책을 업데이트하는 과정에서 새로운 정책이 이전 정책과 지나치게 큰 차이를 보이지 않도록 제한하는 방법인 clipping을 통하여 학습 과정에서 발생할 수 있는 불안정성을 줄이고 복잡성을 완화하여 효율적인 학습을 할 수 있다.

III. PPO 기반 변침점 최적화를 고려한 AUV 제어

그림1은 최대 90도로 회전할 수 있는 AUV의 특성에 따른 행동을 설명한다. AUV는 일정 시간 당 같은 거리만을 기동할 수 있으므로 구 공간으로 다음으로의 기동을 나타낼 수 있으며, 그림1에서의 초록색의 구 위의 한 점으로의 기동이 가능하다. 이때 회전 각도는 이전 기동 각도인 M_{t-1} 과 현재의 기동 벡터인 M_t 의 사이 각도로 계산할 수 있다. 이때 M_t 는 AUV의 행동인 XY평면에서의 회전 각도 θ 와 YZ평면에서의 회전 각도 ψ 를 바탕으로 정의된다. 변침점 최적화를 위한 AUV의 보상함수는 다음과 같이 수식 1과 같이 정리된다.

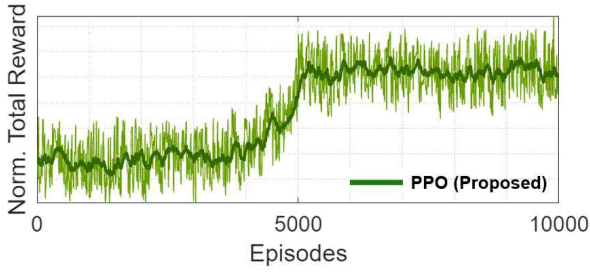


그림 2 에피소드 진행에 따른 누적 보상

$$R = \beta \left(\sqrt{\frac{\alpha^2 (1 - n_i^2)}{K^2}} \right) + \beta \quad (1)$$

보상함수는 $(0, \alpha)$ 와 $(K, 0)$ 을 지나는 타원의 방정식으로 설계하였으며, 변침점의 누적 개수 n_i 가 증가할수록 보상은 작아진다. 이때 α 는 변침점의 생성에 따른 에민도에 해당하는 값으로 0에서 1까지의 값을 가질 수 있다. 해당 실험 환경에서는 0.8로 설정했다. K 는 시스템에서 설정한 최대 변침점의 개수로, 본 환경에서는 14로 설정했다. 만약 n_i 가 K 이상이라면 보상은 -1로 부여하였다. β 는 AUV의 목표지점 도달의 여부로 목표지점에 도달하면 1의 보상을 받도록 설정했다.

IV. 실험 및 성능 평가

본 논문은 실제 해양 환경에서처럼 3차원 환경에서 성능 평가를 진행하였으며, 8개의 무작위 크기와 위치를 갖는 구 형태의 장애물을 구성한 환경에서 실험을 진행하였다. 추가적으로 impact angle을 통하여 실제의 해양 환경에서 임무 수행 성공률을 높일 수 있는 제약 조건을 추가한 환경을 설정하였다. Impact angle은 목표 지점에서의 기동 시 만족해야 하는 AUV의 회전 각도로 이를 설정함으로써 목표 타겟의 AUV의 노출을 최소화하므로 임무 수행의 성공률을 높일 수 있다. 에피소드는 총 10000으로 설정하였으며, 그림 2에서의 에피소드의 진행에 따른 누적 보상 그래프를 통하여 PPO 기반 AUV 제어를 위한 알고리즘의 안정적인 수렴성을 확인할 수 있다. 또한, 그림 3의 시뮬레이션 결과를 통하여 PPO 기반 변침점 최적화를 고려한 AUV 제어의 성능을 확인할 수 있다. AUV가 환경의 가운데 지점에서 목표 지점까지의 기동하는 것을 확인할 수 있으며, 노란색 점으로 표현되는 변침점의 개수는 3개로 변침점을 최소화하면서 기동함을 확인할 수 있다. 이를 통하여 제안한 PPO 기반 변침점 최적화를 고려한 AUV 제어 알고리즘의 우수성과 실제 해양 환경에서의 활용 가능성을 확인할 수 있다.

V. 결론

본 논문에서는 강화학습 알고리즘 중 PPO를 활용하여 AUV의 변침점 최적화를 통한 목표 지점까지의 효율적인 경로 산출 및 제어에 관한 연구를 수행하였다. 또한 실제 복잡하고 시간에 따라 변화하는 해양 환경에서의 지형물을 고려한 3차원 환경에서 실험을 진행하였으며, 실험 결과를 통하여 본 알고리즘의 안정성과 실제 해양 환경에서의 활용 가능성을 확인했다.

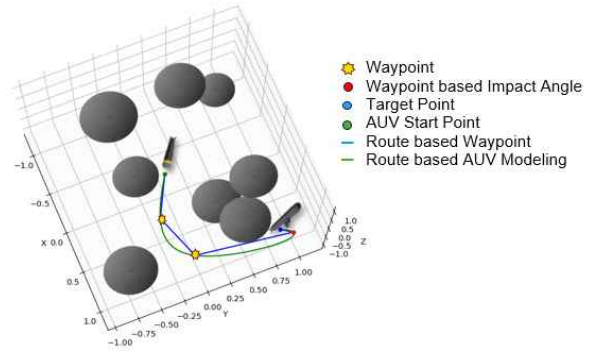


그림 3 PPO 기반 변침점 최적화를 고려한 AUV 제어

ACKNOWLEDGMENT

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2022-0-00907, (2세부) AI Bots 협업 플랫폼 및 자기조직 인공지능 기술개발). 본 논문의 교신저자는 박수현임.

참고 문헌

- [1] A. Sahoo, S. K. Dwivedy, and P. Robi, "Advancements in the field of autonomous underwater vehicle," *Ocean Engineering*, vol. 181, no. 11, pp. 145 - 160, June 2019.
- [2] Y. Zhang, Q. Wang, Y. Shen, and B. He, "An online path planning algorithm for autonomous marine geomorphological surveys based on AUV," *Engineering Applications of Artificial Intelligence*, vol. 118, no. 3, p. 105548, February 2023.
- [3] Cao, C. Sun, and M. Yan, "Target search control of AUV in underwater environment with deep reinforcement learning," *IEEE Access*, vol. 7, no. 7, pp. 96549 - 96559, July 2019.
- [4] P. Yao, L. Qiu, J. Qi, and R. Yang, "AUV path planning for coverage search of static target in ocean environment," *Ocean Engineering*, vol. 241, no. 23, p. 110050, December 2021.
- [5] S. Park, C. Park, and J. Kim, "Learning-Based Cooperative Mobility Control for Autonomous Drone-Delivery," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 4, pp. 4870-4885, April 2024.