



Received: 2020/10/20
Revised: 2021/02/01
Accepted: 2021/03/15
Published: 2021/03/27

***Corresponding Author:**

Seungjo Hong

E-mail: hongsj235@naver.com

강화학습 기반 예인선 제어방법에 의한 선박 자율 접안

Reinforcement Learning Based Tugboats Control for Autonomous Ship Berthing

홍승조^{1*}, 김진환²

¹해군 소령/해군제2함대사령부 인천함 지원부장

²KAIST 기계공학과 부교수

Seungjo Hong^{1*}, Jinwhan Kim²

¹Lieutenant Commander/Executive Officer, ROKS Incheon, 2nd Fleet command, ROK Navy

²Associate Professor, Dept. of Mechanical Engineering, Korea Advanced Institute of Science and Technology

Abstract

선박 접안 작업은 모든 유형의 선박에 해당된다. 특히 저속 기동성이 제한된 대형 선박에는 예인선이 필수적이다. 예인선의 접안 작업은 도선사 직관에 의해 수동적으로 이루어지는 작업으로 일반적인 제어 알고리즘을 적용하기 어려운 주제이다. 최근에는 이러한 복잡한 문제를 해결하기 위해 머신 러닝 기술 중 하나인 강화 학습이 적용되고 있다. 본 연구에서는 예인선을 활용한 접안 문제에 강화학습 방법을 적용하여 문제를 해결하였다. 예인선 및 선박의 동적 모델은 연속 및 이산시간의 동적 모델이 상호 작용하는 하이브리드 시스템으로 정의된다. 그리고 강화학습에서 정책 기울기 방식의 대표적인 알고리즘인 근위 정책 최적화(PPO) 알고리즘을 이용하여 접안문제를 해결하였으며 그 결과는 시뮬레이션으로 보여진다.

Berthing operation is essential for all types of ships. In particular, tugboats are required to large vessels with limited actuation capabilities at low speed. Tugboat automation is a difficult topic to apply typical control algorithms. Recently, to solve these problems, reinforcement learning, one of the techniques of machine learning, is being implemented. In this research, we solved the problem of tugboats using reinforcement learning method. The tugboats dynamic model is defined as a hybrid system in which continuous and discrete dynamic models interact. And solve this problem with proximal policy optimization(PPO) algorithm which trust region method based policy gradient algorithm. The result is shown through a simulation.

Keywords

강화학습(Reinforcement Learning),
선박 자율 접안(Autonomous Ship Berthing),
하이브리드 시스템(Hybrid System),
예인선 제어(Tugboat Control)

Acknowledgement

이 논문은 2020년도 한국해군과학기술학회
추계학술대회 발표 논문임.

1. 서론

선박에 있어서 접안 작업은 모든 형태에 선박에 동일하게 적용되는 필수적인 작업이다. 특히 저속에서 기동성이 제한되는 대형 상선과 같은 경우에는 예인선의 지원이 필수적이라고 할 수 있다. 항구에서는 다양한 작업들이 빈번하게 일어나는데 선박이 견선거에 진·출입하거나 좁은 갑문을 통과하는 등 정밀한 작업이 필요할 때는 예인선의 지원이 필수적이다.

지금까지 예인선에 의한 선박 접이안 작업은 도선사와 예인선에 의한 수동적인 작업으로 이루어져 왔다. 하지만 실제 접이안을 위한 예인선의 사용에는 많은 비용이 소모되고 도선사 및 예인 운용인원 등 여러 인력이 반드시 필요하다. 따라서 이러한 제한점을 보완하기 위해서 예인선의 자율 운항은 상당히 필요한 과제이고, 최근 무인선에 대한 연구가 활발히 진행됨에 따라서 발전 가능성이 높은 분야라고 할 수 있다.

선박 접안작업에서 예인선의 자율화 문제는 예인선의 위치이동 문제와 제어 문제가 복합적으로 이루어진 문제이다. 이러한 문제를 해결하기 위해 최근에는 기계학습의 종류 중 하나인 강화학습 방법을 이용한 문제 해결 시도가 늘어나고 있는 추세이다. 강화학습은 매우 복잡한 비선형 문제나 연속적 비연속적 시스템의 상호작용으로 이루어진 하이브리드 시스템 문제와 같이 기존 제어방법으로는 해결하기 어려운 문제에 적용 가능한 장점이 있다.

예인선을 사용한 선박 자율 접안 관련 연구로는 선박의 계류 위치에 추진기가 장착되어 있는 것과 같이 예인선을 모델링 하고, 이를 슬라이딩 모드 제어 방식으로 접안 작업을 하는 연구[2]가 있었으며, 신경망(neural network)을 이용하여 신경망의 출력값을 선박과 예인선의 제어 입력값으로 사용하는 연구[3]가 있었다.

하지만 기존 연구의 대부분은 예인선이 가하는 힘을 선박에 부착된 추진기 형태로 모델링한 후에 기존 제어기법을 적용한 것이어서 예인선이 선박 주변에서 이동해가며 본선을 밀거나 당기는 실제 운용 방식과는 큰 차이가 있다.

따라서 본 연구에서는 실제 예인선의 운용 방식을 최대한 고려하여 정책 경사(policy gradient) 방법의 강화학습 알고리즘 중 하나인 PPO(proximal policy gradient) 알고리즘을 이용하여 예인선의 위치 이동까지 고려한 선박의 접안 문제를 해결하고자 하였다. 예인선이 선박에 계류하는 위치를 2가지 모드로 정의하였으며 각각의 모드에서 예인선 제어 문제를 강화학습으로 해결하였다. 강화학습에서 동역학적 모델의 구성을 위하여 각 모드에서 예인선의 계류 위치와 선박의 기하학적 관계를 이용하여 선박에 작용하는 힘을 계산하되 바람에 의한 외란을 고려한 모델을 구성하였다. 제안하는 강화학습 기반 접안 방법의 유용성을 확인하기 위하여 외란이 없는 환경뿐만 아니라 외란이 강하게 존재하는 환경에서 학습 및 검증을 수행하였다.

2. 시스템 모델링

2.1 선박 모델링

선박 모델은 비선형 3 자유도 선박 모델을 사용하였다. 선박의 상태는 추진 방향의 종방향 운동(serge)과 그 직각 방향인 횡방향 운동(sway) 그리고 선박의 수직축 중심의 회전 운동(yaw)으로 계산되며 운동방정식은 식 (1)과 같다[5].

$$M\dot{v} + Dv = \tau, \quad \eta = R(\psi)v \tag{1}$$

η, v 는 $\eta = [x, y, \psi]^T \in R^3, v = [u, v, r]^T$ 를 나타내며 η 는 배의 위치(x, y)와 선수각(ψ), v 는 x, y 방향의 선형 속도와 선회율을 의미한다. τ 는 선박에 가해지는 힘과 모멘트 값이며, 선박의 제어 입력값으로 사용된다.

$M \in R^{3 \times 3}$ 과 $D \in R^{3 \times 3}$ 은 각각 관성행렬(inertia matrix)과 제동행렬(damping matrix)을 의미하며 M 은 선박의 질량과 유체역학적 추가 질량(hydrodynamic added mass)

으로 계산되고, D 는 선형 제동행렬, $R(\psi)$ 은 회전행렬(rotation matrix)이다. 각각의 행렬은 식 (2)와 같이 계산된다.

$$M = \begin{bmatrix} m - X_{\dot{u}} & 0 & 0 \\ 0 & m - Y_{\dot{v}} & -Y_{\dot{r}} \\ 0 & -N_{\dot{v}} & I_z - N_{\dot{r}} \end{bmatrix}$$

$$D = \begin{bmatrix} -X_u & 0 & 0 \\ 0 & -Y_v & -Y_r \\ 0 & -N_v & -N_r \end{bmatrix} \tag{2}$$

$$R(\psi) = \begin{bmatrix} \cos(\psi) & -\sin(\psi) & 0 \\ \sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

2.2 예인선 모델링

본 연구에서는 Fig. 1과 같이 2대의 예인선을 활용하며, 각 예인선은 지정된 위치에 계류한다. 예인선의 힘의 작용점은 변하지 않는다고 가정한다. 선박에 가해지는 힘 τ 는 각 예인선의 힘이 선박의 중심에 작용하는 힘의 기하학적 관계로 식 (3)과 같이 계산된다.

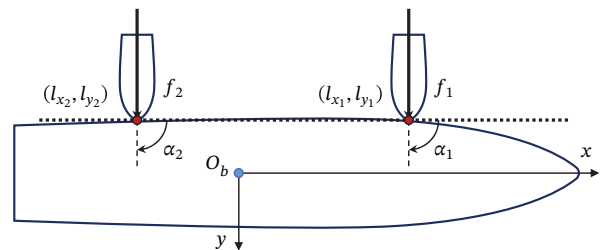


Fig. 1. Ship motion by tugboats

$$\tau = \begin{bmatrix} \cos(\alpha_1) \sin(\alpha_1) - l_{y1} \cos(\alpha_1) + l_{x1} \sin(\alpha_1) \\ \cos(\alpha_2) \sin(\alpha_2) - l_{y2} \cos(\alpha_2) + l_{x2} \sin(\alpha_2) \end{bmatrix}^T \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \tag{3}$$

f_1, f_2 는 두 대의 예인선이 작용하는 힘, α_1, α_2 는 예인선의 힘의 방향을 나타낸다. 이는 본체 고정 좌표(body fixed frame)의 x 축을 기준으로 한다. l_x, l_y 는 선박의 중심에서부터 i 번째 예인선의 작용점까지의 거리를 의미한다.

2.3 하이브리드 시스템 모델링

선박 접안 문제는 예인선의 제어를 나타내는 연속 시간 동적모델과 예인선의 위치변화에 따른 이산 시간 동적모델이 조합되어 구성된 형태이며, 해당 문제의 모델링을 위하여 연속적, 비연속적 동적모델이 상호작용하는 하이브리드 시스템 개념을 활용하여 문제를 정의하였다.

본 연구에서는 예인선의 위치이동을 정의하기 위해서

예인모드와 접안모드의 두 가지로 설정하였다. 예인모드에서는 예인선이 선박의 선수, 선미 방향에서 서로 당기며, 접안모드에서는 예인선이 선박의 한 현측에 계류하여 선박을 밀고 당기는 역할을 한다.

Fig. 2는 예인선 전환 하이브리드 시스템을 도식화한 것이다. 예인선은 q 값에 의해서 모드가 전환되며 q_1 은 예인 모드(towing mode), q_2 는 접안모드(berthing mode)를 나타낸다. 모드가 전환될 때에는 예인선의 위치 변화가 이루어진다. 예인선은 선박에서 이탈하여 계류 위치를 변경하고, 변경하는 시간 T 동안에는 제어입력 τ 가 0이 되도록 하였다. G 함수는 변수 q 값이 이전 값과 같으면 해당 예인선 모드를 유지하며, 이전 값과 다르면 모드 전환이 이루어진다. 새로운 모드가 시작될 때는 R 함수에 의해 제어 입력값이 초기화된다.

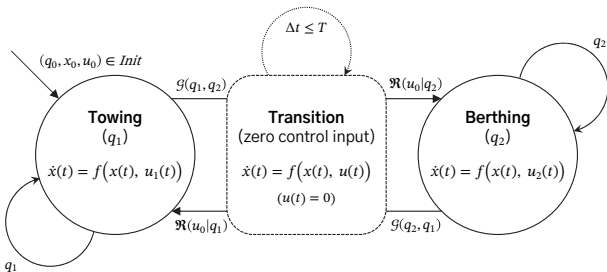


Fig. 2. Tugboat transition hybrid system

2.4 외란 모델링

해양환경에서는 바람, 조류, 파도에 의한 외란이 작용하나, 선박 접안이 이루어지는 항구에서는 바람에 의한 영향이 조류, 파도에 의한 영향에 비해서 지배적으로 작용한다. 따라서 본 연구에서는 바람에 의한 외란의 영향을 고려하였다.

선박에 미치는 바람의 힘은 선박의 측면 투영면적과 풍력 계수로 식 (4)와 같이 표현된다[1].

$$\tau_{wind} = \frac{1}{2} \rho_a V_{rw}^2 \begin{bmatrix} C_X(\gamma_{rw}) A_{Fw} \\ C_Y(\gamma_{rw}) A_{Lw} \\ C_N(\gamma_{rw}) A_{Lw} L_{oa} \end{bmatrix} \quad (4)$$

V_{rw}, γ_{rw} 는 각각 바람의 상대속력과 상대방향, ρ 는 기온에 따른 공기밀도를 의미한다. A_{Fw}, A_{Lw} 는 각각 선박의 전방, 측면 투영 영역의 면적, L_{oa} 는 선박의 길이를 의미한다. C_X, C_Y, C_N 은 바람의 상대방향에 따른 바람 계수이다. 바람의 속력과 방향은 평균값을 기준으로 가우시안 분포(Gaussian density distribution)에서 추출하여 사용한다.

3. 강화학습 알고리즘

3.1 정책 경사기반 근위 정책 최적화(PPO)

본 연구에서는 에이전트의 연속 행동공간(continuous action space)에 적합하고, 안정적이고 높은 성능으로 최근 강화학습에 가장 많이 쓰이고 있는 PPO 알고리즘을 사용하였다.

PPO 알고리즘은 신뢰 영역 기반의 정책 경사 알고리즘으로 연속 행동 집합인 정책을 업데이트할 때 그 범위를 제한하여 안정적인 업데이트를 보장한다. 손실함수는 식 (5)와 같이 계산된다.

$$L(\theta) = E_t [\min(r_t(\theta), \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)) \hat{A}_t], \quad (5)$$

$$r_t(\theta) = \frac{\pi_\theta(a|s)}{\pi_{\theta_{old}}(a|s)}$$

θ 는 정책을 의미하며 π_θ 는 정책 θ 를 따랐을 때 행동 a 를 취할 확률값이다. $r_t(\theta)$ 는 이전 정책과 현재의 정책의 확률값의 비율이고, 이 값과 ϵ 값에 따라서 clipping 된 값 중에서 작은 값을 손실함수 값으로 취하게 된다. \hat{A}_t 는 시간 t 에서 이득값을 의미하고[4], 보상함수 식 (6)으로 계산되는 보상값 R 을 이용하여 계산된다.

3.2 학습 네트워크

PPO 알고리즘에서는 가치 네트워크와 정책 네트워크를 공동으로 사용한다. 공동 네트워크는 2개의 hidden layer를 사용하며 각 layer는 64 노드로 구성되어 있으며 활성화 함수(activation function)로는 Tanh를 사용하였다.

Fig. 3은 본 연구에서 사용된 학습 네트워크이며, 네트워크의 출력은 예인선의 제어 입력값으로 사용한다. 네트워크의 출력값은 각 예인선의 힘의 변화량과 힘의 방향의 변화량 그리고 예인선 모드를 선택하는 총 5개의 값($\Delta f_1, \Delta f_2, \Delta \alpha_1, \Delta \alpha_2, k$)으로 구성되어 있다. 이 중 k 값은 예인선의 이산화된 모드 변화를 나타내기 위하여 -1에서 1까지의 연속적인 action 값을 2개로 이산화하였다. $-1 \leq k < 0$ 이면 q_1 (예인모드), $0 \leq k \leq 1$ 이면 q_2 (접안모드)로 예인선 모드를 선택하도록 하였다.

네트워크 입력은 관찰 공간(observation space)으로 선박의 상태를 나타내는 x, y, ψ, u, v, r 과 보상함수 계산에 사용되는 e_d, e_ψ, e_v (거리오차, 방위오차, 속도오차) 그리고 이

전 단계의 출력값을 현재 단계에서의 네트워크 입력값으로 설정하였다. 네트워크 출력값이 제어입력의 변화량이기 때문에 네트워크가 이를 학습하게 만들기 위해서 이전 단계의 출력값을 다음 단계 입력값으로 정해주었다.

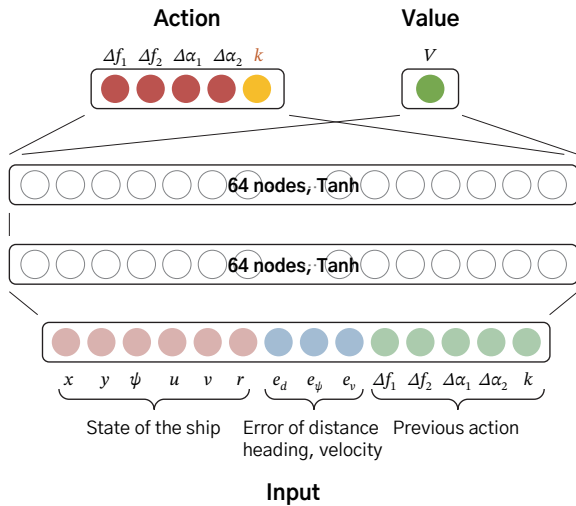


Fig. 3. Policy network and value network configuration

3.3 보상함수

보상함수는 목표하는 선박의 상태를 기준으로 현재 상태와의 오차를 계산하여 오차가 최소가 되었을 때 보상값 1을 받도록 설계하였으며 식 (6)과 같이 정의된다. e_d 는 목표 지점과 현재 위치와의 오차, e_p 는 목표 선수각과 현재 선수각과의 오차를 의미하며 e_v 는 선박의 목표 속력과 현재 속력의 오차를 의미한다. 실제 접안에서는 선박이 저속으로 기동하기 때문에 목표 속력을 1.5 m/s로 설정하였다.

$$R = \frac{-\|e_d\| - \|e_p\| - c\|e_v\| + \alpha}{\alpha} + p \quad (6)$$

세 개의 오차값 중에서 속도 오차값은 상대적으로 작기 때문에 그 비율을 고려하여 상수 c 를 곱해주었고, α 값은 정규화 인자(normalization factor)로 보상값을 0에서 1 사이의 값으로 만들어주는 상수이다. p 값은 벌점으로 선박이 부두에 부딪히면 받는 값으로 -1로 설정하였다.

보상함수 값은 매 시간 스텝마다 계산되어 네트워크가 증치를 업데이트하기 위한 손실함수 계산에 사용된다.

4. 시뮬레이션 결과

시뮬레이터는 Python 기반의 pygame 모듈을 사용하

여 만들었다. 강화학습 환경은 가장 보편적으로 쓰이는 OpenAI Gym 환경을 사용하였다. Fig. 4는 시뮬레이터 화면을 보여준다. 이 시뮬레이터를 기반으로 학습을 진행하였다.

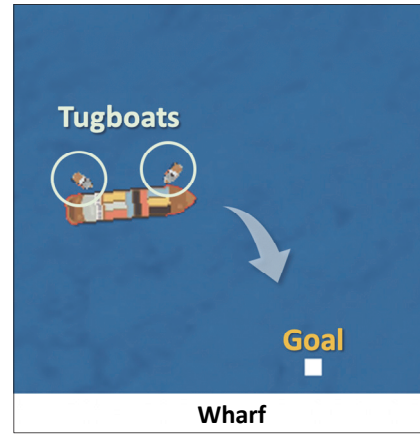


Fig. 4. Ship berthing simulator

4.1 예비 시뮬레이션 결과

본 시뮬레이션에 앞서 예인선 지원 선박 모델의 운동을 확인하기 위하여 Fig. 4의 시뮬레이터 환경에서 학습을 진행하였다. 선박 모델은 총 톤수 6,000톤, 길이 76.2 m의 선박 모델이며 초기위치와 목표위치까지 거리는 182 m이다. 선박이 시뮬레이터 창밖으로 벗어나거나 부두에 부딪히면 에피소드가 종료되고 재시작하도록 하였으며 제어주기는 1 Hz이다.

학습결과 예인선 두 대에 의해 목표 지점에 선박이 자율 접안을 하였으며, Fig. 5는 선박의 궤적을 나타낸다.

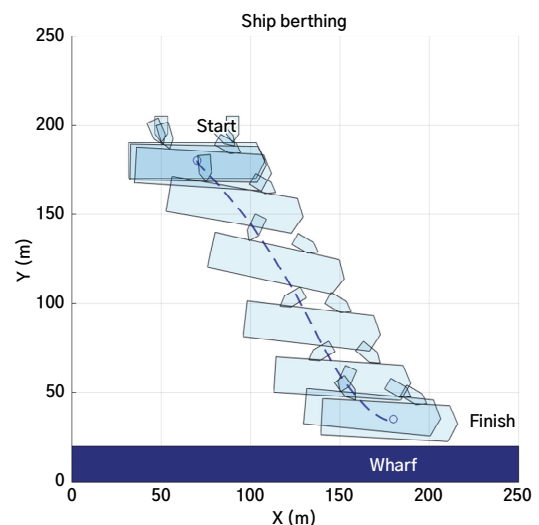


Fig. 5. Trajectory of berthing ship

Fig. 6(a), Fig. 6(b)는 학습된 모델로 검증한 결과를 나타낸다. 각각은 두 대의 예인선에 대한 힘과 힘의 방향에 대한 그래프이다.

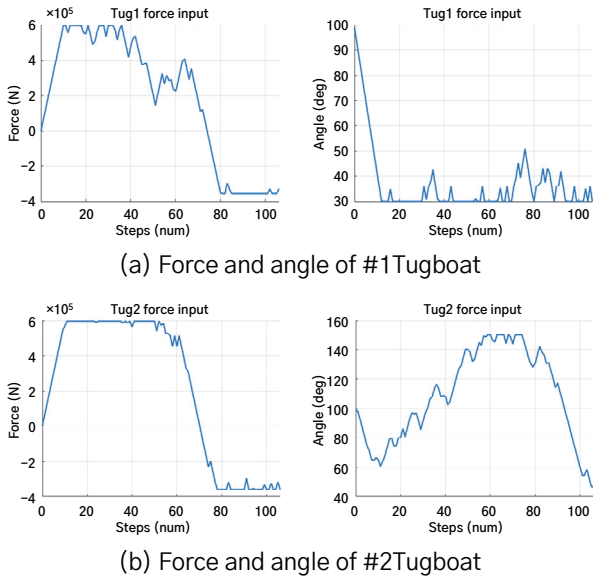


Fig. 6. The value of tugboat force and angle

4.2 본 시뮬레이션 결과

앞서 확인한 예비 시뮬레이션 결과를 바탕으로 본 시뮬레이션을 구현하였다. 본 시뮬레이션에서는 Fig. 7과 같이 실제 인천 내항의 부두 입항 시나리오를 구성하였다. 좁은 갑문에서 시작하여 원하는 부두 위치에 선박이 접안하도록 하였으며 앞서 정의한 하이브리드 시스템 문제를 해결하였다.

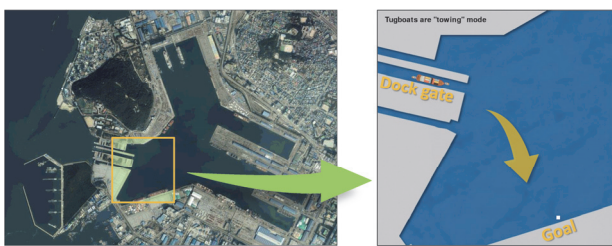


Fig. 7. Image of Incheon Port and simulator

제어입력의 제약조건은 각 예인선 모드에 대해서 설정하였다. 각 모드에서의 제약조건은 Table 1과 같이 설정하였다.

학습 초기에는 좁은 갑문을 통과하기 때문에 선박이 부두에 부딪히거나 예인선이 부두에 부딪혀서 에피소드가 일찍 종료되었지만 학습이 진행됨에 따라서 선박이 갑문을 통과하였고 이후 예인선 모드 전환이 여러 번 일어났다. 최

종 학습 단계에서는 예인선이 예인모드로 갑문을 통과한 이후 한 번의 모드전환이 이루어지고, 최종 위치에 선박이 도착하도록 학습모델이 수렴한 것을 확인할 수 있었다

Table 1. Constraints of control input

Sort	Towing mode		Berthing mode		Unit
	Min	Max	Min	Max	
f_1	-596	0.0	-358	596	kN
f_2	-358	0.0	-358	596	kN
α_1	-240	-120	30	150	°(deg)
α_2	-60	60	30	150	°(deg)
Δf	-59.6	59.6	-59.6	59.6	kN
$\Delta \alpha$	-1.9	-1.9	-5.7	-5.7	°(deg)

학습 초기에는 좁은 갑문을 통과하기 때문에 선박이 부두에 부딪히거나 예인선이 부두에 부딪혀서 에피소드가 일찍 종료되었지만 학습이 진행됨에 따라서 선박이 갑문을 통과하였고 이후 예인선 모드 전환이 여러 번 일어났다. 최종 학습 단계에서는 예인선이 예인모드로 갑문을 통과한 이후 한 번의 모드전환이 이루어지고, 최종 위치에 선박이 도착하도록 학습모델이 수렴한 것을 확인할 수 있었다.

Fig. 8은 학습 경과에 따른 예인선 모드 전환을 나타낸 것이다. 학습 초기에는 좁은 갑문을 통과하기 때문에 선박이 부두에 부딪혀 에피소드가 일찍 종료되었지만 학습이 진행됨에 따라 선박이 갑문을 통과하였으며, 이후 단계에서는 모드전환이 2회 이상 일어났다.

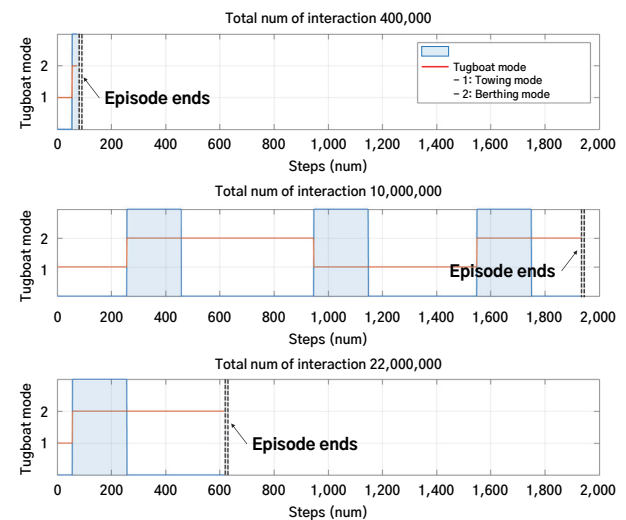


Fig. 8. Tugboat mode transition according to learning progress

최종 학습 단계에서는 단 한 번의 모드 전환이 실행되고 접안모드로 최종 위치에 도달하도록 학습모델이 수렴하였다.

Fig. 9은 학습된 결과를 시뮬레이터 화면상에 나타낸 것이다. 빨간색, 노란색 궤적은 각각 외란이 작용할 때와 작용하지 않을 때 선박의 궤적을 의미하며 녹색 구간에서 예인선 모드 전환이 이루어졌다. 외란은 풍속 11 m/s - 13m/s, 풍향은 205° - 215°로 설정하였다. 강한 외란에서도 강화학습 기반의 제어 방법으로 자율 접안이 가능하다는 유용성을 확인하였다.

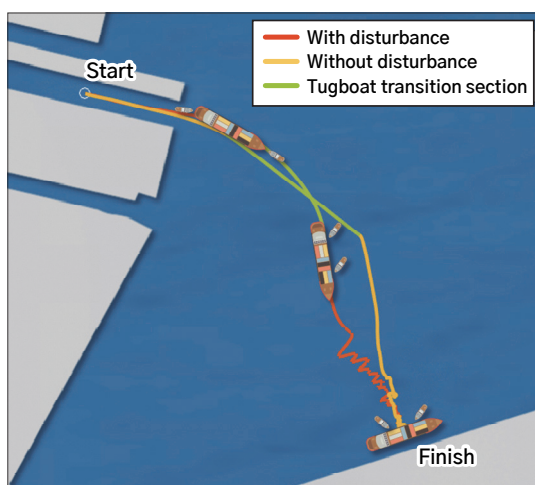


Fig. 9. Training result on simulator

5. 결론

본 논문에서는 정책 경사 방법의 강화학습 알고리즘 중 하나인 PPO 알고리즘을 이용하여 예인선의 위치 이동까지 고려한 선박의 접안 문제 해결을 시도하였다. 예인선 2대에

대해 선박에 계류하는 위치에 따라 예인모드와 접안모드 2가지로 예인선 모드를 정의하였으며 각 모드에서 예인선의 제어문제를 강화학습을 이용하여 해결하였다.

예인선을 이용한 접안 제어 문제를 다루되 강화학습 기법을 활용하여 예선의 이동과 운용 판단의 자율화를 시도하고 그 가능성을 확인한 국내 최초의 연구 사례이며 이후 지속적인 연구를 통해 연구 내용과 결과를 구체화할 예정이다.

참고문헌

- [1] Fossen, Thor I. Handbook of marine craft hydrodynamics and motion control. John Wiley & Sons, 2011.
- [2] Kawai, Hideki, Young Bok Kim, and Kwon Soon Lee. "A ship berthing system design with four tug boats." Journal of mechanical science and technology 25.5, 2011.
- [3] Im, Namkyun. "A study on ship automatic berthing with assistance of auxiliary devices." International Journal of Naval Architecture and Ocean Engineering 4, No. 3, pp 199-210, 2012
- [3] Van Der Schaft, Arjan J., and Johannes Maria Schumacher. An introduction to hybrid dynamical systems, Vol.251. London: Springer, 2000.
- [4] Schulman, John et al. "Proximal policy optimization algorithms." arXiv preprint arXiv:1707.06347, 2017.
- [5] Perez, Tristan, and Thor I. Fossen. "A matlab toolbox for parametric identification of radiation- force models of ships and offshore structures." 2009, 1-15.