Journal of the KIMST, Vol. 24, No. 1, pp. 61-69, 2021
ISSN 1598-9127(Print) · 2636-0640(Online)
DOI https://doi.org/10.9766/KIMST.2021.24.1.061

Research Paper 센서·신호처리 부문

무인수상정 경로점 추종을 위한 강화학습 기반 Dynamic Window Approach

허진영[1] · 하지수[2] · 이준식[2] · 유재관[2] · 권용진*[,1]

[1] 아주대학교 산업공학과
[2] LIG넥스원㈜ 무인체계개발단

Dynamic Window Approach with path-following for Unmanned Surface Vehicle based on Reinforcement Learning

Jinyeong Heo[1] · Jeesoo Ha[2] · Junsik Lee[2] · Jaekwan Ryu[2] · Yongjin Kwon*[,1]

[1] *Department of Industrial Engineering, Ajou University, Korea*
[2] *Unmanned Systems, LIG Nex1 Co., Ltd., Korea*

(Received 6 October 2020 / Revised 26 January 2021 / Accepted 29 January 2021)

Abstract

Recently, autonomous navigation technology is actively being developed due to the increasing demand of an unmanned surface vehicle(USV). Local planning is essential for the USV to safely reach its destination along paths. the dynamic window approach(DWA) algorithm is a well-known navigation scheme as a local path planning. However, the existing DWA algorithm does not consider path line tracking, and the fixed weight coefficient of the evaluation function, which is a core part, cannot provide flexible path planning for all situations. Therefore, in this paper, we propose a new DWA algorithm that can follow path lines in all situations. Fixed weight coefficients were trained using reinforcement learning(RL) which has been actively studied recently. We implemented the simulation and compared the existing DWA algorithm with the DWA algorithm proposed in this paper. As a result, we confirmed the effectiveness of the proposed algorithm.

Key Words : Dynamic Window Approach(동적 창 접근), Path-following(경로 추종), Local Planning(지역 경로), Unmanned Surface Vehicle(무인수상정), Reinforcement Learning(강화학습)

1. 서 론

최근 남북 간 불확실한 국제 정세 및 해안경계 초소 철수 등 해안 경계능력 약화로 이에 대응하기 위한 수단으로 무인수상정(USV, Unmanned Surface Vehicle) 개발 및 보급이 시급한 실정이다.

해외에서는 미국의 Spartan Scout USV, ASW USV,

* Corresponding author, E-mail: yk73@ajou.ac.kr

UISS, 영국의 C-Sweep, 프랑스의 Mk2, 캐나다의 Dolphin, 이스라엘의 Protector 등이 개발되었으며 국내에서는 최근 LIG넥스원의 해검I에 이어 해검II가 개발 중에 있다[1].

Fig. 1. Sea Sword II being developed by LigNex1

무인수상정 수요가 급증함에 따라 이를 자율적으로 운용하기 위한 자율 운항 기술, 장애물 인식 및 충돌 회피 기술 등의 개발이 활발하게 진행되고 있으며, 무인수상정은 자율 운항 시 사용자가 설정한 경로를 따라 안전하게 목적지까지 도착해야한다. 경로 계획은 크게 전역 경로와 지역 경로계획으로 나눌 수 있다. 전역 경로는 무인수상정이 환경에 대한 정보를 지도 형태로 미리 알고 있다고 가정하고 지도에 표기된 지형 및 인공구조물과 충돌하지 않는 최적 경로를 생성하는 것이고 지역 경로는 시간에 따라 변하는 동적 환경에서 충돌을 회피하기 위해 전역 경로를 크게 벗어나지 않으면서 안전한 회피 경로를 계획하는 것이다[2-3].

Dynamic Window Approach(DWA) 알고리즘은 지역 경로 계획으로 이동체의 속도, 방향 및 센서 정보로부터 얻는 장애물과의 거리를 토대로 최적의 선속도와 각속도를 도출하여 장애물을 회피하고 목적지에 도달하는 방법이다[3]. 결과적으로 DWA는 적은 연산량에 의한 빠른 수행속도로 실제 환경에서 우수한 장애물 충돌 회피 성능을 보인다. 그러나 기존의 DWA는 목적지 간의 직선으로 이루어지는 경로선에 대한 추종은 고려하지 않아 충돌 회피 기동 이후, 경로선에서 벗어난 운항을 할 수 있다[3]. 이러한 문제는 경로선 추종을 위한 개선된 DWA 알고리즘(2017)에 의해 보완할 수 있다. 하지만 여전히 경로선 인근에 장애물이 많거나 장애물 사이의 좁은 영역을 통과하는 경우, 진입하지 못해 크게 우회하거나 지나친 감속으로 운항이 지연된다. 또한 DWA의 목적함수는 목적지에 대한 방향, 속도, 장애물과의 거리와 관련된 함수의 합으로 표현되며, 해당 함수에 가중치를 조절함에 따라 선박의 운동 특성이 결정되기 때문에 경로선 추종에 있어 최적의 가중치가 주변 환경에 따라 달라진다. 즉, 특정 경로선 추종 시 크게 우회하거나 좁은 영역을 가로질러야하는 경우 등 모든 상황에서 최적의 가중치를 알 수 있다면 효율적인 경로 계획이 가능할 것이다. 그러나 휴리스틱하게 결정한 가중치를 모든 상황에서 실험을 통해 도출하기 어렵다. 이러한 문제는 강화학습 기법을 활용하여 학습을 통해 시뮬레이션 상에서 가중치를 결정할 수 있다[4].

본 논문에서는 위에 언급한 문제점들을 보완하여 충돌 위험이 높은 환경에서 장애물을 회피하고 지정된 경로선 추종을 잘 수행할 수 있는 강화학습 기반 DWA 알고리즘을 제안한다. 먼저 기존의 DWA 알고리즘과 제시된 알고리즘을 설명하고 본론에서 시뮬레이션 결과를 통한 비교 검증을 수행하고자 한다.

2. Dynamic Window Approach

DWA 알고리즘은 현재 선박의 속도에서 다음 시간까지 취할 수 있는 목표 선속도와 각속도의 입력 범위를 나타내는 Dynamic Window를 생성하고 이 영역 안 선속도(v)와 각속도(w) 쌍 중에 목적함수의 값이 최대가 되는 목표 선속도와 각속도의 쌍을 도출한다.

Dynamic Window 생성 범위는 아래 Fig. 1로 표현할 수 있다[5].

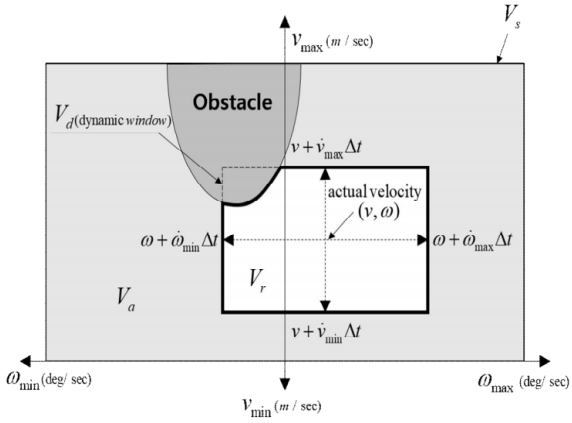

Fig. 2. Dynamic window[5]

위 Fig 1에서 V_d 영역은 다음 시간까지 취할 수 있는 선속도와 각속도의 영역을 나타낸 것으로 식 (1)로 표현되며, 선박의 가감 속도와 회전 속도 사양을 반영할 수 있다. V_s는 선박이 위치한 공간 내에 취할 수 있는 최대, 최소 영역을 나타내며, V_a는 선박의 현재 속도에서 장애물과 충돌하지 않고 제동 가능한 속도 범위이며, 최종적으로 식 (2)과 같이 3개 영역의 교집합에 해당하는 V_r 내의 제어 입력을 사용한다[5].

$$V_d = (v,w) | v \in [v - \dot{v}_{\min}\Delta t,\ v + \dot{v}_{\max}\Delta t], \quad w \in [w - \dot{w}_{\min}\Delta t,\ w + \dot{w}_{\max}\Delta t] \tag{1}$$

$$V_r = V_s \cap V_a \cap V_d \tag{2}$$

V_r 영역에서 취한 v, w은 DWA 알고리즘의 목적함수인 식 (3)에서 대입되어 목적함수 $G(v,w)$가 최대가 되는 v, w 쌍을 선박의 목표 선속도와 각속도로 결정한다.

$$G(v,w) = \alpha \times heading(v,w) + \beta \times clearance(v,w) + \gamma \times velocity(v,w) \tag{3}$$

$heading$과 $clearance$, $velocity$항은 선박의 방향, 충돌 및 속도에 관련된 함수이고 α, β, γ는 해당 함수의 가중치이다. $heading$은 제어 입력 시 선박과 목표점과의 방향 차이를 나타내며, 목표점을 향해 가려는 성질을 가진다. $clearance$는 제어 입력 이후 선박 위치에서 가장 가까운 장애물과의 거리로 우회의 특성을 가지며 $velocity$는 선박의 최대 속도 대비 제어 입력 이후의 속도로 취할 수 있는 가장 높은 속도를 선택한다[5].

2.1 선박 운동

현재 선박의 위치 x, y, θ에서 v, w로 운동 할 때 Δt 시간 후의 선박 위치는 식 (4)와 같이 정의 할 수 있다. 각속도가 0인 경우에는 현재 방향에서 Δt 동안 속도 v로 x, y 각 축으로 이동한 만큼 가산한다[6].

$$\begin{pmatrix} x' \\ y' \\ \theta' \end{pmatrix} = \begin{pmatrix} x \\ y \\ \theta \end{pmatrix} + \begin{pmatrix} -\frac{v}{w}sin\theta + \frac{v}{w}sin(\theta + w\Delta t) \\ \frac{v}{w}cos\theta - \frac{v}{w}cos(\theta + w\Delta t) \\ w\Delta t \end{pmatrix} \tag{4}$$

2.2 경로선 추종을 위한 개선된 DWA 알고리즘

개선된 DWA에서는 선박이 충돌회피 이후에도 경로선 추종을 위해 선박의 위치와 경로선과의 거리를 평가하는 함수 d_{line}와 이에 대한 가중치 δ를 추가하여 아래 식 (5)와 같이 정의된다[6].

$$G(v,w) = \alpha \times heading(v,w) + k \times \beta \times clearance(v,w) + \gamma \times velocity(v,w) + (1-k) \times \delta \times d_{line}(v,w) \tag{5}$$

여기서, k는 장애물 감지 유무에 대한 계수로, 장애물 감지 시 1, 비 감지 시 0으로 결정된다. d_{line}은 Δt 시간 후에 선박을 경로에 근접시키는 v, w 조합이 최고가 되게 하여 선박을 경로에 근접하도록 유도한다[6]. 그러나 Fig. 3에서 보듯이 단순히 센서에서 장애물 감지 시 k가 1이 되면 d_{line}이 상쇄되어, 기존의 DWA 목적함수와 동일한 구조가 되기 때문에 장애물이 많고 이동 가능한 영역이 좁은 경우, 경로선 추종이 가능함에도 불구하고 크게 우회하는 지역 경로를 생성한다. 또한 추가된 d_{line} 함수의 가중치 인자 δ에 대해서도 여전히 실험을 통해 값을 설정해야하기 때문에 지역 최적화에 빠질 수 있다.

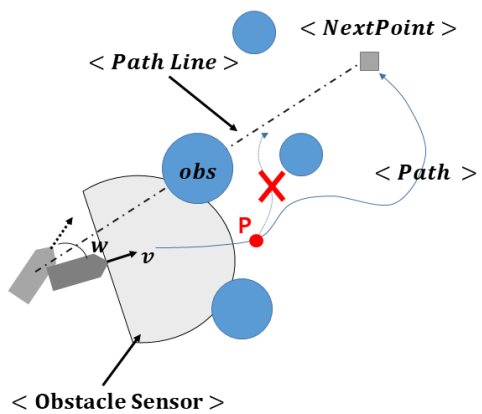

Fig. 3. Path-following of improved DWA algorithm in a complex environment

본 논문에서는 장애물과의 안전거리가 확보된 경우, 감지 여부와 관계없이 경로선 추종이 가능한 방법을 제안하고 기존의 휴리스틱하게 결정했던 가중치 인자들을 강화학습 기법을 통해 학습시켜 경로 계획 시 최적의 선속도와 각속도를 도출하고자 한다.

3. 경로선 추종을 위한 강화학습 기반 Dynamic Window Approach

개선된 DWA에서 단순히 장애물 감지 여부에 따라 경로 추종 함수가 반영되지 않는 문제점을 보완하기 위해 식 (5)의 장애물 감지 유무 계수 k를 충돌 회피 여부 계수 λ로 변경하여 식 (6)과 같이 정의하였다.

$$\lambda = (\sqrt{(x_{obs}-x')^2+(y_{obs}-y')^2} < d_{brake}+d_{safe}) \quad (6)$$

$$d_{brake} = v^2/2a \quad (7)$$

λ는 제어 입력 후에 선박의 위치 x', y'에서 가장 가까운 장애물의 거리가 제동거리 d_{brake}와 안전거리 d_{safe}의 합 보다 작으면 1, 크면 0을 반환한다. d_{brake}는 식 (7)에서 선박의 현재 선속도와 가속도 v, a를 통해 간단히 계산되며 d_{safe}는 장애물을 안전하게 회피하기 위한 여유 값을 나타낸다. 즉, 장애물이 감지되어도 λ가 1이 아닌 경우, 경로선 추종이 가능해진다.

최종적으로 위 내용을 적용한 목적함수 $G(v,w)$는 식 (8)과 같이 표현된다.

$$\begin{aligned} G(v,w) = {} & \alpha_{learn} \times heading(v,w) \\ & + \lambda \times \beta_{learn} \times clearance(v,w) \\ & + \gamma_{learn} \times velocity(v,w) \\ & + |1-\lambda| \times \delta_{learn} \times d_{line}(v,w) \end{aligned} \quad (8)$$

가중치 인자 α, β, γ, δ와 안전거리 d_{safe}는 강화학습에 적용하기 위해 식 (9)와 같이 학습 파라미터로 정의하였다. 무인수상정은 어떠한 경우에도 장애물과 충돌하지 않아야하며, 장애물이 없는 환경에서는 주어진 경로선을 추종해야하기 때문에 방향이나 속도 관련 가중치들의 비해 높은 값을 가져야 한다. 본 연구에서는 가중치 β, δ는 1로 지정하였다.

$$\begin{cases} \alpha_{learn} \leftarrow \alpha, \ 0 < \alpha \le 0.5 \\ \beta_{learn} \leftarrow \beta, \ \ \beta = 1 \\ \gamma_{learn} \leftarrow \gamma, \ \ 0 < \gamma \le 0.5 \\ \delta_{learn} \leftarrow \delta, \ \ \delta = 1 \\ d_{learn} \leftarrow d_{safe}, \ 10 \le d \le 50 \ (m) \end{cases} \quad (9)$$

본 논문에서는 Fig. 4와 같이 제안된 DWA 알고리즘강화학습 구조에 추가하여 식 (9)의 가중치 인자를 학습하였다. 강화학습이란 주어진 환경(*Environment*)에서 객체(*Agent*)가 환경의 상태(*State*)를 인식해 선택하는 행동(*Action*)에 따라 보상(*Reward*)을 최대화하는 최적의 정책을 학습하는 방법이다[7].

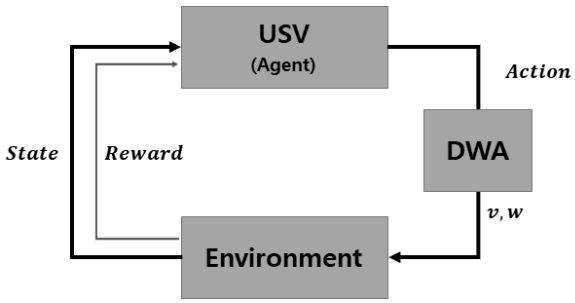

Fig. 4. Reinforcement learning structure with DWA

학습을 수행하기 전에 *State*, *Action* 그리고 *Reward*를 정의해야한다. 식 (10)에서 *State*는 상태 관측 값을 의미하며, 실시간 탐지센서로부터 장애물 위치에 따라 세 가지 상황을 정수형태로 분류하였다. *Action*은 식 (9)의 가중치 인자를 의미한다. 벡터 형태로 DWA 목적함수에 전달되어 선속도(v)와 각속도(w)를 계산하고 무인수상정의 위치를 갱신한다.

$$\begin{aligned} State &= \begin{cases} 0 : \text{장애물 없음} \\ 1 : \text{경로선 인근 장애물 탐지} \\ 2 : \text{경로선 상 장애물 탐지} \end{cases} \\ Action &= [\alpha_{learn}, \beta_{learn}, \gamma_{learn}, \delta_{learn}, d_{learn}] \end{aligned} \quad (10)$$

Table 1. Definition of rewards

(+) Reward
• 목표지점 도착 : +2 • 경로선 접근 횟수 : 총 횟수 × 0.5
(−) Reward
• 장애물 충돌 : -2 • 경로맵 이탈 : -2 • 시간 소요 : -0.01 (0.1초당)

최종적으로 환경으로부터 Table 1에 정의된 보상규칙에 따라 *Reward*를 받는다. 양의 보상은 무인수상정이 목표지점에 도착하거나 경로선 추종 시 접근 횟수에 따라 주어진다. 음의 보상의 경우, 충돌이 발생하거나 임무 영역을 이탈했을 때 주어지며, 시간에 따

른 약한 음의 보상을 부여함으로써 비효율적인 경로 계획을 방지하였다.

3.1 PPO 알고리즘

본 논문에서 강화학습 방식은 Proximal Policy Optimization(PPO) 알고리즘을 사용하였다. 2017년 OpenAI 팀에 의해 도입된 기법으로 PPO는 객체가 환경과 상호작용을 통해 데이터를 샘플링 하는 것과 확률적 기울기 상승을 사용해 대리 목표 함수를 최적화하는 것을 반복하여 학습하는 기법이다[7]. 오픈 소스 플러그인 Unity ML-Agent를 사용하였으며, ML-Agent에서 시뮬레이션 환경 및 에이전트를 설계하면 자체적으로 강화학습 수행하며 신경망 모델을 생성해준다[8]. Fig. 5와 같이 Unity3D Learning 환경에서 수집된 변수 값들을 외부 프로세스로 전송하고 PPO 알고리즘으로 학습된 결과를 다시 Unity3D로 전송해주면서 학습이 수행된다.

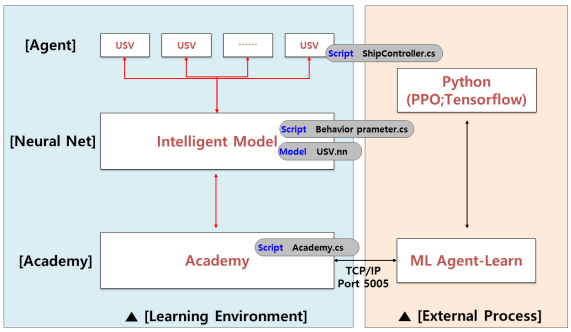

Fig. 5. Unity3D ml-agent structure

Algorithm PPO

1 For iteration=1,2,… do
2 For actor=1,2,…,N do
3 Run policy $\pi_{\theta_{old}}$ in environment for T timesteps
4 Compute advantage estimates $\hat{A}_1, \ldots, \hat{A}_T$
5 End For
6 Optimize surrogate L wrt θ, with K epochs and minibatch size $M \leq NT$
7 $\theta_{old} \leftarrow \theta$
8 End For

Fig. 6. PPO algorithm[9]

Fig. 6은 PPO 알고리즘의 간단한 의사코드를 나타낸다. 학습의 주체 actor는 시뮬레이션 수행 동안 주어진 환경에서 policy $\pi_{\theta_{old}}$에 따라 행동을 취하고 T시간 동안 수집한 결과에 추정 보상을 계산한다. 추정 보상을 토대로 policy가 갱신되며 설정된 반복 횟수(iteration) 동안 이러한 프로세스로 학습이 수행된다. 본 논문에서는 PPO에 관한 내용은 간단한 의사코드만 언급하고 참고문헌으로 대체한다[9].

3.2 학습 수행

Unity3D ML-Agent 플러그인을 사용하기 위해 동일한 환경에서 시뮬레이터를 구현하였다. Fig. 7과 같이 학습효과를 높이고 학습시간을 줄이기 위해 다양하게 환경을 구성하였다. 여기서 장애물 감지 범위는 반경 200 m, 시야 각(Field Of View) ±60°로 설정하였고 경로맵의 크기는 1000 m × 1000 m로 격자 한 칸은 100 m이다.

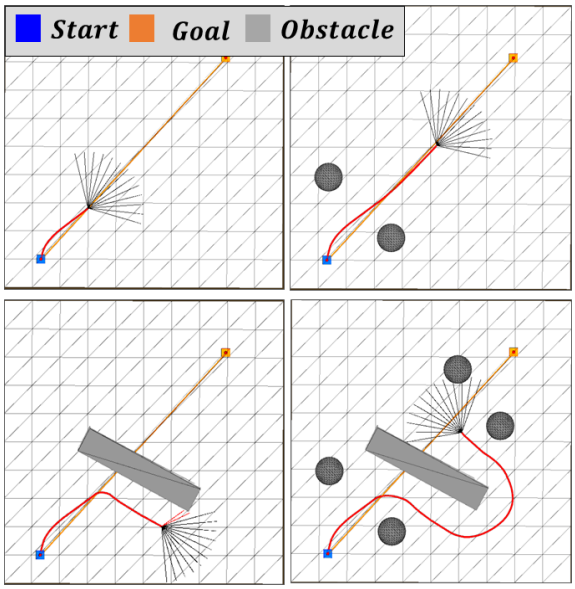

Fig. 7. Simulator for training

학습은 총 30000번, 학습률 0.003으로 수행되었으며 학습 속도 향상을 위해 식 (9)의 인자 범위를 0.1, 0.2, 0.3과 같이 이산 값으로 지정하였다. 아래 Fig 8에서 보듯이 학습이 진행됨에 따라 누적 보상의 증가와 손실 함수의 감소하는 그래프를 통해 학습이 잘 수행되었음을 확인하였다. 학습이 진행되는 동안 다양한 환경을 구성한 것은 식 (9)의 목적함수를 구성하는 가중치 계수 값이 고정되지 않고 상황에 따라 유연한 값을 가지도록 유도하였다.

Fig. 8. A graph of the results of training

4. 시뮬레이션 수행

본 논문에서 제안된 강화학습 기반 DWA 알고리즘과 기존의 DWA알고리즘을 비교 검증하기 위해 Unity3D기반 시뮬레이터를 개발하였다. 본 시뮬레이터는 경로 생성 알고리즘 검증 수행을 위한 것으로 해상환경과 선체 거동은 고려하지 않고 생성 경로를 선박이 완전히 추종한다는 가정 하에 모의를 진행하였다. 경로선은 주황색 점선, 이동 궤적은 실선으로 표시하였고 모든 고정 장애물은 회색으로 표시하였다. 여기서 기존의 알고리즘 DWA와 개선된 DWA는 편의상 Method A, Method B로 표기하였다.

비교 검증 시 알고리즘 특성을 달리하기 위해 개선된 DWA(2017)에서 사용된 가중치 비율 이외에 추가로 식 (5)의 방향 관련 가중치 α값을 다르게 설정하여 Method B – 1, 2로 아래와 같이 설정하였다.

$$Method\,B-1\,(\alpha=0.05,\beta=1.0,\gamma=0.1,\delta=1.0)$$
$$Method\,B-2\,(\alpha=0.1,\beta=1.0,\gamma=0.1,\delta=1.0)$$

먼저 시뮬레이터 구동 양상을 확인하기 위해 Method A, B 알고리즘을 구현하고 장애물 회피와 경로추종 기능을 Fig. 9와 같이 시뮬레이션 상에서 확인하였다. 기존의 방식 Method A와 달리 경로선 추종을 고려한 Method B는 장애물 회피 이후 다시 경로선으로 복귀하여 본 시뮬레이터가 잘 동작함을 확인하였다.

본 연구에서 제안한 강화학습 기반 DWA 알고리즘과 Method B – 1, 2의 비교 검증을 위해 경로선 상 장애물 Obs-1와 복귀 경로에 인접한 Obs-2를 회피하는 간단한 시나리오 A와 2개 이상 경로점 추종 시 다수의 장애물이 경로선 상에 위치한 복잡한 상황의 시나리오 B를 설정하였다.

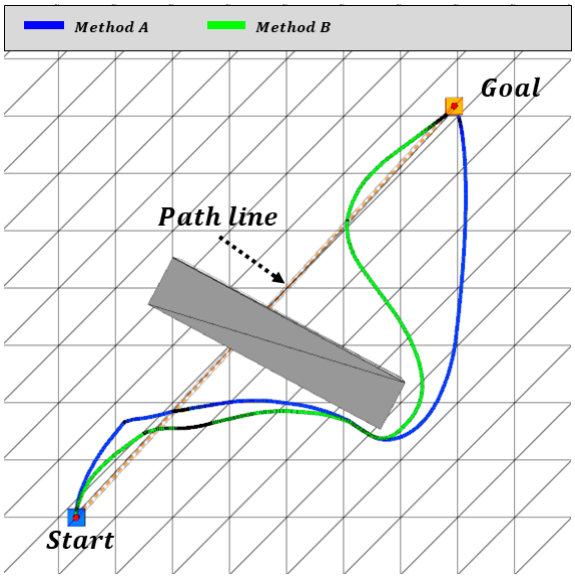

Fig. 9. Simulator for algorithm verification

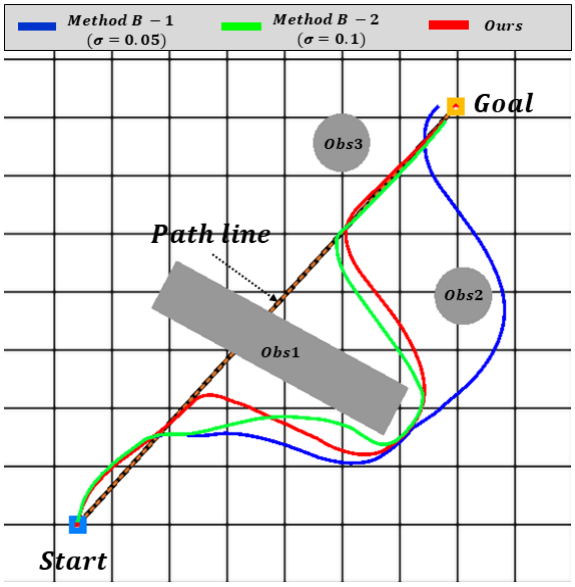

Fig. 10. Simulation results of scenario A

Fig. 10, 11은 시나리오 A, B에 대한 시뮬레이션 결과를 나타내고 Fig. 12, 13은 이동 궤적과 경로선 간의 일치 정도를 나타내는 그래프로 본 논문에서는 경로 불일치성(PI, Path-Inconsistency)이라 정의하였다. PI는 일정구간 마다 이동 궤적과 경로선의 떨어진 수직거리를 경로맵 반경(500 m)으로 나눈 값으로 떨어진

거리가 크면 클수록 1에 가까워지고 가까울수록 0에 가까운 값을 가진다. 모든 PI의 합을 경로 구간으로 나눈 평균 PI를 경로선 추종 지표로 사용하였으며, 상대적으로 낮은 값이 경로선을 잘 추종했다고 해석할 수 있다. Table 2는 시나리오 A, B에서 각 알고리즘에 대한 평균 PI와 소요시간을 나타낸다.

Table 2에서 보듯이 시나리오 A에서 제안된 알고리즘이 평균 PI 0.141, 소요시간 47.5초로 Method B – 1, 2에 비해 빠른 이동속도와 정확한 경로 추종을 나타냈다. 마찬가지로 경로점이 많은 시나리오 B에서도 Method B – 2와 유사한 경로 추종결과를 보였지만 소요시간 114초로 월등한 성능을 나타낸 것을 확인하였다.

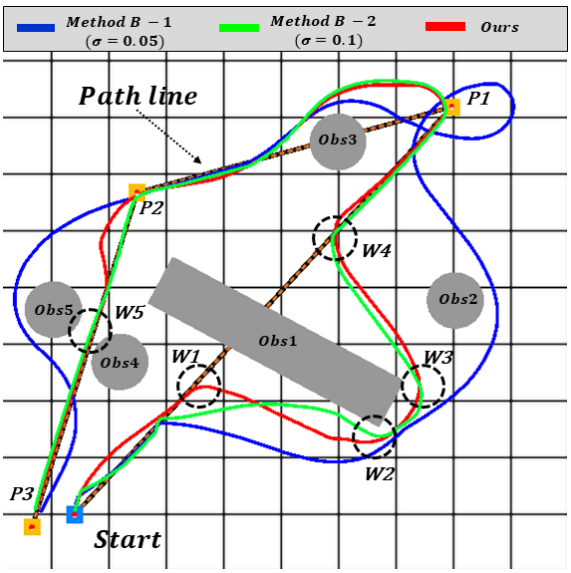

Fig. 11. Simulation results of scenario B

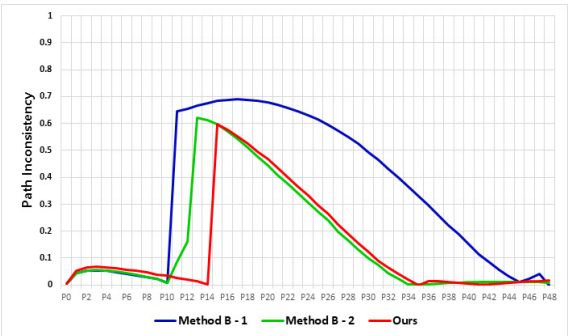

Fig. 12. Path-Inconsistency of scenario A

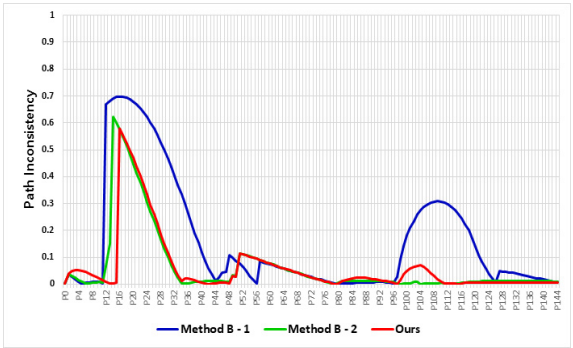

Fig. 13. Path-Inconsistency of scenario B

Table 2. Results of scenario A, B

Method	Path-Inconsistency		Arrival Time	
	A	B	A	B
Method B - 1	0.338	0.160	61.5s	130s
Method B - 2	0.159	0.063	88.5s	177s
Ours	**0.141**	**0.061**	**47.5s**	**114s**

5. 시뮬레이션 결과분석

DWA 알고리즘은 목적함수의 가중치 값에 따라 운동특성이 달라진다. Fig. 10, 11에서 Method B – 1은 속도 가중치(γ)가 방향 가중치(α)보다 높아 에서 보듯이 장애물 Obs1 회피 후 감속하여 경로선 방향으로 진입하지 못하고 빠른 속도로 장애물 Obs2를 크게 우회하는 결과를 나타냈다. 또한 경로점 P1 통과 시 부정확한 방향성을 보였다. 반면에 Method B – 2는 본 논문에서 제안된 알고리즘과 유사한 경로 추종을 보였으나 장애물이 없는 구간에도 방향과 속도에 대한 가치를 동일하게 고려하기 때문에 모든 시나리오에서 가장 느린 결과를 나타냈다.

이처럼 DWA 목적함수에서 고정된 가중치 인자를 갖는 경우, 다양한 상황에서 최적 경로를 계획하기 어렵다. 따라서 장애물 탐지 조건에 따라 최적의 가중치를 DWA 목적함수에 적용한다면 효율적인 경로계획이 가능하다. 본 논문에서는 강화학습을 통해 이러한 문제점을 해결했으며 시나리오 B에서 장애물 탐지 조건에 따라 가중치 인자 값이 변화하는 것을 Fig. 14에서 확인 할 수 있다. Fig. 11의 W1~5 위치에서 가중치

값을 보면 방향 전환 시 속도 가중치(γ)가 최솟값을 가지고 직선 경로에서는 최댓값을 가지는 것을 확인할 수 있다. 또한 속도 가중치(γ)가 작은 값을 가질 때 안전거리 d_{safe}의 값이 대체적으로 큰 값을 가지는 것으로 보아 장애물이 인접한 구간임을 유추할 수 있다.

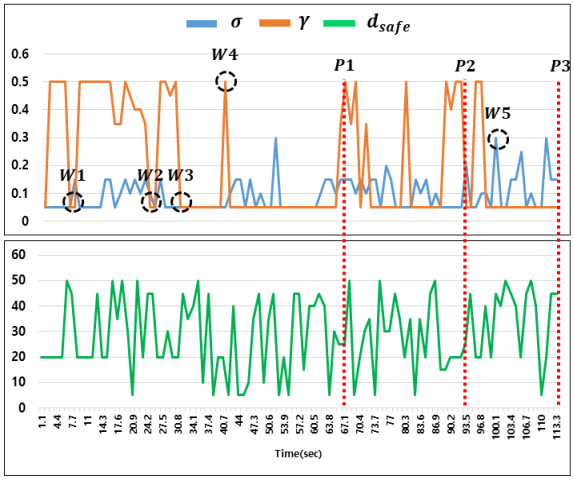

Fig. 14. Change of weight factors over time of DWA based on RL in scenario B

6. 결 론

무인수상정은 유사시 장애물 충돌 회피는 물론 특정 경로선을 따라 정찰 감시를 수행할 수 있는 지역 경로계획 기능이 필요하다. Dynamic Window Approach (DWA) 알고리즘은 이동체의 동적상태를 반영하는 지역 경로계획 기법으로 목표지점에 대한 방향, 속도, 장애물과의 거리를 토대로 최적의 속도와 각속도를 도출 할 수 있다. 하지만 DWA 알고리즘 특성상 목적함수의 가중치를 휴리스틱하게 결정해야하기 때문에 비교적 좁은 장애물 사이를 통과하거나 장애물을 큰 반경으로 우회하는 경우 등 모든 상황에 대해 적절한 최적의 가중치를 반영하기 어렵다. 따라서 본 논문에서는 강화학습을 통한 DWA 알고리즘을 제안하였고 다양한 장애물 탐지 조건에 따라 적절한 가중치 인자를 결정하여 효율적인 경로선 추종이 가능함을 확인하였다. 그러나 고정 장애물에 대한 충돌 회피만 고려했기 때문에 동적 환경에서는 적용이 어려우며 DWA 알고리즘 특성상 충돌에 대한 회피가 가장 큰 부분을 차지하기 때문에 회피 경로가 전역 최적경로인지 확인이 필요하다. 따라서 향후 이러한 문제를 해결하기 위한 추가 연구를 진행하여 본 논문에서 제안한 강화학습 기반 DWA 알고리즘의 효용성을 제고시키고자 한다.

후 기

본 논문은 민군기술 실용화연계사업 “연안경계 및 신속대응 무인경비정 실용화연계” 사업의 지원으로 작성되었습니다.

References

[1] Hyogon Kim, Sung-Jo Yun, Young-Ho Choi, Jae-Kwan Ryu, Byong-Jae Won, Jin-Ho Suh, “Improved Dynamic Window Approach with Ellipse Equations for Autonomous Navigation of Unmanned Surface Vehicle,” Journal of Institute of Control, Robotics and Systems, 26(8), pp. 624-629, 2020.

[2] Kwimi Kim, Jungmok Ma, “A Study on the Research Trends in Unmanned Surface Vehicle using Topic Modeling,” Korea Academy Industrial Cooperation Society, 21(7), pp. 597-606, 2020.

[3] Jong-Gyu Ham, Joong-Tae Park, Jae-Bok Song, “Mobile Robot Navigation based on Global DWA with Optimal Waypoints,” Journal of Institute of Control, Robotics and Systems, 13(7), pp. 624-630, 2007.

[4] Kim Jee-Seon, “Local Collision Avoidance Algorithm in Dynamic Environment using Collision Probability,” The Graduate School of Ewha Womans University, 2020.

[5] D. Fox, W. Burgard, S. Thrun, “The Dynamic Window Approach to Collision Avoidance,” IEEE Robotics & Automation Magazine, Vol. 4, No. 1, pp. 23-33, 1997.

[6] Hyogon Kim, Sung-Jo Yun, Young-Ho Choi, Jung-Woo Lee, Jae-KWan Ryu, Byong-Jae Won, Jin-Ho Suh, “Improved Dynamic Window Approach With Path-Following for Unmanned Surface Vehicle,”

Journal of Embedded Systems and Applications, 12(5), pp. 295-301, 2017.

[7] Suyeong Jang et. al., “Research Trends on Deep Reinforcement Learning,” Electronics and Telecommunications Trends, Vol. 34 No. 4, pp. 1-14, 2019.

[8] Dong-Ham Kim, Sung-Uk Lee, Jong-Ho Nam, Yoshitaka Furukawa, “Determination of Ship Collision Avoidance Path using Deep Deterministic Policy Gradient Algorithm,” Journal of the Society of Naval Architects of Korea, 56(1), pp. 58-65, 2019.

[9] Yi-Hong Liang, Sin-Jin Kang, Sung Hyun Cho, “A Study about the Usefulness of Reinforcement Learning in Business Simulation Games using PPO Algorithm,” Journal of Korea Game Society, 19(6), pp. 61-70, 2019.