creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

**저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.**

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

공학박사 학위논문

# 심층강화학습을 이용한 무인수상선의 충돌회피

Collision Avoidance for an Unmanned Surface Vehicle Using Deep Reinforcement Learning

2018 년 2 월

서울대학교 대학원

조선해양공학과

우 주 현

# 심층강화학습을 이용한 무인수상선의 충돌회피

Collision Avoidance for an Unmanned Surface Vehicle Using Deep Reinforcement Learning

지도교수 김 용 환

이 논문을 공학박사 학위논문으로 제출함

2018년 2월


서울대학교 대학원

조선해양공학과

우 주 현


우주현의 공학박사 학위논문을 인준함

2018년 2월

| | | |
|---|---|---|
| 위 원 장 | 김 태 완 | (인) |
| 부위원장 | 김 용 환 | (인) |
| 위 원 | 노 명 일 | (인) |
| 위 원 | 최 현 택 | (인) |
| 위 원 | 안 진 형 | (인) |

# 초록

무인수상선이 할당된 임무를 성공적으로 수행하기위해서는 선박 및 장애물에 대한 충돌회피 성능이 필수적으로 확보 되어야한다. 이는 무인수상선 자선의 안전성을 확보하기 위함과 동시에 무인수상선 근처에서 운항하는 일반 선박의 인명 및 재산상의 손실을 방지하기 위함이다. 이와 같은 사회적 요구에 따라 무인수상선의 충돌회피 능력을 향상시키기 위한 연구가 활발히 이루어져 왔다.

본 논문에서는 기존에 수행되었던 선박의 충돌회피와 관련된 선행연구들의 한계점을 분석하고 이들을 극복하는 방향으로 연구목표를 설정하였다. 본 연구에서 설정한 네 가지의 목표는 다음과 같다. 첫째, 충돌위험도를 추론할 때, 자선과 장애물 간의 조우상황을 단순화 하지 않고, 직접 고려할 수 있어야한다. 둘째, 다수의 선박에 대한 충돌위험성을 인식할 수 있어야 한다. 셋째, 자선의 조종성능이나 환경하중의 영향을 고려할 수 있어야 한다. 넷째, 조우 상황별로 상이한 기준을 이용하여 충돌회피를 수행할 수 있어야 한다.

무인수상선의 충돌회피는 일반적으로 탐지- 의사결정 - 경로계획 - 제어의 단계를 거친다. 각종 장애물 탐지센서를 이용해 얻은 장애물의 위치 및 운동정보를 바탕으로 현재 조우상황을 판단하고, 이를 통해 회피여부에 대한 의사결정을 진행한다. 장애물회피 여부가 결정되면 충돌을 방지하기 위한 경로계획을 수립하고, 수립된 명령에 따라 선박을 제어하여 충돌을 회피한다. 본 연구에서는 이들 단계 중 충돌회피 여부 및 시점을 판단하는 의사결정 단계에 대한 연구를 심층강화학습이라는 기법을 이용하여 진행하였다. 강화학습은 학습자가 스스로 주변환경과

상호작용을 통해 최적의 행동을 찾아 나가는 기계학습의 한 종류이다. 본 연구에서는 학습자를 무인수상선으로 선정하고, 반복적인 충돌회피 시뮬레이션을 통해서 무인수상선 스스로 최적의 충돌회피 시점을 찾아나가는 접근방법을 사용하였다. 시뮬레이션을 통해 강화학습을 수행하기 위해서는 무인수상선의 동역학 정보가 필요한데, 이를 위해 본 연구에서는 무인수상선 실선에 대한 시스템 식별을 수행하였다. 본 연구에서 제안하는 심층강화학습 기반 충돌회피 알고리즘은 국제 해상충돌 예방규칙을 준수하도록 구성하였고, 각 조우상황별로 상이한 요구조건을 만족할 수 있도록 구성하였다. 또한, 무인수상선이 반복적으로 충돌회피 시뮬레이션을 통해 주변환경과 행동에 대한 반응을 학습하므로, 경험적으로 자선의 조종성능과 환경하중에 대한 영향을 반영 할 수 있다.

구성된 충돌회피 알고리즘을 검증하기 위하여 다양한 시나리오에 대한 충돌회피 시뮬레이션을 수행하였다. 이로부터 얻어진 자선과 타선의 궤적, 침로각 정보, 상대거리 정보, 충돌 회피 시점 등을 통해 본 연구에서 제안하는 기법의 효과성을 확인하였다. 학습된 충돌회피 알고리즘은 무인수상선 실선에 탑재되어 실해역 충돌회피 실험을 통해 실제 환경에서의 충돌회피 능력을 확인하였다.

**주요어**: 무인수상선, 충돌 회피, 강화 학습, 인공 지능

**학번**: 2012-23312

# 목차

**초록** **i**

**1 서론** **1**

1.1 연구 배경 . . . . . 1

1.2 문헌 조사 및 기존 연구사례 . . . . . 4

1.2.1 장애물 탐지 . . . . . 4

1.2.2 충돌회피 시점 추론 . . . . . 5

1.2.3 충돌회피 경로 계획 . . . . . 6

1.3 연구 목표 . . . . . 6

1.4 연구 내용 및 한계 . . . . . 8

**2 심층강화학습** **11**

2.1 강화학습 . . . . . 11

2.1.1 강화학습 문제 . . . . . 11

2.1.2 마코프 의사결정 과정 . . . . . 13

2.1.3 강화학습 알고리즘 . . . . . 16

2.2 심층강화학습 . . . . . 22

2.2.1 Deep Q Network . . . . . 22

2.2.2 심층강화학습 기법 . . . . . 24

2.2.3 심층강화학습 기반 연구 . . . . . 24

**3 무인수상선의 충돌회피** **26**

3.1 선박의 충돌 회피 . . . . . 26

3.1.1 국제 해상충돌 예방규칙 . . . . . 26

3.1.2 조우상황의 분류 . . . . . 28

3.1.3 회피 행동의 결정 . . . . . 32

3.2 충돌회피 알고리즘 . . . . . 35

3.2.1 충돌회피 시점 추론 알고리즘 . . . . . 35

3.2.2 속도 장애물 기법 . . . . . 41

**4 심층강화학습 기반 충돌 회피** **45**

4.1 심층강화학습 신경망 . . . . . 47

4.1.1 SMDP의 구성 . . . . . 47

4.1.2 심층강화학습 알고리즘의 동작과정 . . . . . 59

4.1.3 심층강화학습 신경망 학습 . . . . . 60

**5 성능 검증** **66**

5.1 대상 무인수상선 . . . . . 66

5.1.1 Wave Adaptive Modular Vessel 16' . . . . . 66

5.1.2 무인수상선의 동역학 시뮬레이터 . . . . . 69

5.2 시뮬레이션 결과 . . . . . 87

5.2.1 충돌회피 알고리즘 분석 . . . . . 87

5.2.2 충돌회피 성능에 대한 분석 . . . . . 94

5.2.3 조종성능의 영향에 대한 분석 . . . . . 109

5.2.4 다수 선박에 대한 충돌회피 . . . . . 116

5.3 자유 항주 실험 결과 . . . . . 126
5.3.1 실험 환경 . . . . . 126
5.3.2 충돌 회피 실험 결과 . . . . . 127

**6 결론** **142**

**A 무인수상정의 시스템 식별** **145**
A.1 무인수상선 동역학의 시스템 식별 . . . . . 145

**Abstract** **163**

# 표 목차

1.1 Statistical data of marine accident in Korea according to types of accident in recent five years . . . . . 3

3.1 Abbreviation for encounter type . . . . . 30

5.1 Principal specification of Seoul National University WAM-V ASV 67

5.2 Principal properties for wind load calculation . . . . . 75

5.3 Beaufort number and corresponding wind speed. Table from [49] 78

5.4 Avoidance behavior begin time according to encounter situation and collision risk assessment method . . . . . 98

5.5 Avoidance behavior begin time according to encounter situation and collision risk assessment method . . . . . 102

5.6 Avoidance behavior begin time with respect to wind direction and collision risk assessment method . . . . . 116

# 그림 목차


1.1 Different types of Unmanned surface vehicle . . . . . 2

1.2 Stage of the autonomous surface vehicle collision avoidance problem 8

2.1 Three main types of machine learning . . . . . 12

2.2 Interaction between learning agent and environment in reinforcement learning problem . . . . . 12

2.3 Iterative iteraction between agent and environment . . . . . 12

2.4 Diagram of tabular Q learning algorithm. Figure from [5] . . . . 20

2.5 Structure of the Deep Q Network. Figure from [28] . . . . . 23

2.6 Deep Reinforcement learning application to locomotion problem. Figure from [30] . . . . . 25

3.1 Categorization of obstacle by relative position. Figure from [40] . 29

3.2 Categorization of obstacle by relative heading angle. Figure from [40] . . . . . 30

3.3 Chart used to determine the encounter type. Figure from [40] . . 31

3.4 Desired avoidance behavior of ASV in case of Head on, according to COLREGs . . . . . 32

3.5 Desired avoidance behavior of ASV in case of Crossing, according to COLREGs . . . . . 34

3.6 Desired avoidance behavior of ASV in case of Overtaking according to COLREGs . . . . . 34
3.7 Ship domain for collision risk assessment, suggested by Fujii, Coldwell and Goodwin . . . . . 36
3.8 Conceptual description of TCPA and DCPA, Figure from [48] . . 37
3.9 Characteristics of DCPA and TCPA . . . . . 38
3.10 Collision risk estimation result using Lee and Rhee. Figure from [49] . . . . . 40
3.11 Collision risk estimation result using Ahn et al. Figure from [49] 40
3.12 Constructed velocity obstacle using (a)Minkowski sum and (b) under assumption of disk shape. Figure from [37] . . . . . 42
3.13 Velocity obstacle region and selectable velocity space set. Figure from [37] . . . . . 43

4.1 Schematic diagram of the deep reinforcement learning based collision risk assessment system . . . . . 46
4.2 Concept of semi-markov decision process . . . . . 47
4.3 Operation display screen of Furuno 3 Lucero radar plotting system 49
4.4 Encounter situation map representation . . . . . 51
4.5 Environment load estimation method for deep reinforcement learning input . . . . . 52
4.6 Conceptual description of the vector field guidance method for linear path follow guidance . . . . . 54

4.7 Collision avoidance guidance method using Velocity Obstacle method, with respect to avoidance direction . . . . . 55
4.8 Structure of te deep reinforcement learning network by Wang et al (2016), Image from [43] . . . . . 61
4.9 Proposed structure of te deep reinforcement learning network . . 61
4.10 Training result of the deep reinforcement learning network . . . . 63

5.1 Seoul National University WAM-V autonomous surface vehicle during Maritime RobotX Chellenge 2016 (ASV competition), Hawaii, U.S. . . . . 68
5.2 Schematic description of the ASV coordinate system . . . . . 70
5.3 Coordinate system for environmental disturbance . . . . . 74
5.4 Projected Area of the SNU ASV . . . . . 76
5.5 Turning test trajectory under 0 deg wind load(Beaufort No.2) . . 79
5.6 Turning test trajectory under 90 deg wind load(Beaufort No.2) . 79
5.7 Turning test trajectory under 270 deg wind load(Beaufort No.2) 80
5.8 Turning test trajectory under 0 deg wind load(Beaufort No.5) . . 80
5.9 Turning test trajectory under 90 deg wind load(Beaufort No.5) . 81
5.10 Turning test trajectory under 270 deg wind load(Beaufort No.5) 81
5.11 Turning test trajectory under 0 deg 0.1m/s current load . . . . . 84
5.12 Turning test trajectory under 90 deg 0.1m/s current load . . . . 84
5.13 Turning test trajectory under 270 deg 0.1m/s current load . . . . 85
5.14 Turning test trajectory under 0 deg 0.3m/s current load . . . . . 85
5.15 Turning test trajectory under 90 deg 0.3m/s current load . . . . 86

5.16 Turning test trajectory under 270 deg 0.3m/s current load . . . . 86
5.17 Estimated collision risk value with respect to target ship's position 88
5.18 Estimated collision Risk of the ASV, based on proposed method 89
5.19 Saliency map of the grid map with respect to obstacle existance . 91
5.20 Trajectory of collision avoidance simulation using proposed method (Head on) . . . . . 93
5.21 Advantage value of the DRL network during the collision avoidance simulation . . . . . 93
5.22 Course angle and desired course angle during the collision avoidance simulation . . . . . 94
5.23 Collision avoidance simulation result using different collisin risk assessment method (Head on) . . . . . 96
5.24 Collision avoidance simulation result using different collisin risk assessment method (Crossing, give way) . . . . . 97
5.25 Schematic description of the Monte-Carlo simulation environment 99
5.26 State flow of the target ship decision making process during the Monte-Carlo simulation . . . . . 100
5.27 Histogram of the Monte-Carlo simulation result for Head on encounter situation . . . . . 104
5.28 Histogram of the Monte-Carlo simulation result for Crossing (give way) encounter situation . . . . . 105
5.29 Histogram of the Monte-Carlo simulation result for Head on and Crossing encounter situation . . . . . 106

5.30 Probability of the collision avoidance with respect to collision risk assessment methods . . . . . 108
5.31 Collision avoidance simulation result with respect to ship with different maneuverability (Head on) . . . . . 110
5.32 Collision avoidance simulation result with respect to ship with different maneuverability (Crissong, give way) . . . . . 111
5.33 Collision avoidance simulation result using different collisin risk assessment method under 6.7m/s wind from West . . . . . 113
5.34 Collision avoidance simulation result using different collisin risk assessment method under 6.7m/s wind from East . . . . . 114
5.35 Collision avoidance simulation result using different collisin risk assessment method under 6.7m/s wind from North . . . . . 115
5.36 Example of complex multiple ship encounter situation . . . . . 117
5.37 Trajectory of multiple ship collision avoidance simulation using DRL . . . . . 119
5.38 Snapshot of multiple ship collision avoidance simulation using DRL 120
5.39 Trajectory of multiple ship collision avoidance simulation using CPA . . . . . 121
5.40 Snapshot of multiple ship collision avoidance simulation using CPA 122
5.41 Example of complex multiple ship encounter situation (II) . . . . 123
5.42 Trajectory of multiple ship collision avoidance simulation case (II) using DRL . . . . . 124

5.43 Snapshot of multiple ship collision avoidance simulation case (II) using DRL . . . . . 125
5.44 Test environment in Lake Pyoungtaek) . . . . . 126
5.45 Collision avoidance free running test and simulation data (Head on) . . . . . 129
5.46 Snap shot of collision avoidance test and simulation data (Head on) . . . . . 130
5.47 Snapshot photo of collision avoidance free running test (Head on) 131
5.48 Trajectory of collision avoidance free running test (Overtaking) . 133
5.49 Snapshots of collision avoidance free running test (Overtaking) . 134
5.50 Snapshot photo of collision avoidance free running test (Overtaking) . . . . . 135
5.51 Collision scenario of multiple ship avoidance . . . . . 136
5.52 Trajectory of multiple ship collision avoidance free running test . 138
5.53 Snapshot of multiple ship collision avoidance free running test (1st target ship) . . . . . 139
5.54 Snapshot of multiple ship collision avoidance free running test (2nd target ship) . . . . . 140
5.55 Snapshot of multiple ship collision avoidance free running test (3rd target ship) . . . . . 141

A.1 Relation between electric motor throttle and steady state RPM response . . . . . 145
A.2 Simplified block diagram of the RPM model . . . . . 146

A.3 Step response test result of an electric thruster motor of the ASV 146

A.4 Bollard-pull test of the ASV . . . . . 148

A.5 Static input test result of the SNU ASV . . . . . 151

A.6 Acceleration-decceleration test of the SNU ASV . . . . . 152

A.7 Zigzag test of the SNU ASV . . . . . 153

A.8 Acceleration-decceleration test of the SNU ASV . . . . . 155

# 제 1 장 서론

## 1.1 연구 배경

최근 들어 무인시스템에 대한 사회적 요구가 증대되고, 관련기술의 성능 및 신뢰도가 향상됨에 따라, 기존에 인간이 운용하던 시스템을 로봇이 대체하는 무인화가 사회 곳곳에서 전반적으로 이루어지고 있다. 이러한 경향은 선박 분야에서도 확인할 수 있는데, 선박의 무인화를 통해 인간이 수행하기 힘든 임무를 무인수상선이 대체하여 수행하거나, 운용비용을 절감하는 노력이 이루어지고 있다. Fig.1.1은 최근 개발되거나, 현재 개발중인 다양한 종류의 무인수상선을 나타낸다. Fig.1.1(a)는 미국방성 국방고등연구사업국(Defense Advanced Research Projects Agency, DARPA)에서 장시간동안 적군의 잠수함을 추적하기 위해 개발중인 무인수상선이다. Fig.1.1(b)는 이스라엘 군수업체에서 페르시아만 연안의 경비를 목적으로 개발한 무인수상선 Protector의 사진이다. Fig.1.1(c)는 무인수상선 C-Enduro를 나타내고 있는데, 장기간동안 해상 데이터 수집을 목적으로 만들어졌다. 선박에 장착된 풍력발전기와 태양광패널을 이용하여 자체적으로 운항에 필요한 전기를 생산할 수 있다는 특징이 있다. 이 외에도 해상운송을 위해 DNV GL이 개발중인 Revolt(Fig.1.1(d)) 등 기존의 선박을 무인화하려는 노력이 민간, 군, 산업 등 다양한 분야에서 이루어지고있는 실정이다.

앞서 나열했던 여러 종류의 무인수상선들은 각기 다른 목적을 가지고 다른 임무를 수행하지만, 해당 임무를 성공적으로 수행하기 위해서는 장애물을 스스로 인지하고 회피하는 충돌 회피 능력을 보유하고 있어야 한다는 공통점을 가진다.

(a) Sea Hunter (DARPA)

(b) Protector (Rafael)

(c) C-enduro (ASV Global)

(d) Revolt(DNV GL)

Fig 1.1: Different types of Unmanned surface vehicle

선박의 충돌사고는 선박의 구조적인 충격뿐 아니라 재산상의 손실, 기름유출로 인한 환경오염, 심지어 인명피해까지 야기시킬 수 있기 때문에 무인수상선은 반드시 인간 수준, 혹은 그 이상의 충돌회피 성능을 확보해야 한다. Table.1.1은 국내 해양수산부 산하기관인 중앙해양안전심판원에서 집계한 최근 5년간 국내에서 발생한 선박사고의 유형을 정리한 표이다. 충돌사고는 전체 선박사고의 13.5%를 차지하는데, 이는 매 년 약 200건 이상의 충돌사고를 의미한다. Campbell et al (2012)[4]의 연구에 따르면, 선박의 충돌사고중 대부분은 경계소홀, 충돌예방규칙 미준수 등의 인적과오(human error)에 의해 발생하는 것으로 조사되었다. 무인수상선의 경우 이러한 인적과오의 영향을 배제시킬 수 있으므로, 효과적이고 신뢰성 높은

| 년도 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|
| 충돌 | 260 (14.3%) | 196 (12.5%) | 175 (16.0%) | 180 (13.5%) | 235 (11.2%) |
| 접촉 | 32 | 33 | 23 | 19 | 28 |
| 좌초 | 120 | 113 | 91 | 96 | 84 |
| 전복 | 58 | 39 | 32 | 35 | 32 |
| 화재/폭발 | 84 | 105 | 79 | 97 | 100 |
| 침몰 | 65 | 41 | 21 | 19 | 31 |
| 기관손상 | 652 | 486 | 290 | 339 | 703 |
| 인명사상 | 84 | 60 | 45 | 113 | 144 |
| 안전운항저해 | 234 | 227 | 191 | 205 | 331 |
| 기타 | 220 | 270 | 146 | 227 | 413 |
| 계 | 1809 | 1573 | 1093 | 1330 | 2101 |

Table 1.1: Statistical data of marine accident in Korea according to types of accident in recent five years

무인수상선의 충돌회피 알고리즘의 개발을 통해 해상환경에서의 충돌사고 확률을 현저히 감소시킬 수 있을것으로 기대된다.

## 1.2 문헌 조사 및 기존 연구사례

무인수상선의 충돌회피와 관련된 연구주제는 장애물 탐지에 대한 연구, 탐지된 장애물 정보를 이용해 충돌회피의 시점을 추론하는 연구, 장애물의 회피 경로를 계획하는 연구로 구분지을 수 있다. 각 주제에 해당하는 대표적인 선행연구사례는 다음과 같다.

### 1.2.1 장애물 탐지

무인수상선이 장애물을 탐지하는데 사용하는 센서는 주로 레이더(Radar), 라이더(Lidar), 영상센서, 소나 등이 있다. Almeida et al (2009)[2]는 무인수상선에 장착된 레이더를 이용하여 장애물을 탐지하고, CPA(the closest point of approach, 최근접점)정보를 이용하여 장애물의 위험도를 분류하는 연구를 수행한 바 있다. Schuster et al (2014)[34]는 AIS를 대체하기위해 저가형 레이더를 기반으로 무인수상선의 장애물 탐지에 대한 연구를 수행 한 바 있다. 레이더 센서는 원거리에 위치한 장애물에 대한 탐지성능은 높지만, 상대적으로 가까운 장애물이나 침로변경 여부를 파악하기 어렵다는 점, 파도의 탐지 및 수면의 반사도 문제 등의 한계점을 제시한 바 있다.

Wang et al (2011)[19]은 스테레오 및 단안카메라를 이용한 무인수상선의 영상기반 장애물 탐지에 대한 연구를 수행한 바 있다. 영상내 존재하는 관심영역(region of intererst)을 결정하기위해 RANSAC 알고리즘을 이용하여 수평면을 추출하고, 특징점을 탐지하여 추적하는 방식으로 장애물을 탐지하였다. 무인수상선의 영상기반 장애물탐지는 근거리(최대 100m)에 위치한 장애물을 대상으로 한다.

Lebbad and Nataraj[26]는 라이더 센서를 이용하여 장애물을 탐지하고, 이를 이용해 영상내 장애물의 영역을 추출하는 연구를 수행한 바 있다. 이들 연구에 의

하면 환경하중 등의 영향에 의해 발생하는 무인수상선의 자세변화에 따라 라이더 이미지가 왜곡되는 현상을 확인하였다.

### 1.2.2 충돌회피 시점 추론

충돌회피 시점 추론에 대한 연구는 초기에 운항자의 충돌회피 의사결정을 돕기 위한 형태로 선박의 충돌회피 분야에서 1970년대부터 활발하게 이루어져 왔다. 회피 시점의 추론은 크게 두가지 방식으로 구분되어질 수 있는데, 이는 선박안전영역을 이용한 방법과 CPA를 이용한 방법으로 구분된다. 선박안전영역은 Fujii and Tanaka (1971)[16]가 제안한 방법으로, 자선을 기준으로 가상의 안전영역을 설정하고, 해당영역에 장애물이 침범하면 충돌회피를 수행하는 개념이다. 이후 Goodwin (1975)[17], Coldwell (1983)[9], Davis (1980)[12]에 의해 선박 안전영역에 국제해상충돌예방규칙을 고려하려는 시도가 있었다. 한편, 선박안전영역에 상대속도를 고려하기 위한 접근 (Colley et al (1983)[10], 퍼지(Fuzzy) 추정자 또는 신경회로망을 이용하여 기존 연구의 기법적 한계를 극복하려는 연구가 수행된 바 있다.([47],[39],[22])

CPA를 이용한 충돌회피 시점 추론 기법은 Iwasaki and Hara (1986)가 최초로 제안한 방법으로, DCPA(Distance at closest point of approach, 최단접근거리)와 TCPA(Time to closest point of approach, 최단접근시간)의 조합으로 충돌위험도를 추론하는 방법이다. Hasegawa and Kouzuki (1987)[18]에 의해 운항자의 경험을 반영하려는 시도가 있었고, Ahn et al (2012)[1]에 의해 조종성능 및 환경인자 등을 고려한 충돌회피 시점 추론이 연구되어진 바 있다.

### 1.2.3 충돌회피 경로 계획

장애물 회피를 수행하기 위한 경로계획 기법은 크게 두 종류로 구분되는데, 이는 장애물 지도를 기반으로 충돌을 회피하는 경로를 생성하는 전역 경로 계획 (global path planning) 기법과, 실시간 센서 정보를 이용하여 국부적으로 경로를 재생성하는 지역 경로 계획 기법(local path planning)이 존재한다. 무인수상선의 충돌회피를 위해 이루어진 전역경로계획 기법으로는, Larson et al (2007)[24]이 제안한 격자 지도에 대한 A*알고리즘을 이용한 경로생성기법, Casalino et al (2009)[6]가 제안한 가시성 그래프(visibility graph, VG)을 이용한 기법이 대표적이다. Xie (2014)[46]는 모바일 로봇의 충돌회피에 자주 사용되는 가상 포텐셜 필드 (artificial potential field)기법을 이용하여 전역경로를 수행한 바 있다.

무인수상선의 지역경로계획 기법에 대한 대표적인 연구는 Kuwata et al (2014)[23]가 수행한 velocity obstacle을 이용한 충돌회피 기법으로, 추후 Myre (2016)[29]와 Stenersen (2015)[37]의 무인수상선 충돌회피 연구에 영향을 끼쳤다. 이외에도 무인수상선이 이동방향에 대한 후보 경로를 생성하고, 각각의 경로에 대한 평가를 수행하여 최적의 경로를 찾는 Morphin 알고리즘 [25], 경계박스 (bounding box) 개념을 이용해 움직이는 장애물에 대한 접근을 방지하는 연구[3](Blaich et al, 2015) 등이 이루어 진 바 있다.

## 1.3 연구 목표

본 연구에서는 무인수상선의 충돌회피와 관련된 선행연구들의 연구결과를 토대로 다음의 네가지 연구목표를 수립하였다.

첫째, 자선과 장애물간의 충돌위험도 추론을 위해 다양하고 복잡한 인자들을 동시에 고려할수 있는 기법을 제안하는 것이다. 대표적인 충돌시점 결정방법인 충

돌위험도(Hawegawa and Kouzuki, 1987)나 선박안전영역(Davis et al, 1980) 등의 기법은 자선과 타선의 조우상황을 각각 DCPA 와 TCPA ,상대거리 와 상대속도 등의 인자들로 지나치게 단순화 하는 문제가 있었다. 선박의 충돌회피는 이러한 요소 뿐만 아니라 자선과 타선의 속도비율, 조우상황, 해상운항법 등의 요소를 동시에 고려해야하는 복잡한 문제이므로, 이러한 요소를 반영할 수 있는 기법을 제안하고자 한다.

둘째, 다수의 선박이 존재하는 복잡한 조우상황에 대한 위험성을 인식할 수 있는 기법을 제안하는 것이다. 대부분의 충돌회피 알고리즘의 경우, 다수 선박에 대한 조우상황 평가를 위해, 다수 선박에 대한 조우상황을 1:1 조우상황을 반복적으로 고려하는 방식으로 단순화 하여 해결하고자 하였다. 본 연구에서는 이러한 기존연구의 한계성에서 벗어나서, 다수 선박이 존재하는 조우상황 자체에 대한 위험성을 평가할 수 있는 기법을 제안하고자 한다.

셋째, 자선의 조종성능이나 환경하중의 영향을 경험적으로 고려할 수 있는 기법을 제안하고자 한다. 자선의 조종성능이나 환경하중은 선박의 침로제어에 직접적인 영향을 끼치는 요소로, 이는 선박의 충돌 회피 시점 결정에 있어서 매우 중요하게 고려되어야 할 요소이다.

넷째, 타선과의 조우상황에 따라 상이한 기준을 이용하여 충돌회피를 수행할 수 있어야한다. 국제 해상충돌 예방규칙에 의하면, 자선과 타선의 피항의무 여부는 두 선박간의 조우상황에 의해 결정된다. 때문에, 각각의 조우상황에 따라 이에 알맞는 회피행위를 수행할 수 있어야 한다.

## 1.4 연구 내용 및 한계

무인수상선의 충돌회피는 일반적으로 Fig.1.2와 같은 4개의 단계를 거치며 이루어진다. 탐지(sensing)단계에서는 AIS, Radar, Lidar, Vision 센서 등과 같은 장애물 탐지센서를 이용하여 장애물의 위치를 탐지하고 운동정보를 추정한다. 장애물의 정보가 파악되면, 의사결정(decision making)단계에서 현재 조우상황의 위험도를 파악하여 충돌회피 여부를 결정한다. 충돌회피 여부가 결정되면, 경로계획(path planning)단계에서 회피를 수행하기위한 명령이 생성되고, 무인수상선이 명령을 따르도록 제어(control)함으로써 충돌회피가 이루어진다.

Fig 1.2: Stage of the autonomous surface vehicle collision avoidance problem

본 연구에서는 위의 4단계 중 의사결정 모듈에 해당하는 부분을 심층강화학습 기법을 통해 개발하였다. 강화학습이란 학습을 수행하는 학습자가 스스로 주변 환경과 상호작용(interaction)하며 최적의 행동을 찾는 기계학습의 한 종류이다. 심층강화학습을 이용하여 충돌회피 알고리즘을 개발하기 위하여 무인수상선의 동역학을 반영하는 시뮬레이터를 개발하고, 이를 이용해 반복적인 충돌회피 시뮬레이션을 수행, 최적의 충돌회피 시점을 찾도록 하였다. 본 연구에서는 실선 무인수상선의 동역학을 시뮬레이터에 반영하기 위해 무인수상선 동역학에 대한 시스템 식별 실험을 수행하였고, 이를 통해 얻은 조종운동모델을 적용하여 강화학습을 진행하였다. 강화학습을 통해 얻은 충돌회피 알고리즘은 시뮬레이션과 실선실험을 통해 검증하였다. 충돌회피 알고리즘에 대한 검증을 위하여 단일선박에 대한 충

돌시나리오 뿐만 아니라 다수의 선박이 복잡하게 운항하는 상황에 대한 충돌회피 시나리오를 통해 제안하는 기법의 성능을 검증하였다.

본 연구는 다음과 같은 3가지의 한계점을 갖는다. 우선 첫째, 무인수상선과 환경하중에 대한 모델링오차가 존재한다. 무인수상선의 동역학을 반영하기 위해 시스템식별 실험을 수행하였지만, 선형시스템에 대한 시스템식별의 기법적 한계로 인하여 비선형 모델링오차가 존재한다. 또한, 환경하중의 모델링에 대한 오차 또한 존재한다. 특히, 본 연구에서는 환경하중의 영향을 모두 정상(steady)상태로 가정하였는데, 실제 충돌회피상황에서는 이러한 가정이 만족되지 않는다. 둘째, 저속 선박에 대한 검증만이 이루어졌다. 본 연구에서 활용한 무인수상선은 최대 속도가 5 knots이고 길이가 5 m인 저속의 소형 무인수상선이다. 실선을 통한 검증을 수행하는 것을 목표로 하였기때문에 불가피하게 저속의 무인수상선에 대한 검증만이 이루어졌다. 셋째, 강화학습이 오직 시뮬레이션을 통해서만 이루어진다. 본 연구에서는 실제 무인수상선에 강화학습 기반 충돌회피 알고리즘이 탑재되어 충돌회피 실험이 이루어지지만, 이는 기존에 학습된 알고리즘을 사용하는 것일 뿐, 실시간으로 학습을 진행하는 것은 아니다. 만일 실선에 대한 학습이 이루어진다면, 앞서 설명했던 모델링오차에 대한 불확실성을 감소시킬 수 있는 장점이 존재한다. 하지만 본 연구에서는 문제의 단순화와 시간 및 비용의 효율성을 위하여 시뮬레이션 환경에서만 강화학습을 진행하였다.

본 논문은 다음과 같은 구성을 갖는다. 2장에서는 심층강화학습 기법에 대한 이론적 배경을 설명하였다. 심층강화학습의 배경이 되는 강화학습 이론과 마코프 의사결정 과정에 대하여 자세히 설명하였다. 3장에서는 충돌회피를 위해 고려해야하는 국제 해상충돌 예방규칙과 이에 따른 행동공간의 구성에 대하여 설명하고, 무인수상선의 충돌회피에 사용된 지역경로계획 기법인 속도 장애물(velocity

obstacle)기법에 대하여 설명하였다. 4장에서는 본 연구에서 제안된 심층강화학습 기반 충돌회피 기법에 대하여 설명하고, 그 특성에 대한 분석을 진행하였다. 심층강화학습을 위한 세미-마코프 의사결정과정을 구성하였고 충돌회피를 위해 사용된 유도법칙과 제어기에 대하여 설명하였다. 5장에서는 심층강화학습 기반 충돌회피 기법에 대한 검증을 수행하였다. 충돌회피 성능검증을 위한 대상 무인수상선 실선과 운동모형에 대하여 설명하고, 시뮬레이션과 자유항주실실험을 통한 성능분석에 대하여 기술하고있다.

# 제 2 장 심층강화학습

## 2.1 강화학습

### 2.1.1 강화학습 문제

강화학습은 학습자(agent)가 주변환경과 상호작용을 하며 특정 목적을 이루기위한 최선의 행동을 찾아가는 기계학습(machine learning)의 한 종류이다. 감독학습(supervised learning)의 경우, 학습단계에서 입력과 그에 대한 정답(label)이 수동적으로 주어지지만, 강화학습은 학습자가 능동적으로 환경에 대한 행동(action)을 하고, 이 행동에 대한 평가(보상, reward)를 받아 학습이 이루어진다.

이때, 보상은 현재 학습자가 처해 있는 상태(state)가 얼마나 높은 효용값을 갖는지를 평가하는 스칼라 값이다. 주변환경과 학습자 간의 상호작용은 Fig.2.2에 도식화되어 있다. 매 시간 $t$마다 에이전트는 현재 상태 $s_t$에 대한 보상 $r_t$를 받고, 학습된 규칙을 통해 행동 $a_t$를 수행한다. 주변환경은 에이전트의 행동에 대한 반응으로 에이전트의 상태를 $s_{t+1}$로 변화시키게 되고, 에이전트는 새로운 보상값 $r_{t+1}$를 받게된다. 이러한 순차적 의사결정 과정은 강화학습이 진행됨에 따라 반복된다.

강화학습과정에서 학습자의 행동은 정책함수(policy function) $\pi(s_t)$에 의해서 결정이되는데, 강화학습의 목표는 최적의 정책함수 $\pi^*$를 찾음으로써 미래에 받을 보상의 합 $G = r_{t+1} + r_{t+1} + ... = \sum_{k=1}^{\infty} r_{t+k+1}$을 최대화 하는 것이다. 강화학습의 한가지 특징은 이러한 정책함수를 학습하는 단계에서 exploration(탐험)과 exploitation(이용)이라는 두가지 상이한 단계를 거친다는 점이다. 여기서 exploration 과정은, 학습자가 행동을 선택할 때 이전에 가보지 않은 영역의 행동

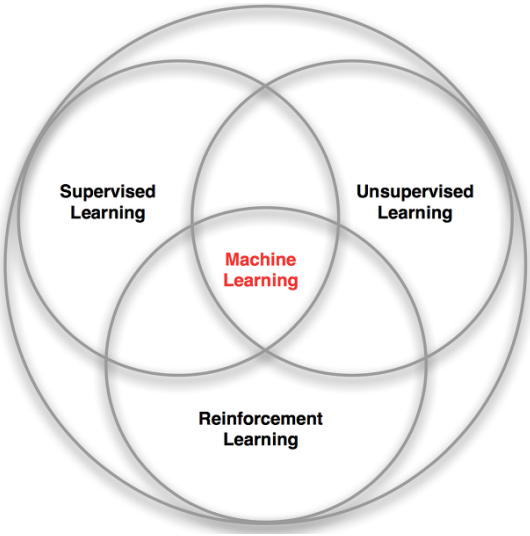



Fig 2.1: Three main types of machine learning

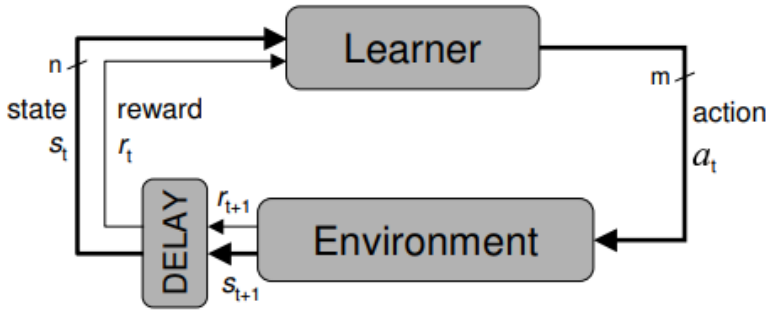



Fig 2.2: Interaction between learning agent and environment in reinforcement learning problem

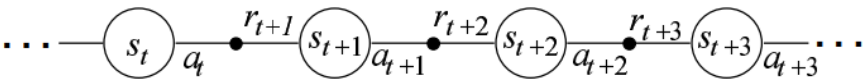



Fig 2.3: Iterative iteraction between agent and environment

을 수행함으로써 다양한 공간에 대해 학습하는 것을 의미하고, exploitation 과정은 과거에 수행했던 행동들 중 효과적이였던 행동값을 다시 선택하는 것을 의미한다. 강화학습은 이들 중 한가지의 과정만을 사용하여 문제를 해결하는 것이 아니라, exploration과 exploitation을 적절히 조화롭게 사용하여 올바른 정책을 찾아나간다.

강화학습의 또다른 특징은, 현재의 보상값을 최대화하는 것이 목표가 아니라 미래의 보상값의 합을 최대화 하는 것이 목표라는 점이다. 즉, 현재 시점에서 선택하는 행동이 즉시 받게되는 보상(immediate reward) 뿐만 아니라 이 행동을 선택함으로써 미래에 받을 수 있는 보상값에도 영향을 미친다는 특징이다. 이는 강화학습이 다른 기계학습과 구분되는 중요한 특성이다.

### 2.1.2 마코프 의사결정 과정

마코프 의사결정 과정(Markov decision process)은 의사결정과정에 대한 수학적 표현방법으로, 일반적으로 강화학습이론은 마코프 의사결정 과정을 기반으로 한다. 마코프 의사결정 과정은 모든 환경이 마코프 속성(Markov property)을 갖는다는 가정하에 이루어지는데, 마코프 속성이란 어떤 상태가 다음 결정을 내리기에 필요한 모든 정보를 가지고 있음을 의미한다. 이를 수학적으로 표현하면 식 2.1와 같다. 마코프 속성에 따르면, 미래의 보상값은 과거의 상태 또는 행동에 영향을 받지 않는다. 따라서, 마코브 속성을 가지면, 다음 시간의 상태값과 보상값을 계산하기 위해서는 현재 시점 이전에 대한 정보는 불필요하고, 대신 오직 현재의 상태와 행동값만이 필요하다.

$$P(s_{t+1}r_{t+1}|s_t, a_t) \;=\; P(s_{t+1}r_{t+1}|s_t, a_t, r_t, s_{t-1}, a_{t-1}, r_{t-1}, ...) \tag{2.1}$$

상태공간과 행동공간이 주어지면, 학습자의 거동에 대한 확률적 모형은 $P^a_{ss'}$

라는 상태전이함수 (state transition function)로 표현될 수 있다. 이는 현시점의 상태 $s$에 행동 $a$가 가해질 때, 다음시점에서의 상태가 $s'$일 확률을 의미한다. 상태 전이함수는 수학적으로 식 2.2로 표현된다.

$$P^a_{ss'} = P\{s_{t+1} = s' | s_t = s, a_t = a\} \tag{2.2}$$

특정 상태에 대한 보상값을 결정하는 함수를 강화함수(reinforcement function)라고 하는데, 강화함수는 결정론적(deterministic) 함수로 정의되고, $R^a_{ss'}$로 표현된다. 상태 $s$에서 $a$행동을 수행하여 $s'$상태가 되었을때 학습자가 받는 보상값은 식 2.3에 의해 표현된다.

$$R^a_{ss'} = E\{r_{t+1} | s_t = s, a_t = a, s_{t+1} = s'\} \tag{2.3}$$

강화학습의 목표는 미래에 받을 보상값을 최대화하는 것이다. 이를 수식화하기위해 미래에 받을 보상을 의미하는 반환(return) $G$를 식 2.4와 같이 정의한다. 이때, $\gamma$는 현재 시점의 보상을 미래 시점의 보상에 비해 높이 평가하기 위해 도입된 할인 인자(discount factor)로, 0과 1사이의 값을 갖는다.

$$G_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + .. = \sum_{k=1}^{\infty} \gamma^k r_{t+k+1} \tag{2.4}$$

학습자의 행동을 결정하는 정책 $\pi$은 해당 정책을 유지하였을 때 미래에 받을 것으로 예상되는 보상값을 이용하여 평가한다. 이 값을 가치 함수(value function)라고 하는데, 현재 상태에 대한 가치함수를 상태 가치함수 $V^\pi(s)$라고 정의한다. 상태가치함수 $V^\pi(s)$는 현재상태 $s$에서 정책 $\pi$를 유지할때 기대되는 할인된 보상값의 합 (discounted sum of rewards) 으로 정의된다. 현재 상태와 행동에 대한 가치함수는 행동가치함수 $Q^\pi(s, a)$로 정의되고, 이는 현재 상태 $s$에서 행동 $a$를 수행하고, 이후에는 정책 $\pi$를 따를 때 기대되는 할인된 보상값의 합을 의미한다.

두 가치함수는 식 2.5와 식 2.6로 정의된다.

$$V^{\pi}(s) = E_{\pi}\{G_t|s_t = s\} = E_{\pi}\left\{\sum_{k=1}^{\infty}\gamma^k r_{t+k+1}|s_t = s\right\} \tag{2.5}$$

$$Q^{\pi}(s,a) = E_{\pi}\{G_t|s_t = s, a_t = a\} = E_{\pi}\left\{\sum_{k=1}^{\infty}\gamma^k r_{t+k+1}|s_t = s, a_t = a\right\} \tag{2.6}$$

강화학습의 목표는 학습을 통해 미래에 받을 보상값을 최대화 하는 최적의 정책 $\pi^*$을 찾는것이다. 최적 정책 $\pi^*$는 가치함수의 정의에 따라 모든 정책 $\pi$중, 모든 상태와 행동에 대하여 최대값을 갖는 가치함수 ($V^{\pi}$혹은 $Q^{\pi}$)를 의미한다. 이는 수학적으로 다음과 같이 표현된다.

$$V^{\pi^*}(s) = max_{\pi}V^{\pi}(s) \tag{2.7}$$

$$Q^{\pi^*}(s,a) = max_{\pi}Q^{\pi}(s,a) \tag{2.8}$$

최적의 가치함수를 찾기위해서 Bellman 방정식을 사용하는데, Bellman 방정식은 특정 상태 $s$의 가치값과 다음 상태에서의 가치값간의 관계를 나타내는 수식이다. 두 상태간의 관계를 도출하기 위해 상태전이함수 $P^a_{ss'}$를 사용하면, 상태가치함수 $V$와 행동가치함수 $Q$에 대한 Bellman 최적식(Bellman optimality equation)을 식 2.9와 식 2.10과 같이 정리할 수 있다.

$$V^{\pi^*}(s) = max_a \sum_{s'} P^a_{ss'}\left[R^a_{ss'} + \gamma V^*(s')\right] \tag{2.9}$$

$$Q^{\pi^*}(s,a) = max_a \sum_{s'} P^a_{ss'}\left[R^a_{ss'} + \gamma Q^*(s',a')\right] \tag{2.10}$$

연속적인 두 상태에 대한 상태전이함수 $P^a_{ss'}$를 알면, 벨만 최적식을 통해 최적가치함수 $V^*$와 $Q^*$를 찾을수 있다. 최적가치함수가 구해지면, 이를 이용하여 최적정책 $\pi^*$를 구할 수 있는데, 이는 곧 강화학습 문제의 해결을 의미한다. 최적

상태 가치함수 $V^*$를 알고 있다면, 최적정책은 현재상태 $s$에서 전이될 수 있는 상태 중 최대의 상태함수값을 갖는 $s'$로 전이시키는 행동 $a$을 선택함으로써 구할 수 있다. 이때의 행동 $a$는 각 상태 $s'$값이 갖는 상태가치함수값을 비교함으로써 계산할 수 있다. 최적의 정책을 찾기위해서는 오직 가치함수값의 상대적인 값만을 비교하면 되는데, 이는 가치함수의 정의 자체가 이미 장기적(long-term) 보상값을 의미하는 반환(식 2.4)값으로 정의되었기 때문이다. 최적 행동 가치함수 $Q^*$를 알고있는 경우, 최적의 행동은 현재상태가 $s$일때 선택할수 있는 행동 $a$중 최대값의 행동가치함수 $Q^*(s,a)$값을 갖는 행동을 의미한다. 이는 식 2.11으로 표현할 수 있다.

$$\pi^*(s) = argmax_{a \in A(s)} Q^{\pi^*}(s,a) \tag{2.11}$$

### 2.1.3 강화학습 알고리즘

마코프 의사결정 과정을 기반으로 정의되는 강화학습문제를 해결하기 위하여 다양한 기법들이 개발되어 왔다. 본 절에서는 이러한 다양한 기법 중 대표적인 3가지 기법인 동적 프로그래밍 기법, 몬테카를로 기법, 시간차 기법에 대하여 설명한다.

- **동적 프로그래밍 기법 (dynamic programming, 이하 DP기법)**

  DP기법은 환경에 대한 정보인 $P^a_{ss'}$와 $R^a_{ss'}$에 대한 사전정보가 주어질 때 사용할 수 있는 기법으로, Bellman 최적식을 반복적으로 계산하여 최적정책을 도출하는 기법이다. DP기법은 온라인(online)기법으로, 학습자가 환경과 상호작용을 함과 동시에 학습을 수행하는 기법이다. DP기법에서는 매 계산스텝에서 현재 상태 $s$와 가능한 모든 다음상태 $s'$에 대한 가치함수값을 업데이트시킴으로써 학습이 이루어진다. 가치함수를 갱신하기위해 사용하

는 방정식은 식 2.12와 같다.

$$V_{k+1}(s) \leftarrow \sum_{a} \pi(s,a) \sum_{s'} P^{a}_{ss'} \left[ R^{a}_{ss'} + \gamma V_k(s') \right] \tag{2.12}$$

식 2.12에 의하면 $k+1$시점에서의 상태가치함수를 갱신하기위하여 $k$시점의 상태가치함수가 이용되는 것을 확인할 수 있다. 이때, 최적정책을 알지못하는 상황이므로, $V^{\pi}(s')$값 대신 현시점에서의 상태가치함수 $V_k(s')$를 사용한다. 이와같은 과정을를 부트스트랩(bootstrap) 이라 하는데, 이를 통해 DP 알고리즘에서는 온라인으로 학습을 진행할 수 있다. 학습을 반복적으로 수행하면, 정책 $\pi$에 대한 개선이 이루어지고, $V_k(s')$는 $V^{\pi}(s')$에 가까운 값으로 개선되는 특징이 있다.

- **몬테 카를로 기법 (Monte-Carlo method, 이하 MC기법)**

  MC기법은 DP기법과는 다르게 환경에 대한 사전정보가 필요없다. 대신, 가치함수를 갱신하기 위해 학습자가 환경과 상호작용하며 얻게되는 상태의 변화, 보상값이 학습에 사용된다. MC기법은 경험과 학습이 동시에 이루어지는 온라인 기법이 아닌, 현재의 정책 $\pi$를 따라 환경과의 상호작용을 하고, 사건(episode)이 끝난 후에 얻은 경험값들을 이용해 일괄적으로 학습을 수행하는 오프라인(offline) 학습법이라는 특징이 있다. 학습이 끝나게 되면, 반환값을 얻기위해 들렀던 모든 상태 $s$값에 대하여 가치함수 갱신 (value update)를 수행한다. 이 과정을 반복적으로 수행하면, 특정 정책 $\pi$에 대한 가치함수가 계산된다. 식 2.13은 상태가치함수를 갱신하는데 일반적으로 사용되는 방정식을 나타내고있다. 이때, $\alpha$는 학습의 속도를 결정짓는 학습 속도(learning rate) 변수이고, $r_t$는 사건을 수행하며 얻은 보상값이다.

$$V^{\pi}_{k+1}(s_t) \leftarrow V^{\pi}_{k}(s_t) + \alpha \left[ r_t - V^{\pi}_{k}(s_t) \right] \tag{2.13}$$

MC기법에서는 현재 정책에 대한 평가(evaluation)가 진행되고, 이 정책에 대한 개선(improvement)이 이루어진다. 이러한 과정들을 정책 반복(policy iteration)이라 하는데, 이 과정들을 반복하면서 최적의 정책을 도출해 내는 방법이 MC기법이다. MC기법의 단점은 반환 $G$를 얻기위해 사건이 끝날 때까지 기다려야 한다는 점이다. 이러한 단점 때문에 MC기법은 연속적인 학습이 요구되는 문제에는 적합하지 않다는 특징이 잇다.

- **시간차 기법 (temporal difference method, 이하 TD기법)**

  TD기법은 타 기법들의 단점을 보완한 기법이다. MC기법처럼 상호작용을 통한 경험값을 이용하기때문에 환경에 대한 사전정보가 필요하지 않다는 장점이 있고, DP기법처럼 미지의 정책을 추정하기위해 부트스트랩 기법을 사용하였기때문에 온라인으로 학습이 가능하다는 장점이 존재한다. TD기법에서 상태가치함수의 갱신은 식 2.14를 통해 이루어진다. TD기법은, MC기법과 유사하게 상태가치함수 $V^{\pi}_{k+1}(s_t)$를 갱신하기위하여 현재의 가치함수와 미래에 대한 할인된(discounted) 보상값을 사용한다. 다만, MC기법과 다른점은 미래에 대한 반환값 $G$를 바로 사용하는 것이 아니라, 이를 현재 받은 보상값(immediate reward) $r_t$와 할인된 다음상태의 가치함수값 $\gamma V^{\pi}_{k}(s_t+1)$으로 표현한다는 점이다. 이러한 부트스트랩을 통해 TD기법은 온라인 학습이 가능해지는 기법적 특징이 있다.

$$V^{\pi}_{k+1}(s_t) \leftarrow V^{\pi}_{k}(s_t) + \alpha\big[r_t + \gamma V^{\pi}_{k}(s_t+1) - V^{\pi}_{k}(s_t)\big] \tag{2.14}$$

TD기법은 환경에 대한 사전정보가 없고, 온라인으로 학습이 필요한 문제에 적합한 강화학습 기법으로, 대부분의 강화학습을 통한 문제에 이 TD기법이 활용된다. 다양한 종류의 강화학습 알고리즘이 TD기법을 기반으로 제안되

었는데, 대표적인 알고리즘은 액터-크리틱(Actor-Critic) 알고리즘과 Q-학습 알고리즘이다.

본 연구에서는 TD기법 알고리즘 중 Q-학습 기반 알고리즘을 이용하여 무인수상선의 충돌위험도를 추론하였다. 다음의 내용은 본 연구에서 사용한 알고리즘의 기반이 되는 Q-학습 기법에 대한 자세한 설명을 포함하고 있다.

**Q-학습(Q-learning)**

Q-학습은 Watkins and Dayan,(1992)[44]에 의해 제안된 TD기법 기반 알고리즘으로, 행동가치함수 $Q^{\pi}$를 이용하여 최적정책 $\pi^{*}$를 찾는 기법이다. 행동가치함수 $Q^{\pi}(s,a)$는 가치함수 $V^{\pi}$와는 다르게 최적의 행동가치함수 $Q^{\pi^{*}}(s,a)$가 한번 정해지면, 환경에대한 정보가 필요없이 곧바로 최적의 정책을 추출해 낼 수 있다는 장점이 존재한다. Q-학습의 또 다른 중요한 특징은 off 정책(off policy) 학습기법이라는 점인데, 이는, 최적 정책 $Q^{*}$을 도출하기 위해 특정 정책을 항상 따르지 않아도 주어진 환경을 균일하게 방문하기만 하면 최적정책으로 수렴한다는 것을 의미한다. Q-학습의 행동가치함수 $Q(s_t,a_t)$의 갱신은 식 2.15에 의하여 이루어진다.

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha\big[r_{t+1} + \gamma \cdot max_{a_{max}} Q(s_{t+1},a_{max}) - Q(s_t,a_t)\big] \quad (2.15)$$

앞서 설명하였듯이 $Q$-학습이 최적정책으로 수렴하기위해서는 상태-행동 공간 내에 있는 모든 상태-행동 쌍에 균일하게 방문하는 것이 필요하다. 이를 위해 2.1.1절에서 설명한 exploration 과정을 거치는 것이 일반적이다. 가보지 않은 영역의 행동을 선택하는 exploration과 현재 정책에서 최선의 행동을 선택하는 exploitation은 서로 상반되는 개념이지만 이둘을 조화롭게 사용하면 $Q$-학습의 정책을 최적정책으로 수렴시킬 수 있다. $Q$-학습에서 사용하는 exploration 기법 중 가장 대표적인

방법은 가장 $\epsilon$-greedy 방법이다. $\epsilon$-greedy방법은 $\epsilon$의 확률로 exploration을 수행하는 기법으로, 학습이 진행됨에 따라 이 $\epsilon$값을 점근적으로 감소시키는 방법이다. 즉, 학습의 초기단계에는 환경에 대한 경험 값이 부족하므로 exploration에 집중하는 반면에, 학습이 진행됨에 따라 exploration보다는 exploitation의 빈도를 높여 최적 정책에 수렴하도록 유도하는 방법이다.

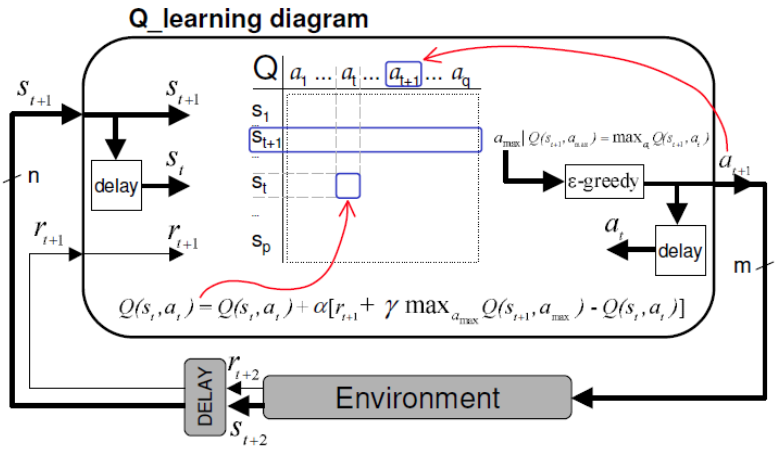



Fig 2.4: Diagram of tabular Q learning algorithm. Figure from [5]

Fig.2.4는 가장 단순한 형태의 $Q$-학습 알고리즘의 동작을 도식화 한 그림이다. Fig.2.4에 의하면, 행동가치함수 $Q(s_t, a_t)$의 갱신을 위하여 테이블의 $s_t$행 $a_t$열의 $Q$값을 찾고, 다음 상태의 최대 Q값을 이용하여 $Q(s_t, a_t)$값을 갱신하는 것을 확인할 수 있다.

Fig.2.4에서 보여주는 단순한 형태의 $Q$학습 알고리즘은 이산적인 상태 및 행동공간에 대해 정의되었다. 하지만, 대부분의 로봇공학과 관련된 문제, 혹은 실제 현실 문제(real-world problem)들은 연속적인 상태 및 행동공간에 대한 고려를 필요로 한다. 이러한 문제를 해결하기 위해서는 상태 및 행동공간을 더 작은 단위로

이산화시키는 방식으로 접근할 수 있는데, 만일 상태공간이나 행동공간의 차원(dimension)이 고차원인 경우 지나치게 많은 이산상태가 생성되게 된다. 결과적으로, Fig.2.4의 테이블의 크기가 매우 커지게 되어 Q-학습을 수행하는데 필요한 시간이 기하급수적으로 증가하게 된다. 이러한 문제는 강화학습이 실제 현실 문제에 적용되는 것을 어렵게 만드는 요인으로, 이러한 문제를 *일반화 문제*라 한다.

**신경 Q-학습 (neural Q Learning)**

Q-학습의 *일반화 문제*를 해결하기 위한 다양한 접근방식이 개발되었는데, 가장 대표적인 방법은 신경회로망 등의 함수근사기를 사용하여 상태가치함수$V(s,a)$나 행동가치함수 $Q(s,a)$를 근사화 하는 방법이다. 신경회로망은 비선형 함수를 근사화하는데 탁월한 성능을 보이는 함수근사기(function approximator) 이고, 연산이 쉽다는 장점이 존재한다. 또한, 신경회로망을 구성하는 데 필요한 파라미터의 개수가 행동공간이나 상태공간의 크기에 종속적이지 않다는 장점이 있다. 신경망 Q-학습에서는 동일한 Q-학습의 갱신 방정식 2.15를 이용하여 행동가치함수를 갱신한다. 다만, 행동가치함수값 을 계산하기 위하여 신경회로망을 이용하여 Q값을 근사화 한다는 점이 기존의 Q-학습과 다른 점이다.

신경 Q-학습 기법도 기법적 한계가 존재하는데, 학습이 이루어짐에 따라 신경망의 안정성(stability)이 떨어진다는 점이다. Q-학습에는 크게 2종류의 학습이 존재하는데, Q값을 근사화 시키기위한 신경망의 학습이 존재하고, 두번째는 행동가치함수를 갱신시켜 최적정책을 찾아가는 학습이 존재한다(Q-학습). 이 두 단계의 학습은 서로를 기반으로 상호의존적으로 작동한다. 예를 들면, 식 2.15에 의하면, $Q(s_t,a_t)$값을 갱신하기위해서는 다음 상태에서의 행동가치함수 값이 필요로 하게되는데, 이는 $Q(s,a)$에 대한 근사기인 신경망의 학습을 위해 자기 자신의

함수를 이용하여 $Q(s', a_{max})$값을 계산해야 한다. 이러한 상호 의존성은, 신경 Q-학습 알고리즘을 매우 불안정하게 만드는 특성이 있으며, 이때문에 학습의 안정성이 감소하 게되고, 결국 학습의 성능과 수렴성에 부정적인 영향을 미친다. Weaver et al (1998)[45]는 이러한 문제를 *간섭 문제*라고 정의하였다.

## 2.2 심층강화학습

심층강화학습은 일반적으로 딥러닝이라고 지칭되는 심층신경망(deep neural network)을 강화학습과 결합하여 이미지와 같은 시각적 정보에 대한 강화학습을 수행하는 기법이다. 시각적 정보에 대한 처리를 하기위해 합성곱 신경망(convolutional neural network)이 일반적으로 사용된다.

합성곱 신경망을 이용하여 시각적 정보에 대한 강화학습을 시도는 이전에도 다수 존재했지만, Mnih 이 2015년에 발표한 DQN(Deep Q-Network) [28] 알고리즘이 등장한 이후, 심층강화학습에 대한 연구가 폭발적으로 증가하였다. Mnih은 DQN을 이용하여 컴퓨터가 스스로 비디오게임인 Atari game을 학습하는 연구를 수행하였다.

### 2.2.1 Deep Q Network

Mnih et al (2015)[28] 이 고안한 DQN 알고리즘의 구조는 Fig.2.5와 같다. DQN은 입력단에 합성곱층이 포함되어있는데, 이는 게임 이미지 픽셀에 대한 합성곱 연산을 수행하여 이미지 각 영역의 관계를 파악하여 영상정보를 처리하기 위함이다. DQN은 크게 두 가지 종류의 기법을 도입하여 신경망기반 Q-학습 알고리즘에서 빈번히 발생하는 간섭문제를 해결하였는데, 이는 경험 재생(experience replay)기법과 목표신경망의 분리(seperate target network)기법이다.

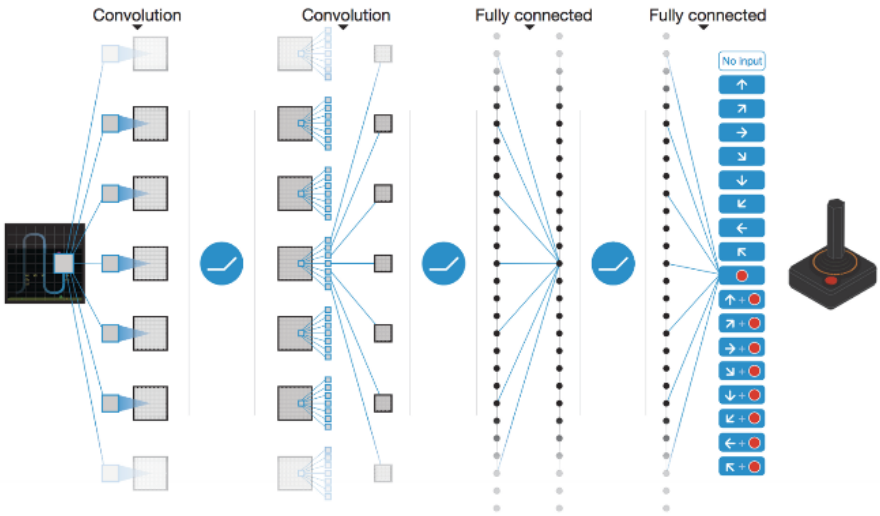



Fig 2.5: Structure of the Deep Q Network. Figure from [28]

경험 재생은, 단순히 학습자가 환경과 상호작용하여 경험을 얻으면 이 경험을 곧바로 학습에 사용하는 것이 아니라, 경험의 표본을 데이터베이스에 저장해두고 매 순간 학습을 해야할때 데이터베이스 내에 있는 표본을 랜덤하게 추출하여 학습을 수행하는 방법이다. 강화학습의 학습 데이터는 일반적으로 연속적인 상태공간을 움직이면서 생성되기때문에 학습 데이터들 간의 연성(coupling)이 심하다는 특징이 존재한다. 하지만 경험 재생 기법을 사용하면, 학습데이터에 대한 연성을 제거할 수 있으며, 이를 통해 과적합(overfitting)을 억제하고 안정적인 학습을 가능하게 한다.

두 번째 기법은 목표신경망(target network)을 분리하는 방법인데, 이는 앞서 설명하였던 간섭문제의 원인을 해소하기 위함이다. 신경망을 이용한 강화학습의 경우 두 가지 학습이 동시에 이루어지기 때문에 불안정한 특성을 보이는데, 목표 Q 값을 계산하기위한 신경망을 분리하여, 두 가지 학습의 간섭을 차단하는 방법이다.

### 2.2.2 심층강화학습 기법

DQN이 발표 된 이후, 여러가지 종류의 심층강화학습 기법들이 추가적으로 발표되었는데, 이들 기법을 통해 보다 안정적이고 빠른 속도의 학습이 가능해졌다.

Double DQN기법은 Hasselt et al (2016)[42]이 제안한 기법으로, DQN에서 종종 발생하는 행동가치함수(Q값)에 대한 과대평가를 억제하는 기법이다. Double DQN에서는 두 개의 신경망을 사용하는데, 한개의 신경망은 빠르게 업데이트 되면서 행동을 선택하도록 구성하고, 나머지 신경망은 상대적으로 천천히 업데이트를 수행하며 다른 네트워크의 Q값이 지나치게 커지는것을 방지하는 구조이다.

Dual DQN기법은 Wang et al (2016)[43]이 제안한 기법으로, 식 2.16의 정의에 따라, 행동가치함수($Q$)의 학습을 위해 상태가치함수($V$)와 어드벤티지($A$)를 각각 분리하여 학습을 수행하는 구조이다.

$$Q(s,a) = V(s) + A(s,a) \tag{2.16}$$

이 때, 상태가치함수는 현재 상태에 대한 효용가치를 판단하는 반환(return) 값을 의미하고, 어드벤티지($A$)는 각 행동에 대한 상대적인 효용값을 의미한다. 기존에 $Q$값 만을 추정하던 구조를 $V$와 $A$에 대해 분리하여 추정하는 이유는, 학습자가 얻게 되는 보상값이 현재 상태에 의한 것인지 이전에 수행한 행동에 의한 것인지 여부를 분리할 수 있기때문이다. 이를 통해 기존기법에 비해 더 강건하게 상태가치함수를 추정할 수 있다.

### 2.2.3 심층강화학습 기반 연구

심층강화학습의 대표적인 연구들은 Atari 비디오 게임을 통해 개발 되어져 왔지만, 실제 다양한 분야에서 심층강화학습에 대한 적용연구들 또한 활발히 이루

어지고 있는 실정이다. 자율주행자동차(Sallab et al, 2017)[32], 보행로봇(Peng et al, 2017)[30], 무인항공기(Coates et al, 2019)[8], 로봇팔 (Levine et al, 2016)[27] 등의 광범위한 분야에서 심층강화학습기법이 사용되고있다.

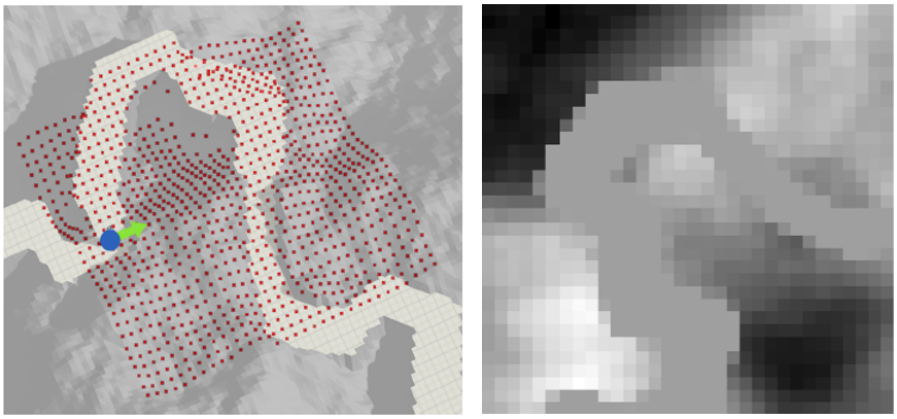


Fig 2.6: Deep Reinforcement learning application to locomotion problem. Figure from [30]

Fig.2.6은 시뮬레이션 환경에서 인간의 2족 보행을 학습하는 연구를 수행한 바 있다. 보행을 위한 지형정보는 Fig.2.6의 우측그림과 같은 형태로 격자지도 형태로 제공하였다.

# 제 3 장 무인수상선의 충돌회피

## 3.1 선박의 충돌 회피

### 3.1.1 국제 해상충돌 예방규칙

국제 해상충돌 예방규칙(International regulations for preventing collisions at sea, COLREGs) 은 1972년 국제 해사 기구(IMO)가 제정한 규칙으로, 해상에서 운항하는 선박의 운항자가 선박의 충돌사고를 예방하기위해 필수적으로 준수해야 하는 국제조약을 의미한다. COLREGs는 각 조우상황에 대하여 선박의 피항의무 여부에 대하여 규정하고있고, 피항의무선(give way vessel)과 진로유지선(stand on vessel)의 바람직한 운항방법에 대하여 기술하고 있다. COLREGs는 국제적으로 통용되는 충돌예방규칙임과 동시에 세계 각국의 해상교통안전법과 밀접한 관련이 있으므로, 운항자가 탐승하는 유인선박뿐만 아니라 무인수상선의 경우에도 이를 반영하여 행동계획을 수립해야 한다. COLREGs는 총 5장으로 구성되어 있는데, 본 연구에서는 항법규정에 해당하는 Part B의 규칙들이 주로 활용되었다. 아래의 항목들은 본 연구에 활용된 주요 COLREGs 규칙을 나열한 것이다.[21]

**제 8조. 충돌 회피를 위한 동작**

(a) 충돌을 피하기 위한 모든 동작은 적극적이고, 충분한 시간을 두고 그리고 적절한 선박운용술에 따라 행하여야 한다.

(b) 충돌을 피하기 위한 침로와 속력의 동시변경, 침로 또는 속력의 변경은, 사정이 허락하는 한, 육안이나 레이더에 의하여 **관찰하고 있는 타선에게 즉시**

**명백하도록 충분히** 하여야 한다. 연속적인 작은 침로와 속력의 변경, 침로 또는 속력만의 변경은 피하여야 한다.

(c) 만일 충분한 해면이 있고 적시에 충분하게 행하고 다른 또 하나의 근접상태가 형성되지 아니한다면, 침로만의 변경도 근접 상황을 피하는 가장 유효한 동작이 될 수 있다.

(d) 타선과의 충돌을 피하기 위하여 취하는 동작은 안전한 거리를 두고 통과하도록 하여야 한다. 수행한 피항행위의 효과는 타선이 완전히 항과할 때까지 주의깊게 확인하여야 한다.

**제 14조. Head on 상황**

(a) 충돌의 위험이 내포되도록 2척의 동력선이 반대되는 방향 또는 거의 반대되는 방향으로 마주치는 경우에는 각 선박은 서로 다른 선박의 좌현측을 통과할 수 있도록 각기 우현측으로 변침하여야 한다.

**제 15조. Crossing 상황**

- 두 척의 동력선이 서로 진로를 횡단할 경우에 충돌의 위험이 있을 때에는 다른 선박을 우현측에 두고 있는 선박이 다른 선박의 진로를 피하여야 하며, 사정이 허락하는 한, 다른 선박의전방을 횡단하여서는 안된다

**제 16조. 피항선(give way vessel)의 동작**

- 이 규칙에 의하여 다른 선박의 진로를 피하여야 할 선박은 타선을 확실하게 피할 수 있도록 가능한 한 조기에 충분한 동작을 취하여야 한다.

**제 17조. 유지선(stand on vessel)의 동작**

(a) (i) 두 선박 중 한 선박이 다른 선박의 진로를 피하여야 할 경우 다른 선박은 그 침로 및 속력을 유지하여야 한다.

(ii)그러나 유지선은 진로를 피하여야 할 선박이 이 규칙에 따른 적절한 동작을 취하지 않음이 분명해지는 즉시, 자선의 조종만으로서 충돌 회피 행위를 취할 수 있다.

(b) 양선이 아주 가까이 접근하여 피항선의 동작만으로 충돌을 피할 수 없다고 판단할 때에는 유지선도 피항을 위한 협력동작을 취하여야 한다.

(c) 횡단 상태에서 다른 동력선과 충돌을 피하기 위하여 이 조문 (a)항 (ii)의 규정에 따라 동작을 취하 는 선박은 상황이 허락하는 한, 자선의 좌현측에 있는 선박을 피하기 위하여 좌현측으로 변침하여서 는 아니된다.

**제 13조. Overtaking 상황**

(a) 추월선은 제2장 제1절에 있는 규칙의 여하한 규정에도 불구하고 추월당하는 선박의 진로를 피하여 야 한다.

(d) 두 선박간의 방위가 그 후에 여하히 변경되더라도 추월선이 본규칙상의 의미에 있어서의 횡단선으로 되는 것은 아니며 또한 추월선은 완전히 앞질러 멀어질 때까지 추월당하는 선박의 진로를 피하여야 할 의무를 벗어나지 못한다.

### 3.1.2 조우상황의 분류

본 연구에서는 자선과 타선의 조우상황을 판단하기위하여 Tam and Bucknall (2010) [40]이 제안한 조우상황 분류기법을 사용하였다. Tam and Bucknall은 조우상황의 판단을 위하여 자선과 타선의 위치 정보와 선수각 정보를 사용하였다.

타선과 자선의 상대위치가 계산되면, 타선은 Fig.3.1의 6개의 영역(region)중 1개의 영역으로 분류할 수 있다. 이때, $\{R1, R2, R3, R4, R5, R6\}$ 6개의 영역은 각각 $\{\frac{1}{8}\pi, \frac{1}{2}\pi, \frac{15}{8}\pi, \frac{5}{8}\pi, \frac{3}{2}\pi, \frac{11}{8}\pi\}$ 각도에 의하여 구분된다. 이때, $R_4$를 구분하기위한 각도는 Overtaking에 대한 COLREGs 규칙 (13조)에 의하여 결정되지만, $R_1$을 구분하기위한 각도는 COLREGs규칙에 명시되어있지 않다. 다만, Tam and Bucknall은 타선의 조우상황이 너무 민감하게 변화하는 상황을 억제시키고, 불확실성에 대한 고려를 수행하기위해 임의로 충분히 큰 각도를 사용하였다.

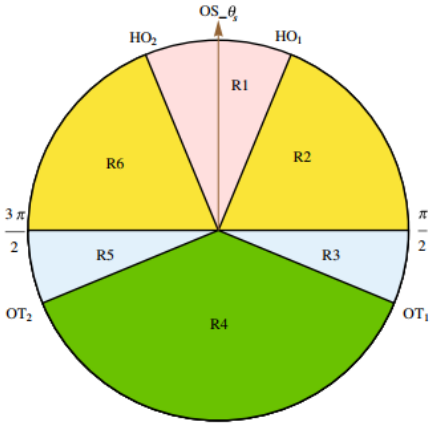



Fig 3.1: Categorization of obstacle by relative position. Figure from [40]

타선이 위치한 영역이 정해지면, 자선에 대한 타선의 상대 선수각을 계산하고, 이에따라 타선을 $\{TSR1, TSR2, TSR3, TSR4, TSR5, TSR6\}$ 6개의 영역 중 1개의 영역으로 분류할 수 있다. Fig.3.3은 타선이 상대 위치와 선수각에 의하여 분류되었을때, 해당영역이 의미하는 조우상황을 도식화 하고 있다. 각 조우상황에 대한 설명은 Table.3.1에 기술되어있다. Tam and Bucknall이 제안하는 조우상황 분류 기법은 자선과 장애물이 각각의 선수각 방향으로 운동한다는 가정하에 움직이는 동적(dynamic)장애물에 대한 조우상황을 분류하고 있다. 움직이지 않는

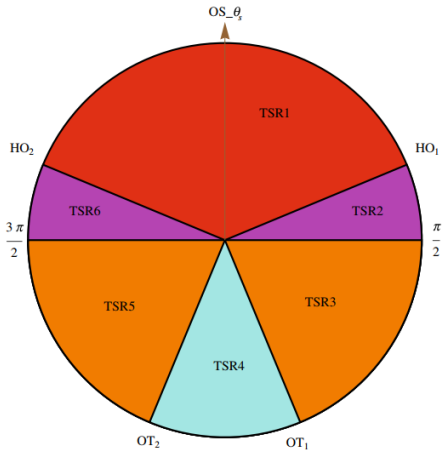



Fig 3.2: Categorization of obstacle by relative heading angle. Figure from [40]

정적(static) 장애물의 경우 본 연구에서는 조우상황을 따로 분류하지 않고, 임의의 선박안전영역(ship domain)내에 정적 장애물이 존재하면, 모두 회피대상으로 인식한다고 가정하였다.

| **Abbreviation** | **Encounter type** |
|---|---|
| HO | Head on encounter |
| OT | Overtaking encounter |
| SO | Crossing encounter (stand on) |
| SF | Safe encounter |
| GW | Crossing encounter (give way) |

Table 3.1: Abbreviation for encounter type

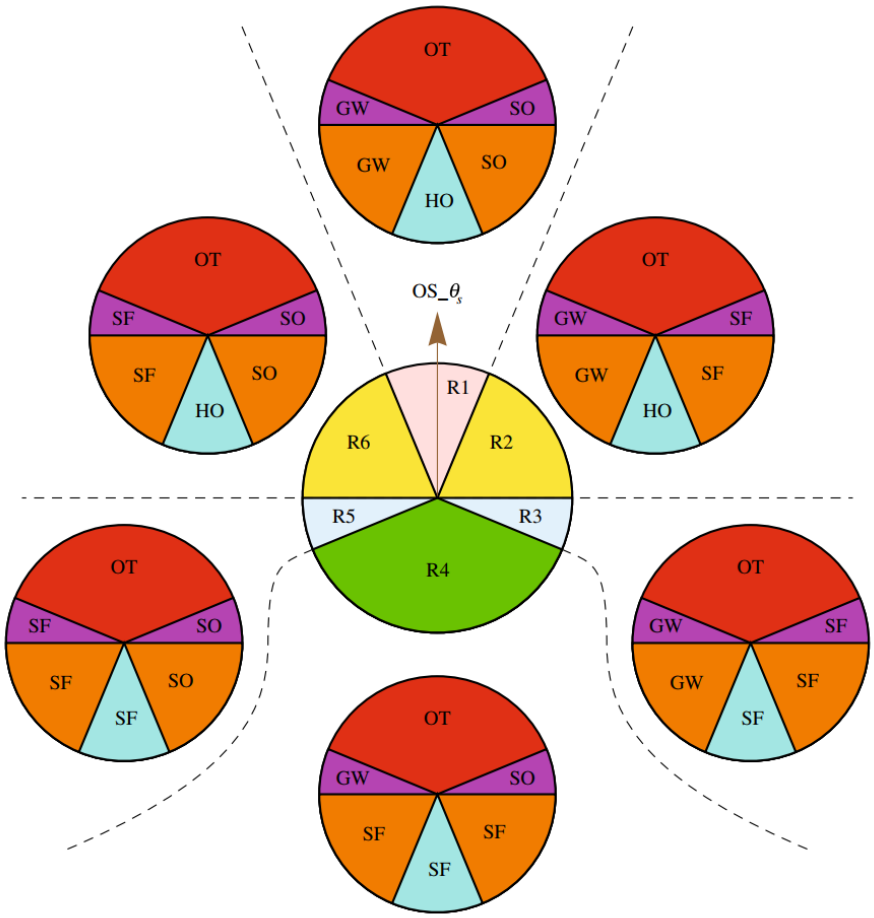



Fig 3.3: Chart used to determine the encounter type. Figure from [40]

### 3.1.3 회피 행동의 결정

3.1.2절에 의해 회피 대상 장애물과의 조우상황이 결정되면, 장애물에 대한 회피행동을 수행해야 한다. 이때, 회피행위는 COLREGs에 의거하여 현재 자선이 처한 조우상황에서 권고되는 회피행동이어야 한다. 다음의 내용은 각 조우상황에서 선박의 피항의무 여부와 COLREGs 규정을 기반으로 선정된 바람직한 피항행위를 나타내고있다.

- **Head on 상황** - 자선과 타선이 서로 마주보며 접근하는 Head on 조우상황에서는 자선과 타선이 모두 피항의 의무를가지고, 두 선박이 모두 우현변침을 수행하여 충돌을 회피해야 한다. (COLREGs 제 14조) 이때, 자선의 피항여부에 대한 타선의 가관측성(observability)을 향상시키기 위하여 충분히 큰각도의 변침을 통해 피항행위를 수행해야 한다. (COLREGs 제 8조)

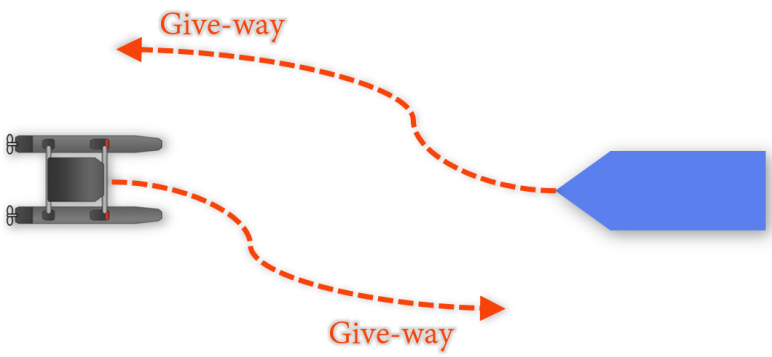


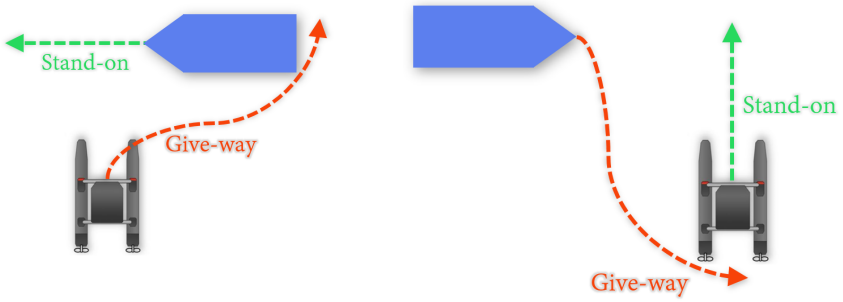

Fig 3.4: Desired avoidance behavior of ASV in case of Head on, according to COLREGs

- **Crossing (give way)** - 자선의 우현측에 타선을 두고 타선의 진로방향을

향해 횡단하는 Crossing (give way) 조우상황에서는 자선이 피항의무선의 지위를 갖고, 타선은 진로유지선의 지위를 갖는다. 이때, 자선은 타선의 선수방향을 횡단 할 수 없고, 우현변침을 통해 타선의 선미방향으로 피항행위를 수행해야 한다. (COLREGs 제 15조) 진로유지선인 타선은 피항의무선인 자선의 피항여부에 따라 향후 협력동작의 필요성 여부를 결정한다. 진로유지선의 의사결정과정의 불확실성을 감소시키고 피항의무선에 대한 가관측성을 향상시키기 위해 자선은 조기상황에 충분히 큰 피항행위를 수행해야 한다. (COLREGs 제 16조)

- **Crossing (stand on)** - 자선의 좌현측에 타선을 두고 타선의 진로방향을 향해 횡단하는 Crossing (stand on) 조우상황에서는 자선은 진로유지선의 지위를 갖고, 타선이 피항의무선의 지위를 갖는다. 이때, 약속되지 않은 진로유지선의 침로변경 또는 속력변경은 오히려 충돌의 위험도를 증가시킬 여지가 존재 하므로, 자선은 침로와 속력을 유지해야 한다. 하지만 피항의무선이 충돌의 회피를 위한 적절한 동작을 취하지 않으면, 자선은 진로유지선임에도 불구하고 충돌의 회피를 위한 적절한 피항행위를 수행해야 한다. (COLREGs 제 17조(a))

- **Overtaking** - 자선이 타선을 추월하는 Overtaking 조우상황에서는 자선이 피항의무선의 지위를 가지고, 추월당하는 선박은 진로유지선의 지위를 갖는다. Overtaking에 대한 COLREGs 규칙(제13조)은 Head on이나 Crossing의 경우처럼 피항의무선의 회피방향에 대하여 특정하여 명시하고 있지는 않지만, 일반적으로 운항자들은 우현변침을 통한 회피를 선호하는 것으로 알려져있다. [23]




(a) Crossing(give way)

(b) Crossing(stand on)

Fig 3.5: Desired avoidance behavior of ASV in case of Crossing, according to COLREGs

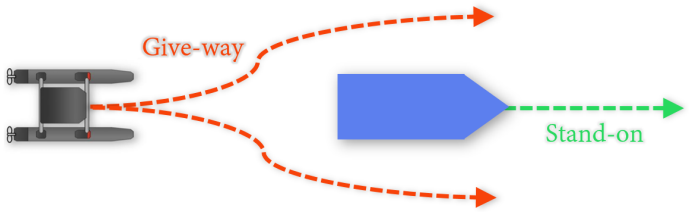



Fig 3.6: Desired avoidance behavior of ASV in case of Overtaking according to COLREGs

- **Static obstacle** - 2.2.1절의 COLREGs는 움직이는 선박간의 조우상황 및 회피행동에 대하여 기술하고 있다. 따라서, 대상 장애물이 움직이지 않는 정적(static) 장애물인 경우, 회피계획단계에서 COLREGs를 적용하는 것은 유효하지 않다. 이 경우 자선은 상황에 알맞은 적절한 피항행위(좌현변침 또는 우현변침)를 수행하여 장애물과의 충돌을 회피해야 한다.

## 3.2 충돌회피 알고리즘

이 절에서는 본 연구를 수행하는데 사용되거나 비교대상으로 선정된 충돌회피 알고리즘에 대한 이론적 배경에 대하여 기술한다. 먼저, 3.2.1절에서는 선박의 충돌회피 시점을 추론하는 데 사용되는 충돌위험도의 기본 개념과 이로부터 발전된 다양한 충돌위험도 추론 방법에 대하여 소개하고 있다. 3.2.2절에서는 본 연구에서 피항행위를 결정하는 데 사용한 속도 장애물(velocity obstacle) 기법에 대하여 설명하고 있다.

### 3.2.1 충돌회피 시점 추론 알고리즘

자선과 타선이 조우하는 상황에서 충돌 회피기동의 필요성을 판단하고, 회피 시점을 결정하는 방법론은 충돌회피 분야의 한 주제로 꾸준히 연구되어져 왔다. 충돌회피의 시점이 지나치게 빠르면 불필요한 회피기동이 발생하게 되고, 지나치게 늦으면 충돌위험도와 자선의 안전성에 대한 불확실성이 증가하게 된다. 이러한 이유로 충돌회피 여부의 판단 및 회피행위의 시점에 대한 적절한 의사결정은 충돌회피의 효과성을 확보하기 위한 중요한 요소이다. 충돌회피 시점 추론 알고리즘의 경우 크게 선박안전영역에 기반한 기법과 최근접점을 이용한 기법으로 분류할 수 있다.

#### 선박 안전영역을 이용한 충돌위험도

먼저, 자선의 위치를 기준으로 가상의 선박안전영역(ship domain)을 설정하고, 해당 영역에 대한 타선의 침범여부를 통해 충돌의 위험성을 판단하는 방법이다. Fujii and Tanaka (1971)[16]가 처음으로 타원형의 선박안전영역을 제시한 후, 선박안전영역에 국제 해상충돌 예방규칙의 요소를 반영하기 위한 노력이 Coldwell

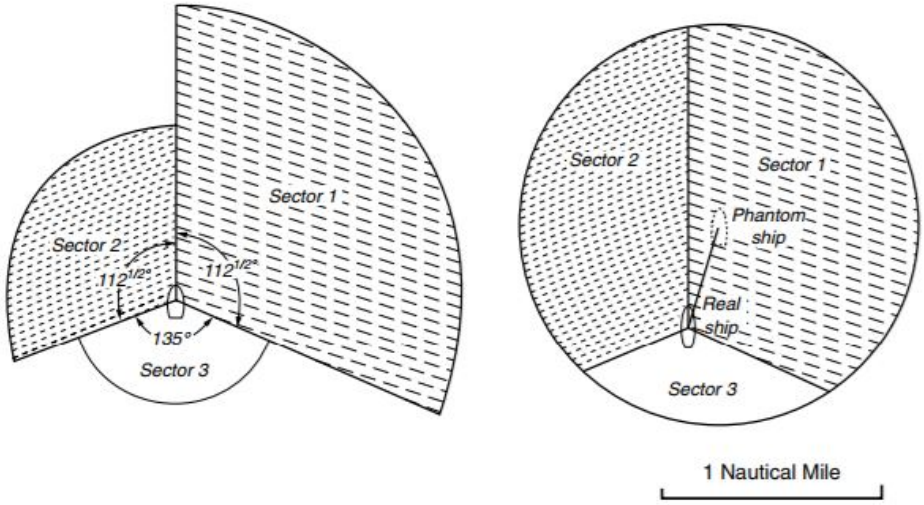



Fig 3.7: Ship domain for collision risk assessment, suggested by Fujii, Coldwell and Goodwin

[9], Goodwin [17], Davis [12]에 의해 이루어졌다. COLREGs에 의하면 회피기동은 대부분 우현변침을 통해 이루어져야 하므로, 이들 연구에서는 선박안전영역을 구성할 때 선박의 우측에 더 많은 안전거리를 확보하기위하여 좌우 비대칭 형태의 안전영역을 제안했다. Fig.3.7은 각각 Fujii, Coldwell, Goodwin이 제안한 선박안전영역의 형상을 나타내고 있다. 이 외에도 신경회로망(neural network)기법을 이용하여 선박의 안전영역에 대한 모델링을 수행한 연구(Zhu, 2001 [47]), 최근접점의 개념을 도입하여 선박안전영역의 크기를 결정하는 연구 등이 수행되었다.

Myre (2016)는 자선이 중앙에 위치한 원형의 선박안전영역을 정의하고, 이를 set based guidance(SBG) 조건으로 사용하여 무인수상선의 충돌회피에 대한 연구를 수행한 바 있다.

### 최근접점을 이용한 충돌위험도

최근접점(closest point of approach, 이하 CPA)의 개념은 선박의 충돌회피에서 충돌의 위험도를 추론하는데 사용되는 개념으로, 자선과 타선이 현재의 진로와 속도를 유지했을때 타선이 가장 가까워지는 순간의 점을 의미한다. CPA에 도달했을때, 자선과 타선의 거리를 최단접근거리(distance at closest point of approach, 이하 DCPA)라고 정의하고, CPA점 까지 도달하는데 걸리는 시간을 최단접근시간(time to closest point of approach, 이하 TCPA)라고 정의한다. Fig.3.8은 각각 CPA, DCPA, TCPA의 개념에 대하여 도식적으로 설명하고있고, 식 3.1은 이들에 대한 수학적 정의를 나타내고있다. 이때, $\vec{p}_A$ 와 $\vec{p}_B$는 각각 자선과 타선의 위치, $\vec{v}_A$ 와 $\vec{v}_B$ 는 각각 자선과 타선의 속도를 의미한다.

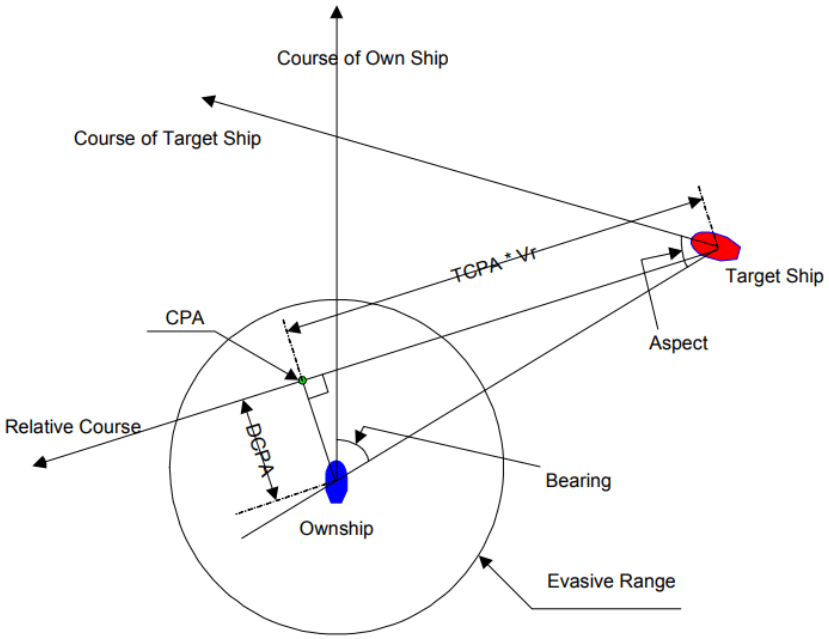



Fig 3.8: Conceptual description of TCPA and DCPA, Figure from [48]

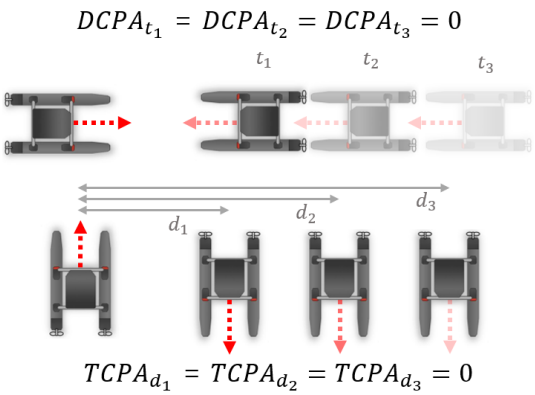



Fig 3.9: Characteristics of DCPA and TCPA

$$TCPA = \begin{cases} 0, & \text{if } ||\vec{v}_A - \vec{v}_B|| \leq \epsilon \\ \frac{(\vec{p}_A - \vec{p}_B)\cdot(\vec{v}_A - \vec{v}_B)}{||\vec{v}_A - \vec{v}_B||^2}, & \text{otherwise} \end{cases} \tag{3.1}$$

$$DCPA = ||(\vec{p}_A + \vec{v}_A \cdot TCPA) - (\vec{p}_B + \vec{v}_B \cdot TCPA)||$$

DCPA와 TCPA는 각각 충돌상황의 공간적 요소와 시간적 요소를 분리하여 표현한다는 특징이 있다. 예를 들어, 자선과 타선이 서로 정면을 바라보고 접근하고있는 Head on 상황의 경우, 충돌까지 걸리는 시간과는 무관하게 DCPA는 항상 0에 가까운 값을 갖는다. TCPA의 경우, 자선과 타선의 상대거리와는 무관하게 최근접점에 도달하는데 걸리는 시간에 의해 결정된다. Fig.3.9은 이러한 현상에 대하여 도식적으로 표현하고 있다. 선박의 충돌회피는 시간적 요소와 공간적 요소 동시에 고려해야 하므로 DCPA와 TCPA를 모두 고려하여 충돌위험도를 추정하는 것이 일반적이다.

자선과 타선간의 충돌 상황에 대한 위험도는 다양한 원인 변수에 의해 결정된다. 이들 변수들을 나열해보면, 선박간의 상대거리, DCPA, TCPA, 조우상황, 상대

속도, 속도비율, 환경하중, 선박의 조종성능, 제한수역 여부, 시계 등을 예로 들수 있다. 이러한 변수들을 모두 고려한 충돌위험도 모델을 구성하는 것은 매우 복잡하고 어려운 작업이기에, 충돌위험도를 단순화하는 연구들이 진행되었다. Kearon (1977)은 충돌위험도에 영향을 미치는 다양한 인자중 DCPA와 TCPA를 고려한 충돌위험도 모델을 제시하였다. Kearon의 모델에서는 DCPA와 TCPA의 제곱에 가중치를 곱하는 형태로 충돌위험도를 판단하였다. Kearon의 모델과 유사한 충돌위험도 추론 방법으로는 Imazu and Koyama의 모델과 Smeaton and Coenen의 모델이 있다. 이들 모형은 Kearon의 모형과 유사하게 DCPA 또는 TCPA의 다항식으로 충돌위험도 모형이 구성되어있지만, 자선과 타선의 상대거리에 대한 항을 추가하여 두 선박간 반드시 유지해야하는 최소거리의 개념을 도입하였다.

최근접점을 이용한 충돌위험도에 대한 추론 기법중 대표적인 연구는 Iwasaki and Hara (1986)가 제안한 퍼지추론(fuzzy inference)를 이용한 기법이다. 앞서 설명한 기법들은 DCPA나 TCPA에 대한 다항식을 통해 충돌위험도를 추론한 반면에, Iwasaki and Hara (1986)는 DCPA와 TCPA를 입력변수로 갖는 퍼지추정자를 통해 충돌위험도를 추정하였다. Hasegawa and Kouzuki (1987)[18]는 퍼지추론을 위한 입력변수 DCPA와 TCPA의 소속함수를 결정하기위해 숙련된 운항자와의 인터뷰 결과를 이용하였다. Lee and Rhee (2001)는 Hasegawa and Kouzuki (1987)의 충돌위험도 추론 기법을 이용하여 선박의 충돌회피 시뮬레이션을 수행한 바 있는데, Fig.3.10은 Lee and Rhee (2001)의 연구에서 사용한 DCPA와 TCPA에 대한 충돌위험도의 추론결과이다. Fig.3.10에서 확인할 수 있듯이 Hasegawa and Kouzuki의 모델은 DCPA와 TCPA에 대하여 선형적으로 증가하는 현상을 확인할 수 있는데, 이는 기존의 연구에서 고려되었던 비선형성을 반영하지 못한 결과였다. 또한, Hasegawa and Kouzuki의 모델은 선박마다 상이한 조종성능의 영향이나

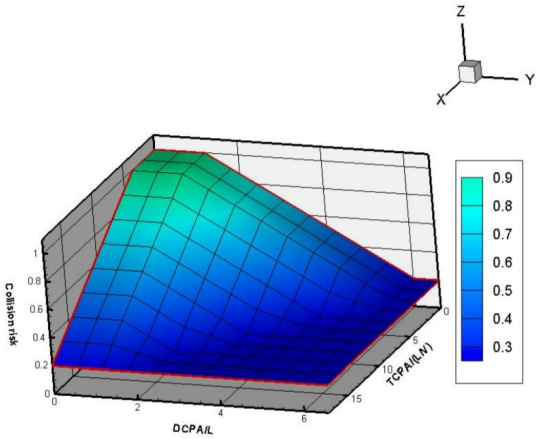



Fig 3.10: Collision risk estimation result using Lee and Rhee. Figure from [49]

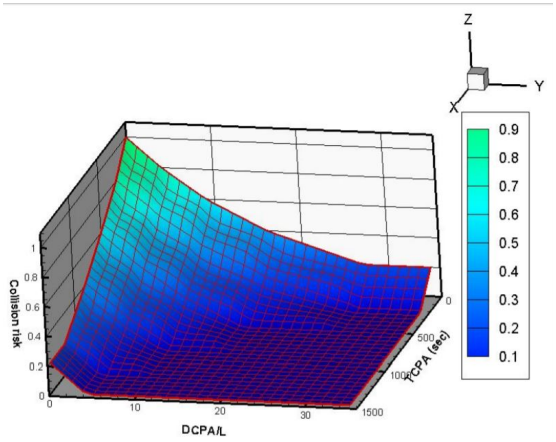



Fig 3.11: Collision risk estimation result using Ahn et al. Figure from [49]

운항환경에 대한 요소들을 고려하지 못한다는 기법적 한계가 존재하였다. Ahn et al (2012)[1]은 기존 기법에 대한 이러한 한계를 인식하고, 선박의 조종성능을과 운항환경요소를 고려하는 모델을 제안하였다. Ahn et al은 조종운동시뮬레이션을 통해 DCPA와 TCPA의 소속함수를 결정하였고, 제한수역 여부 및 시계 등의 요소를 신경회로망의 입력요소로 반영하여 충돌위험도의 추론과정에 포함시켰다.

Fig.3.11은 Ahn et al의 기법을 이용하였을 때 DCPA와 TCPA에 대한 충돌위험도의 추론결과를 나타내고있다.

### 3.2.2 속도 장애물 기법

속도 장애물 기법은 로봇공학 분야에서 장애물과의 충돌회피를 위한 행동을 결정하는데 흔히 사용되는 기법으로, 1989년 Tychonievich가[41] 선박의 운항을 위한 조종판(maneuvering board) 의 개념을 제안한 이후, 충돌 원뿔(collision cone) 기법[7], 속도 지도(velocity map)기법[11]등의 이름으로 지속적으로 개선되고있는 기법이다. 최근들어 Stenersen (2015)[37], Myre (2016)[29], Kuwata et al (2014)[23]에 의해 무인수상선의 충돌회피 분야에도 적용된 바 있다.

속도 장애물 기법에서 속도 장애물(velocity obstacle) $VO$는 자선이 선택할 수 있는 속도 중, 움직이는 장애물의 상대속도를 고려하였을 때, 충돌을 야기시키는 속도의 집합을 의미한다.[23] 이를 수학적으로 표현하면 식 3.2 와 같다.

$$VO_B^A(\vec{v}_B) = \{\vec{v}_A \mid \lambda(\vec{p}_A, \vec{v}_A - \vec{v}_B) \ \cap \ (\mathcal{B} \oplus -\mathcal{A}) \neq \varnothing\} \tag{3.2}$$

이때, 식 3.3는 $p$의 위치에서 $v$의 방향으로 향하는 직선으로 정의된다.

$$\lambda(p, v) = \{p + tv \mid t \geq 0\} \tag{3.3}$$

식 3.2에 사용된 두개의 연산(operation)은 다음과 같이 정의된다.

$$\textbf{Minkowski 합 연산} \ : \quad \mathcal{A} \oplus \mathcal{B} = \{a + b \mid a \in \mathcal{A}, b \in \mathcal{B}\} \tag{3.4}$$

$$\textbf{반사(reflection) 연산} \ : \quad -\mathcal{A} = \{-a \mid a \in \mathcal{A}\} \tag{3.5}$$

식 3.2에서 자선과 장애물의 형상을 원판 형태로 가정하면(Fig.3.12), 식 3.2는 식 3.6으로 정리될 수 있다. 이때, $r_{AB}$는 자선과 장애물의 반지름의 합으로

정의되고, $D(x, r)$은 중심이 $x$ 이고, 반지름이 $r$인 원판을 나타낸다.

$$VO_B^A(\vec{v}_B) = \{\vec{v}_A \mid \lambda(\vec{p}_A, \vec{v}_A - \vec{v}_B) \ \in D(\vec{p}_B - \vec{p}_A, r_{AB}) \tag{3.6}$$

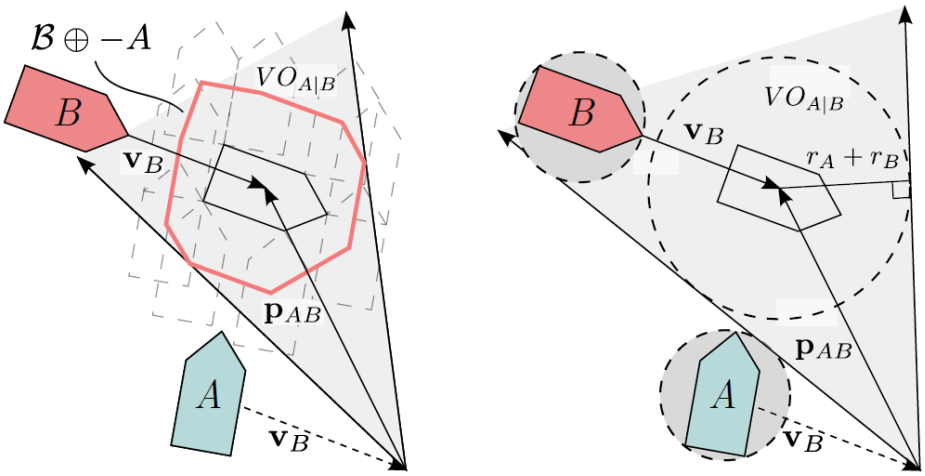



Fig 3.12: Constructed velocity obstacle using (a)Minkowski sum and (b) under assumption of disk shape. Figure from [37]

식 3.2 또는 3.6 을 적용할 경우, 타선에 대한 자선의 속도 장애물 집합이 계산된다. 이는 Fig.3.12와 같은 원뿔 형태의 위험영역으로 표현할 수 있는데, 속도장애물 기법의 정의에 따라, 만일 자선이 위험영역에 해당하는 속도를 선택하고 유지하면 이는 곧 장애물과의 충돌야기시키는 방향으로 운동하고 있음을 의미한다. 반대로, 장애물과의 충돌회피는 자선의 속도를 선택할 때 위험영역 밖에 있는 임의에 속도를 선택함으로써 이루어 질 수 있다. Fig.3.13은 속도장애물 영역과 충돌회피를 위해 자선이 선택할 수 있는 속도집합 $\mathcal{V}_1$, $\mathcal{V}_2$, $\mathcal{V}_3$ 을 나타내고 있다. 각각의 속도집합은 수식 3.7에 의해 정의된다.

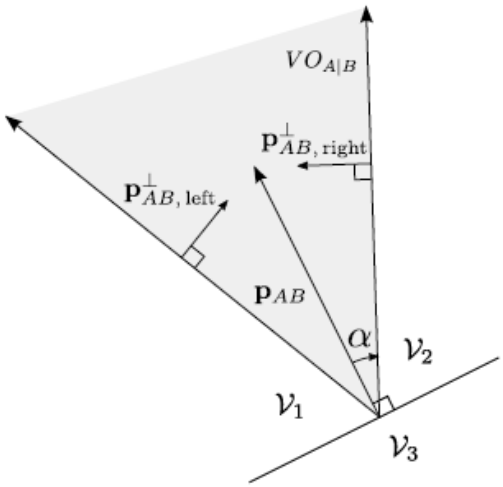



Fig 3.13: Velocity obstacle region and selectable velocity space set. Figure from [37]

$$
\begin{aligned}
\mathcal{V}_1 &= \{v_A \mid v_A \ \notin \ VO^A_B \ \cup \ \mathcal{V}_3 \ \wedge \ [\ p_{AB} \times v_{BA}]_{-}z < 0\}, \\
\mathcal{V}_2 &= \{v_A \mid v_A \ \notin \mathcal{V}_1 \cup \mathcal{V}_3 \cup VO^A_B\}, \\
\mathcal{V}_3 &= \{v_A \mid p_{AB} \ \cdot \ v_{BA} \ < \ 0\}
\end{aligned}
\tag{3.7}
$$

이때, 각각의 속도집합은 고유한 회피행동의 특성을 갖는데, $\mathcal{V}_1$에 해당하는 속도집합은 자선의 시점에서 장애물의 왼쪽을 통과하며 회피하는 행위를, $\mathcal{V}_2$는 장애물의 오른쪽을 통과하며 회피하는 행위를, $\mathcal{V}_3$는 장애물을 통과하지 않고 장애물로부터 멀어지는 회피행위를 의미한다. 선박간의 충돌회피에서는 대부분의 조우상황에서 선박의 좌현변침을 통한 충돌회피를 원천적으로 금지하고 있기 때문에, 자선의 속도를 선정하는 단계에서 이를 고려하여 $\mathcal{V}_1$를 제외하고 속도를 선택할 수 있다.

Kuwata et al (2014) [23]는 속도장애물을 구성하는 단계에서 충돌위험도가 높

지 않은 장애물에 대한 불필요한 회피행위를 방지하기 위하여 식 3.8을 만족하는 장애물에 대해서만 속도장애물을 생성하였다. 식 3.8은 회피행동을 수행하기위한 대상 장애물과의 최소한의 충돌위험도 조건을 의미한다. 이때, $TCPA_i$는 장애물 $i$에 대한 최단접근시간을, $DCPA_i$는 장애물 $i$에 대한 최단접근거리를 의미한다.

$$\begin{aligned} 0 \leq TCPA_i \leq TCPA_{max} \\ DCPA_i \leq DCPA_{min} \end{aligned} \tag{3.8}$$

# 제 4 장 심층강화학습 기반 충돌 회피

4장에서는 본 연구에서 제안하는 심층강화학습 기반 충돌회피 기법에 대하여 구체적으로 설명한다. 일반적으로 무인수상선의 충돌회피 알고리즘은 탐지(detection) - 의사 결정(decision making) - 경로 계획(path planning) - 제어(control)의 4단계로 구성되는데, 본 연구에서는 이 중 의사결정에 해당하는 단계를 효과적으로 수행하기위한 심층강화학습 기반 충돌회피 의사결정 알고리즘에 대한 연구를 제안한다.

탐지 단계에서 무인수상선이 조우하고있는 장애물의 정보가 AIS, Radar, Lidar 등의 센서로부터 획득되면, 이들 정보는 바로 충돌 의사결정 모듈로 전달된다. 전달된 정보는 바로 현재의 조우상황을 표현하는 격자지도(grid map)형태로 가공되어지는데(Fig.4.1의 1번), 이 격자지도에 대해서는 4.1.1절에서 (Fig.4.4)에서 자세히 설명하고 있다. 전달받은 장애물의 운동정보를 바탕으로, 탐지된 장애물 중 충돌의 위험성이 존재하는 장애물을 보수적으로 판단한다.(Fig.4.1의 2번) 이를 위해서, 본 연구에서는 [23]에서 제안하였던 DCPA와 TCPA를 이용한 방법(식 3.8)을 사용하였다. 만일 탐지된 장애물 중 식 3.8에 해당하는 장애물이 존재한다면, 이 정보는 강화학습 기반 충돌회피기동 전환기(Fig.4.1의 3번)에 전달된다. 충돌회피기동 전환기 모듈은 자선과 타선간의 충돌위험성이 존재하는 상황에서 충돌회피 여부, 회피방향, 그리고 회피시점을 결정하는 모듈이다. 일반적으로 동적장애물에 대한 충돌회피는 대부분의 경우의 우현변침을 통해 회피가 이루어지지만, 동적 장애물 중 Overtaking 조우상황이나 정적장애물에 대한 장애물 회피는 우현변침과 좌현변침중 유용한 방법을 선택할 수 있도록 구성하였다. 따라서, 충돌회피기동

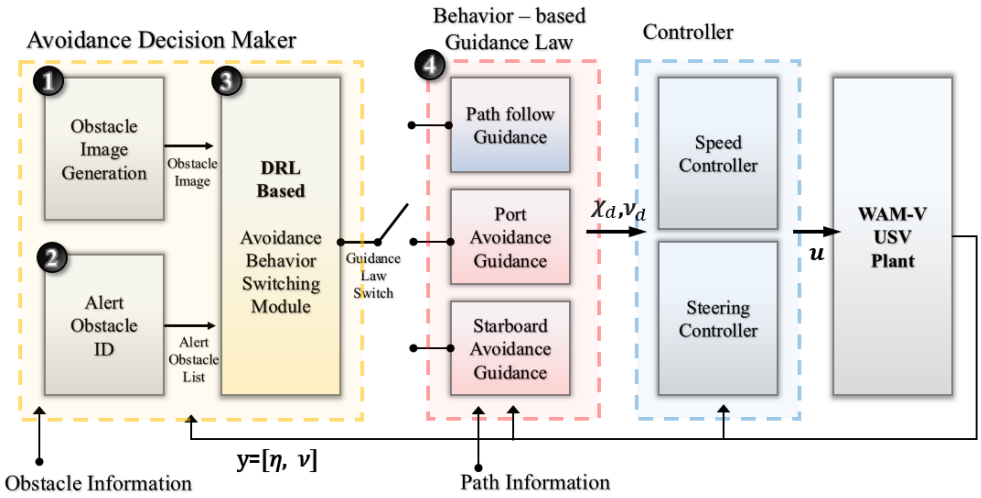



Fig 4.1: Schematic diagram of the deep reinforcement learning based collision risk assessment system

전환기(Fig.4.1의 3번)는 총 3가지의 행동기반 유도기를 선택할 수 있도록 구성되어있다. 먼저, 충돌회피를 수행하지 않고 현재의 경로를 추종하는 유도기, 좌현변침을 통해 충돌회피기동을 수행하는 유도기, 우현변침을 이용하여 충돌회피를 수행하는 유도기 로 구분된다. 각각의 유도기에 대한 설명은 다음 절에서 자세히 기술하고 있다. Fig.4.1에서는 본 연구에서 제안하고자 하는 심층강화학습 기반 충돌회피 의사결정 알고리즘의 개략적인 구성도를 나타내고있다.

## 4.1 심층강화학습 신경망

본 연구에서 제안하는 충돌회피 의사결정 모듈은 Minh et al (2015)[28]이 제안한 심층강화학습을 기반으로 구성되어져 있다. 본 절에서는 심층강화학습을 구성하는 세미-마코프 의사결정 과정의 구성에 대한 설명과, 심층강화학습을 구성하는 신경회로망의 학습에 대하여 설명하고 있다.

### 4.1.1 SMDP의 구성

세미 마코프 의사결정 모델(Semi-Markov Decision Process, SMDP)은 1999년 Sutton et al[38]이 제안한 개념으로, 기존의 강화학습에 사용되었던 의사결정 모델인 마코프 의사결정모델에 사용되는 행동(action) 의 개념을 시간적으로 확장(temporally extend)한 개념이다.

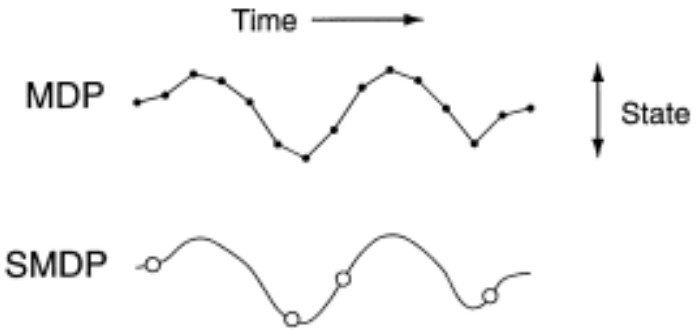



Fig 4.2: Concept of semi-markov decision process

Fig.4.2에서 확인할 수 있듯이, MDP의 경우 학습자가 환경에 작용하는 행동이 매 시점(time step)에 대하여 이루어지는 하위단계(low-level)의 행동이지만, SMDP의 경우 보다 복잡한 상위단계(high-level)의 작업이 이루어진다. 이를 무인수상선에 대입하여 설명하면, MDP에서 사용하는 행동(action)변수는 매 순간순

간의 침로각 명령(course angle command)으로 표현할 수 있는 반면, SMDP에서 정의하는 행동은 경로추종, 우현 변침을 통한 회피, 좌현 변침을 통한 회피 등의 보다 복잡하고 시간적으로 확장된 개념의 행동기반 명령을 사용한다. 본 연구에서는 이러한 SMDP의 행동을 정의하기위해 각종 기동을 수행하기위한 유도법칙(guidance law)를 선정하였고, 이를 강화학습의 행동변수로 사용하였다.

**상태(state)의 표현**

심층강화학습 기반 충돌회피 의사결정 모듈이 회피여부 또는 회피방향에 대한 판단을 수행하기 위해서는 현재 무인수상선이 처해있는 장애물과의 조우상황에 대한 충분한 정보를 전달받아야 한다. 무인수상선의 충돌회피에 필요한 정보는 임무수행을 위해 추종하는 경로에 대한 정보, 장애물의 상대위치 및 운동정보, 환경하중 정보, 현재 무인수상선의 속도정보 등이 있다. 본 연구에서는 이들 정보 중 무인수상선과의 기하학적 관계로 표현되는 정보들에 대해서는 격자지도(grid map)로 형태로 상태변수를 가공하여 심층강화학습의 입력변수로 사용하였다. 이에 해당하는 정보들은 다음과 같다.

- 무인수상선이 추종하는 경로정보
- 동적(dynamic) 장애물 정보
- 정적(static) 장애물 정보

이들 장애물의 정보를 격자지도 형태로 구성하는 이유는, 경로 정보 및 장애물 정보는 근본적으로 선박과 해당 정보와의 기하학적 관계가 의미를 갖는 정보들이기 때문이다. 인간이 자선과 타선간의 충돌위험성 여부를 판단할 때, 저차원(low dimension)의 방위각과 거리정보 (예를들어, 방위각 6도, 0.5마일) 보다 Fig.4.3과

같은 레이더 화면에 의존하는 것도, Fig.4.3과 같은 고차원(high dimension)의 격자지도 형태의 정보가 자선과 타선가의 조우상황을 판별하기 위한 보다 직관적(intuitive)인 정보를 제공하기 때문이다.

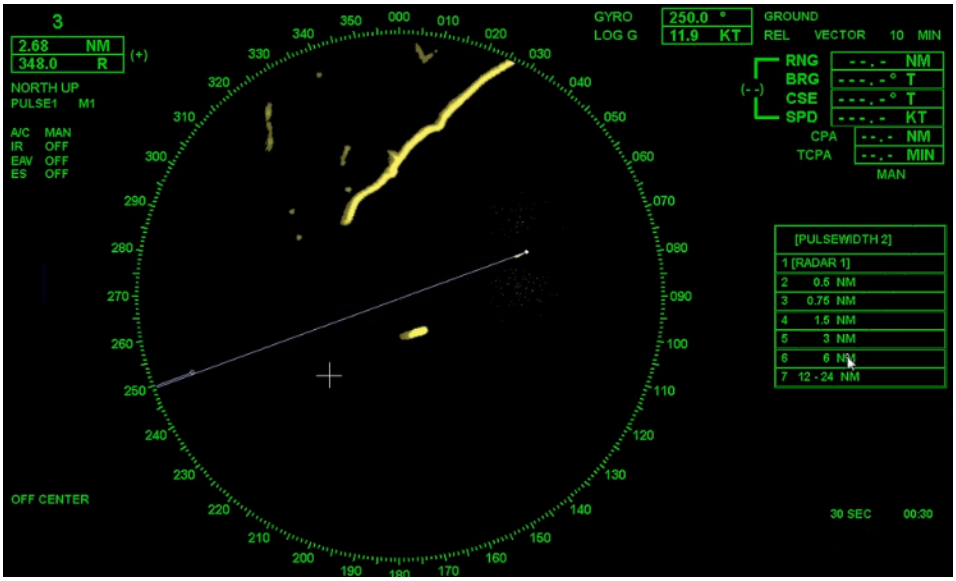


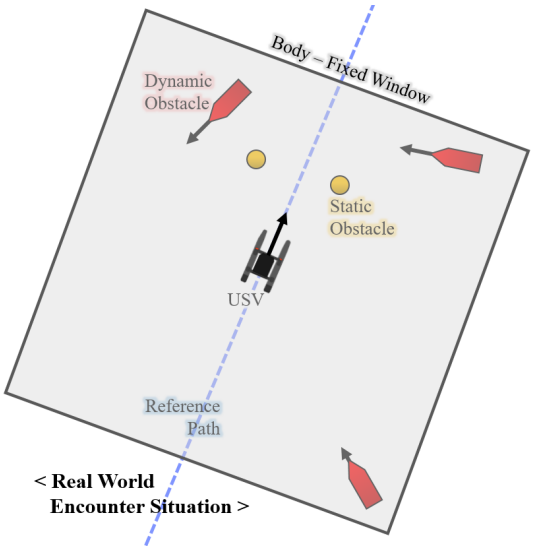

Fig 4.3: Operation display screen of Furuno 3 Lucero radar plotting system

격자지도를 이용하여 무인수상선이 처해있는 조우상황을 표현하는 방법에 대해서는 Fig.4.4에서 자세하게 설명하고 있다. Fig.4.4a가 무인수상선이 처해있는 조우상황이라면, 이와같은 조우상황을 본 연구에서는 Fig.4.4b와 같은 3층(layer)의 (84×84 픽셀) 격자지도 형태로 표현하였다. 각 층은 경로, 동적장애물의 위치와 속도, 정적장애물의 상대위치를 의미하고 있다. 조우환경은 Fig.4.4a에 보이는 body-fixed window (336×336 미터)내에 있는 장애물 및 경로에 대한 정보를 표현하고있는데, 이 창은 무인수상선을 중심에 두도록 좌표계가 설정되어있고, 무인수상선에 따라 함께 회전하는 좌표계에 의해 정의된다. 동적 장애물 격자지도의 경우, 장애물의 위치와 현재 장애물의 속도를 함께 표현하고있는데, 장애물의 속도

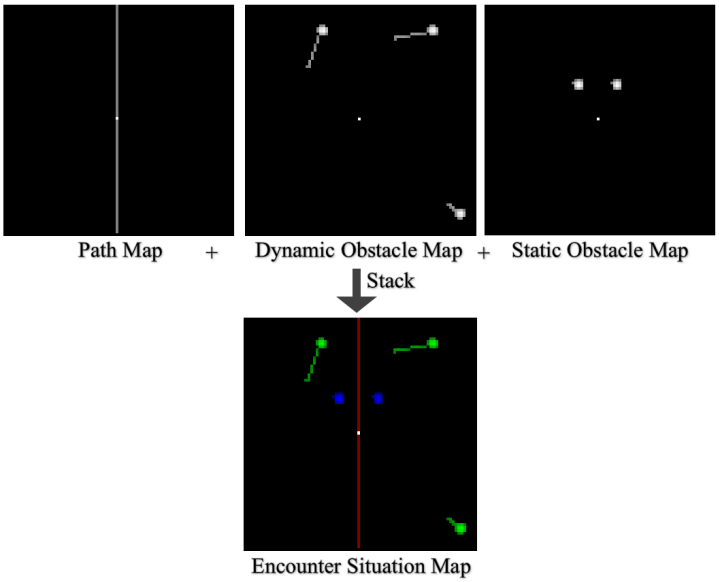
는 장애물의 중점으로부터 시작되는 직선의 길이에 비례하도록 표현하였다.이는, Fig.4.3의 레이더 체계의 운용화면에서 장애물의 속도를 표현하는과 방법을 차용한것으로, 조우상황 격자지도를 통해 장애물의 속도를 동시에 인식할 수 있도록 구성하였다. 각각의 격자지도는 84×84 픽셀의 크기를 갖는 행렬로 표현된다. 격자지도 중 경로를 나타내는 경로지도(path map)는, 경로에 해당하는 영역이 가장 밝게 표현되고(밝기 255), 그 이외의 영역은 0의 밝기로 표시되도록 구현하였다. 경로의 형상은 3픽셀의 굵기로 표현되고, 경로로부터 떨어질수록 어두운 밝기로 표현된다. 동적장애물 지도는 움직이는 장애물의 정보를 밝기로 표현하는 격자지도로, 장애물이 존재하는 위치의 영역이 밝게 표현된다. 장애물의 크기는 지름이 5픽셀인 원형으로 표현되고, 장애물의 위치에 가까울수록 밝게 표현된다. 장애물의 속도는 장애물에서 운동방향으로 그려지는 직선의 길이로 표현된다. 직선의 두께는 1픽셀이고, 길이는 최대 30픽셀까지 장애물의 속도에 선형비례하도록 구현하였다. 정적 장애물은 동적 장애물과 동일하게 표현된다. 장애물의 지름은 5픽셀이고, 장애물의 위치에 가까울수록 밝게 표현된다.

격자지도형태로 장애물의 위치를 표현하는경우 장애물의 크기나 형상에 대한 고려를 동시에 할 수 있다는 장점이 존재한다. 하지만 본 연구에서는 문제의 단순화를 위해 동적 및 정적 장애물을 크기가 고정된 단순한 원형의 장애물로 가정하고 상태를 표현하였다.

충돌회피에 사용되는 정보 중 격자지도 형태로 표현할 수 없는 변수들은 추가적인 저차원 상태벡터를 구성하여 강화학습의 입력변수로 사용하였다. 이에 해당하는 정보는 현재 무인수상선 자선의 속도정보, 각 추진기의 RPM정보, 무인수상선이 받고있는 환경하중 정보이다. 환경하중은 조류 및 풍하중의 영향을 고려하였는데, 이때 조류나 풍하중의 정보(입사각, 하중의 속도)을 바로 입력하는 방식이




(a) Real world encounter situation of ASV




(b) Composition of the encounter situation map

Fig 4.4: Encounter situation map representation

아닌 무인수상선이 환경 하중에 의해 편향된 상태값을 계산하고, 이를 강화학습의 입력변수로 사용함으로써 고려하였다. 편향값을 계산하기 위해서는 5.1.2절에서 설명할 무인수상선의 동역학 모델과 이전 시점에서의 제어명령을 이용하여 전방 시뮬레이션(forward-simulation)을 통해 얻은 상태변수값과 현 시점에서 계측을 통해 얻은 상태변수와의 차이를 계산하여 얻는다. Fig.4.5은 환경하중에 의해 편향된 상태값을 얻는 방법에 대하여 설명하고 있다.

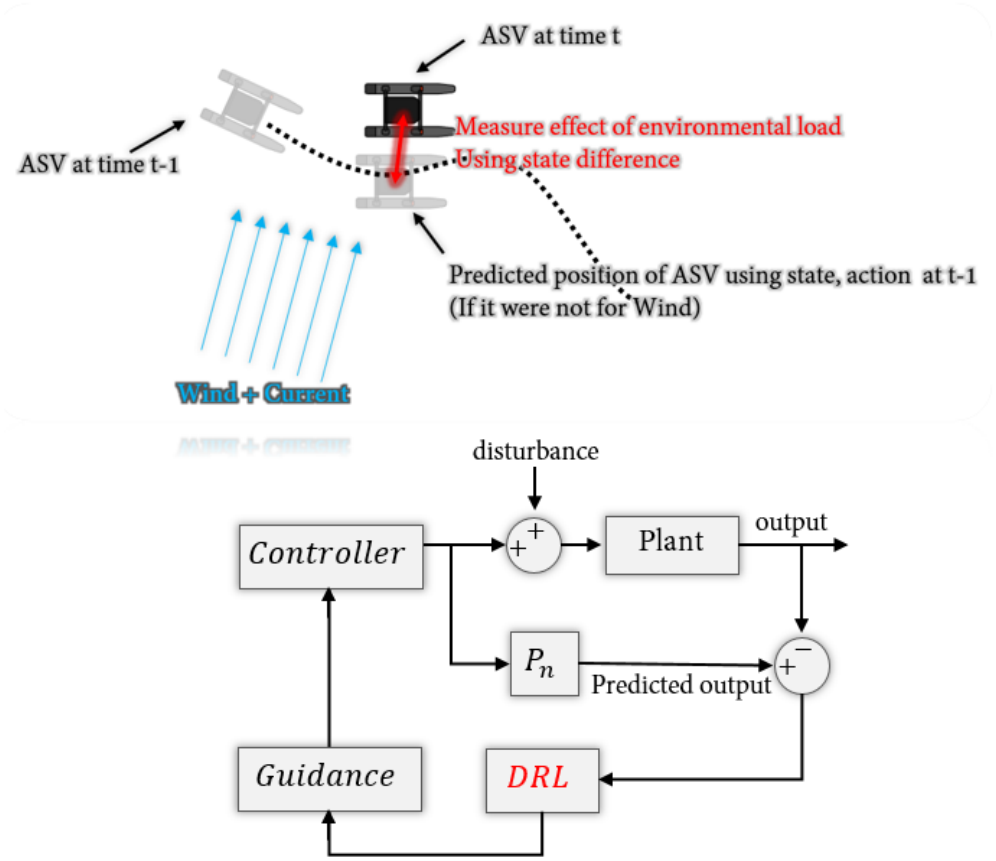



Fig 4.5: Environment load estimation method for deep reinforcement learning input

**행동 변수**

본 연구에서 제안하는 충돌회피 의사결정 시스템은, 무인수상선이 현재 처해있는 조우상황을 입력받아 무인수상선이 따라야 하는 유도명령의 종류를 출력값으로 반환하는 형태로 이루어져있다. 이는 곧 무인수상선이 따라야 하는 유도기를 전환(switch)하는 역할을 수행하게 된다. 본 연구에서 사용하는 유도기는 총 3가지 종류로 다음과 같은 역할을 수행한다.

- 무인수상선의 경로추종
- 좌현변침을 통한 장애물 회피
- 우현변침을 통한 장애물 회피

**경로 추종**을 위해서는 본 연구에서는 벡터 필드 유도기(vector field guidance, 이하 VFG)를 사용하였다. VFG는 직선 또는 원형경로를 따라가기위해 무인운동체가 추종해야하는 선수각정보를 사전에 벡터필드 형태로 구성하고, 현재 무인운동체의 위치에 따른 벡터방향을 목표 선수각으로 사용하는 유도기법이다. Fig.4.6은 무인수상선의 직선 경로 추종을 위해 정의된 벡터필드에 대하여 표현하고 있다. 수식 4.1은 직선경로 추종을 위해 VFG를 사용하였을 때 계산되는 목표침로각에 대한 수식을 나타내고있다. 이때, $y$는 경로와 무인수상선간의 수직 오차(cross track error)를 의미하고, $\chi^{\infty}$는 목표침로각의 범위를 정하는 계수를, $k$는 침로각의 변화속도를 조절하는 파라미터이다. $\chi_{path}$는 관성좌표계에서 정의된 직선경로의 각도로 정의된다.

$$\chi^{d}(y) = \chi^{\infty}\frac{2}{\pi}tan^{-1}(ky) + \chi_{path} \tag{4.1}$$

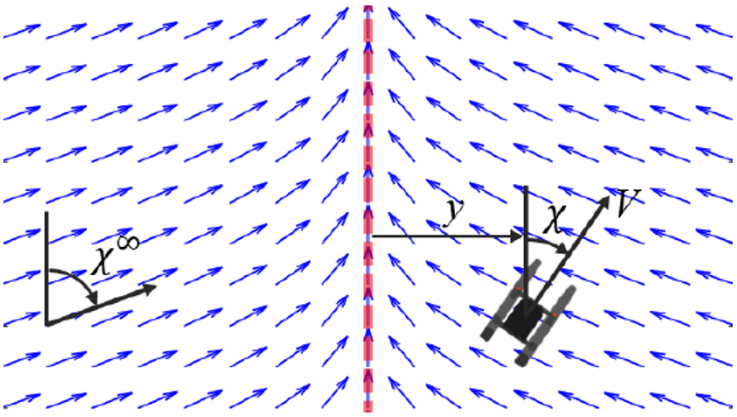



Fig 4.6: Conceptual description of the vector field guidance method for linear path follow guidance

**무인수상선의 장애물 회피**를 위해서는 3.2.2절에서 설명한 속도장애물 기법을 이용하였다. 속도장애물 기법의 경우, 장애물의 상대위치와 속도에 따른 충돌원뿔(collision cone)을 구성하고, 이 원뿔 밖의 영역에 해당하는 속도를 선택함으로써 충돌을 회피하는 기법이다. 본 연구에 사용된 충돌회피 행동 유도기는 회피방향에 따라서 구분되는데, 좌현변침 또는 우현변침에 대한 행동유도기는 이는 충돌원뿔을 기준으로 좌현방향 또는 우현방향에 금지속도 영역(prohibited velotic region)을 설정함으로써 정의하였다. Fig.4.7은 이에 대한 도식적인 설명을 나타내고있다. 회피를 위한 속도 및 침로각을 선택할 때, 속도장애물에 해당하는 속도와 더불어 금지속도 영역에 대한 속도를 배제시킴으로써 단방향에 대한 회피를 수행하는 유도기를 구성할 수 있다.

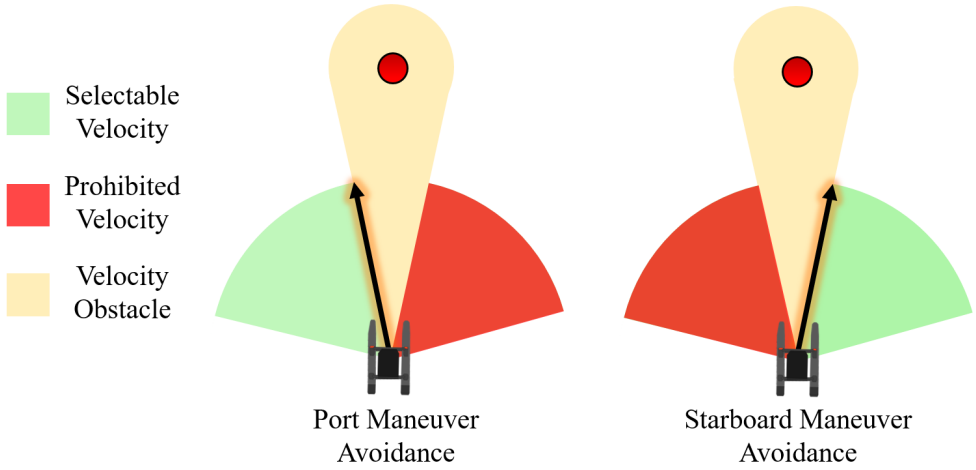



Fig 4.7: Collision avoidance guidance method using Velocity Obstacle method, with respect to avoidance direction

### 보상 함수

심층강화학습 알고리즘의 경우 학습자와 주변환경이 상호작용을 하며 학습이 이루어지는데, 이때 현재 학습자가 수행한 행동이 얼마나 바람직한 행동인지의 여부를 보상(reward)값을 이용하여 제공해주어야한다. 심층강화학습 기반 충돌회피 의사결정 기법의 학습을 위해 사용된 보상함수는 다음의 3종류의 보상함수로 구성된다.

- 경로 추종에 따른 양(positive)의 보상
- 충돌 회피 성공에 따른 양(positive)의 보상
- 충돌에 따른 음(negative)의 보상

무인수상선이 불필요한 회피기동을 수행하거나 지나치게 이른 시점에 회피기동을 하는 것을 방지하기위해서, **경로추종에 대한 양의 보상값**을 제공해주는 것이

필요하다. 경로추종에 대한 양의 보상값은 경로에 대한 수직추종오차 (cross track error)가 0이 될 경우에 최대값을 갖도록 식 4.2와 같은 형태로 구성하였다. 경로추종에 대한 보상값은 현재 무인수상선이 선택한 행동이 경로추종과정인 경우에만 전달된다. 식 4.2의 $k_{r_{path}}$는 경로추종 보상값의 민감도를 조절하는 계수이고, $e_{cross}$는 경로에 대한 수직추종오차를 의미한다.

$$r_{path} = \exp\left(k_{r_{path}} \cdot abs(e_{cross})\right) \tag{4.2}$$

**충돌 회피를 성공적으로 수행하였을 때, 양의 보상값**을 전달받도록 세미-마코프 의사결정과정을 구성하면 무인수상선이 충돌에 임박했을 경우, 경로추종 행동를 포기하고 충돌회피 행동을 수행하도록 만들 수 있다. 이때, 중요한 점은 충돌회피 시점에 따라서 다른값의 보상값을 제공해주어야 한다는 점인데, 이는 각 조우상황, 환경하중상황, 선박의 종류에 따라 바람직한 충돌회피시점이 다르기때문에 학습을 통해 이들 시점을 스스로 파악하도록 적절한 보상값을 전달해주어어야한다. 충돌회피에 대한 보상함수는 앞서 3.1.3절에서 정의한 각 조우상황별 회피전략을 기반으로, 바람직한 회피기동을 수행하였을 때 최대값의 보상값이 전달되도록 구성되었다. 충돌 회피의 성공적 수행에 따른 보상값은 무인수상선이 충돌회피에 대한 행동을 하였을 경우(좌현 또는 우현변침을 통한)에만 제공되고, 경로추종을 수행하였을 때는 제공되지 않는다.

- Head on 조우상황

  Head on 조우상황의 경우, 회피시점결정에서 중요하게 고려해야 할 요소는 타선이 자선의 회피기동에 대해 가관측성(observability)을 가질 수 있도록 명백한 회피행위를 수행하는 것이다.(COLREGs 8조 (b)에 의거) VO알고리즘을 이용하여 충돌회피를 수행할 때, 지나치게 이른 시점에 충돌회피를

수행하면 회피기동여부가 뚜렷하게 구분되지 못하는 모호한 회피기동을 수행하는 특성이 존재하는데,(낮은 가관측성) Myre (2016)[29]는 이러한 기법적 특성이 COLREGs에 위배된다고 지적한 바 있다. 따라서 본 연구에서는 무인수상선의 회피시점을 현재 추종하는 침로각을 기준으로 30도 이상의 변침이 발생하는 시점을 Head on조우상황의 충돌회피 시점으로 정의하였다. 이를 위해 학습자에게 제공되는 보상값은 식 4.3과 같다. 이때, $\hat{\chi}_{avoid}$는 충돌회피를 수행하였을 때, 기존 추종각을 기준으로 계산되는 최대변침각을 의미한다. $k_{r_{avoid}}$는 최대변침각도 오차에 대한 민감도를 설정하는 계수이다.

$$r_{avoid} = \exp\left(k_{r_{avoid}} \cdot abs(\hat{\chi}_{avoid} - 30)\right) \tag{4.3}$$

- Crossing (stand on) 조우상황

  Crossing (stand on) 조우상황에서는 자선이 진로유지선의 지위를 가지고있으므로, 충돌회피시점을 최대한 미루고, 타선의 피항행위 여부를 판단하는 과정이 필요하다. 따라서, Head on과 유사한 형태의 보상값 수식의 형태를 갖지만 최대 변침각도의 범위를 75도로 정의하였다. 이는, COLREGs 17조 (b)에서 의미하는 피항의무선이 피항의 의무를 수행하지 않아 자선의 피항을 수행해야하는 시점에 대한 정량적 수치를 의미한다. Crossing (stand on) 조우상황의 보상함수는 다음과 같이 결정된다.

$$r_{avoid} = \exp\left(k_{r_{avoid}} \cdot abs(\hat{\chi}_{avoid} - 75)\right) \tag{4.4}$$

- Crossing (give way) 조우상황

  Crossing (give way) 조우상황에서는 자선이 피항의무를 가지고 있고, 타선은 자선의 피항여부를 주시하며 침로유지여부를 판단하고 있는 상황이다.

때문에, Crossing (give way) 조우상황에서는 최대한 빠르게 회피기동을 수행하여 타선에 회피의사를 표명하는작업이 필요하다. 심층강화학습에 전달되는 장애물의 경우, DCPA와 TCPA가 임계값 이내인 어느정도 위험성이 존재하는 장애물 인데, (Fig.4.1의 2번 모듈로부터) 만일 장애물이 Crossing (give way)조우상황으로 판단되면, 자선은 즉시 회피기동을 수행하여 회피의사를 밝히도록 보상함수를 다음과 같이 결정하였다.

$$r_{avoid} = \exp(k_{r_{avoid}} \cdot abs(TCPA_{now} - TCPA_{Thresh})) \tag{4.5}$$

- Overtaking 조우상황

  Overtaking의 경우, Crossing (give way) 조우상황과 동일하게 자선이 피항의무를 가지고있으며, 타선은 진로유지를 하며 자선의 피항행위 여부를 관측하고 있으므로, 최대한 빨리 회피기동을 수행하여 회피의사를 밝히는것이 중요하다. 따라서, Crossing (give way)의 경우와 동일하게, 임계 TCPA, 위험대상 장애물이 발견되는 즉시 회피기동을 수행한다. 보상함수는 Crossing (give way) 의 보상함수 식 4.5를 사용한다.

- 정적 장애물에 대한 회피

  정적 장애물의 경우, Head on과 유사한 형태로 자선이 회피기동을 수행하면서 최대 변침각이 45도 이상 일어나는 시점을 회피시점으로 결정하였다. 이는 수식으로 다음과 같이 정의된다.

$$r_{avoid} = \exp\left(k_{r_{avoid}} \cdot abs(\hat{\chi}_{avoid} - 45)\right) \tag{4.6}$$

  정적 장애물의 경우, 동적 장애물과는 다르게 좌현 또는 우현방향으로 회피방향을 선택할 수 있다는 차이가 존재한다. 회피방향 선택은 강화학습의

가치함수값을 비교하여 더 높은 가치함수값을 갖는 회피방향으로 회피기동을 수행하도록 구성하였다.

**충돌에 따른 음의 보상값**은 무인수상선이 장애물과 충돌하였을경우 최우선적으로 고려하는 보상값으로, 만일 무인수상선이 장애물과 사전에 정해놓은 임계거리값 이하로 가까워지면, -1(최소 보상값)의 보상값을 갖도록 정의하였다. 충돌에 대한 보상값은, 현재 무인수상선이 수행하고있는 행동과 상관없이 충돌에 따른 보상값으로 현재의 보상값이 결정되며, 보상을 받는순간 해당 학습 에피소드를 종료한다.

### 4.1.2 심층강화학습 알고리즘의 동작과정

심층강화학습 알고리즘의 동작과정은 Mnih et al[28]가 제안한 Deep Q Network, 또는 Wang et al[43] 이 제안한 Dual DQN 구조와 유사한 방식으로 작동한다. 이들 기법 모두 2장에서 설명한 Q-learning을 기반으로 최적의 행동을 결정하는 기법이므로, 특정 상태가 입력되었을때, 해당 상태에서 선택할 수 있는 모든 행동의 행동 가치함수(Q-value)값을 추정하는 역할을 수행한다. 행동 가치함수에 대하여 다시한번 설명하면, 식 4.7와 같이 표현할 수 있다. 행동 가치함수는 결국 특정 상태($s_t = s$)에서 특정 행동($a_t = a$)을 수행하였을 때 얻을것으로 기대되는 반환(return)값을 의미한다. 반환값은, 미래에 얻을것으로 기대되는 보상(reward)값의 총 합을 의미한다. (식 4.8)

$$Q^{\pi}(s,a) = E_{\pi}\{G_t|s_t = s, a_t = a\} = E_{\pi}\left\{\sum_{k=1}^{\infty}\gamma^k r_{t+k+1}|s_t = s, a_t = a\right\} \quad (4.7)$$

$$G_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + .. = \sum_{k=1}^{\infty}\gamma^k r_{t+k+1} \quad (4.8)$$

심층 신경망의 학습 단계에서는 이 행동 가치함수의 학습을 위하여 식 4.9과 같이 Bellman Equation을 사용하여 가치 반복(Value iteration)을 수행하여 행동 가치함수값을 추정한다.

$$Q_{i+1}(s,a) = E\left\{ r + \gamma max_{a'} Q_i(s',a') | s,a \right\} \tag{4.9}$$

행동 가치함수값이 추정되면, 행동 가치함수의 함수근사기를 학습시키기위한 목적함수를 정의한다. 이 목적함수는 목표 행동 가치함수값과 추정된 행동 가치함수값의 차이의 제곱의 기대값으로 식 4.10과 같이 정의된다.

$$L(w) = E\left\{ [r + \gamma max_{a'} Q_i(s',a',w) - Q(s,a,w)]^2 \right\} \tag{4.10}$$

이때, $w$는 함수근사기의 파라미터를 의미하고, 신경망의 학습은, 최적화 기법을 통하여 $L(w)$이 0이 되는 $w$를 찾는것을 목표로 한다.

학습이 완료 될 경우, 다음의 과정을 통해 충돌회피를 위한 유도기를 선택한다. 우선, 매순간 신경망에 현재 상태를 나타내는 격자지도와 저차원(low dimension) 상태값들을 입력한다. 이때, 신경망에서는 선택가능한 행동에 대한 행동 가치함수값이 출력된다. 이들 행동 가치함수값(Q-value) 들을 비교하여 그들 중 행동 가치함수의 값이 가장 큰 유도기법을 선택한다. 이러한 정책을 greedy 정책이라고 표현하는데, 본 연구의 검증단계에서 수행한 자유항주실험에서는 이 greedy 정책을 이용하여 충돌회피를 진행하였다.

### 4.1.3 심층강화학습 신경망 학습

#### 심층강화학습 신경망의 구성

강화학습에 사용된 신경망은 Fig.4.8과 같이 구성되어있다. 이는 Minh et al (2015)[28]이 제안한 Deep Q Network구조에 Wang et al (2016)[43]이 제안한 이

중구조를 결합한 형태이다. 이들 구조에는 고차원의 이미지만을 입력받는데 반해, 본 연구에서는 저차원의 상태값들 (RPM, 속도, 환경하중정보 등)을 벡터 형태로 입력받을수 있도록 구조를 변경하였다. (Fig.4.9)

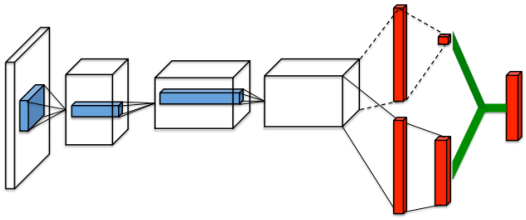


Fig 4.8: Structure of te deep reinforcement learning network by Wang et al (2016), Image from [43]

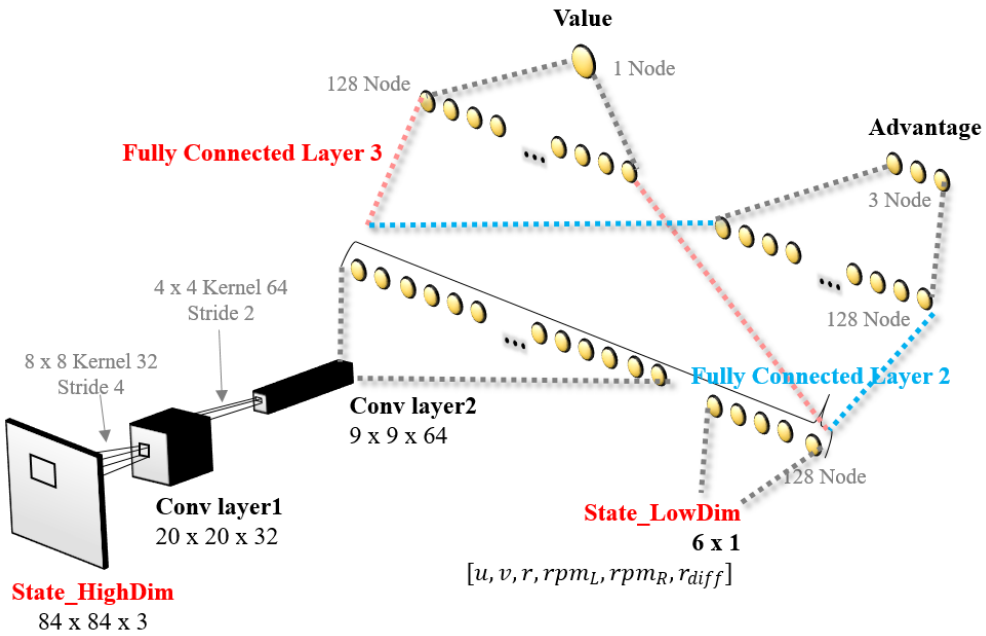



Fig 4.9: Proposed structure of te deep reinforcement learning network

**학습환경의 설정**

본 연구에서 제안하는 충돌의사결정 모듈을 학습시키기위해 강화학습 시뮬레이션을 수행하였다. 실선 무인수상선의 동역학을 식별한 조종운동모델을 이용하여 무인수상선의 시뮬레이터를 구성하였고, (5.1.2절)이를 토대로 강화학습을 진행하였다. 학습된 결과는 추후 실선에 탑재되어 자유항주실험을 통해 충돌회피 성능을 확인할 것이기 때문에, 강화학습은 최대한 실선과 유사한 환경에서 학습이 이루어져야 한다. 본 연구에서는 각종 센서와 컴퓨터간의 시간지연(time delay)현상과 각 센서의 계측주기를 고려하여 무인수상선 시뮬레이터를 구성하고, 이를 통해 심층강화학습을 시뮬레이션 상에서 진행하였다.

**Exploration**

본 연구에서는 세미-마코프 의사결정 모형을 도입하여 무인수상선의 행동(예를들어, 침로각) 대신 행동의 순차적 나열(예를들어, 우현변침 또는 경로추종을 위한 유도기명령 집합)을 선택하는 접근방식을 사용하였다. 또한, 1:1회피 시나리오의 경우 오직 하나의 에피소드에서 한번의 회피기동만이 수행되므로, 기존 강화학습에서 사용하는 방식의 exploration기법을 사용할 경우 회피시점에 대한 균일한 경험을 할 수 없다는 단점이 존재한다. 따라서, 본 연구에서는 충돌회피 시점결정을 위한 새로운 exploration기법을 고안하였다. 그 방법은 아래와 같다.

1. Epsilon-greedy 방법과 동일하게 epsilon값을 선택한다. epsilon값은 학습이 진행됨에 따라 단조 감소하는 형태로 업데이트시킨다.

2. 초기에 장애물이 위험장애물로 분류되는 순간, 자선과의 $TCPA_{init}$를 계산한다.

3. $TCPA_{init}$에 현재의 epsilon값을 곱해 $TCPA_{exp}$를 계산한다.

4. 자선과 타선간의 TCPA가 $TCPA_{exp}$보다 클 경우, 경로추종을 수행하고, $TCPA_{exp}$보다 작아지는순간 장애물 회피를 수행한다. 선택가능한 회피행위가 여러개 일 경우, 랜덤하게 회피행위를 선택한다.

이와같은 방식의 exploration을 수행하면 행동공간에 균일한 경험 뿐만 아니라 시간에 대한 균일한 경험을 수행할 수 있다는 장점이 있다.

**학습 결과**

앞서 설명한 내용들을 기반으로 강화학습 신경망을 학습하였다. 다양한 조우 상황에 대하여 무인수상선의 충돌회피 학습을 진행하기위해 단일선박 회피(Head on, Crossing, Overtaking), 다수선박 회피, 정적장애물과 동적장애물 복합회피 총 6가지 종류의 보조 사건(sub-episode)으로 구성되는 사건(episode)를 구성하여 이에 대한 충돌회피 강화학습을 진행하였다. 선박의 초기 위치와 방향은 주어진 범위내에서 랜덤하게 변화하고, 환경하중의 방향과 크기도 사건에 따라 랜덤하게 설정하여 학습환경을 구성하였다.

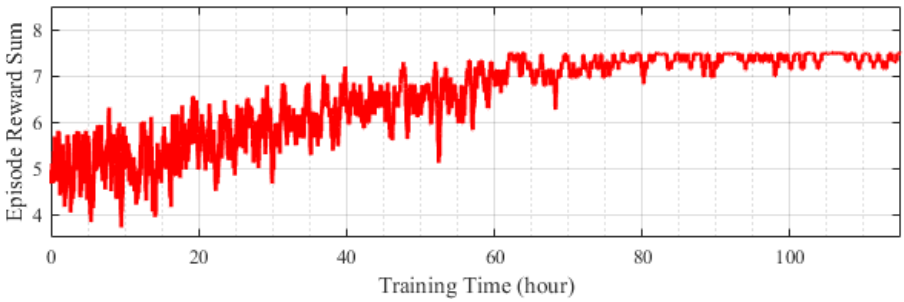



Fig 4.10: Training result of the deep reinforcement learning network

심층강화학습을 통해 충돌회피 시나리오에 대한 학습결과는 Fig.4.10과 같다. 그림상의 x축은 학습이 진행되는 시간을 의미하고, y축은 각 사건의 보상값을 나타낸다. 이 값은 충돌회피가 얼마나 성공적으로 수행되는지를 평가하는 성능지표로 사용된다. 각 사건의 보상값은 6종류의 보조사건의 보상값의 합으로 정의하였다. 학습 결과, 학습이 진행됨에 따라 충돌회피 성능이 증가하는 것을 확인할 수 있고, 특정시점 이후부터는 성능지수의 기울기값이 0에 가까워지며 최종 정책(policy)으로 수렴하는 현상을 확인할 수 있다. 심층강화학습 신경망의 학습은 Tensorflow를 이용하여 수행되었으며, Adam Optimizer를 이용하여 학습이 이루어졌다. 학습에는 약 115시간이 소요되었으며, 학습에 사용된 PC의 사양은 6세대 i7 3.40Ghz의 CPU에 Ram은 16GB, GPU는 GeForce GTX 970 를 사용하였다.

**기존 기법과의 차별성**

앞서 언급하였듯이, 본 연구에서 사용한 심층강화학습의 구조는 Wang et al (2016)[43]이 제안한 구조를 기반으로 하고 있다. 하지만, 무인수상선의 충돌회피라는 특수한 문제에 맞추어 기존 구조를 수정 및 보완하여 사용하였다. 본 연구에서 사용한 심층강화학습의 모델과 Wang et al이 제안하는 강화학습 모델은 크게 3가지 관점에서 차이점이 있다. 우선, Wang et al이 제안한 구조의 경우 시각적인 정보만을 이용하여 상태를 표현하였다. 하지만 무인수상선의 충돌회피 문제에서는 시각적인 정보로 충돌회피를 수행하기위해 필요한 모든 정보를 전달 할 수 없다. 예를들어, 자선의 속도나 환경하중의 크기정보, 자선의 RPM상태 등의 요소들은 시각적인 정보로 변형하기 어려운 정보들이다. 따라서, 본 연구에서는 이들 정보를 신경망에 직접 입력할 수 있도록 저차원(low dimension) 상태변수의 입력층을 구성하였다. 두번째 큰 차이점은 기반 수학모델이 다르다는 점이다. Wang et al이

제안한 심층신경망은 MDP(Markov decision process)를 기반으로 구성되어있지만 본 연구에서 사용하는 심층강화학습은 SMDP(Semi-Markov decision process) 모델을 기반으로 한다. 따라서, 시간에 대하여 확장된 행동변수를 사용하여 충돌회피를 수행한다는 특징이 있다. 마지막 차이점은 exploration 방식의 차이점이다. 기존 방식의 경우, 행동공간에 대하여 Epsilon-greedy exploration 기법을 활용하여 수행하였지만, 본 연구에서는 무인수상선의 충돌회피의 문제 특성상 회피기동은 1대의 선박에 대하여 1번밖에 수행을 하지 못하는 특징이 존재한다. 또한, 학습의 목적이 회피행위의 시점 추론이라는 점에서 기존 기법과 차이가 있다. 따라서, 행동공간 뿐만 아닌, 시간 공간에 대해서도 exploration을 수행할 수 있는 기법을 제안하여 심층강화학습 기반 충돌회피 알고리즘의 학습에 사용하였다.

# 제 5 장 성능 검증

5장에서는 본 연구에서 제안한 충돌회피기법의 성능 검증과정에 대하여 기술하고 있다. 제안하는 기법의 성능 검증을 위하여 무인수상선 시뮬레이션과 실선에 대한 자유항주실험을 수행하였다. 5.1절은 시뮬레이션과 자유항주실험에서 사용된 대상선박에 대하여 설명하고있으며, 5.2절과 5.3절은 각각 시뮬레이션과 자유항주실험을 통한 무인수상선의 충돌회피성능 검증에 대하여 기술하고 있다.

## 5.1 대상 무인수상선

5.1절은 본 연구에서 대상 무인수상선으로 삼고있는 WAM-V 16'선형에 대한 설명과 무인수상선 동역학 모델에 대한 구체적인 설명을 포함 하고 있다. 강화학습을 이용한 충돌회피 시점 추론 알고리즘은 무인수상선의 동역학에 대한 어떠한 사전정보도 필요로 하지 않지만, 본 연구에서는 무인수상선의 강화학습이 실선 실험을 통해 이루어진 것이 아니라 무인수상선의 시뮬레이터를 통해 이루어지므로, 시뮬레이터의 구성을 위한 무인수상선의 동역학 정보가 필요하다. 5.1.1절은 대상 무인수상선의 주요재원 및 시스템의 구성에 대하여 설명하고 있으며, 5.1.2절에서는 강화학습에 사용된 무인수상선의 동역학 시뮬레이터에 대하여 기술하고 있다.

### 5.1.1 Wave Adaptive Modular Vessel 16'

Wave Adaptive Modular Vessel(WAM-V)형 무인수상선은 Marine Advanced Research, Inc. 에서 개발한 쌍동선 형태의 무인수상선 플랫폼으로, 고무튜브 형

태의 본 선체와 상부데크로 구성되어져 있다. WAM-V 형 무인수상선의 가장 큰 특징은 본 선체와 데크사이에 현가 장치(suspension system)가 설치되어있다는 점인데, 이는 파랑중 운항시 발생하는 충격을 흡수하여 상부데크의 운동을 최소화 시키는 역할을 한다. 본 연구에서 사용된 서울대학교의 WAM-V 선체는 길이 4.88m, 선폭 2.5m의 크기를 가지며 좌우 선체에 각각 설치되어 있는 전기 모터를 이용하여 추진력을 발생시킨다. 운항가능한 최고속력은 약 5 knots이고, 3 knots로 운항시 약 3시간동안 운항가능하다. 전기모터는 무인수상선의 양 선미에 고정되어 있고, 추진기의 방위각 조절이 불가능한 형태이다. 때문에, 선회를 하기위해서는 양 추진기의 추력차를 발생시켜 선회에 필요한 제어모멘트를 얻는다. Fig.5.1은 본 연구에 사용된 대상 무인수상선의 사진이고, Table.5.1은 무인수상선의 상세제원을 나타내고 있다.

| Parameter | Value |
|---|---|
| Length (m) | 4.88 |
| Beam (m) | 2.50 |
| Draft (m) | 0.30 |
| Mass (kg) | 151 |
| Moment of Inertia (kg · m) | 257 |

Table 5.1: Principal specification of Seoul National University WAM-V ASV

본 연구에서 제안하는 충돌회피 시스템을 무인수상선 실선에 적용하여 검증하기위해서는 무인수상선 스스로 경로추종 및 충돌회피 기동을 수행하는 자율운항성능이 요구된다. 이를 위해 무인수상선에는 자선의 위치 및 자세를 계측할 수 있는 GPS 센서와 AHRS 센서를 탑재시켰고, 충돌회피 명령을 계산하기위한 컴퓨터,

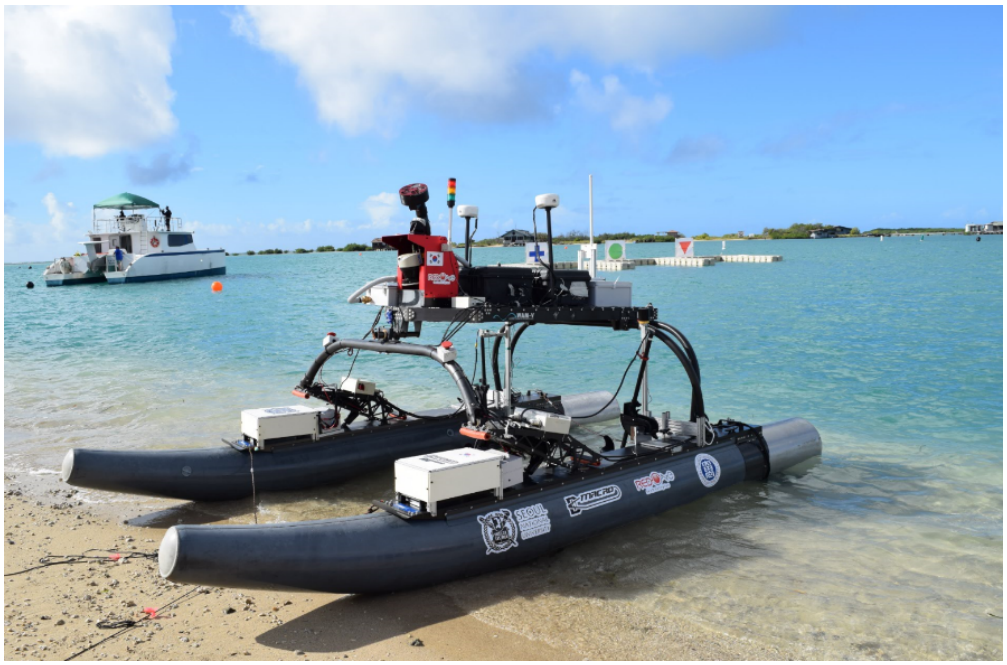


Fig 5.1: Seoul National University WAM-V autonomous surface vehicle during Maritime RobotX Chellenge 2016 (ASV competition), Hawaii, U.S.

제어명령을 추진기에 전달시키는 제어컴퓨터 등을 탑재시켰다.

### 5.1.2 무인수상선의 동역학 시뮬레이터

자선의 조종성능 및 환경하중에 대한 영향을 강화학습과정에 포함시키기 위하여 본 연구에서는 무인수상선 운동 시뮬레이터를 구현하였다. 시뮬레이터는 무인수상선의 조종성을 반영하기위한 조종운동 모형과 환경하중을 고려하기 위한 조류 및 풍하중 모형을 포함하고 있다. 본 연구의 최종목적은 시뮬레이터에서 학습한 충돌회피알고리즘을 실선에 적용하여 알고리즘의 효과성을 입증하는 것이기 때문에, 시뮬레이터에 사용하는 무인수상선의 동역학 모형은 실선의 동역학적 특성을 충분히 반영할 수 있는 모형이어야한다. 이를 위해 본 연구에서는 실선 무인수상선에 대한 동역학 모델 시스템 식별(system Identification) 실험을 수행하였고, 이를 통해 실선의 동역학적 특성을 추출하여 강화학습을 위한 시뮬레이션에 적용하였다.

#### 무인수상선의 조종운동 모형

무인수상선의 조종운동을 표현하기위하여 본 연구에서는 무인수상선을 3자유도 수평면 운동을 하는 강체(rigid body)로 가정하였고, 6자유도 운동 중 전후동요(surge), 좌우동요(sway), 선수동요 (yaw) 운동성분만을 고려하였다. Fig.5.2는 조종운동모형에 사용된 좌표계를 개략적으로 표현하고 있다. 여기서 $x_i$와 $y_i$는 각각 지구고정좌표계에서 측정된 북쪽(North)과 동쪽(East) 위치성분을 나타내고, $u$, $v$와 $V$는 각각 무인수상선의 전후동요 속도, 좌우동요 속도 그리고 전속도(total speed)를 의미한다. 무인수상선의 선수각(heading angle)은 $\psi$로 표현되고, 편류각(drift angle)과 침로각(course angle)은 각각 $\beta$와 $\chi$ 로 표현된다. 무인수상선의 편류각은 수식 $\beta = \arcsin(\frac{v}{V})$ 에 의해 계산되고, 침로각은 선수각과 편류각의 합으로 식 $\chi = \psi + \beta$ 에 의해 계산된다. Fossen [14, 15]의 표기법을 사용하여

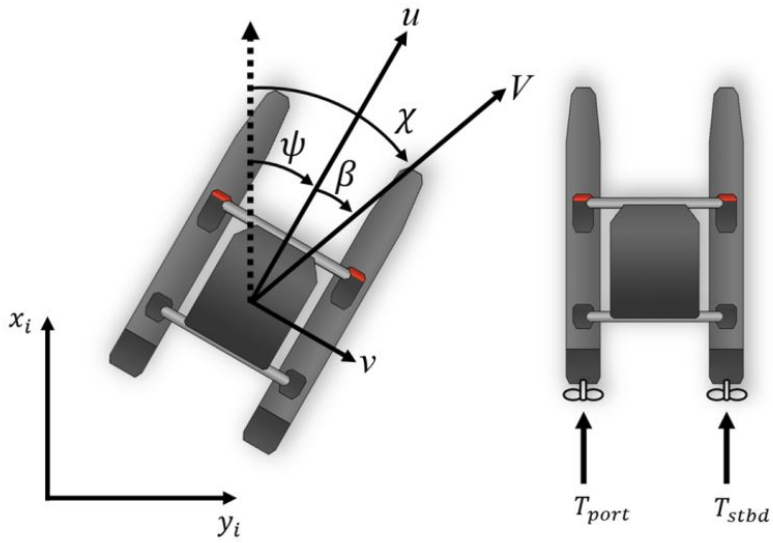



Fig 5.2: Schematic description of the ASV coordinate system

지구고정좌표계에서 표현된 무인수상선의 위치 및 자세벡터를 $\eta = [x, y, \psi]^T$로 정의하고, 선체고정좌표계에서 표현된 무인수상선의 선속도 및 회전속도 벡터를 $\nu = [u, v, r]^T$로 정의하면 이들은 식 5.1과 같은 관계를 갖고, 이를 통해 무인수상선의 동역학 방정식을 5.3과 같이 정의할 수 있다.

$$\boldsymbol{\eta} = \dot{\boldsymbol{R}}(\boldsymbol{\eta})\boldsymbol{\nu} \tag{5.1}$$

$$\boldsymbol{R}(\boldsymbol{\eta}) = \begin{bmatrix} cos\psi & -sin\psi & 0 \\ sin\psi & cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5.2}$$

$$\boldsymbol{M}\dot{\boldsymbol{\nu}} + \boldsymbol{C}(\boldsymbol{\nu})\boldsymbol{\nu} + \boldsymbol{D}(\boldsymbol{\nu})\boldsymbol{\nu} = \boldsymbol{f} \tag{5.3}$$

이 때, $\boldsymbol{M}$은 무인수상선의 질량성분과 부가질량성분으로 구성되는 질량(mass) 행렬이고, $\boldsymbol{C}$는 코리올리(Coriolis)힘과 구심력 성분이 포함된 행렬, $\boldsymbol{D}$는 감쇠(damping) 행렬을 의미한다. $\boldsymbol{f}$는 무인수상선에 설치되어있는 추진기를 통해 발생시키는 제어력과 모멘트를 나타낸다. 본 연구에 사용된 무인수상선의 추진기는 전방위 추진기(azimuth thruster)처럼 추력의 방향을 바꿀수 있는 형태가 아닌, Fig.5.2와 같이 선체 길이방향의 추력만을 발생시킬 수 있는 형태이므로 추진기에 의한 제어력과 제어모멘트는 수식 5.4과 같이 표현된다. 각 성분을 고려한 3자유도 조종운동방정식[36] 은 식 5.5 - 5.7 으로 정의된다.

$$\boldsymbol{f} = \begin{bmatrix} \tau_X \\ \tau_Y \\ \tau_N \end{bmatrix} = \begin{bmatrix} T_{port} + T_{stbd} \\ 0 \\ (T_{port} + T_{stbd})B/2 \end{bmatrix} \tag{5.4}$$

$$\begin{aligned} (m - X_{\dot{u}}\dot{u}) - m(x_G r^2 + vr) + Y_{\dot{v}}vr + \frac{Y_{\dot{r}} + N_{\dot{v}}}{2}r^2 + \\ X_u u + X_{|u|u}|u|u = \tau_X \end{aligned} \tag{5.5}$$

$$\begin{aligned} (m - Y_{\dot{v}}\dot{v}) + (mx_G - Y_{\dot{r}})\dot{r} + (m - X_{\dot{u}})ur + Y_v v + Y_r r + \\ Y_{|v|v}|v|v + Y_{|r|r}|r|r = \tau_Y \end{aligned} \tag{5.6}$$

$$\begin{aligned} (mx_G - N_{\dot{v}})\dot{v} - (I_{ZZ} - N_{\dot{r}})\dot{r} + mx_G ur - Y_{\dot{v}}uv - \frac{Y_{\dot{r}} + N_{\dot{v}}}{2}ur + \\ X_{\dot{u}}uv + N_v v + N_r r + N_{|v|v}|v|v + N_{|r|r}|r|r = \tau_N \end{aligned} \tag{5.7}$$

여기서, $X_{(\cdot)}$, $Y_{(\cdot)}$, $N_{(\cdot)}$ 은 전후동요 힘, 좌우동요 힘, 선수동요 모멘트에 대한 유체력 미계수를 의미하며, $x_G$는 선체고정좌표계의 원점에서부터 무게중심까지의 길이를 의미한다.

### 무인수상선의 시스템 식별

앞 절에서 정리한 무인수상선의 조종운동방정식은 다수의 비선형 조종미계수를 포함한다. 이러한 고차수(high order) 방정식에 대한 파라미터 식별은 센서의 계측잡음 또는 외란에 대한 높은 민감도를 갖는 특성이 있다. 따라서, 비교적 최근 수행된 다수의 연구에서도 무인수상선의 동역학 계수를 추출하기위하여 모형을 선형화 하는 등 단순한형태의 조종 모형으로 변화시킨 뒤, 이 계수들을 식별하는 기법이 사용되고 있다. ([35], [36]) 본 연구에서도 이러한 접근방법을 선택하여 수식 5.5 - 5.7에 대한 선형화를 진행하고, 선형방정식의 계수를 추출하여 실선 무인수상선의 동역학을 모사하였다. 선형모델의 구성을 위한 선형화 점(linearization point)은 충돌회피 실험 시 운용속도로 사용할 조건인 $u = 1.5\,m/s, v = 0\,m/s, r = 0\,deg/s$, 조건을 토대로 선정하였다.

$$u = 1.5\ m/s, \quad v = 0\ m/s, \quad r = 0\ deg/s, \tau_x = \tau_{x_0}\ N, \quad \tau_N = \tau_{N_0}\ N \cdot m \quad (5.8)$$

모형의 단순화를 위하여 전후동요(surge)에 대한 동역학이 좌우동요 및 선수동요와 비연성(decouple)되어있다고 가정하고, 수식 5.8의 선형화점에 대하여 동역학 방정식 5.5 - 5.7을 선형화 하면 식 5.9와 같은 선형 시불변(time invariant) 상태공간(state space) 방정식으로 무인수상선의 동역학을 표현할 수 있다.

$$\begin{bmatrix} \dot{u} \\ \dot{v} \\ \dot{r} \end{bmatrix} = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ r \end{bmatrix} + \begin{bmatrix} b_{11} & 0 \\ 0 & b_{22} \\ 0 & b_{32} \end{bmatrix} \begin{bmatrix} \tau_X \\ \tau_N \end{bmatrix} \quad (5.9)$$

시스템 식별의 목적은 무인수상선 실선의 운항데이터를 기반으로 실선의 운항특성을 반영하는 미지수 8개($a_{11}, a_{22}, a_{23}, a_{32}, a_{33}, b_{11}, b_{22}, b_{32}$)를 추정하는 것이다. 본 연구에서 활용된 무인수상선 실선의 운동모형을 획득하기위해, 본 연구에서

는 구동기 동역학에 대한 시스템 식별 실험과 조종운동에 대한 시스템 식별 실험을 다수 시행하였다. 하지만, 시스템 식별 실험에 대한 내용은 무인수상선의 충돌회피라는 본 논문의 주제와 다소 상이하므로, 이에 대한 내용은 본 논문의 부록에 수록하고, 본 장에서는 시스템식별의 결과만을 도입하여 사용하였다. 시스템식별을 통해 최종적으로 도출된 선형시불변 시스템 상태공간 방정식은 식 5.10와 같다.

$$\begin{bmatrix} \dot{u} \\ \dot{v} \\ \dot{r} \end{bmatrix} = \begin{bmatrix} -1.0191 & 0 & 0 \\ 0 & 0.0161 & -0.0052 \\ 0 & 8.2861 & -0.9860 \end{bmatrix} \begin{bmatrix} u \\ v \\ r \end{bmatrix} + \begin{bmatrix} 0.0028 & 0 \\ 0 & 0.0002 \\ 0 & 0.0307 \end{bmatrix} \begin{bmatrix} \tau_X \\ \tau_N \end{bmatrix} \quad (5.10)$$

**환경하중의 고려**

무인수상선의 충돌회피 시점 추종에 환경하중의 영향을 고려하기 위해서는 강화학습이 이루어지는 무인수상선 시뮬레이터에서 환경하중 모형을 고려해야 한다. 무인수상선이 운항할 때 선체에 작용하는 환경하중은 조류하중, 풍하중, 파랑하중 등이 있다. 무인수상선의 충돌회피 문제에서는 환경하중에 의해 발생하는 표류(drift) 운동이 관심 대상이므로, 본 연구에서는 파랑하중을 제외한, 조류하중과 풍하중에 대한 영향만을 고려하였다. 물론, 파랑하중에 의한 2차 파강제력도 표류운동을 야기시키지만, 동일 해상상태에서 무인수상선에 작용하는 표류력은 상대적으로 조류하중과 풍하중의 영향이 지배적이므로, 본 연구에서는 파랑하중에 의한 영향은 고려하지 않았다. 풍하중의 방향 $\psi_w$ 와 조류하중의 방향 $\psi_c$ 는 Fig.5.3과 같이 지구고정좌표계에서 정의하였다. $V_w$, $V_c$는 각각 풍하중과 조류하중의 속력을 의미한다.

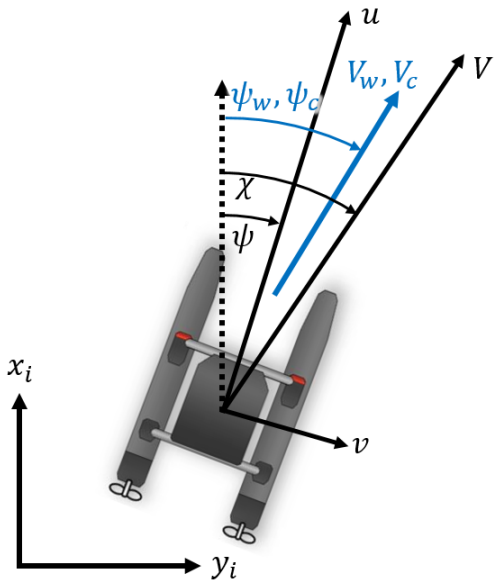



Fig 5.3: Coordinate system for environmental disturbance

### 풍 하중

지구고정좌표계에서 풍하중의 방향 $\psi_w$가 정의되면, 무인수상선 선체에 작용하는 바람의 상대속도는 식 5.11과 같이 정리할 수 있다.

$$
\begin{aligned}
u_{rel} &= V_w cos(\psi_w - \psi) + u \\
v_{rel} &= V_w sin(\psi_w - \psi) + v \\
V_{rel} &= \sqrt{u_{rel}^2 + v_{rel}^2} \\
\psi_{rel} &= \tan^{-1}(\frac{v_{rel}}{u_{rel}})
\end{aligned}
\tag{5.11}
$$

이때, $u_{rel}$ 와 $v_{rel}$ 는 전후동요 및 좌우동요 방향에 대한 바람의 상대속도를 나타내고 $V_{rel}$ 는 상대속도의 전속도를, $\psi_{rel}$는 바람의 상대속도의 각도를 나타낸다.

계산된 바람의 상대속도는 각 자유도에 대하여 풍하중의 힘과 모멘트를 계산하는데 사용된다. 선박에 대한 풍하중 모형은 Gould, Inoue, Fossen등에 의해 연구되어진 바 있다. 본 연구에서는 Fossen의 풍하중 모형을 사용하여 무인수상선에 작용하는 풍하중의 크기를 추정하였다. 무인수상선의 선체에 작용하는 풍하중의 힘과 모멘트는 다음과 같이 표현된다.

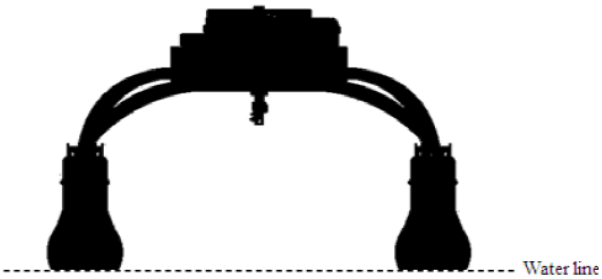
$$\tau_w = \begin{bmatrix} \tau_{w_X} \\ \tau_{w_Y} \\ \tau_{w_N} \end{bmatrix} = \frac{1}{2}\rho_a V_{rel}^2 \times \begin{bmatrix} C_X(\psi_{rel})A_{FW} \\ C_Y(\psi_{rel})A_{LW} \\ C_N(\psi_{rel})A_{LW}L_{aa} \end{bmatrix} \tag{5.12}$$

이때, $\rho_a = 1.2kg/m^3$ 는 공기의 밀도를 나타내고, $C_X(\psi_{rel}), C_Y(\psi_{rel}), C_N(\psi_{rel})$ 는 풍하중 계수를 나타낸다. $A_{FW}$ 와 $A_{LW}$는 각각 전투영면적(frontal projected area)과 측투영면적(lateral projected area)을 나타내는데, WAM-V 16' 선형에 대한 3D모델을 기반으로 CAD프로그램을 이용하여 계산한다.(Fig.5.4) Table. 5.2는 풍하중 계산을 위한 무인수상선 실선의 주요 제원을 나타내고있다.

| Parameter | Value |
|---|---|
| Loa [m] | 4.88 |
| Lpp [m] | 4.58 |
| $A_{FW}$ $[m^2]$ | 1.53 |
| $A_{LW}$ $[m^2]$ | 2.31 |

Table 5.2: Principal properties for wind load calculation

일반적으로 풍하중계수는 풍동시험을 통해 입사각에 따른 하중계수를 계측하는 방법으로 구한다. 하지만 본 연구에서 사용하는 WAM-V선형에 대한 풍동시




(a) Frontal projected area, $A_{FW}$

(b) Lateral projected area, $A_{LW}$

Fig 5.4: Projected Area of the SNU ASV

험을 수행하지 못하였기때문에 Fossen [14]이 제안한 방법을 사용하여 식 5.13을 이용하여 풍하중계수를 근사화하였다.

$$\begin{aligned} C_X(\psi_{rel}) &= -\,c_x\ cos(\psi_{rel}) \\ C_Y(\psi_{rel}) &= c_y\ sin(\psi_{rel}) \\ C_N(\psi_{rel}) &= c_z\ sin(2\psi_{rel}) \end{aligned} \tag{5.13}$$

Fossen [14]에 의하면, $c_x$, $c_x$, $c_x$ 는 각각 식 5.14 의 집합에 속하는 값으로 근사화 할 수 있다.

$$c_x \in (0.5, 0.9), \quad c_y \in (0.7, 0.95), \quad c_z \in (0.05, 0.2) \tag{5.14}$$

Qu et al (2015)[31]은 Fossen [14]의 가정에 따라, WAM-V 선형에 대한 풍하중계수를 $c_x = 0.7$, $c_y = 0.8$, $c_z = 0.1$로 가정하고 풍하중에 대한 앞먹임 제어(feedforward control)를 수행한 바 있다. 본 연구에서는 Qu et al이 사용한 풍하중계수를 이용하여 풍하중의 영향을 모사하였다.

일반적으로 선박이 운항할 때, 바람의 세기는 Beaufort number를 이용하여 표현된다. Table.5.3은 Beaufort number에 따른 풍속의 범위와 각 경우에 해당하는 해상상태(sea state)의 값을 Pierson-Moskowitz 스펙트럼의 해상 상태 표를 기반으로 정리하였다. Fig.5.5 - 5.10은 각각 풍하중의 입사각이 0°, 90°, 270°이고 Beaufort Number가 2, 5 인 조건하에서 무인수상선이 $\delta n_m = 1$, $\delta n_d = \pm 1$ 의 명령을 가지며 선회시험을 수행한 시뮬레이션 결과를 나타내고있다. 바람의 입사각 및 풍속에 따른 무인수상선의 선회 표류각(turning drift angle)과 표류거리 (drifting distance)를 바탕으로 앞서 가정한 풍하중 모형에 대한 타당성을 확인할 수 있다. 시뮬레이션 결과, 선회시험이 진행될수록 풍하중의 영향으로 선회궤적이 표류하는 현상을 확인할 수 있다. 선회궤적의 표류 방향은 선회방향과 동일한 방향인것으로 확인되었다. 본 연구에서는 무인수상선의 충돌회피 시점 추론에 풍하중이 미치는 영향을 고려하기위하여 Beaufort Number 5의 풍하중을 실험조건으로 선정하였다.

| Beaufort Number | Description | Wind Speed (m/s) | Representative value (m/s) | Sea State |
|---|---|---|---|---|
| 0 | Calm | 0 - 0.3 | 0.0 | 0 |
| 1 | Light air | 0.3 - 1.6 | 0.9 | 0 |
| 2 | Light breeze | 1.6 - 3.4 | 2.5 | 1 |
| 3 | Gentle breeze | 3.4 - 5.4 | 4.4 | 2 |
| 4 | Moderate breeze | 5.5 - 7.9 | 6.7 | 3 |
| 5 | Fresh breeze | 8.0 - 10.0 | 9.0 | 4 |
| 6 | String breeze | 10.8 - 13.8 | 12.3 | 5 |
| 7 | Moderate gale | 13.9 - 17.1 | 15.50 | 6 |
| 8 | Fresh gale | 17.2 - 20.7 | 18.95 | 7 |
| 9 | Strong gale | 20.8 - 24.4 | 22.60 | 7 |
| 10 | Storm | 24.5 - 28.4 | 26.45 | 8 |
| 11 | Violent storm | 28.5 - 32.6 | 30.55 | 90 |

Table 5.3: Beaufort number and corresponding wind speed. Table from [49]

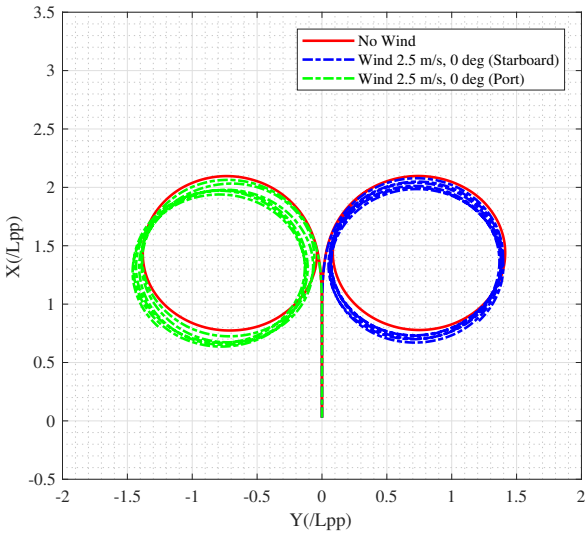



Fig 5.5: Turning test trajectory under 0 deg wind load(Beaufort No.2)

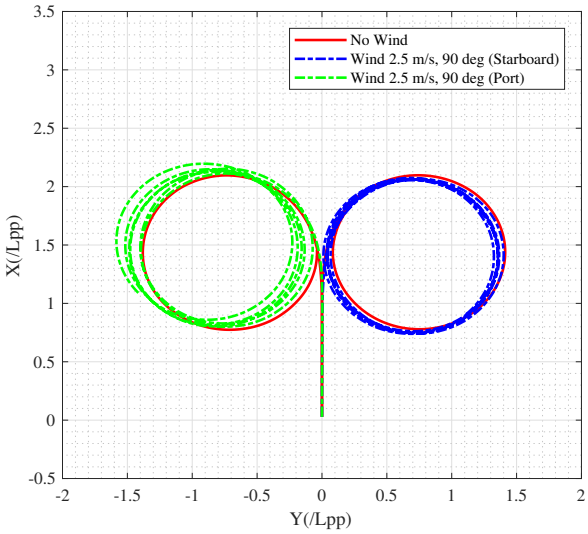



Fig 5.6: Turning test trajectory under 90 deg wind load(Beaufort No.2)

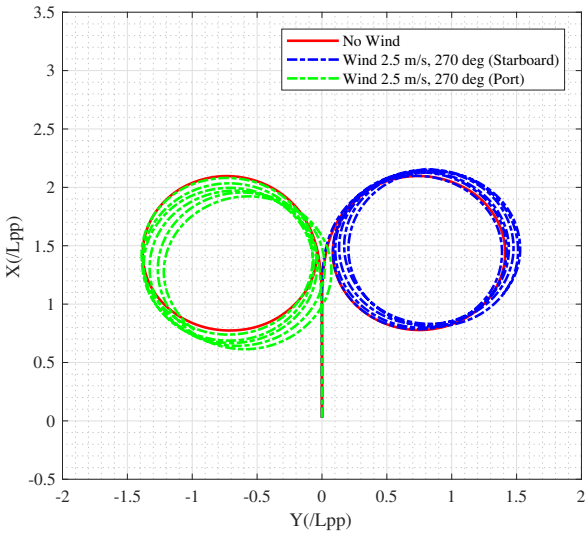



Fig 5.7: Turning test trajectory under 270 deg wind load(Beaufort No.2)

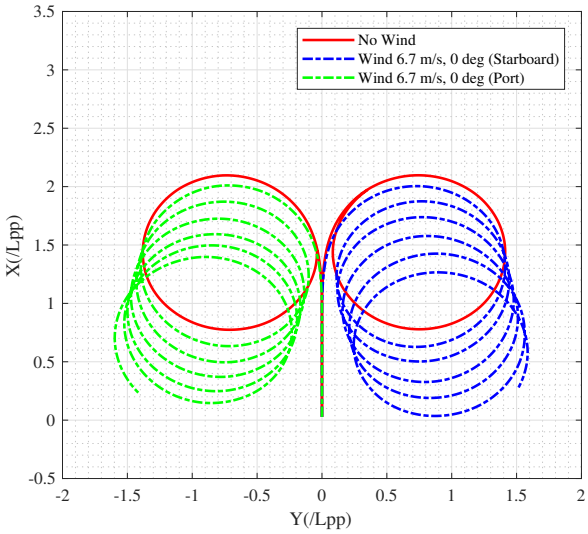



Fig 5.8: Turning test trajectory under 0 deg wind load(Beaufort No.5)

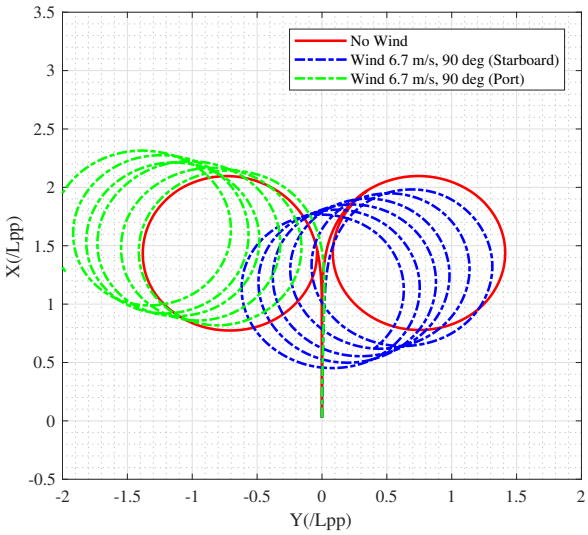



Fig 5.9: Turning test trajectory under 90 deg wind load(Beaufort No.5)

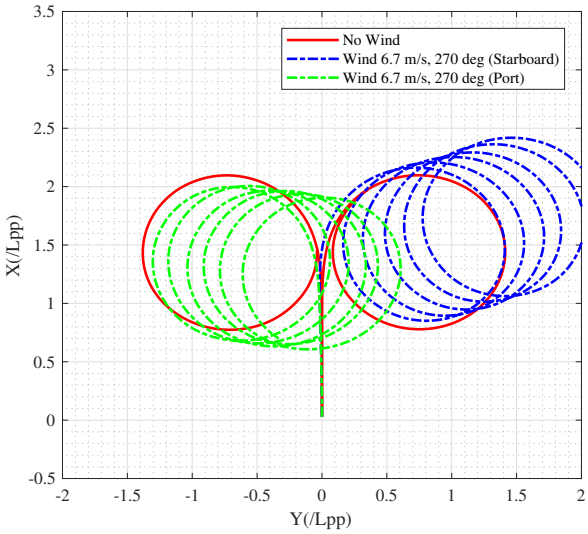



Fig 5.10: Turning test trajectory under 270 deg wind load(Beaufort No.5)

### 조류 하중

본 연구에서는 무인수상선에 작용하는 조류하중을 고려하기위해 Hwang [20]의 방법을 사용하였다. 무인수상선 선체에 작용하는 조류의 상대속도와 상대가속도는 각각 식 5.15와 식 5.16와 같이 정리된다.

$$\begin{aligned} u_r &= u - V_c cos(\psi - \psi_c) \\ v_r &= v + V_c sin(\psi - \psi_c) \end{aligned} \tag{5.15}$$

$$\begin{aligned} \dot{u}_r &= \frac{du_r}{dt} \\ &= \frac{d}{dt}(u - V_c cos(\psi - \psi_c)) \\ &= \dot{u} - \dot{V}_c cos(\psi - \psi_c) - V_c sin(\psi - \psi_c) \cdot \frac{d}{dt}(\psi - \psi_c) \\ &= \dot{u} + V_c sin(\psi - \psi_c) \cdot \frac{d}{dt}(\psi - \psi_c) \\ &= \dot{u} + V_c \cdot r \cdot sin(\psi - \psi_c) \\ \\ \dot{v}_r &= \frac{dv_r}{dt} \\ &= \frac{d}{dt}(v + V_c sin(\psi - \psi_c)) \\ &= \dot{v} + \dot{V}_c sin(\psi - \psi_c) + V_c cos(\psi - \psi_c) \cdot \frac{d}{dt}(\psi - \psi_c) \\ &= \dot{v} + V_c cos(\psi - \psi_c) \cdot \frac{d}{dt}(\psi - \psi_c) \\ &= \dot{v} + V_c \cdot r \cdot cos(\psi - \psi_c) \end{aligned} \tag{5.16}$$

식 5.15와 식 5.16에서 구한 상대속도와 상대가속도를 무인수상선의 조종운동방정식에 대입하면 일정한 방향과 속도를 갖는 조류하중하에 선체에 작용하는

동유체력을 계산할 수 있다. (식 5.17)

$$
\begin{aligned}
X_c &= m(\dot{u}_r - v_r r - x_G r^2) \\
Y_c &= m(\dot{v}_r - u_r r - x_G \dot{r}) \\
N_c &= I_{zz}(\dot{r} + m x_G(\dot{v}_r + u_r r)
\end{aligned}
\tag{5.17}
$$

Fig.5.11 - 5.16은 무인수상선에 0.1m/s 와 0.3m/s의 속도에 입사각이 0°, 90°, 270°인 조류조건 하에서 $\delta n_m = 1$, $\delta n_d = \pm 1$ 의 명령을 가지며 선회시험을 수행한 시뮬레이션 결과를 나타내고있다. 풍하중에 대한 시뮬레이션결과와 유사하게 일정한 방향으로 선회궤적이 표류하는 것을 확인할 수 있다. 다만, 선회궤적의 표류방향은 풍하중에 대한 시뮬레이션 결과와는 반대로 선회 반대방향으로 표류하는 현상을 확인하였다. 본 연구에서는 무인수상선의 충돌회피 시점 추론에 조류하중이 미치는 영향을 고려하기위하여 0.1 m/s의 조류하중이 선체에 작용하는 조건을 실험조건으로 선정하였다.

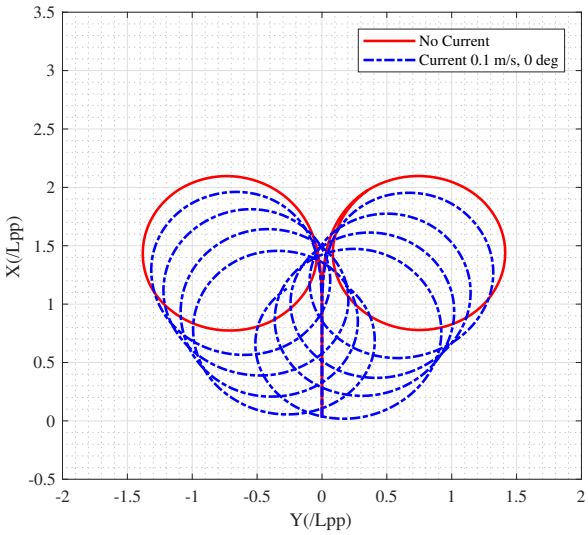



Fig 5.11: Turning test trajectory under 0 deg 0.1m/s current load

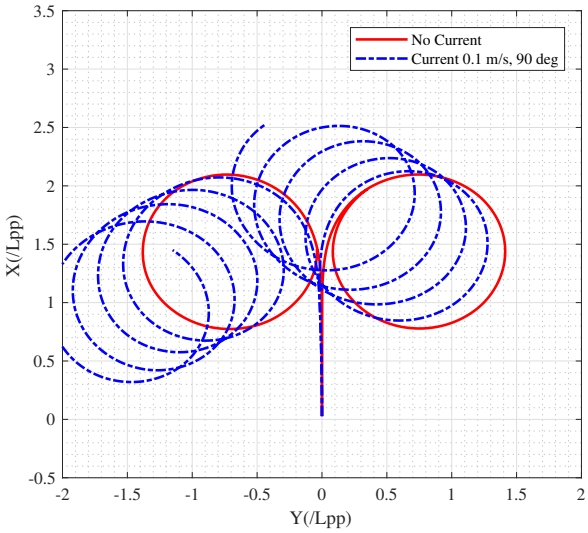



Fig 5.12: Turning test trajectory under 90 deg 0.1m/s current load

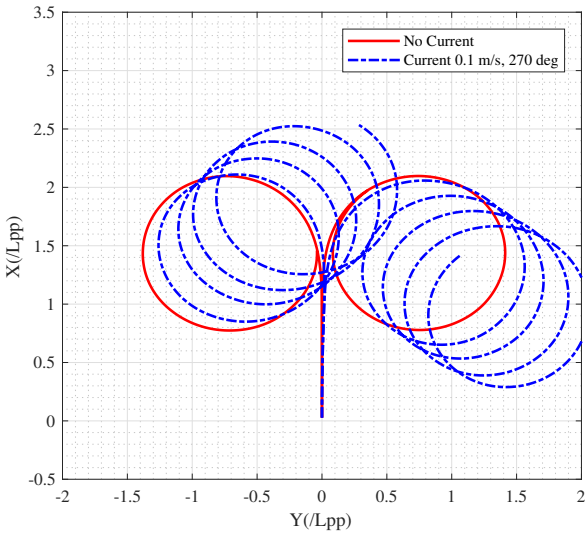



Fig 5.13: Turning test trajectory under 270 deg 0.1m/s current load

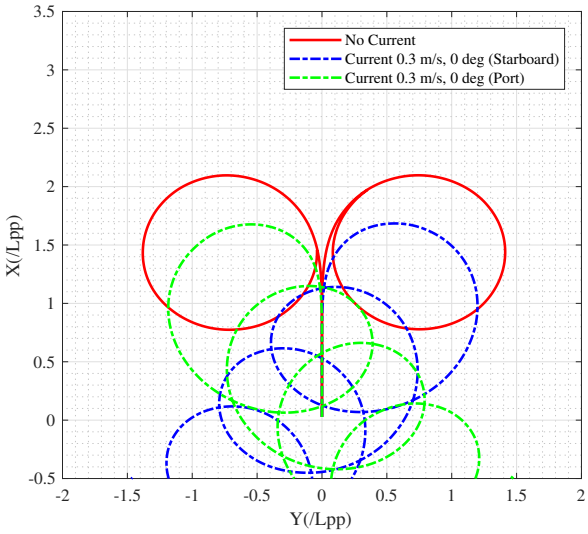



Fig 5.14: Turning test trajectory under 0 deg 0.3m/s current load

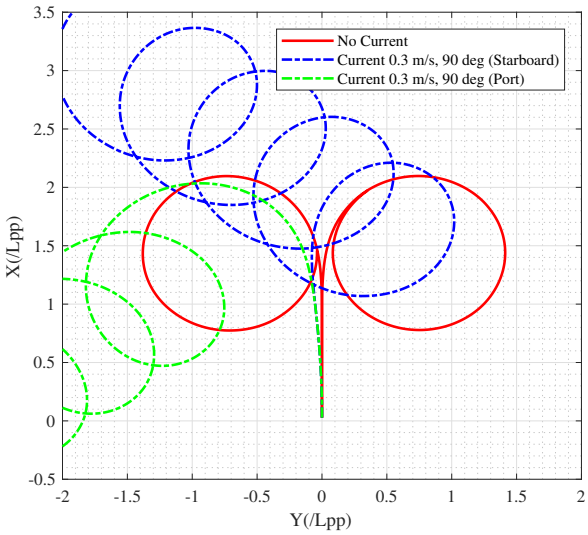



Fig 5.15: Turning test trajectory under 90 deg 0.3m/s current load

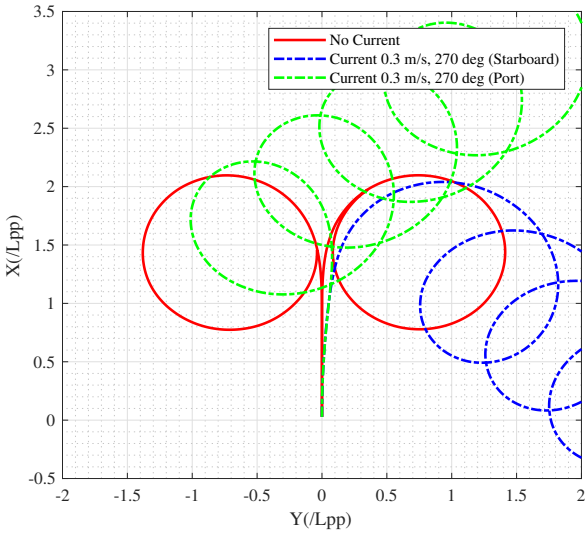



Fig 5.16: Turning test trajectory under 270 deg 0.3m/s current load

## 5.2 시뮬레이션 결과

### 5.2.1 충돌회피 알고리즘 분석

본 연구에서 제안하는 기법을 이용하여 충돌회피 시뮬레이션을 수행하기에 앞서, 다양한 접근방법을 통해 알고리즘의 성능 및 효용성을 분석하고자 하였다. 우선, 충돌위험도의 추론성능을 확인하여 기존에 연구되었던 충돌위험도 추론방식과 유사한 경향을 보이는지 확인하였다. 또한, 장애물이 주변에 없는 상황과 존재하는 상황의 격자지도(grid map)를 입력하였을 때, 각각의 경우에 신경망이 이미지의 어떤 부분에 관심을 가지는지, 현저 지도(saliency map)[13] 기법을 통하여 확인해보았다. 마지막으로, 충돌회피가 이루어지는 각 단계에 대한 분석을 진행하였다.

#### 충돌위험도 추론

충돌위험도는 현재 조우상황이 얼마나 위험한 상황인지에 대하여 정량적으로 평가하는 지표이다. 강화학습에서는 이와 유사한 개념이 존재하는데, 이는 가치함수(value function)로, 현재 상황에서 미래에 받을것으로 예상하는 보상(reward) 값의 합을 의미하는 개념이다.

본 연구에서는 임의의 상대위치에 장애물을 위치시킨 격자지도(grid map)를 신경망에 입력하고, 이에따라 출력되는 가치함수값의 결과를 분석하여 해당 위치의 장애물에 대한 신경망의 충돌위험도를 분석하는 가상실험을 수행하였다. 가상실험을 위해 장애물을 균등하게 분포시키고 각 상황에 대한 격자지도를 생성하였다. 장애물과의 횡방향 상대거리는 -168에서 168사이의 값을 100등분하여 위치시켰고, 종방향 상대거리는 0부터 168 사이의 값을 50등분 하여 총 5000 개의 후보점

에 대한 충돌위험도를 계산하였다. 장애물의 속도와 침로각 또한 중요한 변수가 될 수 있는데, 본 가상실험에서는 -y축의 방향을 향하고 1.5m/s의 속도로 운항한다고 가정하였다. 신경망에서 출력되는 가치함수는 미래상황에 대한 긍정도를 표현하는 값이고, 충돌위험도는 위험성에 대한 표현지표이기때문에, 가치함수의 부호를 반전시키고, 최대값을 1로 정규화 하여 추론위험도값을 계산하였다.

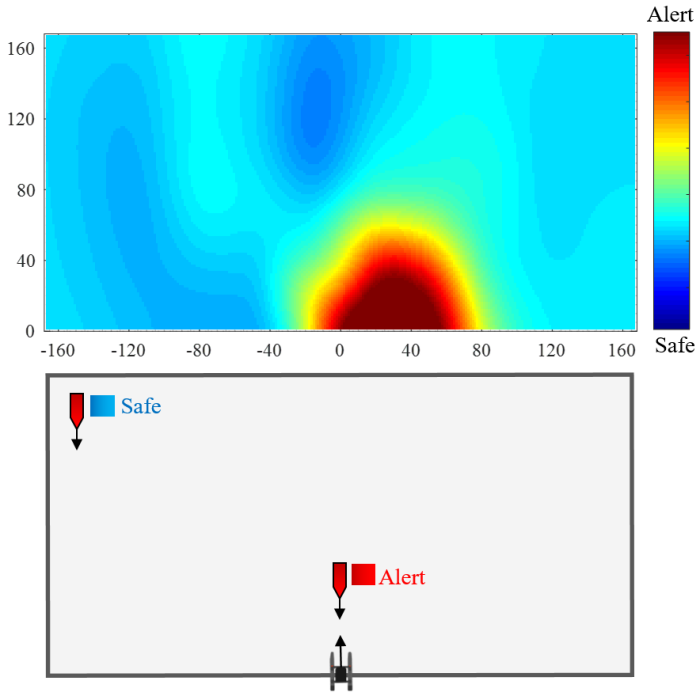



Fig 5.17: Estimated collision risk value with respect to target ship's position

Fig.5.17은 가상실험의 결과를 나타낸다. 실험 결과, 자선과 장애물간의 상대거리가 작을 수록 높은 충돌위험도를 갖는현상을 확인할 수 있었고, 좌측에 위치한 장애물보다는 우측에 위치한 장애물에 대해 더 높은 충돌위험도를 갖는것을 확인할 수 있었다. 이는 Coldwell[9], Goodwin[17], Davis[12]의 우측으로 편향(bias)된 선박의 안전영역과 같이, 자선기준 우측에 존재하는 장애물이 상대적으로 높은

충돌위험도를 갖는 선박 충돌회피의 특징을 반영하는 현상이다. 이러한 현상의 원인은 바로 COLREGs 때문인데, 선박 충돌회피의 경우 대부분의 조우상황에 대하여 우현변침을 통해 충돌회피를 수행해야 한다는 특징이 있다. 하지만 만일 장애물이 우측에 존재하면 회피에 소요되는 시간이 길어지고, 상대적으로 위험한 상대거리에 노출되는 시간이 길어지는 현상을 반영한 것이다. 이러한 현상에 대한 어떠한 정보도 학습단계에서 제공하지 않았지만, 강화학습기반 충돌회피 시스템은 COLREGs를 준수하는 충돌회피를 스스로 수행하면서 각 조우상황에 대한 충돌위험도를 스스로 학습할 수 있는것을 확인할 수 있었다.

Fig.5.17의 충돌위험도 결과를 이용하여, 고전적인 충돌위험도 추론기법인 TCPA와 DCPA에 대한 충돌위험도 함수를 구성해보았다. 대표점들의 위치와 속도를 이용하여 TCPA와 DCPA를 각각 계산하고, 이에따른 충돌위험도 함수를 구성한 결과는 Fig.5.18과 같다. 추론된 충돌위험도 함수는 선행연구의 충돌위험도 모형

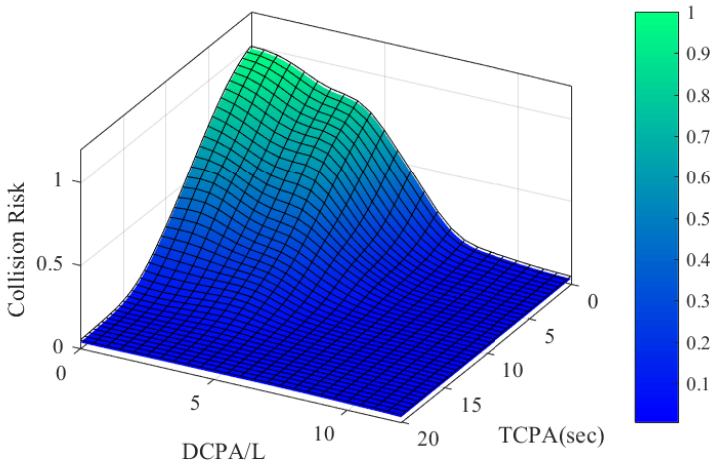



Fig 5.18: Estimated collision Risk of the ASV, based on proposed method

(Fig.3.10, Fig.3.11)과 유사하게 TCPA와 DCPA가 감소할수록 최대값으로 증가하

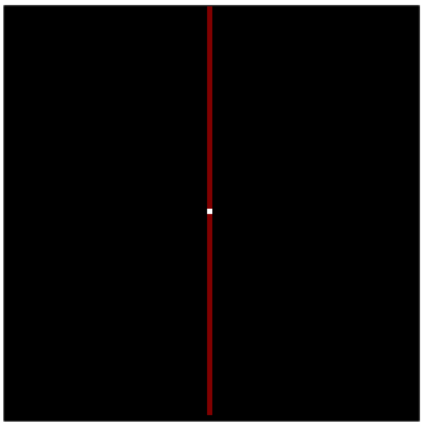
는 경향을 보이고 있으며, TCPA나 DCPA에 대하여 비선형적으로 충돌위험도가 변화하는 현상을 확인할 수 있었다.

**장애물에 대한 인식**

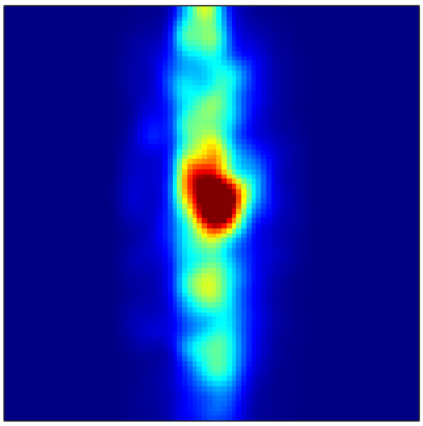
본 연구에서 사용한 심층강화학습 신경망의 구조는 Fig.4.8과 같이 합성곱 신경망 (convolutional neural network)이 입력층(input layer)에 연결되어있는 구조이다. 현저 지도(saliency map)기법은 합성곱 신경망을 통해 패턴을 인식하는 문제에서 현재 신경망이 이미지의 어느 영역에 집중을 하고있는지 여부를 분석하는 도구이다. 이 기법을 이용하면, 현재 신경망이 집중하고 있는 이미지의 영역을 가시적으로 확인할 수 있기때문에, 구성된 신경망의 동작과정에 대하여 간접적으로 추론할 수 있는 틀을 제공한다.

Engel et al (2001)[13]이 제안한 현저 지도 계산 방법은 가치함수값에 대한 이미지 각 픽셀의 1차 미분값을 계산하여 이루어진다. 미분값이 큰 픽셀은 곧 가치함수에 영향을 크게 미치는 영역임을 의미하기때문에, 신경망이 집중하고있는 영상영역이라고 표현할 수 있다.

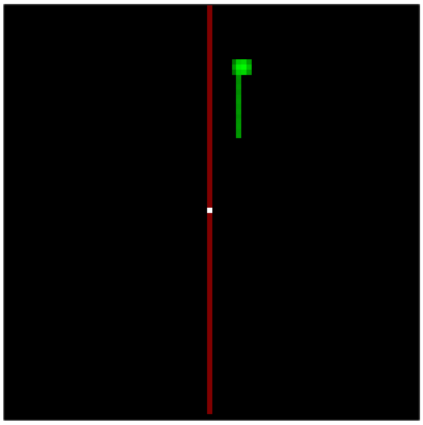
현저 지도를 이용한 신경망 분석을 수행하기위하여 신경망에 장애물이 존재하는 경우의 격자지도와 존재하지 않는 경우의 격자지도를 각각 입력하고, 현저 지도를 계산해본 결과는 Fig.5.19와 같다.



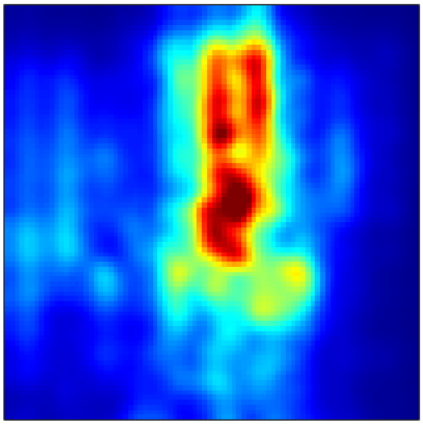
(a) State grid map without obstacle



(b) Saliency map without obstacle



(c) State grid map with obstacle



(d) Saliency map with obstacle

Fig 5.19: Saliency map of the grid map with respect to obstacle existance

분석 결과, 장애물이 존재하지 않는 상태에서는 주로 자선의 위치(Fig.5.19(a)의 흰 점)와 경로의 형상에 집중하고있는 현상을 확인할 수 있었다, 하지만 Fig.5.19(b)와 같은 장애물이 존재하는 상황에서는 자선의 위치, 경로, 장애물이 위치한 영역에서 높은 집중이 일어나는 현상을 화인할 수 있었다. 이를 통해 학습된 신경망

구조가 장애물의 등장에 따라 해당 영역이 활성화(activation)되고, 이 정보가 강화학습의 가치함수에 영향을 끼치는 과정을 확인할 수 있었다.

**충돌 회피 과정**

심층강화학습 기반 충돌회피 기법을 이용하여 충돌회피 시뮬레이션을 수행한 결과는 Fig.5.20 - Fig.5.21과 같다. 시뮬레이션 조건은 자선은 (0,0)의 위치에서 90도의 각도로 항주하고, 타선은 (0,200)의 초기위치에서 -90도의 각도로 접근하는 Head on상황에 대하여 고려하였다. 자선과 타선 모두 1.5m/s의 속도로 운항하도록 설정하였다. 본 연구에서는 타선에 대한 위치 및 운동정보는 AIS, Radar, Lidar, Vision센서 등을 통해 탐지 및 식별이 완료되었다고 가정하고, 전달받은 장애물의 위치, 자세, 속도정보를 활용하여 충돌회피 시뮬레이션을 수행하였다.

시뮬레이션 결과, 약 31초 이후부터 충돌회피가 이루어지는데, 이 회피기동은 강화학습 신경망에서 추종하는 유도기를 전환시켰기 때문이다. 유도기의 전환은 강화학습 신경망에서 출력되는 어드벤티지(advantage)값에 의해 결정이 되는데, 이 advantage값은 행동가치함수(action value function)에서 상태 가치함수(state value function)을 뺀 값이다. 즉, 현재 상황에 대한 판단은 배재하고 오로지 특정 행동에 대한 기대 효용값을 나타내는 값이다. 만일 경로추종유도기에 대한 advantage 값보다 우현변침을 통한 회피기동유도기의 advantage값이 높으면, 해당순간에 유도기가 전환되고, 장애물을 회피할 때까지 해당 유도기가 작동하여 충돌회피가 이루어진다. Fig.5.21에서 이러한 현상을 확인할 수 있는데, 각 단계별로 추종된 advantage값을 비교해보면, avoidance 단계에서는 회피기동에 대한 advantage값이 경로추종의 advantage값보다 커지면서 회피기동이 시작되는 현상을 확인할 수 있다.

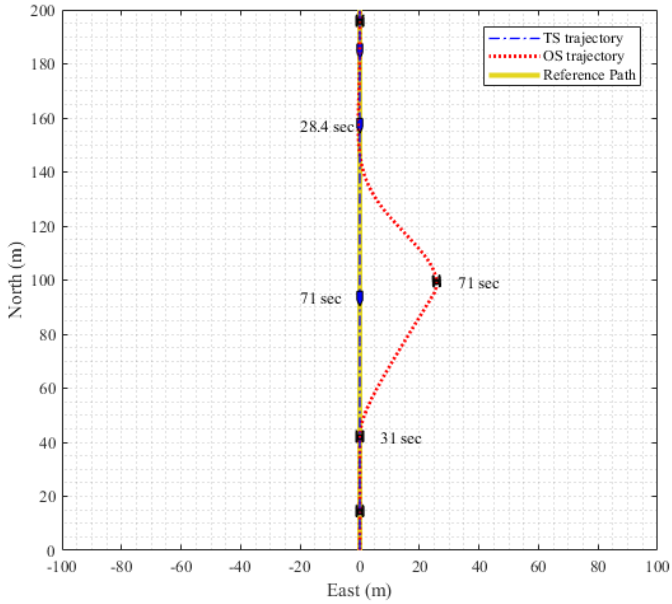



Fig 5.20: Trajectory of collision avoidance simulation using proposed method (Head on)

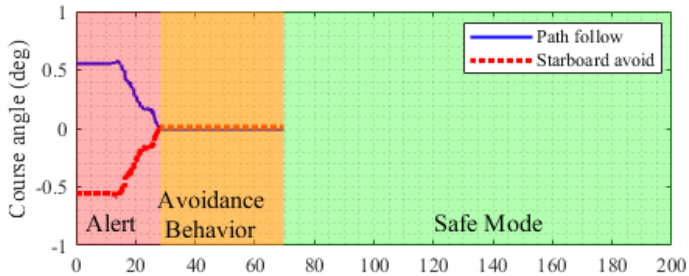



Fig 5.21: Advantage value of the DRL network during the collision avoidance simulation

이렇게 행동기반 유도기로부터 유도기가 선택되면, 해당 유도기는 경로추종 또는 충돌회피를 수행하기위한 유도명령을 생성한다. 유도명령은 목표침로각과

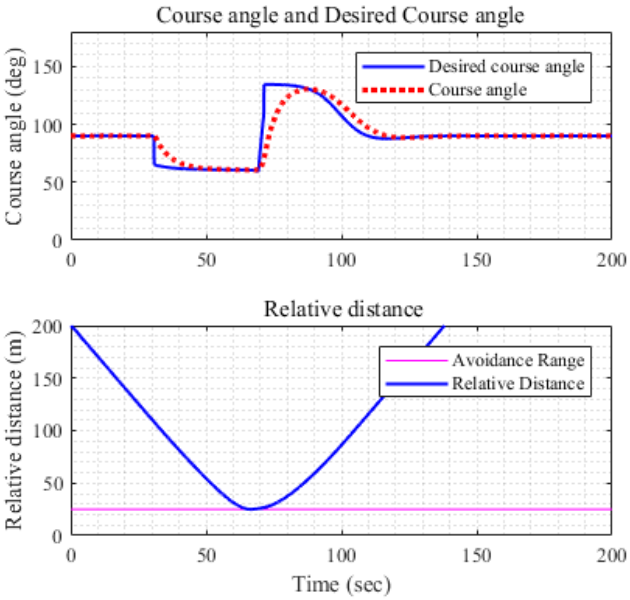



Fig 5.22: Course angle and desired course angle during the collision avoidance simulation

목표선속으로 구분되는데, Fig.5.22는 이에 따른 목표 선수각의 시계열 자료를 나타내고있다. Fig.5.22와 같이 충돌회피 명령에 따라 침로각 명령이 생성되면, 침로각 제어를 통해 충돌 회피가 이루어진다. 침로각의 제어를 위하여, PID 제어기를 설계하여 사용하였다.

### 5.2.2 충돌회피 성능에 대한 분석

본 연구에서 제안하는 심층강화학습 기반 충돌회피 기법의 효과성을 검증하기 위하여 타 충돌회피 시점추론 알고리즘과의 비교 분석을 진행하였다. 비교대상으로 선정한 알고리즘은 Kuwata et al (2014)[23]가 제안한 CPA를 이용한 방법과, Myre(2016) 가 제안한 Set Based Guidance 기법 [29]을 선정하였다. 세 연구 모두

속도 장애물(velocity obstacle, 이하 VO)기법을 이용하여 충돌회피 행동을 수행하였다는 특징이 있다. 본 연구를 제외한 나머지 두 기법에 대한 설명은 3장에 기술되어있다. 본 연구에서는 CPA를 이용한 기법의 경우, TCPA가 60sec 이내에 들어오면 충돌회피를 시작하도록 설정했고, SB기법의 경우 90m 이내에 타선이 들어오면 충돌회피를 시작하도록 설정하였다. 회피시점 결정에 사용되는 이들 파라미터의 결정은 CPA기법의 경우 Crossing상황을, SBG기법의 경우 Head on 상황을 기준으로 해당 조우상황에서 바람직한 회피기동이 이루어지도록 선정하였다.

본격적인 충돌회피 성능 분석에 앞서, 각 기법의 특성을 파악하고자 Head on 조우상황과 Crossing 조우상황에 대한 시뮬레이션을 수행하였다. 각각의 경우 타선은 1.5m/s의 속도로 접근하고, 초기위치는 각각 (0,200)과 (100,100)으로 설정하여 (0,100)의 위치에서 조우하는 상황을 구현하였다. Fig.5.23과 Fig.5.24는 각각 Head on과 Crossing (give way) 조우상황에 대한 시뮬레이션의 예시 결과를 나타내고있으며, 각 조우상황에서 기법별 충돌회피 시작 시간은 Table.5.4에서 확인할 수 있다. Head on 상황의 경우, COLREGs에 의거하여 타선이 자선의 회피를 관측할 수 있는 명백한 회피명령을 생성해야하는데, CPA를 이용한 방법의 경우, 충돌회피의 시점이 지나치게 이른 시점에 이루어져서 변침이 상대적으로 작게 이루어지는것을 확인할 수 있다. 반면, 심층강화학습 기법이나 SB기법의 경우, 30도 이상의 변침각이 확보되는현상을 확인할 수 있다.

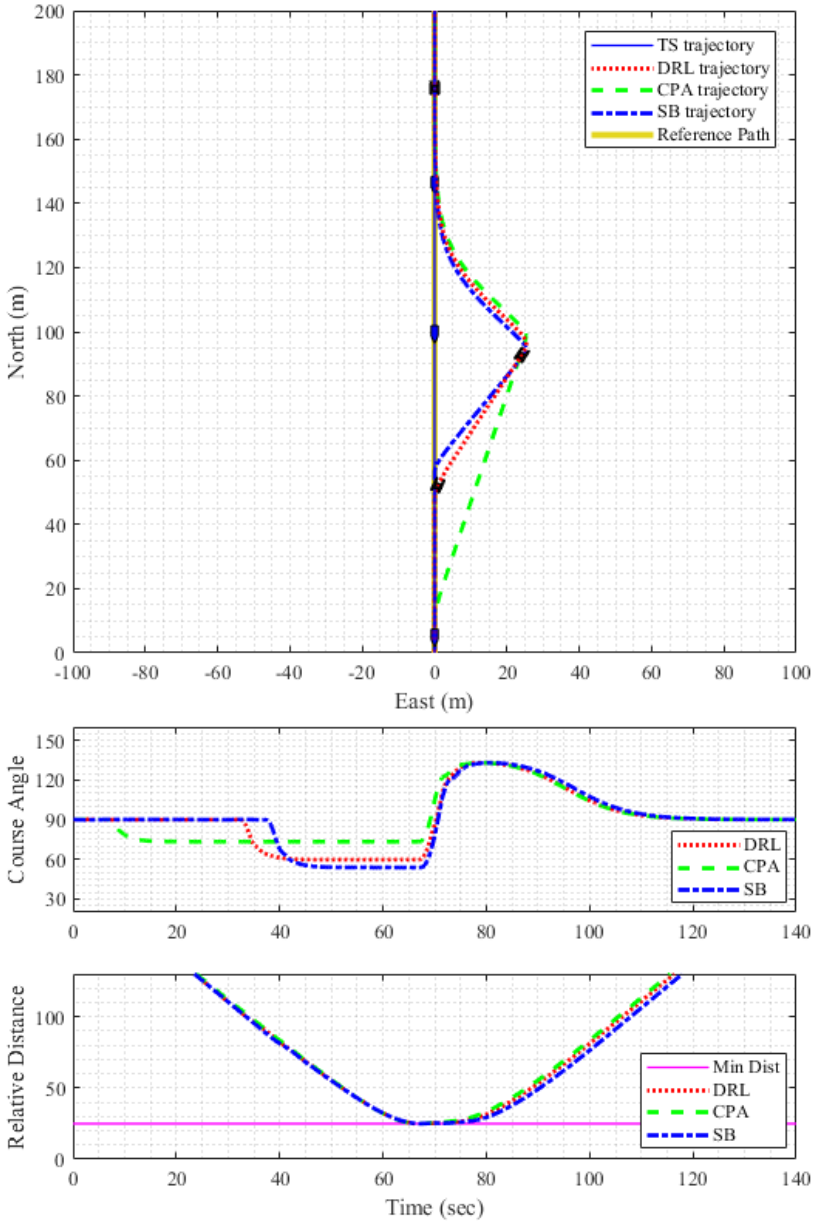



Fig 5.23: Collision avoidance simulation result using different collisin risk assessment method (Head on)

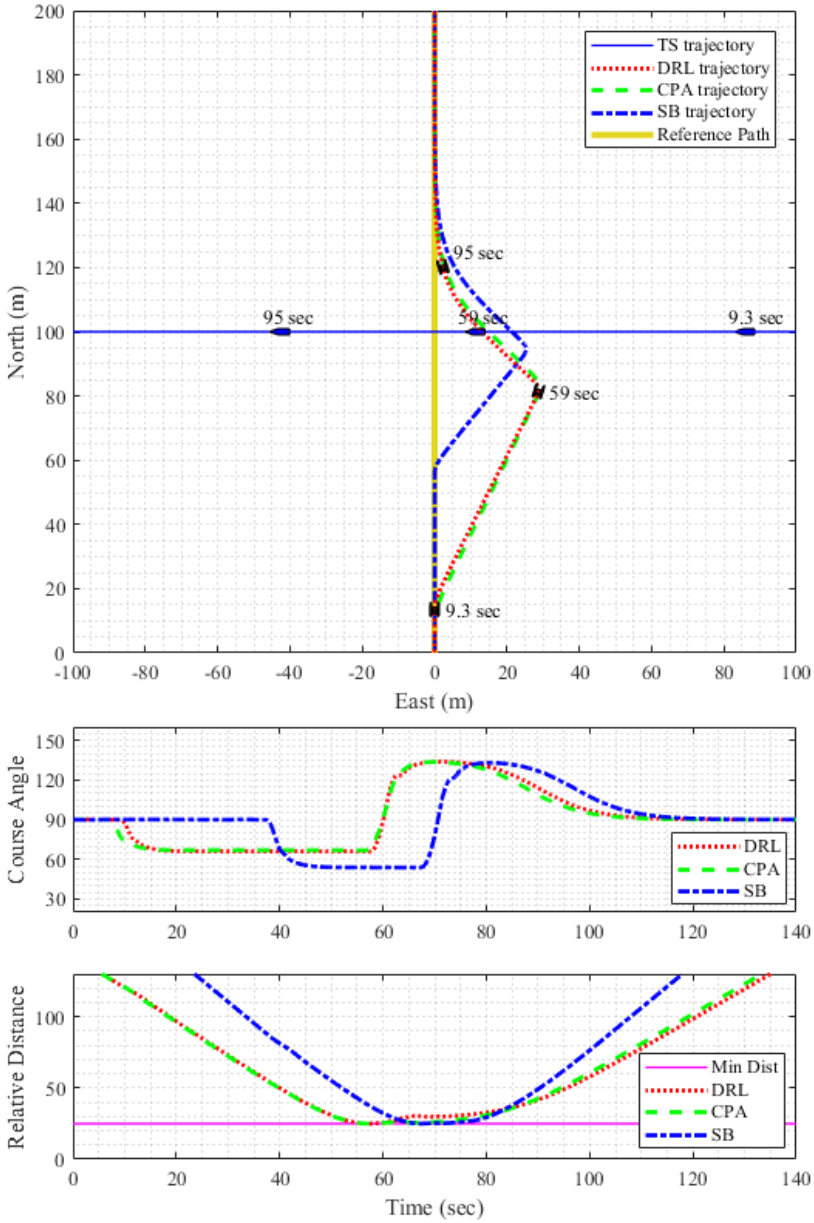



Fig 5.24: Collision avoidance simulation result using different collisin risk assessment method (Crossing, give way)

| Encounter Situation | CPA | Set Based | DRL (proposed) |
|---|---|---|---|
| Head on | 7.2 sec | 37.3 sec | 31 sec |
| Crossing (give way) | 7.2 sec | 24.0 sec | 7.2 sec |

Table 5.4: Avoidance behavior begin time according to encounter situation and collision risk assessment method

Crossing (give way)에 대한 충돌회피 시뮬레이션 결과, CPA와 DRL을 이용한 경우, 두 경우 모두 7.2초에 피항행위를 시작하는 것을 확인할 수 있다. Crossing (give way) 조우상황에서는 자선이 피항의무선박이고 타선이 진로유지선이므로, 타선에게 자선의 회피기동여부에 대하여 최대한 빠른 시점에 표현해야 한다. 본 연구에서는 이를 고려하기 위해 타선 장애물이 위험장애물로 분류됨과 동시에 곧바로 충돌회피를 시작하도록 보상함수를 식 4.5 와 같이 정의하였다. CPA와 DRL 기법과는 다르게, SB를 이용한 충돌회피 알고리즘은 타선이 자선으로부터 90m 이내에 접근을 한 이후에야 회피기동을 수행하게된다. 따라서 Crossing (give way)과 같은 조우상황에서는 충돌회피 시점이 상대적으로 늦어지게 되는데, 이는 곧 진로유지선에게 명확한 충돌회피 의사를 보이지 못하는 것으로 간주되어 타선의 혼란을 야기시킬 수 있다.

심층강화학습을 이용한 충돌회피 추론 알고리즘의 경우, 각각의 조우상황에 대하여 다른 보상함수를 사용하여 각 조우상황에 알맞는 충돌회피 기준을 사용할 수 있다. 그 결과, Head on 상황에서 SB기법이 갖는 장점과 (명확한 경로변침에 따른 회피행동에 대한 가관측성 확보) Crossing (give way)상황에서 CPA기법이 갖는 장점 (최대한 빠른시점에 회피기동을 수행하여 회피의지 표명, 진로유지선의

혼란을 방지) 을 동시에 확보할 수 있다.

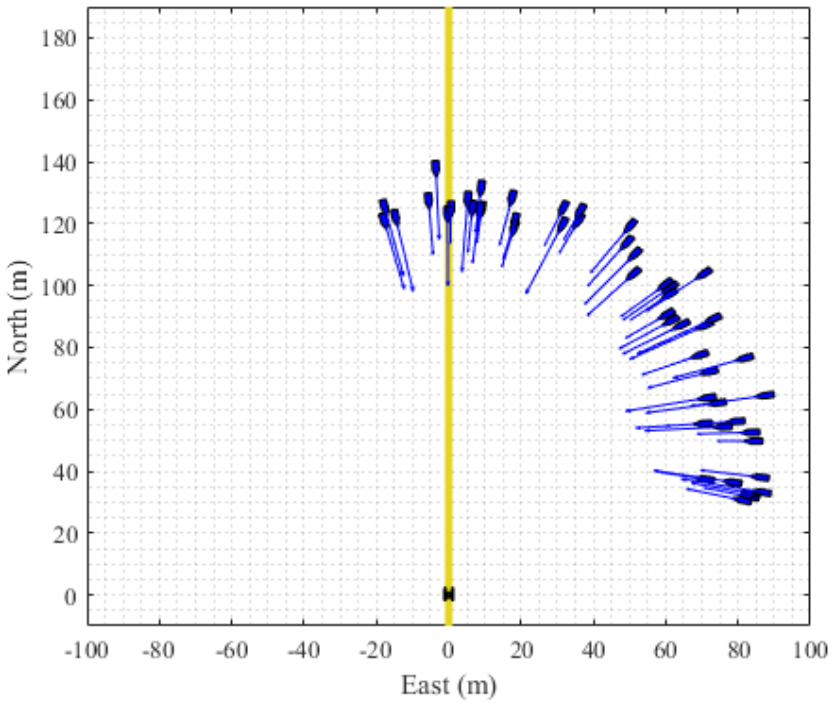



Fig 5.25: Schematic description of the Monte-Carlo simulation environment

심층강화학습 기반 충돌회피 의사결정기법에 대한 성능 분석을 수행하기 위하여, 본 연구에서는 타선의 위치, 자세, 속도를 랜덤하게 갖는 다양한 케이스에 대한 몬테카를로 시뮬레이션 (Monte-Carlo Simulation)을 수행하였다. 충돌회피 상황은 타선과의 1:1 조우상황으로 가정하였고, Head on 상황과 Crossing (give way) 상황을 가정하여 임의의 조건에 대한 충돌회피 성능을 분석하고자 하였다. Fig.5.25는 본 절에서 수행한 몬테카를로 시뮬레이션의 초기조건을 나타내고있다. Fig.5.25의 파란색 선박은 각 실험 케이스에 대한 타선의 초기조건을 의미한다. Head on 상황에 대하여 1000개의 초기조건과 Crossing (Give way) 상황에 대하

여 1000개의 초기조건을 선정하여 시뮬레이션의 초기조건을 구성하였다. Head on 상황의 경우, 초기 각도는 자선이 경로를 추종하였을 때, CPA점에서 타선이 자선의 선수각 기준 $-25°$ - $25°$의 범위내에서 조우하도록 균등분포로 정의하였고, Crossing (Give way) 상황에서는 $25°$ - $115°$의 범위를 갖도록 초기조건을 가정하였다.

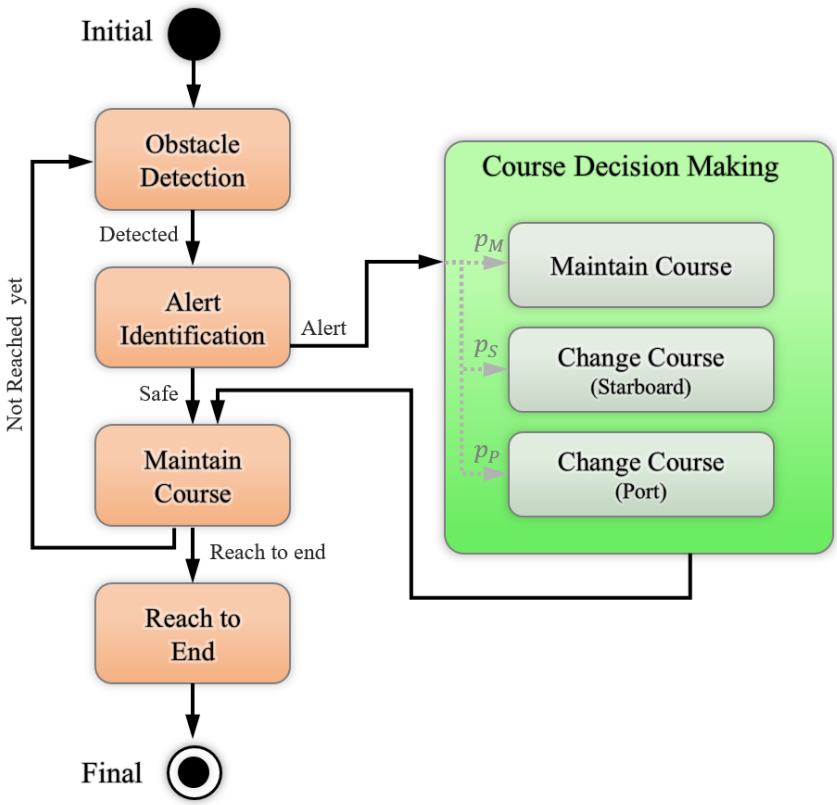



Fig 5.26: State flow of the target ship decision making process during the Monte-Carlo simulation

보다 사실적인 충돌회피 환경을 구성하기위해, 본 시뮬레이션에서는 타선의 회피행위를 고려하였다. 타선이 선택할 수 있는 행동변수는 자선과 동일하게 1)

현재 침로의 유지, 2) 우현 변침, 3) 좌현 변침 3가지로 정의하였고, 현재 타선의 상황에 따라 확률적으로 회피행위를 수행하도록 고려하였다. 타선이 회피행위를 수행하는 단계는 Fig.5.26과 같은 상태를 따라 이루어진다. 장애물을 탐지하고, 장애물의 위험성 여부가 식별되면, 타선은 장애물에 대한 회피여부를 판단하게된다. 본 연구에서는 타선의 행동선택에 대한 의사결정을 위해 현재 탐지되는 자선의 가관측성을 고려하고자하였다. 즉, Head on 상황의 경우, 상대선박이 충분히 큰 각도로 변침을 수행하여 회피여부에 대한 상황이 명료할 경우를 명백한 (obvious) 상태로 구분하고, 그렇지 않은 경우를 모호한(ambiguous) 상태로 구분하여 각각의 경우에 대하여 다른 회피행위에 대한 확률분포를 부여하였다. Crossing (give way) 상황의 경우,(타선의 기준으로는 Crossing (stand on) 조우상황) 자선이 최대한 빠른 시점에서 회피행위를 수행하여 회피의지를 표현하는경우를 명백한 상태로, 그렇지 않은 경우를 모호한 상태로 구분하였다. 타선의 행위결정을 위해 사용된 각 행동에 대한 확률변수는 Table.5.5와 같다. 각각의 값은 타선이 해당 행위를 수행할 확률을 의미한다. 예를들어, Head on (obvious) 상황에서, 타선은 20% 의 확률로 현재 침로를 유지하고, 70%의 확률로 우현변침을 수행하며, 10%의 확률로 좌현변침을 수행한다. COLREGs에 따르면, 해당 조우조건에서 우현변침을 통해 회피를 수행하는 것이 바람직하므로 $p_S$(starboard side turn)값을 $p_M$(maintain course) 또는 $p_P$(port side turn) 값에 비해 높게 가정하였다. 반면에, $p_P$값은 COLREGs에서 권고하는 회피방향과 반대로 충돌회피를 수행하는 것을 의미하는데, 현재 가관측성의 상태가 모호한(ambiguous)경우, 장애물의 회피행위 여부에 대한 불확실이 증가하므로, 이에 대한 반영을 위해 $p_P$값을 명백한(obvious) 조우상황에 비해 높게 가정하였다. 이와 같이 COLREGs에 부합하지 않는 회피행위까지 고려하는 것은 실제로 대부분의 선박 충돌사고가 전방주시의 의무를 다하지 않거나,

선박충돌예방규칙을 준수하지 않는 등의 인적과오에 의해서 발생하기때문이다. 중앙해양안전심판원에 따르면, 해양충돌사고의 주요원인은 경계소홀로 잠재원인은 졸음운항, 운항 중 다른 업무 수행 및 레이더 조작 미숙 등에 의해 발생한다고 조사되어진 바 있다. 또한, 선박 충돌사고중 가장 많은 비율을 차지하는 어선 충돌사고의 경우, 운항자 대부분이 해기지식이 부족하거나 생계형이라는 이유로 무면허 운항이 많기 때문에 안전관리가 취약하다고 조사되어진 바 있다.

| Probability | Head on (Obvious) | Head on (Ambiguous) | Crossing (SO) (Obvious) | Crossing (SO) (Ambiguous) |
|---|---|---|---|---|
| $p_M$ | 0.2 | 0.3 | 0.7 | 0.5 |
| $p_S$ | 0.7 | 0.5 | 0.2 | 0.3 |
| $p_P$ | 0.1 | 0.2 | 0.1 | 0.2 |

Table 5.5: Avoidance behavior begin time according to encounter situation and collision risk assessment method

Fig.5.27과 Fig.5.28은 각각 Head on과 Crossing (give way) 조우상황에 대한 Monte-Carlo 시뮬레이션 결과를 나타내고있다. 충돌회피의 성능을 나타내는 지수로 본 연구에서는 충돌회피상황에서 자선과 타선이 가장 가까워지는 순간의 거리값을 사용하였다. 만일 이 값이 특정 값보다 작다면 (기하학적 형상을 고려할 때, 본 연구에서는 15 m로 가정) 이는 물리적으로 충돌이 발생함을 의미한다. Fig.5.27과 Fig.5.28은 각 조우상황에 대한 1000번의 Monte-Carlo 시뮬레이션의 결과로 계산된 자선과 타선의 최소거리를 히스토그램 형태로 표현한 그림이다. 시뮬레이션 결과에 따르면, 대부분의 시뮬레이션의 경우 약 25m의 최소거리를 갖는것을

확인할 수 있는데, 이는 본 연구에서 충돌회피 경로기법인 velocity obstacle 알고리즘에서 25m를 최소거리로 설정하여 충돌회피를 수행하였기 때문이다. Head on 조우상황에 대한 시뮬레이션 결과(Fig.5.27)를 보면, 심층강화학습(DRL)을 이용한 경우와 Set based Guidance(SBG)을 이용한 경우에는 약 1000회의 시뮬레이션 중 약 980회 이상 최소거리 25m 이상의 결과를 얻는것을 확인하였다. 반면에, CPA를 이용한 경우, 25m 이내의 최소거리를 갖는 시뮬레이션 케이스가 다수 존재하는 것을 확인할 수 있는데, 이는 CPA를 이용할 경우, 타선의 가관측성의 저하에 따라 모호한(ambiguous) 상태로 인식되어 타선이 높은 확률로 COLREGs에 위배되는 회피행위 (침로유지 또는 좌현변침)을 수행하였기 때문인것으로 파악된다. 25m 이내의 최소거리를 갖는것은, 충돌회피 알고리즘에서 설정한 최소유지거리를 위반하는 것으로, 상대적으로 위험한 조우상황을 의미한다. 시뮬레이션 결과, 상대선박이 진로를 유지하거나 COLREGs에 부합하는 충돌회피를 수행하는경우, 타선과 자선의 최소거리가 25m 이내로 계측되는 상황은 단 한번도 발생하지 않았다. 타선과 자선의 최소거리가 25m 이내가 되는 경우는 모두 타선이 COLREGs에서 권고하는 반대방향으로 충돌을 회피하는 상황에서 발생하는 것을 확인하였다.

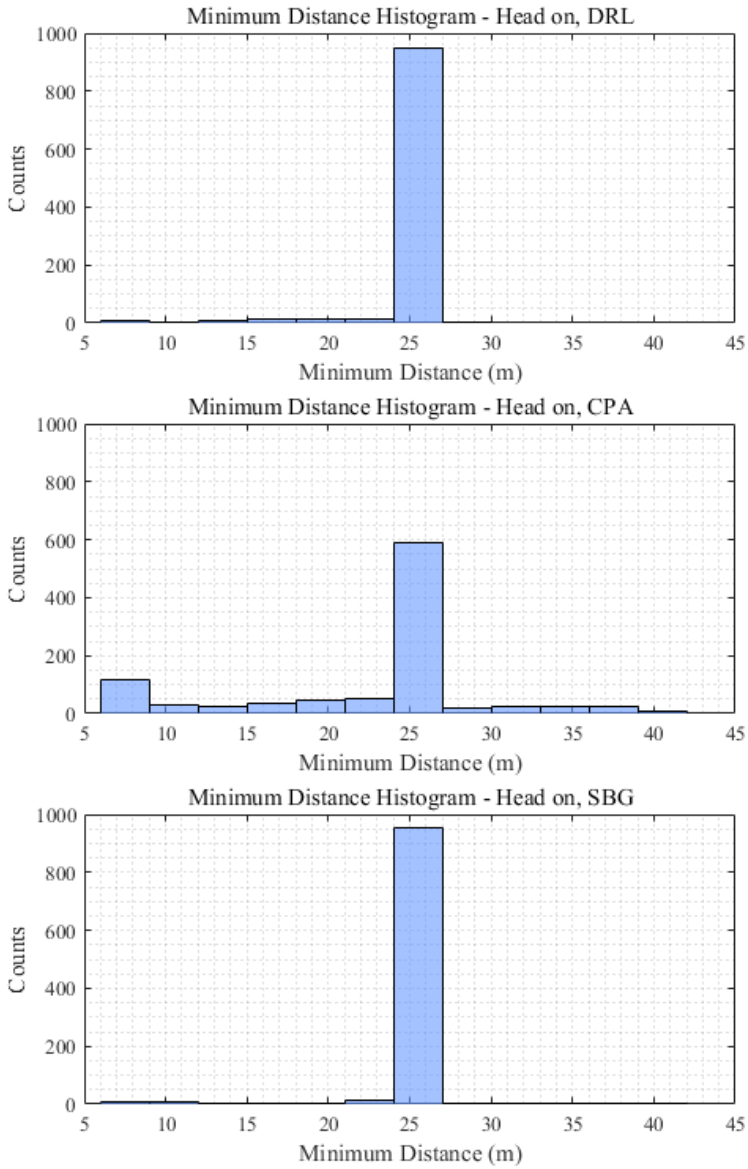



Fig 5.27: Histogram of the Monte-Carlo simulation result for Head on encounter situation

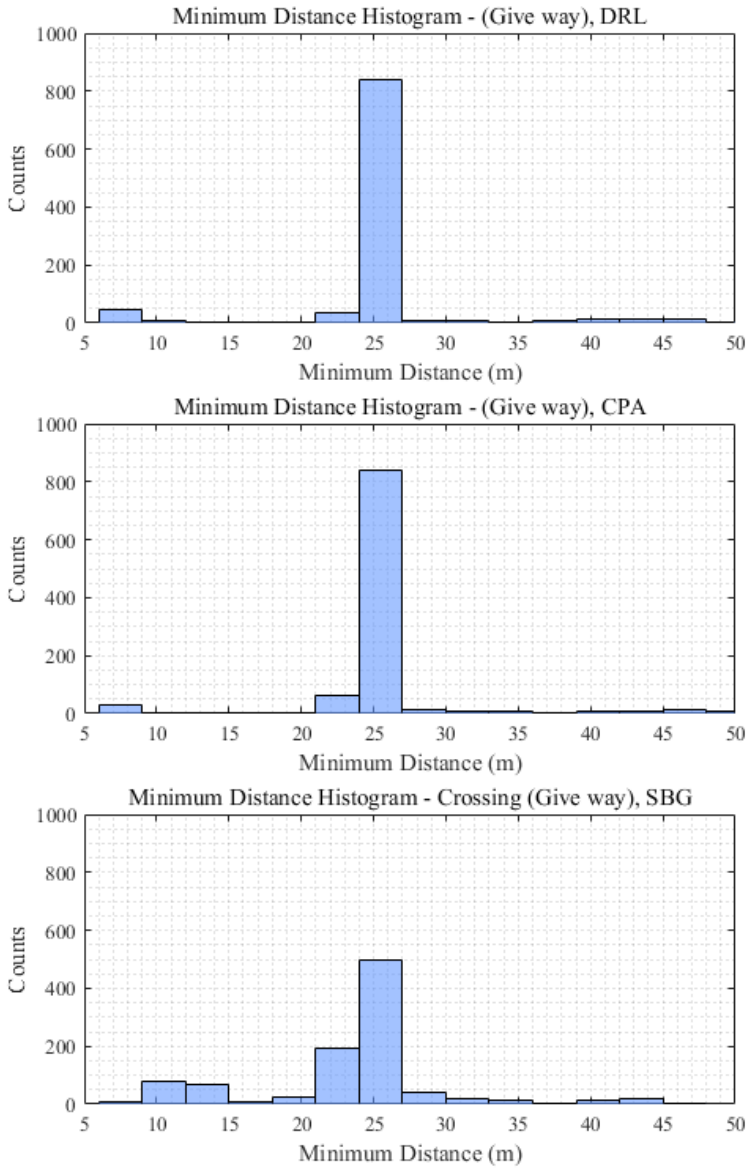



Fig 5.28: Histogram of the Monte-Carlo simulation result for Crossing (give way) encounter situation

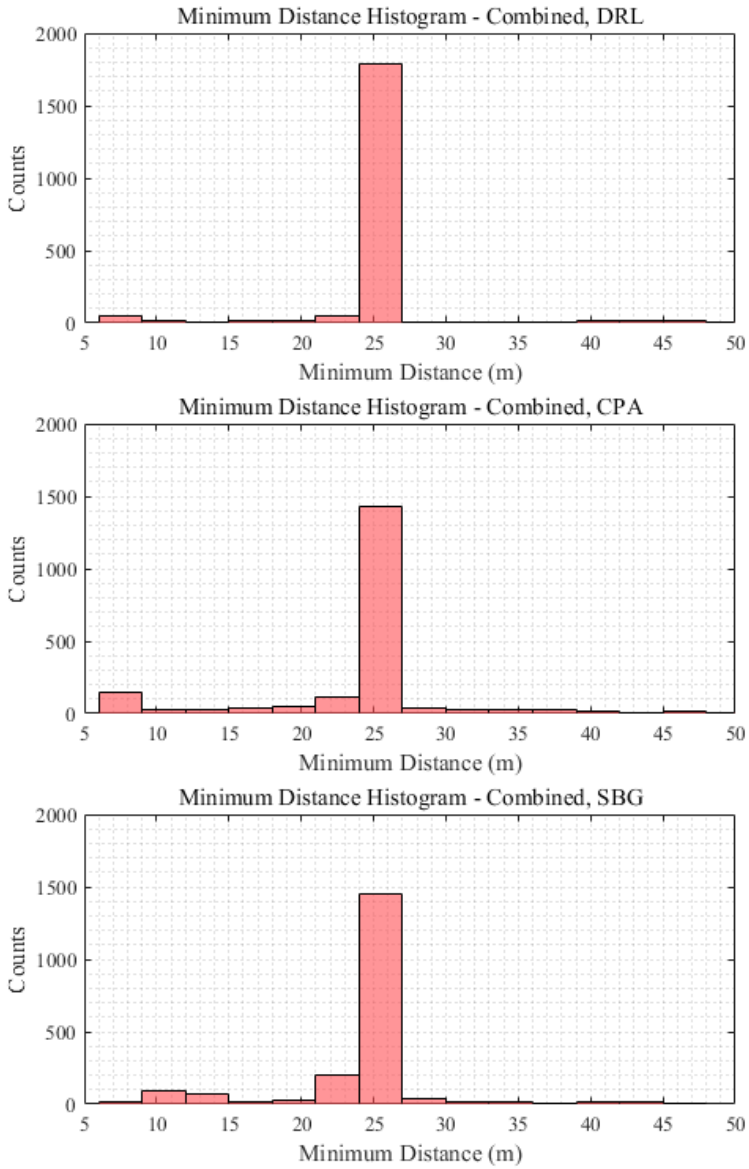



Fig 5.29: Histogram of the Monte-Carlo simulation result for Head on and Crossing encounter situation

이와 같은 현상은 Crossing (give way) 조우상황의 시뮬레이션 결과에서도 확인할 수 있다. SBG알고리즘을 사용할 경우, 심층강화학습(DRL) 또는 CPA를 이용한 경우에 비해 25m 이내의 최소거리를 갖는 빈도수가 증가하는 것을 확인할 수 있었다. 이는, SBG의 경우 상대적으로 늦은 자선의 충돌회피 시점으로 인하여 타선이 자선의 충돌회피 의사를 명확하게 인식하지 못하였기 때문이다. 즉, 현재 상태를 모호한 조우상태로 인식되어 COLREGs에 위배되는 회피행위를 보다 빈번히 선택하였기 때문인 것으로 판단된다.

Fig.5.29는 Head on과 Crossing (give way)상황의 시뮬레이션 결과를 누적하여 만든 히스토그램을 나타내고있다. 심층강화학습 기반 충돌회피기법의 경우, 조우상황에 따라 다른 기준을 적용하여 회피를 수행하고, 타선의 자선에 대한 가관측성을 고려하여 충돌회피를 명확한 단계로 유지시킬수 있음을 확인하였다. 그에 대한 결과로, Fig.5.29에서 DRL기법이 25m이상의 최소거리를 가장 많이 발생시킨것을 확인할 수 있다.

Fig.5.30은 Monte-Carlo 시뮬레이션을 통해 도출된 충돌회피 성공확률 결과이다. 앞서 계산된 최소거리정보를 이용하여, 최소거리가 15m이내인 경우, 자선과 타선이 충돌했다고 가정하고 총 충돌한 사건의 빈도수를 계산하여 확률을 도출하였다. Head on 상황의 경우, DRL, CPA, SBG기법의 충돌회피 성공확률이 98.2%, 82.6%, 98.4%로 계산되었다. Crossing (give way)상황의 경우 DRL, CPA, SBG 기법의 충돌회피 성공확률은 94.5%, 96.5%, 83.9%로 계산되었고, Head on 과 Crossing (give way)상황을 누적하여 얻은 충돌회피 성공확률은 DRL, CPA, SBG 기법 각각 96.3%, 89.5%, 91.15%로 도출되었다.

Monte-Carlo 시뮬레이션 결과, 심층강화학습을 이용한 경우(DRL) 가장 높은 충돌회피 성공확률을 얻는것을 확인할 수 있었는데, 이에 대한 가장 큰 원인은 심

층강화학습을 사용한 경우, 조우상황별로 각기 다른 기준을 적용하여 충돌회피를 수행함으로써 각 조우상황에서 타선의 자선에 대한 가관측성의 영향을 고려할 수 있기 때문인것으로 판단된다. 이로 인해 타선이 자선을 인식할 때, 조우상황을 모호한 상황이 아닌 명확한 상황으로 인식하도록 유도하여, 불확실성을 감소시키고 타선의 불안전한 회피행위 빈도를 저감시켜 충돌회피를 수행한 결과로 해석할 수 있다.

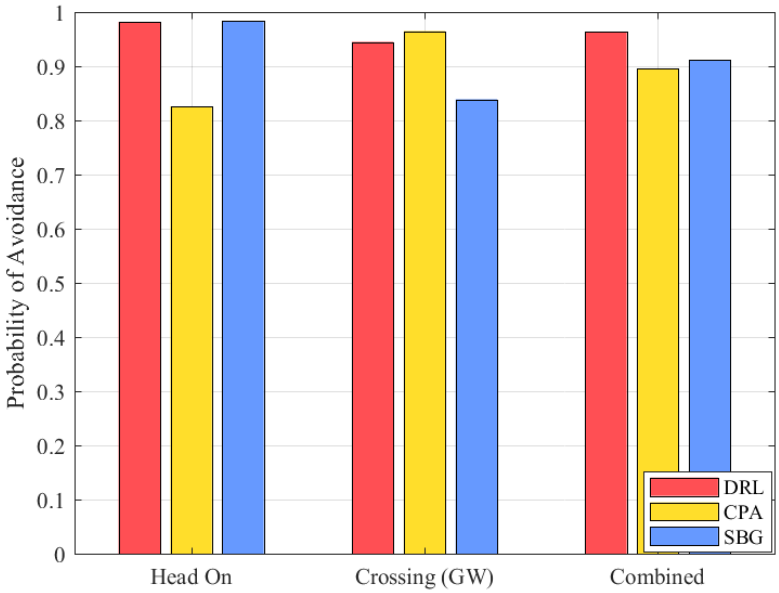



Fig 5.30: Probability of the collision avoidance with respect to collision risk assessment methods

### 5.2.3 조종성능의 영향에 대한 분석

**자선의 조종성능의 영향**

기존의 충돌위험도는 장애물과의 TCPA와 DCPA만을 고려하기 때문에, 자선의 조종성능이나 회피능력을 고려할수 없다. 반면에, 심층강화학습 기반 충돌회피 기법은 환경과의 상호작용을 통해 조종성능이 고려된 환경에서 학습이 이루어지므로, 충돌회피 시점을 결정시 자선의 조종성능을 반영시킬 수 있다. 이에 대한 검증을 위해, 각기 다른 조종성을 가지는 선박에 대한 충돌회피 시뮬레이션을 수행하였다. 실선의 조종모델을 Model.A라 정의하고, Model.A의 제어력을 발생시키는데 사용되는 계수에 0.5의 상수배를하여 조종성능을 저감 시킨 모델을 Model.B라고 정의하였다. 조종성능에 따른 차이를 확인하기위해, Model.B에 대한 학습을 수행하고, 각 모델에 대하여 충돌회피 시뮬레이션을 수행하였다. Fig.5.31은 각기다른 조종성능을 가지고 있는 두 종류의 선박이 Head on 조우상황에 대하여 충돌회피 시뮬레이션을 수행한 결과이다. 시뮬레이션 결과, Model.A의 경우 31초에, Model.B의 경우 27.5초에 충돌회피가 이루어지는것을 확인할 수 있었다. 이는 Model.B의 경우 부족한 자선의 조종성능을 고려하여 회피행위를 미리 시작하여 보상함수를 최대화 하는 방향으로 학습이 수행된것을 의미한다. Crossing (stand on)상황에 대해서도 충돌회피 시뮬레이션을 수행하였다.(Fig.5.32) 시뮬레이션의 조건은 타선이 (-66,100)의 위치에서 시작하여 1m/s으로 0도의 침로각도를 가지며 접근하는 상황이고, 자선은 (0,0)의 위치에서 시작하여 90도의 침로각도로 항해하는 상황이다. 시뮬레이션 결과, Model.A의 경우 충돌회피가 48.5초에 시작된 반면, Model.B의 경우 44.3초에 충돌회피가 시작되었다. 이는 부족한 조종성능을 반영하여 Model.A보다 앞선 시점에 회피기동을 수행하는 것을 의미한다.

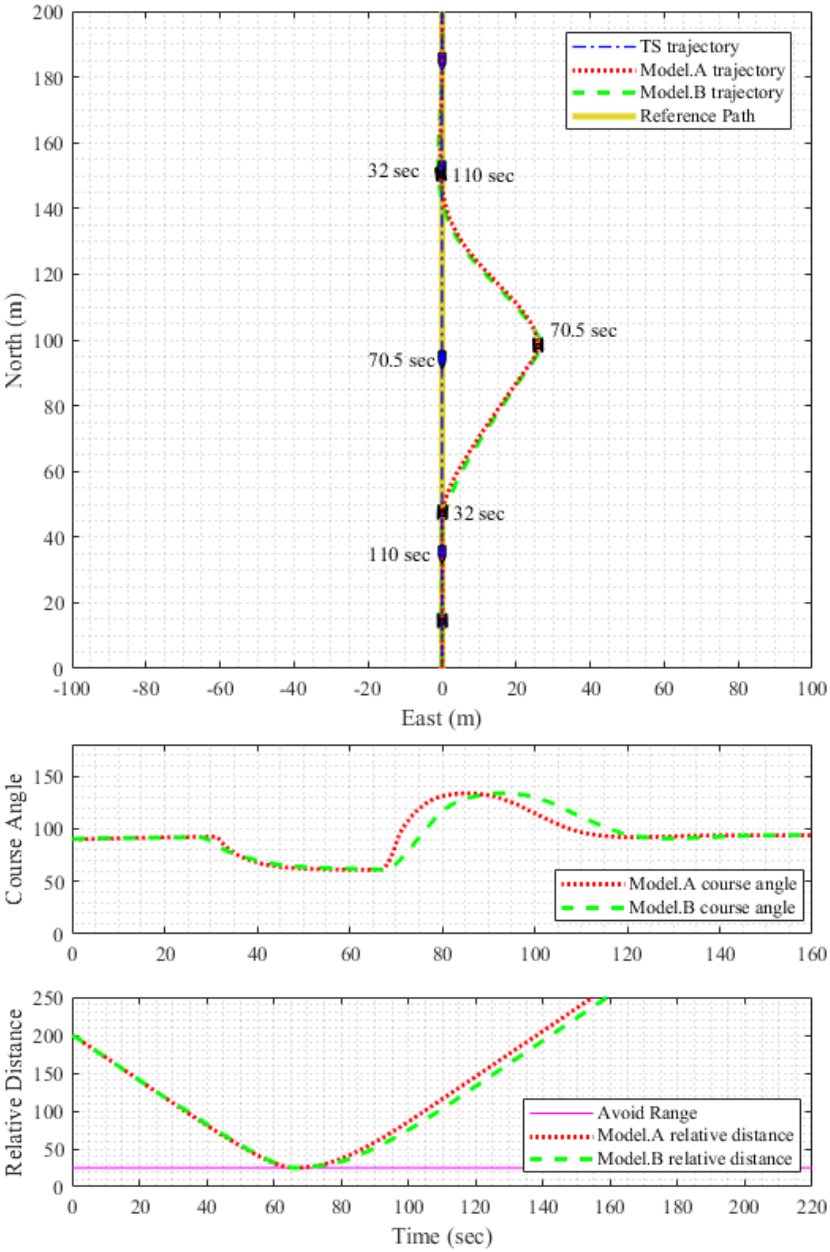



Fig 5.31: Collision avoidance simulation result with respect to ship with different maneuverability (Head on)

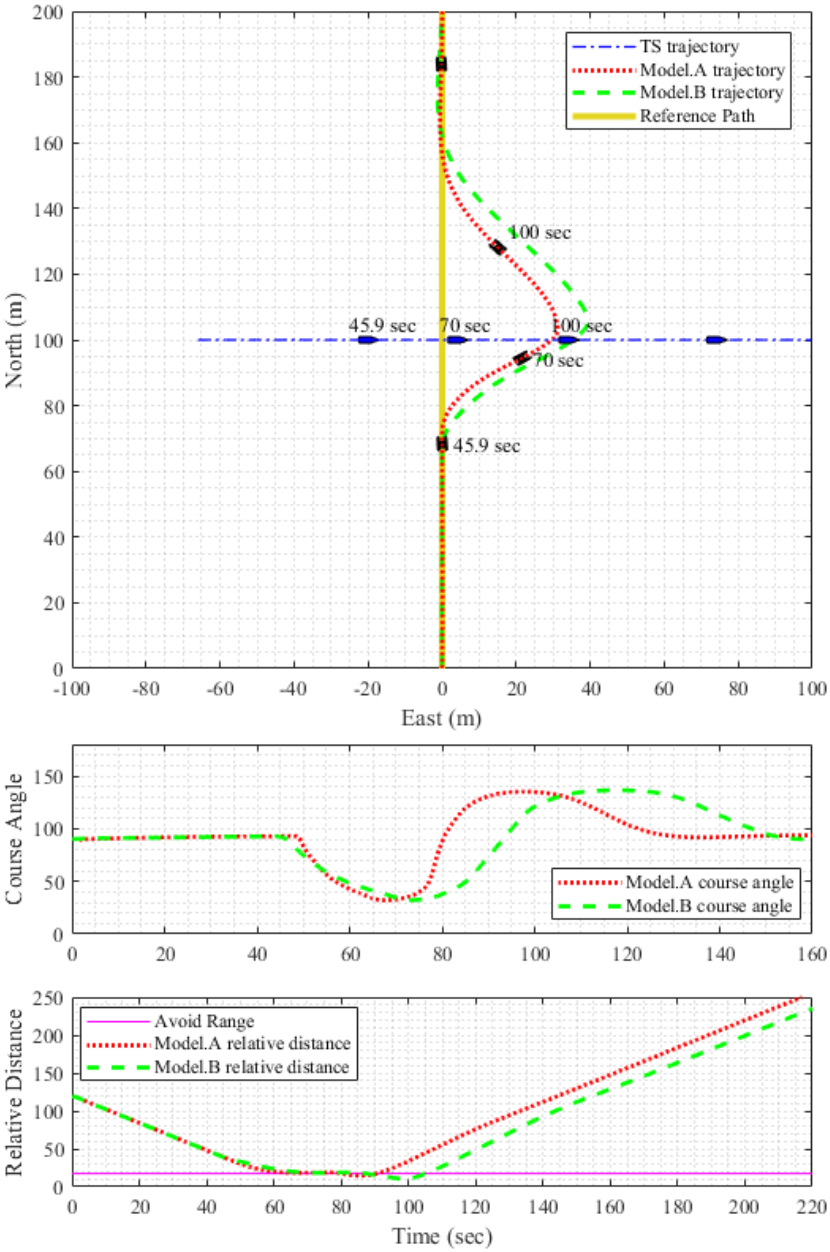



Fig 5.32: Collision avoidance simulation result with respect to ship with different maneuverability (Crissong, give way)

**환경하중의 영향**

심층강화학습 기반 충돌회피 기법은 충돌회피 시점결정 시 환경하중의 영향에 대한 요소를 고려할 수 있다. 환경하중에 대한 요소를 고려하는 과정은 조종성능에 대한 영향을 고려하는 과정과 유사한 방식으로 이루어지는데, 환경하중이 작용하는 환경하에서 충돌회피 강화학습을 반복적으로 수행함으로써 이루어진다. 환경하중이 충돌회피 시점에 미치는 영향을 확인하기 위해서, 여 6.7m/s의 풍하중이 각각 동, 서, 북쪽 방향에서 입사하는 환경하에 Head on 시나리오에 대한 충돌회피 시뮬레이션을 수행하였다. Table.5.6은 시뮬레이션을 통해 얻어진 각 알고리즘의 충돌회피 시점을 나타내고있고, Fig.5.33, Fig.5.34, Fig.5.35는 풍향이 서쪽, 동쪽, 북쪽 인 경우에 대한 시뮬레이션 결과를 나타내고있다. 시뮬레이션 결과, CPA를 이용하여 충돌회피를 수행하였을 때 충돌회피는 풍향에 관계없이 약 7.2초에 시작되는것을 확인할 수 있었다. 풍향에 관계없이 일정한 시점으로 추론되는 이러한 현상은 SB기법을 이용하여 회피시점을 추론한 결과에서도 동일하게 확인할 수 있었다. 반면, 심층강화학습을 이용한 결과는 풍향에 따라서 회피시점이 결정되는것을 확인할 수 있었는데, 이는 충돌회피시점 결정과정에서 자선의 침로변경능력을 경험적으로 반영한 결과라고 볼 수 있다. 환경하중이 작용하는 방향이 회피기동에 도움을 주는 경우(Fig.5.33), 이른 시점에 충돌회피가 수행되는것을 확인할 수 있지만, 환경하중의 방향이 회피기동성능을 저해하는 방향에서 작용하는경우 (Fig.5.34) 주어진 회피조건(30도 이상의 변침각)을 만족하기위해 일정시간 경로추종을 유지한 뒤, 변침이 시작되는현상을 확인할 수 있었다.

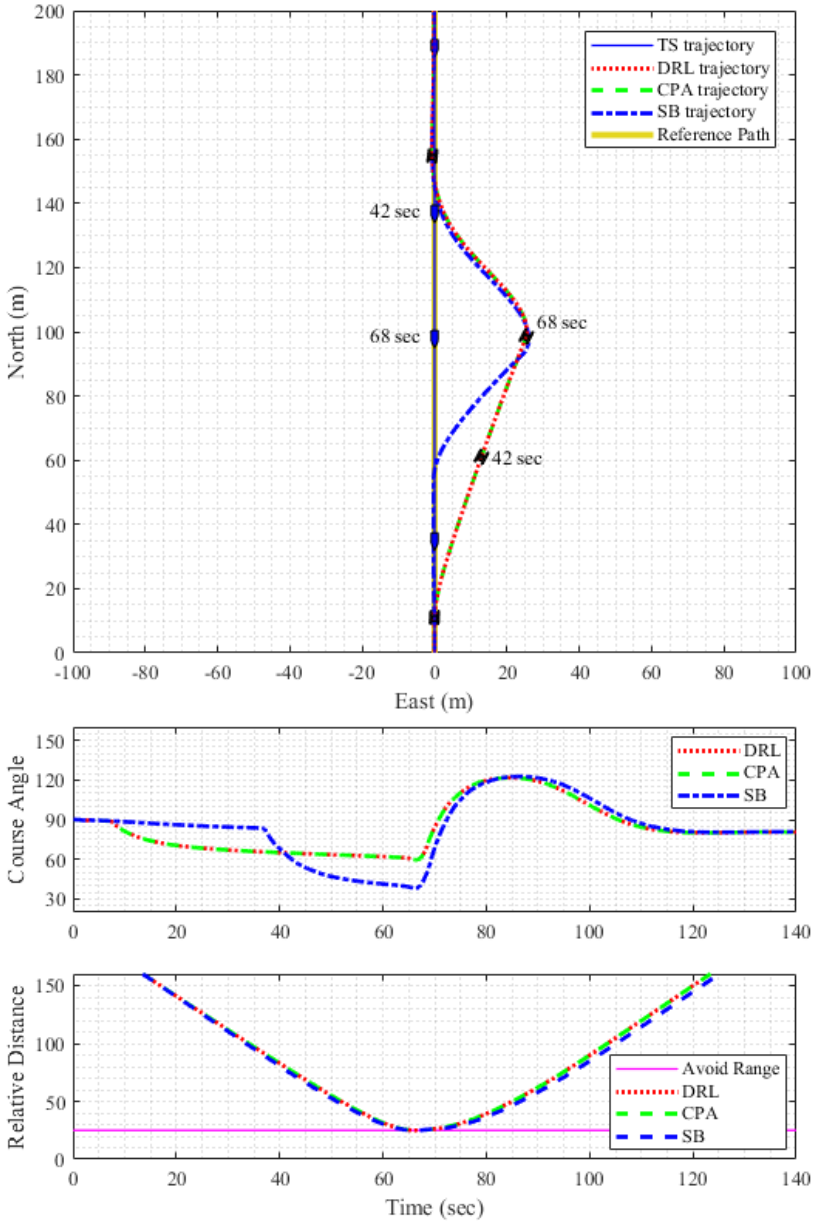



Fig 5.33: Collision avoidance simulation result using different collisin risk assessment method under 6.7m/s wind from West

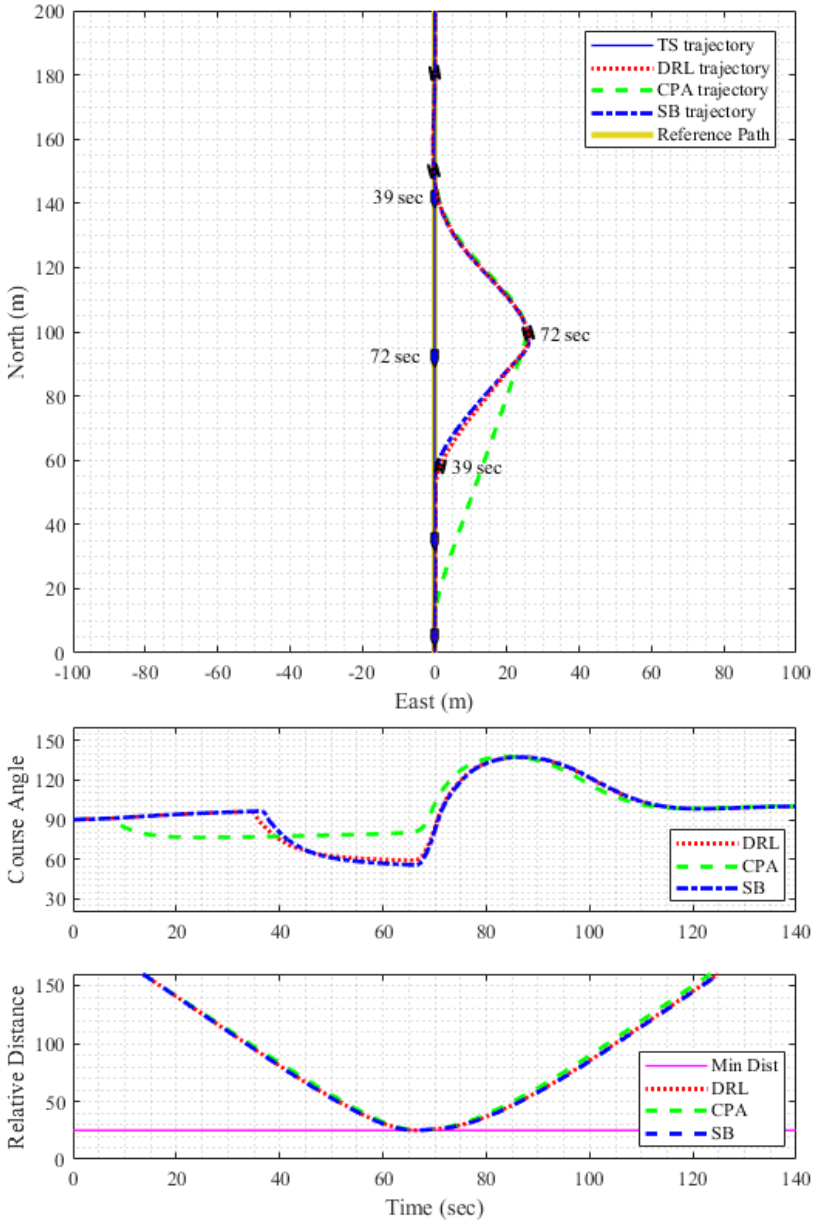



Fig 5.34: Collision avoidance simulation result using different collisin risk assessment method under 6.7m/s wind from East

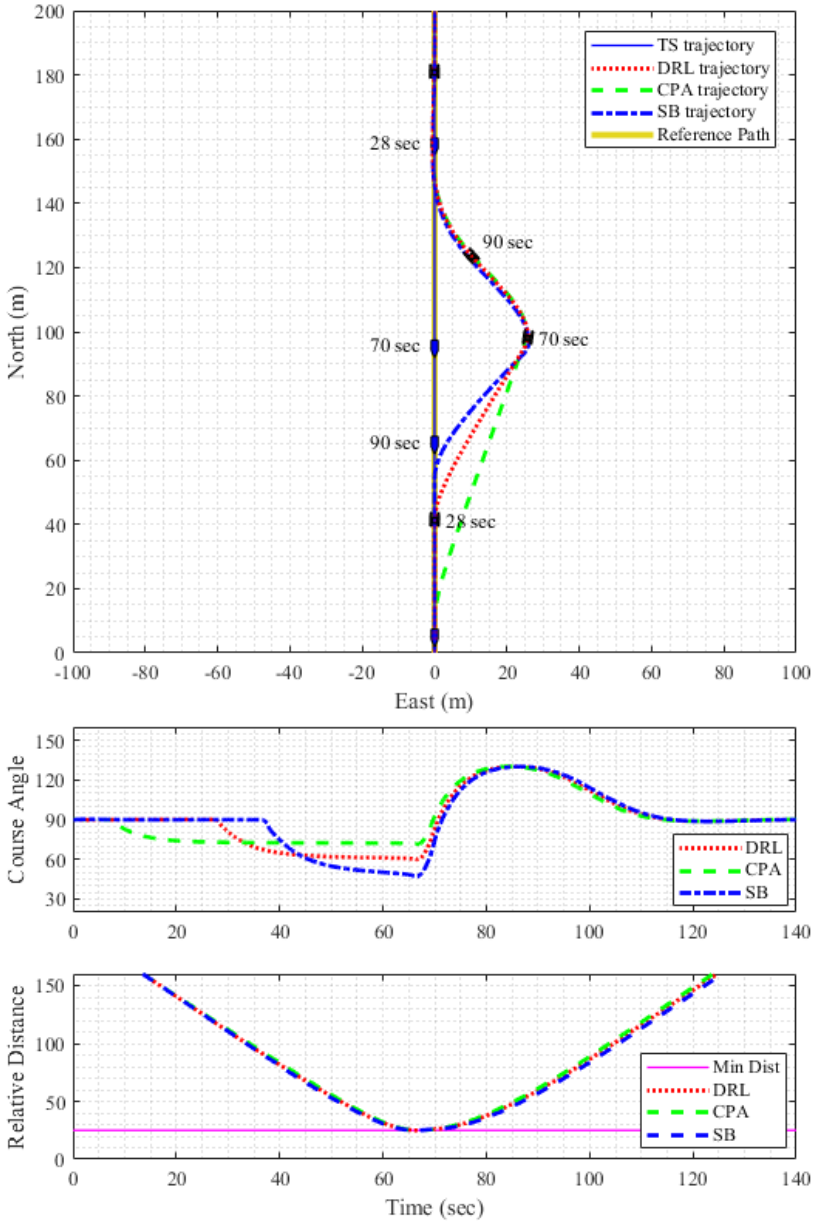



Fig 5.35: Collision avoidance simulation result using different collisin risk assessment method under 6.7m/s wind from North

| Method | 6.7 m/s Wind from West | 6.7 m/s Wind from North | 6.7 m/s Wind from East |
|---|---|---|---|
| CPA | 7.2 sec | 7.2 sec | 7.2 sec |
| SB | 36.6 sec | 36.6 sec | 36.6 sec |
| DRL | 7.2 sec | 27.5 sec | 34.5 sec |

Table 5.6: Avoidance behavior begin time with respect to wind direction and collision risk assessment method

### 5.2.4 다수 선박에 대한 충돌회피

선박의 충돌회피 의사결정은 조우상황의 복잡도에 의존적인 특성을 갖는다. 즉, 회피 대상선박이 한 대가 아닌 다수가 존재하고, 보다 많은 상황적 인자들을 고려해야 할 필요성이 있는경우 단일 선박에 비해 높은 불확실성을 갖는 특성이 있다. 이는 과거시점에 결정한 의사결정이 미래의 영향을 끼치는 의사결정 문제의 특징에 기인하는데, 현재시점에서 바람직한 충돌회피 행위가 미래시점에서는 불안전한 상황을 야기시킬 수 있음을 의미한다. 이러한 요소들을 고려하기위해서는 복잡한 조우상황에 대한 다양한 정보들이 요구된다.

Fig.5.36과 같은 조우상황을 예를 들어보면, 자선의 정면에는 타선 1대가 자선과 동일한 방향으로 운항하고 있다. 경로를 안정적으로 추종하기위해서는 타선을 추월해야 하는데, 이는 Overtaking (give way)의 조우상황이므로 좌현 또는 우현방향으로 변침이 모두 가능하다. 대상선박의 위치가 경로를 기준으로 좌측에 위치하므로, 대상선박만을 고려한다면 우현변침을 통해 회피를 수행하는 것이 바람직한 의사결정일 것이다. 하지만, 우현변침을 수행한다면 우측상단에서 접근하

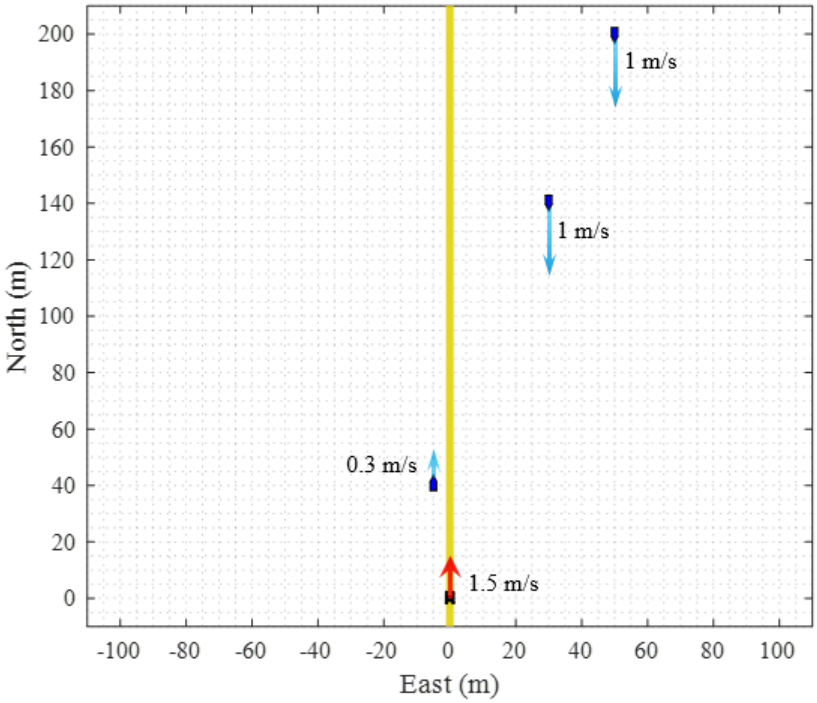



Fig 5.36: Example of complex multiple ship encounter situation

고있는 선박과 조우하게되고, 이를 회피하기위해 다시 연속적으로 충돌회피단계를 수행해야 한다. 이와 같이, 다수 선박에 대한 충돌회피 문제의 경우, 현재 시점에서 바람직한 의사결정이 미래 시점에서는 오히려 선박의 안전성을 저해시키는 결과를 야기시킬 수 있다는 특징이 존재한다. 이러한 복잡성에 대응하기위해서는 복잡한 조우상황에 대한 정보를 바탕으로 높은수준의 의사결정을 수행해야 한다. 하지만 기존 충돌위험도 추론기법 등의 경우, 이러한 조우상황을 지나치게 단순화 하여 의사결정에 필요한 정보가 부족한 현상이 발생하기도 한다. 예를들면, DCPA와 TCPA를 이용하여 Fig.5.36의 조우상황을 표현하면, 단 6개의 숫자로 조우상황을 표현하게되는데, 이 6개 숫자의 조합을 만족하는 조우상황은 유일하지 않고 무수

히 많은 특징이 존재한다. 즉, 6개의 특징은 현시점의 고유한 조우상황을 표현하지 못한다. 반면에, 본 연구에서 제안하는 심층강화학습 기반 충돌회피 기법은 조우상황 자체를 의사결정단계의 입력변수로 사용하므로 현재의 복잡한 조우상황을 고유하게 표현할 수 있다는 장점이 있다.

Fig.5.39와 Fig.5.37은 각각 CPA 기법과 심층강화학습 기반 기법을 이용하여 Fig.5.36에서 표현한 예제에 대하여 충돌회피를 수행한 결과를 나타내고있다. CPA를 이용한 경우(Fig.5.39), 회피초기단계에 전방에서 조우한 타선을 회피하기위해 우현변침을 수행하였다. 이는 타선의 위치가 회피시점을 기준으로 좌현방향에 위치하므로, 회피를 위해 최소노력이 소요될 것으로 판단되는 우현변침을 통해 회피를 수행한것이었다. 하지만, 우현변침을 통한 회피 중 다른 타선들을 조우하게되고, 이들을 회피하기위해 연속적으로 충돌회피기동을 수행하는 것을 확인할 수 있다. 이때, Overtaking 상황에서 우현변침을 선택함으로써 최대 횡방향 오차는 78.2 m로 계산되었고, 171초동안 경로를 벗어나있었다. 심층강화학습일 이용한 경우(Fig.5.37), 회피를 결정하는 시점(6sec)에서 우측 상단의 장애물에 대하여 고려하여, 좌현변침을 수행하여 Overtaking 조우상황에 대한 충돌회피를 수행하는 것을 확인할 수 있었다. 이를 통해 최대 횡방향 오차는 31m로 계산되었고, 약 87초동안 경로를 벗어나 있었던것으로 파악되었다. Fig.5.38과 Fig.5.40은 각각 CPA와 심층강화학습을 이용한 경우 각 시점에서의 자선과 타선의 위치 및 궤적을 표현하고 있다.

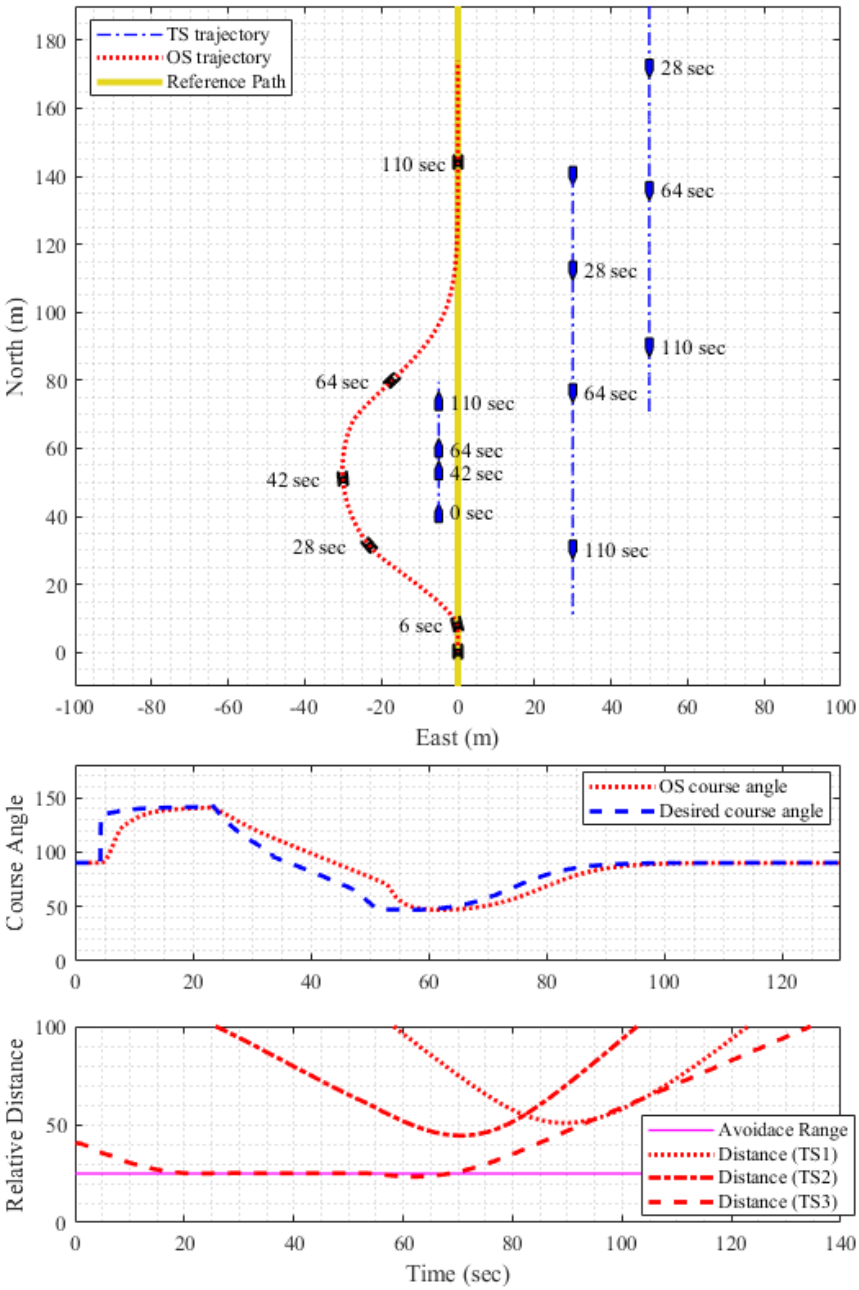



Fig 5.37: Trajectory of multiple ship collision avoidance simulation using DRL

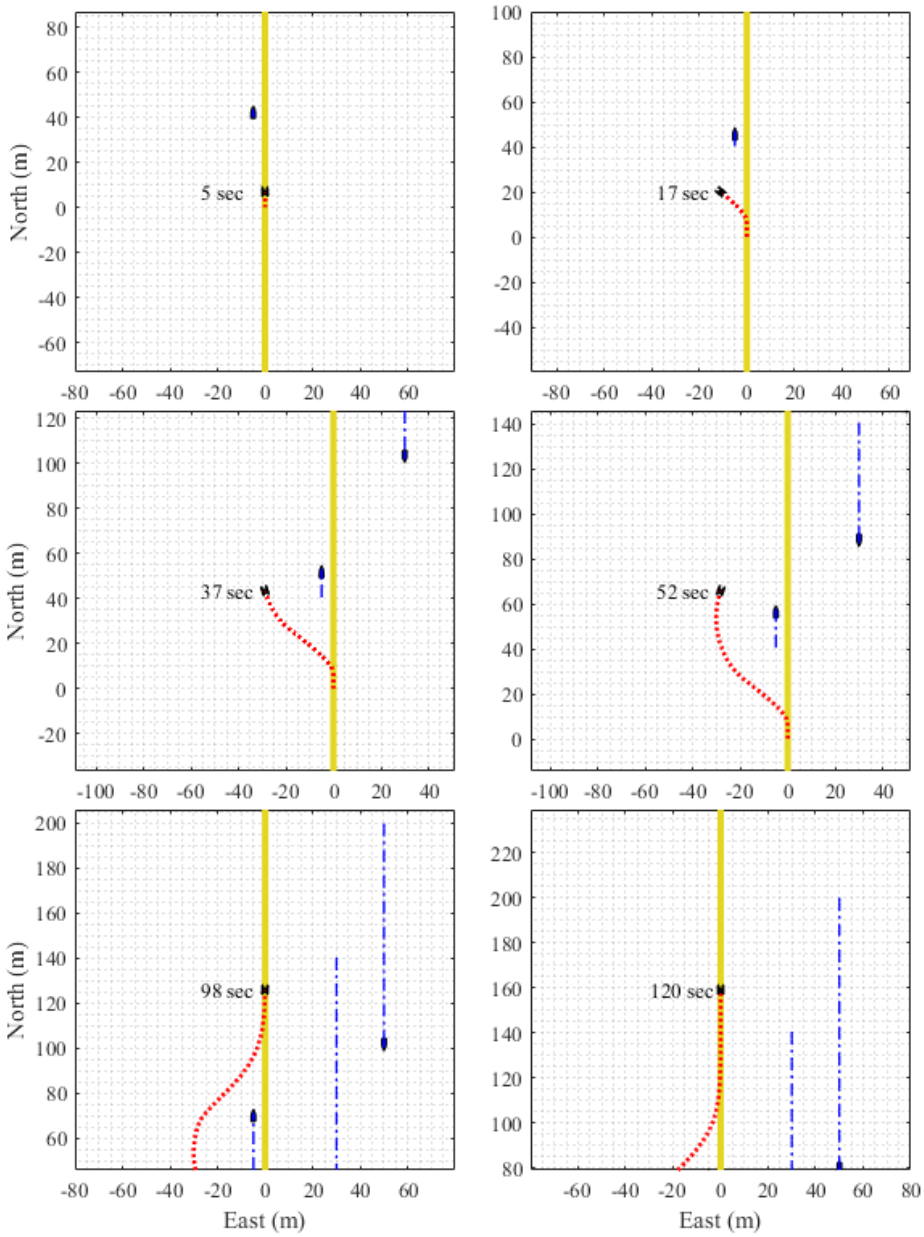



Fig 5.38: Snapshot of multiple ship collision avoidance simulation using DRL

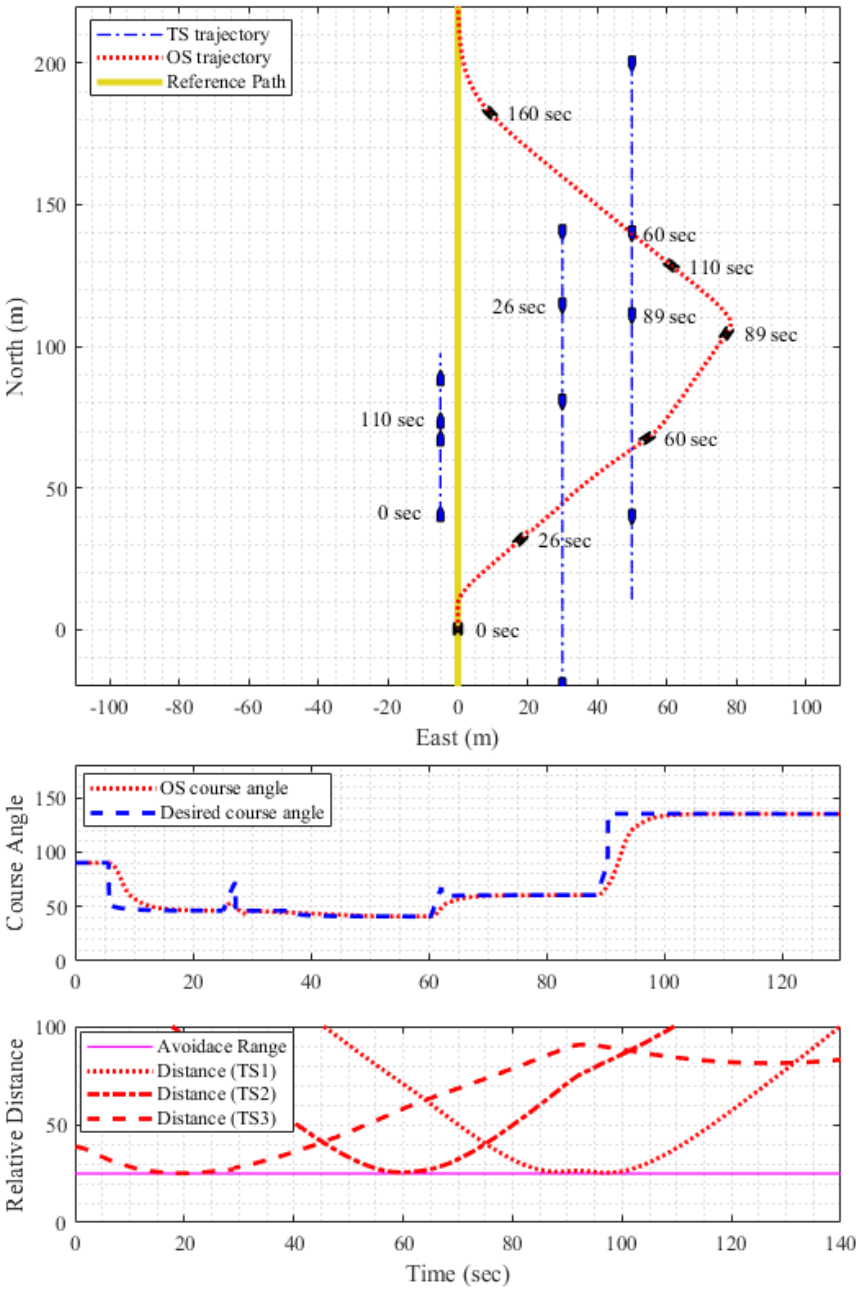



Fig 5.39: Trajectory of multiple ship collision avoidance simulation using CPA

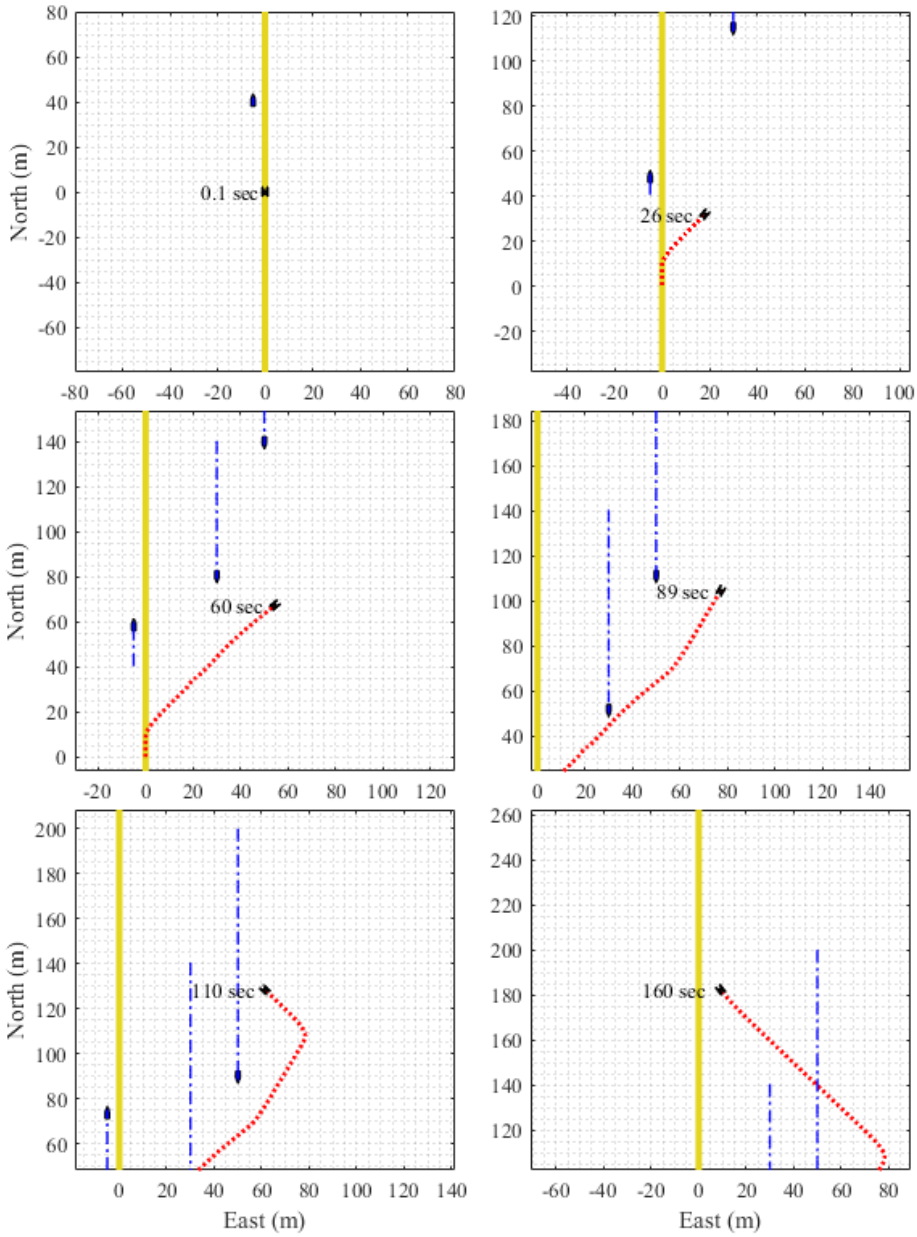



Fig 5.40: Snapshot of multiple ship collision avoidance simulation using CPA

Fig.5.41은 다수 선박이 존재하는 새로운 조우상황을 나타내고있다. 조우상황의 초기에는 자선의 우현에서 접근중인 선박을 Crossing (give way) 조우상황으로 간주하고 회피를 수행하게된다. 충돌회피를 수행하면, Fig.5.41의 우측 상단에 위치해있던 타선들이 자선에 접근하게되고, 이를 회피하기위해 연속적으로 우현변침을 수행해야하는 조우상황이다. 해당 예제는 이전 예제와는 다르게, 모든 조우상황에 대하여 회피방향이 COLREGs에 의해 결정되므로, 좌현 또는 우현변침 여부를 선택할 수는 없다. 다만, Fig.5.42의 결과를 확인해보면, 연속적으로 장애물이 존재하는 상황에서도 이들을 인식하여 충돌회피대상을 변경할 수 있음을 확인할 수 있다. Fig.5.42는 전체 시뮬레이션의 궤적을, Fig.5.43은 특정 시점에 대한 자선과 타선의 위치와 궤적을 나타내고있다.

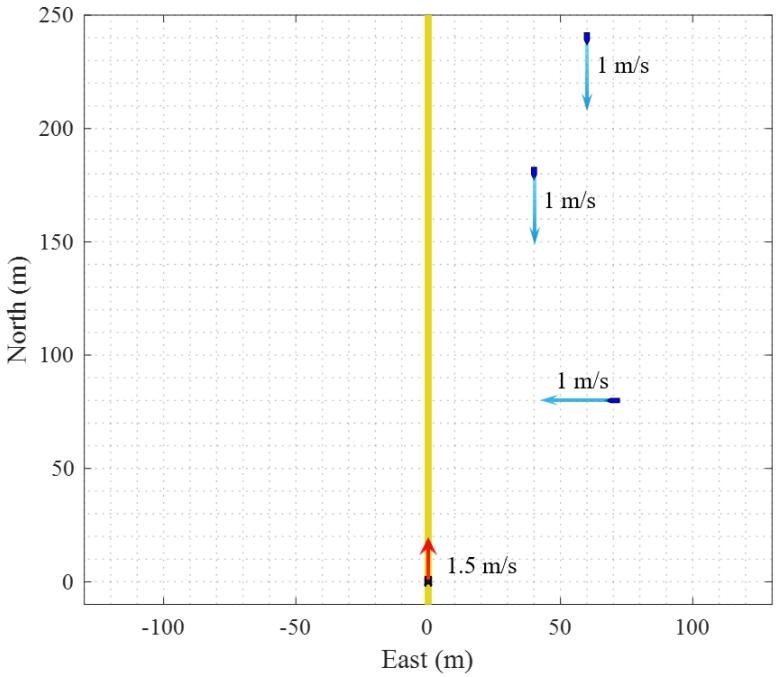



Fig 5.41: Example of complex multiple ship encounter situation (II)

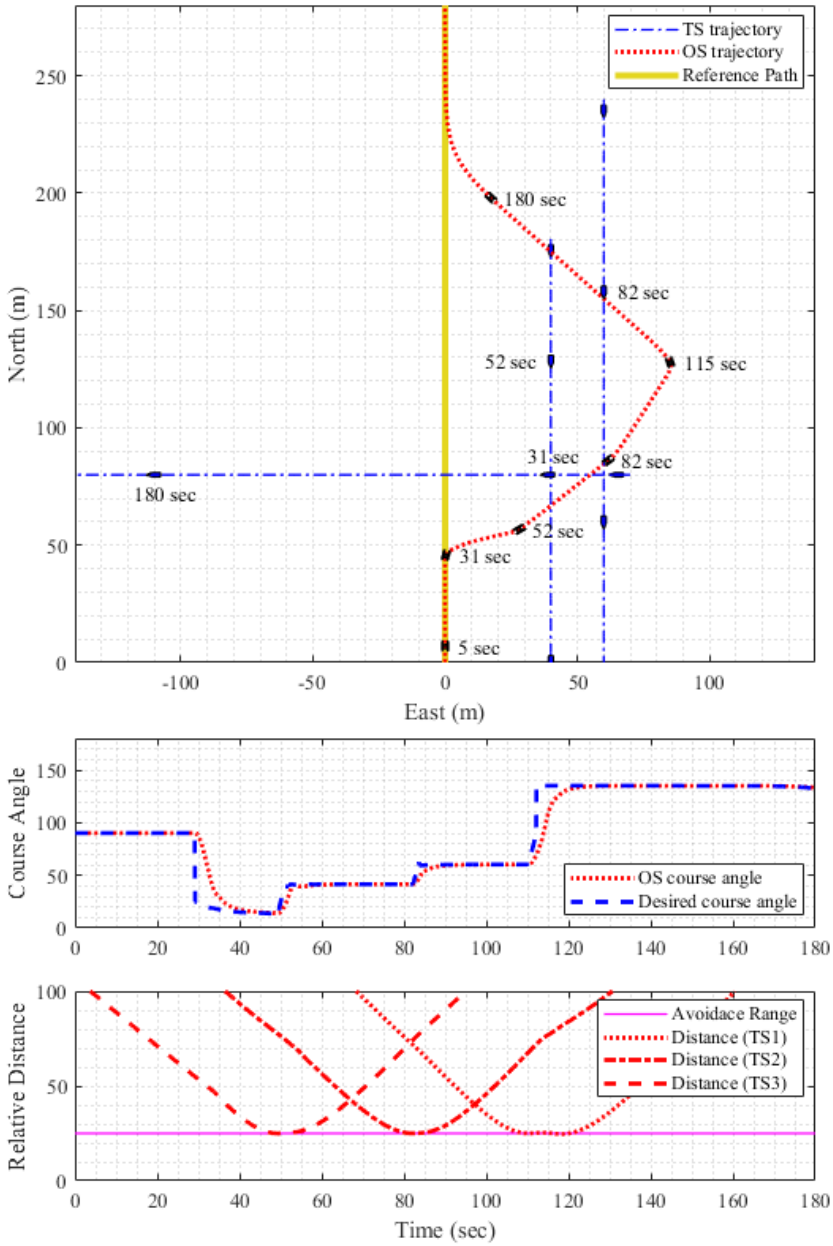



Fig 5.42: Trajectory of multiple ship collision avoidance simulation case (II) using DRL

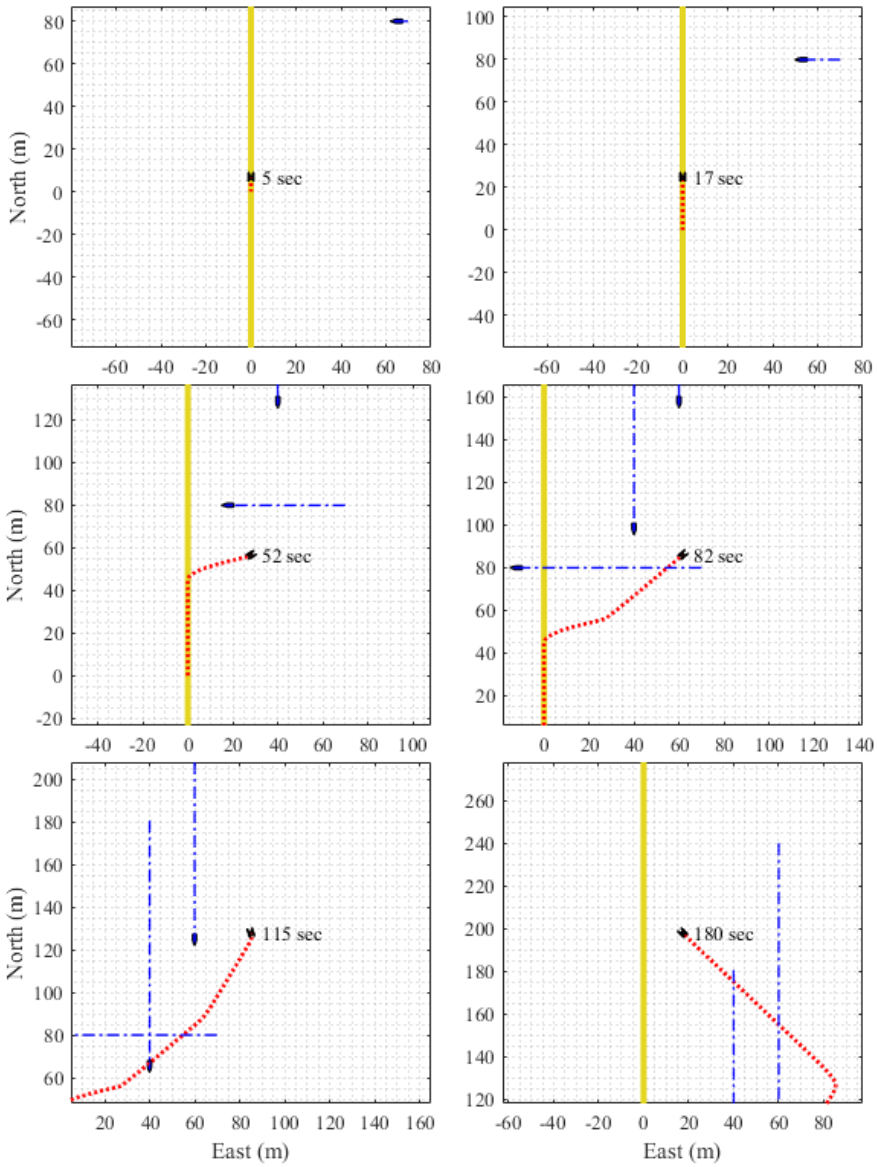



Fig 5.43: Snapshot of multiple ship collision avoidance simulation case (II) using DRL

## 5.3 자유 항주 실험 결과

### 5.3.1 실험 환경

Fig 5.44: Test environment in Lake Pyoungtaek)

본 절에서는 5.2절 에서 시뮬레이션을 통해 검증하였던 심층강화학습 기반 충돌회피 알고리즘에 대한 실험적 검증을 수행하였다. 실험을 위하여 시뮬레이션 상에서 학습된 심층강화학습 모형을 추출하여 무인수상선 실선 시스템에 적용시켰다. 추출된 심층강화학습 모형은 간단한 행렬연산을 통해 각 행동에 대한 행동가치함수를 계산할 수 있다. 자선의 항법정보는 무인수상선에 탑재된 항법센서를 통해 계측하였고, 장애물 정보의 경우, 실험의 단순화를 위하여 가상의 장애물이 접근한다고 가정하였다. 즉, 장애물의 위치, 자세, 속도정보는 안정적으로 탐지, 추적 되어 그 결과값이 바로 제공된다고 가정하였다. 실제 실험에서는 장애물 정보를

시간의 함수로 사전에 정의하고, 무인수상선에 탑재된 컴퓨터에 의해서 장애물의 위치, 자세 정보를 계산하여 충돌회피에 활용하였다. 심층강화학습 모듈에 의하여 유도명령이 선택되면, 이들 명령을 수행하기위한 속도제어기와 침로각 제어기가 제어명령을 생성하고, 이를 무인수상선의 선외기에 전달하여 운동을 발생시켰다. 실험은 경기도 평택호에서 수행되었으며, 5.1절에서 설명한 무인수상선 실선을 이용하였다. Fig.5.44는 실험환경을 나타내고있다.

### 5.3.2 충돌 회피 실험 결과

충돌회피 성능에 대한 실험적 검증을 수행하기위하여 단일 선박에 대한 충돌회피 실험을 수행하였다. 실험 환경은 각각 Head on 과 Overtaking 조우상황, 다수선박에 대한 조우상황으로 설정하였다.

#### Head on 조우상황

Head on 조우상황은 타선이 자선과 마주보며 0.8 m/s의 속도로 접근하고 있는 상황을 선정하였다. Head on 조우상황에 대한 실험 결과는 Fig.5.45에서 확인할 수 있다. 결과 그래프의 빨간색 점선은 자유항주실험결과를 나타내고, 초록색 실선은 자유항주실험과 동일한 초기조건 하에서 수행된 시뮬레이션의 결과를 나타내고있다. 실험결과, 자선과 타선간의 상대거리가 약 65m가 되는 시점에서 충돌회피가 이루어지는 것을 확인할 수 있었고, 회피단계에서 계측된 최대 변침각은 46° 인 것으로 측정되었다. Head on 조우상황에서 바람직한 변침각은 30도로 설정되었지만, 실제 계측된 변침각은 이보다 높은 값을 가지는것을 확인할 수 있었는데, 이는 시뮬레이션상에서 고려하지 못했던 제어기의 특성때문인것으로 파악되는데, 실험결과의 16초 부근에서 발생한 침로각의 오버슈트(overshoot)가 최대침로각

의 크기를 증가시키는 원인으로 작용하였다. 본 연구에서 구현한 심층강화학습 모듈은 오로지 유도명령만을 전환(switching)시키는 역할을 수행하기 때문에, 시뮬레이션과 실험의 환경을 동일하게 구현하려면 제어기가 포함된 폐루프 시스템(closed loop system)에 대한 식별을 수행해야 한다. 하지만 본 연구에서는 선박 및 구동기의 동역학모델에 대한 개루프 시스템(open loop system)에 대한 식별만을 수행하였고 이에 따른 결과가 충돌회피 실험 과 시뮬레이션 의 차이를 발생시킨것으로 분석된다. 실험결과의 30초 부근에서는, 시뮬레이션에서는 발생하지 않았던 선수각 동요(oscillation) 운동이 발생하는 것을 확인할 수 있었다. 이는 실선과 시뮬레이션의 모델링오차의 영향 또는 실험환경에서 발생하는 외란에 의한 영향일 수도 있을것으로 판단된다. 실험과 시뮬레이션 결과, 충돌회피가 종료되고, 원경로로 복귀하는 시점상에 약 4초정도의 시간차가 존재하는 것을 확인할 수 있었는데, 이는 자유항주실험의 경우 반복적인 선수동요 운동으로 인하여 선속저하가 크게 발생하였고, 이때문에 타선과 조우하는 시점이 늦어지고, 회피침로각을 추종하는 속도가 늦어진 영향인 것으로 판단된다. 하지만 이러한 차이에도 불구하고, 두 경우 모두 사전에 입력된 최소유지거리 25m를 유지하며 타선과의 충돌을안정적으로 회피하는 것을 확인할 수 있었다. Fig.5.46은 Head on 상황에서 자선과 타선의 궤적을 시점별로 표현한 그림을 나타내고, Fig.5.47은 드론으로 촬영한 충돌회피 실험의 스냅샷 사진이다.

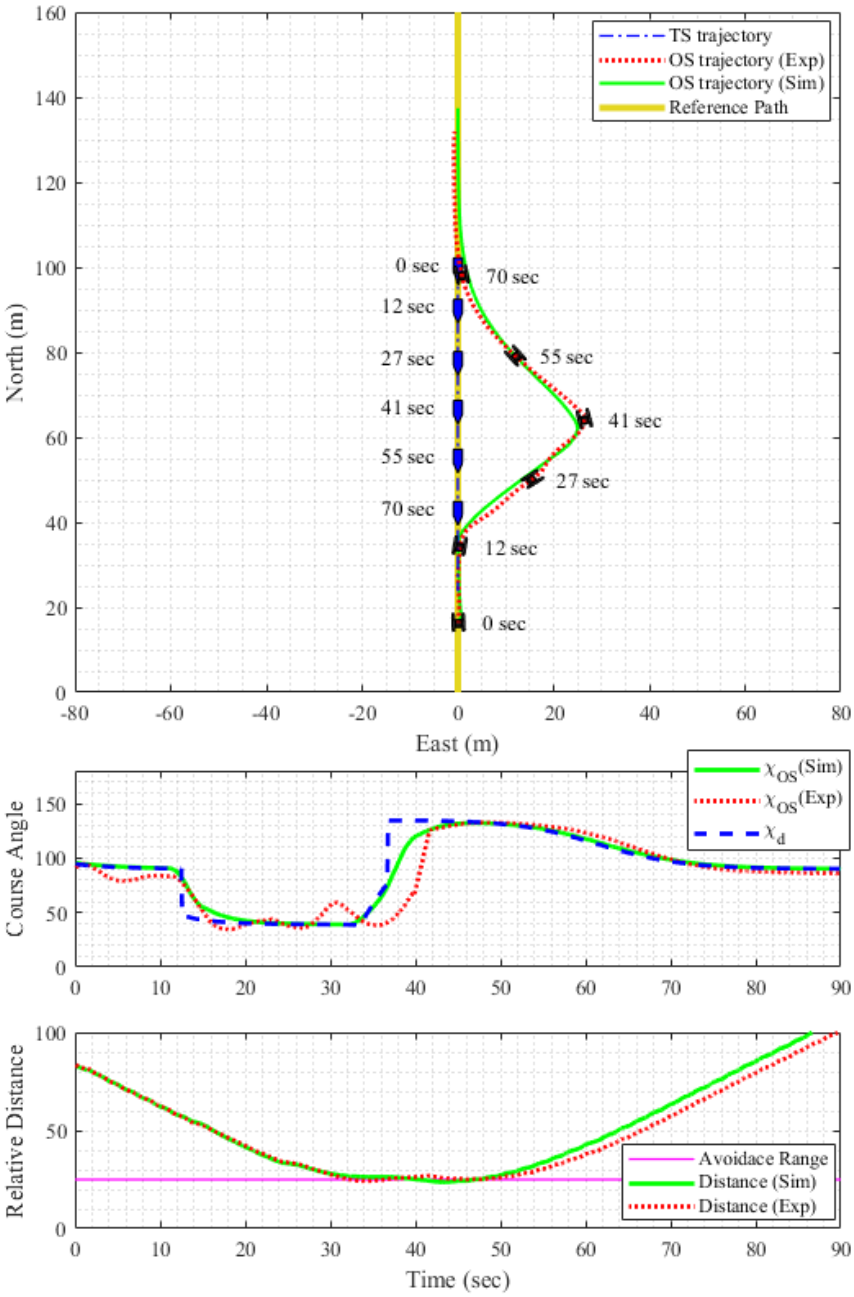



Fig 5.45: Collision avoidance free running test and simulation data (Head on)

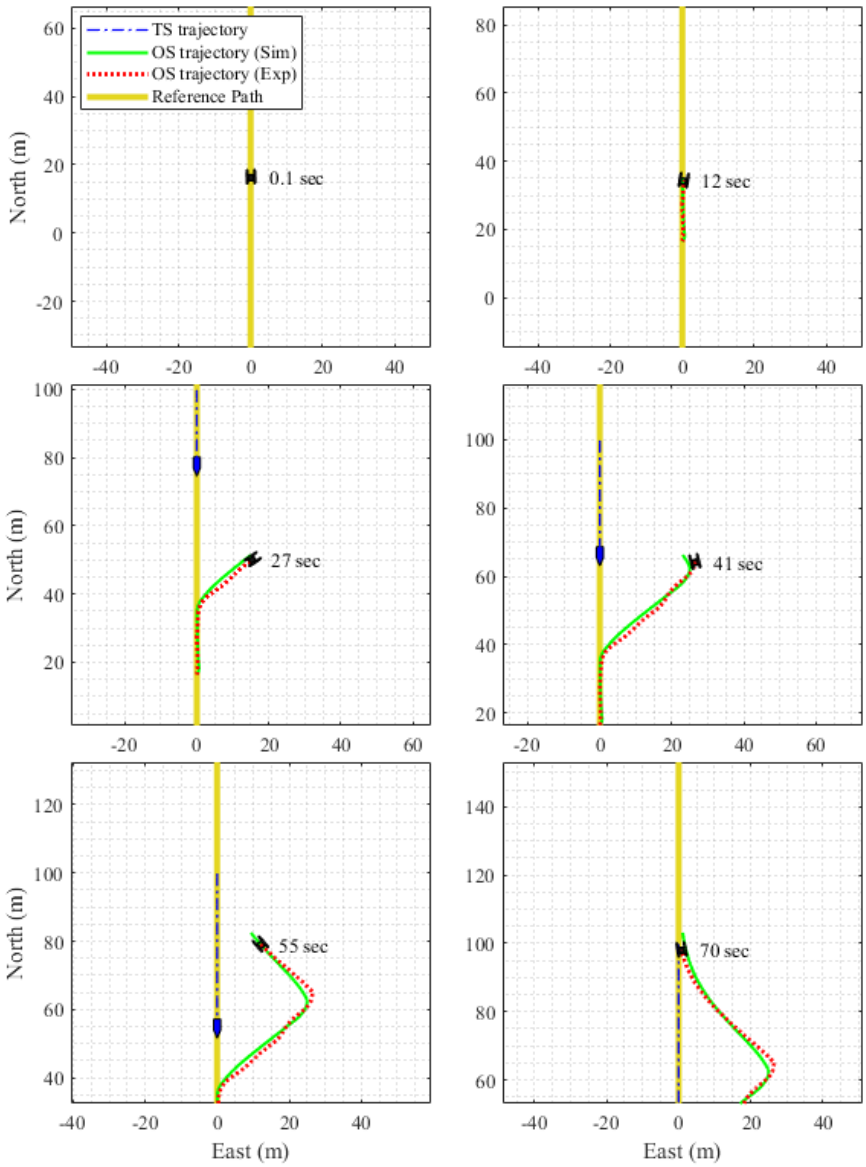



Fig 5.46: Snap shot of collision avoidance test and simulation data (Head on)

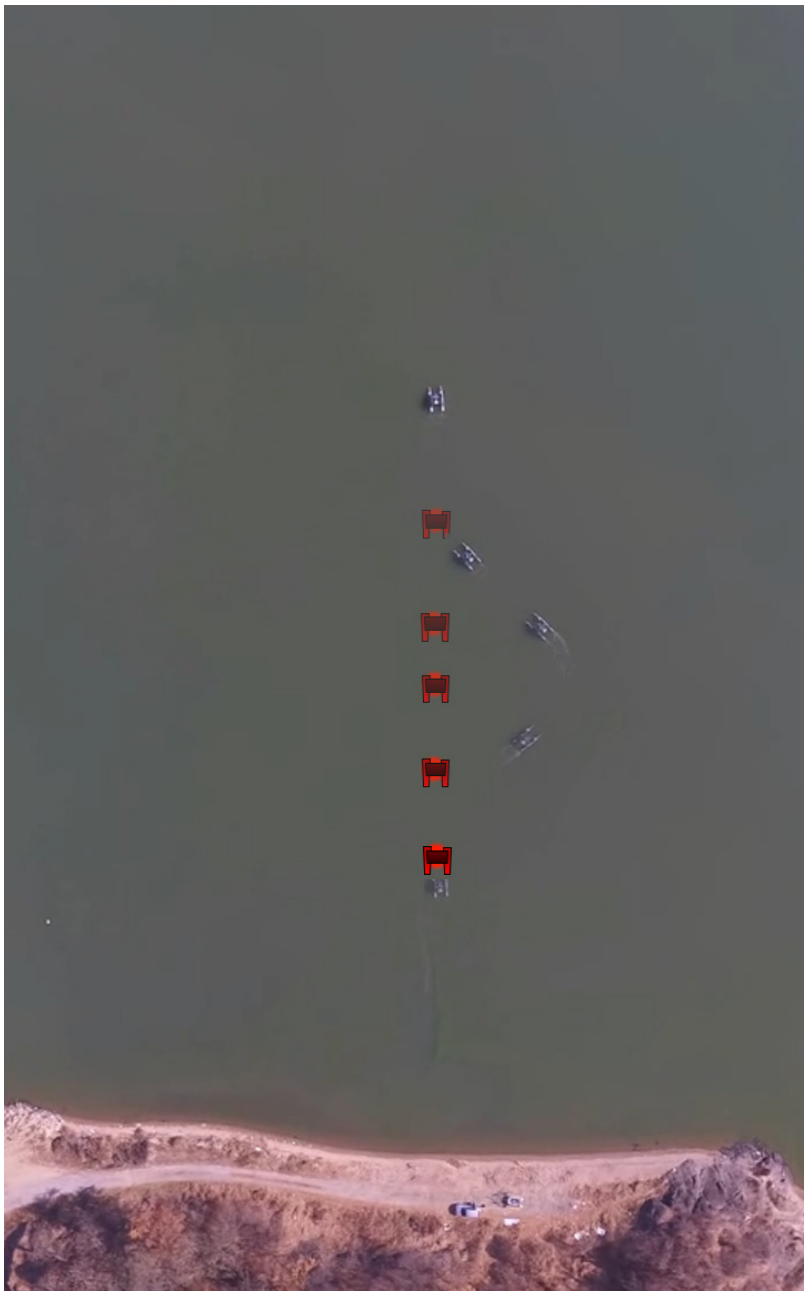


Fig 5.47: Snapshot photo of collision avoidance free running test (Head on)

**Overtaking 상황**

Overtaking 조우상황에 대한 충돌회피 실험 결과는 Fig.5.48에 나타내고 있다. Overtaking 조우상황의 경우, 타선은 자선과 동일한 90°를 향해 항주하고 있고 속도는 0.5m/s로 설정하였다. Overtaking 조우상황에 대한 충돌회피의 경우, 낮은 상대속도 때문에 기존 조우상황에 비해 긴 시간동안 넓은 영역을 거치며 회피기동이 수행되는 것을 확인할 수 있었다. 실험결과 Fig.5.48의 침로각 시계열 자료에서도 Head on 상황에서 발견한 침로각명령에 대한 오버슈트를 확인할 수 있었는데, 그럼에도 불구하고 안전거리 25m를 안정적으로 유지하며 충돌회피행위를 수행하는 것으로 확인할 수 있다. Overtaking 조우상황의 경우, 자선과 타선의 상대속도가 상대적으로 작고, 이때문에 다른 실험조건에 비하여 시뮬레이션과 실험간의 궤적오차가 작은것을 확인할 수 있다. Fig.5.48의 125sec 시점에서의 궤적을 보면, 최종 종료시점에서 실험과 시뮬레이션의 위치오차가 약 6m정도 발생하는 것을 확인할 수 있다. 이는 Head on 상황과 유사한 원인에 의하여 기인한것으로 판단되는데, 자유항주실험시 발생하는 제어기의 오버슈트에 의하여 선속저하가 크게 발생하고, 이때의 속도오차가 누적되어 이와같은 위치오차가 발생한것으로 판단된다. Fig.5.49는 Overtaking 상황에서 자선과 타선의 궤적을 시점별로 표현한 그림이고, Fig.5.50은 드론으로 촬영한 충돌회피 실험의 스냅샷 사진이다.

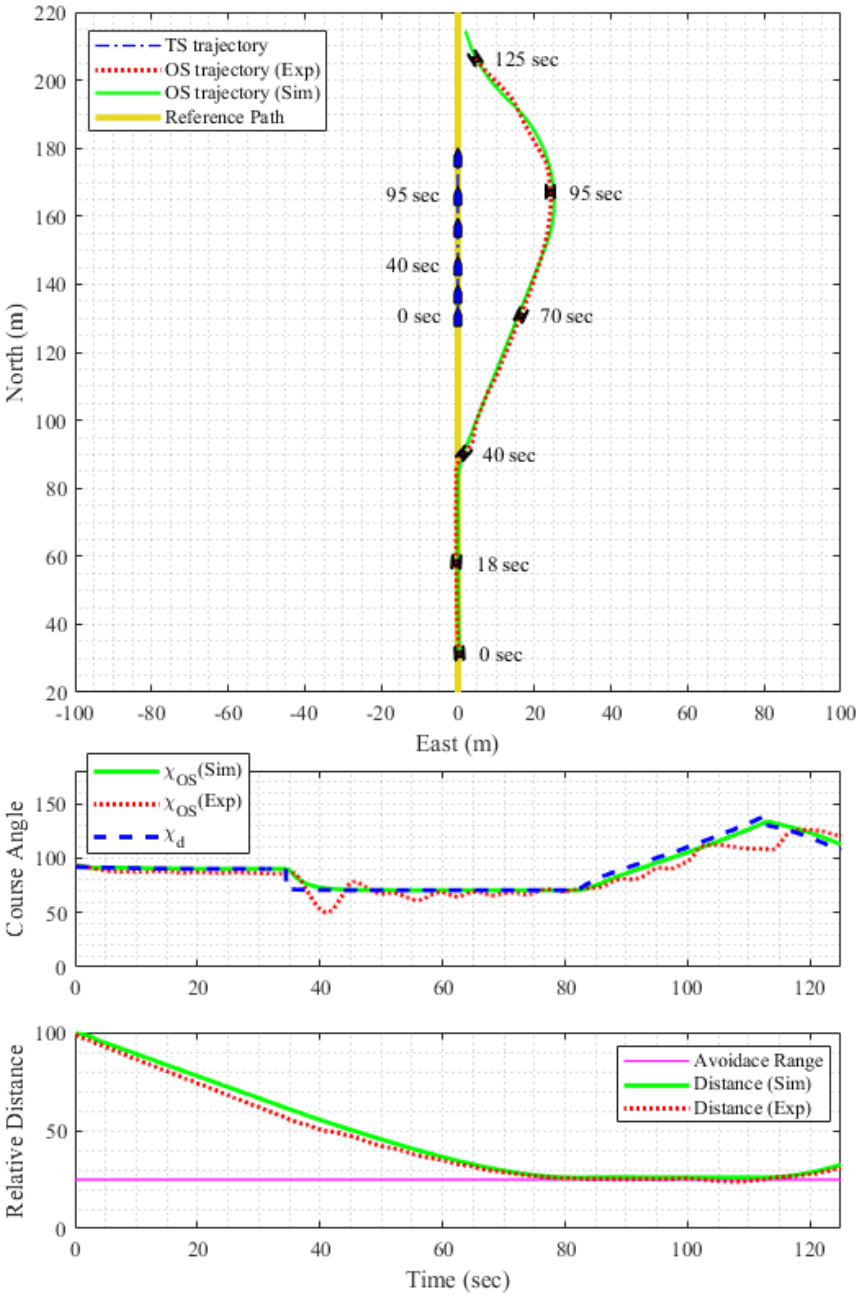



Fig 5.48: Trajectory of collision avoidance free running test (Overtaking)

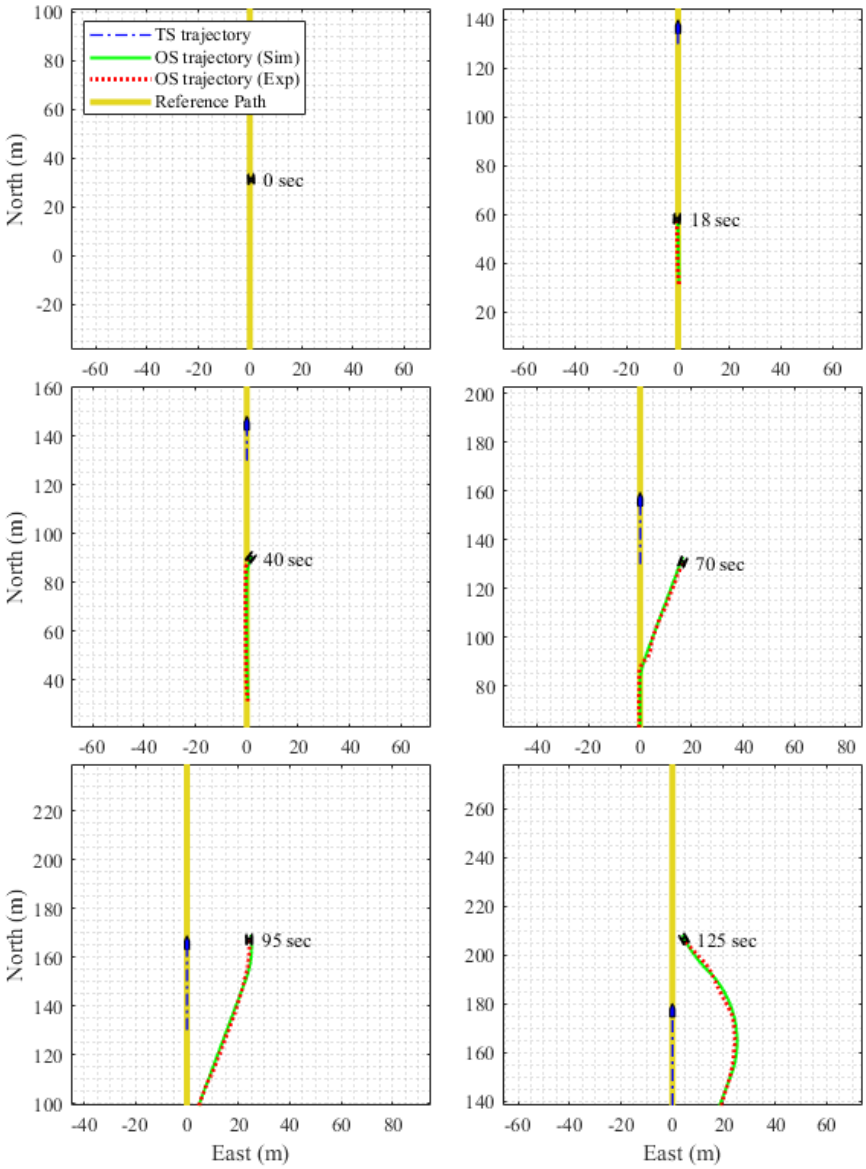



Fig 5.49: Snapshots of collision avoidance free running test (Overtaking)

Fig 5.50: Snapshot photo of collision avoidance free running test (Overtaking)

**다수 선박 조우상황**

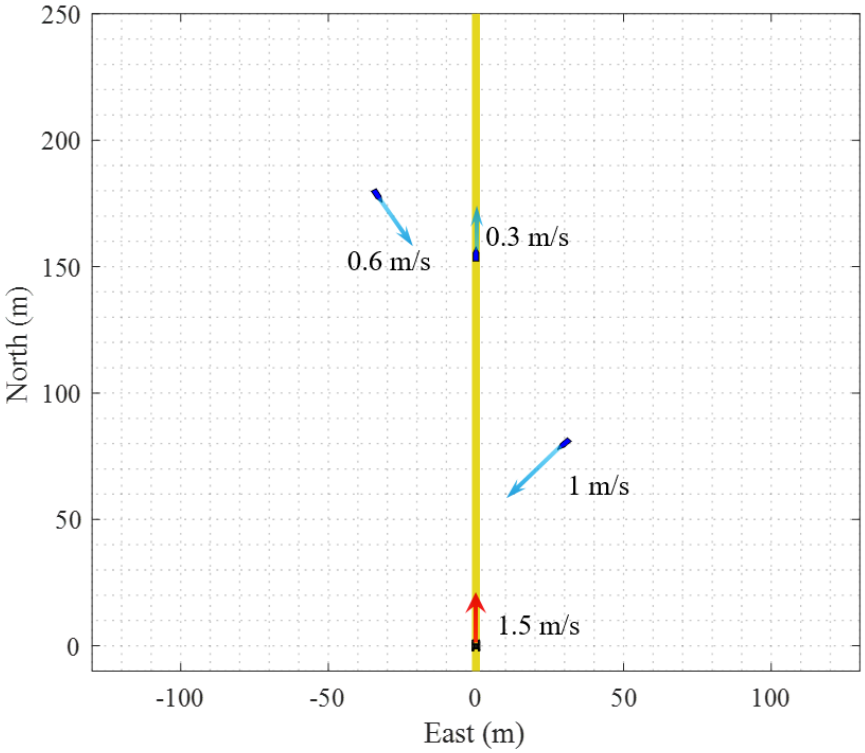



Fig 5.51: Collision scenario of multiple ship avoidance

Fig.5.51은 다수 선박 조우상황의 초기조건을 나타내고있다. 본 조우상황에서는 총 3대의 타선에 대한 충돌회피를 수행한다. 초기에 자선의 우현측에 존재하는 선박을 우현변침을 통해 회피하고 추종경로로 복귀하는 중 좌현측에서 또다른 Crossing 조우상황을 회피하게 된다. 이후 추종경로로 복귀하고, 다시 Overtaking 조우상황에 대하여 회피를 수행하는 시나리오이다. Fig.5.52의 자유항주실험 및 시뮬레이션 결과를 보면, 초기 조우상황에서는 약 10초 경 첫번째 선박에 대한 충돌회피가 시작된다. 시뮬레이션과 자유항주실험에서 동일한 시점에 회피를 시

작하는 것을 확인할 수 있다. 이후 약 31초 경 추종경로로 복귀기동이 시작한다. 약 70초 경에 경로에 도착하게되는데, 이때 자유항주실험과 시뮬레이션상의 거리 오차가 약 5m정도 나는것을 Fig.5.53을 통해 확인할 수 있다. 이는, 복귀경로에서 실험과 시뮬레이션의 궤적 차이 때문에 발생한 오차인 것으로 판단된다. 이후, 약 80초 경 까지 경로를 추종하다 두번째 조우상황을 만나게 되는데, 이 때 시뮬레이션의 결과가 실험의 결과에 비해 앞서있는 상태이므로 이에대한 영향으로 보다 이른시점에 회피가 이루어지는것을 확인할 수 있다. 실험과 시뮬레이션 결과에서 확인할 수 있는 회피경로 형상의 차이는 이와같은 충돌회피 시점 차이, 선박의 운동 모델링오차, 외란 의 영향 등에 의해 발생할 수 있다. 두 번째 회피가 이루어지고 원 경로로 복귀하면 (약 140초) 전방에 타선이 저속으로 운항하고 있음을 확인할 수 있다. 이를 추월하기위하여 우현변침을 수행하는데, 이때도 시뮬레이션의 경우가 실험데이터에 비하여 이른 시점에 충돌회피가 이루어지는것을 확인할 수 있다. 충돌회피의 시점 차이는 충돌회피 경로의 형상으로부터도 추론 할 수 있다. Fig.5.53 - Fig.5.55는 본 시나리오를 진행함에 따라 조우하게되는 세가지 조우상황에 대하여 자선과 타선의 궤적을 시점별로 표현한 그림을 나타내고있다.

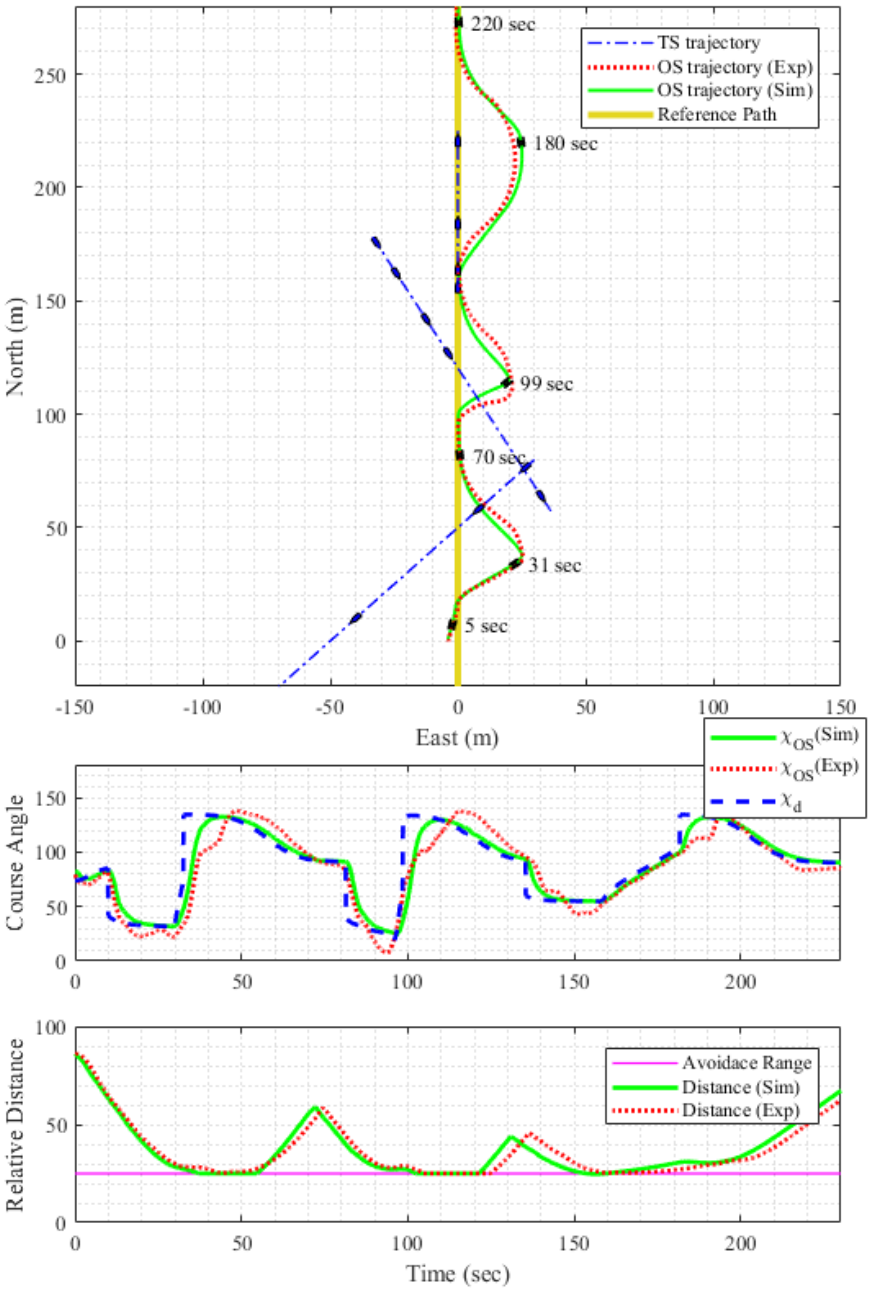



Fig 5.52: Trajectory of multiple ship collision avoidance free running test

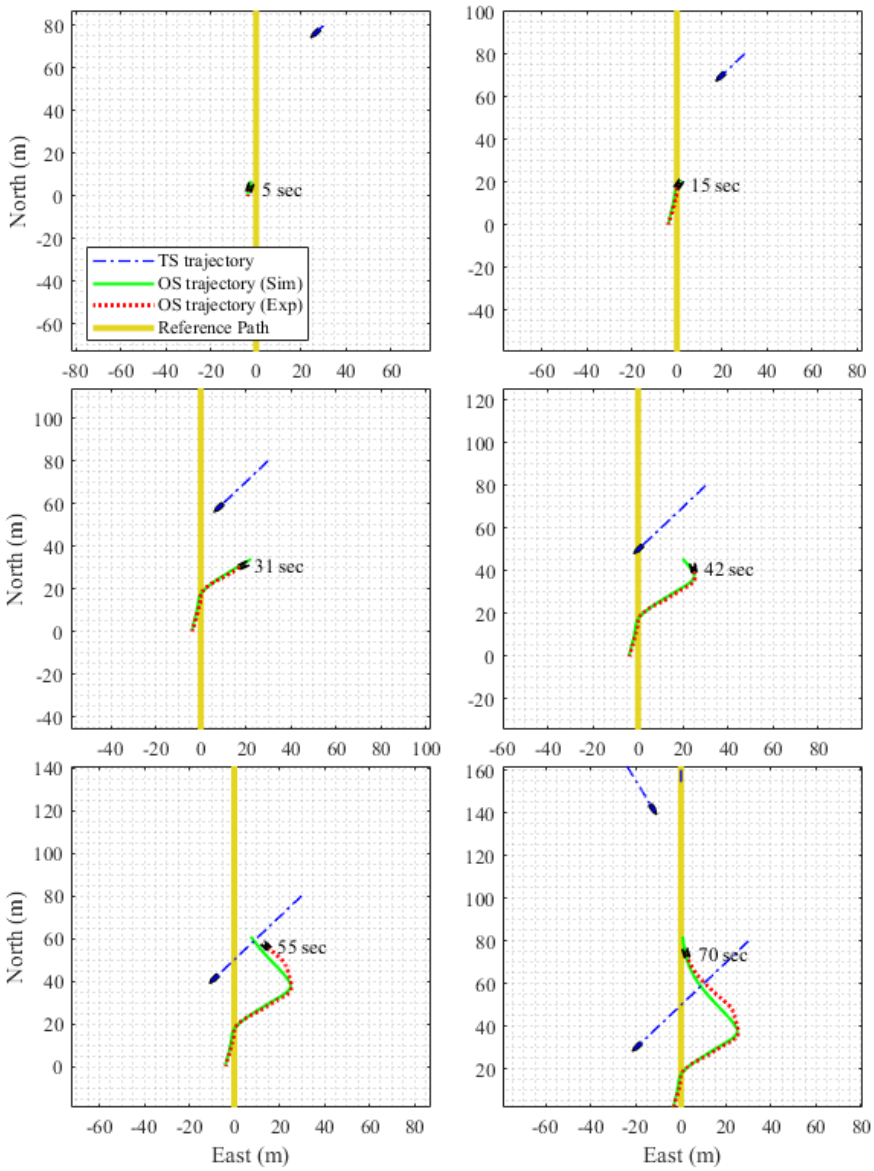



Fig 5.53: Snapshot of multiple ship collision avoidance free running test (1st target ship)

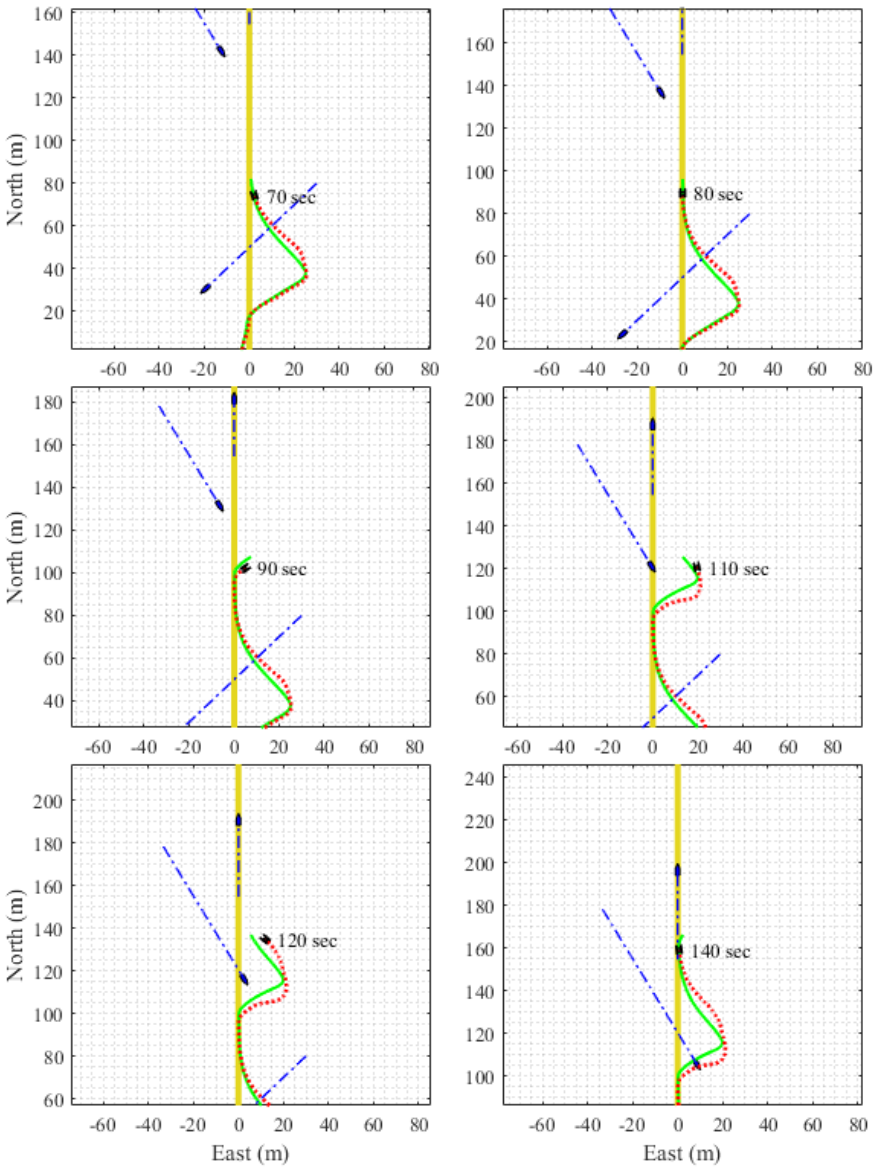



Fig 5.54: Snapshot of multiple ship collision avoidance free running test (2nd target ship)

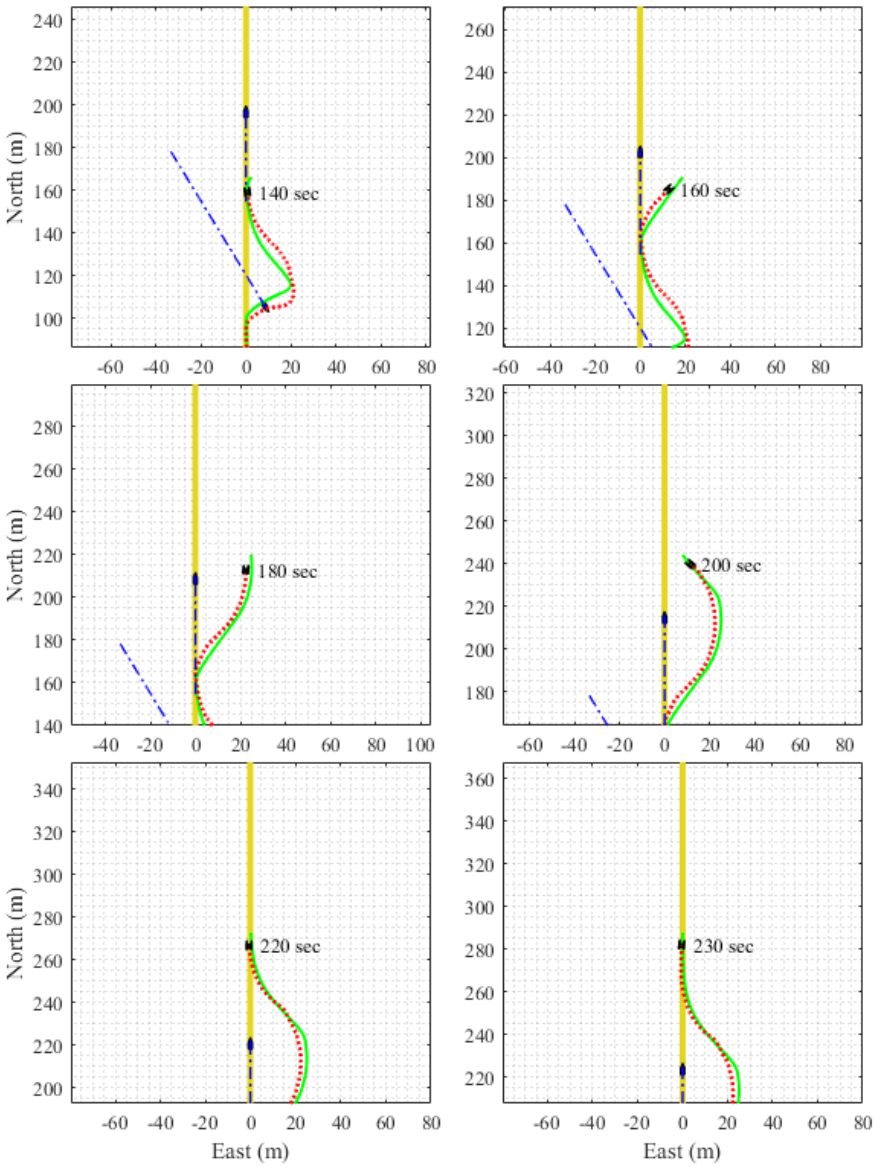



Fig 5.55: Snapshot of multiple ship collision avoidance free running test (3rd target ship)

# 제 6 장 결론

본 논문에서는 무인수상선의 자율운항을 위한 충돌회피 알고리즘을 제안하고, 시뮬레이션과 무인수상선 실선 실험을 통하여 해당 알고리즘의 성능을 검증하였다. 무인수상선 실선에 대한 강화학습을 시뮬레이션 상에서 수행하기위해서는 높은 정확도를 갖는 무인수상선의 동역학 모델이 필요하다. 이를 위해 무인수상선의 실선 조종 실험을 수행하고, 실험데이터를 기반으로 시스템 식별을 수행하였다. 구성된 심층강화학습 기법은 조우상황이 표현된 격자지도(grid map) 이미지 형태로 자선과 장애물간의 조우상황을 인식한다. 이는 자선이 처해있는 조우상황을 시각적으로 표현하여, 합성곱신경망(convolutional neural network)이 이에 대한 직관(intuition)을 학습하게 하기 위함이다. 심층강화학습의 신경망은 최종적으로는 무인수상선의 제어에 사용하게될 유도기법을 선택하도록 구성되었는데, 이는 각각 경로추종, 좌현변침을 통한 회피, 우현변침을 통한 회피를 수행하기위한 유도기법으로 구성되었다. 선박의 충돌회피는 국제 해상충돌 예방규칙(COLREGs)을 고려하여 수행해야하고, 조우상황에 따라 개별적인 요구사항이 존재한다. 이러한 요소들을 충돌회피에 반영하기위해 심층강화학습의 보상함수를 조우상황에 따라 다르게 구성하였다. 구성된 심층강화학습 기법을 이용하여 시뮬레이션 상에서 다양한 조우상황에대한 강화학습을 진행하였고, 이를 통해 선박충돌회피에 대한 경험적 요소를 충돌회피 알고리즘에 반영시킬 수 있었다. 본 연구를 바탕으로 다음과 같은 네 가지 결론을 도출할 수 있었다.

첫째, 심층강화학습을 이용하여 학습된 신경망을 분석해본 결과, 기존에 연구된 기법들과 유사한 방식으로 장애물에대한 충돌위험도를 평가할 수 있었다.

선박의 충돌회피의 특징인 좌우 영역에 대한 위험도의 비대칭성([9], [17], [12])을 스스로 학습을 통해 터득할 수 있음을 확인하였고, 격자지도 이미지에 대한 현저 지도(saliency map)분석을 통해 충돌위험도의 추론이 장애물 정보에 기인하는 것을 확인하였다.

둘째, 다수 선박이 존재하는 복잡한 상황에 대한 충돌 회피를 수행할 수 있었다. 특히, 제안하는 충돌회피 기법을 통해 현재 시점의 충돌위험도 뿐만 아니라, 특정 행위를 통해 야기되는 미래 시점의 충돌위험도까지 고려하여 충돌회피 시점 및 방향을 선택할 수 있었다.

셋째, 강화학습을 통해 충돌회피에 필요한 경험적 요소를 충돌회피 시점추론에 반영할 수 있었다. 이는, 선박의 조종성능이나 환경하중이 미치는 영향을 고려하여 충돌회피 시점을 결정할 수 있는 것을 의미한다. 시뮬레이션 결과, 조종성능이 상대적으로 낮은 선박은 그렇지 않은 선박에 비해 보다 이른시점에 충돌회피기동을 수행하는 현상을 확인할 수 있었다. 환경하중의 경우, 환경하중이 회피기동에 도움을 주는경우 회피가 더 이른 시점에 이루어지는 현상을 발견하였다.

넷째, 조우상황별로 적합한 회피전략을 통해 충돌을 회피할 수 있었다. 국제해상충돌 예방규칙에 따르면, 각 조우상황별로 상이한 요구사항이 존재하는데, 본 알고리즘은 조우상황별로 강화학습의 보상함수를 독립적으로 정의하여 각 조우상황에 맞는 회피전략을 통해 충돌회피를 수행할 수 있었다.

본 연구에서 설계된 심층강화학습 기반 충돌회피 알고리즘은 고차원 격자 이미지에 대한 강화학습을 반복적으로 수행해야하므로, 학습시간이 매우 오래걸린다는 단점이 존재한다. 이를 단축 시키기위한 학습의 효율성을 증대시키거나, 과적합(over fitting)을 억제시켜 높은 학습계수(learning rate)로 강화학습을 수행하는 기법에 대한 추가 연구가 필요할 것이다. 또한, 본 연구에서 처럼 시뮬레이션을 통

해 강화학습을 진행하는 것이 아니라, 실제 무인수상선을 운용하면서 강화학습을 진행한다면, 충돌회피 의사결정과정에서 무인수상선의 모델링오차 등의 요소가 미치는 영향을 최소화 할 수 있을 것으로 기대된다.

# 부록 A 무인수상정의 시스템 식별

## A.1 무인수상선 동역학의 시스템 식별

### 구동기 동역학 모델 식별

Fig.A.1은 추진기 모터에 입력되는 명령값에 대한 프로펠러 RPM의 정상상태 응답 (steady state response)을 나타내는 그래프이다. 측정 결과, 실선 추진기의 RPM값은 최대 2380의 RPM으로 회전할 수 있으며 모터입력값에 대하여 선형적으로 RPM이 변화하는 것을 확인할 수 있었다.

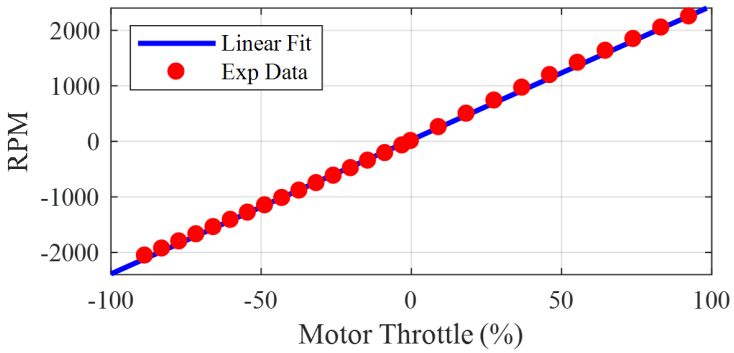



Fig A.1: Relation between electric motor throttle and steady state RPM response

모터 명령과 RPM간의 정상상태 응답이 도출되면, 모터명령값과 실제 모터RPM간의 과도응답(transient response)에 대한 정보가 필요한데, 이를 위하여 RPM의 변화량에 대한 포화(saturation)를 고려하는 단순한 형태의 타동역학 모형을 도입하였다. Fig.A.2와 같은 구동기 동역학을 고려함으로써, RPM명령 $\delta n_{cmd}$와

이에대한 응답인 RPM $\delta n$간의 시간 지연(time delay) 현상을 모사할 수 있다. Fig.A.3은 $\delta n_{cmd}$에 대한 $\delta n$의 응답시험 결과를 나타내고있다. 실험결과, $\dot{\delta n}_{max}$ 값은 1510 rpm/s 으로 계산되었다.

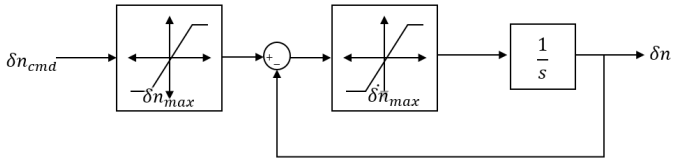



Fig A.2: Simplified block diagram of the RPM model

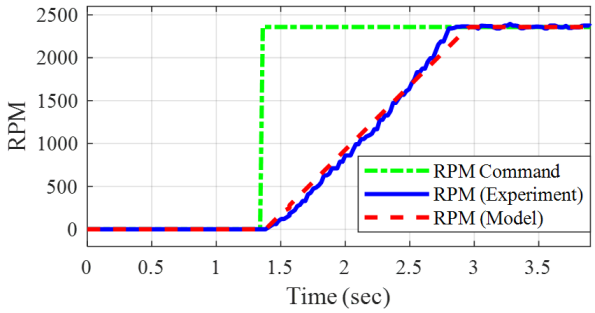



Fig A.3: Step response test result of an electric thruster motor of the ASV

RPM 모델에 대한 식별이 완료되면 RPM과 추진기가 발생시키는 추력사이의 관계를 식별해야 한다. 이를 위하여 본 연구에서는 실선에 대한 Bollard-pull 실험을 진행하였다. Bollard-pull 실험은 로프을 이용하여 무인수상선을 부두에 구속시키고, 구속된 상태로 추진기를 가동시켜 로프에 걸리는 장력을 계측하는 실험이다. 이때 로프에 걸리는 장력은 추진기에 의해 발생한 힘이므로 Bollard-pull 실험을 통해 RPM에 대한 추력곡선을 얻을 수 있다. FigureA.4은 실선 무인수상선에 대한 Bollard-pull실험 결과를 나타내고있다. 실험 결과, 단일 추진기의 RPM

$\delta n$에 대한 추력은 식 A.1 - A.2와 같은 2차함수로 표현된다.

**정방향 RPM**

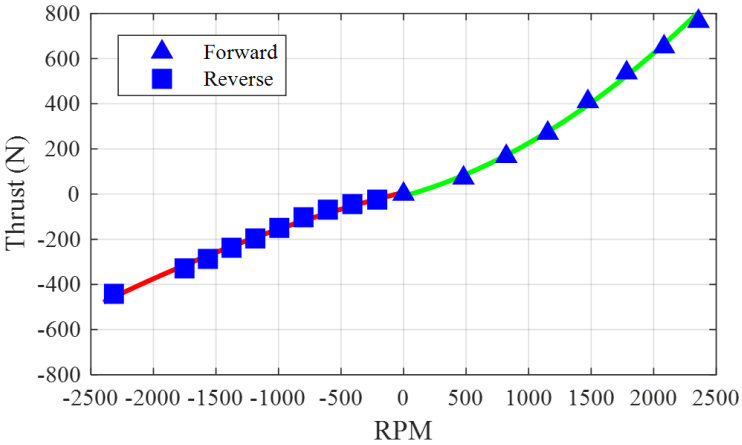
$$T = 3.54 \times 10^{-5}\delta n^2 + 0.084\delta n - 3.798 \tag{A.1}$$

**역방향 RPM**

$$T = -1.189 \times 10^{-5}\delta n^2 + 0.071\delta n + 4.331 \tag{A.2}$$

이론적으로 프로펠러에 의해 발생하는 추력은 RPM과 프로펠러 전진 속도(advanced speed)의 함수이다. 하지만 Bollard-pull 실험의 경우 선체를 부두에 고정해야만 로프의 장력을 측정할수 있기때문에 추력에 대한 전진 속도의 영향을 계측할 수 없다는 실험적 한계가 존재한다.

(a) SNU ASV during the bollard-pull test




(b) Relationship between motor RPM and thrust force

Fig A.4: Bollard-pull test of the ASV

### 시스템 식별 기법

본 연구에서는 Sonnenburg et al [36] 가 제안한 시스템식별 기법을 사용하여 식 5.9의 파라미터들을 추정하였다. 본 기법은 일정한 입력 변수를 갖는 정적 입력 실험 (static input test)을 통해 상태변수와 입력변수 간의 선형 구속조건을 구성하고, 이 구속조건과 동적 입력 실험(dynamic input test) 데이터를 이용하여 파라미터를 식별하는 기법이다. 식 5.9 과같은 선형 시불변 시스템은 식 A.3과

같이 표현할 수 있다.

$$\dot{x} = Ax + Bu \tag{A.3}$$

이때, $x(t) \in \mathbb{R}^n$ 이고, $u(t) \in \mathbb{R}^m$이다. 시스템 행렬 $A$가 Hurwitz라고 가정하고, 시스템 입력변수 $u_{ss}$가 시스템에 입력되면, 시스템의 상태변수는 정상상태 $x_{ss}$에 도달하게 된다. 이때, 충분히 많은 입력-정상상태 응답을 계측하면, 식 A.4와 같은 정상상태 구속조건식을 구성할 수 있다.

$$x_{ss} = \Phi u_{ss} \tag{A.4}$$

이때, $\Phi$는 정적 입력 실험을 통해 얻은 구속조건 행렬$(n \times m)$로, $\Phi$ 행렬의 성분은 $u_{ss}$ 와 $x_{ss}$의 선형보간을 통해 얻는다. 정상상태 하에서 식 A.4를 식 A.3에의해 구속조건 $\Phi$는 A.5와 같이 정리된다.

$$\Phi = -A^{-1}B \tag{A.5}$$

정적 입력 실험을 통해 구속조건식이 도출되면, 동적 입력 실험의 시계열 데이터를 통해 시스템 방정식의 파라미터를 추정한다. 동적 입력 실험 데이터는 상태변수 시계열 데이터 $x(t)$ 와 입력변수의 시계열 데이터 $u(t)$를 포함하고 있다. 계측된 상태변수 시계열 데이터에 평탄화(smoothening)를 위한 신호 필터를 적용하고, 미분을 수행하면 $\dot{x}(t)$데이터를 계산할 수 있다. 식 A.3에 식 A.5을 대입하면, 식 A.6와 같이 정리된다.

$$\dot{x}(t) = A(x(t) - \Phi u(t)) \tag{A.6}$$

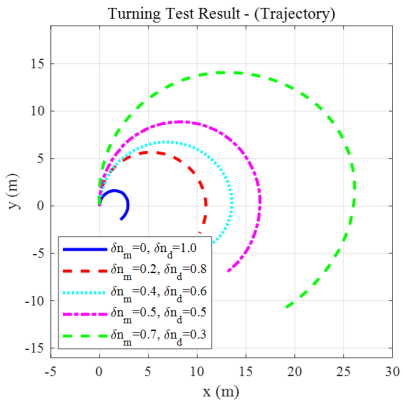
식 A.6 에 포함된 변수 중 시스템 행렬 $A$를 제외한 모든 변수는 계측되거나 계산된 변수이므로, 최소자승법을 통해 시스템 행렬 $A$ 파라미터를 추정할 수 있다. 시스템 행렬 $A$가 추정되면, 식 A.5에 의하여 시스템의 입력행렬 $B$의 파라미터를 계산할 수 있다.

**무인수상선 동역학 식별**

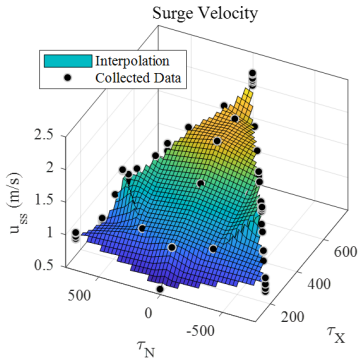
정상 상태 응답 구속조건을 계산하기위하여 정적 입력 실험을 수행하였다. 입력 변수의 단순화를 위하여 제어 명령 변수 $\delta_{n_m}$과 $\delta_{n_d}$를 정의하였다. 두 변수 모두 정규화(normalization)가 된 변수로 -1에서 1사이의 값을 갖는다. $\delta n_m$은 좌현과 우현 추진기의 RPM 합을 정규화된 값을 의미하고, $\delta n_d$는 좌현과 우현 추진기의 RPM 차이값의 정규화된 값을 의미한다.

$$\begin{aligned} \delta_{n_m} &= (\delta_{n_{port}} + \delta_{n_{stbd}})/2\delta_{n_{max}} \\ \delta_{n_d} &= (\delta_{n_{port}} - \delta_{n_{stbd}})/2\delta_{n_{max}} \end{aligned} \tag{A.7}$$

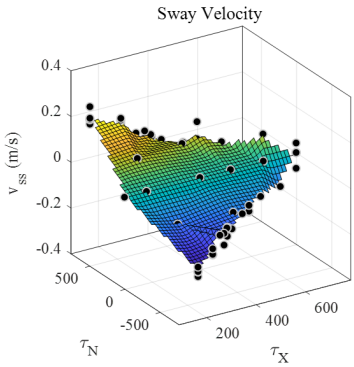
Fig.A.5는 실선 무인수상선을 이용하여 수행한 정적 입력 시험결과를 나타내고있다. Fig.A.5a는 다양한 입력조건하에서 실험을 수행하였을 때 무인수상선이 지나간 궤적을 나타내고있고, Fig.A.5b - A.5d는 60개의 정적 입력 조건 하에서 입력변수에 대한 각 자유도의 상태변수의 정상상태 응답 을 나타내고있다.



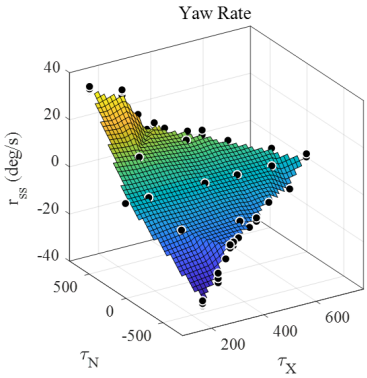

(a) Static input test trajectory result




(b) Surge velocity steady state response




(c) Sway velocity steady state response




(d) Yaw rate steady state response

(e) Video snapshot during the turning test

Fig A.5: Static input test result of the SNU ASV

정적 입력 실험을 통해 얻은 각 입력에 대한 정상 상태 응답 계측값들을 바탕으로, 식 A.4의 구속조건식을 구성하면 다음과 같이 표현된다.

$$u_{ss} = \Phi_u \tau_{X_{ss}}, \qquad v_{ss} = \Phi_v \tau_{N_{ss}}, \qquad r_{ss} = \Phi_r \tau_{N_{ss}} \tag{A.8}$$

Fig.A.5b - A.5d의 실험계측값의 입력변수와 상태변수간의 선형보간을 통해 구속조건식 A.8의 파라미터는 다음과 같이 추정할 수 있다.

$$\Phi_u = 0.00210, \qquad \Phi_v = 0.00022, \qquad \Phi_r = 0.34101 \tag{A.9}$$

구속조건의 파라미터가 추정되면, 구속조건과 동적 입력 실험데이터들을 이용하여 시스템 행렬을 구할 수 있다. 동적 입력 실험으로는 6가지 조건에 대한 가감속 실험 (FigureA.6)과 50가지 조건에 대한 지그재그 실험(Fig.A.7)을 수행하였다.

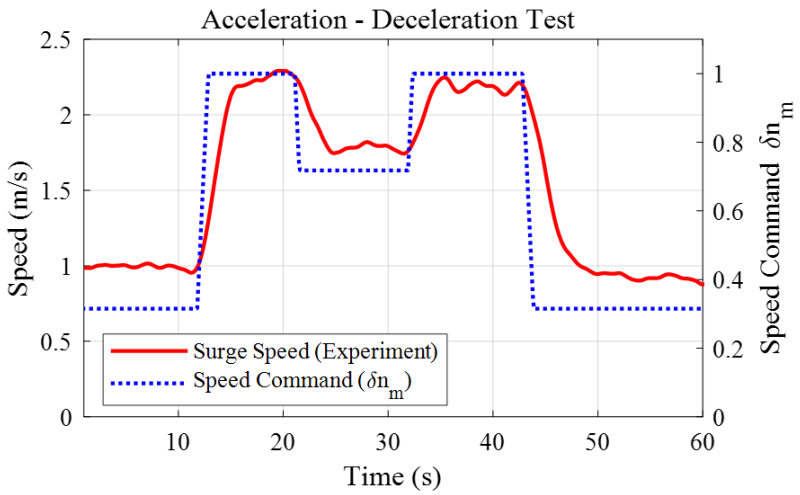
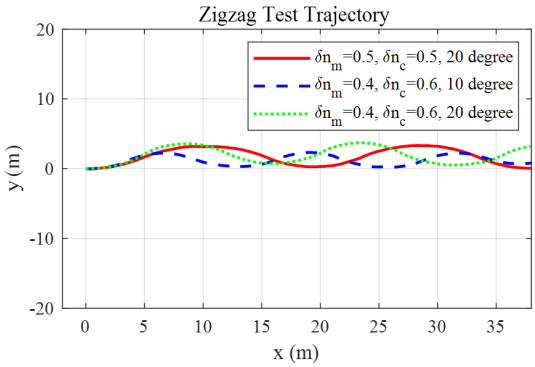
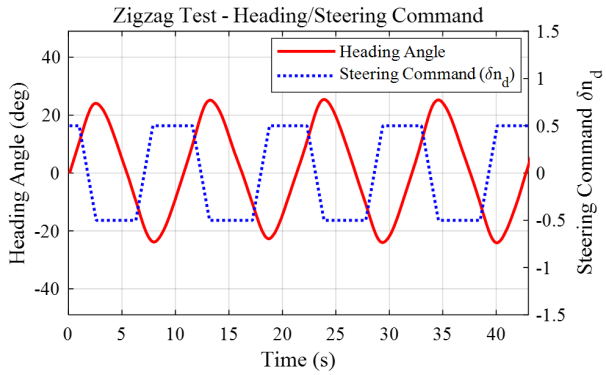



Fig A.6: Acceleration-decceleration test of the SNU ASV




(a) Zigzag test trajectory result with various test condition




(b) Time history of heading angle and steering command during the zigzag test

(c) Video snapshot during the zigzag test

Fig A.7: Zigzag test of the SNU ASV

동적실험의 시계열 데이터들은 무인수상선에 탑재된 GPS센서와 AHRS센서로부터 20Hz의 주기로 계측된다. 상태변수값의 미분값을 구할 때, 센서의 계측값을 바로 사용하게되면 데이터에 포함되어있는 계측잡음이 증폭되는 현상이 발생한다. 이를 방지하기위하여 본 연구에서는 Savitsky-Golay 필터[33]를 시계열 신호에 적용하고, 필터링 이후의 신호를 미분하여 가속도 성분의 시계열 자료를 획득한다. 식 A.6 에 의해, 동적 입력 실험데이터들과 식 A.9의 구속조건 파라미터들을 사용하면 시스템 행렬$A$의 파라미터를 추정할 수 있고, 이를 이용하여 입력행렬 $B$의 파라미터 또한 추정이 가능하다. 시스템 식별 결과, 실선 실험을 통한 무인수상선의 선형 동역학 모델은 식 A.10과 같이 표현할 수 있다.

$$\begin{bmatrix} \dot{u} \\ \dot{v} \\ \dot{r} \end{bmatrix} = \begin{bmatrix} -1.0191 & 0 & 0 \\ 0 & 0.0161 & -0.0052 \\ 0 & 8.2861 & -0.9860 \end{bmatrix} \begin{bmatrix} u \\ v \\ r \end{bmatrix} + \begin{bmatrix} 0.0028 & 0 \\ 0 & 0.0002 \\ 0 & 0.0307 \end{bmatrix} \begin{bmatrix} \tau_X \\ \tau_N \end{bmatrix} \quad \text{(A.10)}$$

FigureA.8은 $\delta n_m = 0.3$, $\delta n_d = 0.8$ 조건의 지그재그 시험을 수행하였을 때, 식 A.10의 동역학 모델과 실제 무인수상선의 거동을 비교한 결과이다. 실제 무인수상선 실험에 사용된 시스템 입력을 시스템 식별을 통해 얻은 선형 동역학 모델에 입력하였을 때의 좌우동요와 선수동요 각속도성분을 비교하였다. 비교 결과, 실험을 통해 계측한 상태변수의 거동과 시스템식별을 통해 얻은 동역학 모델의 상태변수의 거동이 유사한것을 확인할 수 있었다. 시스템 식별에 사용되지 않은 20가지 조건의 실험 데이터에 대하여 좌우동요 속도와 선수동요 각속도의 오차값을 계산해본 결과, 좌우동요의 경우 0.053 $m/s$의 RMS값을, 선수동요의 경우 4.68 °의 RMS값을 갖는것을 확인하였다. 식별된 무인수상선 동역학은 강화학습을 수행하기위한 시뮬레이터를 구현하는데 사용된다.

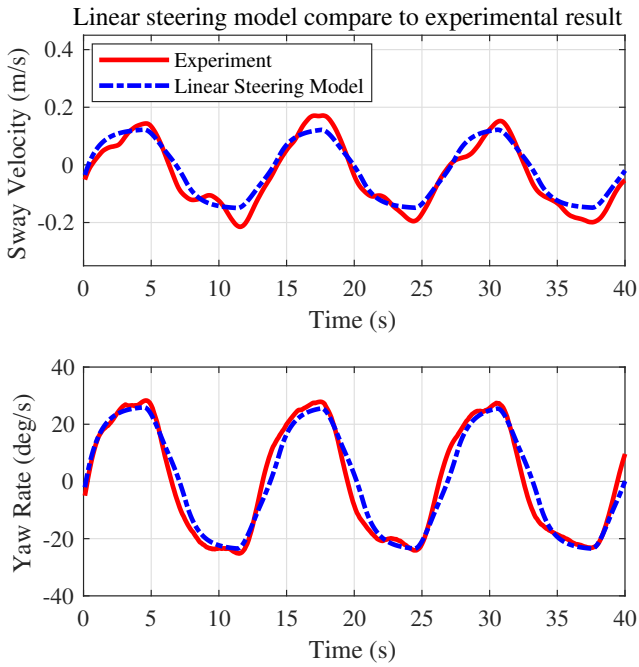



Fig A.8: Acceleration-decceleration test of the SNU ASV

# 참고문헌


[1] Jin-Hyeong Ahn, Key-Pyo Rhee, and Young-Jun You. A study on the collision avoidance of a ship using neural networks and fuzzy logic. *Applied Ocean Research*, 37:162–173, 2012.

[2] Carlos Almeida, Tiago Franco, Hugo Ferreira, Alfredo Martins, Ricardo Santos, José Miguel Almeida, João Carvalho, and Eduardo Silva. Radar based collision detection developments on usv roaz ii. In *Oceans 2009-Europe*, pages 1–6. IEEE, 2009.

[3] Michael Blaich, Steffen Köhler, Michael Schuster, Thomas Schuchhardt, Johannes Reuter, and Thomas Tietz. Mission integrated collision avoidance for usvs using laser range finder. In *OCEANS 2015-Genova*, pages 1–6. IEEE, 2015.

[4] Sable Campbell, Wasif Naeem, and George W Irwin. A review on improving the autonomy of unmanned surface vehicles through intelligent collision avoidance manoeuvres. *Annual Reviews in Control*, 36(2):267–283, 2012.

[5] Marc Carreras Pérez et al. *A proposal of a behavior-based control architecture with reinforcement learning for an autonomous underwater robot.* Universitat de Girona, 2003.

[6] Giuseppe Casalino, Alessio Turetta, and Enrico Simetti. A three-layered architecture for real time path planning and obstacle avoidance for surveillance usvs operating in harbour fields. In *Oceans 2009-Europe*, pages 1–8. IEEE, 2009.

[7] Animesh Chakravarthy and Debasish Ghose. Obstacle avoidance in a dynamic environment: A collision cone approach. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 28(5):562–574, 1998.

[8] Adam Coates, Pieter Abbeel, and Andrew Y Ng. Apprenticeship learning for helicopter control. *Communications of the ACM*, 52(7):97–105, 2009.

[9] TG Coldwell. Marine traffic behaviour in restricted waters. *The Journal of Navigation*, 36(3):430–444, 1983.

[10] BA Colley, RG Curtis, and CT Stockel. Manoeuvring times, domains and arenas. *The Journal of Navigation*, 36(2):324–328, 1983.

[11] Bruno Damas and José Santos-Victor. Avoiding moving obstacles: the forbidden velocity map. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 4393–4398. IEEE, 2009.

[12] PV Davis, MJ Dove, and CT Stockel. A computer simulation of marine traffic using domains and arenas. *The journal of Navigation*, 33(2):215–222, 1980.

[13] Yaakov Engel and Shie Mannor. Learning embedded maps of markov processes. In *in Proceedings of ICML 2001*. Citeseer, 2001.

[14] Thor I Fossen. *Guidance and control of ocean vehicles*. John Wiley & Sons Inc, 1994.

[15] Thor I Fossen. *Marine control systems: guidance, navigation and control of ships, rigs and underwater vehicles*. Marine Cybernetics, 2002.

[16] Yahei Fujii and Kenichi Tanaka. Traffic capacity. *The Journal of Navigation*, 24(4):543–552, 1971.

[17] Elisabeth M Goodwin. A statistical study of ship domains. *The Journal of Navigation*, 28(3):328–344, 1975.

[18] Kazuhiko Hasegawa and Akihiko Kouzuki. Automatic collision avoidance system for ships using fuzzy control. *Journal of the Kansai Society of Naval Architects*, (205), 1987.

[19] Terry Huntsberger, Hrand Aghazarian, Andrew Howard, and David C Trotz. Stereo vision–based navigation for autonomous surface vessels. *Journal of Field Robotics*, 28(1):3–18, 2011.

[20] Wei-Yuan Hwang. *Application of system identification to ship maneuvering*. PhD thesis, Massachusetts Institute of Technology, 1980.

[21] IMO. Convention on the international regulations for preventing collisions at sea, 1972 (colregs), year = 1972, url = http://www.imo.org/en/about/conventions/listofconventions/pages/colreg.aspx.

[22] Zhao Jingsong, Wu Zhaolin, and Wang Fengchen. Comments on ship domains. *The Journal of Navigation*, 46(3):422–436, 1993.

[23] Yoshiaki Kuwata, Michael T Wolf, Dimitri Zarzhitsky, and Terrance L Huntsberger. Safe maritime autonomous navigation with colregs, using velocity obstacles. *IEEE Journal of Oceanic Engineering*, 39(1):110–119, 2014.

[24] Jacoby Larson, Michael Bruch, and John Ebken. Autonomous navigation and obstacle avoidance for unmanned surface vehicles. Technical report, SPACE AND NAVAL WARFARE SYSTEMS CENTER SAN DIEGO CA, 2006.

[25] Jacoby Larson, Michael Bruch, Ryan Halterman, John Rogers, and Robert Webster. Advances in autonomous obstacle avoidance for unmanned surface vehicles. Technical report, SPACE AND NAVAL WARFARE SYSTEMS CENTER SAN DIEGO CA, 2007.

[26] Anderson Lebbad and C Nataraj. A bayesian algorithm for vision based navigation of autonomous surface vehicles. In *Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM), 2015 IEEE 7th International Conference on*, pages 59–64. IEEE, 2015.

[27] Sergey Levine, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, 17(39):1–40, 2016.

[28] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

[29] Helene Myre. Collision avoidance for autonomous surface vehicles using velocity obstacle and set-based guidance. Master's thesis, NTNU, 2016.

[30] Xue Bin Peng, Glen Berseth, KangKang Yin, and Michiel Van De Panne. Deeploco: Dynamic locomotion skills using hierarchical deep reinforcement learning. *ACM Transactions on Graphics (TOG)*, 36(4):41, 2017.

[31] Huajin Qu, Edoardo I Sarda, Ivan R Bertaska, and Karl D von Ellenrieder. *Wind Feed-forward Control of a USV*. IEEE, 2015.

[32] Ahmad EL Sallab, Mohammed Abdou, Etienne Perot, and Senthil Yogamani. Deep reinforcement learning framework for autonomous driving. *Electronic Imaging*, 2017(19):70–76, 2017.

[33] Abraham Savitzky and Marcel JE Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639, 1964.

[34] Michael Schuster, Michael Blaich, and Johannes Reuter. Collision avoidance for vessels using a low-cost radar sensor. *IFAC Proceedings Volumes*, 47(3):9673–9678, 2014.

[35] Jongho Shin, Dong Jun Kwak, and Young-il Lee. Adaptive path following control for an unmanned surface vessel using an identified dynamic model. *IEEE/ASME Transactions on Mechatronics*, 2017.

[36] Christian Sonnenburg, Aditya Gadre, Doug Horner, Sean Kragelund, A Marcus, Daniel J Stilwell, and CA Woolsey. Control-oriented planar motion modeling of unmanned surface vehicles. In *OCEANS 2010*, pages 1–10. IEEE, 2010.

[37] Thomas Stenersen. Guidance system for autonomous surface vehicles. Master's thesis, NTNU, 2015.

[38] Richard S Sutton, Doina Precup, and Satinder Singh. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112(1-2):181–211, 1999.

[39] Rafal Szlapczynski. A unified measure of collision risk derived from the concept of a ship domain. *The Journal of Navigation*, 59(3):477–490, 2006.

[40] CheeKuang Tam and Richard Bucknall. Collision risk assessment for ships. *Journal of marine science and technology*, 15(3):257–270, 2010.

[41] Lou Tychonievich, David Zaret, John Mantegna, Robert Evans, Eric Muehle, and Scott Martin. A maneuvering-board approach to path planning with moving obstacles. In *Proceedings of the 11th international joint conference on Artificial intelligence-Volume 2*, pages 1017–1021. Morgan Kaufmann Publishers Inc., 1989.

[42] Hado Van Hasselt, Arthur Guez, and David Silver. Deep reinforcement learning with double q-learning. In *AAAI*, pages 2094–2100, 2016.

[43] Ziyu Wang, Tom Schaul, Matteo Hessel, Hado Van Hasselt, Marc Lanctot, and Nando De Freitas. Dueling network architectures for deep reinforcement learning. *arXiv preprint arXiv:1511.06581*, 2015.

[44] Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.

[45] Scott Weaver, Leemon Baird, and Marios Polycarpou. An analytical framework for local feedforward networks. *IEEE Transactions on Neural Networks*, 9(3):473–482, 1998.

[46] Shaorong Xie, Peng Wu, Yan Peng, Jun Luo, Dong Qu, Qingmei Li, and Jason Gu. The obstacle avoidance planning of usv based on improved artificial potential field. In *Information and Automation (ICIA), 2014 IEEE International Conference on*, pages 746–751. IEEE, 2014.

[47] Xiaolin Zhu, Hanzhen Xu, and Junqing Lin. Domain and its model based on neural networks. *The Journal of Navigation*, 54(1):97–103, 2001.

[48] 안진형. *뉴로-퍼지 技法을 利用한 船舶의 衝突 回避에 관한 硏究*. PhD thesis, 서울대학교 대학원, 2005.

[49] 유영준. *충돌비율을 이용한 선박충돌회피 알고리즘의 개발*. PhD thesis, 서울대학교 대학원, 2013.

# Abstract

Unmanned Surface Vehicle's ability to prevent collision with nearby obstacle is one of the most critical ability to ensure safety and effectiveness during mission. This is not only for it's own safety, but also to protect life and property of any nearby manned ships. According to this social demands, collision avoidance of an unmanned surface vehicle has been actively researched recently.

In this paper, we have analyzed related research, in order to find out limitation of the previous work. Then, according to the previous work, we have set four goals for the research. Firstly, in estimating collision risk, own ship and obstacle's complex encounter situation should not be simplified too much. Secondly, one should be able to recognize collision risk of the multiple ships. Thirdly, maneuverability of own ship or environmental load should be considered for avoidance maneuver. Fourthly, there should be distinguished standards for different encounter situation.

Generally, collision avoidance of the unmanned surface vehicle follows four stages below. (Detection – Decision making – Path planning – Control) Using obstacle detection sensor, obstacle's position and motion information are identified. Using the information, decision making for collision avoidance is made. Then, path planning is accomplished and controller of the vehicle generate actuator command in order to make avoidance maneuver. In this thesis, deci-

sion making process of the unmanned surface vehicle, using deep reinforcement learning is mainly studied. Reinforcement learning is a method that an agent learns to find out best behavior to accomplish its goal, based on its interaction experience with the environment. In this study, an unmanned surface vehicle is set as a learning agent, and repeated collision avoidance simulation is performed to find out best collision avoidance direction and avoidance time.

In order to perform reinforcement learning, an accurate maneuvering model for the unmanned surface vehicle is needed. For this, we have performed system identification for the dynamic model of the vehicle. Proposed collision avoidance algorithm is built to follow international regulation for ship collision avoidance(COLREGs), designed to meet distinguished standards for different encounter situation. In addition, using the repeated collision avoidance simulation experience, own vehicle's maneuverability and effect of environmental load can be considered. In order to validate collision avoidance, collision avoidance simulation with various scenario have been performed. Using the trajectory, course angle, relative distance, avoidance time information, effectiveness of the proposed collision avoidance method is validated. Deep reinforcement learning based collision avoidance method is then implemented to 16 feet long unmanned surface vehicle, and autonomous collision avoidance test of the vehicle have been performed to validate the proposed method in real world condition.

**Keywords**: Unmanned Surface Vehicle, Collision Avoidance, Reinforcement Learning, Artificial Intelligence

**Student Number**: 2012-23312