



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](#)

공학박사 학위논문

# Velocity Obstacles와 심층강화학습을 이용한 VLCC급 유조선의 충돌회피 방법

A Method for Collision Avoidance of VLCC Tankers  
using Velocity Obstacles and Deep Reinforcement Learning



지도교수 남 중 호

2019년 2월

한국해양대학교 대학원

조선해양시스템공학과

김 동 함

본 논문을 김동함의 공학박사 학위논문으로 인준함.

위원장 : 이 성 욱

위원 : 남 중 호

위원 : 이 승 재

위원 : 윤 현 규

위원 : 정 현



2018 년 12 월 19 일

한국해양대학교 대학원

# 목 차

List of Tables .....	iv
List of Figures .....	vi
Abstract .....	x

<b>1. 서 론</b>	
1.1 연구 배경 .....	1
1.2 연구 사례 .....	3
1.3 연구 목표 .....	6
<b>2. 선박 조종운동의 수학적 모델 및 수치계산법</b>	
2.1 조종운동의 수학적 모델 .....	8
2.2 외력의 수학적 모델 .....	12
2.2.1 주선체에 작용하는 유체력 .....	12
2.2.2 프로펠러에 기인하는 유체력 .....	13
2.2.3 조타에 기인하는 유체력 .....	13
2.3 조종운동 수치계산법 .....	15
2.4 자동 경로 추종 방법 .....	16

<b>3. 선박의 충돌회피 방법</b>	
3.1 국제해상충돌예방규칙 .....	19
3.2 조우상황의 분류 .....	21
3.2.1 조우상황 결정표의 문제점 .....	24
3.2.2 새로운 조우상황 결정표 .....	26
3.3 선박 안전영역 .....	28
3.3.1 Kijima의 선박 안전영역 모델 .....	30
3.4 Velocity obstacles .....	32
3.5 충돌위험도 .....	36
<b>4. 심층강화학습을 이용한 충돌회피</b>	
4.1 심층강화학습 .....	38
4.1.1 DQN(Deep Q-network) .....	41
4.2 충돌회피를 위한 마르코프 결정 과정 정의 .....	44
4.2.1 상태 .....	44
4.2.2 행동 .....	45
4.2.3 보상 .....	47
4.3 DQN 적용 .....	49
4.4 심층강화학습을 이용한 충돌회피 결과 및 고찰 .....	51
<b>5. Velocity obstacles(VO) 이론을 이용한 충돌회피</b>	
5.1 선박의 가용속도 .....	54
5.2 휴리스틱 접근법을 통한 목표 속도 결정 .....	56
5.3 목표 속도 도달을 위한 제어 방법 .....	60
5.3.1 회피 경로를 이용한 방법 .....	61
5.3.2 위치 기반 VO .....	61
5.4 VO를 이용한 충돌회피 결과 및 고찰 .....	67

<b>6. VO와 심층강화학습 기반 충돌회피</b>	
6.1 탐지단계 .....	78
6.1.1 회피 행동 영역 .....	78
6.1.2 상태 .....	81
6.2 의사결정 단계 .....	81
6.2.1 보상 .....	82
6.2.2 심층강화학습의 신경망 학습 .....	84
6.3 행동 단계와 제어 단계 .....	86
6.3.1 행동 정의 .....	86
6.3.2 다수선박에서 행동 정의 .....	89
6.3.3 제어 단계 .....	90
6.4 VO와 심층강화학습 기반 충돌회피 결과 및 고찰 .....	91
6.4.1 다른 방법들과 비교 .....	106
6.4.2 선박 대체 결과 .....	116
<b>7. 결 론</b>	
<b>참고문헌</b> .....	122

## List of Tables

Table 1.1 Occurrences of marine accident according to kinds of accident (2012~2017) .....	2
Table 2.2 Fuzzy rule for autopilot first phase (Furukawa et al., 2004) ....	18
Table 2.3 Fuzzy rule for autopilot second phase (Furukawa et al., 2004)	18
Table 3.1 Abbreviations of encounter types .....	24
Table 3.2 Fuzzy rule for collision risk (Lee and Rhee, 2001) .....	37
Table 4.1 Deep Q-learning with experience replay (Mnih et al., 2015) ....	43
Table 4.2 Principal dimensions of KVLCC .....	49
Table 4.3 Hydrodynamic derivative values .....	50
Table 4.4 Initial conditions for encounter scenario .....	50
Table 5.1 Procedure for target steering angle search .....	66
Table 5.2 Initial condition of 180° head-on situation .....	68
Table 5.3 Initial condition of 90° crossing situation .....	69
Table 5.4 Initial condition of another 90° crossing situation .....	71
Table 5.5 Initial condition of 0° overtaking situation .....	73
Table 6.1 Procedure of detection step .....	79
Table 6.2 Procedure of target velocity selection for multiple ship avoidance .....	90
Table 6.3 Initial condition of multiple ship collision avoidance 1 .....	96
Table 6.4 Initial condition of multiple ship collision avoidance 2 .....	99
Table 6.5 Initial condition of multiple ship collision avoidance 3 .....	101
Table 6.6 Initial condition of multiple ship collision avoidance 4 .....	103
Table 6.7 Initial condition of multiple ship collision avoidance 5 .....	105

Table 6.8 Quantitative measures of collision avoidance to compare with other methods ..... 109

Table 6.9 Quantitative measures of risk distance and extra operation distance ..... 115

Table 6.10 Principal dimensions of SR108 ..... 117

Table 6.11 Hydrodynamic derivative values of SR108 ..... 117



## List of Figures

Fig. 2.1 Coordinate system .....	8
Fig. 2.2 Fuzzy inference process for a rudder angle (Furukawa et al., 2004) .....	17
Fig. 2.3 Parameters used in fuzzy inference for a rudder angle .....	17
Fig. 2.4 Membership functions of fuzzy sets for autopilot first phase (Furukawa et al., 2004) .....	17
Fig. 2.5 Membership functions of fuzzy sets for autopilot second phase (Furukawa et al., 2004) .....	18
Fig. 3.1 Regions used to categorize the position of the obstacle (Tam & Bucknall, 2010) .....	22
Fig. 3.2 Regions used to categorize the heading of the obstacle (Tam & Bucknall, 2010) .....	22
Fig. 3.3 Chart used to determine the encounter type (Tam & Bucknall, 2010) .....	23
Fig. 3.4 Example of crossing situation for encounter decision .....	25
Fig. 3.5 Example of overtaking situation for encounter decision .....	26
Fig. 3.6 Corrected chart to determine the encounter type .....	27
Fig. 3.7 Examples of ship domain (Wang, 2009) .....	29
Fig. 3.8 Ship domains evoked in head-on situation (Wang, 2009) .....	30
Fig. 3.9 Situation of Two moving obstacles (Fiorini & Shiller, 1998) .....	33
Fig. 3.10 Set of collision cone (Fiorini & Shiller, 1998) .....	33
Fig. 3.11 Set of velocity obstacle (Fiorini & Shiller, 1998) .....	34
Fig. 3.12 Set of reachable velocities (Fiorini & Shiller, 1998) .....	35
Fig. 3.13 Set of reachable avoidance velocities (Fiorini & Shiller, 1998) ..	35

Fig. 3.14 Membership functions of fuzzy sets for collision risk (Lee and Rhee, 2001) .....	36
Fig. 4.1 Interaction between environment and agent .....	38
Fig. 4.2 State of agent at encounter situation .....	45
Fig. 4.3 Action of agent at encounter situation .....	46
Fig. 4.4 Trajectories for changing weight .....	52
Fig. 4.5 Time history of speed for changing weight .....	52
Fig. 4.6 Time history of heading for changing weight .....	53
Fig. 5.1 Velocity vector after 2sec .....	55
Fig. 5.2 KVLCC velocity vector after 40sec .....	57
Fig. 5.3 RV in crossing situation ( $\Delta t=40\text{sec}$ ) .....	58
Fig. 5.4 Example of heuristic search .....	58
Fig. 5.5 VO and avoidance direction in each encounter .....	59
Fig. 5.6 Selection of target velocity using heuristic search .....	60
Fig. 5.7 Avoidance course toward target velocity .....	61
Fig. 5.8 Trajectory by steering angle $-10^\circ$ .....	62
Fig. 5.9 PBVO at crossing situation .....	64
Fig. 5.10 Trajectories by steering angle .....	64
Fig. 5.11 Trajectories by target steering angles .....	66
Fig. 5.12 Trajectories of $180^\circ$ head-on situation using VO .....	68
Fig. 5.13 Time histories of $180^\circ$ head-on situation using VO .....	69
Fig. 5.14 Trajectories of $90^\circ$ crossing situation using VO .....	70
Fig. 5.15 Time histories of $90^\circ$ crossing situation using VO .....	70
Fig. 5.16 Trajectories of another $90^\circ$ crossing situation using VO .....	72
Fig. 5.17 Time histories of another $90^\circ$ crossing situation using VO .....	72
Fig. 5.18 Trajectories of $0^\circ$ overtaking situation using VO .....	74
Fig. 5.19 Time histories of $0^\circ$ overtaking situation using VO .....	74

Fig. 6.1 Schematic diagram of collision avoidance method using VO and deep reinforcement learning .....	77
Fig. 6.2 Steps for collision avoidance action .....	80
Fig. 6.3 State of agent for detection step .....	82
Fig. 6.4 Imazu problem .....	85
Fig. 6.5 Action of agent for action step .....	87
Fig. 6.6 Selection method of target velocity for avoidance to port .....	87
Fig. 6.7 Selection method of target velocity for target point following ...	88
Fig. 6.8 Problem of action selection in overtaking situation .....	88
Fig. 6.9 Selection method of target velocity at multiple encounter situation .....	89
Fig. 6.10 Trajectories of 180° head-on situation using VO & DRL at 792sec and 3,168sec .....	91
Fig. 6.11 Trajectories of 180° head-on situation using VO & DRL at 3,960sec and 6,336sec .....	92
Fig. 6.12 Time histories of 180° head-on situation using VO & DRL .....	92
Fig. 6.13 Trajectories of 90° crossing situation using VO & DRL .....	93
Fig. 6.14 Time histories of 90° crossing situation using VO & DRL .....	94
Fig. 6.15 Trajectories of 0° overtaking situation using VO & DRL .....	95
Fig. 6.16 Time histories of 0° overtaking situation using VO & DRL .....	96
Fig. 6.17 Trajectories of multiple ship collision avoidance 1 .....	97
Fig. 6.18 Time histories of multiple ship collision avoidance 1 .....	98
Fig. 6.19 Trajectories of multiple ship collision avoidance 2 .....	99
Fig. 6.20 Time histories of multiple ship collision avoidance 2 .....	100
Fig. 6.21 Trajectories of multiple ship collision avoidance 3 .....	101
Fig. 6.22 Time histories of multiple ship collision avoidance 3 .....	102
Fig. 6.23 Trajectories of multiple ship collision avoidance 4 .....	103

Fig. 6.24 Time histories of multiple ship collision avoidance 4 .....	104
Fig. 6.25 Trajectories of multiple ship collision avoidance 5 .....	105
Fig. 6.26 Time histories of multiple ship collision avoidance 5 .....	106
Fig. 6.27 Collision avoidance results to compare with other methods in head-on situation .....	108
Fig. 6.28 Collision avoidance results to compare with other methods in crossing situation .....	108
Fig. 6.29 Collision avoidance results to compare with other methods in overtaking situation .....	109
Fig. 6.30 Bar graph for comparing max collision risk .....	110
Fig. 6.31 Bar graph for comparing average overlapped area .....	110
Fig. 6.32 Bar graph for comparing extra operation distance .....	110
Fig. 6.33 Collision avoidance results to compare with VO2 and other methods in overtaking situation .....	112
Fig. 6.34 Bar graph to compare with VO2 and other methods in overtaking situation .....	112
Fig. 6.35 Example of risk distance recognized by own ship and target ship .....	114
Fig. 6.36 Bar graph for comparing total quantitative measure ( $w_{total}=0.6$ ) .....	115
Fig. 6.37 Bar graph for comparing total quantitative measure ( $w_{total}=0.8$ ) .....	115
Fig. 6.38 Collision avoidance results to compare with KVLCC and SR108 in head-on situation .....	118
Fig. 6.39 Collision avoidance results to compare with KVLCC and SR108 in crossing situation .....	118
Fig. 6.40 Collision avoidance results to compare with KVLCC and SR108 in overtaking situation .....	119

# A Method for Collision Avoidance of VLCC Tankers using Velocity Obstacles and Deep Reinforcement Learning

Kim, Dong-Ham

Department of Naval Architecture and Ocean Systems Engineering  
Graduate School of Korea Maritime and Ocean University

## Abstract

본 연구에서는 VLCC급 유조선들 간에 충돌회피 시, 기본적인 운항 안전성을 보장하면서 운항효율이 높은 회피 경로를 찾기 위해 VO(Velocity Obstacles)와 심층강화학습을 이용하였다. 선박들 간의 조우상황에서 회피 행동 선택 시 선박안전영역에 근거한 안전거리를 유지함으로써 운항 안전성을 우선적으로 보장하며 목표 운항 경로와 속도를 유지하지 못하여 발생하는 추가 운항 거리를 줄임으로써 운항효율을 높인다. 또한 선박의 회피 행동은 국제해상충돌예방규칙에 근거하여 선택된다.

심층강화학습은 현재 또는 미래에 받을 보상을 최대화하는 최적 행동 정책을 찾는 학습 방법이다. 심층강화학습을 선박의 충돌회피에 활용하여 운항 안전성을 보장하고 운항효율을 높이는 최적 회피 행동 정책을 찾게 된다. 우선 두 선박의 조우상황을 마르코프 결정 과정을 통해 수학적으로 정의하였다. 상태 집합은 자선과 타선의 상태를 잘 나타내는 매개변수를 선정하여 정의하였으며 행동 집합은 목표 운항 경로에 수직한 회피 운항 경로를 생성하고 자선에서 회피 운항 경로까지 수직 거리로 정의하였다. 정의된 마르코프 결정 과정과 DQN(Deep Q-network)을 이용하여 교차 조우상황에 대한 최적 행동 정책을 학습하였으며 이때 목표 운항 경로 이탈 거리와 감쇄 속력에 대한 영향력의 비율을 조절함으로써 해당 비율이 회피 경로에 미치는 영향을 확인하였다. 또

한 학습과 적절한 하이퍼파라미터를 찾는 데 걸리는 시간이 상당한 것을 확인하였다.

VO는 속도를 갖는 물체들 간의 충돌회피를 위한 방법으로 주로 로봇이나 무인수상선에 활용되어져 왔다. 하지만 VLCC급 유조선의 경우 로봇이나 무인수상선과는 달리 조타명령에 의한 응답이 상당히 늦다. 이와 같은 특성을 반영하여 VO를 VLCC급 유조선에 적용하기 위한 방안을 마련한다. VO의 목표 회피 속도를 선정하기 위해서 휴리스틱 접근법이 사용되었으며 선박안전영역을 침범하지 않고 추가 운항 거리를 최소화하는 휴리스틱 조건에 의해 충돌 회피가 가능하였다. 하지만 회피 시작 시점을 지정해야 한다는 단점이 존재하였다.

마지막으로 학습에 사용되는 시간을 줄이고 최적 회피 시작 시점을 유연하게 결정하기 위해 VO와 심층강화학습을 함께 이용하여 VLCC급 유조선의 충돌회피 방법을 제시하였다. VO를 기반으로 국제해상충돌예방규칙에 근거한 행동 종류를 경로추종, 우현 회피, 좌현 회피, 목표지점 추종의 네 가지로 선정하였다. 심층강화학습은 현재 조우상황에서 최적 회피 행동을 판단하고 선택하도록 학습되며, VO를 통해 다수 선박의 조우상황에서도 회피 행동을 할 수 있도록 한다. 학습된 최적 행동 정책은 2척, 3척, 4척의 조우상황과 크기와 형상이 다른 선박으로 대체한 테스트를 수행하였다.

**KEY WORDS:** Collision avoidance 충돌회피; Velocity obstacles 속도 장애물; Deep reinforcement learning 심층강화학습; Deep Q-network 심층 큐 네트워크; COLREGs 국제해상충돌예방규칙.

# 제 1 장 서 론

## 1.1 연구 배경

최근 인공지능의 발전과 자율주행 자동차가 실용화 되고 있는 가운데 무인 시스템에 관한 관심이 커지고 있다. 해상에서도 무인선 관련 기술이 발전함에 따라 무인선의 안전성에 대한 국제 기준과 지침 마련을 위한 요구가 있어 왔다. 영국은 IMO(International Maritime Organization: 국제해사기구) MSC(Maritime Safety Committee: 해사안전위원회) 95차에서 무인선 규정 제정에 대한 필요성을 제기하는 정보문서를 제출하였으며, IMO MSC 98차 회의(2017년 6월)에서 처음으로 MASS(Maritime Autonomous Surface Ships)라는 용어를 사용하였다(박한선, 2018). MASS는 “바다 표면을 자율운항 형태의 무인 항해가 가능한 일반화물선”으로 정의(이윤석, 2018)할 수 있다.

MASS의 등급은 단계적 범제화를 고려하여 IMO MSC 제99차 회의에서 4단계로 구분하였다(이윤석, 2018). 1단계는 자동화된 절차 및 의사결정 시스템을 갖추고 있으나 선원이 운용하고 시스템을 제어하는 선박이다. 2단계는 원격으로 제어되거나 선원이 승선하여 필요할 때 운항에 관여하는 선박이며, 3단계는 선원 없이 원격으로만 제어되는 선박이다. 4단계는 완전한 자율운항선박으로 스스로 선박을 운용하는 수준이다.

현재 노르웨이의 야라 인터네셔널(Yara International)과 콩스버그 그루펜(Kongsberg Gruppen)은 배터리와 전기모터를 이용한 자율운항선박 야라 버클랜드(Yara Birkeland)를 개발하여 2020년 첫 항해를 계획하고 있다(Yara, 2018). 영국의 롤스 로이스(Rolls-Royce)는 구글(Google)과 함께 인공지능 기반 물체인식시스템을 개발하고 있으며 2020년까지 선박 원격조종 기술의 상용화를 목표로 하고 있다(Rolls-Royce, 2017). 우리나라 역시 자율운항선박에 관한 연구를 수행해 왔으며, 선박해양플랜트연구소(KRISO)에서는 “다목적 기능형 무인선 국산화 개발 사업” (김선영 등, 2014)으로 아라곤 II 호를 개발한 바 있다.

선박 운항의 자동화에 있어 가장 비중 있게 고려되고 있는 사항 중 하나는 선박이 항내로 입항 또는 출항할 때 지켜져야 하는 안전성이다. 항내에서 발생할 수 있는 해양사고의 원인으로 여러 가지 요소가 있지만, 그중 선박들 간의 충돌 및 좌초가 특히 중요한 요소로 고려된다. 실제로 최근 5년(2012~2017) 사이 중앙해양안전심판원 통계 자료(중앙해양안전심판원, 2017)에 따르면 Table 1.1과 같이 충돌 및 좌초와 관련된 국내 해양사고의 96.5%는 항해자의 판단 착오 등의 운항 과실에 의해 발생하고 있으며, 이들 중 97%가 출항 후 직무별 과실 및 안전 수칙 미준수와 같은 직접적인 인적 과실(김동진 & 박수용, 2011)에 기인하는 것으로 보고되었다.

**Table 1.1** Occurrences of marine accident according to kinds of accident (2012~2017)

해양사고의 원인		충돌	좌초
운항과실	출항준비불량	1	1
	침로의 선정 유지불량	3	-
	선위확인 소홀	-	1
	조선 부적절	16	16
	경계소홀	704	14
	황천대비, 대응 불량	4	2
	묘박 계류의 부적절	3	-
	항행법규 위반	128	-
	당직근무 태만	17	1
	운항과실 기타	6	4
<b>운항과실 계</b>		<b>882</b>	<b>39</b>
<b>취급 불량 및 결함</b>		<b>4</b>	<b>5</b>
<b>기타</b>		<b>24</b>	<b>5</b>
<b>합계</b>		<b>910</b>	<b>49</b>

MASS가 발전함에 따라 인적 과실에 의한 사고 원인이 줄어들 것으로 기대하고 있다. 이를 위해 MASS가 갖추어야 할 중요 기능 중 하나는 자동 충돌 회피 시스템이며, 스스로 위험을 감지하고 피항 판단을 내릴 수 있는 인공지능 기능을 갖춘 충돌 회피 시스템이 필요하다. 본 연구에서는 이에 대한 기초 연구로 심층강화학습을 이용한 충돌 회피 방법을 제시하고자 한다.

## 1.2 연구 사례

선박안전영역을 이용하여 충돌회피에 이용하는 연구가 진행되어져 왔다. 선박안전영역의 개념은 Fujii and Tanaka(1971)와 Goodwin(1975)에 의해 처음 정의되었으며, 회피 시작 시점에 대한 기준이 된다. 즉, 타선이 자선의 선박안전영역을 침범하였을 때 회피 행동을 시작해야 안전한 회피가 가능하다 여겨지는 것으로 선박안전영역은 이후 조우 각도, 자선과 타선의 속도 등을 고려해 다양한 형상에 관한 연구가 수행되어졌다(Davis et al., 1980; Coldwell, 1983; Zhu et al., 2001; Smierzchalski, 2001; Kijima & Furukawa, 2003; Pietrzykowski & Uriasz, 2004). 선박안전영역은 위험도 평가, 충돌회피, 해상 교통 시뮬레이션, 최적 경로 계획 등과 관련하여 중요한 요소로 사용된다(Wang et al., 2009).

충돌회피 시스템은 자동항해 시스템을 구성하는 주요 요소로 이와 관련된 다양한 연구가 진행되었다. Kose et al.(1998)은 충돌위험도(collision risk)에 의한 위험 분포를 지도에 나타냄으로써 안전한 항로를 결정할 수 있도록 보조하는 시스템을 개발하였다. 충돌위험도는 자선과 타선과의 충돌 위험성을 수치적으로 나타낸 것으로 DCPA(distance at closest point of approach, 최단접근거리)와 TCPA(time at closest point of approach, 최단접근시간)를 이용하여 계산하는 것이 일반적이다. Lee and Rhee (2001), Kijima and Furukawa (2002), 손남선 등(2009), Ota et al.(2016)은 모두 DCPA, TCPA를 퍼지(fuzzy)이론에 적용하여 충돌위험도를 산출하였다. Lee and Rhee는 충돌위험도를 기반으로 전문가 시스템과 A\* 탐색법을 이용하여 충돌 회피 시스템을 개발하였고, Kijima and Furukawa는 충돌위험도에 따라 회피 방향의 범위를 변화시키는 방법을 제안하였다. Lee and Rhee와 Kijima and Furukawa의 방법은 회피 행동공간을 재구성

할 때 행동공간 사이의 변화를 고려하기 힘들다는 단점이 있으며, 손남선 등은 가변공간 탐색법을 이용하여 그 단점을 극복하고자 하였다. 최근 Park and Kim(2017)은 확률 흐름(probability flow) 개념을 이용하여 근사해석적(semianalytical) 방법으로 충돌 확률을 정량적으로 평가하고 충돌위험도로 사용하였다.

앞서 언급한 선박안전영역과 충돌위험도 기반 충돌 회피 시스템은 회피 시점이 정형화되어 있다. 선박안전영역에 의한 방법은 타 선박의 안전영역이 자선의 안전영역을 침범하면 회피를 시작하게 되며, 충돌위험도에 의한 방법은 임계값을 지정하고 충돌위험도가 임계값을 넘으면 회피를 시작한다. 그렇기 때문에 복잡하고 다양한 조우상황에서 유연하게 최적 회피 시점을 결정하기 어렵다. Ota et al.(2016)은 최적의 회피경로를 찾기 위해 회피 시작 시점 및 회피 종료 시점, 회피 경로, 속도의 네 가지 항목들에 대해 각각 몇 가지 행동을 선정하고, 강화학습을 활용하여 조우상황에 따라 최적의 행동을 결정하였다. 하지만 Ota et al.은 타선의 경로 변경과 같은 실시간 조우상황에 대한 대처 행동은 고려하고 있지 않기 때문에 선박의 충돌 회피 문제를 순차적 행동 결정 문제로 정의하고 상황에 따른 최적의 행동 정책을 찾는 방법이 필요하다.

충돌 회피에 관한 연구는 USV(unmanned surface vehicles)와 같은 무인수상선을 대상으로도 수행되었다. Kuwata et al.(2014)은 VO(velocity obstacles)를 이용한 무인수상선의 충돌회피 알고리즘을 제시하였으며 무인수상선이 회피해야 하는 방향을 국제해상충돌예방규칙에 근거하여 분류하고 선택한다. 또한 레이더, 카메라, 라이다(lidar)와 같은 센서로부터 타선의 위험한 움직임을 탐지할 때 정보의 불확실성을 고려하여 VO영역을 확장한 WVO(velocity obstacle with the worst case uncertainty)영역을 정의하고 사용하였다. Kuwata et al.의 방법은 기존 방법과 같이 DCPA 또는 TCPA 임계값에 의해 회피 시작 시점이 결정되기 때문에 여전히 조우상황에 따른 회피 시점을 유연하게 결정하기 어렵다. 또한 조우상황에 따른 피항 각도의 범위가 정해져 있어서 특정한 피항 각도를 향하여 회피하기 때문에 다양한 조우상황에 따라 최적 피항 각도를 유연하게 결정할 수 없다.

한편 박정홍 등(2018)은 무인수상선을 대상으로 충돌확률을 고려한 충돌회피 방법을 수행하였다. 박정홍 등은 무인수상선의 항법 및 탐지 센서 오차에 의한 불확실성을 고려하여 2차원 가우시안 분포(Gaussian distribution)를 이용한 확률 타원을 나타내었으며, 확률 흐름 개념을 이용한 충돌위험도를 활용한다. 그리고 확률 타원과 선박안전영역이 겹쳐지는 충돌 위험 영역을 예측 시점마다 누적하여 표현하고 이를 기반으로 회피 경로를 계획한다. 충돌 확률은 타선의 운동정보에 대한 불확실성을 더 자세히 반영하기 위해서 활용되고 있으며 센서의 불확실성이 클 경우 또는 고속으로 운항하는 무인수상선의 경우 안전한 회피를 위해 이를 고려하는 것이 의미가 있다. 하지만 VLCC급 유조선의 경우 고속 운항을 하는 것이 아니기 때문에 Kuwata et al.이 제시한 WVO를 적용하는 것만으로도 불확실성을 충분히 반영할 수 있다. 그리고 충돌 확률을 이용한 방법 역시 기존 방법과 마찬가지로 충돌 회피 시작을 위한 충돌 확률의 임계값이 정해져 있다. 또한 우주현(2018)은 심층강화학습을 이용한 무인수상선의 충돌회피 방법을 제시하였다. 무인수상선의 충돌회피 방법을 경로추종, 우현회피, 좌현회피 세 가지로 구분하였으며, 선박들 간의 조우상황을 이미지로 표현함으로써 단일 선박 뿐 아니라 다중 선박의 조우상황에서도 최적 행동을 찾을 수 있도록 하였다. 우주현은 회피 시작 시점에 대한 임계값이 지정되어 있는 기존의 방법들과 달리 최적의 회피 시작 시점을 유연하게 결정하기 위해 심층강화학습을 이용하였다. 하지만 Kuwata et al.의 방법과 같이 조우상황에 따른 회피 각도가 지정되어 있으므로 과하게 회피하는 경향이 여전히 존재한다.

우주현, Kuwata et al.과 같이 VO를 적용한 충돌회피 관련 연구는 주로 길이 10m 내외의 무인수상선을 대상으로 하고 있다. VLCC급 유조선은 무인수상선에 비해 크기가 크고 흘수(draft)가 깊기 때문에 유체로부터 받는 저항이 크며, 이로 인해 제어입력(조타명령)에 대한 응답이 무인수상선에 비해 상당히 느리다. VO는 로봇공학 분야에서 로봇의 충돌 회피를 위해 제안된 방법이므로 기본적으로 제어입력에 대한 빠른 응답이 나타나는 물체를 대상으로 연구되었기 때문에 이를 VLCC급 유조선에 적용하기 위해서는 타각 변경에 따른 선박의 속도 변화를 고려할 필요가 있지만 아직 이에 대한 검토가 이루어지지 않았다. 따라

서 본 연구에서는 상기 기술한 이전 연구들의 단점을 극복하기 위해 VO와 심층강화학습을 이용한 충돌회피 방법에 관한 연구를 수행하는 것을 목적으로 한다.

### 1.3 연구 목표

본 연구에서는 로봇의 충돌회피를 위한 방법들 중 하나인 VO와 기계학습을 구성하고 있는 요소기술 중 하나인 강화학습(reinforcement learning)을 이용하여 VLCC급 유조선 충돌회피 방법을 제시하려 한다. 연구 목표는 다음과 같다.

첫째, 기존의 충돌회피 방법들은 국제해상충돌예방규칙에 근거하여 주로 선박의 안전성을 중시하는 회피 방법을 제시하였다. 그동안 충돌 사고와 인적 피해를 방지하기 위해 안전성을 중시하였으나 MASS의 발전과 함께 사람의 개입이 점차적으로 줄어들게 되면 기본적인 안전성을 보장하되 운항효율까지 고려한 충돌회피 방법이 필요하다. 이를 위한 기초 연구로 VO와 심층강화학습을 이용한 충돌회피 방법을 제시한다. 선박의 조우상황은 다양한 패턴이 존재하며 충돌회피 행동을 결정할 때 선박들의 상태와 다양한 회피 행동에 따른 수많은 경우의 수가 존재하기 때문에 안전성과 운항효율을 고려한 충돌회피 경로를 찾는 것은 어려운 문제다. 이와 같이 다양한 패턴과 수많은 경우의 수를 내포하고 있는 문제로부터 최적 행동을 찾는 것은 어려운 문제로 인식되어 왔는데, 최근 심층강화학습을 이용해 이를 해결하려는 연구가 수행되고 있다. Silver et al.(2016)은 바둑과 같이 다양한 패턴과 수많은 경우의 수가 존재하는 문제를 심층강화학습을 이용해 해결할 수 있다는 것을 보여주었으며 선박의 충돌회피 문제 역시 심층강화학습을 적용해 해결 가능할 것으로 판단하였다. 하지만 심층강화학습을 이용한 방법은 학습시간이 상당히 소요되기 때문에 학습시간을 단축할 필요가 있으며 이를 위해 VO 방법을 함께 적용하여 선박의 회피 행동을 제약할 수 있다.

둘째, VO 및 심층강화학습을 이용한 충돌회피 연구는 주로 무인수상선을 대상으로 수행되고 있다(Kuwata et al., 2014; 우주현, 2018). 특히 VO는 로봇 공학 분야에서 로봇의 충돌회피를 위해 개발되었으며 많은 로봇들의 충돌회피에

도 좋은 결과를 보인다. 하지만 VLCC급 유조선은 로봇이나 무인수상선과 같은 운동체와 달리 제어입력에 대한 응답이 매우 늦다는 특징이 있기 때문에 VLCC급 유조선을 대상으로 VO 방법을 적용해 봄으로써 그 문제점을 확인하고 대안을 제시한다.

셋째, 심층강화학습의 목적은 환경으로부터 받는 보상들의 합이 최대가 되는 최적의 행동 양식을 학습하는 것이다. 본 연구는 기존의 연구와는 달리 정형화된 회피 시작 및 종료 시점, 회피 최적 경로 설정을 하는 것이 아니라 많은 반복 행동과 상벌 정책에 따라 경험적 최적 행동을 선택한다. 그러므로 최적 회피 시작 및 종료 시점, 최적 회피 경로는 자선과 타선의 조우상황과 상태에 따라 경험을 기반으로 한 최적 행동을 선택함으로써 만족한다.

선박 조우상황에 대한 심층강화학습의 현실적 환경을 구축하기 위해 조종운동 수학모델이 필요하기 때문에 2장에서는 선박 조종운동 수학모델 및 수치계산법을 설명한다. 3장에서는 본 연구에서 적용되는 국제해상충돌예방규칙과 선박안전영역, VO, 충돌위험도와 같은 충돌회피 방법들에 관해 기술한다. 4장에서는 선박의 충돌회피 문제를 마르코프 의사 결정 방법을 통해 수학적으로 정의하고 심층강화학습을 적용하였으며 교차 조우상황에 대하여 테스트하였다. 5장에서는 VO 방법을 이용해 선박의 충돌회피를 수행하였으며 VLCC급 유조선의 조타명령에 의한 응답이 매우 늦기 때문에 나타나는 특징을 검토하였다. 6장에서는 4장과 5장에서 다루었던 심층강화학습과 VO 방법의 한계를 극복하기 위해 VO와 심층강화학습을 함께 이용한 선박의 충돌회피 방법을 제시한다. 마지막으로 7장에서는 본 논문의 결과를 기술한다.

## 제 2 장 선박 조종운동의 수학적모델 및 수치계산법

본 연구에서는 심층강화학습 및 VO를 이용한 선박의 충돌회피 방법을 구현하는데 있어서 선박 조종운동 시뮬레이션 계산을 이용한다. 평수 중 조종운동 방정식과 외력을 구하는 방법에 대해 설명한다. 조종운동 수학적모델은 Kijima et al.(1990)이 제안한 식을 사용하였다.

### 2.1 조종운동의 수학적모델

평수 중 항주 시에 선박의 조종운동은 일반적으로 2차원 평면에서의 종운동(surge), 횡운동(sway), 선회운동(yaw)에 대한 연동 운동으로 표현된다. 본 연구에서 선박의 운동은 Fig. 2.1과 같은 좌표계를 이용하여 나타내며,  $G-xyz$ 는 선체 고정 좌표계로  $x$ 축,  $y$ 축 방향의 힘을  $X$ ,  $Y$ 로 두고,  $z$  축 주위의 모멘트를  $N$ 으로 한다. 선속  $U$ , 편각  $\beta$ , 회두각  $\psi$ , 회두각속도  $r$ , 타각  $\delta$ 는 각각 Fig. 2.1에 표시된 방향으로 정의한다.

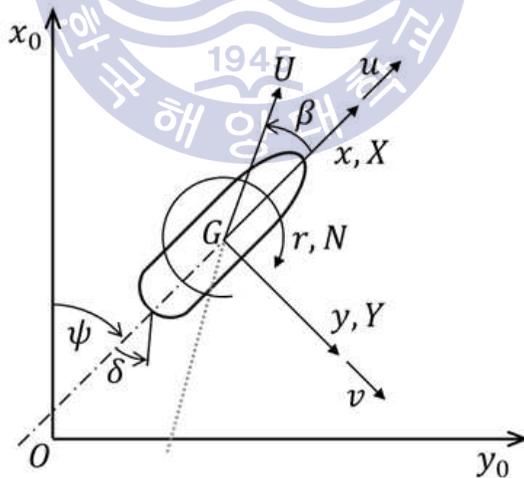


Fig. 2.1 Coordinate system

Fig. 2.1의 관성 좌표계  $O-x_0y_0$ 에서 평면 운동하고 있는 선박의 무게 중심  $G$ 의 가속도를  $\ddot{x}_0$ ,  $\ddot{y}_0$ , 회두 각가속도를  $\dot{r}$ 라 하면 관성 좌표계에서  $x$ ,  $y$ 축 방향

유체력  $X_0$ ,  $Y_0$ 와  $z$ 축 주위의 유체력  $N$  사이에 식 (2.1)과 같은 관계식이 성립한다.

$$\begin{aligned} m\ddot{x}_0 &= X_0 \\ m\ddot{y}_0 &= Y_0 \\ I_{zz}\ddot{\psi} &= N \end{aligned} \tag{2.1}$$

여기서  $m$ 은 선박의 질량,  $I_{zz}$ 는  $z$ 축 주위의 관성 모멘트이다. 선박에 작용하는 유체력을 선체 고정 좌표계 성분으로 표현하면 식 (2.2)와 같으며, 식 (2.1)에 식 (2.2)를 대입하면 식 (2.3)과 같다.

$$\begin{aligned} X &= X_0 \cos\psi + Y_0 \sin\psi \\ Y &= Y_0 \cos\psi - X_0 \sin\psi \end{aligned} \tag{2.2}$$

$$\begin{aligned} X &= m(\ddot{x}_0 \cos\psi + \ddot{y}_0 \sin\psi) \\ Y &= m(\ddot{y}_0 \cos\psi - \ddot{x}_0 \sin\psi) \end{aligned} \tag{2.3}$$

관성 좌표계에서 속도  $\dot{x}_0$ ,  $\dot{y}_0$ 와 선체 고정 좌표계에서 속도  $\dot{x}$ ,  $\dot{y}$ 의 관계식은 좌표 변환 행렬에 의해 식 (2.4)와 같이 나타낼 수 있으므로 미분 공식에 의해 관성 좌표계에서의 가속도  $\ddot{x}_0$ ,  $\ddot{y}_0$ 는 식 (2.5)와 같이 표현 가능하다.

$$\begin{aligned} \dot{x}_0 &= u \cos\psi - v \sin\psi \\ \dot{y}_0 &= v \cos\psi + u \sin\psi \end{aligned} \tag{2.4}$$

$$\ddot{x}_0 = \dot{u} \cos\psi - u\dot{\psi} \sin\psi - \dot{v} \sin\psi - v\dot{\psi} \cos\psi \quad (2.5)$$

$$\ddot{y}_0 = \dot{v} \cos\psi - v\dot{\psi} \sin\psi + \dot{u} \sin\psi + u\dot{\psi} \cos\psi$$

여기서  $u$ ,  $v$ 는  $x$ 축,  $y$ 축 방향의 속도성분이다. 식 (2.5)를 식 (2.3)에 대입하면 식 (2.6)과 같이 선박의 관성 좌표계에 대한 오일러 운동 방정식을 얻을 수 있다.

$$m(\dot{u} - v\dot{\psi}) = X$$

$$m(\dot{v} + u\dot{\psi}) = Y \quad (2.6)$$

$$I_{zz}\ddot{\psi} = N$$

선박의 운동은 선체의 주기 운동에 의한 유체반력이 존재하며 이 중 선체의 운동 가속도에 비례하는 부가질량력을 고려하면 식 (2.7)과 같다.

$$(m + m_x)\dot{u} - (m + m_y)v\dot{\psi} = X$$

$$(m + m_y)\dot{v} + (m + m_x)u\dot{\psi} = Y \quad (2.7)$$

$$(I_{zz} + i_{zz})\dot{r} = N$$

여기서  $m_x$ ,  $m_y$ 은 선체의 중량 및  $x$ ,  $y$ 축 방향의 부가질량,  $i_{zz}$ 는  $z$ 축 주위의 부가 관성 모멘트이다.  $u$ ,  $v$ 는 선속과 편각을 이용해 식 (2.8)과 같이 표현할 수 있으며, 식 (2.8)을 식 (2.7)에 대입하면 식 (2.9)과 같다.

$$u = U \cos\beta \quad (2.8)$$

$$v = -U \sin\beta$$

$$\begin{aligned}
(m+m_x)(\dot{U}\cos\beta-U\dot{\beta}\sin\beta)+(m+m_y)Ur\sin\beta &= X \\
-(m+m_y)(\dot{U}\sin\beta+U\dot{\beta}\cos\beta)+(m+m_x)Ur\cos\beta &= Y \\
(I_{zz}+i_{zz})\dot{r} &= N
\end{aligned} \tag{2.9}$$

식 (2.9)를 무차원화하여 식 (2.10)과 같은 조종운동 방정식을 얻을 수 있다. 기호 옆에 표기된 홀따옴표(')는 무차원화된 값을 의미하며, 무차원화는 식 (2.11)과 같은 방식을 따른다.

$$\begin{aligned}
(m'+m'_x)\left(\frac{L}{U}\right)\left(\frac{\dot{U}}{U}\cos\beta-\dot{\beta}\sin\beta\right)+(m'+m'_y)r'\sin\beta &= X' \\
-(m'+m'_y)\left(\frac{L}{U}\right)\left(\frac{\dot{U}}{U}\sin\beta+\dot{\beta}\cos\beta\right)+(m'+m'_x)r'\cos\beta &= Y' \\
(I'_{zz}+i'_{zz})\left(\frac{L}{U}\right)^2\left(\frac{\dot{U}}{L}r'+\frac{U}{L}\dot{r}'\right) &= N'
\end{aligned} \tag{2.10}$$

$$\begin{aligned}
[m', m'_x, m'_y] &= [m, m_x, m_y] / \left(\frac{1}{2}\rho L^2 d\right) \\
[I'_{zz}, i'_{zz}] &= [I_{zz}, i_{zz}] / \left(\frac{1}{2}\rho L^4 d\right) \\
[X', Y'] &= [X, Y] / \left(\frac{1}{2}\rho L d U^2\right)
\end{aligned} \tag{2.11}$$

$$N' = N / \left(\frac{1}{2}\rho L^2 d U^2\right)$$

$$r' = r L / U$$

여기서  $\rho$ 는 유체밀도,  $L$ 은 선체길이(수선간장),  $d$ 는 흘수이다.

## 2.2 외력의 수학적 모델

앞 절에서 유도한 식 (2.10)의 조종운동 방정식 우변에 있는 외력은 식 (2.12)와 같이 MMG(manoeuvring mathematical model group) 모델(Kijima & Nakiri, 2003)로 표현 가능하다.

$$\begin{aligned} X' &= X'_H + X'_P + X'_R \\ Y' &= Y'_H + Y'_R \\ N' &= N'_H + N'_R \end{aligned} \quad (2.12)$$

여기서 아래첨자  $H, P, R$ 은 각각 선체, 프로펠러, 타와 관련된 항을 의미한다. 프로펠러에 의한 외력 중, 횡력( $Y'_P$ )과 모멘트( $N'_P$ )는 다른 유체력에 비해 매우 작으므로 생략하였다.

### 2.2.1 주선체에 작용하는 유체력

주선체에 작용하는 유체력  $X'_H, Y'_H, N'_H$ 는 식 (2.13)과 같다.

$$\begin{aligned} X'_H &= X'_{\beta r} r' \sin\beta + X'_{uu} \cos^2\beta \\ Y'_H &= Y'_\beta \beta + Y'_r r' + Y'_{\beta\beta} \beta|\beta| + Y'_{rr} r'|r'| + (Y'_{\beta\beta r} \beta + Y'_{\beta r r} r')\beta r' \\ N'_H &= N'_\beta \beta + N'_r r' + N'_{\beta\beta} \beta|\beta| + N'_{rr} r'|r'| + (N'_{\beta\beta r} \beta + N'_{\beta r r} r')\beta r' \end{aligned} \quad (2.13)$$

여기서  $X'_{\beta r}, Y'_\beta, N'_\beta, Y'_r, \dots, N'_{\beta r r}$ 은 유체력 미계수로 Mori(1995)가 제안한 선미선형을 나타내는 파라미터를 사용하여 표현된다(Kijima & Nakiri, 2003).

### 2.2.2 프로펠러에 기인하는 유체력

프로펠러에 기인하는 유체력  $X'_p$ 은 식 (2.14)와 같다.

$$X'_p = (1 - t_p)n^2 D_p^4 K_T(J_p) / \left(\frac{1}{2} L d U^2\right) \quad (2.14)$$

여기서  $t$ 는 추력감소계수,  $n$ 은 프로펠러 회전수,  $D_p$ 는 프로펠러 직경,  $J_p$ 는 프로펠러 전진계수,  $K_T(J_p)$ 는 추력계수를 나타낸다.  $K_T(J_p)$ 와  $J_p$ 는 식 (2.15)와 같이 나타낼 수 있다.

$$K_T(J_p) = C_1 + C_2 J_p + C_3 J_p^2 \quad (2.15)$$

$$J_p = U \cos \beta (1 - w_p) / (n D_p)$$

여기서  $w_p$ 는 프로펠러 위치에서 유효반류계수,  $C_1, C_2, C_3$ 는 프로펠러 단독 성능을 나타내는 상수이다.

### 2.2.3 조타에 기인하는 유체력

조타에 기인하는 유체력은 식 (2.16)과 같이 나타낼 수 있다.

$$\begin{aligned} X'_R &= -(1 - t_R) F'_N \sin \delta \\ Y'_R &= -(1 + a_H) F'_N \cos \delta \\ N'_R &= -(x'_R + a_H \cdot x'_H) F'_N \cos \delta \end{aligned} \quad (2.16)$$

여기서  $F'_N$ 는 타수직력(rudder normal force),  $t_R$ 은 조향에 의한 저항 증가 보정 계수,  $x'_R$ 는 타수직력의 착력점의 x좌표,  $a_H$ 는 선체에 작용하는 타의 간섭력을 나타내는 계수,  $x'_H$ 는 선체에 작용하는 타의 간섭력 작용점의 x좌표를 나타

낸다.  $t_R$ 은 식 (2.17)과 같은 관계가 있으며,  $F'_N$ 은 식 (2.18)과 같이 나타낼 수 있고 식 (2.18)에서 사용되는  $C_N$ ,  $U_R^2$ ,  $\alpha_R$ 는 식 (2.19)와 같다.

$$(1-t_R) = 0.28 C_B + 0.55 \quad (2.17)$$

$$F'_N = (A_R/Ld) C_N U_R^2 \sin \alpha_R \quad (2.18)$$

$$C_N = 6.13 K_R / (K_R + 2.25)$$

$$U_R^2 = (1-w_R)^2 1 + C \cdot g(s)$$

$$g(s) = \eta K \{2 - (2-K)s\} s / (1-s)^2$$

$$\eta = D_P / h_R$$

$$K = 0.6(1-w_P) / (1-w_R)$$

$$s = 1.0 - (1-w_P) U \cos \beta / nP$$

$$w_P = w_{P0} \cdot \exp(-4.0 \beta_P^2) \quad (2.19)$$

$$\beta'_P = \beta - x'_P \cdot r'$$

$$w_R = w_{R0} \cdot w_P / w_{P0}$$

$$w_{P0} = 0.5 C_B - 0.05$$

$$\alpha_R = \delta - \gamma \cdot \beta'_R$$

$$\beta'_R = \beta - 2x'_R \cdot r'$$

$$x'_P \approx -0.5$$

$$x'_R \approx -0.5$$

여기서  $A_R$ 는 타의 면적,  $K_R$ 는 타의 종횡비,  $h_R$ 는 타의 높이,  $U_R$ 는 유효 유입 속도,  $\alpha_R$ 는 유효 유입 각,  $C$ 는 좌우 조향 차에 따른 계수로 우현 회전 시 0.935, 좌현 회전 시 1.065를 그 값으로 한다.  $\gamma$ 는 정류계수,  $P$ 는 프로펠러 피치,  $w_{P0}$ 는 정상 직진 시  $w_P$ 의 값,  $w_R$ 은 타 위치에서 유효반류계수,  $w_{R0}$ 는 정상 직진 시  $w_R$ 의 값,  $C_B$ 는 방형계수,  $s$ 는 프로펠러의 슬립 비,  $x_p$ 는 프로펠러에 의해 작용하는 힘의 x축 방향 위치로 선체 길이에 의해 무차원화된다.

### 2.3 조종운동 수치계산법

앞 절에서 유도한 식 (2.10)의 조종운동 방정식을  $\dot{\beta}$ ,  $\dot{r}$ ,  $\dot{U}$ 에 대해 정리하면 식 (2.20)과 같다.

$$\begin{aligned} \frac{d\beta}{dt} &= \frac{U}{L} \left\{ \frac{m' + m'_x}{m' + m'_y} r' \cos^2 \beta + \frac{m' + m'_y}{m' + m'_x} r' \sin^2 \beta \right. \\ &\quad \left. - \frac{1}{m' + m'_x} \sin \beta \cdot X' - \frac{1}{m' + m'_y} \cos \beta \cdot Y' \right\} \\ \frac{dr'}{dt} &= \frac{U}{L} \left\{ \frac{N' - Y' \cdot \overline{CG}}{I'_{zz} + i'_{zz}} \right\} - r' \frac{U}{r'} \left\{ \frac{X'}{m' + m'_x} \cos \beta - \frac{Y'}{m' + m'_y} \sin \beta \right. \\ &\quad \left. - \left( \frac{m' + m'_y}{m' + m'_x} - \frac{m' + m'_x}{m' + m'_y} \right) r' \sin \beta \cos \beta \right\} \\ \frac{dU}{dt} &= \frac{U^2}{L} \left\{ \frac{X'}{m' + m'_x} \cos \beta - \frac{Y'}{m' + m'_y} \sin \beta \right. \\ &\quad \left. - \left( \frac{m' + m'_y}{m' + m'_x} - \frac{m' + m'_x}{m' + m'_y} \right) r' \sin \beta \cos \beta \right\} \end{aligned} \quad (2.20)$$

여기서  $\overline{CG}$ 는 선체 중앙(midship)과 무게중심 G 사이의 거리를 의미한다. 회두각( $\psi$ ), 관성좌표계에서 선체 중심의 위치( $x_0$ ,  $y_0$ ), 타각( $\delta$ ), 프로펠러 회전수( $n$ )의 시간에 따른 변화율과 관련된 식은 식 (2.21)과 같이 나타낼 수 있다.

$$\frac{d\psi}{dt} = \frac{Ur'}{L}$$

$$\frac{dx_0}{dt} = U\cos(\psi - \beta)$$

$$\frac{dy_0}{dt} = U\sin(\psi - \beta) \tag{2.21}$$

$$\frac{d\delta}{dt} = D_R$$

$$\frac{dn}{dt} = D_N$$

여기서  $D_R$ 은 타각의 변화율,  $D_N$ 은 프로펠러 회전수의 변화율이다.  $\beta$ ,  $r$ ,  $U$ ,  $\psi$ ,  $x_0$ ,  $y_0$ ,  $\delta$ ,  $n$ 을 변수로 하는 연립미분방정식은 컴퓨터에 의한 수치적분법 Runge-Kutta-Gill을 이용하여 그 해를 구한다.

## 2.4 자동 경로 추종 방법

자동 경로 추종 방법은 Furukawa et al.(2004)의 퍼지를 이용한 방법을 활용한다. 전체 과정은 Fig. 2.2와 같으며 크게 두 단계의 과정을 거쳐 경로추종을 위한 타각을 산출한다. 이때 타각 산출을 위해 입력되는 데이터는 목표경로에서 자선까지 수직거리( $l$ )와 목표경로와 이루는 각도( $\phi$ ), 회두각속도( $r'$ )이며, Fig. 2.3을 참고한다.

Fig. 2.2의 첫 번째 단계(first phase)에서 활용되는 퍼지집합은 Fig. 2.4와 같고, 퍼지규칙은 Table 2.2와 같다. 두 번째 단계(second phase)에서 입력되는 실제 회두각속도와 목표각속도의 차이는 식 (2.22)와 같이 산출하며, 퍼지집합은 Fig. 2.5, 퍼지규칙은 Table 2.3과 같다.

$$\Delta r' = r'_d - r' \tag{2.22}$$

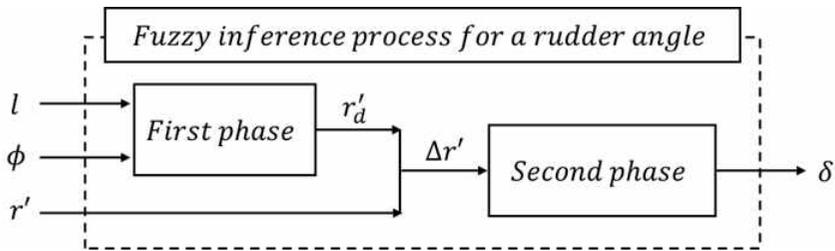


Fig. 2.2 Fuzzy inference process for a rudder angle  
(Furukawa et al., 2004)

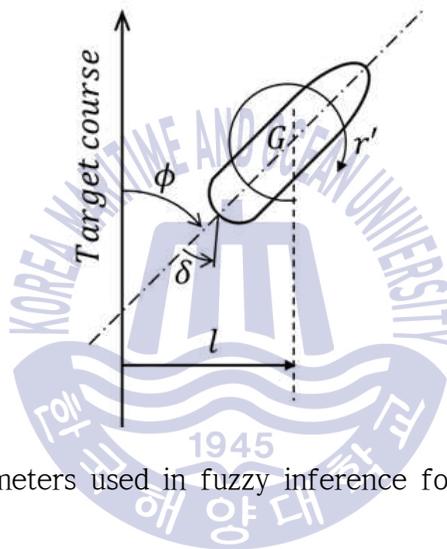


Fig. 2.3 Parameters used in fuzzy inference for a rudder angle

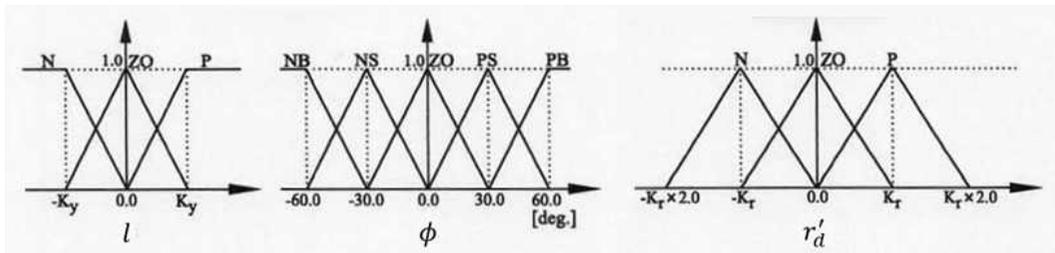
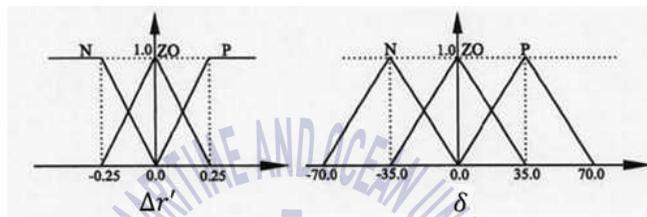


Fig. 2.4 Membership functions of fuzzy sets for autopilot first phase  
(Furukawa et al., 2004)

**Table 2.2** Fuzzy rule for autopilot first phase (Furukawa et al., 2004)

		$\phi$				
		NB	NS	ZO	PS	PB
$l/L$	N	P	P	P	ZO	N
	ZO	P	P	ZO	N	N
	P	P	ZO	N	N	N



**Fig. 2.5** Membership functions of fuzzy sets for autopilot second phase (Furukawa et al., 2004)

**Table 2.3** Fuzzy rule for autopilot second phase (Furukawa et al., 2004)

$\Delta r'$		
N	ZO	P
N	ZO	P

퍼지규칙에서 사용되는 언어변수는 “N: negative”, “B: big”, “S: small”, “ZO: zero”, “P: positive” 를 의미한다. 이와 같은 퍼지집합과 퍼지규칙으로부터 퍼지추론을 하는 방법은 맘다니(Mamdani) 추론법(Mamdani & Assilian, 1975)을 활용한다.

## 제 3 장 선박의 충돌회피 방법

선박 간 조우관계에 있어서 모든 선박에게 적용되는 국제해상충돌예방규칙과 본 연구를 수행하는데 활용되는 충돌회피 방법들에 관해 기술한다. 3.1절은 국제해상충돌예방규칙을 설명하며, 3.2절은 조우관계에서 자선의 운항 방법을 국제해상충돌예방규칙을 기반으로 결정하는 방법을 소개한다. 3.3절은 선박 안전 영역, 3.4절은 velocity obstacles, 3.5절은 충돌위험도 방법을 설명한다.

### 3.1 국제해상충돌예방규칙

해상에서 선박들 간의 충돌을 예방하기 위해 국제해사기구가 1927년 국제해상충돌예방규칙 협약(COLREGs, convention on the international regulation for preventing collisions at sea)을 채택하였으며 COLREGs는 1977년 이후에 발효되었다. COLREGs는 수상에 있는 모든 선박에 적용되며 크게 총칙, 선박의 항법, 등화·형상물, 음향·발광신호에 관한 규정으로 구성되어 있다. 선박의 항법에 관한 규정에서 조우상황에 대한 운항 방법을 기술하고 있다.

COLREGs에 따르면 선박의 조우상황은 크게 마주침(head on), 교차(crossing), 추월(overtaking)의 3가지로 구분할 수 있다. 마주침 상황에서 각 선박은 서로 다른 선박을 좌현에 두고 회피할 수 있도록 우현 측으로 변침해야 한다. 교차 상황은 다른 선박을 우현 측에 두고 있는 선박이 다른 진로를 피해야 하는 피항선(give way vessel)이 되며 다른 선박을 좌현 측에 두고 있는 선박은 유지선(stand on vessel)이 된다. 추월 상황은 추월 방향을 규정하고 있지 않으며, 추월하는 선박은 추월당하는 선박의 진로를 피해야 한다.

COLREGs 내용에서 선박 간 조우관계에 관련된 규정의 일부 내용은 아래와 같다.

#### RULE 13 추월

- (a) 추월선은 제2장 제1절 및 제2절에 있는 규정에도 불구하고 추월당하는 선박의 진로를 피하여야 한다.

- (b) 다른 선박의 정횡 후 22.5도를 넘는 후방 측, 추월당하고 있는 선박과의 관계에서 야간에는 그 선박의 선미등만 볼 수 있고 현등을 볼 수 없는 선박은 추월선으로 보아야 한다.
- (c) 다른 선박을 추월하고 있는지에 관해 의문이 있는 선박은 자선이 추월하고 있는 경우로 생각하고 이에 합당한 동작을 취하여야 한다.
- (d) 두 선박간의 방위가 그 후에 변경되더라도 추월선이 본 규칙상의 의미에서 횡단선으로 되는 것이 아니며 추월선은 완전히 앞질러 멀어질 때까지 추월당하는 선박의 진로를 피하여야 할 의무를 벗어나지 못한다.

**RULE 14 정면으로 마주치는 상태**

- (a) 충돌의 위험이 존재하는 2척의 동력선이 서로 반대되는 침로 또는 거의 반대되는 침로에서 마주치는 경우 각 선박은 서로 다른 선박의 좌현측을 통과할 수 있도록 각기 우현 측으로 변침하여야 한다.
- (b) 서로 다른 선박을 선수방향 또는 거의 선수방향에서 보는 경우 즉, 야간에는 다른 선박의 두 개의 마스트 정부 등을 일직선 또는 거의 일직선상에서 보며 동시에 양현등을 볼 수 있는 경우, 그러한 마스트 정부등이나 또는 양현등만 볼 수 있는 경우, 그리고 주간에 있어서는 다른 선박의 상응하는 면을 보는 경우에는 정면으로 마주치는 상태가 존재한다고 보아야 한다.
- (c) 그러나 상태가 존재하는지 여부에 관해 의문이 있는 선박은 정면으로 마주치는 상태에 있다 생각하고 행동하여야 한다.

**RULE 15 횡단상태**

- (a) 두 척의 동력선이 서로 진로를 횡단할 경우, 충돌의 위험이 있을 때는 다른 선박을 우현 측에 두고 있는 선박이 다른 선박의 진로를 피하여야 하며, 사정이 허락하는 한 다른 선박의 전방을 횡단하면 안 된다.

**RULE 16 피항선의 동작**

- (a) 이 규칙에 따라 다른 선박의 진로를 피하여야 할 선박은 타선을 확실하게

피할 수 있도록 가능한 조기에 충분한 동작을 취하여야 한다.

### RULE 17 유지선의 동작

- (a) (i) 두 선박 중의 한 선박이 다른 선박의 진로를 피하여야 할 경우 다른 선박은 그 침로 및 속력을 유지하여야 한다.  
(ii) 그러나 진로를 피하여야 할 선박이 이 규칙에 따른 적절한 동작을 취하지 아니하고 있음이 분명하여지는 즉시로 유지선은 자신의 조종만으로 충돌을 피하기 위한 동작을 취할 수 있다.
- (b) 이유를 불문하고 침로와 속력을 유지하여야 할 선박은 양선이 아주 가까이 접근하였기 때문에, 피항선의 동작만으로 충돌을 피할 수 없다고 판단할 때에는 충돌을 피하기 위한 최선의 협력 동작을 취하여야 한다.
- (c) 횡단 상태에서 다른 동력선과 충돌을 피하기 위하여 이 조문 (a)항 (ii)의 규정에 따라 동작을 취하는 선박은 상황이 허락하는 한, 자신의 좌현 측에 있는 선박을 피하기 위하여 좌현 측으로 변침하여서는 안 된다.
- (d) 이 조문은 피항선에게 진로를 피하여야 할 의무를 면제하는 것은 아니다.

### 3.2 조우상황의 분류

Tam and Bucknall(2010)은 Fig. 3.1, Fig. 3.2, Fig. 3.3과 같이 간단한 그림들을 제시하여 COLREGs 규칙에 따른 조우상황을 판단할 수 있도록 하였다. Fig. 3.1은 자선을 원의 중앙에 두고 자선의 선수각(heading,  $OS_{\theta_s}$ )을 기준으로 타선이 위치하고 있는 구역을 분류하고 있는 그림이며 모두 여섯 구역(R1~6)으로 나누어져 있다. 구역을 구분하는 각도  $\{HO_1, HO_2, OT_1, OT_2\}$ 는  $\{\pi/8, 15/8\pi, 5/8\pi, 11/8\pi\}$  값을 제시하고 있으며,  $OT_1, OT_2$ 는 COLREGs Rule 13 (b)에서 명시하고 있지만 마주침(head on) 상황을 구분하는 각도  $HO_1, HO_2$ 는 COLREGs Rule 14에서 명시하고 있지 않기 때문에  $\pi/8, 15/8\pi$  값을 추천하고 있다(Tam & Bucknall, 2010). Fig. 3.2는 자선을 원의 중앙에 두고 자선의 선수각( $OS_{\theta_s}$ )을 기준으로 타선의 선수각을 여섯 구역(TSR1~6)으로 분류하고 있다. TSR 각 구역은 Table 3.1과 같은 조우상황 항목들 중 하나로 정의하며, 각 R 구역에

대한 모든 TSR 구역들의 조우상황은 Fig. 3.3과 같은 조우상황 결정표로 정의하고 있다.

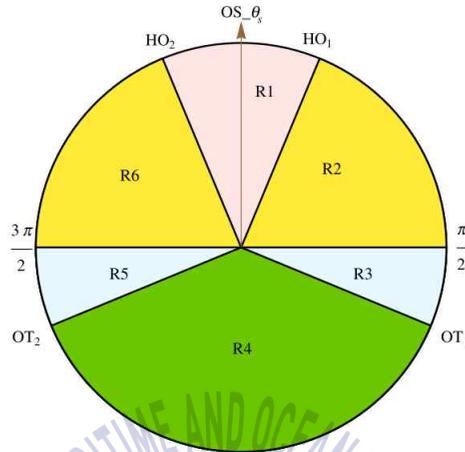


Fig. 3.1 Regions used to categorize the position of the obstacle (Tam & Bucknall, 2010)

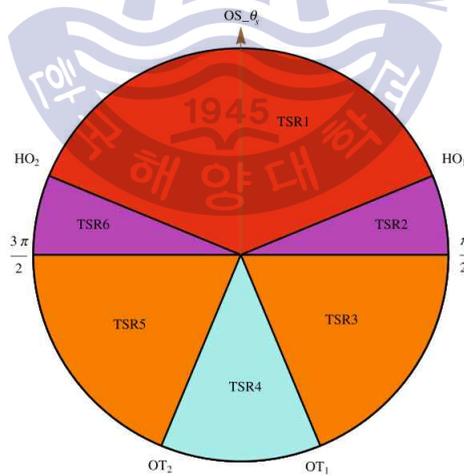


Fig. 3.2 Regions used to categorize the heading of the obstacle (Tam & Bucknall, 2010)

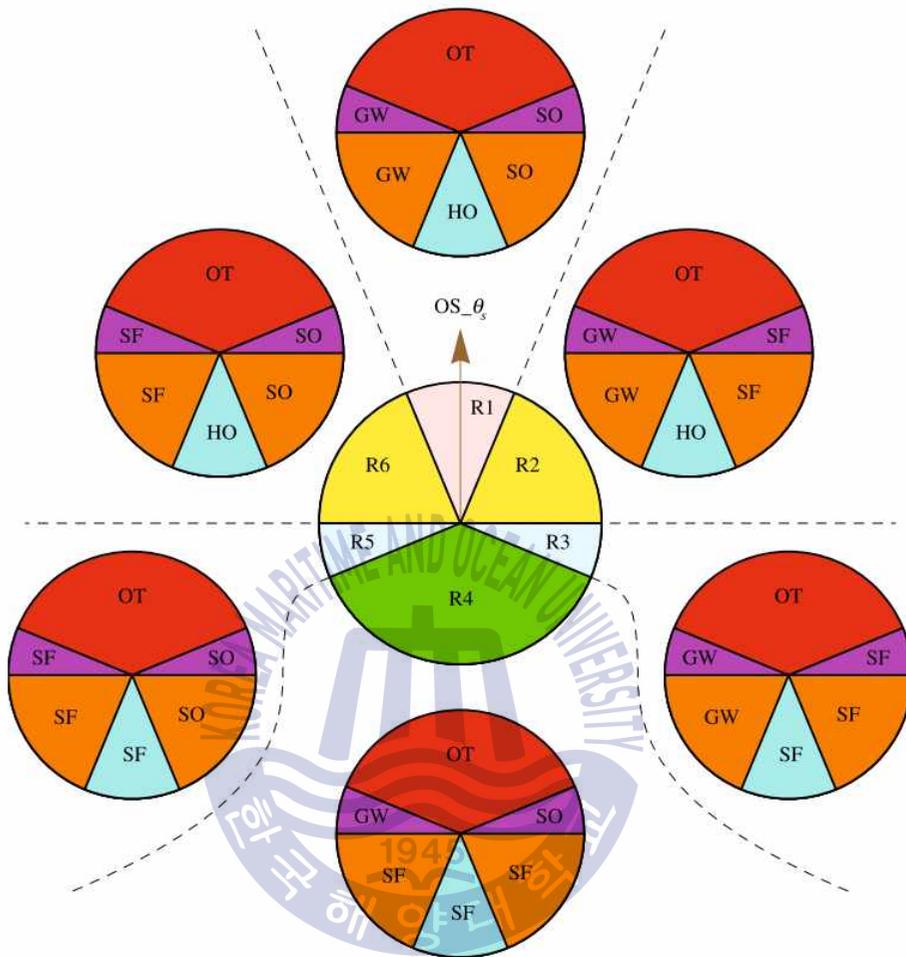


Fig. 3.3 Chart used to determine the encounter type (Tam & Bucknall, 2010)

**Table 3.1** Abbreviations of encounter types

약어	설명
HO	마주침(head-on) 조우상황
OT	추월(overtaking) 조우상황
SO	유지선(stay-on) 조우상황
SF	안전(safe)한 조우상황
GW	피항선(give-way) 조우상황
ST	정적 장애물(static obstacle)

동일 선상에서 마주침 조우상황을 가정하여 예로 들면, Fig. 3.1에서 자선의 선수각( $OS_{\theta_s}$ ) 방위에 타선이 존재하며 R1 구역에 해당한다. Fig. 3.2에서 타선의 선수각은 마주침 조우상황이기 때문에 자선의 선수각( $OS_{\theta_s}$ )과 반대방향 즉, 선수각을 기준으로 180도 회전한 각도이기 때문에 TSR4 구역에 해당한다. Fig. 3.3의 R1 구역에서 TSR4는 HO 즉, 마주침 조우상황으로 정의되어 있기 때문에 자선은 마주침 선박이 된다.

### 3.2.1 조우상황 결정표의 문제점

Fig. 3.4와 같이 선박의 45도 교차 조우상황에서 Tom & Bucknall이 제안한 Fig. 3.3 조우상황 결정표에 따르면 자선(ownship, OS)은 타선(target ship, TS)에 대해 추월선(OT)의 의무를 갖게 된다. COLREGs Rule 13 (b)에 의해 추월선은 타선의 정횡선(Fig. 3.4의 1점 쇄선) 후방에 위치해야 한다. 하지만 45도 교차 조우상황에서 Fig. 3.4와 같이 자선은 타선의 정횡선 전방에 위치하기 때문에 추월선으로 볼 수 없으며, 이는 교차 조우상황으로 자선은 피항선(GW)의 의무를 가진다. Tam & Bucknall이 제안한 조우상황 결정표의 경우 추월 조우상황에서 특정 시간 간격마다 조우상황을 실시간으로 판단하기 위해 TSR1 구역을 모두 추월선(OT)으로 정의한 것으로 보인다.

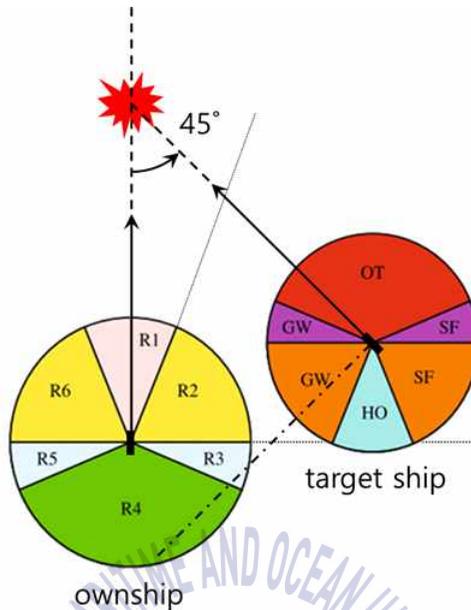


Fig. 3.4 Example of crossing situation for encounter decision

또한 Fig. 3.3 조우상황 결정표는 추월선과 추월당하는 선박 모두 추월선(OT)으로 분류되어져 있다. Fig. 3.5는 추월 조우상황을 나타내고 있으며, Fig. 3.5에서 자선 주변의 TSR은 타선의 입장에서 자선에 대한 TSR 분류표(R4)이며, 타선 주변의 TSR은 자선의 입장에서 타선에 대한 TSR 분류표(R1)이다. 이와 같이, 기존 조우상황 결정표에 따르면 자선과 타선 모두 추월선(OT)으로 분류한다. COLREGs는 Rule 13에서 추월당하는 선박인 피추월선에 대한 언급이 없기 때문에 피추월선의 행동을 정의하기 어렵지만 피추월선은 침로와 속력을 유지해야 추월선과의 충돌 가능성을 줄일 수 있다. 윤점동(2015)은 피추월선의 행동에 대해 침로와 속력을 유지해야 할 것은 틀림이 없으나 횡단 상태의 유지선이 행하여야 하는 정도로 침로와 속력을 엄격하게 유지해야 하는 것은 아닌 뜻으로 받아들여야 된다고 권고하고 있다. 본 연구에서 피추월선은 유지선(SO)의 의무를 갖는 것으로 분류하였으며, 유지선의 경로가 바뀌더라도 추월선이 충분히 대응할 수 있도록 한다.

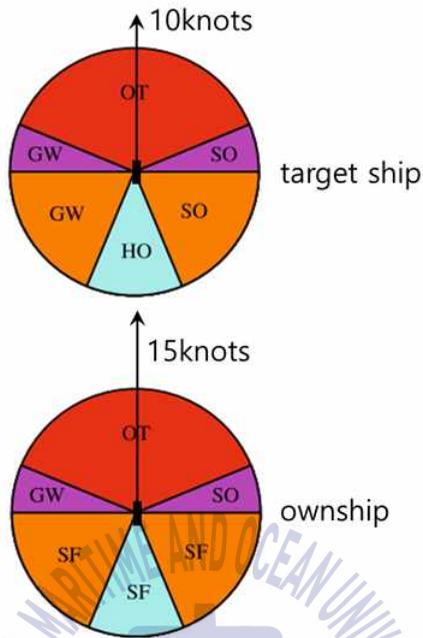


Fig. 3.5 Example of overtaking situation for encounter decision

### 3.2.2 새로운 조우상황 결정표

Tam & Bucknall(2010)이 제안한 조우상황 결정표 Fig. 3.3은 앞서 서술한 것과 같이 두 가지 문제를 가지고 있다. 첫째, 교차 조우상황에서 피항선(GW) 임에도 불구하고 추월선(OT)으로 분류되는 경우가 있다. 둘째, 추월선과 피추월선 모두 추월선(OT)으로 분류한다.

두 가지 문제점을 해결하기 위해 본 논문에서는 한 가지 가정과 함께 Fig. 3.6과 같은 새로운 조우상황 결정표를 제시한다. 가정은 타선이 자선을 기준으로 하는 특정 영역을 침범하였을 때부터 그 영역을 벗어날 때까지 자선의 조우상황은 변경되지 않는다는 것이다. 즉, 특정 영역 안에서 자선 또는 타선의 경로가 변경되어도 조우상황은 바뀌지 않으며, 이는 추월 상황에서 COLREGs Rule 13 (d)의 규정을 만족시킨다. 본 논문에서는 해당 규정을 마주침과 교차 조우상황까지 확대 적용한다.

Fig. 3.6에서 변경된 항목들을 보면 R2, R3에 대한 TSR1 영역은 피항선(GW)으로 분류하고 R5, R6에 대한 TSR1 영역은 유지선(SO)으로 분류하였다. 이는 Fig. 3.4에서와 같이 COLREGs 기반의 조우상황을 만족시키기 위해 변경되었다. R1과 R4에 대한 TSR1 영역은 자선과 타선의 속력에 따라 추월선(OT)과 유지선(SO)이 결정되도록 한다. 속력이 상대적으로 느린 선박은 속력이 빠른 선박을 추월할 수 없기 때문에 속도를 비교하여 조우상황을 분류하였다.

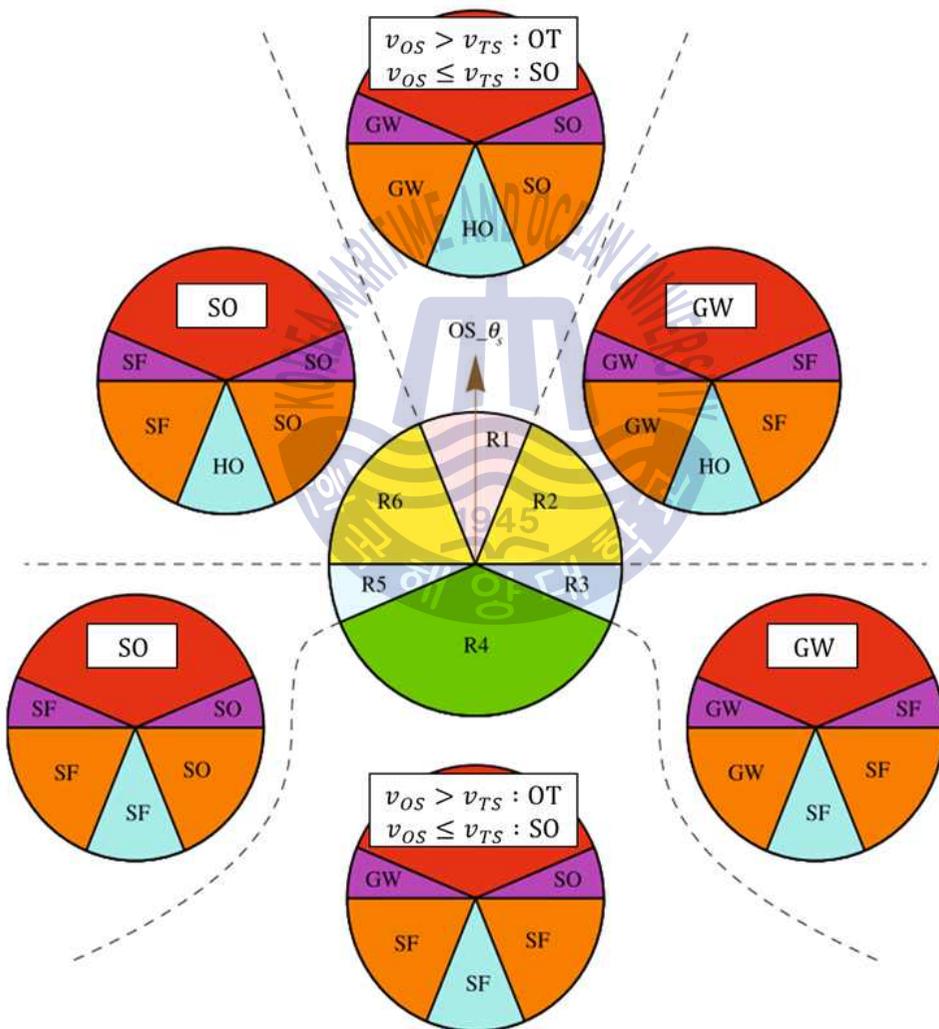


Fig. 3.6 Corrected chart to determine the encounter type

### 3.3 선박 안전영역

Wang(2009)은 David, Zhu, Pietrzykowski, Goodwin, Kijima, Coldwell, Fujii가 제안한 7개의 다양한 선박 안전영역을 비교, 분석하였다. 각 안전영역에 대한 파라미터들과 형상은 Fig. 3.7과 같다. 선박 안전영역은 기하학적인 특징에 따라 원(circle), 다각형(polygon), 타원(ellipse) 형상으로 구분할 수 있다. Goodwin과 Davis 모델은 원, Smierzchalski, Pietrzykowski 모델은 다각형, Fujii, Coldwell, Kijima 모델은 타원의 형상을 보인다.

Wang(2009)은 마주침, 교차, 추월 상황에 대해 시뮬레이션 분석을 수행하였으며 Fig. 3.8은 마주침 조우상황에서 각 안전영역 모델을 비교한 것이다. Fig. 3.8에서 Davis 모델이 가장 큰 영역으로 나타나고 Fujii 모델이 가장 작은 영역으로 나타난다. 작은 영역으로 나타나고 있는 Fujii 모델과 Coldwell 모델은 충돌 회피 행동을 취하기에 너무 위험해 보이기 때문에 활용 시 조심할 필요가 있다 (Wang, 2009).

본 연구에서는 VO 및 심층강화학습을 이용한 충돌회피를 위해 회피 경계를 판단하는 기준으로 선박 안전영역을 활용하였다. Fig. 3.8과 같이 Wang이 비교한 각 모델의 형상에서 Davis, Zhu, Pietrzykowski 모델은 회피 경계로 활용하기에 너무 과도하며 Fujii, Coldwell은 상대적으로 작은 경계를 갖고 있기 때문에 위험성을 수반한다. Goodwin 모델은 대부분 우현 변침 회피를 규정하고 있는 COLREGs에 근거하여 우현 방향 안전거리를 더 확보하기 위하여 비대칭 형태의 안전영역을 제안하고 있다. Goodwin은 회피 경계로 활용하기에 적절한 크기를 갖고 있으나 비대칭 영역을 나타내고 있기 때문에 추후 VO 활용에 어려움이 있다. 따라서 본 논문에서는 영역 크기의 안전성과 형상의 대칭성, 연속성을 만족시키는 Kijima 모델을 회피 경계로 활용한다.

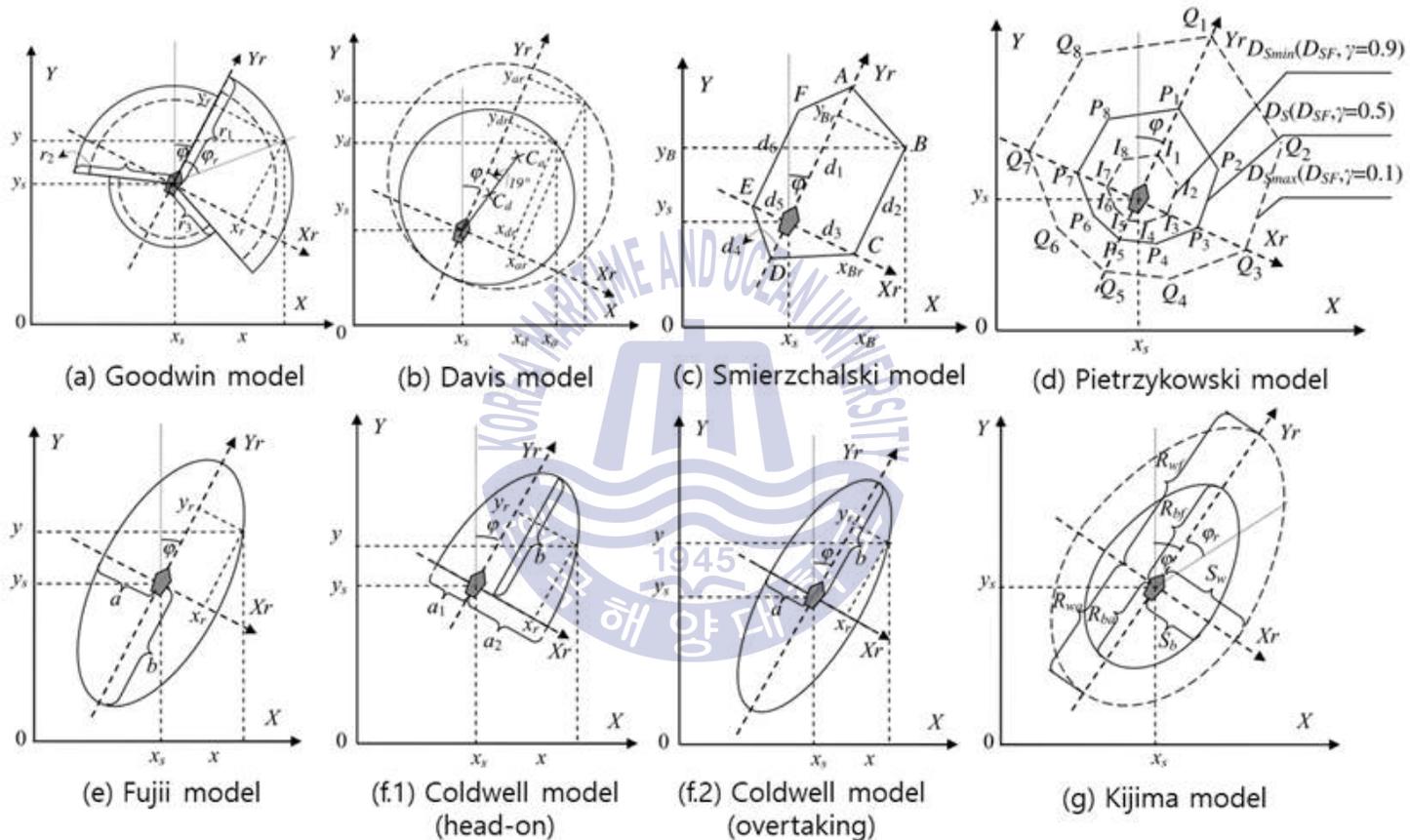


Fig. 3.7 Examples of ship domain (Wang, 2009)

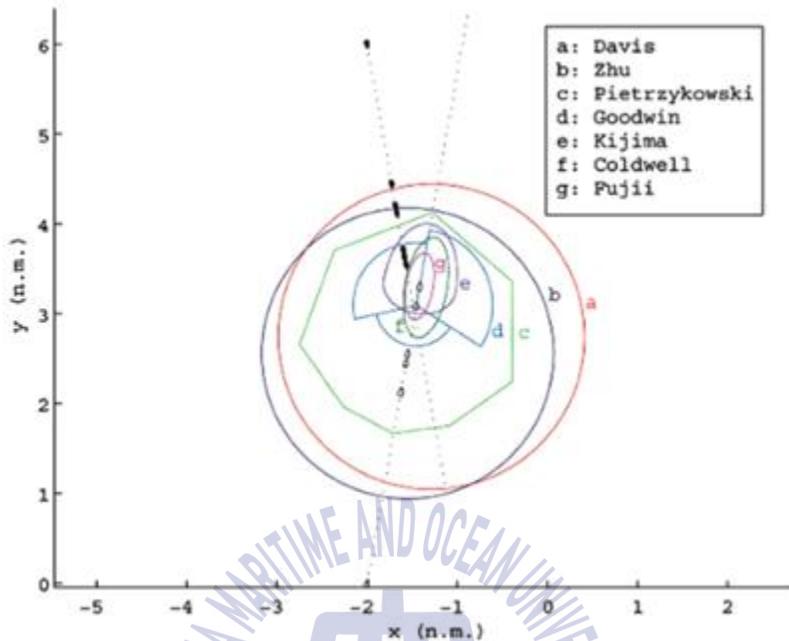


Fig. 3.8 Ship domains evoked in head-on situation (Wang, 2009)

### 3.3.1 Kijima의 선박 안전영역 모델

Kijima and Furukawa(2003)의 선박 안전영역 모델은 Fig. 3.7 (g)와 같이 두 개의 타원을 붙여놓은 형상을 하고 있으며 선수( $R_{bf}$ ,  $R_{wf}$ )와 선미( $R_{ba}$ ,  $R_{wa}$ ), 횡축( $S_b$ ,  $S_w$ ) 방향의 3가지 파라미터에 의해 타원 크기가 결정된다. 또한 Fig. 3.7 (g)에서 실선으로 나타낸 폐색영역(blocking area)과 점선으로 나타낸 감시영역(watching area)의 두 가지 영역을 제안하고 있다. 폐색영역은 타선의 감시영역이 폐색영역을 침범했을 때 자선이 현재 경로를 변경 또는 유지 할 것인지 여부를 결정하는 영역이다. 폐색영역과 감시영역의 형상을 결정하는 파라미터의 계산식은 각각 식 (3.1), 식 (3.2)와 같다.

$$R_{bf} = L + (1 + s) T_{90} U$$

$$R_{ba} = L + T_{90} U \quad (3.1)$$

$$S_b = B + (1 + t) D_T$$

$$R_{wf} = L + 2(R_{bf} - L)$$

$$R_{wa} = L + 2(R_{ba} - L) \quad (3.2)$$

$$S_w = B + 2(S_b - B)$$

여기서  $L$ ,  $B$ ,  $U$ 는 각각 선박의 길이, 폭, 노트 단위 속력을 나타내고,  $T_{90}$ 는 90도 선수각까지 회전하는데 걸리는 시간,  $D_T$ 는 선회지름(tactical diameter),  $s$ 와  $t$ 는 조우상황의 영향을 고려한 계수로 식 (3.3)과 같고  $U_T$ 는 타선의 속력이다.

$$s = 2 - (U - U_T)/U; \quad t = 1 \quad \text{while head-on encounter}$$

$$s = 2 - \alpha/\pi; \quad t = \alpha/\pi \quad \text{while crossing encounter} \quad (3.3)$$

$$s = 1; \quad t = 1 \quad \text{while overtaking encounter}$$

$T_{90}$ 과  $D_T$ 는 식 (3.4)와 같은 추정식을 사용하여 구한다(Kijima & Furukawa, 2003).

$$T_{90} \simeq (0.67/U) \cdot \sqrt{A_D^2 + (D_T/2)^2}$$

$$A_D = L \exp(0.3591 \log U + 0.0952) \quad (3.4)$$

$$D_T = L \exp(0.5441 \log U - 0.0795)$$

### 3.4 Velocity obstacles

VO는 로봇공학 분야에서 주로 활용되고 있으며 로봇들의 충돌회피를 위한 알고리즘이다. VO란 용어는 Fiorini and Shiller(1993)가 처음 사용하였으며, 장애물의 크기와 속도를 기반으로 충돌 가능성이 존재하는 영역을 생성하고 해당 영역을 피해 행동을 결정하도록 한다.

Fig. 3.9와 같은 상황에서 로봇 A가 VO를 이용하여 속도가 있는 장애물 B를 피해 행동을 결정하는 방법을 순서대로 정리하면 다음과 같다.

① 충돌 원뿔(collision cone,  $CC_{A,B}$ ) 집합 정의

$$CC_{A,B} = \{ \mathbf{v}_{A,B} \mid \lambda_{A,B} \cap \hat{B} \neq \emptyset \} \quad (3.5)$$

식 (3.5)에서  $\mathbf{v}_{A,B}$ 는 B에 대한 A의 상대속도고  $\lambda_{A,B}$ 는  $\mathbf{v}_{A,B}$ 의 확장선이다. Fig. 3.10과 같이 A는 한 점으로 줄이고 A의 크기를 B에 덧붙인다. 즉, B의 크기는 B의 반지름에 A의 반지름을 더한 크기  $\hat{B}$ 가 되며, 이는 A와 B의 크기를 고려하여 두 물체가 부딪치지 않도록 VO를 고려하기 위한 함이다.  $CC_{A,B}$ 는 Fig. 3.10의 회색 영역과 같이 A를 원점으로  $\hat{B}$ 를 외접하는 두 탄젠트 벡터( $\lambda_f, \lambda_r$ ) 사이 영역이다.

② 속도 장애물(velocity obstacle,  $VO_B$ ) 집합 정의

$$VO_B = CC_{A,B} \oplus \mathbf{v}_B \quad (3.6)$$

식 (3.6)에서  $\oplus$ 는 민코프스키(Minkowski) 벡터합 연산자이며,  $VO_B$ 는 Fig. 3.11과 같이 민코프스키 벡터합 연산자에 의해  $CC_{A,B}$ 를  $\mathbf{v}_B$ 에 의해 이동시킨 것과 같다. 민코프스키 벡터합은 유클리드 공간에 존재하는 X와 Y의 두 벡터 집합에서 X에 있는 모든 벡터를 각각 Y에 있는 벡터에 더해 새로운 벡터 집합을 생성하는 연산이다.

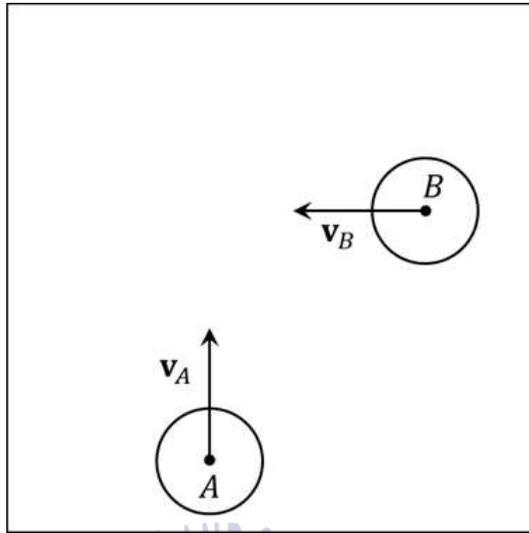


Fig. 3.9 Situation of Two moving obstacles  
(Fiorini & Shiller, 1998)

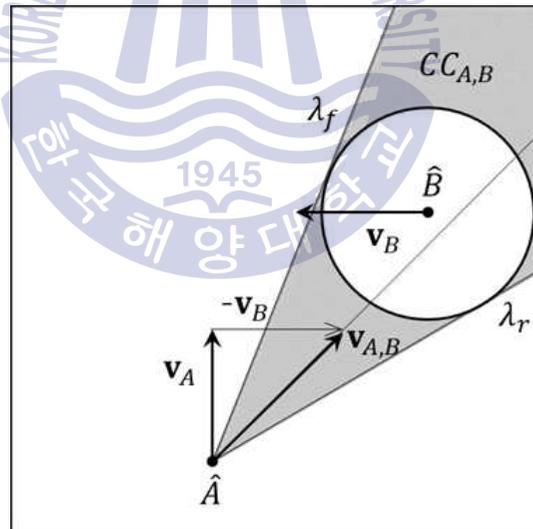


Fig. 3.10 Set of collision cone (Fiorini & Shiller, 1998)

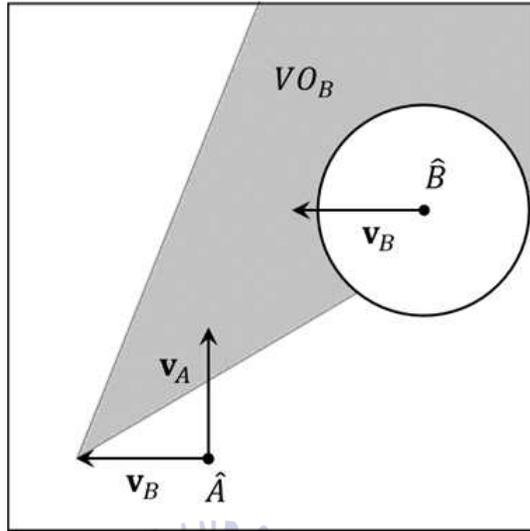


Fig. 3.11 Set of velocity obstacle  
(Fiorini & Shiller, 1998)

③ 실현 가능한 가속도(feasible accelerations,  $FA$ ) 집합 정의

$$FA(t) = \{ \ddot{\mathbf{x}} \mid \dot{\mathbf{x}} = f(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{u}), \mathbf{u} \in U \} \quad (3.7)$$

식 (3.7)에서  $\mathbf{x}$ 는 위치 벡터,  $f(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{u})$ 는 로봇 동역학 함수,  $\mathbf{u}$ 는 액추에이터(actuator)의 변화에 따른 벡터이고,  $U$ 는 로봇에서 구현 가능한 제어들의 집합이다. 즉, 현재 상태에서 제어기의 작동 정도에 따라 나타날 수 있는 모든 가속도의 집합을 의미한다.

④ 가용속도(reachable velocities,  $RV$ ) 집합 정의

$$RV(t + \Delta t) = \{ \mathbf{v} \mid \mathbf{v} = \mathbf{v}_A(t) \oplus \Delta t \cdot FA(t) \} \quad (3.8)$$

Fig. 3.12와 같이 미소시간  $\Delta t$ 에 대한  $FA(t)$ 를 현재 A의 속도에 더한 영역이  $RV$ 가 된다.

⑤ 가용회피속도(reachable avoidance velocities,  $RAV$ ) 집합 정의

$$RAV(t + \Delta t) = RV(t + \Delta t) \ominus VO(t) \quad (3.9)$$

식 (3.9)에서  $\ominus$ 는 여집합을 의미하며 RAV는 Fig. 3.13와 같이 RV 영역에서 VO 영역을 제외시킨 회색 영역과 같다.

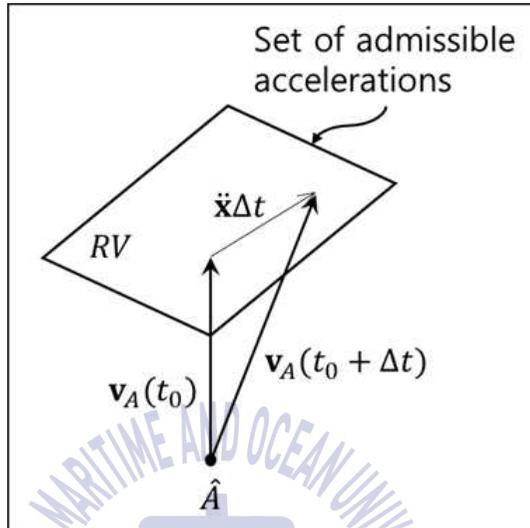


Fig. 3.12 Set of reachable velocities  
(Fiorini & Shiller, 1998)

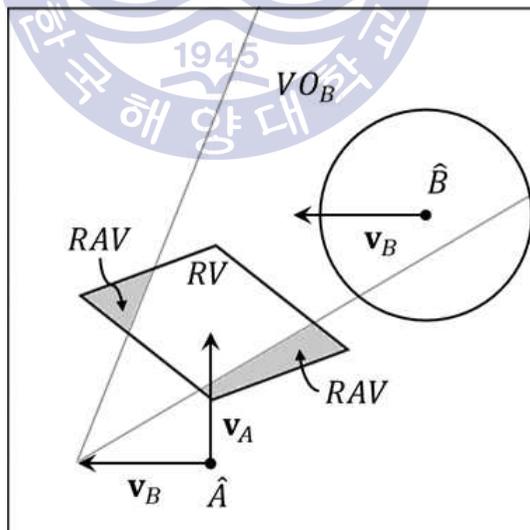


Fig. 3.13 Set of reachable avoidance velocities  
(Fiorini & Shiller, 1998)

VO에서 충돌을 회피하기 위해 RAV 영역 안에 들어가는 속도 벡터들 중 하나를 선택한다. 선택을 위한 방법은 크게 전체 탐색(global search)과 휴리스틱 탐색(heuristic search)이 있다. 전체 탐색은 오프라인 적용 방법으로 RAV 영역에서 노드(node)들을 추출하고 모든 경우의 수를 구하여 목적 함수에 따라 최적 경로를 찾는다. 휴리스틱 탐색은 휴리스틱 조건에 의한 최적 경로를 선택한다.

### 3.5 충돌위험도

충돌위험도는 Lee and Rhee(2001)의 퍼지를 이용한 방법을 활용한다. 충돌위험도 산출에 활용되는 퍼지집합은 Fig. 3.14와 같고, 퍼지규칙은 Table 3.2와 같다. 퍼지집합에서 TCPA(time to the closest point of approach)는 두 선박이 현재 위치에서 선속과 침로를 유지하며 항해하였을 경우, 두 선박이 가장 근접한 거리가 되기까지 걸리는 시간이다. DCPA(distance to the closest point of approach)는 이때의 가장 근접한 거리이다.

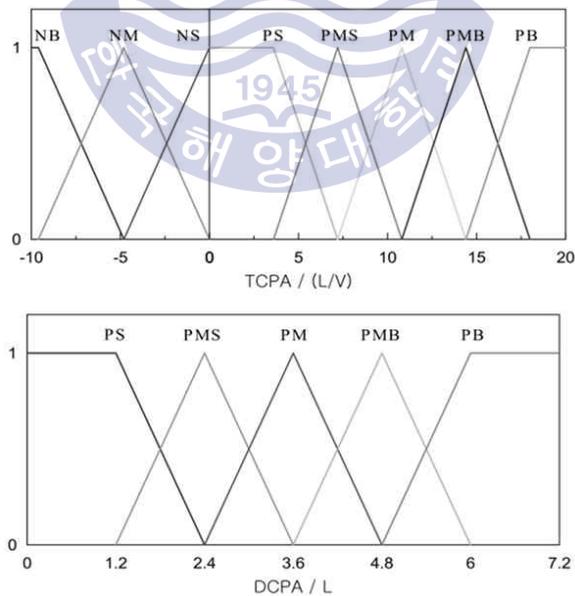


Fig. 3.14 Membership functions of fuzzy sets for collision risk (Lee and Rhee, 2001)

Table 3.2 Fuzzy rule for collision risk (Lee and Rhee, 2001)

		<i>DCPA</i>				
		PS	PMS	PM	PMB	PB
<i>TCPA</i>	NB	-0.2	-0.2	-0.2	-0.2	-0.2
	NM	-0.6	-0.2	-0.2	-0.2	-0.2
	NS	-1.0	-0.6	-0.2	-0.2	-0.2
	PS	1.0	0.8	0.6	0.4	0.2
	PMS	0.8	0.6	0.4	0.2	0.2
	PM	0.6	0.4	0.2	0.2	0.2
	PMB	0.4	0.2	0.2	0.2	0.2
	PB	0.2	0.2	0.2	0.2	0.2

퍼지규칙에서 추가된 언어변수는 “M: middle” 을 의미한다. 이와 같은 퍼지 집합과 퍼지규칙으로부터 최종 충돌위험도를 구하는 방법은 식 (3.10)과 같다. 여기에서  $\alpha_i$ 는 n개의 퍼지집합 함수 중, i번째 함수에 의해 결정된 적합도이며,  $CR_i$ 는 Table 3.2 퍼지규칙 상에서 출력을 계산하는데 사용되는 소속값이다.

$$Collision\ Risk = \frac{\sum_{i=1}^n CR_i \cdot \alpha_i}{\sum_{i=1}^n \alpha_i} \quad (3.10)$$

## 제 4 장 심층강화학습을 이용한 충돌회피

심층강화학습 방법에 대한 이론과 함께 심층강화학습을 이용한 충돌회피 방법에 대해 설명한다. 선박의 충돌회피 상태를 마르코프 의사 결정 방법에 따라 수학적으로 정의하고 심층강화학습을 통해 최적 행동 정책을 찾는다.

### 4.1 심층강화학습

강화학습은 기계학습의 한 분야로 강화의 개념은 행동학자인 스키너(B.F.Skinner)가 스키너 상자를 이용하여 쥐가 보상을 통해 행동과 그 결과 사이의 관계를 학습하는 것을 확인함으로써 처음 사용하였다. 강화학습에서 에이전트(agent)는 어떠한 행동을 통해 환경으로부터 보상을 받게 되는데 현재 또는 미래에 보상을 최대로 받을 수 있는 최적의 행동을 학습하는 것이 강화학습의 목적이다 (Fig. 4.1).

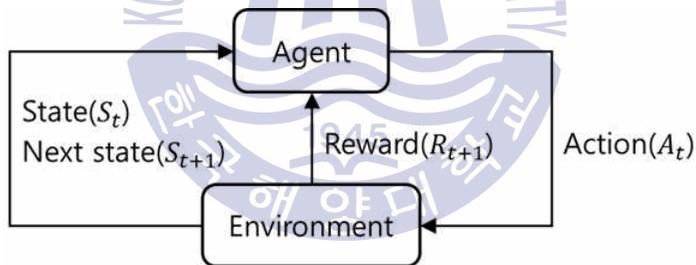


Fig. 4.1 Interaction between environment and agent

선박의 충돌 회피 상황에서 각 선박은 에이전트가 되며 에이전트는 해상이라는 환경에 존재한다. 강화학습을 통해 에이전트의 최적 행동을 구하기 위해서는 우선 충돌 회피 상황을 순차적 행동 문제로 나타내는 것이 필요하다. MDP(Markov decision process)를 이용하여 순차적 행동 결정 문제를 수학적으로 정의할 수 있으며 MDP의 구성요소는 상태(states,  $S$ ), 행동(actions,  $A$ ), 보상(reward,  $R$ ), 상태 변환 확률(state transition probability), 감가율( $\gamma$ , discount factor)이다.

**상태( $S$ )**는 에이전트가 환경에서 관찰하고 있는 자신의 상태 집합이다. 선박의 조우상황에서 각종 센서로부터 받는 현재 자선의 상태 또는 타선의 상태, 해상 상태 등이 상태 집합으로 고려되어 질 수 있으며 상태 집합을 표현할 때 이미지 또는 매개변수가 이용된다. Mnih et al.(2013), 우주현(2018)은 에이전트와 환경의 상태를 이미지로 나타내고 이미지의 각 픽셀에 대한 수치 정보를 상태 집합으로 사용하였다. 반면 Chae et al.(2017)은 에이전트의 현재 상태를 잘 표현할 수 있는 매개변수를 정의하고 매개변수를 상태집합으로 사용하였다. 그리고 **행동( $A$ )**은 에이전트가 할 수 있는 가능한 모든 행동의 집합이다. 선박의 조우상황에서 타각 및 추력의 제어가 행동 집합이 될 수 있으며 이때 본 연구에서 사용하는 DQN(Deep Q-network)의 행동 집합은 이산적으로 표현되어야 하기 때문에 제어 입력을 이산화하는 과정이 필요하다. **보상( $R$ )**은 에이전트가 행한 행동으로 인해 환경이 에이전트에게 주는 상벌이며 선박의 조우상황에서 COLREGs 위반과 같은 행동에 대한 상벌을 고려할 수 있다. MDP에서 **상태 변환 확률**은 현재 상태에서 어떤 행동을 할 경우 다음 상태에 도달할 수 있는 확률이다. 상태 변환 확률은 자유모델 기반 강화학습 방법에서 신경망(neural network) 학습 시에 반영되기 때문에(Li, 2017) 자유모델 기반 강화학습 방법인 DQN은 상태 변환 확률을 생략할 수 있다. 자유모델은 에이전트가 환경에 대해 알지 못해도 탐사(exploration)를 통해 무작위 행동을 수행해 봄으로써 그 보상 결과를 통해 행동 정책을 학습하는 방식이다. 조타명령에 따른 응답이 느린 VLCC급 선박의 경우 자유모델 기반 강화학습 방법을 적용하면 조타명령에 대한 응답이 느리더라도 행동 정책 학습이 가능하다. **감가율( $\gamma$ )**은 보상의 현재 지닌 가치가 시간이 지남에 따라 감소하는 것을 고려하여 수학적으로 표현한 것으로 0에서 1 사이의 상수로 결정된다.

정의된 MDP에서 최적 행동 정책을 찾기 위해 벨만(Bellman) 기대 방정식(식 4.1)과 벨만 최적 방정식(식 4.2)이 활용된다. 기본적으로 벨만 방정식들은 다이나믹 프로그래밍을 이용하여 풀 수 있으며 이는 각종 문제점을 보완하며 살사(SALSA), 큐러닝(Q-learning)으로 발전하였다(이응원 등, 2017). 선박의 조우상황에서 최적 행동 정책을 찾는다는 것은 제어 행동에 대한 보상을 최대화하는 회

피 경로를 찾는다는 의미와 동일하다.

$$V_{\pi}(s) = \mathbf{E}_{\pi}[R_{t+1} + \gamma V_{\pi}(S_{t+1}) | S_t = s] \quad (4.1)$$

$$V^*(s) = \max_a \mathbf{E}[R_{t+1} + \gamma V^*(S_{t+1}) | S_t = s, A_t = a] \quad (4.2)$$

$\pi$ 는 정책(policy)으로 현재 상태( $S_t$ )에서 에이전트가 취할 수 있는 어떤 행동( $A_t$ )의 확률이며  $V_{\pi}$ 는 정책을 고려한 가치함수로 앞으로 받게 될 보상에 대한 기댓값( $\mathbf{E}$ )을 의미하고 감가율( $\gamma$ )에 대한 영향이 포함되어 있다. 벨만 기대 방정식(식 4.1)은 현재 상태에서의 가치함수와 다음 상태에서의 가치함수 사이의 관계식을 나타내고 있으며 벨만 최적 방정식(식 4.2)은 앞으로 받게 될 보상에 대한 기댓값이 최대가 되는 최적 정책을 따라 행동했을 때( $\max_a$ ) 가치함수를 나타내고 있다.

벨만 기대 방정식과 벨만 최적 방정식은 큐함수(Q function,  $Q_{\pi}(s, a)$ )를 이용하여 나타낼 수 있으며 식 (4.3), 식 (4.4)와 같다. 큐함수는 현재 상태( $s$ )에서 한 행동( $a$ )에 따라 앞으로 받을 보상의 합을 출력하는 일종의 함수로 상태와 행동, 두 가지 변수를 가진다. 반환값 형태로 표현된 현재 상태에서의 가치함수는 반환값의 특성상 종료시점까지 행동해야지만 계산 가능하다. 반면 큐함수는 현재 상태에서 한 행동에 의해 앞으로 받게 될 보상을 미리 알고 있으며, 가치함수를 대신하여 사용된다.

$$Q_{\pi}(s, a) = \mathbf{E}_{\pi}[R_{t+1} + \gamma Q_{\pi}(S_{t+1}, A_{t+1}) | S_t = s, A_t = a] \quad (4.3)$$

$$Q^*(s, a) = \mathbf{E}[R_{t+1} + \gamma \max_{a'} Q^*(S_{t+1}, a') | S_t = s, A_t = a] \quad (4.4)$$

큐함수에 대한 벨만 최적 방정식(식 4.4)에서 최적 큐함수( $\max_{a'} Q^*(S_{t+1}, a')$ )는 가장 큰 큐함수를 갖는 다음 행동( $a'$ )을 선택하는 것이다. 여기서 홑따옴표(')는 다음 시간( $t+1$ )에 대한 값을 의미한다.

### 4.1.1 DQN(Deep Q-network)

DQN을 설명하기 위해 우선 큐러닝에 대한 이해가 필요하다. 큐러닝은 식 (4.5)와 같이 큐함수를 업데이트하며 다음 행동에 대한 큐함수( $Q(S_{t+1}, a')$ )가 현재 행동에 대한 큐함수( $Q(S_t, A_t)$ )에 영향을 주지 않는 오프폴리시 시간차 제어(off-policy temporal-difference control) 방법이다. 식 (4.5)는 현재 행동에 대한 큐함수를 다음 행동에 대한 큐함수를 이용하여 업데이트하고 있으며 이때 다음 행동에서의 큐함수는 실제 다음 행동( $A_{t+1}$ )에 대한 큐함수가 아니라 최대의 큐함수( $\max_a Q$ ) 값을 갖는 행동( $a'$ )에 의해 결정된다. 즉, 실제 다음 행동은 큐함수 업데이트에 영향을 주지 않고 최적 행동이 큐함수 업데이트에 영향을 준다. 식 (4.5)의  $\alpha$ 는 큐함수를 업데이트하는 정도를 의미하는 상수다.

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha(R + \gamma \max_a Q(S_{t+1}, a') - Q(S_t, A_t)) \quad (4.5)$$

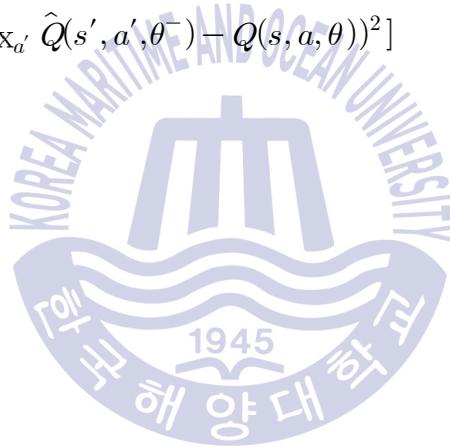
DQN은 앞서 설명한 큐러닝과 인공신경망을 함께 사용함으로써 인공신경망을 통해 Q 함수를 학습하는 방법이다. 이를 위해 Mnih et al.(2015)은 큐러닝에서 두 가지 새로운 기능을 추가하였다. 첫 번째는 경험 리플레이(experience replay) 기능이다. 에이전트는 환경을 탐험하며 현시점의 상태와 행동, 보상과 관련된 샘플( $s, a, r, s'$ )을 얻게 되며, 이를 컴퓨터 메모리에 저장한다. 저장된 샘플들 중 무작위로 샘플들을 추출하여 학습에 사용하는 것이 경험 리플레이 방법이다. 강화학습의 특성 상 DQN 이전 방법들은 연속된 시간상에서 일어나는 연관된 샘플들을 이용하여 학습되기 때문에 국부 최적화에 빠지기 쉬웠다. 경험 리플레이 방법에서 무작위로 추출된 샘플들은 서로 상관관계가 작기 때문에 국부 최적화에 잘 빠지지 않고 목표에 잘 수렴한다. 두 번째는 목표 신경망(target network)을 분리한 것이다. 신경망 업데이트를 위한 손실함수(loss function)는 평균 제곱 오차(mean squared error, MSE)를 사용하며 식 (4.6)과 같다. 손실함수에서  $y$ (식 4.7)는 정답 역할을 하고  $Q(s, a, \theta)$ 는 예측의 역할을 하고 있으며, 신경망은 정답과 예측의 차이인 손실함수를 최소화하도록 신경망을 업데이트한다. 하지만 여기에서 정답이 변하면 안 되지만 정답과 예측에 사

용되는 큐함수( $Q$ )가 동일하기 때문에 큐함수가 업데이트되면서 정답도 같이 변하게 된다. 이를 방지하기 위해 식 (4.8)과 같이 목표 신경망( $\hat{Q}(s', a', \theta^-)$ )을 따로 두고 목표 신경망은 주기적으로 인공신경망( $Q(s, a, \theta)$ )으로 업데이트된다. 이로써 목표 신경망은 일정 기간 변하지 않고 정답 역할을 수행한다. DQN의 절차는 Table 4.1과 같다.

$$L = \mathbf{E}[(y - Q(s, a, \theta))^2] \quad (4.6)$$

$$y = \mathbf{E}[R_{t+1} + \gamma \max_{a'} Q(s', a', \theta) | S_t = s, A_t = a] \quad (4.7)$$

$$L = \mathbf{E}[(R_{t+1} + \gamma \max_{a'} \hat{Q}(s', a', \theta^-) - Q(s, a, \theta))^2] \quad (4.8)$$



**Table 4.1** Deep Q-learning with experience replay (Mnih et al., 2015)

```
Initialize replay memory  $D$  to capacity  $N$ 
Initialize action-value function  $Q$  with random weights  $\theta$ 
Initialize target action-value function  $\hat{Q}$  with weights  $\theta^- = \theta$ 
for episode = 1,  $M$  do
    Initialize sequence  $s_1$  and preprocessed sequence  $\phi_1 = \phi(s_1)$ 
    for  $t = 1, T$  do
        With probability  $\varepsilon$  select a random action  $a_t$ 
        otherwise select  $a_t = \operatorname{argmax}_a Q(\phi(s_t), a; \theta)$ 
        Execute action  $a_t$  in emulator and observe reward  $r_t$  and
        image  $x_{t+1}$ 
        Set  $s_{t+1} = s_t, a_t, x_{t+1}$  and preprocess  $\phi_{t+1} = \phi(s_{t+1})$ 
        Store transition  $(\phi_t, a_t, r_t, \phi_{t+1})$  in  $D$ 
        Sample random minibatch of transitions  $(\phi_j, a_j, r_j, \phi_{j+1})$  from  $D$ 
        Set  $y_j = \begin{cases} r_j & \text{if episode terminates at step } j+1 \\ r_{j+1} + \gamma \max_{a'} \hat{Q}(\phi_{j+1}, a'; \theta^-) & \text{otherwise} \end{cases}$ 
        Perform a gradient descent step on  $(y_j - Q(\phi_j, a_j; \theta))^2$  with
        respect to the network parameters  $\theta$ 
        Every  $C$  steps reset  $\hat{Q} = Q$ 
    end for
end for
```

## 4.2 충돌회피를 위한 마르코프 결정 과정 정의

선박의 충돌회피를 위한 순차적 행동 결정 문제를 MDP를 통해 수학적으로 나타낸다. 에이전트는 각 선박이 되고 환경은 2장에서 정의한 선박의 운동조종 수학모델을 기반으로 한 해상 환경이다. 에이전트는 목표 경로를 추종하며 운항하는데, 다른 선박과 충돌 위험이 있을 때 회피 행동을 한다. 이때 안전성과 운항효율을 고려한 최적의 행동 정책을 찾고자 한다.

### 4.2.1 상태

상태( $\mathcal{S}$ )는 MDP의 한 요소로 두 선박의 조우상황에서 에이전트가 처해 있는 상태를 수치적으로 표현한 집합이다(Fig. 4.2). 본 논문에서는 식 (4.9)와 같이 상태 집합을 정의하였다.

$$\mathcal{S} = \{v_{i1}, v_{j1}, r_1, l, \psi_s, \delta, i_{rel}, j_{rel}, v_{i_{rel}}, v_{j_{rel}}\} \quad (4.9)$$

여기서 아래첨자  $i$ 와  $j$ 는 목표 경로 방향과 목표 경로에 수직인 방향을 의미한다.  $v_{i1}$ ,  $v_{j1}$ 은 자선의  $i$ ,  $j$  방향 속도,  $r_1$ 은 자선의 선회 각속도,  $l$ 은 목표 경로에서 자선까지 수직하게 떨어진 거리,  $\psi_s$ 는 목표 경로와 선수방향이 이루는 각도,  $\delta$ 는 타각,  $i_{rel}$ ,  $j_{rel}$ 은  $i$ 와  $j$ 방향으로 타선의 상대 거리,  $v_{i_{rel}}$ ,  $v_{j_{rel}}$ 는  $i$ 와  $j$ 방향으로 타선의 상대속도( $v_{i2} - v_{i1}$ ,  $v_{j2} - v_{j1}$ )다.

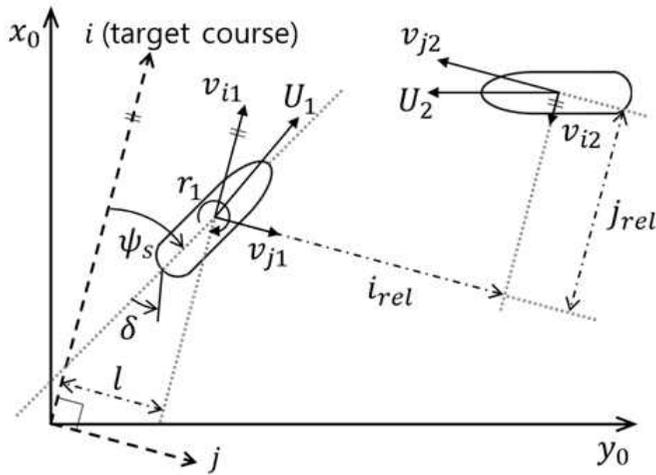


Fig. 4.2 State of agent at encounter situation

#### 4.2.2 행동

본 연구에서는 Fig. 4.3과 같이 목표 경로(target course)를 추종하고 있는 자선이 타선과 충돌 가능성이 있을 경우, 새로운 회피경로(avoidance course)를 생성하고 추종함으로써 충돌을 회피하도록 한다. 회피경로는 목표 경로에 평행하고 자선을 중심으로 정의된다. 행동( $a$ )은 자선의 중심으로부터 회피경로까지 수직 거리로 정의한다. DQN에서 에이전트가 할 수 있는 행동은 이산적으로 표현되기 때문에 회피경로의 최대, 최소 수직 거리( $\pm a_{\max}$ )를 정하고 이를 11개로 분할하였으며, 각 분할된 행동은 식 (4.10)과 같이 행동 집합의 원소로 정의한다. 행동 원소의 수가 많을수록 더 섬세한 제어가 가능하지만 선택할 수 있는 행동이 늘어나기 때문에 최적 행동정책을 찾기 위한 학습시간이 더 많이 걸리며, 행동 원소의 수를 줄이면 학습시간은 줄어들지만 제어가 어려워진다. 각 원소의 값은 식 (4.11)에 의해 계산되며, 목표 경로 추종 시 자선이 목표 경로에 가까울수록 세밀한 제어가 필요하기 때문에 Fig. 4.3과 같이 자선 근처에서 거리의 간격이 좁고 멀수록 넓어지도록 하였다.

$$A = \{a_0, a_1, a_2, \dots, a_{10}\} \quad (4.10)$$

$$a_i = \begin{cases} a_{\max} \cdot (i/5)^2 & \text{if } i < 6 \\ -a_{\max} \cdot ((i-5)/5)^2 & \text{else} \end{cases} \quad (4.11)$$

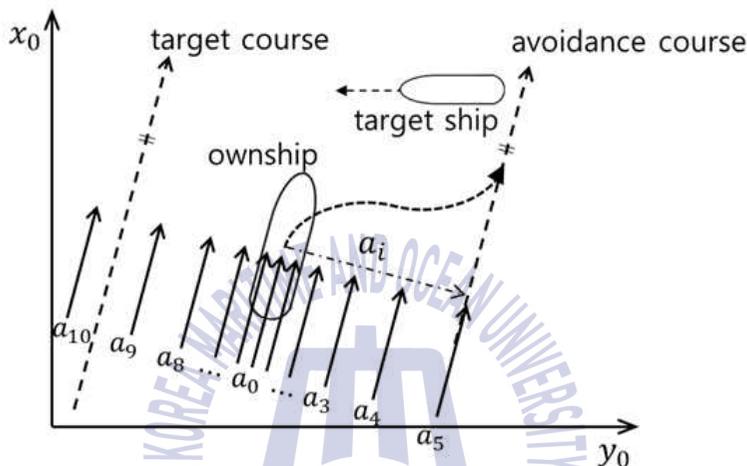


Fig. 4.3 Action of agent at encounter situation

자동 경로추종 제어기는 목표 경로를 설정하였을 때 배가 목표 경로를 따라 운항하도록 타를 자동 조종한다. 자동 경로추종 제어기는 PID(proportional - integral-derivative)제어를 이용한 방법과 퍼지제어를 이용한 방법이 연구되어 왔으며, PID 제어를 이용한 방법을 사용하면 적절한 게인 값을 찾아야 하는 어려움이 있으므로 본 연구에서는 2.4절의 퍼지를 이용한 자동 경로추종 제어를 활용하였다.

Fig. 4.3과 같이 타선이 나타날 경우 목표 경로는 회피 경로로 대체되고, 자선이 회피 경로를 향하도록 자동 경로추종 제어기가 작동한다. 자선과 타선 사이에 충돌 가능성이 없을 때는 자선이 목표 경로를 유지할 수 있도록 회피 경로를 목표 경로와 일치시킨다. 강화학습 시 처음에는 무작위로 행동을 선택하게 되며 학습을 반복할수록 무작위 행동을 줄이고 학습된 행동을 선택하게 된다. 강화학습 과정에서 충돌과 회피를 반복하다 보면 타선을 회피하는 것이 높은

보상을 받는다는 것을 학습하게 되고 회피 경로를 이동하여 목표 경로를 벗어나도록 변칙하는 피항 행동을 하게 된다. 자선이 타선을 회피한 후, 목표 경로에 가깝게 운항하는 행동을 선택하는 것이 높은 보상을 받는다는 것을 학습하게 되면 자선은 목표 경로로 복귀하는 행동을 하게 된다.

### 4.2.3 보상

강화학습에서 에이전트는 현재 처한 상태( $S_t$ )에서 어떤 행동( $A_t$ )을 했을 때, 다음 상태( $S_{t+1}$ )가 되면서 환경으로부터 보상( $R_{t+1}$ )을 받는다(Fig. 4.1). 보상을 최대한으로 받거나 받게 될 행동을 학습하는 것이 강화학습의 목표이기 때문에 보상은 강화학습 적용에서 가장 중요한 요소이다. 본 연구에서는 두 선박의 조우 상황에서 운항효율과 안전성을 고려한 최적의 회피 경로를 찾도록 한다.

식 (4.12)과 같이 보상은 운항효율과 안전성에 관련된 두 가지 보상( $R_e$ ,  $R_s$ )으로 나눌 수 있다. 운항효율은 항해 거리를 줄여 목표 경로에서 벗어나는 거리를 최소화하는 것과 피항 행동으로 인한 속도 감쇄를 최소화함으로써 항해 속력을 가능한 유지하는 것과 밀접한 관계가 있다. 즉, 피항을 위해 목표 경로에서 자선의 위치가 수직인 거리( $l$ )만큼 떨어지게 되면 운항 거리는 늘어나고 변칙에 의한 속력( $U$ ) 감쇄로 운항 시간이 늘어난다. 그러므로 자선이 가능한 목표 경로에 가까이 위치하고 속력을 유지할수록 운항효율과 관련된 보상( $R_e$ )이 큰 값을 갖도록 한다.

$$R = R_e + R_s \quad (4.12)$$

식 (4.13)의 운항효율과 관련된 보상의 첫 번째 항( $D_1$ )은 자선이 목표 경로에서 벗어난 수직 거리( $l$ ) 만큼 음의 보상을 받게 하며 이는 자선이 목표 경로에 가능한 가깝게 피항 행동을 하도록 한다. 또한 자선의 속력( $U$ )이 목표 항해속력( $U_{init}$ )보다 증가하면 연료 소모가 많아지고, 감소하면 운항시간 늘어남으로 운항효율이 떨어진다. 목표 항해속력을 기준으로 속력의 증·감소에 비례하여 음의 보상을 받게 하며, 선회 시에 가능한 자선의 속력이 느려지지 않도록 학

습한다. 세 번째 항은 조건 변수  $c_e$ 를 두어 자선이 목표 경로의 좌측에 위치하도록( $l < 0$ ) 좌현 변침할 경우 음의 보상 받게 하며 COLREGs 규칙과 같이 자선의 좌현에 타선을 두고 우현 방향으로 피항 행동을 하도록 한다.  $w_r$ 은 회피 거리에 대한 보상( $D_1$ )과 속력의 감쇄에 대한 보상( $D_2$ )의 비를 조절하기 위한 가중치로 0에서 1 사이의 값을 가진다. 거리와 속력과 관련된 두 보상의 비율에 따라 회피 행동의 변화를 관찰하기 위해  $w_r$  가중치를 사용하였다.

$$R_e = w_r \cdot D_1 + (1 - w_r) \cdot D_2 + c_e \times l \quad \begin{cases} \text{if}(l < 0), & c_e = 1 \\ \text{otherwise,} & c_e = 0 \end{cases} \quad (4.13)$$

$$D_1 = -|l|$$

$$D_2 = -|U_{init} - U| \cdot \Delta t$$

안전성과 관련된 보상( $R_s$ )은 식 (4.14)와 같다.  $w_s$ 는 충돌 시 받게 되는 가중치 상수로 5를 그 값으로 하였으며,  $c_s$ 는 충돌 여부에 대한 조건 변수로 선박이 충돌할 경우 1이 되며 그렇지 않으면 0이다. 실제 충돌은 두 선박이 부딪친 상태지만 본 논문에서는 회피 거리의 안전성을 보장하기 위해 자선의 선체 윤곽이 타선의 폐색영역(blocking area)을 침범하면 충돌하였다 간주한다. 폐색영역은 3.3.1절에서 언급하였으며 식 (3.1)을 통해 얻을 수 있다.

$$R_s = -w_s \times c_s \quad \begin{cases} \text{if}(collision), & c_s = 1 \\ \text{otherwise,} & c_s = 0 \end{cases} \quad (4.14)$$

강화학습 과정에서 두 선박이 충돌할 경우, 운항효율 보상( $R_e$ )보다 안전성 보상( $R_s$ )이 상대적으로 큰 음의 값을 가지기 때문에 타선과 충돌하지 않는 범위 내에서 운항효율 보상을 최대화하는 피항 행동을 하도록 학습된다. 즉, 자선은 타선이 폐색영역에 들어오지 않으면서 최대한 폐색영역에 가깝게 거리를 유지하며 돌러가도록 학습한다.

### 4.3. DQN 적용

강화학습에서 탐험(exploration)과 이용(exploit)은 최적의 행동 정책을 찾기 위한 중요 요소이다. 탐험을 충분히 수행하지 않으면 다양한 경험으로부터 최적 행동 정책을 찾기 어렵고 국부적인 최적 해를 찾게 될 가능성이 크다. 이를 해결하기 위해 가장 많이 활용되는 방법이 엡실론-탐욕(epsilon-greedy) 정책이다. 엡실론-탐욕 정책은 처음에는 무작위 행동을 많이 수행함으로써 에이전트가 탐험을 많이 할 수 있도록 하며, 시간이 갈수록 학습된 행동을 선택하는 비중을 높인다. 한 에피소드를 수행할 때마다 엡실론( $\epsilon$ )은 식 (4.15)와 같이 업데이트 된다. 엡실론이  $\epsilon_{\min}$ 보다 크면  $\epsilon_{\text{decay}}$ 만큼 감소하고, 그렇지 않으면  $\epsilon_{\min}$  값이 된다.

$$\epsilon = \begin{cases} \epsilon - \epsilon_{\text{decay}} & \text{if, } \epsilon > \epsilon_{\min} \\ \epsilon_{\min} & \text{otherwise,} \end{cases} \quad (4.15)$$

본 연구에서는 Table 4.2와 같은 선박해양플랜트연구소 공개 선형인 KVLCC 주요 제원을 활용하였으며, 유체력 미계수는 Kijima & Nakiri(2003)의 추정식을 통해 산출된 Table 4.3의 값을 사용한다. 시나리오는 두 선박이 15노트의 속도로 90도 각도에서 교차 조우하는 상황이며 시나리오의 시작 조건은 Table 4.4와 같다. Table 4.1의 DQN 절차를 통해 학습을 수행하며 Table 4.4의 시나리오가 반복된다.

**Table 4.2** Principal dimensions of KVLCC

Item	Value
Length	320 m
Breadth	58 m
Draft	20.8 m
Block Coefficient	0.8101

Table 4.3 Hydrodynamic derivative values

Item	Value	Item	Value
$X'_{uu}$	-0.02050873	$X'_{\beta r}$	-0.032517508
$Y'_{\beta}$	0.36208424	$N'_{\beta}$	0.125635713
$Y'_r$	0.113776326	$N'_r$	-0.0533008836
$Y'_{\beta\beta}$	0.843024015	$N'_{\beta\beta}$	0.0141946375
$Y'_{rr}$	0.0761068165	$N'_{rr}$	-0.0286193416
$Y'_{\beta\beta r}$	-0.362524033	$N'_{\beta\beta r}$	-0.191315025
$Y'_{\beta rr}$	0.423484325	$N'_{\beta rr}$	-0.0610068738

Table 4.4 Initial conditions for encounter scenario

Own ship		Target ship	
$(x_1, y_1)$	(-19200, 0) m	$(x_2, y_2)$	(0, 19200) m
$\psi_1$	0 deg	$\psi_2$	90 deg
$U_1$	7.97 m/s (15.5 knots)	$U_2$	7.97 m/s (15.5 knots)

심층강화학습을 이용해 최적 행동 정책을 찾기 위해서 하이퍼파라미터를 선정하고 무작위 탐색법(Bergstra & Bengio, 2012)을 이용하여 하이퍼파라미터 값을 결정한다. 무작위 탐색법은 결과의 정확도를 살펴가며 하이퍼파라미터 값의 범위를 줄여가는 과정을 반복함으로써 최적 하이퍼파라미터 값을 찾는 방법이다. 선정된 하이퍼파라미터는 감가율(discount factor), 학습률(learning rate), 배치크기(batch size), 은닉층(hidden layer)의 수, 뉴런(neuron) 수, 보상 식 (4.10)의 가중치( $w_r$ )로 두었다.

#### 4.4 심층강화학습을 이용한 충돌회피 결과 및 고찰

보상 식 (4.13)에서 가중치( $w_r$ )에 따라 선박의 자취가 달라지는 경향을 비교해 보았다. 가중치는 목표 경로에서 벗어난 수직 거리와 조타에 의한 속도 감쇄의 비를 조절하기 위한 상수로 거리와 속도 중 어떤 것의 영향력을 중시할 것인지 결정한다. 가중치는 0.2에서부터 0.2 간격으로 0.8까지 모두 4가지 케이스를 테스트하였으며, 각 가중치에 대한 학습 결과는 Fig. 4.4, Fig. 4.5, Fig. 4.6과 같다. Fig. 4.4에서 타선(TS, target ship)의 자취는 파선으로 표시하였으며 자세한 비교를 위해 원점을 중심으로 자취를 확대하였다. 자취는 길이로 무차원화하여 나타내었다.

모든 케이스에서 최대 회피 수직 거리는 약 6L로 비슷한 것(Fig. 4.4)을 알 수 있다. 이는 자선의 폐색영역 안에 타선이 침범하지 않도록 자선의 회피 거동을 최적화하기 때문이다. 가중치 0.8 케이스는 속도에 비해 거리의 영향력을 중시하기 때문에 목표 경로에서 벗어나는 것에 대한 불이익이 크고 속도의 감쇄에 따른 불이익은 상대적으로 작다. 따라서 가중치 0.8 케이스는 4가지 케이스들 중에서 자취, 속도(Fig. 4.5), 선수각(Fig. 4.6) 모두 가장 급격한 변화를 보이며 누적 회피 거리가 가장 작고 속도의 감쇄는 가장 크다. 반면 가중치 0.2 케이스는 거리에 비해 속도의 영향력이 중시되기 때문에 가장 완만한 변화를 나타내고 있다. 누적 회피 거리는 가장 크고 속도의 감쇄는 가장 작게 나타난다. 가중치에 따라 어떤 선박의 거동이 적절한 것인지는 선박을 운항하는 사람마다 다를 수 있기 때문에 확정하기 어렵다. 회피 시작 및 종료 시점, 회피 및 복귀 선수각, 속도 감쇄를 고려하여 적절한 가중치를 선정할 수 있다.

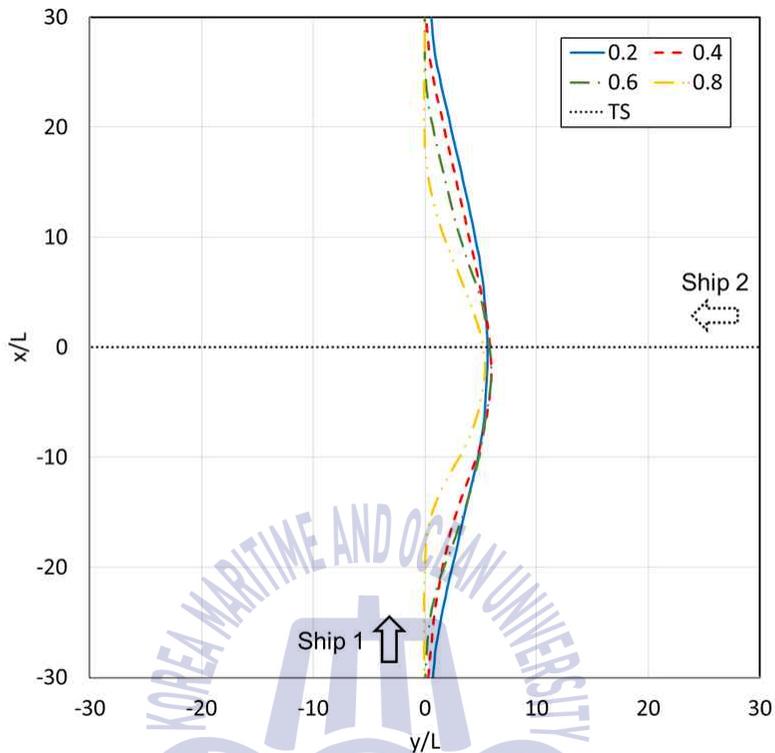


Fig. 4.4 Trajectories for changing weight

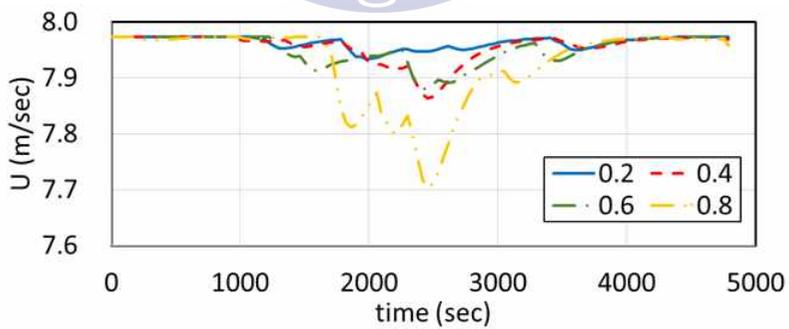


Fig. 4.5 Time history of speed for changing weight

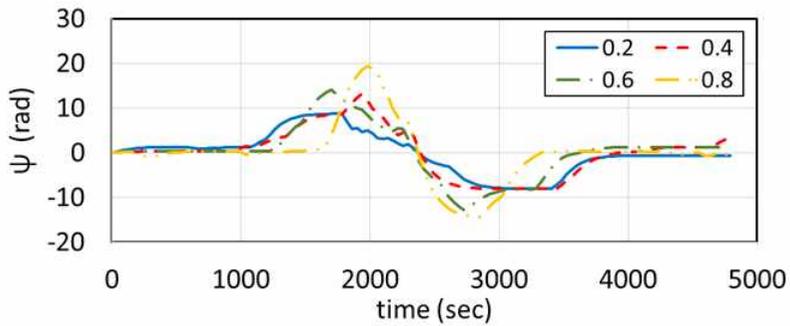


Fig. 4.6 Time history of heading for changing weight

본 연구에서 학습에 사용된 컴퓨터 사양은 CPU Intel i7-7820X, 메모리 32GB, GPU NVIDIA Geforce GTX 970을 사용하였으며 심층강화학습 구현을 위한 라이브러리로 Tensorflow를 이용하였다. 하이퍼파라미터 1세트에 대해 10,000번의 에피소드를 훈련하였을 때 약 12시간이 수행되었으며 하이퍼파라미터 100세트를 수행에 약 50일이 소요되었다. 에피소드는 한 번의 시나리오를 수행하는 것을 의미한다. 90도 교차 조우상황이 반복되는 간단한 시나리오에도 불구하고 많은 시간이 소요되었음을 알 수 있었으며 다양한 시나리오를 반영하게 되면 더 많은 시간이 필요하게 된다. 다양한 시나리오를 반영하고 학습에 소요되는 시간을 줄이기 위해 선박 충돌회피를 위한 행동 선택 문제에서 발생할 수 있는 경우의 수를 줄일 방법이 필요하다.

## 제 5 장 Velocity obstacles(VO) 이론을 활용한 충돌회피

VO 이론은 로봇의 자동 충돌회피와 관련된 연구에 많이 활용되고 있으며 복수의 동적 장애물들의 회피에도 적용할 수 있다. VO 이론에서 정의하고 있는 VO 영역을 피해 이동한다면 로봇이 장애물들에 부딪히지 않고 목표지점까지 이동 가능하다. 또한 RAV(가용회피속도) 영역에서 사용자 정의 가치함수에 따라 이동 방향을 선택한다면 최적 경로를 찾을 수 있다. 선박의 충돌회피 역시 복수의 동적 장애물들을 충돌 없이 회피하는 문제와 동일하기 때문에 VO 이론을 VLCC급 유조선의 충돌회피에 응용하고자 할 때 VLCC급 유조선의 조종 응답 특성에 따른 VO 이론의 특징을 파악한다.

### 5.1 선박의 가용속도

VO는 로봇공학에서 처음 제안된 방법이기 때문에 로봇의 운동학적 특성에 의해 RV(3.4절 참고)가 결정된다. 하지만 VLCC급 유조선은 자체 관성과 유체로부터 받는 저항이 크기 때문에 조타명령에 의한 응답이 로봇이나 자동차 등에 비해 많이 늦기 때문에 3.4절에서 언급한 RV(가용속도) 형상을 가질 수 없다. VLCC급 유조선을 대상으로 VO를 적용하기 위해서는 VLCC급 선박의 조종 응답 특성에 맞는 새로운 RV 형상을 정의하거나 RV 대체 방안을 마련해야 한다.

조타에 따른 선박의 RV 형상을 파악하기 위해 15.5knots 정상 상태에서 타각에 따른 시간 간격( $\Delta t$ ) 이후 속도를 계산한다. 일반적으로 VLCC급 선박은 긴급 상황이 아니면  $\pm 10$ 도 이상 타를 사용하지 않기 때문에 최대 타각은  $-10$ 도에서  $10$ 도로 가정한다. RV의 정의는 Fig. 3.12와 같이 시간 간격( $\Delta t$ ) 이후에 허용 가능한 속도의 변화 범위를 의미한다. 선박이 정상 상태를 유지하였을 때와  $10$ 도,  $-10$ 도로 각각 목표 타각을 주었을 때 시간 간격 2초 이후의 속도 벡터를 그래프로 나타내면 Fig. 5.1과 같다.

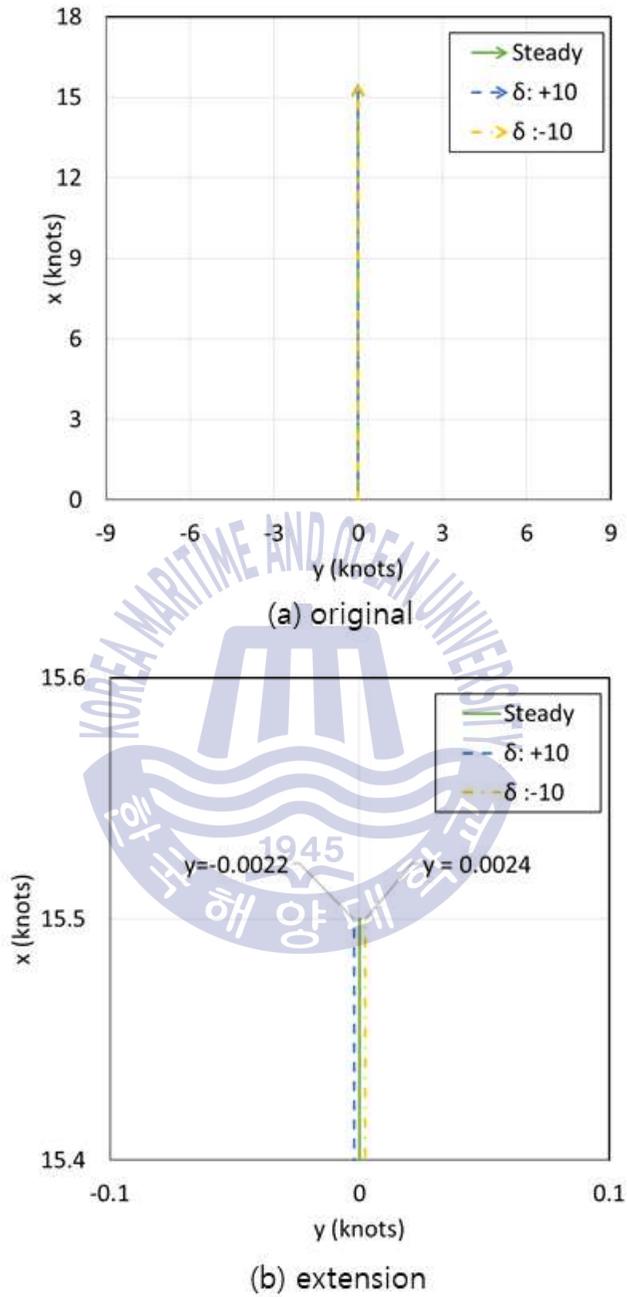


Fig. 5.1 Velocity vector after 2sec

세 가지 상태에 따른 속도 벡터의 차이가 거의 없기 때문에 Fig. 5.1의 (a)는 속도벡터가 겹쳐있다. 속도 벡터의 끝 부분을 확대하여 나타내면 Fig. 5.1의 (b)와 같다. 우현으로 선회하도록 10도 명령타각을 주었으나 왼쪽 파선과 같이 조타 방향의 반대 방향으로 선체가 이탈하는 현상을 관찰할 수 있다. 좌현 선회도 마찬가지로 좌현 변침을 위한 -10도 명령타각을 주었으나 오른쪽 1점 쇄선과 같이 우현 방향의 속도 벡터가 나타난다. 시간 간격 2초에서 선박의 속도 변화는 거의 일어나지 않는다.

시간 간격 40초의 속도를 그래프로 나타내면 Fig. 5.2와 같다. 40초는 4장에서 활용한 KVLCC의 축척된 모델이 선박의 길이( $L$ )만큼 직진 이동할 수 있는 시간이다. 추력 변화에 의한 속도 크기의 변화를 고려하였을 때 Fig. 5.2의 점선과 같은 RV 형상을 예상할 수 있다. 40초 이후 속도 벡터의 변화가 나타나기는 하지만 해당 RV는 Fig. 5.2 (a)와 같이 매우 작은 영역을 형성하고 있기 때문에 VO와의 차집합 영역으로부터 적절한 속도 위치를 선정하기 어렵다.

Fig. 5.3에서는 320m 길이 선박들의 90도 각도 조우상황을 가정하여 VO를 도식화함으로써 RV 영역의 상대적 크기를 확인한다. 자선은 (0,0) 좌표에서 x축 방향의 15.5knots 속도  $V_A$ 로 운항 중이고 타선은 (1600, 1600) 좌표에서 음의 y축 방향의 15.5knots 속도  $V_B$ 로 운항 중이다. 이때, 타선에 대한 VO 영역은  $VO_B$ 와 같으며 앞서 가정한 점선 RV는 모두 VO 영역에 포함된다. 조타에 의한 선박의 속도 변화를 고려하였을 때 RV는 활용이 어렵다.

## 5.2 휴리스틱 접근법을 통한 목표 속도 결정

VO에서 회피 경로를 결정하기 위한 온라인 방법인 휴리스틱 접근법(heuristic search)은 현재 상태에서 가장 휴리스틱에 적합한 목표 속도를 결정하는 것이다. 예를 들면, 목표지점까지 직진 거리를 휴리스틱으로 설정하게 되면, 현재 시점에서 목표 속도( $V_{VO}$ )는 Fig. 5.4와 같이 RV 영역 안에서 목표지점까지 직진 거리 상에 위치하도록 결정한다. Fig. 5.4의 마름모 도형은 RV 영역, 마름모 도형 안에 색칠된 영역은 VO 영역이고 나머지 영역이 RAV 영역이다. 본 논문에서는 다음과 같은 세 가지 휴리스틱 조건을 적용하여 RAV 영역 없이 목표 속

도를 결정할 수 있도록 한다.

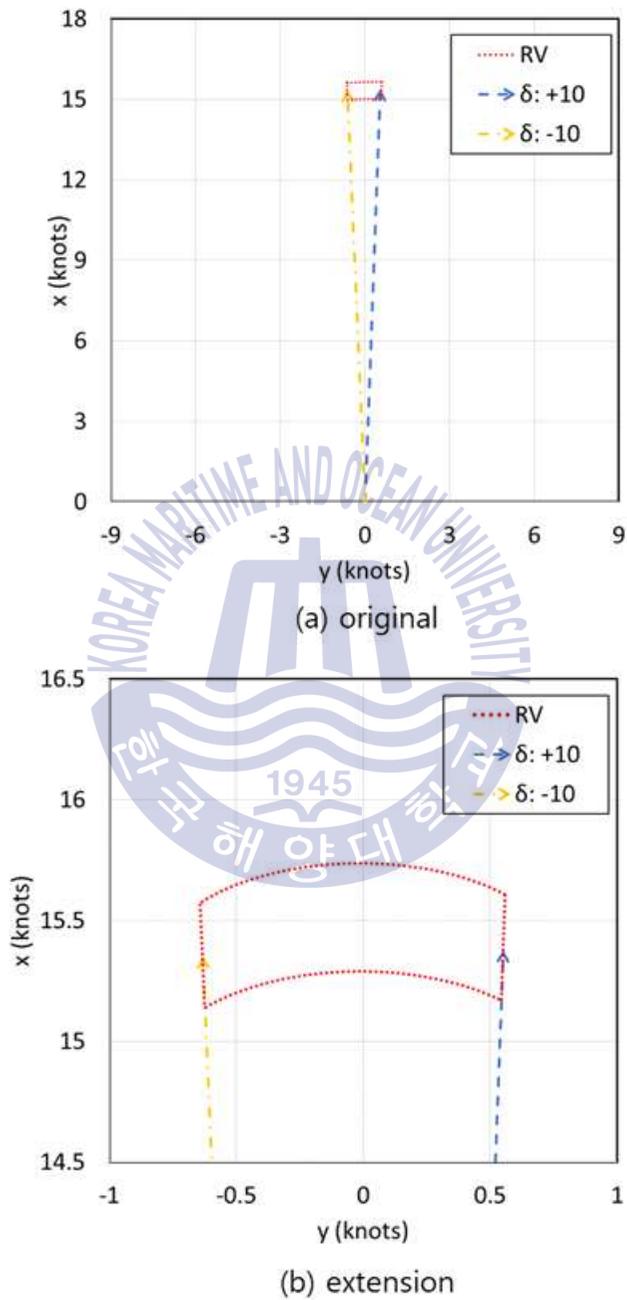


Fig. 5.2 KVLCC velocity vector after 40sec

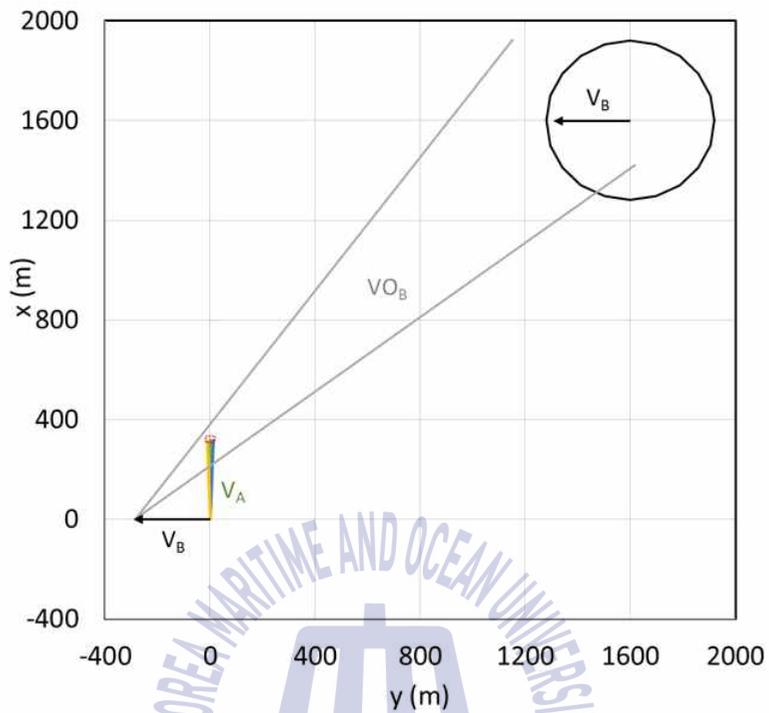


Fig. 5.3 RV in crossing situation ( $\Delta t=40\text{sec}$ )

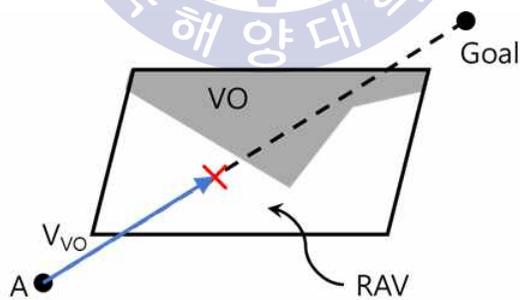
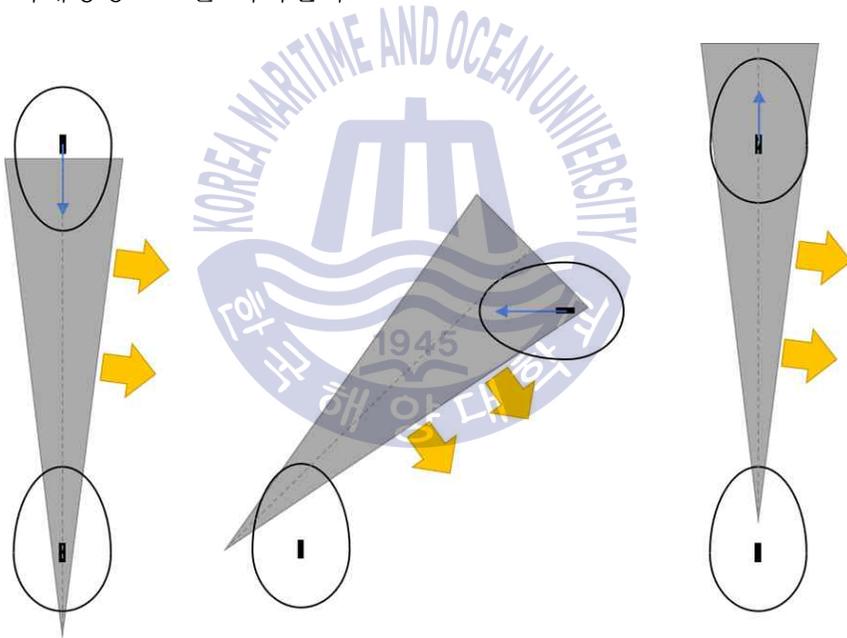


Fig. 5.4 Example of heuristic search

- A. 모든 조우상황에서 우현 변침을 통해 회피한다.
- B. VO 영역의 경계를 따라 이동한다.
- C. 추력의 변화는 없다.

Fig. 5.5는 왼쪽부터 마주침(head-on), 교차(crossing), 추월(overtaking) 조우상황에서 VO 영역(색칠된 영역)과 COLREGs 기반 충돌회피 방향을 화살표로 나타내고 있다. COLREGs에 따르면 추월 상황에서 좌현 변침 추월 또는 우현 변침 추월이 모두 가능하다. 추월 조우상황에서 우현 변침 추월만 가능하도록 제약하면 COLREGs에서 제안하고 있는 세 가지 조우상황에 대한 모든 회피 방향이 동일하게 된다. 즉, Fig. 5.5의 화살표 방향과 같이 VO 영역의 중앙 점선을 기준으로 시계방향으로만 회피한다.



**Fig. 5.5** VO and avoidance direction in each encounter

휴리스틱 조건 A에 조건 B를 더하면 Fig. 5.6의 1점 쇄선과 같이 시계 방향에 위치한 VO 경계선으로 회피 속도 위치 결정 범위를 제한할 수 있다. VO의 경계선을 따라가는 것은 장애물과의 회피 간격을 최소화하는 것을 의미한다. 본 연구에서는 폐색영역을 회피 경계로 두고 있으며 폐색영역 경계선과의 거리를 최소화는 조건이 된다.

최종적으로 휴리스틱 조건 C를 적용하여 속도 벡터를 결정할 수 있다. 추력의 변화 없이 현재 속도를 유지한다고 가정하였을 때, Fig. 5.6의 파선과 같이 자선의 속력을 반경으로 하는 원(circle)이 허용 가능한 속도 벡터들(RV)의 경계가 된다. VO의 시계 방향 경계(Fig. 5.6의 1점 쇄선)와 원의 교차점(Fig. 5.6의 X 지점)을 향하는 속도 벡터가 목표 속도( $V_{VO}$ )가 된다.

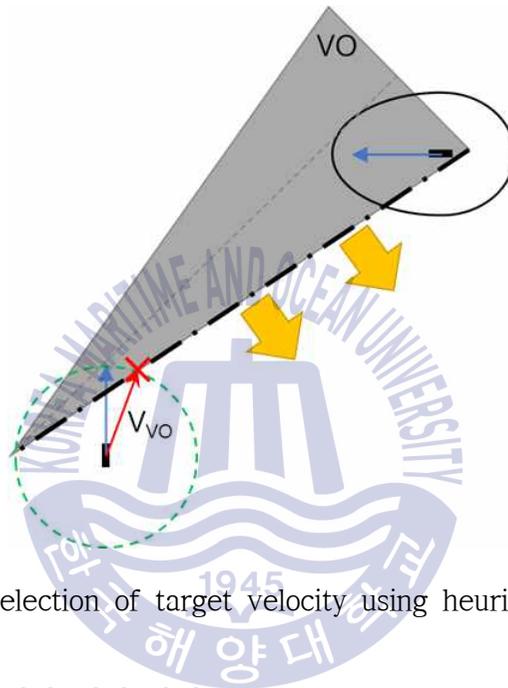


Fig. 5.6 Selection of target velocity using heuristic search

### 5.3 목표 속도 도달을 위한 제어 방법

목표 속도가 정해지면 목표 속도에 도달하기 위한 제어 방법이 필요하다. 특정 시간 후 목표 속도에 도달할 수 있는 타각을 알 수 없기 때문에 제어가 용이하지 않다. 이를 위해 속도 기반의 VO 영역을 위치 기반으로 표현하여 목표 속도를 목표 위치로 변경하고 목표 위치에 도달하기 위한 타각을 데이터베이스화하여 활용하는 방법을 소개한다. 하지만 회피 경로를 이용하여 간단히 목표 속도에 도달할 방법이 있으며 본 연구에서는 회피 경로를 이용한 방법을 활용하였다.

### 5.3.1 회피 경로를 이용한 방법

목표 속도를 향하는 회피 경로(Fig. 5.7의 회피경로)를 새로 지정하고 2.4절의 퍼지를 이용한 경로추종 제어를 이용하면, 목표 속도는 경로추종을 통해 달성된다. VLCC급 선박의 조종 응답 특성상 즉각적인 속도 변경은 불가능하지만 회피 경로에 변화가 없다면 수 초 후에 목표 속도를 만족하게 된다.

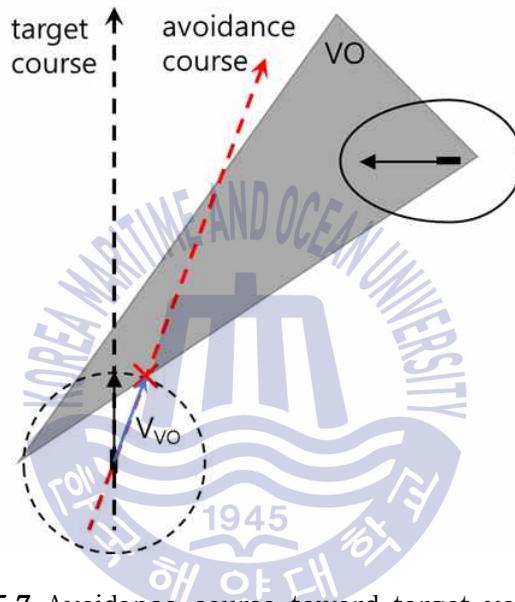


Fig. 5.7 Avoidance course toward target velocity

### 5.3.2 위치 기반 VO

RV를 활용하지 않고 휴리스틱에 의해 속력 벡터를 결정할 수 있어도 VLCC급 선박의 특성 상 정상 상태에서 미소 시간 간격( $\Delta t$ ) 이후 속도의 변화가 작기 때문에 선택된 속도 벡터를 만족시키는 타각을 알기 어렵다. 또한 선박이 선회 중일 때 선택된 속도 벡터에 의해 선회 방향을 변경해야 할 경우 반응은 더 늦어진다.

시간에 따른 선회 특성을 확인하기 위해 x축 방향 15.5knots 정상 상태에서 -10도의 목표 타각을 주었을 때 선박의 자취를 확인해 보았다(Fig. 5.8). 선박은

4장에서 정의한 KVLCC를 활용하였다. 자선의 선수방향이 약 90도가 될 때까지 자취를 확인하였으며, Fig. 5.8의 원형 표식은 원점 (0, 0)에서 시작하여 40초 간격의 선박 위치를 나타낸다. 선박의 활동 범위를 넓게 고려하기 위해서 상당히 오랜 시간 이후의 상태를 예측해야 한다.

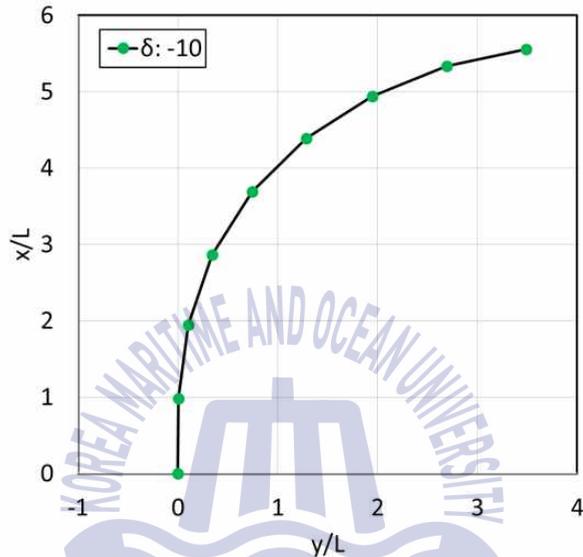


Fig. 5.8 Trajectory by steering angle  $-10^\circ$

VO에서 속도 즉, 단위 시간 당 이동 거리를 기반으로 회피 지점을 찾지 않고 위치를 기반으로 회피 지점을 찾으면 오랜 시간 이후 상태까지 고려할 수 있다. 식 (5.1)은 VO의 정의이며,  $t^*$  이후의 위치를 기반으로 회피 위치를 찾기 위한 VO 즉, PBVO(position based VO)는 식 (5.2)와 같이 변경하였다.

$$VO = CC_{A,B} \oplus \mathbf{v}_B \quad (5.1)$$

$$PBVO = CC_{A,B} \oplus (\mathbf{v}_B \cdot t^*) \quad (5.2)$$

PBVO를 90도 교차 조우상황으로 예로 들면 Fig. 5.9와 같다. 두 선박 모두 15.5knots로 항해하고 있으며 이는 7.97m/s 이다. 현재 속도로 320초를 항해하였을 때 이동 거리는 2526m이며,  $PBVO_B$ 는 식 (5.2)에 의해  $CC_{A,B}$  영역을 타선의 속도 방향으로 2526m 이동( $\mathbf{v}_B \cdot t^*$ )시킨 것과 같다. 또한 자선이 현재 속도로

향해하였을 때 320초 후에 도달하는 위치( $v_A \cdot t^*$ )는  $PBVO_B$  영역 안에 들어있기 때문에 타선과 충돌할 가능성이 존재한다. 이 때문에 자선은 320초 후에  $PBVO_B$  영역을 벗어난 위치로 향하도록 타각을 변경해야 한다.

5.2절에서 언급한 휴리스틱 조건들을 적용하면 목표 회피 지점을 선택할 수 있다. 이때 목표 회피 지점( $P_{PBVO}$ )은 식 (5.3)과 같이 VO영역에서 선택한 목표 회피 속도( $V_{VO}$ )에 시간( $t^*$ )을 곱한 지점과 같다. 현 시점에 회피를 위한 명령 타각은 320초 뒤에 목표 회피 지점( $P_{PBVO}$ )에 도달하게 해 주는 타각이어야 한다.

$$P_{PBVO} = V_{VO} \cdot t^* \quad (5.3)$$

### 회피 지점과 타각

회피 지점을 선정하였지만 회피 지점에 도착하기 위한 타각은 알고 있지 않다. 식 (5.4)와 같이 현재 선박의 상태들과 자선에서 회피 지점( $P_{PBVO}$ )까지의 거리( $x_{PBVO}, y_{PBVO}$ )를 변수로 하여 목표 타각( $\delta^*$ )을 산출하는 추정 식을 세워야 한다.

$$\delta^* = f(\beta, r, U, n_p, \delta, x_{PBVO}, y_{PBVO}) \quad (5.4)$$

모든 변수를 고려한 추정식을 세우는 것은 어렵기 때문에 정상 상태를 가정하여 변수를 간략화한다. 정상 상태를 가정하면 프로펠러 회전수가 독립변수가 되고 선박의 상태들( $\beta, r, U$ )이 종속 변수가 되기 때문에 식 (5.5)와 같이 종속 변수들은 생략할 수 있다.

$$\delta^* = f(n_p, \delta, x_{PBVO}, y_{PBVO}) \quad (5.5)$$

추정 식을 세우기 위해 우선 타각에 따른 선박의 자취를 확인하였다. 타각은 정상 상태에서의 타각( $\delta_{steady}$ )에서  $\pm 10$ 도로 제한 범위를 두었으며, 모두 41개의 타각을 식 (5.6)과 같이 산출하였다. 타각을 구할 때 제곱식을 사용함으로써  $t^*$  이후 선박의 위치가 특정 구간에서만 크게 차이 나는 현상을 줄였다. 15.5knots 정상 상태에서 각 목표 타각에 의한 선박의 자취는 Fig. 5.10과 같다.

선수각이 처음보다  $\pm 90$ 도를 넘어가면 중단하였다.

$$\delta_i = \begin{cases} \delta_{steady} - 10 \times (i/20)^2 & , \text{if } i \leq 20 \\ \delta_{steady} + 10 \times ((i-40)/20)^2 & , \text{if } i > 20 \end{cases} \quad (i = 0, 1, \dots, 40) \quad (5.6)$$

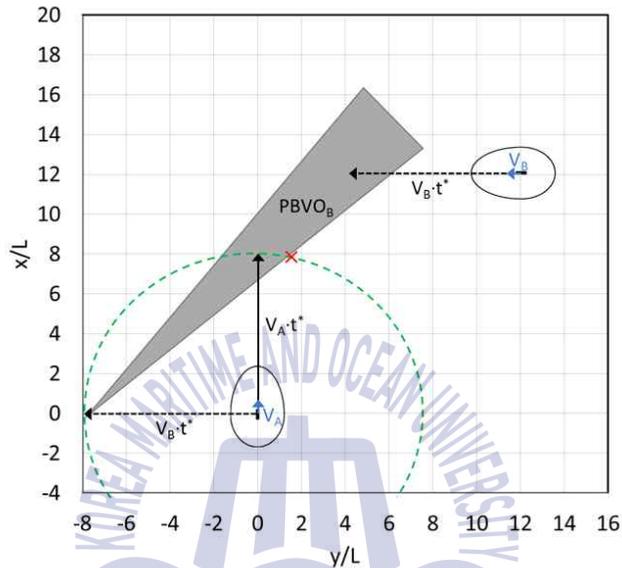


Fig. 5.9 PBVO at crossing situation

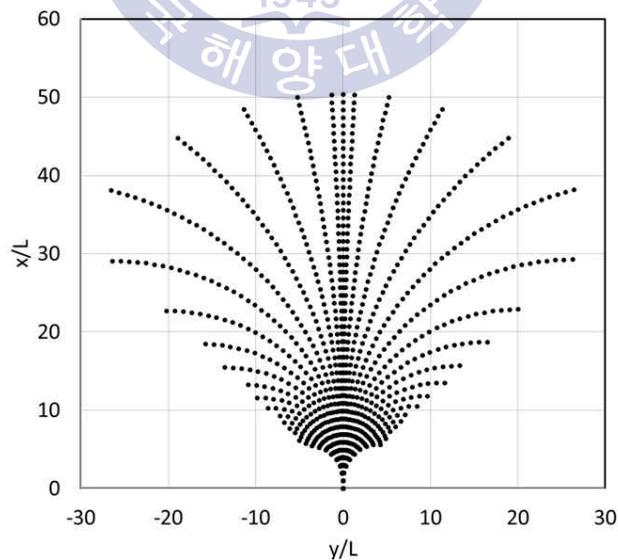


Fig. 5.10 Trajectories by steering angle

Fig. 5.10에서 각 타각에 대한 선박의 자취는 타원의 4분의 1 형태를 보인다. 식 (5.7)과 같은 타원의 방정식으로 회귀식을 세우고 각 타각에 대한 선박의 자취 데이터를 이용하여 곡선 맞춤(curve fitting)을 통해 파라미터 a, b 값을 구한다. 식 (5.7) 타원의 방정식은 원점을 지나고 타원의 중심이 y축 선상에 항상 존재한다.

$$\frac{(y-a)^2}{a^2} + \frac{x^2}{b^2} = 1 \quad (5.7)$$

프로펠러 회전수( $n_p$ ), 현재 타각( $\delta$ ), 목표 타각( $\delta^*$ )에 대한 타원(식 (5.7))의 a, b 파라미터를 데이터베이스화하면 식 (5.8)과 같이 선박의 자취가 예측할 수 있다. 선박의 현재 상태( $n_p, \delta$ )는 정해져 있기 때문에 목표 회피 지점( $P_{PBVO}$ )에 t\* 시간 뒤 도착하게 되는 목표 타각( $\delta^*$ )을 찾는다. Fig. 5.11에서 목표 회피 지점( $P_{PBVO}$ )의 목표 타각  $\delta_i^*$ 로 인한 자취, 즉 타원 내부에 목표 회피 지점이 있는지 여부를 Table 5.1과 같이 순서대로( $i=1,2,\dots,40$ ) 확인함으로써 목표 회피 지점 근처를 지나가는 목표 타각을 찾을 수 있다. 식 (5.7) 타원의 식에 목표 회피 지점 ( $x, y$ )를 입력하였을 때, 목표 회피 지점은 식 (5.7) 좌현의 계산 결과가 1을 초과하면 타원 밖에 위치하고 1 미만이면 타원 안에 존재한다. 이를 통해 목표 회피 지점이 목표 타각  $\delta_i^*, \delta_{i+1}^*$  사이에 존재하는지 여부를 확인할 수 있다. 보다 자세한 타각을 구하기 위해  $\delta_i^*, \delta_{i+1}^*$  사이의 타각을 10등분( $\delta_j^*, \delta_{j+1}^*$ )하고 Table 5.1의 절차를 반복( $j=1,2,\dots,9$ )한다. 최종적으로 목표 회피 지점과 타원과의 거리를 선형 보간 하여 최종 목표 타각을 구할 수 있다.

$$x_{PBVO}, y_{PBVO} = f(n_p, \delta, \delta^*) \quad (5.8)$$

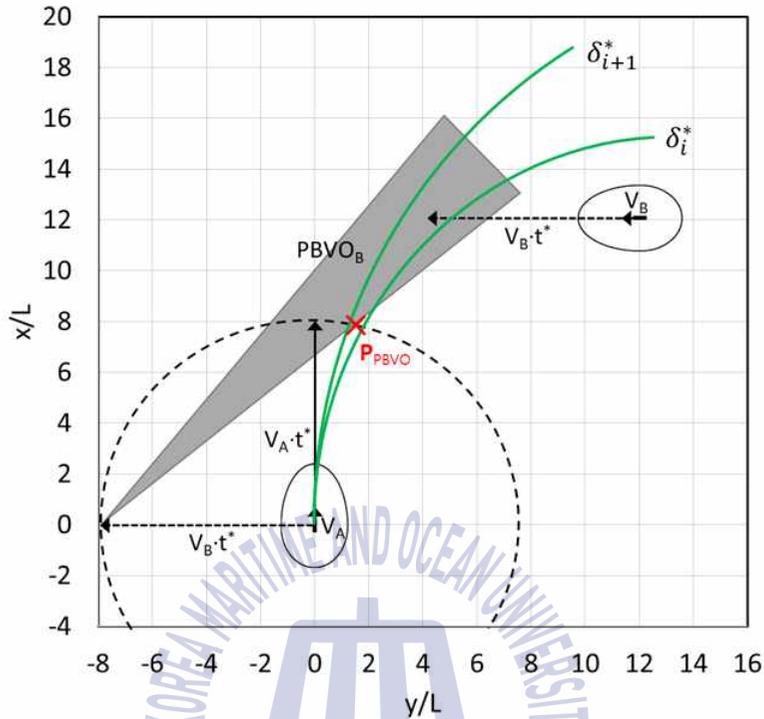


Fig. 5.11 Trajectories by target steering angles

Table 5.1 Procedure for target steering angle search

procedure TARGET STEERING ANGLE SEARCH

$x_{PBVO}, y_{PBVO} = V_{PBVO}$ 의 좌표

for  $i = 1, 40$  do

$a_i, b_i = \text{Get\_parameters\_of\_ellipse}(n_p, \delta, \delta_i^*)$

$a_{i+1}, b_{i+1} = \text{Get\_parameters\_of\_ellipse}(n_p, \delta, \delta_{i+1}^*)$

if  $\frac{(y_{PBVO} - a_i)^2}{a_i^2} + \frac{x_{PBVO}^2}{b_i^2} \geq 1$  and  $\frac{(y_{PBVO} - a_{i+1})^2}{a_{i+1}^2} + \frac{x_{PBVO}^2}{b_{i+1}^2} \leq 1$  then

$(x_{PBVO}, y_{PBVO})$  lies between the ellipses by  $\delta_{i+1}^*$  and  $\delta_i^*$

end if

end for

## 5.4 VO를 이용한 충돌회피 결과 및 고찰

VO를 VLCC급 유조선의 충돌 회피를 위해 적용하였을 때, 그 특징을 확인하기 위해 간단한 테스트를 수행하였다. 테스트에 사용된 선박의 목표 속도 벡터는 아래와 같은 2가지이다.

- ① 항해 속력을 유지하고 목표지점을 향하는 속도 벡터
- ② 5.2절의 휴리스틱 조건에 의한 우현 회피 속도 벡터

목표지점 방향 속도 벡터(①)가 우선순위를 가지며, 해당 속도 벡터가 VO 영역 안에 포함된 경우 우현 회피 속도 벡터(②)를 목표 속도로 한다. 이와 같은 조건으로 마주침, 교차, 추월 조우상황에 대해 각각 VO를 이용한 충돌회피 테스트를 하였다.

### 마주침 조우상황

마주침 조우상황은 180도 각도로 수행하였다. 초기조건은 Table 5.2와 같으며 결과는 Fig. 5.12, Fig. 5.13과 같다. 그림에서 실선은 Ship 1의 자취이며, 점선은 Ship 2의 자취이다. Ship 1과 Ship 2의 거리가 가장 가까울 때 Ship 1의 폐색영역이 표시되어 있으며 이때, Ship 2의 위치는 십자 표시식으로 나타내었다.

마주침 조우상황에서 두 선박 모두 회피의 의무가 있기 때문에 VO를 이용한 회피 동작을 수행한다. 먼 거리에도 불구하고 시작과 동시에 회피 동작을 수행하며, 이는 목표지점 방향 속도 벡터가 VO 영역에 포함되어 있기 때문에 다음 행동으로 우현 회피를 수행하기 때문이다. 타선이 자선의 좌현으로 지나가게 되면 약 3,100초에서 목표지점 방향 속도 벡터는 VO 영역을 벗어나게 되며 선박은 목표지점을 향하게 된다. 목표지점은 목표 경로의 끝점으로 지정되어 있으며, 목표 경로는 초기조건 위치의 2배 거리에서 시작해 원점의 반대편 위치에서 끝나도록 하였다. 즉, Ship 1의 목표 경로는 (-51200, 0)을 시작점으로 하고 (51200, 0)이 끝점이 된다. 약 3,100초에 선박은 목표지점으로 향하기 위해

타를 사용하며 선수각이 목표지점을 향하면 더 이상 타를 사용하지 않는다(Fig. 5.13). 선수각에 의해 선박은 목표경로에 조금씩 가까워진다.

Table 5.2 Initial condition of 180° head-on situation

Ship 1		Ship 2	
$(x_1, y_1)$	(-25600, 0) m	$(x_2, y_2)$	(25600, 0) m
$\psi_1$	0 deg	$\psi_2$	180 deg
$U_1$	7.97 m/s (15.5 knots)	$U_2$	7.97 m/s (15.5 knots)

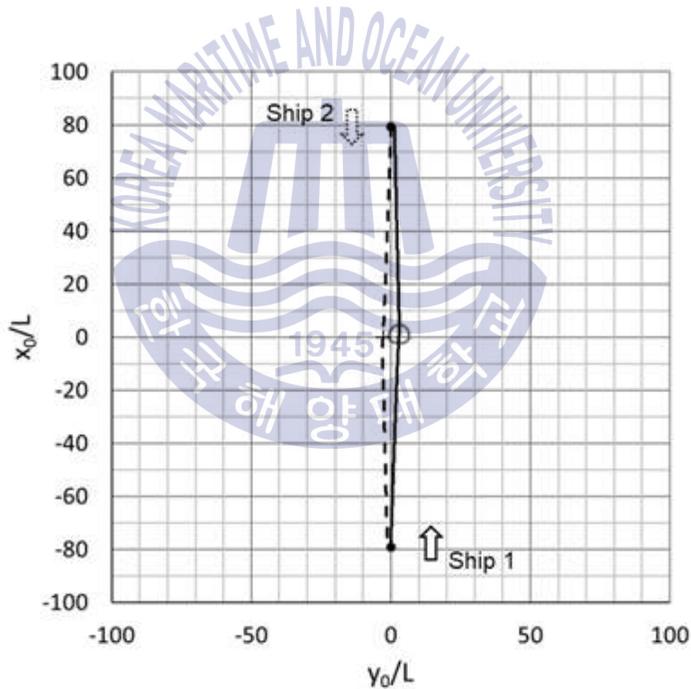


Fig. 5.12 Trajectories of 180° head-on situation using VO

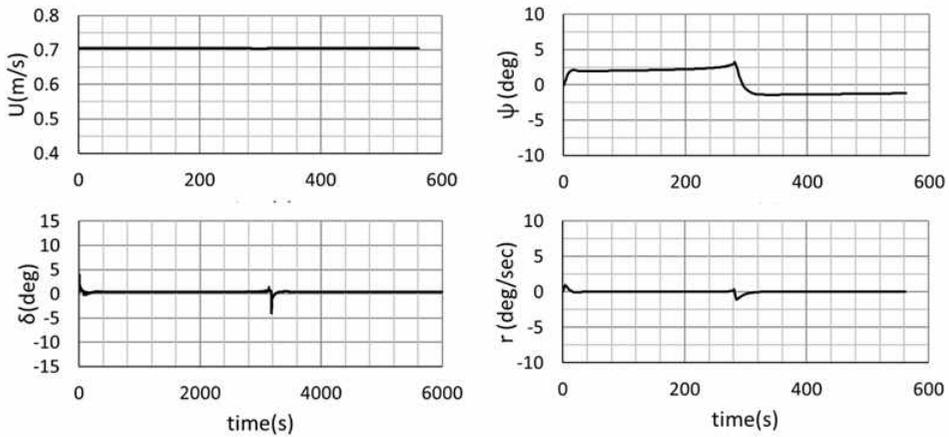


Fig. 5.13 Time histories of 180° head-on situation using VO

### 교차 조우상황

교차 조우상황은 90도 각도로 수행하였다. 초기조건은 Table 5.3과 같으며 결과는 Fig. 5.14, Fig. 5.15와 같다. 교차 조우상황에서 Ship 1은 피항선의 의무를 갖고 Ship 2는 유지선의 의무를 갖기 때문에 VO를 이용한 회피 동작은 Ship 1이 수행하며 Ship 2는 정상상태를 유지한다. 마주침 조우상황과 마찬가지로 목표지점 방향 속도 벡터가 VO영역에 포함되어 있기 때문에 시작과 동시에 회피 동작을 수행한다.

Table 5.3 Initial condition of 90° crossing situation

Ship 1		Ship 2	
$(x_1, y_1)$	$(-25600, 0)$ m	$(x_2, y_2)$	$(0, 25600)$ m
$\psi_1$	0 deg	$\psi_2$	90 deg
$U_1$	7.97 m/s (15.5 knots)	$U_2$	7.97 m/s (15.5 knots)

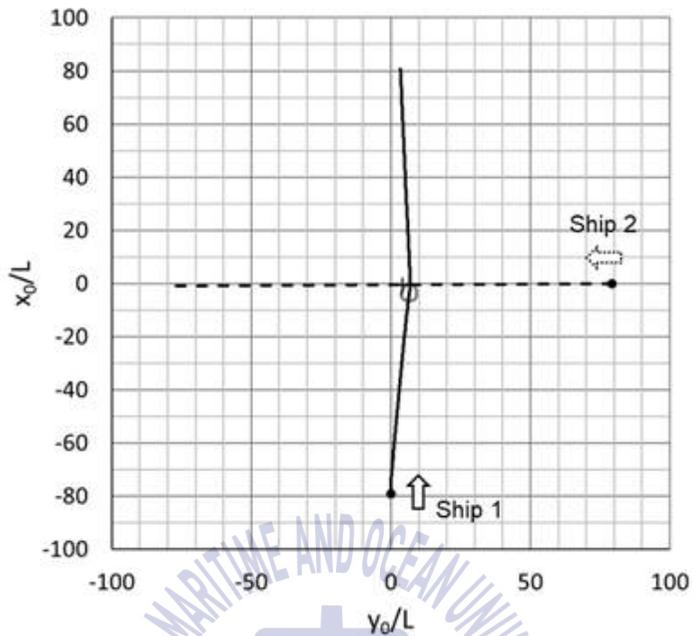


Fig. 5.14 Trajectories of 90° crossing situation using VO

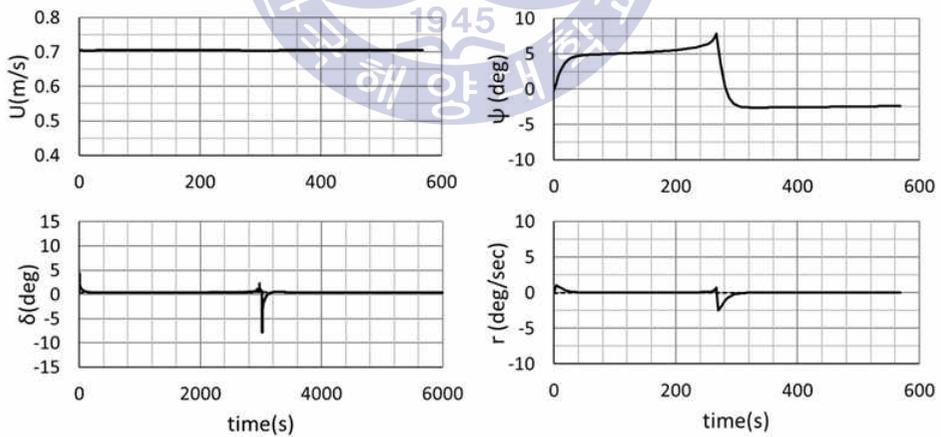


Fig. 5.15 Time histories of 90° crossing situation using VO

마주침 조우상황과 유사한 패턴으로 회피동작이 수행되며, 차이점은 회피 선수각이 크기 때문에 Ship 1의 회피동작이 더 크게 나타난다. 교차 상황은 약 5도 정도의 선수각을 유지하며 회피동작을 수행(Fig. 5.15)하고, 이는 마주침 조우상황에 비해 약 2배 정도 큰 것을 알 수 있다. 마주침 조우상황은 두 선박 모두 회피 동작을 수행하는데 반해 교차 조우상황에서는 피항선만 회피 동작을 수행하기 때문에 2배 정도 차이가 발생한다.

먼 거리에서부터 충돌회피를 수행하면 속력의 큰 감소 없이 안정적인 회피가 가능하다. 두 선박이 가까운 거리에서 피항 행동을 수행했을 때 VO를 이용한 충돌회피 결과를 확인하기 위해 Table 5.4와 같은 조건을 테스트하였다. 결과는 Fig. 5.16, Fig. 5.17과 같다.

충돌회피를 위해 Ship 1은 급격하게 우현 방향으로 회피를 수행하며 Fig. 5.17에서 타각의 변화가 큰 것을 알 수 있다. 타각의 사용을  $\pm 10$ 도로 제한하였기 때문에 10도를 넘어가지 않았으며, 타각의 변화가 심하기 때문에 속도 감쇄가 많이 발생한 것을 알 수 있다. 선수방향을 최대 약 50도까지 틀며 유지선을 회피하고, 회피 후 타각을 반대로 사용하며 목표지점을 향하게 되면서 선수방향 각도가 줄어든다. 먼 거리에서 회피를 수행하면 속도 감쇄와 선수각의 변화가 적고 가까운 거리에서 회피를 수행할수록 속도 감쇄와 선수각의 변화가 커지게 된다. 회피 시작 지점을 조절하여 적절한 회피 방법을 찾는 것이 필요하다.

**Table 5.4** Initial condition of another 90° crossing situation

Ship 1		Ship 2	
$(x_1, y_1)$	(-3840, 0) m	$(x_2, y_2)$	(0, 3840) m
$\psi_1$	0 deg	$\psi_2$	90 deg
$U_1$	7.97 m/s (15.5 knots)	$U_2$	7.97 m/s (15.5 knots)

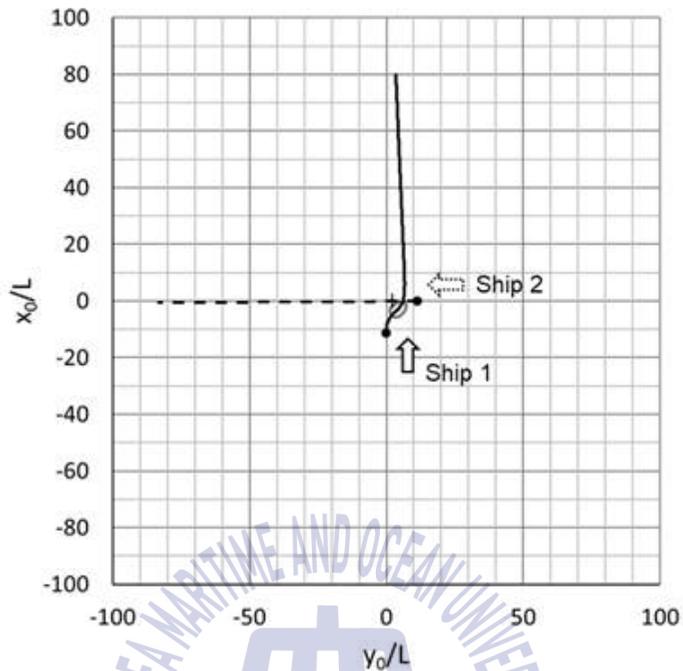


Fig. 5.16 Trajectories of another 90° crossing situation using VO

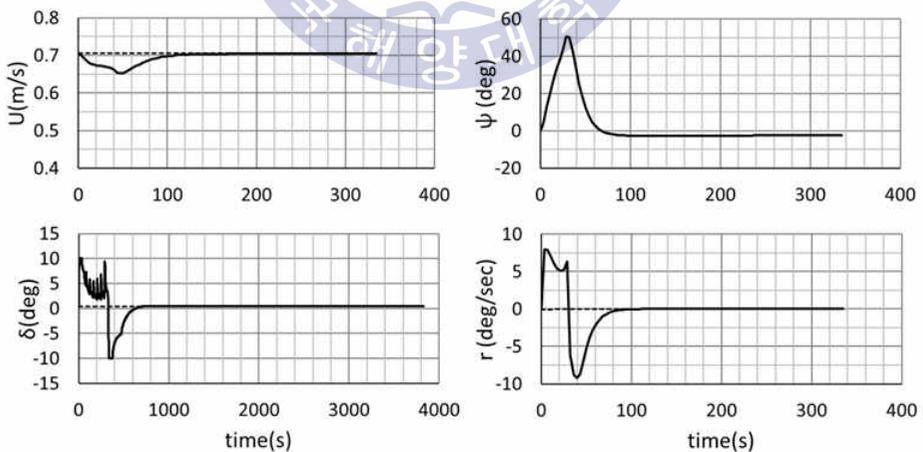


Fig. 5.17 Time histories of another 90° crossing situation using VO

## 추월 조우상황

추월 조우상황은 0도 각도로 수행하였으며 두 선박의 속도 차이는 5knots다. 초기조건은 Table 5.5과 같으며 결과는 Fig. 5.18, Fig. 5.19와 같다. 추월 조우상황에서 Ship 1은 추월선의 의무를 갖고 Ship 2는 유지선의 의무를 갖기 때문에 VO를 이용한 회피 동작은 Ship 1이 수행하며 Ship 2는 정상상태를 유지한다. Fig. 5.18에서 Ship 1은 앞서 다른 조우상황과 마찬가지로 시작과 동시에 회피 동작을 수행한다.

추월 조우상황에서는 Fig. 18과 같이 약 5도의 선수각으로 추월 행동을 한다. 특이점은 약 3,000초에서 목표지점을 향하기 위해 선수각을 틀지만 곧 다시 회피 동작을 수행하며 약 3,200초부터 또 다시 목표지점을 향한다. 이는 약 3,000초에 추월선의 목표지점 방향 속도 벡터가 VO 영역을 벗어나서 목표지점을 향하지만 다시 목표지점 방향 속도 벡터가 VO 영역에 들어가게 됨으로써 우현 회피 동작을 수행하는 것이다.

두 선박이 가장 가까워지는 지점의 시간은 3,000초이다. Fig. 5.18은 두 선박이 가장 가까워지는 지점에서 추월선이 유지선을 완전히 추월하지 않았음을 보여주고 있다. 추월선이 유지선을 완전히 추월하기도 전에 목표지점 방향 속도 벡터는 VO 영역을 벗어나기 때문에 추월선이 유지선 쪽으로 향하면서 다시 VO 영역에 들어가게 된다.

Table 5.5 Initial condition of 0° overtaking situation

Ship 1		Ship 2	
$(x_1, y_1)$	(-25600, 0) m	$(x_2, y_2)$	(-17280, 0) m
$\psi_1$	0 deg	$\psi_2$	0 deg
$U_1$	7.97 m/s (15.5 knots)	$U_2$	5.4 m/s (10.5 knots)

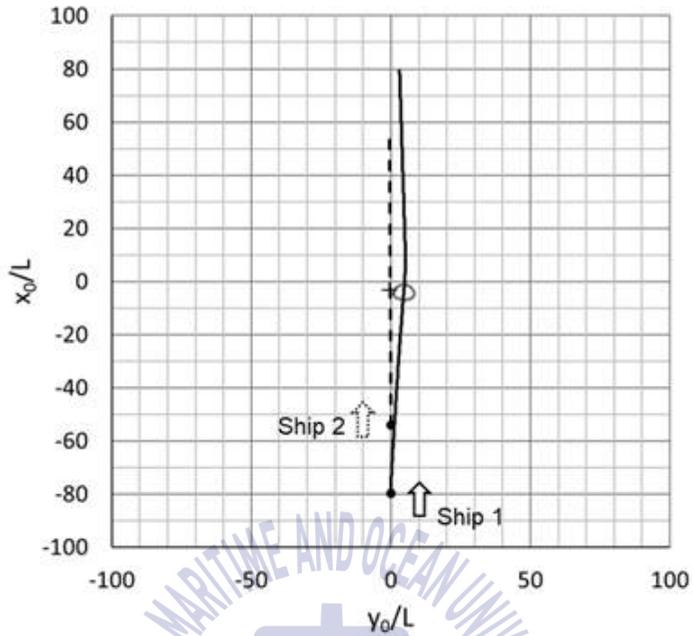


Fig. 5.18 Trajectories of 0° overtaking situation using VO

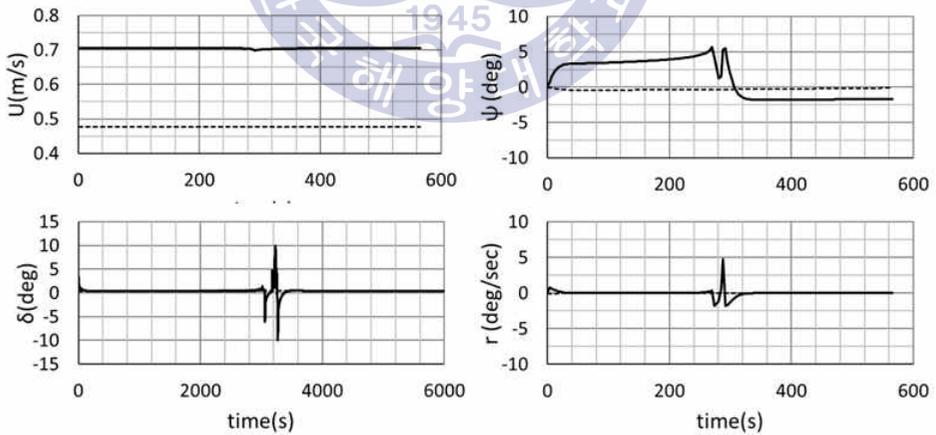


Fig. 5.19 Time histories of 0° overtaking situation using VO

VLCC급 유조선을 대상으로 충돌회피를 위해 VO를 적용하였을 때 선박의 운동 특성으로 인해 RV 정의가 어려움을 확인하였다. RV를 이용하지 않고 목표 회피 속도를 찾기 위해 휴리스틱 조건을 세웠으며 이를 통해 우현 회피가 가능하였다. 휴리스틱 조건과 함께 목표지점을 향하는 행동을 추가하여 마주침, 교차, 추월 조우상황에 대해 테스트 하였다. 테스트에서 나타난 VO의 특징은 VO를 적용함과 동시에 자선은 회피를 시작하기 때문에 회피 시작 시점의 선정이 필요하다는 것이다. 이를 해결하기 위한 방법으로 Fiorini(1998)는 회피 시작 시간을 임계값으로 두었으며 Kuwata(2014)는 TCPA와 DCPA 임계값을 회피 시작 시점으로 사용하였다. 하지만 이와 같은 방법들은 움직이는 물체마다 운동 특성이 다르기 때문에 적절한 임계값을 찾아야 하는 어려움이 있다. 앞선 테스트에서 나타난 또 다른 특징은 추월 상황에서 추월선이 유지선을 완전히 추월하기 전에 목표지점 방향 속도 벡터가 VO영역을 벗어나게 되는 것이다. 그러므로 추월하기 전에 목표 경로로 복귀하려는 위험한 움직임을 취하게 되며 추월이 완전히 이루어진 상태에서 목표 경로로 복귀하기 위한 회피 종료 시점을 선정하는 것이 필요하다. 이와 같은 특징을 통해 각 조우상황에서 최적 회피 시작 시점과 종료시점을 결정해야 할 필요성을 확인하였다. 하지만 각 조우상황에 따라 최적 임계값이 다르기 때문에 임계값을 고정하는 것은 좋은 방법이 아니다.

## 제 6 장 VO와 심층강화학습 기반 충돌회피

앞서 4장에서는 심층강화학습을 이용한 충돌회피 방법을 선박의 90도 조우상황에서 테스트하였으며, 5장에서는 VO를 이용한 VLCC급 유조선의 조종 응답 특성을 확인하였다. 심층강화학습을 이용할 경우 다양한 충돌회피 패턴으로부터 최적 행동 정책을 학습하기 위해 많은 시간이 필요하며, 하이퍼파라미터(hyper-parameter) 최적화를 위한 시간까지 고려하면 소요시간이 기하급수적으로 늘어난다. VO는 VLCC급 유조선에 대한  $RV$ (reachable velocity)를 정의하기 어렵고 VO를 적용함과 동시에 거리가 멀어도 회피 동작을 수행하는 특징이 있다. 6장에서는 VO와 심층강화학습을 함께 활용함으로써 두 방법의 단점을 극복하고자 한다.

일반적으로 충돌회피 알고리즘은 탐지(detection), 의사 결정(decision making), 경로 계획(path planning), 제어(control)의 단계를 거쳐 수행된다(우주현, 2018). 본 연구에서는 VO를 기반으로 회피 행동들을 정의하고 심층강화학습을 통해 최적 행동 정책을 찾음으로써 의사 결정 단계와 경로 계획 단계를 대체하는 모듈을 제안한다. VO와 심층강화학습 기반 충돌회피 방법은 탐지, 의사 결정, 행동, 제어의 단계를 거치며 실시간으로 모든 단계가 반복된다.

본 논문에서 제안하는 선박의 충돌회피 방법의 개략적인 구성은 Fig. 6.1과 같다. 가장 먼저 탐지 단계(①)에서 자선 주변의 장애물 정보가 선박 자동식별장치(AIS, Automation Information System), 레이더 등으로부터 획득된다. 탐지된 장애물의 COLREGs에 근거한 조우상황 유형은 3.2.2절에서 제안한 조우상황 결정표에 의해 판단한다. 탐지된 장애물 정보는 MDP으로 표현된 상태 집합으로 나타낼 수 있으며, 상태 집합은 다음 단계의 입력이 된다.

자선의 감시영역(watching area)을 타선의 감시영역이 침범하면 의사 결정 단계(②)가 된다. 이때 탐지 중이던 조우상황 유형은 감시 영역을 벗어날 때까지 변경되지 않는다. 의사 결정은 회피 행동 변환 모듈이 하며, 이 모듈은 심층강화학습을 통해 결정된 최적 행동 정책에 의해 회피 행동을 판단한다. 행동 단

계(③)에서 선택된 행동에 따라 VO를 기반으로 회피 경로와 목표 추력을 산출한다. 해당 정보는 선박 조종운동 제어를 위한 제어 단계(④)의 입력이 된다.

VO는 회피 방향이 구분된 행동 집합들을 제공하기 때문에 COLREGs에 근거하여 회피 방향을 선택할 때 활용하기가 용이하다(Kuwata et al., 2014; 우주현, 2018). 본 논문에서 회피 방향의 구분은 VO를 기반으로 하였으며, 목표 회피 속도는 5.2절의 휴리스틱(heuristic) 탐색 방법을 활용한다. 휴리스틱 조건은 자선과 타선의 폐색영역(blocking area)을 침범하지 않는 범위에서 타선과의 거리를 최소화하는 것을 기준으로 한다.

심층강화학습은 회피 행동 선택 모듈로써 자선과 타선의 현재 상태에서 어떤 행동이 최적 행동인지 판단한다. 본 논문에서 선택할 수 있는 행동은 목표 경로추종, 우현 회피, 좌현 회피, 목표지점 추종으로 모두 4종류다. 반복 학습을 통해 최적 행동 정책을 찾고 이를 기반으로 현재 상태에서 최적 행동을 선택하게 된다.

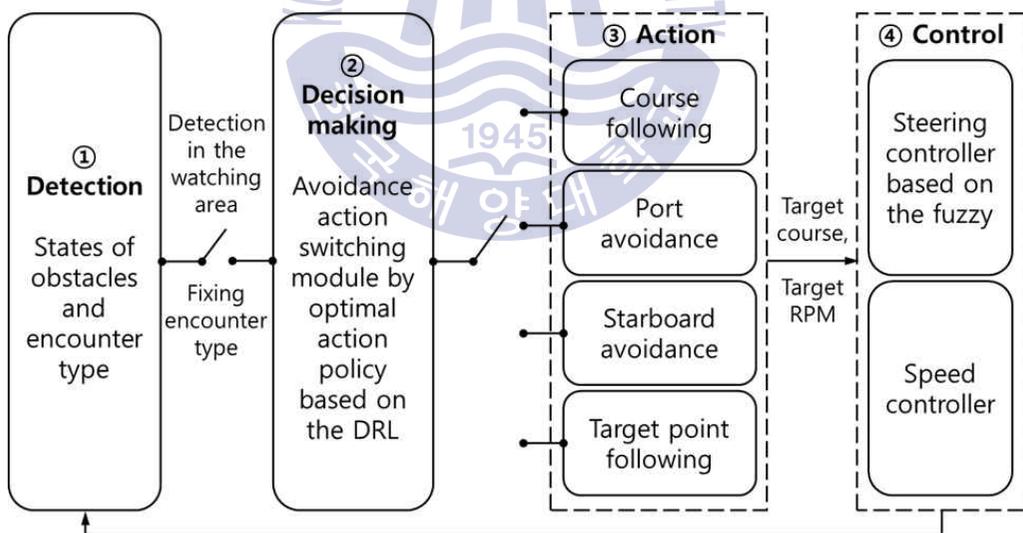


Fig. 6.1 Schematic diagram of collision avoidance method using VO and deep reinforcement learning

## 6.1 탐지 단계

탐지 단계에서는 MDP(마르코프 의사 결정) 상태 집합 생성과 조우상황 유형 결정의 2가지 작업을 수행하며 절차는 Table 6.1과 같다. AIS 및 레이더, 감지기 등으로부터 자선과 타선의 현재 상태에 대한 정보를 MDP 상태 집합 형태로 생성한다. MDP 상태 집합은 다음 의사 결정 단계의 입력이 된다. 의사 결정 단계로의 진행 조건은 자선의 감시영역과 타선의 감시영역이 중첩될 경우이며, 그때까지 자선은 목표 경로를 추종한다. 조우상황 유형 판단은 3.2.2절의 조우상황 결정표를 이용하여 두 감시영역이 중첩될 때까지 반복하며, 중첩될 경우 조우상황 유형은 서로의 감시영역을 벗어날 때까지 달라지지 않는다. Table 6.1에서 WA는 감시영역, BA는 폐색영역, OS는 자선, TS는 타선, HO는 마주침 선박, OT 추월선, GW는 피항선을 의미한다.

### 6.1.1 회피 행동 영역

본 연구의 충돌회피 방법은 자선과 타선의 거리에 따라 단계적으로 달라지며 단계를 구분하는 거리의 기준은 3.3.1절 Kijima의 안전영역 모델에 근거하였다. Kijima가 제안한 폐색영역과 감시영역의 중첩 여부에 따라 Fig. 6.2와 같이 4단계로 회피 행동 영역을 구분한다. Fig. 6.2에서 실선은 폐색영역을 나타내고 점선은 감시영역을 나타낸다.

첫 번째 단계는 두 선박이 서로의 감시영역을 간섭하지 않고 운항하는 경우(㉠)로 조우상황 유형을 판단하는 단계이다. 두 번째는 두 선박의 감시영역이 중첩된 경우(㉡)로 회피 행동 변환 모듈에 의해 회피 여부를 결정하는 단계이다. 세 번째는 두 선박의 폐색영역이 중첩된 경우(㉢)로 COLREGs에 의해 유지선의 의무를 갖는 선박도 회피 행동을 하는 단계이다. 두 선박의 폐색영역이 중첩된 경우는 대부분 회피의 의무를 갖는 선박이 피항하지 않거나 판단 실수로 회피 방향이 잘못된 경우다. 네 번째는 자선이 타선의 폐색영역을 침범한 경우(㉣)로 두 선박이 충돌한 것으로 본다.

Table 6.1 Procedure of detection step

<pre> <b>procedure</b> DETECTION STEP   상태 집합 결정(<math>S</math>)   회피 행동 영역 확인(<math>step</math>)   <b>if</b> 타선 감시영역(<math>WA_{TS}</math>)과 자선 감시영역(<math>WA_{OS}</math>)이 겹치는 경우 <b>then</b>     <b>if</b> HO, OT, GW 조우상황 유형 선박(<math>encounter</math>)일 경우 <b>then</b>       의사 결정 단계 실행     <b>end if</b>     <b>if</b> 타선 폐색영역(<math>BA_{TS}</math>)과 자선 폐색영역(<math>BA_{OS}</math>)이 겹치는 경우 <b>then</b>       의사 결정 단계 실행       <b>if</b> 자선이 타선 폐색영역(<math>BA_{TS}</math>)을 침범한 경우 <b>then</b>         충돌       <b>end if</b>     <b>end if</b>   <b>else</b>     조우상황 결정표로부터 선박의 조우상황 유형 결정(<math>encounter</math>)     목표 경로 추종 행동(<math>a_0</math>)   <b>end if</b> </pre>
--

일반적인 선박 운항 시, 여러 단계의 피항동작을 취해야 하는 거리는 일률적으로 정할 수 없다. 대양에서 교차 조우상황의 경우 유지선은 경로를 유지하고 피항선은 회피 동작을 하는 단계(㉞)의 거리를 5~8마일로 보고, 피항선이 회피 동작을 하지 않을 경우 유지선도 선택적으로 피항 동작을 하는 단계(㉟)의 거리를 2~3마일로 본다는 주장이 있다(윤점동, 2015). 윤점동(2015)은 ㉞ 단계의 동작거리 5~8마일은 약간 큰 감이 있으며 ㉟ 단계의 동작거리 2~3마일은 대단히 타당성 있다 판단하였다.

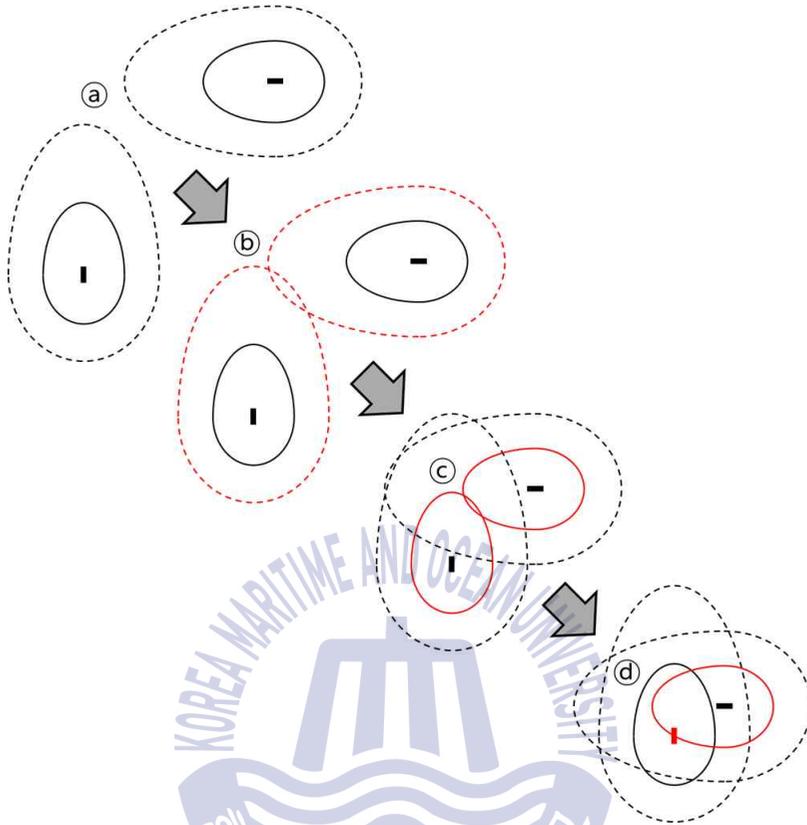


Fig. 6.2 Steps for collision avoidance action

두 선박 모두 길이 320m, 폭 58m에 15.5knots 속력의 90도 교차 조우상황일 경우, 각 선박 폐색영역의 종방향 파라미터( $R_{bf}$ ) 값은 약 0.83마일이고 감시영역의 종방향 파라미터( $R_{wf}$ ) 값은 약 1.47 마일이다. 두 선박의 폐색영역과 감시영역이 겹치기 시작하는 거리를 고려하면 파라미터 값의 2배가 되기 때문에 ㉓ 단계의 동작거리는 1.66마일, ㉔ 단계의 동작거리는 2.94마일이다. Kijima의 안전영역 모델을 이용할 경우, ㉓ 단계의 동작거리는 통상적인 동작거리 2~3마일에 비해 약간 작은 편이며 ㉔ 단계의 동작거리는 통상적인 동작거리 5~8마일이 약간 크다는 것을 고려하더라도 많이 작은 편임을 알 수 있다.

본 연구에서는 심층강화학습을 통해 최적 회피 시점을 찾기 위하여 감시영역의 범위를 식 (6.1)과 같이 확장하였으며, 이는 기존 감시영역(식 (3.2))의 약 2.5

배다.

$$\begin{aligned}R_{wf} &= L + 5(R_{bf} - L) \\ R_{wa} &= L + 5(R_{ba} - L) \\ S_w &= B + 5(S_b - B)\end{aligned}\tag{6.1}$$

### 6.1.2 상태

심층강화학습의 적용을 위해 선박의 조우상황에 대한 문제를 MDP로 나타내야 하며 MDP의 구성요소 중 상태집합은 식 (6.2)와 같다(Fig. 6.3). VO와 심층강화학습을 이용한 충돌회피 문제에서 반드시 필요한 상태 요소들만 정의하였기 때문에 상태 집합이 3장에서 정의한 상태 집합보다 간략해졌다.

$$S = \{l, \psi_s, U, p_{xr}, p_{yr}, v_{xr}, v_{yr}\}\tag{6.2}$$

여기에서  $U$ 는 자선 속력이고,  $p_{xr}, p_{yr}$ 은 자선에서 타선까지 상대거리이며  $v_{xr}, v_{yr}$ 은 자선에서 타선의 상대속도( $\mathbf{v}_{TS} - \mathbf{v}_{OS}$ )다. 집합 요소 중,  $l, \psi_s, U$ 는 자선의 상태를 나타내고  $p_{xr}, p_{yr}, v_{xr}, v_{yr}$ 은 자선과 타선 사이의 관계를 나타낸다. 상태 집합은 주기적으로 갱신되며 의사 결정 단계의 입력 데이터가 된다.

### 6.2 의사결정 단계

의사결정 단계는 선박 조우상황에 대한 상태 데이터를 입력받아 회피 행동을 판단하고 선택하게 된다. 이 단계에서 회피 행동 변환기 모듈은 현재 상태에서 최적의 행동을 선택하며 심층강화학습을 통해 찾은 최적 행동 정책을 기반으로 선택한다. 회피 행동 변환기 모듈을 활용하기 위해서는 사전 학습을 통해 최적 행동 정책을 찾아야 한다.

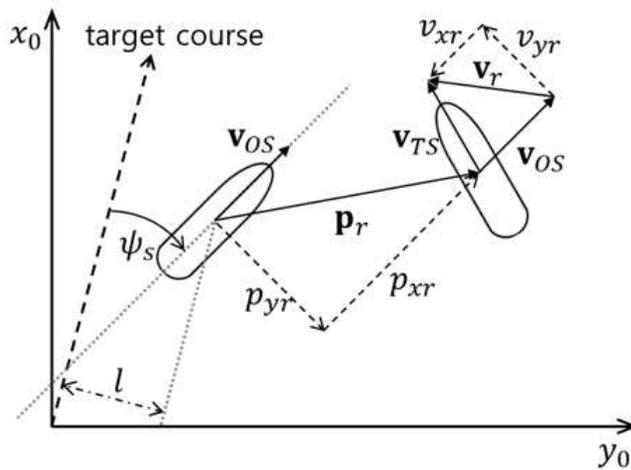


Fig. 6.3 State of agent for detection step

### 6.2.1 보상

강화학습은 현재와 미래에 받게 될 보상의 합이 최대가 될 수 있는 행동을 학습하기 때문에 MDP의 구성요소 중 보상은 최적 행동 정책을 학습하는데 중요한 요소다. 보상은 모두 음(negative)의 값을 가지며 아래와 같이 크게 4가지로 구분된다.

- 목표 경로 이탈 거리와 속도 감쇄에 대한 음의 보상
- COLREGs의 규정을 따르지 않는 행동에 대한 음의 보상
- 두 선박의 폐색영역이 중첩되는 상태(© 단계)에 대한 음의 보상
- 충돌(㉠ 단계)에 대한 음의 보상

첫 번째, 목표 경로 이탈 거리와 속도 감쇄에 대한 보상( $R_{control}$ )은 식 (6.3)과 같다. 회피 시, 목표 경로에 가능한 가깝고 속도 감쇄가 적은 방향으로 운동하도록 한다. 해당 보상은 4장에서 정의한 4.2.3절의 보상과 유사하며 4.3.5절의 가중치 비교 결과를 바탕으로 거리( $l$ )와 속도( $U$ ) 항의 비를 동일하게 선정하여 가중치는 생략되었다.

$$R_{control} = -|l| - |U_{init} - U| \cdot \Delta t \quad (6.3)$$

두 번째, COLREGs의 규정을 따르지 않는 행동에 대한 보상( $R_{COLREGs}$ )은 식 (6.4)와 같다. 여기서  $p_{COLREGs}$ 는 COLREGs 규정을 위반하였을 때 받게 되는 벌칙 상수로  $2^2$  값을 사용하였다.  $c_{COLREGs}$ 는 COLREGs 조건 변수로 COLREGs 규정의 위반 여부를 나타내며, 0 또는 1 값을 가진다.

$$R_{COLREGs} = -p_{COLREGs} \cdot c_{COLREGs} \quad (6.4)$$

COLREGs 조건 변수가 1의 값이 되는 조건은 COLREGs에 근거하여 현재 조우상황과 행동에 따라 달라진다. 마주침(head-on) 조우상황에서 두 마주침선(HO)과 교차(crossing) 조우상황에서 피항선(GW)은 모두 COLREGs에서 우현 방향으로만 회피할 것을 명시하고 있다. 또한 유지선(SO)은 침로 및 속력을 유지할 것을 명시하고 있기 때문에 두 선박의 폐색영역이 겹치기 전까지 회피행동을 하지 않는다. COLREGs 조건 변수가 1이 되는 COLREGs 위반 조건은 마주침선과 피항선이 좌현 회피 행동을 하였을 경우와 유지선이 회피 거리 단계 © 이전에 우현 또는 좌현 회피 행동을 하였을 경우다. 추월선의 경우, COLREGs에서 우현 또는 좌현 회피를 허용하고 있기 때문에 제외되었다.

세 번째, 두 선박의 폐색영역이 중첩되는 상태에 대한 보상( $R_{BA}$ )은 식 (6.5)와 같다. 여기서  $p_{BA}$ 는 두 선박의 폐색영역이 중첩되었을 때(© 단계) 받게 되는 벌칙 상수로  $2^4$  값을 사용하였다.  $c_{BA}$ 는 폐색영역 조건 변수로 두 폐색영역의 중첩 여부를 나타내며, 0 또는 1 값을 가진다. 마주침선, 피항선, 추월선의 경우 회피의 의무를 갖기 때문에 폐색영역이 중첩되면 조건 변수가 1이 되지만, 유지선의 경우 회피의 의무가 없기 때문에 폐색영역이 중첩되더라도 조건 변수가 1이 되지 않는다.

$$R_{BA} = -p_{BA} \cdot c_{BA} \quad (6.5)$$

네 번째, 충돌에 대한 보상( $R_{collision}$ )은 식 (6.6)와 같다. 여기서  $p_{collision}$ 는 자선이 타선의 폐색영역을 침범하였을 때(㉔ 단계) 받게 되는 벌칙 상수로  $2^6$  값을

사용하였다.  $c_{collision}$  는 충돌 조건 변수로 충돌 여부를 나타내며, 0 또는 1 값을 가진다. 마주침선, 피항선, 추월선의 경우 회피의 의무를 갖기 때문에(COLREGs Rule 17 (d)) 충돌 시 충돌 조건 변수가 1이 되지만, 유지선의 경우 ㉠ 동작거리 단계에서 회피 행동을 하더라도 근본적인 회피의 의무는 없기 때문에 충돌 조건 변수는 0이다.

$$R_{collision} = -p_{collision} \cdot c_{collision} \quad (6.6)$$

## 6.2.2 심층강화학습의 신경망 학습

VO와 심층강화학습을 이용한 충돌회피 방법에서 사용되는 심층강화학습 방법은 DQN이다. DQN의 경험적 리플레이 방법을 적용한 알고리즘은 Table 4.1과 같다. 심층강화학습의 신경망 학습에서 중요 이슈는 다음과 같다.

- 시작 조건
- 탐험(exploration)-활용(exploitation) 문제
- 하이퍼파라미터(hyper-parameter) 최적화

학습 시작 조건은 충돌회피 방법에서 얼마나 다양한 충돌 상태를 학습할 수 있는지 여부를 결정하는 중요 조건이다. 학습 시 다양한 충돌 조우상황을 경험하지 못하면 에이전트가 학습 때 경험하지 못한 상태에 처했을 경우 잘못된 행동을 선택할 수 있다. 6.1.2절에서 정의한 상태 집합의 경우 단일선박에 대한 조우상태로 한정하였으며, 조우상황에 대한 상태를 파라미터화하여 나타내었다. 다수선박의 조우상황을 상태집합으로 표현할 경우 단일선박 조우상황에 비해 매우 다양한 패턴들이 존재하기 때문에 모든 경우의 수를 고려한 시작조건을 만들기 어려울 뿐만 아니라 학습시간도 오래 걸리게 된다.

선박은 목표 경로를 설정하고 그 경로를 추종하며 향해가기 때문에 목표 경로의 시작 조건이 다양할 필요가 있으며 이를 위해 이마즈(Imazu) 문제를 활용하였다. 이마즈 문제는 도쿄 대학의 이마즈가 실선 데이터를 바탕으로 발생 가능한 조우상황을 Fig. 6.4와 같이 요약한 것이다. 이마즈 문제는 단일선박 조우

상황 뿐 아니라 다수선택 조우상황까지 포함되어 있으며 충돌 회피 시스템 평가에 사용되어지고 있다(Cai & Hasegawa, 2013). 시작 조건은 이마즈 문제에서 요약한 케이스와 각 선택의 조우 각도를 간섭되지 않는 선에서 무작위로 선정하였다.

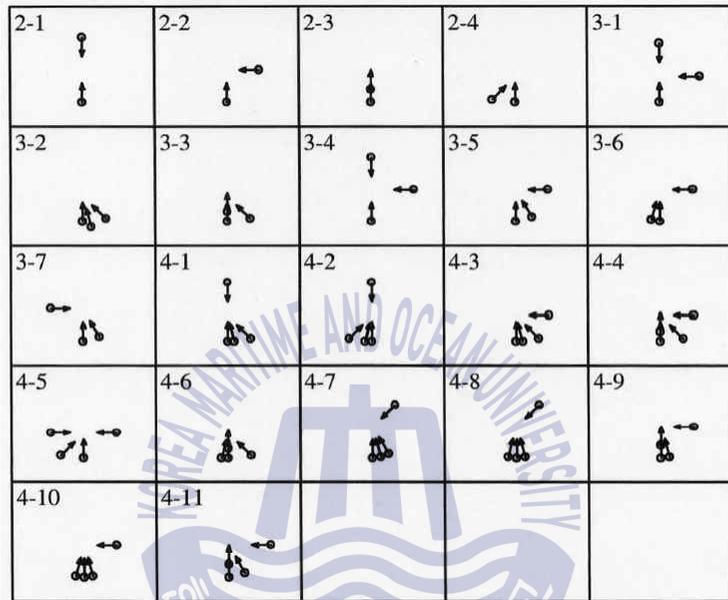


Fig. 6.4 Imazu problem

강화학습의 중요 요소인 탐험(exploration)-활용(exploitation)을 위한 방법은 엡실론-탐욕( $\epsilon$ -greedy) 정책을 사용하였다. 처음에는 무작위로 행동을 선택함( $\epsilon = 1.0$ )으로써 다양한 탐험을 수행하며, 학습이 진행됨에 따라 엡실론 값을 줄임으로써 점차 무작위 행동은 줄이고 학습된 행동을 활용한다. 활용을 위한 행동 선택 방법은 Q 함수로부터 값이 가장 큰 행동을 수행하는 탐욕(greedy)적 방법이다. 또한 엡실론이 임계값( $\epsilon = 0.1$ ) 이하로 떨어지지 않게 하여 학습된 행동만 반복하지 않고 낮은 비율로 탐험을 계속 수행할 수 있도록 한다.

심층강화학습을 이용해 최적 행동 정책을 찾기 위해서 하이퍼파라미터 최적화 작업이 필요하다. 학습 결과는 하이퍼파라미터 값에 민감하기 때문에(Metz et al., 2017; Henderson et al., 2017) 오랜 시간이 걸리더라도 하이퍼파라미터

최적화 작업은 수행되어야 한다. 본 논문에서 선정한 하이퍼파라미터는 감가율 (discount factor), 학습률(learning rate), 배치크기(batch size), 은닉층(hidden layer)의 수, 뉴런(neuron) 수로 두고 4장과 마찬가지로 무작위 탐색법을 이용하였다. 본 연구에서 학습에 사용된 컴퓨터 사양은 CPU Intel i7-7820X, 메모리 32GB를 사용하였다. GPU를 활용하는 것이 CPU를 활용하는 것보다 빠르지만 GPU는 개당 하이퍼파라미터 1세트만 훈련이 가능한 반면 CPU는 동시에 8세트 하이퍼파라미터 훈련이 가능하기 때문에 CPU를 이용하였다. 하이퍼파라미터 1 세트에 대해 50,000번의 에피소드를 훈련하였을 때 약 80시간이 수행되었고 하이퍼파라미터 50세트를 수행하는데 소요되는 전체 시간은 약 167일이며 CPU를 이용함으로써 21일이 소요되었다.

### 6.3 행동 단계와 제어 단계

행동 단계는 의사결정 단계의 회피 행동 변환기 모듈이 판단한 행동을 수행하는 단계다. 행동은 VO를 기반으로 정의하였으며 다수선박의 조우상황을 고려하였다. 제어 단계는 행동 단계로부터 회피 경로와 목표 RPM을 입력받아 선박의 조종제어를 수행하는 단계다.

#### 6.3.1 행동 정의

에이전트가 할 수 있는 행동 집합은 식 (6.7)과 같이 나타낼 수 있다. 행동은 Fig. 6.5와 같이 **목표 경로추종**( $a_0$ ), **우현 회피**( $a_1$ ), **좌현 회피**( $a_2$ ), **목표지점 추종**( $a_3$ ) 모두 4가지이다.

$$A = \{a_0, a_1, a_2, a_3\} \quad (6.7)$$

**목표 경로추종 행동**은 2.4절의 퍼지(Fuzzy)를 이용한 목표 경로추종 방법을 이용한다. **우현 회피**는 5.2절의 VO를 기반으로 한 휴리스틱 탐색 방법을 통해 결정된다. 우현 회피가 VO의 중앙선을 기준으로 시계방향에 위치한 VO 경계선 상에 속도 벡터를 선정하였다면, **좌현 회피**는 Fig. 6.6과 같이 반시계방향에 위치한 VO 경계선 상에 속도 벡터를 선정한다. 선박의 항해 속력을 반지름으로

하는 원(Fig. 6.6의 점선)과 반시계방향에 위치한 VO 경계선(Fig. 6.6의 1점 쇄선)의 교점이 좌현 회피를 위한 목표 속도다. 이로써 추월 상황에서 COLREGs가 허용하고 있는 좌현 회피를 통한 추월도 가능해진다. **목표지점 추종**은 목표 경로의 끝 지점을 향하는 속도 벡터를 선택하는 것이다. 목표지점 추종 행동은 Fig. 6.7의 1점 쇄선과 같이 목표지점을 향하고 향해 속력 크기를 갖는 속도 벡터(Fig. 6.7의 파란색 화살표)를 선정한다. 목표 지점추종을 행동에 포함시킨 이유는 추월 조우상황에서 타선의 진로를 방해하지 않는 행동이 필요하기 때문이다. 추월 조우상황에서 Fig. 6.8과 같이 자선이 타선을 지나친 뒤, 목표 경로추종 행동( $a_0$ )과 우현회피 행동( $a_1$ )은 타선의 진로를 방해하게 될 가능성이 크다. 목표지점 추종 행동( $a_3$ )은 자선이 타선의 진로를 가능한 방해하지 않도록 한다.

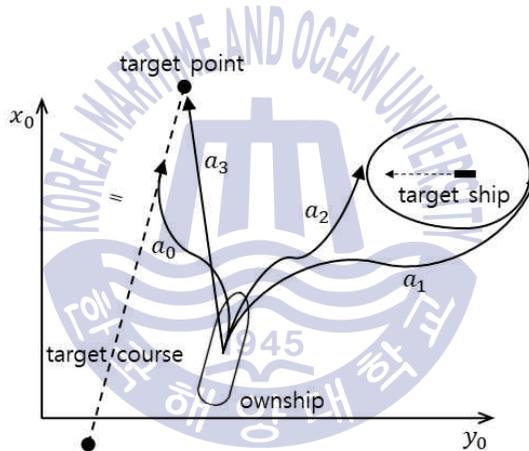


Fig. 6.5 Action of agent for action step

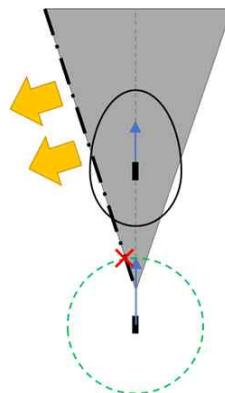


Fig. 6.6 Selection method of target velocity for avoidance to port

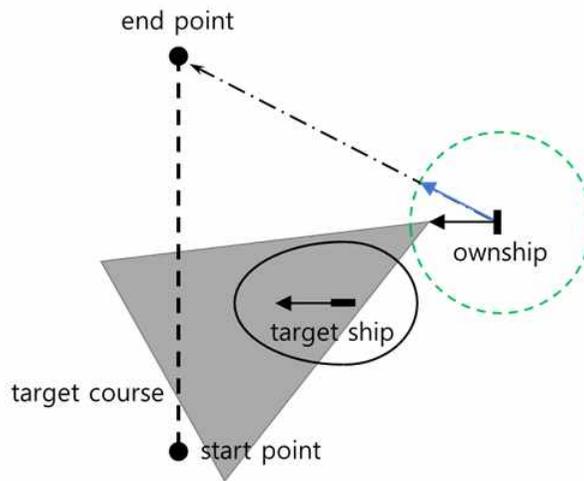


Fig. 6.7 Selection method of target velocity for target point following

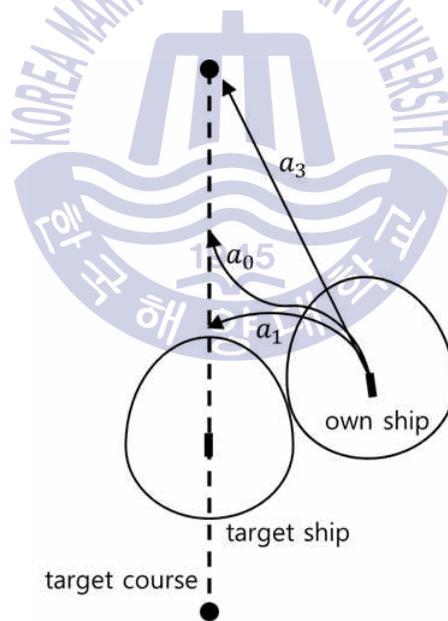


Fig. 6.8 Problem of action selection in overtaking situation

### 6.3.2 다수선박에서 행동 정의

다수선박의 조우상황은 각 타선에 대한 VO를 생성하게 되며 단일선박에 비해 VO 범위가 넓고 복잡하게 된다. 자선의 주변에 다수선박이 존재하더라도 6.1.1에서 정의한 회피 행동 영역 ㉞ 단계에 있는 선박만 고려대상이 된다. ㉞ 단계에 존재하는 선박이 2대일 경우 Fig. 6.9와 같은 상황을 가정하였다. OS(자선)는 TS1(타선1)을 추월하는 중에 교차 조우상황의 TS2(타선2)를 만나는 경우다. 우선 TS1과 TS2의 충돌위험도를 계산하며 충돌위험도는 3.5절의 계산 방법을 이용하였다. TS1의 충돌위험도가 더 높고 회피 행동 변환기 모듈이 우현회피 행동( $a_1$ )을 결정했다 가정하며, 자선은 휴리스틱 방법에 따라 결정되는 교차점(Fig. 6.9의 붉은색 X 표시)을 목표 속도로 선정해야 한다. 하지만 VO2 범위 안에 있기 때문에 TS2와 충돌 가능성이 존재며, 이때 속력을 조절함으로써 VO2 범위를 피한다. 목표 속도 벡터와 VO2 경계의 교차점을 새로 구하고 새 교차점을 향하는 목표 속도 벡터(파란색 화살표)를 선택한다. 속력은 선택한 최종 목표 속도 벡터의 크기가 된다. 목표속도를 구하는 절차를 정리하면 Table 6.2와 같다.

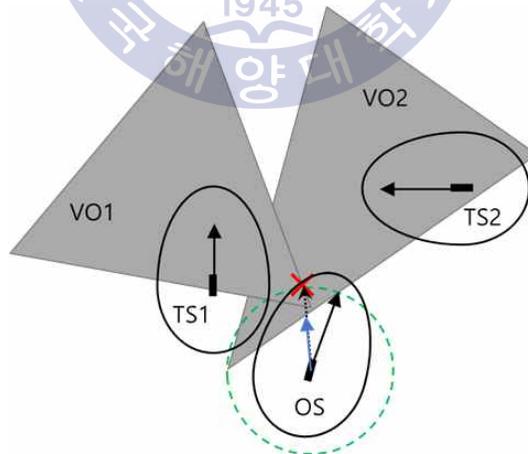


Fig. 6.9 Selection method of target velocity at multiple encounter situation

**Table 6.2** Procedure of target velocity selection for multiple ship avoidance

```

procedure 다수선박 조우상황 시 목표 속도 선택
  for  $id = 1, T$  do
    각 선박에 대한 충돌위험도 계산( $CR_{id}$ )
  end for
  최대 충돌위험도( $CR_{max}$ )를 갖는 선박의 ID 선택( $ID_{TS}$ )
  선택된 선박( $ID_{TS}$ )의  $VO_{ID_{TS}}$  경계선으로 부터 목표 속도 결정( $v^*$ )
  for  $id = 1, T$  do
    if  $id == ID_{TS}$  then
      continue
    end if
    if 목표 속도  $v^*$ 가  $VO_{id}$ 의 영역에 있는지 여부 확인 then
       $VO_{id}$ 의 경계선으로 부터 새로운 목표 속도 결정( $v_{new}^*$ )
    end if
  end for

```

### 6.3.3 제어 단계

제어 단계는 행동 단계에서 결정된 속도벡터로부터 회피 경로와 목표 RPM을 입력 데이터로 받는다. 회피 경로는 속도벡터의 방향으로부터 결정되고 목표 RPM은 속도벡터의 크기로부터 결정된다. 회피 경로는 퍼지(Fuzzy) 조타 제어기에 입력되어 목표타각을 산출하며 목표 RPM과 함께 선박의 운동에 반영된다. 퍼지 조타 제어기는 2.4절의 자동 경로추종 제어기를 활용하였다.

## 6.4 VO와 심층강화학습 기반 충돌회피 결과 및 고찰

### 마주침 조우상황

마주침 조우상황에서 초기 조건은 5장의 Table 5.2 초기 조건과 동일하고 결과는 Fig. 6.10, Fig. 6.11, Fig. 6.12와 같다. Fig 6.10과 Fig, 6.11은 특정 시간에서 자선과 타선의 이동 경로를 타나내며 3,168초 시간의 그래프에서 X 표시는 두 선박의 감시영역이 겹치는 순간에 선박의 위치를 나타낸다. 두 선박은 감시영역이 겹치더라도 즉시 회피하는 것이 아니라 얼마 동안 목표 경로 추종 후 회피 동작을 하고 있으며, COLREGs 규칙에 따라 모두 상대선을 좌현에 두고 우현으로 회피하였다. Fig. 6.12는 왼쪽 위부터 시계방향으로 속도, 선수각, 회두각속도, 타각에 대한 시계열 그래프를 나타내고 있으며, 자선과 타선의 회피운동이 동일하기 때문에 시계열 그래프가 겹쳐 있다. 시계열 그래프에서 선수각은  $-180\sim 180$ 도 범위의 값을 갖기 때문에 선수각이  $-180$ 도 미만과  $180$ 도 이상의 값을 갖게 될 경우 부호가 바뀌는 현상을 없애기 위해 초기 선수각을  $0$ 도로 두고 그래프를 나타낸다.

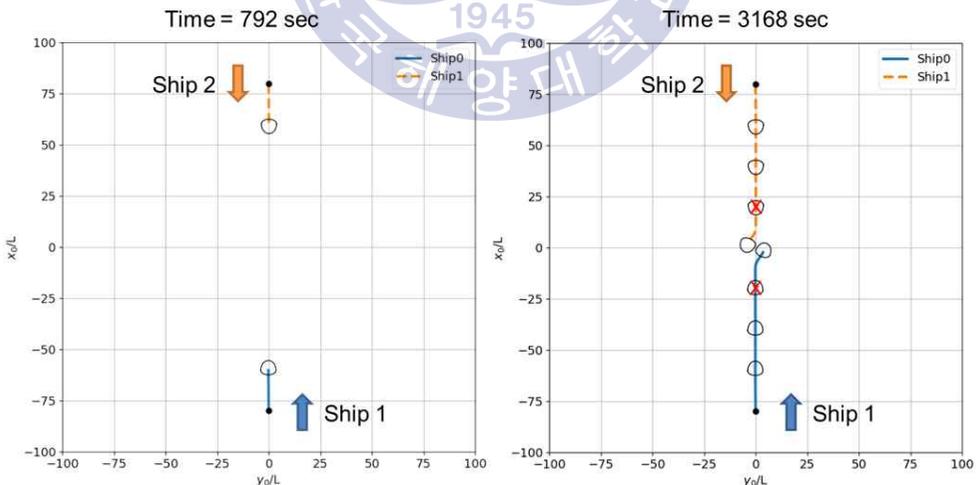


Fig. 6.10 Trajectories of  $180^\circ$  head-on situation using VO & DRL at 792sec and 3,168sec

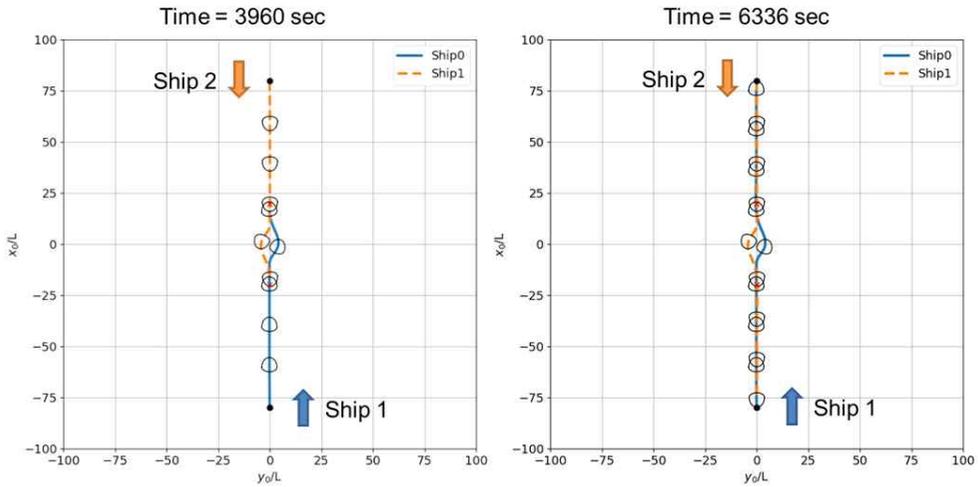


Fig. 6.11 Trajectories of 180° head-on situation using VO & DRL at 3,960sec and 6,336sec

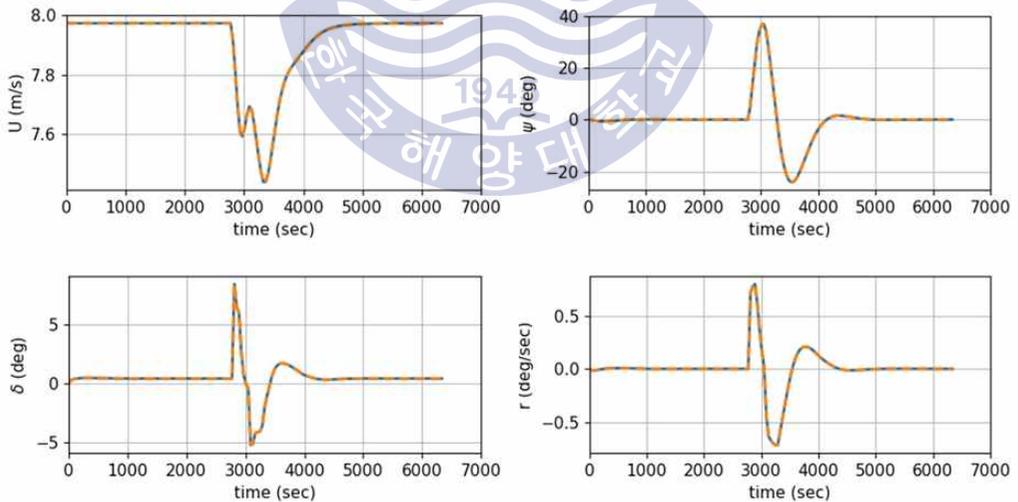


Fig. 6.12 Time histories of 180° head-on situation using VO & DRL

## 교차 조우상황

교차 조우상황에서 초기 조건은 Table 5.3과 같고 결과는 Fig. 6.13, Fig. 6.14와 같다. COLREGs 규칙에 따라 타선을 우측에 두고 있는 선박(Ship 1)은 피항선, 타선을 좌측에 두고 있는 선박(Ship 2)은 유지선이 되며, 이와 같은 규칙을 준수하고 있다. 두 선박의 감시영역이 겹치는 순간 각 선박의 위치가 Fig. 6.13의 3,168초 시간의 그래프 상에 X 표시로 나타난다. 감시영역이 겹치더라도 회피 동작을 바로 수행하지 않으며 학습을 통해 보상을 최대화하는 충돌 회피 시작지점이 선정되었다.

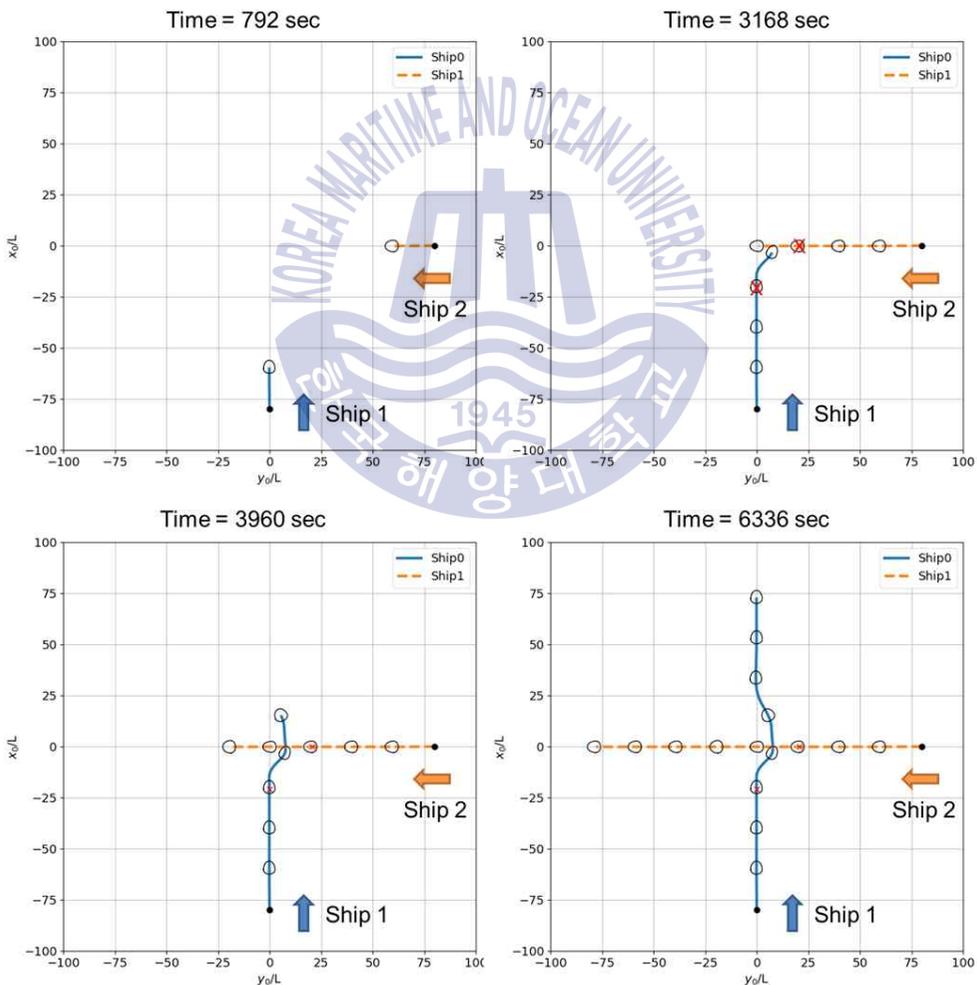


Fig. 6.13 Trajectories of 90° crossing situation using VO & DRL

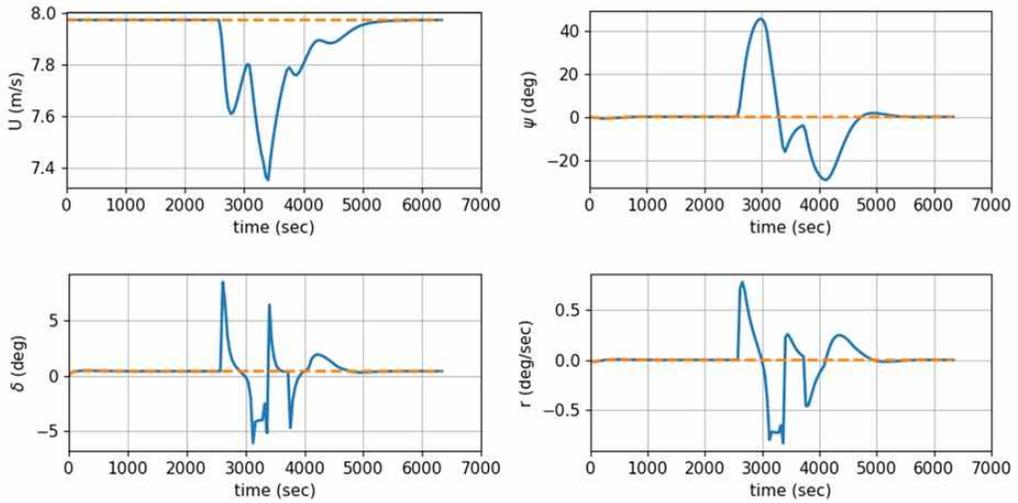


Fig. 6.14 Time histories of 90° crossing situation using VO & DRL

### 추월 조우상황

추월 조우상황에서 초기 조건은 Table 5.5와 같고 결과는 Fig. 6.15, Fig. 6.16과 같다. COLREGs 규칙에 따르면 추월 방향은 우현방향 또는 좌현방향 모두 가능하지만 Fig. 6.15 추월선의 자취를 보면 우현방향으로 회피하도록 학습된 것을 알 수 있다. 또한 추월선은 피추월선을 추월한 후에 목표 침로로 바로 복귀하지 않고 Fig. 6.15의 3,960초 그래프와 같이 피추월선의 폐색영역과 겹치지 않도록 피해가는 것을 볼 수 있다.

앞서 마주침과 교차 조우상황 결과와는 달리 추월 조우상황의 Fig. 6.16 시계열 그래프에서 타각이 빈번히 변경되는 것을 확인할 수 있다. 추월 조우상황의 경우 다른 조우상황에 비해 회피 시작 시점이 빠르고 회피 종료 시점은 느리다. 그 때문에 목표 경로 이탈 거리와 속도 감쇄에 대한 보상 식 (6.3)에서 목표 경로 이탈 거리에 의해 누적된 음의 보상 값이 크므로 속도 감쇄에 대한 음의 보상을 최소화하는 방향으로 학습되었기 때문이다. 마주침과 교차 조우상황에서 속도의 최저 값은 각각 7.44m/s, 7.35m/s 인 반면 추월 조우상황에서는 7.83m/s로 초기 속도 7.97m/s로부터 가장 작은 속도 감쇄를 보인다. 즉, 추월

조우상황에서는 두 선박의 폐색영역이 겹치지 않게 하도록 목표 경로 이탈 거리가 커지기 때문에 이에 대한 누적된 음의 보상 값이 커지게 되므로 속도 감쇄에 대한 음의 보상을 최소화하는 행동을 학습하게 되며 타각을 빈번히 변경함으로써 속도 감쇄를 최소화하고 있다.

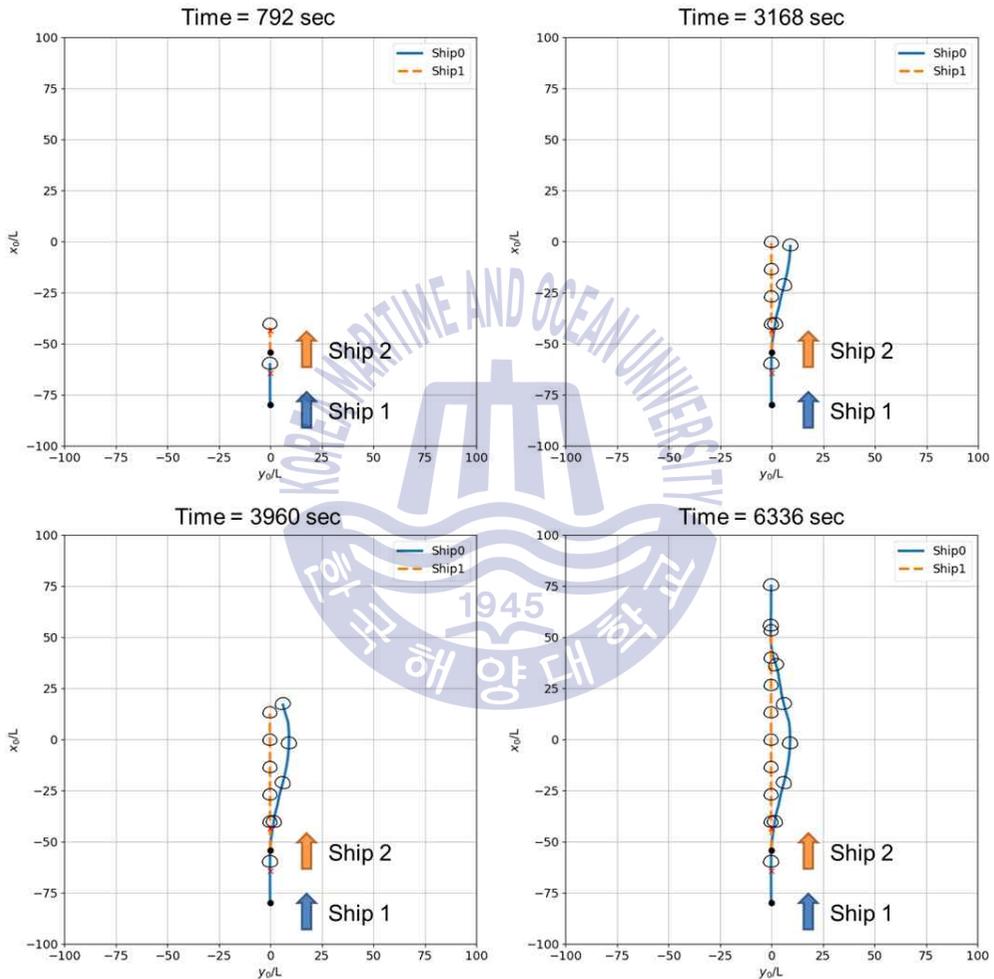


Fig. 6.15 Trajectories of  $0^\circ$  overtaking situation using VO & DRL

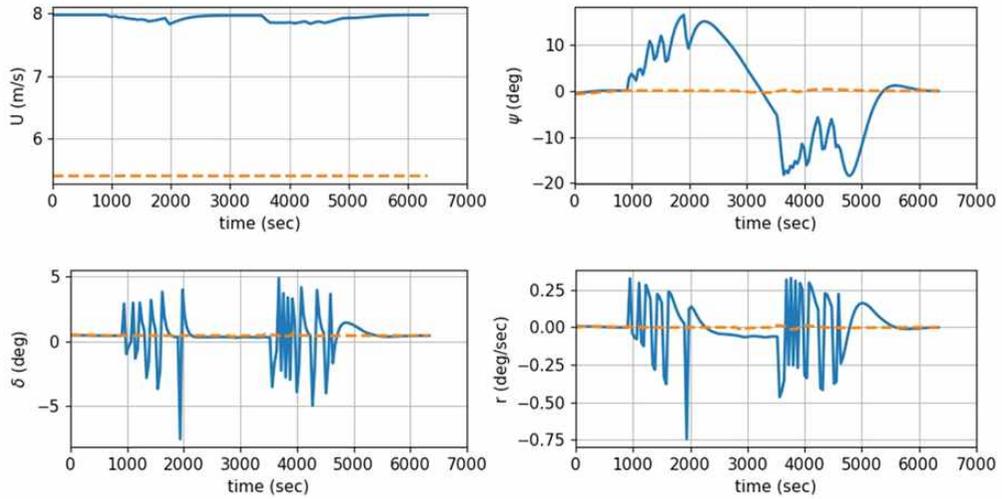


Fig. 6.16 Time histories of 0° overtaking situation using VO & DRL

### 다수선박의 조우상황 1 (GW-HO)

3척 선박이 조우하는 상황에 대해 VO와 심층강화학습을 이용한 방법을 테스트하였으며 각 선박의 초기 조건은 Table 6.3과 같다. Ship 1과 Ship 2는 피항선-유지선 의무 관계에 있으며 Ship 1이 피항선, Ship 2는 유지선의 의무가 있다. Ship1과 Ship 3은 서로에게 마주침선의 의무를 가지며 Ship 2와 Ship 3은 피항선-유지선 의무 관계에 있다. 결과는 Fig. 6.17, Fig. 6.18과 같으며 모든 선박이 서로 간에 피항선 또는 마주침선의 의무가 하나씩 있기 때문에 Fig. 6.17에서 모든 선박이 우회회피를 통해 피항하고 있는 것을 확인할 수 있다.

Table 6.3 Initial condition of multiple ship collision avoidance 1

	Ship 1	Ship 2	Ship 3
$(x, y)$	(-25600, 0) m	(0, 25600) m	(25600, 0) m
$\psi$	0 deg	90 deg	180 deg
$U$	6.69 m/s (13 knots)	7.97 m/s (15.5 knots)	7.97 m/s (15.5 knots)

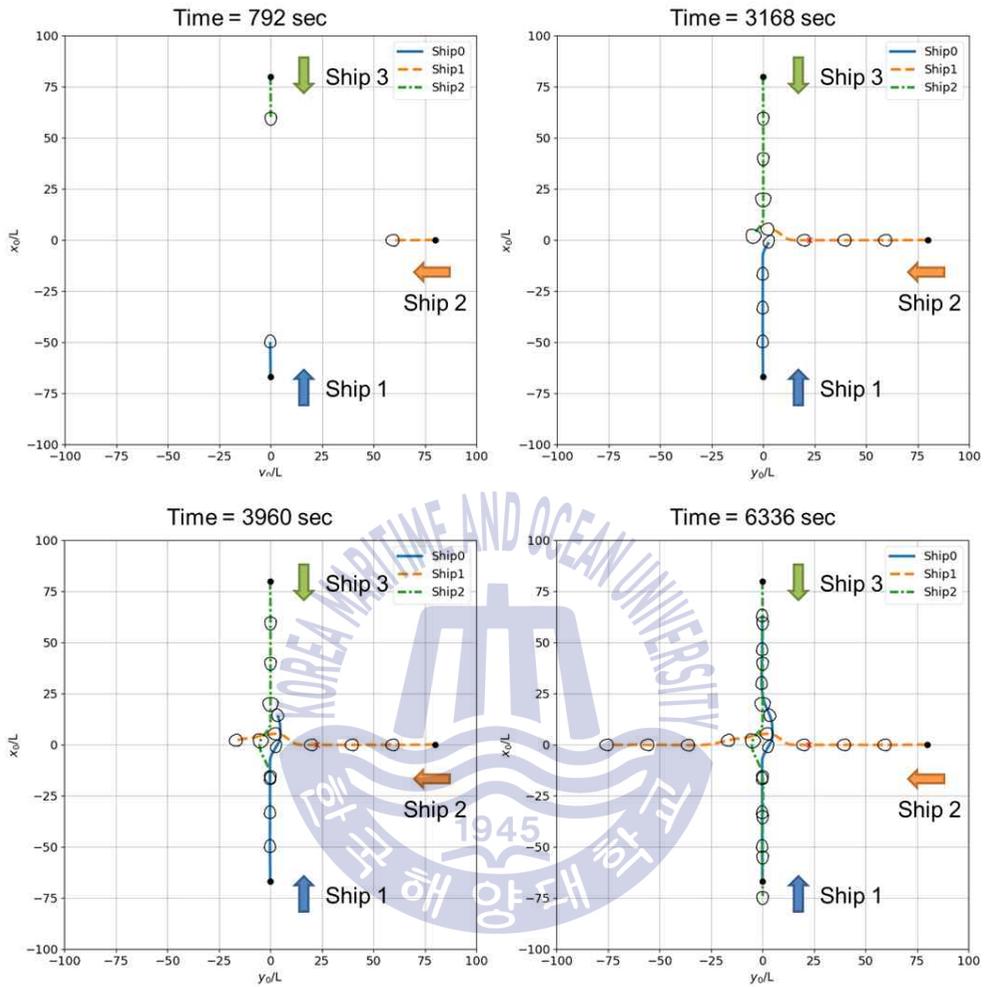


Fig. 6.17 Trajectories of multiple ship collision avoidance 1

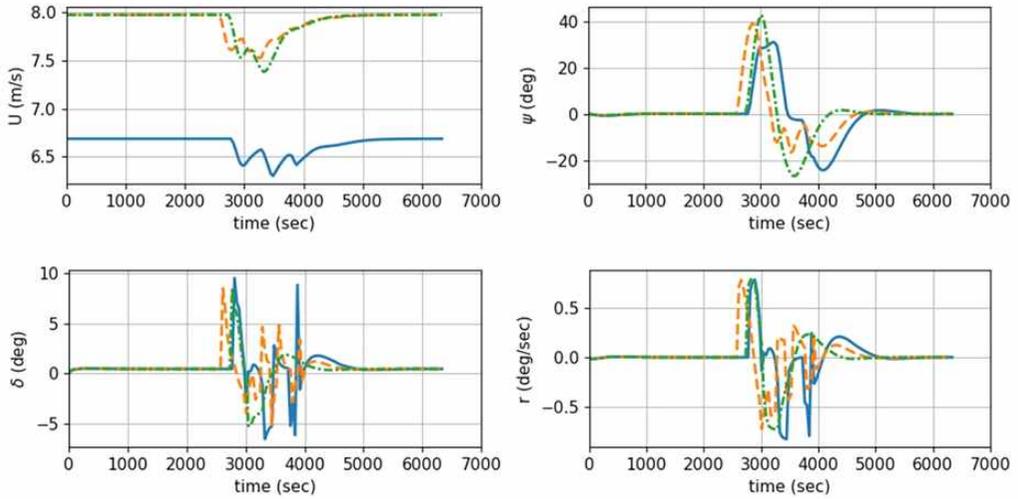


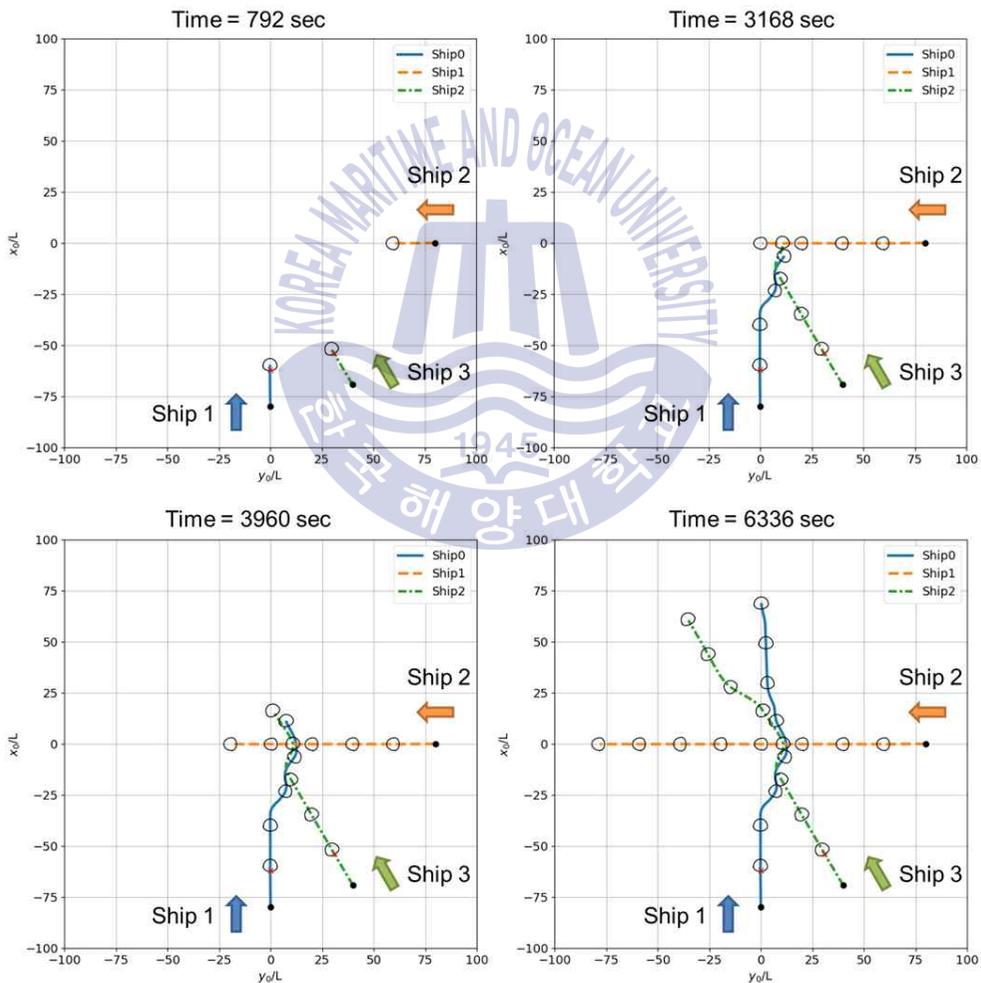
Fig. 6.18 Time histories of multiple ship collision avoidance 1

### 다수선박의 조우상황 2 (GW-SO)

3척이 조우하는 상황으로 각 선박의 초기 조건은 Table 6.4와 같고 결과는 Fig. 6.19, Fig. 6.20과 같다. Ship 2는 모두에게 유지선의 의무를 가지며 Ship 3과 Ship 1은 Ship 2에게 피항선이다. Ship 1과 Ship 3은 피항선-유지선 의무 관계를 가진다. Fig. 6.19의 3,168초 그래프에서 Ship 1이 Ship 3을 피항하기 위해 우회회피를 시도하며 이후 Ship 3도 Ship 2를 피항하기 위해 우회회피를 수행한다. Ship 3이 우회회피를 하자 Ship 1은 목표 경로로 복귀하는 듯하지만 Ship 2를 피항하기 위해 다시 우회회피를 수행한다. 3,620초 그래프에서 Ship 1과 Ship 3은 Ship 2를 완전히 회피하였으며 이후 각 선박은 목표경로를 추종한다.

**Table 6.4** Initial condition of multiple ship collision avoidance 2

	Ship 1	Ship 2	Ship 3
$(x, y)$	$(-25600, 0)$ m	$(0, 25600)$ m	$(-34637, 9984)$ m
$\psi$	0 deg	90 deg	30 deg
$U$	7.97 m/s (15.5 knots)	7.97 m/s (15.5 knots)	7.97 m/s (15.5 knots)



**Fig. 6.19** Trajectories of multiple ship collision avoidance 2

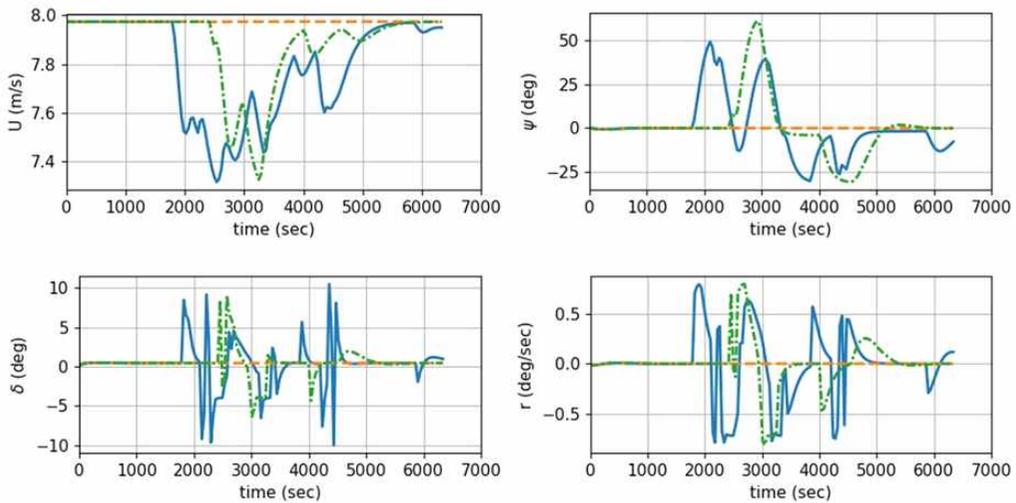


Fig. 6.20 Time histories of multiple ship collision avoidance 2

### 다수선박의 조우상황 3 (HO-GW)

4척 선박이 조우하는 상황으로 각 선박의 초기조건은 Table 6.5와 같고 결과는 Fig. 6.21, Fig. 6.22와 같다. Ship 1과 Ship 2 그리고 Ship 2와 Ship 4는 서로 마주침선의 의무를 갖고 Ship 1과 Ship 3, Ship 1과 Ship 4, Ship 4와 Ship 3, Ship 3과 Ship 2는 모두 피항선-유지선 의무 관계에 있다. Fig. 6.21의 3,168초 그래프에서 Ship 1은 Ship 4를 피항하기 위해 우회회피를 하고 Ship 4는 Ship 3을 피항하기 위해 우회회피를 한다. Fig. 6.21의 3,960초 그래프에서 Ship 3은 Ship 2를 피항하고 있으며 모든 선박은 회피가 끝난 후 목표 경로로 복귀한다. 여기서 Ship 1과 Ship 2, Ship 2와 Ship 4는 마주침 선박이지만 Ship 1과 Ship 4가 Ship 3을 피항하기 위해 우현으로 이동하였기 때문에 Ship 2는 유지선인 것처럼 목표 경로만 추종한다.

Table 6.5 Initial condition of multiple ship collision avoidance 3

	Ship 1	Ship 2	Ship 3	Ship 4
$(x, y)$	(-25600, 0)m	(0, 25600)m	(-21478, 21414)m	(-22605, 18099)m
$\psi$	0 deg	180 deg	45 deg	15 deg
$U$	6.69 m/s (13 knots)	7.97 m/s (15.5 knots)	7.97 m/s (15.5 knots)	6.69 m/s (13 knots)

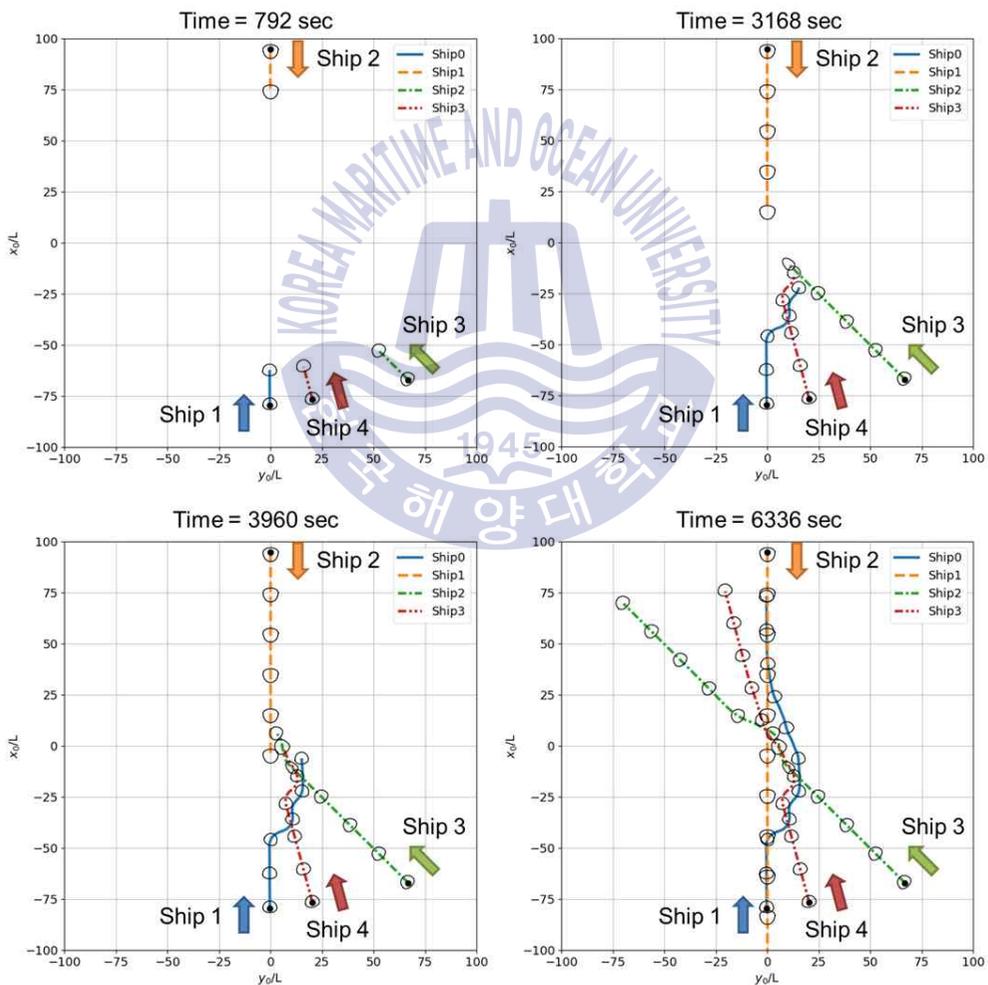


Fig. 6.21 Trajectories of multiple ship collision avoidance 3

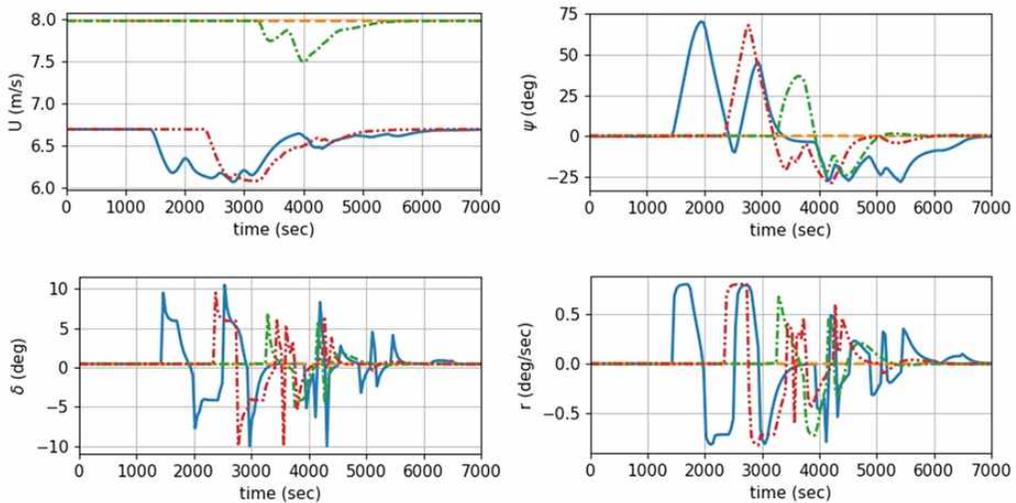


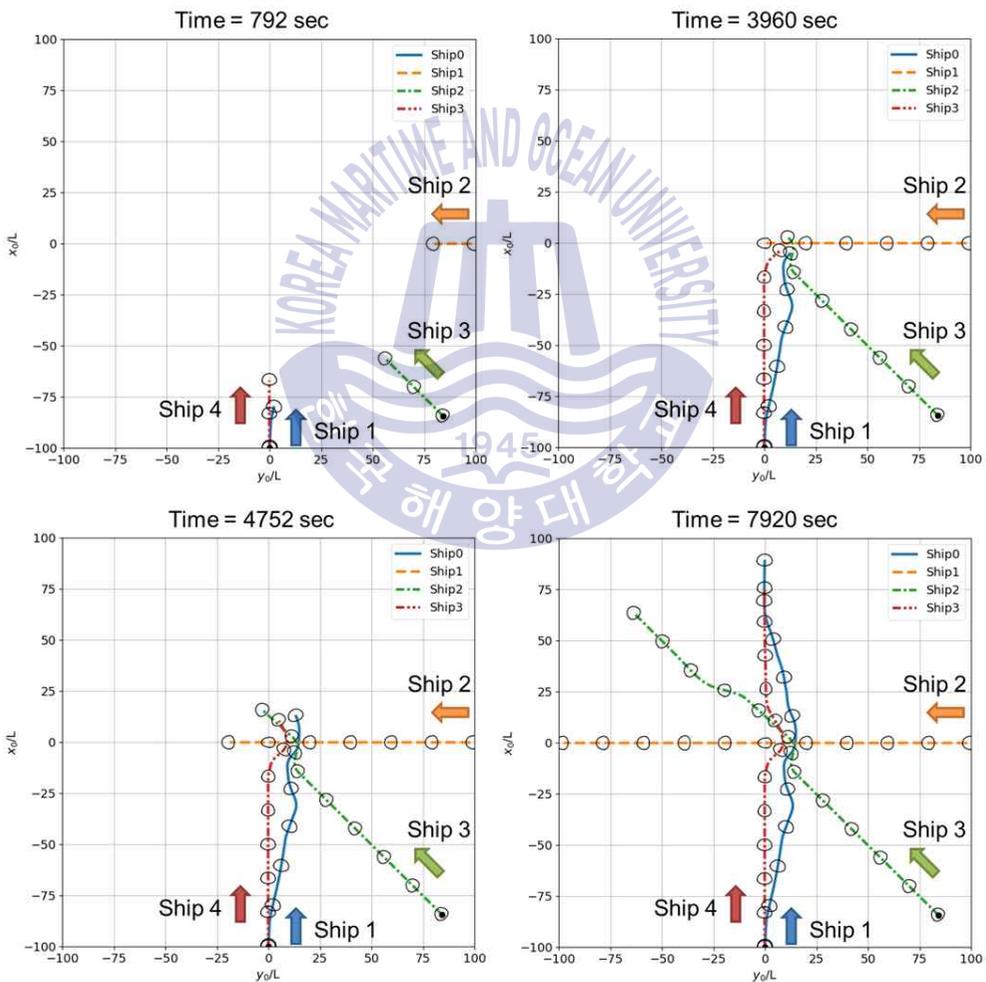
Fig. 6.22 Time histories of multiple ship collision avoidance 3

#### 다수선박의 조우상황 4 (SO-GW-OT)

4척 선박이 조우하는 상황으로 각 선박의 초기조건은 Table 6.6과 같고 결과는 Fig. 6.23, Fig. 6.24와 같다. Ship 1과 Ship 4는 추월선-피추월선 의무 관계에 있으며 Ship 2는 모든 선박에게 유지선이 되고 Ship 1과 Ship 3, Ship 4는 Ship 2에게 피항선의 의무를 가진다. Fig 6.23의 3,960초 그래프에서 Ship 1이 Ship 4를 추월하는 중 Ship 1의 입장에서 유지선이어야 하는 Ship 4가 Ship 2를 피항하기 위해 우현 회피를 한다. 이때 Ship 4가 Ship 1에게 접근하기 때문에 Ship 1은 Ship 4를 피하기 위해 우현 방향으로 회피하게 된다. 이 과정에서 두 선박의 폐색영역이 겹치는 것을 확인할 수 있지만 유지선임에도 다른 선박을 피항하기 위해 회피를 수행할 경우 이를 피하기 위한 대처 행동을 학습했다는 것을 알 수 있다. Ship 2는 모든 선박에게 유지선의 의무를 갖기 때문에 어떠한 변칙도 하지 않았다.

**Table 6.6** Initial condition of multiple ship collision avoidance 4

	Ship 1	Ship 2	Ship 3	Ship 4
$(x, y)$	$(-32000, 0)\text{m}$	$(0, 32000)\text{m}$	$(-22605, 18099)\text{m}$	$(-26747, 0)\text{m}$
$\psi$	0 deg	90 deg	45 deg	0 deg
$U$	7.97 m/s (15.5 knots)	7.97 m/s (15.5 knots)	7.97 m/s (15.5 knots)	6.69 m/s (13 knots)



**Fig. 6.23** Trajectories of multiple ship collision avoidance 4

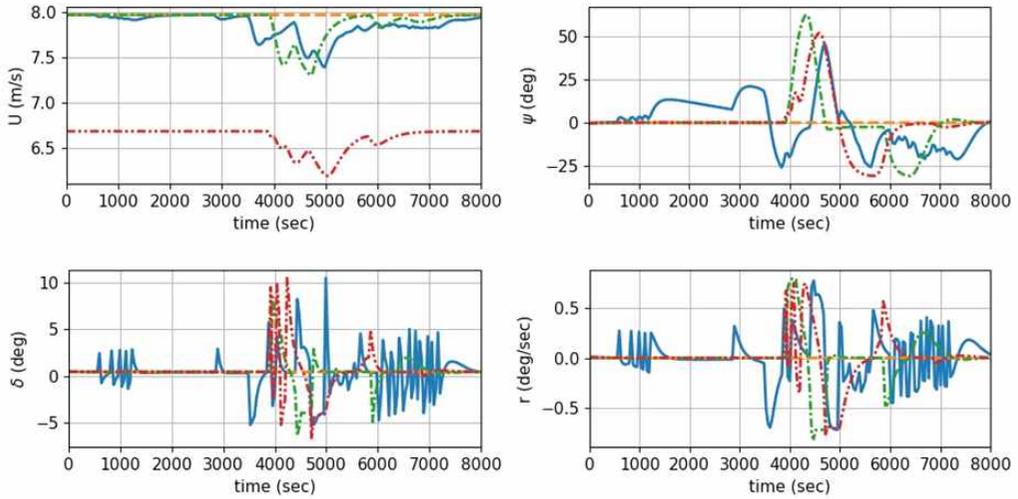


Fig. 6.24 Time histories of multiple ship collision avoidance 4

#### 다수선박의 조우상황 5 (SO-GW)

4척 선박이 조우하는 상황으로 각 선박의 초기조건은 Table 6.7과 같고 결과는 Fig. 6.25, Fig. 6.26과 같다. Ship 2는 다른 모든 선박들에 대하여 유지선의 의무를 갖고 있으며 Ship 1, Ship 3, Ship 4는 Ship 2에 대해 피항선의 의무를 가진다. 그리고 Ship 1과 Ship 4, Ship 4와 Ship 3, Ship 1과 Ship 3은 모두 피항선-유지선 의무 관계에 있다. Fig. 6.25의 3,168초 그래프에서 Ship 1은 Ship 4를 피항하기 위해 우현 회피를 시작하지만 Ship 4 역시 Ship 3을 피항하기 위해 우현회피를 시작한다. Ship 1의 입장에서 유지선이어야 할 Ship 4가 변침하여 멀어지기 때문에 Ship 1은 회피 대상을 잃고 목표 침로를 찾아가는 것을 볼 수 있다. 그리고 목표 침로를 추종하던 Ship 1은 Ship 3과 Ship 2를 차례로 피항하기 위해 우현 회피를 다시 수행한다. 이후 모든 선박들은 회피가 끝나고 다시 목표 침로를 추종하게 된다.

Table 6.7 Initial condition of multiple ship collision avoidance 5

	Ship 1	Ship 2	Ship 3	Ship 4
$(x, y)$	(-25600, 0)m	(18048, 18112)m	(-22182, 12762)m	(-24742, 6502)m
$\psi$	0 deg	135 deg	30 deg	15 deg
$U$	7.97 m/s (15.5 knots)	7.97 m/s (15.5 knots)	7.97 m/s (15.5 knots)	7.97 m/s (15.5 knots)

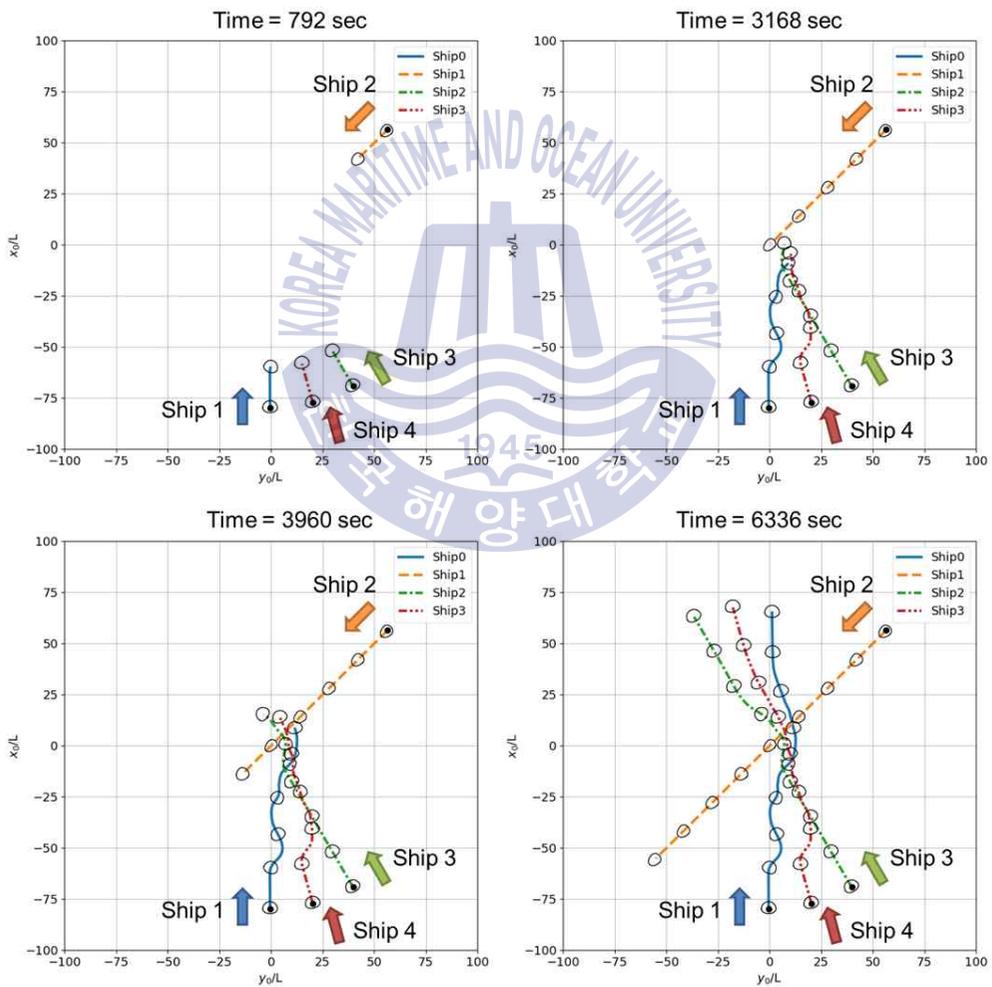


Fig. 6.25 Trajectories of multiple ship collision avoidance 5

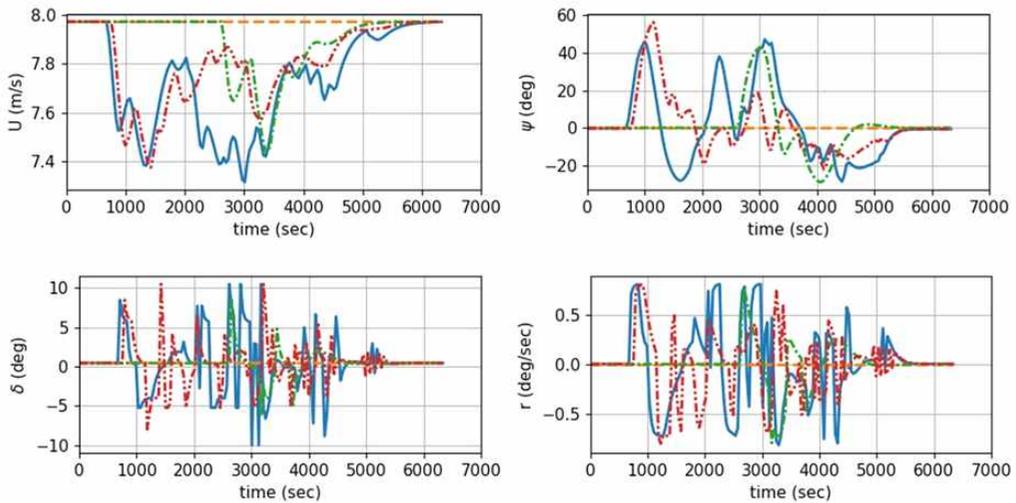


Fig. 6.26 Time histories of multiple ship collision avoidance 5

#### 6.4.1 다른 방법들과 비교

Velocity Obstacles와 심층강화학습을 이용한 충돌회피 방법의 효과를 검증하기 위해 다른 충돌회피 알고리즘과 비교한다. 비교 대상으로 선정한 충돌회피 알고리즘은 Kuwata et al.(2014)의 VO를 기반으로 한 충돌회피 방법과 본 논문의 3.5절에서 설명된 충돌위험도 기반 충돌회피 방법이다. 본 연구에서 제시하고 있는 방법은 “DRL”, Kuwata et al.의 VO 기반 방법은 “VO”, 충돌위험도 기반 방법은 “CR”로 표기한다. “CR” 방법은 충돌위험도가 0.6 이상일 경우 회피를 시작하며 충돌위험도가 0.8을 넘으면 매우 위험한 상황임을 의미한다. “VO” 방법에서 회피를 시작하는 TCPA 임계값은 679초로 설정하였으며 VO 영역의 정의를 “DRL” 방법과 동일하게 함으로써 같은 조건으로 비교한다. “CR” 방법과 “VO” 방법은 TCPA가 0보다 작을 경우 목표경로를 추종하도록 하였다.

다른 방법들에 대한 비교는 Table 5.2, Table 5.3, Table 5.5와 같이 마주침, 교차, 추월 조우상황에서 수행한다. 각 조우상황에 대한 수행 결과는 Fig. 6.27, Fig. 6.28, Fig. 6.29와 같으며, 왼쪽 그림은 Ship 1 위치에 대한 자취를 나타내

고 있으며 오른쪽 그림은 위쪽부터 속도, 충돌위험도, 두 선박의 폐색영역 겹침 비율에 대한 시간이력을 나타낸다. 폐색영역 겹침 비율( $O$ )은 식 (6.8)과 같이 자선의 전체 폐색영역 대비 겹친 폐색영역의 비율을 의미한다.

$$O = (\text{overlapped area} / \text{total area}) \times 100 \quad (6.8)$$

다른 방법들과 정량적인 비교를 위해 3가지 정량적인 지표를 마련한다. 첫째, 충돌회피 과정에서 산출되는 최대 충돌위험도를 확인함으로써 어떤 방법이 충돌위험이 가장 큰 방법인지 판단한다. 둘째, 충돌회피 과정에서 두 선박의 폐색영역이 겹치기 시작하는 시간부터 폐색영역이 겹치지 않게 되는 시간까지 폐색영역 겹침 비율을 평균하여 나타냄으로써 회피 과정 중 자선의 폐색영역이 평균적으로 중첩되는 정도를 정량적으로 확인한다. 셋째, 운항효율을 평가하기 위한 지표로 식 (6.9)와 같이 자선이 피항하는 동안 목표 경로에서부터 떨어진 수직 거리의 변화량( $\Delta l$ )과 속도 감쇄에 따른 손실 거리( $|U_{target} - U| \cdot \Delta t$ )의 누적 합을 그 값으로 하며 선박의 길이( $L$ )에 의해 무차원화 된다. 즉, 충돌회피 과정에서 발생하는 추가 운항 거리( $l'_{extra}$ )를 의미하며 값이 클수록 추가 운항 거리가 늘어난 것이기 때문에 운항효율이 떨어지는 것을 의미한다. Table 6.8은 마주침, 교차, 추월 조우상황에서 각 방법에 대한 3가지 정량적 지표를 비교하고 있으며 Fig. 6.30, Fig. 6.31, Fig. 6.32는 각 정량적 지표를 막대 그래프로 나타내고 있다.

$$l'_{extra} = \left( \sum (\Delta l + |U_{target} - U| \cdot \Delta t) \right) / L \quad (6.9)$$

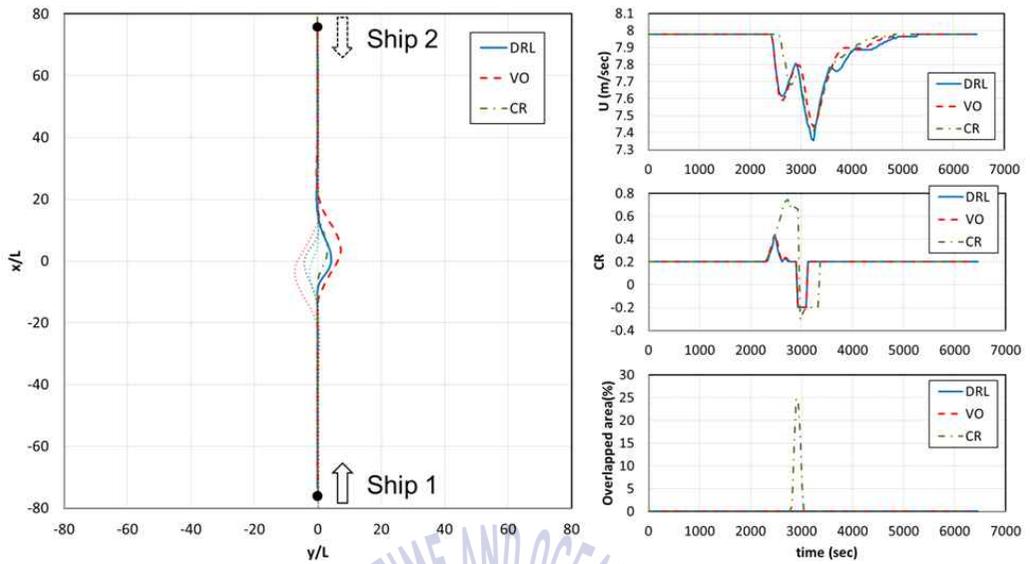


Fig. 6.27 Collision avoidance results to compare with other methods in head-on situation

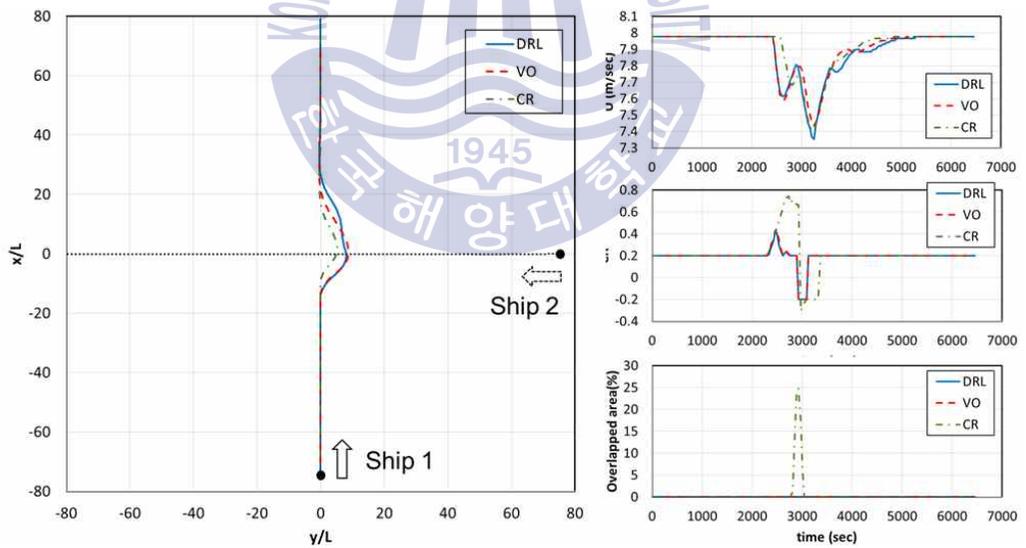


Fig. 6.28 Collision avoidance results to compare with other methods in crossing situation

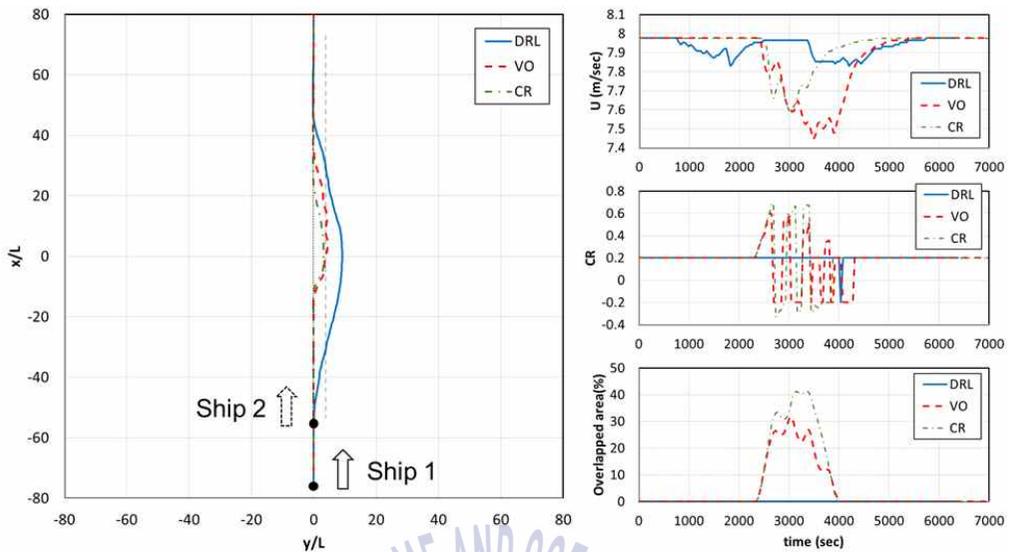


Fig. 6.29 Collision avoidance results to compare with other methods in overtaking situation

Table 6.8 Quantitative measures of collision avoidance to compare with other methods

max collision risk			
	DRL	VO	CR
head-on	0.65	0.41	0.85
crossing	0.42	0.44	0.75
overtaking	0.20	0.61	0.70
average overlapped area(%)			
	DRL	VO	CR
head-on	0	0	18
crossing	0	0	14
overtaking	0	18	26
extra operation distance			
	DRL	VO	CR
head-on	10.6	16.4	7.5
crossing	18.6	20.1	12.7
overtaking	19.8	12.5	8.3

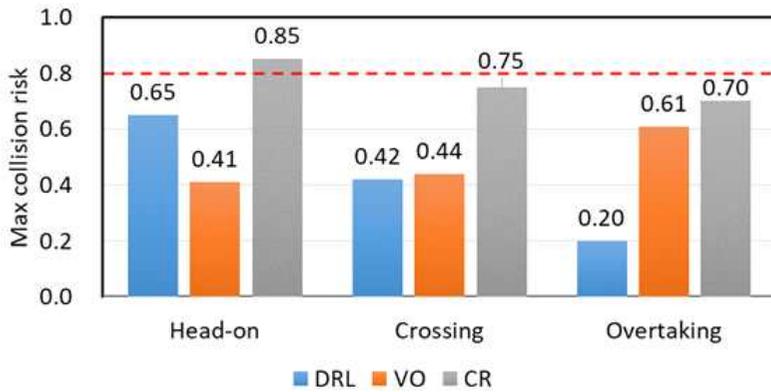


Fig. 6.30 Bar graph for comparing max collision risk

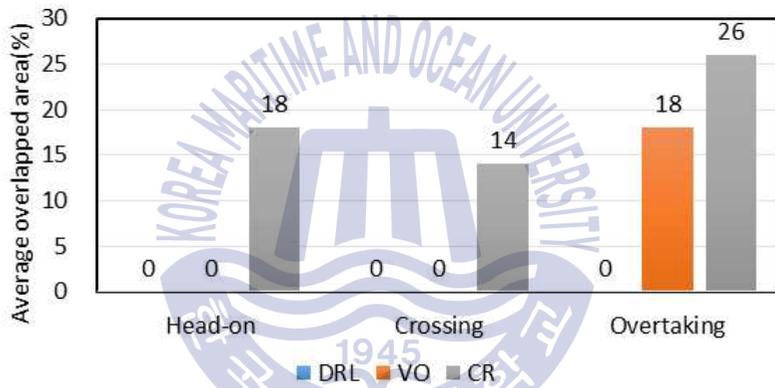


Fig. 6.31 Bar graph for comparing average overlapped area

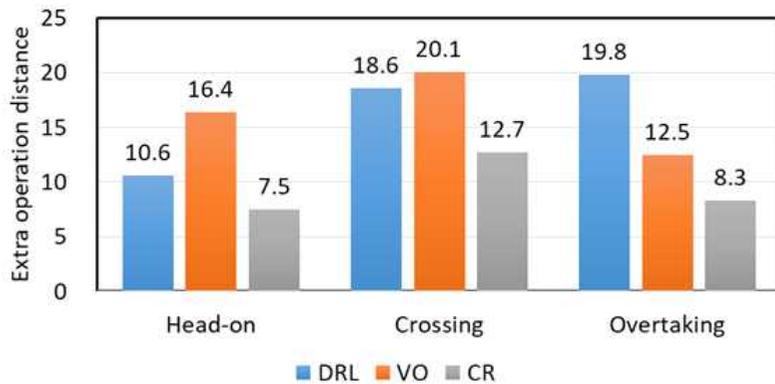


Fig. 6.32 Bar graph for comparing extra operation distance

Fig. 6.30은 왼쪽부터 “DRL”, “VO”, “CR” 방법에 대한 막대 수치가 차례로 나타나고 있으며 마주침, 교차, 추월 조우상황에 대해 각 방법의 최대 충돌위험도를 비교하고 있다. “CR” 방법은 각 조우상황에서 최대 충돌위험도가 가장 높으며 심지어 마주침 조우상황에서 매우 위험한 상황임을 나타내는 0.8 값을 넘었다. “VO” 방법과 “DRL” 방법은 모두 최대 충돌위험도가 0.8 이하로 나타나기 때문에 안전한 회피를 하고 있다. Fig. 6.31은 각 방법에 대한 평균 폐색영역 중첩 비율을 비교하고 있으며 “DRL” 방법은 회피를 수행하는 동안 중첩 영역이 전혀 나타나지 않는다. “VO” 방법은 마주침과 교차 상황에서 중첩 영역이 나타나지 않지만 추월 상황에서 중첩 영역이 나타난다. “CR” 방법은 15~25%의 평균 중첩 비율을 보이며 회피 중 선박 안전영역에 근거한 안전거리를 유지하지 못하고 두 선박의 폐색영역이 중첩된다. Fig. 6.32는 추가 운항거리를 비교하고 있으며 마주침과 교차 조우상황에서 “CR” 방법이 가장 수치가 작고 “VO” 방법이 가장 수치가 크다. 추월 조우상황에서는 “CR” 방법이 수치가 가장 작고 “VO”, “DRL” 순서로 값이 커진다.

“CR” 방법은 목표 경로에 가장 가깝게 회피하기 때문에 세 방법 중 가장 위험한 회피 경로를 보이는 반면 운항효율은 가장 높다. 모든 조우상황에서 “DRL” 방법과 “VO” 방법은 두 선박의 폐색영역이 겹치지 않도록 VO를 정의하기 때문에 안전영역이 겹치지 않도록 회피함으로써 기본적인 안전성을 보장하지만 운항효율은 “CR” 방법에 비해 떨어진다. 마주침과 교차 조우상황에서 “DRL” 방법이 “VO” 방법보다 추가 운항 거리가 짧으므로 운항 효율이 더 높은 반면 추월 조우상황에서는 “VO” 방법이 “DRL” 방법보다 추가 운항 거리가 더 짧기 때문에 운항 효율이 더 높다. 추월 조우상황에서 “VO” 방법은 중첩영역이 발행하기 때문에 목표 경로와 타선에 가깝게 회피를 하게 됨으로써 운항 효율이 높아지게 된다.

추월 조우상황에서 “VO” 방법은 중첩영역이 발생하기 때문에 운항효율은 안전성이 동일하지 않은 조건에서 “DRL” 방법과 비교된다. 안전성을 동일하게 하고 비교하기 위해 “VO” 방법에서 회피 시작 시간을 TCPA 1,131초로 변경함으로써 두 폐색영역이 겹치지 않도록 하였다. 해당 방법은 “VO2”로 표기하며

비교 결과는 Fig. 6.33, Fig. 6.34와 같다.

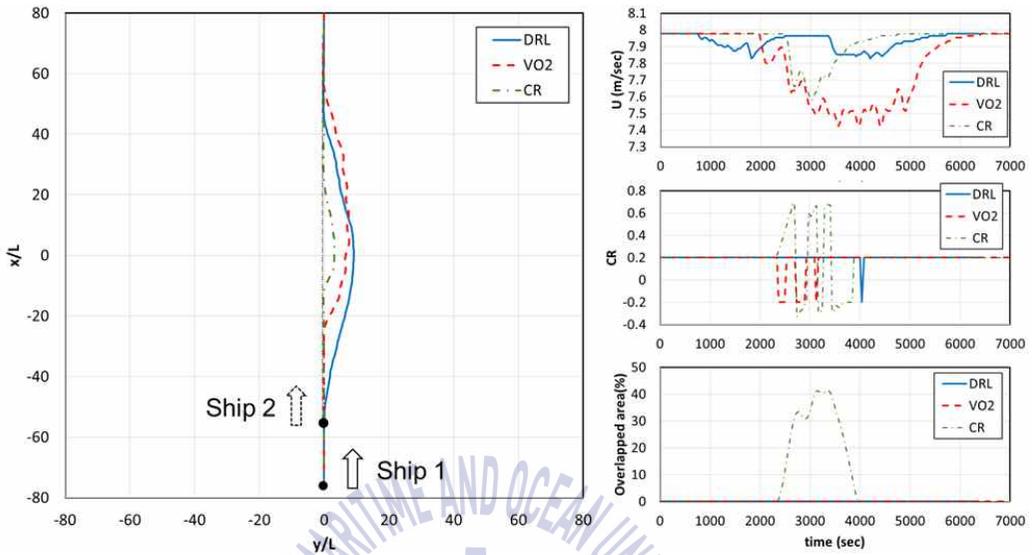


Fig. 6.33 Collision avoidance results to compare with VO2 and other methods in overtaking situation

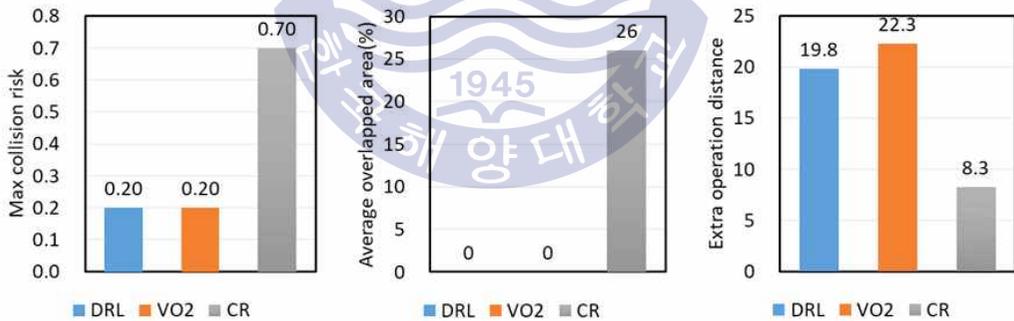


Fig. 6.34 Bar graph to compare with VO2 and other methods in overtaking situation

Fig. 6.34와 같이 “VO” 방법을 폐색영역이 중첩되지 않도록 안전성 조건을 동일하게 하였을 경우 마주침과 교차 조우상황과 마찬가지로 추월 조우상황에서도 “DRL”의 운항 효율이 “VO” 방법보다 높게 나타난다. “CR” 방법은 회피 시작 충돌위험도 값을 조절하여도 두 선박의 폐색영역이 겹치지 않도록 할 수 없기 때문에 안전성 조건을 다른 두 방법과 동일하게 두고 비교할 수 없었다.

안전성 조건이 동일할 경우 “VO” 방법은 회피 시작 시점과 회피 각도가 고정되어 있는 반면 “DRL” 방법은 학습을 통해 찾은 최적 행동 정책에 따라 회피 시작 시점과 회피 행동이 결정되기 때문에 운항 효율이 더 높게 나타난다.

### 안전성과 운항효율의 통합지표를 통한 비교

앞서 “DRL”과 “VO” 방법이 가장 좋은 안전성을 보이며 “CR” 방법은 운항 효율이 가장 높은 것으로 확인되었다. 이처럼 최대 충돌위험도와 폐색영역 평균 중첩 비율, 추가 운항 거리는 안전성과 운항효율을 각각 비교할 수 있지만 “DRL”, “VO”, “CR” 중 어떤 방법이 가장 좋은 방법인지 알기 어렵다. 세 가지 방법을 비교하기 위해 안전성과 운항효율을 함께 나타낼 수 있는 통합지표 ( $c_{total}$ )를 정의하고 이를 통해 세 가지 방법을 비교한다.

$$c_{total} = w_{total} \cdot l'_{risk} + (1 - w_{total}) \cdot l'_{extra} \quad (6.10)$$

통합지표( $c_{total}$ )는 식 (6.10)과 같이 안전성을 대표하는 지표( $l'_{risk}$ )와 운항효율을 대표하는 지표( $l'_{extra}$ )를 이용하여 산출한다. 두 선박의 조우상황에서 선장마다 자선이 안전하다 느끼는 기준이 다르기 때문에 이를 고려하기 위해 가중치 ( $w_{total}$ )를 두었다. 가중치는 0에서 1의 값을 가지며 값이 클수록 안전성에 대한 영향력이 큰 통합지표를 산출하게 되고 값이 작을수록 운항효율에 대한 영향력이 큰 통합지표를 산출한다.  $l'_{extra}$  (식 (6.9))는 운항효율을 나타내는 단일 지표이기 때문에 통합지표 산출 시 운항효율을 대표하는 지표로 활용한다. 하지만 안전성 평가를 위해 정의한 최대 충돌위험도 또는 폐색영역 평균 중첩 비율은  $l'_{extra}$ 와 단위가 달라 값의 크기를 동등하게 비교할 수 없기 때문에  $l'_{risk}$ 로 활용할 수 없다. 안전성과 운항효율의 영향을 모두 고려한 통합지표를 산출하기 위해 식 (6.11)과 같이  $l'_{risk}$ 를 새로 정의할 필요가 있다.

$$l'_{risk} = \left( \sum (\Delta d_{OS} + \Delta d_{TS}) \right) / L \quad (6.11)$$

$l'_{risk}$ 는 두 선박의 폐색영역이 중첩됨으로써 자선과 타선이 인지하는 위험 거리로 정의할 수 있으며  $d_{OS}$ 는 자선이 타선에게 인지하는 위험 거리,  $d_{TS}$ 는 타

선이 자선에게 인지하는 위험 거리이다. 6.1.1절에서 두 선박의 폐색영역이 겹치면 유지선도 회피하는 영역이기 때문에  $l'_{risk}$ 는 Fig. 6.35와 같이 중첩된 두 폐색영역 사이에서 정의할 수 있다. 자선이 타선에게 인지하는 위험 거리  $d_{OS}$ 를 구하기 위해 Fig. 6.35 왼쪽 그림의 원 표식(●)과 같이 우선 자선(OS)에서 가장 가까운 타선(TS) 폐색영역의 경계 지점을 찾는다. 해당 지점으로부터 타선까지 점선을 그었을 때 자선 폐색영역 경계까지 거리가  $d_{OS}$ 이며 Fig. 6.35 왼쪽 그림에서 화살표로 표시한 거리와 같다. 마찬가지로 방법으로 Fig. 6.35 오른쪽 그림과 같이 타선이 자선에게 인지하는 위험 거리  $d_{TS}$ 를 구한다.

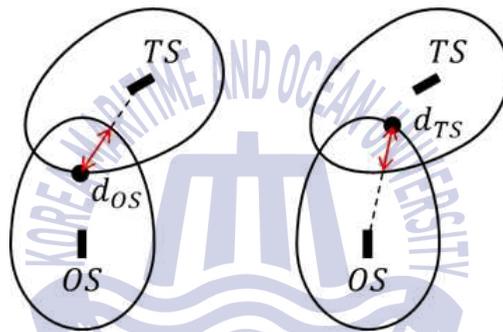


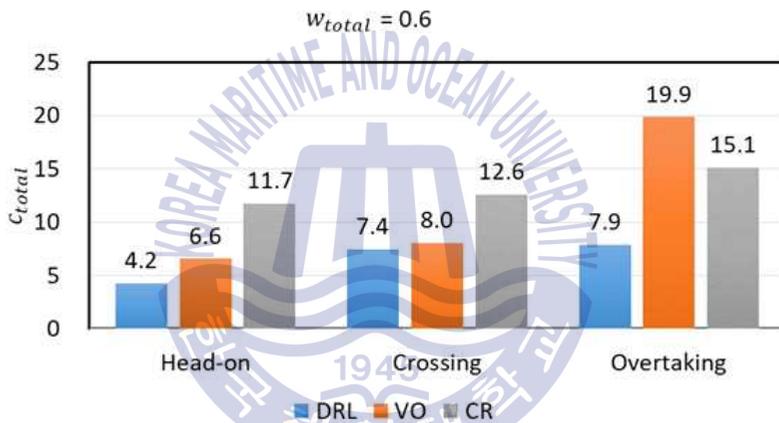
Fig. 6.35 Example of risk distance recognized by own ship and target ship

통합지표 산출을 위해 식 (6.10)의 가중치( $w_{total}$ )를 결정해야 한다. 가중치는 안전성과 운항효율의 영향력을 조절하는 값으로 0.5의 경우 영향력이 동일한 것을 의미한다. 본 연구에서는 타선과의 거리가 가까워지더라도 운항효율을 중요하게 생각하는 사람과 운항효율보다는 타선과의 거리를 유지하는 것이 중요하다 생각하는 사람을 가정하고 가중치 값을 0.6과 0.8로 선정하여 테스트한다. 충돌 회피에서 안전성을 유지하는 것이 운항효율을 높이는 것 보다 중요하기 때문에 운항효율을 중요시하는 사람일지라도 안전성에 대한 영향력을 운항효율보다 낮거나 같도록 하지 않을 것으로 판단하여 가중치 값을 0.6으로 선정하였다. 마주침, 교차, 추월 조우상황에서 산출된 위험 거리( $l'_{risk}$ )와 추가 운항 거리( $l'_{extra}$ )는 Table 6.9와 같고 가중치 0.6과 0.8 값으로 산출된 통합지표는 Fig.

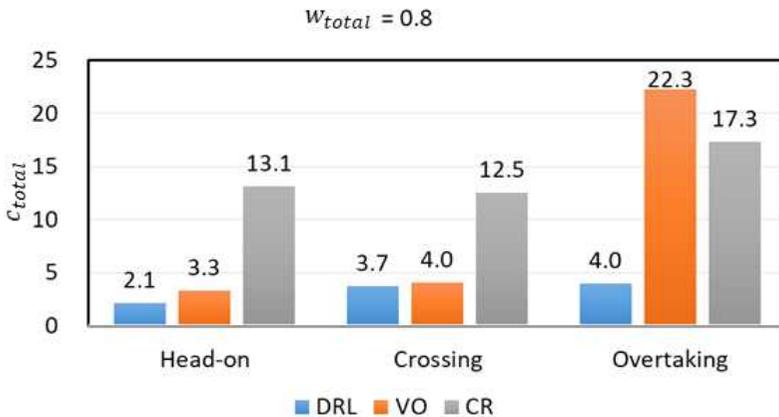
6.36, Fig. 6.37과 같다.

**Table 6.9** Quantitative measures of risk distance and extra operation distance

	DRL		VO		CR	
	$l'_{risk}$	$l'_{extra}$	$l'_{risk}$	$l'_{extra}$	$l'_{risk}$	$l'_{extra}$
head-on	0.0	10.6	0.0	16.4	14.5	7.5
crossing	0.0	18.6	0.0	20.1	12.5	12.7
overtaking	0.0	19.8	24.7	12.6	19.6	8.3



**Fig. 6.36** Bar graph for comparing total quantitative measure ( $w_{total}=0.6$ )



**Fig. 6.37** Bar graph for comparing total quantitative measure ( $w_{total}=0.8$ )

통합지표( $c_{total}$ ) 값은 안전성, 운항효율과 반비례 관계에 있기 때문에 값이 작을수록 안전성과 운항효율이 높으며 Fig. 6.36, Fig. 6.37에서 “DRL” 방법이 모든 조우상황에서 가장 작은 값을 나타내고 있기 때문에 가장 좋은 결과를 보임을 알 수 있다. “DRL” 방법은 폐색영역에 의한 안전거리를 확보하고 있기 때문에 위험 거리( $l'_{risk}$ )에 대한 영향이 없으며 위험 거리에 대한 영향이 없는 방법들 중 추가 운항 거리( $l'_{extra}$ )가 가장 짧기 때문에 통합지표 값이 가장 작게 나타난다. 마주침과 교차 조우상황에서는 중첩 폐색영역이 나타나지 않는 “VO” 방법의 통합지표가 “CR”보다 작은 값을 가진다. 이는 “CR” 방법이 위험 거리( $l'_{risk}$ )에 대한 영향이 크기 때문이며 이로 인해 가중치 값이 커질수록 통합지표 값도 커진 것을 확인할 수 있다. 추월 조우상황에서는 “VO”와 “CR” 방법 모두 중첩 폐색영역이 발생하며 “CR” 방법의 통합지표가 “VO”보다 작은 값을 가진다. “VO”의 평균 중첩 폐색영역 비율은 “CR”보다 작지만(Fig. 6.29) 평균 중첩 폐색영역 비율의 변화량이 많기 때문에 “VO”의 통합지표가 “CR”보다 높게 나타난다.

가중치는 값이 작을수록 위험 거리에 대한 영향이 줄어들기 때문에 운항효율이 가장 높은 “CR” 방법이 “DRL” 방법보다 통합지표 값이 작아지는 구간이 존재한다. 가중치 값이 0.36 이하일 경우 추월 조우상황에서 “CR” 방법이 “DRL” 방법보다 통합지표 값이 작아진다. 하지만 앞서 언급한 대로 충돌 회피에서 안전성을 유지하는 것이 운항효율을 높이는 것보다 중요하므로 운항효율을 중요시하는 사람일지라도 안전성에 대한 영향력을 운항효율보다 낮거나 같도록 하지 않을 것이기 때문에 가중치 값을 0.36 이하로 설정하기 어려우며 가중치 값이 0.5를 초과한다고 보았을 때 “DRL” 방법이 가장 좋은 결과를 나타낸다.

## 6.4.2 선박 대체 결과

VO와 심층강화학습을 이용한 충돌회피 방법을 다른 선박에 적용하였을 경우 충돌회피 행동의 변화를 확인하기 위해 크기와 종류가 다른 SR108 컨테이너선 선형을 이용하여 비교 테스트를 수행한다. SR108의 주요제원을 Table 6.10과 같이 KVLCC와 동일하게 축척하였으며, 유체력미계수는 Table 6.11과 같다.

SR108을 대상으로 Table 5.2, Table, 5.3, Table 5.5와 같은 마주침, 교차, 추월 조우상황에서 수행한 결과는 Fig. 6.38, Fig. 6.39, Fig. 6.40과 같으며 두 선박의 폐색영역이 중첩되는 구간은 나타나지 않았다.

**Table 6.10** Principal dimensions of SR108

Item	Value
Length	175 m
Breadth	25.4 m
Draft	9.5 m
Block Coefficient	0.5717

**Table 6.11** Hydrodynamic derivative values of SR108

Item	Value	Item	Value
$X'_{uu}$	-0.01465564	$X'_{\beta r}$	-0.08631303
$Y'_{\beta}$	0.22255084	$N'_{\beta}$	0.09870571
$Y'_r$	0.01394765	$N'_r$	-0.02823273
$Y'_{\beta\beta}$	0.66037595	$N'_{\beta\beta}$	0.02542121
$Y'_{rr}$	0.08251527	$N'_{rr}$	-0.0541023
$Y'_{\beta\beta r}$	0.0334646	$N'_{\beta\beta r}$	-0.5706726
$Y'_{\beta rr}$	0.98769	$N'_{\beta rr}$	0.0364034

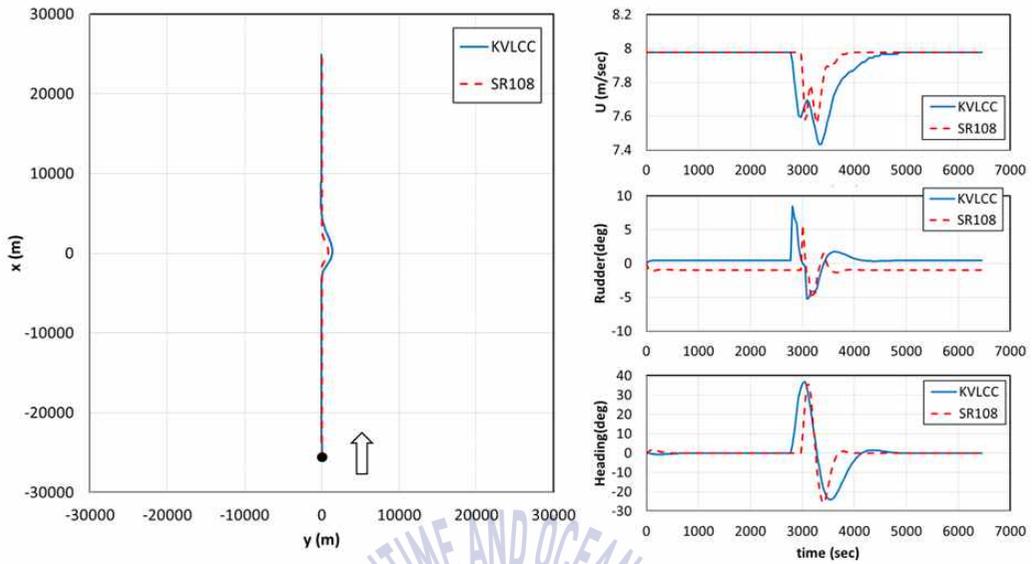


Fig. 6.38 Collision avoidance results to compare with KVLCC and SR108 in head-on situation

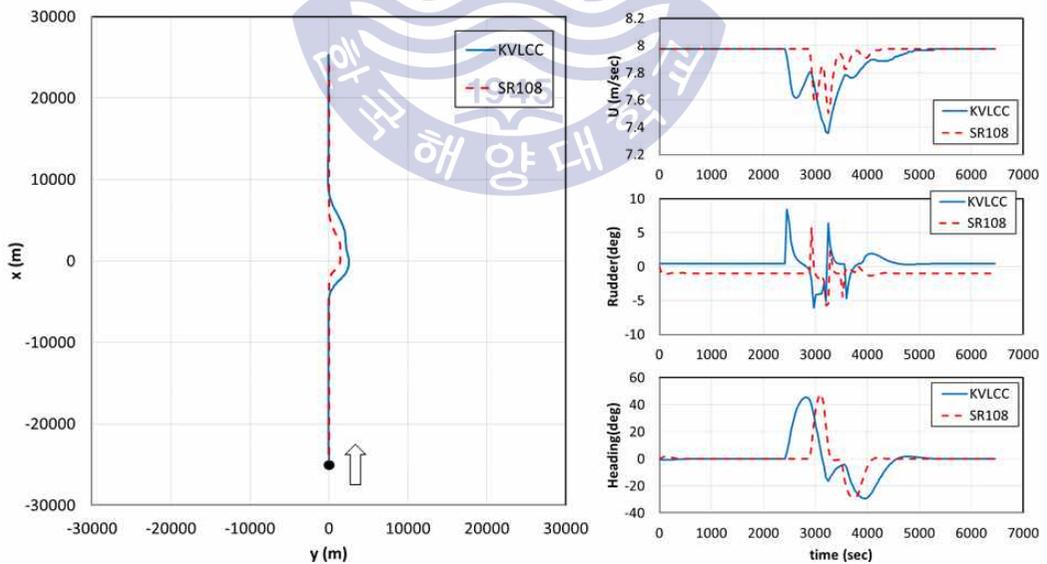


Fig. 6.39 Collision avoidance results to compare with KVLCC and SR108 in crossing situation

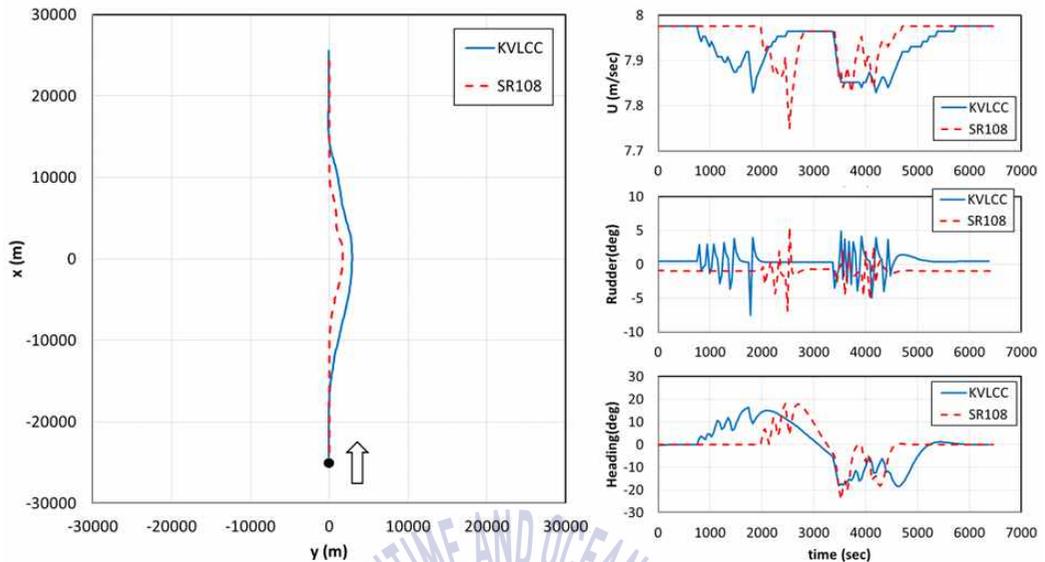


Fig. 6.40 Collision avoidance results to compare with KVLCC and SR108 in overtaking situation

SR108의 전장이 KVLCC보다 짧아 폐색영역 크기가 작기 때문에 회피 시작 시점이 늦고 회피 종료 시점은 빠르다. 하지만 SR108을 이용한 결과는 KVLCC를 활용할 때와 선택된 행동이 동일하기 때문에 유사한 운동 경향을 보인다. 심층강화학습을 통해 학습된 결과를 대상 선박이 변경되어도 활용할 수 있다면 학습에 걸리는 시간을 줄일 수 있다. 본 논문에서 심층강화학습을 통해 훈련된 최적 행동 정책은 대상 선박이 변경되어도 활용할 수 있음을 확인하였다. Fig. 6.38, Fig. 6.39, Fig. 6.40의 결과에서 KVLCC와 SR108이 직진 시에 타각 0도를 유지하지 않는 것은 프로펠러 회전 방향이 일정함에 따라 비대칭 횡력이 발생함으로써 선박이 조금씩 선회하기 때문에 이를 상쇄하기 위함이다.

## 제 7 장 결 론

본 논문에서는 VO와 심층강화학습을 이용한 VLCC급 유조선의 충돌 회피 방법을 제안하였다. 우선 기존의 안전성 중심의 충돌 회피 방법에서 벗어나 심층강화학습을 이용해 기본적인 안전성을 보장하면서 운항효율을 높이는 회피 방법을 제시하였다. 심층강화학습은 현재 또는 미래의 보상이 최대가 되는 최적 행동 정책을 학습하는 것으로 안전성과 운항효율을 보상에 고려해 줌으로써 경험적 방법에 따라 안전성과 운항효율이 반영된 회피 경로를 결정하는 것을 확인하였다. 선박의 충돌 회피 문제를 MDP를 통해 수학적으로 표현하였으며 최적 행동 정책을 찾기 위해 DQN를 활용하여 심층강화학습을 수행하였다. 또한 VLCC급 유조선에 VO를 적용할 때 조타명령에 의한 응답이 매우 느리다는 특성으로 인해 발생하는 문제점을 확인하였다. 문제점을 해결하기 위해 몇 가지 휴리스틱 조건을 이용하여 대안을 제시하였으며 VO를 이용하여 선박의 우현 회피가 가능하게 하였다. 이를 간단한 행동 방법을 이용하여 단일선박의 조우 상황에 대해 수치 시뮬레이션을 통해 테스트하였다.

마지막으로 VO와 심층강화학습을 함께 이용한 충돌 회피 방법을 제시하였다. 기존의 충돌 회피 방법과 다른 개선점은 다음과 같다. 첫째, 충돌 시작 시점 및 회피 선수각에 대한 임계값 없이 경험을 통해 상황에 따른 최적의 회피 시작 시점 및 회피 선수각을 찾는다. 둘째, 회피 경로는 보상에 COLREGs와 운항 안전성 및 운항효율을 고려함으로써 COLREGs를 준수하고 안전성 및 운항효율을 고려한 회피 경로를 설정한다.

본 연구의 한계는 첫째로 실선 적용을 하지 못한 것이다. 선박의 운동 수학적 모델을 이용하여 테스트하였지만 실선 또는 모형선에 적용하여 검증하는 작업이 진행되어야 할 것이다. 둘째는 피항선임에도 불구하고 피항하지 않는 선박에 대한 고려가 되어있지 않다. 두 선박의 폐색영역이 겹치게 되면 유지선도 피항을 위한 행동을 하는 구역임을 정의하였다. 하지만 본 연구는 폐색영역이 겹치지 않도록 보상정책을 마련하였으며 피항선임에도 불구하고 피항하지 않는 선박에 대한 훈련 시나리오가 마련되어 있지 않다. 이를 고려하기 위한 연구가

추후 수행되어야 할 것으로 사료된다. 셋째로 심층학습을 이용한 방법은 하이퍼파라미터 값에 민감하며 하이퍼파라미터 최적화에 걸리는 시간이 상당하다는 것이다. 이와 같은 한계를 해결할 수 있는 연구가 진행되어야 할 것으로 사료된다.



## 참 고 문 헌

- Bergstra, J. & Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13, pp.281-305.
- Cai, Y. & Hasegawa, K., 2013. Evaluating of Marine Traffic Simulation System through Imazu Problem. *Conference proceedings, the Japan Society of Naval Architects and Ocean Engineers*, 17, pp.191-194.
- Chae, H., Kang, C.M., Kim, B., Kim, J., Chung, C.C. & Choi, J.W., 2017. Autonomous Braking System via Deep Reinforcement Learning. *arXiv preprint*, arXiv:1702.02302.
- Coldwell, T. G., 1983. Marine traffic behaviour in restricted waters. *The Journal of Navigation*, 36, pp.431-444.
- Davis, P. V., Dove, M. J. & Stockel, C. T., 1980. A computer simulation of marine traffic using domains and arenas. *The Journal of Navigation*, 33, pp.215-222.
- Fiorini, P. & Shiller, Z., 1993. Motion planning in dynamic environments using the relative velocity paradigm. *Proceedings IEEE International Conference on Robotics and Automation*, Atlanta, USA, May 2-6, 1993, pp.560-565.
- Fiorini, P. & Shiller, Z., 1998. Motion Planning in Dynamic Environments Using Velocity Obstacles. *The International Journal of Robotics Research*, 17(7), pp.760-772.
- Fujii, Y. & Tanaka, K., 1971. Traffic capacity. *The Journal of Navigation*, 24, pp.543-552.
- Furukawa, Y., Kijima, K. & Ibaragi, H., 2004. Development of automatic course modification system using fuzzy inference. *International Federation of*

- Automatic Control Proceedings*, 37(10), pp.77-82.
- Goodwin, E. M., 1975. A statistical study of ship domains. *The Journal of Navigation*, 28, pp.329-341
- Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D. & Meger, D., 2017. Deep Reinforcement Learning that Matters. *arXiv preprint*, arXiv:1709.06560
- Kijima, K. & Furukawa, Y., 2002. Development of Collision Avoidance Algorithm Using Fuzzy Inference. *Proceedings of ISOPE Pacific/Asia Offshore Mechanics Symposium*, Daejeon, Korea, November 17-20, 2002, pp.123-130.
- Kijima, K. & Furukawa, Y., 2003. Automatic collision avoidance system using the concept of blocking area. *International Federation of Automatic Control Proceedings*, Girona, Spain, 36(21), pp.223-228.
- Kijima, K., Katsuno, T., Nakiri, Y. & Furukawa, Y., 1990. On the Manoeuvring Performance of a Ship with the Parameter of Loading Condition. *Journal of Society of Naval Architects of Japan*, 168, pp.141-148.
- Kijima, K. & Nakiri, Y., 2003. On the Practical Prediction Method for Ship Manoeuvring Characteristics. *Transaction of the West-Japan Society of Naval Architects*, 105, pp.21-31.
- Kose, K., Hirono, K., Sugano, K. & Sato, I., 1998. A New Collision-Avoidance-Supporting-System and Its Application to Coastal-Cargo-Ship "SHOYO MARU". *International Federation of Automatic Control Proceeding*, Fukuoka, Japan, 31, pp.263-268.
- Kuwata, Y., Wolf, M.T., Zarzhitsky, D., & Huntsberger, T.L., 2014. Safe Maritime Autonomous Navigation With COLREGS, Using Velocity Obstacles. *IEEE Journal of Oceanic Engineering*, 39(1), pp.110-119.
- Lee, H.J. & Rhee, K.P., 2001. Development of collision avoidance system by

- using expert system and search algorithm. *International Shipbuilding Progress*, 48, pp.197-212.
- Li, Y., 2017. Deep Reinforcement Learning: An Overview. *arXiv preprint*, arXiv:1701.07274.
- Mamdani, E.H. & Assilian, S., 1975. An experiment in linguistic synthesis with a fuzzy logic controller, *International Journal of Man-Machine Studies*, 7(1), pp.1-13.
- Metz, L., Ibarz, J., Jaitly, N. & Davidson, J., 2017. Discrete Sequential Prediction of Continuous Actions for Deep RL. *arXiv preprint*, arXiv:1705.05035.
- Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D. & Riedmiller, M., 2013. Playing Atari with Deep Reinforcement Learning. *Neural Information Processing Systems, Lake Tahoe, USA*, 9 December 2013.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S. & Hassabis, D., 2015. Human-level control through deep reinforcement learning. *Nature*, 518, pp.529-533.
- Mori, S., 1995. Note of Ship Form Design(24). *FUNE-NO-KAGAKU*, 48, pp.40-49. (in Japanese)
- Ota, D., Masuyama, T., Furukawa, Y. & Ibaragi, H., 2016. Development of Automatic Collision Avoidance System for Ships Using Reinforcement Learning. *Proceedings of 7th PAAMES and AMEC2016*, Hong Kong, 13-14 October 2016.
- Park, J.H. & Kim, J.W., 2017. Predictive Evaluation of Ship Collision Risk Using the Concept of Probability Flow. *IEEE Journal of Oceanic Engineering*,

42(4), pp.836-845.

Pietrzykowski, Z. & Uriasz, J., 2004. The ship domain in a deep-sea area. *Proceeding of the 3rd International Conference on Computer and IT Applications in the Maritime Industries*, Sigüenza, Spain.

Rolls-Royce, 2017. *Rolls-Royce joins forces with Google Cloud to help make autonomous ships a reality* [Online] Available at: <https://www.rolls-royce.com/media/press-releases/2017/03-10-2017-rr-joins-forces-with-google-cloud-to-help-make-autonomous-ships-a-reality.aspx> [Accessed 31 October 2018].

Silver, D., Huang, A., Maddison, C., Guez, A., Sifre, L., Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Graepel, T., Lillicrap T., Leach, M., Kavukcuoglu, K. & Hassabis, D., 2016. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), pp.484-489.

Smierzchalski, R., 2001. On-line trajectory planning in collision situation at sea by evolutionary computation-experiments. *Proceeding of IFAC Conference on Computer Applications in Marine Systems*, Glasgow, U.K.

Tam, C. & Bucknall, R., 2010. Collision risk assessment for ships. *Journal of Marine Science and Technology*, 15(3), pp. 257-270

Wang, N., Meng, X., Xu, Q. & Wang, Z., 2009. A unified analytical framework for ship domains. *The Journal of Navigation*, 62, pp.643-655.

Yara International, 2018, *The first ever zero emission, autonomous ship* [Online] Available at: <https://www.yara.com/knowledge-grows/game-changer-for-the-environment/> [Accessed 31 October 2018].

Zhu, X., Xu, H. & Lin, J., 2001. Domain and its model based on neural networks. *The Journal of Navigation*, 54, pp.97-103.

- 김동진, 곽수용, 2011. 국내 해양선박사고의 인적 오류의 요인 평가. *대한인간공학회지*, 30(1), pp.87-98.
- 김선영, 손남선, 김성환, 2014. 다목적 지능형 무인선 국산화 개발. *대한조선학회지*, 51(2), pp.9-12.
- 박정홍, 최진우, 최현택, 2018. 경로 예측의 불확실성을 고려한 무인수상선의 충돌 위험 평가 및 회피 경로 계획. *제어로봇시스템학회 논문지*, 24(7), pp.608-616
- 박한선, 2018, 한국의 자율운항선박(MASS) 대응정책 방향. *Telecommunications Technology Association Journal*, 178, pp.60-65
- 손남선, 요시타카 후루카와, 김선영, 가쓰로 기지마, 2009. 가변공간 탐색법을 이용한 다중선박의 충돌회피 알고리즘에 관한 연구. *한국해양환경공학회지*, 12(1), pp.15-22.
- 우주현, 2018. *심층강화학습을 이용한 무인수상선의 충돌회피*. 박사학위논문. 서울: 서울대학교 조선해양공학과.
- 윤점동, 2015. *국제해상충돌예방규칙 및 관련된 국내법규 해설*. 2015년 개정판. 부산: 세종출판사.
- 이용원, 양혁렬, 김건우, 이영무, 이의령, 2017. *파이썬과 케라스로 배우는 강화학습*. 파주: 위키북스.
- 이윤석, 2018, 자율운항선박(MASS) 개발 동향 분석. *Telecommunications Technology Association Journal*, 178, pp.20-26
- 중앙해양안전심판원, 2017, *사고종류별 해양사고 원인현황(재결분, 2013~2017 총괄)* [Online] Available at: <https://www.kmst.go.kr/kmst/statistics/annualReport/selectAnnualReportList.do> [Accessed 31 October 2018].