Autonomous Collision Avoidance Control Using Deep Reinforcement Learning for Maritime Autonomous Surface Ships

P Lee, G Theotokatos, E Boulougouris

University of Strathclyde, Glasgow, UK

E-mail: p.lee@strath.ac.uk

Abstract. The maritime industry has been progressing towards autonomous shipping with the main barrier and scepticism being on the safety assurance of the next-generation autonomous ships. This study aims to enhance the safety of the autonomous ships by developing an intelligent agent that makes evasive decisions considering the ship domain as a safety zone. The proposed approach is demonstrated by considering the case study of a short sea shipping cargo ship. An intelligent reinforcement learning agent is trained to manoeuvre the investigated ship in restricted sea area. The results of this study verify the agent's ability to make safe evasive decisions and control the autonomous collision avoidance for autonomous ships in known and unknown environments.

1. Introduction

The quest for more sustainable and efficient operations of the shipping industry has resulted in a number of initiatives for developing and building autonomous ships, also known as Maritime Autonomous Surface Ships (MASS). However, the safety and functinality of MASS systems and its subsystems need to be verified prior to their full-scale implementation. According to [1], the realization of MASS lays upon the successful development of the Autonomous Navigation System and the Autonomous Machinery System, also known as the "Artificial Captain" and "Artificial Chief Engineer" respectively. The Autonomous Collision Avoidance System (ACAS) is one of the core subsystems of the Artificial Captain that is responsible for making and controlling machine-based collision avoidance decisions and actions without human intervention, which is related to ships with degrees of autonomy four according to IMO [2].

The pertinent literature proposed various control methods for MASS and currently, more research has been focused on the machine learning based control methods, such as Deep Reinforcement Learning (DRL) algorithms. [3] proposed two controllers using Deep Deterministic Policy Gradient (DDPG) algorithm to steer a ship through a narrow gate. [4] developed a Deep Q-Network (DQN) algorithm for the collision avoidance control among static obstacles. [5] investigated the steering control problem using DDPG for straight-path following under the influence of sea current. The curved-path following control problem was studied in [6].

This study aims to propose an Autonomous Collision Avoidance (ACA) control method for MASS using DDPG and ship domain. Range laser sensors are used to calculate the distance of the ship from the obstacles and a reward function based on ship domain is developed. An intelligent agent based on

DDPG algorithm is developed that controls the rudder angle and revolution of the propulsion to navigate in a restricted area without violating the ship domain.

The remainder of this study is structured as follows. Section 2 describes the developed methodology and its rationale. Section 3 delineates the case study details. Section 4 presents and discusses the derived results. Lastly, Section 5 summarises the main findings and the conclusions of this study.

2. Methodology

The followed methodology to accomplish the aim of this study consists of six phases as presented in Figure 1. In the first phase, the case study and the MASS reference system is defined, from where the critical and main components are digitally modelled in the second phase.

In Phase 3, an intelligent ACA agent is developed using Deep Deterministic Policy Gradient (DDPG) algorithm. DDPG is an off-policy, model-free, and online reinforcement learning algorithm suitable continuous control action and observation space. It learns to make the optimal control action that maximize the expected cumulative long-term reward.

During Phase 4, the agent is trained in a navigating sea environment with the aim of converging to the optimal solution and generalizing its knowledge. Finally, the developed controller is simulated, and its effectiveness is verified in known and unknown environments.

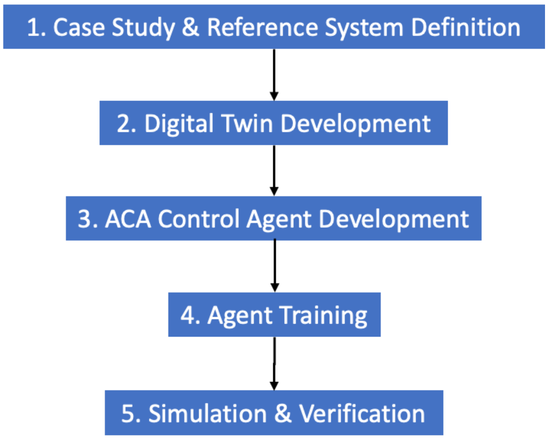

Figure 1 Methodology flowchart.

3. Case Study

3.1. The collision avoidance scenario

The case study includes the navigation of an autonomous ship through a restricted sea area heading towards North and the exit through a gateway without colliding with other static obstacles as seen in Figure 2. Specifically, the MASS is required to navigate safely without violating the safety zone by intersecting with other obstacles.

The investigated vessel is considered to be a next-generation autonomous container SSS vessel, servicing cargo transport between Europe's main ports at the North Sea. The length of the vessel is 175 metres, single propeller, with maximum rudder angle of 40 degrees and maximum propeller revolution of 160 rpm. The ship is considered unmanned equipped with autonomous decision-making systems, such as the ACAS, that satisfy the degree of autonomy four according to IMO [2]. The main components of the ACAS are the range sensors that calculate the distance of the ship from the obstacles and the intelligent agent that makes the evasive decision and action commands, which are the rudder angle command and the propeller revolution command.

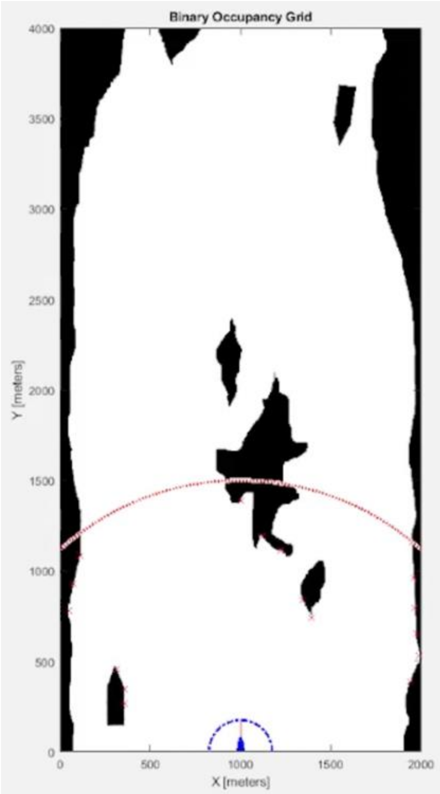

Figure 2 The collision avoidance scenario for the investigated vessel.

3.2. Digital twin development

A four degrees of freedom Manoeuvring Model Group (MMG) model was used to model the manoeuvrability of the ship. The model used in the training was taken from the Marine Systems Simulator (MSS) toolbox [7], specifically considering the 4-DOF container vessel that has been validated and tested in other studies [6, 8]. Laser sensors were used as range sensors of up to 32 points and 1.5 kilometres detection range. Safety zone was set to be 175 metres, which is a circular area around the ship. The digital twin and all the training process was conducted on the MATLAB 2021b version.

3.3. ACA-DM agent development

The DDPG algorithm consists of two Deep Neural Networks (DNN), the Actor network and the Critic network. The Actor network approximates an action from the current state, while the Critic network updates the parameters of the Actor network by generating a Q-value, which is an estimate of the discounted long-term reward at the start of each episode, given the initial observation of the environment [4].

The Actor network consists of an input layer from the environment observations, an output layer that produces the action commands, and two hidden layers in between with 400 units each. The Critic network consists of an input layer from the environment observations and action commands, an output layer that generates the Q-value, and two hidden layers in between with 300 units each. The input consists of the ship states, including the real-time x-position, y-position, heading angle, and the distance measures from the laser sensors. The output consists of the action commands regarding the rudder angle and the propeller's revolution.

3.4. Agent training

The agent was trained to navigate in a 2 x 4 kilometres map area. The agent was trained for up to 1800 episodes until the episode rewards showed good convergence and proximity to the expected Q0 value as seen in Figure 3.

The defined rewards were a reward for maintaining the original heading direction of 90 degrees, a reward for keeping the vessel away from the safety zone, and a reward for the approaching distance closer to the exit.

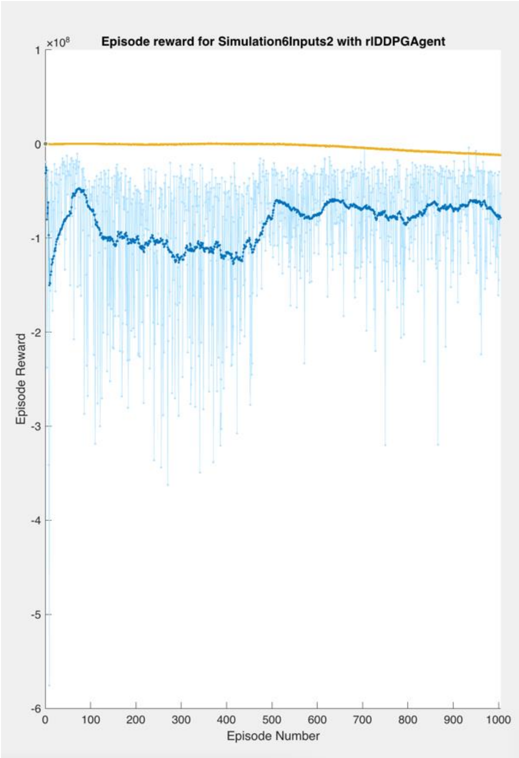

Figure 3 Training process of the agent (light blue: episode reward, dark blue: average reward, orange: episode Q0).

4. Results & Discussion

After the completion of the training process the agent was simulated in the pretrained map environment. Results from the simulations verified the agent's ability to make evasive decisions and control the ship without violating the safety zone as shown in Figure 4. Also, it was verified that the integration of the ship domain led to the evasive decision to exit from the broader East narrow gate although the West narrow gate was the closest gateway, which is the desired decision from the safety perspective as shown in Figure 4.

To verify the generalization of the agent's learning and the agent's capability to control in unknown and more challenging environments, the agent was tested in new maps that the agent has not seen before. The maps were amended by either switching the order of the gateways or minimizing the width difference between the two gateway solutions as shown in Figure 5. Simulation results verified the agent's generalized learning and the capabilities as an ACA controller.

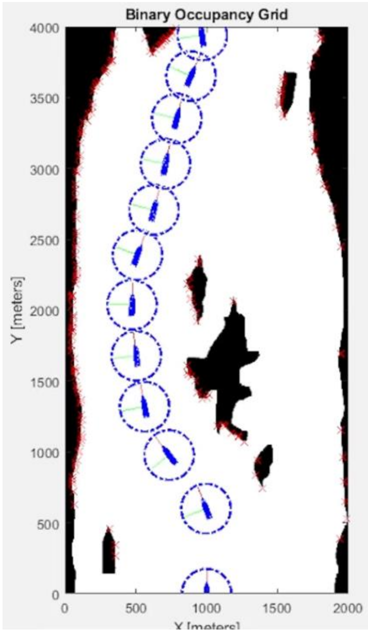

Figure 4 Simulation results in the trained environment.

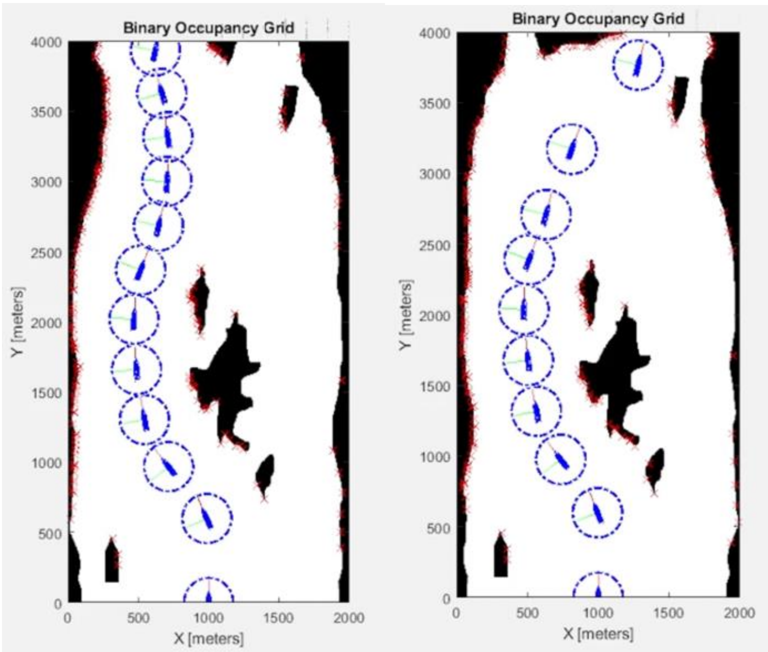

Figure 5 Simulation results in unknown environments.

5. Conclusions

In this study, a machine learning controller based on reinforcement learning was developed that controls the collision avoidance of a MASS while considering the ship domain. A collision avoidance case study and a reference ship model were defined. Main components of the reference system were modelled including the ship characteristics and manoeuvrability, the range sensors, and the intelligent agent. For the intelligent agent, a DDPG based algorithm was developed that makes evasive decisions and controls the action commands by considering the readings from the range sensors and the ship domain.

The main findings of this study are the following:

- The integration of ship domain leads to more safe evasive decisions.

- Intelligent agents, such as reinforcement learning based algorithms, have great generalisation capabilities suitable for real-time and unknown environments.

However, it should be mentioned that despite its strengths, some drawbacks of this methods include the time demanding regarding the reward function adjustment, neural network hyper-parameters tuning, and training process. Some limitations of this study include the training using a specific MMG model that is suitable for a specific ship. However, a well-trained agent can be used to learn a new MMG model as a “black-box” and converge with much less training episodes as mentioned in [6].

Acknowledgements
The authors greatly acknowledge the funding from DNV AS and RCCL for the Maritime Safety Research Centre (MSRC) establishment and operation. The opinions expressed herein are those of the authors and should not be construed to reflect the views of DNV AS and RCCL. Part of this study was carried out in the framework of the AUTOSHIP project, which is funded by the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 815012.

References
[1] Ø. J. Rødseth *et al.*, "AUTOSHIP deliverable D3.1: Autonomous ship design standards," Revision 1.1, 2021.
[2] IMO. "Autonomous shipping." https://www.imo.org/en/MediaCentre/HotTopics/Pages/Autonomous-shipping.aspx (accessed.
[3] A. Layek, N. A. Vien, and T. Chung, "Deep reinforcement learning algorithms for steering an underactuated ship," in *2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 2017: IEEE, pp. 602-607.
[4] Y. Cheng and W. Zhang, "Concise deep reinforcement learning obstacle avoidance for underactuated unmanned marine vessels," *Neurocomputing,* vol. 272, pp. 63-73, 2018.
[5] A. B. Martinsen and A. M. Lekkas, "Straight-path following for underactuated marine vessels using deep reinforcement learning," *IFAC-PapersOnLine,* vol. 51, no. 29, pp. 329-334, 2018.
[6] A. B. Martinsen and A. M. Lekkas, "Curved path following with deep reinforcement learning: Results from three vessel models," in *OCEANS 2018 MTS/IEEE Charleston*, 2018: IEEE, pp. 1-8.
[7] T. Fossen and T. Perez, "Marine systems simulator (mss)," *URL http://www. marinecontrol. org,* 2004.
[8] K.-H. Son and K. Nomoto, "5. on the coupled motion of steering and rolling of a high-speed container ship," *Naval Architecture and Ocean Engineering,* vol. 20, pp. 73-83, 1982.